ESCUELA POLITECNICA NACIONAL

ESCUELA DE INGENIERÍA

"SISTEMA PARA PRONOSTICAR EL COMPORTAMIENTO DE CONSTRUCCIÓN DE VIVIENDAS EN EL PAÍS A TRAVÉS DEL TIEMPO APLICANDO TÉCNICAS DE INTELIGENCIA ARTIFICIAL"

PROYECTO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO INFORMÁTICO MENCIÓN SOFTWARE

JUAN CARLOS DELGADO LOYOLA

DIRECTOR: ING. CARLOS MONTENEGRO ARMAS.

Quito, Marzo 2007

INTRODUCCIÓN

La predicción del comportamiento de la construcción de viviendas es uno de los temas primordiales del desarrollo socioeconómico del país, ya que se encuentra dentro de los problemas que necesitan de herramientas efectivas que engendren soluciones óptimas para la toma de decisiones. La aplicación de técnicas de inteligencia artificial responde a esta necesidad, ya que descubre relaciones ocultas y persistentes en los datos en forma automática.

Se conoce que los métodos aplicados para predecir fenómenos son en su mayoría de tipo estadístico, como el análisis univariante o el multivariante, con resultados relativamente satisfactorios, pero con la inconveniencia de que se construyen sobre hipótesis que limitan el espacio de búsqueda y las posibilidades de asociación y distribución de las variables explicativas. De ahí el propósito de usar técnicas de inteligencia artificial, ya que éstas realizan en forma acertada una mayor exploración de las soluciones en forma heurística.

Luego de un minucioso estudio, se ha seleccionado la técnica de Los algoritmos genéticos, ya que estos buscan el mejor resultado y se aplican a problemas de clasificación y predicción. Dentro de las técnicas estudiadas, las que más se aproximan a ésta son: las Redes Neuronales Artificiales, los Sistemas Expertos, la Inducción de Reglas y los Árboles de Decisión.

El sistema desarrollado para este efecto ha sido aplicado a casos de estudio reales con datos resultado de las encuestas anuales de edificaciones (permisos de construcción) procesadas por el Instituto Nacional de Estadísticas (INEC) en base a los permisos aprobados por los municipios del país. Las variables investigadas son: el tipo de obra, los materiales predominantes en los cimientos, estructura, paredes y cubierta o techo, los usos de la edificación y el origen del financiamiento. Como parte complementaria al análisis y exploración inicial de los datos, se incluyen otras variables secundarias tales como indicadores socioeconómicos relacionados al fenómeno, con las cuales es posible analizar correlaciones con la variable permisos de construcción.

Los resultados son reglas de asociación clasificadas por el algoritmo con valores de cobertura y confianza para una conjunto de casos de ejemplo. El parámetro de confianza se evalúa por medio de pruebas de contraste de hipótesis aplicado a proporciones, mecanismo frecuentemente usado en estadística paramétrica. Para determinar si una regla es válida dentro de un rango de aceptación del parámetro, se selecciona muestras aleatorias de ejemplos distintas a la utilizada por el algoritmo genético durante el proceso de descubrimiento de reglas, y de esta forma se tiene un grado de certeza relativamente aceptable de que las reglas aplicadas son persistentes y predicen el comportamiento futuro de la conformación de las viviendas en el país.

ÍNDICE DE CONTENIDO

Pág

INTRODUCCIÓN ……..... 2
INDICE DE CONTENIDO 4

CAPITULO 1. EL COMPORTAMIENTO DE LA CONSTRUCCIÓN DE VIVIENDAS EN EL PAÍS 12

1.1 DESCRIPCIÓN GENERAL DEL FENÓMENO DE LA CONSTRUCCIÓN DE VIVIENDAS EN EL PAÍS..... 12
1.1.1 GENERALIDADES 12
1.1.2 PRINCIPALES FACTORES QUE EXPLICAN EL COMPORTAMIENTO DE LA CONSTRUCCIÓN DE VIVIENDAS EN EL PAÍS 13
1.1.2.1 Factores del medio físico y natural 13
1.1.2.2 Factores demográficos y humanos 14
1.1.2.3 Factores económicos 15
1.2 DESCRIPCIÓN Y ANÁLISIS DE LAS VARIABLES INVESTIGADAS..... 19
1.2.1 DESCRIPCIÓN DE LAS VARIABLES INVESTIGADAS 19
1.2.1.1 Descripción de las variables contenidas en los permisos de construcción para edificaciones de viviendas 19
1.2.1.2 Descripción de las relaciones existentes entre las variables contenidas en los permisos de construcción para edificación de viviendas 24
1.2.2 ANALISIS DE LAS VARIABLES INVESTIGADAS 27
1.2.2.1 Generalidades sobre el análisis de las variables investigadas 27
1.2.2.2 Análisis del comportamiento del número de permisos de construcción y el índice de metros cuadrados a construirse..... 28
1.2.2.3 Análisis del comportamiento de los materiales predominantes en las viviendas..... 30
1.2.2.4 Análisis del comportamiento del tipo de obra, uso, origen y destino del financiamiento, superficie del terreno y área a construir..... 32

CAPITULO 2. TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA REALIZAR PRONÓSTICOS 36

2.1 INVESTIGACIÓN BIBLIOGRÁFICA DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA REALIZAR PRONÓSTICOS 36
2.1.1 GENERALIDADES SOBRE LA APLICACIÓN DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA REALIZAR PRONÓSTICOS 36
2.1.1.1 Aplicación de Técnicas de Inteligencia Artificial a la resolución de problemas predictivos 36
2.2. ANÁLISIS DE LAS TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA REALIZAR PRONÓSTICOS 38
2.2.1 TECNICAS BASADAS EN ALGORITMOS GENÉTICOS 38
2.2.2 TÉCNICAS BASADAS EN ÁRBOLES DE DECISIÓN 41
2.2.3 TÉCNICAS BASADAS EN REDES NEURONALES ARTIFICIALES 42
2.2.4 TÉCNICAS BASADAS EN VECINDAD Y RAZONAMIENTO BASADO EN CASOS 46

Pág

2.3 SELECCIÓN DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA REALIZAR PRONÓSTICOS..... 49
2.3.1 ESTRATEGIAS DE SELECCIÓN DE TECNICAS DE INTELIGENCIA ARTIFICIAL PARA REALIZAR PRONÓSTICOS 49
2.3.1.1 Selección de la Técnica de Algoritmos Genéticos para realizar pronósticos por clasificación mediante extracción de reglas de asociación 49
2.3.1.2 Otras estrategias de selección de técnicas de inteligencia artificial para realizar pronósticos 56
2.3.2 PRINCIPALES VENTAJAS Y DESVENTAJAS DE LA SELECCIÓN DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA REALIZAR PRONÓSTICOS 58
2.3.2.1 Ventajas y Desventajas de Selección de Técnicas de Algoritmos Genéticos para realizar pronósticos frente a otras técnicas..... 58
2.3.2.2 Ventajas y Desventajas de Selección de Otras técnicas para realizar pronósticos 59

CAPITULO 3. IMPLEMENTACIÓN DE UN SISTEMA DE PREDICCIÓN Y ANÁLISIS DE DATOS APLICADO AL COMPORTAMIENTO DE LOS CASOS DE CONSTRUCCIÓN DE VIVIENDAS EN EL PAÍS. 61

3.1 ANALISIS..... 61
3.1.1. ESPECIFICACIÓN DE REQUERIMIENTOS DE SOFTWARE 61
3.1.1.1 Introducción..... 61
3.1.1.2 Descripciones 63
3.1.1.3 Requerimientos específicos 65
3.1.1.4 Restricciones del Sistema..... 67
3.1.2 DESCRIPCION DEL PROBLEMA 68
3.1.3 MODELO DE OBJETOS 69
3.1.3.1 Descripción del modelo de objetos 69
3.1.3.2 Diagrama del modelo de objetos 72
3.1.3.3 Dicccionario de Datos 73
3.1.4 MODELO DINÁMICO 76
3.1.4.1 Identificación de escenarios y eventos 76
3.1.4.2 Diagrama de transición de estados 80
3.1.5 MODELO FUNCIONAL 81
3.1.5.1 Diagrama de Flujo de datos 81
3.1.5.2 Especificación de operaciones y funciones 82
3.1.5.3 Realización de casos de uso aplicados al modelo funcional 85
3.2. DISEÑO 86
3.2.1 DISEÑO DEL SISTEMA 86
3.2.1.1 Descripción de Subsistemas..... 86
3.2.1.2 Arquitectura del Sistema 96
3.2.2 DISEÑO DE OBJETOS 98
3.2.2.1 Diccionario de Objetos 98
3.2.2.2 Modelo Lógico basado en objetos: Modelo Entidad-Relación 99
3.2.2.3 Modelo Relacional basado en registros: Modelo Físico 100
3.2.2.4 Diseño de Interfaz de Usuario 101

Pág

3.3.1.1 Descripción del entorno de desarrollo..... 110
3.3.1.2 Creación del espacio de trabajo 110
3.3.1.3 Creación del destinatario de la aplicación 110
3.3.2 CREACIÓN DE LA BASE DE DATOS RELACIONAL 111
3.3.2.1 Creación de la Base de Datos Relacional 111
3.3.2.2 Configuración de los Objetos principales de la Base de Datos 111
3.3.3 DESARROLLO DE LA INTERFAZ DE USUARIO 113
3.3.3.1 Creación de clases de objetos visuales 113
3.3.3.2 Creación de clases de objetos no visuales 114
3.3.3.3 Creación de clases de objetos de interfaz 115
3.4 IMPLEMENTACIÓN..... 117
3.4.1 IMPLEMENTACIÓN DEL SISTEMA SPACV 117
3.4.1.1 Implementación de las Librerías 117
3.4.1.2 Implementación de los Subsistemas 118
3.4.1.3 Implementación de clases, operaciones y módulos 119
3.5 PRUEBAS 125
3.5.1 PRUEBAS DE UNIDAD 125
3.5.1.1 Plan de pruebas de unidad..... 125
3.5.1.2 Descripción de los casos de prueba de unidad 125
3.5.1.3 Ejecución de las pruebas de unidad y evaluación de resultados 126
3.5.2 PRUEBAS DE INTEGRACIÓN 130
3.5.2.1 Plan y estrategias de integración 130
3.5.2.2 Descripción de las fases de integración 130
3.5.2.3 Ejecución de las pruebas de integración y evaluación de resultados 131
3.5.3 PRUEBAS DE VALIDACIÓN 131
3.5.3.1 Plan de validación 131
3.5.3.2 Desarrollo de las pruebas de aceptación 132
3.5.3.3 Ejecución de las pruebas de validación y evaluación de resultados 132
3.5.4 PRUEBAS DEL SISTEMA 134
3.5.4.1 Pruebas de recuperación 134
3.5.4.2 Pruebas de seguridad 134
3.5.4.3 Pruebas de resistencia y rendimiento 135

CAPITULO 4. APLICACIÓN DEL SISTEMA A UN CASO DE ESTUDIO..... 136

4.1 SELECCIÓN DE EJEMPLOS DE CASOS DE PREDICCIÓN CON LAS VARIABLES INVESTIGADAS 136
4.1.1. DELIMITACIÓN Y DESCRIPCIÓN DE LOS CASOS DE EJEMPLO..... 136
4.1.1.1 Delimitación de los casos de ejemplo por localidad geográfica 136
4.1.1.2 Descripción de los casos de ejemplo por variables investigadas y reglas de asociación 136
4.2 INTRODUCCIÓN DE DATOS AL SISTEMA Y PROCESAMIENTO 137
4.2.1 INTRODUCCIÓN DE DATOS DE EDIFICACIONES Y SERIES HISTÓRICAS 137
4.2.1.1 Introducción de datos de Localidad Geográfica y Tiempo 137
4.2.1.2 Introducción de variables de edificaciones 138
4.2.1.3 Introducción de datos históricos de edificaciones..... 138

Pág

4.2.2.1 Introducción de casos de predicción 139
4.2.2.2 Introducción de parámetros del algoritmo genético 139
4.2.2.3 Introducción de tipos de reglas de asociación y parámetros de ejecución 140
4.2.2.4 Proceso de ejecución del algoritmo genético para descubrimiento de reglas de asociación para casos de predicción..... 141
4.3 ANALISIS DE RESULTADOS 142
4.3.1 PRUEBAS DE CONTRASTE DE HIPÓTESIS APLICADO A LAS REGLAS DE ASOCIACIÓN DESCUBIERTAS 142
4.3.1.1 Descripción de la metodología de pruebas de contraste de hipótesis 142
4.3.1.2 Aplicación de las pruebas de contraste de hipótesis a las reglas de asociación descubiertas 144
4.3.2 RESULTADOS DE LAS PRUEBAS DE CONTRASTE DE HIPÓTESIS ... 144
4.3.2.1 Opciones de contraste de hipótesis aplicadas en el sistema 144
4.3.2.2 Obtención de resultados de las pruebas de constraste de hipótesis 145

