

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA DE SISTEMAS

DESARROLLO DE UN SISTEMA DE INDICADORES PARA MONITOREAR EL PROCESO DE APRENDIZAJE CASO DE ESTUDIO TALLER DE ROBÓTICA DE LA UNIDAD EDUCATIVA VERBO

**TESIS PREVIA A LA OBTENCIÓN DEL GRADO DE MÁSTER (Msc) EN
SISTEMAS DE INFORMACIÓN MENCIÓN INTELIGENCIA DE NEGOCIOS
Y ANALÍTICA DE DATOS MASIVOS**

NORMA ELIZABETH ARCOS PILALUISA

norma.arcos@epn.edu.ec

DIRECTOR: JULIÁN ANDRES GALINDO LOSADA

julian.galindo@epn.edu.ec

Quito, enero 2021

APROBACIÓN DEL DIRECTOR

Como director del trabajo de titulación DESARROLLO DE UN SISTEMA DE INDICADORES PARA MONITOREAR EL PROCESO DE APRENDIZAJE CASO DE ESTUDIO TALLER DE ROBÓTICA DE LA UNIDAD EDUCATIVA VERBO desarrollado por Norma Elizabeth Arcos Pilaluisa, estudiante de la Maestría en Sistemas de Información, Mención en INTELIGENCIA DE NEGOCIOS Y ANALÍTICA DE DATOS MASIVOS, habiendo supervisado la realización de este trabajo y realizado las correcciones correspondientes, doy por aprobada la redacción final del documento escrito para que prosiga con los trámites correspondientes a la sustentación de la Defensa oral.



Dr. Julián Andrés Galindo Losada

DIRECTOR

DECLARACIÓN DE AUTORÍA

Yo, Norma Elizabeth Arcos Pilaluisa declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.



Norma Elizabeth Arcos Pilaluisa

DEDICATORIA

Dedico este trabajo a Dios y a mi familia a quienes les debo la oportunidad de concluir una etapa más de crecimiento en mi vida.

Normy

AGRADECIMIENTO

A Dios por permitirme despertar cada día y seguir adelante en la vida.

A mis papitos Anita y Luis, quienes son un ejemplo de vida, de lucha, de trabajo y de amor, de quienes me siento muy orgullosa ya que con su amor, esfuerzo, dedicación, comprensión y apoyo incondicional me ha dado las lecciones más valiosas en mi vida.

A mi ñaño Vinny gracias por tu compañía, por tus consejos y tu apoyo.

A mi tío Vicente que, gracias a su apoyo, consejos y enseñanzas de vida que permitieron ser una mejor persona cada día.

A mis tíos y tías gracias por sus consejos y apoyo incondicional.

A mis sobrinitas Emili y Dani que con sus travesuras inspiraban a su tía para darle más fuerza para alcanzar este logro.

Al Dr. Julián Galindo por su ayuda y aporte en este proyecto.

Al personal de la Unidad Educativa Cristiano Verbo y en especial al director Mario Cando quienes me dieron las facilidades necesarias para el desarrollo de este proyecto.

Normy

ÍNDICE DE CONTENIDO

1.	INTRODUCCIÓN	1
1.1.	Descripción del problema.....	1
1.1.1	Historia de la empresa	1
1.2.	Objetivo general	1
1.3.	Objetivos específicos	2
1.4.	Alcance	2
1.5.	Marco teórico.....	2
1.5.1	Descripción de la situación actual.....	2
1.5.2	Justificación de la metodología.....	4
1.5.3	Selección de metodología.....	6
1.5.4	Selección de herramientas.....	7
1.5.5	Bases Teóricas.....	15
2.	METODOLOGÍA.....	29
2.1	Análisis de requerimientos	29
2.1.1	Diagnóstico de la Información	29
2.1.2	Análisis de la experiencia de usuario	35
2.1.3	Diseñar el sistema de indicadores.....	38
2.1.4	Modelo conceptual.....	40
2.2	Análisis de OLTP	41
2.2.1	Conformación de indicadores	41
2.2.2	Establecer correspondencias	43
2.2.3	Nivel de granularidad.....	45
2.2.4	Modelo conceptual ampliado	50
2.3	Modelo lógico del DW	51
2.3.1	Tablas de Dimensiones	51
2.3.2	Tablas de Hechos	51
2.3.3	Diseño del DW	52
2.3.4	Uniones.....	54
2.4	Integración de datos.....	55
2.4.1	Implementar la integración de datos.....	60

2.4.2	Extracción	61
2.4.3	Transformación.....	69
2.4.4	Limpieza de datos	73
2.4.5	Simulación del Modelo de Predicción	92
2.5	Visualización.....	95
3.	RESULTADOS Y DISCUSIÓN	112
3.1	Carga inicial	112
3.2	Respuestas a las preguntas planteadas	112
3.3	Evaluar el uso del sistema de indicadores	119
4.	CONCLUSIONES Y RECOMENDACIONES.....	121
4.1	Conclusiones	121
4.2	Recomendaciones	123
	REFERENCIAS BIBLIOGRÁFICAS.....	125
	SIGLAS.....	129
	DEFINICIONES.....	130
	ANEXOS	131

LISTA DE FIGURAS

Figura 1.1 Pasos de Metodología HEFESTO con Visualización [1]	7
Figura 1.2 Preferencias Editor Workbench	9
Figura 1.3 SQL Execution - Editor Workbench	9
Figura 1.4 Workbench: Restauración de base de datos	10
Figura 1.5 Panel de Azure -Cuenta Estudiante.....	10
Figura 1.6 Panel de Servicio Azure.....	11
Figura 1.7 Comparación de Herramientas para la creación del DW [21]	12
Figura 1.8 Plataforma de Análisis Empresarial Pentaho [22]	13
Figura 1.9 Pentaho y Visualizer [24]	15
Figura 1.10 DW, características [1].....	17
Figura 1.11 Modelo circuplejo de afecto [27]	18
Figura 1.12 Algoritmos Estadísticos para Modelos de Data Science [28].....	19
Figura 1.13 Proceso de Ciencia de Datos [29].....	20
Figura 1.14 Top de métodos y algoritmos usados en Ciencia de Datos y Aprendizaje Automático en el 2018/9 vs 2017 [31]	21
Figura 1.15 Curva de Comparación ROC.....	24
Figura 1.16 Proceso de Consulta y Análisis [1]	27
Figura 2.1 Correo de acuerdo de las preguntas planteadas.....	31
Figura 2.2 Modelo Conceptual.....	40
Figura 2.3 Entendimiento de los datos.....	41
Figura 2.4 TalleresRE, Diagrama de Entidad Relación.....	43
Figura 2.5 Talleres RE, correspondencia	44
Figura 2.6 Talleres RE, Modelo Conceptual Ampliado.....	50
Figura 2.7 Taller RE, tabla de dimensión "PERSONA"	52
Figura 2.8 Taller RE, tabla de dimensión "TALLER"	52
Figura 2.9 Taller RE, tabla de dimensión "EXPERIENCIA"	53
Figura 2.10 Taller RE, tabla de dimensión "FASE"	53
Figura 2.11 Taller RE, tabla de dimensión "TIEMPO"	53
Figura 2.12 Taller RE, tabla de dimensión "IMAGEN"	54
Figura 2.13 Taller RE, tabla de dimensión "EMOCION"	54
Figura 2.14 Taller RE, Uniones.....	55
Figura 2.15 Transformación para subida de archivos JSON.....	61
Figura 2.16 Spoon: Resultado de la carga de archivos JSON	62
Figura 2.17 Spoon: Parametrización de campos de un archivo JSON	62
Figura 2.18 Spoon: Extracción de Excel a base de datos	63
Figura 2.19 RM: Extracción de CSV a Base de datos.....	63
Figura 2.20 OLTP: Tabla TALLER	63
Figura 2.21 OLTP-Tabla PERSONA	64
Figura 2.22 OLTP-Tabla REGISTROT	64
Figura 2.23 Importación de archivo CSV a Excel	64
Figura 2.24 Archivo con caracteres especiales	65
Figura 2.25 Uso de codificación UTF8 en Excel	65
Figura 2.26 Spoon: Uso de Replace in String	65
Figura 2.27 Tiempo aleatorio entre en un rango de horas en Excel	66
Figura 2.28 RM-Turbo Prep: Generar columna	66
Figura 2.29 RM-Turbo Prep: Auto Modelado	67
Figura 2.30 RM-Turbo Prep: Auto Modelado, preparación variable objetivo	67
Figura 2.31 RM-Turbo Prep: Rangos Variable Objetivo.....	68
Figura 2.32 RM-Turbo Prep: Vista General de los Modelos Predictivos	68
Figura 2.33 Curvas ROC	68
Figura 2.34 Spoon:Opción concatenar campos.....	69
Figura 2.35 Spoon: Manejo de fechas	69
Figura 2.36 Excel: Manejo de cantidades como cadenas.....	70
Figura 2.37 RM: Cambio de tipo de dato	70
Figura 2.38 Spoon: Cambio de tipo de dato.....	71
Figura 2.39 Spoon: Reemplazo de grupo de string a numérico	71
Figura 2.40 Spoon: Renombrar atributos.....	71

Figura 2.41	Spoon: Actualización de la edad con el dato de la tabla dim_persona	72
Figura 2.42	Uso de la moda para la edad del estudiante	72
Figura 2.43	RM-Turbo Prep: Función Pivot.....	73
Figura 2.44	Spoon: Retirar datos que no son requeridos	73
Figura 2.45	RM-TurboPrep:Remover columnas	74
Figura 2.46	Spoon: Filas únicas	74
Figura 2.47	Excel: Quitar duplicados	75
Figura 2.48	Spoon: Database join para selección del id de la fase conociendo el nombre de la misma.....	75
Figura 2.49	RM: Descartar faceld en "n.a."	76
Figura 2.50	RM - Turbo Prep: Descartar faceld en "n.a."	76
Figura 2.51	Spoon: Descartar faceld en "n.a."	76
Figura 2.52	Cálculo de moda en edad de la persona.....	77
Figura 2.53	Spoon: Actualización del dato edad.....	77
Figura 2.54	Spoon: Reemplazo en cadena	78
Figura 2.55	Spoon: Filtrado de filas para descartar adultos	78
Figura 2.56	RM-Turbo Prep: Descarte de adultos	79
Figura 2.57	RM-TurboPrep: Reemplazo de valores nulos por 0	79
Figura 2.58	RM-Turbo Prep: Reemplazo de caracteres especiales por 0	80
Figura 2.59	RM-Turbo Prep: Reemplazo de valores nulos por ND- No detectado	80
Figura 2.60	RM-TurboPrep: Reemplazo de un mismo tipo de lentes.....	80
Figura 2.61	RM-Turbo Prep: Caracteres Perdidos	81
Figura 2.62	RM-TurboPrep: Replace Missing Values	81
Figura 2.63	RM-Turbo Prep: Resultado del uso de Replace Missing Values	81
Figura 2.64	Spoon: Transformación para determinar la experiencia de usuario	82
Figura 2.65	Spoon: Actualización de la predicción	82
Figura 2.66	Taller RE, Carga Inicial	83
Figura 2.67	Taller RE, Carga de Dimensión TIEMPO	84
Figura 2.68	Taller RE, Carga de Dimensión PERSONA.....	84
Figura 2.69	Spoon: Obtener datos del estudiante del OLTP	85
Figura 2.70	Taller RE, Carga de Dimensión TALLER	85
Figura 2.71	Spoon: Obtener datos del taller del OLTP	85
Figura 2.72	Taller RE, Carga de Dimensión FASE.....	86
Figura 2.73	Spoon: Obtener datos de las fases del OLTP	86
Figura 2.74	Taller RE, Carga de Dimensión IMAGEN	86
Figura 2.75	Spoon: Obtener datos para la imagen del OLTP	87
Figura 2.76	Taller RE, Carga de Dimensión Emociones.....	87
Figura 2.77	Taller RE, Carga de Dimensión Experiencia	88
Figura 2.78	Spoon: Obtener datos para la experiencia de usuario del OLTP	88
Figura 2.79	Taller RE, Trabajo para la carga de Tabla de Hechos h_registroTaller.....	88
Figura 2.80	Taller Re, Estructura de transformaciones para la carga a la Tabla de Hechos H_REGISTROTALLER.....	89
Figura 2.81	Taller RE, Trabajo para la carga de Tabla de Hechos h_prediccion.....	90
Figura 2.82	Taller RE, Transformación para la carga pivot de la experiencia y predicción del usuario	91
Figura 2.83	Spoon: Obtener idPersona e idTaller para h_prediccion.....	91
Figura 2.84	Taller RE, Transformación para la experiencia y predicción del usuario.....	92
Figura 2.85	Spoon: Obtener datos para la experiencia y predicción de la tabla h_prediccion	92
Figura 2.86	Simulación RM-Turbo Prep- Optimización 1.....	93
Figura 2.87	Correlación Positiva y Negativa para Emoción Neutral	93
Figura 2.88	RM-TurboPrep- Optimización 2.....	94
Figura 2.89	Saiku: Cubo Registro Taller.....	95
Figura 2.90	Saiku: Relación Hechos y Dimensiones.....	96
Figura 2.91	Saiku: Indicador Total Estudiantes	96
Figura 2.92	Saiku: Indicador Total Emociones.....	97
Figura 2.93	Saiku: Indicador Promedio Emociones.....	97
Figura 2.94	Saiku: Cubo Predicción.....	98
Figura 2.95	Saiku: Relaciones para Cubo Predicción	98
Figura 2.96	Saiku: Indicador Total Experiencia Positiva	99

Figura 2.97 Saiku: Indicador Total Experiencia Negativa	99
Figura 2.98 Saiku: Indicador Total Experiencia Neutral	100
Figura 2.99 Saiku: Dimensión Tiempo	101
Figura 2.100 Saiku: Indicador para "Emocion"	101
Figura 2.101 Saiku: Indicador para Número de Estudiantes	102
Figura 2.102 Saiku: Medidas	103
Figura 2.103 Saiku Reportes: Emoción por género.....	103
Figura 2.104 Saiku Reportes: Emoción por fecha	104
Figura 2.105 Saiku Reportes: Emociones por taller	104
Figura 2.106 Saiku Reportes: Emociones por fecha y taller.....	104
Figura 2.107 Saiku Reportes: Emoción por estudiante y fecha	105
Figura 2.108 Saiku Reportes: Emoción por taller, género, grupo, estudiante y fecha.....	105
Figura 2.109 Saiku Reportes: Emoción por taller, género, grupo, estudiante, edad, fase y fecha..	106
Figura 2.110 Saiku Reportes: Emoción por género y año	106
Figura 2.111 Saiku Reportes: Emociones por fecha y taller.....	106
Figura 2.112 Saiku Reportes: Emociones por fases de un taller	107
Figura 2.113 Saiku Reportes: Emociones por fase y estudiante	107
Figura 2.114 Saiku Reportes: Emoción por estudiante.....	107
Figura 2.115 Saiku Reportes: Emoción por fase	108
Figura 2.116 Diseño Dashboard.....	109
Figura 2.117 Visualizador Datafor: Diseño Dashboard.....	109
Figura 2.118 Visualizador Datafor: Diseño dashboard aplicando principios de diseño	110
Figura 2.119 Visualizador Datafor: Versión final del dashboard de emociones	110
Figura 2.120 Visualizador Datafor: Dashboard de predicción de experiencia de usuario	111
Figura 3.1 Indicadores del dashboard.....	112
Figura 3.2 Total y promedio de Emociones por fotografía.....	113
Figura 3.3 Tabla de Emociones por Estudiante, Edad y Género	113
Figura 3.4 Filtrado de Emoción positivas para un estudiante.....	114
Figura 3.5 Emociones positivas y negativas por estudiante	114
Figura 3.6 Emoción felicidad por un estudiante	114
Figura 3.7 Total y promedio de emociones por taller	115
Figura 3.8 Número de estudiantes por taller y género	115
Figura 3.9 Total y promedio de emociones por género.....	116
Figura 3.10 Total y promedio de emociones por género y fase	116
Figura 3.11 Experiencia Vs Prediccion por Taller.....	117
Figura 3.12 Experiencia vs Predicción por estudiante	117
Figura 3.13 Experiencia vs Predicción por Género	118
Figura 3.14 Experiencia vs Predicción por Fase.....	118
Figura 3.15 Experiencia vs Predicción por Taller y Fase.....	119
Figura 3.16 Puntuación de aceptabilidad [35].....	120
Figura 0.1 Programa utilizado por Azure para reconocimiento facial	147
Figura 0.2 Cara Objetivo - Mujer sonriendo [38].....	148
Figura 0.3 Caras Candidatas - Personas Sonriendo [38]	148
Figura 0.4 Atributos Faciales [38].....	149
Figura 0.5 Diagrama de cara - puntos de referencia etiquetados [19]	150
Figura 0.6 Cabeza con ejes de rotación alrededor del eje x (pitch), de rotación y de rotación alrededor del eje y (yaw) etiquetados [19].....	151
Figura 0.7 Spoon: Arranque.....	158
Figura 0.8 Spoon: Entorno	159
Figura 0.9 Spoon: Conexión a Base de Datos	159
Figura 0.10 Spoon: Transformación.....	160
Figura 0.11 Spoon: Inicio.....	162
Figura 0.12 Spoon: Generar Filas	163
Figura 0.13 Spoon: Secuencia.....	164
Figura 0.14 Spoon: Calculadora	165
Figura 0.15 Spoon: Selección Valores.....	167
Figura 0.16 Spoon: Tabla de Entradas	170
Figura 0.17 Spoon: Tabla Salida	172
Figura 0.18 Spoon: Fórmula	175

Figura 0.19 Spoon: Actualización	176
Figura 0.20 Spoon: Ordenar filas.....	177
Figura 0.21 Filas Únicas.....	178
Figura 0.22 Spoon: Unión de Base de Datos.....	180
Figura 0.23 Spoon: Concatenar Campos	181
Figura 0.24 Spoon: Archivo CSV de Entrada.....	183
Figura 0.25 Spoon: Archivo Json de Entrada.....	185
Figura 0.26 Spoon: Agregar constantes	188
Figura 0.27 Spoon: Reemplazo en cadena	188
Figura 0.28 RM: Arranque.....	191
Figura 0.29 RM: Carga de Datos.....	191
Figura 0.30 RM: Resultado de conjunto de datos.....	192
Figura 0.31 RM: Tipos de datos del conjunto de datos.....	192
Figura 0.32 RM: Medidas de calidad del conjunto de datos	193
Figura 0.33 Visualizaciones del conjunto de datos.....	193
Figura 0.34 Turbo Prep: Selección del conjunto de datos	194
Figura 0.35 Turbo Prep: Carga de Datos.....	194
Figura 0.36 Turbo Prep: Importación de datos.....	194
Figura 0.37 Turbo Prep: Localización de datos	195
Figura 0.38 Turbo Prep: Especificación del formato de datos.....	195
Figura 0.39 Turbo Prep: Opciones para columnas de datos	196
Figura 0.40 Turbo Prep: Almacenamiento de los datos	196
Figura 0.41 Turbo Prep: Conjunto de datos cargados	197
Figura 0.42 Turbo Prep: Transformación de columnas.....	197
Figura 0.43 Turbo Prep: Función limpieza de datos.....	198
Figura 0.44 Turbo Prep: Generar columna.....	198
Figura 0.45 Turbo Prep: Función pivot.....	198
Figura 0.46 Turbo Prep: Modelado.....	199
Figura 0.47 Turbo Prep: Historial de cambios.....	199
Figura 0.48 Turbo Prep: Commit de transformación	199
Figura 0.49 Turbo Prep: Opción Modelado	200
Figura 0.50 Turbo Prep: Carga de datos para el modelado	200
Figura 0.51 Turbo Prep: Selección de tarea.....	201
Figura 0.52 Turbo Prep: Preparación de objetivos	201
Figura 0.53 Turbo Prep: Definición de costos	202
Figura 0.54 Automodelado: Detalle de las columnas	203
Figura 0.55 Automodelado: Selección de entradas	204
Figura 0.56 Auto Modelado: Tipos de modelos	204
Figura 0.57 Auto Modelado: Resultados.....	205
Figura 0.58 Automodelado:Comparación de modelos	205
Figura 0.59 Automodelado:Modelo Random Forest.....	206
Figura 0.60 Automodelado:Random Forest -Pesos	206
Figura 0.61 Automodelado: Random Forest - Simulador.....	207
Figura 0.62 Automodelado: Encontrar entradas óptimas.....	207
Figura 0.63 Auto Modelado: Definir limitaciones y constantes.....	208
Figura 0.64 Auto Modelado: Optimización de Parámetros	209
Figura 0.65 Automodelado:Factores de correspondencia a variables.....	209
Figura 0.66 Automodelado: Corrida de la optimización	210
Figura 0.67 Automodelado: Rendimiento de Random Forest	210
Figura 0.68 Automodelado: Tabla de Elevación	211
Figura 0.69 Automodelado: Random Forest - Predicción	211
Figura 0.70 Despliegue Random Forest.....	211
Figura 0.71 Despliegue del modelo.....	212
Figura 0.72 Localización del despliegue.....	212
Figura 0.73 Nueva localización de despliegue.....	213
Figura 0.74 Nueva carpeta de almacenamiento de despliegue	213
Figura 0.75 Monitoreo de despliegue.....	214
Figura 0.76 Puntuación - Despliegue.....	214
Figura 0.77 Pentaho Server Ce: Arranque	215

Figura 0.78	Pentaho Server Ce: Consola.....	216
Figura 0.79	Pentaho Server Ce: Ubicación Logs Tomcat.....	216
Figura 0.80	Comandos para verificar puertos en Ubuntu	216
Figura 0.81	Marketplace - Saiku Analytics	216
Figura 0.82	Meteorite bi.....	217
Figura 0.83	Error licencia Saiku.....	217
Figura 0.84	Ruta de licencia de Saiku	217
Figura 0.85	Inicio Manage Data Source	218
Figura 0.86	Creación de Data Source.....	218
Figura 0.87	Pentaho Server Ce: Conexión a base de datos	219
Figura 0.88	Pentaho Server Ce: Prueba de Conexión.....	219
Figura 0.89	Asistente para creación de Data Source	220
Figura 0.90	Pentaho Server: Tablas de hechos y dimensiones	220
Figura 0.91	Uniones hechos y dimensiones.....	221
Figura 0.92	Versiones Pentaho Server y Conector Mysql	221
Figura 0.93	Rutas para los drives de base de datos	222
Figura 0.94	Ubicación de drivers de mysql.....	222
Figura 0.95	Confirmación de la creación de un recurso de datos.....	222
Figura 0.96	Editor de modelo de fuente de datos.....	223
Figura 0.97	Inicio de Saiku Analytics	223
Figura 0.98	Crear una nueva consulta en Saiku.....	223
Figura 0.99	Entorno Saiku Analytics	224
Figura 0.100	Saiku: Selección de medidas y dimensiones.....	224
Figura 0.101	Saiku: Visualización.....	225
Figura 0.102	Inicio de Visualizador - Datafor	226
Figura 0.103	Entorno Visualizador - Datafor	226
Figura 0.104	Visualizador - Datafor: Barra de herramientas	227
Figura 0.105	Visualizador - Datafor: Visualización de Página.....	227
Figura 0.106	Visualizador - Datafor: Componentes de página.....	228
Figura 0.107	Visualizador - Datafor: Componentes Gráficos	228
Figura 0.108	Visualizador - Datafor: Componentes de mapas.....	229
Figura 0.109	Visualizador - Datafor: Componentes para asistencia.....	229
Figura 0.110	Visualizador - Datafor: Componentes para filtros.....	229
Figura 0.111	Visualizador - Datafor: Datos de las medidas.....	230
Figura 0.112	Visualizador - Datafor: Opciones para el estilo de las medidas.....	231
Figura 0.113	Visualizador - Datafor: Título del cubo de datos.....	231
Figura 0.114	Visualizador - Datafor: Fondo y borde del componente	232
Figura 0.115	Visualizador - Datafor: Color de los datos.....	232
Figura 0.116	Fuente del componente de medida	232
Figura 0.117	Visualizador - Datafor: Formato para datos de comparación.....	232
Figura 0.118	Visualizador - Datafor: Colores de comparación	233
Figura 0.119	Visualizador - Datafor: Marco de diseño	233
Figura 0.120	Visualizador - Datafor: Menú de componentes	233

LISTA DE TABLAS

Tabla 1.1 Comparativo de Metodologías Inmon, Kimball y Hefesto [13]	5
Tabla 1.2 Ejemplos de Errores Absolutos y Relativos	22
Tabla 2.1 Factibilidad de responder vs datos disponibles	31
Tabla 2.2 Prioridad de Respuestas.....	35
Tabla 2.3 Comparativa de Modelos en RM - Automodel.....	36
Tabla 2.4 Atributos de Relevancia en Modelo Random Forest	37
Tabla 2.5 Indicadores y perspectivas	39
Tabla 3.1 Registros con errores encontrados.....	112
Tabla 3.2 Consolidado de puntaje SUS	120
Tabla 0.1 Funciones Calculadora	167

LISTA DE ANEXOS

Anexo I Convocatoria para el Taller de Robótica Educativa	132
Anexo II Talleres de Robótica Educativa.....	133
Anexo III Fotografías de Talleres	134
Anexo IV Archivo JSON- Reconocimiento Facial.....	135
Anexo V Archivos de Excel o CSV de Reconocimiento Facial	136
Anexo VI Encuesta en Google Forms para director y maestros	137
Anexo VII Preguntas Recolectadas en Video Conferencia y encuestas con Usuarios Rector, Profesor, Padre de Familia y Estudiante.....	144
Anexo VIII Servicio Azure Face.....	147
Anexo IX Ejemplos de fotografías con ruido	154
Anexo X Tutorial de Pentaho Data Integration. Kettle- Spoon.....	158
Anexo XI Tutorial RapidMiner.....	190
Anexo XII Tutorial de Pentaho BI-Server	215
Anexo XIII Tutorial DATAFOR.....	226

RESUMEN

Este proyecto tiene como objetivo desarrollar un Sistema de Indicadores para monitorear el proceso de aprendizaje, trayendo como caso de estudio Talleres de Robótica de la Unidad Educativa Verbo y utilizando Hefesto como metodología de desarrollo de software en Inteligencia de Negocios.

En el primer capítulo se describe la situación actual de la Unidad Educativa Verbo en cuanto a Talleres de Robótica y de los procesos que se implementarán en el proyecto. Además, se justifica el uso de la metodología y de las herramientas de desarrollo.

En el segundo capítulo se realiza el análisis de los requerimientos, la aplicación de la metodología Hefesto, se recolectan los requerimientos de información específicos de los diferentes usuarios, se identifican indicadores y perspectivas de análisis, se observa el Procesamiento de Transacciones en Línea (OLTP) para construir los indicadores y se emplea un modelo para realizar la predicción de la experiencia de usuario para la aplicabilidad del taller de RE. Se confeccionan las tablas de dimensiones y las tablas de hechos, posteriormente utilizando técnicas de limpieza y calidad de datos, procesos ETL, se definen estrategias para la construcción y carga del Data Warehouse (DW). Además, se construyen dashboards de emociones, conocimiento por cada fase antes, durante y después de la aplicación de los talleres.

En el tercer capítulo, como resultados se puede apreciar que en los talleres predomina la emoción neutral con un valor de 1552.589, seguida de felicidad en 283.618, lo que de forma general indica que el taller de RE tuvo un efecto positivo en los estudiantes y de acuerdo a los resultados obtenidos en la escala de usabilidad SUS, el sistema presenta un nivel aceptable mayor al 70%.

Finalmente, se presentan las conclusiones y recomendaciones obtenidas durante el desarrollo del proyecto.

Palabras clave: análisis de sentimientos, proceso de aprendizaje, reconocimiento facial, emociones, aplicación web, robótica educativa.

ABSTRACT

The objective of this project is to develop a System of Indicators to monitor the learning process, using Robotics Workshops of the Verbo Educational Unit as a case study and using Hefesto as a software development methodology in Business Intelligence.

The first chapter describes the current situation of the Verbo Educational Unit in terms of Robotics Workshops and the processes that will be implemented in the project. In addition, the use of the methodology and development tools is justified.

In the second chapter, the requirements analysis is carried out, the application of the Hefesto methodology, the specific information requirements of the different users are collected, indicators and analysis perspectives are identified, the Online Transaction Processing (OLTP) to build the indicators and a model is used to make the prediction of the user experience for the applicability of the RE workshop. The dimension tables and fact tables are made, later using cleaning techniques and data quality, ETL processes, strategies are defined for the construction and loading of the Data Warehouse (DW). In addition, dashboards of emotions and knowledge are built for each phase before, during and after the application of the workshops.

In the third chapter, as results it can be seen that neutral emotion predominates in the workshops with a value of 1,552,589, followed by happiness in 283,618, which generally indicates that the RE workshop had a positive effect on students and According to the results obtained in the SUS usability scale, the system presents an acceptable level greater than 70%.

Finally, the conclusions and recommendations obtained during the development of the project are presented.

Keywords: sentiment analysis, learning process, facial recognition, emotions, web application, educational robotics.

1. INTRODUCCIÓN

1.1. Descripción del problema

La institución Educativa Particular Cristiano Verbo reportó a través de reuniones con las autoridades que el proceso de monitoreo del aprendizaje es complicado, toma mucho tiempo, se realiza de forma manual, involucra muchos profesores y hay problemas de consistencia o estándar al agrupar la información puesto que se usan diferentes formatos para recopilar la información. Por tanto, este reporte evidencia que el monitoreo es manual y que el tiempo en la toma de decisiones en el proceso de aprendizaje es cuantioso. Esto nos da un primer síntoma en la etapa inicial de levantamiento de requerimientos y problemas en un sistema de Inteligencia de Negocio (BI) ([1], etapa1).

Se puede apreciar por el indicio anteriormente expuesto que la causa del problema es que no existe un instrumento (sistema) para monitorear el aprendizaje de Talleres de Robótica Educativa (RE) y que sea accesible a las autoridades de la institución.

El problema de no cuantificar el efecto en un individuo al momento de aplicar una nueva tecnología en su aprendizaje no permite determinar la relevancia de las actividades empleadas que podrían ser consideradas como una buena práctica durante la enseñanza.

Considerando el contexto ecuatoriano actual, donde el uso de pedagogías lúdicas robóticas digitales en las escuelas y colegios es bajo, es necesario considerar el seguimiento pedagógico, psicológico, y técnico de este modelo de aprendizaje [2].

1.1.1 Historia de la empresa

La Unidad Educativa Cristiana Verbo es una comunidad educativa particular, cristiana e inclusiva, orientada a formar niños, niñas y adolescentes con una educación de calidad y calidez, basada en principios y valores bíblicos; manteniendo el más alto nivel académico, descubriendo la individualidad y potencial de cada estudiante, para que sus dones y talentos sean usados en la sociedad a la luz de la verdad de Dios.

1.2. Objetivo general

Desarrollar un sistema de indicadores para monitorear el proceso de aprendizaje Caso de Estudio Taller de Robótica de la Unidad Educativa Verbo.

1.3. Objetivos específicos

Los objetivos para la consecución del objetivo general para monitorear el proceso de aprendizaje a través de la robótica educativa empleada en el Taller Colaborativo de Robótica Educativa aplicado en la Unidad Educativa Particular Cristiano Verbo son:

- Diagnosticar la información disponible para medir los indicadores
- Diseñar el sistema de indicadores
- Implementar la integración de datos
- Evaluar el uso del sistema de indicadores

1.4. Alcance

En este proyecto se realiza un sistema de indicadores para monitorear el proceso de aprendizaje a través del Taller Colaborativo de Robótica Educativa aplicado en la Unidad Educativa Particular Cristiano Verbo.

Se utiliza la Metodología Hefesto para satisfacer a la gama de usuarios junto a sus necesidades específicas en cuanto al análisis de información.

1.5. Marco teórico

1.5.1 Descripción de la situación actual

El uso de robótica educativa atrae un gran interés de los educadores e investigadores [2] como una herramienta [3] disponible para desarrollar habilidades cognitivas y sociales para los estudiantes de diferentes niveles de educación y se puede aplicar para el aprendizaje [4] en ciencias, matemáticas tecnología, informática y otras disciplinas [5].

El sistema educativo ecuatoriano ha mantenido un desarrollo desigual y diferenciado fruto de inequidades socioeconómicas y una ineficaz cobertura de servicios públicos en el territorio [6].

Considerando el contexto ecuatoriano actual, donde el uso de pedagogías lúdicas robóticas digitales en las escuelas y colegios es bajo, es necesario considerar el seguimiento pedagógico, psicológico, y técnico del modelo de aprendizaje.

Sin embargo, en la actualidad, no se ha realizado un análisis de los datos obtenidos en la aplicación de pedagogías lúdicas robóticas digitales en las escuelas y colegios.

Es necesario realizar la evaluación de la experiencia de usuario con el uso de las emociones como variables para medir el impacto en el aprendizaje del usuario por medio del uso de la RE [7].

La RE también ha sido usada para motivar a los estudiantes en el estudio de materias relacionadas con las TICs. Por ejemplo, la universidad de Georgia [8] enfrentó la deserción y el desinterés de los jóvenes por estudiar ingeniería de sistemas gracias a

la RE reduciendo el porcentaje de aceptación en más del 60%. Principalmente, esta universidad decidió implementar la robótica en el curso de introducción. Aquí los estudiantes desarrollaron capacidades en los fundamentos básicos de programación gracias al uso de un robot personalizado. El software de programación fue basado en JAVA (Myro) y por el lado de hardware fue el Scribbler Parallax, un robot de bajo costo, pero de estructura muy robusta. Por tanto, la RE puede ser una estrategia pedagógica para motivar el aprendizaje de las ciencias informáticas en los estudiantes en tercer nivel.

En la interacción hombre-computadora, las emociones de los usuarios son reconocidas como importantes en el diseño y procesos de evaluación [9]. En este contexto, el campo de la computación afectiva es sinónimo de "computación que se relacionan con, surgen de, o influyen en las emociones". De esta noción, hay modelos y herramientas tales como Affdex (análisis facial para la detección de emociones) [10] para medir los efectos en términos de emociones básicas de Modelo de Ekman (sorpresa, felicidad, tristeza, asco, ira, miedo) y neutral.

Además, también hay estudios para explorar la relación de las emociones con las características de otros usuarios, como la capacidad de aprendizaje, el rendimiento y la comunicación en algunos ámbitos específicos.

Las emociones se miden e interpretan en términos de respuestas a estímulos ("algo que despierta o incita a la actividad"; entonces, sobre la base de las mediciones, se decide en relación con las acciones aplicables; estas acciones se ejecutan; y el bucle comienza de nuevo con nuevas medidas de emociones.

La detección de emociones podría permitir un monitoreo del aprendizaje del usuario (alumno) como complemento de las observaciones explícitas (métodos antes mencionados). Aunque son estudiados en estos contextos, su uso en RE para el monitoreo y control de emociones no está siendo explorado a profundidad.

Uso de BI en el contexto actual referido a la educación

La inteligencia de negocios (Business Intelligence - BI), permite que el proceso de toma de decisiones esté fundamentado sobre un amplio conocimiento de la información, propicia que las organizaciones puedan traducir sus objetivos en indicadores de estudio, y que estos puedan ser analizados desde diferentes perspectivas, con el fin de encontrar información que no solo se encargue de responder a preguntas de lo que está sucediendo o ya sucedió, sino también, que posibilite la construcción de modelos, mediante los cuales se podrán predecir eventos futuros.

Cabe destacar, que la aplicación de soluciones BI no es solo para grandes-medianas empresas, sino para quien desee tomar decisiones a través del análisis de sus datos.

[1]

En la actualidad, BI cuenta con una variedad de tecnologías que están diseñadas para dar potencial al análisis de la información de una organización dependiendo de sus características, necesidades, presupuesto, tamaño, actividad y la relevancia de los datos.

Las unidades educativas particulares cuentan con una infraestructura suficiente para poder llevar a cabo la recolección de datos y análisis de la información adecuada para elaborar una implementación BI.

El uso Business Intelligence permite optimizar gastos, financiación, recursos, tiempos, así como la elaboración de un plan estratégico para mejorar la experiencia de los estudiantes.

Las escuelas pueden aumentar y mejorar la gestión de sus recursos al conocer las necesidades de sus estudiantes.

Implementación del taller de RE

Se realizó el levantamiento y recopilación de la información mediante la ejecución de un Taller Colaborativo de RE para monitorear el proceso de aprendizaje basado en emociones como enseñanza lúdica e inclusiva a través de la tecnología en niños de 7 a 10 años. Un grupo de 16 alumnos participaron en talleres de RE como son los siguientes: Ordenar para Reciclar y Metamorfosis de la Rana, pero al momento se desconocen los resultados de aplicar estos talleres para realizar una toma de decisiones respecto a resultados emocionales, experiencia de usuario para de esta forma continuar o no con su uso.

Los datos disponibles del taller de RE se manejan hojas de excel y pdf, la información tanto de los estudiantes que participaron como los talleres realizados constan en los ANEXO I y ANEXO II respectivamente.

Esta información corre el riesgo de perderse, presentar inconsistencias y volverse inútil de darle un seguimiento adecuado.

1.5.2 Justificación de la metodología

Análisis de metodologías de inteligencia de negocios

Existen varias metodologías que pueden ser utilizadas para el desarrollo de un Data Warehouse; sin embargo, las que se imponen entre la mayoría son las metodologías de Inmon, Kimball y Hefesto.

Inmon propone una metodología top-down que transfiere la información de los diferentes procesamientos de transacciones en línea (OLTP) a un repositorio centralizado (Inmon, 2005) [11]. Kimball & Ross (2002) por su parte, propone una metodología bottom-up que permite construir un Data Warehouse de forma escalonada considerando los Data Marts ya creados para construir el DW. [12]

Hefesto es una metodología que combina las mejores prácticas de metodologías existentes como por ejemplo Inmon y Kimball, y permite crear un Data Warehouse a partir de la recolección de requerimientos, seguido de un proceso de extracción, transformación y carga (ETL), hasta definir un esquema lógico sea este un Data Warehouse o Data Mart (Bernabeu, 2010). [13]

En la Tabla 1.1, se presenta un cuadro comparativo de las metodologías mencionadas con sus características más relevantes. [14]

Tabla 1.1 Comparativo de Metodologías Inmon, Kimball y Hefesto [13]

CARACTERÍSTICAS	INMON	KIMBALL	HEFESTO
Enfoque	Top-down	Bottom up	Flexible Top-down o Bottom up
Complejidad	Alta	Baja	Baja
Orientación de la data	Maneja data o áreas	Orientado a procesos	Flexible maneja data , orientado a procesos
Herramientas	Tradicional(Entidad-Relación-ER y Data Item Sets-DIS)	Modelo dimensional	ER, DIS, Modelo Dimensional
Accesibilidad del usuario final	Bajo	Alto	Alto
Objetivos	Ofrecer una buena solución técnica basada en métodos probados	Ofrecer una solución que sea fácil para usuarios finales, consulta de datos y tener una tasa de respuesta razonable	Ofrecer una buena solución técnica que sea fácil para usuarios finales , consulta de datos y tener una tasa de respuesta razonable
Naturaleza de los requerimientos	Estratégica	Táctica	Estratégica/Táctica
Requerimientos de integración de datos	Organización	Áreas de negocio individual	Organización/Áreas de negocio individual.
Estructura de datos	No métricas	Métricas de negocio, rendimiento y cuadros de mando	Métricas de negocio, rendimiento y cuadros de mando.
Persistencia de la data	Alta tasa de cambio de sistemas fuentes	Los sistemas fuentes son relativamente estables.	Los sistemas fuentes son relativamente estables.
Requerimientos de personal y habilidades	Equipos grandes de especialistas	Equipos pequeños generalistas	Equipos pequeños o grandes de especialistas.
Tiempo de entrega	Los requerimientos de la organización permiten la puesta en marcha con más tiempo	La necesidad de la primera aplicación DW es urgente	Puede iniciar con un data mart y continuar al DW.
Costo de implementación	Altos costos de inicio con menores costos de desarrollo de proyectos posteriores	Menor costo de arranque, con cada proyecto posterior el costo es el mismo.	Bajo Costo, de acuerdo a posibilidades de la empresa.

1.5.3 Selección de metodología

Al analizar estas características correspondientes a algunas metodologías como Inmon, Kimball y Hefesto, se observa que Hefesto es la más adecuada para los objetivos buscados en este proyecto, ya que se adapta debido a que su aspecto fundamental lo constituyen los requerimientos de los usuarios, permitirá la construcción de un DW [15] de forma sencilla, ordenada e intuitiva y con varios casos de éxito implementados. También, se lo elige debido a que reduce la resistencia al cambio, al involucrar a los usuarios finales en cada etapa de la construcción, y además se entregan resultados inmediatos.

Se utilizará la Metodología Hefesto para satisfacer a la gama de usuarios junto a sus necesidades específicas en cuanto al análisis de información [1].

De acuerdo con Hefesto y aplicado al presente trabajo el proceso está dividido en cuatro fases:

En la primera fase es donde se deberán recolectar los requerimientos de información específicos de los diferentes usuarios, así como entender sus diversas necesidades, para que luego en conjunto con ellos se generen las preguntas que les ayudarán a alcanzar sus objetivos. Luego, se deben identificar los indicadores resultantes de los interrogativos y sus respectivas perspectivas de análisis.

En la segunda fase, se analizarán los OLTP para determinar cómo se construirán los indicadores, señalar las correspondencias con los datos fuentes y para seleccionar los campos de estudio de cada perspectiva.

En la tercera fase se pasará a la construcción del modelo lógico del depósito DW, en donde se definirá cuál será el tipo de esquema que se implementará. Seguidamente, se confeccionarán las tablas de dimensiones y las tablas de hechos, para luego efectuar sus respectivas uniones.

Por último, utilizando técnicas de limpieza y calidad de datos, procesos ETL, se definirán políticas y estrategias para la carga inicial del DW y su respectiva actualización.

El diseño de un sistema de indicadores para monitorear el aprendizaje de niños a través de la robótica empleada en el Taller Colaborativo de RE aplicado en la Unidad Educativa Particular Cristiano Verbo ayudará tanto a autoridades, profesores, padres de familia de dicha entidad a conocer el impacto de la aplicación de la robótica educativa.

Con la información recolectada y siguiendo la metodología Hefesto, se construirán dashboards de emociones, conocimiento por cada fase antes, durante y después de la aplicación de los talleres.

Los dashboards permitirán a los diferentes actores involucrados monitorear [7], encontrar respuestas a preguntas planteadas y tomar decisiones respecto a la aplicación de la RE.

En la Figura 1.1 La metodología HEFESTO puede resumirse a través del siguiente gráfico:



Figura 1.1 Pasos de Metodología HEFESTO con Visualización [1]

1.5.4 Selección de herramientas

Las herramientas que se han elegido para el desarrollo del presente proyecto, han sido seleccionadas en base a parámetros generales como son costo de licencia, costo de implementación, crecimiento de la organización; así como, facilidad de uso, acceso y son las siguientes:

- **Características del servidor:** Se trabaja en una máquina virtual con las características siguientes:
VMware Workstation 16Pro
Version:10.0.0
Memoria: 8GB
Procesadores: 4
Disco duro:120MB
SO: 5.7.31-0ubuntu0.18.04.1
- **Gestor de Base de Datos:** Mysql-community versión.
- **Herramienta visual de diseño de bases de datos:** MySQL.Workbenchy.0.3 Community.
- **Hosting Público para Imágenes:** <https://cloudinary.com/>. Ver Anexo XIV
- **Servicio Azure Face:** Atributos y detección de caras.
- **ETLs:** Pentaho Data Integration Stable Release -Version 4.3.0, Kettle-spoon. Ver Anexo X Kettle-spoon.
- **Herramientas para Inteligencias de Negocio:** Pentaho -community versión.
PENTAHO BI SERVER: pentaho-server-ce-8.3.0.0-371. Ver Anexo XII Sección Tutorial de Pentaho BI-Server
- **BI olap cubes:** SAIKU. Ver Anexo XII Sección Tutorial de Saiku.
- **Calidad de Datos:** RapidMiner(RM) Studio Versión 9.8. Educational Edition. Ver Anexo XI RapidMiner
- **Data Mining:** Ver Anexo XI RapidMiner, Sección TurboPrep
- **BI Dashboards:** Datafor. Ver Anexo XIII Tutorial Datafor

Mysql Workbench 0.3 Community

MySQL Workbench es una herramienta visual de diseño de bases de datos que integra desarrollo de software, administración de bases de datos, diseño de bases de datos, gestión y mantenimiento para el sistema de base de datos MySQL.

En la Figura1.2, se pueden apreciar algunas de las funcionalidades que fueron requeridas en el presente proyecto, así como dificultades presentadas por este gestor de base de datos fueron las siguientes:

- Instalación de Workbench. [16]
- RespalDOS. [17]
- Parametrizar el editor para que permita sentencias condicionales de actualización y eliminación sin restricciones.

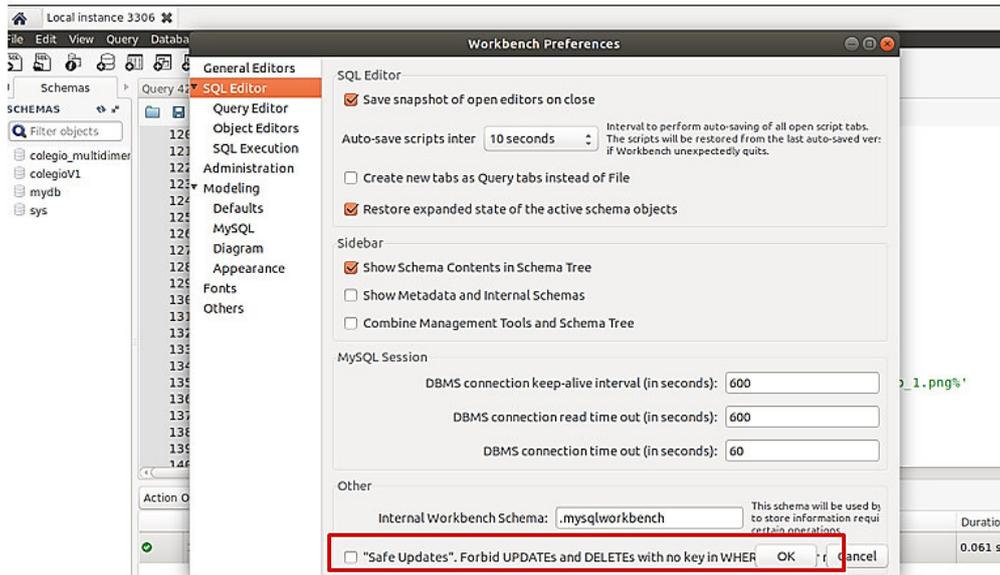


Figura 1.2 Preferencias Editor Workbench

- Cambio de las parametrizaciones, en la Figura 1.3, se indica la opción del editor de Workbench donde por defecto limita las consultas de registros a 1000. [18]

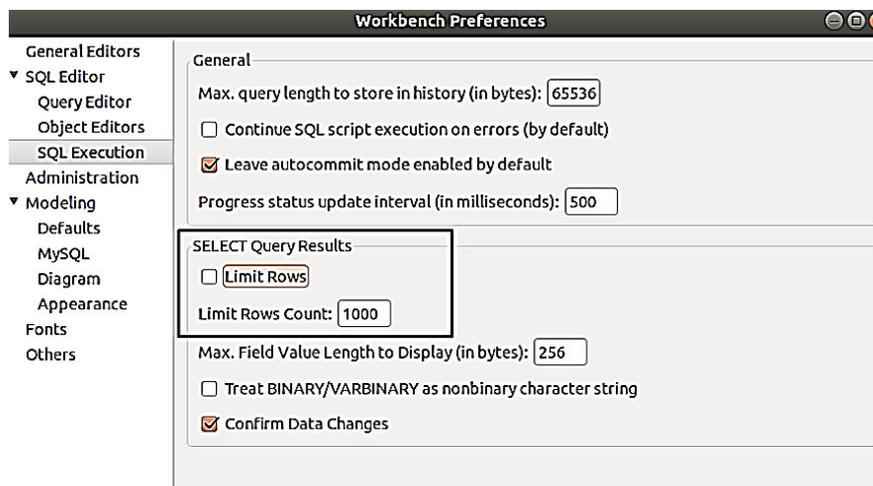


Figura 1.3 SQL Execution - Editor Workbench

- Para restaurar bases de datos en workbench:
 1. Ir al panel Management
 2. Opción Data Import/Restore, Figura 1.4.
 3. Activar la opción Import from self-contained file
 4. Seleccionar el archivo backup.sql
 5. Empezar la importación

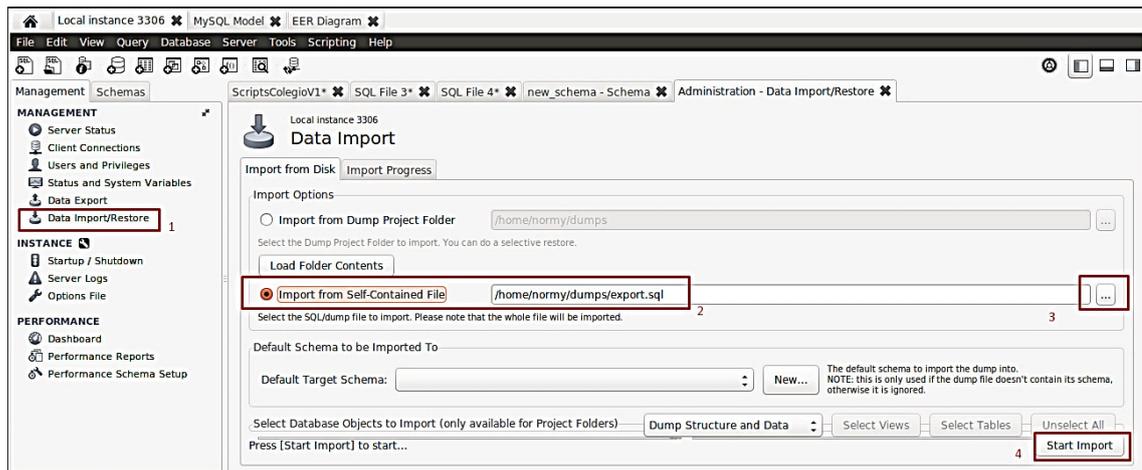


Figura 1.4 Workbench: Restauración de base de datos

Microsoft Azure y Hosting público

Para realizar el análisis de una imagen con Azure Face REST API es necesario realizar una suscripción en Microsoft Azure, en este caso se la realiza con una cuenta de estudiante, Figura 1.5.

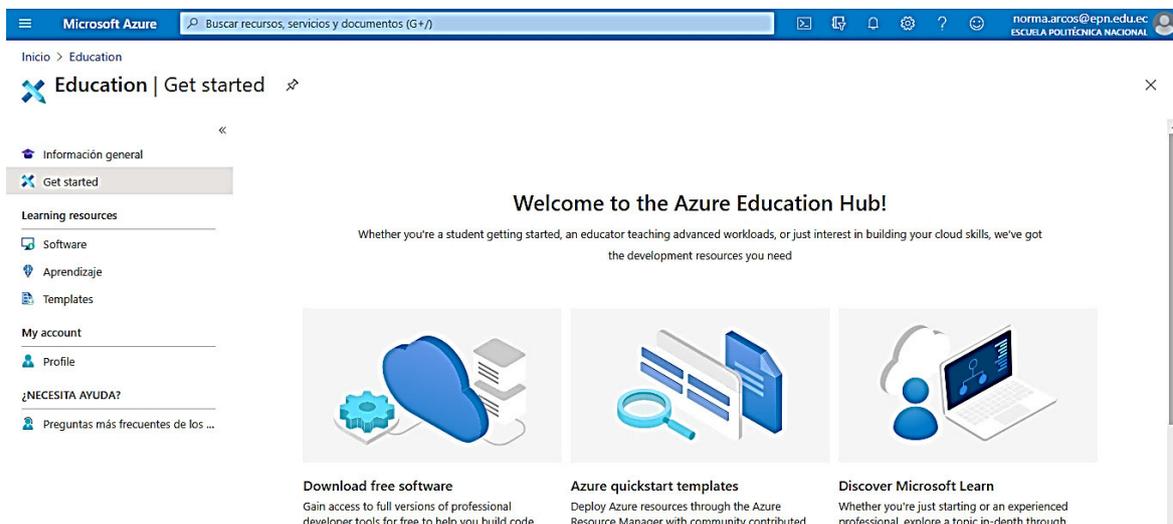


Figura 1.5 Panel de Azure -Cuenta Estudiante

Con Face de Azure no se crea el reconocimiento facial desde cero, se utiliza una API [20] para ofrecer esta funcionalidad en sus aplicaciones. Ver Anexo VIII.

Pentaho

Existen varias herramientas para la realización de ETLs y creación de DWs, en la Figura 1.7 se presenta un cuadro comparativo entre dos herramientas representativas y por lo cual se justifica el uso de pentaho en el presente trabajo.

	JasperSoft	Pentaho
Tipo de Licencia	JasperSoft LGPL V3 iReport GPL V3	Pentaho Reporting LGPL V2.1
Lenguaje de desarrollo	Modelo normalizado Java, Perl	Java
Desarrollo de Plug-in	Eclipse Plug-in Available	Eclipse Plug-in Available
JDBC Driver	6 conexiones a fuentes	37 conexiones a fuentes
Compilación de Reporte	SI	NO
Plataformas soportadas	Windows, Linux, Mac OS X	Windows, Linux, Mac OS X
Servidor de Aplicaciones	JBoss	JBoss
Servicios web	Tomcat	Tomcat
Reporte/Gráficos	Si (JasperReport)	Si (Pentaho Report Designer)
Cuadros de Mando	Si (JFreeChart)	Si (JFreeChart)
Herramienta ETL	Si (JasperETL)	Si (Pentaho Data Integration)
Soporte para Minería de Datos	No	Si (Weka)

Figura 1.7 Comparación de Herramientas para la creación del DW [21]

Pentaho es una plataforma de análisis e integración de datos de código abierto, habilita la automatización, orquestación e integración de flujos de datos con volúmenes de datos masivos y heterogéneos de manera sencilla en una sola herramienta.

Los usuarios pueden capturar, mezclar, limpiar y preparar datos de distintas fuentes. Pentaho elimina la codificación, la complejidad, y brinda la mejor calidad de datos para las áreas de TI y BI con herramientas visuales para una muy buena experiencia de usuario.

En la Figura 1.8 se puede apreciar la Pentaho Suite: Business Analytics, Data Integration y Reporting Services que permiten diseñar arquitecturas empresariales de datos de nueva generación, altamente escalables, integrables y lo más importante: siendo costo-efectivos.

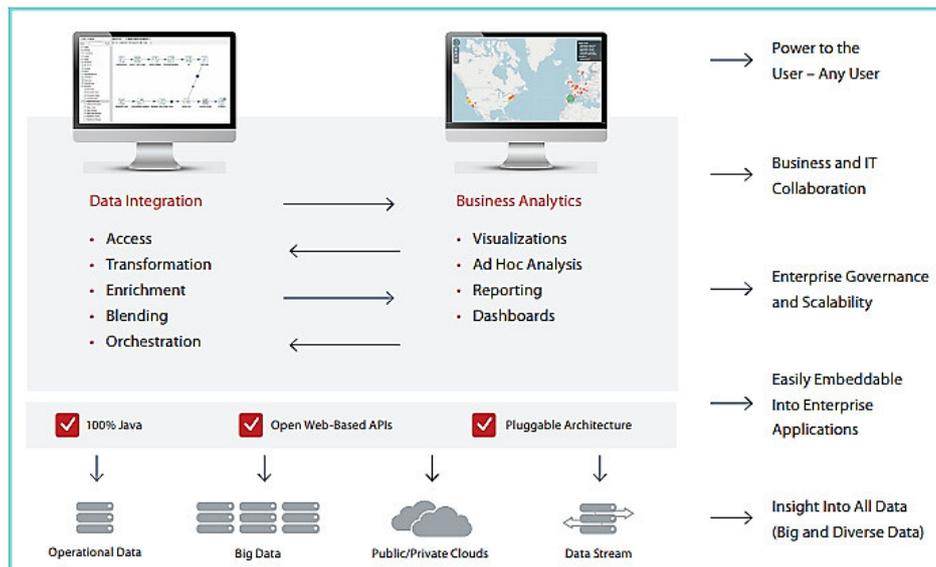


Figura 1.8 Plataforma de Análisis Empresarial Pentaho [22]

Saiku

Saiku permite a los usuarios empresariales explorar fuentes de datos complejas, utilizando una interfaz familiar de arrastrar y soltar y una terminología empresarial fácil de entender, todo dentro de un navegador. Se pueden seleccionar los datos de interés, mirarlos desde diferentes perspectivas, profundizar en detalles y una vez se tenga una respuesta guardar los resultados, compartirlos, exportarlos a Excel o PDF, directamente desde el navegador. Saiku se conecta a una amplia gama de fuentes de datos, lo que le permite explorar los datos en tiempo real directamente desde la fuente. [23]

RapidMiner

RapidMiner Studio (RM) es el programa de minería de datos ofrecido por la compañía que lleva el mismo nombre. La iniciativa nació en 2001, en la unidad de Inteligencia artificial de la Universidad Tecnológica de Dortmund en Alemania.

Diseñado inicialmente como un software libre, ha evolucionado hasta convertirse en una plataforma cruzada que combina minería de datos, con minería de texto, aprendizaje de máquina, Inteligencia de negocios y análisis de negocios. RM permite procesar variables numéricas y no numéricas a la vez, con gráficos, respuesta rápida que soporta varias fuentes de datos, potencia los procesos de negocios. Con este software se puede acelerar la adquisición de datos y el análisis para tomar acciones rápidamente. El problema es que no es una bodega de datos y solo ofrece 1 Gb de RAM gratis, de ahí en adelante es pagado.

Cabe mencionar que la versión académica RapidMiner de duración de un año es la que se utiliza en el presente proyecto.

Turbo Prep

Mejora el proceso clásico de RM, está diseñado para facilitar la preparación de datos de forma guiada sin la necesidad de generar un proceso paso a paso por su cuenta. Proporciona una interfaz de usuario en la que los datos siempre están visibles al frente y al centro, donde se puede realizar cambios paso a paso y ver los resultados al instante o refinarlos según sea necesario, con una amplia gama de funciones de apoyo para preparar los datos para la creación de modelos o su presentación. Permite analizar los datos por medio de gráficos o verificando los indicadores de calidad. Es posible acceder directamente al Auto modelado para la generación de modelos de análisis de datos.

Visualizer - Datafor

Datafor (Shanghai Shuwei Information Technology Co., Ltd.) es una empresa de tecnología dinámica fundada en 2015. La empresa se centra en el desarrollo de productos y el diseño de soluciones en los campos de visualización de big data, visualización de datos de series de tiempo, inteligencia empresarial y análisis de big data. Después de que la empresa Hitachi adquirió Pentaho, Datafor se convirtió en socio de Big Data de Hitachi Vantara, recibiendo el "Premio a la Mejor Innovación 2019".

Pentaho, una innovadora plataforma de análisis integrado de big data, es la plataforma de análisis integrado de big data de código abierto más popular del mundo. El visualizador de Datafor es el complemento de la visualización y análisis de datos en la plataforma Pentaho PBA.

Algunas de sus características son las siguientes:

- Dispone de herramientas ágiles de visualización y análisis de datos
- Complemento para la plataforma de análisis de inteligencia empresarial
- Aplicable a Pentaho Business Edition y Community Edition

Las capacidades de análisis y visualización de datos del visualizador complementan y mejoran las funciones "Agile BI" y visualización de datos de series temporales de Pentaho.

Las potentes funciones de la plataforma de Pentaho y el motor de análisis multidimensional respaldan la implementación y las capacidades de acceso a datos de Visualizer, Figura 1.9.

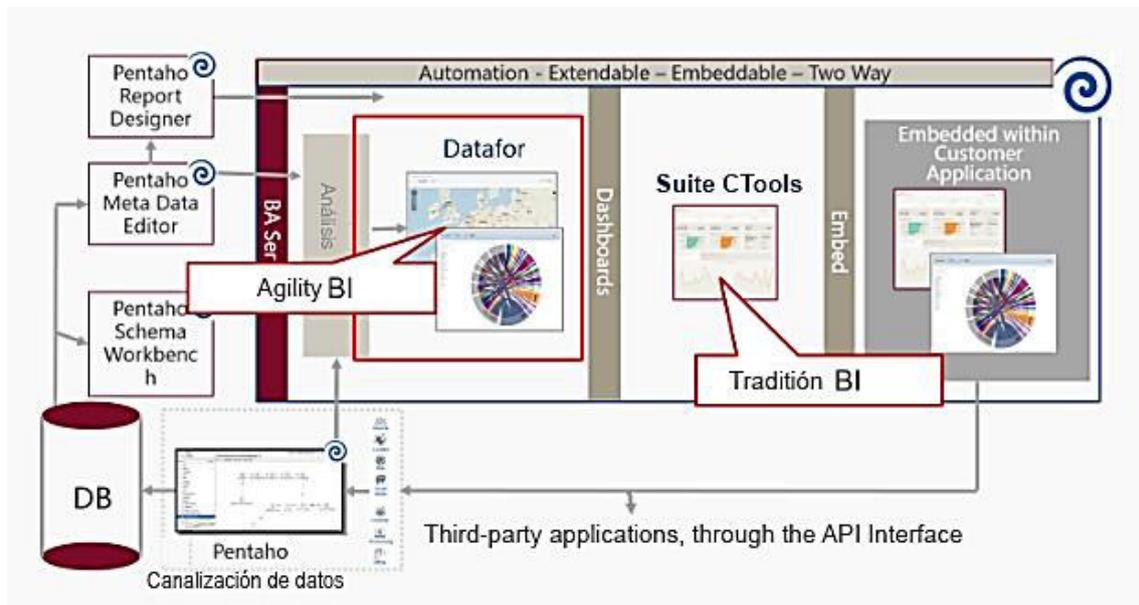


Figura 1.9 Pentaho y Visualizer [24]

1.5.5 Bases Teóricas

Es necesario realizar la revisión general ciertas definiciones que intervienen a lo largo de este proyecto.

Indicadores

Un indicador es una característica específica, observable y medible, se aprecian los siguientes:

- Estáticos: se obtienen por consultas SQL y permiten apreciar cambios en cuanto a cantidades.
- Dinámicos: se obtienen por modelamiento basado en data mining, clustering, prediction, clasificación de progresos que está provocando la aplicación del taller de RE en las emociones de estudiante y que permite mejorar la enseñanza de una materia específica.

OLTP

Representa toda aquella información transaccional que genera la empresa en su accionar diario, además, de las fuentes externas con las que puede llegar a disponer. Estas fuentes de información, son de características muy disímiles entre sí, en formato, procedencia, función, etc.

Los OLTP son diseñados para soportar el procesamiento de información diaria de las empresas, y el énfasis recae en maximizar la capacidad transaccional de sus datos. Su estructura es altamente normalizada permite brindar mayor eficiencia a sistemas con muchas transacciones que acceden a un pequeño número de registros y están fuertemente condicionadas por los procesos operacionales que deben soportar, para la óptima actualización de sus datos. Esta estructura es ideal para llevar a cabo el proceso transaccional diario, brindar consultas sobre los datos cargados y tomar decisiones diarias.

Entre los OLTP más habituales que pueden existir en cualquier organización se encuentran:

- Archivos de textos.
- Hipertextos.
- Hojas de cálculos.
- Informes semanales, mensuales, anuales, etc.
- Bases de datos transaccionales

Es aquí en donde se realiza el proceso de extraer desde las diferentes fuentes de información de la empresa, tanto internas como externas, los datos que serán necesarios para encontrar las respuestas a las preguntas planteadas en el paso anterior.

Una vez que los datos son seleccionados y extraídos, se guardan en un almacenamiento intermedio, lo cual permite, entre otras ventajas:

- Manipular los datos sin interrumpir ni paralizar los OLTP, ni tampoco el DW.
- No depender de la disponibilidad de los OLTP.
- Almacenar y gestionar los metadatos que se generarán en los procesos ETL.
- Facilitar la integración de las diversas fuentes, internas y externas.

El almacenamiento intermedio constituye en la mayoría de los casos una base en donde la información puede ser acumulada por ejemplo en tablas auxiliares, tablas temporales, etc. Los datos de estas tablas serán los que finalmente (luego de su correspondiente transformación) poblarán el DW.

Modelo lógico de DW

Debido a que, para llevar a cabo BI, es necesario gestionar datos guardados en diversos formatos, fuentes y tipos, para luego depurarlos e integrarlos, además de almacenarlos en un solo destino o base de datos que permita su posterior análisis y

exploración, es imperativo y de vital importancia contar con un proceso que satisfaga todas estas necesidades. Este proceso se denomina Data Warehousing.

Los datos son integrados, transformados y limpiados, para luego ser cargados en el DW para poder llevar a cabo procesos de consulta y análisis para luego tomar decisiones estratégicas y tácticas de alto nivel, Figura 1.10.



Figura 1.10 DW, características [1]

El almacenamiento de datos históricos, es lo que permite al DW desarrollar pronósticos y análisis de tendencias y patrones, a partir de una base estadística de información.

Principalmente, la información del DW se estructura en cubos multidimensionales, ya que estos preparan esta información para responder a consultas dinámicas con un buen rendimiento.

A continuación, se confecciona el modelo lógico de la estructura del DW, teniendo como base el modelo conceptual.

Se define el tipo de modelo a utilizar y luego se lleva a cabo las acciones para diseñar las tablas de dimensiones y de hechos. Finalmente, se realizan las uniones entre estas tablas. [1]

Tipo de Modelo Lógico del DW

Las bases de datos multidimensionales implican variantes posibles de modelamiento, que permiten realizar consultas de soporte de decisión: En este caso se hace uso del modelo de constelación.

No es recomendable usar copo de nieve a pesar de que se desee orden, tampoco el modelo estrella ya que se requiere de más de una tabla de hechos para trabajar con los indicadores dinámicos, por lo que se adopta el modelo Constelación.

El modelo Constelación, permite tener más de una tabla de hechos, por lo cual se podrán analizar más aspectos claves del negocio con un mínimo esfuerzo adicional de diseño.

Contribuye a la reutilización de las tablas de dimensiones, ya que una misma tabla de dimensión puede utilizarse para varias tablas de hechos. Como desventaja se tiene que este modelo no es soportado por todas las herramientas de consulta y análisis. [25]

Data Mart (DM)

Un Data mart es la implementación de un DW con alcance restringido a un área funcional, problema en particular, departamento, tema o grupo de necesidades.

Dentro de las ventajas de aplicar un DM a un negocio, se tienen las siguientes:

- Son simples de implementar.
- Conllevan poco tiempo de construcción y puesta en marcha.
- Permiten manejar información confidencial.
- Reflejan rápidamente sus beneficios y cualidades.
- Reducen la demanda del depósito de datos.

Experiencia de usuario

La experiencia de usuario evalúa las interacciones de los usuarios y los productos tecnológicos cuyo resultado es la generación de una percepción positiva o negativa de un servicio, producto o dispositivo. Se verá condicionada por expectativas y experiencias previas del usuario. Ofrece una perspectiva más amplia acerca del uso y consumo de productos interactivos, y por tanto más acorde con la realidad. [26]

Se preocupa por factores poco considerados como el comportamiento emocional del usuario, la importancia del diseño y la estética en este comportamiento.

La teoría dimensional conocida como modelo circunplejo del afecto tiene características que la hacen útil para la evaluación continua del estado emocional.

El modelo circunplejo de afecto, propuesto originalmente por Schlosberg en 1941, postula que la estructura subyacente de la experiencia afectiva puede caracterizarse como un ordenamiento de estados afectivos en la circunferencia de un círculo, como se muestra en la Figura 1.11.

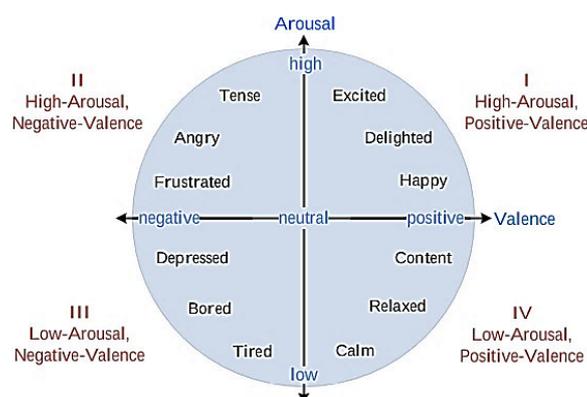


Figura 1.11 Modelo circunplejo de afecto [27]

Una escala para juzgar las expresiones faciales. Una descripción de la calidad afectiva atribuida a los entornos.

A una separación de 90°, dos estados afectivos no deben estar correlacionados entre sí. A medida que la separación se acerca a los 180 °, los estados afectivos deberían tener correlaciones negativas cada vez mayores entre sí.

Existe familiaridad con muchas emociones, así, la ira es útil para impedir que una situación continúe; el disgusto permite corregir algo que no contribuya con la satisfacción de las necesidades; el temor es un mecanismo de defensa en contra de las amenazas; la alegría es una recompensa cuando se alcanza una meta; la sorpresa es activada en presencia de un objeto novedoso, y la tristeza se manifiesta cuando no se está en condición de satisfacer una necesidad.

De acuerdo al modelo, la felicidad y la tristeza parecen dependientes de tal manera que cuando una aumenta, la otra disminuye y viceversa, mientras que no se puede establecer directamente una relación entre las emociones sorpresa y felicidad.

Predicción

La Ciencia de los Datos trata de extraer el conocimiento y las percepciones de los mismos en cualquier formato, ya sea estructurados o no estructurados, y aquí es donde es similar a la minería de datos. Además, utiliza técnicas de varios otros campos para lograr esto, Figura 1.12. Campos como las Matemáticas, la Estadística, la Informática (estructuras de datos y algoritmos y aquí es donde se correlaciona con el aprendizaje automático) y la Ciencia de la Información (aquí es donde la toma de decisiones se lleva a cabo a través de la comunicación de los resultados, la visualización y la presentación de informes) [28]

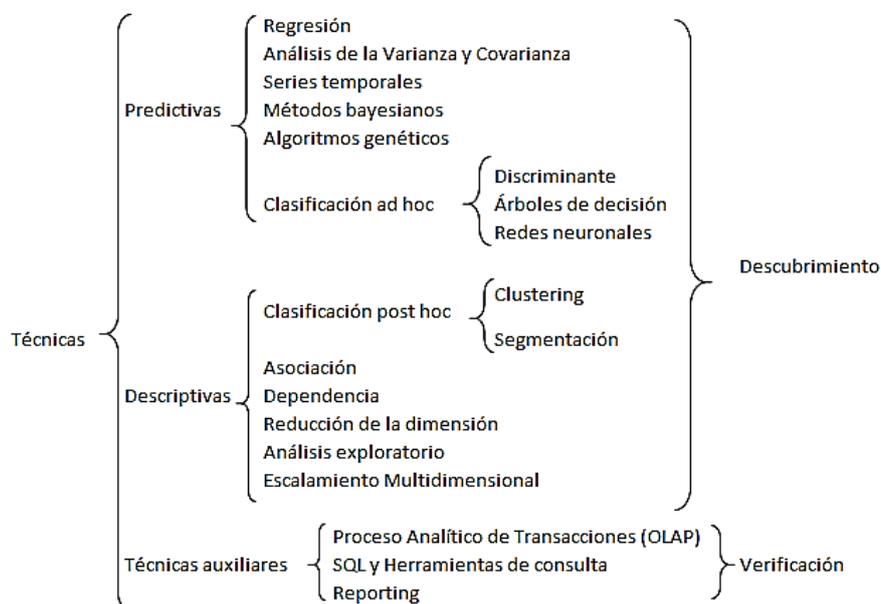


Figura 1.12 Algoritmos Estadísticos para Modelos de Data Science [28]

En la Figura 1.13, se presenta el proceso de la ciencia de los datos:

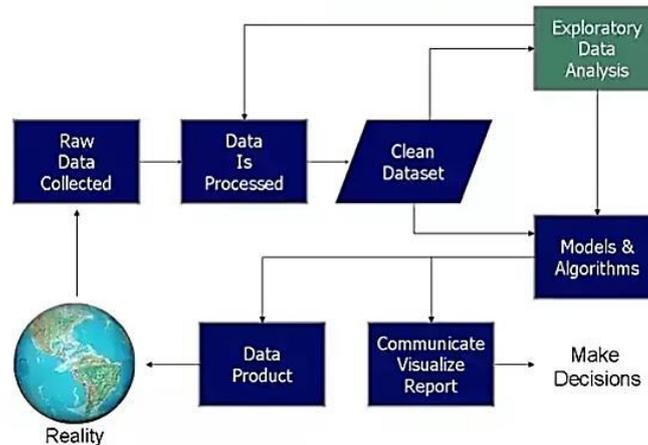


Figura 1.13 Proceso de Ciencia de Datos [29]

El análisis de datos, también conocido como analítica de datos, es un subdominio de la ciencia de los datos porque utiliza casi el mismo proceso con algunas exclusiones. Cuando se realiza el análisis de datos los procesos que se llevan a cabo son: inspeccionar, depurar, transformar y modelar los datos para encontrar información útil que sugiera conclusiones para apoyar la toma de decisiones.

La minería de datos es una técnica de análisis de datos, pero utilizada con un propósito diferente. No se utiliza con fines descriptivos en absoluto, sino que se emplea para el descubrimiento de conocimientos, para la elaboración de modelos y para fines de predicción.

Los sistemas de predicción son técnicas que construyen y estudian nuevas previsiones a través de una rama de la inteligencia artificial denominada aprendizaje automático (machine learning) [29]

El aprendizaje automático ofrece la capacidad de aprendizaje a una máquina para lograr realizar predicciones con precisión sobre sistemas nuevos que no se hayan dado en situaciones anteriores, incorpora los principios de la minería de datos, pero también puede crear correlaciones automáticas y aprender de ellas para aplicarlas a nuevos algoritmos. Hace uso de diferentes técnicas como modelos estadísticos, redes neuronales (ANN), máquinas de vectores soporte (SVM) o herramientas de agrupamiento (clusters) para poder predecir situaciones en base a la experiencia obtenida. [30] Por otra parte, los sistemas Data Mining se desarrollan bajo lenguajes de última generación basados en Inteligencia Artificial y utilizan métodos matemáticos tales como:

- Redes Neuronales.

- Sistemas Expertos.
- Programación Genética.
- Árboles de Decisión.

En una encuesta de minería de datos sobre el top de los métodos y algoritmos de ciencia de datos / aprendizaje automático usados desde 2017 a 2019 para una aplicación del mundo real se obtuvieron los resultados de la Figura 1.14.

