

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA DE SISTEMAS

**SISTEMA DE INTELIGENCIA DE NEGOCIOS DEL
MONITOREO DE LA EXPERIENCIA DE USUARIO EN UN
TALLER DE ROBÓTICA EDUCATIVA**

**PROYECTO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN
SISTEMAS DE INFORMACIÓN**

RICARDO ANDRÉS ORTIZ VILLACRÉS

ricardo.ortiz02@epn.edu.ec

DIRECTOR: DR.JULIÁN ANDRES GALINDO LOSADA

julian.galindo@epn.edu.ec

CODIRECTOR: MARCO OSWALDO SANTORUM GAIBOR

marco.santorum@epn.edu.ec

Quito, septiembre 2021

DECLARACIÓN

Yo, Ricardo Andrés Ortiz Villacrés 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.

A través de la presente declaración cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

A handwritten signature in black ink, appearing to read 'Ricardo Ortiz Villacrés', is written over a horizontal line. The signature is stylized and somewhat cursive.

Ricardo Andrés Ortiz Villacrés

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Ricardo Andrés Ortiz Villacrés, bajo mi supervisión.



Dr. Julián Andrés Galindo Losada

DIRECTOR DE PROYECTO

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Ricardo Andrés Ortiz Villacrés, bajo mi supervisión.

A handwritten signature in blue ink, consisting of several overlapping loops and a central scribble, positioned above a horizontal line.

Dr. Marco Oswaldo Santorum

CODIRECTOR DE PROYECTO

AGRADECIMIENTO

Quiero agradecer principalmente a mis padres, que sin importar las circunstancias me apoyaron en todo lo que necesité y no dejaron que me falte nada, me dieron un gran ejemplo de vida y me enseñaron a perseverar y trabajar duro para lograr mis sueños.

Agradezco a mis hermanos Milagros, Alejandro y Nadia por siempre estar ahí dándome su apoyo y permanente compañía y alegrándome cada día para salir adelante.

Doy gracias a mi novia Sofía que me apoyó desde el primer momento me dio la fuerza y confianza en mi mismo para llegar a dónde estoy.

Estoy profundamente agradecido con el Dr. Julián Galindo por todo el apoyo, conocimiento y guía que me ha brindado durante la elaboración de este proyecto.

Finalmente, agradezco a la Escuela Politécnica Nacional por darme las herramientas, el conocimiento y los valores con los que podré ser un gran profesional para este país.

DEDICATORIA

Dedico este trabajo a mi familia y a mi novia, a quienes les debo el haber llegado al lugar donde estoy, me llenaron de sueños y esperanzas y me ayudaron a crecer como persona y profesional.

INDICE DE CONTENIDO

Contenido

1. CAPÍTULO 1. MARCO DE REFERENCIA	11
1.1 DESCRIPCIÓN DEL PROBLEMA	11
1.1.1 HISTORIA DE LA INSTITUCIÓN	11
1.1.2 DESCRIPCIÓN CASO DE ESTUDIO.....	12
1.1.3 DESCRIPCIÓN DE LOS TALLERES	14
1.2 DESCRIPCIÓN DE LA SITUACIÓN ACTUAL DEL PROBLEMA	17
1.3 OBJETIVOS.....	18
1.3.1 Objetivo General.....	18
1.3.2 Objetivos Específicos	18
1.4 JUSTIFICACIÓN TEÓRICA.....	18
1.5 JUSTIFICACIÓN DE LA METODOLOGÍA.....	23
1.5.1 ANALIZAR METODOLOGÍAS DE INTELIGENCIA DE NEGOCIOS	24
1.5.2 SELECCIÓN DE METODOLOGÍA.....	26
1.6 JUSTIFICACIÓN PRÁCTICA.....	31
1.7. SELECCIÓN DE HERRAMIENTAS.....	35
1.8. ARQUITECTURA GENERAL DEL SISTEMA.....	38
2. CAPÍTULO 2. DESARROLLO DEL SISTEMA.....	39
2.1 ANÁLISIS DE REQUERIMIENTOS	39
2.1.1 PLANEAR EL ANÁLISIS DE REQUERIMIENTOS	39
a) Caso de estudio	39
b) Análisis de información recolectada.....	40
c) Identificar preguntas necesarias.....	40
d) Identificar indicadores óptimos y perspectivas que se requieren.	42
2.1.2 REALIZAR EL ANÁLISIS DE REQUERIMIENTOS.....	42
a) Elaboración de encuestas profesores y autoridades del plantel.....	42
b) Validar indicadores y perspectivas para diseñar el sistema.	42
c) Realizar modelo conceptual	43
2.1.3 ANALIZAR DATOS OBTENIDOS Y VERIFICAR VIABILIDAD	43
2.2 ANÁLISIS DE OLTP	43
2.2.1 DISEÑO EL ANÁLISIS DE OLTP.....	44
2.2.2 IMPLEMENTACIÓN ANÁLISIS DE OLTP	44

2.2.3 CONFORMAR INDICADORES	45
2.2.4 ESTABLECER CORRESPONDENCIAS Y VERIFICAR NIVEL DE GRANULARIDAD.....	46
2.3 MODELO LÓGICO DEL DW	53
2.3.1 IMPLEMENTACIÓN DEL MODELO LÓGICO DEL DW	54
2.3.2 CREACIÓN DE TABLAS DE DIMENSIONES	54
2.3.3 CREACIÓN DE TABLA DE HECHOS.....	55
2.3.4 CREACIÓN DE UNIONES Y VERIFICACIÓN DEL MODELO.....	56
2.4 INTEGRACIÓN DE DATOS Y VISUALIZACIÓN	57
2.4.1 DISEÑO DE INTEGRACIÓN DE DATOS.....	57
2.4.2 IMPLEMENTACIÓN DE LA INTEGRACIÓN DE DATOS.....	57
a) Calidad de datos.....	57
b) Generación de Cluster en los datos	59
c) Carga de datos en el sistema	62
2.4.3 IMPLEMENTACIÓN Y VISUALIZACIÓN DE DATOS	67
a) Análisis de datos obtenidos.....	67
b) Interpretación de datos obtenidos, indicadores y perspectivas	71
3. CONCLUSIONES Y RECOMENDACIONES.....	73
3.1 CONCLUSIONES.....	73
3.2 RECOMENDACIONES	73
4. BIBLIOGRAFÍA.....	75
5. ANEXOS.....	78
I. Glosario	78
II. Modelo del JSON generado con el que se trabajó inicialmente.....	80
III. Ejemplos de fotografías	82
IV. Encuesta en Google Forms para Director y Maestros	84

ÍNDICE DE TABLAS

Tabla 1	- Comparación de metodologías de BI	25
Tabla 2	- Software que se utilizará.....	37
Tabla 3	- Tabla de preguntas para los profesores	41
Tabla 4	- Indicadores óptimos que se dedujeron de las preguntas.....	42
Tabla 5	- Clústers exportados de Rapid Miner	61
Tabla 6	- Tabla de valores de Centroides de los Clúster	62
Tabla 7	- Tabla del primer análisis del cubo	67
Tabla 8	- Tabla del segundo análisis del cubo	68
Tabla 9	- Tabla del tercer análisis del cubo.....	69
Tabla 10	- Tabla del cuarto análisis del cubo.....	69
Tabla 11	- Tabla del quinto análisis del cubo.....	70

ÍNDICE DE ILUSTRACIONES

Ilustración 1	- Fases de WeDo 2.0	16
Ilustración 2	- Fases del Datawareing.....	19
Ilustración 3	- Características de BI	21
Ilustración 4	- Proceso de Minería de datos	22
Ilustración 5	- Clasificación de procesos de Minería de Datos	23
Ilustración 6	- Metodología Kimball.....	25
Ilustración 7	- Metodología Crisp DM.....	26
Ilustración 8	- Metodología Hefesto	27
Ilustración 9	- Carga de datos OLTP	29
Ilustración 10	- Características Data warehouse	30
Ilustración 11	- Modelo Estrella	33
Ilustración 12	- Ejemplo de tablas de dimensiones	34
Ilustración 13	- Ejemplo de tabla de hechos	35
Ilustración 14	- Ejemplo de un DWH realizado	35
Ilustración 15	- Arquitectura del sistema de inteligencia de negocios	38
Ilustración 16	- Modelo Conceptual desarrollado	43
Ilustración 17	- Modelo de base de datos	46
Ilustración 18	- Correspondencias entre la base y el modelo	47
Ilustración 19	- Tabla Dimensión Fase.....	54
Ilustración 20	- Tabla Dimensión Taller	55
Ilustración 21	- Tabla Dimensión Estudiante.....	55
Ilustración 22	- Tabla Dimensión Tiempo	55
Ilustración 23	- Tabla de Hechos experiencia del usuario	56
Ilustración 24	- Modelo del Datawarehouse completo	56
Ilustración 25	- Diseño de la operación de calidad de datos	57
Ilustración 26	- Datos generados sobre la calidad de los datos	58
Ilustración 27	- Histograma de estabilidad de datos	58
Ilustración 28	- Histograma de text-ness	59
Ilustración 29	- División por Clústers.....	60
Ilustración 30	- Mapa de calor de Clústers	61
Ilustración 31	- Árbol de decisiones de Clúster	62
Ilustración 32	- Extracción de datos ETL de la dimensión	62
Ilustración 33	- Extracción de datos ETL de la dimensión Estudiante	63
Ilustración 34	- Extracción de datos ETL de la dimensión Fase	63
Ilustración 35	- Extracción de datos ETL de la dimensión Tiempo.....	63
Ilustración 36	- Extracción de datos ETL de la tabla de hechos	64
Ilustración 37	- Esquema de un cubo multidimensional	65
Ilustración 38	- Dimensiones del cubo.....	65
Ilustración 39	- Jerarquías dentro de las dimensión Estudiante.....	65
Ilustración 40	- Jerarquías dentro de las dimensión Fase	66
Ilustración 41	- Jerarquías dentro de las dimensión Taller	66
Ilustración 42	- Jerarquías dentro de las dimensión Tiempo	66
Ilustración 43	- Medidas establecidas dentro del cubo.....	67
Ilustración 44	- Histograma del primer análisis del cubo	68
Ilustración 45	- Histograma del segundo análisis del cubo.....	68
Ilustración 46	- Histograma del Tercer análisis del cubo.....	69
Ilustración 47	- Histograma del cuarto análisis del cubo.....	70
Ilustración 48	- Histograma del quinto análisis del cubo.....	71

1. CAPÍTULO 1. MARCO DE REFERENCIA

1.1 DESCRIPCIÓN DEL PROBLEMA

Se busca desarrollar un sistema de Inteligencia de Negocios utilizando un software libre o gratuito. Este sistema permitirá monitorear el aprendizaje basado en emociones de estudiantes que participan en talleres de Robótica Educativa. Adicionalmente, se debe realizar una debida planificación para satisfacer las necesidades de las autoridades de la Unidad Educativa.

1.1.1 HISTORIA DE LA INSTITUCIÓN

Este proyecto se realiza en la Unidad Educativa Cristiana Verbo la cual es una institución educativa particular, cristiana e inclusiva, cuyo principal objetivo es formar seres humanos con gran ética, una excelente educación y conciencia por el mundo que los rodea. La educación de los alumnos busca tener como base los principios y valores inculcados por la religión, las enseñanzas bíblicas y fomentando el pensamiento crítico de cada alumno.

De esta manera, la Unidad Educativa para estar a la vanguardia educativa, implementa nuevas herramientas para la enseñanza, entre ellas la realización de diferentes cursos de varios temas como el que se analizará en este caso, un taller de robótica.



1.1.2 DESCRIPCIÓN CASO DE ESTUDIO

Históricamente, el sistema educativo ecuatoriano ha mantenido un desarrollo desigual y diferenciado fruto de inequidades socioeconómicas y un ineficaz acceso a recursos informáticos y tecnológicos. Por otra parte, se puede observar que en Ecuador se reduce cada vez más los recursos públicos destinados a la educación, dificultando enormemente la labor docente de enseñanza e impidiendo que la educación en el país evolucione.

Al ser la falta de presupuesto el principal enemigo de la educación, se han tenido que ingeniar diferentes soluciones para compensar esto. Durante las últimas décadas, se ha ido mejorando el acceso a la tecnología y la velocidad del internet, pero no ha sido suficiente ya que no es solamente el problema de no tener los instrumentos, sino que no hay una adecuada capacitación a los educadores ni una adecuada integración la tecnología en la enseñanza[1].

De esta manera, los profesores han ido integrando a la educación diferentes metodologías y herramientas por cuenta propia, apoyándose en las TICs para realizar sus labores de enseñanza. Así, se despegan de los viejos métodos de enseñanza y abren camino a las mentes jóvenes a nuevas carreras que contengan ciencia, robótica, programación, etc. De hecho, han nacido nuevas tendencias como gamificación (aprendizaje con videojuegos), el aula invertida(mayor autonomía del estudiante y enseñanza a distancia.

Ya ha sido demostrado que los recursos multimedia (texto, audio, imagen, animación, vídeo y ejercicios interactivos) contribuyen a elevar la calidad y eficiencia del aprendizaje ya que motivan y despiertan el interés de los estudiantes por el aprendizaje y desarrolla habilidades cognitivas y tecnológicas únicas que les servirán en la vida profesional en las carreras actuales y futuras carreras que surjan[2].

En cuanto a la implementación de este tipo de enseñanza, se lo debe hacer de una manera gradual y controlada ya que, al no tener una capacitación

adecuada los docentes, estos nuevos tipos de enseñanza podrían abrumar y confundir a los estudiantes. Por ello, es ideal que se vaya monitoreando la implementación de estas nuevas metodologías y las TICs en la enseñanza.

Por esto, la educación necesita el apoyo de una disciplina que pueda registrar, procesar y analizar la información que se vaya monitoreando en los procesos de enseñanza. Así, se puede implementar la conocida como Inteligencia de Negocios o (BI) la cual es una disciplina que combina análisis de negocios, minería, visualización, herramientas e infraestructura de datos, además de prácticas recomendadas para ayudar a las empresas a tomar decisiones basadas en los datos[3].

Uno de los objetivos básicos de los sistemas de información es proporcionar a la organización la información necesaria para controlar sus actividades. Cuando un responsable tiene que tomar una decisión pide o busca la información necesaria, la cual permite reducir la incertidumbre. Sin embargo, aunque todos tengan acceso a estos datos, no todos los responsables recogen la misma información y tampoco la interpretan igual, depende de muchos factores, como pueden ser su experiencia, formación, disponibilidad, etc.

Del mismo modo, los responsables pueden necesitar recoger más o menos información dependiendo de su mayor o menor aversión al riesgo. De esta manera, al emplear BI, se consideran varios factores a la vez, se analiza de mejor manera la información y nos proporciona datos que se desconocían hasta el momento, ayudando a tomar decisiones de una manera mucho más eficaz y eficiente, corriendo con menos riesgo.

En la actualidad, el manejar grandes cantidades de datos y no conocer la manera de utilizarlos es un gran problema, a una autoridad se le puede presentar toda esta información y no podrá entenderlo. Sin el uso de BI se pasa demasiado tiempo analizando la causa del problema o entendiendo los requerimientos del cliente. De esta manera, con el tiempo se va evidenciando

que solo se sabrá si uno realmente comprende a las personas al manejar la información de manera adecuada y que entregue un valor real de lo que se necesita.

En vista de que los niños y jóvenes son actores económicos en el presente y sus decisiones influyen en el futuro la sociedad, es una prioridad preocuparse por su educación y mediante este proyecto se busca ayudar a los profesores a entender mejor lo que sienten sus alumnos al medir sus emociones e interpretar estos datos y que de esta manera puedan guiarlos de la mejor manera posible, mediante el uso de la información que se obtiene en las clases, presentada de una manera clara, para que los docentes sepan su significado y con su experiencia tomen la mejor decisión de cómo mejorar la educación para los niños[4].

Así, con toda esta información, los docentes sabrán como manejar estas clases prácticas y de ciencia para maximizar el aprendizaje y que los alumnos muestren interés por estas y en el futuro abrir más cursos de este tipo o expandir los que ya se tiene para que los alumnos desde temprana edad aprendan distintas destrezas que los guiarán a estudiar diferentes carreras técnicas o científicas.

1.1.3 DESCRIPCIÓN DE LOS TALLERES

A continuación, se detallará la información sobre el taller de robótica que se ha aplicado a los niños de la Unidad Educativa Verbo, para tener un mejor contexto cómo se obtuvo la información y cómo ayudará el análisis de esta información.

a) Descripción general

- Inicio del taller(20min)

En esta etapa se explicará a los estudiantes las actividades, se conforma los equipos, se les hará conversar sobre el tema y se les estimulará a que realicen el trabajo.