CAPITULO 5. CONCLUSIONES Y RECOMENDACIONES..... 149

5.1 CONCLUSIONES..... 149
5.2 RECOMENDACIONES 152

BIBLIOGRAFÍA 154

ANEXOS 159

ANEXO 1 . MANUALES 160

1.1 MANUAL DE INSTALACIÓN 160
1.1.1 REQUERIMIENTOS MÍNIMOS 160
1.1.1.1 Requerimientos de Hardware 160
1.1.1.2 Requerimientos de Software 160
1.1.2 GUIA DE INSTALACIÓN 161
1.1.2.1 Proceso de instalación 161
1.1.2.1.1 Tareas previas a la instalación 161
1.1.2.1.2 Instalación de la parte del Servidor 162
1.1.2.1.3 Instalación de la parte del Cliente 166
1.2 MANUAL DEL USUARIO..... 168
1.2.1 INTRODUCCIÓN 168
1.2.1.1 Acerca del Sistema 168
1.2.1.2 Objetivos y Alcance del Sistema 168
1.2.1.3 Terminología 169
1.2.1.3.1 Términos especiales 169
1.2.1.3.2 Abreviaturas 170
1.2.1.3.3 Siglas 170
1.2.2 GENERALIDADES 171
1.2.2.1 Seguridad y Administración de Usuarios 171
1.2.2.1.1 Descripción de Usuarios 171

Pág

1.2.2.1.4 Roles y Permisos 173
1.2.2.2 Breve descripción del Sistema 174
1.2.3 GUIA DE OPERACIÓN 175
1.2.3.1 Uso de los dispositivos de Entrada 175
1.2.3.2 Uso de los dispositivos de Salida 176
1.2.3.3 Acceso a la Aplicación 176
1.2.3.4 Descripción de Módulos y Pantallas 177
1.2.3.4.1 Módulo de Datos y Administración de Usuarios 178
1.2.3.4.2 Módulo de Proceso 187
1.2.3.4.3 Acceso a pantalla de instalación de datos de prueba y mantenimiento del sistema 193
1.2.3.4.4 Acceso a pantalla de ayuda 195
1.2.3.5 Ejecución de un caso de estudio 196

ÍNDICE DE TABLAS

Pág

Tabla 1.1 Ciudades más pobladas …… 15
Tabla 1.2 Viviendas ocupadas, desocupadas y en construcción. …… 16
Tabla 1.3 Uso Residencial de la Edificación ……... 22
Tabla 1.4 Variables principales contenidas en los permisos de construcción ……. 22
Tabla 1.5 Uso No Residencial de la Edificación ……. 23
Tabla 1.6 Origen del Financiamiento …….. 23
Tabla 1.7 Material Predominante en Techo o Cubierta ……... 30
Tabla 1.8 Material Predominante en las paredes exteriores …… 31
Tabla 1.9 Financiamiento en viviendas proyectadas …….. 34
Tabla 1.10 Origen del Financiamiento 34
Tabla 2.1 Técnicas aplicadas a la resolución de problemas predictivos …… 36
Tabla 3.1 Análisis: Especificación de Requerimientos – Definiciones ……... 62
Tabla 3.2 Análisis: Especificación de Requerimientos – Acrónimos …….. 62
Tabla 3.3 Análisis: Especificación de Requerimientos – Abreviaturas ……... 63
Tabla 3.4 Análisis: Modelo de Objetos – Descripción de Atributos ……... 75
Tabla 3.5 Análisis: Modelo de Objetos – Descripción de Operaciones …….. 76
Tabla 3.6 Análisis: Modelo Dinámico – Escenario para introducir casos de predicción ……..... 77
Tabla 3.7 Análisis: Modelo Dinámico – Escenario para procesamiento de caso de predicción …… 78
Tabla 3.8 Análisis: Modelo Dinámico – Escenario para ejecución del algoritmo genético …… 79
Tabla 3.9 Análisis: Modelo Funcional – Identificación de actores ……. 85
Tabla 3.10 Composición Genética Cromosomas de materiales predominantes de la construcción ……. 91
Tabla 3.11 Parámetros algoritmo de predicción genético ……. 92

Pág

Tabla 3.13 Lista de Entidades …… 99
Tabla 3.14 Lista de Relaciones …….. 99
Tabla 3.15 Lista de Tablas ……. 100
Tabla 3.16 Lista de Referencias ……. 100
Tabla 3.17 Opciones de menú del Sistema …… 109
Tabla 3.18 Perfil de Usuarios …… 109
Tabla 3.19 Objetos visuales básicos …….. 114
Tabla 3.20 Clases de ventana ancestro y contenedoras ……. 115
Tabla 3.21 Clases de objetos de interfaz ……... 117
Tabla 3.22 Implementación de los Subsistemas de SPACV ……. 118
Tabla 3.23 Resumen final de clases, operadores y módulos – Subsistema AG..... 124
Tabla 3.24 Casos de prueba de unidad – Subsistema: Mantenimiento de Edificaciones, Indicadores ……... 125
Tabla 3.25 Casos de prueba de unidad – Subsistema: Casos de Predicción …….. 126
Tabla 3.26 Casos de prueba de unidad – Subsistema: Algoritmo Genético …….. 126
Tabla 3.27 Casos de prueba de unidad – Subsistema: Mantenimiento Edificaciones, Indicadores…… 126
Tabla 3.28 Resultados de las Pruebas de Unidad – Subsistema: Casos de Predicción y Análisis ……... 127
Tabla 3.29 Resultados de las Pruebas de Unidad – Subsistema Algoritmo Genético Caja Negra ……... 128
Tabla 3.30 Resultados de las Pruebas de Unidad – Subsistema Algoritmo Genético Caja Blanca …….. 129
Tabla 3.31 Resultados de Integración del Sistema …… 131
Tabla 3.32 Identificación de los requerimientos de usuario para las pruebas de aceptación …… 132
Tabla 3.33 Identificación de los criterios de aceptación para las pruebas …… 132
Tabla 3.34 Resumen de pruebas de validación …….. 132
Tabla 3.35 Resultados de las pruebas de aceptación ……. 133
Tabla 3.36 Resultados de las pruebas de recuperación …….. 134
Tabla 3.37 Resultados de las pruebas de seguridad ……... 135
Tabla 3.38 Resultados de las pruebas de resistencia ……. 135
Tabla 3.39 Resultados de las pruebas de rendimiento …….. 135
Tabla 4.1 Delimitación de los casos de ejemplo por localidad geográfica ……. 136
Tabla 4.2 Descripción de variables seleccionadas para los casos de ejemplo ……. 137
Tabla 4.3 Resumen de localidades geográficas ingresadas al sistema …… 137
Tabla 4.4 Resumen de registros de tiempo ingresados al sistema 137
Tabla 4.5 Resumen de registros de variables de edificaciones ingresadas al sistema ……. 138
Tabla 4.6 Resumen de registros de edificaciones ingresados al sistema ……. 139
Tabla 4.7 Resumen de registros de series de indicadores ingresados al sistema …. 139
Tabla 4.8 Descripción de casos de prueba o predicción …….. 140
Tabla 4.9 Parámetros del algoritmo genético ingresados para los casos de estudio 140
Tabla 4.10 Descripción de estructura de las reglas de asociación a descubrir para el caso de estudio ……. 140
Tabla 4.11 Parámetros de ejecución del algoritmo genético aplicado a los casos de prueba …….. 141

Pág

Tabla 4.13 Aceptación de hipótesis para comparar proporciones …… 144
Tabla 4.14 Parámetros de contraste de hipótesis aplicados a las reglas descubiertas 145
Tabla 4.15 Resultados de aplicación de contraste de hipótesis a reglas descubiertas del tipo 1 ……. 146
Tabla 4.16 Resultados de aplicación de contraste de hipótesis a reglas descubiertas del tipo 2 …….. 147
Tabla 4.17 Resultados de aplicación de contraste de hipótesis a reglas descubiertas del tipo 3 …….. 148

ÍNDICE DE FIGURAS

Pág.

Figura 1.1 Número de Permisos por Vivienda ……. 29
Figura 1.2 Índice de metros cuadrados a construirse …… 29
Figura 2.1 Operadores Genéticos ……. 41
Figura 2.2 Modelo generalizado para una neurona ……... 44
Figura 2.3 Arquitectura general de un Sistema Basado en casos (SBC)……... 48
Figura 2.4 Algoritmo genético aplicado al descubrimiento de reglas 54
Figura 2.5 Variación del algoritmo genético aplicado al descubrimiento de reglas.. 55
Figura 3.1 Diagrama del modelo de objetos ……. 72
Figura 3.2 Escenario para introducir casos de predicción al sistema …… 77
Figura 3.3 Escenario para Procesamiento de caso de predicción …….. 78
Figura 3.4 Escenario para ejecución del algoritmo genético …… 79
Figura 3.5 Diagrama de estados para clase caso de predicción …… 80
Figura 3.6 Diagrama de estados para clase individuo …….. 80
Figura 3.7 Diagrama de Flujo de datos …… 81
Figura 3.8 Flujo de Control de Algoritmo Genético …… 84
Figura 3.9 Caso de Uso del Sistema de Predicción y Análisis …… 85
Figura 3.10 Diagrama de Clases del módulo Edificación ……. 86
Figura 3.11 Diagrama de Clases del módulo Indicador Económico ……. 87
Figura 3.12 Diagrama de Clases del módulo Predicción …….. 88
Figura 3.13 Diagrama de Clases del módulo Algoritmo Genético …… 89
Figura 3.14 Representación genética del antecedente de un individuo Cromosoma-Regla …… 90
Figura 3.15 Activación de atributos del antecedente de un individuo Cromosoma-Regla …… 91
Figura 3.16 Pseudocódigo del Algoritmo Genético …… 95
Figura 3.17 Diagrama de Arquitectura del Sistema SPACV …….. 96
Figura 3.18 Barra de herramientas para mantenimiento ……. 101
Figura 3.19 Iconos de acceso directo a opciones del sistema ……. 102
Figura 3.20 Menú principal de opciones del sistema …….. 102
Figura 3.21 Formato de Pantalla ……. 103
Figura 3.22 Diseño de Ventana de acceso (Login) …… 104
Figura 3.23 Diseño de Ventana Mantenimiento / Presentación de datos …… 104
Figura 3.24 Diseño de Ventana Petición de Parámetros ……. 105
Figura 3.25 Diseño de Ventana de petición de parámetros de impresión ……... 105

Pág

Figura 3.27 Diseño de Ventana Interfaz de Casos de Predicción: Análisis..... 106
Figura 3.28 Diseño de Ventana Interfaz de Casos de Predicción……..... 106
Figura 3.29 Diseño de Ventana Interfaz Sección Predicción AG: Parámetros 107
Figura 3.30 Diseño de Ventana Interfaz Sección Predicción AG: Proceso 107
Figura 3.31 Diseño de Ventana Interfaz Sección Predicción AG: Resultados ……... 108
Figura 3.32 Código de creación de la Base de Datos SPACV..... 112
Figura 3.33 Código de creación de la Tabla Caso_Predicción 112
Figura 3.34 Código de agregación de claves foráneas de tiempo y localidad en la tabla Caso_Predicción 112
Figura 3.35 Usuarios configurados en la base de datos SPACV a través del administrador ……... 113
Figura 3.36 Desarrollo de la aplicación SPACV – Creación de Librerías y objetos principales …… 117
Figura 3.37 Fases de Integración de módulos ……..... 130
Figura 3.38 Zona de aceptación para contraste unilateral de hipótesis …….. 143

ÍNDICE DE FÓRMULAS

Pág.

Fórmula 2.1 Desorden promedio…….. 42
Fórmula 2.2 Distancia euclídica ……... 46
Fórmula 2.3 Función de adaptación generalizada para A.G descubrimiento de reglas 51
Fórmula 2.4 Función de adaptación ponderada para A.G descubrimiento de reglas ... 52
Fórmula 2.5 Función de adaptación con reducción de error …… 52
Fórmula 2.6 Soporte y confianza de evaluación de reglas de asociación …… 53
Fórmula 3.1 Expresión Resultado del cruce …… 93
Fórmula 3.2 Expresión de adaptabilidad del individuo ……... 94
Fórmula 4.1 Estimado de proporción …….. 143
Fórmula 4.2 Estadístico Z …… 143
Fórmula 4.3 Promedio ponderado entre proporciones ……. 143

CAPITULO 1.
EL COMPORTAMIENTO DE LA CONSTRUCCIÓN DE VIVIENDAS EN EL PAIS

1.1. DESCRIPCIÓN GENERAL DEL FENÓMENO DE LA CONSTRUCCIÓN DE VIVIENDAS EN EL PAÍS.

1.1.1. GENERALIDADES.

El problema de la vivienda es un elemento clave de diagnóstico para medir el grado de crecimiento socio-económico de un país.