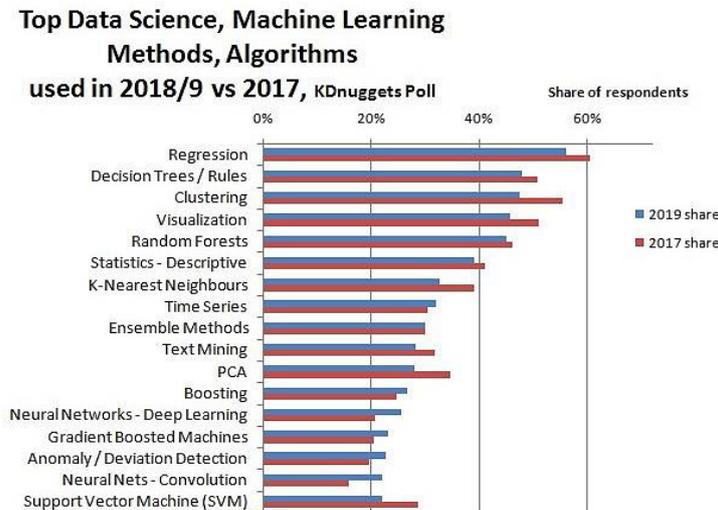


Figura 1.14 Top de métodos y algoritmos usados en Ciencia de Datos y Aprendizaje Automático en el 2018/9 vs 2017 [31]

Los parámetros que se deben considerar al momento de seleccionar un modelo para la predicción son los siguientes:

Error de Clasificación

Una clasificación binaria hace predicciones donde el resultado tiene dos valores posibles positivos y negativos. Además, la predicción para cada ejemplo puede ser correcta o incorrecta, lo que lleva a una matriz de confusión de 2x2 con 4 entradas:

- TP: el número de "verdaderos positivos", ejemplos positivos que se han identificado correctamente
- FP: el número de "falsos positivos", ejemplos negativos que se han identificado incorrectamente
- FN: el número de "falsos negativos", ejemplos positivos que se han identificado incorrectamente
- TN: el número de "verdaderos negativos", ejemplos negativos que se han identificado correctamente

$$\text{classification error} = (\text{Incorrect predictions}) / (\text{Number of Examples}) = (\text{FP} + \text{FN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN})$$

Errores Absoluto

El error absoluto de cada medida será la diferencia entre cada una de las medidas y ese valor tomado como exacto (la media aritmética).

Errores Relativo

El error relativo de cada medida será el error absoluto de la misma dividido por el valor tomado como exacto (la media aritmética).

En la Tabla 1.2, se presenta un ejemplo utilizando medidas de tiempo de un recorrido efectuadas por diferentes alumnos: 3,01 s; 3,11 s; 3,20 s; 3,15 s

Se calcula el valor que se considera exacto: $(3,01 + 3,11 + 3,20 + 3,15) / 4 = 3,12$

Los errores absoluto y relativo de cada medida:

Tabla 1.2 Ejemplos de Errores Absolutos y Relativos

MEDIDAS	ERRORES ABSOLUTOS	ERRORES RELATIVOS
3,01 s	$3,01 - 3,12 = - 0,11$ s	$-0,11 / 3,12 = - 0,036$ (- 3,6%)
3,11 s	$3,11 - 3,12 = - 0,01$ s	$-0,01 / 3,12 = - 0,003$ (- 0,3%)
3,20 s	$3,20 - 3,12 = + 0,08$ s	$+0,08 / 3,12 = + 0,026$ (+ 2,6%)
3,15 s	$3,15 - 3,12 = + 0,03$ s	$+0,03 / 3,12 = + 0,010$ (+ 1,0%)

Correlación

Una entrada de un dato se interpreta como un valor absoluto. Por lo que se aplica tanto a correlaciones positivas como negativas. Por ejemplo, si los sujetos más altos pesan más y los más bajitos pesan menos, entre peso y altura tendremos una correlación positiva: a mayor altura, mayor peso; mientras que, si los de más edad corren más despacio y los más jóvenes corren más deprisa, entre edad y velocidad tendremos una correlación negativa; a mayor edad, menor velocidad.

Precisión - Accuracy

La precisión se calcula tomando el porcentaje de predicciones correctas sobre el número total de ejemplos. Predicción correcta significa los ejemplos donde el valor del atributo de predicción es igual al valor del atributo de etiqueta.

$$\text{accuracy} = (\text{Correct predictions}) / (\text{Number of Examples}) = (TP + TN) / (TP + FP + FN + TN)$$

Si la precisión es menor o igual al 50% ese modelo no será útil ya que sería como lanzar una moneda al aire para tomar decisiones. Si alcanzamos un 90% o más podremos tener una buena confianza en los resultados que nos otorga el modelo.

Desviación Estándar

La desviación estándar es la medida de dispersión más común, que indica qué tan dispersos están los datos con respecto a la media. Mientras mayor sea la desviación estándar, mayor será la dispersión.

Ganancia

Calcula la ponderación de los modelos con respecto al atributo de clase utilizando la ganancia de información. Cuanto mayor sea el peso de un modelo, más relevante se considerará.

Tiempo Total

Tiempo que toma la construcción del modelo.

Tiempo de Entrenamiento

Tiempo que toma el entrenamiento del modelo

Tiempo de Anotación

Tiempo que toma la escritura y anotaciones del modelo

ROC Comparison-Matriz diagonal de costo beneficio

La curva de Características de Funcionamiento del Receptor (ROC) es un gráfico de la sensibilidad, o tasa de verdaderos positivos, frente a la tasa de falsos positivos (uno menos la especificidad o la tasa de verdaderos negativos) , para un sistema de clasificador binario, ya que su umbral de discriminación varía. El ROC también se puede representar de manera equivalente trazando la fracción de verdaderos positivos de los positivos (TPR = tasa de verdaderos positivos) frente a la fracción de falsos positivos de los negativos (FPR = tasa de falsos positivos).

Las curvas ROC se calculan ordenando primero los ejemplos clasificados por confianza. Posteriormente, todos los ejemplos se tienen en cuenta con una confianza decreciente para trazar la tasa de falsos positivos en el eje x y la tasa de verdaderos positivos en el eje y.

Para crear un gráfico ROC, Figura 1.15, y calcular el área bajo la curva (AUC), varía el umbral y se traza un punto (x, y) para cada valor de umbral:

- Eje y - tasa de verdaderos positivos = (verdaderas predicciones positivas) / (número de ejemplos positivos) = $TP / (TP + FN)$
- Eje x - tasa de falsos positivos = (predicciones de falsos positivos) / (número de ejemplos negativos) = $FP / (FP + TN)$

Se sugiere tomar los algoritmos que estén balanceados que presenten sus valores máximos en la diagonal de la matriz de falsos positivos, es decir el que se encuentra en la parte superior izquierda de la gráfica donde se encuentra el mejor valor de la diagonal de costo beneficio.

ROC Comparison

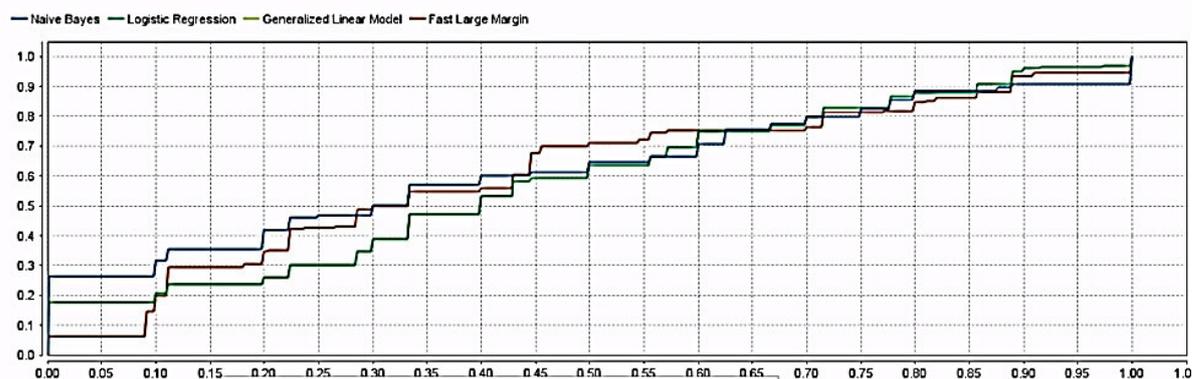


Figura 1.15 Curva de Comparación ROC

Random Forest

Es un método versátil de aprendizaje automático capaz de realizar tanto tareas de regresión como de clasificación. También lleva a cabo métodos de reducción dimensional, trata valores perdidos, valores atípicos y otros pasos esenciales de exploración de datos. Es un tipo de método de aprendizaje por conjuntos, donde un grupo de modelos débiles se combinan para formar un modelo poderoso.

En un problema de aprendizaje Supervisado donde se tiene la variable independiente y dependiente.

En Random Forest se ejecutan varios algoritmos de árbol de decisiones en lugar de uno solo. Para clasificar un nuevo objeto basado en atributos, cada árbol de decisión da una clasificación y finalmente la decisión con mayor número de votos es la predicción del algoritmo.

Las ventajas que tiene este algoritmo son las siguientes:

- Puede resolver ambos tipos de problemas, es decir, clasificación y regresión, y realiza una estimación decente en ambos casos.

- Unos de los beneficios que más llama la atención es el poder de manejar grandes cantidades de datos con mayor dimensionalidad. Puede manejar miles de variables de entrada e identificar las variables más significativas, por lo que se considera uno de los métodos de reducción de dimensionalidad. Además, el modelo muestra la importancia de la variable, que puede ser una característica muy útil.
- Tiene un método efectivo para estimar datos faltantes y mantiene la precisión cuando falta una gran proporción de los mismos.

A su vez la desventaja que tiene este algoritmo que puede parecer como una caja negra, ya que se tiene muy poco control sobre lo que hace el modelo. Es posible, en el mejor de los casos, probar diferentes parámetros y datos aleatorios.

Principios de Diseño

Se hace uso de ciertos principios del diseño gráfico para crear una composición efectiva y atractiva.

Numero de colores. En el diseño se debe limitar el número de colores diferentes utilizados a aquellos que el ojo humano puede procesar de un vistazo: alrededor de cinco colores dependiendo de la complejidad del diseño.

Codificación de información. Otra forma de codificar información mediante el color es incluir leyendas que especifiquen qué significa cada color. Este uso se apoya en el principio de la Gestalt que determina que aquellos elementos que comparten atributos gráficos (como el color) son percibidos intuitivamente como pertenecientes a un mismo grupo.

Combinación de colores. La teoría sobre percepción del color - en concreto la teoría sobre los procesos oponentes - afirma que la visión canaliza el color en tres canales diferentes: uno codifica la luminancia (blanco-negro) y otros dos el color (rojo-verde y azul-amarillo). Es una teoría que surge de la apariencia subjetiva de la luz, en concreto de la existencia de pares de colores que no pueden verse al mismo tiempo: rojo-verde, azul-amarillo y negro-blanco (acromático). No son visibles, por ejemplo, el verde rojizo o el azul amarillento.

Saturación. El uso de colores saturados es especialmente recomendable cuando se quiere destacar visualmente algún elemento concreto sobre el resto. No obstante, los colores saturados provocan fatiga visual, por lo que hay que usarlos con cautela.

Eficiencia. Cada segundo ahorrado al usuario mejora su experiencia de uso. Conforme más eficientemente pueda el usuario completar su tarea interactiva, mayor será su

satisfacción y valoración del producto. Para ello el sistema o producto debe responder a cada acción del usuario en el menor tiempo posible, idealmente de forma instantánea.

Minimizar el número de campos. Solicitar al usuario únicamente los datos imprescindibles, ya que un formulario con muchos campos no solo requiere más esfuerzo, sino que también desmotiva al usuario.

Valores por defecto. Si es posible establecer como valor por defecto aquel que la mayoría de usuarios seleccionarán. De esta forma se ahorrará a muchos de los usuarios tener que llevar a cabo esta selección.

Tamaño de los campos de texto. Al introducir información textual, si el campo es muy pequeño y no permite hacerlo de forma cómoda, exigirá más esfuerzo al usuario y provocará que cometa más errores.

Alineación de las etiquetas. Cuando las etiquetas se encuentran próximas a los campos a los que describen, resulta para el usuario menos costoso relacionarlos visualmente.

Organización de los campos. Los formularios con los campos posicionados en varias columnas son más complejos de completar que los que se ordenan en una única columna.

Ley de Fitts. Conforme menor es el área clickable de un elemento del formulario, mayor es el esfuerzo y tiempo requerido para accionarlo.

Ayuda contextual. Cuando el usuario pueda dudar acerca del significado de un campo o sobre cómo introducir los datos, se debe ofrecer ayuda contextual.

Balance. El balance como concepto aplicado al diseño busca — al igual que en la física — igualar la tensión entre elementos. Un diseño balanceado es percibido como más estético, transmite armonía. A alcanzar el balance en diseños asimétricos. Para ello hay que jugar con el peso visual de los elementos hasta lograr un estado de equilibrio entre ambos lados del eje central. El peso visual de un elemento viene determinado por su tamaño, color, intensidad, proximidad a otros elementos o margen.

Espacio en blanco. El espacio en blanco o espacio negativo (no necesariamente el color debe ser blanco) se refiere al espacio y márgenes entre elementos de la interfaz. Este espacio en blanco facilita al usuario diferenciar visualmente elementos o grupos de elementos, permite una exploración visual más calmada, pero también, incrementa la simplicidad, elegancia y apariencia estética del diseño. [32]

Creación de Dashboards

Los Dashboards se pueden entender como una colección de reportes, consultas y análisis interactivos en el que se reflejan las principales métricas o KPI que intervienen en la consecución de los objetivos de una estrategia.

Proceso de consulta y análisis

El mismo, se lleva a cabo a través de seis pasos sucesivos:

1. Los usuarios seleccionan o establecen que datos desean obtener del DW, mediante las interfaces de la herramienta que utilice.
2. La herramienta recibe el pedido de los usuarios, construye la consulta (utilizando la metadata) y la envía al Query Manager.
3. El Query Manager ejecuta la consulta sobre la estructura de datos con la que se esté trabajando (cubo multidimensional, Business Model, etc.).
4. El Query Manager obtiene los resultados de la consulta.
5. El Query Manager envía los datos a la herramienta de consulta y análisis.
6. La herramienta presenta a los usuarios la información requerida.

Estos pasos se aprecian en la siguiente Figura 1.10.

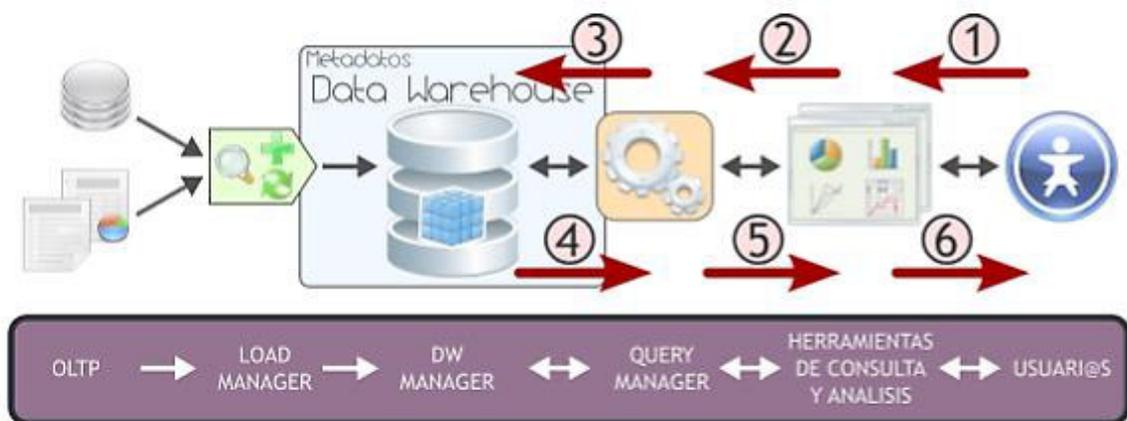


Figura 1.16 Proceso de Consulta y Análisis [1]

Existen diversas maneras de diseñar un Dashboard, cada una de las cuales tiene sus objetivos particulares, pero a modo de síntesis se expondrán algunas características generales que suelen poseer:

- Presentan la información altamente resumida.
- Se componen de consultas, reportes, análisis interactivos, gráficos (de torta, barras, etc.), semáforos, indicadores causa-efecto, etc.
- Permiten evaluar la situación de la empresa con un solo golpe de vista.
- Poseen un formato de diseño visual muy llamativo.

Visualizador (data for)

Un cuadro de mando o dashboard es una herramienta que proporciona información a la empresa sobre su productividad y rendimiento, utilizando para ello una serie de métricas alineadas con los objetivos marcados en la estrategia empresarial (KPI).

Con un cuadro de mando, los gerentes y responsables de la empresa pueden monitorizar todas las operaciones de los distintos departamentos de forma gráfica y sencilla que facilita la interpretación y así acelerar el proceso de toma de decisiones.

Los usuarios acceden a los cubos multidimensionales, Business Models (u otro tipo de estructura de datos) del DW utilizando diversas herramientas de consulta, exploración, análisis, reportes, etc.

Las herramientas de consulta y análisis son sistemas que permiten a los usuarios realizar la exploración de datos del DW, son el nexo entre el depósito de datos y los usuarios.

Estas herramientas emplean conexiones a bases de datos (JNDI, JDBC, ODBC), para obtener la información deseada.

A través de una interfaz gráfica y una serie de pasos, los usuarios generan consultas que son enviadas desde la herramienta de consulta y análisis al Query Manager, este a su vez realiza la extracción de información al DW Manager y devuelve los resultados obtenidos a la herramienta que se los solicitó. Luego, estos resultados son expuestos a los usuarios.

2. METODOLOGÍA

En este capítulo se describen los pasos realizados al seguir la metodología Hefesto.

En el Análisis de Requerimientos, se realiza un diagnóstico de la información disponible de los talleres de RE, se elaboran preguntas que permitan explicar los objetivos de la Institución Educativa Verbo, se identifican en conjunto con los usuarios los indicadores y perspectivas que son tomadas en cuenta para la construcción del DW y se crea un modelo conceptual.

En el Análisis del OLTP, se utilizan las fuentes OLTP para los cálculos de los indicadores y se establecen las respectivas correspondencias entre el modelo conceptual creado en el paso anterior.

En la sección del Modelo Lógico de DW, se establece su estructura, teniendo como base el modelo conceptual que ya ha sido creado y se diseñan las tablas de dimensiones y de hechos con sus relaciones.

Una vez construido el modelo lógico, en la Integración de Datos se procede a poblarlo con datos, utilizando técnicas de limpieza y calidad de datos, procesos ETL, etc., mediante el uso de las herramientas: Spoon de Pentaho, RM y Turbo Prep dependiendo de la necesidad.

Luego de la generación de los cubos que se realiza utilizando la herramienta de la suite de Pentaho denominada Saiku, para la visualización se crea los dashboards para el director y profesores para que puedan monitorizar la información obtenida de las emociones del taller de RE de forma gráfica y sencilla.

2.1 Análisis de requerimientos

Detectar, definir y documentar correctamente los requisitos, es el primer paso, para esto es necesario identificar en conjunto con los usuarios los indicadores y perspectivas que serán tomadas en cuenta para la construcción del DW. Finalmente se confeccionará un modelo conceptual en donde se podrá visualizar el resultado obtenido en este primer paso.

Es muy importante tener en cuenta que la metodología Hefesto se puede utilizar para construir un data warehouse o un data mart a la vez.

Se identificará los requerimientos a través de preguntas que permitan explicar los objetivos de la institución.

2.1.1 Diagnóstico de la Información

Talleres RE

La información de la que se dispone sobre los talleres impartidos de RE es la siguiente:

- Fotografías por grupos de niños. Ver Anexo III.

- Actas de los talleres expuestos donde se indica fecha y nombre de los estudiantes. Ver Anexo I.
- Archivos JSON y CSV o XLSX del análisis de reconocimiento facial realizado. Ver Anexos IV y V respectivamente. Cabe mencionar que estos archivos presentan información faltante como por ejemplo faceld, nombre de la imagen, fase, etc.

Indicadores

Para realizar la medición de indicadores se requiere lo siguiente:

- Relacionar las fotografías con el nombre de los niños, fases, taller y grupo a los que pertenecen.
- Verificar que se encuentren el género, edad y accesorios como lentes correctos del estudiante.
- Verificar que se encuentren registradas las fechas de las fotografías tomadas.
- Realizar el cálculo de la experiencia positiva, negativa o neutral que presentó un estudiante de acuerdo a las emociones obtenidas de las fotografías.
- Realizar predicciones de acuerdo a la experiencia de usuario obtenida.

El análisis de los requerimientos de los diferentes usuarios, es el punto de partida de la metodología Hefesto, ya que ellos son los que deben guiar la investigación hacia un desarrollo que refleje claramente lo que se espera del depósito de datos, en relación a sus funciones y cualidades.

El objetivo principal de esta fase, es la de obtener e identificar las necesidades de información clave de alto nivel, que es esencial para llevar a cabo las metas y estrategias de la empresa, y que facilitará una eficaz y eficiente toma de decisiones.

Elaboración de encuestas para padres, autoridades y maestros delegados

Se elabora mediante Google Forms el cuestionario de preguntas para socializarlo con las autoridades del colegio. Ver Anexo VI.

Las preguntas elaboradas se encuentran en el siguiente link:
<https://docs.google.com/forms/d/e/1FAIpQLScxeZKcsJfswOZkk3N5RIP8SYcSn7x85QmL3jcvpLKgBnvALQ/viewform>

En la Figura 2.1 se encuentra la socialización de las preguntas elaboradas con las autoridades.



MARIO CANDO SALGUERO

para mí

Lun., 5 oct. 11:53 (hace 10 días) ★ ↵

Buenos días

Mario Cando le saluda, en vista a la conversación mantenida con usted y recibida las orientaciones sobre el la importancia del feedback para conseguir los resultados oportunos y óptimos con la implementación de los talleres, y después de haber revisado las preguntas estoy de acuerdo

Atentamente

Mgs. Mario Cando

DIRECTOR

UNIDAD EDUCATIVA "CRISTIANO VERBO"

2467564 / 2468237



Figura 2.1 Correo de acuerdo de las preguntas planteadas

Con las encuestas realizadas se comienza recolectando las necesidades de información de los usuarios y se obtienen las preguntas claves del negocio.

Identificar preguntas

Las preguntas identificadas en esta fase se compararon con los datos disponibles de los Talleres de RE y luego de validar su factibilidad se obtuvieron las enunciadas en la Tabla 2.1.

Tabla 2.1 Factibilidad de responder vs datos disponibles

IDENTIFICADOR	PREGUNTAS	USUARIO Interesado R-Rector P-Profesor Pd-Padre	FACTIBILIDAD
Q1	¿Cuál fue la emoción de los estudiantes en cada una de las fases del taller?	P	si
Q2	¿Cuál es tu argumentación en lo que has realizado?	P	no
Q3	¿Qué dificultad se presentó con el uso del lego? Opcional: Sentimientos negativos	P	no
Q4	¿Qué sentimiento presentaron con el uso del lego? (Promedio por edad y género)	P	si
Q5	¿Qué actitud tenía el niño cuando comenzó a crear su ingenio con los legos? (imaginativos, interpretativos, unos saben que hacer otros no, innovan, inventan)	P	no

IDENTIFICADOR	PREGUNTAS	USUARIO Interesado R-Rector P-Profesor Pd-Padre	FACTIBILIDAD
Q6	¿Cómo podremos afianzar los conocimientos adquiridos de motricidad fina con la aplicación del taller de robótica	P	no
Q7	¿El taller de robótica sirvió para captar la atención del estudiante? Opcional: Total de felicidad por taller.	P	si
Q8	¿El estudiante consiguió expresar ideas a partir de la reflexión constructiva sobre sí mismo y sus experiencias, mediante creaciones artísticas y prácticas corporales propias de su entorno cultural?	P	no
Q9	¿Cuántos niños crearon el producto más confiable y conciso?	P	no
Q10	¿Qué características tienen los niños según el grupo en el que se desarrollaron? Opcional: estados de ánimo, ¿quién demostró entusiasmo y quién no?	P	si
Q11	¿Se pudo apreciar cambios de carácter entre fases?	P	si
Q12	¿En qué momento y en quienes se evidenciaron actitudes positivas o negativas por fase, edad, género?	P, Pd	si
Q13	¿Existió falta de apego en la realización de actividades motrices del taller? Opcional: Existió disgusto	P, Pd	si
Q14	¿A quiénes les costó conseguir resultados? Opcional: Emociones negativas	P	si

IDENTIFICADOR	PREGUNTAS	USUARIO Interesado R-Rector P-Profesor Pd-Padre	FACTIBILIDAD
Q15	¿En quienes se pudo evidenciar la seriación de colores? Opcional: Creación secuencias lógicas matemáticas.	P	no
Q16	¿Quiénes hicieron secuencias de patrones según la forma, tamaño?	P	no
Q17	¿Desarrolló el pensamiento computacional - lógica de programación?	P	no
Q18	¿Se inspiró a los estudiantes en búsqueda, diseño, desarrollo, construcción y prueba de soluciones? Opcional: Estructurar soluciones	P	no
Q19	¿Desarrollaron la habilidad de planificación de su trabajo?	P	no
Q20	¿Desarrollaron la habilidad de toma de decisiones?	P	no
Q21	¿Desarrollaron el uso de recursos didácticos como mapas mentales?	P	no
Q22	¿Desarrollaron abstracción?	P	no
Q23	¿Desarrollaron habilidad para esquematizar ideas?	P	no
Q24	¿Quiénes son los estudiantes más motivados? Opcional: Felicidad - Dominio por fase del taller	P	si
Q25	¿Quiénes son los mejores estudiantes? Opcional: Felicidad -Dominio	P	si
Q26	¿Qué destreza es mayor en un estudiante que en otro?	P	no
Q27	¿Cuál es el taller que más agrada, impacta y llama la atención a los niños,	P	si

IDENTIFICADOR	PREGUNTAS	USUARIO Interesado R-Rector P-Profesor Pd-Padre	FACTIBILIDAD
	a quienes, en qué fase, por género, edad?		
Q28	¿Qué habilidades se ven más desarrolladas y en qué taller?	P	no
Q29	¿Los estudiantes si están aprendiendo?	R, P	si
Q30	¿Sirve este taller para enseñar otras materias?	R, P	si
Q31	¿El niño documenta y usa su razonamiento en relación con la pregunta o el problema?	P	no
Q32	¿El niño resuelve el problema o pregunta construyendo, programando su modelo y haciendo cambios cuando era necesario?	P	no
Q33	¿Documentó ideas y pruebas importantes a lo largo de su proyecto y dio lo mejor cuando lo presentó a otros? Opcional: En la fase de compartir se sintió a gusto.	P	si
Q34	¿Cuándo y quienes mostraron timidez?	P	si
Q35	¿Qué niños tienen un determinado nivel de emocional?	P	si
Q36	¿Hice mi mejor trabajo para resolver el problema o pregunta construyendo y programando mi modelo y haciendo cambios cuando era necesario? A	P	no
Q37	¿Documenté ideas y pruebas importantes a lo largo de mi proyecto y di lo mejor de mí cuando presenté a otros?	P	no
Q38	¿Una cosa que el estudiante hizo realmente bien? Opcional: En caso de	P	si

IDENTIFICADOR	PREGUNTAS	USUARIO Interesado R-Rector P-Profesor Pd-Padre	FACTIBILIDAD
	aplicar nuevamente el taller, ¿Cuántos estudiantes felices por fase?		
Q39	¿Algún aspecto que se puede mejorar para la próxima vez? Opcional: Número de estudiantes con emociones negativas por fase, género, edad	P	si

Las preguntas recolectadas inicialmente se las puede observar en el Anexo VII

En la Tabla 2.2, se establece la prioridad de las respuestas:

Tabla 2.2 Prioridad de Respuestas

Tipo de Indicador	
estáticos	dinámicos
Q1, Q4, Q7, Q10, Q11 Q12, Q13, Q14, Q24, Q25 Q27, Q29, Q33, Q34, Q35	Q30, Q38, Q39

Como se puede apreciar, las necesidades de información expuestas están acorde a los objetivos y estrategias de la institución educativa, ya que es precisamente esta información la que proveerá un ámbito para la toma de decisiones [1], puesto que permitirá analizar el comportamiento de los usuarios en un taller de RE. Comprendiendo las necesidades de los diferentes usuarios, se pretende satisfacer ampliamente sus inquietudes, para así lograr obtener una ventaja competitiva y maximizar su servicio a la comunidad por medio del uso de BI. Por este motivo se requiere conocer variables a futuro que se abordan a continuación.

2.1.2 Análisis de la experiencia de usuario

El cálculo de experiencia de usuario en el presente proyecto se basa en la teoría del modelo circular de afecto de Schlosberg, realizando un promedio de las emociones negativas y comparando su resultado con la emoción positiva (felicidad) y neutral.

Problemas detectados

En vista de que no se tiene un identificador secuencial en las fotografías para realizar el análisis de la sorpresa, a la sorpresa no se le da un tratamiento.

No se puede hacer suma de emociones para considerar la experiencia de usuario en lo referente a la emoción sorpresa, no se tiene un sustento científico justificado para hacerlo, al sumar emociones se obtiene otra emoción y se presentan problemas de ordenamiento.

El tratamiento futuro de la sorpresa deriva en una mejor técnica que se base en un análisis de cada iteración y al momento no se dispone de un orden cronológico preciso de las fotos para realizar la comparación ordenada que se requiere en cada iteración para verificar la variación de sorpresa entre cada iteración por estudiante, lo cual implica un mejor ordenamiento en la toma de los datos y por este motivo la sorpresa se deja como perspectiva en esta tesis ya que amerita un análisis más profundo con otras técnicas. Se plantea como solución realizar las iteraciones ordenadas por usuario y mejorar el ordenamiento en la toma de los datos.

Justificación del modelo para la predicción

En la Tabla 2.3, se presenta el cuadro comparativo de los modelos, el mismo que fue obtenido desde RM.

Tabla 2.3 Comparativa de Modelos en RM - Automodel

Modelo	Error de Clasificación	Precisión	Desviación Estándar	Ganancia	Tiempo Total	Tiempo de Entrenamiento (1,000 Rows)	Puntuación de Tiempo (1,000 Rows)
Naive Bayes	3,08%	96,90%	1,40%	138	3 min 13 s	435 ms	7 s
Random Forest	0,55%	99,50%	0,50%	154	3 min 10 s	1 s	3 s
Logistic Regression	3,26%	96,70%	2,40%	136	7 min 35 s	3 s	6 s
Deep Learning	0,91%	99,10%	0,60%	148	3 min 46 s	6 s	2 s
Decision Tree	0,91%	99,10%	0,60%	150	2 min 46 s	293 ms	2 s
Generalized Linear Model	0,73%	99,30%	0,80%	152	19 min 52 s	1 s	12 s
Gradient Boosted Trees	1,27%	98,70%	0,50%	152	11 min 27 s	4 s	6 s
Support Vector Machine	14,67%	85,30%	1,60%	0	1 hours 1 min 54 s	36 s	20 s

Para comprobar si el modelo generado funciona se divide el conjunto de datos depurados 1983 en dos, de entrenamiento 1190 equivalente al 60% y de testing 793 equivalente al 40%.

Los de prueba o validación son los que se reservan para verificar que a partir de los datos de entrenamiento el modelo funciona. Es decir, si las respuestas predichas por el modelo para un caso totalmente nuevo son acertadas o no.

El modelo Random Forest (RF) es el óptimo a usar para la predicción ya que muestra un menor error de clasificación de 0,55% y una exactitud del 99,50%. En la Tabla 2.4, se describen los pesos de atributos y la relevancia de variables del modelo RF mayor al 50%, obteniendo así un top 5:

Tabla 2.4 Atributos de Relevancia en Modelo Random Forest

ATRIBUTOS	PESO (%)	DEFINICIÓN DE LA VARIABLE	RELEVANCIA
neutral	9	Intensidad de la emoción neutral.	Ausencia de estímulo con una emoción (condicionamiento), surge aprendizaje al verse acompañado de una emoción de felicidad. [33]
Nombre de la fase	9	Fase del Taller de RE	La emoción que presenta el estudiante depende de las actividades que realiza en cada fase del taller.
felicidad	8	Intensidad de la emoción felicidad.	En situaciones de bienestar emocional, aumenta la eficiencia del proceso cognitivo. [33]
tristeza	6	Intensidad de la emoción tristeza.	Cuando la situación evaluada despierta emociones negativas, la activación del córtex prefrontal se produce eminentemente en su parte derecha. [34]
desprecio	5	Intensidad de la emoción desprecio.	Los estados emocionales negativos, surgidos a partir de situaciones de miedo o estrés, propician una activación de la amígdala que deriva en una liberación de adrenalina, noradrenalina y glucocorticoides (cortisol). Así mismo, la presencia de estas hormonas provoca, a su vez, una serie de eventos corporales como el aumento de la tasa de pulsaciones y de transpiración. [33]
lentes	5	Si la cara determinada tiene gafas. Los valores posibles son NoGlasses, ReadingGlasses, Sunglasses y Swimming Goggles.	Los defectos visuales no tratados en los niños, implica un bajo rendimiento escolar, dificultades en el aprendizaje y retraso en su desarrollo psicosocial. [34]

Simular predicción

Por medio del simulador de RM se busca descartar experiencias negativas que derivan en timidez e introversión. Así, por ejemplo, una persona tímida es aquella que no teme a la interacción social, aunque manifiesta preferencia por la soledad, tiene escasa confianza en sí mismo, pobres habilidades sociales y sentimientos de vergüenza que le llevan a evitar el contacto interpersonal, manifiesta temores ante la posibilidad de no alcanzar sus expectativas sociales. [33] Por otro lado, la introversión no se limita únicamente a estas situaciones, sino que tiene un carácter más estable y más independiente de la situación.

Se busca optimizar las condiciones al conseguir experiencias positivas, las cuales son fuentes de motivación. De los resultados obtenidos se puede mencionar que el estado emocional neutral es un estado que se puede aprovechar para meditar sobre mejores decisiones y no derivar en esos extremos emocionales. [34] La emoción neutral facilita los mecanismos atencionales para procesar el significado de la situación y da paso a la reacción emocional apropiada para dicha situación, acompañada de emociones positivas, prepara a la persona para captar información en general y provoca un rápido aumento de la actividad cognitiva.

2.1.3 Diseñar el sistema de indicadores

Una vez que se han establecido las preguntas de negocio, se debe proceder a su descomposición para descubrir los indicadores que se utilizarán y las perspectivas de análisis que intervendrán. Se debe tener en cuenta que los indicadores en general son valores numéricos y representan lo que se desea analizar concretamente así, por ejemplo: los profesores quieren conocer si a un niño le gusta una determinada materia y de esta forma predecir que a un estudiante le pueda ir bien o mal en ciertas asignaturas, inquietudes que resultan relevantes al momento de aplicar una nueva tecnología en la enseñanza como en este caso de estudio lo es la RE.

Las perspectivas se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de responder a las preguntas planteadas [1]. En la Tabla 2.5, se puede visualizar los indicadores y perspectivas identificados.

Tabla 2.5 Indicadores y perspectivas

IDENTIFICADOR	INDICADOR	PERSPECTIVA	USUARIO INTERESADO R-RECTOR P-PROFESOR PD-PADRE
1	<u>Emoción</u> que presenta un estudiante	en cada una de las <u>fases</u> del <u>taller</u> , por <u>edad</u> , <u>fecha</u> y <u>género</u>	P, Pd
2	<u>Total de emociones</u> que se presentaron con el uso del lego	en cada una de las <u>fases</u> del <u>taller</u> , por <u>edad</u> , <u>estudiante</u> y <u>género</u> .	P
3	<u>Promedio de emociones</u> que se presentaron con el uso del lego	en cada una de las <u>fases</u> del <u>taller</u> , <u>edad</u> y <u>género</u>	P
4	<u>Total Experiencia Positiva</u> o <u>Negativa</u>	en cada una de las <u>fases</u> del <u>taller</u> , por <u>edad</u> , <u>fecha</u> y <u>género</u>	P
5	<u>Promedio de Experiencia Positiva</u> o <u>Negativa</u>	en cada una de las <u>fases</u> del <u>taller</u> , por <u>edad</u> y <u>género</u>	R
6	<u>Total de Experiencia Positiva</u> o <u>Negativa</u> de un estudiante	en cada una de las <u>fases</u> del <u>taller</u> , por <u>edad</u> por <u>género</u>	P, Pd
7	<u>Predicción de Experiencia Positiva</u> o <u>Negativa</u> de un estudiante	en cada una de las <u>fases</u> del <u>taller</u> , <u>edad</u> y <u>género</u>	R,P, Pd
8	<u>Total de Estudiantes</u>	por <u>taller</u> , por <u>edad</u> , <u>grupo</u> y <u>género</u>	P
9	<u>Emociones que presenta</u>	<u>la imagen</u> de un estudiante	P

Las preguntas seleccionadas hacen explícitos los objetivos estratégicos planteados por la organización, a la vez que enfocan las necesidades de información en los procesos principales de la institución educativa.

En síntesis, los indicadores son:

Estáticos:

- Total de estudiantes
- Total por emoción (enfado, desprecio, asco, temor, felicidad, neutral, tristeza, sorpresa)
- Promedio de emociones (enfado, desprecio, asco, temor, felicidad, neutral, tristeza, sorpresa)
- Total experiencia (positiva, negativa)
- Promedio experiencia (positiva, negativa)

Dinámico:

- Predicción de la experiencia de usuario(positiva, negativa)

Y las perspectivas de análisis son:

- Fases
- Emociones
- Experiencia de Usuario
- Estudiante (nombre, foto, edad)
- Taller
- Imagen

2.1.4 Modelo conceptual

A través de este modelo, se podrá observar con claridad cuáles son los alcances del proyecto, para luego poder trabajar sobre ellos, además el poseer un alto nivel de definición de los datos, permite su presentación ante los usuarios y explicarlos con facilidad.

Em la Figura 2.2 se presenta la representación gráfica del modelo conceptual.

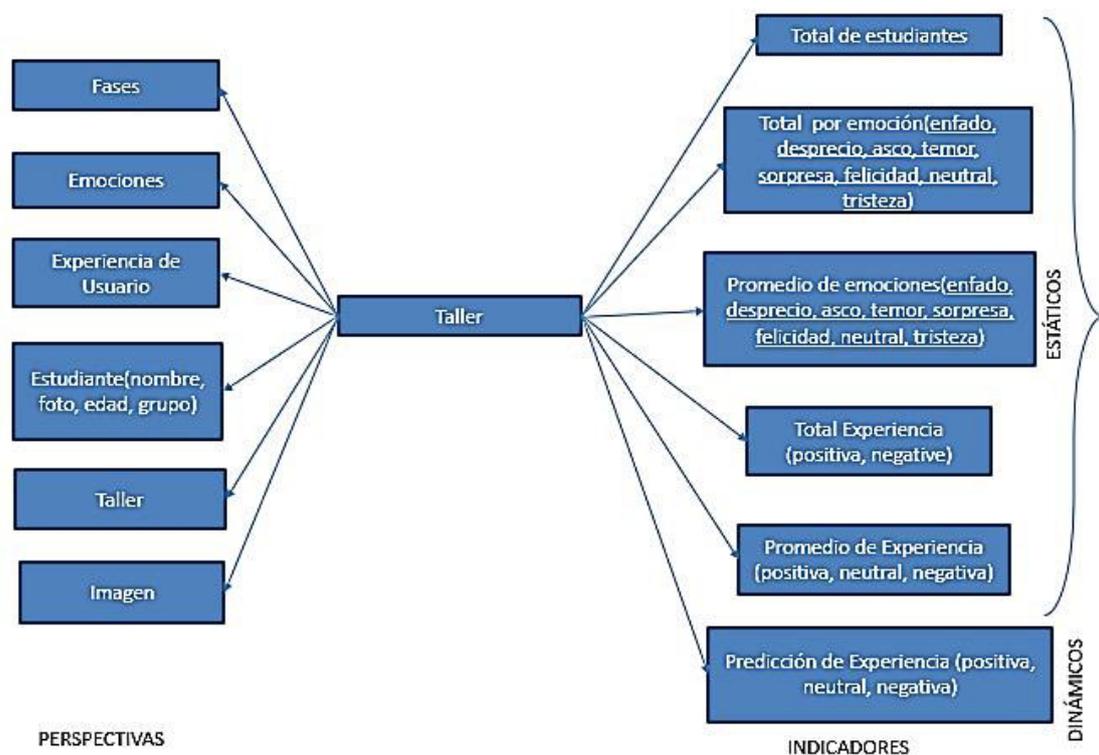


Figura 2.2 Modelo Conceptual

Como puede apreciarse en la figura anterior, el modelo conceptual permite de un solo vistazo y sin poseer demasiados conocimientos previos, comprender cuáles serán los resultados que se obtendrán, cuáles serán las variables que se utilizarán para analizarlos y cuál es la relación que existe entre ellos. Cabe mencionar entonces que

es necesario el entendimiento, validez fiabilidad de los datos, que se puede resumir en la figura siguiente:

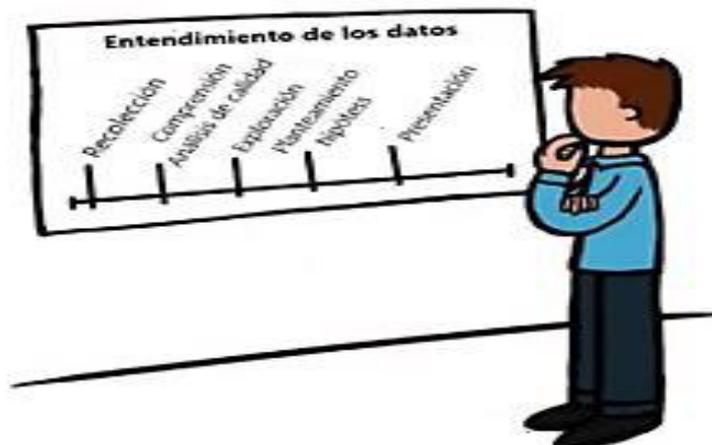


Figura 2.3 Entendimiento de los datos

2.2 Análisis de OLTP

Se analizarán las fuentes OLTP para determinar cómo serán calculados los indicadores y para establecer las respectivas correspondencias entre el modelo conceptual creado en la sección 2.1.2.3 y las fuentes de datos. Luego, se van a definir los campos que se incluirán en cada perspectiva. Finalmente, se ampliará el modelo conceptual con la información obtenida.

2.2.1 Conformación de indicadores

Los indicadores se calcularán de la siguiente manera:

- Estáticos:
 - Total de Estudiantes
 - Hechos1: IdPersona
 - Función: Conteo de estudiantes (COUNT)
 - Aclaración: el indicador "Total de Estudiantes" representa el conteo de los estudiantes que participaron en los talleres de RE por edad, fase, taller, fecha y grupo.
 - Total de Emociones
 - Hechos1: enfado, disgusto, desprecio, temor, felicidad, neutral, tristeza, sorpresa
 - Función: Sumarización de cada una de las emociones (SUM)
 - Aclaración: el indicador "Total de Emociones" representa la sumatoria de las emociones que se presentan en los talleres de RE por edad, fase, taller, fecha y grupo.

- Promedio de Emociones
 - Hechos1: enfado, disgusto, desprecio, asco, temor, felicidad, neutral, tristeza, sorpresa
 - Función: Promedio de cada una de las emociones (AVG)
Aclaración: el indicador “Promedio de Emociones” representa el promedio de cada una de las emociones que se presentan en los talleres de RE por edad, fase, taller, fecha y grupo.

- Total de Experiencia (positiva, neutral, negativa)
 - Hechos2: enfado, disgusto, desprecio, temor, felicidad, tristeza, sorpresa
 - Función: Sumarización de las emociones negativas: enfado, disgusto, desprecio, temor (SUM) y Cálculo del mayor valor (MAYOR) en emoción positiva, negativa y neutral
Aclaración: el indicador “Total Experiencia” representa comparativa del total de las emociones negativas (enfado, disgusto, desprecio, temor), positiva (felicidad), neutral y que se presentan en los talleres de RE por edad, género, fase, taller, fecha y grupo.

- Promedio de Experiencia (positiva, neutral, negativa)
 - Hechos2: enfado, disgusto, desprecio, asco, temor, felicidad, tristeza, sorpresa
 - Función: Promedio de experiencia positiva, neutral y negativa del usuario (AVG)
Aclaración: el indicador “Promedio de Experiencia” representa el promedio la experiencia positiva(felicidad), neutral y negativa(enfado, disgusto, desprecio, temor) del usuario en los talleres de RE por edad, género, fase, taller, fecha y grupo.

- Dinámicos:
 - Predicción de Experiencia de Usuario
 - Hechos2: prediccionExperienciaUsuario
 - Función: Predicción - Algoritmo Random Forest y gradient Forest
Aclaración: el indicador “Predicción de Experiencia de Usuario” representa el resultado de la aplicación del algoritmo Random Forest para la predicción de la experiencia de usuario en los talleres de RE por edad, género, fase, taller y grupo.

2.2.2 Establecer correspondencias

Con los datos disponibles se creó un modelo entidad relación, Figura 2.4. Se examinó el OLTP disponible que contiene la información requerida, sus características para identificar las correspondencias entre el modelo conceptual y las fuentes de datos.

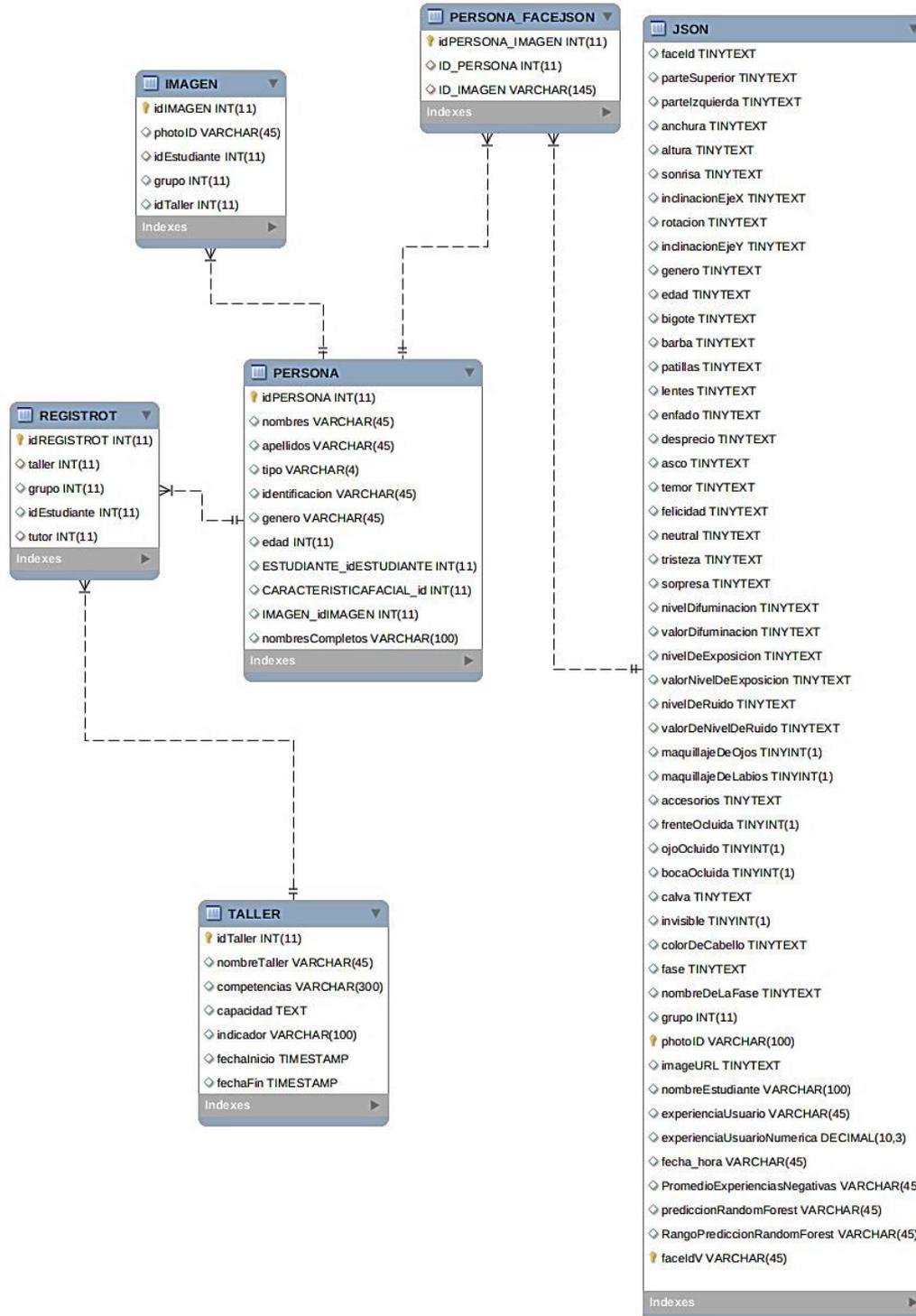


Figura 2.4 TalleresRE, Diagrama de Entidad Relación

En la Figura 2.5, se expone la correspondencia entre los dos modelos.

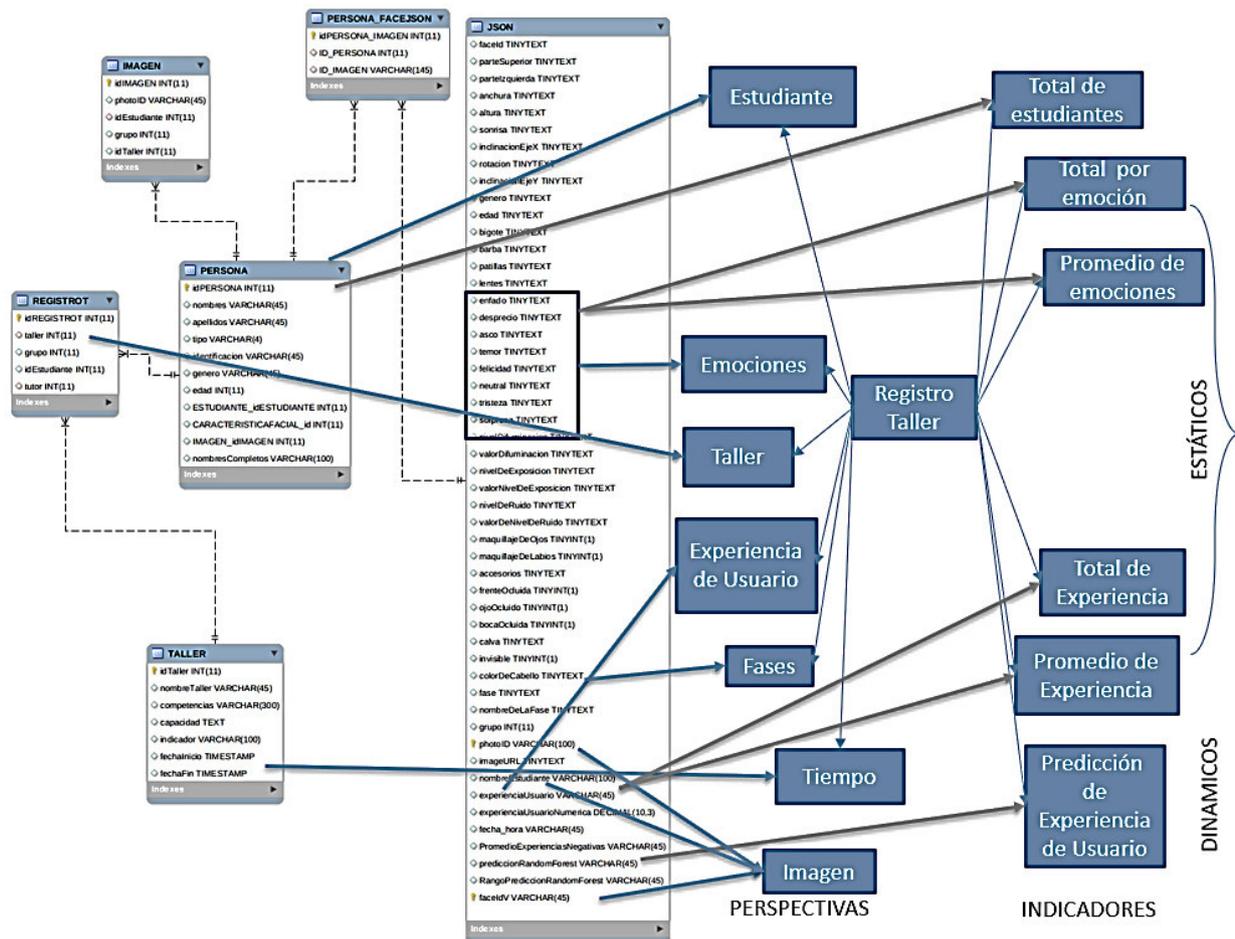


Figura 2.5 Talleres RE, correspondencia

Las relaciones identificadas fueron las siguientes:

- La tabla “Persona” con la perspectiva “Estudiante”.
- Los campos “enfado”, “desprecio”, “asco”, “temor”, “felicidad”, “neutral”, “tristeza” y “sorpresa” de la tabla “Json” con la perspectiva “Emociones”.
- La tabla “Taller” con la perspectiva “RegistroT”.
- El campo “fase” de la tabla “Json” con la perspectiva “Fases”.
- Los campos “fechaInicio” y “fechaFin” con la perspectiva “Tiempo”.
- El conteo del campo “idPersona” de la tabla “Persona” con el indicador “Total de estudiantes”.
- La sumatoria de los campos “enfado”, “desprecio”, “asco”, “temor”, “felicidad”, “neutral”, “tristeza” y “sorpresa” de la tabla “Json” con el indicador “Total por emoción”.
- El promedio de los campos “enfado”, “desprecio”, “asco”, “temor”, “felicidad”, “neutral”, “tristeza” y “sorpresa” de la tabla “Json” con el indicador “Promedio de emociones”.

- Los campos "photoID", "nombreEstudiante" y "faceIdV" de la tabla "Json" con la perspectiva "Imagen".
- El campo "experienciaUsuario" de la tabla "Json" con la perspectiva "Experiencia de Usuario".
- La suma total del campo "experienciaUsuario" de la tabla "Json" con el indicador "Total de Experiencia".
- El promedio del campo "experienciaUsuario" de la tabla "Json" con el indicador "Promedio de Experiencia".
- El campo "prediccionRandomForest" de la tabla "Json" con el indicador "Predicción de Experiencia"

2.2.3 Nivel de granularidad

De acuerdo a las correspondencias establecidas, se analizaron los campos residentes en cada tabla de las que se hizo referencia. Se examinó la base de datos para intuir los significados de cada campo y luego se consultó con el profesor y el director que participaron de los talleres los aspectos de los cuales no se comprendía su sentido.

Con respecto a la perspectiva "Estudiante", los datos disponibles son los siguientes:

- idPersona: es la clave primaria de la tabla "Persona", y representa unívocamente a un estudiante en particular.
- nombres: primero y segundo nombre del estudiante.
- apellidos: primero y segundo apellido del estudiante.
- tipo: identificación del grupo al que pertenece un estudiante.
- genero: género del estudiante, con valores masculino y femenino.
- edad: edad del estudiante, para el caso del taller valor comprendido entre 7 a 10 años.
- nombresCompletos: se refiere a los nombres y apellidos del estudiante.
- grupo: número de grupo de niños en la que se desarrolló un determinado estudiante.
- idEstudiante: identificador del estudiante.
- tutor: identificador del tutor que estuvo realizando seguimiento del taller.

Con respecto a la perspectiva "Emociones", los datos disponibles son los siguientes:

- faceIdV: identificador único del rostro detectado.
- parteSuperior: área rectangular para la ubicación de la cara en la imagen, parte superior.

- partelzquierda: área rectangular para la ubicación de la cara en la imagen, parte izquierda.
- anchura: área rectangular para la ubicación de la cara en la imagen, ancho.
- altura: área rectangular para la ubicación de la cara en la imagen, alto.
- sonrisa: intensidad de la sonrisa, un número entre $[0,1]$, cero para no sonriente y uno para muy sonriente.
- inclinacionEjeX: ángulos de rotación alrededor del eje x, valores son de -90 a 90 grados.
- rotación: ángulos de rotación alrededor del eje z, valores son de -90 a 90 grados.
- inclinacionEjeY: ángulos de rotación alrededor del eje y, valores son de -90 a 90 grados.
- genero: estimado de la cara determinada. Los valores posibles son hombre, mujer(male or female) y sin sexo definido.
- edad: es la edad estimada en años de una cara determinada.
- bigote: longitud de retorno en el área de vello facial: bigote.
- barba: longitud de retorno en el área de vello facial: barba.
- patillas: longitud de retorno en el área de vello facial: patillas.
- lentes: Si la cara determinada tiene gafas. Los valores posibles son NoGlasses, ReadingGlasses, Sunglasses y Swimming Goggles.
- enfado: Intensidad de la emoción enfado.
- desprecio: Intensidad de la emoción desprecio.
- asco: Intensidad de la emoción disgusto.
- temor: Intensidad de la emoción temor.
- felicidad: Intensidad de la emoción felicidad.
- neutral: Intensidad de la emoción neutral.
- tristeza: Intensidad de la emoción tristeza.
- sorpresa: Intensidad de la emoción sorpresa.
- nivelDifuminacion: la cara está borrosa o no. Devoluciones de nivel: 'Low', 'Medium' or 'High'.
- valorDifuminacion: la cara está borrosa o no. El valor devuelve un número entre $[0,1]$, cuanto más grande es la más borrosa.
- nivelDeExposicion: nivel de exposición facial. Devoluciones de nivel 'GoodExposure', 'OverExposure' or 'UnderExposure' -'Buena exposición', 'Sobreexposición' o 'Subexposición'.
- valorDeExposicion: nivel de exposición facial. El valor devuelve un número entre $[0,1]$, cuanto más grande es mayor el nivel de exposición.

- nivelDeRuido: nivel de ruido de los píxeles de la cara. Devoluciones de nivel 'Low', 'Medium' y 'High'.
- valorDeNivelDeRuido: nivel de ruido de los píxeles de la cara. El valor devuelve un número entre [0,1], cuanto más grande, más ruidoso.
- maquillajeDeOjos: El valor devuelve un número [0,1] ya sea que las áreas de los ojos estén maquilladas o no.
- maquillajeDeLabios: El valor devuelve un número [0,1] ya sea que las áreas de los labios estén maquilladas o no.
- accesorios: accesorios alrededor de la cara, incluidos 'sombreros', 'gafas' y 'máscara'. Matriz vacía significa que no se han detectado accesorios. Tenga en cuenta que esto es después de que se detecta una cara. Una máscara grande puede resultar en que no se detecte ningún rostro.
- frenteOcluida: si el área facial de la frente está ocluida.
- ojoOcluido: si el área facial de los ojos está ocluida.
- bocaOcluida: si el área facial de la boca está ocluida
- calva: grupo de valores de cabello que indican si el rostro detectado presenta calva.
- invisible: grupo de valores de cabello que indican si el cabello es visible
- colorDeCabello: grupo de valores de cabello que indican el color del cabello si es visible.
- photoID: nombre de la imagen capturada durante el taller de RE.
- imageURL: dirección o repositorio en el cual se subió la imagen para su análisis.
- promedioExperienciasNegativas: promedio de las emociones negativas (enfado, desprecio, asco, temor, tristeza) detectadas en un rostro.
- experienciaUsuario: experiencia predominante de acuerdo al promedio de emoción. Devoluciones de experiencia de usuario 'positiva', 'negativa' y 'neutral'.
- experienciaUsuarioNumerica: valor de la experiencia predominante de acuerdo al promedio de emoción ('positiva', 'negativa' y 'neutral').
- fecha_hora: fecha y hora en que se obtuvo la captura del rostro.
- prediccionRandomForest: predicción de la experiencia de usuario obtenida mediante el modelo Random Forest. Devoluciones de la predicción de Random Forest 'positiva' y 'neutral'.
- rangoPrediccionRandomForest: rango de valores en el que se encuentra la predicción obtenida mediante el modelo Random Forest. Devoluciones de rango de predicción de Random Forest 'range1' [0-0,347), 'range2' [0,347-0,674) y 'range3' [0,674-1,00]

Con respecto a la perspectiva "Taller", los datos que se pueden utilizar son los siguientes:

- idTaller: identificador único del taller realizado.
- nombreTaller: nombre del taller realizado.
- competencias: descripción del conjunto de capacidades, habilidades, destrezas y actitudes vinculadas con el taller.
- capacidad: descripción de las habilidades relacionadas al conocimiento adquirido mediante el taller.
- indicador: unidades de medición que permiten evaluar el rendimiento del niño en el taller.
- fechaInicio: fecha de inicio del taller.
- fechaFin: fecha de finalización del taller.
- idRegistroT: identificador único del registro de un taller realizado.
- taller: identificador del taller realizado.

En la perspectiva "Experiencia de Usuario", los datos que se pueden utilizar son los siguientes:

- experienciaUsuario: experiencia predominante de acuerdo al promedio de emoción. Devoluciones de experiencia de usuario 'positiva', 'negativa' y 'neutral'.

En la perspectiva "Fases", los datos que se pueden utilizar son los siguientes:

- fase: Identificador de la fase. Por ejemplo: 1->Explorar, 2->Crear, 3->Compartir
- nombreDeLaFase: Nombre de la fase del taller. Por ejemplo: Crear, Explorar, Compartir

Con respecto a la perspectiva "Tiempo", que es la que determinará la granularidad del depósito de datos, los datos que pueden emplearse son los siguientes:

- Año.
- Número de mes.
- Número de día.

Una vez que se recolectó toda la información pertinente y se consultó con los usuarios sobre los datos de interés para analizar los indicadores ya expuestos, los resultados obtenidos fueron los siguientes:

- Perspectiva "Persona":
 - "idPersona" de la tabla "Persona", ya que representa a un estudiante.
 - "genero" de la tabla "Persona", ya que referencia al género del estudiante.
 - "edad" de la tabla "Persona", ya que referencia al género del estudiante.
 - "grupo" de la tabla "Persona", ya que referencia al grupo en el cual trabajó el estudiante.
 - "nombresCompleto" de la tabla "Persona", ya que referencia a los nombres y apellidos del estudiante.

- Perspectiva "Emociones":
 - "enfado" de la tabla "Json", ya que hace referencia a la emoción enfado.
 - "desprecio" de la tabla "Json", ya que hace referencia a la intensidad de emoción desprecio.
 - "asco" de la tabla "Json", ya que hace referencia a la intensidad de la emoción disgusto.
 - "temor" de la tabla "Json", ya que hace referencia a la intensidad de la emoción temor.
 - "felicidad" de la tabla "Json", ya que hace referencia a la intensidad de la emoción felicidad.
 - "neutral" de la tabla "Json", ya que hace referencia a la intensidad de la emoción neutral.
 - "tristeza" de la tabla "Json", ya que hace referencia a la intensidad de la emoción tristeza.
 - "sorpresa" de la tabla "Json", ya que hace referencia a la intensidad de la emoción sorpresa.

- Perspectiva "Taller":
 - "taller" de la tabla "RegistroT", ya que referencia al identificador del taller.
 - "nombreTaller" de la tabla "Taller", ya que referencia a nombre del taller. Este campo es obtenido a través de la unión con la tabla "RegistroT"
 - "fechaInicio" de la tabla "Taller", ya que referencia a la fecha de inicio del taller.
 - "fechaFin" de la tabla "Taller", ya que referencia a la fecha de finalización del taller.

- Perspectiva "Experiencia de Usuario":
 - "experienciaUsuario" de la tabla "Json", ya que hace referencia a la experiencia positiva, neutral o negativa del estudiante.
- Perspectiva "Fases":
 - "fase" de la tabla "Json", ya que hace referencia al identificador de la fase.
 - "nombreDeLaFase" de la tabla "Json", ya que hace referencia al nombre de la fase del taller.
- Perspectiva "Tiempo":
 - Año.
 - Número de mes.
 - Número de día.

2.2.4 Modelo conceptual ampliado

En la Figura 2.6 se desarrolla el modelo conceptual, colocando bajo cada perspectiva los campos seleccionados y bajo cada indicador su respectiva fórmula de cálculo.

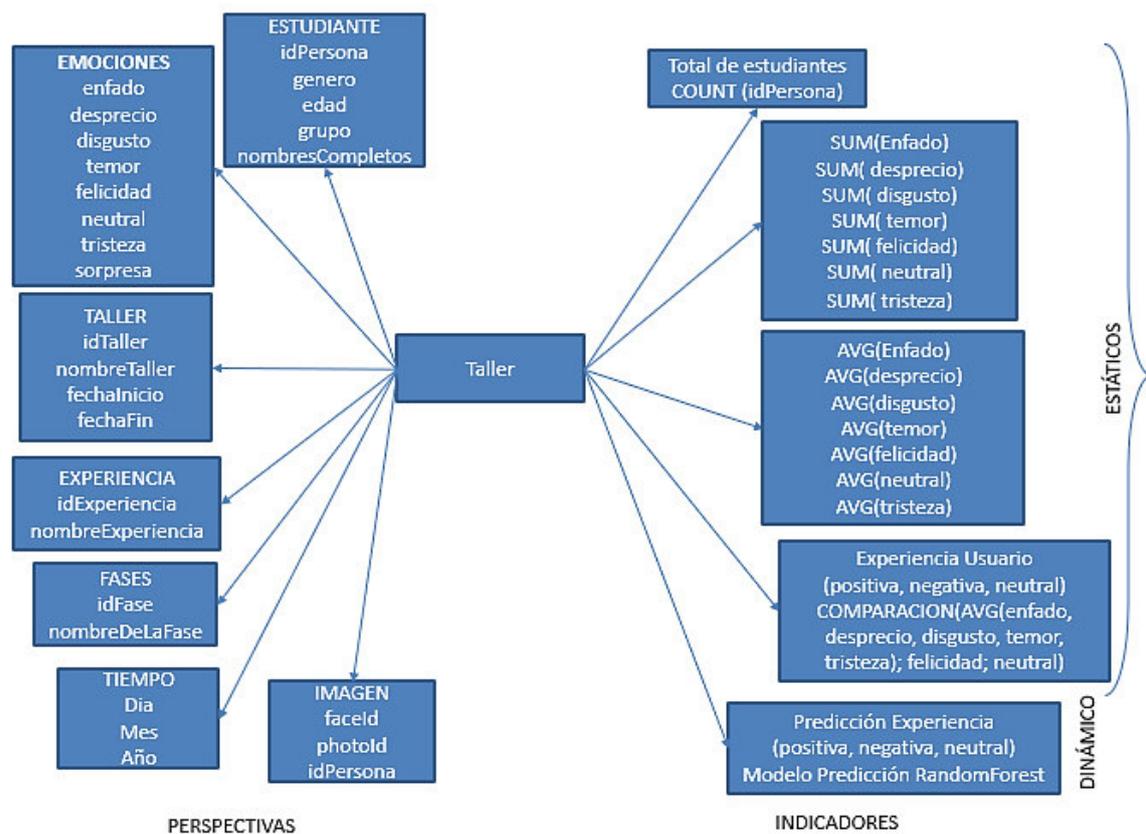


Figura 2.6 Talleres RE, Modelo Conceptual Ampliado

2.3 Modelo lógico del DW

2.3.1 Tablas de Dimensiones

Las tablas de dimensiones definen como están los datos organizados lógicamente y proveen el medio para analizar el contexto del negocio. Contienen datos cualitativos. Representan los aspectos de interés, mediante los cuales los usuarios podrán filtrar y manipular la información almacenada en la tabla de hechos. Cada tabla de dimensión podrá contener los siguientes campos:

- Clave principal o identificador único.
- Claves foráneas.
- Datos de referencia primarios: identifican la dimensión.
- Datos de referencia secundarios: complementan la descripción de la dimensión.

No siempre la clave primaria del OLTP, se corresponde con la clave primaria de la tabla de dimensión relacionada. Es recomendable manejar un sistema de claves subrogadas en el DW totalmente diferente al de los OLTP, ya que, si estos últimos son recodificados, el almacén quedaría inconsistente y debería ser poblado nuevamente en su totalidad.

En un DW, la creación y el mantenimiento de una tabla de dimensión Tiempo es obligatoria, y la definición de granularidad y estructuración de la misma depende de la dinámica del negocio analizado. Toda la información dentro del depósito, posee su propio sello de tiempo que determina la ocurrencia de un hecho específico, representando de esta manera diferentes versiones de una misma situación.

2.3.2 Tablas de Hechos

Las tablas de hechos contienen, precisamente, los hechos que serán utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones. Contienen datos cuantitativos. Los hechos son aquellos datos que residen en una tabla de hechos y que son utilizados para crear indicadores, a través de sumalizaciones preestablecidas al momento de crear un cubo multidimensional. El cubo almacena los datos de forma multidimensional, es decir, a través de tablas de hechos y tablas de dimensiones.

Las sumalizaciones no están referidas solo a sumas, sino también a promedios, mínimos, máximos, totales por sector, porcentajes, fórmulas predefinidas, etc., dependiendo de los requerimientos de información del negocio. [1]

En este paso, se definen las tablas de hechos, que son las que contienen los hechos a través de los cuales se construyen los indicadores de estudio.

Para el esquema constelación, se realiza lo siguiente:

- Las tablas de hechos se confeccionan teniendo en cuenta el análisis de las preguntas realizadas a los usuarios en pasos anteriores con sus respectivos indicadores y perspectivas.
- Se asigna un nombre a cada tabla de hechos que represente la información analizada, área de investigación, negocio enfocado, etc.
- Se define su clave primaria, que se compone de la combinación de las claves primarias de cada tabla de dimensión relacionada.
- Se crea tantos campos de hechos como indicadores se hayan definido en el modelo conceptual y se les asigna los mismos nombres que estos.

2.3.3 Diseño del DW

Se procede con el diseño de las tablas de dimensiones:

- Perspectiva "Persona":
 - La nueva tabla de dimensión tendrá el nombre "dim_persona".
 - Se le agregará una clave principal con el nombre "idPersona".

Se puede apreciar el resultado de estas operaciones en la Figura 2.7.

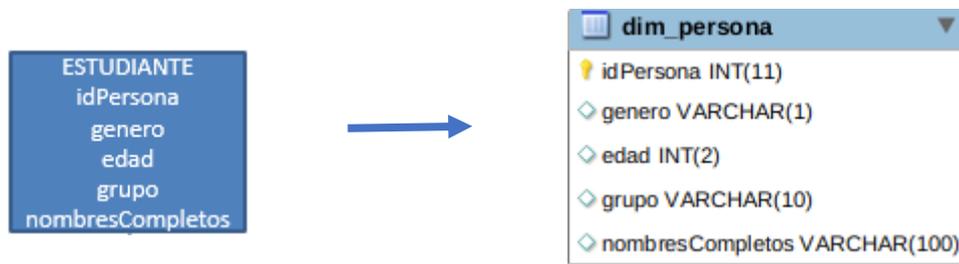


Figura 2.7 Taller RE, tabla de dimensión "PERSONA"

- Perspectiva "Taller":
 - La nueva tabla de dimensión tendrá el nombre "dim_taller".
 - Se le agregará una clave principal con el nombre "id_taller".

Se puede apreciar el resultado de estas operaciones en la Figura 2.8.

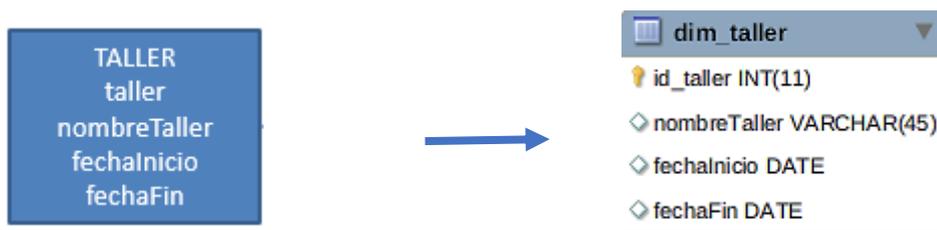


Figura 2.8 Taller RE, tabla de dimensión "TALLER"

- Perspectiva "Experiencia":
 - La nueva tabla de dimensión tendrá el nombre "dim_experiencia".
 - Se le agregará una clave principal con el nombre "idExperiencia".

Se puede apreciar el resultado de estas operaciones en la Figura 2.9.

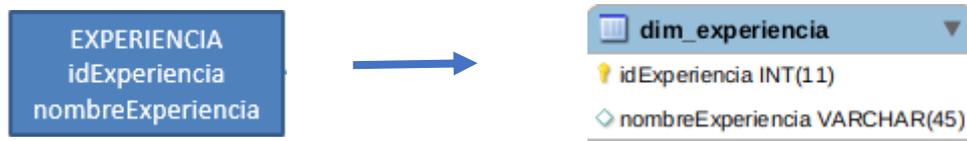


Figura 2.9 Taller RE, tabla de dimensión "EXPERIENCIA"

- Perspectiva "Fases":
 - La nueva tabla de dimensión tendrá el nombre "dim_fase".
 - Se le agregará una clave principal con el nombre "idFase".

Se puede apreciar el resultado de estas operaciones en la Figura 2.10.

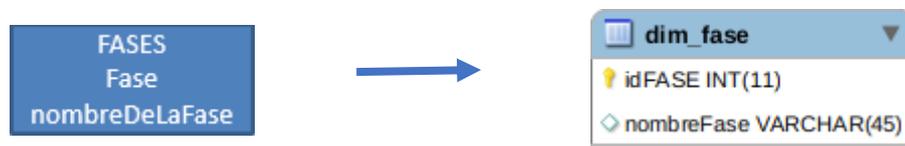


Figura 2.10 Taller RE, tabla de dimensión "FASE"

- Perspectiva "Tiempo":
 - La nueva tabla de dimensión tendrá el nombre "dim_tiempo".
 - Se le agregará una clave principal con el nombre "fecha_tie".

Se puede apreciar el resultado de estas operaciones en la Figura 2.11.

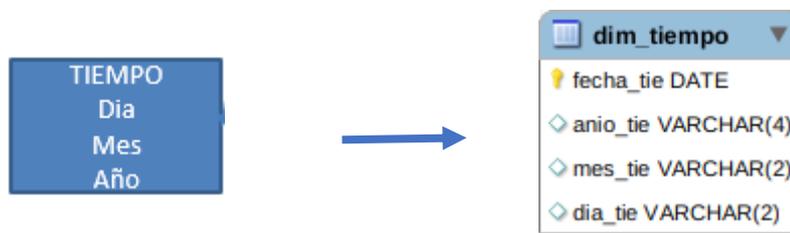


Figura 2.11 Taller RE, tabla de dimensión "TIEMPO"

- Perspectiva "Imagen":
 - La nueva tabla de dimensión tendrá el nombre "dim_imagen".
 - Se le agregará una clave principal con el nombre "id_imagen".

Se puede apreciar el resultado de estas operaciones en la Figura 2.12.

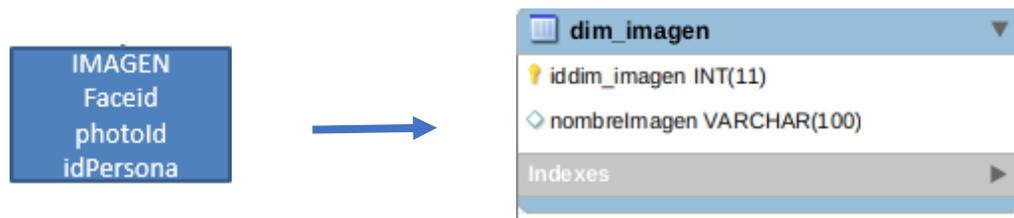


Figura 2.12 Taller RE, tabla de dimensión "IMAGEN"

- Perspectiva "Emocion":
 - La nueva tabla de dimensión tendrá el nombre "dim_emocion".
 - Se le agregará una clave principal con el nombre "iddim_emocion".

Se puede apreciar el resultado de estas operaciones en la Figura 2.13.

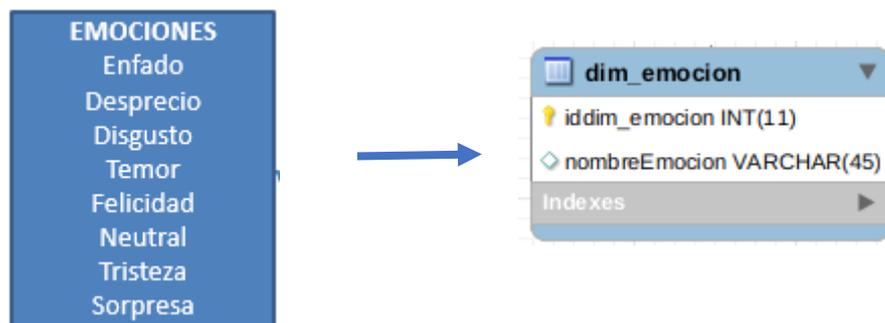


Figura 2.13 Taller RE, tabla de dimensión "EMOCION"

Se crearán dos hechos, que se corresponden con los dos indicadores y serán renombrados, "h_registroTaller", "valor" por "Total Estudiantes", "Total Emociones" y "h_prediccion", "valorExperiencia" por "ExperienciaUsuario", "Predicción Experiencia". La tabla de hechos tendrá el nombre "h_registroTaller". Su clave principal será la combinación de las claves principales de las tablas de dimensiones antes definidas: "idPersona", "id_taller", "idFase", "fecha_tie", "id_emocion" e "id_imagen".