- Sesión 1(80min)
Se va guiando como empezar, se conversa sobre lo que creen que construirán, se enseña imágenes de exploradores a los estudiantes y se los invita a recrearlos con legos, pudiendo hacer sus creaciones con manuales o con las indicaciones del equipo.
- Sesión 2 (60min)
Se organiza a los estudiantes formando diferentes grupos y se entrega el kit WeDo, se incentiva que los estudiantes conversen sobre lo que se va a hacer, se los orienta y acompaña en la construcción y se implementan las partes movibles.
- Cierre (60min)
Se realizan exposiciones orales sobre lo que se hizo, el profesor realiza preguntas sobre la actividad y como se sienten los estudiantes respecto a todo lo que hicieron.

b) Descripción de la metodología del taller

La metodología usada en este caso fue la de LEGO Education WeDo 2.0, la cual es una solución que desarrolla prácticas científicas mediante un sistema de aprendizaje basado en robots. De esta manera, su software se caracteriza por ser amigable, inspirando a los estudiantes a descubrir el mundo en torno a ellos y adentrándolos a los conceptos básicos de robótica y programación.

Con la aplicación de WeDo 2.0 los alumnos pueden explorar, crear y compartir diferentes conocimientos, mientras aprenden a programar, analizar y construir mezclando la ciencia, ingeniería y tecnología.

Se busca que esta metodología se aplique principalmente en niños de más de 7 años, aunque esto puede ser diferente según el país. De este modo, existen diversos cursos y talleres para aplicar WeDo 2.0, pero en forma general se compone de 3 fases:

- **Fase Explorar.-** Es la fase de socialización con los participantes. Tiene una duración de 30 a 60 minutos, se socializan ideas generales de cómo funciona el taller. Los estudiantes conectan con una pregunta científica o un problema de ingeniería, establecen una línea de investigación y consideran las posibles soluciones. Los pasos de la fase Explorar son conectar y debatir.
- **Fase Crear.-** En esta etapa los participantes del taller proceden a construir y programar. Tiene una duración de 45 a 60 minutos. Los estudiantes construyen, programan y modifican un modelo LEGO®. Los proyectos pueden ser de tres tipos: investigar, diseñar soluciones y usar modelos. En función del tipo de proyecto, la fase Crear diferirá de un proyecto a otro. Los pasos de la fase Crear son construir, programar y modificar.
- **Fase Compartir.-** En esta etapa los participantes del taller presentan y comparten sus resultados a otros participantes. Tiene una duración de 45 minutos o más. Los estudiantes presentan y explican sus soluciones con sus modelos LEGO, así como el documento que han elaborado con sus hallazgos mediante la herramienta integrada de documentación. Los pasos de la fase Compartir son documentar y presentar[5].

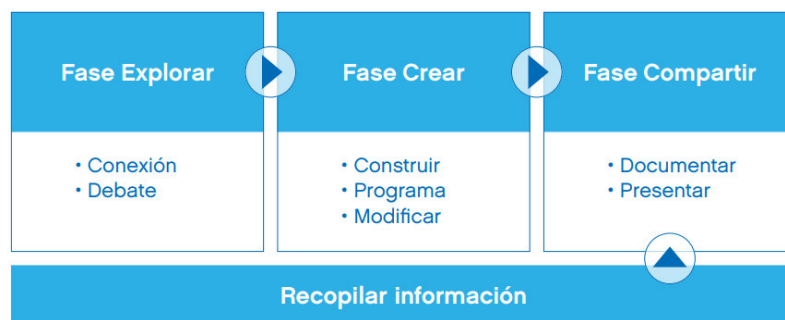


Ilustración 1.- Fases de WeDo 2.0

1.2 DESCRIPCIÓN DE LA SITUACIÓN ACTUAL DEL PROBLEMA

El trabajar con datos recopilados puede traer varios problemas y dificultades. Actualmente, utilizamos solo un 20% de nuestro tiempo en analizar datos para tomar decisiones. El 80% restante, se lo utiliza en búsqueda, limpieza y organización de grandes volúmenes de datos, siendo una estrategia de trabajo ineficiente. Estos problemas pueden resumirse en 5 tipos a los cuales se debe enfrentar y superar.

El primero es la recopilación de datos ya que no se sabe si fueron obtenidos de una manera correcta, si están actualizados o si son precisos. El segundo problema es la calidad de los datos, no se conoce la consistencia, completitud o validez de la información. El tercero es el análisis y visualización, el problema se da que muchas veces no se sabe cómo representar de manera adecuada los datos y no se puede analizar su significado. El cuarto problema es compartir, no se puede comunicar los resultados de una manera entendible o atractiva. Y finalmente, el quinto problema es que no se puede predecir el futuro, ya que los datos no están debidamente trabajados[6].

De esta manera, la mejor forma de evitar estos problemas es estableciendo un esquema ideal de obtención de datos y emplear una metodología de Inteligencia de Negocios (BI), para superar estas adversidades y ayudar a entregar información real y útil sobre el desarrollo de los talleres a los profesores. Además, con los datos cumpliendo ciertos parámetros, se puede analizarlos e interpretarlos mejor, para que los miembros de la Unidad Educativa puedan mejorar en la manera de impartir clases.

1.3 OBJETIVOS

1.3.1 Objetivo General

- Implementar un sistema de inteligencia de negocios del monitoreo de la experiencia de usuario en un taller de robótica educativa.

1.3.2 Objetivos Específicos

- Analizar los requerimientos de la unidad educativa para la implementación del sistema de inteligencia para diseñar un sistema adecuado para sus necesidades.
- Recolectar los datos obtenidos en los talleres para su posterior análisis e interpretación y mostrar resultados de manera clara para los profesores.
- Ofrecer un reporte a los profesores sobre los problemas encontrados en los resultados obtenidos del análisis de los talleres e identificar los puntos a mejorar.

1.4 JUSTIFICACIÓN TEÓRICA

Como se ha mencionado, la mejor manera de dar solución es haciendo uso de varios de los recursos que ofrece la Inteligencia de Negocios

Se propone usar esta metodología porque ofrece grandes beneficios a lo largo de un proyecto ya que los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender, además de adaptarse bien a un proyecto como este debido a que su aspecto fundamental lo constituyen los requerimientos de los usuarios.

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. Además, utiliza modelos conceptuales y modelos lógicos, los cuales son sencillos de interpretar y analizar y es independiente del software/hardware que se utilicen para su implementación, lo cual otorga una amplia gama de posibilidades para solucionar el problema.

Esta es una metodología fundamentada en una muy amplia comparación de metodologías existentes, experiencias de varios usuarios. Su gran ventaja es que está en continua evolución, y se han tenido en cuenta, como gran valor agregado, toda la retroalimentación que han aportado quienes han utilizado esta metodología en diversos países y con diversos fines.

Otro beneficio es que el DW puede adaptarse muy bien al desarrollo de software. Lo que se busca, es entregar una primera implementación que satisfaga una parte de las necesidades de los usuarios, para demostrar las ventajas del DW y motivarlos a seguir con el proyecto[7].

Dentro del DW se puede identificar las siguientes fases:



Ilustración 2.- Fases del Data Warehousing

- FASE 1: Dirigir y Planear. En esta fase inicial 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.
- FASE 2: Recolección de Información. 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.

- FASE 3: Procesamiento de Datos. 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.
- FASE 4: Análisis y Producción. Ahora, se procederá a trabajar sobre los datos extraídos e integrados, utilizando herramientas y técnicas propias de la tecnología BI, para crear inteligencia. Como resultado final de esta fase se obtendrán las respuestas a las preguntas, mediante la creación de reportes, indicadores de rendimiento, cuadros de mando, gráficos estadísticos, etc.
- FASE 5: Difusión. Finalmente, se les entregará a los usuarios que lo requieran las herramientas necesarias, que les permitirán explorar los datos de manera sencilla e intuitiva[7].

Beneficios

Entre los beneficios más importantes que BI proporciona a las organizaciones, vale la pena destacar los siguientes:

Reduce el tiempo mínimo que se requiere para recoger toda la información relevante de un tema en particular, ya que la misma se encontrará integrada en una fuente única de fácil acceso.

Automatiza la asimilación de la información, debido a que la extracción y carga de los datos necesarios se realizará a través de procesos predefinidos.

- Proporciona herramientas de análisis para establecer comparaciones y tomar decisiones.
- Cierra el círculo que hace pasar de la decisión a la acción.
- Permite a los usuarios no depender de reportes o informes programados, porque los mismos serán generados de manera dinámica.

- Posibilita la formulación y respuesta de preguntas que son claves para el desempeño de la organización.
- Permite acceder y analizar directamente los indicadores de éxito.
- Se pueden identificar cuáles son los factores que inciden en el buen o mal funcionamiento 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 usuarios podrán consultar y analizar los datos de manera sencilla e intuitiva.

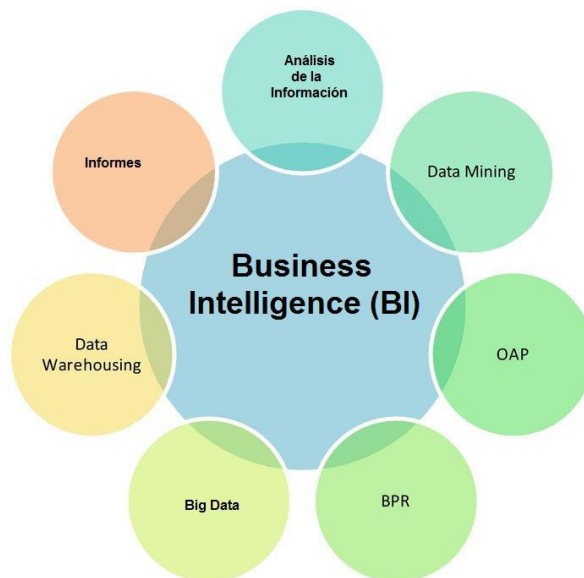


Ilustración 3.-Características de BI

Es necesario definir el tipo de datos con los que se va a trabajar, ya que se tienen 2 tipos de indicadores. Primero, tenemos del tipo estático que sirven solo para hacer consultas teniendo información histórica y por este motivo no pueden modificarse con el tiempo, simplemente utilizadas como un control o para explicar un comportamiento de una temática determinada en un período o lapso determinado. Por otro lado, tenemos los indicadores dinámicos los cuales utilizan datos del presente y del día a día, por lo que tienen que ser actualizados periódicamente, además de que es propenso a modificaciones frecuentes,

siendo un claro ejemplo de ello un Control de Stock, un Inventario de un negocio o cualquier otra DB similar[8].

El proceso que se va a llevar a cabo es una minería de datos, la cual se define como el proceso de hallar anomalías, patrones y correlaciones en grandes conjuntos de datos para predecir resultados. Esto se lo realiza utilizando diferentes técnicas y de esta manera se puede incrementar sus ingresos, recortar costos, mejorar sus relaciones con clientes, reducir riesgos y más[9].

Max Bramer en 2007 describió el proceso básico de minería de datos que consta de tres pasos: preparación de datos, análisis de datos y proceso de evaluación. El análisis de datos se divide en 2 tipos: descripción y predicción.



Ilustración 4.- Proceso de Minería de datos

Así, los modelos de construcción de clasificación que representan y distinguen clases o nociones para predicciones futuras con valores numéricos desconocidos o faltantes. Introducimos técnicas básicas para la clasificación de datos, como cómo construir clasificadores de árboles de decisión, clasificadores bayesianos, redes de creencias bayesianas y clasificadores basados en reglas[10].

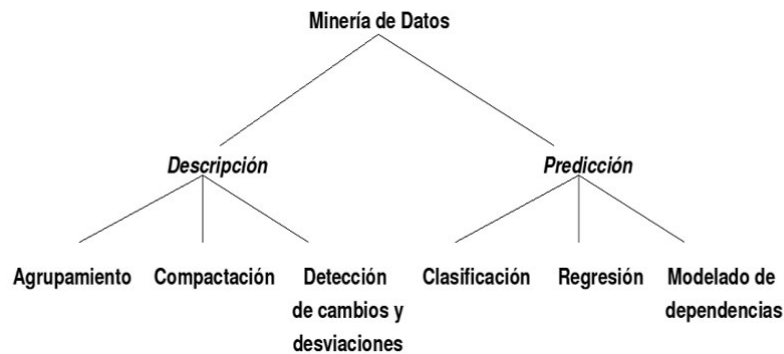


Ilustración 5 Clasificación de procesos de Minería de Datos

Dentro de la minería descriptiva, como un método no supervisado existe el Clustering o agrupamiento, el cual emplea el algoritmo k-means que es uno de los métodos de particionamiento más populares, ya que su uso es sumamente sencillo. De esta forma, se agrupa los datos en un número de clústeres definido por el usuario, k. Cada observación se asigna al grupo más cercano, cuyo centro está definido por el valor medio aritmético (promedio) de sus puntos, el centroide[11].

1.5 JUSTIFICACIÓN DE LA METODOLOGÍA

La experiencia en distintos proyectos hace evidente que para cada proyecto de BI y para cada organización debe seleccionarse la metodología que más posibilidades de éxito tenga.

L.T. Moss en un artículo divulgativo propone las siguientes características que debería cumplir una metodología para este tipo de sistemas decisionales:

- 1) Debe estar orientada al cambio y no a la consecución de un producto final.
- 2) La gestión del proyecto debe ser de forma global y transversal a toda empresa.
- 3) Hay que poder manejar múltiples subproyectos a la vez y en paralelo.
- 4) Es importante que se tenga en cuenta todas las tareas/procesos, sean o no críticos.
- 5) Debe basarse en la gestión de los caminos críticos del flujo de trabajo empresarial.

6) Hay que orientarlo las personas y a las relaciones entre ellas.

7) Y lo más importante, debe estar alineada con las necesidades de negocio de la organización.

Estas características se han sintetizado a partir de la experiencia práctica y la realidad. Por ello, se debe tener en cuenta estos aspectos para la selección de las metodologías más adecuadas para la gestión de un proyecto BI. También, es de vital importancia basarse en un proceso de análisis de factores críticos de éxito de todo proyecto de BI.

Por eso es necesario, en primer lugar, analizar cuáles son los factores de éxito de los proyectos de BI similares o también reunirse con las partes interesadas y por medio de técnicas de investigación como entrevistas o encuestas, conocer mejor el panorama y contexto del lugar, además de comprender las necesidades que se tiene.[12].

1.5.1 ANALIZAR METODOLOGÍAS DE INTELIGENCIA DE NEGOCIOS

CARACTERÍSTICA	CRISP–dm	KIMBALL	HEFESTO
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
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 ItemSets-DIS)	Modelo dimensional	ER, DIS, Modelo Dimensional
NATURALEZA DE LOS REQUERIMIENTOS	Estratégica	Táctica	Estratégica/Táctica

REQUERIMIENTOS DE INTEGRACION 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 datawarehouse es urgente	Puede iniciar con un Datamart y continuar al Datawarehouse
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.
ACCESIBILIDAD DEL USUARIO FINAL	Bajo[13]	Alto[14]	Alto[7]

Tabla 1 Comparación de metodologías de BI

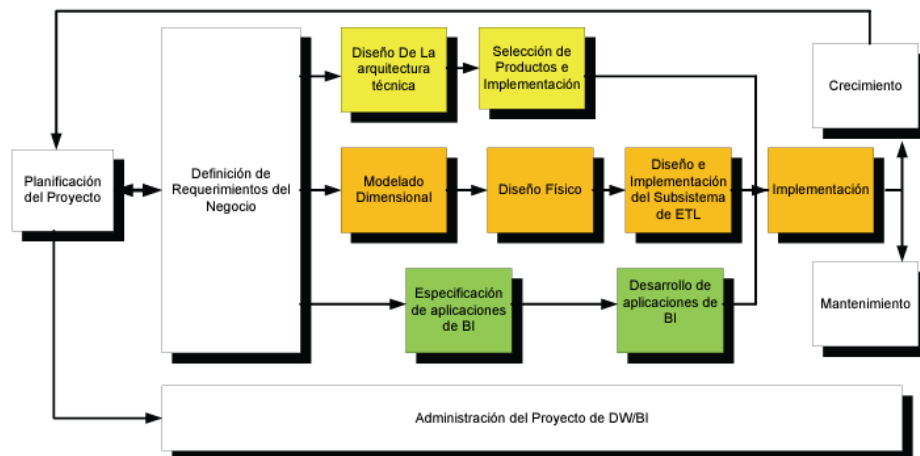


Ilustración 6 Metodología Kimball

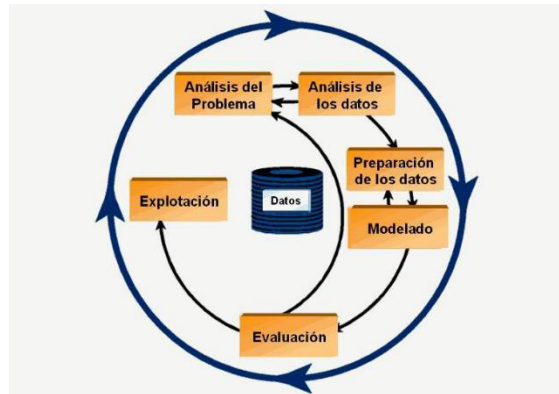


Ilustración 7 Metodología Crisp DM

1.5.2 SELECCIÓN DE METODOLOGÍA

HEFESTO

HEFESTO es una metodología que permite la construcción de cubos de datos de forma sencilla, ordenada e intuitiva.