Conocer las tendencias actuales en la conformación de las viviendas, así como también los materiales que predominan en las construcciones, el uso de la edificación y sus fuentes de financiamiento, es tarea primordial para plantear retos hacia el futuro, especialmente aquellos que tienen que ver con dotar de una vivienda digna a los habitantes.

El Ecuador, se encuentra actualmente en un proceso de consolidación de las tendencias generales de urbanización de la economía, como resultado de la modernización capitalista que se inició con el proceso acelerado de producción y exportación petrolera. La construcción de viviendas, en este escenario, requiere de una mayor atención para llegar a una industrialización efectiva, que utilice poca inversión y estimule el ahorro. ([1]) ([2])

Descubrir las reglas de comportamiento que tienen las construcciones actuales, es, pues, fundamental para proyectar las tendencias positivas de crecimiento económico del sector hacia el futuro.

---

[1] CARRION FERNANDO., **op.cit**.
[2] DÁVALOS SOTELO R,, RICALDE M., **op.cit**.

1.1.2. PRINCIPALES FACTORES QUE EXPLICAN EL COMPORTAMIENTO DE LA CONSTRUCCIÓN DE VIVIENDAS EN EL PAIS.

El comportamiento de la construcción de viviendas en el país, en lo que se refiere a los cambios estructurales de las edificaciones, su uso, origen y financiamiento, depende de factores físicos o naturales, demográficos, humanos y económicos. Algunos son coyunturales y reflejan la realidad del momento, otros muestran su permanencia en el tiempo y sólo pueden ser explicados a través de un conocimiento profundo de la realidad, resultado de años de investigación.

1.1.2.1 Factores del medio físico y natural.

La ubicación geográfica de una construcción o vivienda, es decir, el lugar en el cual está situado el terreno, en conjunción con la benignidad del clima y otras ventajas como la calidad del suelo y el acceso a recursos naturales, hacen que tanto los materiales empleados como el uso de la edificación se vean afectados.

Los resultados de los censos de población y vivienda recientes indican que la mayor concentración de viviendas en el país se produce en las zonas urbanas y especialmente en las ciudades, puesto que el acceso a los servicios, el atractivo urbano, el grado de accesibilidad y proximidad con las fuentes de trabajo, además, las condiciones de salud y educación que se ofrecen, son factores decisivos en la elección del lugar para conseguir y construir vivienda, sin dejar de lado la enorme influencia del medio físico y natural.

La superficie total del país es de 256,370 km^2, de los cuales la región costa con sus 15,954 km^2, es la que se ha caracterizado por cubrir la mayor parte del área a construir del país. Del total nacional de superficie a construir proyectada para el año 2005, estimada en 5,103,144 m^2, en la que se incluye área de planta baja y pisos, solamente para la provincia del Guayas se han declarado 1,399,625 m^2 en área a construir.

Las condiciones ambientales y de accesibilidad son un factor determinante en las construcciones de la región amazónica, ya que con una superficie de 17,947 km^2, superior al de las regiones costa y sierra, su mayor área a construir se encuentra en la provincia de Morona Santiago, con apenas 54,207 m^2, según la declaración de permisos de construcción del año 2005.

1.1.2.2 Factores demográficos y humanos.

Los cambios estructurales en la composición de la población, el crecimiento demográfico, las diferencias sociales y culturales, los recursos de las familias y las motivaciones de las personas a elegir su vivienda, las ofertas de las compañías constructoras y los planes de vivienda adaptados a sus preferencias, hacen que el desarrollo humano sea el eje central sobre el cual gira la industria de la construcción.

Se conoce que el desarrollo urbano en el Ecuador tiene como ejes potenciales a las ciudades de Quito, Guayaquil, Cuenca y Machala, ya que son las más pobladas y además el 61% de la población nacional corresponde al área urbana. El crecimiento poblacional alcanza actualmente el 1.74%. (INEC, CEPAR: Estimación y proyecciones de población). ([3]),([4]).

La población urbana crece más aceleradamente que las proyecciones realizadas, ya que influyen otros factores coyunturales en el crecimiento poblacional y urbano con diferencias marcadas entre ciudades de la Sierra y la Costa. Mientras las ciudades intermedias tienen tasas bajas de crecimiento, las secundarias en su mayoría tienen crecimiento negativo que se explica por la inmigración interna y el acelerado crecimiento de los centros poblados.

La conformación de las viviendas obedece a la composición de los hogares en el Ecuador y puede clasificarse en dos grandes grupos: *Unifamiliares y*

---

[3] CARRION, F., **op.cit.**, p 170.
[4] VILLAVICENCIO, G , **op.cit**, pp 2-10

Multifamiliares. El primer grupo comprende la categoría de viviendas para una familia y el segundo comprende las viviendas para dos, tres o más familias.

20 CIUDADES MÁS POBLADAS					
Ciudad	**Hab.**	**Ciudad**	**Hab.**	**Ciudad**	**hab.**
Guayaquil:	1.985.379	Ambato	154.095	Babahoyo	76.869
Quito:	1.399.378	Riobamba	124.807	Sangolquí	56.794
Cuenca:	277.374	Quevedo	120.379	Latacunga	51.689
Machala:	204.578	Loja	118.532	Tulcán	47.359
Santo Domingo:	199.827	Milagro	113.440		
Manta	183.105	Ibarra	108.535		
Eloy Alfaro (Durán)	174.531	Esmeraldas	95.124		
Portoviejo	171.847	La Libertad	77.646		

Tabla 1.1. Ciudades más pobladas. Fuente : INEC: 2001 VI Censo Nacional de Población y V. De Vivienda.

El tamaño familiar varía entre las distintas zonas y regiones del país. Así en las zonas urbanas, las familias son menos numerosas (4.3 miembros), especialmente en la región de la sierra, mientras que, en las zonas rurales dispersas, el tamaño familiar promedio se eleva a casi cinco miembros por hogar, e incluso más en la región del oriente del país

1.1.2.3. Factores económicos

La conformación de las viviendas, el predominio de un material sobre otro, los cambios en el uso dado a la edificación y el origen del financiamiento, pueden explicarse en términos de crecimiento económico.

El crecimiento de los indicadores económicos no necesariamente posibilita el desarrollo a todos los estratos sociales. Según el Instituto Nacional de Estadística y Censos, el déficit de vivienda llega a 1,4 millones de unidades. Aún no se lo puede cubrir aún, debido a que los costos financieros son muy altos para la mayoría de la población por las altas tasas de interés y la inflación. Según la Cámara de Construcción de Guayaquil, el país necesita no menos de 70’000.000

m^2 de viviendas (déficit) y no menos de 4'000.000 m^2 anuales para evitar que crezca el déficit. ([5])

VIVIENDAS OCUPADAS, DESOCUPADAS Y EN CONSTRUCCIÓN.				
AREA	TOTAL VIVIENDAS	OCUPADAS	DESOCUPADAS	EN CONSTRUCCIÓN
TOTAL	3.456.103	3.027.249	314.705	109.071
URBANA	2.033.560	1.881.638	88.180	60.581
RURAL	1.422.543	1.145.611	226.525	48.490

Tabla 1.2. Fuente : INEC 2001 ; Viviendas ocupadas, desocupadas.. , VI Censo de Población y V de Vivienda.

El déficit de viviendas se divide en déficit cuantitativo (en cantidad) y déficit cualitativo (en calidad), siendo mayor el primero. El déficit cuantitativo ha llegado a 850.000 unidades y el déficit cualitativo está en las 350.000 unidades, lo que demuestra que hay mayor necesidad de producción de viviendas, que la de mejorar las condiciones de las existentes (ampliaciones y reconstrucciones). Además, cerca del 50% de la población urbana se encuentra en situación deficitaria de servicios. ([6]) ([7]) ([8])

Al reducirse las familias de altos ingresos, también se reduce la conformación de sus viviendas, lo cual confirma la regla de que a mayores ingresos, tienden a ser más reducidas las familias, el hacinamiento es menor y la capacidad de compra de vivienda es mayor. ([9]) .

La demanda real por vivienda está determinada por la capacidad de pago de las familias y el valor de la vivienda que se encuentra en el mercado. La capacidad de compra de vivienda por parte de las familias también depende de la oferta de vivienda en el mercado. ([10]).([11]) ([12])

---

[5] CIDEIBER , **op.cit**.
[6] BID., **op.cit** p.18 ; INEC., **Encuesta de Condiciones de Vida** (ECV), **op.cit**.
[7] CARRION FERNANDO., **El Proceso Urbano en el Ecuador**, **op_cit**. p.125
[8] INEC. **Censo de Población y Vivienda**; BID., **op.cit** p. 18
[9] OMS, **op.cit**. pp 12-13
[10] OIT., **op.cit**. pp. 35-41.
[11] VASQUEZ. L, **op.cit**. p.316
[12] BID., **op. cit** . p 19.

La mayor fuente de financiamiento para vivienda es la que corresponde a recursos propios de particulares, con valores superiores al 80%. En los últimos años, gran parte de los recursos propios provienen de familias de emigrantes que han llegado a estabilizarse en el exterior o regresan al país para destinar la mayor parte de sus ahorros a vivienda. Las remesas recibidas de los emigrantes constituyen actualmente uno de los puntales de sostenimiento de la economía nacional, ocupando el segundo lugar en el ingreso nacional, sólo después del petróleo. ([13]). Las viviendas construidas por los emigrantes en el país además, están revolucionando la conformación tradicional en el número de cuartos, dormitorios y número de pisos con mayor predominio de madera y otros materiales en la construcción, especialmente en la estructura y las paredes.

En cuanto al financiamiento a crédito, la banca ha tenido un papel protagónico, aunque la mayor parte se ha destinado al consumo en detrimento al sector productivo nacional. Para el año 2004, apenas 269 millones de dólares son para la construcción, mientras que para el consumo en tarjetas de crédito se ha destinado 2,863 millones. ([14]).

Aunque en décadas anteriores, el gobierno asumió por varios años el crédito otorgado a vivienda, como promotor, constructor y actor financiero, su participación actual es mínima, menos del 50% corresponden a viviendas producidas con intervención del sector público. ([15])([16]) ([17]).

En promedio, desde el año 2000 se ha destinado un valor cercano al 7.% del PIB a la Industria de la Construcción. La economía del Ecuador entre 1990 y 1997 creció a una tasa anual promedio de 3.2%, mientras que en el período 1997-2000 se redujo en 1.9% promedio anual.

La inflación incide directamente en el incremento de los precios, especialmente en los precios de los materiales predominantes de las construcciones, y por ende, en

---

[13] ACOSTA, A, **op.cit.** p. 10
[14] VASQUEZ, L., **op.cit**., p.261.
[15] BID, **op.cit**. p 17-18
[16] O.I.T., **op.cit** p. 46
[17] INEC, **Encuesta ECV**., **op.cit**.

el precio del metro cuadrado de construcción. La inflación llegó a su máximo nivel en el 2000, durante la dolarización, llegando a 107.9%. Y desde fines de ese año la tendencia cambia produciéndose una desaceleración que continúa desde 22.4% en el 2002 hasta llegar al 2.3% en el 2005. Esta inflación aún es superior a la de economías vecinas sin dolarizar. ([18]) ([19])

La deuda externa es la que más ha afectado la economía ecuatoriana en los últimos años con detrimento del desarrollo del sector productivo y la industria, incluida la construcción. Los desembolsos por deuda superan al ingreso, y el monto de la deuda va en aumento. ([20])([21])

Todas las actividades económicas, especialmente el sector de la construcción se ven afectadas por las crisis económicas. Entre 1995 y 2000 creció el número de pobres de 34% al 71%, y la pobreza extrema dobló su número Las consecuencias de esta crisis fueron el masivo desempleo y subempleo; la caída de los ingresos y posibilidad de inversión en vivienda, por la caída vertiginosa de la confianza en el país. ([22]).

Sin embargo, en épocas de recuperación de crisis producidas por desastres naturales es posible que el crecimiento del sector de la construcción experimente un ligero crecimiento, tal es el caso del fenómeno de El Niño, que en 1998 tuvo un crecimiento del 6%, tras reactivarse la actividad de la construcción. ([23]).