La tabla de hechos tendrá el nombre "h_prediccion". Su clave principal será la combinación de las claves principales de las tablas de dimensiones antes definidas: "idExperiencia", "id_taller", "idFase" e "id_imagen".

2.3.4 Uniones

En la Figura 2.14 se encuentra el esquema con las uniones correspondientes entre sus tablas de dimensiones y sus tablas de hechos.

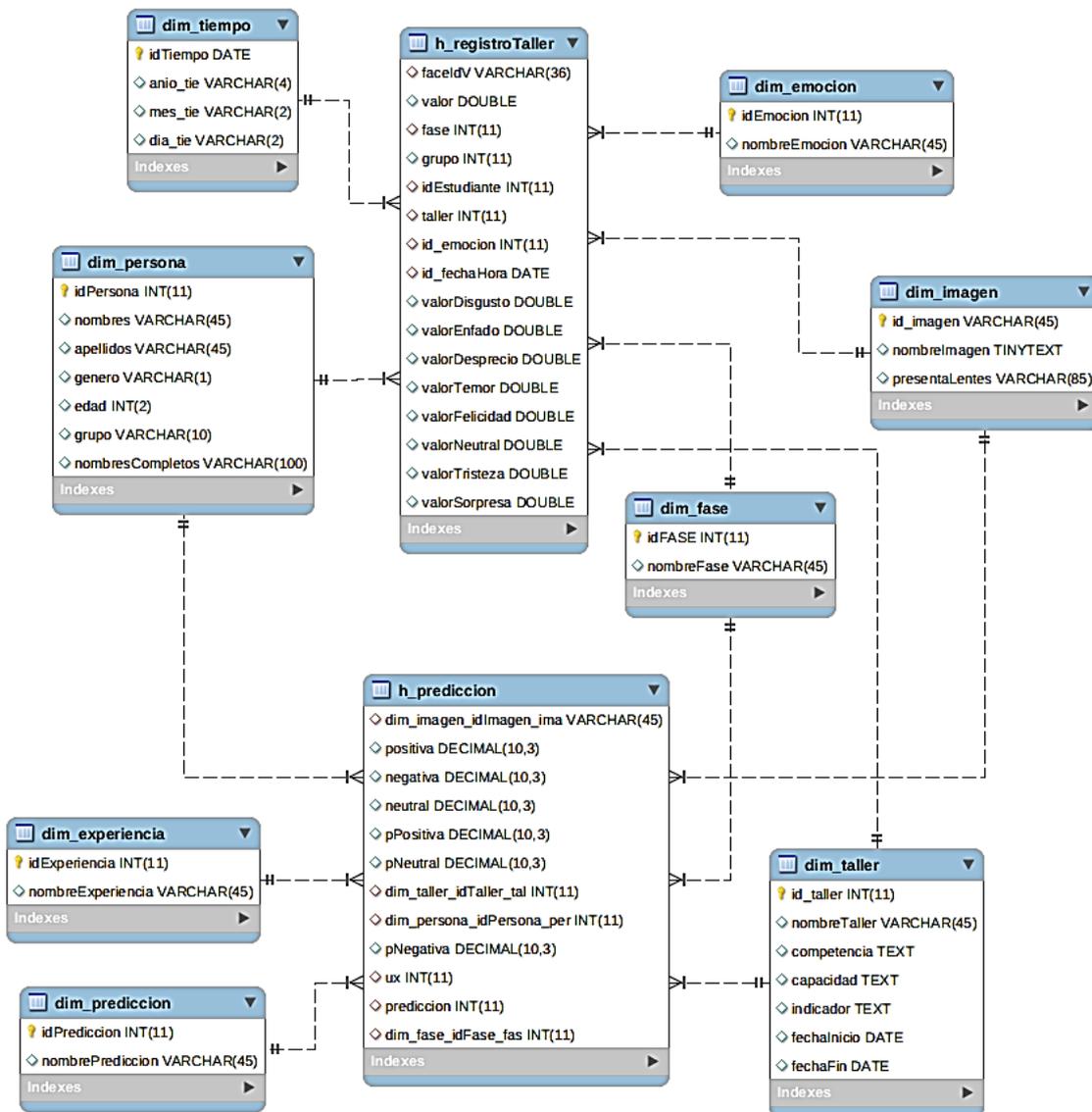


Figura 2.14 Taller RE, Uniones

2.4 Integración de datos

La integración implica que todos los datos de diversas fuentes que son producidos por distintos departamentos, secciones y aplicaciones, tanto internos como externos, deben ser consolidados en una instancia antes de ser agregados al DW, y deben por lo tanto ser analizados para asegurar su calidad y limpieza, entre otras cosas. La Integración de Datos agrupa una serie de técnicas y subprocesos que se encargan de llevar a cabo todas las tareas relacionadas con la extracción, manipulación, control, integración, depuración de datos, carga y actualización del DW, etc. Es decir, todas las tareas que se realizarán desde que se toman los datos de los diferentes OLTP hasta que se cargan en el DW. Una de estas técnicas son los procesos ETL: Extracción, Transformación y Carga de Datos (Extraction, Transformation and Load).

En el proceso de Extracción se tienen técnicas enfocadas a tomar solo los datos indicados y mantenerlos en un almacenamiento intermedio, en el proceso de transformación se analizan los datos para verificar que sean correctos y válidos; en el proceso de Carga de Datos se agrupan técnicas propias de la carga y actualización del DW. [1]

Extracción

Es aquí en donde se realiza el proceso de extraer desde las diferentes fuentes de información de la institución, tanto internas como externas, los datos que serán necesarios para encontrar las respuestas a las preguntas planteadas en el punto 2.1.2.1 “Elaboración de encuestas para padres, autoridades y maestros delegados”. [1] La extracción de datos resuelve diferentes tipos de problemas relacionados con las convenciones de nombres, unidades de medidas, codificaciones, fuentes múltiples, etc. Los datos se encuentran en fuentes externas textuales, hipertextuales, hojas de cálculos y la obtención de los mismos presenta dificultades ya que se tiene que realizar cambios de formato y volcado de información a partir de alguna herramienta específica. Una vez que los datos son seleccionados y extraídos, se guardan en un almacenamiento intermedio, lo cual permite:

- Manipular los datos sin interrumpir ni paralizar los OLTP, ni tampoco el DW.
- No depender de la disponibilidad de los OLTP.
- Almacenar y gestionar los metadatos que se generen en los procesos ETL.
- Facilitar la integración de las diversas fuentes, internas y externas.

El almacenamiento intermedio constituye una base de datos en donde la información puede ser almacenada en tablas auxiliares y tablas temporales.

Los datos de estas tablas serán los que finalmente (luego de su correspondiente transformación) poblarán el DW.

Transformación

En esta fase es donde se integran y cargan los datos en crudo en un formato utilizable para el análisis. Esta actividad puede realizarse mediante la creación de una nueva base de datos, agregando datos a una base de datos ya existente o bien consolidando la información. [1]

Los casos más comunes en los que se deberá realizar transformación, son los siguientes:

Codificación. Una inconsistencia muy típica que se encuentra al intentar integrar varias fuentes de datos, es la de contar con más de una forma de codificar un atributo en común.

- Medida de atributos. Los tipos de unidades de medidas utilizados para representar los atributos de una entidad, varían considerablemente entre sí, a través de los diferentes OLTP.
- Convenciones de nombramiento. Usualmente, un mismo atributo es nombrado de diversas maneras en los diferentes OLTP
- Fuentes múltiples. Un mismo elemento puede derivarse desde varias fuentes. En este caso, se debe elegir aquella fuente que se considere más fiable y apropiada.

Además de lo antes mencionado, esta función se encarga de realizar los procesos de Limpieza de Datos (Data Cleansing) y Calidad de Datos, su objetivo principal es el de realizar distintos tipos de acciones contra el mayor número de datos erróneos, inconsistentes e irrelevantes. Las acciones más típicas que se pueden llevar a cabo al encontrarse con Datos Anómalos (Outliers) o Datos Faltantes (Missing Values) son:

- Reemplazar el valor.
- Discretizar los valores de las columnas.
- Eliminar la columna.
- Filtrar la columna.
- Filtrar la fila errónea, ya que a veces su origen, se debe a casos especiales.
- Reemplazar el valor.
- Esperar hasta que los datos faltantes estén disponibles.

Un punto muy importante que se debe tener en cuenta al elegir alguna acción, es el de identificar la causa de la anomalía, para luego actuar en consecuencia, con el fin de evitar que se repitan, agregándole de esta manera más valor a los datos de la organización.

Carga

Esta función se encarga, por un lado, de realizar las tareas relacionadas con la Carga Inicial (Initial Load) y actualización o mantenimiento periódico (siempre teniendo en cuenta un intervalo de tiempo predefinido para tal operación).

El objetivo de esta tarea es añadir al depósito aquellos datos nuevos que se fueron generando desde la última actualización.

Políticas de Carga

El proceso de carga tiene la tarea de mantener la estructura del DW. Antes de realizar una nueva actualización, es necesario identificar si se han producido cambios en las

fuentes originales de los datos recogidos, desde la fecha del último mantenimiento, a fin de no atentar contra la consistencia del DW. Los datos antes de moverse al almacén, deben ser analizados con el propósito de asegurar su calidad.

Para efectuar esta operación, se pueden realizar las siguientes acciones:

- Cotejar las instancias de los OLTP involucrados.
- Utilizar disparadores en los OLTP.
- Recurrir a Marcas de Tiempo (Time Stamp), en los registros de los OLTP.
- Comparar los datos existentes en los dos ambientes (OLTP y DW).
- Ejecutar y definir las políticas de particionamiento: El objetivo de realizar esto, es conseguir una mayor eficiencia y performance en las consultas al no tener que manejar todo el grueso de los datos. Esta política debe aplicarse sobre la tabla de hechos que, como se explicará más adelante, es en la que se almacena toda la información que será analizada.
- Realizar copias de resguardo incrementales o totales de los datos del DW.

Planear la Integración de Datos

De acuerdo a lo mencionado en la sección anterior de la Integración de Datos se determina las siguientes tareas para la elaboración de ETLs.

Extracción

1. Subir archivos de json y csv a la base de datos.
2. Subir información de los talleres incluyendo nombres completos de los estudiantes que asistieron y fechas de realización que se encuentran en pdf.
3. Estandarizar las entradas (CSV y archivos JSON) orden de campos, (formatos) para la tabla JSON
4. Quitar caracteres especiales al pasar de la base de datos a un csv y viceversa.
5. Verificar la fecha de las fotografías y horas de acuerdo al acta de asistencia para los registros.

Para la experiencia y predicción:

6. Calcular de la experiencia de usuario.
7. Obtener la predicción de la experiencia del usuario para esto ver la tarea a realizar y la variable objetivo. Decidir el tipo de problema a resolver y los rangos de análisis.
8. Decidir el mejor algoritmo de predicción.

Transformación

1. Codificar: En un inicio el faceld tenía el valor de "n.a" por eso no se podía considerar clave primaria luego se identifica los rostros de los estudiantes en

cada imagen por eso la clave primaria en primera instancia se compone de los atributos: "photoId", "faceId" e "idPersona"; luego se pasó a considerar nombre de la imagen.

2. Validar la medida de atributos: considerar el manejo de fecha y formatear la misma, verificar cantidades que se las considera como texto de los archivos json, cvs o Excel ya que no se pueden realizar cálculos y se requiere números con tres cifras decimales en emociones.
3. Estandarizar nombramientos: considerar el manejo de nombre de grupos, fases, nombres de los estudiantes, para los nombres de los atributos se puede cambiar etiquetas de columna y trasladarlos de inglés a español para mayor comprensión.
4. Elegir de fuentes múltiples: tomar en cuenta la fuente de la cual se considera la edad y sexo del estudiante, tabla de la base de datos o del json de programa de reconocimiento facial.

Para la experiencia y predicción:

5. Usar Auxiliares: Se utiliza archivos xlsx y CSV para utilizar la función pivot en RM para traspaso de predicciones a la tabla de hechos h_prediccion.

Limpieza de Datos:

1. Considerar la eliminación de las columnas que no son relevantes en el presente estudio.
2. Eliminar registros duplicados.
3. Incluir id o nombre de fases faltantes en los registros de los talleres, esto se hace de acuerdo a la revisión de la fotografía o verificando con alguno de los demás datos existentes.
4. Filtrar Columna: verificar la fase, considerar los faceId diferentes de "n.a.", edad del estudiante menor a 20, (de acuerdo a la ley de Pareto se toma el 20% de los registros recopilados para verificar que edad es la adecuada a tomar en cuenta para descartar o no registros, de 1983 se toman 397 en los cuales se verifica que la edad que se debe considerar es los menores a 20 años, el mismo principio se aplicó para relacionar el nombre del estudiante de acuerdo a la fotografía)
5. Indicar la edad verdadera y el género de los estudiantes.
6. Verificar que los campos de cualquiera de los atributos no presenten valores nulos o vacíos como por ejemplo "idFase", calva, accesorios, edad, etc.
7. Filtrar la fila errónea: No Existe (NE). El NE se aplica en el nombre del estudiante cuando en la foto se encuentra un intruso, es decir un adulto.

8. Reemplazar el valor: tratamiento de valores nulos, cambiarlos a polinomial, reemplazar por valor 0 o descartarlos, por ejemplo: accesorios, invisible, calva, fase, "nombreEstudiante", fecha_hora, cabello, en valores detectados como perdidos por ejemplo emociones reemplazar por el valor 0, ya que los valores de 0 en emociones en un registro determinado afectan al momento de calcular promedios
9. Esperar hasta que los datos faltantes estén disponibles: faceld no había en un inicio, por lo que se utiliza el programa de reconocimiento facial para generar un "faceld".

Carga

En esta etapa se lleva adelante una serie de tareas básicas, tales como limpieza de datos, calidad de datos, procesos ETL.

2.4.1 Implementar la integración de datos

Para la integración de datos se emplean las herramientas Spoon de Pentaho, RM y Turbo Prep dependiendo de la necesidad.

Se puede apreciar como diferencias que en la herramienta Spoon de Pentaho, ciertos procedimientos son más engorrosos al realizar el ETL que en RM sin embargo el consumo de procesador es menor.

Primero se realiza la carga de los datos de las dimensiones y luego los de las tablas de hechos, teniendo en cuenta la correcta correspondencia entre cada elemento.

Problemas en la integración de datos

Se presentan problemas al utilizar la Api Rest de Azure como los siguientes:

- Podría haber una diferencia entre la edad y el género de la persona a la que se muestra en la fotografía, no existe precisión en este dato, por ejemplo, no se distingue bien la edad detecta de 25 años al rostro de una niña. La evidencia se aprecia en el Anexo IX Ejemplo de fotografías con ruido. Por este motivo se prioriza el uso de la tabla de base datos de persona donde están los datos correctos de los niños.
- Existen fotos pixeladas de baja calidad, por lo que se pierden datos. No detecta el rostro y los accesorios de la imagen, para disminuir los errores de reconocimiento deben asegurarse las condiciones explicadas en el Anexo VIII sección Datos de Entrada.

- Fue necesario realizar nuevamente el reconocimiento de emociones en las fotografías ya que en los archivos de Excel no se contaba con el nombre de las fotografías para poder asociar a la persona a quien pertenece. Se verificaron características como posición de la cara, claridad de la fotografía, género, accesorios, rostro más visible y mayormente cubierto en la fotografía para colocar los nombres. Luego se genera un documento de Excel donde se asoció dichas fotos con los nombres de los estudiantes, para esto también se pudo aplicar Machine Learning o distribuir los archivos CSV y fotografías para que se ingresen manualmente los nombres por cada registro o separar las fotografías por cada niño para realizar nuevamente el análisis incluyendo taller, fase y fechas.
- En un inicio no se dispone de todos los faceld por lo que se generó el nombre de las imágenes concatenando datos conocidos y se lo utilizó como identificador. El atributo faceld cambia en cada análisis por tal motivo no se puede empatar rostros ya identificados en un análisis con otro análisis, pero la ventaja que se presenta es que se generan más datos. Posteriormente al usar la Api Rest de Azure ya se obtuvo un “idFace” único el cual fue utilizado como el identificador definitivo.

2.4.2 Extracción

1. En la Figura 2.15 se observa el paso de los datos del archivo JSON a la base de datos en MySQL – Base de Datos: colegioV1, TABLA JSON.

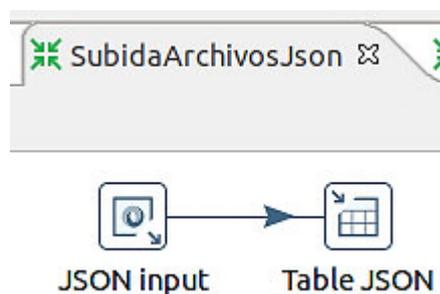


Figura 2.15 Transformación para subida de archivos JSON

En la Figura 2.16 se aprecia el resultado del trabajo anterior.

#	faceId	parteSuperior	partelzquierda	anchura	altura	sonrisa	tono	rodar	guinio	genero	edad	bigote
3	f09ed68a-673d-4aad-b9f7-95ad9ad9e594	247	663	72	72	0	-12.1	5.2	35.8	female	7	0
4	94fef1a7-6fbd-46f3-934a-1ec9041c0048	239	657	72	72	0	-8.7	2.5	49.9	female	8	0
5	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
6	c6a4acab-62ea-43bc-ad1a-bde647d2e5dd	246	583	75	75	0.997	-0.2	-7.2	-13.4	female	12	0

bigote	barba	patillas	lentes	enfado	desprecio	asco	temor	felicidad	neutral	tristeza	sorpresa	nivelDeDesenfoque	valorNivelDesenfoque
0	0	0	SwimmingGoggles	0	0	0	0	0.995	0.005	0	0	low	0.11
0	0	0	ReadingGlasses	0	0	0	0	0.998	0	0.001	0	low	0
n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
0	0	0	ReadingGlasses	0	0	0	0	0.997	0.003	0	0	medium	0.29
0	0	0	NoGlasses	0	0	0	0	0.999	0	0	0	medium	0.47

nivelDeExposicion	valorNivelDeExposicion	nivelDeRuido	valorDeNivelDeRuido	maquillajeDeOjos	maquillajeDeLabios	accesorios	fi:
overExposure	0.84	low	0	0	0	[[{"type":"glasses","confidence":0.97}]]	0
overExposure	0.76	low	0	0	1	[[{"type":"glasses","confidence":0.98}]]	0
n.a.	n.a.	n.a.	n.a.	0	0	n.a.	0
goodExposure	0.74	low	0.26	0	0	[[{"type":"glasses","confidence":0.99}]]	0
overExposure	0.97	low	0.15	0	0	[[{"type":"glasses","confidence":0.99}]]	0

Figura 2.16 Spoon: Resultado de la carga de archivos JSON

Para que pentaho pueda leer los datos de un archivo json se debe validar la estructura del archivo JSON y realizar la parametrización de los campos como se aprecia en la Figura 2.17, caso contrario trae información incompleta o la información no se detecta.

Step name

File	Content	Fields	Additional output fields
Name	Path	Type	Format
1	faceId	\$.faceId	String
2	top	\$.top	String
3	left	\$.left	String
4	width	\$.width	String
5	height	\$.height	String
6	smile	\$.smile	String
7	pitch	\$.pitch	String
8	roll	\$.roll	String
9	yaw	\$.yaw	String
10	gender	\$.gender	String
11	age	\$.age	String
12	moustache	\$.moustache	String
13	beard	\$.beard	String
14	sideburns	\$.sideburns	String
15	glasses	\$.glasses	String

Select fields

Help OK Preview rows Cancel

Figura 2.17 Spoon: Parametrización de campos de un archivo JSON

En la Figura 2.18 se presenta el trabajo para extraer información de archivos de Excel a la tabla JSON.

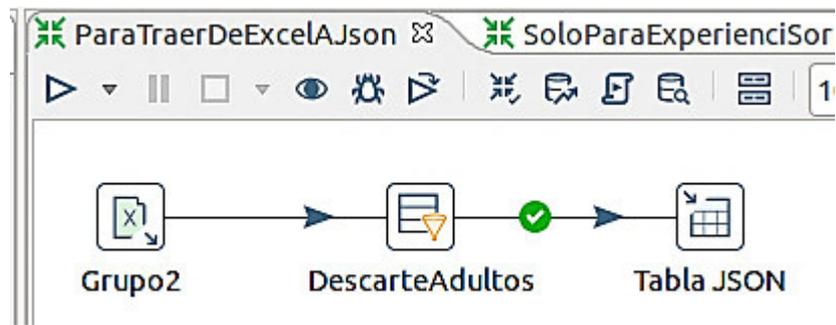


Figura 2.18 Spoon: Extracción de Excel a base de datos

En la Figura 2.19 se realiza el mismo proceso en RM.

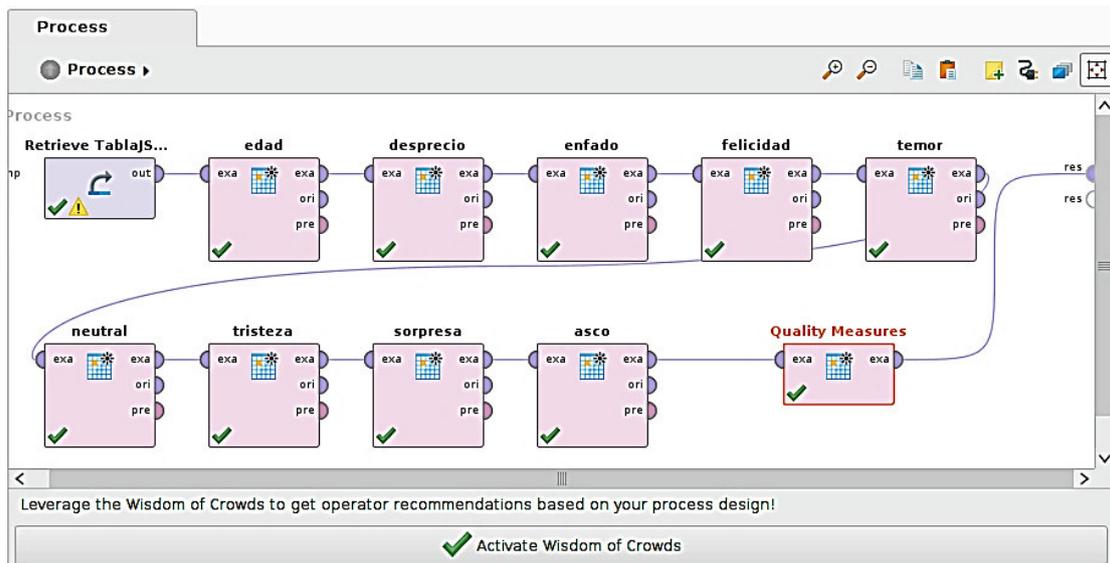


Figura 2.19 RM: Extracción de CSV a Base de datos

- En la Figura 2.20, los datos de los talleres en pdf para una mejor manipulación se pasaron a la base de datos, tabla TALLER.

#	idTaller	nombreTaller	competencias	capacidad	indicador
1	1	Ordenar para reciclar	CIENCIA Y AMBIENTE!...	Comprende y aplica conoci...	Caracteriza el problem...
2	2	Metamorfosis de rana	CIENCIA Y AMBIENTE!...	Comprende y aplica conoci...	Menciona que la meta...
3	3	Moviendo Materiales	CIENCIA FÍSICA! haga ...	Ciencias de la Naturaleza!In...	NE
4	4	Costruyendo un Satélite Espacial(Explorador)	CIENCIA Y AMBIENTE!...	Comprende y aplica conoci...	NE

fechaInicio	fechaFin
2020-01-13 07:30:00	2020-01-15 12:10:00
2020-01-20 07:30:00	2020-01-23 12:12:00
2020-01-27 07:30:00	2020-01-30 12:12:00
2020-02-03 07:30:00	2020-02-06 12:13:00

Figura 2.20 OLTP: Tabla TALLER

En la Figura 2.21 se puede verificar que desde el acta de asistencia al curso que se encontraba en formato pdf se sacaron los datos de los estudiantes en la tabla PERSONA.

#	idPERSONA	nombres	apellidos	tipo	genero	edad	nombresCompleto
5	6	Christofer Andrés	Galindo Rodríguez	GR1	M	10	Galindo Rodríguez Christofer Andrés
6	7	Mateo Alejandro	Fernández Llumigusín	GR1	M	7	Fernández Llumigusín Mateo Aleja...
7	11	Paula Noemi	Freire Proaño	GR1	F	7	Freire Proaño Paula Noemi
8	12	Mylena Tahis	Calderón Viteri	GR2	F	9	Calderón Viteri Mylena Tahis
9	13	Hannah Micaela	Carrillo Coral	GR2	F	9	Carrillo Coral Hannah Micaela

Figura 2.21 OLTP-Tabla PERSONA

En la Figura 2.22 se aprecia que las fotografías de los talleres se listaron e identificaron los niños que habían realizado el taller, esta información se subió a la tabla REGISTROT.

#	idREGISTROT	taller	grupo	idEstudiante
1	1	2	3	15
2	2	2	3	16
3	3	1	2	20
4	4	1	2	19

Figura 2.22 OLTP-Tabla REGISTROT

- En la Figura 2.23 se estandarizan las entradas de archivos CSV y JSON, se verifica el orden de campos, formatos de datos para la tabla JSON valiéndose de Excel, Pentaho y RapidMiner.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	faceld	parteSuperior	partetzquierda	anchura	altura	sonrisa	tono	rodar	guinio	genero	edad	bigote	barba
2	7b6ca55e-d4cd-4fa4-bff7-430eee360f18	268	78	96	96	0.003	-3.1	-5	56	male	12	0	0
3	62a3abc2-3ceb-4297-b432-1412bb550826	254	284	77	77	1	3.3	3.4	18.7	female	7	0	0
4	f09ed68a-673d-4aad-b9f7-95ad9ad9e594	247	663	72	72	0	-12.1	5.2	35.8	female	7	0	0
5	94fe1a7-6fbd-46f3-934a-1ec9041c0048	239	657	72	72	0	-8.7	2.5	49.9	female	8	0	0
6	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
7	c6a4acab-62ea-43bc-ad1a-bde647d2e5dd	246	583	75	75	0.997	-0.2	-7.2	-13.4	female	12	0	0
8	4ef68cd1-f9d8-457e-ac7a-30dd955943e8	223	141	104	104	0	-2.7	-2.1	46.7	male	10	0	0
9	02c57b69-9422-4cf3-8c56-007b45a5a1eb	242	623	75	75	0.014	-1.7	5.8	16.7	female	11	0	0
10	f34aef99-3c6d-4898-afad-2bdc5190abf0	239	634	78	78	0	-2.8	-0.9	31.2	female	6	0	0
11	ac30586c-e56d-4aeb-b633-108c0c8d4202	224	114	129	129	0.001	-11.1	2.6	25.9	male	9	0	0

Figura 2.23 Importación de archivo CSV a Excel

- En la Figura 2.24 se cambia de caracteres especiales.

	AR	AS
1	imageURL	nombreEstudiante
2	https://deusgallet.000webhostapp.com/recursos/AAL_emotions/Grupo1/Explorar/sesionid Galindo Rodríguez Christofer Andrés	
3	https://deusgallet.000webhostapp.com/recursos/AAL_emotions/Grupo1/Explorar/sesionid Freire Proanio Paula Noemi	
4	https://deusgallet.000webhostapp.com/recursos/AAL_emotions/Grupo1/Explorar/sesionid Freire Proanio Paula Noemi	
5	https://deusgallet.000webhostapp.com/recursos/AAL_emotions/Grupo1/Explorar/sesionid Freire Proanio Paula Noemi	
6	https://deusgallet.000webhostapp.com/recursos/AAL_emotions/Grupo1/Explorar/sesionid Freire Proanio Paula Noemi	
7	https://deusgallet.000webhostapp.com/recursos/AAL_emotions/Grupo1/Explorar/sesionid Galindo Rodríguez Christofer Andrés	
8	https://deusgallet.000webhostapp.com/recursos/AAL_emotions/Grupo1/Explorar/sesionid Freire Proanio Paula Noemi	
9	https://deusgallet.000webhostapp.com/recursos/AAL_emotions/Grupo1/Explorar/sesionid Freire Proanio Paula Noemi	
10	https://deusgallet.000webhostapp.com/recursos/AAL_emotions/Grupo1/Explorar/sesionid Galindo Rodríguez Christofer Andrés	
11	https://deusgallet.000webhostapp.com/recursos/AAL_emotions/Grupo1/Explorar/sesionid Freire Proanio Paula Noemi	
12	https://deusgallet.000webhostapp.com/recursos/AAL_emotions/Grupo1/Explorar/sesionid Fernández Llumigüñ Mateo Alejandro	
13	https://deusgallet.000webhostapp.com/recursos/AAL_emotions/Grupo1/Explorar/sesionid Freire Proanio Paula Noemi	

Figura 2.24 Archivo con caracteres especiales

En la Figura 2.25 se muestra el uso de UTF8 al importar el archivo desde Excel.

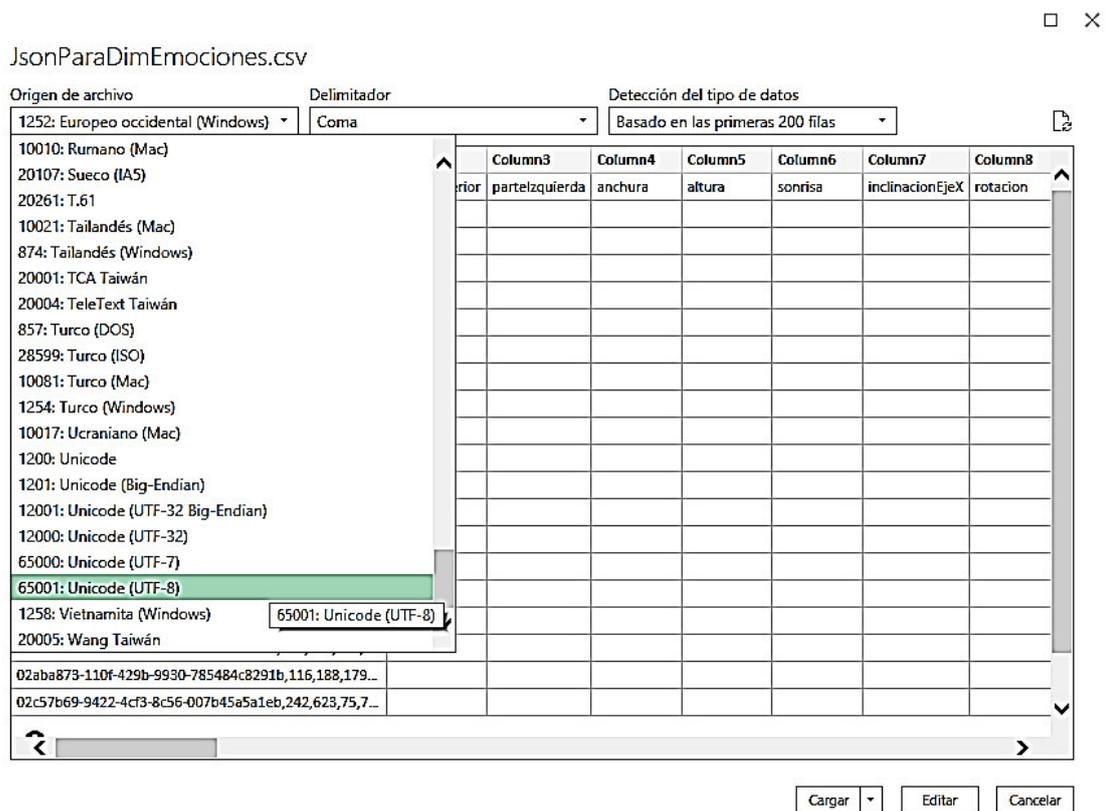


Figura 2.25 Uso de codificación UTF8 en Excel

En la Figura 2.26 se encuentra el cambio de caracteres especiales en Spoon.

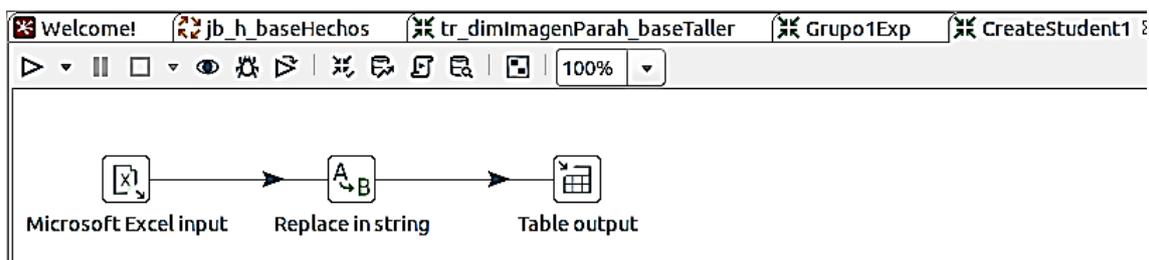


Figura 2.26 Spoon: Uso de Replace in String

5. En la Figura 2.27, al usar Excel para los datos faltantes se incluye un número randómico para la hora de las fotografías de acuerdo a la fecha y horarios del taller.

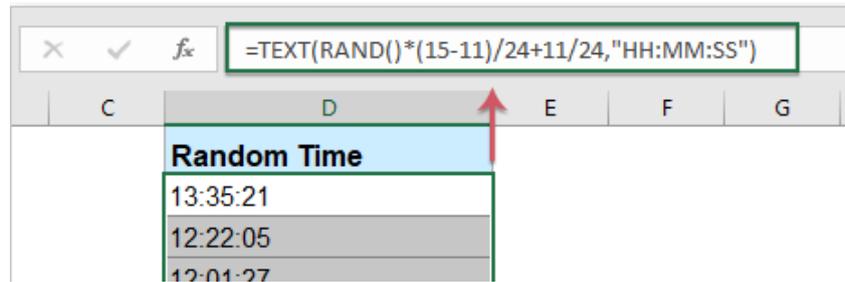


Figura 2.27 Tiempo aleatorio entre en un rango de horas en Excel

6. Cálculo de la experiencia de usuario. En la Figura 2.28 se muestra la comparación para elegir el valor mayor de la experiencia de usuario numérica y de acuerdo al resultado indicar la experiencia de usuario que predomina es positiva, negativa, neutral.

Figura 2.28 RM-Turbo Prep: Generar columna

7. En la Figura 2.29, para predecir la Experiencia de Usuario en la herramienta de modelado de RM se debe seleccionar Predict y la columna objetivo de la predicción "experiencia".

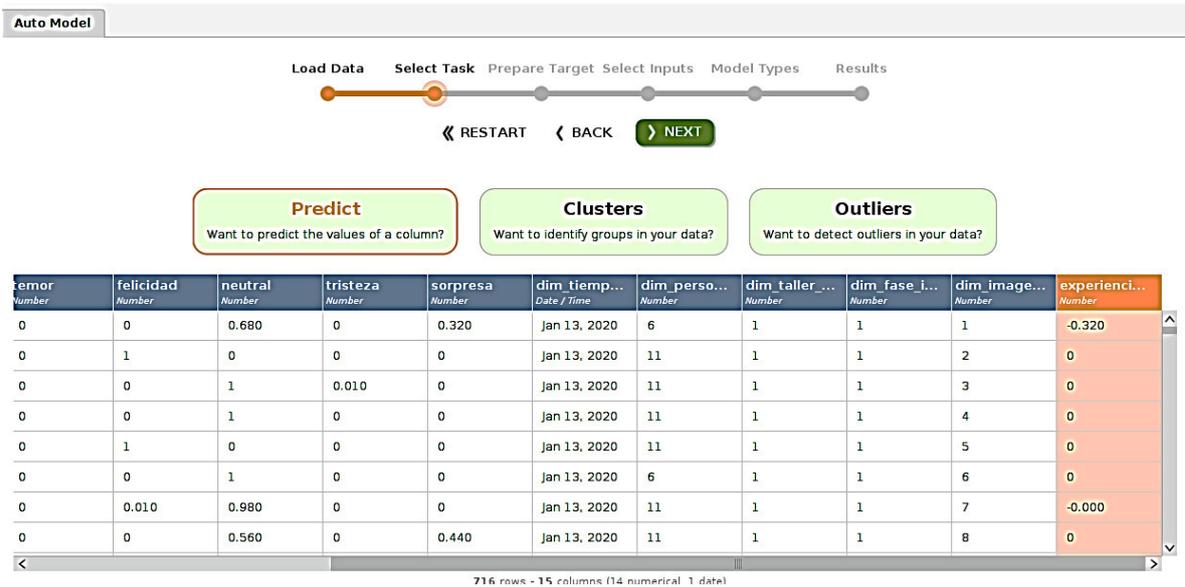


Figura 2.29 RM-Turbo Prep: Auto Modelado

Dado experiencia de usuario tiene 3 valores posibles “positivo”, “negativo” y “neutral” el problema es de clasificación como se indica en la Figura 2.30.

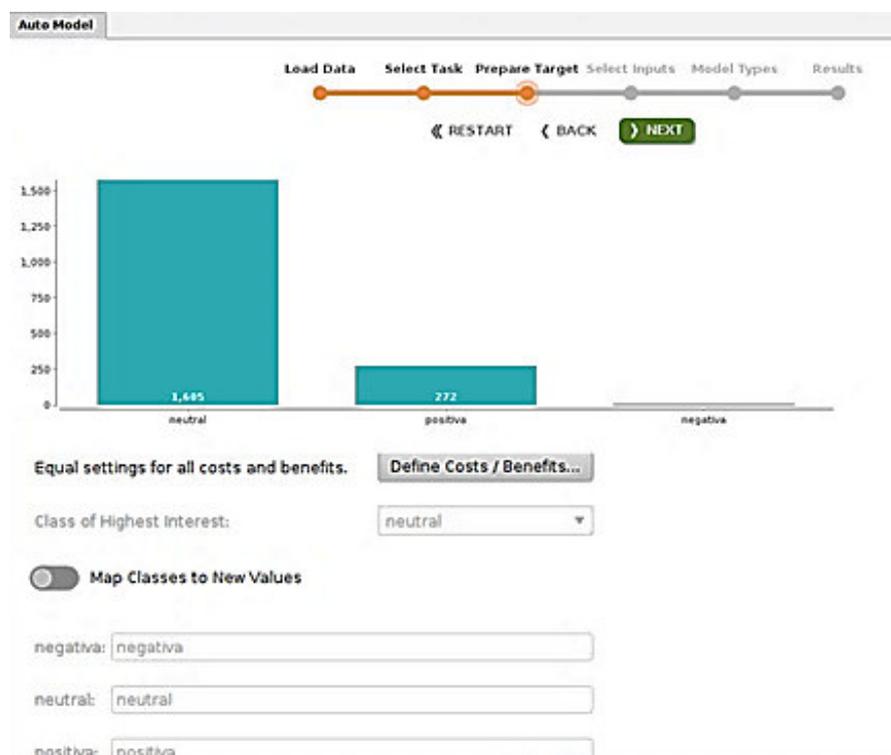


Figura 2.30 RM-Turbo Prep: Auto Modelado, preparación variable objetivo

En la Figura 2.31 se muestran los rangos de la experiencia de usuario numérica.

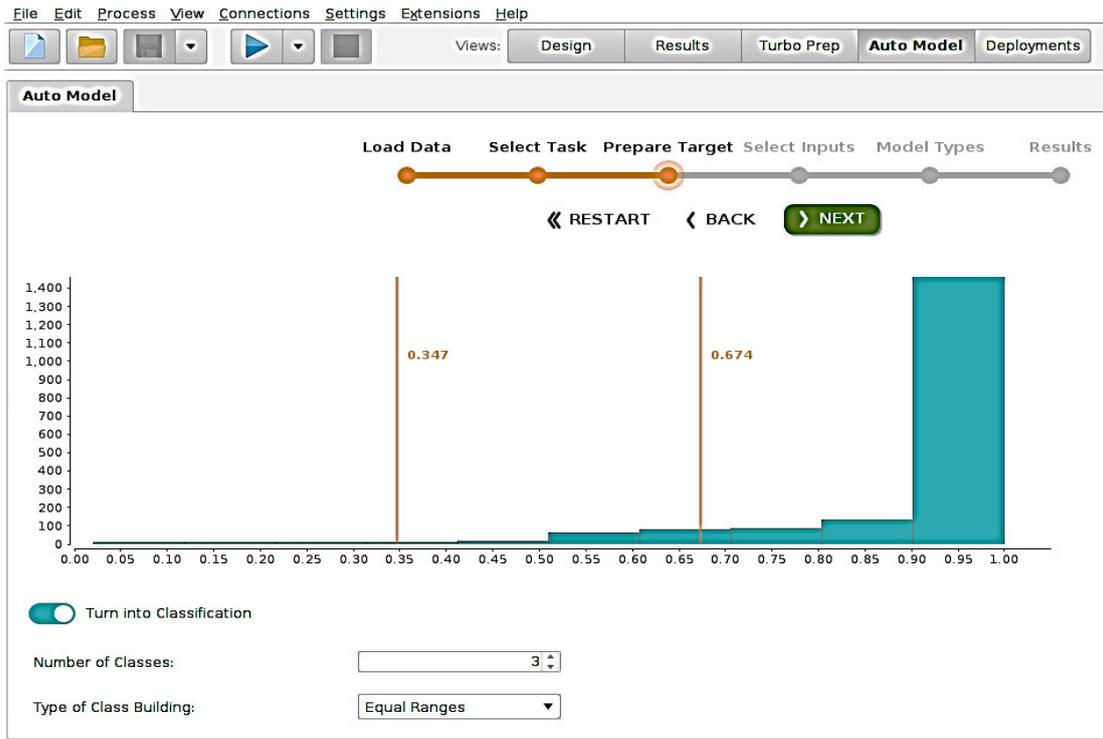


Figura 2.31 RM-Turbo Prep: Rangos Variable Objetivo

Resultados de la predicción

Model	Classification Er...	Standard Deviat...	Gains	Total Time
Random Forest	0.0%	± 0.0%	160	4 min 13 s
Gradient Boosted Trees	0.0%	± 0.0%	160	5 min 44 s
Support Vector Machine	14.3%	± 2.1%	0	47 min 2 s

Figura 2.32 RM-Turbo Prep: Vista General de los Modelos Predictivos

8. Se utiliza la comparación de los modelos generados representada en la Curva Roc, Figura 2.33. Para esto se elige el algoritmo que se encuentra en la parte superior izquierda de la curva.

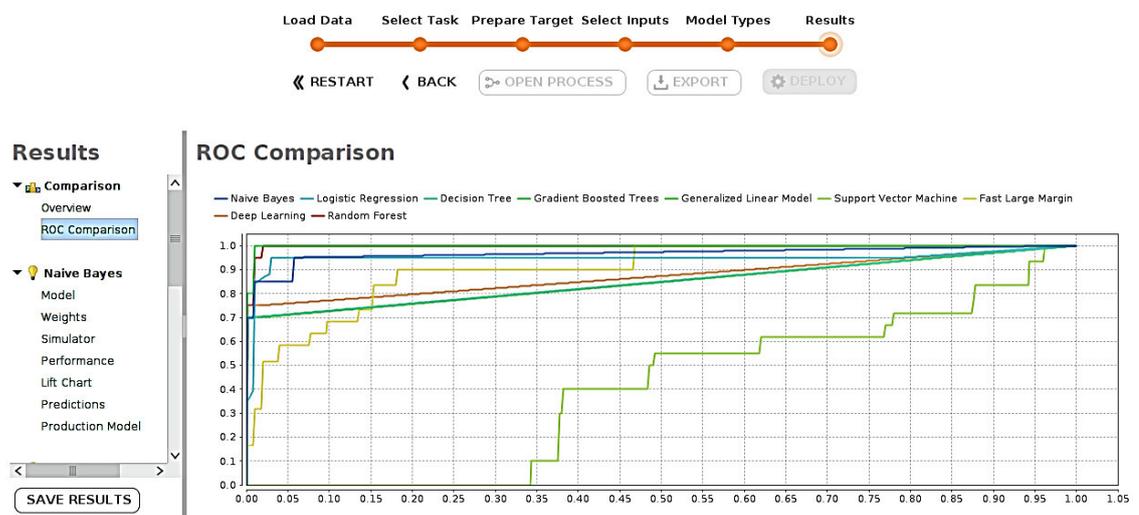


Figura 2.33 Curvas ROC

2.4.3 Transformación

1. En un inicio no se dispone de algunos nombres de las imágenes por lo que se procede a concatenar para el código de imagen el “idPersona”, faceld y “photoId” como se expone en la Figura 2.34.

Concat Fields

Step name:

Target Field Name:

Length of Target Field:

Separator:

Enclosure:

Name	Type	Format	Length	Precision	Currency	Decimal	Group
1 idpersona	String						
2 faceld	String						
3 photoId	String						

Get Fields Minimal width

OK Cancel

Figura 2.34 Spoon: Opción concatenar campos

2. En la Figura 2.35 se muestra el formateo de fechas en Spoon.

Select values

Step name:

Select & Alter Remove Meta-data

Fields to alter the meta-data for:

Fieldname	Rename to	Type	Length	Precision	Binary to Normal?	Format	Date Format Lenient?	Date Locale
1 fecha_calculada		None			N	yyyy-MM-dd	N	

Figura 2.35 Spoon: Manejo de fechas

Se cambia el tipo de datos de valores de texto a numérico. En la Figura 2.36 se indica que al importar de csv a Excel se elige la opción “No detectar tipos de datos” para conservar todo el número como una cadena.

JsonParaDimEmociones.csv

Origen de archivo: 1252: Europeo occidental (Windows) | Delimitador: Coma | Detección del tipo de datos: Basado en las primeras 200 filas

	nombreEstudiante	experienciaUsus		a_hora	Prc
ebhostapp.com/recursos/AAL_...	Fernández Llumigusán Mateo Alejandro	neutral		13/01/2020 12:06:15	
m/dggnilweu/image/upload/v...	Sarmiento Cárdoma Elías Sebastián	neutral		23/01/2020 11:41:02	
ebhostapp.com/recursos/AAL_...	Freire Proaño Paula Noemi	positiva	774	No detectar tipos de datos	
ebhostapp.com/recursos/AAL_...	Carrillo Coral Hannah Micaela	neutral	978	13/01/2020 8:39:33	
ebhostapp.com/recursos/AAL_...	Galindo Rodríguez Christofer Andrés	neutral	943	13/01/2020 7:38:39	
ebhostapp.com/recursos/AAL_...	Fernández Llumigusán Mateo Alejandro	neutral	975	16/01/2020 11:31:22	
googleapis.com/v0/b/imagenes...	Huang Zhou Lishan Karen	neutral	990	20/01/2020 8:48:04	
googleapis.com/v0/b/imagenes...	Ruiz Castillo Ane	neutral	837	23/01/2020 11:53:58	
googleapis.com/v0/b/imagenes...	Ruiz Castillo Ane	positiva	1000	20/01/2020 8:04:58	
googleapis.com/v0/b/imagenes...	Ruiz Castillo Ane	neutral	674	20/01/2020 8:29:46	
googleapis.com/v0/b/imagenes...	Rosero Escobar Isaac Edison	neutral	997	23/01/2020 11:44:13	
googleapis.com/v0/b/imagenes...	Ruiz Castillo Ane	neutral	995	20/01/2020 8:48:04	
ebhostapp.com/recursos/AAL_...	Fernández Llumigusán Mateo Alejandro	positiva	567	13/01/2020 8:21:17	
ebhostapp.com/recursos/AAL_...	Fernández Llumigusán Mateo Alejandro	neutral	994	13/01/2020 8:02:26	
ebhostapp.com/recursos/AAL_...	Galindo Rodríguez Christofer Andrés	neutral	981	13/01/2020 8:15:04	
ebhostapp.com/recursos/AAL_...	Freire Proaño Paula Noemi	neutral	960	16/01/2020 11:44:13	
ebhostapp.com/recursos/AAL_...	Bravo Noboa Jazmán Esmeralda	neutral	952	16/01/2020 11:50:39	
ebhostapp.com/recursos/AAL_...	Cabezas Altamirano Emanuel Ricardo	neutral	715	13/01/2020 11:40:03	
ebhostapp.com/recursos/AAL_...	Freire Proaño Paula Noemi	neutral	984	13/01/2020 7:48:10	
m/dggnilweu/image/upload/v...	Guerron Ríos Sarahi Alejandra	neutral	980	20/01/2020 7:49:58	

Cargar | Editar | Cancelar

Figura 2.36 Excel: Manejo de cantidades como cadenas

En la Figura 2.37 se muestra el cambio de tipo de datos de real a polinomial para la carga de datos en un repositorio local de RM- Turbo Prep para generar el modelo.

Format your columns.

Date format: Enter value... | Replace errors with missing values

	felicidad	neutral	tristeza	sorpresa	grupo	experien...
	real	real	real	real	integer	real
1	0.003	0.676	0.001	0.320	1	0.320
2	1.000	0.000	0.000	0.000	1	0.000
3	0.000	0.995	0.005	0.000	1	0.000
4	0.000	0.998	0.000	0.001	1	0.000
5	0.997	0.003	0.000	0.000	1	0.000
6	0.000	0.999	0.000	0.000	1	0.000
7	0.014	0.984	0.001	0.000	1	0.000
8	0.000	0.561	0.000	0.439	1	0.000
9	0.001	0.998	0.001	0.000	1	0.000
10	0.001	0.993	0.006	0.000	1	0.000
11	0.000	0.788	0.211	0.000	1	0.000
12	0.000	0.953	0.046	0.001	1	-0.001

no problems.

Previous | Finish | Cancel

Figura 2.37 RM: Cambio de tipo de dato

En la Figura 2.38 se presenta el cambio a tipo de dato numérico para las emociones en Spoon.

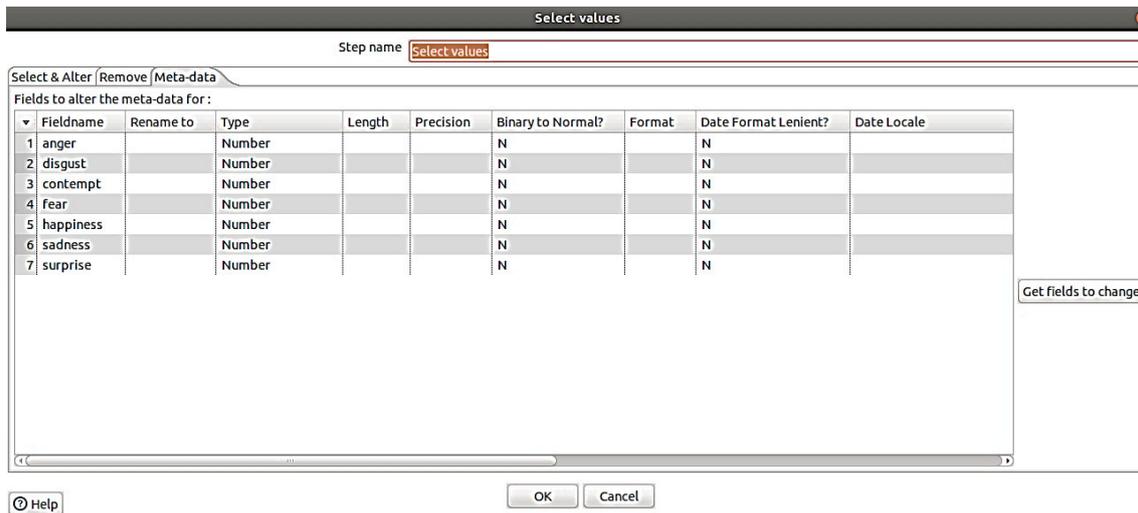


Figura 2.38 Spoon: Cambio de tipo de dato

3. En la Figura 2.39 se expone el reemplazo de la nomenclatura de grupo por número.

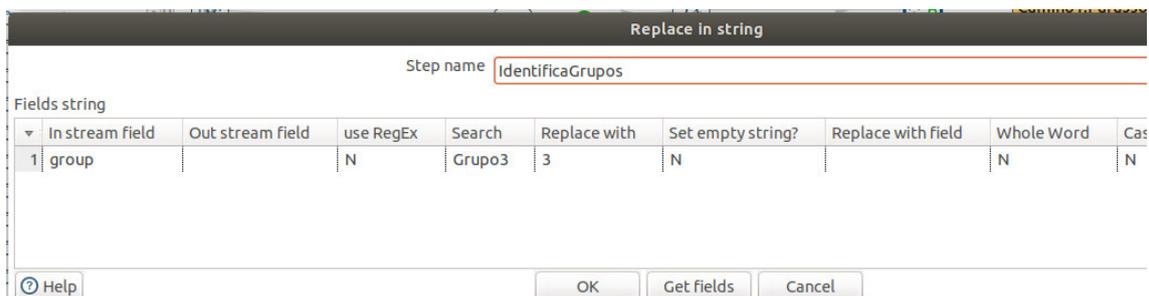


Figura 2.39 Spoon: Reemplazo de grupo de string a numérico

En la Figura 2.40 se presenta el cambio de las etiquetas de las columnas.

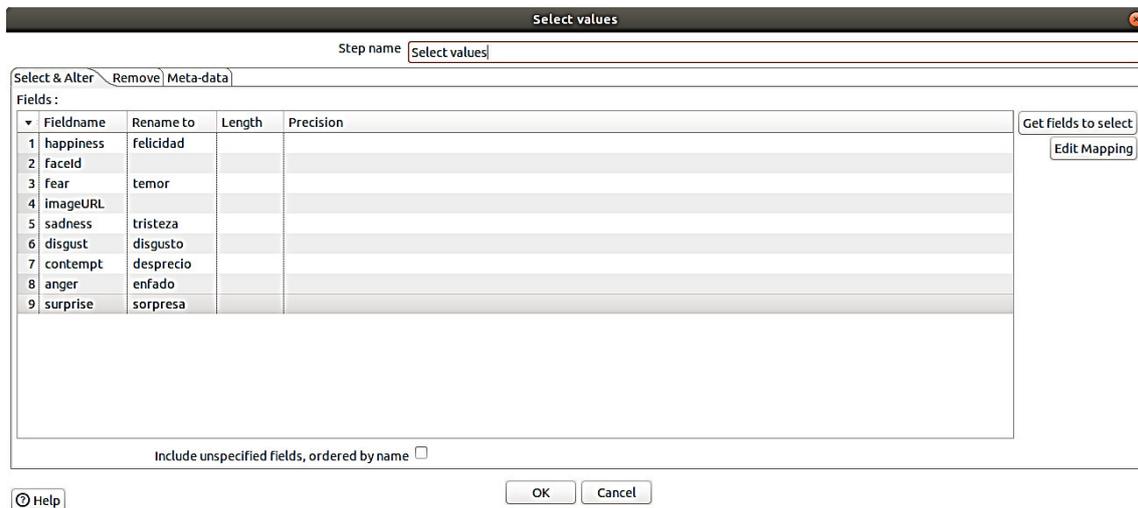


Figura 2.40 Spoon: Renombrar atributos

4. En la Figura 2.41 se expone la tabla de persona desde la cual se selecciona la edad real de los estudiantes.

Update

Step name

Connection

Target schema

Target table

Commit size

Use batch updates?

Skip lookup

Ignore lookup failure? Flag field (key found)

The key(s) to look up the value(s):

▼	Table field	Comparator	Stream field1	Stream field2	Get fields
1	nombresCompleto	=	nombreEstudiante		

Update fields:

▼	Table field	Stream field	Get update fields
1	edad	edad	

Figura 2.41 Spoon: Actualización de la edad con el dato de la tabla dim_persona

De los archivos de JSON y CSV se utiliza la moda para la edad de los estudiantes de los cuales no se dispone de este dato, Figura 2.42.

```
select edad,count(edad)as contador ,nombreEstudiante
from colegioV1.JSON where
nombreEstudiante = ? and edad<>'n.a.' and edad <=20
and nombreEstudiante<>'NE'
group by edad order by contador desc limit 1;
```

Figura 2.42 Uso de la moda para la edad del estudiante

5. En la Figura 2.43 se muestra el Pivot para predicciones.

Turbo Prep

Pivot

JsonUltimoParaprediccion1 (2)

Drag columns from the left to build a pivot table. ⓘ

COMMIT PIVOT CLEAR TABLE CANCEL

Search text

nombreDelAse Category

grupo Number

photoID Category

imageURL Category

nombreEstudiante Category

experienciaUsuario Category

fecha_hora Category

prediccionRandomForest Category

facelidV Category

Group-By

facelidV

Column Grouping

experienciaUsuario

facelidV	experienciaUsuario... count	experienciaUsuari... count	experienciaUs... count
fb20adb6-e282-4a11-8325-931adf80ee39	1	?	?
8b056658-279c-41a0-ad01-37a8b2e11dab	?	1	?
8ed4ef8a-7515-4478-b5d8-e685881a5fb6	1	?	?
a0edabd6-f190-45c8-8cd3-22168d5eabde	1	?	?
f49a702c-4fbe-4288-ac45-7e6230fcd70e	1	?	?
100a0c0e-ce19-48b4-a890-a406a42a61a3	1	?	?
aAf2orff-473a-4h81-b506-6a1f5d4d95518	?	1	?

1 Group-by Column 2 Aggregates Column Header: experienciaUsuario 1,932 rows - 9 columns (5 nominal, 4 numerical)

experienciaUs... count

experienciaUs... mode

Figura 2.43 RM-Turbo Prep: Función Pivot

2.4.4 Limpieza de datos

1. En la Figura 2.44 se observa la eliminación de las columnas “parteSuperior”, “partelzquierda”, anchura, “altura”, sonrisa, “inclinacionEjeX”, “inclinacionEjeY”, “rotacion”, bigote, barba, patillas, nivel y valor de difuminación, nivel y valor de exposición, nivel y valor de ruido, maquillaje de labios y ojos, accesorios, frente, ojo y boca ocluidos, calva, invisible, color de cabello.

Select values

Step name: DatosNoRequeridos

Select & Alter Remove Meta-data

Fields to remove:

Fieldname
1 parteSuperior
2 partelzquierda
3 anchura
4 altura
5 sonrisa
6 tono
7 rodar
8 guinio
9 bigote
10 barba
11 patillas
12 nivelDeDesenfoue
13 valorNivelDesenfoue
14 nivelDeExposicion
15 valorNivelDeExposicion
16 nivelDeRuido
17 valorDeNivelDeRuido
18 maquillajeDeOjos
19 maquillajeDeLabios
20 accesorios
21 frenteOcluida
22 ojoOcluido
23 bocaOcluida

OK Cancel Help

Figura 2.44 Spoon: Retirar datos que no son requeridos

En la Figura 2.45 se muestra la eliminación de columnas que no tienen un valor representativo para la predicción, se excluye bigote barba porque son niños.

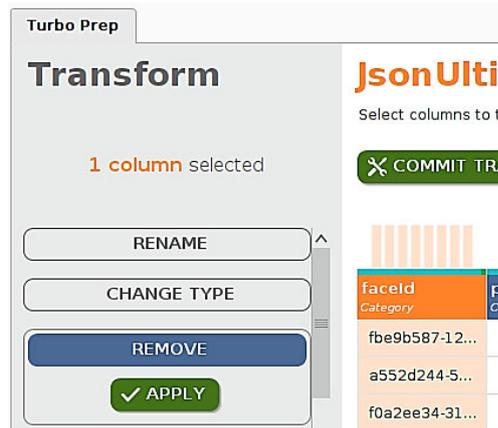


Figura 2.45 RM-TurboPrep:Remover columnas

2. En la Figura 2.46 se presenta la verificación para que no se repitan los faceld y el nombre del estudiante para nombrar a las imágenes con estos atributos.

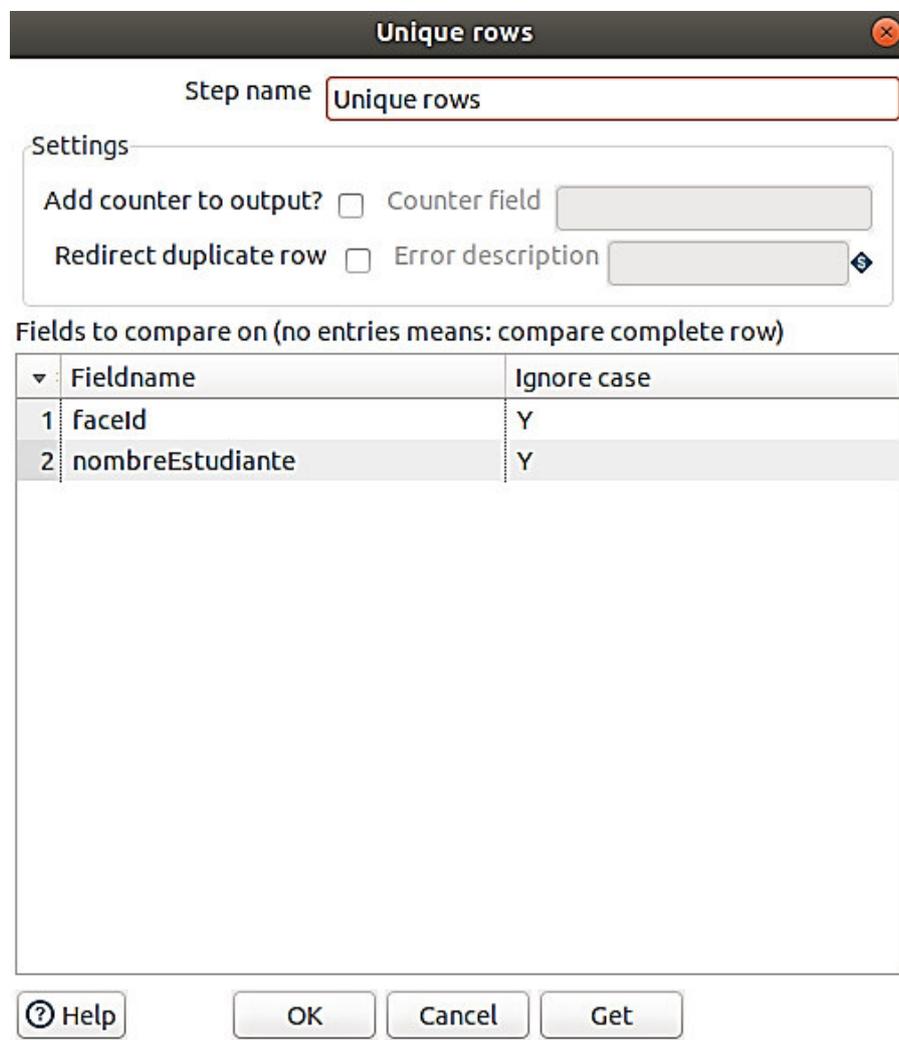


Figura 2.46 Spoon: Filas únicas

En ciertas ocasiones fue necesario el apoyo de la herramienta office Excel, Figura 2.47.

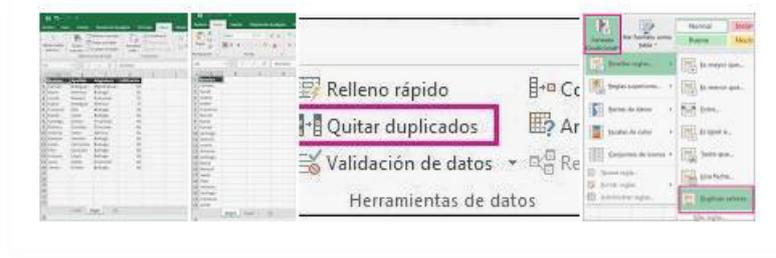


Figura 2.47 Excel: Quitar duplicados

3. En la Figura 2.48 se puede apreciar la sentencia donde se incluyen id de fases en los registros que no tienen este dato.

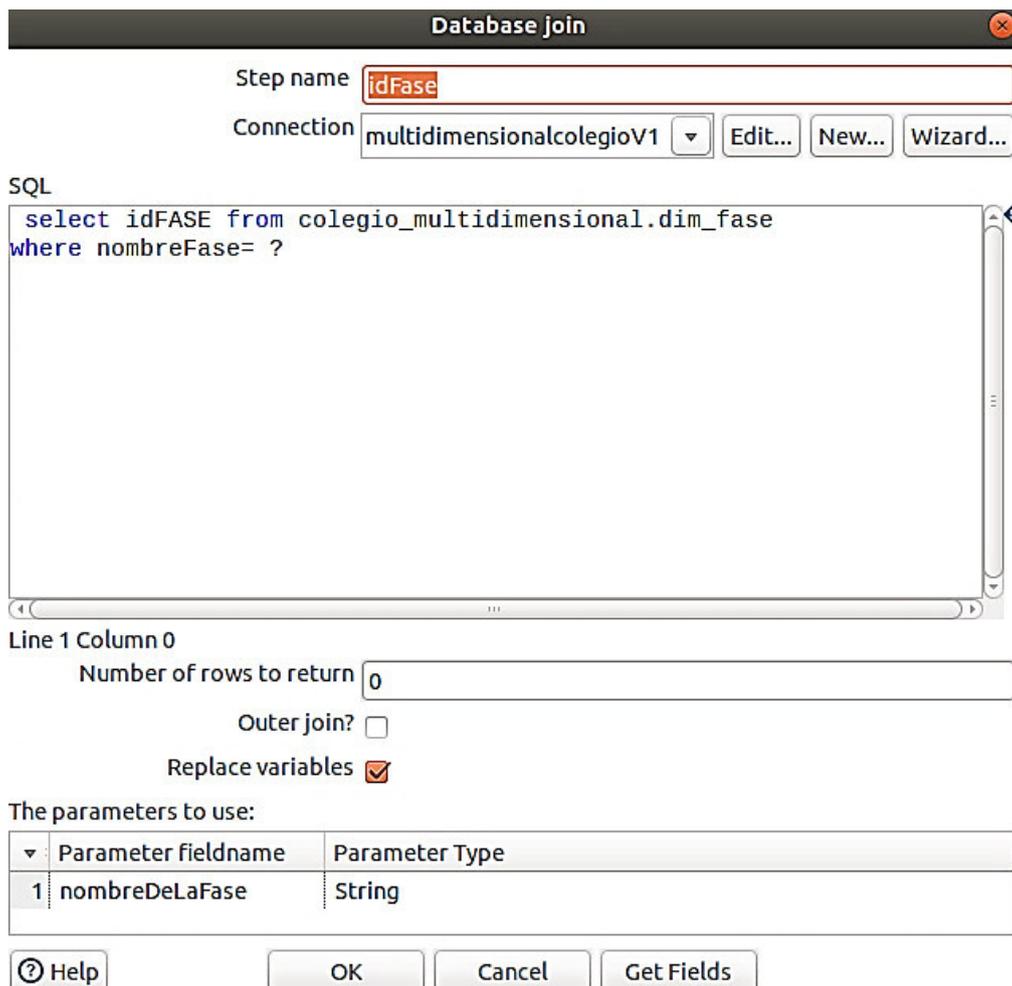


Figura 2.48 Spoon: Database join para selección del id de la fase conociendo el nombre de la misma

4. En la Figura 2.49 se presenta el filtrado de los registros del archivo Json donde la edad del rostro sea mayor a 20 y donde el "faceld" detectado sea "n.a."

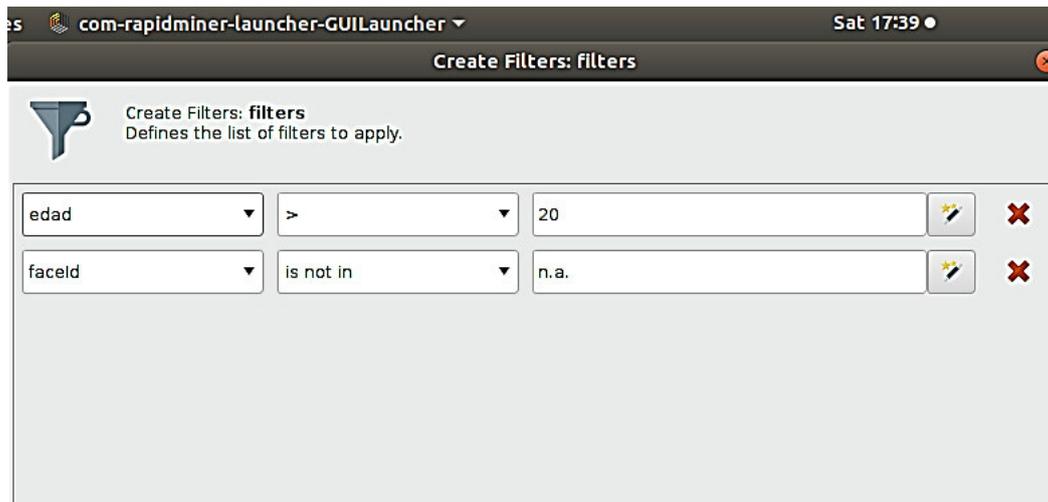


Figura 2.49 RM: Descartar faceld en “n.a.”

En la Figura 2.50 se observa el filtrado de los valores de “n.a.”.

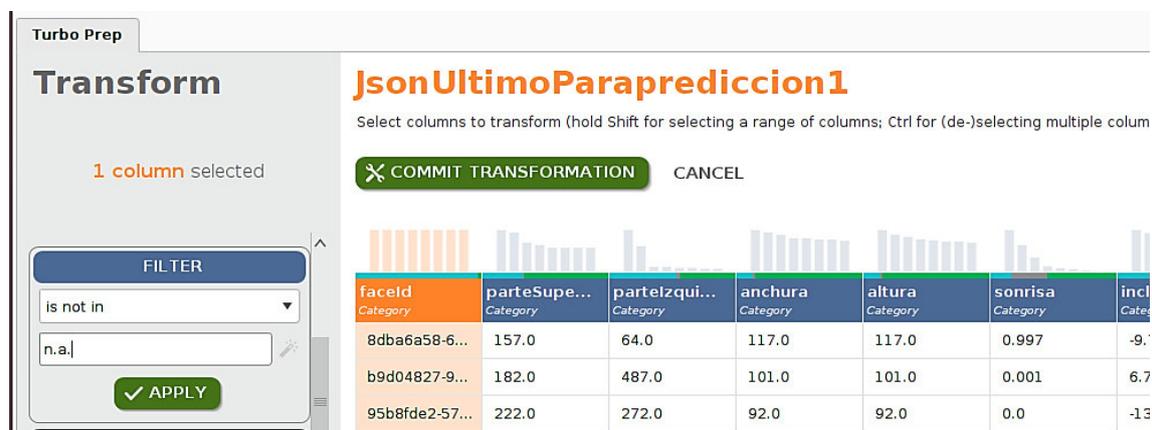


Figura 2.50 RM - Turbo Prep: Descartar faceld en “n.a.”

En la Figura 2.51 indica el filtrado de los faceld que si presentan un código.



Figura 2.51 Spoon: Descartar faceld en “n.a.”

Cuando no se dispone de la edad y género del estudiante se considera un promedio o moda en la edad detectada por el rostro, Figura 2.52.

	A	B	C	D	E
1	12			31	10
2	10			24	11
3	9			17	12
4	12			13	7
5	16			3	14
5	12			2	16
7	6			1	18
3	13			1	5
3	16			5	6
0	13			4	7
1	11			9	8
2	7			9	19

Figura 2.52 Cálculo de moda en edad de la persona

5. En la Figura 2.53 se muestra la actualización de la tabla dim_persona que se llena con la edad precisa de los estudiantes en la tabla persona.

Update ✖

Step name

Connection

Target schema

Target table

Commit size

Use batch updates?

Skip lookup

Ignore lookup failure? Flag field (key found)

The key(s) to look up the value(s):

▼	Table field	Comparator	Stream field1	Stream field	Get fields
1	nombresCompleto	=	nombreEstudiante		
2					

Update fields:

▼	Table field	Stream field	Get update fields
1	edad	edad	

Figura 2.53 Spoon: Actualización del dato edad

6. En la Figura 2.54 se expone el reemplazo de los valores null del atributo calva por el valor de cero.



Figura 2.54 Spoon: Reemplazo en cadena

7. En la Figura 2.55 se indica el filtrado para descartar registros de adultos.

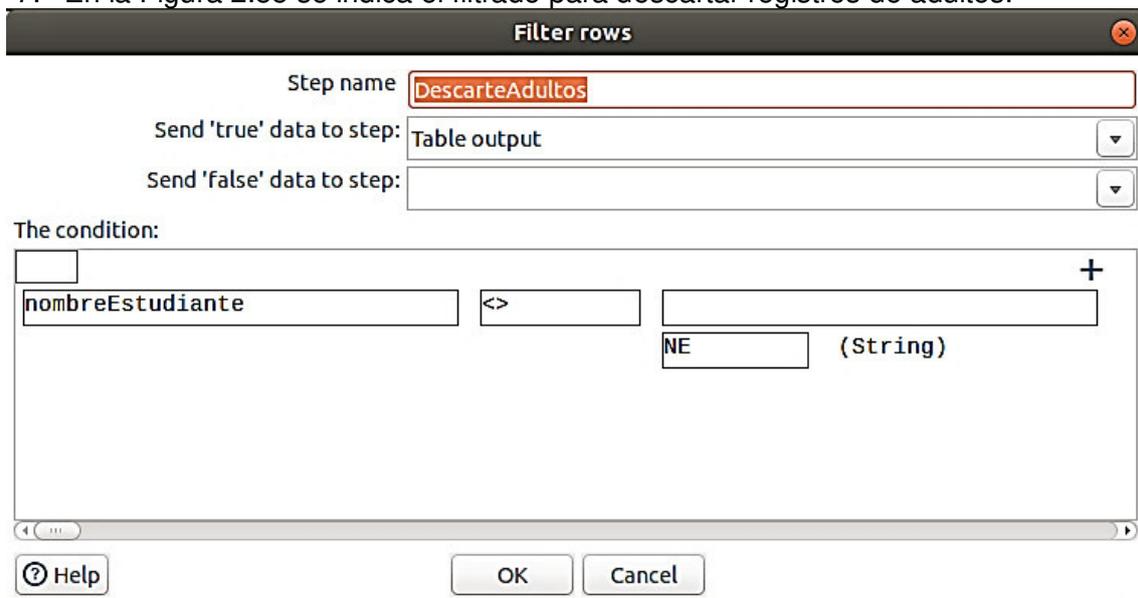


Figura 2.55 Spoon: Filtrado de filas para descartar adultos

En la Figura 2.56 se muestra la eliminación de los registros que no tienen nombre ya que los que no tienen nombre corresponden a personas adultas en las que no se ingresó el nombre.

Turbo Prep

Transform

1 column selected

COPY

FILTER

is not in

NULL

✓ APPLY

RANGE

JsonUltimoParaprediccion1

Select columns to transform (hold Shift for selecting a range of columns; Ctrl for (de-)selecting multiple columns)

✗ COMMIT TRANSFORMATION CANCEL

nombre	nombreDe... Category	grupo Number	photoID Category	imageURL Category	nombreEst... Category	expe Catego
	Explorar	4	5e25a2c444...	https://fireba...	Rosero Esco...	neut
	Explorar	4	5e25a2c444...	https://fireba...	Huang Zhou ...	neut
	Explorar	4	5e25a2c444...	https://fireba...	Ruiz Castillo ...	neut
	Explorar	4	5e25a2d844...	https://fireba...	Rosero Esco...	neut
	Explorar	4	5e25a2d844...	https://fireba...	Huang Zhou ...	neut

Figura 2.56 RM-Turbo Prep: Descarte de adultos

No fue visible el cabello o no presenta calva, por lo cual se reemplaza null por 0, Figura 2.57.