La metodología HEFESTO posee las siguientes características:

- Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- El aspecto fundamental la constituyen los requerimientos de los usuarios.
- Reduce la resistencia al cambio, ya que involucra a los usuarios finales en cada etapa de la construcción, y además se entregan resultados inmediatos.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente del software/hardware que se utilicen para su implementación.
- Cuando se culmina con una fase, los resultados obtenidos se constituyen en la entrada de la fase siguiente.

Puede resumirse esta metodología de la siguiente manera



Ilustración 8 Metodología Hefesto

Actividades de la Metodología HEFESTO

HEFESTO plantea la realización de las siguientes actividades:

Paso 1) Análisis de requerimientos

Se obtiene e identifica las necesidades de información clave para las llegar a las metas y ejecutan las estrategias para facilitar la toma de decisiones.

- **Tarea 1. Identificar las preguntas del negocio.** - Se identifican los requerimientos de los usuarios a través de preguntas complejas sobre el negocio que expliciten los objetivos de su organización.
- **Tarea 2. Indicadores y Perspectivas.** - Se analizarán estas preguntas a fin de proceder a su descomposición para descubrir los Indicadores que se utilizarán y las perspectivas de análisis que intervendrán. Los Indicadores, son valores numéricos y representan lo que se desea analizar concretamente.
- **Tarea 3. Modelo Conceptual.** - Se confeccionará un modelo conceptual a partir de los indicadores y perspectivas, en donde se podrá visualizar el resultado obtenido en este primer paso.

Paso 2) Análisis de fuentes de datos

- **Tarea 1.-Hechos e Indicadores.** - Se analizan los orígenes de datos para determinar cómo serán calculados los Indicadores, definiendo los siguientes conceptos para cada uno de ellos: Hechos: hechos que la componen con su respectiva forma de cálculo. Función de agregación que se utilizará: por ejemplo, SUM, AVG, COUNT, etc.
- **Tarea 2.-Mapeo de orígenes de datos.** - Identificar las fuentes de datos para asegurar que existan los datos requeridos.
- **Tarea 3.-Granularidad.** - Seleccionar los campos que contendrá cada perspectiva y que servirán para analizar los indicadores.
- **Tarea 4. Modelo conceptual ampliado.** - Se amplía el modelo conceptual con la información obtenida en este paso y colocando debajo de cada perspectiva los campos seleccionados y debajo de cada Indicador su respectiva fórmula de cálculo.

Paso 3) Modelo Lógico del DW

- En el tercer paso se define el tipo de modelo lógico a usar (modelo multidimensional) y las tablas de dimensiones y tablas de hechos con sus respectivas relaciones.
- En este paso se incluye la definición de la tipología del modelo multidimensional de datos, definir las tablas de dimensiones y de hechos, o la creación del modelo lógico de los datos.

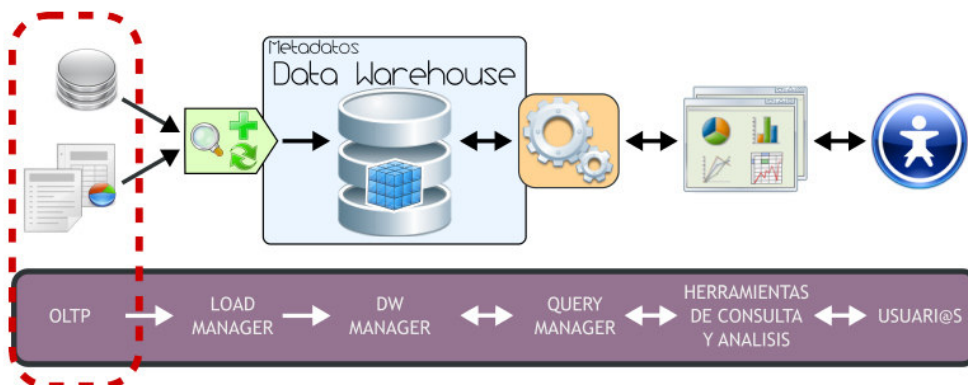


Ilustración 9.- Carga de datos OLTP

Construcción de Indicadores Académicos

Para la identificación de los indicadores académicos que se usarán en los tableros de control se aplicarán el paso 1 y paso 2 de la metodología HEFESTO, de la siguiente manera:

Paso 1. Análisis de requerimientos que incluye:

- **Tarea 1:** Identificar las preguntas del negocio
- **Tarea 2:** Identificar los Indicadores y Perspectivas a tomar en cuenta
- **Tarea 3:** Elaborar el Modelo Conceptual genérico

Paso 2. Análisis de fuentes de datos que incluye

- **Tarea 1:** Hechos e Indicadores
- **Tarea 2:** Mapeo de orígenes de datos
- **Tarea 3:** Granularidad
- **Tarea 4:** Modelo conceptual ampliado

Uso de Datawarehousing

Se debe utilizar un Datawarehouse(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.

Este proceso es el encargado de extraer, transformar, consolidar, integrar y centralizar los datos que una organización que se genera en todos los ámbitos de su actividad diaria (compras, ventas, producción, etc.) y/o información externa relacionada. Así, se permite el acceso y exploración de la información requerida, a través de una amplia gama de posibilidades de análisis multivariantes, con el objetivo final de dar soporte al proceso de toma de decisiones estratégico y táctico.

También, el DWH posibilita la extracción de datos de sistemas operacionales y fuentes externas, permite la integración y homogeneización de los datos de toda la empresa, provee información que ha sido transformada y resumida, para que ayude en el proceso de toma de decisiones estratégicas y tácticas. Además, convertirá entonces los datos operacionales de la empresa en una herramienta competitiva, debido a que pondrá a disposición de los usuarios indicados la información pertinente, correcta e integrada, en el momento que se necesita.



Ilustración 10.- Características Data warehouse

Es importante recordar, que para cumplir los objetivos del DWH, es necesario

que la información que se extrae ,transforma y consolida, sea almacenada de manera centralizada en una base de datos con estructura multidimensional denominada DataWarehouse (DW).

Una de las definiciones más famosas sobre DW, es la del conocido como padre del DWH William Harvey Inmon, quien define: “Un Data Warehouse es una colección de datos orientada al negocio, integrada, variante en el tiempo y no volátil para el soporte del proceso de toma de decisiones de la empresa”.

1.6 JUSTIFICACIÓN PRÁCTICA

La innovación en este estudio también se apoya en la evaluación de la experiencia del usuario con el uso de las emociones como variables para medir el impacto en la persona. Por ejemplo, se puede medir el impacto del uso de una interfaz web en la adaptación de un estudiante en una clase. Este método puede ser aplicado en escuelas y colegios públicos y privados de Quito, los cuales están interesados en promover las capacidades relacionadas con las habilidades digitales en robótica educativa a sus alumnos.

En vista de que los niños y jóvenes son actores económicos en el presente y sus decisiones influyen en el futuro de sus sociedades, mediante este proyecto se busca concientizar y promover a la vez, la responsabilidad social de los profesores con los niños y promover el gusto de los niños por las carreras técnicas y la ciencia en general. Además, como requerimiento pedido por la unidad educativa particular verbo norte, se necesita implementar clases de robótica educativa y el análisis de la toma de decisiones respecto a de su impacto.

Se tiene distintas razones para hacer uso de BI y son numerosas las aplicaciones que se le va a dar:

1. La administración de la unidad educativa y los profesores no pueden hacer mucho con la información obtenida diario de los niños. Así, con la ayuda de BI todos estos datos pueden ser recolectados e interpretados constantemente para ayudar al desarrollo infantil e identificar problemas que surjan en la clase.
2. Se pueden evitar problemas a futuro para la administración y los profesores al brindar al ayuda oportuna y necesaria a los niños.
3. Se facilitará la creación de reportes y se ahorrará tiempo en el análisis e interpretación de estos, además de buscar soluciones de manera más eficiente.
4. Los cambios en los reportes son mucho más fáciles, lo que antes tomaba semanas, meses o años, con BI puede hacerse en cuestión de minutos, siendo muy adaptativo.
5. Es de las mejores maneras de obtener resultados aproximados a la realidad de manera rápida, eficaz y eficiente. Los análisis que podrían tomar una gran cantidad de tiempo, se los puede obtener en segundos, incluso analizados.

La búsqueda de software de BI y análisis de datos puede ser un proceso desalentador y costoso, que requiere un gran tiempo investigación. Las herramientas de BI empresarial más populares a menudo proporcionan más de lo necesario para las organizaciones no empresariales o para proyectos más sencillos, con funcionalidades avanzadas usadas solo por personas conocedoras de estos temas a profundidad. Afortunadamente, hay una serie de herramientas de inteligencia empresarial gratuitas y de código abierto. Algunas de estas soluciones son ofrecidas por proveedores que buscan eventualmente venderle su producto empresarial, y otras son mantenidas y operadas por una comunidad de desarrolladores que buscan democratizar el BI.

De esta manera, se logrará cumplir con los objetivos planteados y los requerimientos de los usuarios sin gastar grandes cantidades de dinero y pudiendo tener ayuda y retroalimentación de la comunidad que maneja y actualiza el software usado.[15]

1.6.1 Estructura Necesaria

a) Base de datos multidimensional

Para este sistema, es necesaria una base de datos multidimensional la cual es una base de datos en donde la información se almacena en forma multidimensional, es decir, a través de tablas de hechos y tablas de dimensiones. Proveen una estructura que permite, a través de la creación y consulta a una estructura de datos determinada, tener acceso flexible a los datos, para explorar y analizar sus relaciones, y consiguientes resultados. Las bases de datos multidimensionales implican tres variantes posibles de modelamiento, que permiten realizar consultas de soporte de decisión:

- Esquema en estrella (Star Scheme).
- Esquema copo de nieve (Snowflake Scheme).
- Esquema constelación o copo de estrellas (Starflake Scheme).

El diseño que mejor se adapta a las necesidades del sistema es el esquema de estrella, que es un tipo de esquema de base de datos relacional que consta de una sola tabla de hechos central rodeada de tablas de dimensiones, ideal para tener como tabla de hechos la experiencia de los usuarios (medición de sentimientos) y los demás parámetros medidos como dimensiones.



Ilustración 11 Modelo Estrella

b) 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.

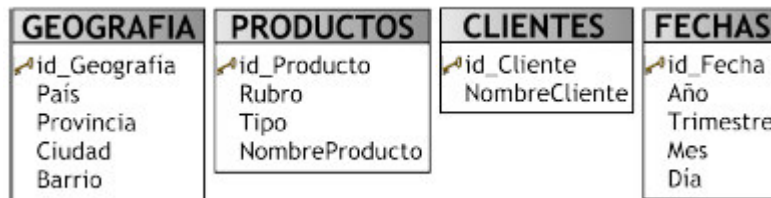


Ilustración 12.- Ejemplo de tablas de dimensiones

Más detalladamente, cada tabla de dimensión podrá contener los siguientes campos:

- Clave principal o identificador único.
- Clave foráneas.
- Datos de referencia primarios: datos que identifican la dimensión. Por ejemplo: nombre del cliente.
- Datos de referencia secundarios: datos que complementan la descripción de la dimensión. Por ejemplo: e-mail del cliente, fax del cliente, etc.

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 que se este analizando

c) Tabla 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 datos instantáneos en el tiempo, que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de dimensiones.

Para ejemplificar este nuevo concepto de hechos, se enumerarán algunos que son muy típicos y fáciles de comprender:

- $\text{ImporteTotal} = \text{precioProducto} * \text{cantidadVendida}$
- $\text{Rentabilidad} = \text{utilidad} / \text{PN}$
- $\text{CantidadVentas} = \text{cantidad}$

- PromedioGeneral = AVG(notasFinales)



Ilustración 13.- Ejemplo de tabla de hechos

De esta manera, se puede diseñar el DWH considerando las necesidades del negocio y respondiendo las preguntas planteadas[7].

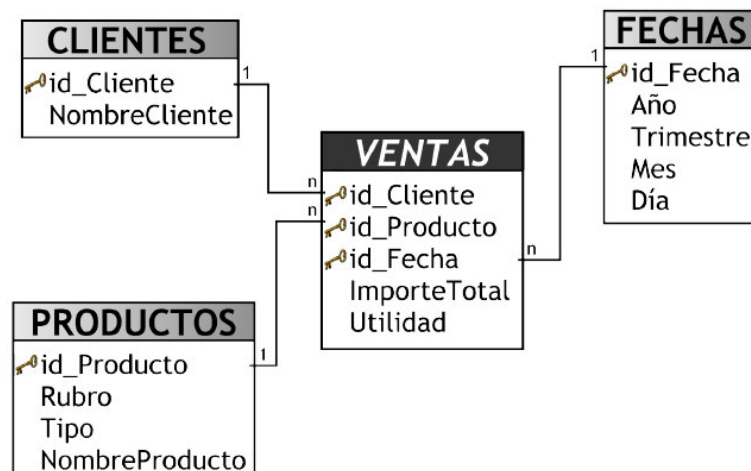





Ilustración 14.- Ejemplo de un DWH realizado

1.7. 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 acceso y son las siguientes:

PROGRAMA	DESCRIPCIÓN	UTILIZACIÓN
<p>Oracle VM VirtualBox V6.1</p> 	<p>Desarrolle usando el VirtualBox, implemente en cualquier lugar Flexibilidad esencial para un mundo de multiplataformas Como el producto de virtualización multiplataforma más popular en el mundo, el Oracle VM VirtualBox permite ejecutar varios sistemas operativos en Mac OS, Windows, Linux, o Oracle Solaris[16].</p>	<p>Se ha creado un ambiente adecuado, con las prestaciones necesarias para el correcto funcionamiento del sistema. Se trabaja en una máquina virtual con las características siguientes:</p> <ul style="list-style-type: none"> • Oracle Virtual Box • Version:6.1.2 • Memory 8GB • Processors 4 • Hard Disk:20 GB • SO: Windows Server 2012
BASE DE DATOS		
<p>MySQL Community Edition V8.0.26</p> 	<p>MySQL es un sistema de gestión de bases de datos que cuenta con una doble licencia. Por una parte es de código abierto, pero por otra, cuenta con una versión comercial gestionada por la compañía Oracle[17].</p>	<p>Diseño e implementación de la base de datos original y posteriormente del datawarehouse, almacenamiento de datos y conexión con los otros programas</p>
CALIDAD DE DATOS		
<p>RAPID MINER</p> 	<p>RapidMiner es una herramienta de Minería de Datos ampliamente usada y probada a nivel internacional en aplicaciones empresariales, de gobierno y academia[18].</p>	<p>Permite asegurar que los datos obtenidos son de utilidad para trabajarlos, además de permitir trabajar con indicadores estáticos y dinámicos.</p>
MANEJO DE ETL/DATAWAREHOUSE		
<p>PENTAHO SPOON PDI V8.3</p> 	<p>También conocido como Kettle, es la herramienta ETL de Pentaho. Como su propio nombre indica (Extract, Transform, Load), nos permitirá la extracción de datos de diferentes fuentes, transformar dichos datos y cargarlos en diferentes destinos[19].</p>	<p>Extrae los datos obtenidos del taller y mediante los procesos diseñados se los inserta en el datawarehouse por medio de una conexión con mysql.</p>

MANEJO DE CUBOS DE INFORMACIÓN		
<p>PENTAHO SCHEMA WORKBENCH V8.3</p> 	<p>Schema Workbench es una herramienta desarrollada con java con altas capacidades de edición y customización, que facilita la creación y corrección de los ficheros XML que componen el esquema de Mondrian a desarrollar[20].</p>	<p>Se conecta con el dataware house y se puede diseñar y modelar los cubos para las futuras consultas, igualmente se conecta con mysql.</p>
PRESENTACIÓN DE DATOS		
<p>PENTAHO BI SERVER V8.3</p> 	<p>Pentaho es una plataforma BI (Business Intelligence) la cual permite transformar los datos en información, y esta información en conocimiento, de forma que sea útil para investigar los eventos y el curso de la empresa[21].</p>	<p>Permite cargar los datos del cubo creado anteriormente y gestionar los plugins para visualizar los datos.</p>
<p>PLUGIN SAIKU</p>  <p>CUTTING EDGE OPEN SOURCE ANALYTICS</p>	<p>Saiku es un excelente visor OLAP que proporciona al usuario final una magnifica herramienta para realizar análisis de forma fácil e intuitiva[22].</p>	<p>Plugin de Pentaho BI Server el cual permite visualizar los cubos que se crearon e interactuar con las medidas, niveles y jerarquías creadas.</p>
<p>PLUGIN DATAFOR</p>  <p>DATAFOR</p>	<p>Es un software de análisis de exploración y visualización de datos lanzado por Shanghai Shuwei Information Technology Co., Ltd. en la plataforma Pentaho. Se implementa en el servicio Pentaho Business Analytics (PBA) en forma de complemento de Pentaho[23].</p>	<p>Permite crear reportes y gráficos con los datos del cubo creado anteriormente.</p>

Tabla 2.- Software que se utilizará

1.8. ARQUITECTURA GENERAL DEL SISTEMA

La arquitectura que se va a manejar consta de forma general de 5 partes:

- 1) Fuentes: Son los datos que se obtuvieron del taller de robótica, puestos en JSON los cuales se pasó dicha información a un CSV.
- 2) Extracción, transformación y carga: En esta parte se trabajará con los datos obtenidos en Pentaho, cargando los datos de interés.
- 3) Almacenamiento: Aquí la información extraída se almacena en una base de datos con dimensiones y tabla de hechos, creando así el datamart.
- 4) Agregación: Los datos del datamart se los analiza mediante cubos de información en Pentaho y se obtiene diferentes análisis.
- 5) Visualización: Los análisis realizados se los puede visualizar como cubos, como dashboards o como reportes.