En los últimos años, (2004 y 2005) la economía ecuatoriana registra notables signos de recuperación. La vigencia de la dolarización, junto al mejoramiento que presenta el balance fiscal, posibilita estabilizar la economía y restablecer la confianza de los agentes económicos. La recuperación del crédito ha sido notoria en áreas de consumo duradero y de vivienda ([24])

---

[18] INEC., **Indice de Precios de la Construcción (IPCO) op.cit**.
[19] VASQUEZ, L, SALTOS, N. **op.cit** . p.315.
[20] BCE, Banco Central del Ecuador, **op.cit**.
[21] ACOSTA, A., **op.cit** p.7
[22] ACOSTA A, **op.cit** p2.
[23] CIDEIBER , **op.cit**.
[24] BID., **op.cit**. p 16.; OLEAS, J, **op.cit**.

1.2 DESCRIPCIÓN Y ANÁLISIS DE LAS VARIABLES INVESTIGADAS.

1.2.1 DESCRIPCIÓN DE LAS VARIABLES INVESTIGADAS.

1.2.1.1 Descripción de las variables contenidas en los permisos de construcción para edificación de viviendas.

La información básica que se dispone sobre los casos de construcción de viviendas es la que surge directamente de los permisos de construcción aprobados por los municipios del país a los proyectos de construcción, reconstrucción o ampliación de edificaciones y que son registrados en la encuesta de edificaciones cuyos resultados se recopilan por meses y las publica cada año el Instituto Nacional de Estadística y Censos (INEC).

Los resultados de la encuesta proporcionan una medida del ritmo de crecimiento de la construcción, la inversión y fuentes de financiamiento en el sector de la construcción, además permiten conocer el número de viviendas ofrecidas, su costo unitario, su superficie y los materiales utilizados.

Sin embargo, cabe señalar que los resultados de esta encuesta no reflejan el número de construcciones efectivas, ya que en algunos casos estas no se llevan a cabo y además no se registran aquellas construcciones que se efectúan sin el permiso correspondiente. Por lo tanto los casos de construcción de viviendas están limitados a las construcciones formales y en proyecto de construirse. ([25])

La información básica de los permisos de construcción es la siguiente:

---

[25] INEC., **Anuario Estadístico del Ecuador 2003** , **op.cit**. p. 59.

Localidad geográfica (provincia, cantón y parroquia):

Cada caso de construcción en los permisos declarados, está circunscrito a una localidad geográfica. La localidad se subdivide en provincia, cantón y parroquia de acuerdo a la división política territorial administrativa.

Tiempo (mes y año de inscripción):

En la fecha de registro del permiso, se declara el mes y el año, llevándose a cabo el proceso de recolección de la información durante todo el año.

Propiedad de la edificación:

La propiedad puede ser privada o pública. La edificación es obra privada cuando el propietario original de la misma es una persona privada (natural o jurídica) y se la ha financiado con recursos del sector privado y/o público. La edificación es obra pública cuando el propietario original y quien financia es el Gobierno Central, Municipio, Consejo Provincial, entidades estatales como el IESS, ISFA, BEV, BEDE, Banco Central, etc.

Tipo de obra:

Un permiso de construcción puede ser solicitado para nuevas construcciones, reconstrucciones o ampliaciones. El tipo de obra es nueva construcción cuando la obra de ingeniería se construye desde sus cimientos, o se ha derrocado una construcción y al mismo tiempo se levanta otra edificación desde sus cimientos. Las reconstrucciones son obras que no afectan ni a cimientos, ni a estructura de la edificación. Puede considerase como reconstrucción a la reparación de la cubierta, trabajos inferiores como la eliminación o reparación de las paredes. La información sólo debe referirse a la labor a realizarse.

Para la ampliación, se especifican dos tipos: ampliación hacia arriba y horizontal. Si la ampliación es hacia arriba, solo se informa el dato de área a construirse, que

corresponde exclusivamente a la parte nueva. No se informa cimientos. En el resto de variables se informa todo lo que corresponda. La ampliación es horizontal cuando existe comunicación directa e interior entre la parte antigua y la nueva. En este caso viene registrada toda la información. En cambio, si entre la construcción existente y la nueva, no existe comunicación directa e interior, la información debe entenderse como nueva construcción.

Superficie del terreno en m^2:

La superficie total del terreno se refiere al tamaño (en m2) del lote. Esta no puede ser menor al área de planta baja. No se consideran áreas destinadas a agricultura.

Área de construcción en m2:

El área total a construirse es la suma de las superficies horizontales (pisos) que tiene la edificación. El área a construirse en planta baja es el área sobre la cual se asienta la edificación. El área para espacios verdes o patio(s) es el área destinada a jardines y espacios recreacionales, como patios, canchas. En el caso de edificaciones en el área rural, se excluye la superficie destinada a producción agrícola como huertas.

Materiales predominantes a ser utilizados en la edificación:

Estos se especifican para cimientos, estructura, paredes y el techo o cubierta.

USO RESIDENCIAL DE LA EDIFICACIÓN	
TIPO DE USO	**CATEGORÍA**
Residencial Unifamiliar	Residencia para una familia
Residencial Multifamiliar	Residencia para dos familias
Residencial Multifamiliar	Residencia para tres o más familias

Tabla 1.3. Uso Residencial de la Edificación. Fuente: INEC

VARIABLES PRINCIPALES CONTENIDAS EN LOS PERMISOS DE CONSTRUCCIÓN						
TIPO DE OBRA	USO DE LA EDIFICACION	ORIGEN DEL FINANCIAMIENTO	MATERIALES PREDOMINANTES EN LA CONSTRUCCION			
			Cimientos	Estructura	Paredes	Cubierta
Nueva Construcción	Unifamiliar	Por recursos propios	Hormigón armado	Hormigón armado	Ladrillo	Hormigón armado
Ampliación	Multifamiliar	A crédito	Hormigón ciclópeo	Metálica	Bloque	Eternit, ardex asbesto-cemento.
Reconstrucción			Sobre pilotes de madera	Madera	Madera	Teja
			Sobre pilotes de hormigón	Otros	Adobe o tapia	Zinc
			Otros		Caña revestida	Otros
					Prefabricada	
					Otros	

Tabla 1.4. Variables principales contenidas en los permisos de construcción. Fuente: INEC

Uso y otras características de la edificación (número de edificaciones a construirse, viviendas, cuartos, dormitorios, pisos y área total a construirse en m2):

Los usos de la edificación son tanto para obras residenciales como para obras no residenciales, pudiendo haber construcciones mixtas.

Para uso residencial de la edificación se especifican: número de viviendas, número de cuartos, y el número de dormitorios. Tanto para uso residencial como para no residencial se tienen número de unidades a construir, número de pisos y el total de área a construir en metros cuadrados. Si ambos usos comparten el mismo piso, éste se menciona en los datos de uno de los dos.

Origen, valor y destino del financiamiento:

El origen del financiamiento se subdivide en recursos propios y a crédito, y tiene diferentes categorías dentro de estos dos grupos.

USO NO RESIDENCIAL DE LA EDIFICACIÓN	
TIPO DE USO	**CATEGORÍA**
No Residencial	Comercial (Almacenes, supermercados, bodegas, tiendas, edificios para oficinas y / o consultorios, edificios de paqueamiento, hoteles, restaurantes)
No Residencial	Industrial o artesanal (Naves industriales y sus oficinas administrativas, Bodegas, Talleres, etc)
No Residencial	Edificio Administrativo (De propiedad del Estado para funcionamiento de entidades públicas como ministerios, institutos, bancos, etc.)
No Residencial	Educación (Centros de Enseñanza, escuelas de formación básica, media y superior)
No Residencial	Cultura (Casa de la Cultura, Museos, Bibliotecas, Teatros, etc.)
No Residencial	Edificaciones en Complejos Recreacionales, Estadios, Coliseos cerrados (Excluye canchas deportivas al aire libre)
No Residencial	Hospitales, clínicas y otros establecimientos de salud
No Residencial	Transportes y Comunicaciones (Terminales terrestres de ferrocarriles, edificaciones en aeropuertos, centrales de teléfonos, telégrafos, télex, etc.
No Residencial	Iglesias, Templos, Conventos y afines.
No Residencial	Garaje (Si se encuentra dentro del edificio, como subsuelo, o si estando fuera del edificio, tiene cubierta.
No Residencial	Otros

Tabla 1.5. Uso No Residencial de la Edificación. Fuente: INEC

ORIGEN DEL FINANCIAMIENTO	
TIPODE RECURSO	**CATEGORÍA**
Recursos Propios	Recursos Propios de particulares
Recursos Propios	Recursos propios del IESS o Seguro de las Fuerzas Armadas
Recursos Propios	Recursos propios del Banco de la Vivienda
Recursos Propios	Recursos propios de las Mutualistas
Recursos Propios	Recursos propios de Construcciones Privadas
Recursos Propios	Recursos propios del Sector Público
Crédito	Préstamo del IESS o Seguro de las Fuerzas Armadas
Crédito	Préstamo de las Mutualistas
Crédito	Préstamo de Otros Bancos
Crédito	Préstamo de Otras Financieras
Crédito	Otros Préstamos
Crédito	Otros Recursos

Tabla 1.6. Origen del Financiamiento. Fuente: INEC

Los datos provenientes de la encuesta se complementan con información calculada, tal como el valor promedio del metro cuadrado a construir, resultado de dividir el área a construir del valor total de la edificación. Este valor permite también calcular el índice de metros cuadrados a construirse, en cuyo cálculo se agrupa y relaciona la información obtenida a partir del total de superficie a construir del año actual con un año base.

1.2.1.2. Descripción de las relaciones existentes entre las variables contenidas en los permisos de construcción para edificación de viviendas.

Entre las variables contenidas en los permisos de construcción se hallan presentes diferentes relaciones. Algunas son de sumarización, otras son de asociación. Pueden además existir relaciones desconocidas en la información, fruto del comportamiento del fenómeno.

Relaciones de sumarización.

La superficie total del terreno debe ser igual a la sumatoria del área planta baja y el área de espacios verdes.

El total de pisos de la edificación esta dado por la sumatoria del número de pisos del área residencial y el número de pisos del área no residencial.

El área total a construirse es equivalente a la suma del área a construir residencial y el área a construir no residencial.

El valor total de la edificación esta dada por la sumatoria de los valores de edificación de cada unidad, correspondientes al valor del terreno y el valor de obra urbanizada.

Relaciones de asociación conocidas:

Cuando se trata de una edificación de propiedad pública, el origen de financiamiento solo puede provenir de: recursos propios o del IESS o ISFFA y recursos propios del Banco de la Vivienda, En el caso inverso, cuando se trata de una obra de propiedad privada, se tiene cualquier origen del financiamiento excepto los considerados para obra pública.

Cuando la ampliación es hacia arriba hay dato únicamente en superficie del terreno, y cuando es una ampliación horizontal, la superficie total del terreno nuca sobrepasa los 40 m^2 y es igual al área de planta baja, sin área de espacios verdes. Si el tipo de obra es reconstrucción, no hay valor en cimientos ni estructura.

Generalmente, la superficie total del terreno no sobrepasa los 3501 m^2 en área residencial y los 10000 m^2 en área no residencial. El área de planta baja es siempre menor o igual a la superficie del terreno y al área total a construir. Siendo el total a construir, en la mayoría de los casos, de alrededor un 30% del producto del área planta baja por el número total de pisos.

Las áreas de espacios verdes y/o patio, generalmente no sobrepasan los 3000 m^2 tanto en área residencial como en área no residencial.. Se excluyen casos de nuevas construcciones de uso industrial.

Cuando la estructura predominante es de hormigón armado o metálica, por lo regular no se tiene cimientos de madera u otros, ni tampoco cubierta de zinc u otros. Aunque el caso de la cubierta, no se descarta que puedan aparecer otros materiales compuestos de madera reforzada.

Cuando la estructura predominante es de madera o de categoría otros, los cimientos ni la cubierta, por regla general, no son de hormigón. Aunque existen unos cuantos casos de construcciones modernas que utilizan estructuras de maderas reforzadas si sus cimientos son de hormigón.

Cuando el uso de la edificación es residencial para una familia, se tiene siempre el valor de 1 en número de viviendas. Si es residencial para dos familias, el valor del número de viviendas es 2, y es mayor o igual a 3 cuando el uso es residencial para tres o más familias.

El número de cuartos, en la mayoría de los casos no sobrepasa a 12 cuando el uso de la edificación es residencial para una familia, no supera los 24 cuando es residencial para dos familias, y cuando es residencial para tres o más familias, no hay restricción y por lo general es mínimo 6.

Con el número de dormitorios sucede algo similar, ya que su número está entre 1 y 6 para casos de residencias para una familia; entre 2 y 20 para residencias de 2 familias y supera los 6 si es residencia para tres o más familias.

El área a construir en residencia para una familia generalmente no llega a 350 m^2.

El número de pisos proyectado para uso de residencia de una familia, particularmente está entre 1 y 2. No se descarta las construcciones modernas, tipo americano con 3 o más pisos.