Turbo Prep

Transform

1 column selected

SAMPLE

SORT

REPLACE

NULL

0

Use regular expressions

✓ APPLY

JsonUltimoParaprediccion1

Select columns to transform (hold Shift for selecting a range of columns)

✗ COMMIT TRANSFORMATION CANCEL

bocaOcluido Number	bocaOcluida Number	calva Category	invisible Category	colorCabello Category
0	0	0	NULL	N
0	0	0	NULL	N
0	0	0	NULL	N
0	0	0	NULL	N
0	0	0	NULL	N

1,931 rows - 4

Figura 2.57 RM-TurboPrep: Reemplazo de valores nulos por 0

En la Figura 2.58 se aprecia el reemplazo por 0 en los registros donde no se detecta el color de cabello.

Transform

1 column selected

SORT

REPLACE

{} x

0 x

Use regular expressions

APPLY

JsonUltimoParaprediccion1

Select columns to transform (hold Shift for selecting a range of columns; Ctrl for (de-)

COMMIT TRANSFORMATION CANCEL

bocaOcluido	bocaOcluida	calva	invisible	colorDeCa...
Number	Number	Category	Category	Category
0	0	0	0	NULL
0	0	0	0	NULL
0	0	0	0	NULL
0	0	0	0	NULL
0	0	0	0	NULL

Figura 2.58 RM-Turbo Prep: Reemplazo de caracteres especiales por 0

En la Figura 2.59 se indica el reemplazo de null en accesorios por ND - No detectado.

REPLACE

NULL x

ND x

Use regular expressions

APPLY

SPLIT

valorNivel...	nivelDeRuido	valorDeNiv...	maquillaje...	maquillaje...	accesorios	fr
Category	Category	Category	Category	Category	Category	Category
0.46	low	0.0	1	0	NULL	C
0.24	low	0.0	1	1	NULL	1
0.45	low	0.22	0	0	NULL	C
0.33	low	0.0	0	0	NULL	C
0.13	medium	0.68	1	1	NULL	C

Figura 2.59 RM-Turbo Prep: Reemplazo de valores nulos por ND- No detectado

En la Figura 2.60 se muestra la corrección del tipo de lentes.

Transform

1 column selected

FILTER

RANGE

SAMPLE

SORT

REPLACE

SwimmingGoggles x

SwimmingGlasses x

Use regular expressions

APPLY

JsonUltimoParaprediccion1

Select columns to transform (hold Shift for selecting a range of columns; Ctrl for (de-)se

COMMIT TRANSFORMATION CANCEL

enfado	temor	genero	lentes
Category	Category	Category	Category
0.0	0.0	male	ReadingGlasses
0.001	0.0	female	NoGlasses
0.0	0.0	female	NoGlasses
0.0	0.0	male	ReadingGlasses
0.0	0.0	female	NoGlasses
0.0	0.0	female	NoGlasses
0.0	0.0	male	ReadingGlasses
0.0	0.0	female	NoGlasses
0.0	0.0	female	NoGlasses
0.0	0.0	male	ReadingGlasses

Figura 2.60 RM-TurboPrep: Reemplazo de un mismo tipo de lentes

8. En RM-Turbo Prep al encontrar valores perdidos, éstos son representados con el carácter "?", en la Figura 2.61 se puede ver que a estos valores se los reemplaza con un valor de 0.

Transform

1 column selected

MySQL Workbench

CHANGE TYPE

REMOVE

COPY

FILTER

RANGE

GruposModeloV1

Select columns to transform (hold Shift for selecting a range of columns; Ctrl for (de-)selecting multiple columns; Alt to select all columns of the same t...)

COMMIT TRANSFORMATION CANCEL

photoID Category	imageURL Category	nombreE... Category	experienciaUsuario Category	experienciaUsuarioNumerica Number	fecha_hora Category	Promedio... Category
sesionid_15...	https://deu...	Galindo Ro...	neutral	0.676	2020-01-13...	0.0002
sesionid_15...	https://deu...	Freire Proa...	positiva	1	2020-01-13...	0.0
sesionid_15...	https://deu...	Freire Proa...	neutral	0.995	2020-01-13...	0.001
sesionid_15...	https://deu...	Freire Proa...	neutral	0.998	2020-01-13...	0.0
sesionid_15...	https://deu...	NULL	NULL	?	2020-01-13...	NULL

Figura 2.61 RM-Turbo Prep: Caracteres Perdidos

En la Figura 2.62 se indica la funcionalidad "Replace".

Data Sets

+ LOAD DATA

GruposModeloV1

//Local Repository/data/ModeloPrediccionV1/GruposModeloV1

Rows: 1,983

Columns: 48

Last Change: None

TRANSFORM CLEANSE GENERATE MERGE

Replace

Transformations

Cleansing

Show Details...

Sort View (Ascending)

Sort View (Descending)

Cancel Sorting

lentes Category	enfado Cat	desprecio Category	temor Category	felicidad Category	neutral Category	tristeza Category	sorpresa Category
NoGlasses	0	0	0	0.003	0.676	0.001	0.32
NoGlasses	0	0	0	1	0	0	0
SwimmingGo...	0	0	0	0	0.995	0.005	0
ReadingGlas...	0	0	0	0	0.998	0	0.001
n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

1,983 rows - 48 columns (41 nominal, 7 numerical)

Figura 2.62 RM-TurboPrep: Replace Missing Values

Se verifica que se han reemplazado los valores perdidos por el 0

Cleanse

1 column selected

REMOVE CORRELATED

REPLACE MISSING

NORMALIZATION

DISCRETIZATION

DUMMY ENCODING

GruposModeloV1

Select a column to clean (hold Shift for selecting a range of columns; Ctrl for (de-)selecting multiple columns; Alt to select all columns of the

COMMIT CLEANSE CANCEL

photoID Category	imageURL Category	nombreE... Category	experienciaUsuario Category	experienciaUsuarioNumerica Number	fecha_hora Category
sesionid_15...	https://deu...	Freire Proa...	positiva	1	2020-01-13...
sesionid_15...	https://deu...	Freire Proa...	neutral	0.995	2020-01-13...
sesionid_15...	https://deu...	Freire Proa...	neutral	0.998	2020-01-13...
sesionid_15...	https://deu...	NULL	NULL	0	2020-01-13...
sesionid_15...	https://deu...	Freire Proa...	positiva	0.997	2020-01-13...

Figura 2.63 RM-Turbo Prep: Resultado del uso de Replace Missing Values

9. Al tener ya los datos disponibles se procede al cálculo de la experiencia de usuario con ose observa en la Figura 2.64.

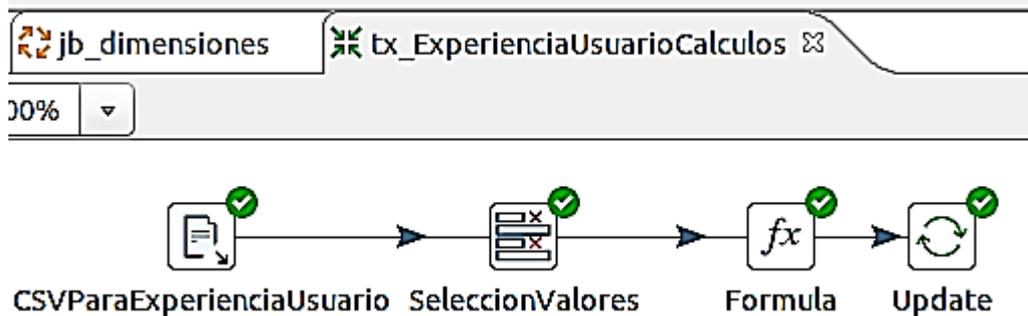


Figura 2.64 Spoon: Transformación para determinar la experiencia de usuario

En la Figura 2.65, se observa la actualización de los datos de predicción obtenidos de RM-Turbo Prep.

The 'Update' dialog box is shown with the following configuration:

- Step name: Update
- Connection: colegioV1
- Target schema: (empty)
- Target table: JSON
- Commit size: 100
- Use batch updates?:
- Skip lookup:
- Ignore lookup failure?: Flag field (key found): (empty)

The key(s) to look up the value(s):

Table field	Comparator	Stream field1
1 desprecio	=	desprecio
2 grupo	=	grupo
3 neutral	=	neutral
4 nombreEstudiante	=	nombreEstudiante

Update fields:

Table field	Stream field
1 prediccionRandomForest	prediction(experienciaUsuario)

Figura 2.65 Spoon: Actualización de la predicción

En la Figura 2.66 se muestra el proceso ETL planteado para la Carga Inicial.

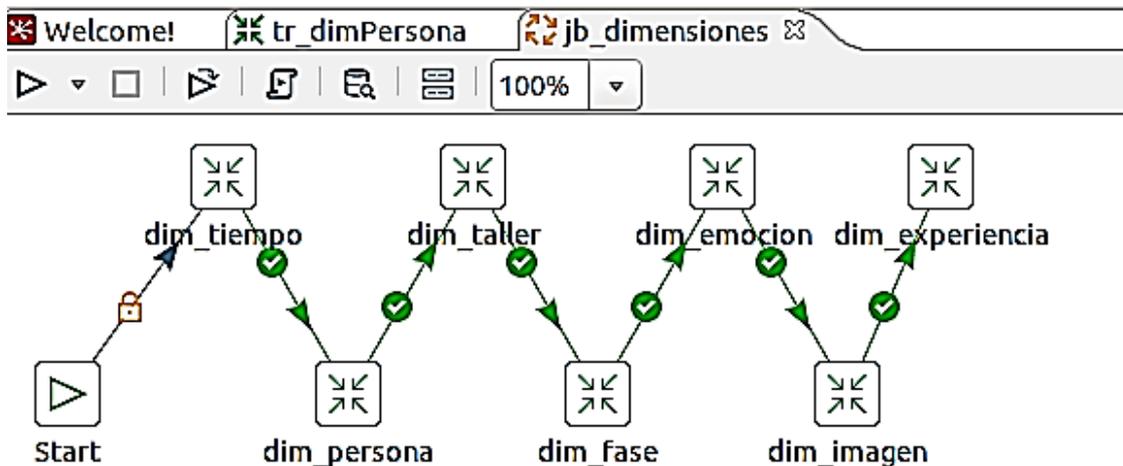


Figura 2.66 Taller RE, Carga Inicial

Las tareas que lleva a cabo este proceso son:

- Inicio: inicia la ejecución de los pasos en el momento en que se le indique.
- Carga de Dimensión TIEMPO: ejecuta el contenedor de pasos que cargará la dimensión TIEMPO, más adelante se detallará el mismo.
- Carga de Dimensión PERSONA: ejecuta el contenedor de pasos que cargará la dimensión PERSONA, más adelante se detallará el mismo.
- Carga de Dimensión TALLER: ejecuta el contenedor de pasos que cargará la dimensión TALLER, más adelante se detallará el mismo.
- Carga de Dimensión FASE: ejecuta el contenedor de pasos que cargará la dimensión FASE, más adelante se detallará el mismo.
- Carga de Dimensión EMOCION: ejecuta el contenedor de pasos que cargará la dimensión EMOCION, más adelante se detallará el mismo.
- Carga de Dimensión IMAGEN: ejecuta el contenedor de pasos que cargará la dimensión IMAGEN, más adelante se detallará el mismo.
- Carga de Dimensión EXPERIENCIA: ejecuta el contenedor de pasos que cargará la dimensión EXPERIENCIA, más adelante se detallará el mismo.

En la Figura 2.67 se presentan las tareas llevadas a cabo por "CargadeDimensiónTIEMPO".

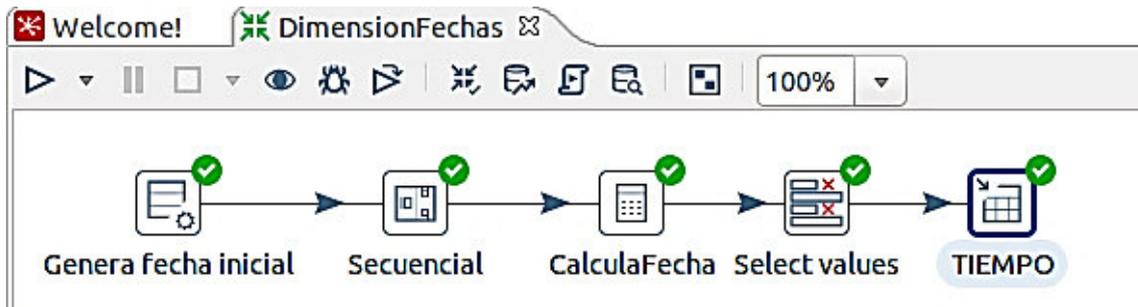


Figura 2.67 Taller RE, Carga de Dimensión TIEMPO

Este paso es un contenedor de pasos, así que incluye las siguientes tareas:

- Genera fecha inicial: Paso que permite generar filas de tipo fecha con el formato yyyy-MM-dd y el valor inicial depende de la fecha de inicio de los datos existentes.
- Secuencial: Genera un número secuencial para la fecha.
- Calcula Fecha: calcula la fecha considerando como datos de entrada la fecha inicial del paso "Genera fecha inicial", realiza la operación suma de días, adicionando los días indicados en el paso "Secuencial". Separa los años, meses y días para definir la granularidad.
- Select values: Para seleccionar los campos requeridos como son: "fecha_calculada", "anio_calculado", "mes_calculado" y "dia_calculado".
- Tiempo: almacena en la tabla de dimensión TIEMPO los datos obtenidos en el paso anterior.

En la Figura 2.68 se observan las tareas llevadas a cabo por "Carga de Dimensión PERSONA".

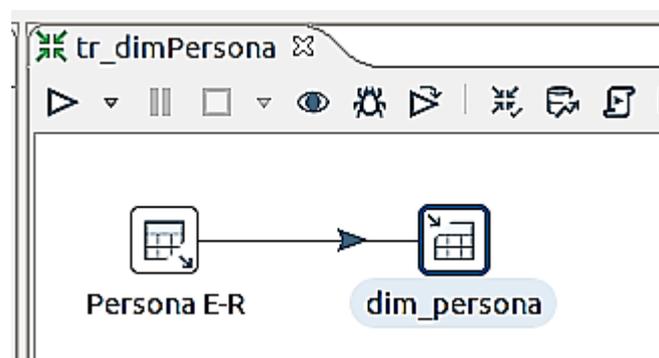


Figura 2.68 Taller RE, Carga de Dimensión PERSONA

Este paso es un contenedor de pasos, así que incluye las siguientes tareas:

- Persona E-R: obtiene a través de una consulta SQL los datos del OLTP necesarios para cargar la dimensión PERSONA.

En la Figura 2.69 , se expone la sentencia SQL que contiene este paso:

```
select nombres,apellidos,nombresCompleto,s, idPersona, edad,
genero,tipo from colegioV1.PERSONA
where tipo like '%GR%'
```

Figura 2.69 Spoon: Obtener datos del estudiante del OLTP

- Dim_persona: almacena en la tabla de dimensión DIM_PERSONA los datos obtenidos en el paso anterior.

En la Figura 2.70 se pueden ver las tareas llevadas a cabo por "Carga de Dimensión TALLER".



Figura 2.70 Taller RE, Carga de Dimensión TALLER

Este paso es un contenedor de pasos, así que incluye las siguientes tareas:

- AtributosTaller: obtiene a través de una consulta SQL los datos del OLTP necesarios para cargar la dimensión TALLER.

En la Figura 2.71, se expone la sentencia SQL que contiene este paso:

```
SELECT
idTALLER, nombreTaller, competencias, capacidad, indicador, fechaInicio, fechaFin
FROM TALLER
```

Figura 2.71 Spoon: Obtener datos del taller del OLTP

- SeleccionValores: selecciona los valores necesarios que se obtuvieron en el paso anterior, cambia al tipo de dato a fecha de las fechas de inicio y fin del taller y permite parametrizar su formato.
- Dim_taller: almacena en la tabla de dimensión DIM_TALLER los datos obtenidos en el paso anterior.

En la Figura 2.72, se muestran las tareas llevadas a cabo por "Carga de Dimensión FASE". Este paso es un contenedor de pasos, así que incluye las siguientes tareas:

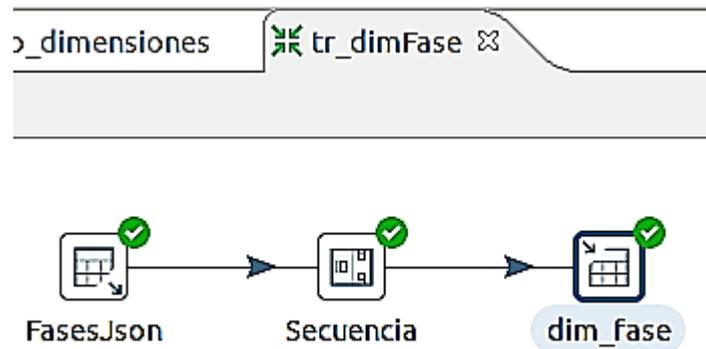


Figura 2.72 Taller RE, Carga de Dimensión FASE

- FasesJson: obtiene a través de una consulta SQL los datos del OLTP necesarios para cargar la dimensión FASE.

En la Figura 2.73 se expone la sentencia SQL que contiene este paso:

```
select distinct(nombreDeLaFase) from colegioV1.JSON
```

Figura 2.73 Spoon: Obtener datos de las fases del OLTP

- Secuencia: genera un número de secuencia para la fase.
- Dim_fase: almacena en la tabla de dimensión DIM_FASE los datos obtenidos en el paso anterior.

En la Figura 2.74 se muestran las tareas llevadas a cabo por "Carga de Dimensión IMAGEN". Este paso es un contenedor de pasos, así que incluye las siguientes tareas:

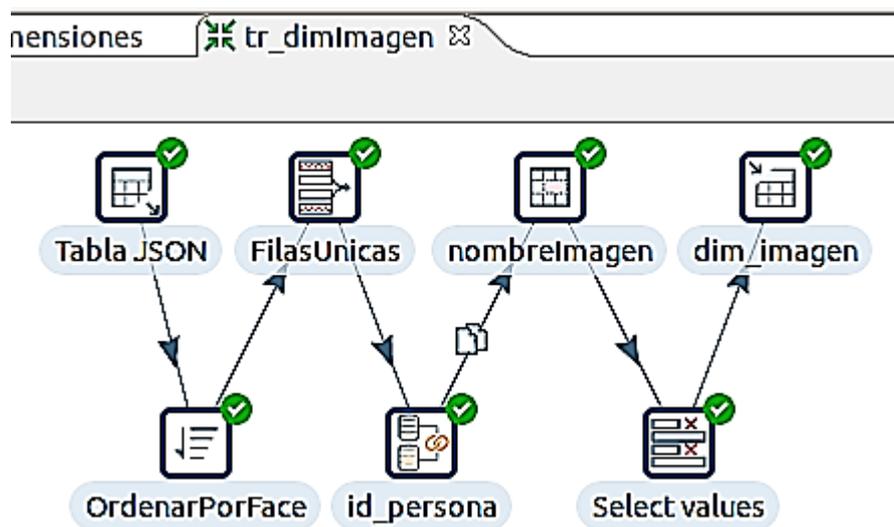


Figura 2.74 Taller RE, Carga de Dimensión IMAGEN

- Tabla JSON: obtiene a través de una consulta SQL los datos del OLTP necesarios para cargar la dimensión IMAGEN.

En la Figura .75 se expone la sentencia SQL que contiene este paso:

```

select *
FROM colegioV1.JSON
  where faceId <> 'n.a.' and nombreEstudiante is not null
and nombreEstudiante <>'NE'

```

Figura 2.75 Spoon: Obtener datos para la imagen del OLTP

- Sort rows: Ordena las filas del paso anterior por el campo faceld.
- Unique rows: Elimina las filas duplicadas del flujo de entrada y filtra solo las filas únicas del paso anterior por faceld y nombre del estudiante.
- Id_persona: Busca el id persona de la tabla dimensión persona teniendo como parámetro el nombre completo del estudiante.
- nombrelimagen: Concatena los campos "idpersona", "faceld" y "photold" intercalados por un guión bajo para designar un nombre que identifique a la fotografía.
- Select values: selecciona los valores faceld, nombrelimagen y lentes del paso anterior.
- Dim_imagen: almacena en la tabla de dimensión DIM_IMAGEN los datos obtenidos en el paso anterior.

En la Figura 2.76 se muestran las tareas llevadas a cabo por "Carga de Dimensión EMOCION". Este paso es un contenedor de pasos, así que incluye las siguientes tareas:

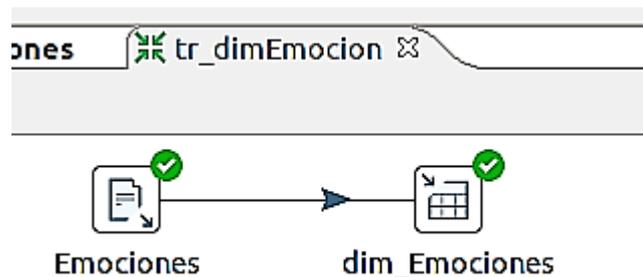


Figura 2.76 Taller RE, Carga de Dimensión Emociones

- Emociones: Lee los datos de las emociones de un archivo csv.
- Dim_Emociones: almacena en la tabla de dimensión DIM_EMOCIONES los datos obtenidos en el paso anterior.

A continuación, se especifican las tareas llevadas a cabo por "Carga de Dimensión EXPERIENCIA". Este paso es un contenedor de pasos, así que incluye las siguientes tareas:

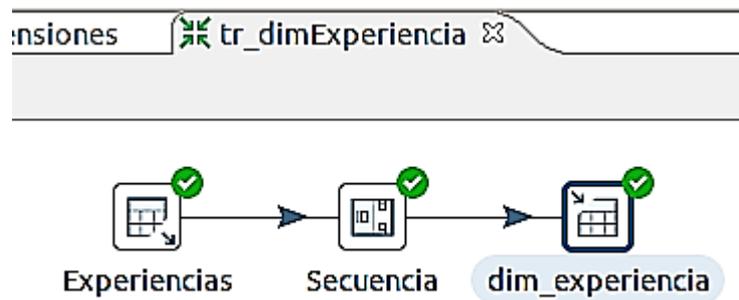


Figura 2.77 Taller RE, Carga de Dimensión Experiencia

- Experiencias: obtiene a través de una consulta SQL los datos del OLTP necesarios para cargar la dimensión experiencia.

En la Figura 2.78 se expone la sentencia SQL que contiene este paso:

```
select distinct(experienciaUsuario) from colegioV1.JSON
where experienciaUsuario is not null
```

Figura 2.78 Spoon: Obtener datos para la experiencia de usuario del OLTP

- Secuencia: genera un número de secuencia para la experiencia de usuario.
- Dim_experiencia: almacena en la tabla de dimensión DIM_experiencia los datos obtenidos en el paso anterior.

En la Figura 2.79 se muestran las tareas llevadas a cabo por "Carga de Tabla de Hechos H_REGISTROTALLER".

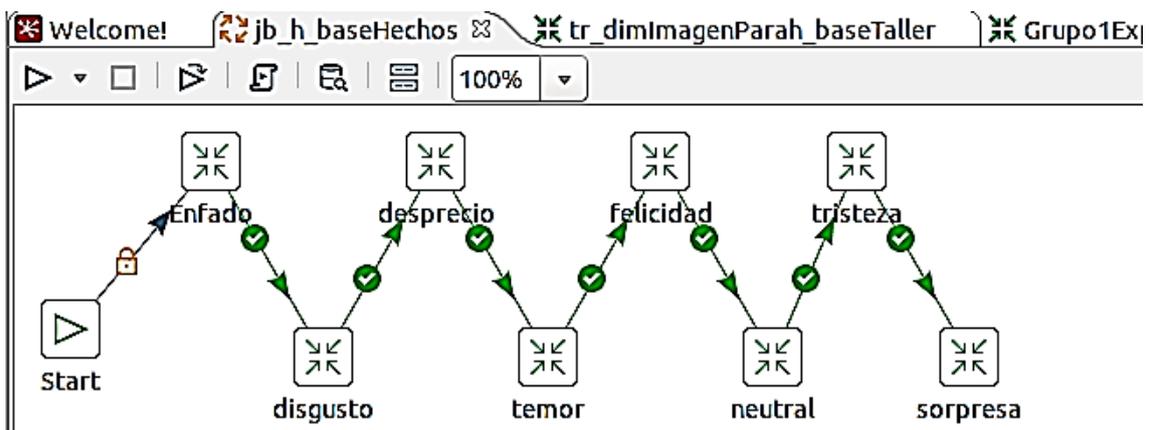


Figura 2.79 Taller RE, Trabajo para la carga de Tabla de Hechos h_registroTaller

Este paso es un contenedor de pasos, así que incluye las siguientes tareas

- Start: inicia la ejecución de los pasos en el momento en que se le indique.
- Enfado: ejecuta el contenedor de pasos que carga los datos de los registros que presentan la emoción enfado, más adelante se detallará el mismo.
- Disgusto: ejecuta el contenedor de pasos que carga los datos de los registros que presentan la emoción disgusto, más adelante se detallará el mismo.

- Desprecio: ejecuta el contenedor de pasos que carga los datos de los registros que presentan la emoción desprecio, más adelante se detallará el mismo.
- Temor: ejecuta el contenedor de pasos que carga los datos de los registros que presentan la emoción temor, más adelante se detallará el mismo.
- Felicidad: ejecuta el contenedor de pasos que carga los datos de los registros que presentan la emoción temor, más adelante se detallará el mismo.
- Neutral: ejecuta el contenedor de pasos que carga los datos de los registros que presentan la emoción neutral, más adelante se detallará el mismo.
- Tristeza: ejecuta el contenedor de pasos que carga los datos de los registros que presentan la emoción tristeza, más adelante se detallará el mismo.
- Sorpresa: ejecuta el contenedor de pasos que carga los datos de los registros que presentan la emoción sorpresa, más adelante se detallará el mismo.

En la Figura 2.80 se pueden ver las tareas llevadas a cabo en la transformación “tr_enfado” para la carga de la “Tabla de Hechos h_registroTaller”.

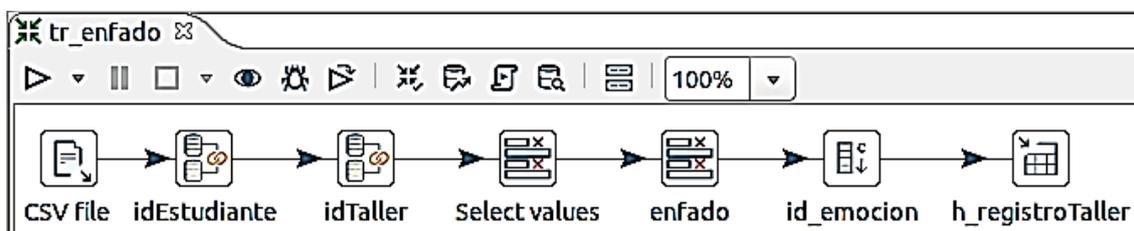


Figura 2.80 Taller Re, Estructura de transformaciones para la carga a la Tabla de Hechos H_REGISTROTALLER

Este paso es un contenedor de pasos que funciona de manera similar para todas las emociones, así que incluye los siguientes:

- CVS file: Lee datos del archivo CSV que contiene todos los registros obtenidos de emociones.
- idEstudiante: Obtiene el idPersona de un estudiante mediante el nombre del estudiante de la tabla DIM_PERSONA.
- idTaller: Obtiene el idTaller de un estudiante mediante el idEstudiante de la tabla registroT.
- SelectValues: Selecciona las emociones de disgusto, desprecio, enfado, felicidad, sorpresa, temor, neutral, tristeza, así como también los valores obtenidos de los pasos anteriores faceld, fase, grupo, taller y fecha_hora.

- Enfado: Selecciona la emoción que corresponde, formatea los datos en el caso de fecha al formato yyyy-MM-dd además de parametrizar la longitud y tipo del idEstudiante.
- Id_emocion: agrega un valor constante a cada emoción.
- h_registroTaller: almacena en la tabla de hechos H_REGISTROTALLER los datos obtenidos en el paso anterior.

En la Figura 2.81 se exponen las tareas llevadas a cabo por "Carga de Tabla de Hechos H_PREDICCION".

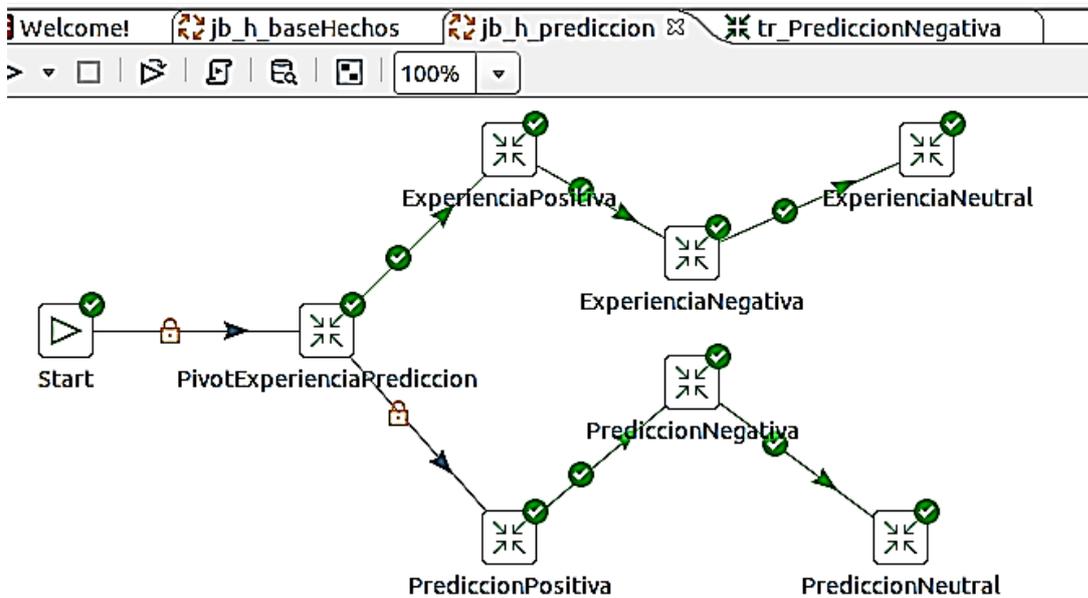


Figura 2.81 Taller RE, Trabajo para la carga de Tabla de Hechos h_prediccion

Este paso es un contenedor de pasos, así que incluye las siguientes tareas:

- Start: inicia la ejecución de los pasos en el momento en que se le indique.
- PivotExperienciaPrediccion: ejecuta el contenedor de pasos que carga los datos de los registros que presentan la experiencia de usuario y la predicción de la misma, más adelante se detallará el mismo.
- ExperienciaPositiva: ejecuta el contenedor de pasos que carga los datos de los registros que presentan la experiencia de usuario positiva, más adelante se detallará el mismo.
- ExperienciaNegativa: ejecuta el contenedor de pasos que carga los datos de los registros que presentan la experiencia de usuario negativa, más adelante se detallará el mismo.
- ExperienciaNeutral: ejecuta el contenedor de pasos que carga los datos de los registros que presentan la experiencia de usuario neutral, más adelante se detallará el mismo.

- PrediccionPositiva: ejecuta el contenedor de pasos que carga los datos de los registros que presentan la predicción de usuario positiva, más adelante se detallará el mismo.
- PrediccionNegativa: ejecuta el contenedor de pasos que carga los datos de los registros que presentan la predicción de usuario negativa, más adelante se detallará el mismo.
- PrediccionNeutral: ejecuta el contenedor de pasos que carga los datos de los registros que presentan la predicción de usuario neutral, más adelante se detallará el mismo.

En la Figura 2.82 se muestran las tareas llevadas a cabo en la transformación "tr_PrediccionPivot" para la carga de la "Tabla de Hechos h_prediccion". Este paso es un contenedor de pasos que funciona de manera similar para todas las emociones, así que incluye los siguientes:

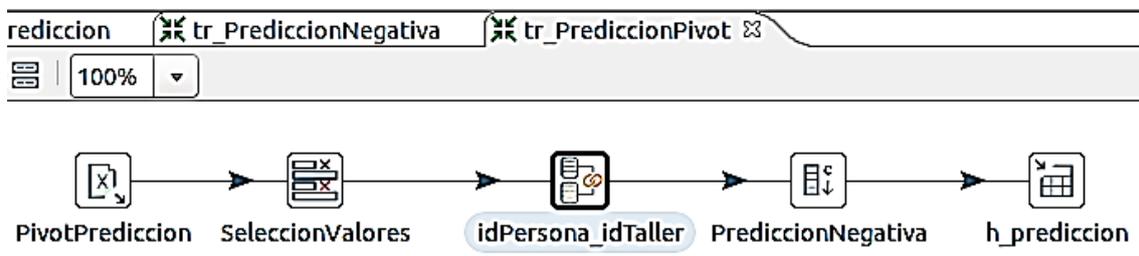


Figura 2.82 Taller RE, Transformación para la carga pivot de la experiencia y predicción del usuario

- PivotPrediccion: Lee datos del archivo xlsx que contiene todos los registros obtenidos de la experiencia del usuario y su predicción.
- SeleccionValores: Selecciona el faceld, experienciaUsuarioPositiva, experienciaUsuarioNegativa, prediccionUsuarioNeutral, prediccionUsuarioPositiva, fase, nombreEsudiante y fecha_hora. Cambia las etiquetas de las columnas y formatea la fecha.
- IdPersona_idTaller: Obtiene mediante una sentencia SQL el idPersona y el idTaller de un estudiante mediante el idEstudiante de la tabla registroT.

En la Figura 2.83 se expone la sentencia SQL que contiene este paso:

```
select p.idPersona,rt.taller from multiV1.dim_persona p
inner join colegioV1.REGISTROT rt on p.idPersona=rt.idEstudiante
where p.nombresCompleto=?
```

Figura 2.83 Spoon: Obtener idPersona e idTaller para h_prediccion

- prediccionNegativa: agrega un valor constante de 0 a la predicción negativa ya que se requiere para cálculos y gráficos. Cabe mencionar que no se aprecia experiencia Negativa.
- h_prediccion: almacena en la tabla de hechos H_PREDICCION los datos obtenidos en el paso anterior.

En la Figura 2.84 se observan las tareas llevadas a cabo en la transformación “tr_ExperienciaPositiva” para la carga de la “Tabla de Hechos h_registroTaller”.

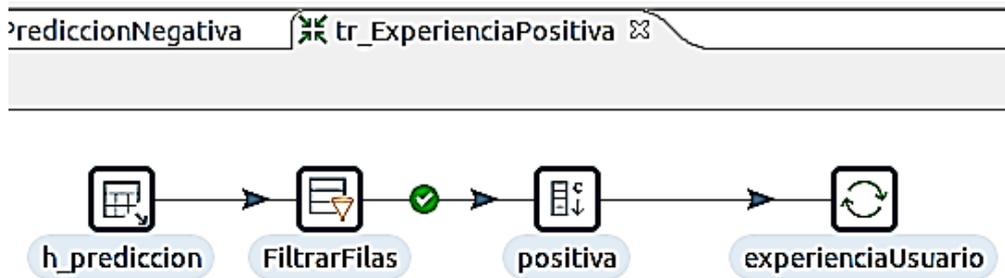


Figura 2.84 Taller RE, Transformación para la experiencia y predicción del usuario

Este paso es un contenedor de pasos que funciona de manera similar para experiencia y predicción, así que incluye los siguientes:

- h_prediccion: obtiene a través de una consulta SQL los datos del OLTP necesarios para cargar la dimensión experiencia.

En la Figura 2.85 se expone la sentencia SQL que contiene este paso:

```
SELECT dim_imagen_idImagen_ima, positiva, negativa, neutral, pPositiva, pNeutral, pNegativa
FROM multiV1.h_prediccion
```

Figura 2.85 Spoon: Obtener datos para la experiencia y predicción de la tabla h_prediccion

- FiltrarFilas: filtra las filas que presentan la experiencia o predicción que se elige en el siguiente paso.
- Positiva: agrega un valor constante a la experiencia o predicción de acuerdo con el código de la tabla de dimensiones dim_experiencia.
- experienciaUsuario: actualiza la experiencia o predicción del usuario por medio de la clave que es el idImagen de la tabla de dimensiones dim_imagen.

2.4.5 Simulación del Modelo de Predicción

En la Figura 2.86 se muestra la funcionalidad de simulación de RM que aprovecha la sabiduría de la información para obtener recomendaciones de operadores basadas en el modelo optimizado.

Primera Optimización:

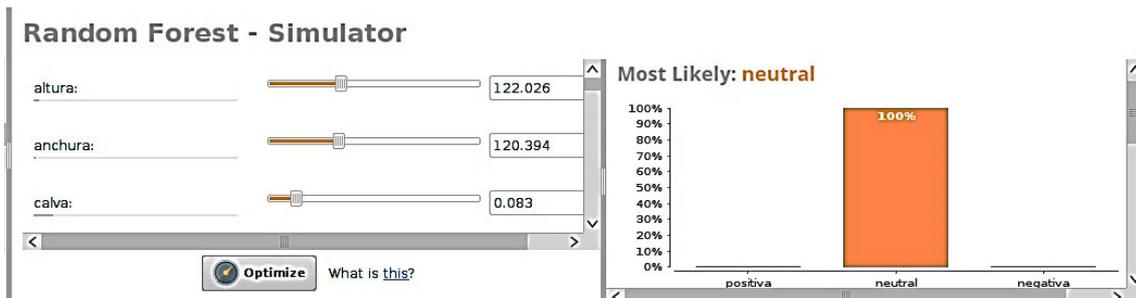


Figura 2.86 Simulación RM-Turbo Prep- Optimización 1

En este caso el promedio corresponde al rostro de un estudiante de dimensiones 122 x120 y que relativamente no tiene calvicie.

El escenario más probable, según el gráfico de barras superior a la derecha, es que el estudiante presentara una emoción neutral en el taller de RE.

En la Figura 2.87 se muestran los factores importantes para que se encuentre la emoción neutral presente, se explica lo que está en su contra: sobre todo, es su experiencia de usuario y la rotación que presenta el rostro, que se muestran como barras verdes. Verde en este contexto implica que la clase experiencia de usuario y la rotación que presenta el rostro está de acuerdo con la predicción de experiencia de usuario, es decir, "neutral". Las barras rojas para la sonrisa y el nivel de difuminación de la imagen implican un desacuerdo con la predicción y, por lo tanto, una correlación positiva con la experiencia de usuario.

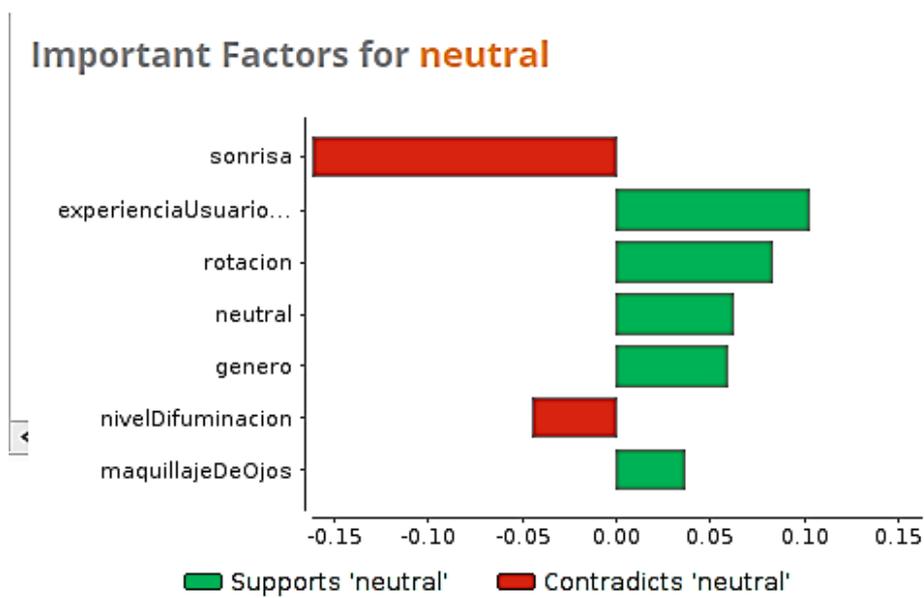


Figura 2.87 Correlación Positiva y Negativa para Emoción Neutral

Segunda Optimización

Para este caso, en la Figura 2.88 se indica el promedio que corresponde al grupo 2, fase Compartir o Explorar, género femenino, usa lentes que no presentará temor, disgusto o enfado.

El escenario más probable, según el gráfico de barras superior a la derecha, es que el estudiante presentará un mayor promedio en emoción neutral 83%, seguido de positiva en 17% en el taller de RE.

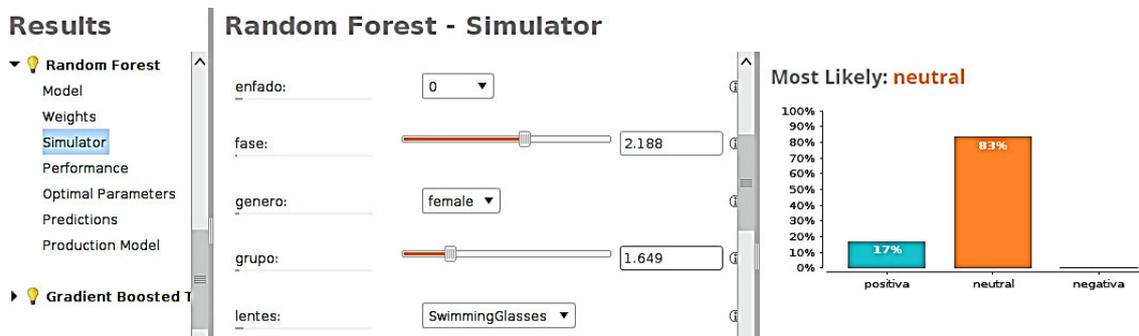


Figura 2.88 RM-TurboPrep- Optimización 2

En el presente caso el promedio corresponde al enfado 0, fase 2, género femenino, grupo 1 que presenta problemas de visión.

El escenario más probable, según el gráfico de barras superior a la derecha, es que el estudiante presentará un mayor promedio en emoción neutral 83%, seguido de positiva en 17% en el taller de RE.

Actualización

Cuando se haya cargado en su totalidad el DW, se deben establecer sus políticas y estrategias de actualización o refresco de datos.

Las políticas de actualización que se sugieren para que sean aplicadas por los usuarios son las siguientes:

La información se refrescará cada fin de mes a las doce de la noche. Los datos de las tablas de dimensiones "DIM_PERSONA", "DIM_MAGEN" y tabla de hechos "H_REGISTROTALLER" serán cargados totalmente cada mes, mientras que la carga a la tabla "H_PREDICCION" se recomienda realizarla anualmente. Los datos de la tabla de dimensión "DIM_TIEMPO" se cargarán de manera incremental teniendo en cuenta la fecha de la última actualización. Los datos de la tabla de hechos que corresponden al último mes (30 días) a partir de la fecha actual, serán agregados cada vez que se requiera. Estas acciones se realizarán durante un periodo de prueba, para analizar cuál

es la manera más eficiente de generar las actualizaciones, basadas en el estudio de los cambios que se producen en los OLTP y que afectan al contenido del DW.

2.5 Visualización

Olap Cubes (saiku)

La generación de los cubos se realiza utilizando la herramienta de la suite de Pentaho denominada Saiku. Esta herramienta permite crear esquemas con los cubos deseados, compuestos por dimensiones con sus respectivos atributos almacenados en el modelo multidimensional.

Se crea un cubo multidimensional que se llama "Cubo de RegistroTaller" y que está basado en el modelo lógico diseñado en el caso práctico de la metodología Hefesto.

Se selecciona la tabla de hechos "H_REGISTRO_TALLER" y las de dimensiones como se observa en la Figura 2.89.

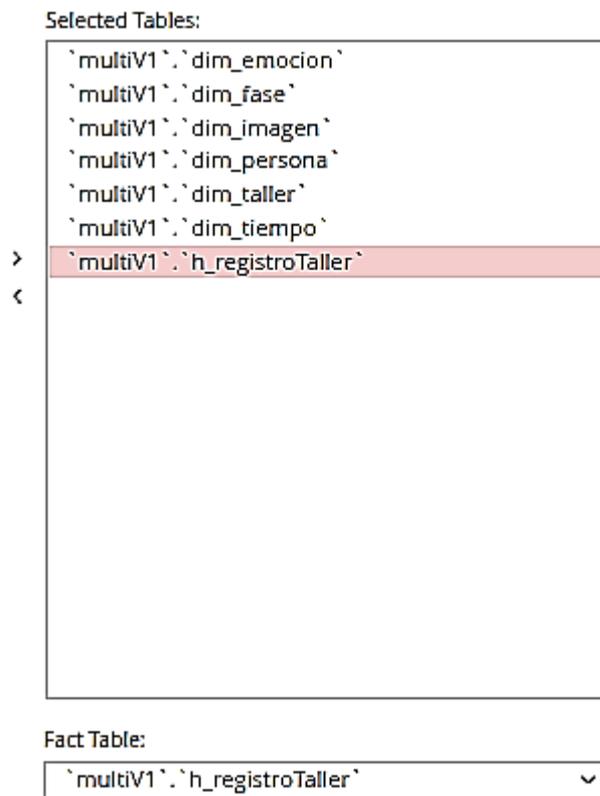


Figura 2.89 Saiku: Cubo Registro Taller

En la Figura 2.90 se indica los joins entre la tabla de hechos h_registroTaller y dimensiones:

Left Table: `multiV1`.`h_registroTaller`
 Right Table: `multiV1`.`dim_imagen`

Key Field: id_emocion, taller, idEstudiante, grupo, fase, valor, **faceldV**

Key Field: presentalentes, nombrelimagen, **id_imagen**

Join(s):
 `multiV1`.`h_registroTaller`.id_emocion - INNER JOIN - `multiV1`.`dim_emocion`.idEmoi
 `multiV1`.`h_registroTaller`.id_fechaHora - INNER JOIN - `multiV1`.`dim_tiempo`.idTiem
 `multiV1`.`h_registroTaller`.taller - INNER JOIN - `multiV1`.`dim_taller`.id_taller
 `multiV1`.`h_registroTaller`.idEstudiante - INNER JOIN - `multiV1`.`dim_persona`.idPers
 `multiV1`.`h_registroTaller`.fase - INNER JOIN - `multiV1`.`dim_fase`.nombreFase

Figura 2.90 Saiku: Relación Hechos y Dimensiones

Creación de Indicadores

En la Figura 2.91 se puede observar que se realiza un conteo del hecho “IdEstudiante” de la tabla de hechos “H_REGISTROTALLER”, para crear el indicador denominado “Total Estudiantes”.

Data Source Model Editor

Available: Dim tiempo (Anio tie, Dia tie, IdTiempo, Mes tie), H registroTaller (FaceldV, Fase, Grupo, Id emocion, Id fechaHora, **IdEstudiante**, Taller, Valor)

Measures: Fase, Grupo, Id emocion, IdEstudiante, Taller, Valor, **TotalEstudiantes**

Dimensions: Dim emocion, Dim fase, Dim imagen, Dim persona, Dim taller, Dim tiempo

Properties: Source Column: IdEstudiante, Display Name: TotalEstudiantes, Default Aggregation: COUNT, Format: 0.00

Figura 2.91 Saiku: Indicador Total Estudiantes

En la Figura 2.92 se puede observar como de la tabla de hechos “H_REGISTROTALLER”, se resume el hecho “valor” para crear el indicador denominado “Total de Emociones”.

La fórmula utilizada para crear este indicador es la siguiente:

- “Total de Emociones” = SUM(H_REGISTROTALLER.VALOR).

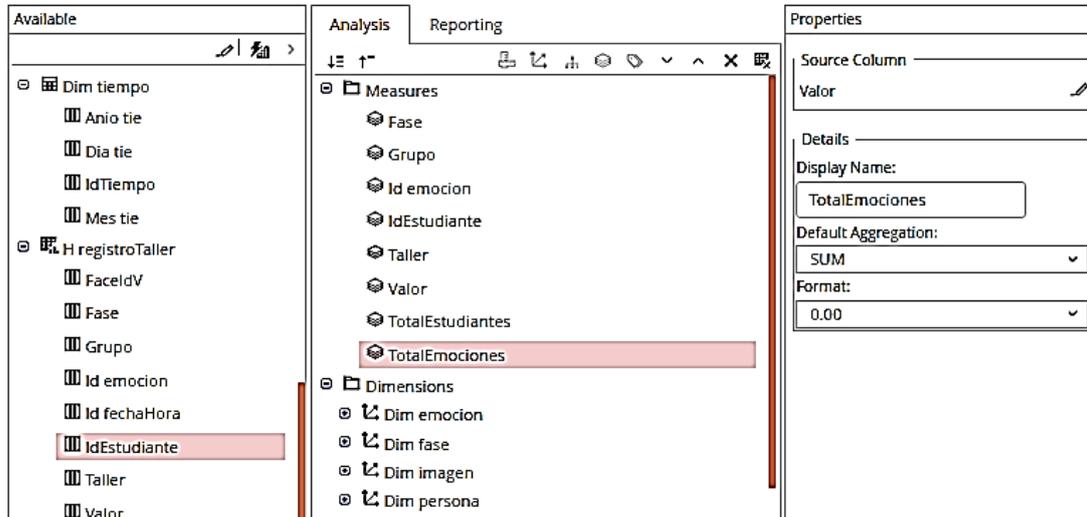


Figura 2.92 Saiku: Indicador Total Emociones

En la Figura 2.93 se puede observar como de la tabla de hechos “H_REGISTROTALLER”, se resume el hecho “valor” para crear el indicador denominado “Promedio de Emociones”.

La fórmula utilizada para crear este indicador es la siguiente:

“Promedio de Emociones” = AVERAGE(H_REGISTROTALLER.VALOR).

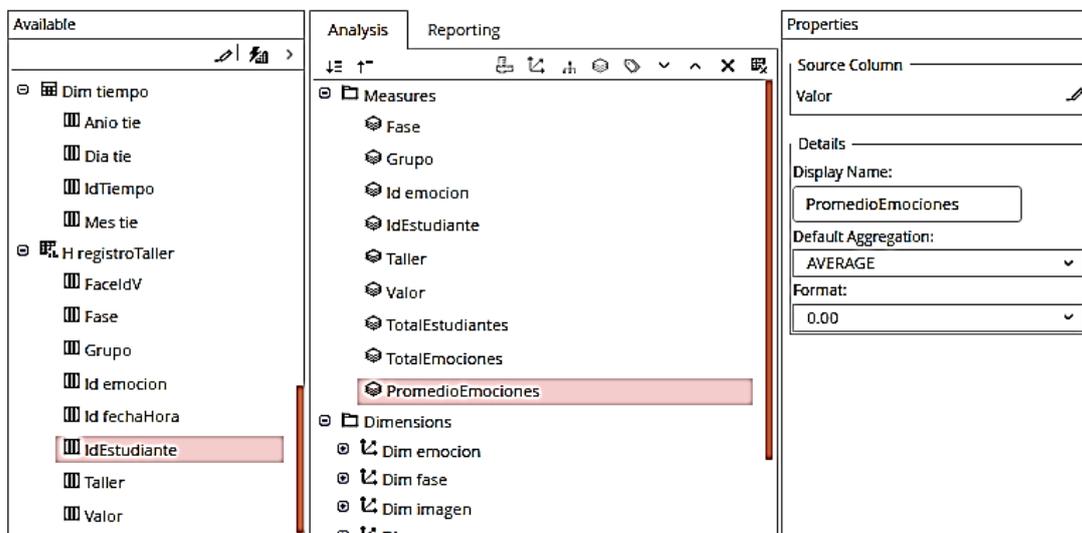


Figura 2.93 Saiku: Indicador Promedio Emociones

En la Figura 2.94 se muestra la creación del cubo multidimensional que se llama "Cubo de Predicción" y que está basado en el modelo lógico diseñado en el caso práctico de la metodología Hefesto.

Se selecciona la tabla de hechos "H_PREDICCIÓN" y las de dimensiones:

Selected Tables:

`multiV1`.`dim_emocion`
`multiV1`.`dim_experiencia`
`multiV1`.`dim_fase`
`multiV1`.`dim_imagen`
`multiV1`.`dim_persona`
`multiV1`.`dim_taller`
`multiV1`.`dim_tiempo`
`multiV1`.`h_prediccion`

Fact Table:

`multiV1`.`h_prediccion`

Figura 2.94 Saiku: Cubo Predicción

En la figura 2.95 se muestran las uniones entre la tabla de hechos h_prediccion y dimensiones.

Left Table:	Right Table:
`multiV1`.`h_prediccion`	`multiV1`.`dim_taller`
Key Field:	Key Field:
<ul style="list-style-type: none"> id_fechaHora dim_fase_idFase_fas prediccion ux pNegativa dim_persona_idPersona_per dim_taller_idTaller_tal 	<ul style="list-style-type: none"> fechaFin fechaInicio indicador capacidad competencia nombreTaller id_taller
Join(s):	
<pre>`multiV1`.`h_prediccion`.prediccion - INNER JOIN - `multiV1`.`dim_experiencia`.idExper `multiV1`.`h_prediccion`.dim_fase_idFase_fas - INNER JOIN - `multiV1`.`dim_fase`.noml `multiV1`.`h_prediccion`.dim_imagen_idImagen_ima - INNER JOIN - `multiV1`.`dim_ima; `multiV1`.`h_prediccion`.dim_persona_idPersona_per - INNER JOIN - `multiV1`.`dim_pei `multiV1`.`h_prediccion`.id_fechaHora - INNER JOIN - `multiV1`.`dim_tiempo`.idTiempoc</pre>	

Figura 2.95 Saiku: Relaciones para Cubo Predicción

En la Figura 2.96 se puede observar como de la tabla de hechos “H_PREDICCION”, se resume el hecho “Positiva” para crear el indicador denominado:

- “Total de Experiencia Positiva”.

La fórmula utilizada para crear este indicador es la siguiente:

- “Total de Emociones” = SUM(H_PREDICCION.POSITIVA).

Data Source Model Editor

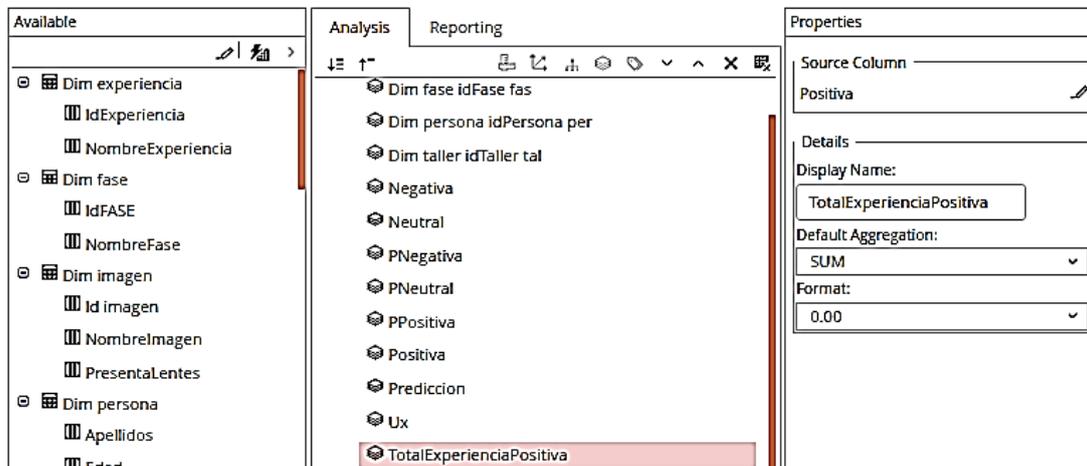


Figura 2.96 Saiku: Indicador Total Experiencia Positiva

En la Figura 2.97 se puede observar como de la tabla de hechos “H_PREDICCION”, se resume el hecho “Negativa” para crear el indicador denominado:

- “Total de Experiencia Negativa”.

La fórmula utilizada para crear este indicador es la siguiente:

- “Total de Emociones” = SUM(H_PREDICCION.NEGATIVA).

Data Source Model Editor

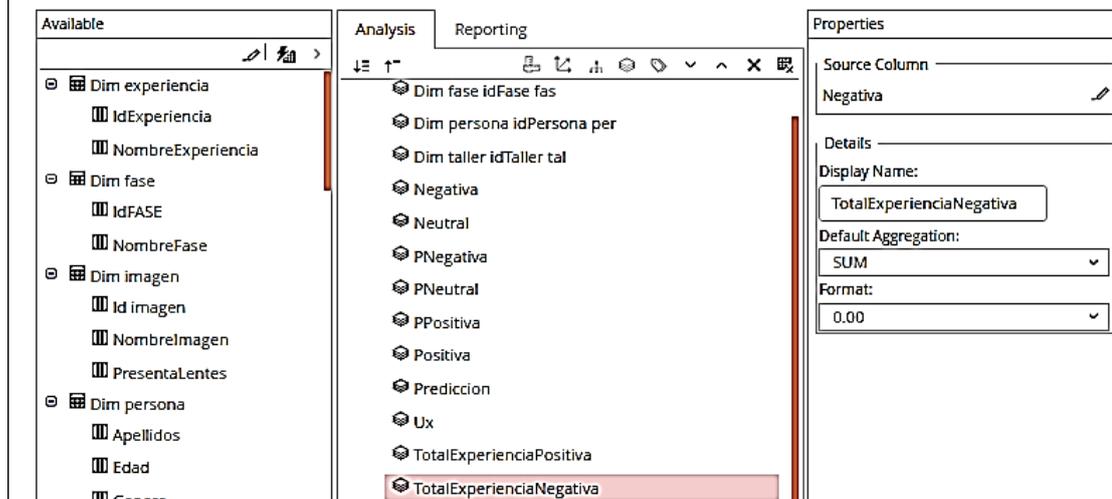


Figura 2.97 Saiku: Indicador Total Experiencia Negativa

En la Figura 2.98 se puede observar como de la tabla de hechos “H_PREDICCION”, se resume el hecho “Neutral” para crear el indicador denominado:

- “Total de Experiencia Neutral”.

La fórmula utilizada para crear este indicador es la siguiente:

- “Total de Emociones” = SUM(H_PREDICCION.NEUTRAL).

Data Source Model Editor

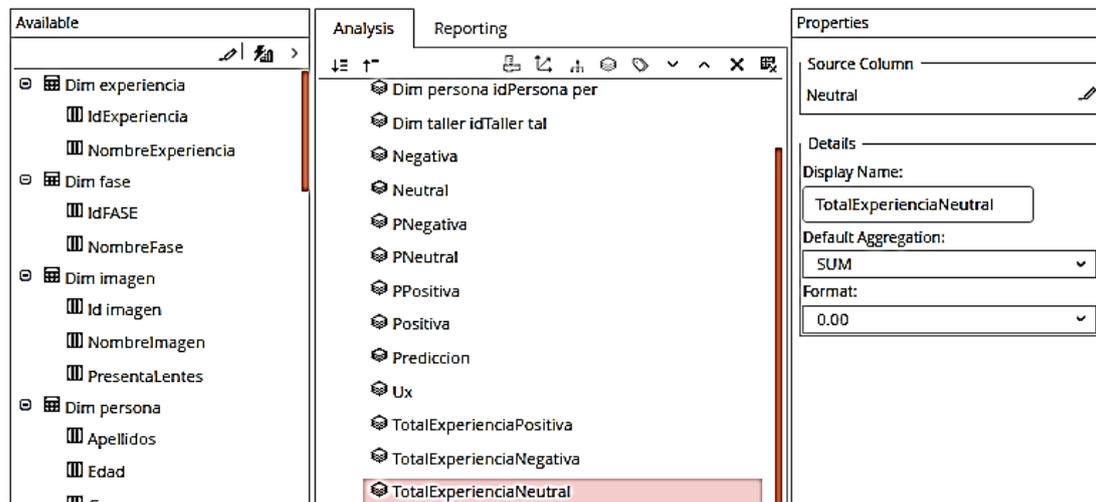


Figura 2.98 Saiku: Indicador Total Experiencia Neutral

Creación de Atributos

En la Figura 2.99 se muestra la creación de la Dimensión Tiempo y se agrega al cubo seis atributos:

- De la tabla de dimensión “DIM_TIEMPO”, se tomará el campo “anio-tie” para la creación del atributo denominado:
 - “Años”.
- De la tabla de dimensión “DIM_TIEMPO”, se tomará el campo “mes-tie” para la creación del atributo denominado:
 - “Meses”.
- De la tabla de dimensión “DIM_TIEMPO”, se tomará el campo “dia-tie” para la creación del atributo denominado:
 - “Dias”.

Data Source Model Editor

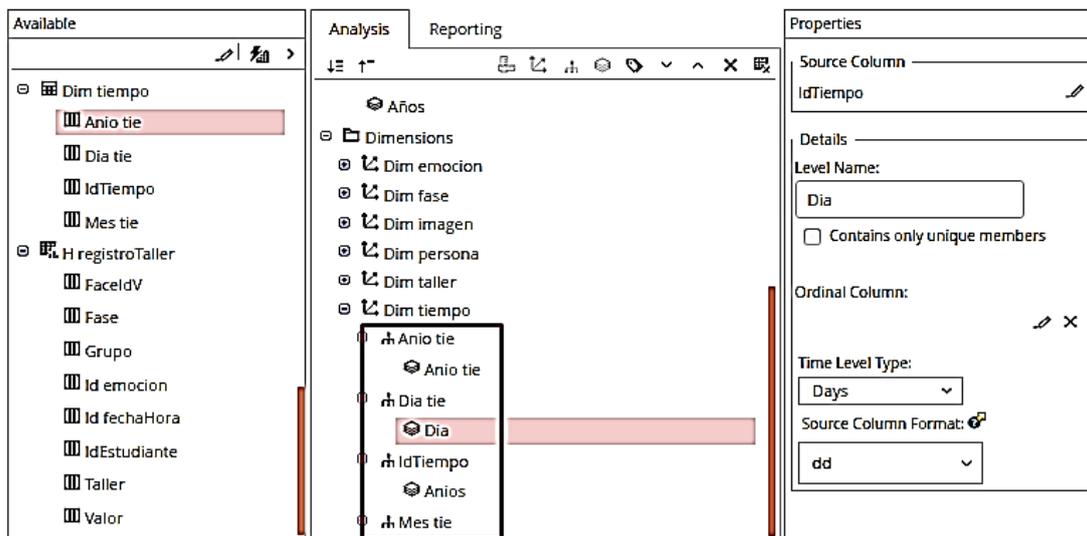


Figura 2.99 Saiku: Dimensión Tiempo

- En la Figura 2.100 se puede observar como de la tabla de dimensión “DIM_EMOCION”, se tomará el campo “nombreEmocion” para la creación del atributo denominado:
 - “Emocion”.

Data Source Model Editor

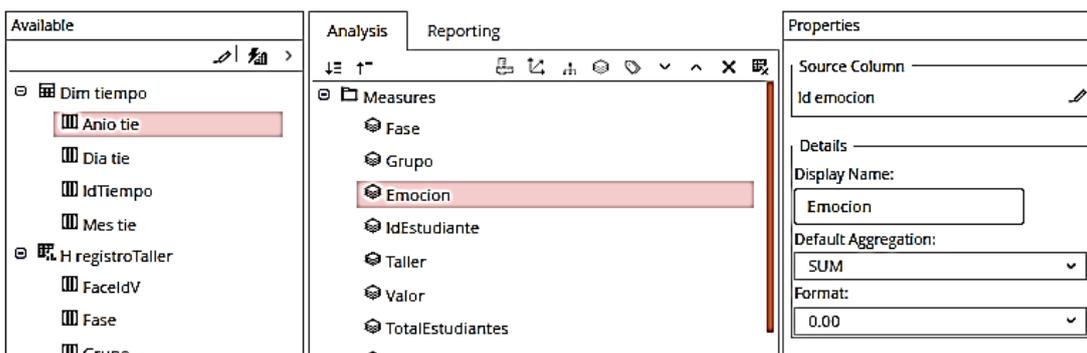


Figura 2.100 Saiku: Indicador para “Emocion”

- En la Figura 2.101 se puede observar como de la tabla de dimensión “DIM_PERSONA”, se tomará el campo “nombresCompletos” para la creación del atributo denominado:
 - “Estudiante”.

Data Source Model Editor

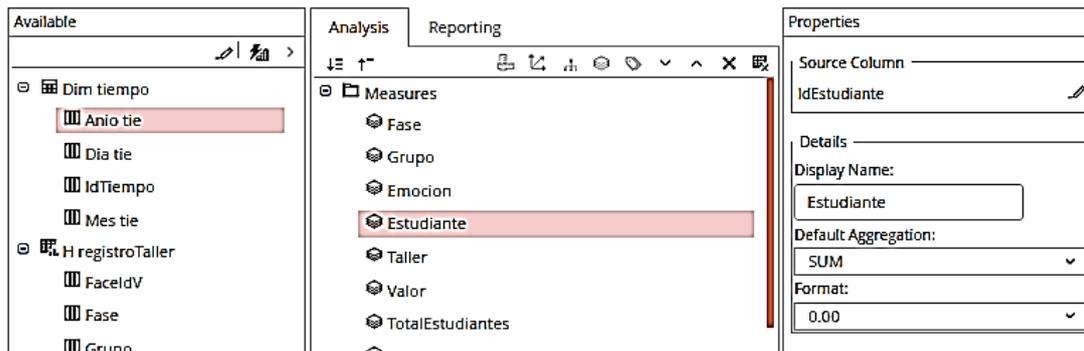


Figura 2.101 Saiku: Indicador para Número de Estudiantes

- De la tabla de dimensión “DIM_PERSONA”, se tomará el campo “genero” para la creación del atributo denominado:
 - “Género”.
- De la tabla de dimensión “DIM_PERSONA”, se tomará el campo “edad” para la creación del atributo denominado:
 - “Edad”.
- De la tabla de dimensión “DIM_PERSONA”, se tomará el campo “grupo” para la creación del atributo denominado:
 - “Grupo”.
- De la tabla de dimensión “DIM_FASE”, se tomará el campo “nombreFase” para la creación del atributo denominado:
 - “Fase”.
- De la tabla de dimensión “DIM_TALLER”, se tomará el campo “nombreTaller” para la creación del atributo denominado:
 - “Taller”.
- De la tabla de dimensión “DIM_TALLER”, se tomará el campo “fechaInicio” para la creación del atributo denominado:
 - “FechaInicio”.
- De la tabla de dimensión “DIM_TALLER”, se tomará el campo “fechaFin” para la creación del atributo denominado:
 - “FechaFin”.
- De la tabla de dimensión “DIM_IMAGEN”, se tomará el campo “nombreImagen” para la creación del atributo denominado:
 - “Imagen”.
- De la tabla de dimensión “DIM_TALLER”, se tomará el campo “presentaLentes” para la creación del atributo denominado:

- “Lentes”.

En la Figura 2.102 se presentan todas las medidas creadas.

Data Source Model Editor

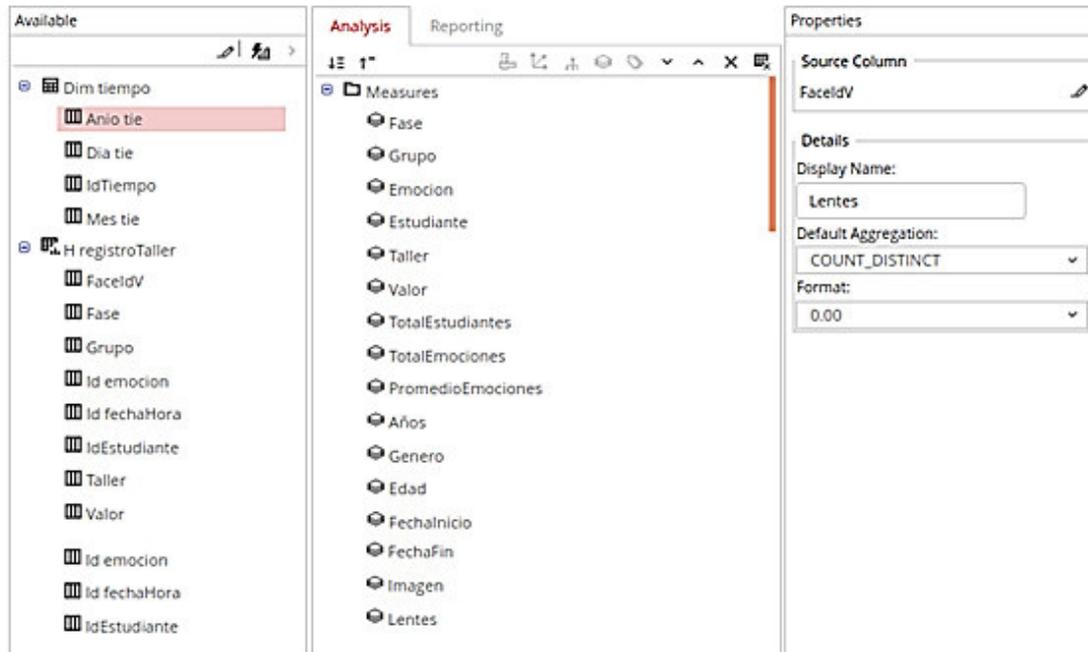


Figura 2.102 Saiku: Medidas

Visualización y reportería en Saiku

Después de generar los cubos y publicarlos en el Business Intelligence Server de Pentaho se pueden generar diferentes reportes mediante la selección de las dimensiones deseadas para responder las preguntas planteadas como requisitos, por ejemplo:

En la Figura 2.103 se muestra la Emoción por género.

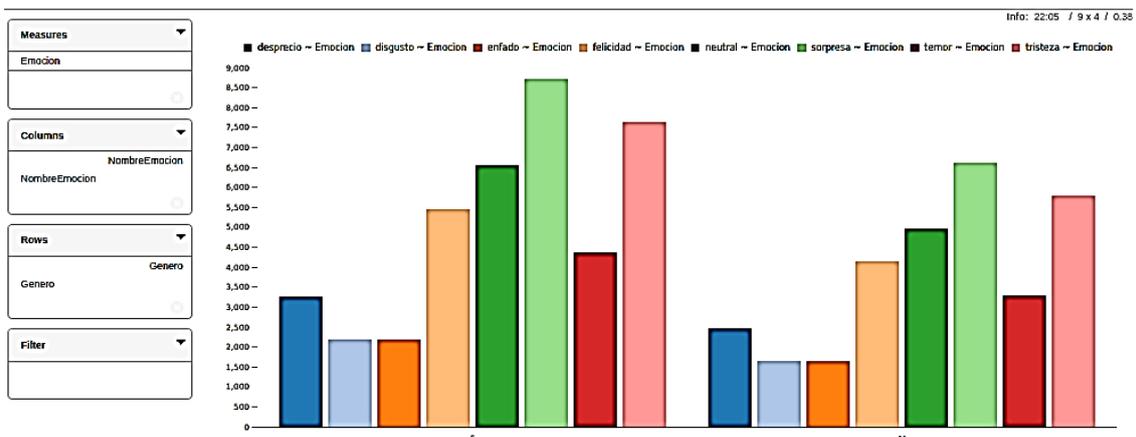


Figura 2.103 Saiku Reportes: Emoción por género

En la Figura 2.104 se muestra la Emoción por Fecha.

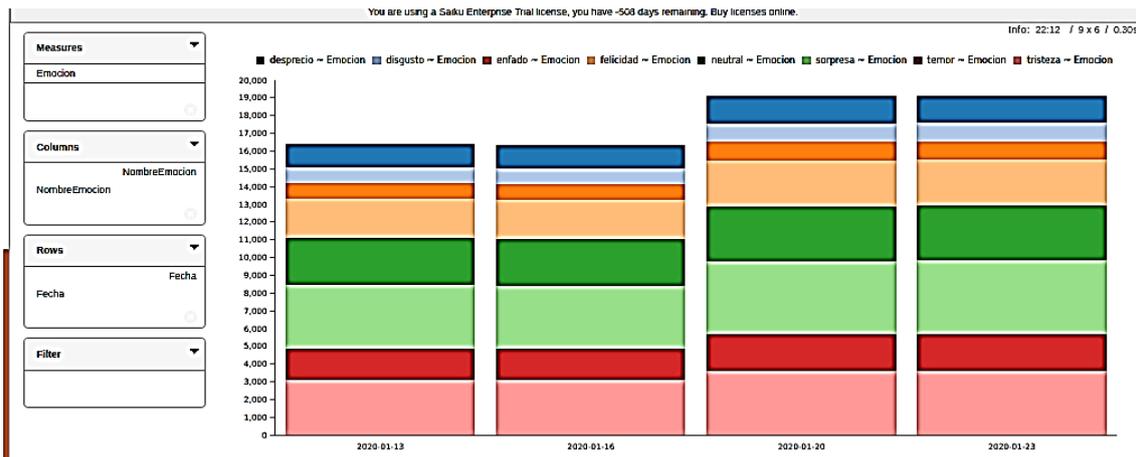


Figura 2.104 Saiku Reportes: Emoción por fecha

En la Figura 2.105 se muestran las Emociones por taller.

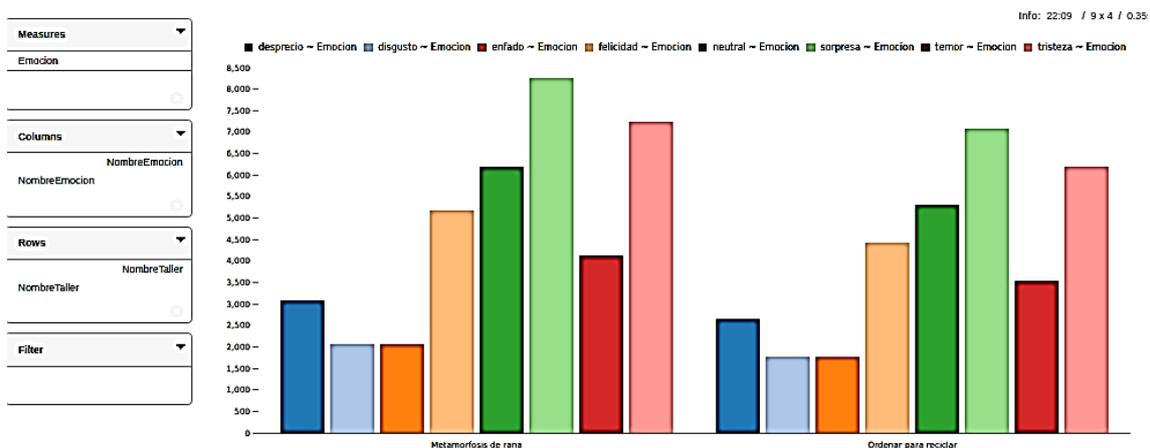


Figura 2.105 Saiku Reportes: Emociones por taller

En la Figura 2.106 se muestran las Emociones por fecha y taller.