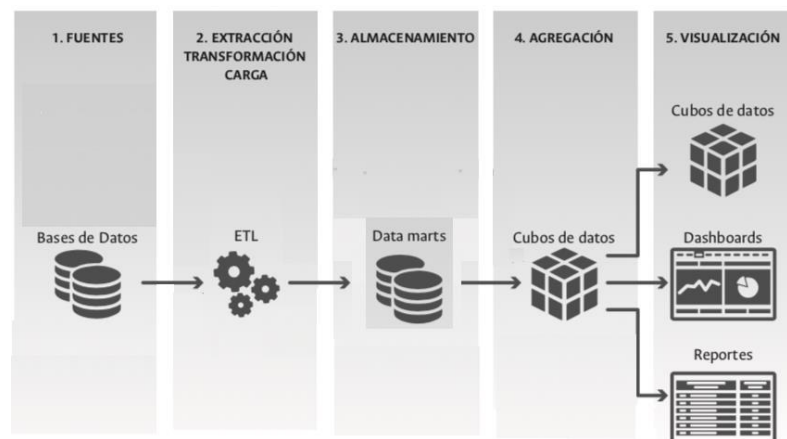


Ilustración 15 Arquitectura del sistema de inteligencia de negocios

2. CAPÍTULO 2. DESARROLLO DEL SISTEMA

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 con HEFESTO se puede utilizar para construir un Data Warehouse o un Data Mart a la vez.

Un Datamart (DM) 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 Datamart 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.

2.1.1 PLANEAR EL ANÁLISIS DE REQUERIMIENTOS

a) Caso de estudio

Una de las bases fundamentales dentro de la metodología Hefesto es el análisis de los requerimientos de los diferentes usuarios, ya que esto determinará el rumbo que tome la investigación hacia un desarrollo que permita obtener los resultados deseados y que sean de utilidad para los involucrados.

Se toma como partida los datos obtenidos, concentrándose en los datos de emociones alegría, ira, miedo, asco, tristeza, neutral, sorpresa, desprecio; con estos datos se procede a analizar los requerimientos a cumplir.

Se va a implementar los requerimientos que involucren el análisis

b) Análisis de información recolectada

Dentro de las actividades previas, se tienen las siguientes:

- Concatenar 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.
- Obtener la medición de la experiencia positiva o negativa que presentó un estudiante de acuerdo con las emociones obtenidas de las fotografías.
- Clasificar a los alumnos en grupos de acuerdo con la experiencia de usuario obtenida.

c) Identificar preguntas necesarias

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 siguientes:

IDENTIFICADOR	PREGUNTAS	Usuario R-Rector P->Profesor	FACTIBILIDAD
P1	¿Consideraría útil dividir los resultados de los niños en grupos calculados mediante el sistema?	P ,R	si
P2	¿Cuál cree que es el sentimiento más importante para conocer si el niño está disfrutando el taller?	P ,R	si
P3	¿Cuál cree usted que es el parámetro que indica mayor interés en el taller?	P ,R	si

P4	¿Cuál cree que es el parámetro que mejor refleje que el niño tiene dificultades en el taller?	P ,R	si
P5	¿Cuál cree que es el mejor parámetro para agrupar y analizar las mediciones del taller?	P ,R	si
P6	¿Cuál cree que es el parámetro que muestra mejor el desinterés de los niños en el taller?	P ,R	si
P7	¿Cree usted que el nivel de neutralidad del niño es un índice importante para considerarlo en una medición?	P ,R	si
P8	¿Consideraría de utilidad para el análisis crear un parámetro EMOCIONES POSITIVAS al promediar los sentimientos FELICIDAD, SORPRESA Y NEUTRALIDAD?	P ,R	si
P9	¿Usted cree que sea una buena idea modificar el taller con los resultados obtenidos en el análisis?	P ,R	si
P10	¿De que manera cree que influya el los cambios que se podrían hacer a los cursos de tipo WeDo 2.0?	P ,R	si

Tabla 3.- Tabla de preguntas para los profesores

d) Identificar indicadores óptimos y perspectivas que se requieren.

Identificador	Indicador	Perspectiva
P1	mayor emoción positiva detectada	Clúster, fase
P2	mayor emoción positiva detectada	Clúster, fase
P3	promedio de las emociones	Clúster, fase
P4	promedio de las emociones	Clúster, fase
P5	emociones detectadas	Clúster, fase
P6	Clúster	Clúster, fase
P7	experiencia positiva o negativa, lentes	Clúster, fase

Tabla 4 Indicadores óptimos que se dedujeron de las preguntas

2.1.2 REALIZAR EL ANÁLISIS DE REQUERIMIENTOS

a) Elaboración de encuestas profesores y autoridades del plantel.

Para asegurar el cumplimiento de requerimientos, se ha optado por hacer una encuesta a los profesores y el rector del plantel, para verificar que los indicadores obtenidos sean de utilidad para la institución y los resultados obtenidos representen un beneficio.

b) Validar indicadores y perspectivas para diseñar el sistema.

Los datos que se han obtenido han sido se los ha distribuido de la siguiente manera para un mejor análisis.

Datos obtenidos

- Emociones Negativas (promedio de IRA, DESPRECIO, MIEDO , ASCO)
- Emociones Positivas (promedio ALEGRIA, NEUTRAL, SORPRESA)

Perspectivas: Estudiante, Fase, Taller.

Indicadores: Mayor y menor emoción positiva, promedio emociones detectadas, mejor, peor fase.

c) Realizar modelo conceptual

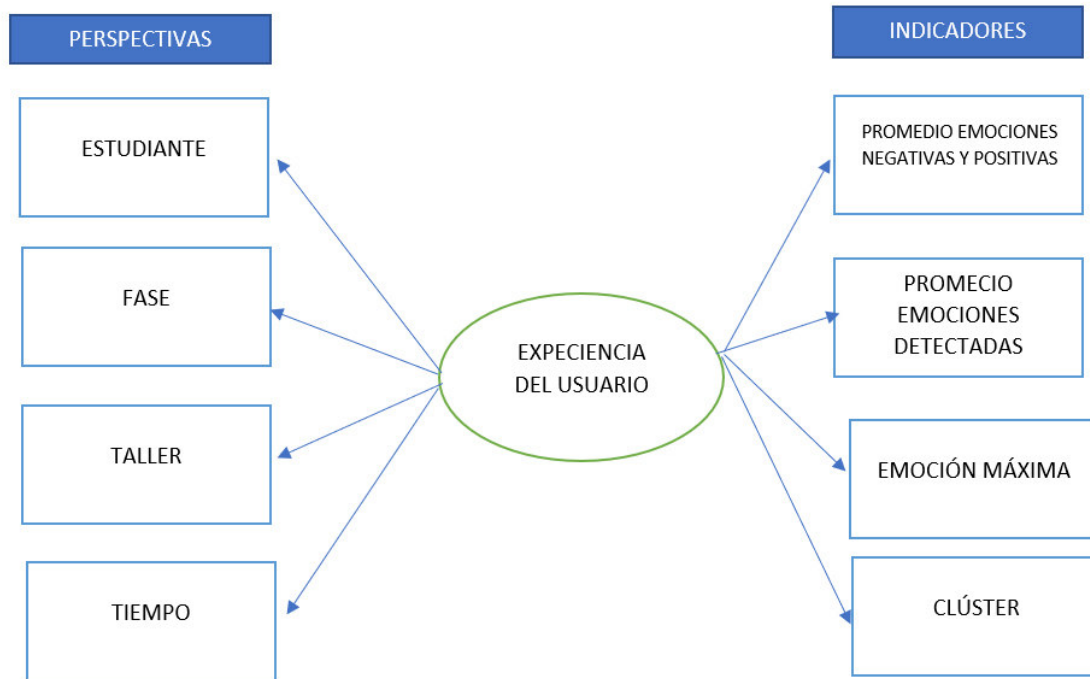


Ilustración 16.- Modelo Conceptual desarrollado

2.1.3 ANALIZAR DATOS OBTENIDOS Y VERIFICAR VIABILIDAD

Los indicadores obtenidos, y las perspectivas obtenidas se han comprobado con los miembros de la institución como útiles y de gran importancia, por lo que se ha procedido a implementarlas en los modelos diseñados y se los ha considerado para la presentación final de datos.

2.2 ANÁLISIS DE OLTP

OLTP son los sistemas conocidos como On-Line Transactional Processing. Estos procesan las transacciones de tiempo real de un negocio. Contienen estructuras de datos optimizadas para la introducción y a la adición de los datos.

Su principal desventaja es que proporciona capacidades muy limitadas para la toma de decisiones. Se podría decir que estos sistemas definen el comportamiento operacional de un entorno operacional de gestión:

- Altas, bajas, modificaciones, consultas.
- Consultas rápidas y encuestas.
- Poco volumen de información.
- Transacciones rápidas.
- Gran volumen de concurrencia.

2.2.1 DISEÑO EL ANÁLISIS DE OLTP

Dentro de las estructuras de datos, se dice que los OLTP son objetos diseñados para soportar el procesamiento de información en grandes cantidades, y el énfasis recae en maximizar la capacidad transaccional de sus datos sin perder su calidad.

De esta manera, 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 tiene gran utilidad al manejar varios datos que se obtienen continuamente.

2.2.2 IMPLEMENTACIÓN ANÁLISIS DE OLTP

En la implementación, serán calculados los indicadores de la manera previamente descrita para establecer las respectivas correspondencias de forma que satisfagan los requisitos previamente mostrados. Así, se definen de manera correcta las perspectivas y ampliará el modelo conceptual para reflejar de mejor manera lo que se va a representar.

Con los diferentes tipos de información que se tiene, se procede a extraer los datos que serán necesarios para encontrar las respuestas a las preguntas que se determinaron.

Posteriormente a estos procesos, se almacenará la información y se gestionará los metadatos para la generación de los procesos ETL.

Para definir un ETL, se dice que es un tipo de integración de datos que hace referencia a los tres pasos (extraer, transformar, cargar) que se utilizan para mezclar datos de múltiples fuentes.

Se utiliza a menudo para construir un almacén de datos. Durante este proceso, los datos se toman (extraen) de un sistema de origen, se convierten (transforman) en un formato que se puede almacenar y se almacenan (cargan) en un datawarehouse u otro sistema. Extraer, cargar, transformar (ELT) es un enfoque alternativo pero relacionado diseñado para canalizar el procesamiento a la base de datos para mejorar el desempeño. De esta manera, se busca facilitar la integración de las diversas fuentes, internas y externas[24].

2.2.3 CONFORMAR INDICADORES

Los indicadores se calcularán de la siguiente manera:

- Total de Emociones.- Representa la sumatoria de las emociones que se presentan en los talleres por edad, fase, taller y grupo.
 1. Hechos: enfado, disgusto, desprecio, temor, felicidad, neutral, tristeza, sorpresa
 2. Función: Suma de cada una de las emociones (SUM)

- Promedio de Emociones.- Es el promedio de cada una de las emociones que se presentan en los talleres por edad, fase, taller y grupo.
 1. Hechos: enfado, disgusto, desprecio, asco, temor, felicidad, neutral, tristeza, sorpresa
 2. Función: Promedio de cada una de las emociones (AVG)

- Variable de Clúster
 1. Hechos: Clúster
 2. Función: Utilización de la función Clúster de Rapidminer, que tiene varias consideraciones de otros campos para su generación.

2.2.4 ESTABLECER CORRESPONDENCIAS Y VERIFICAR NIVEL DE GRANULARIDAD

Teniendo los datos obtenidos como JSON y las fotos, se generó un modelo entidad relación. Posteriormente, se analizó los campos para la generación del OLTP con la información requerida, sus características para identificar las correspondencias entre el modelo conceptual y las fuentes de datos.

Teniendo el siguiente modelo entidad relación:

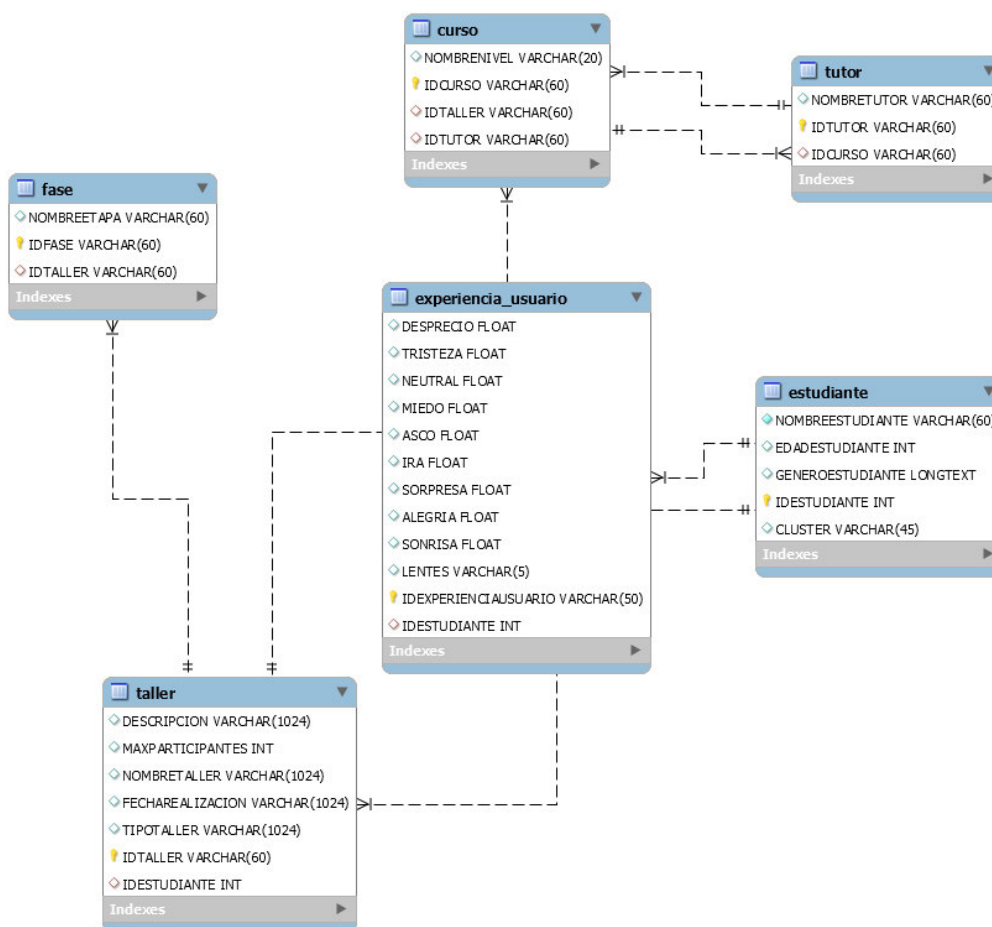


Ilustración 17.- Modelo de base de datos

El modelo entidad relación, establece ciertas correspondencias con las perspectivas e indicadores encontrados, como se observa en la siguiente imagen:

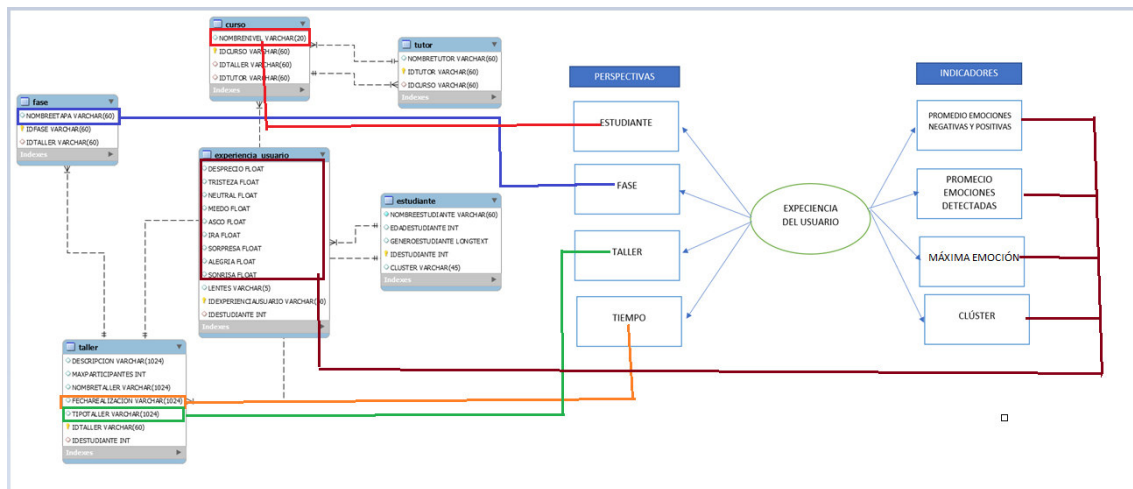


Ilustración 18.- Correspondencias entre la base y el modelo

Las relaciones identificadas fueron las siguientes:

- La tabla “Estudiantes” con la perspectiva “Estudiante”.
- Los campos “enfado”, “desprecio”, “asco”, “temor”, “felicidad”, “neutral”, “tristeza” y “sorpresa” de la tabla “Experiencia de usuario” con la perspectiva “Estudiantes”.
- La tabla “Taller” con la perspectiva “Fase”.
- El campo grupo de la tabla “Taller” con la perspectiva “Taller”.
- El campo “fase” de la tabla “Estudiantes” con la perspectiva “Fases”.
- La sumatoria de los campos “enfado”, “desprecio”, “asco”, “temor”, “felicidad”, “neutral”, “tristeza” y “sorpresa” de la tabla “Estudiantes” con el indicador “Emociones detectadas”.
- El promedio de los campos “enfado”, “desprecio”, “asco”, “temor”, “felicidad”, “neutral”, “tristeza” y “sorpresa” de la tabla “Estudiantes” con el indicador “Promedio de emociones”.
- Los campos “photoID”, “nombreEstudiante” y de la tabla “Estudiantes” con la perspectiva “Estudiante”.
- El campo “experienciaUsuario” con la perspectiva “Experiencia de Usuario”.
- La suma total del campo “experienciaUsuario” de la tabla “Estudiante” con el indicador “Total de Experiencia”.

- El promedio del campo “experienciaUsuario” de la tabla “Estudiante” con el indicador “Promedio”.

Nivel de granularidad

Con los datos previamente presentados en cada tabla de las que se hizo referencia. se procedió a validar la estructura presentada, teniendo los siguientes resultados:

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

- idEstudiante: es la clave primaria de la tabla “Estudiante”, 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.
- género: género del estudiante, con valores masculino y femenino.
- edad: edad del estudiante
- tutor: identificador del tutor que estuvo realizando seguimiento del taller, se validó que solo se utilizará un nombre genérico.

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.
- parteIzquierda: 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.
- género: género del estudiante, valores posibles son hombre, mujer 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.

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' y 'negativa' .

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.

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

- Perspectiva "Estudiante":
 - "idEstudiante" de la tabla "Estudiante", ya que representa a un estudiante.
 - "genero" de la tabla "Estudiante", ya que referencia al género del estudiante.
 - "edad" de la tabla "Estudiante", ya que referencia al género del estudiante.
 - "nombresCompletos" de la tabla "Estudiante", ya que referencia a los nombres y apellidos del estudiante.
- Perspectiva "Experiencia del usuario":
 - "enfado" de la tabla "Experiencia del usuario", ya que hace referencia a la emoción enfado.
 - "desprecio" de la tabla "Experiencia del usuario", ya que hace referencia a la intensidad de emoción desprecio.
 - "asco" de la tabla "Experiencia del usuario", ya que hace referencia a la intensidad de la emoción disgusto.
 - "temor" de la tabla "Experiencia del usuario", ya que hace referencia a la intensidad de la emoción temor.
 - "felicidad" de la tabla "Experiencia del usuario", ya que hace referencia a la intensidad de la emoción felicidad.
 - "neutral" de la tabla "Experiencia del usuario", ya que hace referencia a la intensidad de la emoción neutral.
 - "tristeza" de la tabla "Experiencia del usuario", ya que hace referencia a la intensidad de la emoción tristeza.
 - "sorpresa" de la tabla "Experiencia del usuario", 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 “Taller”
- Perspectiva “Fases”:
 - “fase” de la tabla “Experiencia del usuario”, ya que hace referencia al identificador de la fase.
 - “nombreDeLaFase” de la tabla “Experiencia del usuario”, ya que hace referencia al nombre de la fase del taller[7].

2.3 MODELO LÓGICO DEL DW

El proceso BI nos pide gestionar datos guardados en diversos formatos, fuentes y tipos, para luego depurarlos e integrarlos, para su posterior análisis y exploración, por lo que es imprescindible que satisfaga todas las necesidades vistas previamente para tener un proceso de datawearing óptimo.

De esta manera, los datos que se recolectan pueden ser integrados, transformados y limpiados, para luego ser cargados en el DW y de esta manera llevar a cabo procesos de consulta y análisis para luego tomar decisiones estratégicas y tácticas de alto nivel.

La parte importante del DW son sus datos históricos, ya que es lo que permite desarrollar pronósticos y análisis de tendencias y patrones, a partir de una base estadística de información.

Una de las maneras de guardar datos del DW es en cubos multidimensionales, ya que estos preparan esta información para responder a consultas dinámicas con un buen rendimiento. De esta manera, se procede a definir 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[7].

2.3.1 IMPLEMENTACIÓN DEL MODELO LÓGICO DEL DW

Uno de los componentes de este modelo son las tablas de dimensiones, donde los datos son organizados lógicamente y proveen el medio para analizar el contexto del negocio. De esta manera, se pueden obtener datos cualitativos los cuales pueden mostrar aspectos de interés, mediante los cuales los usuarios podrán filtrar, manipular, visualizar y analizar la información almacenada en la tabla de hechos.

Cada tabla de dimensión podrá contener los siguientes campos:

- (PK) Clave principal.
- (FK) Claves foráneas.(referencia a otras tablas)
- Datos de referencia primarios: datos que identifican la dimensión.
- Datos de referencia secundarios: datos que complementan la descripción de la dimensión.

Es muy importante tener en cuenta la definición de granularidad y estructuración de la misma depende de la dinámica del negocio que se esté analizando. 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[25].

2.3.2 CREACIÓN DE TABLAS DE DIMENSIONES

En el modelo planteado se contempla crear 4 tablas dimensionales, las cuales se presenta a continuación:

La tabla de la Dimensión Fase tiene el campo de la PK “DIMFASE” y el nombre de la Fase

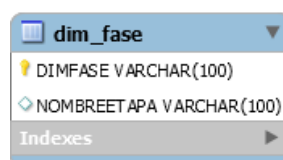


Ilustración 19 Tabla Dimensión Fase

La tabla de la Dimensión Taller tiene el campo de la PK “DIMITALLER” y los campos de el nombre del taller, el curso y la fase.

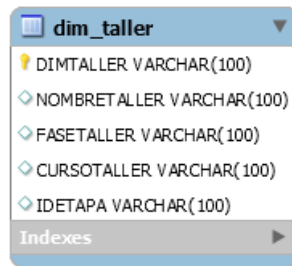


Ilustración 20 Tabla Dimensión Taller

La tabla de la Dimensión Estudiante tiene el campo de la PK “DIMESTUDIANTE” y el nombre del estudiante, su edad, género y el clúster calculado.

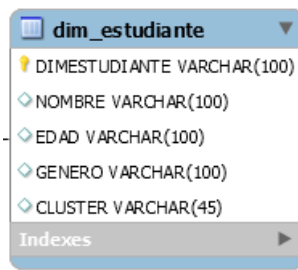


Ilustración 21.- Tabla Dimensión Estudiante

La tabla Dim tiempo tiene la PK ID_TIEMPO, además de la fecha de realización del taller y una división del año, mes y día.

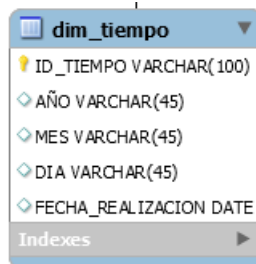


Ilustración 22.- Tabla Dimensión Tiempo

2.3.3 CREACIÓN DE TABLA DE HECHOS.

La tabla de hechos de experiencia del usuario consta de todas las medidas de emociones de los datos recolectados. Un ID_EXPERIENCIA que es la PK , el promedio de emociones positivas y negativas, si tienen lentes y los 4 campos de FK de las otras tablas.

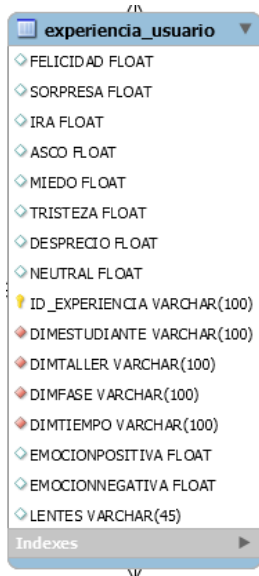


Ilustración 23 Tabla de Hechos experiencia del usuario

2.3.4 CREACIÓN DE UNIONES Y VERIFICACIÓN DEL MODELO

Se crean las uniones correspondientes y se valida que el modelo sea funcional.

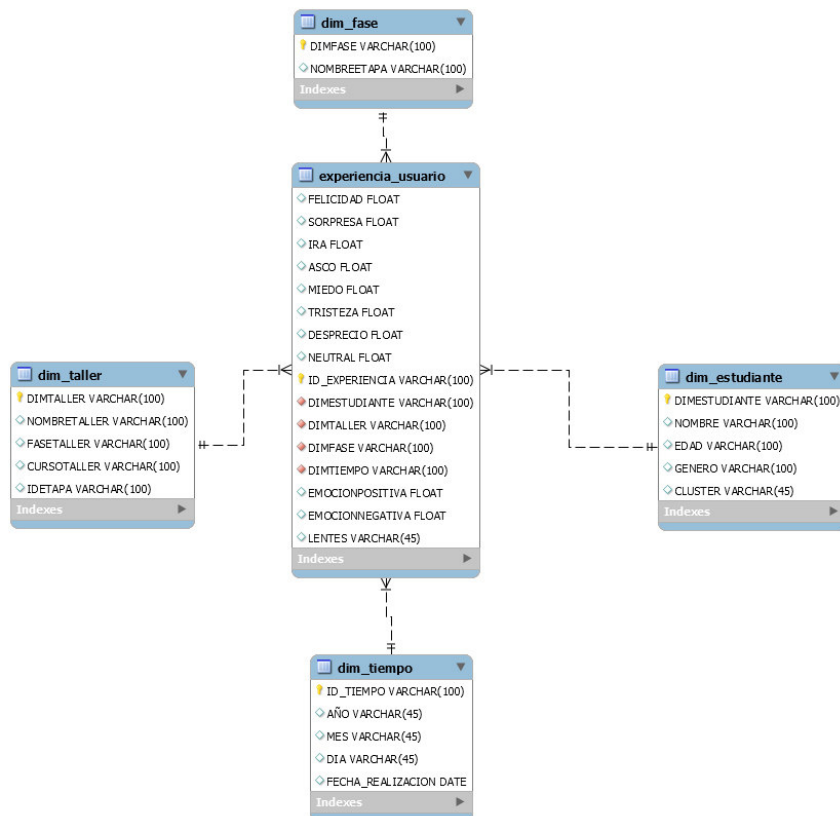


Ilustración 24 Modelo del Datawarehouse completo

2.4 INTEGRACIÓN DE DATOS Y VISUALIZACIÓN

En esta sección se presentarán y visualizarán los datos recolectados y trabajados a lo largo del proyecto.

2.4.1 DISEÑO DE INTEGRACIÓN DE DATOS

Dentro de estos modelos, 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 estrella.

Se debe hacer un exhaustivo análisis de los modelos que se tiene. Así se puede tener conclusiones como que no es recomendable usar copo de nieve a pesar de que se desee orden, o el modelo constelación, permite tener más de una tabla de hechos, por lo cual no es útil en este proyecto.

El modelo óptimo en este caso sería el modelo estrella, ya que su simplicidad y adaptabilidad es perfecto para el tipo de análisis que se necesita.

2.4.2 IMPLEMENTACIÓN DE LA INTEGRACIÓN DE DATOS

a) Calidad de datos

Para tener una idea de la calidad de nuestros datos, se ha decidido utilizar el software RapidMiner, para verificar esto, donde se cargan todos los datos obtenidos de los niños.

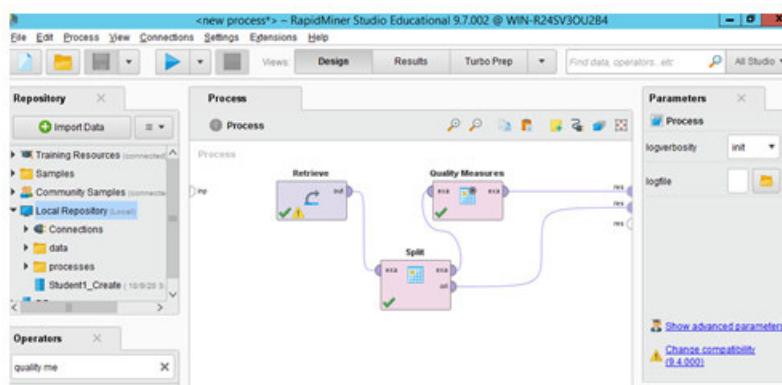


Ilustración 25.- Diseño de la operación de calidad de datos

Se puede observar en la tabla de quality measure los siguientes puntos:

- 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, text-ness),
- Columnas donde casi todos los valores son idénticos (Stability),
- Columnas con valores perdidos (Missing).

Row No.	Attribute	ID-ness	Stability	Missing	Text-ness
1	id	1	0.007	0	0
2	faceld	1	0.007	0	0.498
3	student	0.007	1	0	0
4	phase	0.007	1	0	0
5	anger	0.007	0.945	0	0
6	contempt	0.007	0.568	0	0
7	disgust	0.007	0.911	0	0
8	fear	0.007	0.904	0	0
9	happiness	0.014	0.274	0	0
10	neutral	0.007	0.151	0	0
11	sadness	0.007	0.295	0	0
12	surprise	0.007	0.479	0	0
13	age	0.082	0.322	0	0
14	gender	0.014	0.863	0	0.030

ExampleSet (25 examples, 0 special attributes, 5 regular attributes)

Ilustración 26.- Datos generados sobre la calidad de los datos

Se puede representar dicha estabilidad de forma gráfica, observando cierta homogeneidad en los datos obtenidos con valores altos y varios datos en valores altos.