El área de garaje por lo general no sobrepasa los 1001 m^2 y es mayor a 8 m^2, en usos residenciales. Pueden considerarse parquemientos para uso comercial a los garajes que superen este límite. El total de pisos en localidades urbanas, generalmente no supera a 25.

Respecto al valor del metro cuadrado de construcción, resultado de dividir el valor de la edificación entre el total del área a construir, por lo general da como resultado valores no menores a $20 ni mayores a $600 de acuerdo a la ubicación geográfica. Sin embargo hay localidades urbanas que superan este valor considerablemente, llegando hasta los $1,000 el m^2. ([26])

[26] INEC - **Encuesta anual de edificaciones**, **op.cit,** Años 2000-2005

1.2.2 ANÁLISIS DE LAS VARIABLES INVESTIGADAS.

1.2.2.1 Generalidades sobre el análisis de las variables investigadas.

La construcción de modelos de predicción requiere de un estudio previo del comportamiento de las variables independientes o explicativas del fenómeno frente a las variables dependientes o predictivas. Con este fin se debe buscar los mecanismos y técnicas que nos permitan encontrar relaciones entre las variables y de esta manera construir el modelo. Las variables, una vez asociadas de manera inteligente, pueden revelar secretos ocultos que conduzcan a explicar y pronosticar el fenómeno ([27])

Para analizar el comportamiento de las variables continuas pueden aplicarse parámetros de series de tiempo descritos a continuación:

Comportamiento: Se refiere en general al conjunto de datos históricos observados en diferentes períodos de tiempo y a su manera de reflejar un determinado movimiento, que puede ser regular, sistemático (con variaciones cada ciertos tiempo) o irregulares. ([28])

Tendencia: Muestra el desplazamiento gradual de la serie de tiempo y que por lo común es el resultado de factores a largo plazo, como cambios en las políticas económicas, crisis o desastres naturales, o variaciones en otros aspectos demográficos, de mercado, tecnológicos, etc. ([29])

Ciclo: Cuando hay una secuencia de puntos arriba y debajo de la línea de tendencia y que se repiten a través de cierto tiempo, se dice que se tiene un componente cíclico de la serie. El ciclo significa cambios de tendencia. ([30])

---

[27] MAKRIDAKIS - WHEELWRIGHT, **op.cit**. p.61.
[28] PEREZ BELLO, C , **op.cit**. p.5.,
[29] ANDERSON D., SWEENEY D., **op.cit**. pp. 756-759.
[30] PEREZ BELLO, **op.cit**., p.6.; ANDERSON D., SWEENEY D, **op.cit**. p. 758.

Estacionalidad: Cuando existen movimientos en las series de tiempo que ocurren de nuevo por la misma época del año se tiene este nuevo componente de la serie conocido como estacionalidad. Para que haya este componente los datos deben recolectarse trimestral, mensual o incluso semanalmente. ([31])

Aleatoriedad: Cuando hay variaciones irregulares o aleatorias producidas por sucesos inusuales y que producen movimientos sin un patrón discernibles. Pueden ser producidas por eventos a corto plazo e imprevisibles tales como guerras, inundaciones, terremotos, cambios políticos, crisis económicas, etc. ([32])

Para las variables discretas se pueden usan dos estrategias: la primera consiste en determinar en la distribución porcentual. La segunda consiste en realizar cruces y asociaciones con otras variables. Estas estrategias pueden aplicarse también a las variables continuas, previo un proceso de discretización.

1.2.2.2 Análisis del comportamiento del número de permisos de construcción y el índice de metros cuadrados a construirse.

Su comportamiento es irregular con variaciones marcadas de 1992 a 1999 y de tendencia ascendente en los cinco años siguientes. De 1990 a 1993 los datos tienden a ascender, pero en ese año sufren una desaceleración con un ligero ascenso en el año 1994, la desaceleración continúa hasta 1997 y a partir de ese año vuelven a ascender, ascenso que se mantiene hasta el último año con una ligera pero marcada baja en el año 1999. Los datos observados no presentan ciclo, pues los cambios son esporádicos y se deben a un factor de aleatoriedad de la serie. Los movimientos que experimenta la variable podrían deberse a muchas causas ocultas. La tendencia a disminuir los permisos para viviendas residenciales de 1993 a 1997 pudo deberse a la crisis económica agravada por el fenómeno de “El Niño”, que justamente se inicia en el año 1993.

---

[31] WEBSTER ALLEN L., **op.cit**. p. 414.
[32] ANDERSON D., SWEENEY D., **op.cit** . p 759.; WEBSTER ALLEN L., **op.cit**. p 414;

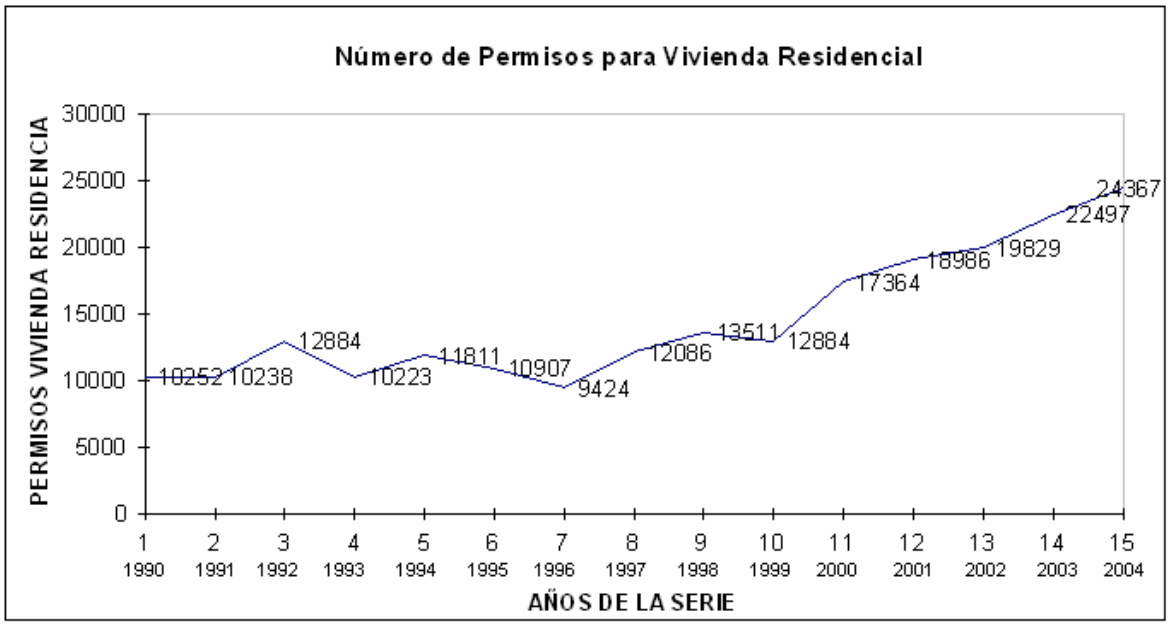


Figura 1.1. Número de permisos para vivienda residencial. Fuente: INEC. Encuesta de Edificaciones.

La repentina desaceleración hacia el año 1999 estaría relacionada con la crisis financiera bancaria y la marcada tendencia ascendente iniciada en el año 2000 puede deberse al fenómeno dolarización, las fuertes remesas de inmigrantes invertidas en la construcción de viviendas y la tendencia a bajar la inflación.

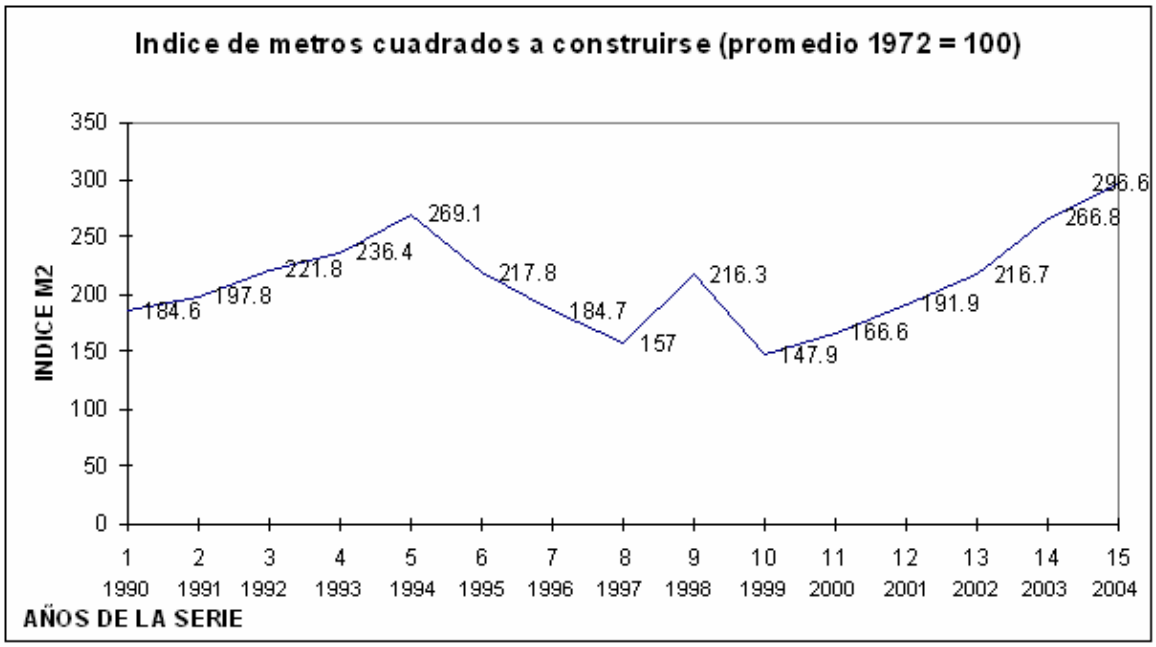


Figura 1.2. Indice de metros cuadrados a construirse. Fuente: INEC. Encuesta de Edificaciones

El comportamiento del índice de metros cuadrados a construir es similar. Durante la década pasada y los albores de este siglo ha sido ligeramente irregular aunque en cierta medida predecible respecto a los ascensos de inicios y fines del período que producen una línea de tendencia ascendente en todo el período.

1.2.2.3 Análisis del comportamiento de los materiales predominantes en las viviendas.

Un indicador de las condiciones físicas de la vivienda son los materiales con los que está hecha la vivienda, información que nos puede dar una idea de la calidad de la vivienda. ([33])

Al comparar los resultados de los censos de población y vivienda recientes, se tiene que en la Provincia del Guayas, como el resto del país, el material que aún predomina en el techo o cubierta de las viviendas, es el "zinc" con el 67,14 y el 63,15 por ciento, de 1990 al 2001 respectivamente; mientras que para Pichincha, mayor proporción tienen las viviendas cuyos techos o cubiertas son de "losa de hormigón" que ya llega al 56,06 por ciento en 2001.

VIVIENDAS PARTICULARES OCUPADAS CON PERSONAS PRESENTES, SEGÚN MATERIALES PREDOMINANTES EN TECHO O CUBIERTA, CENSOS 1990, 2001.								
MATERIALES PREDOMINANTES EN TECHO O CUBIERTA	GUAYAS				PICHINCHA			
	1990		2001		1990		2001	
Losa de Hormigón	80,658	15.73	146,722	19.14	182,711	45.87	342,320	56.06
Asbesto	62,522	12.20	98,204	12.81	58,718	14.73	68,130	11.16
Zinc	344,187	67.14	484,161	63.15	76,040	19.09	117,364	19.22
Teja	8,907	1.74	10,024	1.31	78,894	19.81	67,836	11.11
Paja	14,795	2.88	4,510	0.58	1,614	0.41	962	0.15
Otros	1,595	0.31	23,084	3.01	351	0.09	14,056	2.30
Total	512,664	100.00	766,705	100.00	398,328	100.00	610,668	100.00

Tabla 1.7. Material Predominante en Techo o cubierta, Censo 1990. Fuente: INEC.

[33] INEC , **op.cit**. p.53.,

En Pichincha se aprecia además que la "teja", la cual se mantenía en segundo orden de importancia con el 19,81 en 1990, ya para el 2001 ha sido desplazada por el "zinc" con el 19,22 por ciento. Esta tendencia se ha mantenido con cierta regularidad, si se toma en cuenta los resultados de la encuesta de edificaciones.

Para el año 2005, de los 7,662 permisos de construcción para Guayas, la losa de hormigón (55,05 %) ha reemplazado al zinc (12.88 %), de un total de 24,556 permisos a nivel nacional.