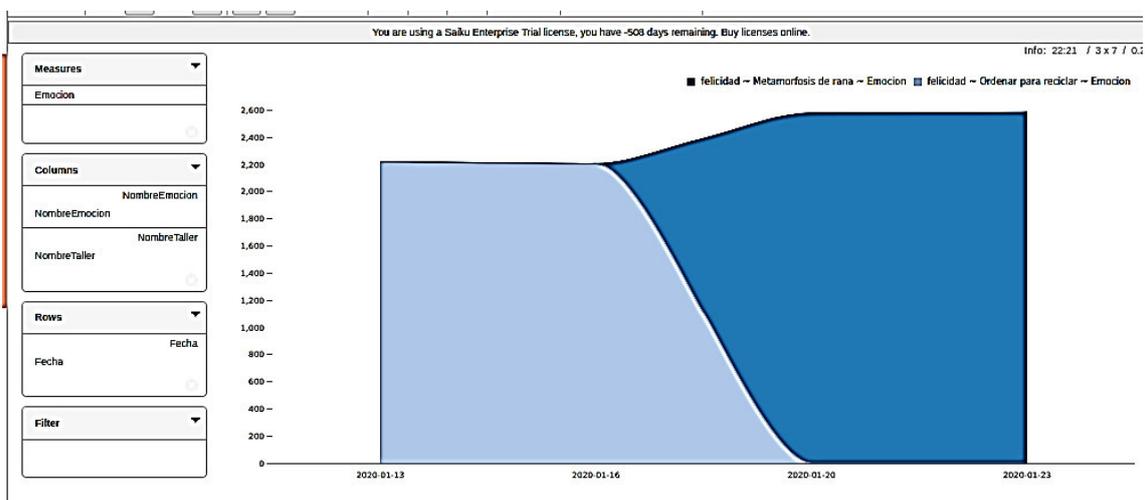


Figura 2.106 Saiku Reportes: Emociones por fecha y taller

En la Figura 2.107 se muestra la Emoción por estudiante y fecha.

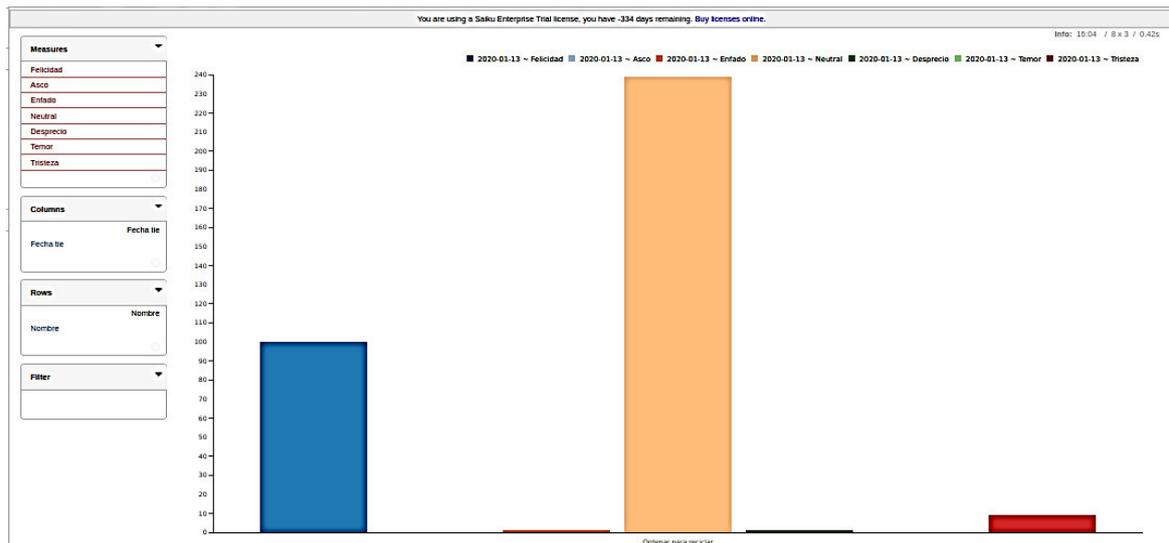


Figura 2.107 Saiku Reportes: Emoción por estudiante y fecha

En la Figura 2.108 se muestra la Emoción por taller, género, grupo, estudiante y fecha.

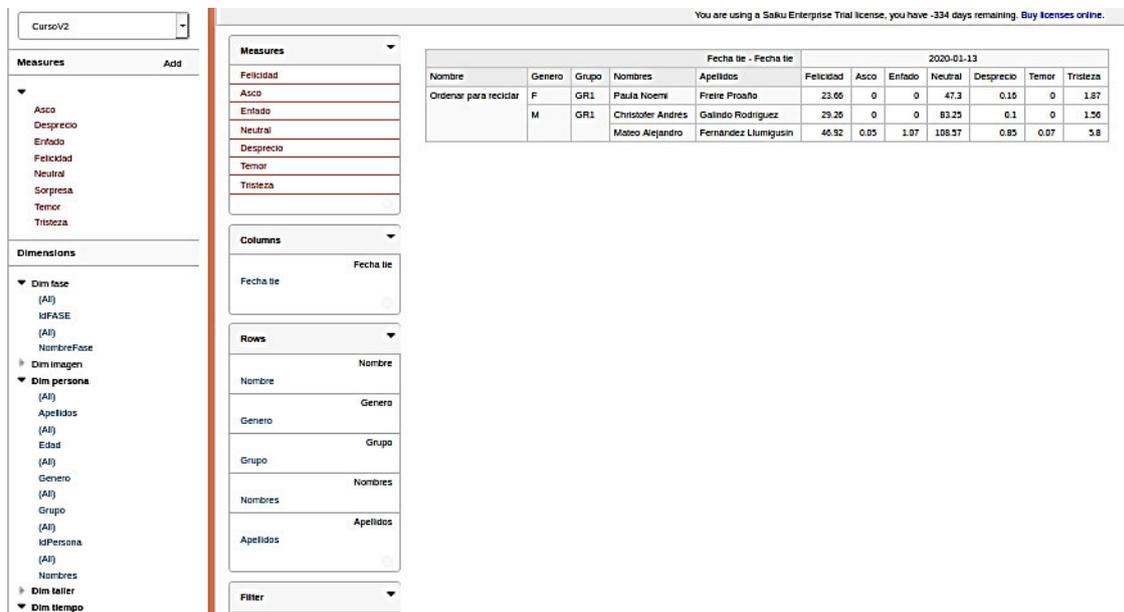


Figura 2.108 Saiku Reportes: Emoción por taller, género, grupo, estudiante y fecha

En la Figura 2.109 se muestra la Emoción por taller, género, grupo, estudiante, edad, fase y fecha.

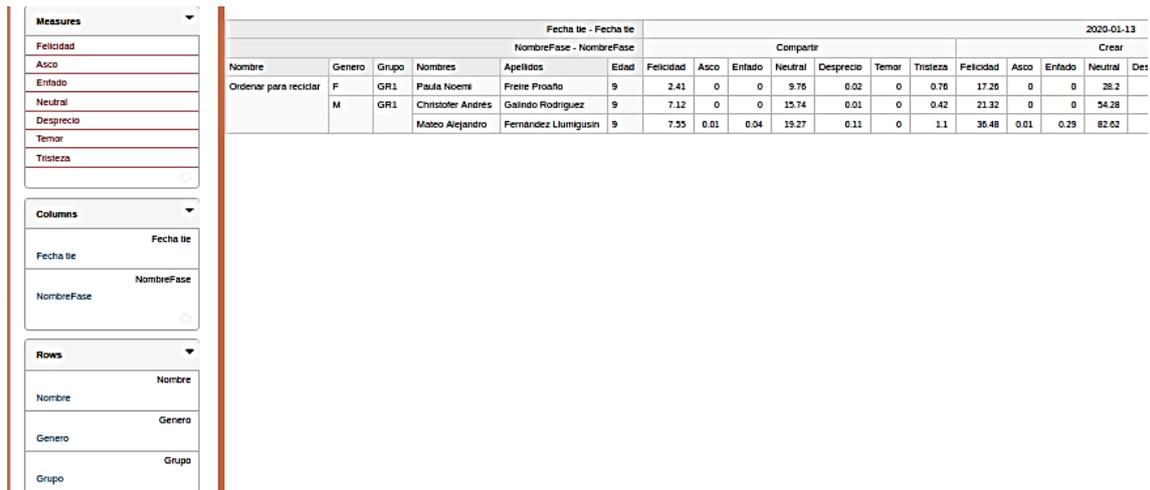


Figura 2.109 Saiku Reportes: Emoción por taller, género, grupo, estudiante, edad, fase y fecha

En la Figura 2.110 se muestra la Emoción por género y año.

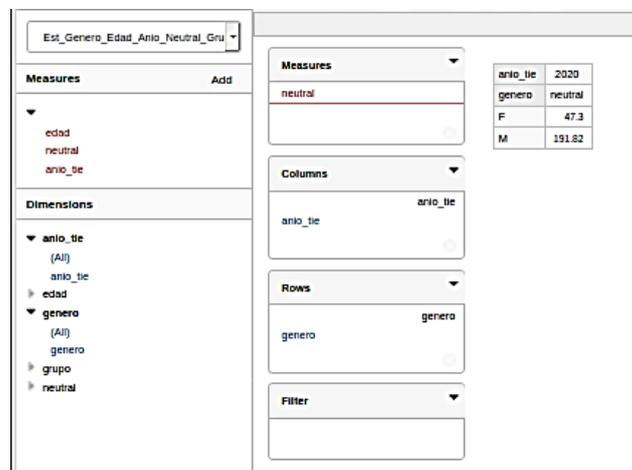


Figura 2.110 Saiku Reportes – Emoción por género y año

En la Figura 2.111 se muestran las Emociones por fecha y taller.

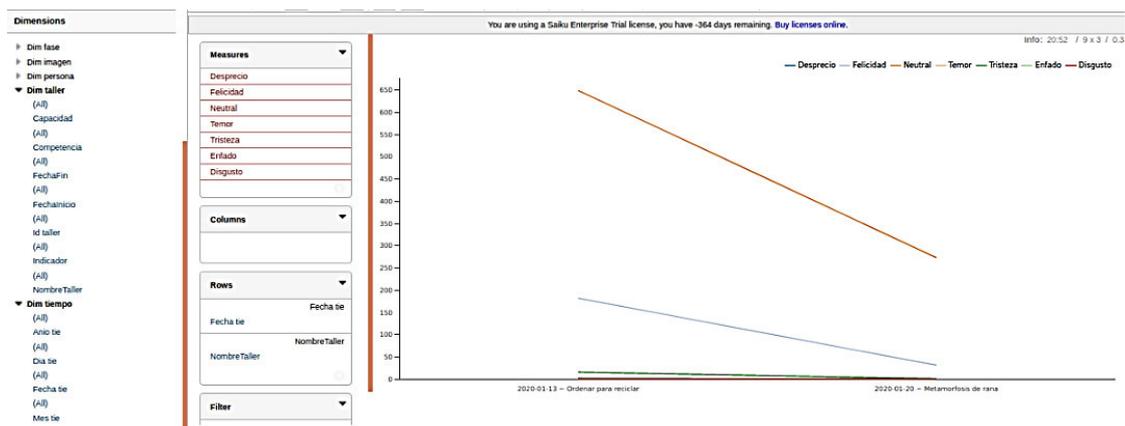


Figura 2.111 Saiku Reportes: Emociones por fecha y taller

En la Figura 2.112 se muestran las Emociones por fases de un taller.

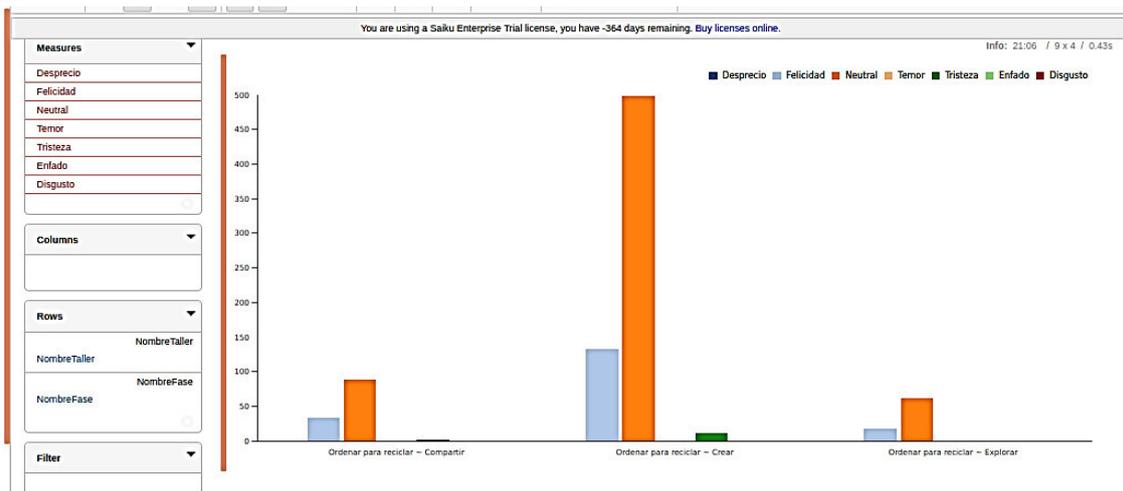


Figura 2.112 Saiku Reportes: Emociones por fases de un taller

En la Figura 2.113 se muestran las Emociones por fase y estudiante

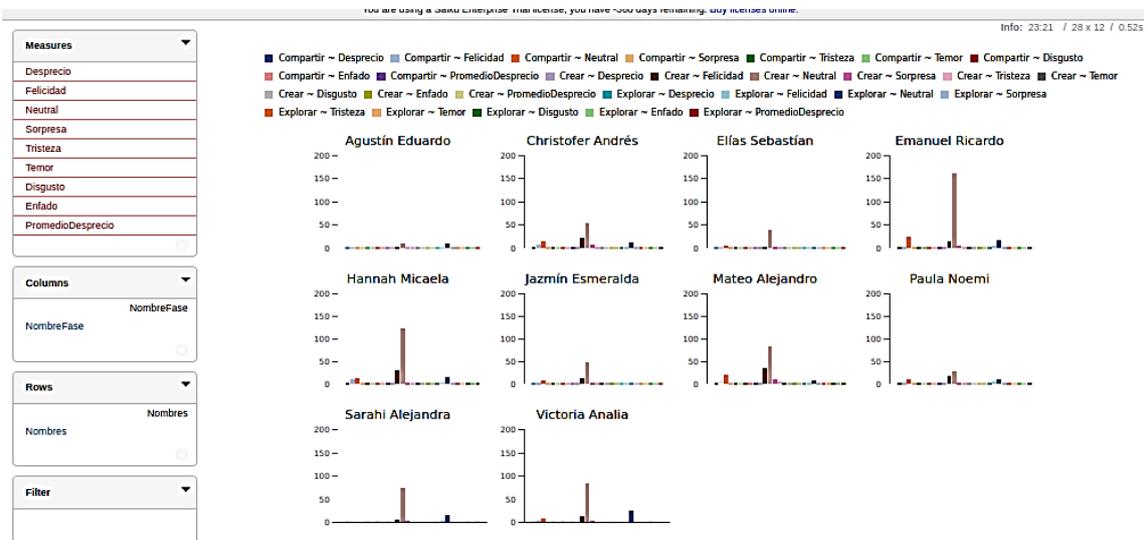


Figura 2.113 Saiku Reportes: Emociones por fase y estudiante

En la Figura 2.114 se muestra la Emoción por estudiante

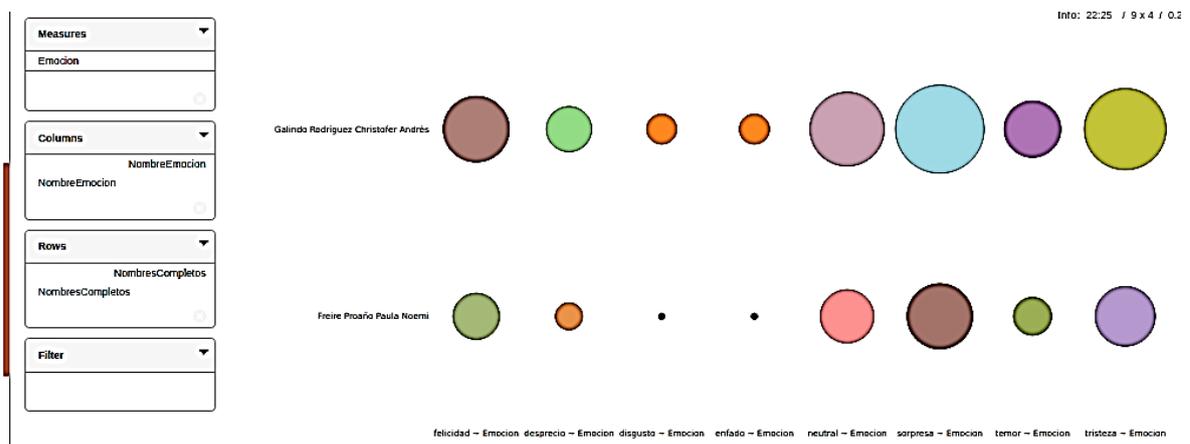


Figura 2.114 Saiku Reportes: Emoción por estudiante

En la Figura 2.115 se muestra la Emoción por Fase.

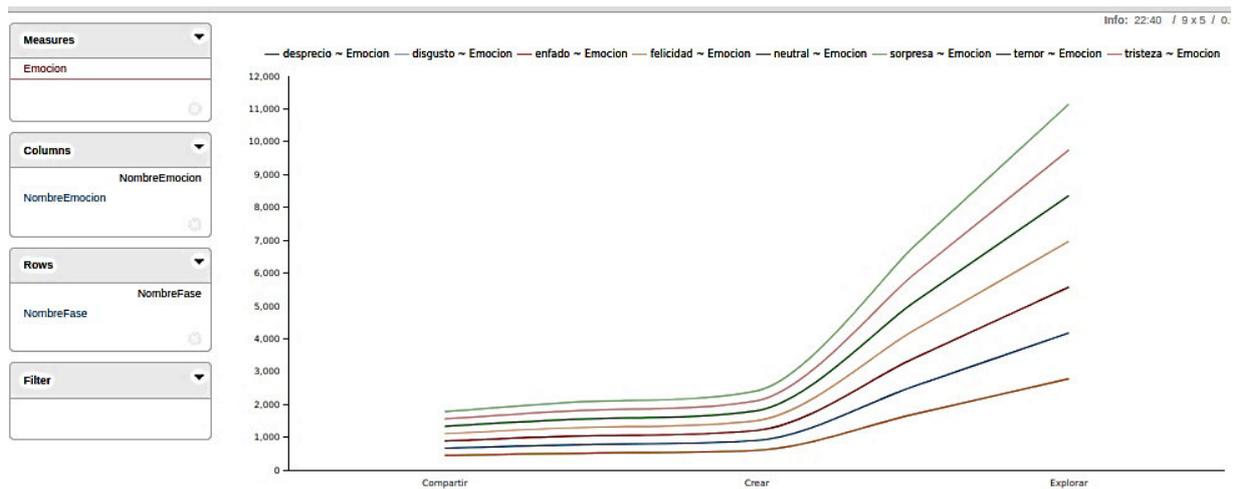


Figura 2.115 Saiku Reportes: Emoción por fase

El cubo realizado sobre el modelo multidimensional de Registro Taller cumple con los requerimientos solicitados, permitiendo generar reportes dinámicos según las necesidades de los usuarios. Además, vale recalcar que a pesar de que no se pueda generar consultas sobre dos cubos a la vez, se puede encontrar la información deseada accediendo directamente a los datos del modelo multidimensional mediante consultas SQL.

Diseño e implementación de Dashboards

Finalmente, se les entregará a los usuarios que lo requieran, las herramientas necesarias que les permitirán explorar los datos de manera sencilla e intuitiva lo que permite obtener las respuestas a las preguntas planteadas, mediante el uso de reportes, indicadores de rendimiento, cuadros de mando, gráficos estadísticos, etc.

De acuerdo a las perspectivas e indicadores identificados se inicia con el diseño de la Figura 2.116.

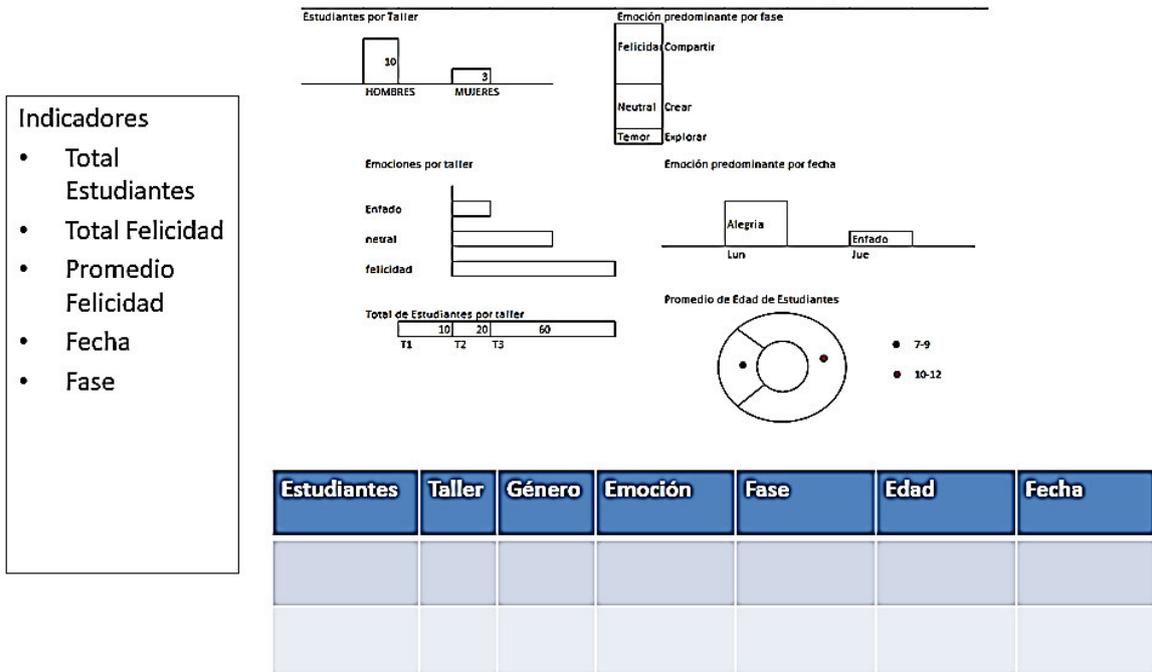


Figura 2.116 Diseño Dashboard

En la Figura 2.117 se puede ver el primer dashboard utilizando la herramienta de visualización:

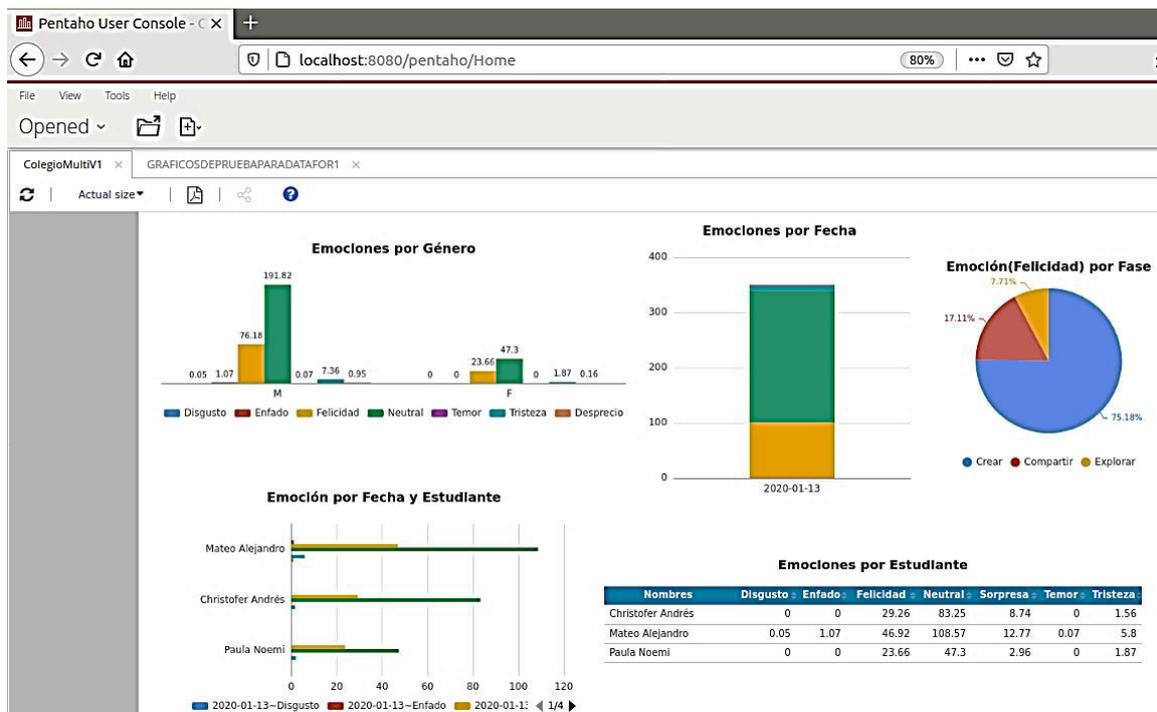


Figura 2.117 Visualizador Datafor: Diseño Dashboard

En la Figura 2.118 se muestra el inicio del dashboard aplicando principios de diseño:

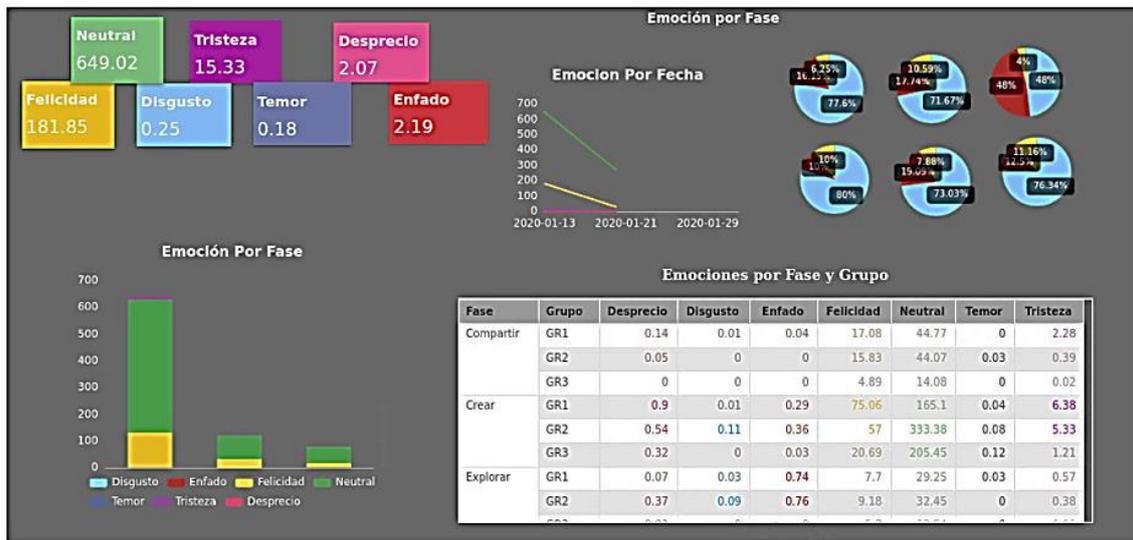


Figura 2.118 Visualizador Datafor: Diseño dashboard aplicando principios de diseño

En la Figura 2.119 se presenta la versión final del dashboard de emociones.

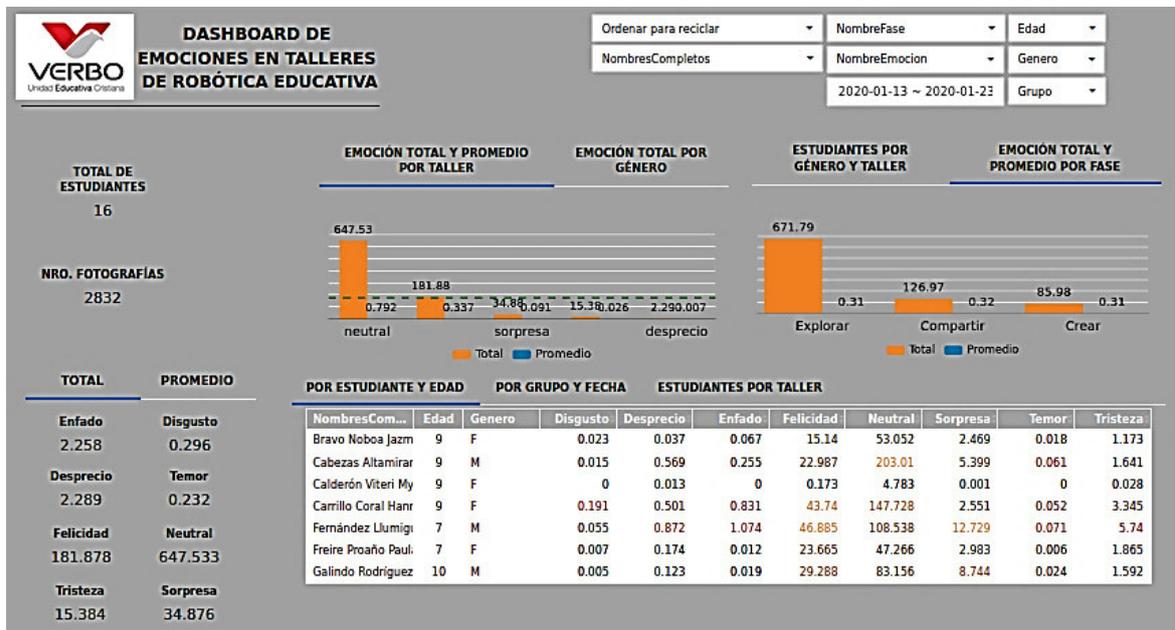


Figura 2.119 Visualizador Datafor: Versión final del dashboard de emociones

En la Figura 2.120 se presenta la versión final del dashboard de predicción de la experiencia de usuario.



Figura 2.120 Visualizador Datafor: Dashboard de predicción de experiencia de usuario

3. RESULTADOS Y DISCUSIÓN

Luego de la implementación del Sistema de Indicadores para Unidad Educativa Verbo, se describe los datos empleados en la carga inicial, así como los errores detectados en ellos, se da respuesta a las preguntas planteadas de los usuarios del sistema y finalmente se evalúa el uso del sistema de indicadores.

3.1 Carga inicial

Se procede a realizar la carga inicial del DW, poblando el modelo de datos que se ha construido anteriormente.

Se trabajó con la muestra que contenía 15344 registros de los cuales 1983 eran datos validos a los restantes se les dio un tratamiento especial para convertirlos en información útil y así cargar el Data Warehouse.

En la Tabla 3.1 se resumen el número de registros procesados con los errores encontrados:

Tabla 3.1 Registros con errores encontrados

Campo	Nro. Registros
accesorios	888
calva	888
invisible	888
colorDeCabello	888
nombreEstudiante	7
experienciaUsuario	57
fecha_hora	12

3.2 Respuestas a las preguntas planteadas

En el sistema a se puede apreciar los siguientes datos:

- Número total de asistentes, número de fotografías tomadas y los totales de las emociones detectadas en las fotografías, Figura 3.1.

TOTAL DE ESTUDIANTES 16 NRO. FOTOGRAFÍAS 5752	TOTAL	PROMEDIO
	Enfado 2.735	Disgusto 0.406
	Desprecio 4.992	Temor 0.436
	Felicidad 283.618	Neutral 1552.589
	Tristeza 22.477	Sorpresa 50.198

Figura 3.1 Indicadores del dashboard

Resultado:En las fotografías predomina la emoción neutral con un valor de 1552.589

- Mediante el uso de filtros se puede apreciar el total y promedio general de emociones en las fotografías de los talleres de Metamorfosis de la rana y Ordenar para reciclar, además de la media entre la que se encuentran las emociones, Figura 3.2.

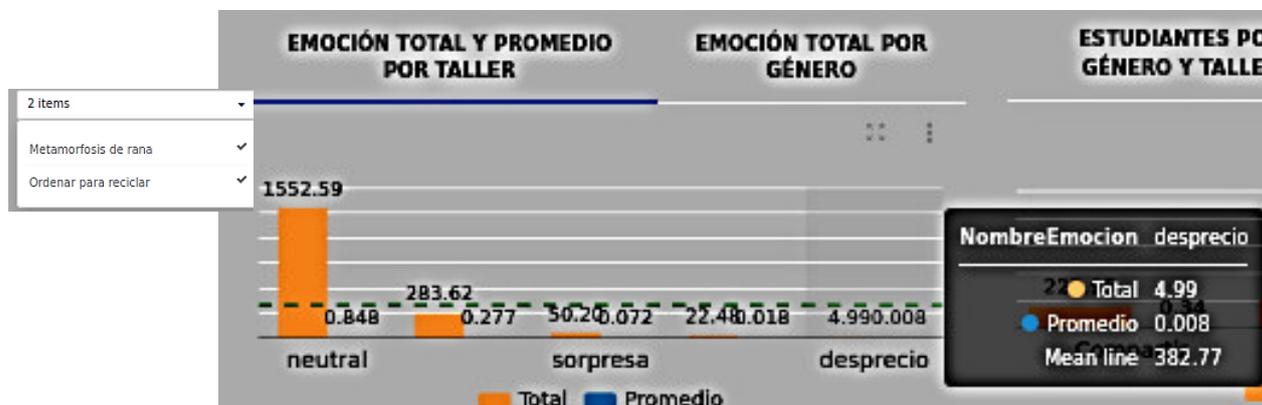


Figura 3.2 Total y promedio de Emociones por fotografía

Resultado: En general en la aplicación de los talleres predomina la emoción neutral con un valor de 1552.59 y su promedio es de 0.848, .

- Se puede apreciar emociones por estudiante , edad y género, Figura 3.3.

POR ESTUDIANTE Y EDAD			POR GRUPO Y FECHA			ESTUDIANTES POR TALLER				
NombresCom...	Edad	Genero	Disgusto	Desprecio	Enfado	Felicidad	Neutral	Sorpresa	Temor	Tristeza
Carrillo Coral Hanr	9	F	0.191	0.501	0.831	43.74	147.728	2.551	0.052	3.345
Fernández Llumign	7	M	0.055	0.872	1.074	46.885	108.538	12.729	0.071	5.74
Freire Praña Paul	7	F	0.007	0.174	0.012	23.665	47.266	2.983	0.006	1.865
Galindo Rodriguez	10	M	0.005	0.123	0.019	29.288	83.156	8.744	0.024	1.592
Guerron Ríos Sara	9	F	0	0.141	0.008	6.292	37.155	1.263	0.037	0.089
Huang Zhou Lisha	9	F	0.003	0.461	0.04	29.625	211.533	2.384	0.013	1.86
Morales Espín San	9	M	0	0.072	0	0.087	8.776	0.001	0	0.061
Naranjo Herrera M	9	F	0.038	0.408	0.218	3.144	68.904	1.217	0.019	2.028
Romo Gallardo Vic	9	F	0.001	0.035	0.002	9.931	62.056	2.476	0.019	0.461

Figura 3.3 Tabla de Emociones por Estudiante, Edad y Género

Resultado: Se puede apreciar que Micaela Carrillo tuvo un nivel de disgusto de 0.191, Mateo Fernández muestra uno de los valores considerables en desprecio: 0.872, temor: 0.071, tristeza: 3.345 y enfado: 1.074, a estos valores se les puede dar seguimiento para analizar las posibles causas.

Sin embargo, Mateo Fernández también muestra un considerable valor en felicidad: 46885 y tiene un valor alto en sorpresa: 12.729, sería importante analizar las fases en las cuales se obtuvo estos niveles.

- Emociones Negativas y Positivas

Para el caso de Mateo Fernández se filtran emociones, Figura 3.4.



Figura 3.4 Filtrado de Emoción positivas para un estudiante

Para el caso de Mateo Fernández se filtran emociones positivas y negativas por fases, Figura 3.4 y Figura 3.5.

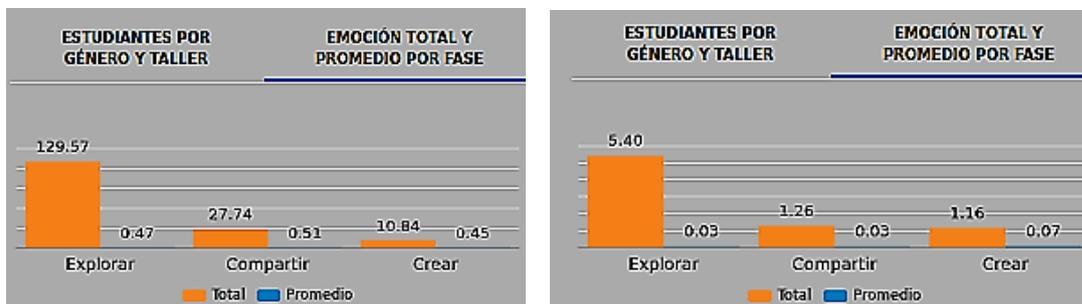


Figura 3.5 Emociones positivas y negativas por estudiante

Resultado: Se puede apreciar que las emociones como felicidad, neutral: 27.74 y sorpresa en un valor de 129.57, superan a las negativas: 5.40.

- En la Figura 3.6 se muestra la Emoción del estudiante por fase.



Figura 3.6 Emoción felicidad por un estudiante

Resultado: Para el caso de Emanuel Cabezas se visualiza que en la fase de explorar se obtuvo un alto nivel de felicidad de 14.83.

- En la Figura 3.7 se aprecia las emociones total y promedio por cada uno de los talleres, a la izquierda Metamorfosis de la rana y a la derecha Ordenar para reciclar.

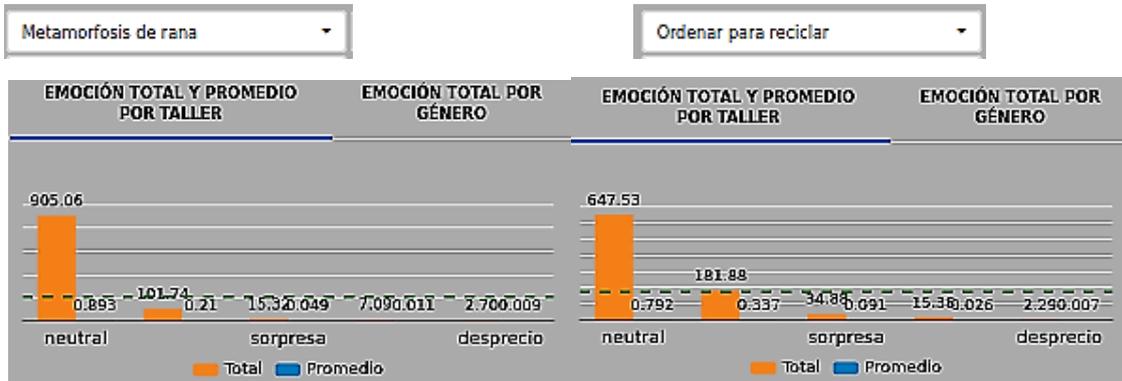


Figura 3.7 Total y promedio de emociones por taller

Resultado: Se puede apreciar que en ambos talleres predomina la emoción neutral con los valores de: 905.06 y 647.53, seguida de felicidad: 101.74 y 181.88 y sorpresa: 15.32 y 34.88 respectivamente, lo que de forma general indica que el taller de RE tuvo un efecto positivo en la emoción de los niños.

- En la Figura 3.8 se expone el número de estudiantes por taller y género

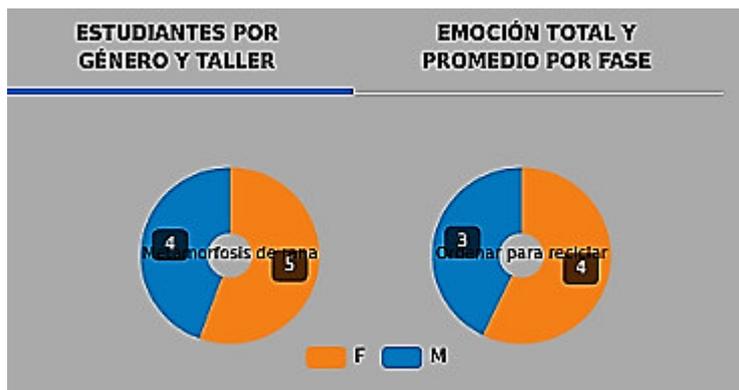


Figura 3.8 Número de estudiantes por taller y género

Resultado: En ambos talleres existió mayor número de niñas 5 y 4 respectivamente.

- En la Figura 3.9 se presenta la Emoción por género.

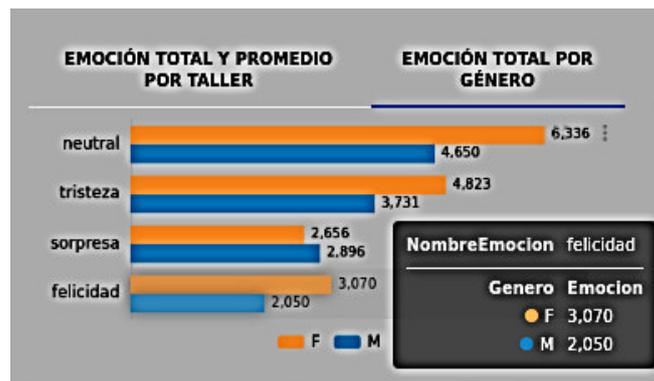


Figura 3.9 Total y promedio de emociones por género

Resultado: En lo que se refiere a género predomina la emoción neutral en las niñas con un valor de 0.336, seguido de un grupo donde la tristeza es mayor en las niñas que en los niños en una cantidad de 4.650, sería interesante ver por fase donde se centra esa tristeza, la sorpresa es mayor en niños que en niñas y se puede apreciar también un grupo donde la tendencia de felicidad es mayor en las niñas en 3.070 que los niños.

- En la Figura 3.10 se observa la Emoción total por género y fase

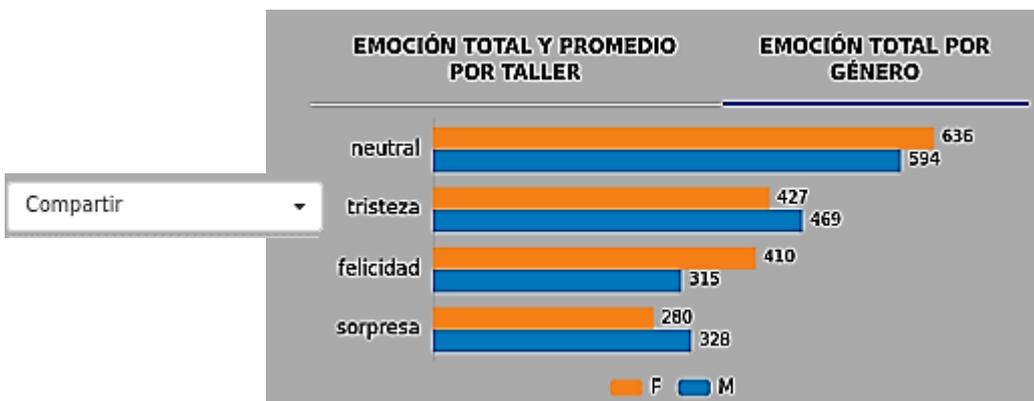


Figura 3.10 Total y promedio de emociones por género y fase

Resultado: En la fase de compartir en las niñas se muestra predominante la emoción neutral en 636, la tristeza al compartir es 427 para las niñas y 469 para los niños, la diferencia muestra que la tristeza es mayor en los niños que en las niñas, la felicidad en las niñas es de 410 y 315 en niños, por lo que la felicidad es mayor en las niñas que los niños.

Lo cual permite deducir que los estudiantes se sintieron a gusto compartiendo sus conocimientos.

- En la Figura 3.11 se muestran los valores de la experiencia de usuario versus la predicción.



Figura 3.11 Experiencia Vs Predicción por Taller

Resultado: En el Taller Ordenar para Reciclar se puede apreciar que bajo las mismas condiciones en las que se realizó el taller, la experiencia negativa reduce en puntaje de 2 a 0, la experiencia neutral se incrementa de 700 a 710 aunque esto signifique la disminución en experiencia positiva de 183 a 175 puntos.

- En la Figura 3.12 se puede ver la experiencia versus la predicción por estudiante.

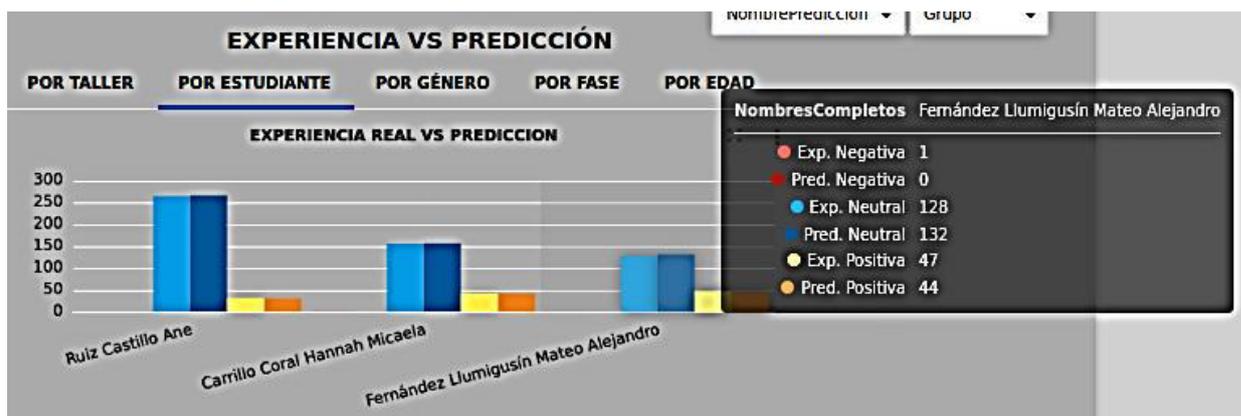


Figura 3.12 Experiencia vs Predicción por estudiante

Resultado: Se puede observar que la experiencia negativa en el caso del estudiante Mateo Fernández disminuye de 1 a 0 puntos, aumenta la experiencia neutral de 128 a 132 puntos, aunque la predicción positiva disminuye de 47 a 44 puntos.

En la Figura 3.13 se puede apreciar la experiencia de usuario versus la predicción por género.



Figura 3.13 Experiencia vs Predicción por Género

Resultado: Para el caso de los estudiantes varones se puede apreciar que la experiencia negativa disminuye de 1 a 0 puntos, la experiencia neutral incrementa de 706 a 715 puntos, aunque se mantiene la tendencia de disminución en la predicción positiva de 120 a 112 puntos.

En la Figura 3.14 se muestra la experiencia versus predicción por fase.



Figura 3.14 Experiencia vs Predicción por Fase

Resultado: En la Figura 3.14 se puede apreciar que la experiencia y predicción neutral tienen un puntaje alto de 1203 y 1212 respectivamente en la fase Explorar, disminuye en la fase de Crear a 265 y 268, para llegar al nivel mínimo en la fase de compartir 172 y 174.

Cabe mencionar que los datos pueden ser exportados en Excel, CVs o Imagen como se muestra en la Figura 3.15.

POR FASE Y TALLER		POR ESTUDIANTE					
NombreTaller	NombreFase	UX Negativa	Pred. Negat...	UX Neutral	Pred. Neutral	UX Positiva	Pred. Positiva
Metamorfosis de	Compartir	0	0	79	79	17	17
Metamorfosis de	Crear	0	0	197	198	19	18
Metamorfosis de	Explorar	0	0	664	667	57	54
Ordenar para rec	Compartir	0	0	93	95	34	32
Ordenar para rec	Crear	2	0	68	70	16	16
Ordenar para rec	Explorar	0	0	539	545	133	127

Figura 3.15 Experiencia vs Predicción por Taller y Fase

Resultado: Se puede exportar los datos de la experiencia versus la predicción por taller y por fase. Se puede apreciar que la experiencia negativa es 2 mientras la predicción negativa es 0. La experiencia neutral es mayor en Metamorfosis de la rana con 664 y la predicción se mantiene alta con 667 y la experiencia positiva disminuye de 133 a 127 en la predicción, en la fase de explorar en el taller Ordenar para reciclar.

3.3 Evaluar el uso del sistema de indicadores SUS Usability

Para medir la usabilidad del sistema se aplica SUS USABILITY, en ingeniería de sistemas, la escala de usabilidad del sistema es una escala Likert que consiste en una escala de calificación que se utiliza para cuestionar a una persona sobre su nivel de acuerdo o desacuerdo con una declaración. La misma contiene diez ítems que ofrece una visión global de las evaluaciones subjetivas de la usabilidad y son los siguientes:

1. Creo que usaría este sistema frecuentemente
2. Encuentro este sistema innecesariamente complejo
3. Creo que el sistema fue fácil de usar
4. Creo que necesitaría ayuda de una persona con conocimientos técnicos para usar este sistema
5. Las funciones de este sistema están bien integradas
6. Creo que el sistema es muy inconsistente
7. Imagino que la mayoría de la gente aprendería a usar este sistema en forma muy rápida
8. Encuentro que el sistema es muy difícil de usar
9. Me siento confiado al usar este sistema
10. Necesité aprender muchas cosas antes de ser capaz de usar este sistema

En la Figura 3.11 se puede ver los puntajes del SUS que se clasifican en una variedad de categorías: mejor imaginable, excelente, bueno, bueno, malo y peor imaginable. Depende del diseñador determinar un curso de acción de seguimiento. Esto puede incluir pruebas detalladas para encontrar la raíz del problema y los puntos débiles existentes, o un replanteamiento de la solución de diseño por completo.

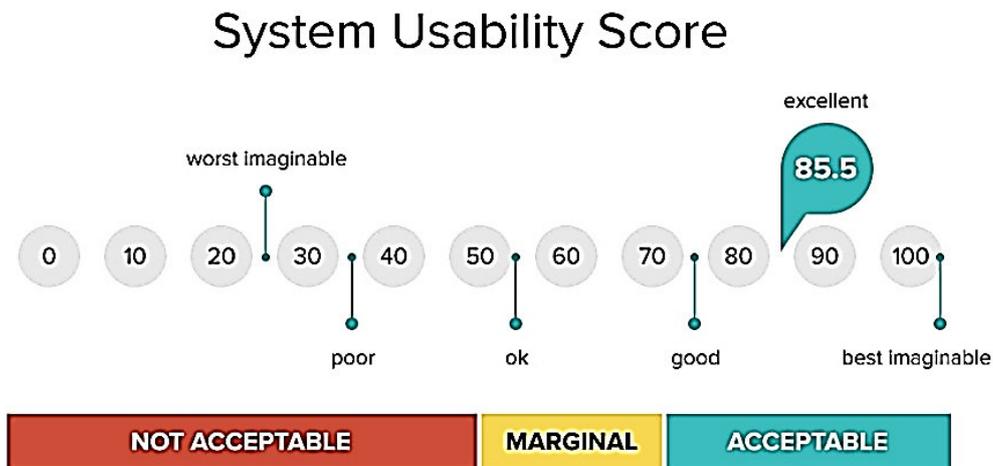


Figura 3.16 Puntuación de aceptabilidad [35]

Se aplica la encuesta un grupo específico de 5 personas. En la Tabla 3.2 se indica los resultados obtenidos de la misma y son los siguientes:

Tabla 3.2 Consolidado de puntaje SUS

PARTICIPANT	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS SCORE
Rubí Arias	4	3	4	3	4	1	4	2	3	2	70,0
Sara Cabrera	5	5	5	1	3	5	5	1	4	2	70,0
Julián Galindo	5	5	3	4	5	1	4	1	4	1	72,5
Wendy Rosero	4	4	4	2	4	2	4	2	4	2	70,0
Francisco Casa	5	5	3	3	4	1	3	1	5	1	72,5

De acuerdo a los resultados obtenidos el sistema presenta un nivel aceptable mayor al 70%.

4. CONCLUSIONES Y RECOMENDACIONES

4.1 Conclusiones

- Se ha desarrollado de un sistema de indicadores para monitorear el proceso de aprendizaje para el Caso de Estudio Taller de Robótica de la Unidad Educativa Verbo usando la metodología Hefesto que se basa en los requerimientos de los usuarios por lo cual su estructura es capaz de adaptarse ante cambios en el negocio, siendo independiente del tipo de ciclo de vida que se emplee para el desarrollo y de las estructuras físicas que contengan el Data Warehouse.
- Se proporciona una herramienta de análisis para establecer comparaciones y tomar decisiones permitiendo pasar de forma ágil de la decisión a la acción, por otro lado, el uso de un visualizador de la información permite conocer cuáles son los factores que inciden en el buen o mal funcionamiento de la aplicación de una nueva tecnología ya que permite acceder y analizar directamente los indicadores de éxito.
- Se automatiza la asimilación de la información de emociones, talleres, estudiantes para que sea integrada en una fuente única de fácil acceso mediante el uso de dos DWH que contienen la información de talleres con las emociones detectadas y de predicción de emociones.
- El sistema de indicadores que posibilita la formulación y respuesta de preguntas que son claves para el desempeño de la institución abarca indicadores estáticos: estudiantes, emociones, experiencia de usuario y un dinámico que es la predicción de la experiencia de usuario.
- La extracción, transformación y carga de los datos necesarios se realiza a través de procesos predefinidos usando etls con la herramienta pentaho data integration y RapidMiner
- Para la evaluación del sistema se aplica SUS Usability a 5 usuarios, el resultado refleja que el sistema presenta un nivel de aceptación mayor al 70%. En este proceso los usuarios consultan y analizan los datos de manera sencilla e intuitiva, se reduce el tiempo mínimo que se requiere para recoger la información relevante de la aplicación de una nueva tecnología, se proporciona herramientas de análisis para establecer comparaciones y tomar decisiones. Los usuarios no dependen de reportes o informes programados, porque el sistema de indicadores los genera de manera dinámica.
- Como contribución en el presente trabajo se incorpora la visualización ya que la metodología Hefesto no incluye esta fase. Se debe mencionar que dicha

implementación se la realiza con una herramienta de software libre de Datafor; que, a pesar de sus limitantes en cuanto al uso de componentes, permite un trabajo limpio para obtener resultados intuitivos y de fácil uso para los usuarios.

- Entre los beneficios más importantes que este proyecto de BI proporciona a la Institución Educativa Verbo, se destacan los siguientes:
 - Posibilita la formulación y respuesta de preguntas que son claves para el desempeño de la institución.
 - Permite acceder y analizar directamente a los indicadores de éxito.
 - Se pueden identificar cuáles son los factores que inciden en los resultados al aplicar una nueva tecnología para el aprendizaje en la institución.
 - Se podrán detectar situaciones fuera de lo normal.
 - Permitirá predecir el comportamiento futuro con un alto porcentaje de certeza, basado en el entendimiento del pasado.
 - Al contar con la información exacta y en tiempo real, es posible, aparte de lo ya mencionado, reconocer y corregir situaciones antes de que se conviertan en problemas y en potenciales pérdidas de los estudiantes incluso en años lectivos, pudiendo conseguir nuevas oportunidades o readaptarse frente a la ocurrencia de sucesos inesperados al aplicar una nueva tecnología.
 - Posibilita la formulación y respuesta de preguntas que son claves para el desempeño de la organización.
 - Se podrán detectar situaciones fuera de lo normal.
 - Permitirá predecir el comportamiento futuro con un alto porcentaje de certeza, basado en el entendimiento del pasado.
 - Los beneficios del DW son apreciados en el mediano y largo plazo. No todos los usuarios confiarán en el DW en una primera instancia, pero sí lo harán una vez que comprueben su efectividad y ventajas. Además, su correcta utilización surge de la propia experiencia.
- El tratamiento futuro de la sorpresa deriva en una mejor técnica que se base en un análisis de cada iteración y al momento no se dispone de un orden cronológico preciso de las fotos para realizar la comparación ordenada que se requiere en cada iteración y verificar la variación de sorpresa por estudiante por lo que se deja como perspectiva en esta tesis ya que amerita un análisis más profundo con otras técnicas. Se plantea como solución realizar las iteraciones ordenadas por usuario y mejorar el ordenamiento en la toma de los datos.

- Se consideraron las variables específicas: neutral, nombre de la fase, felicidad, tristeza, desprecio y lentos; cuyos pesos, estaban arriba del 50% para obtener un preciso y mejor algoritmo de clasificación.
- La metodología de Hefesto usualmente se la emplea con KPI estáticos y no con dinámicos. En el presente proyecto se realiza la predicción de la variable experiencia de usuario como indicador dinámico. Se utiliza RapidMiner para la predicción.
- RapidMiner usa técnicas que construyen y estudian nuevas previsiones a través de una rama de la inteligencia artificial denominada aprendizaje automático (machine learning) incorporando los principios de la minería de datos para el descubrimiento de conocimientos y la elaboración de modelos.
- Se verifica que el modelo Random Forest es el óptimo para realizar la predicción ya que muestra un menor error de clasificación de 0,55% y una exactitud del 99,50%. La predicción del indicador experiencia de usuario permite apreciar el resultado de la aplicación del taller de RE en las emociones de estudiante lo que permite mejorar la enseñanza de una materia específica.
- Un análisis y seguimiento adecuado de los KPIs estáticos y dinámico ayudarán a tomar decisiones a corto y largo plazo.
- De lo expuesto anteriormente permite concluir que la herramienta DW incide significativamente en la planificación de la aplicación de una nueva tecnología en el aprendizaje en la institución ya que el manejo de la información confiable ayuda a tomar las mejores decisiones al director y profesores principalmente.
- De forma general se puede visualizar que los talleres de RE presentan una emoción positiva en los estudiantes lo cual contribuye en su aprendizaje.
- Finalmente, se implementa el sistema por BI en el Caso de Estudio Taller de Robótica de la Unidad Educativa Verbo usando la metodología Hefesto, incluyendo el visualizador de Datafor con KPIs estáticos y dinámico para monitorear el proceso de aprendizaje para de esta manera corregir y mejorar a tiempo la experiencia del niño con el fin de optimizar los niveles de aprendizaje.

4.2 Recomendaciones

- Existen varios puntos claves a considerar y mejorar: el primero es poder contar con una aplicación que gestione la información a ingresar de los estudiantes, talleres, análisis de fotografías, indicadores e integrarlos. El segundo es realizar planes de mantenimientos ante el crecimiento del tamaño de la información, planes de mejora y crecimiento. El tercero es crear una línea de

capacitación a los usuarios funcionales y técnicos para evitar errores por desconocimiento del uso de las herramientas.

- Este conocimiento debe ser oportuno, relevante, útil y debe estar adaptado al contexto de la organización.
- Se puede realizar una tesis donde se considere la inclusión de videos donde se pueda trabajar de forma interactiva con las emociones de los estudiantes.
- Es necesario recopilar la mayor cantidad de atributos que contribuyan a entrenar de mejor manera el método de aprendizaje automático utilizado y de esta forma obtener un mejor nivel de precisión.

REFERENCIAS BIBLIOGRÁFICAS

- [1] R. D. Bernabeu, *Hefesto Data Warehousing: Investigación y Sistematización de Conceptos Hefesto: Metodología para la Construcción de un Data Warehouse*. 2010.
- [2] Noriaki Ando, S. Balakirsky, T. Hemker, M. y Reggiani, and O. Stryk, *Simulación, modelado y programación para robots autónomos: Segunda Conferencia Internacional*. 2010.
- [3] C. Loh, L. Chu, and L. Yoke, *Intelligent Robotics Systems: Inspiring the next*, Springer H. 2013.
- [4] M. Saleiro, C. Bruna, and J. Rodrigues, "A Low-Cost Classroom-Oriented Educational Robotics System," *Int. Publ. Switz.*, pp. 76–81, 2013.
- [5] D. Alimisis, "Educational robotics: Open questions and new challenges," *Themes Sci. Technol. Educ.*, vol. 6, no. 1, pp. 63–71, 2013.
- [6] L. Garcés, A. Mora, M. Aguirre, M. Espinosa, C. Vera, and A. Guardera, "Plan Nacional de Desarrollo 2017-2021-Toda una Vida," 2017. www.planificacion.gob.ec.
- [7] I. Jormanainen and E. Sutinen, "Using data mining to support teacher's intervention in a robotics class," *2012 IEEE Fourth Int. Conf. Digit. Game Intell. Toy Enhanc. Learn.*, pp. 39–46, 2012.
- [8] F. Guasmayan, N. Vargas, and J. Guerrero, "Estado del arte de redes educativas para el intercambio de conocimientos en robótica educativa," *Ing. e Innovación*, vol. 7, no. 2, p. 5, 2019.
- [9] A. Mohanty, "Affective pedagogical agent in e-learning environment: A reflective analysis," *Creat. Educ.*, vol. 7, no. 04, p. 586, 2010.
- [10] D. Vikrant and M. Pirouz, "A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals," *Big Data*, 2020.
- [11] J. L. Tufiño López, "Desarrollo e Implantación del Datamart para el Sistema Nacional de Vigilancia Tecnológica de Software Libre." QUITO, 2011., 2011.
- [12] G. E. S. Peñafiel, V. M. Z. Yáñez, K. P. M. Guamán, and L. M. T. Padilla, "Análisis de metodologías para desarrollar Data Warehouse aplicado a la toma de decisiones," *Cienc. Digit.*, vol. 3, no. 3.4., pp. 397–418, 2019.
- [13] V. A. Haro Valle, W. R. Pérez Rocano, L. Siguenza Guzman, D. Cattrysse, and V. H. Saquicela Galarza, "Diseño e implementación de un sistema de soporte de decisiones para el Centro de Documentación Regional 'Juan Bautista Vázquez,'" *Maskana*, vol. 5, no. 2, pp. 245–256, 2014.

- [14] S. L. Morales Cardoso, "Metodología para procesos de Inteligencia de Negocios con mejoras en la extracción y transformación de fuentes de datos, orientado a la toma de decisiones," 2019.
- [15] F. Martínez Abad, A. Gamazo, and M. J. Rodríguez, "Big Data in Education: Detection of ICT Factors Associated with School Effectiveness with Data Mining Techniques," *Sexta Conf. Int. Salamanca*, pp. 3–5, 2018.
- [16] Z. Memon, "How To Install And Use MySQL Workbench On Ubuntu – Linux Hint," 2020. https://linuxhint.com/install_mysql_workbench_ubuntu/ (accessed Apr. 12, 2021).
- [17] C. CSharp, "(159) Respalda base de datos con Mysql Workbench 0.3 - YouTube," 2010. https://www.youtube.com/watch?v=B1_rmqCVwSE (accessed Apr. 12, 2021).
- [18] EnMiMaquinaFunciona, "¿Cómo eliminar el límite de 1000," 2011. <https://www.enmimaquinafunciona.com/pregunta/25329/como-quitar-1000-limite-de-fila-en-mysql-workbench-consultas> (accessed Apr. 12, 2021).
- [19] Microsoft, "Conceptos de atributos y detección de caras - Azure Cognitive Services," 2021. <https://docs.microsoft.com/es-es/azure/cognitive-services/face/concepts/face-detection> (accessed Apr. 12, 2021).
- [20] Microsoft, "Quickstart: Detect faces in an image using the REST API and JavaScript - Azure Cognitive Services," 2021. <https://docs.microsoft.com/en-us/azure/cognitive-services/face/quickstarts/javascript> (accessed Jun. 04, 2020).
- [21] A. E. VILLACÍS VERA and P. A. FLORES AVENDAÑO, "Análisis comparativo de las herramientas de big data," PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR, 2017.
- [22] Hitachi Vantara, "Pentaho Business Analytics Platform," Santa Clara, 2018. Accessed: Apr. 14, 2021. [Online]. Available: <https://cdw-prod.adobecqms.net/content/dam/cdw/on-domain-cdw/brands/hitachi/pentaho-business-analytics-platform-data-sheet.pdf>.
- [23] M. BI, "Hogar de Saiku Analytics," 2021. <https://www.meteorite.bi/products/saiku/> (accessed Apr. 14, 2021).
- [24] DATAFOR, "Visualización y análisis de datos adaptados al futuro," 2015. <http://www.datafor.com.cn/> (accessed Apr. 16, 2021).
- [25] M. Tamayo and F. J. Moreno, "Análisis del modelo de almacenamiento MOLAP frente al modelo de almacenamiento ROLAP," *Ing. e Investig.*, vol. 26, no. 3, pp. 135–142, 2000.
- [26] J. D. R. Antolines, "Modelo circunplejo del afecto aplicado al control de sistemas dinámicos," *Respuestas*, vol. 14, no. 1, pp. 5–15, 2009.

- [27] R. Bisquerra Alzina, "Psicopedagogía de las emociones," 2009.
- [28] B. Clarke, E. Fokoue, and H. H. Zhang, *Principles and theory for data mining and machine learning*. Springer Science & Business Media, 2009.
- [29] B. Williamson, *Big data en educación: el futuro digital del aprendizaje, la política y la práctica*. Ediciones Morata, 2018.
- [30] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics," *Electronics*, vol. 8, no. 8, p. 832, 2019.
- [31] KDnuggets, "Top Data Science and Machine Learning Methods Used in 2018, 2019," 2021. <https://www.kdnuggets.com/2019/04/top-data-science-machine-learning-methods-2018-2019.html> (accessed May 29, 2021).
- [32] Y. H. Montero, "Experiencia de usuario: principios y métodos," *Exp. Usuario Principios y Métodos*, vol. 9, 2015.
- [33] F. Gordillo, M. Á. Pérez, L. Mestas, J. Salvador, J. M. Arana, and R. M. López, "Diferencias en el Reconocimiento de las Emociones en Niños de 6 a 11 Años," *Acta Investig. psicológica*, vol. 5, no. 1, pp. 1846–1859, 2015.
- [34] J. M. Sánchez Beisel and E. M. Ruetti, "Desarrollo de las emociones en las niñas y los niños. Revisión de los principales factores moduladores," 2018.
- [35] Ideas de Adobe XD, "La escala de usabilidad del sistema y cómo se usa en UX," 2019. <https://xd.adobe.com/ideas/process/user-testing/sus-system-usability-scale-ux/> (accessed Jun. 08, 2021).
- [36] F. J. Ariza-López, J. Rodríguez-Avi, and V. Alba-Fernández, "Control estricto de matrices de confusión por medio de distribuciones multinomiales," *GeoFocus. Rev. Int. Cienc. y Tecnol. la Inf. Geográfica*, no. 21, pp. 215–226, 2018.
- [37] M. de E. del Ecuador, "EDUCACIÓN CULTURAL Y ARTÍSTICA." Accessed: Apr. 19, 2021. [Online]. Available: <https://educacion.gob.ec/wp-content/uploads/downloads/2016/08/ECA-completo.pdf>.
- [38] "Quickstart: Detect faces in an image with the Azure REST API and Python - Azure Cognitive Services | Microsoft Docs." <https://docs.microsoft.com/en-us/azure/cognitive-services/face/quickstarts/python> (accessed Jun. 04, 2020).
- [39] Microsoft, "Cognitive Services APIs Reference," 2017. <https://westus.dev.cognitive.microsoft.com/docs/services/563879b61984550e40cbbe8d/operations/563879b61984550f30395236> (accessed Apr. 12, 2021).
- [40] H. V. Corporation, "Pentaho Documentation," 2021. https://help.pentaho.com/Documentation/9.1/Products/Add_sequence (accessed May 15, 2021).
- [41] RapidMiner, "Documentación de RapidMiner," 2021. <https://docs.rapidminer.com/9.0/studio/installation/> (accessed May 23, 2021).

- [42] M. Hofmann and R. Klinkenberg, *RapidMiner: Data mining use cases and business analytics applications*. CRC Press, 2010.
- [43] L. Sánchez Nogales, “UOCbuntu: Distribución Ubuntu orientada a estudiantes de Ingeniería Informática,” 2012.
- [44] “Installing MySQL JDBC Drivers in Pentaho Data Integration and BA Server tools – Holowczak.com Tutorials.” <http://holowczak.com/installing-mysql-jdbc-drivers-in-pentaho-data-integration-and-ba-server-tools/> (accessed Feb. 20, 2021).

SIGLAS

- RE Robótica Educativa
- BI Business Intelligence
- DW Datawarehouse
- RM RapidMinner
- OLTP Online transaction processing
- ETL Extract, transform and loading
- DIS Data Item Sets - Conjunto de artículos
- DW Data Warehouse - Almacén de datos
- SO Sistema Operativo
- URL Uniform Resource Locator

DEFINICIONES

- Top-down: Los datos son extraídos de sistemas operacionales por procesos ETL y cargados en un área de almacenamiento donde son validados y consolidados en el DW Corporativo.
- Táctica: son las acciones que se van a implementar para lograr un objetivo.
- Estratégica: es la definición y organización de dichas acciones. La estrategia es el plan global que agrupa todas las tácticas a ejecutar.
- Métrica: Son aquellos datos expresados numéricamente que nos sirven para analizar el rendimiento de una determinada campaña de marketing online. Digamos que, gracias a las métricas, podemos saber si estamos cumpliendo un objetivo. Y, en caso contrario, podremos rectificar en tiempo real, siempre que se realice un seguimiento periódico.
- OLTP: Online transaction processing. Es un tipo de procesamiento que facilita y administra aplicaciones transaccionales, usualmente para entrada de datos y recuperación y procesamiento de transacciones (gestor transaccional). Los paquetes de software para OLTP se basan en la arquitectura cliente-servidor ya que suelen ser utilizados por empresas con una red informática distribuida. El término puede parecer ambiguo, ya que puede entenderse "transacción" en el contexto de las "transacciones computacionales" o de las "transacciones en bases de datos". También podría entenderse en términos de transacciones de negocios o comerciales. OLTP también se ha utilizado para referirse a la transformación en la que el sistema responde de inmediato a las peticiones del usuario. Un cajero automático de un banco es un ejemplo de una aplicación de procesamiento de transacciones comerciales. La tecnología OLTP se utiliza en innumerables aplicaciones, como en banca electrónica, procesamiento de pedidos, comercio electrónico, supermercados o industria.
- clasificación binaria: Es la tarea de clasificar los elementos de un conjunto en dos grupos sobre la base de una regla de clasificación. Los problemas típicos de clasificación binaria incluyen pruebas médicas para determinar si un paciente tiene cierta enfermedad o no.
- matriz de confusión : En el campo de la inteligencia artificial y en especial en el problema de la clasificación estadística, una matriz de confusión es una herramienta que permite la visualización del desempeño de un algoritmo que se emplea en aprendizaje supervisado.[36]

ANEXOS

Anexo I Convocatoria para el Taller de Robótica Educativa

Anexo II Talleres de Robótica Educativa

1. ORDENAR PARA RECICLAR
2. METAMORFOSIS DE LA RANA

Anexo III Fotografías de Talleres

Anexo IV Archivo JSON- Reconocimiento Facial

```

{
  "faceId": "1f2e7702-3c71-448f-8a97-648b050966d7",
  "top": 322,
  "left": 287,
  "width": 118,
  "height": 118,
  "smile": 0.001,
  "pitch": 0.7,
  "roll": 2,
  "yaw": 27.9,
  "gender": "male",
  "age": 5,
  "moustache": 0,
  "beard": 0,
  "sideburns": 0,
  "glasses": "NoGlasses",
  "anger": 0.001,
  "contempt": 0.002,
  "disgust": 0.001,
  "fear": 0,
  "happiness": 0.001,
  "neutral": 0.694,
  "sadness": 0.01,
  "surprise": 0.292,
  "blurLevel": "medium",
  "blurLevelValue": 0.54,
  "exposureLevel": "overExposure",
  "exposureLevelValue": 0.88,
  "noiseLevel": "low",
  "noiseLevelValue": 0.12,
  "eyeMakeup": false,
  "lipMakeup": false,
  "accessories": [],
  "foreheadOccluded": false,
  "eyeOccluded": false,
  "mouthOccluded": false,
  "bald": 0.05,
  "invisible": false,
  "hairColor": [
    {
      "color": "brown",
      "confidence": 0.99
    },
    {
      "color": "black",
      "confidence": 0.78
    },
    {
      "color": "red",
      "confidence": 0.36
    },
    {
      "color": "blond",
      "confidence": 0.23
    },
    {
      "color": "other",
      "confidence": 0.1
    },
    {
      "color": "gray",
      "confidence": 0.08
    }
  ],
  "phase": 3,
  "phaseName": "Compartir",
  "group": 1,
  "kidId": 0,
  "photoID": "sesionid_1578920298096Compartir_foto_1.png",
  "imageUrl":
  "https://deusgallet.000webhostapp.com/recursos/AAL_emotions/Grupo1/Compartir/sesionid_1578920298096Compartir_foto_1.png"
}

```

Anexo V Archivos de Excel o CSV de Reconocimiento Facial

Anexo VI Encuesta en Google Forms para director y maestros

Formulario para Profesores y Autoridades

A continuación encontrará preguntas que se pretenden contestar luego de haber utilizado el taller de robótica y a continuación opciones para definir la prioridad de la pregunta por usuario.

1. ¿Cuál fue la reacción de los estudiantes en cada una de las fases del taller?

1. ALTA- Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores

2. ¿Qué sentimiento presentaron con el uso del lego Promedio por edad y género)

1. ALTA- Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores

3.¿El taller de robótica sirvió para captar la atención del estudiante?

1. ALTA- Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores

4.¿Qué características tienen los niños según el grupo en el que se desarrollaron? (estados de ánimo), ¿Quién demostró entusiasmo y quién no?)

1. ALTA- Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores

5.¿Se pudo apreciar cambios de carácter entre fases?

1. ALTA- Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores

:::

6.¿En qué momento y en quienes se evidenciaron actitudes positivas o Negativas? (por fase, edad, género)

1. ALTA- Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores

7.¿Existió falta de apego en la realización de actividades motrices(taller)?

1. ALTA- Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores

8.¿A quiénes les costó conseguir resultados?

1. ALTA- Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores

9. ¿Quiénes son los estudiantes más motivados? (felicidad-Dominio)

1. ALTA- Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores

10. ¿Quiénes son los mejores estudiantes? (dominio)

1. ALTA- Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores

11. ¿Cuál es el taller que más agrada, impacta y llama la atención a los niños (a quienes, en qué fase, por género, edad)?

1. ALTA- Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores

12. ¿Los estudiantes si están aprendiendo?

1. ALTA- Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores

13. ¿Sirve este taller para enseñar otras materias?

1. ALTA- Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores

14. ¿Documentó ideas y pruebas importantes a lo largo de su proyecto y dio lo mejor cuando lo presentó a otros? (En la fase de compartir se sintió a gusto)

1. ALTA- Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores



15.¿Cuándo y quienes mostraron timidez?

1. ALTA- Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores

16.¿Qué niños tienen un determinado nivel de emocional?

1. ALTA- Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores

17.Una cosa que el niño hizo realmente bien.

1. ALTA- Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores

18. Una cosa el niño debe mejorar para la próxima vez.

1. ALTA - Autoridades
2. MEDIA - Autoridades
3. BAJA - Autoridades
4. ALTA - Profesores
5. MEDIA - Profesores
6. BAJA - Profesores

Es posible acceder a la presente encuesta mediante el link:
<https://docs.google.com/forms/d/1YQkvxJlqlcS4vJif6Hw0-6WJDkYJLE0hw44b1I159Nk/edit?usp=sharing>

Anexo VII Preguntas Recolectadas en Video Conferencia y encuestas con Usuarios Rector, Profesor, Padre de Familia y Estudiante

Profesor:

- La aplicación del taller ayuda al niño en los aspectos siguientes:
 - creación de modelos
 - desarrollo del pensamiento computacional
 - inspiración de los estudiantes en búsqueda y desarrollo de soluciones
 - desarrollo de planificación de su trabajo
 - toma de decisiones
 - desarrollo pensamiento secuencial en paralelo
 - desarrollo en el uso de recursos didácticos como mapas mentales
 - desarrollo de la lógica de programación
 - desarrollo de abstracción
 - estructurar soluciones
 - esquematizar ideas
 - diseñar, construir y probar soluciones (prototipos)
 - prerequisites.
- ¿Cuál es la reacción inicial de los estudiantes al explorar? (agrado, desagrado)
- ¿Cuál es tu argumentación del niño en lo que ha realizado? (de acuerdo a la actitud se puede motivar al estudiante)
- ¿Qué características o qué dificultades se presentaron en motricidad fina cuando encajaban los legos? (dificultad de uso del lego, pide ayuda, se frustra el niño)
- ¿Qué actitud tenía el niño cuando comenzó a crear su ingenio con los legos? (imaginativos, interpretativos, unos saben que hacer otros no, innovan, inventan, ej.: conexión de legos ramas para el docente y para el niño es un edificio, el docente respeta la apreciación del niño, se ve competencia, apertura para explorar, motivación a crear y explorar)
- ¿Cómo podremos afianzar los conocimientos adquiridos de motricidad fina con el uso de legos y el robot? (en base a destrezas se sigue ayudando con ejercicios vivenciales, en medio virtual usar recursos para captar atención del estudiante) (destrezas, ej.: OI.2.11. Expresar ideas a partir de la reflexión constructiva sobre sí mismo y sus experiencias, mediante creaciones artísticas y prácticas corporales propias de su entorno cultural.) [37]
- Cuantos niños crearon el producto más confiable y conciso
- ¿Qué características tienen los niños según el grupo en el que se desarrollaron? (¿estados de ánimo, quien demostró entusiasmo y quién no?)