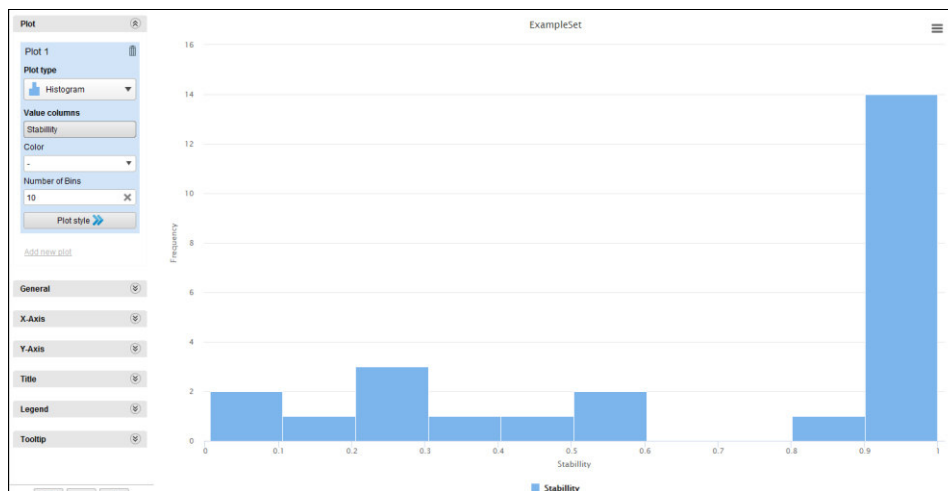


Ilustración 27.- Histograma de estabilidad de datos

De la misma manera, se puede representar el text-ness de forma gráfica, observando que la gran mayoría de datos son similares al ser analizados los datos obtenidos en el taller. Con estos resultados se verifica que los datos obtenidos son de calidad suficiente para implementarlos.

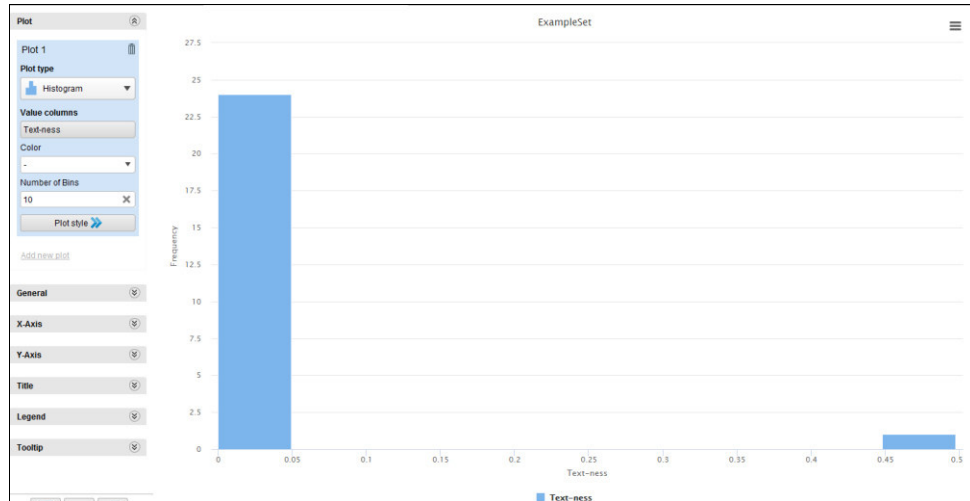


Ilustración 28.- Histograma de text-ness

b) Generación de Cluster en los datos

Agrupación por medio del algoritmo K-means

Este operador realiza la agrupación mediante el algoritmo *k-means* el cual junta los datos que son similares entre sí. Como no es necesario ningún atributo de *etiqueta*, la agrupación en clústeres se puede utilizar en datos sin etiquetar y es un algoritmo de aprendizaje automático no supervisado.

El algoritmo de *k-means* determina un conjunto de *k* grupos y asigna cada ejemplo a un grupo exacto. Los grupos constan de ejemplos similares. La similitud entre los ejemplos se basa en una medida de distancia entre ellos.

Un grupo en el algoritmo de k-means está determinado por la posición del centro en el espacio n-dimensional de los n Atributos del conjunto de datos. Esta posición se llama centroide.

El algoritmo de k-means comienza con un número *k* de puntos que se tratan como el centroide de *k* grupos potenciales. Estos puntos de inicio son la posición de *k* ejemplos extraídos aleatoriamente del conjunto de datos de entrada, o están determinados por la heurística de k-means.

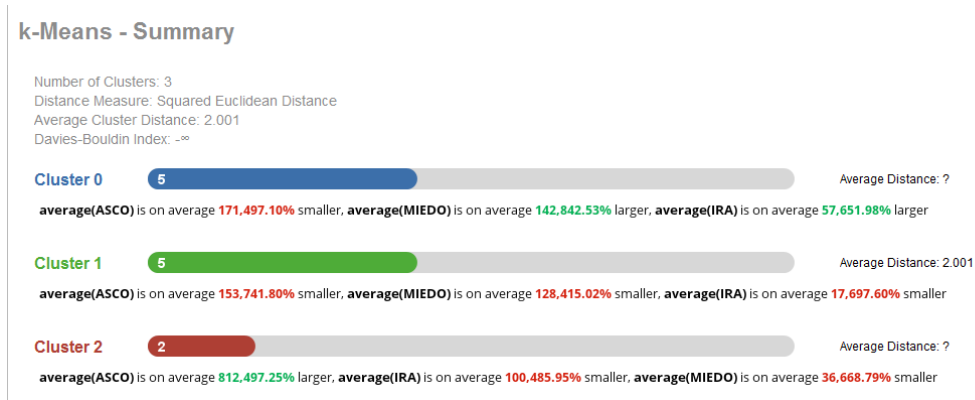


Ilustración 29.- División por Clústers

Todos los ejemplos se asignan a su grupo más cercano (el más cercano se define por el *tipo de medida*). A continuación, los centroides de los grupos se vuelven a calcular promediando todos los ejemplos de un grupo. Los pasos anteriores se repiten para los nuevos centroides hasta que estos ya no se mueven o se alcanzan los *pasos máximos de optimización*. Así, hay que tener en cuenta que no se garantiza que el algoritmo de k-means converja si el tipo de medida no se basa en el cálculo de la distancia euclidiana (porque el recálculo de los centroides promediando supone un espacio euclidiano)[26].

IRA	DESPRECIO	ASCO	MIEDO	FELICIDAD	NEUTRAL	TRISTEZA	SORPRESA	id	cluster
0,006	0,005	0,000	0,000	0,266	0,617	0,033	0,072	1	cluster_0
0,000	0,002	0,000	0,000	0,311	0,622	0,025	0,039	2	cluster_0
0,004	0,003	0,001	0,000	0,217	0,746	0,016	0,012	3	cluster_0
0,000	0,001	0,000	0,000	0,238	0,676	0,013	0,071	4	cluster_0
0,001	0,001	0,000	0,000	0,205	0,742	0,017	0,034	5	cluster_0
0,001	0,002	0,000	0,000	0,098	0,868	0,007	0,023	6	cluster_1
0,000	0,001	0,000	0,000	0,169	0,816	0,013	0,000	7	cluster_1

0,002	0,036	0,000	0,001	0,001	0,870	0,032	0,057	8	cluster_2
0,000	0,009	0,000	0,000	0,052	0,936	0,002	0,002	9	cluster_1
0,000	0,000	0,000	0,000	0,000	0,955	0,044	0,000	10	cluster_1
0,000	0,000	0,000	0,002	0,002	0,974	0,011	0,010	11	cluster_2
0,000	0,000	0,000	0,000	0,333	0,663	0,002	0,000	12	cluster_1

Tabla 5 Clústers exportados de Rapid Miner

De esta manera, primero se ha agrupado por estudiante todas las medidas de emociones registradas, luego a este conjunto reducido de datos se les ha aplicado este algoritmo, para asignar a cada estudiante a un clúster, teniendo que la distribución más equitativa que se pudo lograr es con el uso de 3 clúster.

Mediante este mapa de calor, podemos observar cómo se distribuyeron los datos en los clúster generados.

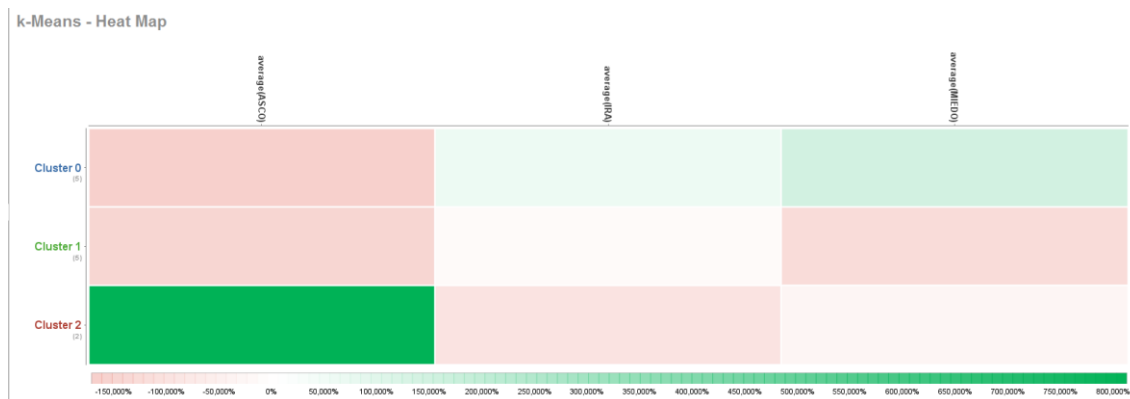


Ilustración 30.-Mapa de calor de Clústers

Por otro lado, en el Árbol de Clúster, se puede observar los criterios que utilizó RapidMiner para crear los Clúster, donde se observa que decidió tomar como principal parámetro el promedio del MIEDO para separar a los del Clúster 2, y según ese valor separar una primera vez, luego tomó como segundo parámetro el promedio de ASCO, para separar a los del Clúster 1 y el último parámetro elegido es del de promedio de ALEGRÍA, donde los separa entre Clúster 0 y Clúster 1.

k-Means - Cluster Tree

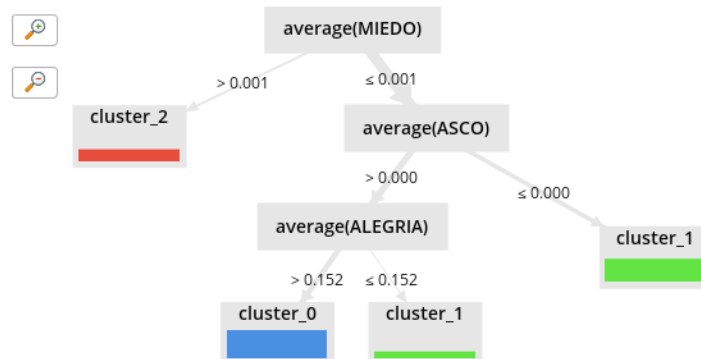


Ilustración 31.- *Árbol de decisiones de Clúster*

Finalmente, se puede observar los valores del centroide de la tabla K-Means, donde se ve los valores de cada uno de los Clústers.

k-Means - Centroid Table

Cluster	average(IRA)	average(DESPRECO)	average(ASCO)	average(MIEDO)	average(ALEGRÍA)	average(NEUTRAL)	average(TRISTEZA)	average(SORPRESA)
Cluster 0	0.003	0.012	0.000	0.001	0.129	0.661	0.027	0.033
Cluster 1	0.001	-0.000	0.000	0.000	0.095	0.848	0.008	0.018
Cluster 2	-0.001	0.002	0.001	0.000	0.389	0.922	0.021	0.035

Tabla 6 *Tabla de valores de Centroides de los Clúster*

c) Carga de datos en el sistema

- **Ingreso de datos dimensión taller**

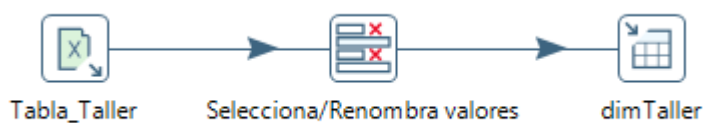


Ilustración 32.- *Extracción de datos ETL de la dimensión*

- Extracción de los datos de tipo excel “taller”, que contiene la tabla taller.
- Selección de los campos de interés en este caso los campos nombreTaller, Curso, Fase
- Almacenamiento en la tabla de Dimensión dim_Taller los datos obtenidos en el paso anterior.
-

- **Ingreso de datos dimensión Estudiante**



Ilustración 33.- Extracción de datos ETL de la dimensión Estudiante

- Extracción de los datos de tipo excel “tabla_Estudiante”, que contiene la tabla estudiante.
- Selección de los campos de interés en este caso los campos nombre, edad, género, clúster
- Almacenamiento en la tabla de Dimensión dim_estudiante los datos obtenidos en el paso anterior.

- **Ingreso de datos dimensión Fase**



Ilustración 34.-Extracción de datos ETL de la dimensión Fase

- Extracción de los datos de tipo excel “tabla_fase”, que contiene la tabla fase.
- Selección de los campos de interés en este caso el campo nombreFase
- Almacenamiento en la tabla de Dimensión dim_Fase los datos obtenidos en el paso anterior.

- **Ingreso datos dimensión tiempo**

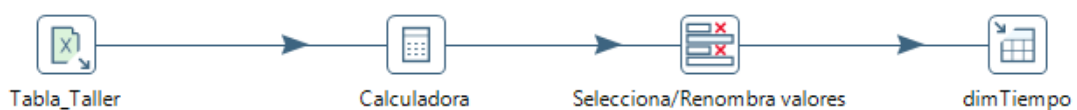


Ilustración 35.-Extracción de datos ETL de la dimensión Tiempo

- Extracción de los datos de tipo excel “tabla_Taller”, que contiene la tabla.
- Se separa la fecha en Año, Mes, Día mediante la calculadora
- Selección de los campos de interés en este caso los campos fecha, Año, Mes, Día
- Almacenamiento en la tabla de Dimensión dim_Tiempo los datos obtenidos en el paso anterior.

- **Ingreso datos Tabla de Hechos**



Ilustración 36.-Extracción de datos ETL de la tabla de hechos

- Extracción de los datos de tipo excel “tabla_Experiencia_Usuario”, que contiene la tabla estudiante.
- Selección de los campos de interés en este caso los campos ira, asco, desprecio, felicidad, tristeza, sorpresa, neutral, emociónPositiva, emociónNegativa
- Almacenamiento en la tabla de Dimensión Tabla_Hechos_UX los datos obtenidos en el paso anterior.

c) Actualización

Después de la primera carga inicial, se procede a activar la acción INSERT/UPDATE, la cual permitirá insertar nuevos registros o actualizar los existentes.

2.4.3 IMPLEMENTACIÓN DE CUBOS MULTIDIMENSIONALES

Una vez que se tiene el DWH lleno con los datos, se procede a crear el cubo de datos, teniendo ya las dimensiones y la tabla de hechos.