VIVIENDAS PARTICULARES OCUPADAS CON PERSONAS PRESENTES, SEGÚN MATERIALES PREDOMINANTES EN PAREDES EXTERIORES, CENSOS 1990, 2001.								
MATERIALES PREDOMINANTES EN LAS PAREDES	GUAYAS				PICHINCHA			
	1990		2001		1990		2001	
Hormigón	339,524	66.23	604,720	78.87	309,849	76.28	521,840	85.45
Adobe	2,319	0.45	3,769	0.49	67,648	16.98	49,102	8.04
Madera	34,773	6.78	57,491	7.50	19,987	5.02	23,653	3.87
Caña Revestida	37,114	7.24	78,693	10.26	2,254	0.57	3,025	0.50
Caña no revestida	97,519	19.02	9,252	1.21	,2661	0.67	1,024	0.17
Otros materiales	1,415	0.28	12,780	1.67	1,929	0.48	12,024	1.97
Total	512,664	100.00	766,705	100.00	398,328	100.00	610,668	100.00

Tabla 1.8. Material Predominante en las paredes exteriores. Fuente: INEC.

El "hormigón, ladrillo o bloque" es el material que predomina en las paredes exteriores de las viviendas particulares ocupadas con personas presentes en las provincias de Guayas, Pichincha y el resto del país ([34])

Se tiene además, que para el año 2005, de los 24,556 permisos de construcción concedidos, aún persiste la tendencia a utilizar el hormigón como material predominante. El 75.1% de las edificaciones han proyectado construir sus cimientos con hormigón; el 94.6% tendría su estructura de hormigón y el 51.6% emplearía hormigón armado en la cubierta o techo. En esta última fase de la construcción, no es tampoco despreciable el 27.9% de edificaciones que utilizarían el eternit, ardex o asbesto-cemento.

[34] INEC , **op.cit**. p.53.,

1.2.2.4 Análisis del comportamiento de tipo de obra, uso, origen y destino del financiamiento, superficie del terreno y área a construir.

Tipo de obra.

Para el año 2004, el 90 % de los permisos han sido otorgados para la ejecución de proyectos de nuevas construcciones, el 8% para ampliaciones y la diferencia para reconstrucciones. Respecto al año 2003, se nota que la tendencia porcentual de las nuevas construcciones se mantiene. De las nuevas construcciones, el 91% se han concedido para proyectos de uso residencial, lo que significa que la tendencia es exactamente igual a la registrada en el año 2003. Y respecto al año 2005 la tendencia porcentual de las nuevas construcciones aún se mantiene con el 88%, con un 9% para ampliaciones y la diferencia en reconstrucciones.

Usos y características de la construcción.

En el año 2004, 24,367 permisos, el 88.6 % del total, fueron otorgados para construcciones con fines residenciales, el 9.3 % para edificaciones no residenciales y el 2.1 % a edificaciones mixtas. En las edificaciones de uso mixto (residencial y no residencial), se construirán también 1.149 viviendas, lo que significa un total nacional de 33.994 soluciones habitacionales, que representan un 12.3 % más que en el año 2003. La tendencia es a aumentar las construcciones residenciales, ya que en el año 2005 el 90.4% se han otorgado permisos para fines residenciales, y solo el 6.6 % para edificaciones no residenciales.

Aunque el número de permisos de construcción concedidos para usos residenciales, en los últimos años ha alcanzado valores altos, el crecimiento de viviendas proyectadas es relativamente menor, pudiendo constatarse que para el año 2001, este crecimiento decayó en el 9% respecto al año anterior, posiblemente debido a la crisis financiera y a los efectos de la dolarización. ([35])

[35] INEC.-, **Encuesta Anual de Edificaciones**., **op.cit**.

El 8.7 % de los permisos residenciales fueron proyectados para edificaciones con dos unidades habitacionales que albergarían a dos familias, es decir 4.230 viviendas. El 88.7 % serían nuevas construcciones, el 29.7 % de las viviendas tendría cinco cuartos (se mantiene la preferencia respecto del 2003) y 48.9 % tres dormitorios.

Si se toma en cuenta el número de dormitorios de las viviendas, se advierte la tendencia a construir con tres dormitorios por vivienda. Además, dentro de los permisos de construcción de uso residencial, se alcanza altos porcentajes destinados a residencias para una familia, superiores al 80%, por lo general. Del total de nuevas construcciones, la tendencia es a tener entre tres y dos dormitorios en forma mayoritaria. Por ejemplo, la tendencia respecto del 2002 cambió con respecto a los años anteriores, ya que en ese año se prefirió dos dormitorios. ([36])

Financiamiento.

El financiamiento de las construcciones con recursos propios tiende cada año a aumentar. Para el año 2004, el 81.2 % de las edificaciones (22.345 permisos), se financió con recursos propios, de los cuales las personas naturales o particulares financiaban el 83.1 % de estos permisos de construcción. Y para el año 2005, se alcanza el 88.9 % en recursos propios, con un financiamiento de las personas particulares del 93.0%, que va en aumento.

Para el financiamiento a crédito, en el 2004 este ocupa el 18.8 % , y en año 2005 desciende al 11.1 %, manteniéndose en segundo orden de importancia al crédito otorgado por el Banco de la Vivienda.

[36] INEC., **Encuesta de Edificaciones**.-**op.cit**.,

FINANCIAMIENTO EN DOLARES AMERICANOS			
AÑOS	PROPIO	CRÉDITO	TOTAL
2001	64,488,349	2,267,163	66,755,512
2002	528,551,788	14,320,197	542,871,985
2003	724,300,788	171,98,627	741,499,415
2004	921,773,310	64,273,934	986,047,244
2005	878,405,901	39,544,323	917,950,224

Tabla 1.9. Financiamiento en viviendas proyectadas, Fuente: INEC.

ORIGEN DEL FINANCIAMIENTO (% valor de la edificación)			
Años	Recursos Propios	Crédito	
1996	83.2%	16.8%	53.3% (Otros Bancos)
1997	89.7%	10.3%	40.2% (Banco Ecuatoriano de la vivienda B.E.V)
1998	92.0%	8%	32.1% (Otros bancos)
1999	93.0%	7%	45.4% (Otros Financiamientos)
2000	94.6%	5.4%	34.5% (Otros Financiamientos)
2001	96.6%	3.4%	41.1% (Otros Financiamientos)
2002	97.4%	2.6%	42.9% (Otros bancos)
2003	86.4%	13.6%	37.2% (Banco de la Vivienda)+ 26.9% (Otras Financieras)
2004	81.2%	18.8%	41.2% Otros Bancos + 40.5% Banco de la Vivienda.
2005	88.9 %	11.1%	37.7% Otras financieras + 26.3% Banco de la Vivienda

Tabla 1.10. Origen del Financiamiento; Fuente: INEC. Encuesta de Edificaciones.

Superficie del terreno, área a construir y valor promedio del metro cuadrado.

De la superficie total del terreno, el porcentaje destinado a la construcción con fines residencias tiene una tendencia a incrementarse en los últimos años. Es así que en el año 2004 de los 7.890.434 m^2 de superficie total del terreno declarado en los permisos, se proyectó la construcción de 5.656.917 m^2. De éstos, el 70.1% fueron destinados para fines residenciales. Y de los 7,112,671 m^2 de superficie del terreno en el año 2005, se ha proyectado la construcción de 5,103,144 m^2, con el 76.1% destinados para fines residenciales, siendo el 23.6 % para usos no residenciales y el 0.3% para construir garajes.

Para el año 2004, la provincia de Pichincha declaró la mayor superficie total del terreno (25.1%) y la mayor superficie a construirse (34.4 %) siguiéndole en orden de importancia la provincia del Guayas. Para el año 2005 es la provincia del Guayas la que ocupa el primer lugar con 25.3 % del terreno y el 27.4% con la mayor superficie a construirse. Como puede apreciarse, son estos dos polos de desarrollo urbano los que año tras año se disputan protagonismo en ocupar la mayor superficie total a construir del país.

Si del total del valor del financiamiento proyectado se divide para el número de metros cuadrados a construirse, se obtiene el valor del metro cuadrado de construcción. El valor promedio para el año 2004, fue de 176 dólares, y el previsto para el año 2005 es de 182 dólares. Si se compara con el valor del año 2003 (153 dólares), se observa que hubo un incremento de $ 23, lo que en términos porcentuales equivale a un crecimiento del 15 % en el primer año, y del 3% en el segundo año. ([37])

[37] INEC., **Encuesta anual de Edificaciones**., **op.cit**. año 2004.

CAPITULO 2.
TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA REALIZAR PRONÓSTICOS

2.1 INVESTIGACIÓN BIBLIOGRÁFICA DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA REALIZAR PRONÓSTICOS.

2.1.1 GENERALIDADES SOBRE LA APLICACIÓN DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA REALIZAR PRONÓSTICOS.

2.1.1.1. Aplicación de Técnicas de Inteligencia Artificial a la resolución de problemas predictivos.

Las técnicas aplicadas a la resolución de problemas predictivos se pueden organizar en dos grandes grupos: el primero corresponde a las que realizan tareas de clasificación y el segundo tiene que ver con tareas de regresión.

TECNICAS APLICADAS A LA RESOLUCIÓN DE PROBLEMAS PREDICTIVOS			
Nombre	Campo de Investigación	PREDICCIÓN	
		Clasificación	Regresión
1. Redes Neuronales	I.A	X	X
2. Árboles de decisión	I.A.; TTD	X	X
3. Vecinos más próximos y	I.A; Estadística.	X	X
4. Algoritmos Genéticos y evolutivos	I.A	X	X
5 Redes bayesianas	I.A; TTD; Estadística.	X	
6 Sistemas Expertos	I.C; I.A	X	X
7 Regresión lineal, logarítmica y otras	Estadística		X
8. Análisis discriminante Multivariante	Estadística	X	

Tabla 2.1. Técnicas aplicadas a la resolución de problemas predictivos.([38])
(I.A = Inteligencia Artificial; TTD = Teoría para la Toma de Decisiones.; I.C = Ingeniería del conocimiento)

Para llevar a cabo la resolución de problemas predictivos se realizan una o varias tareas de minería de datos y se aplican métodos que permitan resolver dichas tareas. Entre las tareas más comunes se encuentran las de clasificación, regresión, agrupamiento y reglas de asociación; las cuales se realizan mediante la extracción de ejemplos sobre un caso particular de predicción.

[38] HERNANDEZ J, RAMIREZ J, FERRI C. **op.cit.**, p.148

Las tareas de clasificación consisten en extraer reglas de asociación entre las características conocidas de un determinado conjunto de datos y usarlas para hacer predicciones sobre la aparición de una característica en el conjunto.

Las tareas de regresión consisten en construir un modelo que permita simular la relación existente entre dos o más variables y utilizarlo para predecir el comportamiento de una de estas variables.

Las reglas de asociación intentan descubrir asociaciones o conexiones entre objetos. Una regla de asociación se expresa de la forma: **A ⟶ B** , donde los objetos A_1.....A_j, tienden a aparecer con los objetos B_1...B_i, dentro de un conjunto de datos.

Los objetivos de la regresión y la clasificación consisten en aprender una función: λ **: E ⟶ S**. , donde E es el conjunto de todos los posibles elementos de entrada, las instancias posibles dentro de E son el conjunto de valores. Mientras que S es el conjunto de valores de salida. La diferencia respecto a la clasificación es que S es numérico, es decir puede ser un valor entero o real. La regresión se conoce con otros nombres: ***interpolación*** (generalmente cuando el valor predicho está en medio de otros) o ***estimación*** (cuando se trata de algo futuro). En clasificación, la función λ se denomina clasificador, y representa a la correspondencia existente en los ejemplos, es decir, para cada valor de E hay un único valor de S. La salida S es nominal, esto quiere decir que puede tomar un conjunto de valores c_1, c_2 .... c_m denominados clases. La función aprendida será capaz de determinar la clase para cada nuevo ejemplo.

Cuando la clasificación se acompaña de una medida de certeza, estimación o fiabilidad de las predicciones, a ésta se la conoce como clasificador suave.

Los problemas de clasificación son similares a los de regresión en cuanto hay que predecir uno o más valores para uno o más ejemplos. Los ejemplos constituyen una muestra tomada de un conjunto de datos con características similares. A éstos se los representa como un conjunto de valores para una serie de atributos

nominales o numéricos. En cada ejemplo se tiene atributos evidencia compuestos por variables independientes o valores de entrada, los cuales vienen acompañados de una salida conocida como clase, categoría o valor numérico representado por al menos una variable dependiente.