- Cambio de carácter, Actitudes Positivas y Negativas, la integración, la falta de apego a la realización de actividades motrices, a quienes les cuesta, a causa matemática la seriación de colores quien escogió el mejor color
- Quien de los niños sabe realizar seriación, (ej. torres azules, negras, blancas, saben q es un número, que es numeral, secuencias lógico matemáticas)
- Quien hizo secuencias de colores, secuencias de patrones, según la forma, tamaño
- Quienes son los mejores estudiantes (H, M)
- Que destreza es mayor en un estudiante que en otro
- cuál es el taller que más agrada y llama la atención a los niños
- que habilidades se ven más desarrolladas y en que taller
- que talleres deben ser los que más se impartan
- Explorar. Documenta y usa su razonamiento en relación con la pregunta o el problema.
- Crear. Resuelve el problema o pregunta construyendo y programando su modelo y haciendo cambios cuando era necesario
- Compartir. Documentó ideas y pruebas importantes a lo largo de su proyecto y dio lo mejor cuando lo presentó a otros
- ¿A qué niño le gusta más una materia?
- ¿Es posible medir la timidez de un niño?
- Que niños tienen un determinado nivel de emocional.
- Temas de conocimiento si está aprendiendo
- experiencia fue positiva negativa Por qué?

Rector:

- Si este modelo de Robótica sirve para ponerlo como materia ver si están aprendiendo
- Sirve este taller para enseñar las materias.

Alumno:

- Crear. Hice mi mejor trabajo para resolver el problema o pregunta construyendo y programando mi modelo y haciendo cambios cuando era necesario
- Compartir. Documenté ideas y pruebas importantes a lo largo de mi proyecto y di lo mejor de mí cuando presenté a otros
- Una cosa que hice realmente bien fue ...

- Una cosa que quiero mejorar para la próxima vez es ...

Padre:

- ¿El niño si está aprendiendo?
- La experiencia fue positiva o negativa ¿Por qué?

Anexo VIII Servicio Azure Face

Incorpora reconocimiento facial a aplicaciones para ofrecer una experiencia de usuario fluida y muy segura. No es necesario tener conocimientos de aprendizaje automático. Las características incluyen: detección de caras que percibe rostros y atributos en una imagen; identificación de la persona que coincide con un individuo de su repositorio privado de hasta 1 millón de personas; reconocimiento de emociones percibidas que detecta una serie de expresiones faciales, como felicidad, desprecio, neutralidad y miedo; y el reconocimiento y la agrupación de caras similares en imágenes. [19]

Es importante tener en cuenta que las expresiones faciales por sí solas no representan el estado interno de las personas.

Con Face de Azure no se crea el reconocimiento facial desde cero, se utiliza una API [20] para ofrecer esta funcionalidad en sus aplicaciones.

En la Figura 0.1 se muestra el programa utilizado por Azure [38] para reconocimiento facial.

```
import requests
import json

#https://docs.microsoft.com/en-us/azure/cognitive-services/face/quickstarts/python

subscription_key = 'dc2be3a61c724d2f9f49023c86643642'

assert subscription_key

face_api_url = 'https://insrostorestudiantes.cognitiveservices.azure.com/face/v1.0/detect'

image_url =
'https://res.cloudinary.com/dggnilweu/image/upload/v1608953989/Grupo3/Lunes/Crear/fotoTomada1
9-2020-8_48_46_ymi9bd.png'

headers = {'Ocp-Apim-Subscription-Key': subscription_key}

params = {

    'returnFaceId': 'true',

    'returnFaceLandmarks': 'false',

    'returnFaceAttributes':
'age,gender,headPose,smile,facialHair,glasses,emotion,hair,makeup,occlusion,accessories,blur,exposu
re,noise',
}

response = requests.post(face_api_url, params=params,
                        headers=headers, json={"url": image_url})

with open('data.json', 'w') as file:
    json.dump(response.json(),file,indent=4)
```

Figura 0.1 Programa utilizado por Azure para reconocimiento facial

Se toma la URL que se genera de la imagen en el hosting público y se la ingresa en el programa de “face detection” en “image_url”.

Atributos y detección de caras

El servicio Azure Face ofrece algoritmos de IA que detectan, reconocen y analizan caras humanas en las imágenes, proporciona varias funciones de análisis facial distintas, que se describen en las secciones siguientes.

Verificación de caras

Verify API realiza una autenticación con dos caras detectadas o a partir de una cara detectada en un objeto personal. Evalúa si dos caras pertenecen a la misma persona.

Búsqueda de caras similares

Find Similar API realiza una coincidencia facial entre una cara objetivo y un conjunto de caras candidatas, y busca un conjunto más reducido de caras parecidas a la cara objetivo. Esto resulta útil para realizar una búsqueda de caras por imagen.

Se admiten dos modos de funcionamiento, matchPerson y matchFace. El modo matchPerson devuelve las caras parecidas tras filtrar por la misma persona mediante Verify API. El modo matchFace ignora el filtro de la misma persona. Devuelve una lista de caras candidatas similares que pueden o no pertenecer a la misma persona.

En la Figura 0.2 se muestra la cara objetivo:



Figura 0.2 Cara Objetivo - Mujer sonriendo [38]

En la Figura 0.3 se observa las imágenes que son las caras candidatas:



Figura 0.3 Caras Candidatas - Personas Sonriendo [38]

Cinco imágenes de personas sonriendo.

Al buscar cuatro caras parecidas, el modo `matchPerson` devuelve a y b, que muestran a la misma persona que la cara objetivo. El modo `matchFace` devuelve a, b, c y d: exactamente cuatro caras candidatas, aunque algunas no sean la misma persona que el objetivo o el nivel de similitud sea bajo.

Detección de caras

El servicio Face detecta rostros humanos en una imagen y devuelve las coordenadas del rectángulo de sus ubicaciones, como se muestra en la Figura 0.4. Si lo desea, la detección de caras puede extraer una serie de atributos faciales, como la posición de la cabeza, el género, la edad, las emociones, el vello facial y las gafas.



Figura 0.4 Atributos Faciales [38]

La detección de caras es la acción de búsqueda de caras humanas en una imagen y, opcionalmente, la devolución de distintos tipos de datos relacionados con las caras. La operación `Face - Detect` se utiliza para detectar caras en una imagen. Como mínimo, cada cara detectada corresponde a un campo `faceRectangle` de la respuesta. Este conjunto de coordenadas de píxeles para la parte izquierda, la superior, el ancho y el alto marca la cara localizada. Con estas coordenadas, puede obtener la ubicación de la cara y su tamaño. En la respuesta de la API, las caras se muestran en orden de tamaño de mayor a menor.

Id. de cara

El identificador de cara es una cadena de identificador único para cada cara detectada en una imagen. Puede solicitar un identificador de cara en su llamada `API Face - Detect` [39], el cual es creado por la API de detección y caducará 24 horas después de la llamada de detección. Para devolver esto, requiere que el parámetro `"returnFaceId"` sea verdadero.

Face Rectangle

De tipo `Object`. Es un área rectangular para la ubicación de la cara en la imagen: Superior, Izquierda, Ancho y Alto.

Puntos de referencia de cara

Los puntos de referencia de cara son un conjunto de puntos fáciles de encontrar en una cara como, por ejemplo, las pupilas o la punta de la nariz. De forma predeterminada, existen 27 puntos de referencia predefinidos, la Figura 0.5 muestra dichos puntos:

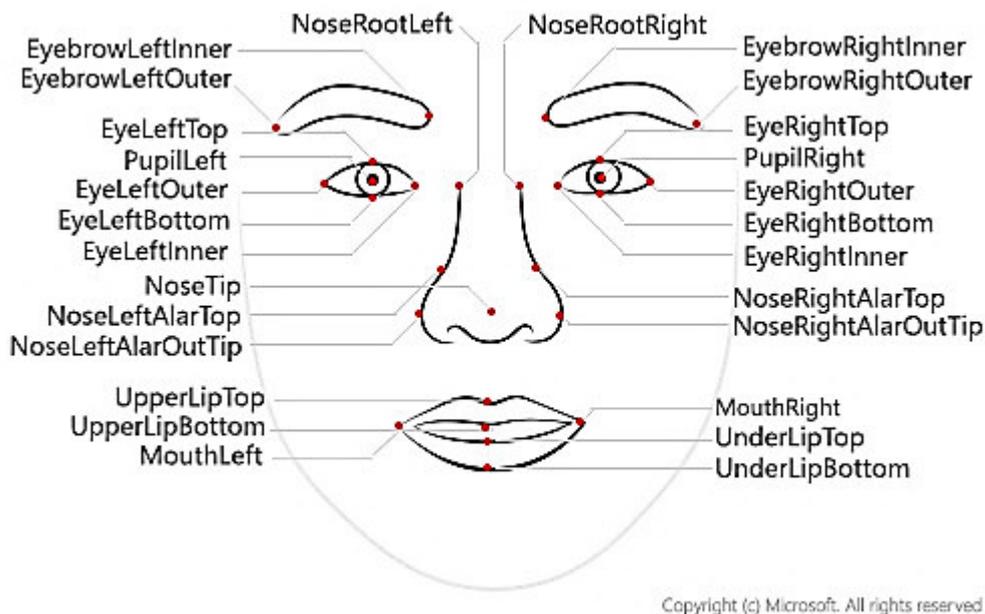


Figura 0.5 Diagrama de cara - puntos de referencia etiquetados [19]

Las coordenadas de los puntos se devuelven en unidades de píxeles.

Atributos

Los atributos son un conjunto de características que la API Face - Detect puede, opcionalmente, detectar. Se pueden detectar los siguientes atributos:

Desenfoco. Desenfoco de la cara de la imagen. Este atributo devuelve un valor entre cero y uno, así como una clasificación informal de bajo, medio o alto.

Emoción. Lista de emociones con una confianza de detección para la cara determinada. Las puntuaciones de confianza están normalizadas y las puntuaciones de todas las emociones suman uno. Las emociones detectadas son felicidad, tristeza, neutralidad, ira, desprecio, asco, sorpresa y temor.

Exposición. Exposición de la cara de la imagen. Este atributo devuelve un valor entre cero y uno, así como una clasificación informal de underExposure, goodExposure o overExposure.

Vello facial. Presencia estimada de vello facial y longitud de la cara determinada.

Pelo. Tipo de pelo de la cara. Este atributo muestra si el pelo está visible, así como la calvicie y el color de pelo que se detecten.

Posición de la cabeza. Orientación de la cara en un espacio 3D. Este atributo se describe mediante los ángulos de rotación alrededor del eje x (pitch), de rotación y de rotación alrededor del eje y (yaw) en grados. Los intervalos de valores son de -90 a 90 grados, de -90 a 90 grados y de -90 a 90 grados, respectivamente. En la Figura 0.6 se muestra el diagrama de asignaciones de ángulos:

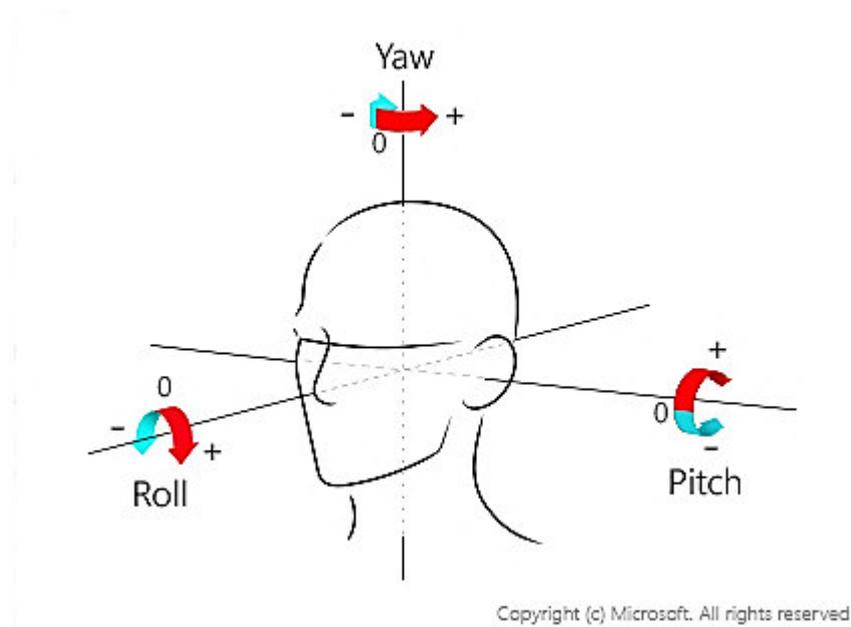


Figura 0.6 Cabeza con ejes de rotación alrededor del eje x (pitch), de rotación y de rotación alrededor del eje y (yaw) etiquetados [19]

Maquillaje. Si la cara está maquillada. Este atributo devuelve un valor booleano para eyeMakeup y lipMakeup.

Ruido. Ruido visual detectado en la cara de la imagen. Este atributo devuelve un valor entre cero y uno, así como una clasificación informal de bajo, medio o alto.

Oclusión. Si hay objetos que bloquean partes de la cara. Este atributo devuelve un valor booleano para eyeOccluded, foreheadOccluded y mouthOccluded.

Los atributos de caras se pueden predecir mediante algoritmos estadísticos. No obstante, es posible que no sean siempre precisos. Se debe tener cuidado al tomar decisiones basadas en datos de atributos.

Datos de entrada

Utilice las siguientes sugerencias para asegurarse de que las imágenes de entrada proporcionan los resultados de detección más precisos:

- Los formatos de imagen de entrada admitidos son JPEG, PNG, GIF para el primer fotograma y BMP.
- El tamaño del archivo de imagen no debe ser superior a 6 MB.
- El tamaño mínimo detectable de la cara es de 36 x 36 píxeles en una imagen no superior a 1920 x 1080 píxeles. Las imágenes de más de 1920 x 1080 píxeles tienen un tamaño mínimo detectable de la cara proporcionalmente mayor. Reducir el tamaño de la cara podría provocar que no se detecten algunas caras, aunque sean mayores que el tamaño mínimo detectable.
- El tamaño máximo de cara detectable es 4096 x 4096 píxeles.
- Las caras fuera del intervalo de tamaño de 36 x 36 a 4096 x 4096 píxeles no se detectarán.
- Es posible que no se puedan detectar algunas caras debido a desafíos técnicos. Unos ángulos de cara extremos (posición de la cabeza) o una oclusión de la cara (objetos como gafas de sol o manos que bloquean parte de la cara) pueden afectar a la detección. Las caras de frente y casi de frente proporcionan los mejores resultados.

Si está detectando caras de una fuente de vídeo, puede mejorar el rendimiento mediante el ajuste de determinados valores de la cámara de vídeo:

- **Suavizado:** muchas cámaras de vídeo aplican un efecto de suavizado. Debe desactivar esta opción, si es posible, porque crea un desenfoque entre fotogramas y reduce la claridad.
- **Velocidad de obturación:** La velocidad de obturación o velocidad de disparo, corresponde al inverso del tiempo de exposición en una fotografía y hace referencia al periodo durante el cual está abierto el obturador de una cámara fotográfica. Una velocidad de obturación más rápida reduce la cantidad de movimiento entre fotogramas y hace que cada fotograma sea más claro. Se recomienda utilizar velocidades de obturación de 1/60 de segundo o más rápidas.
- **Ángulo de obturación:** algunas cámaras especifican el ángulo de obturación en lugar de la velocidad de obturación. Debe usar un ángulo de obturación inferior si es posible. Esto generará unos fotogramas de vídeo más claros.

Una cámara con un ángulo de obturación inferior recibirá menos luz en cada fotograma, por lo que la imagen será más oscura. Deberá determinar el nivel adecuado que vaya a usar.

Anexo IX Ejemplos de fotografías con ruido

1. En esta Fotografía 5.1 se presenta un estudiante de género masculino con lentes sin embargo el algoritmo detecta un rostro de género masculino sin lentes



Fotografía 0.1 Fuente: https://firebasestorage.googleapis.com/v0/b/imagenes-aal.appspot.com/o/imagenes%2Fprofile_4lafm632926?alt=media&token=ad040323-63d8-43c2-b152-50f329b24956

2. En la Fotografía 5.2 hay 4 individuos, pero los rostros de los niños no se pueden apreciar y se encuentra un intruso que es un adulto, el algoritmo lo detecta como un rostro masculino y edad de 35.



Fotografía 0.2 Fuente: sesionid_1578920298096Crear_foto_194.png

3. En la Fotografía 5.3 a pesar de tener 3 rostros no se detecta ninguno.



Fotografía 0.3 Fuente: sesionid_1578920298096Explorar_foto_19.png

4. En la Fotografía 5.4 se dispone de 4 rostros sin embargo como la fotografía se encuentra pixelada y es de baja calidad el algoritmo no reconoce los rostros.



Fotografía 0.4 Fuente:

https://res.cloudinary.com/dggnilweu/image/upload/v1608953967/Grupo3/Lunes/Crear/fotoTomada19-2020-8_19_26_wvrafq.png

5. En la presente fotografía hay 2 niños y el algoritmo detecta dos rostros, pero uno de ellos lo detecta como género femenino. Por este motivo se selecciona al azar las fotografías para su revisión en el reconocimiento facial mediante el uso del Api Rest de Azure y realizar la corrección en el género en el caso de que amerite.



Fotografía 0.5 Fuente: sesionid_1578920298096Explorar_foto_22.png

6. En la fotografía 5.6 hay 3 rostros de un niño y 2 niñas sin embargo se detecta a un rostro de género femenino de 25 años.



Fotografía 0.6 Fuente:

https://res.cloudinary.com/dggnilweu/image/upload/v1608953968/Grupo3/Lunes/Crear/fotoTomada19-2020-8_20_26_r9rsor.png

7. Algunos niños cambian de grupo como se evidencia en la Fotografía 5.7, durante las fases del taller como es el caso de la niña de la fotografía por lo que se afectan los resultados, por ejemplo, en cuanto a emoción por grupo.



Fotografía 0.7 Fuente: sesionid_1578920304418Crear_foto_163.png

8. En la Fotografía 5.8 hay 2 niños, el algoritmo reconoce a uno de ellos y no es posible identificar a que niño corresponde el análisis por lo cual es necesario volver a correr el Api Rest de Azure. Cabe mencionar que la detección de emociones depende de algunos factores como por ejemplo si el rostro está de perfil no se puede leer la emoción, de igual manera si no se presenta la boca, si se presenta un rostro de frente, si se gira no se detectan las emociones, etc.



Fotografía 0.8 Fuente: sesionid_1578920298096Explorar_foto_19.png

Anexo X Tutorial de Pentaho Data Integration. Kettle- Spoon

Spoon es el diseñador gráfico de transformaciones y trabajos del sistema de ETTLS de Pentaho Data Integration (PDI), también conocido como Kettle (acrónimo de "Kettle E.T.T.L. Environment").

Está diseñado para ayudar en los procesos ETLs, que incluyen la Extracción, Transformación, Transporte y Carga de datos. [40]

Instalación de Spoon

Para instalar Spoon seguir las siguientes instrucciones:

1. Instalar Java Runtime Environment (JRE) de Sun Microsystems, versión 1.5 o superior, sitio descarga <http://www.javasoft.com/>.
2. Descargar Kettle, sitio descarga: <http://sourceforge.net/projects/pentaho/files/Data%20Integration/>
3. Descomprimir el archivo recientemente descargado

Inicio de uso de Kettle-spoon

Ingresar a la terminal y digitar los comandos que se muestran en la Figura 0.7.



Figura 0.7 Spoon: Arranque

Seguidamente, se presenta el entorno de trabajo en Spoon, Figura 0.8

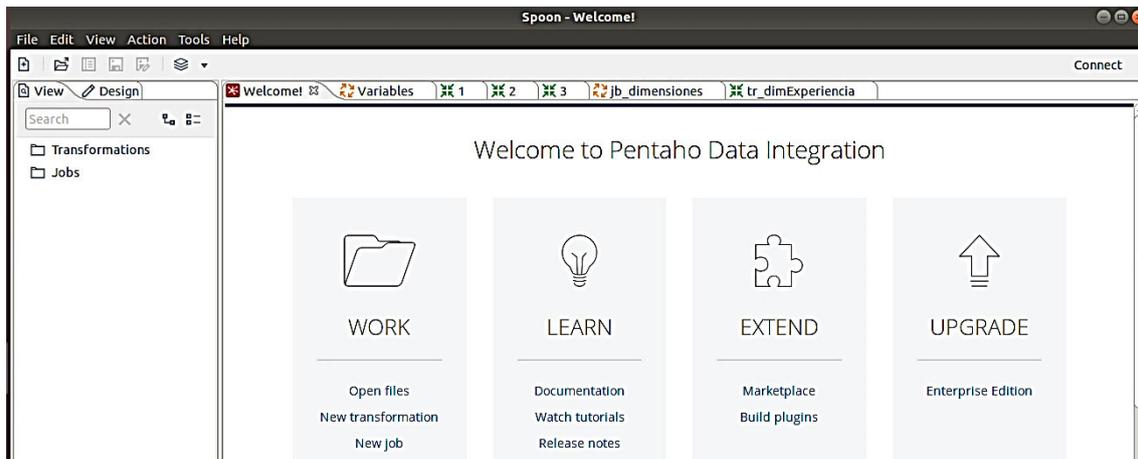


Figura 0.8 Spoon: Entorno

A continuación, se describen algunas de las funcionalidades empleadas en el presente proyecto:

Configurar una conexión a la base de datos

En la Figura 0.9 se muestra la conexión a base de datos, que representa la información de conexión de una base de datos específica y actúa como padre en la jerarquía de todas las tablas físicas y columnas físicas definidas para esa base de datos.

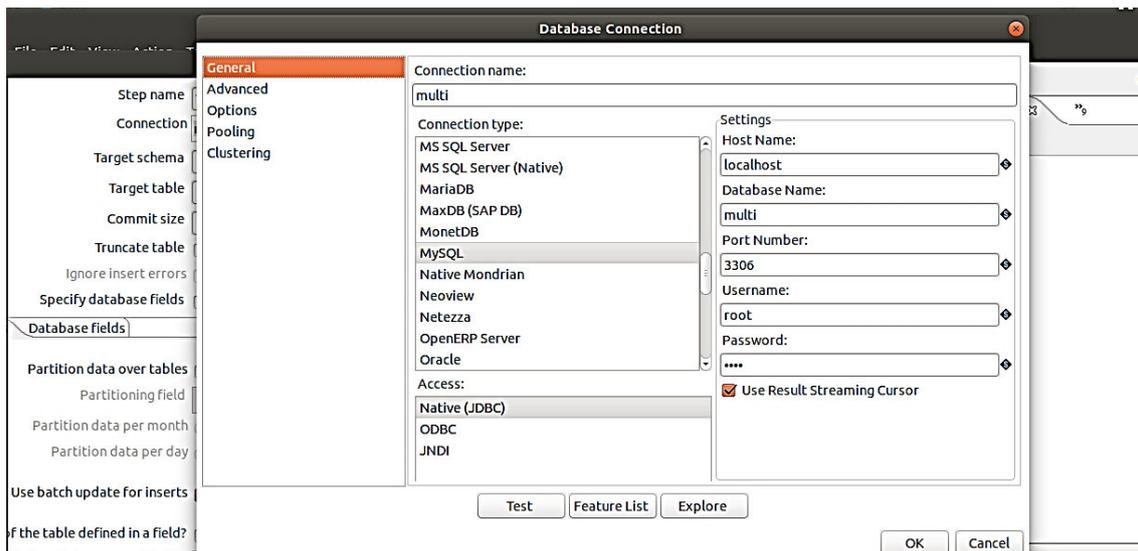


Figura 0.9 Spoon: Conexión a Base de Datos

Para crear una nueva conexión realice lo siguiente:

1. Hacer clic con el botón derecho en Conexiones en el panel de navegación.
2. Seleccionar Nueva conexión . Aparece el cuadro de diálogo Conexión a la base de datos .
3. Ingresar un nombre de conexión .
4. Definir la conexión de su base de datos según sea necesario.

5. Hacer clic en Probar para asegurarse de que sus entradas sean correctas.
6. Hacer clic en Aceptar para guardar su conexión. Aparece la lista de tablas de su base de datos que se van a importar. Si se desea encontrar una tabla específica, ingresar el nombre de la tabla en el cuadro de búsqueda o use una expresión regular. Seleccionar las tablas que se desea importar y hacer clic en Aceptar o Cancelar cuando se haya terminado.

Transformaciones

Es una red de tareas lógicas llamadas pasos, esencialmente flujos de datos o un gráfico dirigido de un conjunto lógico de configuraciones de transformación de datos.

Los nombres de los archivos de transformación tienen la extensión .ktr.

La entrada Transformación ejecuta una transformación previamente definida dentro de un trabajo. En la Figura 0.10 se muestra una Transformación que es el punto de acceso del trabajo a la actividad ETL.

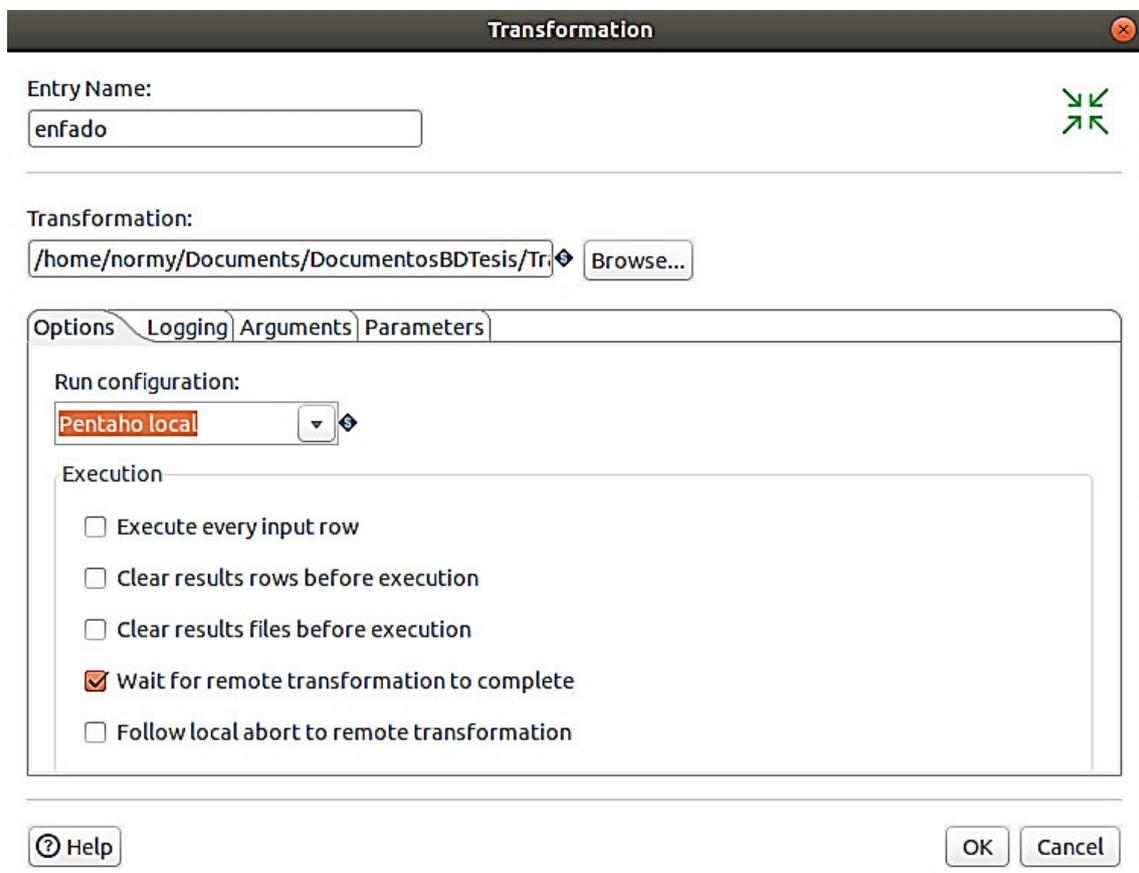


Figura 0.10 Spoon: Transformación

Opciones

Nombre de entrada: Especifica el nombre exclusivo de la entrada de trabajo en el lienzo.

Transformación: Especifique su transformación ingresando en su ruta o haciendo clic en Examinar.

Ejecuta cada fila de entrada: Ejecuta la transformación una vez para cada fila de entrada (bucle).

Borrar filas de resultados antes de la ejecución: Se asegura de que las filas de resultados se borren antes de que comience la transformación.

Borrar archivos de resultados antes de la ejecución: Se asegura de que los archivos de resultados se borren antes de que comience la transformación.

Espere a que termine la transformación remota: Si seleccionó Servidor como su tipo de entorno, elija esta opción para bloquear el trabajo hasta que la transformación se ejecute en el servidor.

Siga el aborto local a la transformación remota: Si seleccionó Servidor como su tipo de entorno, elija esta opción para enviar la señal de cancelación local de forma remota.

Pestaña de Registro

Especificar archivo de registro: Especifica un archivo de registro independiente para ejecutar esta transformación.

Nombre: Especifica el directorio y el nombre base del archivo de registro

Extensión: Especifica la extensión del nombre del archivo.

Nivel de registro : Especifica el nivel de registro para ejecutar la transformación.

Adjuntar archivo de registro: Agrega el archivo de registro en lugar de crear uno nuevo.

Crear carpeta principal: Crea una carpeta principal para el archivo de registro si no existe.

Incluir fecha en el nombre del archivo: Agrega la fecha del sistema al nombre del archivo con el formato AAAAMMDD (_20051231).

Incluir la hora en el nombre del archivo: Agrega la hora del sistema al nombre del archivo con formato HHMMSS (_235959).

Pestaña Argumentos

Copiar resultados a argumentos: Copia los resultados de una transformación anterior como argumentos de la transformación mediante el paso Copiar filas en el resultado.

Argumentos: Especifique qué argumentos de la línea de comandos se pasarán a la transformación.

Pestaña Parámetros

Copiar resultados a parámetros: Copia los resultados de una transformación anterior como parámetros de la transformación utilizando el paso Copiar filas en el resultado.

Pasar los valores de los parámetros a la subtransformación: Pase todos los parámetros del trabajo a la subtransformación.

Parámetros: Especifique el nombre del parámetro pasado a la transformación.

Nombre de la columna de transmisión: Especifique el campo de un registro entrante de una transformación anterior como parámetro.

Valor: Especifique los valores de los parámetros de transformación mediante una de las siguientes acciones:

Obtener parámetros: Obtenga los parámetros existentes ya asociados por la transformación.

Inicio

En la Figura 0.11 se observa el componente Inicio, que permite indicar el agendamiento de horario en el cual se ejecuta el trabajo.

Figura 0.11 Spoon: Inicio

Nombre de Entrada del Trabajo: Especifica el nombre exclusivo de la entrada de inicio.

Repetir: seleccione esta opción si desea repetir la programación automáticamente.

Tipo: elija el tipo de programación (diario, días de la semana, día del mes).

Intervalo en segundos: informa el intervalo en segundos para repetir el proceso.

Intervalo en minutos: informa el intervalo en minutos para repetir el proceso.

Hora del día: informa la hora para ejecutar los procesos.

Día de la semana: informa el día de la semana para ejecutar los procesos.

Día del mes: informa el día del mes para ejecutar los procesos.

Generar filas

En la Figura 0.12 se presenta el componente Generar Filas que muestra como resultado un número específico de filas. Por defecto, las filas están vacías; sin embargo, pueden contener varios campos estáticos. Este paso se utiliza

principalmente con fines de prueba. Puede ser útil para generar un número fijo de filas, por ejemplo, desea exactamente 12 filas durante 12 meses. A veces, puede usar Generar filas para generar una fila que es un punto de inicio para su transformación.

Step name:

Limit:

Never stop generating rows:

Interval in ms (delay):

Current row time field name:

Previous row time field name:

Fields:

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Value	Set empl
1 fechalnicial	Date	yyyy-MM-dd						2020-01-12	N

Buttons: Help, OK, Preview, Cancel

Figura 0.12 Spoon: Generar Filas

Opciones

Nombre del paso: Este nombre debe ser único en una única transformación.

Límite: Establece el número máximo de filas que desea generar.

Nunca dejes de generar filas: Esta opción está destinada a casos de uso en tiempo real en los que nunca desea detener la transformación en ejecución.

Intervalo en ms: El intervalo entre filas generadas en milisegundos.

Nombre del campo de tiempo de fila actual: El nombre de campo * opcional * para el campo Fecha que contiene la hora en la que se generó la fila actual.

Nombre del campo de tiempo de fila anterior: El nombre de campo * opcional * para el campo Fecha que contiene la hora en la que se generó la fila anterior.

Campos: Esta tabla es donde se configura la estructura y los valores de las filas que estás generando (opcional).

Agregar secuencia

En la Figura 0.12 se muestra el componente Agregar secuencia que adiciona una secuencia al flujo PDI. Una secuencia es un valor entero en constante cambio con un valor de inicio e incremento específico. Puede utilizar una secuencia de base de datos para determinar el valor de la secuencia o hacer que la genere PDI. Los valores de

secuencia generados por PDI son únicos cuando se utilizan en la misma transformación y vuelven al mismo valor inicial cada vez que ejecuta la transformación.

The screenshot shows the 'Add sequence' dialog box with the following fields and options:

- Step name:** Secuencial
- Name of value:** Secuencial
- Use a database to generate the sequence:**
 - Use DB to get sequence?**
 - Connection:** [Dropdown menu] [Edit...] [New...] [Wizard...]
 - Schema name:** [Text field] [Schemas...]
 - Sequence name:** SEQ_ [Sequences...]
- Use a transformation counter to generate the sequence:**
 - Use counter to calculate sequence?**
 - Counter name (optional):** [Text field]
 - Start at value:** 1 [Spin button]
 - Increment by:** 1 [Spin button]
 - Maximum value:** 99999999 [Spin button]
- Buttons:** [Help] [OK] [Cancel]

Figura 0.13 Spoon: Secuencia

Opciones

Nombre del paso: Especifique el nombre exclusivo del paso en el lienzo. El nombre del paso se establece en Secuencia de pasos de forma predeterminada.

Nombre del valor: Especifique el nombre del nuevo valor de secuencia para agregar al flujo de PDI. El nombre del valor se establece en valuenam de forma predeterminada.

Secuencia generada en la base de datos

¿Usar DB para generar la secuencia?: Seleccione esta casilla de verificación si desea que la secuencia sea impulsada por una secuencia de base de datos.

Conexión: Seleccione la conexión donde reside la secuencia de su base de datos. Si no tiene una conexión existente, haga clic en Nuevo o Asistente . Si necesita modificar una conexión existente, haga clic en Editar .

Nombre de esquema (opcional): Especifique el nombre de esquema de cualquier tabla relacionada. Haga clic en Esquemas para seleccionar un esquema dentro de la base de datos que especificó para Conexión .

Nombre de la secuencia: Especifique el nombre de la secuencia de la base de datos. Haga clic en Secuencias para seleccionar una secuencia dentro de la base de datos que especificó para Conexión .

Secuencia generada por el contador de transformación PDI

¿Usar contador para calcular la secuencia?: Seleccione esta casilla de verificación si desea que la secuencia sea generada por PDI. Esta opción está configurada de forma predeterminada

¿Usar DB para generar la secuencia?: se selecciona automáticamente si esta casilla de verificación está desactivada.

Nombre del contador (opcional): Especifique el nombre del contador si varios pasos en una transformación generan el mismo nombre de valor para evitar una secuencia única en varios pasos.

Empiece por el valor: Especifique el valor para iniciar la secuencia. Por ejemplo, si el valor Iniciar en se establece en 1, la secuencia comenzará con 1 . Si el valor Iniciar en se establece en 0, la secuencia comenzará con 0 .

Incremento por: Especifique la cantidad en la que la secuencia aumenta o disminuye.

Valor máximo: Especifique el valor a alcanzar antes de regresar la secuencia a la configuración del valor Comenzar en.

Calculadora

En la Figura 0.14 se presenta el componente Calculadora que proporciona funciones predefinidas que puede ejecutar en valores de campo de entrada. Este paso de la Calculadora es una alternativa fácil y rápida al JavaScript personalizado que se usa comúnmente para los cálculos. Para usar, especifique los campos de entrada y el tipo de función a realizar y devolver los resultados. También puede especificar un campo para eliminar del resultado (salida) después de que se hayan calculado todos los valores, lo cual es útil para eliminar valores temporales.

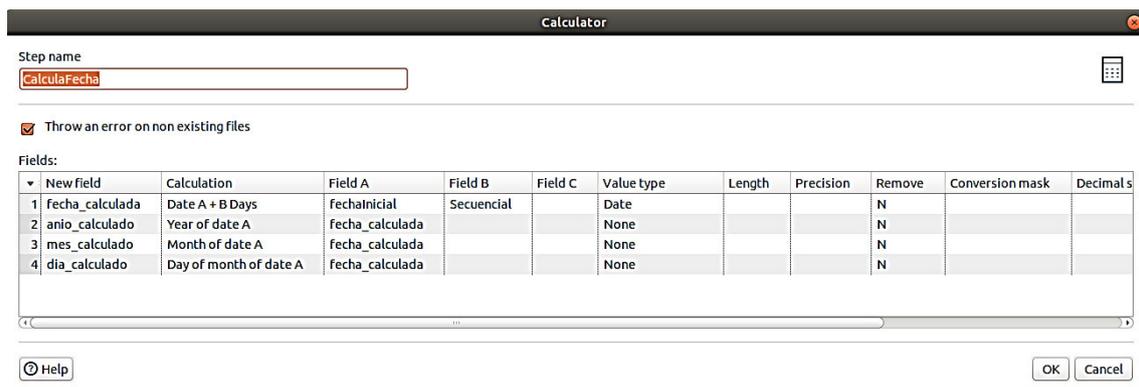


Figura 0.14 Spoon: Calculadora

Opciones

Nombre del paso: especifique el nombre único del paso de la Calculadora en el lienzo. Puede personalizar el nombre o dejarlo como predeterminado.

Nuevo campo: Especifique el nombre del campo.

Cálculo: Cuando hace clic o presiona en este campo, aparece el cuadro de diálogo "Seleccionar el tipo de cálculo". Ingrese la función de calculadora para usar en la transformación. Utilice el campo Filtro para buscar una función específica.

Campo A, B y C: Introduzca los valores para ejecutar la función de calculadora especificada en el campo Cálculo.

Tipo de valor: Seleccione el tipo de datos del campo de la lista desplegable o ingréselo manualmente.

Largo: Especifique la longitud del campo, de acuerdo con los siguientes tipos de campo:

Número: número total de cifras significativas en un número.

Cadena: longitud total de la cadena.

Fecha: longitud de la salida impresa de la cadena (por ejemplo, 'cuatro' es la longitud de un año).

Precisión: Especifique el número de dígitos de coma flotante para campos de tipo numérico.

Eliminar: Elimine este campo del resultado (salida) después de que se hayan calculado todos los valores. Este campo es útil para eliminar valores temporales. Seleccione 'N (o)' o 'Y (es)'. .

Máscara de conversión: Especifique un formato para el campo, como un formato de fecha.

Símbolo decimal: Especifique el símbolo utilizado para representar un punto decimal, ya sea un punto '.' o una coma ', '. Por ejemplo, 5,000.00 o 5.000,00 .

Símbolo de agrupación: Especifique el método utilizado para separar unidades de miles en números de cuatro dígitos o más, ya sea un punto '.' o una coma ', '. Por ejemplo, 5.000 o 5.000 .

Símbolo de moneda: Especifique el símbolo utilizado para representar las monedas, por ejemplo, '\$' o '€' .

El paso de la calculadora también le permite indicar si desea que se muestren errores para la siguiente condición:

Seleccione la casilla de verificación Lanzar un error en archivos no existentes si desea que la transformación genere un error cuando no haya archivos para procesar.

En la Tabla 0.1 se encuentran algunas de las funciones de la calculadora.

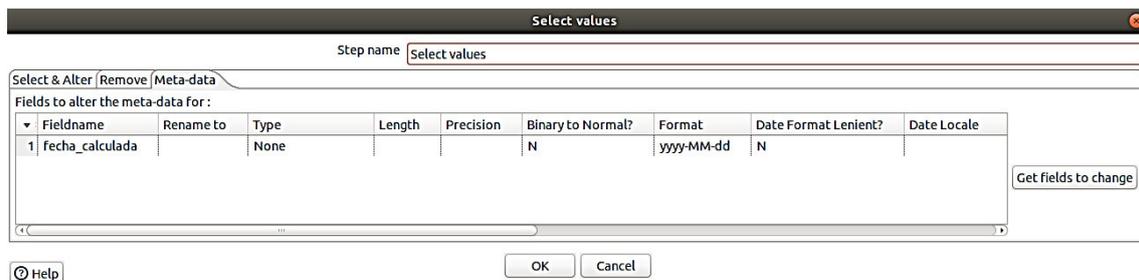
Tabla 0.1 Funciones Calculadora

FUNCIÓN	DESCRIPCIÓN	CAMPOS REQUERIDOS
Fecha A + B Días	Agrega B días al campo Fecha A. <i>Solo se admiten valores enteros para B. Si necesita cálculos que no sean enteros, agregue un segundo cálculo con horas.</i>	A y B
Año de la fecha A	Calcula el año de la fecha A.	A
Mes de la fecha A	Calcula el mes de la fecha A.	A
Día del año de la fecha A	Calcula el día del año (1-365).	A
Día del mes de la fecha A	Calcula el día del mes (1-31).	A

Seleccionar valores

En la Figura 0.15 se presenta el componente “Seleccionar valores” que permite realizar todas las siguientes acciones en los campos del flujo de datos PDI:

- Seleccionar valores
- Quitar valores
- Cambiar el nombre de los valores
- Cambiar tipos de datos
- Configurar longitud y precisión de valores

**Figura 0.15** Spoon: Selección Valores

Opciones

Nombre del paso: Para especificar el nombre exclusivo del paso Seleccionar valores en el lienzo. Puede personalizar el nombre o dejarlo como predeterminado.

Pestaña Seleccionar y alterar

La tabla de esta pestaña contiene las siguientes columnas para los campos que desea seleccionar, renombrar o cambiar la longitud y la precisión.

Nombre del campo: El nombre del campo del flujo de entrada.

Cambiar nombre a: El nuevo nombre del campo. Déjelo en blanco si no desea cambiar el nombre del campo.

Tipo: Seleccione un tipo de dato para el campo

Largo: La longitud del campo

Precisión: El número de dígitos de coma flotante para campos de tipo numérico.

Utilice Obtener campos para seleccionar y agregar todos los campos de los flujos de entrada.

Utilice Editar asignación para abrir el cuadro de diálogo Editar asignación, para definir fácilmente múltiples asignaciones entre los campos de origen y destino.

Seleccione la opción Incluir campos no especificados, ordenados por nombre para incluir campos en los flujos de entrada que no están seleccionados explícitamente en la sección Campos.

El cuadro de diálogo Editar asignación contiene las siguientes opciones:

Campos de origen: Seleccione para resaltar los campos de la transmisión entrante que desea mapear.

Campos de destino: Seleccione para resaltar los campos en la tabla de salida que desea mapear.

Agregar (botón): Haga clic en Agregar para mover una combinación seleccionada de un nombre de campo y un nombre de columna al panel Asignaciones.

Eliminar (botón): Haga clic en Eliminar para mover una combinación mapeada del panel

Asignaciones a los paneles Campos de origen y Campos de destino

Mapeos: El mapeo de campo a columna que designa el campo entrante y la columna de la tabla donde se asignará el campo.

¿Selección automática de objetivos?: Seleccione para que el paso realice la asignación a un objetivo.

¿Ocultar los campos de origen asignados?: Seleccione esta opción para eliminar un campo del panel Campos de origen cuando haya coincidido y se haya movido al panel Asignaciones.

Selección de fuente automática: Seleccione para seleccionar automáticamente un campo del panel Campos de origen cuando se selecciona un nombre de columna.

¿Ocultar los campos de destino asignados?: Seleccione esta opción para eliminar un campo del panel Campos de destino cuando haya coincidido y se haya movido al panel Asignaciones.

Adivinar (botón): Haga clic en Adivinar para realizar una coincidencia automática de todos los campos y la población del panel Asignaciones.

Los resultados de su asignación se muestran en el panel Asignaciones .

Pestaña Eliminar

Esta pestaña le permite eliminar campos del flujo de entrada. Haga clic en Obtener campos para eliminar para agregar campos de la secuencia de PDI entrante.

Las eliminaciones de campos pueden ralentizar el proceso, ya que el orden de los campos se cambia internamente.

Pestaña de metadatos

La tabla de esta pestaña contiene las siguientes columnas para los campos a los que desea cambiar el nombre de los campos de entrada, convertirlos a diferentes tipos de datos y modificar su longitud y precisión.

Nombre del campo: El nombre del campo importado.

Si cambió el nombre de un campo en la pestaña Seleccionar y modificar, el nombre aquí debe coincidir con el valor renombrado.

Cambiar nombre a: El nuevo nombre del campo. Déjelo en blanco si no desea cambiar el nombre del campo.

Tipo: El tipo de datos para este campo.

Largo: La longitud del campo.

Precisión: El número de dígitos de coma flotante para campos de tipo numérico.

¿Binario a normal?: Especifique Y o N para convertir una cadena en un tipo de datos numéricos, cuando corresponda.

Formato: La máscara de formatos numéricos y de fecha válidos comunes.

Formato de fecha ¿Indulgente?: Especifica si el analizador de fechas es estricto o indulgente. Establezca en N para aceptar solo valores de fecha estrictamente válidos. Cuando se establece en Y, el analizador intenta determinar la intención de una fecha incorrecta, si es posible, y corregirla.

Fecha Local: La configuración regional de la fecha que se utilizará para las conversiones y los cálculos de fechas. Déjelo en blanco para usar la codificación predeterminada en su sistema.

Fecha Zona horaria: La zona horaria de la fecha que se utilizará para los cálculos y conversiones de fechas. Déjelo en blanco para usar la codificación predeterminada en su sistema.

Conversión de números indulgente: Especifica si se debe informar de un error al analizar números. Cuando se establece en Y, los números se analizan hasta que se encuentra un valor no numérico, como un guion o una barra. A continuación, el análisis se detiene sin informar de un error. Cuando se establece en N, el analizador informa un error en caso de que haya números no válidos en la entrada.

Codificación: La codificación del archivo de texto que se utilizará. Déjelo en blanco para usar la codificación predeterminada en su sistema. La lista se completa a partir de las codificaciones disponibles en su sistema.

Decimal: El punto decimal como un punto o una coma.

Agrupamiento: El separador para unidades de miles como un punto o una coma.

Divisa: El símbolo de la moneda.

Tabla de Entrada

En la Figura 0.16 se muestra el paso de entrada de tabla que lee información de una base de datos conectada utilizando sentencias SQL. Las sentencias SQL básicas se pueden generar automáticamente haciendo clic en el botón Obtener sentencia de selección SQL.

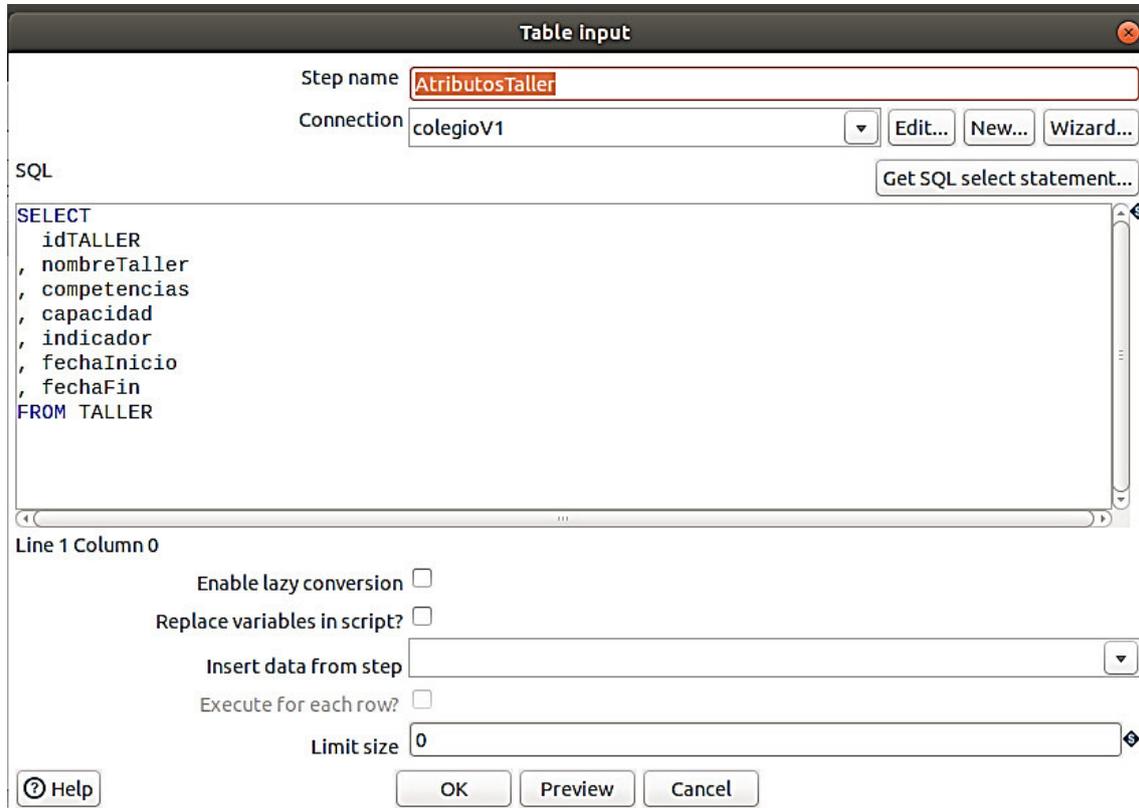


Figura 0.16 Spoon: Tabla de Entradas

Opciones

Nombre del paso: Especifique el nombre exclusivo del paso Entrada de tabla en el lienzo. Puede personalizar el nombre o dejarlo como predeterminado.

Conexión: Seleccione el nombre de una base de datos conectada de la lista desplegable.

Editar (botón): Haga clic para editar su conexión de base de datos actual.

Nuevo (botón): Haga clic para establecer una nueva conexión a la base de datos.

Asistente (botón): Haga clic para abrir una nueva conexión a la base de datos utilizando el Asistente.

SQL: Especifique una declaración sentencias para leer información de la base de datos conectada. También puede hacer clic en el botón Obtener instrucción de selección de

SQL para examinar tablas y generar automáticamente una instrucción de selección básica.

Habilitar la conversión diferida: Seleccione esta opción para habilitar el algoritmo de conversión diferida. Cuando se selecciona, la conversión diferida evita conversiones de tipos de datos innecesarios cuando es posible, lo que puede mejorar significativamente el rendimiento.

¿Reemplazar variables en el script?: Seleccione esta opción para reemplazar variables en el script. Esta función proporciona capacidades de prueba con o sin sustituciones de variables.

Insertar datos del paso: Especifique el nombre del paso de entrada de donde PDI puede esperar que venga la información. Luego, esta información se puede insertar en la declaración SQL. El localizador donde PDI inserta información se indica con un signo de interrogación (?).

¿Ejecutar para cada fila?: Seleccione esta opción para ejecutar la consulta para cada fila individual.

Tamaño límite: Especifique el número de líneas para leer de la base de datos. Un valor de cero ('0') indica leer todas las líneas.

Vista previa (botón): sirve para abrir una nueva ventana y ver un registro de ejecución derivado de una transformación temporal con dos pasos: el paso de entrada de tabla y el paso ficticio. Para ver el registro, haga clic en Registros en la ventana Vista previa que se abre.

Tabla de Salida

Carga datos en una tabla de base de datos. En la Figura 0.17 se muestra el paso de Salida de tabla es equivalente al operador de SQL INSERT y es una solución cuando solo necesita insertar registros. Si solo desea actualizar filas, debe usar el paso Actualizar.

Este paso proporciona opciones de configuración para una tabla de destino y opciones relacionadas con el rendimiento, como Confirmar tamaño y Usar actualización por lotes para inserciones. Hay configuraciones de rendimiento específicas para un tipo de base de datos que se pueden establecer dentro de las propiedades JDBC de conexión de la base de datos. Si inserta un registro en una tabla de base de datos que tiene columnas de identidad, el controlador JDBC devuelve la clave generada automáticamente que utiliza al realizar la inserción.

Table output

Step name: TIEMPO

Connection: multiV1 [Edit... New... Wizard...]

Target schema: [Browse...]

Target table: dim_tiempo [Browse...]

Commit size: 1000

Truncate table:

Ignore insert errors:

Specify database fields:

Main options | Database fields

Partition data over tables:

Partitioning field: [Dropdown]

Partition data per month:

Partition data per day:

Use batch update for inserts:

Is the name of the table defined in a field?:

Field that contains name of table: [Dropdown]

Store the tablename field:

Return auto-generated key:

Name of auto-generated key field: [Text field]

[Help] [OK] [Cancel] [SQL]

Figura 0.17 Spoon: Tabla Salida

Opciones

Nombre del paso: Especifica el nombre exclusivo del paso Salida de tabla en el lienzo. Puede personalizar el nombre o dejarlo como predeterminado.

Conexión: Especifique la conexión a la base de datos donde desea que se escriban sus datos.

Editar (botón): Haga clic para editar su conexión de base de datos actual.

Nuevo (botón): Haga clic para establecer una nueva conexión a la base de datos.

Asistente (botón): Haga clic para abrir una nueva conexión a la base de datos utilizando el Asistente.

Esquema de destino: Especifique el nombre de su esquema de base de datos.

Tabla de destino: Especifique el nombre de la tabla donde desea escribir sus datos.

Comprometer tamaño: Especifique el tamaño del lote de confirmación. El tamaño es el número de instrucciones INSERT que se realizarán antes de enviar un comando COMMIT a la base de datos. La especificación de un tamaño de confirmación puede

cambiar el rendimiento. Las transacciones no son compatibles con todas las plataformas de bases de datos.

Truncar tabla: Seleccione para truncar la tabla antes de que se inserte la primera fila en la tabla. Si ejecutará la transformación en clústeres o con varias copias de este paso, debe truncar la tabla antes de iniciar la transformación.

Ignorar errores de inserción: Seleccione para ignorar todos los errores de inserción, como claves primarias violadas. Se registrarán un máximo de 20 advertencias. Esta opción no está disponible para inserciones por lotes.

Especificar campos de la base de datos: Seleccione para especificar los campos en la pestaña Campos de la base de datos. De lo contrario, todos los campos se insertan de forma predeterminada. Esta opción debe seleccionarse para usar las herramientas de asignación de campos Obtener e Ingresar en la pestaña Campos de la base de datos.

SQL (botón): Haga clic para abrir el Editor SQL simple .

La pestaña de Opciones principales contiene lo siguiente:

Partición de datos sobre tablas: Seleccione para dividir los datos en varias tablas según el valor del campo de fecha especificado en el campo Partición . Cuando se selecciona esta opción, los datos se insertan en tablas cuyos nombres siguen el patrón '<tabla-objetivo> _ <formato-fecha>'. Estas tablas deben crearse manualmente antes de ejecutar la transformación para que los datos se inserten en estas tablas. Consulte Datos de partición por mes / Datos de partición por día para obtener detalles sobre el formato de fecha.

Campo de particionamiento: Especifique el campo de fecha que desea utilizar para determinar cómo dividir los valores en varias tablas. Este valor se utiliza para generar un nombre de tabla fechado en el que se insertarán los datos.

Datos de partición por mes / Datos de partición por día: Determina el formato de fecha utilizado en el nombre de la tabla al particionar la tabla: 'aaaaMM' para 'por mes' y 'aaaaMMdd' para 'por día'.

Usar actualización por lotes para inserciones: Seleccione para usar inserciones por lotes. Esta opción agrupa las declaraciones INSERT, lo que limita los viajes de ida y vuelta a la base de datos. Esta opción está activada de forma predeterminada.

¿El nombre de la tabla está definido en un campo?: Seleccione para especificar el nombre de la tabla de destino en el campo que contiene el nombre de la tabla.

Campo que contiene el nombre de la tabla: Especifique el nombre del campo que se utilizará para el nombre de la tabla cuando la opción "¿Está definido el nombre de la tabla en un campo?" está seleccionado.

Almacenar el campo de nombre de la tabla: Seleccione para almacenar el nombre de la tabla en la salida.

Devolver clave generada automáticamente: Seleccione para devolver la clave que se generó al insertar una fila en la tabla.

Nombre del campo clave generado automáticamente: Especifique el nombre del nuevo campo en las filas de salida que contiene la clave generada automáticamente.

La pestaña de Campos de base de datos contiene lo siguiente:

Campo de tabla: El nombre de la columna de la base de datos en la que se insertan los datos.

Campo de flujo: El campo de flujo que se lee de los pasos de fuente / entrada y se inserta en la base de datos.

Obtener campos (botón): Haga clic para importar los campos de su fuente de datos a la tabla de campos de la base de datos. La opción Especificar campos de la base de datos en el área general debe estar seleccionada para que esta herramienta esté activa.

Ingresar mapeo de campo (botón): Haga clic para abrir la ventana Enter Mapping. La opción Especificar campos de la base de datos en el área general debe estar seleccionada para que esta herramienta esté activa.

En la ventana de mapeo se tienen las siguientes opciones:

Campos de origen: La lista de nombres de campo de la transmisión entrante.

Campos de destino: Los campos de la tabla de salida.

Agregar (botón): Haga clic en Agregar para mover una combinación seleccionada de un nombre de campo y un nombre de columna al panel Asignaciones.

Eliminar (botón): Haga clic en Eliminar para mover una combinación mapeada del panel Asignaciones a los paneles Campos de origen y Campos de destino.

Mapeos: El mapeo de campo a columna que designa el campo entrante y la columna de la tabla donde se asignará el campo.

¿Selección automática de objetivos?: Seleccione para que el paso realice la asignación a un objetivo.

¿Ocultar los campos de origen asignados?: Seleccione esta opción para eliminar un campo del panel Campos de origen cuando haya coincidido y se haya movido al panel Asignaciones.

Selección de fuente automática: Seleccione para seleccionar automáticamente un campo del panel Campos de origen cuando se selecciona un nombre de columna.

¿Ocultar los campos de destino asignados?: Seleccione esta opción para eliminar un campo del panel Campos de destino cuando haya coincidido y se haya movido al panel Asignaciones.

Adivinar (botón): Haga clic en Adivinar para realizar una comparación automática de todos los campos y la población del panel Asignaciones.

Fórmula

Permite calcular expresiones de fórmula dentro de un flujo de datos. Se puede utilizar para crear cálculos simples como [A] + [B] o lógica de negocios más compleja con mucha lógica anidada si / entonces.

Cuando presiona en la entrada de la celda Fórmula, se abre la ventana del editor de fórmulas y le brinda ayuda sobre las funciones disponibles.

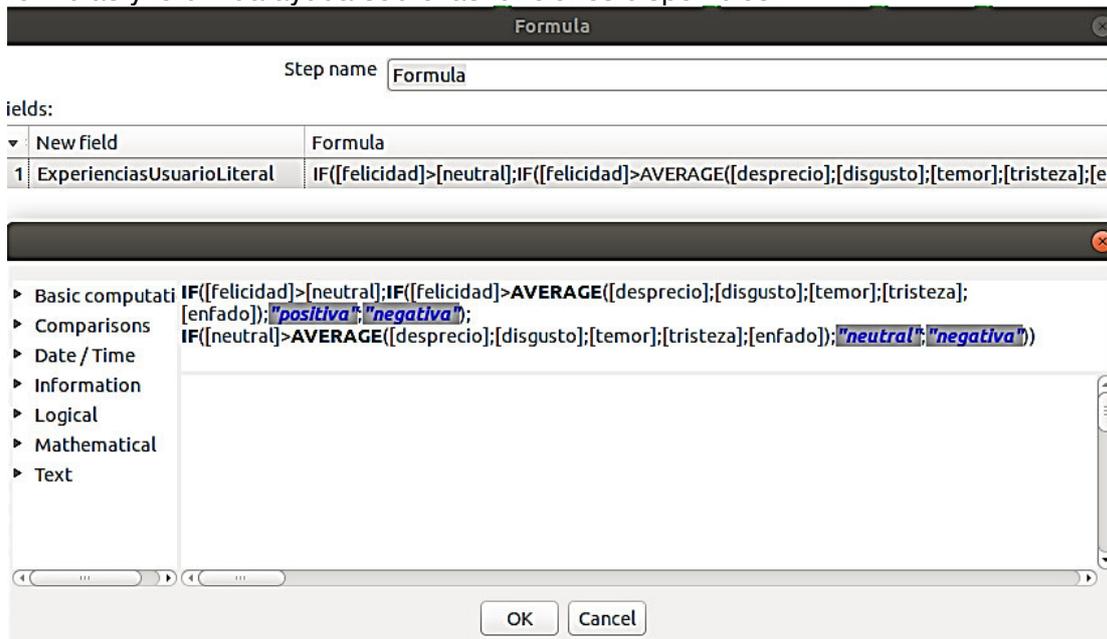


Figura 0.18 Spoon: Fórmula

La ventana del editor de fórmulas realiza una verificación directa de la sintaxis y los nombres de campo proporcionados.

Actualización

En la Figura 0.19 se muestra el paso Insertar / Actualizar, que primero busca una fila en una tabla usando una o más claves de búsqueda.

Update

Step name

Connection Edit... New... Wizard...

Target schema Browse...

Target table Browse...

Commit size

Use batch updates?

Skip lookup

Ignore lookup failure? Flag field (key found)

The key(s) to look up the value(s):

▼	Table field	Comparator	Stream field1	Stream field2	Get fields
1	faceld	=	faceldV		

Update fields:

▼	Table field	Stream field	Get update fields
1	experienciaUsuario1	ExperienciasUsuarioLiteral	

Help OK Cancel SQL

Figura 0.19 Spoon: Actualización

Opciones

Nombre del paso: este nombre tiene que ser único en una sola transformación.

Conexión: La conexión de la base de datos en la que se escriben los datos.

Esquema de destino: El nombre del esquema de la tabla en la que se escriben los datos. Esto es importante para las fuentes de datos que permiten nombres de tablas con puntos.

Tabla de destino: Nombre de la tabla en la que desea realizar la actualización.

Tamaño de Comprometidos: El número de filas para actualizar antes de ejecutar una confirmación.

No realice ninguna actualización: Si está habilitado, los valores en la base de datos nunca se actualizan.

Tabla de búsqueda de claves: Permite especificar una lista de valores de campo y comparadores. Puede utilizar los siguientes comparadores: =, = ~ NULL, <>, <, <=, >, >=, LIKE, BETWEEN, IS NULL, IS NOT NULL.

Actualizar campos: Permite especificar todos los campos de la tabla que desea actualizar, incluidas las claves.

Botón SQL: Para generar el SQL para crear la tabla y los índices para un funcionamiento correcto.

Ordenar Filas

En la Figura 0.20 se aprecia el componente Ordenar Filas que ordena las filas según los campos que especifique y si deben ordenarse en orden ascendente o descendente.

Sort rows

Step name

Sort directory

TMP-file prefix

Sort size (rows in memory)

Free memory threshold (in %)

Compress TMP Files?

Only pass unique rows? (verifies keys only)

Fields:

Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength
1 faceldv	Y	N	N	0

Figura 0.20 Spoon: Ordenar filas

Opciones

Nombre del paso: este nombre debe ser único en una única transformación.

Ordenar directorio: El directorio en el que se almacenan los archivos temporales en caso de que sea necesario; el predeterminado es el directorio temporal estándar para el sistema.

Prefijo de archivo TMP: Elija un prefijo fácilmente reconocible para que pueda identificar los archivos cuando aparezcan en el directorio temporal.

Ordenar tamaño: Cuantas más filas almacene en la memoria, más rápido será el proceso de clasificación porque se deben usar menos archivos temporales y se genera menos E / S.

Umbral de memoria libre (en%): Si el algoritmo de clasificación encuentra que tiene menos memoria libre disponible que el número indicado, comenzará a paginar datos en el disco.

Comprimir archivos TMP: Comprime archivos temporales cuando son necesarios para completar la clasificación.

¿Solo pasa filas únicas?: Habilite si desea pasar filas únicas solo a los flujos de salida.

Tabla de campos: Especifique los campos y la dirección (ascendente / descendente) para ordenar. Puede especificar si desea realizar una ordenación sensible a mayúsculas y minúsculas (opcional)

Obtener campos: Haga clic para recuperar una lista de todos los campos que ingresan en el paso.

Filas únicas

En la Figura 0.21 se aprecia el componente Filas únicas con el cual se elimina las filas duplicadas del flujo de entrada y filtra solo las filas únicas para el siguiente paso.

El flujo de entrada debe ordenarse en un paso antes del paso de Filas únicas; de lo contrario, solo se analizarán y filtrarán correctamente las filas dobles consecutivas.

Step name

Settings

Add counter to output? Counter field

Redirect duplicate row Error description

Fields to compare on (no entries means: compare complete row)

Fieldname	Ignore case
1 faceldV	Y
2 nombreEstudiante	Y

Help OK Cancel Get

Figura 0.21 Filas Únicas

Opciones

Nombre del paso: especifique el nombre exclusivo de la transformación en el lienzo. Puede personalizar el nombre o dejarlo como predeterminado.

El paso de filas únicas requiere definiciones para las siguientes opciones y parámetros:
 ¿Añadir contador a la salida?: Seleccione esta opción para agregar un campo de contador al flujo de salida.

Campo de contador: Especifique un nombre para el campo del contador.

Redirigir fila duplicada: Seleccione esta opción para procesar filas duplicadas como un error y redirigirlas al flujo de errores del paso. Si no selecciona esta opción, se eliminan las filas duplicadas.

Error de descripción: Especifique la descripción de manejo de errores que se muestra cuando el paso detecta filas duplicadas. Esta descripción solo está disponible cuando se selecciona Redirigir fila duplicada .

Campos para comparar tabla: Especifique los nombres de los campos para los que desea buscar valores únicos. Seleccione Obtener para insertar todos los campos del flujo de entrada.

Ignorar caso: Puede optar por ignorar las mayúsculas, minúsculas y títulos, estableciendo esta opción en 'Y'.

Si no especifica ninguna entrada de campo, el paso compara toda la fila.

Unión de base de datos

En la Figura 0.22 se muestra el paso unión de base de datos que permite ejecutar una consulta en una base de datos utilizando datos de pasos anteriores. Los parámetros para esta consulta se pueden especificar de la siguiente manera:

- como signos de interrogación en la consulta SQL.
- como campos en la cuadrícula de datos.

Los dos deben estar en el mismo orden.

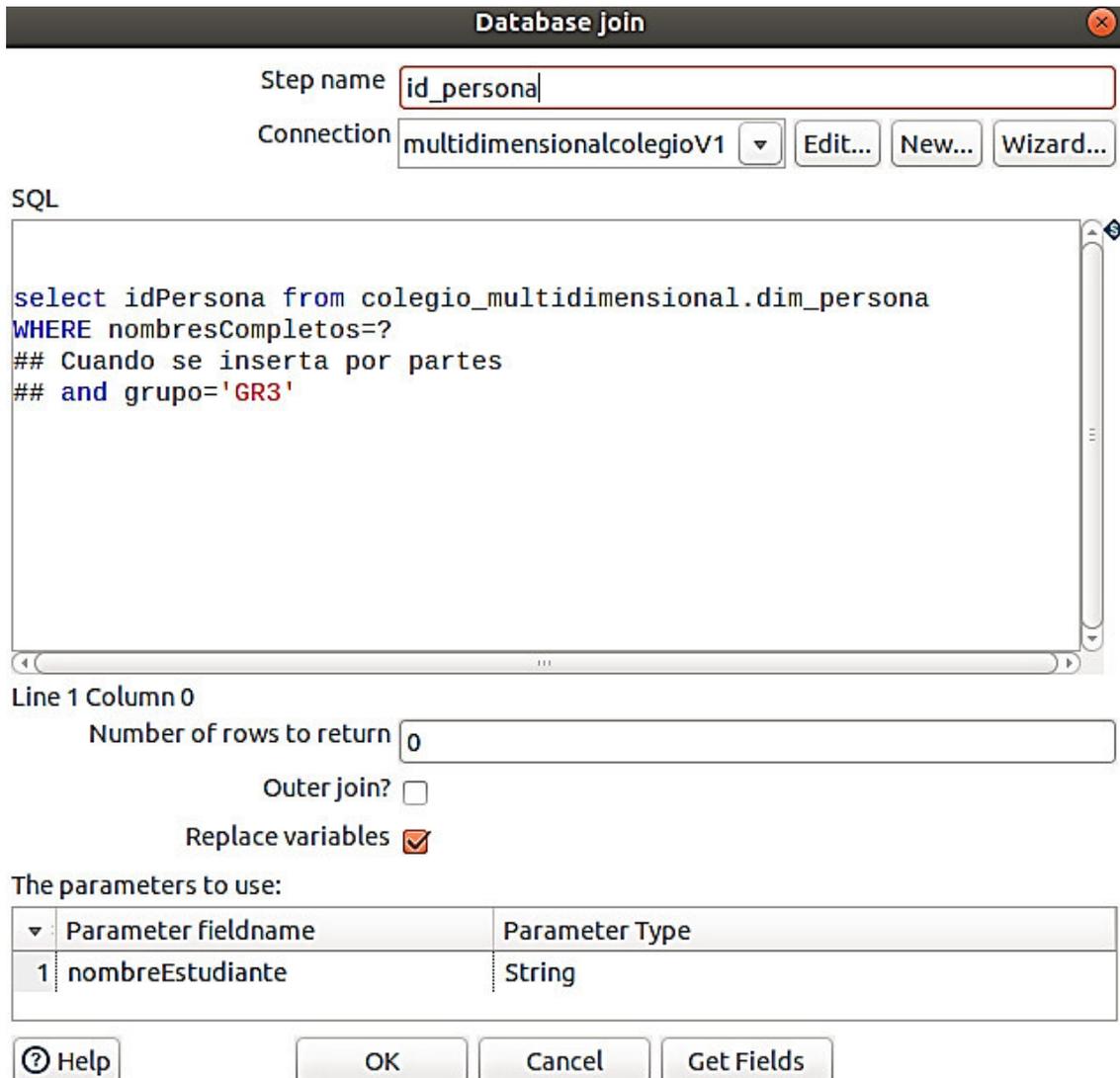


Figura 0.22 Spoon: Unión de Base de Datos

Opciones

Nombre del paso: Este nombre tiene que ser único en una sola transformación.

Conexión: La conexión de la base de datos a utilizar

SQL: Consulta SQL para lanzar hacia la base de datos; utilizar signos de interrogación como marcadores de posición de parámetros.

Número de filas para devolver: Zero (0) devuelve todas las filas; cualquier otro número limita el número de filas

¿Unión externa?: Habilite para devolver siempre un resultado, incluso si la consulta no devolvió un resultado.

Reemplazar variables: Indica que los signos de interrogación que se encuentran en el campo SQL se remplazaran por los valores de la tabla de parámetros.

Tabla de parámetros: Especifique los campos que contienen parámetros y el tipo de parámetro.

Concatenar campos

En la Figura 0.23 se muestra el componente Concatenar campos que se utiliza para concatenar varios campos en un campo de destino. Los campos se pueden separar con un separador y la lógica de este paso es compatible con el paso Salida de archivo de texto.

La compatibilidad con el paso Salida de archivo de texto también permite crear los mismos campos y diseño (incluidos el encabezado, el pie de página y las líneas finales) que podrían almacenarse en otro lugar. Al usar un Grupo por paso con la opción "Concatenar cadenas separadas por" y seleccionar un delimitador de línea (por ejemplo, CR / LF), es posible crear un campo que contenga lo mismo que un campo de salida.

Name	Type	Format	Length	Precision	Currency	Decimal	Group
1 idpersona	String						
2 faceld	String						
3 photoID	String						

Figura 0.23 Spoon: Concatenar Campos

Opciones

Nombre del paso: Este nombre debe ser único en una única transformación.

Nombre del campo de destino: El nombre del campo de destino (tipo de cadena)

Longitud del campo objetivo: La longitud del tipo de cadena (configurando los metadatos del tipo de cadena, esto también es utilizado por la opción Fast Data Dump para optimizar el rendimiento)

Separador: Especifique el carácter que separa los campos en una sola línea de texto. Normalmente esto es; o una pestaña.

Cierre: Un par de cadenas pueden incluir algunos campos. Esto permite caracteres separadores en los campos. La cadena de cierre es opcional.

Pestaña Campos

Es idéntica a la opción de pestaña de campos del paso Salida de archivo de texto y tiene la misma funcionalidad.

Nombre: El nombre del campo.

Tipo: El tipo de campo puede ser Cadena, Fecha o Número.

Formato: La máscara de formato con la que se va a convertir.

Largo: La opción de longitud depende del tipo de campo siguiente:

Precisión: La opción de precisión depende del tipo de campo de la siguiente manera:

Divisa: Símbolo utilizado para representar monedas como \$ 10,000.00 o E5.000,00

Decimal: Un punto decimal puede ser un "." (10,000.00) o "," (5.000,00)

Grupo: Una agrupación puede ser "," (10,000.00) o "." (5.000,00)

Tipo de recorte: El método de recorte que se aplicará a la cuerda. El recorte solo funciona cuando no se proporciona una longitud de campo. (consulte la solicitud de función PDI-2486)

Nulo: Si el valor del campo es nulo, inserte esta cadena en el archivo de texto

Obtener Campos: Haga clic para recuperar la lista de campos de los flujos de campos de entrada

Ancho mínimo: Modifique las opciones en la pestaña de campos de tal manera que el ancho resultante de las líneas en el archivo de texto sea mínimo. Entonces, en lugar de guardar 0000001, escribimos 1, etc. Los campos de cadena ya no se rellenarán con la longitud especificada.

Pestaña avanzada

Eliminar campos seleccionados: Marque esto para eliminar todos los campos seleccionados del flujo de salida.

¿Forzar el recinto alrededor de los campos?: Esta opción obliga a que todos los nombres de campo se incluyan con el carácter especificado en la propiedad Enclosure anterior.

¿Deshabilitar la reparación del gabinete?: Esto es por razones de compatibilidad con versiones anteriores (desde la versión 4.1) relacionadas con envolventes y separadores.

Encabezamiento: Habilite esta opción si desea una fila de encabezado. (Primera línea de la secuencia). Nota: Todos los demás campos de flujo de salida se establecen en Nulo cuando se produce esta línea.

Pie de página: Habilite esta opción si desea una fila de pie de página. (Última línea de la secuencia).

Codificación: Especifique la codificación de cadena que se utilizará.

Campos de almohadilla derecha: Agregue espacios al final de los campos (o elimine los caracteres al final) hasta que tengan la longitud especificada.

Volcado de datos rápido (sin formateo): Mejora el rendimiento al concatenar grandes cantidades de datos al no incluir información de formato.

Dividir cada ... filas: Si este número N es mayor que cero, divida la secuencia resultante en varias partes de N filas.

Agregar línea final de archivo: Le permite especificar una fila final alternativa al flujo de salida.