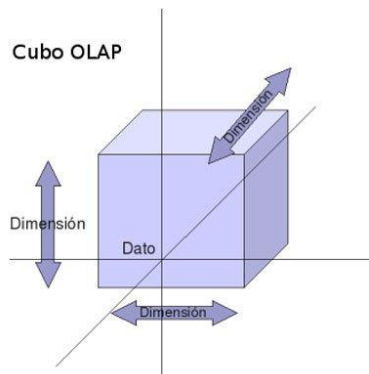


Ilustración 37.- Esquema de un cubo multidimensional

En el caso del sistema, se tiene el cubo dispuesto de la siguiente manera:
Dimensiones: Estudiante, Fase, Taller, Tiempo

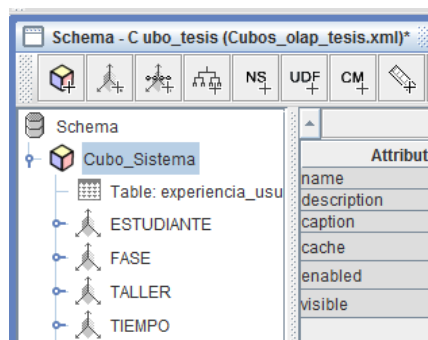


Ilustración 38.- Dimensiones del cubo

Dimensión Estudiante:

-Jerarquías: Estudiante, Clúster, Género

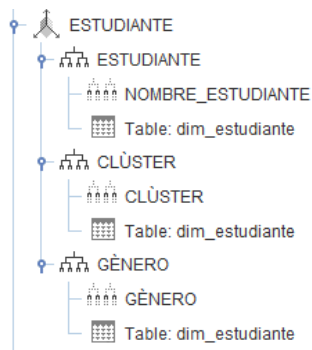


Ilustración 39.- Jerarquías dentro de las dimensión Estudiante

Dimensión Fase

-Jerarquía: Nombre Fase

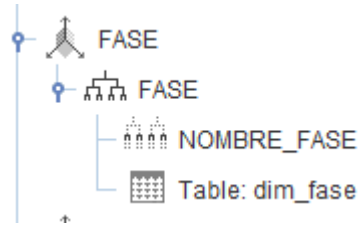


Ilustración 40.-Jerarquías dentro de las dimensión Fase

Dimensión Taller

-Jerarquía: Curso Taller, Nombre_Taller

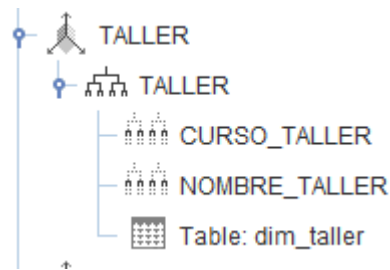


Ilustración 41.- Jerarquías dentro de las dimensión Taller

Dimensión Tiempo

-Jerarquía: Tiempo

-Niveles: Año, Mes, Día, Fecha

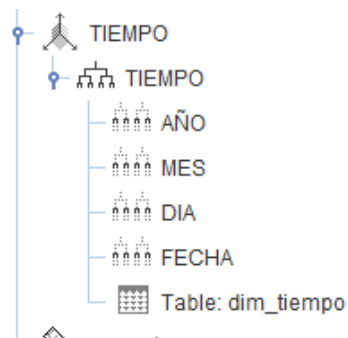


Ilustración 42.-Jerarquías dentro de las dimensión Tiempo

Medidas: promedios y máximos de cada valor



Ilustración 43.- Medidas establecidas dentro del cubo

2.4.3 IMPLEMENTACIÓN Y VISUALIZACIÓN DE DATOS

a) Análisis de datos obtenidos

1. Promedio de emociones positivas detectadas por clúster y por fase

CLÚSTER	cluster_0				cluster_1				cluster_2			
	f	n	s	pos	f	n	s	pos	f	n	s	pos
Compartir	0.24	0.685	0.038	0.321	0.085	0.887	0.016	0.329	0	0.97	0	0.323
Crear	0.268	0.663	0.044	0.325	0.134	0.827	0.025	0.329	0.201	0.721	0	0.307
Explorar	0.089	0.785	0.074	0.316	0.088	0.872	0.026	0.329	0.002	0.783	0.199	0.328

Tabla 7.- Tabla del primer análisis del cubo

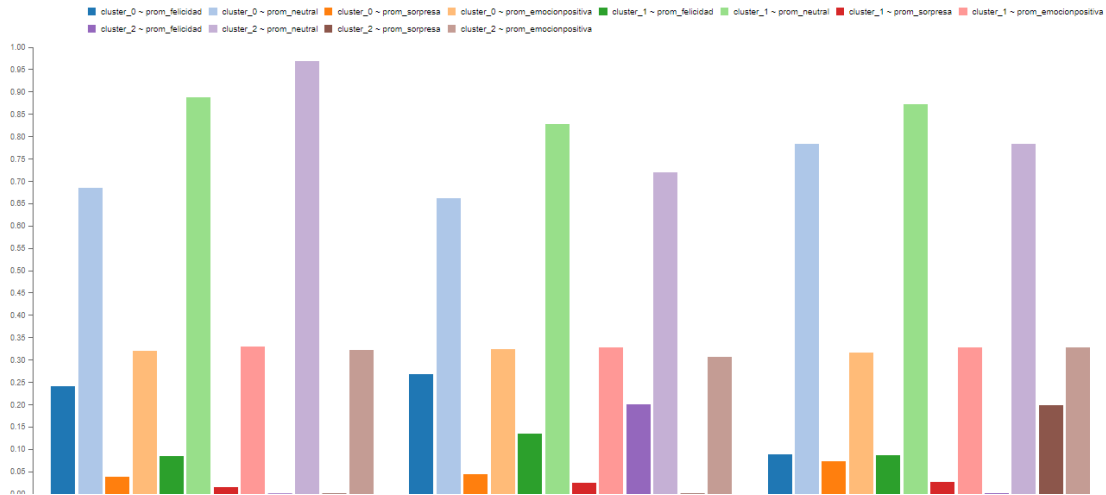


Ilustración 44.- Histograma del primer análisis del cubo

Se puede observar que el clúster con mayores emociones positivas/neutral es el clúster 1. Además, la fase que causa más emociones positivas es la fase de compartir.

2. Promedio emociones negativas detectada por clúster, por fase

CLÚSTER	cluster_0					cluster_1					cluster_2					
	FASE	a	d	i	t	neg	a	d	i	t	neg	a	d	i	t	neg
Compartir		0	0.005	0.004	0.027	0.008	0	0.001	0	0.011	0.003	0	0.002	0	0.028	0.007
Crear		0	0.003	0.002	0.019	0.005	0	0.002	0.001	0.009	0.003	0	0.049	0	0.028	0.007
Explorar		0.002	0.001	0.013	0.036	0.013	0	0.001	0	0.012	0.003	0.001	0.002	0.007	0.004	0.004

Tabla 8.- Tabla del segundo análisis del cubo

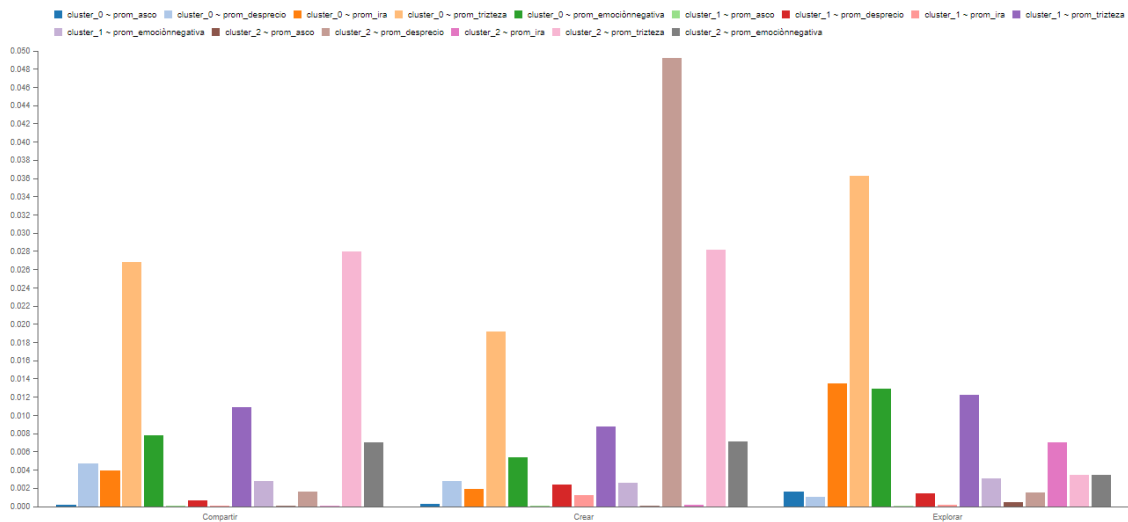


Ilustración 45.- Histograma del segundo análisis del cubo

Se puede observar que el clúster con mayores emociones negativas es el Clúster 0, ya que en los valores de las 3 fases tienen más que el resto. Mientras que la fase con mayores emociones negativas es la fase de Explorar.

2. Mayor valor de ira por clúster por fase

CLÚSTER	cluster_0	cluster_1	cluster_2	Total
NOMBRE_FASE	max_ira	max_ira	max_ira	
Compartir	0,29	0,00	0,00	0,10
Crear	0,74	0,25	0,00	0,33
Explorar	0,76	0,00	0,01	0,26
Total	0,60	0,08	0,01	

Tabla 9.- Tabla del tercer análisis del cubo

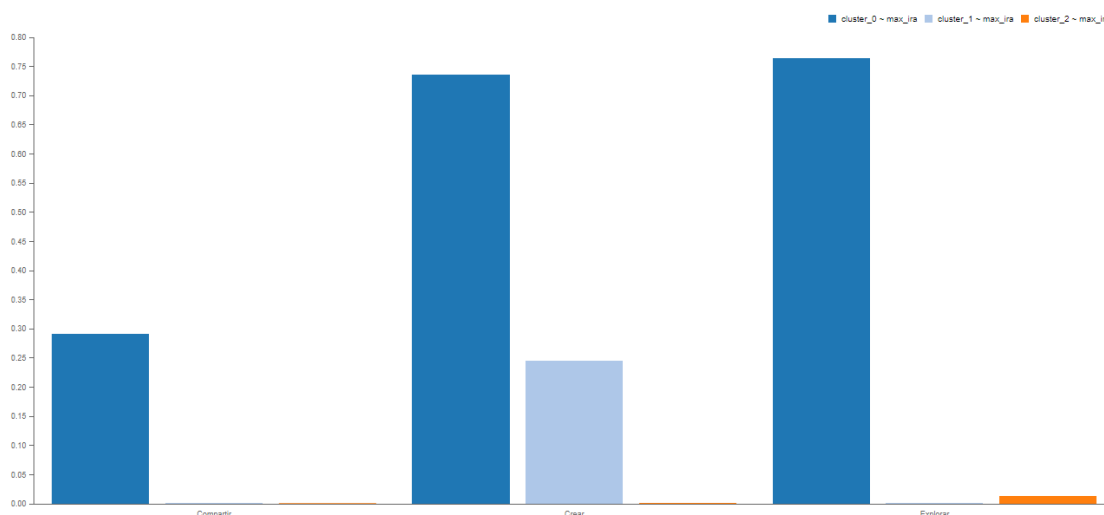


Ilustración 46.- Histograma del Tercer análisis del cubo

El clúster que ha causado el mayor registro de ira/frustración ha sido el Clúster 0, mientras que la fase que ha causado el mayor registro de ira/frustración es la fase de Crear

3. Mayor promedio de sorpresa por clúster por fase

NOMBRE_FASE	Compartir	Crear	Explorar	TOTAL
CLÚSTER	sorpresa	sorpresa	sorpresa	
cluster_0	0,038	0,044	0,074	0,052
cluster_1	0,016	0,25	0,026	0,021
cluster_2	0	0	0,199	0,066
Total	0,018	0,022	0,100	

Tabla 10.- Tabla del cuarto análisis del cubo

El clúster que ha causado el mayor promedio de sorpresa ha sido el Clúster 2, mientras que la fase que ha causado el mayor promedio de sorpresa es la fase de Explorar

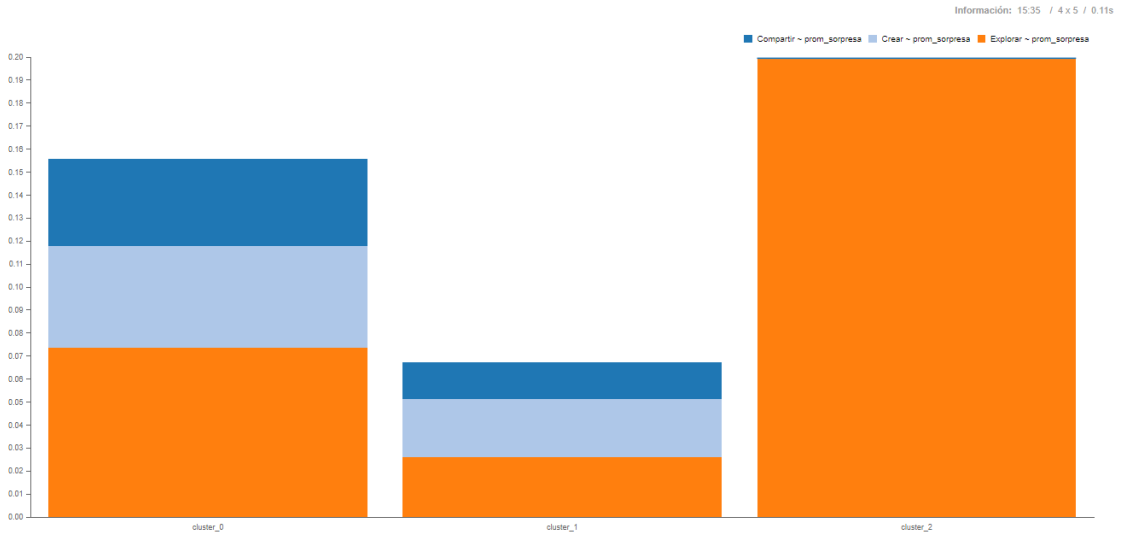


Ilustración 47.- Histograma del cuarto análisis del cubo

5. Desprecio por estudiante por fase.

NOMBRE_FASE	Compartir	Crear	Explorar	TOTAL
CLÚSTER	desprecio	desprecio	desprecio	
cluster_0	0,005	0,03	0,001	0,003
cluster_1	0,001	0,002	0,001	0,001
cluster_2	0,002	0,049	0,002	0,018
TOTAL	0,0027	0,0255	0,0013	

Tabla 11 .- Tabla del quinto análisis del cubo

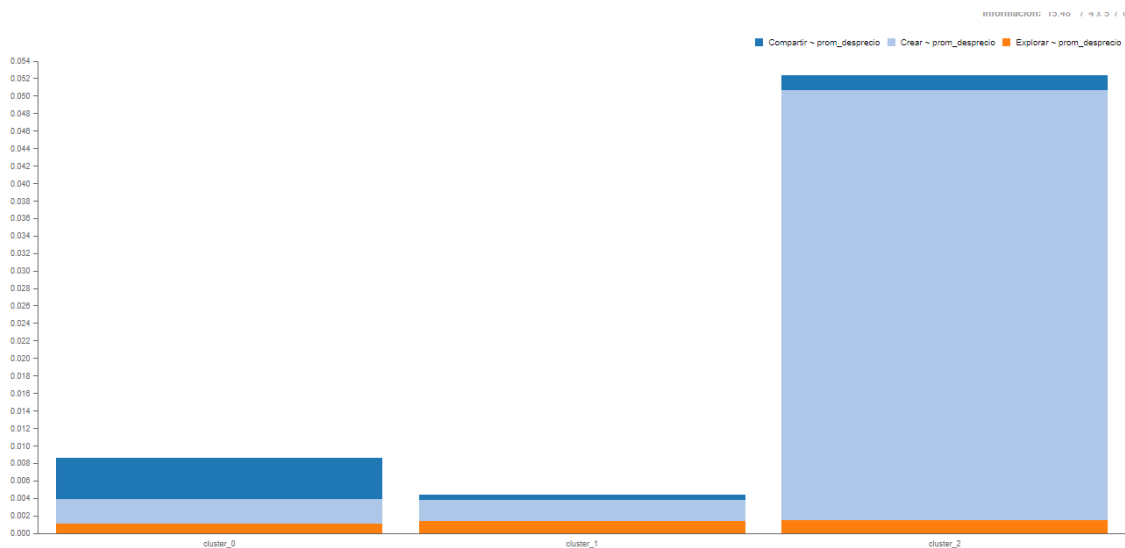


Ilustración 48.-.- Histograma del quinto análisis del cubo

Como se observa, el Clúster que ha causado mayor desprecio/desinterés es el Cluster 2, mientras que la fase que ha causado mucho mayor desprecio/desinterés es la fase crear.

b) Interpretación de datos obtenidos, indicadores y perspectivas

1. Se ha comprobado que el Clúster 1 es el que ha tenido una mejor experiencia, por lo que se debería trabajar con los Clúster 0 y Clúster 2 para mejorar su experiencia, además la Fase compartir es la que proporciona una mejor experiencia, por lo que se puede extrapolar los elementos de esta para mejorar las otras 2 fases.
2. El clúster con mayores emociones negativas es el Clúster 0, por lo tanto hay que poner más atención en este grupo de niños para conocer por qué se sienten mal. Por otro lado la fase con mayores emociones negativas es la fase de Explorar, por lo que hay que evaluar las actividades de esta fase para saber por qué causa estos sentimientos más que las otras fases.
3. El clúster que ha causado el mayor registro de ira/frustración ha sido el Clúster 0, por lo que hay que prestar atención a qué elementos hacen que este grupo de niños se frustren más. Además, la fase que ha causado el

mayor registro de ira/frustración es la fase de Crear, por lo que hay que revisar las razones por las que frustra más que las otras fases.

4. El clúster que ha causado el mayor promedio de sorpresa ha sido el Clúster 2 por lo que sería ideal validar los elementos que hacen que los niños muestren mucho más interés, mientras que la fase que ha causado el mayor promedio de sorpresa es la fase de Explorar, por lo que se debe ver lo replicable de esta en otras fases.