Pueden haber uno o varios métodos o estrategias de resolución para un problema predictivo particular. Sin embargo, en cada método pueden asignarse más de una tarea y una técnicas para producir sistemas híbridos. . ([39])

2.2 ANÁLISIS DE LAS TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA REALIZAR PRONÓSTICOS.

2.2.1. TÉCNICAS BASADAS EN ALGORITMOS GENÉTICOS.

Las Técnicas basadas en algoritmos genéticos (AGs) utilizan programas que están basados en una interpretación de la naturaleza, inspirados en la evolución de los seres vivos como una forma de llegar a definir y aprovechar una estructura de datos que admita todas las posibles soluciones a un problema. Tanto para problemas de regresión como para clasificación se construyen estructuras genéticas que representan a los conjuntos de entrada y de salida, y a través de ciclos de búsqueda se reproducen las mejores y se eliminan las peores soluciones al problema, modificándolas al azar en el proceso hasta que cumplan una condición de convergencia.

Los algoritmos genéticos parten de una población (colección de hipótesis) inicial y en el proceso generan en cada paso, una sucesión de hipótesis, mutando y recombinando las mejores hipótesis del paso anterior, hasta llegar a obtener una población final que contenga las hipótesis más aptas o con mayor probabilidad de acercarse a una solución final del problema. ([40]) ([41]) ([42])

---

[39] HERNANDEZ J, RAMIREZ J, FERRI C., **op.cit.** pp 139-150
[40] GOLDBERG, D.E., **op.cit.** pp. 493-530
[41] DOMÍNGUEZ I.., GABINO G., **op.cit.**
[42] RUSSELL, S., NORVIG P., **op.cit.** pp. 131-134.

Las hipótesis se codifican en cadenas de bits (***cromosomas***), de forma que puedan ser manipuladas por operadores genéticos como el cruce y la mutación. Cada cromosoma tiene varios ***genes*** componentes que corresponden a los parámetros del problema. Los genes se representan por subcadenas de bits, compuestos de 0s y 1s. A todo el conjunto de cromosomas se lo denomina ***genoma***. La secuencia particular de código de bits del genoma de un individuo concreto se conoce como ***genotipo***. La expresión decodificada de un genotipo es la que caracteriza al individuo y se la conoce como ***fenotipo***. ([43]) Los valores que toman los genes determinan el comportamiento del individuo.

El funcionamiento básico del algoritmo genético consiste en actualizar en forma interactiva el conjunto inicial de cada miembro de la población hasta obtener los más aptos, los cuales se evalúan con una ***función de idoneidad,*** que en la mayoría de las aplicaciones es probabilística, es decir, otorga puntuaciones a los miembros en la medida que se acerca a valores óptimos de la función. ([44])

Los pasos para implementar un algoritmo genético básico son los siguientes:

- Llamar a una función principal que borre los datos anteriores e inicialice variables de entorno, entre ellas el número de ejemplares que tendrá cada generación y el tamaño de los cromosomas.
- Generar una población inicial con estructura creada en forma aleatoria.
- Aplicar adaptabilidad y selección. En base al resultado establecer un ranking en base a la convergencia. Evaluar la puntuación de cada uno de los genes, esto es, permitir a cada uno de los individuos reproducirse, de acuerdo con su puntuación. Los cromosomas se reproducen por clonación total de acuerdo a su puntuación, lo cual significa que sólo los más fuertes sobreviven. Los ejemplares seleccionados reemplazan a la población total.
- Aplicar cruce**.** Emparejar los individuos de la nueva población, haciendo que intercambien material genético.

---

[43] MERELO, J., **op.cit.**; WINCHESTER A.M., **op.cit**, pp. 72-74.

[44] DOMINGUEZ I, GABINO G., **op.cit**. pág. 5

- Aplicar mutación. Hacer que alguno de los bits de un gen se vea alterado debido a una mutación espontánea. Ocasionalmente se produce un ejemplar mutante en algún bit.
- Se ejecuta nuevamente el ciclo desde selección hasta que finalmente se llega al estado de convergencia en el sistema, tras lo cual termina el AG.

En el proceso de selección pueden usarse métodos aleatorios como el método de la ruleta (todos compiten) o el método del torneo (compiten por parejas).
A las actuaciones que se realizan sobre las cadenas de bits se les conoce como ***operadores genéticos***.

Los más usuales son: ***Cruce*** (crossover), que simula un cruce, apareamiento ó reproducción sexual; ***Mutación*** (mutation), que correponde a mutación o alteración en los genes; ***Adaptabilidad*** (fitness), que consiste en seleccionar los más aptos; ***Clonación***, que reproduce individuos sin alterarlos; ***Inversión,*** que es el inverso de la clonación. Otros operadores son: ***Reversión***, en donde un cromosoma de baja adaptabilidad, se intercambia a si mismo revirtiendo el orden de los bits.; ***Anulación*** **o Pavlov,** que consiste en anular genes determinados, los cuales se fuerzan a evolucionar por separado y se reinsertan al cromosoma a posteriori.([45]) A la situación en la cual se deja de aplicar operadores genéticos a los individuos más aptos se conoce como **elitismo**.

El algoritmo converge a la solución cuando se cumple una condición de terminación, es decir, cuando se ha llegado a un conjunto n de generaciones prefijada. Para una buena convergencia el algoritmo debe iniciarse con un tamaño de la población suficiente para garantizar diversidad de soluciones.

Como puede verse en la **Fig.2.1**, los operadores genéticos en algunos casos alteran los bits del cromosoma y en otros los reorganizan. El cruce y la mutación sirven para transmitir bits de cromosomas a las nuevas poblaciones (***herencia***). Mientras que la clonación transmite todos los genes a los cromosomas hijos, la reversión los reordena.

---

[45] GONZALEZ BOSQUE M., **op.cit**.

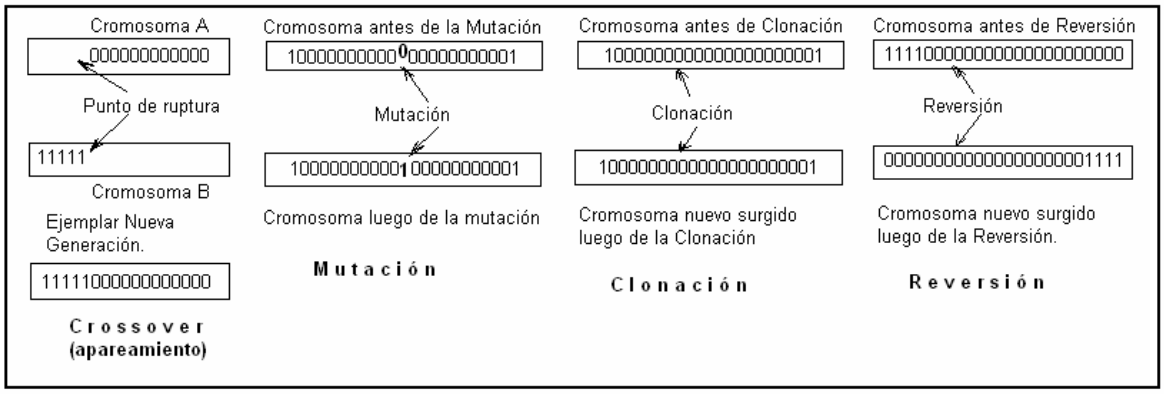


Figura 2.1. Operadores Genéticos. Fuente: [GONZALEZ BOSQUE M]; **op.cit**.

2.2.2. TÉCNICAS BASADAS EN ÁRBOLES DE DECISIÓN

Las Técnicas basadas en árboles de decisión y sistemas de aprendizaje de reglas, usan una representación del conocimiento en forma de reglas y se basan en dos tipos de algoritmos: los denominados "divide y vencerás" como el ID3/C4.5 o el CART, y los denominados "separa y vencerás como el CN2. ([46])

Un árbol de decisión es un conjunto de condiciones o reglas organizadas en una estructura jerárquica, de tal manera que la decisión final se determina siguiendo las condiciones que se cumplen desde la raíz del árbol hasta una de sus hojas. . Toman como entrada un objeto o una situación descrita a través de un conjunto de atributos discretos o continuos y devuelven una **"decisión"**: el valor previsto de la salida dada la entrada. El aprendizaje en árboles de decisión con valores discretos se denomina **"clasificación"**, y el aprendizaje de una función continua se denomina "**regresión**". ([47]).

Un árbol de decisión desarrolla una secuencia de pruebas sobre un valor de un atributo para poder alcanzar una decisión. Cada prueba es un ***nodo*** y representa

[46] HERNANDEZ J, RAMIREZ J, FERRI C., **op.cit,** p.293.
[47] HERNANDEZ J, RAMIREZ J, FERRI C, p.182.; RICH E., KNIGHT K., **op.cit**. pp.455-475

el valor que ha de ser devuelto si dicho nodo es alcanzado, es decir, si se cumple la prueba. ([48]). Si se alcanza el objetivo , se tiene **punto de ruptura** en el árbol.

El más simple es un árbol de decisión booleano, el cual consiste en un vector de atributos de entrada X, y en un único valor de salida booleano y. Mediante un conjunto de ejemplos (x_1,y_1), ... (x_n, y_n) se puede ***inducir*** la estructura del árbol, para lo cual se tienen ejemplos positivos: aquellos en los que la meta esperada es verdadera. Los ejemplos negativos son aquellos en los que la meta es falsa. El conjunto de ejemplos completo se denomina ***conjunto de entrenamiento***.

El algoritmo consiste en hacer pruebas de validez sobre **atributos** (pregunta si el atributo cumple una determinada condición) para luego elegir el mejor atributo (el que produzca el menor desorden y haga más sencillo al árbol), el cual se convierte en un nodo raíz, y luego se continúan las pruebas a partir de este nodo y se siguen creando subárboles, recursivamente, hasta que no queden ejemplos.
Para elegir el mejor atributo, se usa la fórmula del desorden, adoptada de la teoría de la información, dónde p = ejemplos positivos y n = ejemplos negativos.([49]).

$$\textit{Desorden Promedio} = \sum I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) = -\frac{p}{p+n} log_2 \frac{p}{p+n} - \frac{n}{p+n} log_2 \frac{n}{p+n}$$

Fórmula 2.1. Desorden promedio. [NIKOS D]; **op.cit**.

2.2.3 TÉCNICAS BASADAS EN REDES NEURONALES ARTIFICIALES

Las Técnicas basadas en redes neuronales artificiales, son algoritmos que aprenden un modelo mediante el entrenamiento de los pesos que conectan un conjunto de nodos o neuronas. La topología de la red y los pesos de las conexiones determinan el patrón aprendido. El algoritmo comúnmente usado en regresión es el de retropropagación (backpropagation).

---

[48] RUSSELL S, NORVIG P., **op.cit**. pp.744-749.
[49] NIKOS D., **op.cit**.

Las redes neuronales artificiales son un modelo artificial y simplificado del cerebro humano cuyas características básicas son: la capacidad de procesamiento paralelo, la estructura de memoria distribuida y la adaptabilidad.([50])([51])([52]).

El procesamiento en paralelo consiste en utilizar la mayor cantidad posible de unidades de procesamiento a la vez. La estructura de memoria distribuida consiste en realizar conexiones (sinapsis) entre los nodos, para recuperar información a través de la red. La adaptabilidad. consiste en modificar las sinapsis de acuerdo a las necesidades del entorno, lo cual permite trabajar con la generalización de conceptos y responder ante nuevos eventos y cambios futuros.

Los elementos básicos de un sistema neuronal artificial son las ***neuronas***, las cuales tienen una estructura jerárquica similar a las neuronas biológicas. Son dispositivos de cálculo que se organizan en capas con funcionalidad propia. El sistema global de proceso (sistema neuronal), comprende un conjunto capas de redes neuronales artificiales con sus interfaces de entrada y salida. ([53])([54])

La estructura genérica de una neurona artificial según Pitts (1943) se compone de los siguientes elementos: los **nodos**, o unidades de procesamiento conectadas a través de conexiones dirigidas denominadas **sinapsis**; el conjunto de **entradas** o vector de entrada x_i, el cual procede del exterior o de otras entradas ó **dendritas;** y los **pesos sinápticos** w_{ij} de una neurona j, los cuales representan la intensidad de iteracción entre la neurona que proporciona las entradas (presináptica) y la neurona que genera la salida (post-sináptica).

El nivel mínimo que se debe alcanzar para que la neurona se dispare se conoce como potencial post-sináptico y es el resultado de: $\sum w_{ij} x_j - \emptyset$

---

[50] HERNANDEZ O, FERRI C,**op.cit**. pp 327-328;
RUSSELL, S., NORVIG P.,**op.cit**. p. 838.
[51] MARTIN DEL BRÍO B, SANZ M., **op.cit**., pp. 6,7.
[52] GALLEGO A.C., **op.cit.**
[53] Id., **op.cit**, p.13.
[54] RUMELHART D.E, McCLELLAND J.L, **op.cit.**

La ***regla de propagación*** $g(w_{ij}, x_j(t))$, proporciona el valor del potencial post-sináptico $h_i(t) = g(w_{ij}, x_i(t))$ de la neurona i en función de sus pesos y entradas. La función más conocida de tipo lineal se basa en la suma ponderada de las entradas con los pesos sinápticos: $h_i(t) = \sum w_{ij} x_j$, que también puede interpretarse como el producto escalar de los vectores de entrada y los pesos.