Entrada de archivo CSV

En la Figura 0.24 se presenta el paso Entrada de archivo CSV, que lee datos de archivos de texto delimitados en una transformación PDI. Si bien este paso se denomina "Entrada de archivo CSV", también puede utilizar la Entrada de archivo CSV con muchos otros tipos de separadores, como barras verticales, tabulaciones y punto y coma.

The screenshot shows the 'CSV file input' dialog box with the following configuration:

- Step name: Emociones
- Filename: /home/normy/Documents/DocumentosBDTesis/CodigoEmociones.csv
- Delimiter: ,
- Enclosure: "
- NIO buffer size: 50000
- Lazy conversion?:
- Header row present?:
- Add filename to result:
- The row number field name (optional):
- Running in parallel?:
- New line possible in fields?:
- File encoding:

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1 iddim_emocion	Integer	#	15	0	\$.	,	none
2 nombreEmocion	String		9		\$.	,	none

Figura 0.24 Spoon: Archivo CSV de Entrada

Opciones

Nombre del paso: Especifique el nombre exclusivo del paso de entrada de archivo CSV en el lienzo. Puede personalizar el nombre o dejarlo como predeterminado.

Nombre de archivo (o el campo de nombre de archivo): El nombre (nombre de archivo) del archivo de origen CSV.

¿Incluir el nombre del archivo en la salida?: Si la entrada del archivo CSV recibe datos de otro paso, indique si el nombre del archivo de origen de entrada debe incluirse en la salida del paso de entrada del archivo CSV.

Delimitador: Especifique el carácter delimitador de archivo utilizado en el archivo de origen.

Cierre: Especifique el carácter del cierre utilizado en el archivo de origen.

Tamaño de búfer NIO: Especifique el tamaño del búfer de lectura, el número de bytes que se leen a la vez desde el origen.

Conversión perezosa: Indique si se puede utilizar el algoritmo de conversión diferida para mejorar el rendimiento.

¿Fila de encabezado presente?: Indique si el archivo de origen contiene una fila de encabezado que contiene nombres de columna.

Agregar nombre de archivo al resultado: Agrega los nombres de archivo de origen CSV al resultado de esta transformación.

El nombre del campo del número de fila (opcional): Especifique el nombre del campo que contendrá el número de fila en el resultado de este paso.

¿Corriendo en paralelo?: Indique si tendrá varias instancias de este paso en ejecución (copias de los pasos) y si desea que cada instancia lea una parte separada de los archivos CSV.

¿Es posible una nueva línea en los campos?: Indique si los campos de datos pueden contener caracteres de nueva línea.

Codificación de archivos: Especifique la codificación del archivo de origen.

Nombre: Nombre del campo

Tipo: Tipo de campo (cadena, fecha o número)

Formato: Una máscara opcional para convertir el formato del campo original.

Largo: La longitud del campo depende de los siguientes tipos de campo:

Precisión: Número de dígitos de coma flotante para campos de tipo numérico

Divisa: Símbolo utilizado para representar monedas (\$ 5,000.00 o € 5,000,00 por ejemplo)

Decimal: Un punto decimal puede ser un "." o "," (5.000,00 o 5.000,00 por ejemplo)

Grupo: Una agrupación puede ser "," o "" (5,000.00 o 5.000,00 por ejemplo)

Tipo de recorte: El método de recorte para aplicar a una cadena.

Entrada JSON

En la Figura 0.25 se muestra el componente Entrada JSON. Este paso lee datos de estructuras JSON, archivos o campos entrantes utilizando una expresión JSONPath para extraer datos y filas de salida. Las expresiones JSONPath pueden usar notación de puntos o notación de corchetes.

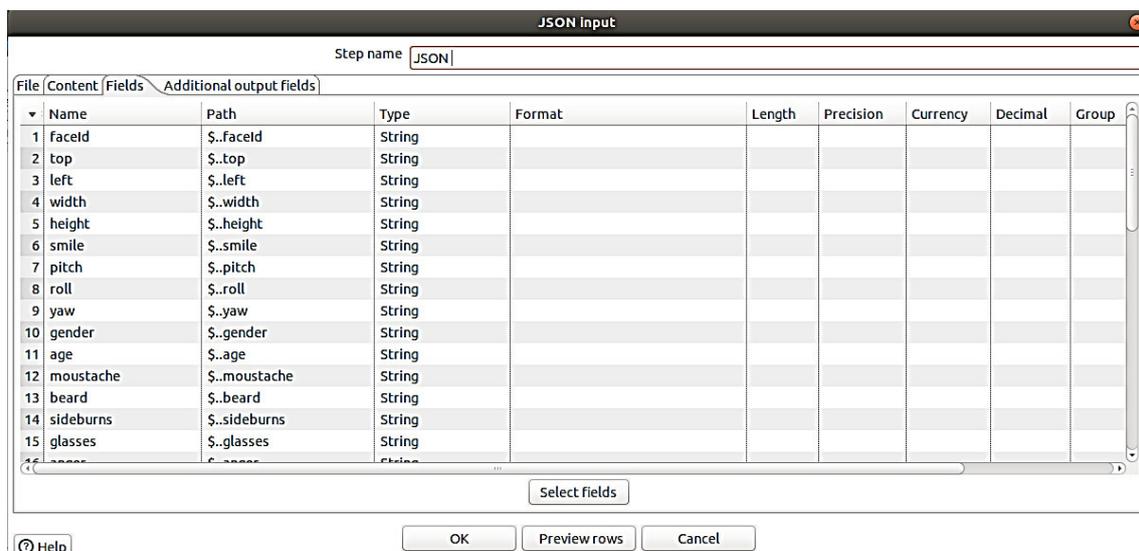


Figura 0.25 Spoon: Archivo Json de Entrada

Opciones

Nombre del paso: especifica el nombre único del paso de transformación de entrada JSON en el lienzo. Puede personalizar el nombre o dejarlo como predeterminado.

Pestaña Archivo

La fuente es de un paso anterior: Seleccione para recuperar la fuente de un campo definido previamente.

Seleccionar campo: Especifique el nombre del campo para usar como fuente de un paso anterior.

Usar campo como nombres de archivo: Seleccione para indicar que la fuente es un nombre de archivo.

Leer fuente como URL: Seleccione para indicar si se debe acceder a la fuente como una URL.

No pase el campo corriente abajo: Seleccione para eliminar el campo de origen del flujo de salida. Esta acción mejora el rendimiento y la utilización de la memoria con grandes campos JSON.

Archivo o directorio: Especifique la ubicación de la fuente si la fuente no está definida en un campo.

Expresión regular: Especifique una expresión regular para hacer coincidir los nombres de archivo dentro de un directorio específico.

Excluir expresión regular: Especifique una expresión regular para excluir nombres de archivo dentro de un directorio específico.

Tabla de archivos seleccionados

La tabla contiene las siguientes columnas:

Directorio de archivos: La ubicación de origen indicada al hacer clic en Agregar después de especificarlo en Archivo o directorio

Comodín (RegExp): Comodines como se especifica en Expresión regular

Excluir comodín: Comodines excluidos como se especifica en Excluir expresión regular

Requerido: Ubicación de la fuente requerida para la entrada

Incluir subcarpetas: Si las subcarpetas están incluidas dentro de la ubicación de origen.

Ficha Contenido

Ignorar archivo vacío: Seleccione para omitir archivos vacíos. Cuando se borran, los archivos vacíos harán que el proceso falle y se detenga.

No genere un error si no hay archivos: Seleccione para continuar cuando no haya archivos disponibles para procesar.

Ignorar la ruta que falta: Seleccione para continuar procesando archivos cuando se produzca un error que indique que (1) ningún campo coincide con la ruta JSON o (2) que todos los valores son nulos. Cuando se borra, no se procesan más filas cuando ocurre un error.

Hoja de ruta predeterminada a nula: Seleccione esta opción para devolver un valor nulo para las rutas faltantes.

Límite: Especifique un límite en el número de registros generados a partir del paso. Los resultados no están limitados cuando se establecen en cero.

Incluir nombre de archivo en la salida: Seleccione para agregar un campo de cadena con el nombre de archivo en el resultado.

Num. de filas en la salida: Seleccione para agregar un campo entero con el número de fila en el resultado.

Agregar nombres de archivo al resultado: Seleccione para agregar archivos procesados a la lista de archivos de resultados.

Pestaña Campos

Nombre: Nombre del campo que se asigna al campo correspondiente en el flujo de entrada JSON.

Camino: Ruta completa del nombre del campo en el flujo de entrada JSON.

Tipo: Tipo de datos del campo de entrada

Formato: Una máscara opcional para convertir el formato del campo original.

Largo: Longitud del campo

Precisión: Número de dígitos de coma flotante para campos de tipo numérico

Divisa: Símbolo de moneda (\$ o € por ejemplo)

Decimal: Un punto decimal puede ser un "." (5,000.00 por ejemplo) o "," (5.000,00 por ejemplo)

Grupo: Una agrupación puede ser "," (10.000,00 por ejemplo) o "." (5.000,00 por ejemplo)

Tipo de recorte: El método de recorte para aplicar a una cadena.

Repetir: El valor correspondiente de la última fila se repite si una fila está vacía

Pestaña Campos de salida adicionales

Campo de nombre de archivo corto: Especifique el campo que contiene el nombre del archivo sin información de la ruta, pero con una extensión.

Campo de extensión: Especifique el campo que contiene la extensión del nombre de archivo.

Campo de ruta: Especifique el campo que contiene la ruta en formato de sistema operativo.

Campo de tamaño: Especifique el campo que contiene el tamaño de los datos.

Es un campo oculto: Especifique el campo que indica si el archivo está oculto o no (booleano).

Campo de última modificación: Especifique el campo que indica la fecha de la última vez que se modificó el archivo.

Campo de Uri: Especifique el campo que contiene el URI.

Campo de raíz uri: Especifique el campo que contiene solo la parte raíz del URI.

Agrega Constantes

El paso Agregar valores constantes que se muestra en la Figura 0.26, es una forma simple y de alto rendimiento de agregar valores constantes al flujo.

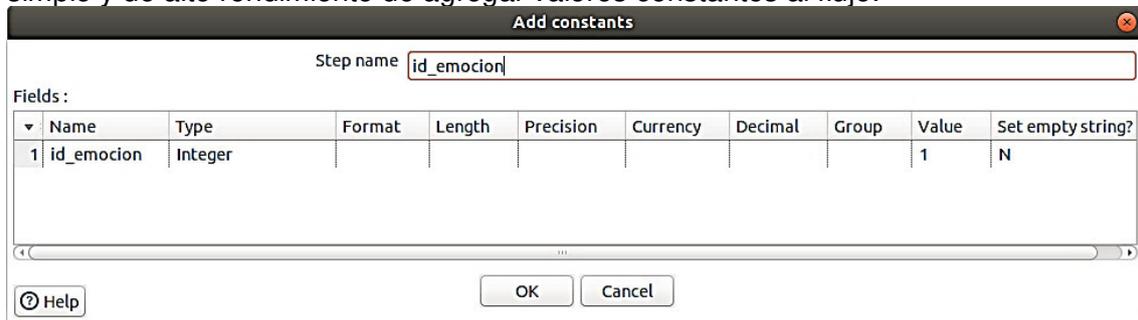


Figura 0.26 Spoon: Agregar constantes

Nombre del paso: Muestra el nombre del paso, este nombre debe ser único en una única transformación.

Campos: Para especificar el nombre, el tipo y el valor en forma de cadena. Luego, especificar los formatos para convertir el valor en el tipo de datos elegido.

Reemplazar en cadena

En la figura 0.27 se muestra el paso Reemplazar en cadena que realiza una búsqueda y reemplazo simples. Este paso admite expresiones regulares y referencias de grupo.

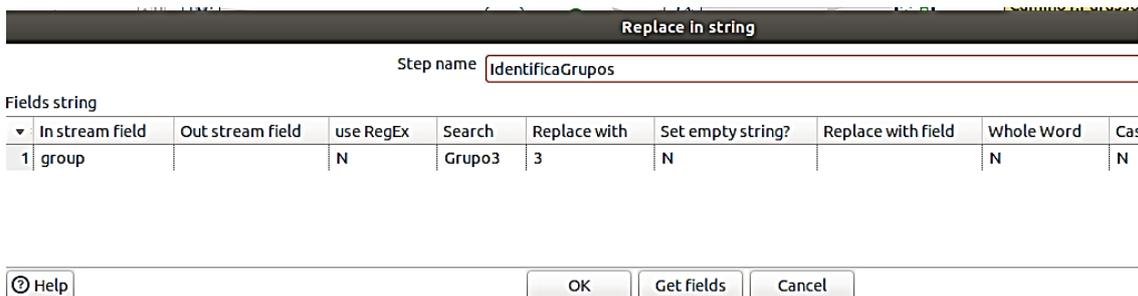


Figura 0.27 Spoon: Reemplazo en cadena

Opciones

Nombre del paso: especifique el nombre único del paso en el lienzo.

En campo de corriente: Especifique el campo para reemplazar del flujo de entrada PDI.

Fuera de campo de corriente: Especifique el nuevo nombre de campo para enviar a la secuencia PDI.

usar RegEx: Seleccione Y (Sí) o N (No) para indicar si se debe usar una expresión regular para buscar valores.

Buscar: Especifique si desea buscar valores coincidentes de esta cadena.

Reemplazar con: Especifique la cadena para reemplazar el valor coincidente.

¿Establecer cadena vacía?: Seleccione Y (Sí) o N (No) para indicar si se deben reemplazar los valores nulos con cadenas vacías.

Reemplazar con campo: Especifique un valor de campo que se utilizará para reemplazar el valor coincidente.

Toda palabra: Seleccione Y (Sí) o N (No) para indicar si se debe reemplazar la palabra completa del valor coincidente.

Distingue mayúsculas y minúsculas: Seleccione Y (Sí) o N (No) para indicar si la búsqueda distingue entre mayúsculas y minúsculas.

Es Unicode: Seleccione Y (Sí) o N (No) para indicar si el valor de búsqueda es una secuencia de caracteres Unicode.

Anexo XI Tutorial RapidMiner

RM es una herramienta para crear modelos y para la realización de análisis predictivos de grandes volúmenes de datos, facilita el autoservicio de análisis predictivo permitiendo una avanzada analítica empleando solamente drag and drop y opcionalmente la generación de código. Acelera la creación de prototipos y la validación de modelos, agiliza la transformación, el desarrollo y la validación de datos. Soporta sofisticadas operaciones de análisis como los sistemas Scoring, aplicaciones de Detección de Desviación y Detección de Fraude, el entorno visual fácil de usar para análisis predictivo. [41]

Requisitos previos

En la Tabla Anexo 1, se incluyen los requisitos mínimos para la instalación de RM. [42]

Tabla A1 Requisitos para instalar RM

CARACTERÍSTICAS	MÍNIMO	RECOMENDADO
CORE	Dual	Quad
Procesador	2GHz	3GHz or faster processor
RAM	4GB	16GB
Espacio Libre en disco	> 1GB	> 100GB
Resolución	1280 X 1024	1280 X 1024
SISTEMA OPERATIVO		PLATAFORMA JAVA
64 bit		64 bit
Windows 7,8,8.1,10		Java 8
Linux		HiDPI support
MacOS X.10.8 o más nuevo		

Instalación

1. Descargar la aplicación en el sitio web de RapidMiner. El driver para RM con mysql se recomienda tomar la versión de jar más alta de mysql.
2. Haga doble clic en el archivo zip descargado (por ejemplo, rapidminer-studio-<version>.zip). La carpeta extraída contiene archivos de script para diferentes sistemas operativos. Asegúrese de que la ruta de la carpeta no contenga +o %caracteres.
3. Abra una ventana de terminal y ejecute ./RapidMiner-Studio.sh desde el directorio que contiene la descarga como se aprecia en la Figura 0.28.

```

normy@ubuntu: ~/Downloads/rapidminer-studio-9.8.0/rapidminer-studio
File Edit View Search Terminal Help
normy@ubuntu:~$ cd /home/normy/Downloads/rapidminer-studio-9.8.0/rapidminer-studio
normy@ubuntu:~/Downloads/rapidminer-studio-9.8.0/rapidminer-studio$ ./RapidMiner-Studio.sh

```

Figura 0.28 RM: Arranque

4. Lea los términos del acuerdo de licencia y haga clic en Acepto para continuar. Se abre RapidMiner Studio.

Problemas RM

- RM no carga archivos json, pero toma datos de un servicio json configurando el driver para esos casos.
- La versión de estudiante permite tener acceso por el período de un año a opciones como Turbo Prep para calidad de datos, esta entre otras no se activan si la licencia de estudiante a caducado.
- Consumo de procesador de RM que no permite llevar a cabo otro proceso mientras se realiza la predicción.
- El tiempo que tomo en generar los modelos alrededor de 2 horas.
- Al descartar valores nulos invisible y fecha se convierte en polinomial, a pesar de hacer el cambio a polinomial para el uso del simulador me aparecen valores nulos.

Tratamiento de Datos

En la Figura 0.29 se muestra la carga de datos, en Importar datos se selecciona el conjunto de datos.

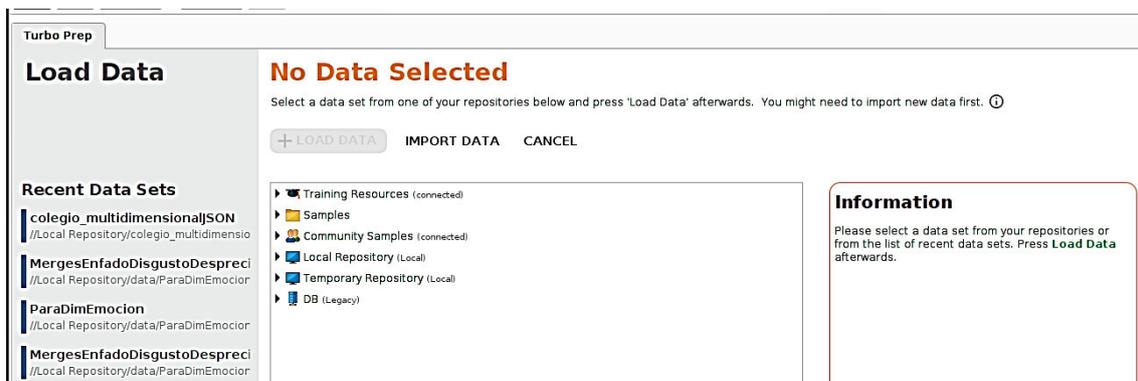


Figura 0.29 RM: Carga de Datos

Una vez se carga el conjunto de datos es posible ver en detalle los valores o ir a Turbo Prep como se aprecia en la Figura 0.30.

Row No.	faceId	parteSu...	partelzq...	anchura	altura	sonrisa	inclinaci...	rot
1	fbe9b587-1...	255.0	558.0	139.0	139.0	0.111	-14.5	5.0
2	a552d244-...	221.0	291.0	115.0	115.0	0.0	-33.7	-8.3
3	f0a2ee34-3...	197.0	72.0	102.0	102.0	0.0	-11.7	3.7
4	7aa0bfd3-2...	195.0	592.0	115.0	115.0	0.0	-2.1	0.4
5	6a265e18-...	225.0	254.0	68.0	68.0	0.002	-13.5	-8.4
6	a3660815-...	212.0	20.0	104.0	104.0	0.0	-16.7	0.6
7	fb20adb6-e...	171.0	574.0	104.0	104.0	0.0	13.7	0.1
8	8b056658-...	192.0	308.0	93.0	93.0	0.619	-7.8	-1.3
9	8ed4ef8a-7...	197.0	97.0	104.0	104.0	0.0	-2.5	5.5
10	a0edabd6-f...	160.0	548.0	103.0	103.0	0.0	21.3	-4.6

Figura 0.30 RM: Resultado de conjunto de datos cargado

En la Figura 0.31 se puede revisar estadísticas, valores perdidos, los valores más frecuentes y los histogramas de estos datos.

Name	Type	Missing	Filter (45 / 45 attributes):
faceId	Polynomial	0	Least ffd8f178 [...] c1c8c (1) Most n.a. (52)
parteSuperior	Polynomial	0	Least 98 (1) Most n.a. (52)
partelzquierda	Polynomial	0	Least 99 (1) Most n.a. (52)
anchura	Polynomial	0	Least 97 (1) Most n.a. (52)
altura	Polynomial	0	Least 97 (1) Most n.a. (52)

Figura 0.31 RM: Tipos de datos del conjunto de datos

En la Figura 0.31 se puede verificar la calidad del set de datos y optimizarla al retirar o reemplazar valores perdidos.

Name	Type	Missing	Filter (5 / 5 attributes):
Nominal	0	Least valorNivelDesenfocue... Most accesorios (1)	Search for Attribut
Real	1	Min 0.002 Max 0.968	Average 0.111
Real	0	Min 0.007 Max 1	Average 0.544
Real	0	Min 0 Max 0.034	Average 0.011
Real	0	Min 0 Max 0.654	Average 0.049

Showing attributes 1 - 5 Examples: 47 Special Attributes: 0 Regular Attributes: 5

Figura 0.32 RM: Medidas de calidad del conjunto de datos

En la Figura 0.33 se puede apreciar la visualización en diferentes tipos de gráficos.

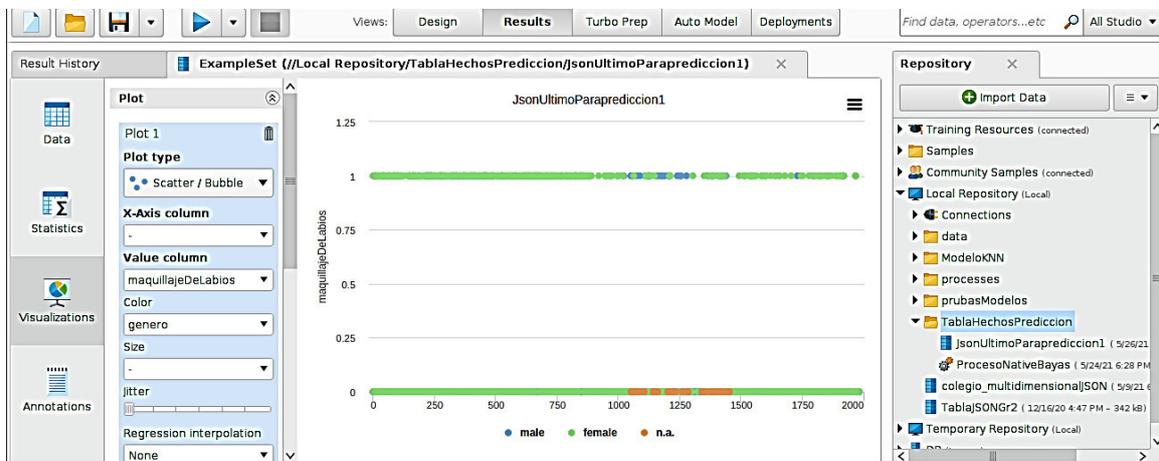


Figura 0.33 Visualizaciones del conjunto de datos

Turbo Prep

Se utiliza cuando los datos se encuentran en un estado disperso o inconsistente, y aún no están listos para la creación de modelos.

Para comenzar se elige Turbo Prep vista presionando el botón en la parte superior de RapidMiner Studio, Figura 0.34.

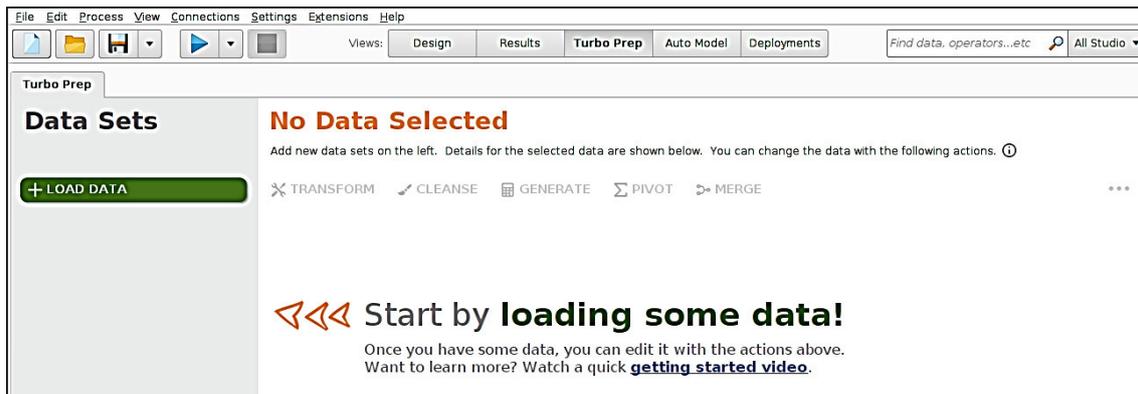


Figura 0.34 Turbo Prep: Selección del conjunto de datos

En la Figura 0.35 se procede a la carga de datos, en Importar datos se selecciona el conjunto de datos.

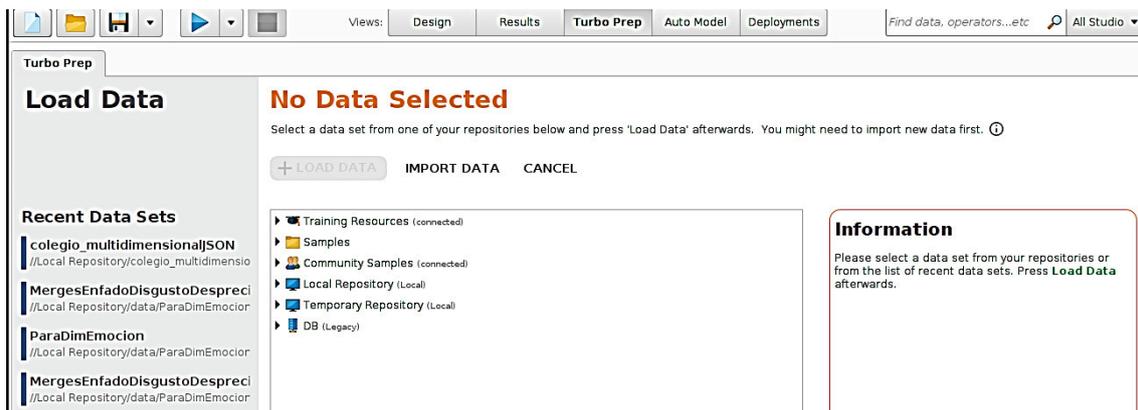


Figura 0.35 Turbo Prep: Carga de Datos

En la Figura 0.36 se busca la ubicación de los datos que puede ser en el computador local o en una base de datos.



Figura 0.36 Turbo Prep: Importación de datos

En la Figura 0.37 se muestra la selección para la ubicación del archivo.

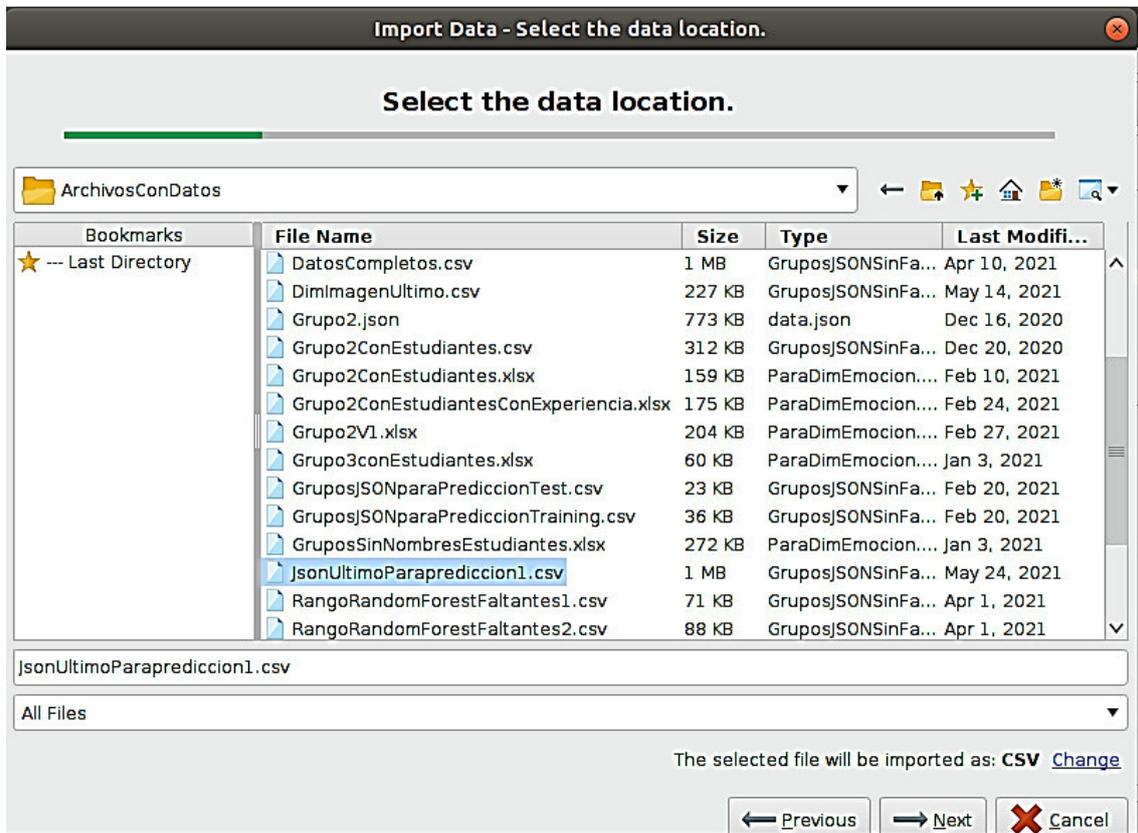


Figura 0.37 Turbo Prep: Localización de datos

En la Figura 0.38 se muestra la especificación del formato de datos.

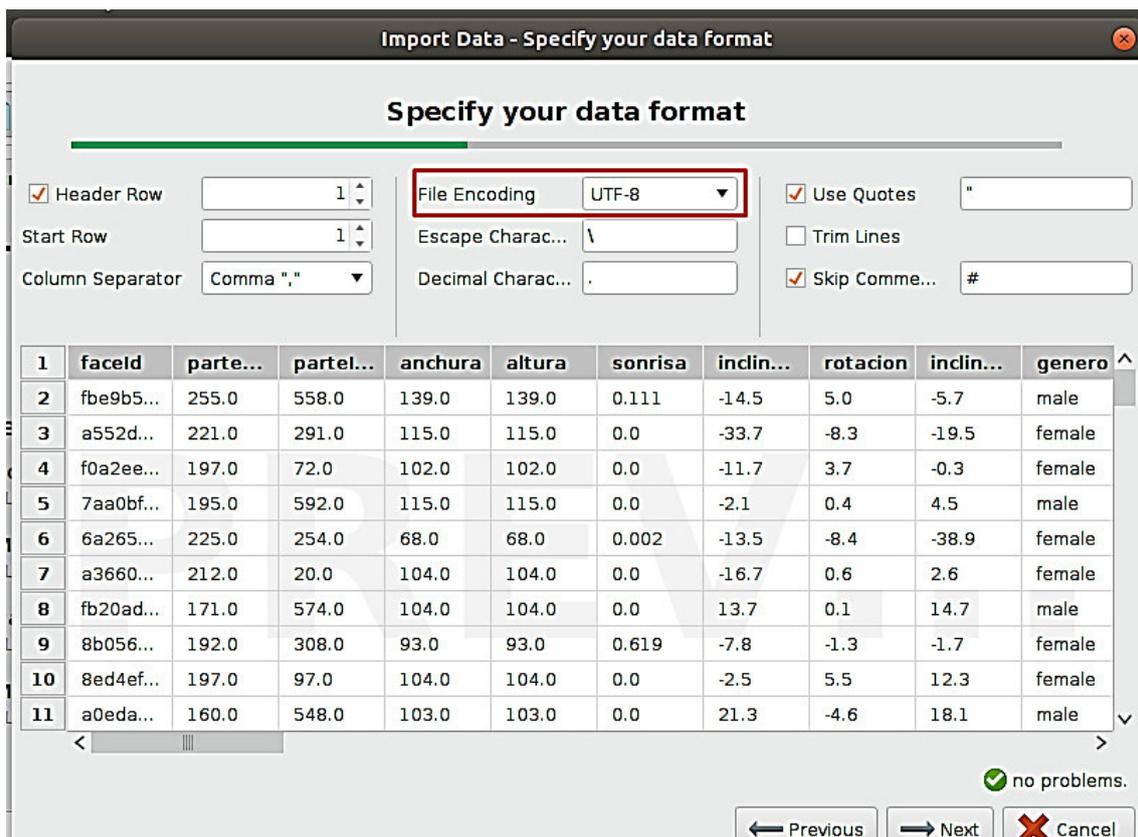


Figura 0.38 Turbo Prep: Especificación del formato de datos

En la Figura 0.39 se muestra como es posible cambiar el tipo, renombrar la columna y excluir columnas.

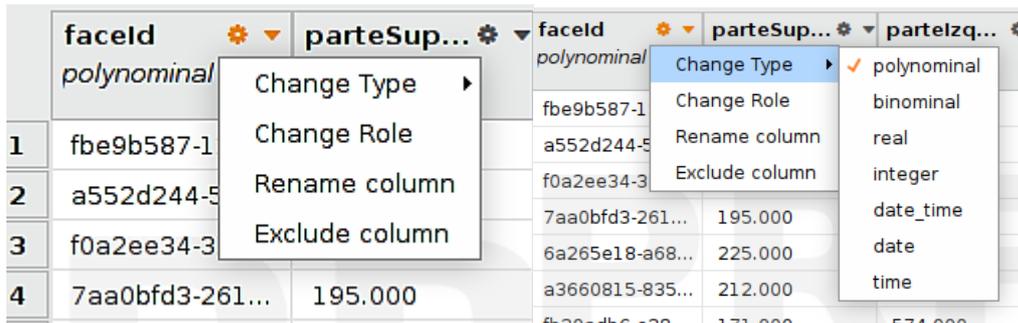


Figura 0.39 Turbo Prep: Opciones para columnas de datos

En la Figura 0.40 se muestra el almacenamiento de los datos en un repositorio local.

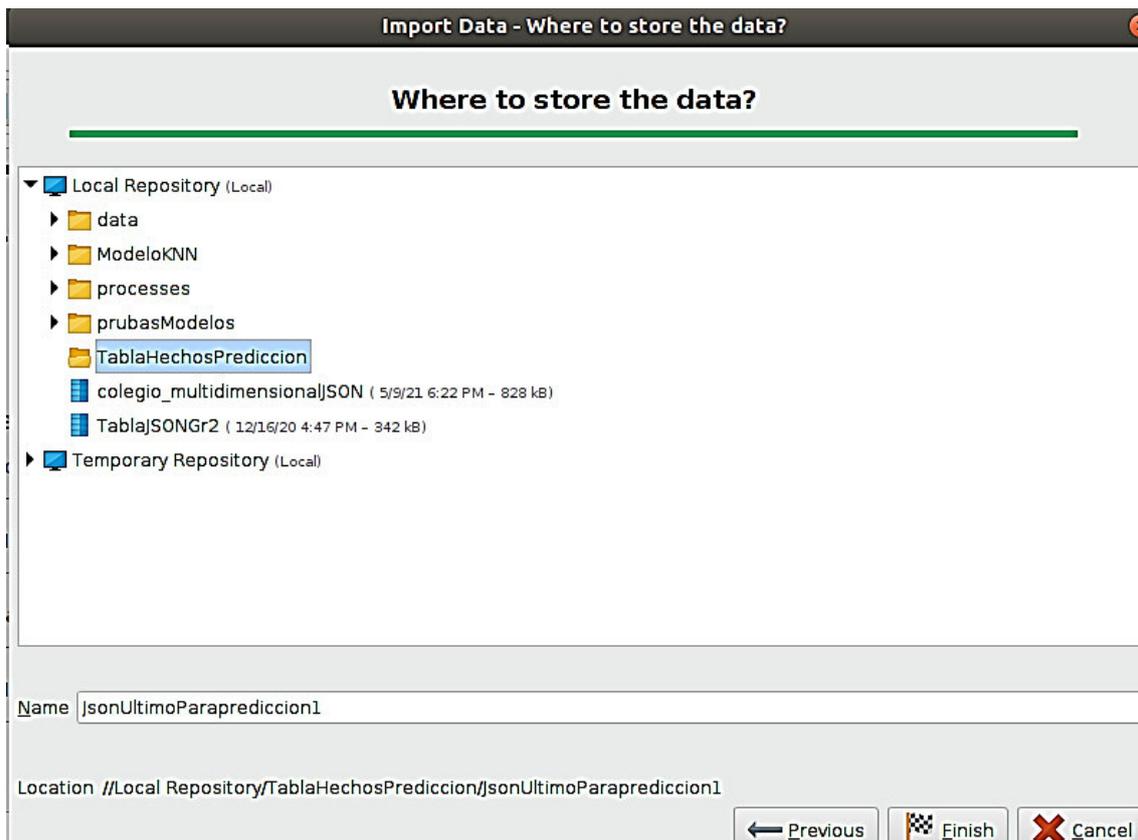


Figura 0.40 Turbo Prep: Almacenamiento de los datos

Luego que se ha verificado que no hay inconvenientes para subir los datos se carga el conjunto de datos, como se muestra en la Figura 0.41.

GruposJSONparaPrediccionTraining

Add new data sets on the left. Details for the selected data are shown below. You can change the data with the following actions: ⓘ

✗ TRANSFORM ✎ CLEANSE 📄 GENERATE ∑ PIVOT ➔ MERGE ...

cod_regist... Number	enfado Number	disgusto Number	desprecio Number	temor Number	felicidad Number	neutral Number	tristeza Number	sorpresa Number
1	0	0	0	0	0	0.680	0	0.320
2	0	0	0	0	1	0	0	0
3	0	0	0	0	0	1	0.010	0
4	0	0	0	0	0	1	0	0
5	0	0	0	0	1	0	0	0
6	0	0	0	0	0	1	0	0
7	0	0	0	0	0.010	0.980	0	0
8	0	0	0	0	0	0.560	0	0.440
9	0	0	0	0	0	1	0	0
10	0	0	0	0	0	0.990	0.010	0

Figura 0.41 Turbo Prep: Conjunto de datos cargados

Las funciones de apoyo de Turbo Prep son:

Transformar: Estas funciones ayudan a crear subgrupos útiles de sus datos, Figura 0.42, permite realizar operaciones de filtrado, rango, ejemplo, retirar o modificar los datos de las columnas individuales, reemplazar.

Transform

1 column selected

REMOVE

COPY

FILTER

equals

Enter value

APPLY

JsonUltimoPar...ccion1

Select columns to transform (hold)

✗ COMMIT TRANSFORM

- Rename
- Change Type
- Remove
- Copy
- Filter
- Range
- Sample
- Sort
- Replace
- Split
- Show Details...
- Sort View (Ascending)
- Sort View (Descending)
- Cancel Sorting

faceld Cat	parteSupe...	anchura Category	altura Category
fb		139.0	139.0
at		291.0	115.0
fo		72.0	102.0
7a		592.0	115.0
6a265e18-a...	225.0	254.0	68.0

1,983 rows - 46 columns (41 nc)

Figura 0.42 Turbo Prep: Transformación de columnas

Limpia: Estas funciones trabajan sobre valores perdidos, duplicados, con la normalización y el agrupamiento. Los datos de baja calidad, del tipo que se describe en Auto Model, se pueden eliminar automáticamente a través de Auto Cleansing como se muestra en la Figura 0.43.

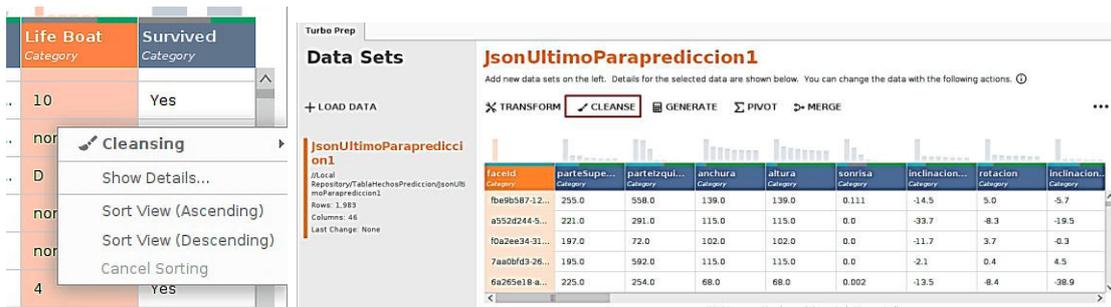


Figura 0.43 Turbo Prep: Función limpieza de datos

Generar: estas funciones lo ayudan a generar nuevas columnas de datos a partir de columnas de datos existentes, Figura 0.44. La gran variedad de operadores lógicos y matemáticos es especialmente útil para la ingeniería de características y para transformaciones de datos más complejas.

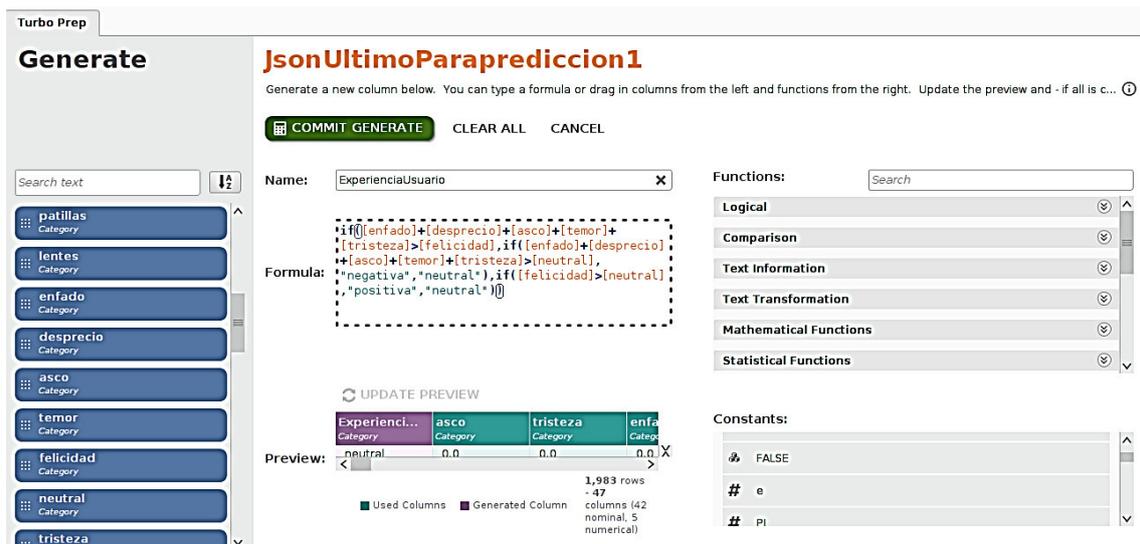


Figura 0.44 Turbo Prep: Generar columna

Pivot: Sus funciones se muestran en la Figura 0.45 las que simplifican la tarea de crear tablas de resumen (tablas dinámicas) a partir de sus datos.

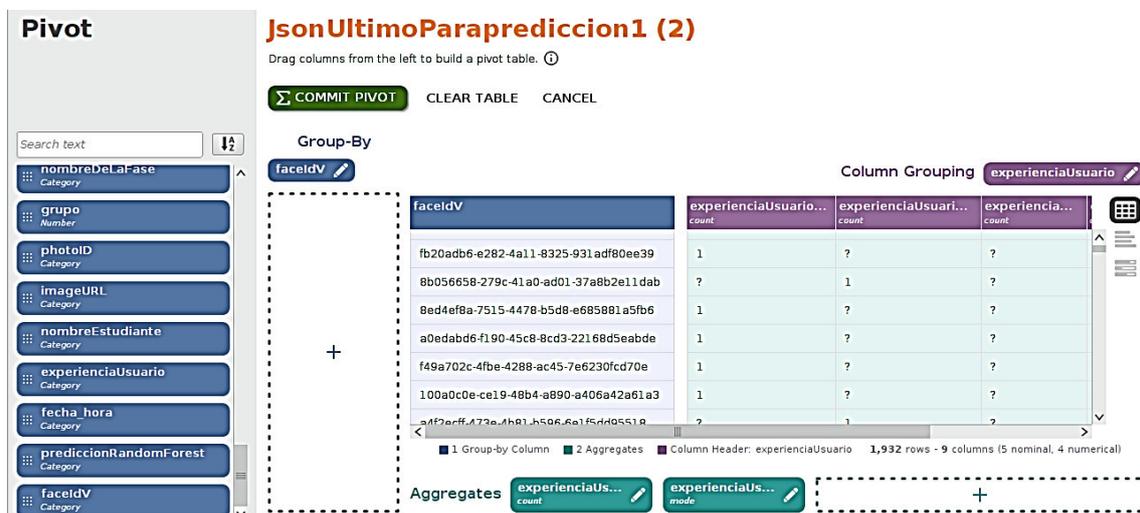


Figura 0.45 Turbo Prep: Función pivot

Fusiona: estas funciones le ayudan a combinar dos o más conjuntos de datos (Unión). Una vez se haya terminado de preparar los datos se puede realizar acciones adicionales, que incluyen:

Modelo: para pasar los datos a Auto Model para construir un modelo ,Figura 0.46.

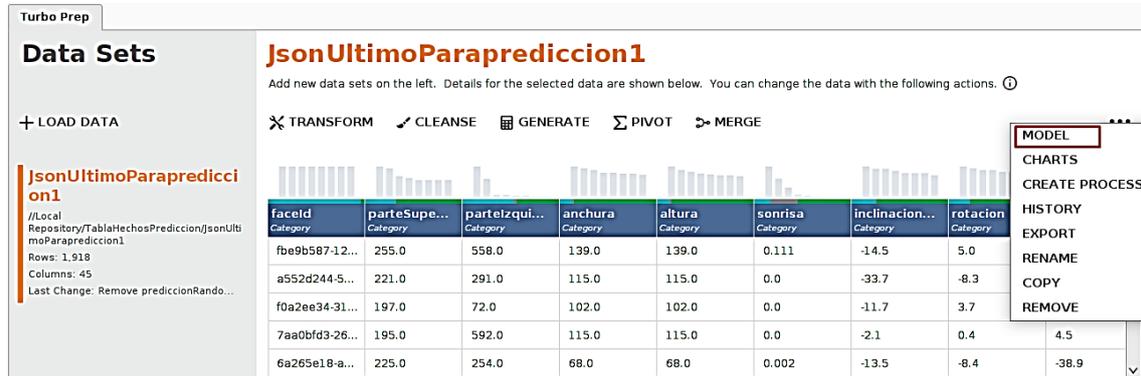


Figura 0.46 Turbo Prep: Modelado

Gráficos: muestra los datos utilizando una variedad de gráficos.

Proceso: guarda los pasos de preparación de datos como un proceso RapidMiner para reutilizarlos más tarde.

Historial: examina el historial de la preparación de sus datos, permite retroceder a un paso anterior y realizar cambios, Figura 0.47.

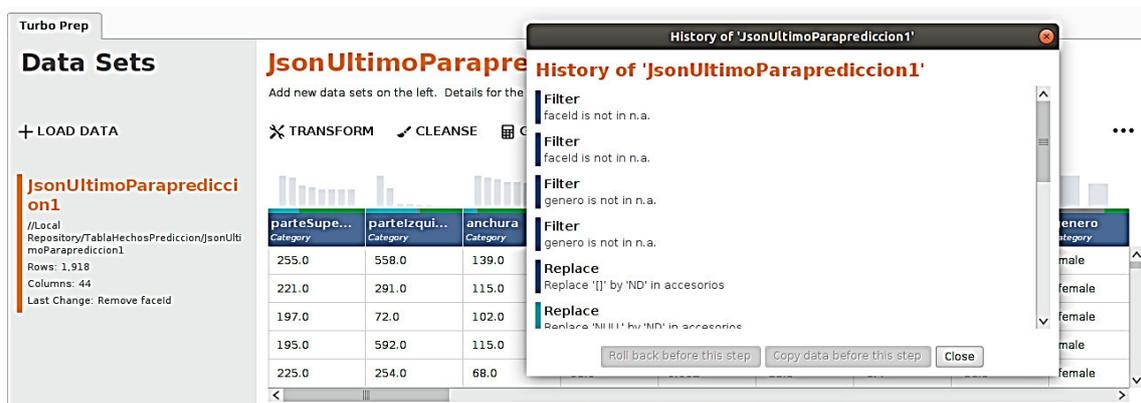


Figura 0.47 Turbo Prep: Historial de cambios

Exportar: para guardar los datos en un archivo o en un repositorio de RapidMiner.

Al terminar los cambios se realiza commit de la transformación para que se ejecuten los cambios en el conjunto de datos, como se muestra en la Figura 0.48.

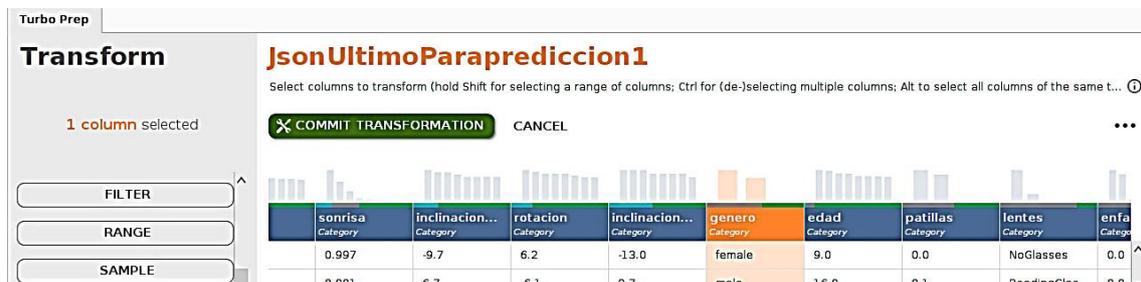


Figura 0.48 Turbo Prep: Commit de transformación

Luego de realizar todas las transformaciones y que se hayan ejecutado, se procede a generar el modelo de predicción, Figura 0.49.

The screenshot shows the Turbo Prep interface. On the left, there's a 'Data Sets' panel with 'JsonUltimoParaprediccion1' selected. The main area shows a data table with the following columns and values:

faceld	parteSupe...	partelzqui...	anchura	altura	sonrisa	inclinacion...	rotacion
Category	Category	Category	Category	Category	Category	Category	Category
f9e9b587-12...	255.0	558.0	139.0	139.0	0.111	-14.5	5.0
a552d244-5...	221.0	291.0	115.0	115.0	0.0	-33.7	-8.3
f0a2ee34-31...	197.0	72.0	102.0	102.0	0.0	-11.7	3.7
7aa0bfd3-26...	195.0	592.0	115.0	115.0	0.0	-2.1	0.4
6a265e18-a...	225.0	254.0	68.0	68.0	0.002	-13.5	-8.4

A dropdown menu is open on the right, showing options: MODEL, CHARTS, CREATE PROCESS, HISTORY, EXPORT, RENAME, COPY, and REMOVE. The 'MODEL' option is highlighted.

Figura 0.49 Turbo Prep: Opción Modelado

Auto Model

Es una extensión de RapidMiner Studio que acelera el proceso de creación y validación de modelos.

Una de sus categorías es la Predicción, que puede resolver problemas de clasificación y regresión. Auto Model ayuda a evaluar los datos, proporciona modelos relevantes para la solución del problema y comparar los resultados de estos modelos, una vez que se completan los cálculos.

Carga de datos

Después de iniciar Auto Model, seleccionar el conjunto de datos de uno de sus repositorios en la sección "Seleccionar Datos para un Nuevo Modelo". Si los datos no están en un repositorio, hacer clic en la sección "Carga de Resultados".

Seleccionar el conjunto de datos nuevo y haga clic en Next en la parte inferior de la pantalla, Figura 0.50.

The screenshot shows the Auto Model interface. At the top, there's a progress bar with steps: Load Data, Select Task, Prepare Target, Select Inputs, Model Types, and Results. The 'Load Data' step is active. Below the progress bar, there are buttons for 'RESTART', 'BACK', and 'NEXT'. The main area is divided into three sections:

- Recent Data Sets:** Shows a list of data sets, including 'GruposJSONparaPrediccion1'.
- Load Results:** A section for loading results, with a note: "No results have been stored so far. Select a data set above to start a new Auto Model run or select a folder with results below." and a button 'SELECT RESULTS FOLDER'.
- Select Data for a New Model:** A tree view showing a folder 'ModeloPrediccionV1' containing several data sets, with 'GruposModeloV1' selected.

On the right, there's an 'Information' panel for 'GruposModeloV1' with the following details:

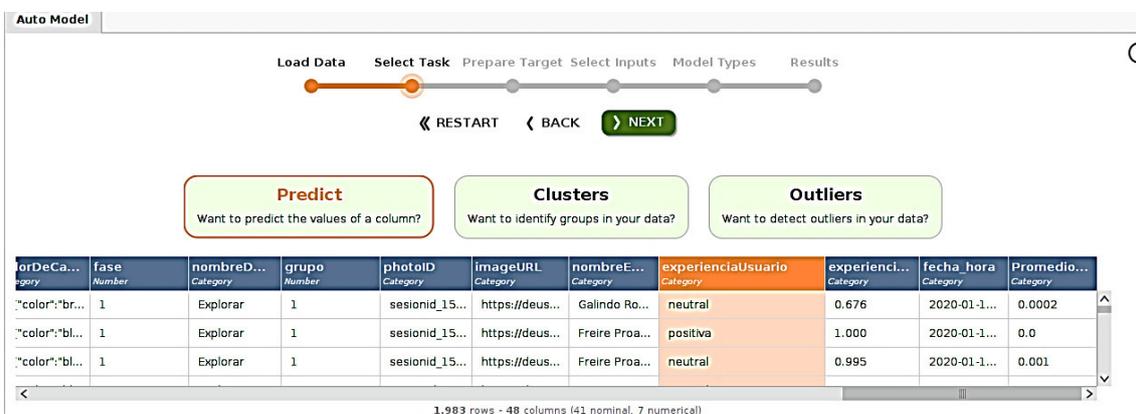
- Name: GruposModeloV1
- Number of rows: 1,983
- Number of columns: 48
- Attributes / Columns

Figura 0.50 Turbo Prep: Carga de datos para el modelado

Seleccionar tarea

Una vez seleccionado un conjunto de datos, se debe decidir qué tipo de problema se va a resolver como se aprecia en la Figura 0.51. Auto Model identifica tres tareas diferentes:

- Predecir
- Clusters
- Valores atípicos



forDeCa... Category	fase Number	nombreD... Category	grupo Number	photoID Category	imageURL Category	nombreE... Category	experienciaUsuario Category	experienci... Category	fecha_hora Category	Promedio... Category
"color": "br...	1	Explorar	1	sesionid_15...	https://deus...	Galindo Ro...	neutral	0.676	2020-01-1...	0.0002
"color": "bl...	1	Explorar	1	sesionid_15...	https://deus...	Freire Proa...	positiva	1.000	2020-01-1...	0.0
"color": "bl...	1	Explorar	1	sesionid_15...	https://deus...	Freire Proa...	neutral	0.995	2020-01-1...	0.001

1.983 rows - 48 columns (41 nominal, 7 numerical)

Figura 0.51 Turbo Prep: Selección de tarea

Preparar objetivo

En la Figura 0.52 se muestra un gráfico de barras con el número de puntos de datos en cada clase, esto es en la sección de Auto Model para problemas de clasificación. Dado experiencia de usuario tiene 3 valores posibles "positivo", "negativo" y "neutral" el problema es de clasificación.

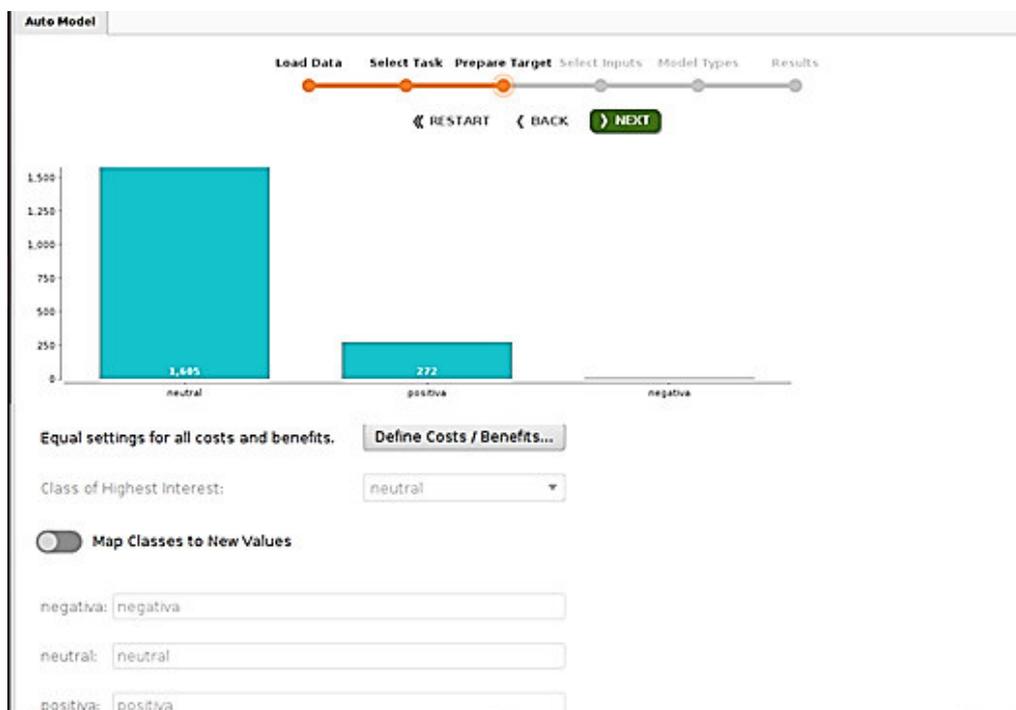


Figura 0.52 Turbo Prep: Preparación de objetivos

Opciones

- Igualdad de condiciones para todos los costos y beneficios: Auto Model incluye una métrica de rendimiento que le permite asignar un costo a cada elemento de la matriz de confusión, de modo que los resultados ya no se identifiquen como meramente verdaderos o falsos, sino en términos de ganancias y pérdidas.
- Definir Costos y Beneficios: Permite acceder a un cuadro de diálogo para definir los costos, éstos se identifican como números negativos y los beneficios como positivos, Figura 0.53.

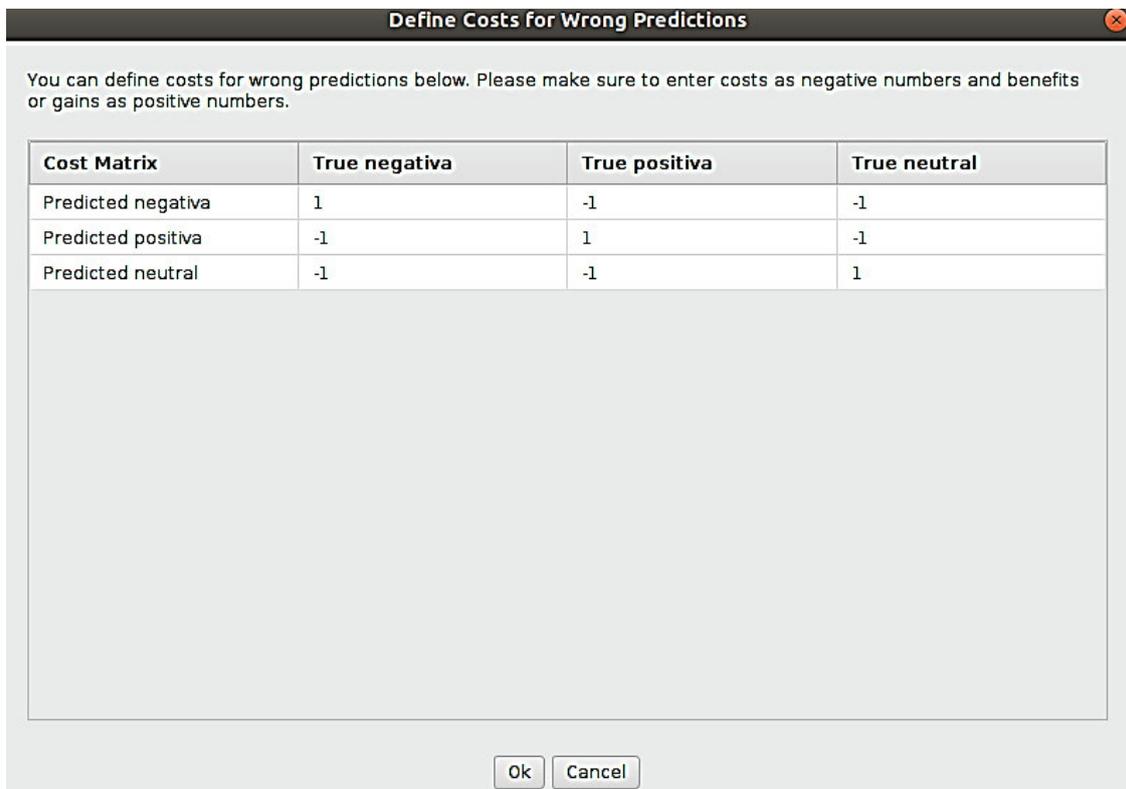


Figura 0.53 Turbo Prep: Definición de costos

- Clase de mayor interés: Permite manejar la predicción a nivel de clasificación para dividir por categoría, se puede tomar las clases necesarias. Este parámetro se vuelve importante más adelante, cuando se presentan los resultados, porque los valores de rendimiento como "Precisión" y "Recuperación" dependen de saber cuál de las clases debe interpretarse como un resultado "positivo". En el presente proyecto, el Class of Highest Interest es "positivo".
- Asignar clases a nuevos valores: Este paso incluye la opción de cambiar el nombre de los valores objetivo de "positivo", "negativo" y "neutral" a algunos otros valores. Cuando hay más de dos clases, esta opción puede ser más útil, porque se puede usar para combinar clases.

Seleccionar entradas

No todas sus columnas de datos ayudan para hacer una predicción. Al descartar algunas de las columnas de datos se puede acelerar el modelo y / o mejorar su rendimiento. Un punto clave para discernir columnas es que se está buscando patrones. Sin alguna variación en los datos y algunos patrones discernibles, es probable que los datos no sean útiles.

En la Figura 0.54 se aprecia un resumen rápido de cosas a tener en cuenta incluye lo siguiente, cuyos valores se muestran junto a las barras de calidad para cada columna de datos como las siguientes:

- Columnas que reflejan demasiado de cerca la columna de destino, o que no reflejan en absoluto (correlación),
- Columnas donde casi todos los valores son diferentes (ID-ness),
- Columnas donde casi todos los valores son idénticos (estabilidad),
- Columnas con valores perdidos (Falta).

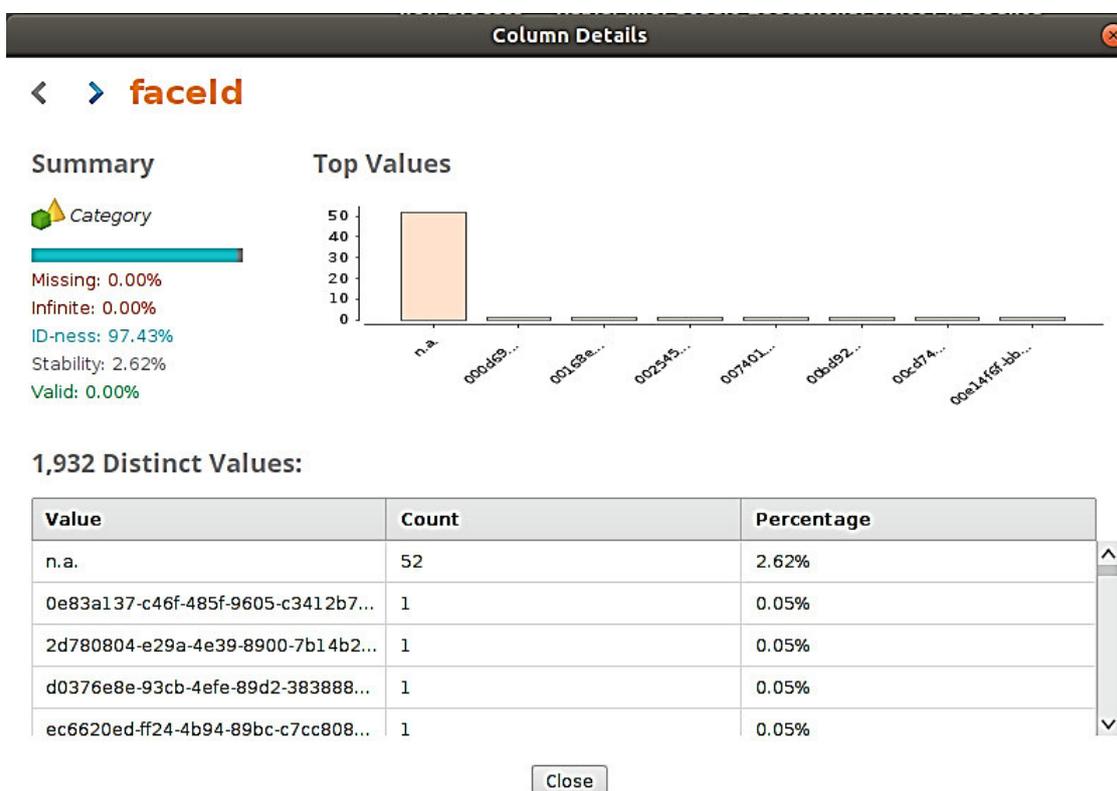


Figura 0.54 Automodelado: Detalle de las columnas

Auto Model resume la situación con una burbuja de estado codificada por colores (rojo / amarillo / verde). Como regla general, es una buena idea anular la selección de al menos aquellas columnas que tienen una burbuja de estado roja que no aportan al modelo, pero se puede anular la selección de las columnas que desee, independientemente de su estado. La entrada para el modelo de aprendizaje automático solo incluirá las columnas seleccionadas, Figura 0.55.

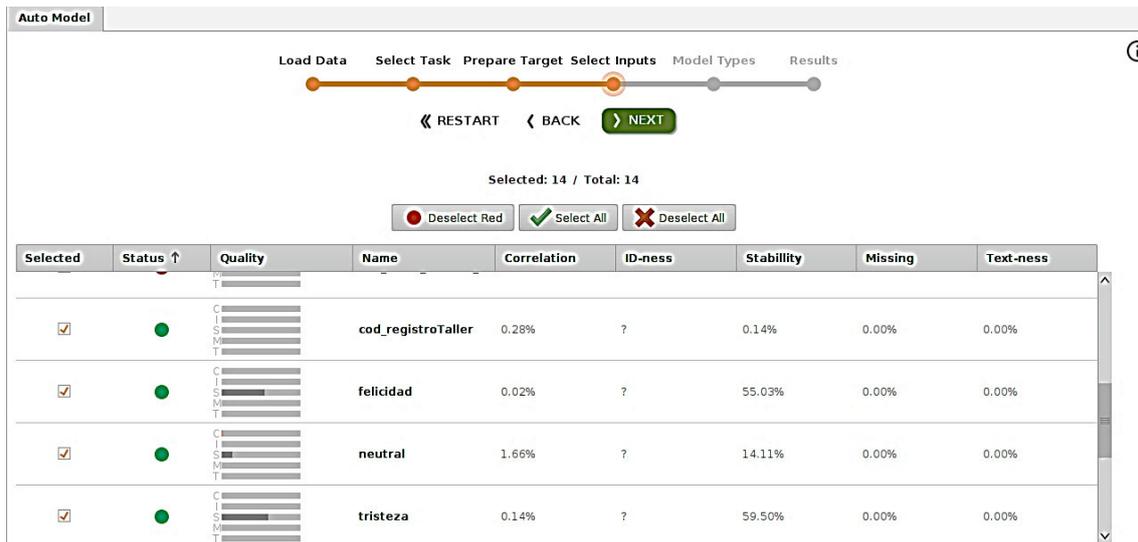


Figura 0.55 Automodelado: Selección de entradas

Tipos de modelos

Auto Model proporciona una selección de modelos, Figura 0.56. Si no se tiene límite de tiempo es mejor construirlos todos y comparar su desempeño. Es recomendable decidir la prioridad entre precisión del modelo terminado o el tiempo que lleva construirlo. Auto Model proporciona los siguientes modelos:

- Bayes Nativo
- Modelo lineal generalizado
- Regresión logística
- Aprendizaje profundo
- Árbol de decisión
- Bosque aleatorio
- Árboles potenciados por degradado (XGBoost)

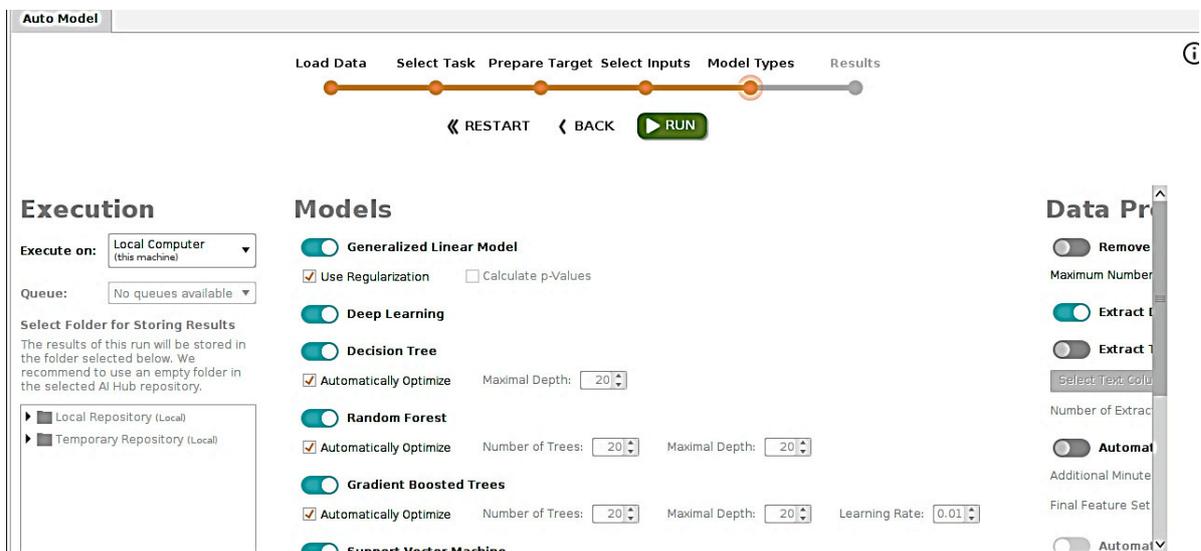


Figura 0.56 Auto Modelado: Tipos de modelos

Resultados

Dependiendo de su conjunto de datos y los modelos seleccionados, es posible que se deba esperar los resultados como se muestra en la Figura 0.57. La barra de progreso en la parte superior rastrea el estado de un cálculo en curso. Es posible detener la construcción de modelos en cualquier momento presionando el botón “Stop”. Los resultados intermedios se muestran a medida que están disponibles, por ejemplo, en Comparison> Overview.

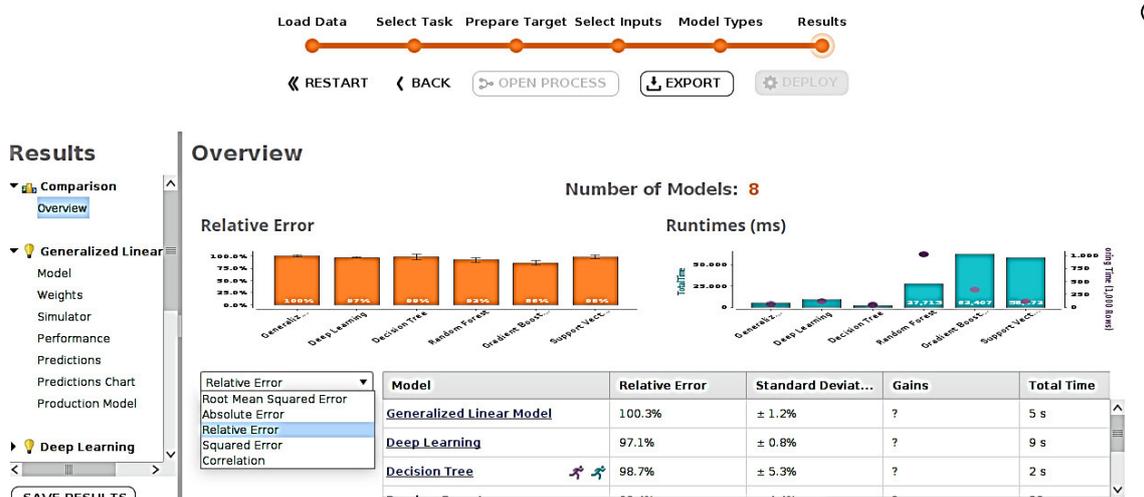


Figura 0.57 Auto Modelado: Resultados

Comparación de modelos

Una vez que haya elegido una métrica de rendimiento, puede utilizar la comparación de modelos para encontrar el mejor modelo, de acuerdo con esa métrica como se muestra en la Figura 0.58.

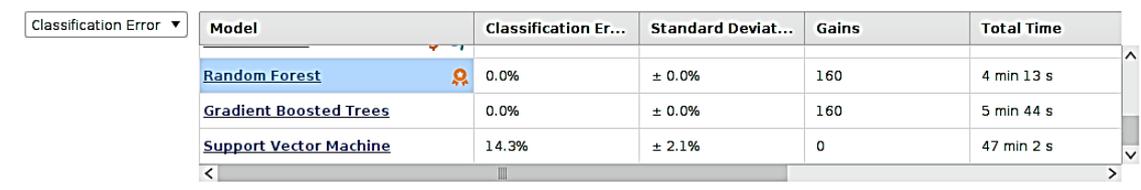


Figura 0.58 Automodelado:Comparación de modelos

Model

Representa de forma gráfica o en una tabla los datos dependiendo del modelo aplicado, en este caso en la Figura 0.59 se muestra la gráfica de árbol del modelo Random Forest.



Figura 0.59 Automodelado:Modelo Random Forest

Pesos

En la Figura 0.60 se puede apreciar los pesos, para obtener más información sobre los datos, independientemente del modelo. Los pesos le indican qué datos de entrada es más probable que afecten la predicción.

Results

- Decision Tree
- Random Forest
 - Model
 - Weights**
 - Simulador
 - Performance
 - Lift Chart
 - Optimal Parameters
 - Predictions
 - Production Model

Random Forest - Weights

Attribute	Weight
experienciaUsuario	0.278
lentes	0.131
neutral	0.099
felicidad	0.092
invisible	0.091
nivelDeRuido	0.087
accesorios	0.086
fase	0.076
identificacionDelNinio	0.061

Figura 0.60 Automodelado:Random Forest -Pesos

Simulador de modelos

Auto Model ayuda a obtener resultados y también a comprender esos resultados. Dispone de una interfaz de usuario con controles deslizantes y listas desplegadas a la izquierda, y gráficos de barras a la derecha, Figura 0.61. Para su estado inicial, Model Simulator se elige valores de datos promedio. En el presente proyecto el promedio corresponde al rostro de un estudiante de dimensiones 122 x120 y que relativamente no tiene calvicie.

El escenario más probable, de acuerdo a la Figura 0.61, es que el estudiante presentara una emoción neutral en el taller de RE.