5. El Clúster que ha causado mayor desprecio/desinterés es el Cluster 2, por lo que hay que acercarse a los niños de este grupo y validar las actividades que más desprecian, mientras que la fase que ha causado mucho mayor desprecio/desinterés es la fase crear, por lo que se debería comparar las actividades con las otras fases y pensar qué se puede hacer para mejorarla.

3. CONCLUSIONES Y RECOMENDACIONES

3.1 CONCLUSIONES

- Se ha logrado implementar de manera efectiva un sistema de inteligencia de negocios dedicado al monitoreo de la experiencia de usuario en un taller de robótica educativa, dónde se permite analizar los datos y todo esto haciendo uso de software libre, por lo cual es una solución rápida, eficaz y sin costo.
- Mediante encuestas se ha logrado analizar los requerimientos de la unidad educativa para la implementación del sistema de inteligencia de acuerdo a las necesidades que presentaban y que les permitirá mejorar la comprensión de los alumnos y tener una perspectiva diferente para mejorar los cursos que impartan.
- Los datos fueron recolectados de manera satisfactoria y posteriormente se ha realizado su respectivo análisis e interpretación, mostrando los resultados de manera clara para los profesores.
- A través del análisis de los datos, mediante gráficos y tablas se ha ofrecido un reporte a los profesores de la experiencia y sentimientos de los estudiantes en diferentes situaciones dentro del taller y estos datos servirán como guía de los aspectos en que se puede mejorar la clase.

3.2 RECOMENDACIONES

- Al no disponer de una gran cantidad de datos, este sistema no se utiliza a su máxima capacidad por lo que es recomendable obtener un mayor volumen de datos, ya que el sistema podría hacer análisis mucho más profundos con una mayor cantidad de registros de niños, incluir más cursos, diferentes talleres y registrar más fechas.

- Sería de gran utilidad agregar a la metodología Hefesto la fase de implementación y visualización ya que permitiría integrar de mejor manera con los pasos anteriores y se podría estandarizar de mejor manera la presentación de los datos al cliente. Además, al estar en una misma metodología sería mucho más fácil la documentación y escalabilidad de un sistema creado a partir de Hefesto.
- Es de vital importancia mantener una comunicación constante con el cliente, ya que este puede brindar retroalimentación de los aspectos faltantes o mejorables del sistema, lo que facilita que el sistema final cumpla con las expectativas y les de la ayuda y apoyo que necesitan.
- Se debe establecer lineamientos al momento de tomar los datos de los alumnos, para evitar problemas en la detección de emociones, registrando el nombre de cada alumno al tomar sus datos, enfocando solo a un alumno a la vez y cuidando que las fotos no sean borrosas, esto facilitará mucho al momento de analizar la información y se tendrá datos más confiables.

4. BIBLIOGRAFÍA

- [1] C. J. Mendoza-Bozada, «Tecnología en la educación ecuatoriana logros, problemas y debilidades», *Dominio de las Ciencias*, vol. 6, n.º 3, pp. 496-516, jul. 2020.
- [2] D. M. G. Ayabaca, J. A. J. Alba, y E. E. E. Guamán, «Implementación de las TIC en el ámbito educativo ecuatoriano», *Sociedad & Tecnología*, vol. 2, n.º 2, Art. n.º 2, jul. 2019, doi: 10.51247/st.v2i2.49.
- [3] Tableau, «¿Qué es la inteligencia de negocios? Guía sobre la inteligencia de negocios y por qué es importante», *Tableau*. <https://www.tableau.com/es-es/learn/articles/business-intelligence> (accedido ago. 11, 2021).
- [4] Y. R. Dixson, «La inteligencia de negocio como apoyo a la toma de decisiones en el ámbito académico», vol. 3, p. 11, 2015.
- [5] L. Education, «WeDo 2.0 Support | Everything You Need», *LEGO® Education*. <https://education.lego.com/en-us/product-resources/wedo-2/troubleshooting/faqs> (accedido ago. 10, 2021).
- [6] Intanis, «5 problemas comunes en el análisis de datos y cómo solucionarlos», *Analítica, Big Data y Transformación Digital*. <https://intanis.com/5-problemas-analisis-de-datos-y-como-solucionarlos/> (accedido ago. 11, 2021).
- [7] R. Bernabeu, «hefesto-v2.pdf». jul. 19, 2010.
- [8] «Base de datos». <https://sistemas.com/base-de-datos.php> (accedido sep. 10, 2021).
- [9] SAS, «¿Qué es la minería de datos?» https://www.sas.com/es_mx/insights/analytics/data-mining.html (accedido ago. 12, 2021).
- [10] W. Said, M. M. Hassan, y A. Fawzy, «Data mining techniques for database prediction: Starting point», *Journal of Theoretical and Applied Information Technology*, vol. 96, pp. 7723-7738, dic. 2018.
- [11] A. B. Román, D. Sánchez-Guzmán, y R. García, «Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo», n.º 4, p. 7, 2014.

- [12] «Factores críticos de éxito de un proyecto de Business Intelligence». <https://www.dataprix.com/es/book/export/html/8002> (accedido sep. 10, 2021).
- [13] SNGULA, «CRISP-DM: La metodología para poner orden en los proyectos», *Sngular*, ago. 02, 2016. <https://www.sngular.com/es/data-science-crisp-dm-metodologia/> (accedido ago. 10, 2021).
- [14] G. R. Rivadera, «La metodología de Kimball para el diseño de almacenes de datos (Data warehouses)», n.º 5, p. 16, 2010.
- [15] J. Galindo losada, «Adaptation de interfaces utilizando emociones», Tesis de doctorado, Université Grenoble Alpes (ComUE), 2019. Accedido: ago. 10, 2021. [En línea]. Disponible en: <http://www.theses.fr/2019GREAM021>
- [16] «Experimente el Oracle VM VirtualBox – Oracle Virtualization». https://go.oracle.com/LP=104602?elqCampaignId=281168&src1=:ad:pas:go:dg:a_lad:71700000086320342-58700007355811342-p65908951384:RC_WWMK201210P00015C0001:SPA&SC=:ad:pas:go:dg:a_lad::RC_WWMK201210P00015C0001:SPA:&gclid=CjwKCAjwvuGJBhB1EiwACU1AiRhi2Pep8H6PBMRbLLilw6DEkYErEEbSIQ7_xRkRBfARRnIj8H1rxoCbplQAvD_BwE&gclsrc=aw.ds (accedido sep. 10, 2021).
- [17] «Qué es MySQL: Características y ventajas», *OpenWebinars.net*, sep. 24, 2019. <https://openwebinars.net/blog/que-es-mysql/> (accedido sep. 10, 2021).
- [18] «Auto Model - RapidMiner Documentation». <https://docs.rapidminer.com/9.3/studio/auto-model/> (accedido sep. 10, 2021).
- [19] «Qué es Pentaho. Sus Productos y Ventajas», *Incentro*, sep. 10, 2021. <https://www.incentro.com/es-es/blog/stories/que-es-pentaho/> (accedido sep. 10, 2021).
- [20] «Schema Workbench», *Auribox Training*, dic. 28, 2017. <https://blog.auriboxtraining.com/business-intelligence/schema-workbench/> (accedido sep. 10, 2021).
- [21] «Qué es Pentaho BI Server», *OpenWebinars.net*, nov. 13, 2019. <https://openwebinars.net/blog/que-es-pentaho-bi-server/> (accedido sep. 10, 2021).

- [22] «Saiku: La Herramienta de Analisis OLAP del Futuro | Dataprix TI». <https://www.dataprix.com/es/planet-it/jortilles/saiku-herramienta-analisis-olap-del-futuro> (accedido sep. 10, 2021).
- [23] «Visualizer简介», *DATAFOR*. http://www.datafor.com.cn/knowledge-base/visualizer_about/ (accedido sep. 10, 2021).
- [24] «¿Qué es ETL? | SAS». https://www.sas.com/es_ar/insights/data-management/what-is-etl.html (accedido sep. 10, 2021).
- [25] D. Tarapues, «METODOLOGÍA HEFESTO», p. 21.
- [26] «k-Means - RapidMiner Documentation». https://docs.rapidminer.com/latest/studio/operators/modeling/segmentation/k_means.html (accedido sep. 10, 2021).

5. ANEXOS

I. Glosario

A

- **Arquitectura, 38:** Un data mart es una base de datos centrada en un ámbito que cumple las exigencias de un grupo concreto de usuarios

B

- **Base de datos, 20:** Conjunto de datos organizado de tal modo que permita obtener con rapidez diversos tipos de información.
- **BI, 13:** Abreviación de Business Intelligence (Inteligencia de Negocios)

C

- **Clustering, 23:** agrupamiento de conjuntos de objetos no etiquetados, para lograr construir subconjuntos de datos conocidos como Clusters
- **correlación, 58:**
Relación recíproca entre dos o más cosas o series de cosas.
- **Cubos de datos, 26:** Estructura de datos que supera las limitaciones de las bases de datos relacionales y proporciona un análisis rápido de datos.

D

- **Datamart, 39:** Base de datos centrada en un ámbito que cumple las exigencias de un grupo concreto de usuarios
- **Datawarehouse, 29:** Es un repositorio unificado para todos los datos que recogen los diversos sistemas para fines analíticos y de acceso.
- **DB, 22:** Abreviación de Database (base de datos).
- **DW, 19:** Abreviación de datawarehouse.

E

- **Esquema, 33:** Es una representación visual de una base de datos, a un conjunto de reglas que rige una base de datos, o bien, a todo el conjunto de objetos que pertenecen a un usuario en particular.
- **ETL, 44:** Del inglés (Extract: extraer, Transform: transformar, Load: cargar) Es un término estándar que se utiliza para referirse a estas operaciones aplicadas a un conjunto de datos.

G

- **Granularidad, 54:** Representa el nivel de detalle al que se desea

almacenar la información sobre el negocio que se esté analizando

I

- **Indicadores de rendimiento, 20:** Es un valor medible que informa si nuestra empresa es eficaz en la consecución de sus objetivos principales
- **Información, 20:** Comunicación o adquisición de conocimientos que permiten ampliar o precisar los que se poseen sobre una materia determinada.
- **Inteligencia de Negocios o (BI), 13:** Es la combinación de tecnología, herramientas y procesos que permiten transformar datos almacenados en información, esta información en conocimiento y este conocimiento dirigido a un plan o una estrategia comercial.

J

- **JSON, 46:** acrónimo de JavaScript Object Notation, 'notación de objeto de JavaScript' es un formato de texto sencillo para el intercambio de datos

K

- **kit WeDo, 15:** Es un kit de robótica recomendado para niños a partir de 7 años. Permite construir y programar robots

M

- **Metodología, 17:** Conjunto de métodos que se siguen en una investigación científica o en una exposición doctrinal

- **Minería de datos, 22:** es un conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos
- **Modelos conceptuales, 19:** Es un modelo de base de datos el cual identifica las relaciones de más alto nivel entre las diferentes entidades
- **Modelos lógicos, 19:** Es un modelo que no es específico de una base de datos que describe aspectos relacionados con las necesidades de una organización para recopilar datos y las relaciones entre estos aspectos

N

- **Necesidades, 19:** Es aquella información que el los usuarios desearían obtener.

O

- **OLAP,37:** El procesamiento analítico en línea (online analytical processing) es un método informático que permite a los usuarios extraer y consultar datos de manera fácil y selectiva para analizarlos desde diferentes puntos de vista
- **OLTP, 43:** El OLTP (Procesamiento de Transacciones En Línea) es una clase de programa que facilita y gestiona aplicaciones orientadas a transacciones, típicamente para ingreso de datos

P

- **Perspectivas, 27:** Las perspectivas son categorías de la organización o del conjunto de datos con las que se

asocian iniciativas, objetivos o metas.

- **Procesos, 23:** Son una sucesión de instrucciones que pretenden llegar a un estado final o que persiguen realizar una tarea concreta.

R

- **Reportes, 20:** Informes que organizan y exhiben la información contenida en una base de datos
- **Retroalimentación, 19:** Es un proceso para controlar continuamente los elementos que intervienen en un determinado proceso, así como los resultados obtenidos, para introducir las modificaciones necesarias.

T

- **Tablas de dimensiones, 28:** Es una tabla de datos almacena información

descriptiva sobre los valores numéricos de una tabla de hechos.

- **Tablas de hechos, 28:** Es una tabla d que almacenan los datos utilizados para calcular las medidas en informes de medidas y se llenan solo a través de transformaciones de ETL
- **Tareas, 23:** Es un proceso individual que se ejecuta dentro de un sistema.
- **TICs, 13:** Las tecnologías de Información y Comunicación (TICs) son el conjunto de herramientas relacionadas con la transmisión, procesamiento y almacenamiento digitalizado de la información.

U

- **usuarios finales, 19 :** Personas que harán uso del sistema. Educativa.

II. Modelo del JSON generado con el que se trabajó inicialmente

```
{
  "jsonFeels": [
    {
      "faceRectangle": {
        "height": 101,
        "top": 159,
        "left": 319,
        "width": 101
      },
      "faceId": "3e453704-fed0-4fd4-8573-135eee19b5c7",
      "faceAttributes": {
        "emotion": {
          "neutral": 0.977,
```



```
"fear": 0.001,  
"sadness": 0.001,  
"contempt": 0.001,  
"anger": 0,  
"happiness": 0,  
"disgust": 0,  
"surprise": 0.02  
},  
"noise": {  
  "noiseLevel": "high",  
  "value": 1  
},  
"gender": "female",  
"blur": {  
  "value": 1,  
  "blurLevel": "high"  
},  
"facialHair": {  
  "beard": 0,  
  "sideburns": 0,  
  "moustache": 0  
},  
"headPose": {  
  "pitch": -5.6,  
  "roll": -3,  
  "yaw": 7.2  
},  
"makeup": {  
  "eyeMakeup": false,  
  "lipMakeup": false  
},  
"glasses": "NoGlasses",  
"occlusion": {  
  "eyeOccluded": false,  
  "mouthOccluded": false,  
  "foreheadOccluded": true  
},  
"hair": {  
  "hairColor": [],  
  "bald": 0,  
  "invisible": true  
},  
"accessories": [  
  {  
    "confidence": 1,  
    "type": "headwear"  
  }  
],
```

```
"age": 18,  
"exposure": {  
  "value": 0.25,  
  "exposureLevel": "underExposure"  
},  
"smile": 0  
}  
},  
{  
  "Etapa": "Explorar",  
  "fecha": "2020-1-23 11:7:25"  
}  
]  
},  
{
```

III. Ejemplos de fotografías

Fotografía borrosa



Fotografía con varios rostros



Fotografía con un tutor detectado



Fotografía válida



IV. Encuesta en Google Forms para Director y Maestros

¿Consideraría útil dividir los resultados de los niños en grupos calculados mediante el sistema? *

- SI
- NO
- ME ES INDIFERENTE

¿Cuál cree que es el sentimiento más importante para conocer si el niño está disfrutando el taller? *

- FELICIDAD
- SORPRESA

⋮

¿Cuál cree usted que es el parámetro que indica mayor interés en el taller? *

- FELICIDAD
- SORPRESA
- SONRISA

¿Cuál cree que es el parámetro que mejor refleje que el niño tiene dificultades en el taller? *

- IRA
- TRISTEZA
- DESPRECIO
- MIEDO



¿Cuál cree que es el mejor parámetro para agrupar y analizar las mediciones del taller? *

- POR TALLER
- POR GRUPO
- POR FECHA
- POR FASE

¿Cuál cree que es el parámetro que muestra mejor el desinterés de los niños en el taller? *

- IRA
- DESPRECIO
- MIEDO
- TRISTEZA

¿Consideraría de utilidad para el análisis crear un parámetro EMOCIONES POSITIVAS al promediar los sentimientos FELICIDAD, SORPRESA Y NEUTRALIDAD? *

- Sí
- No

¿Consideraría de utilidad para el análisis crear un parámetro EMOCIONES NEGATIVAS al promediar los sentimientos IRA, MIEDO, DESPRECIO Y TRISTEZA? *

- Sí
- No

⋮

¿Usted cree que sea una buena idea modificar el taller con los resultados obtenidos en el análisis?

Sugerencias: [Añadir todas](#) | [Sí](#) [No](#) [Tal vez](#)

SI

NO

¿De que manera cree que influya el los cambios que se podrían hacer a los cursos de tipo WeDo 2.0?

POSITIVOS

NEGATIVOS

Se puede acceder al cuestionario mediante el siguiente link :

https://docs.google.com/forms/d/1Ro-uJd5wRwrJb7pouxMdgERS_OfmB6FpnwbWZ8CS6yo/edit