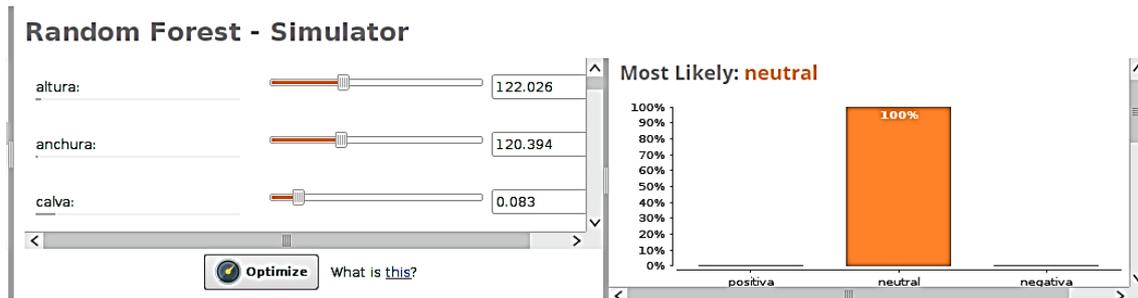


Figura 0.61 Automodelado: Random Forest - Simulador

Optimizar entradas

En el modelado predictivo, se utiliza el mismo para predecir un resultado, dado una entrada. Esta funcionalidad invierte ese procedimiento, comenzando con un modelo y una salida deseada, y prescribiendo una entrada optimizada para lograr el resultado deseado, Figura 0.61.

Este proceso también se denomina análisis prescriptivo, ya que el modelo creado no sólo se utiliza para predecir el futuro, se puede definir un resultado deseado al elegir uno de los siguientes objetivos en el primer paso:

- minimizar la confianza para una clase
- maximizar la confianza para una clase
- acercarse lo más posible a una cierta confianza para una clase
- minimizar la predicción de regresión
- maximizar la predicción de regresión
- acercarse lo más posible a una determinada predicción de regresión

Se tiene un modelo de clasificación con tres posibles resultados y realmente se desea lograr uno de los dos resultados para la situación actual. En este caso, se trataría de maximizar la confianza para la clase que le gustaría lograr, por ejemplo, experiencia positiva.

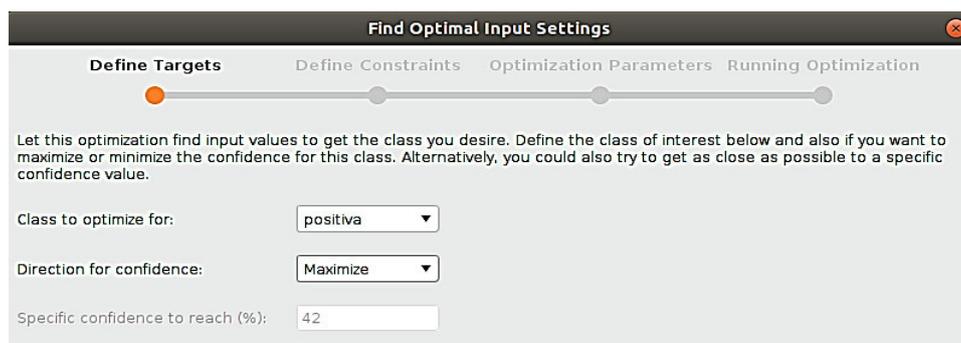


Figura 0.62 Automodelado: Encontrar entradas óptimas

Una vez que se haya especificado el resultado deseado, RapidMiner ejecuta un algoritmo de optimización en segundo plano que intenta encontrar la combinación óptima de entradas para el modelo (la configuración en el lado izquierdo de la pantalla del simulador) de modo que se logre el resultado deseado, Figura 0.63.

En muchos casos, no se desea permitir ciertos valores para las entradas, en estos casos RM presenta un conjunto de opciones para restringir la búsqueda de entradas óptimas:

- mantenerse cerca del promedio de columnas numéricas, dentro de 1/2/3 veces la desviación estándar
- mantenerse por encima del mínimo
- permanecer por debajo del máximo
- mantenerse por encima de un cierto valor
- mantenerse por debajo de un cierto valor

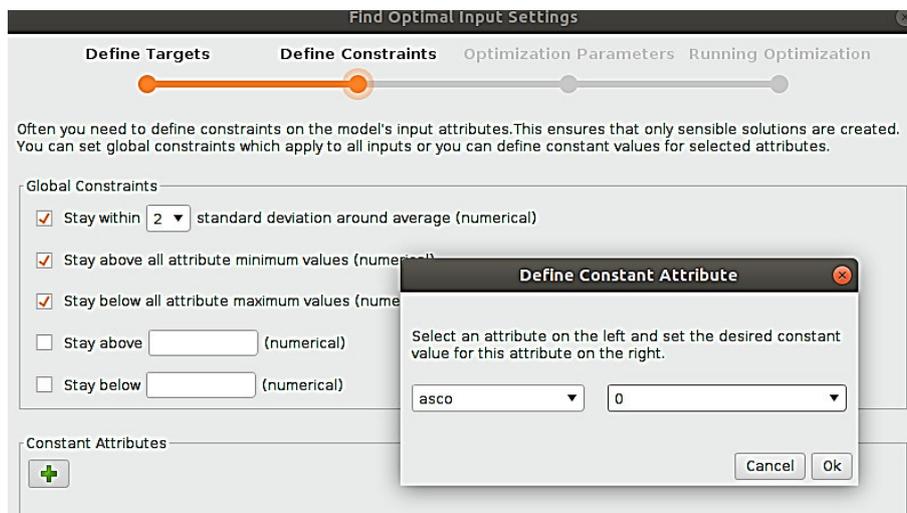


Figura 0.63 Auto Modelado: Definir limitaciones y constantes

Se puede asignar valores constantes a cualquiera de los atributos, anulando las condiciones anteriores, por ejemplo, no permitir experiencia negativa.

Además, es posible definir los parámetros de optimización como el tiempo que se empleará para realizarla, Figura 0.64.

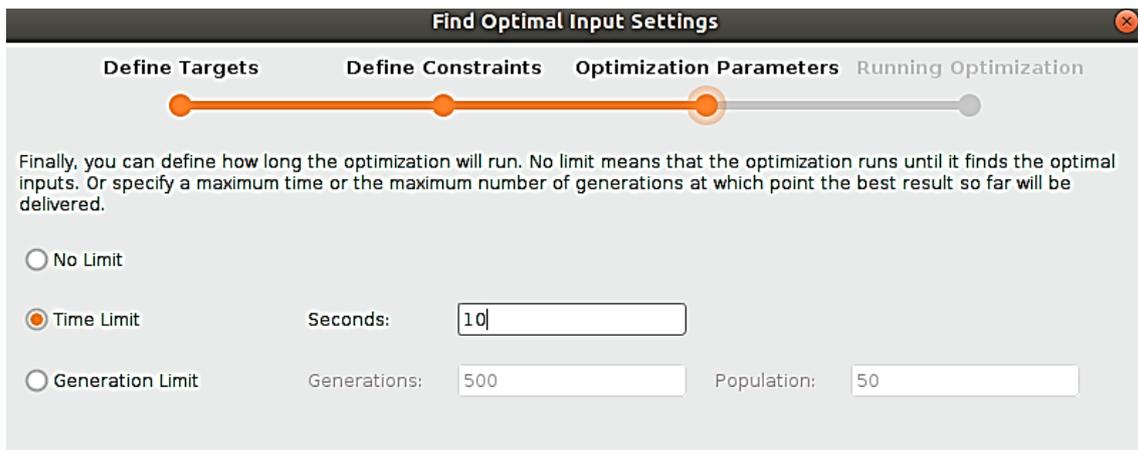


Figura 0.64 Auto Modelado: Optimización de Parámetros

En la Figura 0.65 el gráfico de barras explica los factores importantes para obtener como resultado a la emoción neutral, barras de color verde factores a favor y rojas en contra, la experiencia de usuario y rotación que presenta el rostro se muestran como barras verdes. Verde en este contexto implica que la clase experiencia de usuario y la rotación que presenta el rostro está de acuerdo con la predicción de experiencia de usuario, es decir, "neutral". Las barras rojas de la sonrisa y el nivel de difuminación de la imagen implican un desacuerdo con la predicción y, por lo tanto, una correlación positiva con la experiencia de usuario.

Important Factors for neutral

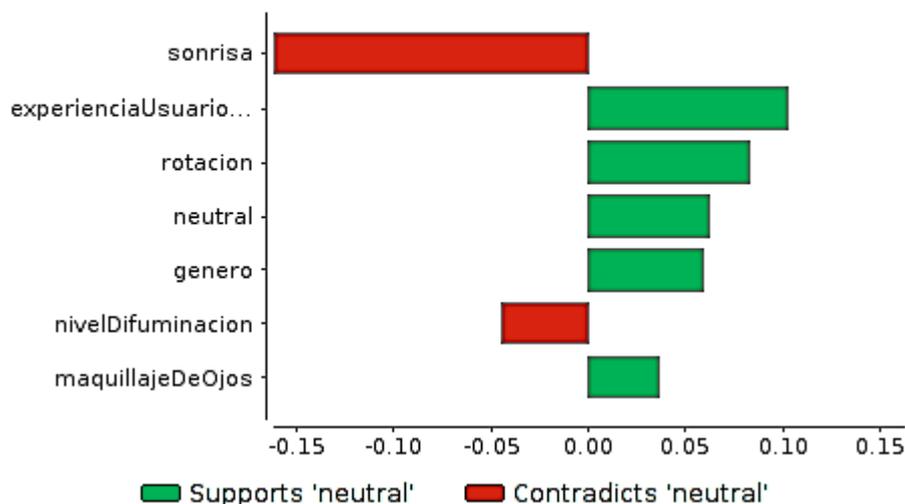


Figura 0.65 Automodelado: Factores de correspondencia a variables

El simulador de modelos es interactivo, por lo que se puede cambiar todos los valores a voluntad y ver inmediatamente el impacto en las predicciones. Al manipular todos los controles deslizantes y listas desplegadas, puede crear rápidamente algo de intuición para el modelo.

Se determina mediante la Figura 0.66 los mejores factores de entrada bajo las limitaciones definidas anteriormente.

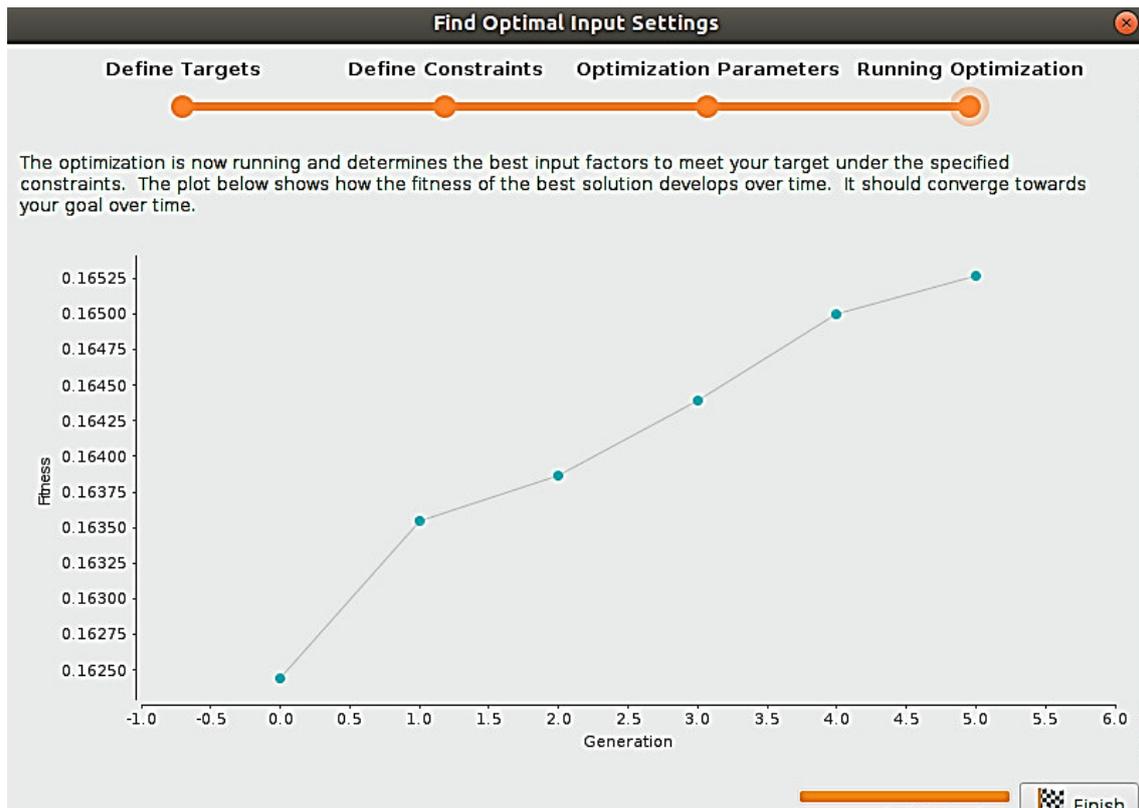


Figura 0.66 Automodelado: Corrida de la optimización

Rendimiento

Permite revisar los valores de los criterios del rendimiento del modelo.

Random Forest - Performance		
Criterion	Value	Standard Deviation
Accuracy	91.7%	± 1.7%
Classification Error	8.3%	± 1.7%
AUC	92.8%	± 3.2%
Precision	93.0%	± 1.8%
Recall	98.2%	± 0.9%
F Measure	95.5%	± 0.9%
Sensitivity	98.2%	± 0.9%
Specificity	31.9%	± 16.0%

Figura 0.67 Automodelado: Rendimiento de Random Forest

Tabla de elevación

Un gráfico de elevación, Figura 0.68, muestra cuanto mejor se desempeña un modelo de aprendizaje automático en comparación con una suposición aleatoria. También le muestra el punto en el que las predicciones se vuelven menos útiles.

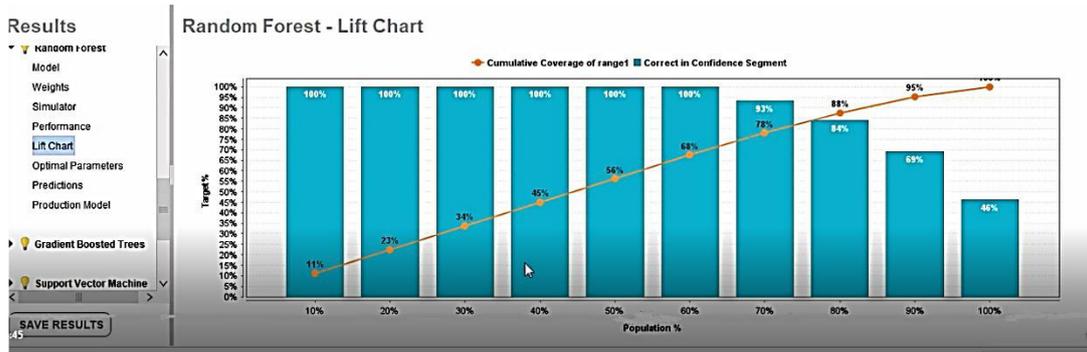


Figura 0.68 Automodelado: Tabla de Elevación

Predicción

En la Figura 0.69 se muestra un esquema de colores de la predicción de Random Forest, si cuando se ejecutó el modelo automático, se marcó la casilla de verificación Explicar predicciones, los datos de puntuación están codificados por colores para indicar su importancia para la predicción, los valores de color verde oscuro respaldan fuertemente la predicción para esa fila de datos, los valores de rojo oscuro respaldan fuertemente una predicción diferente para esa fila de datos, y los colores más claros son menos importantes.

Row No.	experienciaUsuarioNumerica	prediction(experienciaUsuarioNu...	confidence(...	confidence(...	cost	genero	edad
1	range2	range1	0.491	0.509	0.018	male	12
2	range1	range1	0.007	0.993	0.985	female	7
3	range1	range1	0.006	0.994	0.988	female	7
4	range1	range2	0.527	0.473	0.054	female	6
5	range1	range1	0.013	0.987	0.974	female	11
6	range1	range1	0.050	0.950	0.900	male	6

Figura 0.69 Automodelado: Random Forest - Predicción

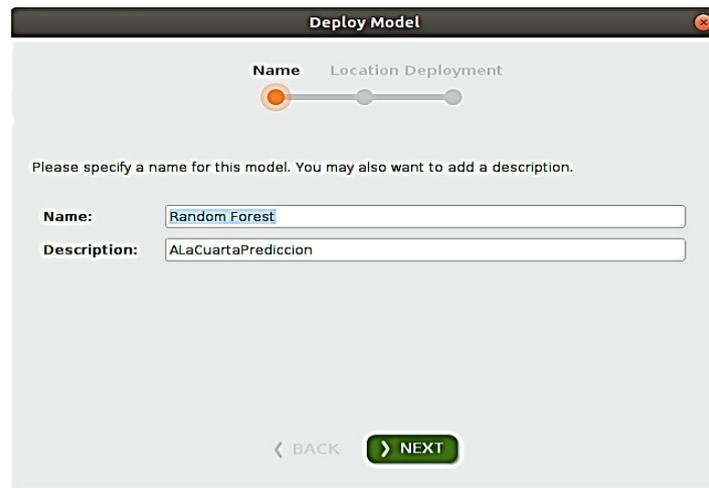
Despliegue

Para obtener el valor total de sus modelos, debe ponerlos en producción como se muestra en la Figura 0.70.



Figura 0.70 Despliegue Random Forest

En la Figura 0.71 se muestran parámetros del despliegue del modelo, se elige un nombre para el modelo.



Deploy Model

Name Location Deployment

Please specify a name for this model. You may also want to add a description.

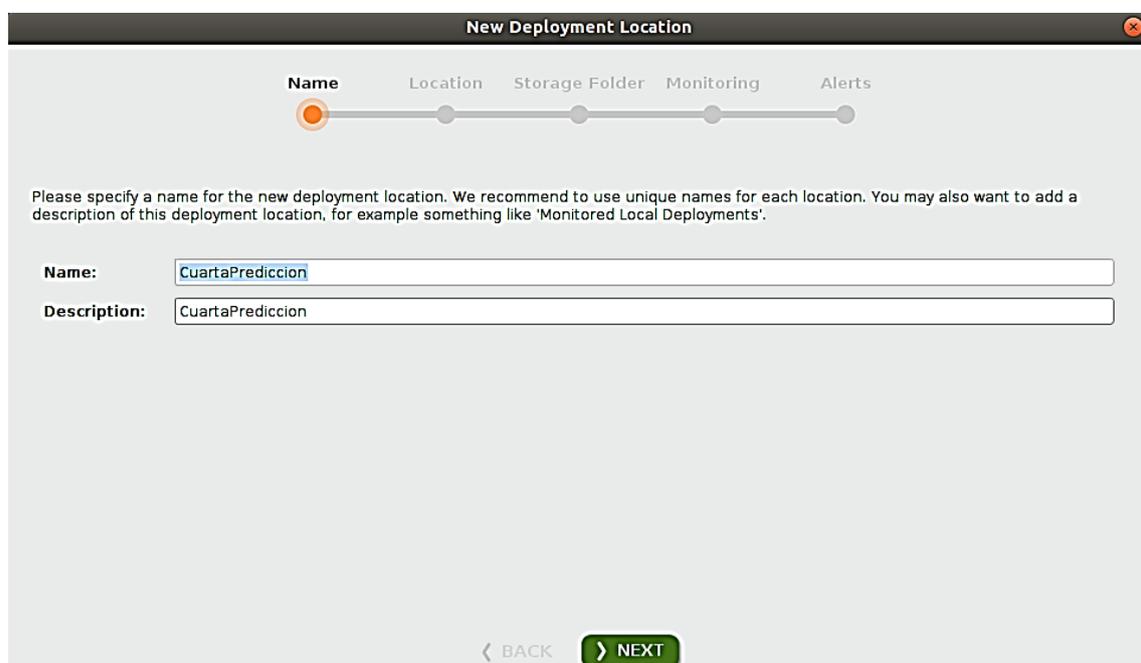
Name: Random Forest

Description: ALaCuartaPrediccion

< BACK > NEXT

Figura 0.71 Despliegue del modelo

En la Figura 0.72 se muestran parámetros del despliegue del modelo como la localización



New Deployment Location

Name Location Storage Folder Monitoring Alerts

Please specify a name for the new deployment location. We recommend to use unique names for each location. You may also want to add a description of this deployment location, for example something like 'Monitored Local Deployments'.

Name: CuartaPrediccion

Description: CuartaPrediccion

< BACK > NEXT

Figura 0.72 Localización del despliegue

Se elige crear una ubicación que puede ser implementación local o remota, Figura 0.73. Las implementaciones remotas incluirán alertas e integraciones. Por el contrario, implementación local no incluye estos componentes.

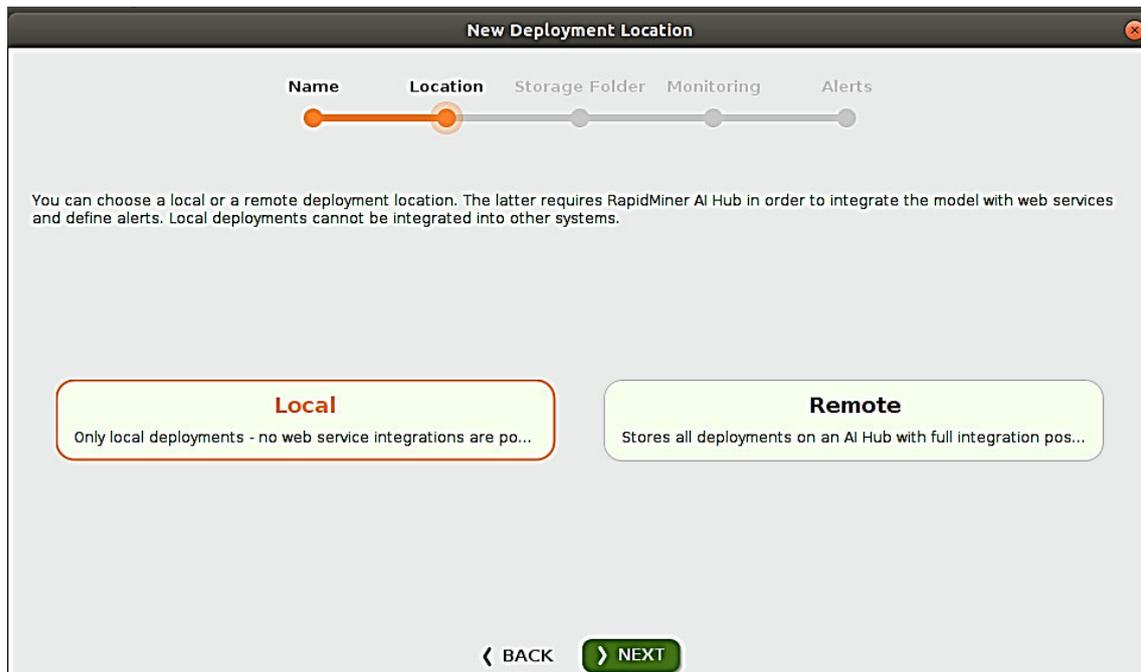


Figura 0.73 Nueva localización de despliegue

Si aún no se ha creado una carpeta vacía dentro del repositorio, puede hacer clic en Crear nueva carpeta, Figura 0.74.

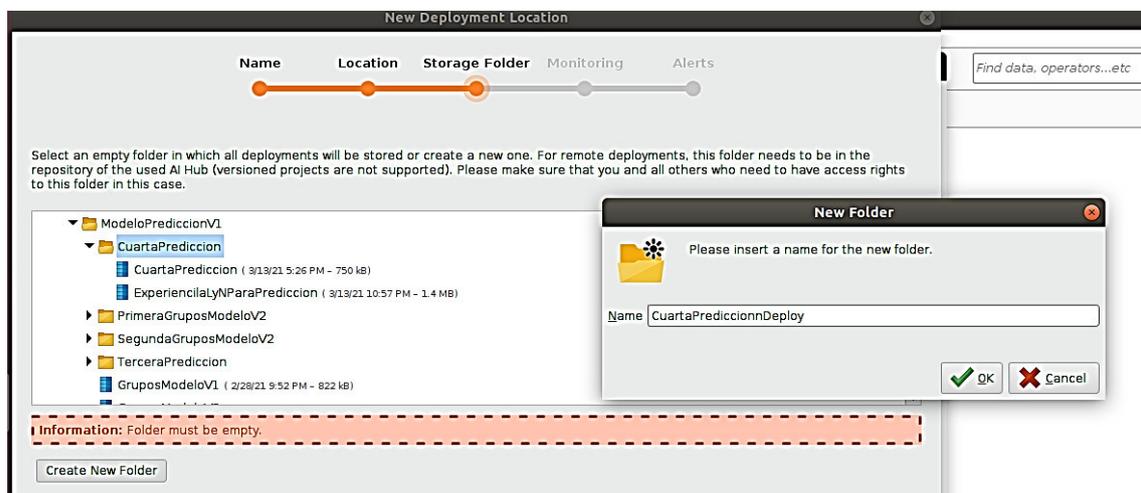


Figura 0.74 Nueva carpeta de almacenamiento de despliegue

Para activar la supervisión, se activa la casilla de Activar Monitoreo y se selecciona una conexión a la base de datos.

Con el monitoreo, Figura 0.75, las implementaciones en esta ubicación incluirán un tablero y resúmenes de desempeño.

En caso de que se haya seleccionado una implementación remota, para activar las alertas de correo electrónico, haga clic en la casilla de verificación y seleccione una conexión Enviar correo.

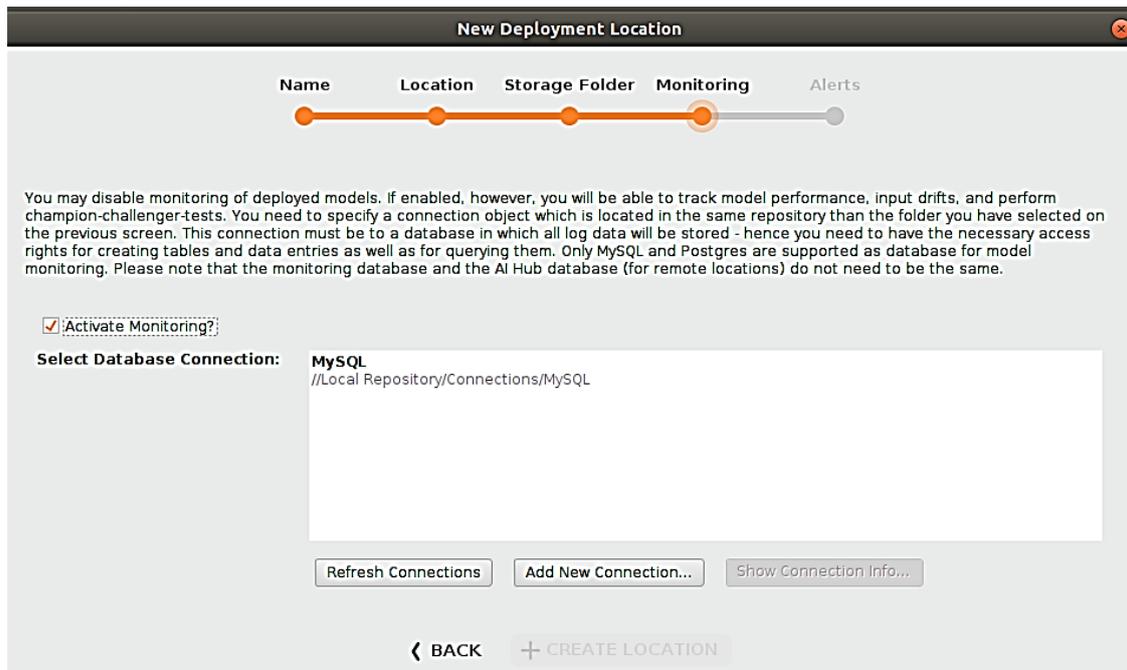


Figura 0.75 Monitoreo de despliegue

Puntuación

Podría decirse que el objetivo principal de una implementación es puntuar los datos, Figura 0.76. Los modelos toman nuevos datos como entrada y devuelven un resultado. Una vez que se ha activado la supervisión, los resultados se recopilan para que pueda mantener estadísticas a largo plazo.

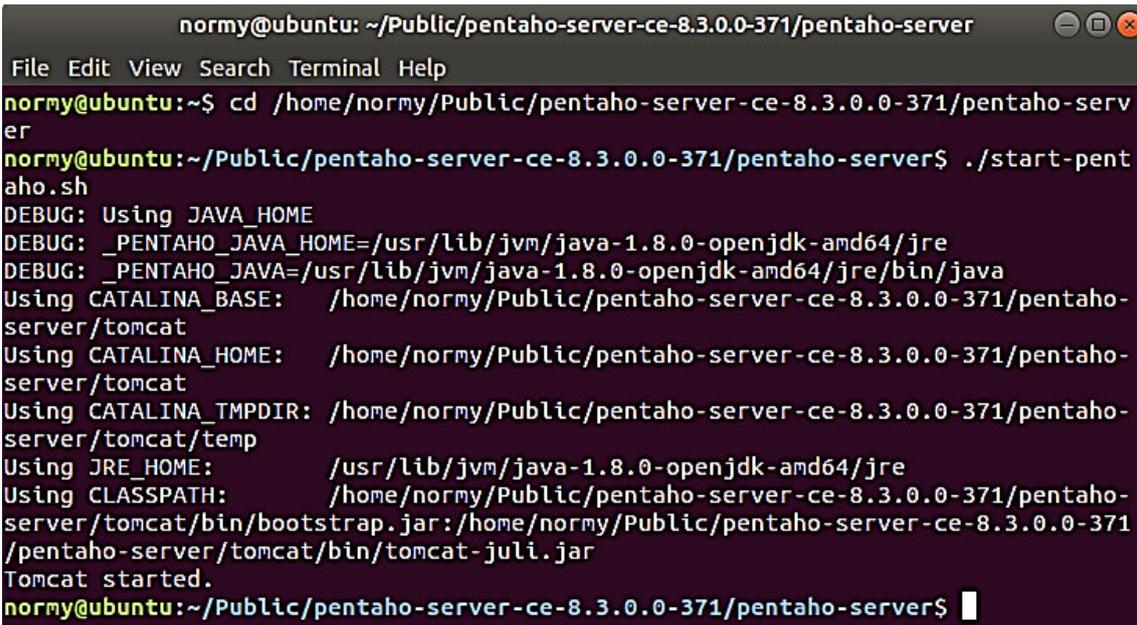
Row No.	cod_regl...	experien...	prediction(experienciaUsua...	dim_tiem...	enfado	disgusto	desprecio	temor	felicidad	neutral
1	717	-0.003	0.001	Jan 13, 2020	0	0	0	0	0	0.990
2	718	-0.003	-0.003	Jan 13, 2020	0	0	0	0	0.020	0.980
3	719	0	0.001	Jan 13, 2020	0	0	0	0	0	1
4	720	0.001	0.001	Jan 13, 2020	0	0	0	0	0	0.820
5	721	0	0.002	Jan 13, 2020	0	0	0	0	0	0.750
6	722	0.001	0.002	Jan 13, 2020	0	0	0.030	0	0	0.960
7	723	0	0.001	Jan 13, 2020	0	0	0	0	0	1
8	724	0	0.002	Jan 13, 2020	0	0	0	0	0	1

Figura 0.76 Puntuación - Despliegue

Anexo XII Tutorial de Pentaho BI-Server

Iniciando Pentaho BI Server.

- Ingresar a la sección de downloads de la web de Pentaho community a efectos de descargar el software: <http://community.pentaho.com/>. Descargar Business Analytics Platform.
- Una vez descargados los archivos, asegurar de tener instalado Java Development Kit (JDK) y Java Runtime Environment (JRE).
- En UNIX se puede instalarlo a través del comando: `sudo apt-get install default-jdk default-jre`.
- Descomprimir el archivo `biserver-ce-X.X.tar.gz`.
- Verificar el archivo `bash.bashrc` que se encuentra en el directorio `/etc` (esto sirve para setear la variable en todos los perfiles). Para ello, abrir una terminal y ejecutar lo siguiente: `sudo nano .bashrc`.
- Agregar en la última línea: `JAVA_HOME=/usr/lib/jvm/jre export JAVA_HOME PATH=$PATH:$JAVA_HOME`. Guardar los cambios para que se actualicen las variables.
- Ingresar en la ruta de Pentaho Server e iniciarlo con comando que se muestra en la Figura 0.77.



```

normy@ubuntu: ~/Public/pentaho-server-ce-8.3.0.0-371/pentaho-server
File Edit View Search Terminal Help
normy@ubuntu:~$ cd /home/normy/Public/pentaho-server-ce-8.3.0.0-371/pentaho-server
normy@ubuntu:~/Public/pentaho-server-ce-8.3.0.0-371/pentaho-server$ ./start-pentaho.sh
DEBUG: Using JAVA_HOME
DEBUG: _PENTAHO_JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64/jre
DEBUG: _PENTAHO_JAVA=/usr/lib/jvm/java-1.8.0-openjdk-amd64/jre/bin/java
Using CATALINA_BASE:  /home/normy/Public/pentaho-server-ce-8.3.0.0-371/pentaho-server/tomcat
Using CATALINA_HOME:  /home/normy/Public/pentaho-server-ce-8.3.0.0-371/pentaho-server/tomcat
Using CATALINA_TMPDIR: /home/normy/Public/pentaho-server-ce-8.3.0.0-371/pentaho-server/tomcat/temp
Using JRE_HOME:      /usr/lib/jvm/java-1.8.0-openjdk-amd64/jre
Using CLASSPATH:     /home/normy/Public/pentaho-server-ce-8.3.0.0-371/pentaho-server/tomcat/bin/bootstrap.jar:/home/normy/Public/pentaho-server-ce-8.3.0.0-371/pentaho-server/tomcat/bin/tomcat-juli.jar
Tomcat started.
normy@ubuntu:~/Public/pentaho-server-ce-8.3.0.0-371/pentaho-server$ █

```

Figura 0.77 Pentaho Server Ce: Arranque

En la Figura 0.78 se muestra el acceso a Pentaho BI Server a través de un navegador mediante la URL: localhost:8080/pentaho/. En el caso de ambientes UNIX debemos cautelar que el archivo tenga permisos de ejecución para el usuario.



Figura 0.78 Pentaho Server Ce: Consola

En caso de existir inconvenientes en el inicio de Pentaho BServer ubicar los logs de Tomcat que se encuentran en la ruta que se muestra n la Figura 0.79.

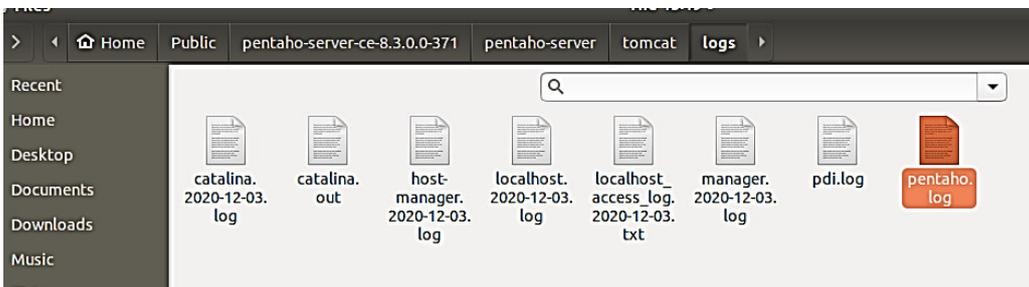


Figura 0.79 Pentaho Server Ce: Ubicación Logs Tomcat

Verificar puertos en Ubuntu disponibles como se aprecia en la Figura 0.80. [43]

```
1 | $ sudo lsof -i -P -n
2 | $ sudo lsof -i -P -n | grep LISTEN
```

Figura 0.80 Comandos para verificar puertos en Ubuntu

SAIKU - Instalación

Ir a Home, Marketplace y buscar la extensión de Saiku Estable, Figura 0.81.

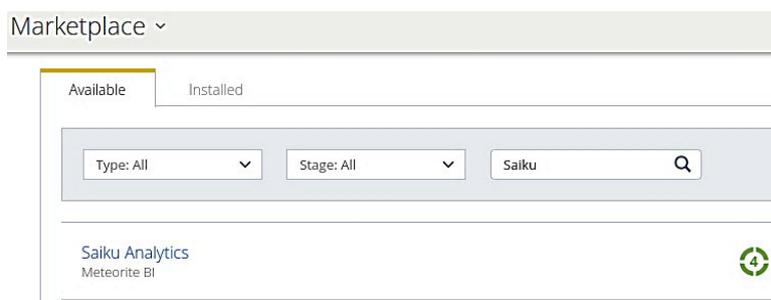


Figura 0.81 Marketplace - Saiku Analytics

En la Figura 0.82 se muestra como descargar una licencia Meteorite.bi <https://www.meteorite.bi/products/saiku/>

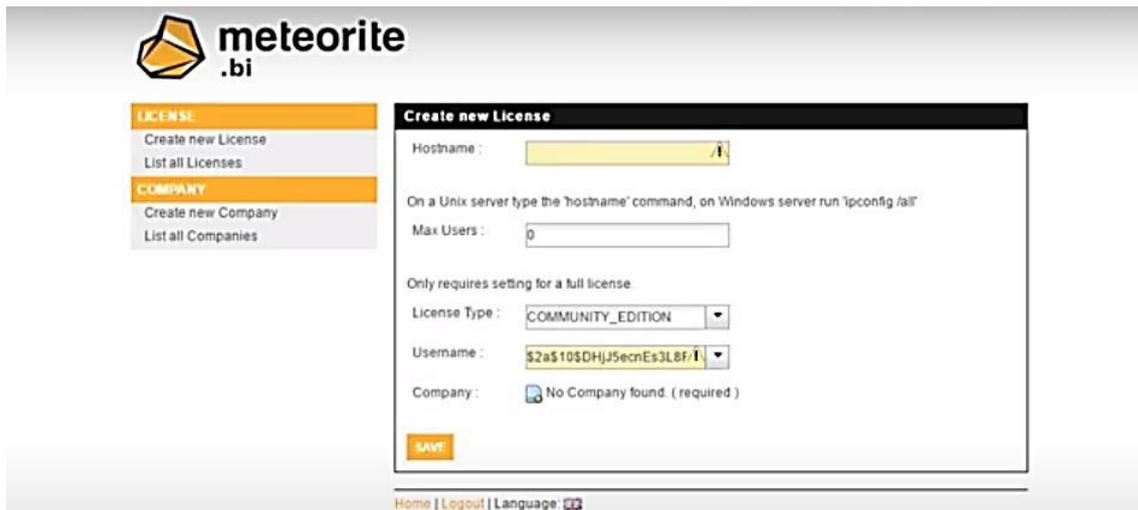


Figura 0.82 Meteorite bi

Mientras no se coloque la licencia en Saiku no funciona e indica el mensaje de error que se muestra en la Figura 0.83.

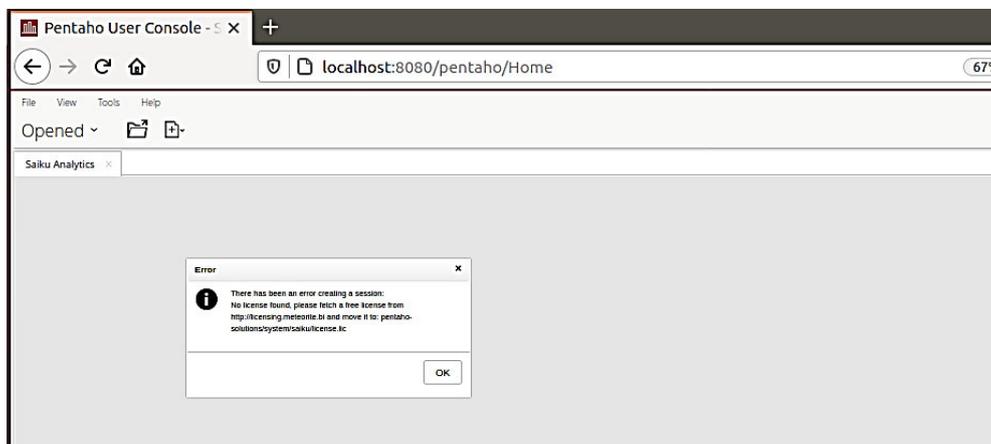


Figura 0.83 Error licencia Saiku

Es necesario parar el pentaho y copiar la licencia en la ruta que se muestra en la Figura 0.84. Posteriormente iniciar el servicio de pentaho.



Figura 0.84 Ruta de licencia de Saiku

Creación de un Data Source

Luego de ingresar en Pentaho Bi-Server, ir a Manage Data Sources como se muestra en la Figura 0.85.

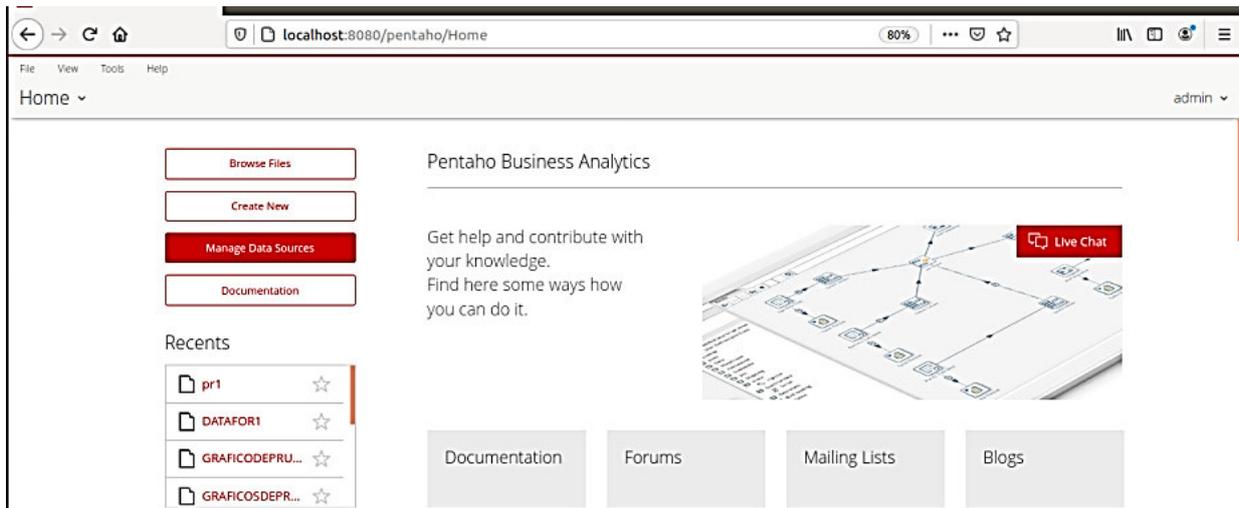


Figura 0.85 Inicio Manage Data Source

Aparece la ventana Manage DataSource, ir a New Data Source, Figura 0.86.

Manage Data Sources

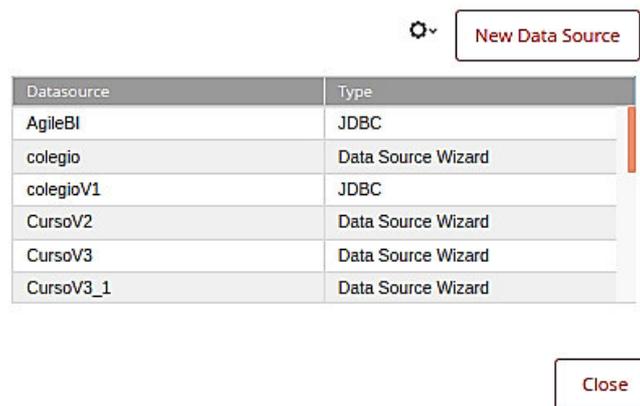


Figura 0.86 Creación de Data Source

Es necesario configurar la conexión a la base de datos, Figura 0.87.

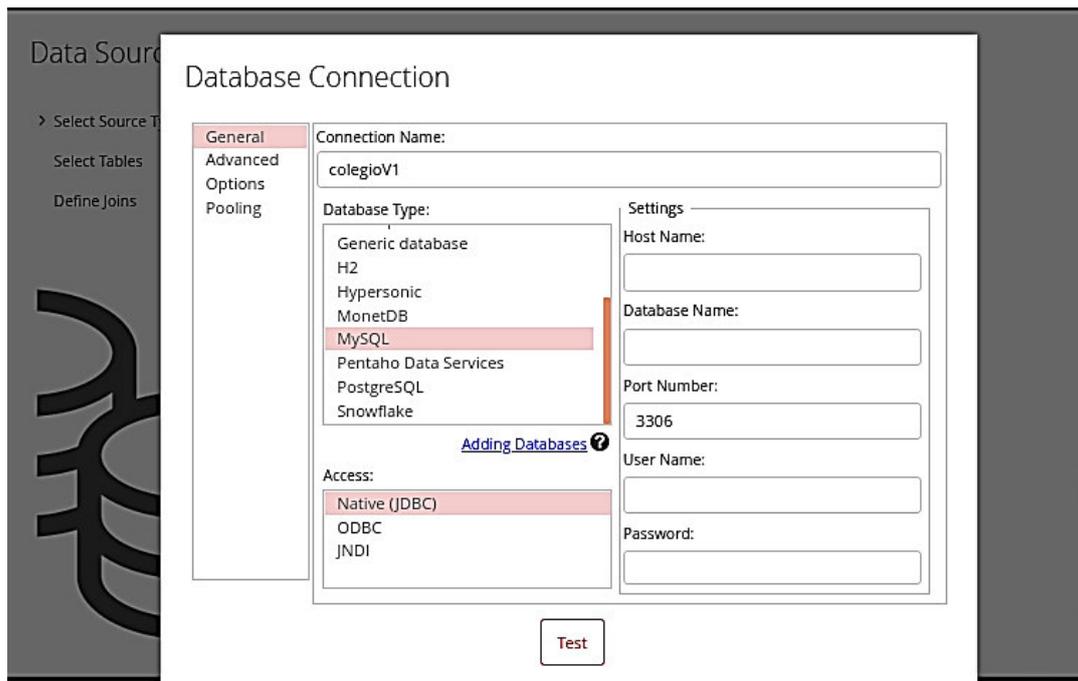


Figura 0.87 Pentaho Server Ce: Conexión a base de datos

Una vez se termine de configurar, en la Figura 0.88 se puede apreciar el test para verificar que la conexión a base de datos sea exitosa.

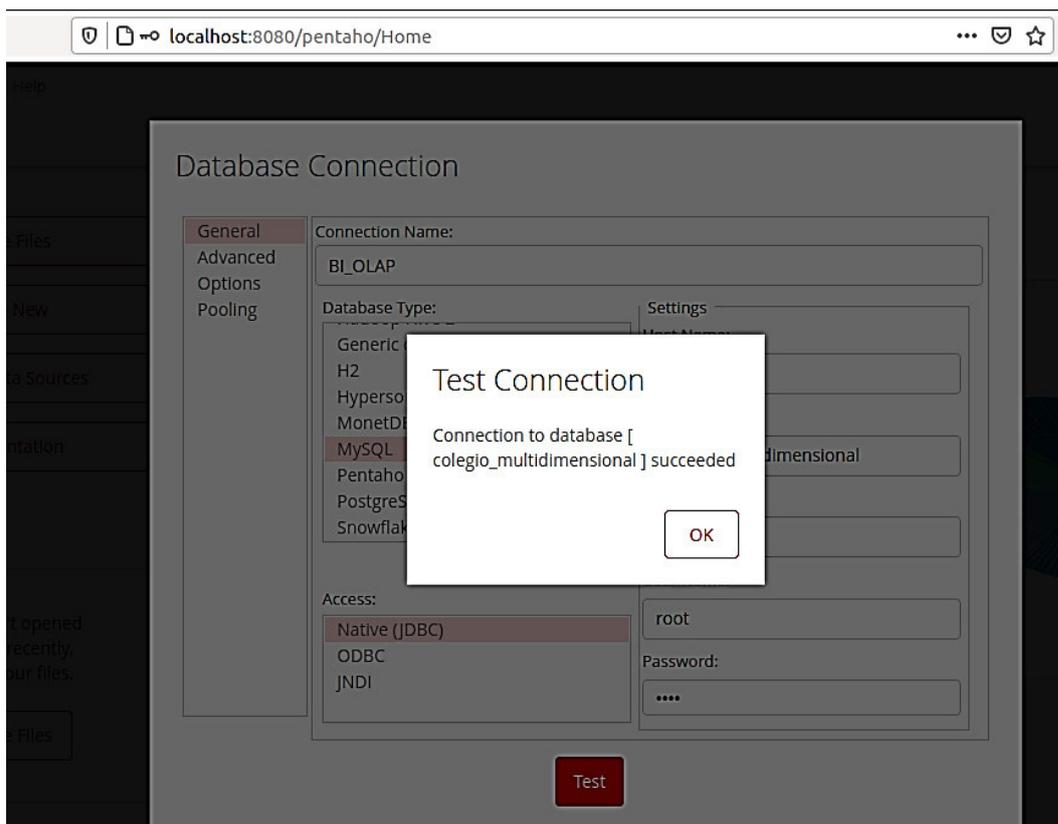


Figura 0.88 Pentaho Server Ce: Prueba de Conexión

Aparece un Wizard para la creación de un Data Source, Figura 0.89, escribir el nombre del Data Source, elegir tablas de la base de datos y la opción Reporte y Análisis (Requiere Esquema Estrella).

Data Source Wizard

> Select Source Type
 Select Tables
 Define Joins

Data Source Name:

Source Type: Database Table(s)

Select a database connection and click Next to choose from a list of the available database tables.

Connection:

- AgileBI
- SampleData
- colegioV1
- multidimensionalcolegioV1

Create data source for:

Reporting only

Reporting and Analysis (Requires Star Schema)

< Back Next > Finish Cancel

Figura 0.89 Asistente para creación de Data Source

Seleccionar las tablas de dimensiones y la tabla de hechos, Figura 0.90.

Data Source Wizard

Select Source Type: Select one table to finish or select multiple tables and click Next to define their joins.

> Select Tables
 Define Joins

Schema: colegio_multidimensional

Available Tables:

Selected Tables:

- 'colegio_multidimensional', 'dim_fase'
- 'colegio_multidimensional', 'dim_imagen'
- 'colegio_multidimensional', 'dim_persona'
- 'colegio_multidimensional', 'dim_taller'
- 'colegio_multidimensional', 'dim_tiempo'
- 'colegio_multidimensional', 'h_registroTaller'

Fact Table: 'colegio_multidimensional', 'h_registroTaller'

< Back Next > Finish Cancel

Figura 0.90 Pentaho Server: Tablas de hechos y dimensiones

- **BA Server:** /pentaho/server/biserver-ee/tomcat/lib/
- **Enterprise Console:** /pentaho/server/enterprise-console/jdbc/
- **Data Integration Server:** /pentaho/server/data-integration-server/tomcat/webapps/pentaho-di/WEB-INF/lib/
- **Data Integration client:** /pentaho/design-tools/data-integration/libext/JDBC/
- **Report Designer:** /pentaho/design-tools/report-designer/lib/jdbc/
- **Schema Workbench:** /pentaho/design-tools/schema-workbench/drivers/
- **Aggregation Designer:** /pentaho/design-tools/agg-designer/drivers/
- **Metadata Editor:** /pentaho/design-tools/metadata-editor/libext/JDBC/

Figura 0.93 Rutas para los drives de base de datos

En la Figura 0.94 se muestra la ubicación de las librerías para pentaho bi.



Figura 0.94 Ubicación de drivers de mysql

Una vez que se han establecido todas las relaciones permite finalizar la construcción del recurso de datos y se pasa a personalizar el recurso de datos.

Data Source Created

The data source was created successfully.

Note

To enable reporting and analysis, a default metadata model was automatically created for this data source. You can edit the data source at any time by going to File > Manage > Data Sources.

- Keep default model
- Customize model now

OK

Figura 0.95 Confirmación de la creación de un recurso de datos

En la Figura 0.96 se presenta el editor del modelo en caso de que se deseen realizar modificaciones o verificar campos de hechos y dimensiones.

Data Source Model Editor

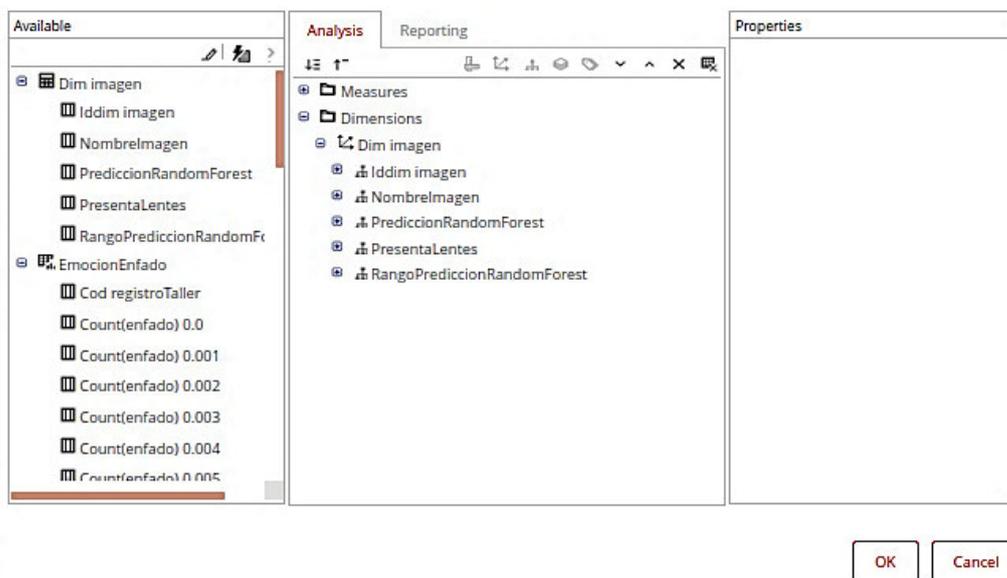


Figura 0.96 Editor de modelo de fuente de datos

Inicio Saiku

Luego de acceder a Pentaho Bi-Server, ir a Home, Create New, Saiku Analytics como se muestra en la Figura 0.97.

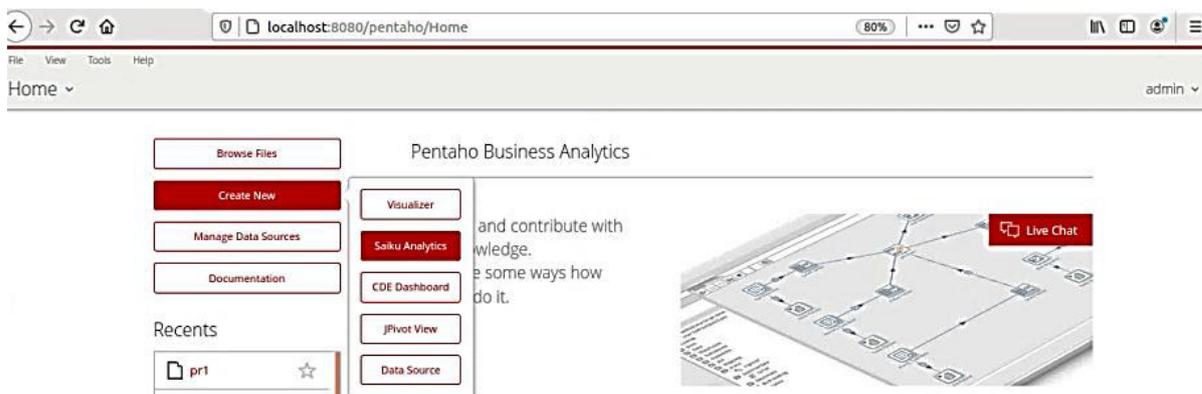


Figura 0.97 Inicio de Saiku Analytics

Ir a Create a new Query. Figura 0.98.

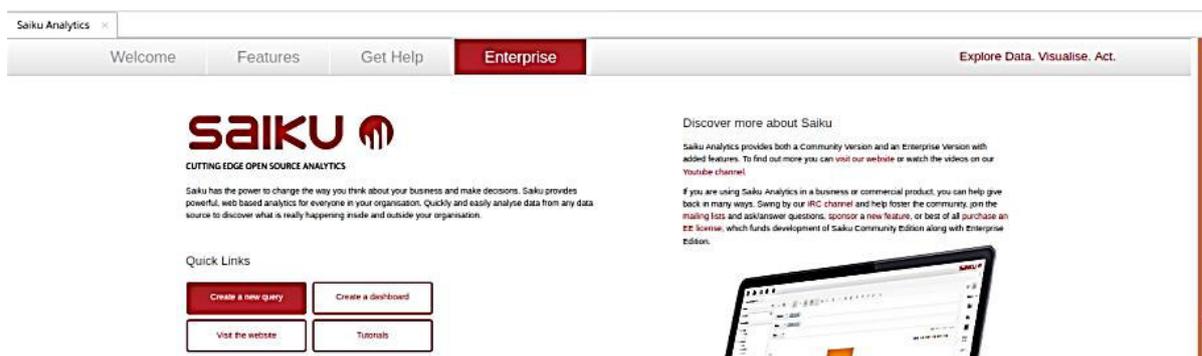


Figura 0.98 Crear una nueva consulta en Saiku

En la sección Cubes, utilizar el cubo anteriormente creado, Figura 0.99.

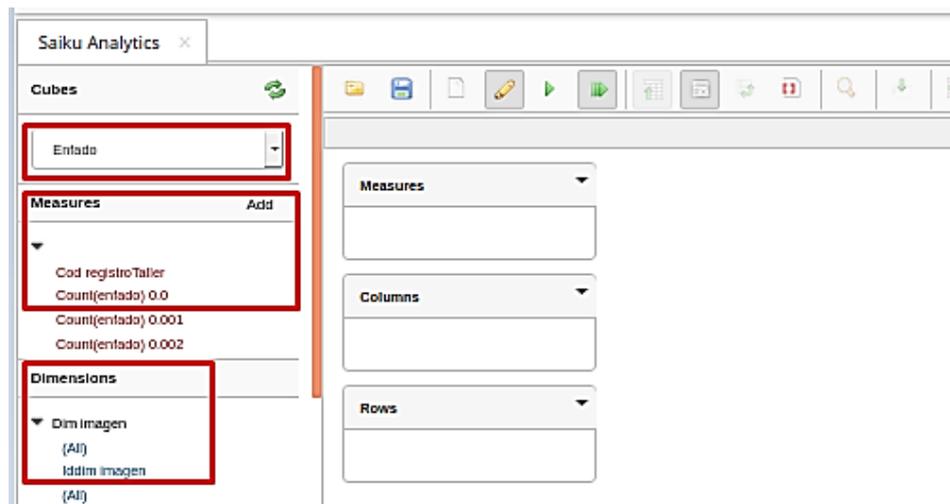


Figura 0.99 Entorno Saiku Analytics

En la Figura 0.100 se muestra como se pueden elegir las medidas (Measures) y dimensiones (Dimensions) deseadas.

Fecha tie - Fecha tie		2020-01-13					
Nombre	Felicidad	Asco	Entado	Neutral	Desprecio	Temor	Tristeza
Ordenar para reciclar	99.84	0.05	1.07	239.12	1.11	0.07	9.23

Figura 0.100 Saiku: Selección de medidas y dimensiones

En la Figura 0.101 se puede verificar de forma gráfica la información del cubo seleccionado.

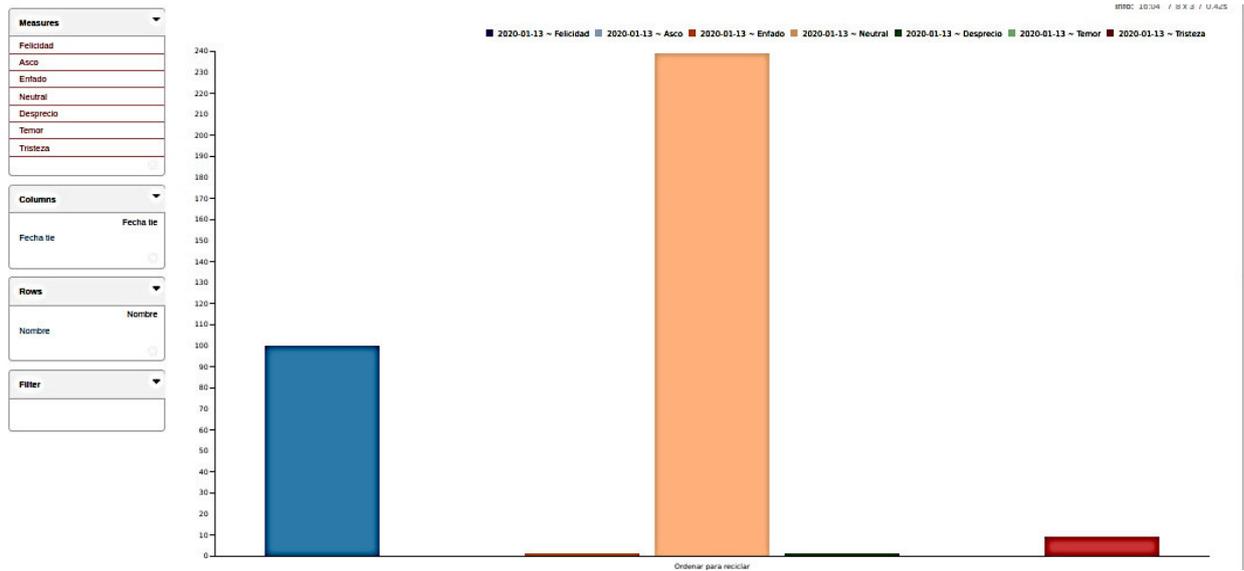


Figura 0.101 Saiku: Visualización

Anexo XIII Tutorial DATAFOR

Visualizer de Datafor permite crear páginas de análisis y visualización de datos, para lo cual se requieren los siguientes pasos:

1. Crear una conexión con una base de datos u otra plataforma de datos.
2. Crear un modelo de análisis utilizando el "Asistente de fuente de datos" o Pentaho Schema Workbench.
3. Crear la página del visualizador y análisis del Visualizador.

Inicio del Visualizador

- Iniciar sesión en la Consola de usuario de Pentaho.
- Seleccionar el botón "Nuevo" en la página principal o seleccionar el submenú "Nuevo" en el menú "Archivo" y seleccionar el elemento del menú "Visualizador" para abrir el diseñador del visualizador, Figura 0.102.

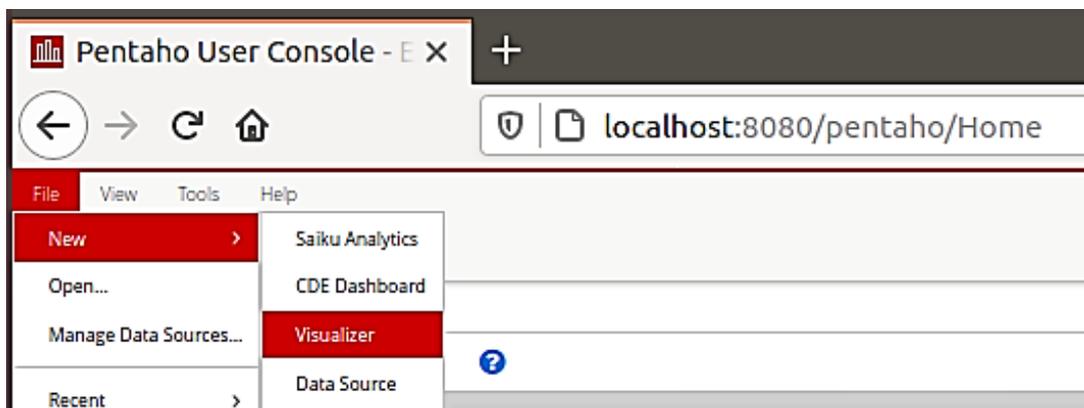


Figura 0.102 Inicio de Visualizador - Datafor

En la Figura 0.103 se muestra el diseñador en pantalla completa.

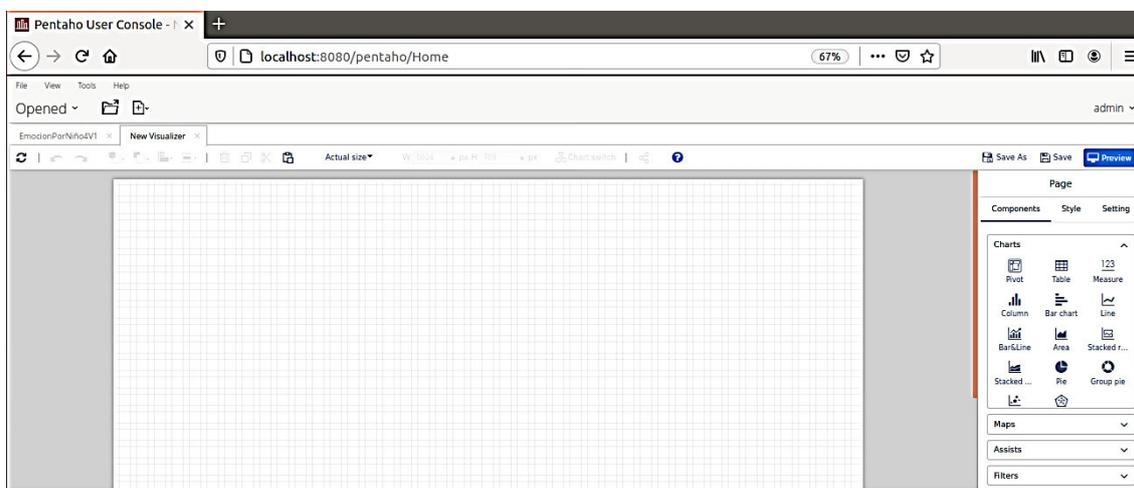


Figura 0.103 Entorno Visualizador - Datafor

Barra de herramientas

En la Figura 0.104 se muestran las opciones de la barra de herramientas del visualizador.

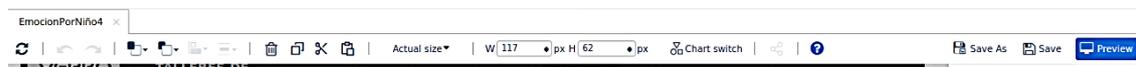


Figura 0.104 Visualizador - Datafor: Barra de herramientas

Presenta las siguientes opciones:

- El título del archivo: Para visualizar el nombre del archivo de visualización de datos.
- Refrescar: Para actualizar los cambios que se hayan realizado en el dashboard
- Botones de hacer y Deshacer: Permiten rehacer o deshacer el trabajo realizado en el dashboard.
- Alineación: Estos botones permiten alinear el contenido horizontal o verticalmente según se desee.
- Borrar: Permite borrar los componentes seleccionados.
- Copiar: Permite copiar los componentes seleccionados.
- Cortar: Permite cortar los componentes seleccionados.
- Tamaño Actual: Contiene opciones para la visualización de la página, Figura 0.105.

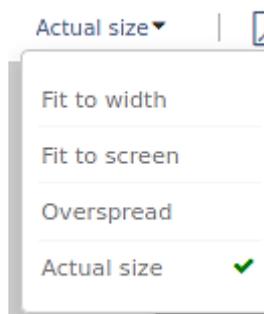


Figura 0.105 Visualizador - Datafor: Visualización de Página

Presenta las siguientes opciones:

- Ajustar al ancho: Para extender el dashboard al ancho de la ventana.
- Ajustar a la pantalla: Para extender el dashboard al ancho de la pantalla.
- Cubrir: Para extender el dashboard a todo el espacio disponible.
- Tamaño actual: Elige un tamaño estándar del dashboard.
- PDF: Permite exportar el trabajo realizado a un archivo PDF
- Preview/Edit: Permite ir a una vista previa del trabajo realizado y regresar a una vista de edición en caso de que se desee realizar cambios
- Save y Save as: Para guardar el archivo con un nombre específico.

Página

Presenta las siguientes opciones:

- Componentes
- Estilo
- Configuraciones

Componentes

Cabe mencionar que para cada componente en la versión de pago se despliegan más opciones, los disponibles en la presente versión se muestran en la Figura 0.106.

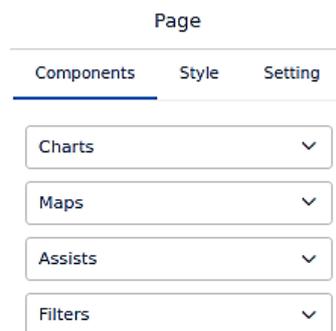


Figura 0.106 Visualizador - Datafor: Componentes de página

Los componentes que se presentan son los siguientes:

- Gráficos
- Mapas
- Asistentes
- Filtros

Gráficos

En la Figura 0.107 se muestran los diferentes tipos de gráficos disponibles.

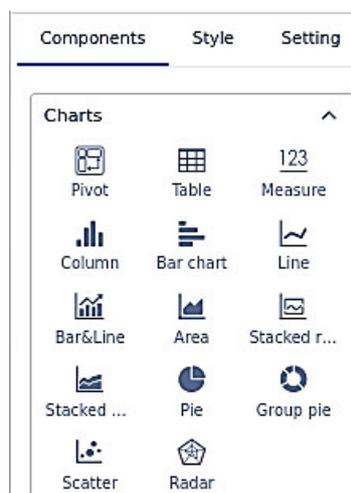


Figura 0.107 Visualizador - Datafor: Componentes Gráficos

Se dispone de los siguientes: Tablas Pivotes Medidas, Gráficos de Columnas, Grafico de Barras, Líneas, Barras y Líneas, Áreas. Pilas Pastel, Pastel de grupos, Dispersión y Radar.

Mapas

En la Figura 0.108 se muestra componentes para la construcción de mapas.

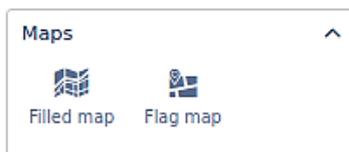


Figura 0.108 Visualizador - Datafor: Componentes de mapas

Se dispone de los siguientes: Mapa Lleno y Mapa de Bandera.

Asistentes

En la Figura 0.109 se muestra componentes que ayudan como asistentes en la visualización.

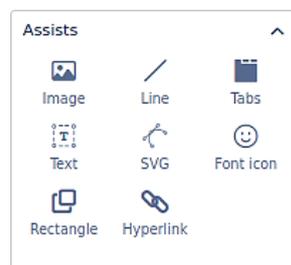


Figura 0.109 Visualizador - Datafor: Componentes para asistencia

Se dispone de los siguientes: Imágenes, Líneas, Pestañas, Texto, Vectoriales, Fuente de Iconos, Rectángulos e Hipervínculo.

Filtros

En la Figura 0.110 se puede ver los filtros que contiene el visualizador para modificar el estilo de la medida.



Figura 0.110 Visualizador - Datafor: Componentes para filtros

Se dispone de los siguientes: Listas, Selector de tiempo, Botones, Combo box Múltiple.

Opciones de Datos de la Medida

En la Figura 0.111 se muestran los datos de la medida, contiene opciones para modificar los datos de la medida.

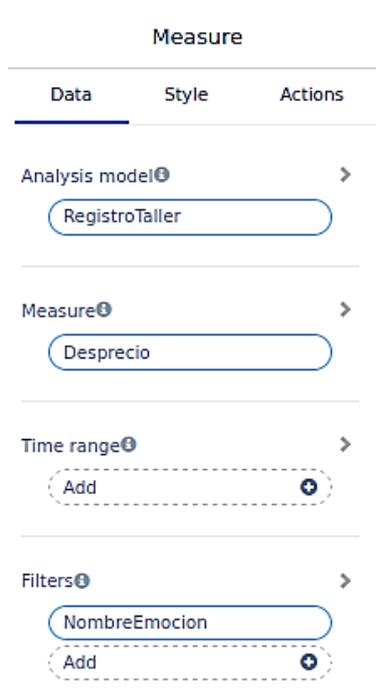


Figura 0.111 Visualizador - Datafor: Datos de las medidas

Esta sección presenta las siguientes opciones:

- Análisis del Modelo: Para elegir el cubo a utilizar para el análisis.
- Medida: Permite elegir el hecho de la tabla de hechos que se desea medir.
- Rango de tiempo: Para definir el rango de tiempo de los datos que son analizados.
- Filtros: Para indicar las condiciones que cumplen los atributos que se presentan en el dashboard.

Opciones del Estilo de la Medida

En la Figura 0.112 se muestran las preferencias del estilo de la medida, contiene opciones para modificar el estilo de la medida.

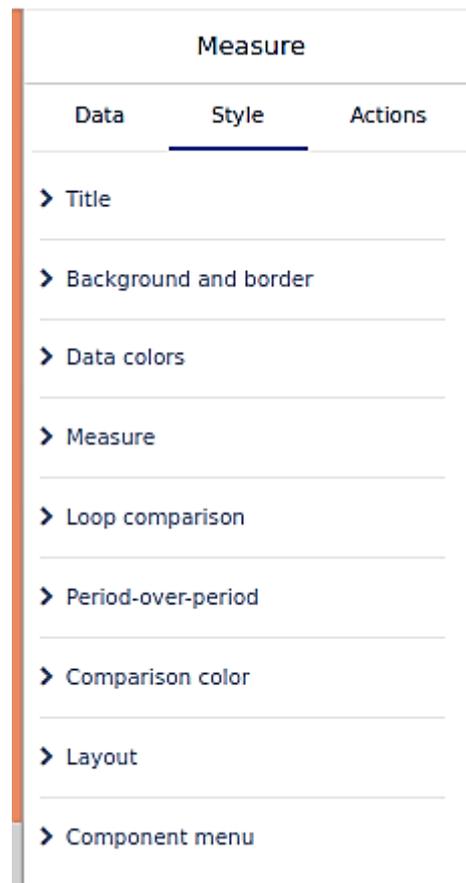


Figura 0.112 Visualizador - Datafor: Opciones para el estilo de las medidas

Cambio título. La Figura 0.113 muestra las preferencias del título de la medida. Cambiar el título , mostrarlo, alineación, Fuente de letra



Figura 0.113 Visualizador - Datafor: Título del cubo de datos

Fondo y bordes. La Figura 0.114 muestra las preferencias del fondo y bordes del panel. Cambio de color de fondo, estilo de bordes y si se muestran en el dashboard.



Figura 0.114 Visualizador - Datafor: Fondo y borde del componente

Color de los datos. La Figura 0.115 muestra las preferencias del color de los datos. Para definir si se emplean los colores por defecto.

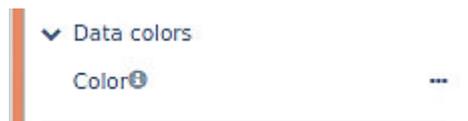


Figura 0.115 Visualizador - Datafor: Color de los datos

Medida. La Figura 0.116 muestra las preferencias de formato de la medida. Define la fuente y alineación del contenido de la medida.



Figura 0.116 Fuente del componente de medida

Bucle de comparación. La Figura 0.117 muestra las preferencias de los períodos de tiempo, unidades de comparación, etc. Permiten definir la fuente, alineación y formato de los campos.

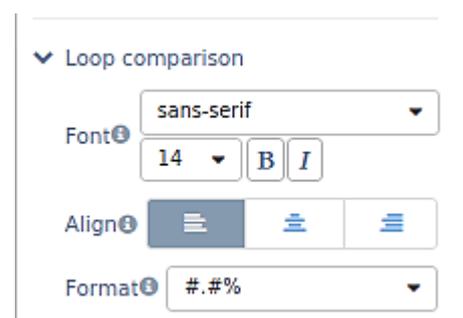


Figura 0.117 Visualizador - Datafor: Formato para datos de comparación

Comparación de Color. La Figura 0.118 muestra las preferencias de comparación de color. Permite comparar el valor utilizado gráficamente para resultados positivos o negativos.



Figura 0.118 Visualizador - Datafor: Colores de comparación

Diseño. La Figura 0.119 muestra las preferencias de diseño. Compara el valor de diseño como marco por Defecto, Derecha o de Fondo.



Figura 0.119 Visualizador - Datafor: Marco de diseño

Componentes de Menú. La Figura 0.120 muestra las preferencias de los componentes del menú. Permite habilitarlos y elegir el color de la barra de herramientas.

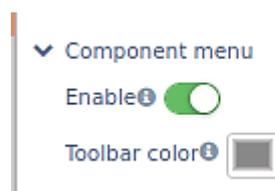


Figura 0.120 Visualizador - Datafor: Menú de componentes