La **función de activación**, $fi(a_i(t-1), h_i(t))$, proporciona el estado de activación actual $a_i(t) = f_i(a_i(t-1))$ de la neurona i, en función de su estado anterior $a_i(t-1)$ y de su potencial post-sináptico actual.

La ***función de salida***, $F_I(a_i(t))$, prorporciona la salida actual $y_i(t) = F_i(a_i(t))$, esto es, el *axon* de la neurona i en función de su estado de activación. De este modo, la operación de la neurona i puede expresarse como:

$.y_i(t) = F_i(f_i[a_i(t-1), g(w_{ij}, x_j(t))])$

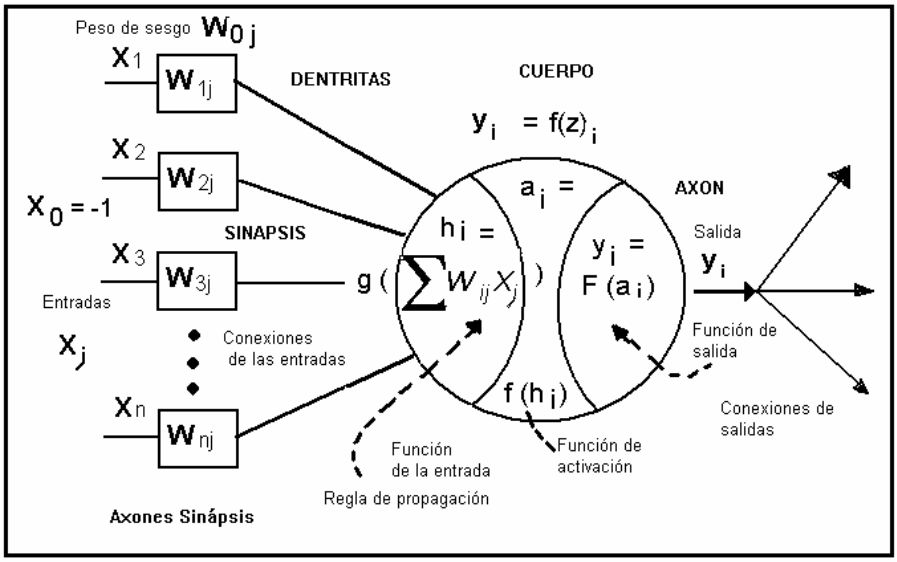


Figura 2.2. Modelo generalizado para una neurona.

Fuente: [MARTIN DEL BRÍO B, SANZ M.]; [RUSSELL S., NORVIG P.], **op.cit**.;

En una ***arquitectura de redes neuronales***, se denomina ***neurona de entrada*** a las neuronas sin sinapsis entrantes y ***neurona de salida*** a las neuronas sin sinapsis salientes. Las que no son de entrada ni de salida se denominan

neuronas ocultas.([55]). La arquitectura de red (**topología**) puede tener una, varias capas ocultas o ninguna.

Según sus conexiones, las estructuras de redes neuronales pueden ser:

- Redes ***acíclicas*** (conexiones hacia delante y conexiones laterales),
- Redes ***cíclicas*** o redes recurrentes. (conexiones hacia atrás).

En las redes neuronales, el conocimiento sobre un problema o fenómeno se lo adquiere mediante el *aprendizaje a partir de ejemplos*.

El aprendizaje de una red neuronal artificial se puede producir de tres formas ([56]):
El ***aprendizaje supervisado***: consiste en introducir una serie de patrones de entrada a la red y a su vez mostrar la salida que se quiere tener. La red muestra la salida memorizada cuando ha sido entrenada a aprender los patrones de entrada mediante el ajuste de los pesos en cada capa. En el ***aprendizaje no supervisado*** se presentan patrones de entrada a la red y ésta los clasifica en categorías según sus rasgos más sobresalientes. En el ***aprendizaje autosupervisado, l***a propia red corrige los errores en la interpretación empleando una realimentación por si misma. En el caso de las redes supervisadas, tanto los valores de entrada como de salida son valores conocidos. En las redes no supervisadas, no se conoce la salida deseada.

Los modelos mas comunmente aplicados a la obtención de pronósticos son:

- ***Adalina.*** (Utiliza función de activación lineal)
- ***Perceptrón multicapa****.* (Utiliza funciones de activación no lineales)

Para realizar aplicaciones que utilizan redes neuronales artificiales, se requiere de ciertas tareas como: seleccionar el conjunto de datos, desarrollar la arquitectura neuronal, normalizar los datos de entrada, definir una función de activación, seguir

[55] MARTIN DEL BRÍO B, SANZ M., **op.cit**., p.24
[56] GALLEGO A.C., **op.cit**.

el proceso de aprendizaje en base a patrones de entrenamiento iniciales, y adaptar la red según los nuevos patrones obtenidos del entrenamiento. ([57])

2.2.4 TÉCNICAS BASADAS EN VECINDAD Y RAZONAMIENTO BASADO EN CASOS.

Las Técnicas basadas en casos o vecindad (vecino más próximo), consisten en encontrar casos similares a nuevos ejemplos, de entre un conjunto conocido. La valoración de cada caso se la hace directamente a través de una fórmula de distancia o de una manera más sofisticada, mediante la estimación de funciones de densidad. Se utiliza aprendizaje por instancias usando analogía entre los valores para una entrada x en su estado final y el que tienen sus estados anteriores similares (vecinos más cercanos). Para un cierto valor x, se estima su densidad de probabilidad desconocida basándose en la idea de que es probable que las propiedades de x sean similares a sus puntos más cercanos de todo el conjunto. ([58])

Esta técnica es muy útil cuando se intenta tomar una decisión sobre nuevos casos presentados y se tienen muchos ejemplos del comportamiento de casos ya conocidos. Es muy común actuar ante una nueva situación como se hizo en situaciones anteriores parecidas o similares (próximas) y en las que se tuvo éxito. Esta técnica utiliza dos métodos de búsqueda: Los métodos que se hacen en el momento en que se presenta un nuevo caso se conocen como ***retardados***, mientras que aquellos en los que se hace una generalización o preprocesamiento antes de que se presenten los nuevos casos son métodos ***no retardados***. ([59])

Los sistemas basados en vecinos más próximos se construyen en base a un conjunto de ejemplos de casos anteriores o individuos llamados **instancias**.

---

[57] SERRANO C., "**op.cit.**

[58] RUSSELL S, NORVIG P., **op.cit**., p.834.
[59] HERNANDEZ J, RAMIREZ J, FERRI C, **op.cit**. p.422.

Para utilizar el concepto de **similitud** entre casos, se utiliza las métricas o medidas de distancia. Esto es, si se quiere saber la similitud entre dos instancias o individuos se debe elegir una **función de distancia**, y se calcula la similitud o distancia entre ambos. La función generalizada más sencilla y utilizada es la distancia euclidea o clásica, aunque existen otras apropiadas para problemas particulares. La ***distancia euclidea***, es la longitud de la recta que une dos puntos en el espacio euclideo (en este caso, el espacio en el que se encuentran los atributos de las instancias, denotados por x_i, y_i).

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

. **Fórmula 2.2**. Distancia Euclídea.

Para el procesamiento es necesario normalizar todos los atributos que tienen las instancias (con valores entre cero y uno por ejemplo), con el objeto de parametrizar el problema y conseguir uniformidad en la distribución de la media.

El funcionamiento básico de un algoritmo basado en distancia de vecinos más próximos es el siguiente: dado un ejemplo de prueba con entrada x, se obtiene la salida y = h(x) a partir de los valores de los k vecinos más cercanos a x. En una fase de evaluación se verificará el éxito de la solución, por ejemplo siendo comprobada en el mundo real. Durante el almacenamiento, la experiencia útil se guarda para futura reutilización, y el caso base se actualiza por un nuevo caso aprendido, o por una modificación de algunos casos existentes. ([60]). ([61])

2.2.5 TÉCNICAS BASADAS EN SISTEMAS EXPERTOS.

Los Sistemas Expertos o Sistemas basados en el conocimiento (SBC), utilizan el conocimiento de expertos humanos para dar respuestas. ([62]). En forma general, un sistema experto responde como un experto a un usuario sobre el tema que conoce. ([63]).

---

[60] HERNANDEZ J, RAMIREZ J, FERRI C, **op.cit**. p.445.
[61] LOZANO L., FERNANDEZ, J, **op.cit**.
[62] BETANZOS A., GUIJARRO B., LOZANO A., **op.cit**. p.12.
[63] SCHILDT H., **op.cit**., p.2.

En un sistema experto se tienen tres componentes: La **base de conocimiento** (BDC), la cual representa el conocimiento de dominio para la solución de problemas específicos, posee información y reglas específicas sobre una materia; el **motor de inferencias**, es una máquina genérica que almacena todos los procesos que efectúa el razonamiento a partir de los datos y utiliza el conocimiento de la BDC, a través de métodos de resolución de problemas; y las **Interfaces**, que son almacenamientos temporales para datos de entrada e intermedios generados durante el proceso de razonamiento, contiene los mecanismos de diálogo hombre-máquina y las explicaciones de las respuestas que ofrece al usuario. ([64]).

El funcionamiento de un sistema experto radica en el motor de inferencias. Pueden ser construidos de tres maneras: Por el ***método de encadenamiento hacia delante***: por el ***método de encadenamiento hacia atrás***; o por el ***método de reglas de producción***. El primero utiliza la información que el usuario le proporciona para moverse a través de reglas IF-THEN hasta que encuentra un punto terminal o respuesta de solución. El encadenamiento hacia atrás parte de una hipótesis (posible solución) y pide información hasta confirmarla o negarla. El método de reglas de producción pide información que tenga una mayor importancia de acuerdo con el estado actual del sistema, este último es un sistema de encadenamiento hacia atrás mejorado. ([65])

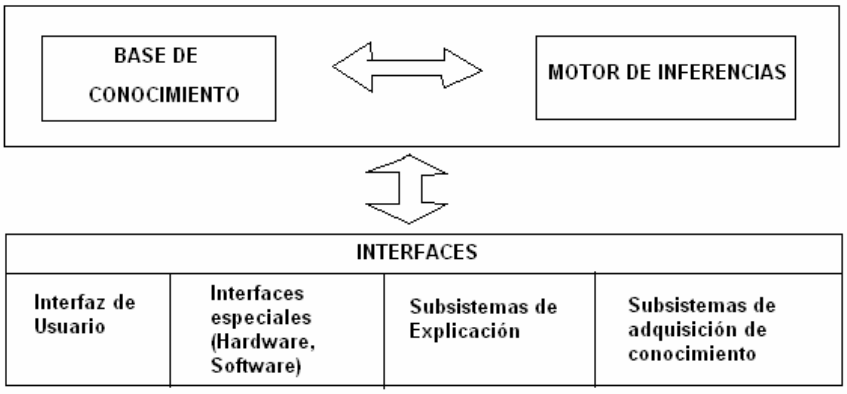


Figura 2.3. Arquitectura general de un SBC. [BETANZOS A.]

[64] BETANZOS A., GUIJARRO B., LOZANO A., **op.cit,** p.14.
[65] SCHILDT. H, **op.cit**. p.61-63.

2.3. SELECCIÓN DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA REALIZAR PRONÓSTICOS.

2.3.1 ESTRATEGIAS DE SELECCIÓN DE TECNICAS DE INTELIGENCIA ARTIFICIAL PARA REALIZAR PRONÓSTICOS.

2.3.1.1 Selección de la Técnica de Algoritmos Genéticos para realizar pronósticos por clasificación mediante extracción de reglas de asociación.

Los algoritmos genéticos son ideales para resolver problemas de extracción del conocimiento, generalmente el asociado al comportamiento de un fenómeno, y son los más utilizados en extracción de reglas.

Para llevar a cabo el proceso de extracción de reglas se debe primero determinar el esquema de representación del fenómeno, el cual es utilizado para codificar cada una de las soluciones, los operadores genéticos y la función de adaptación.

Hay dos maneras de enfocar el esquema de representación, la primera es el de "***Cromosoma = Regla***", en el que cada individuo codifica una única regla, mientras que la del enfoque:"***Cromosoma = Base de Reglas***" , también denominado enfoque Pittsburgh, cada individuo representa un conjunto de reglas.

En el enfoque "***Cromosoma = Regla***" existen dos propuestas genéricas:
- El enfoque Michigan en que cada individuo codifica una única regla pero la solución final será la población final o un subconjunto de la misma, por lo que es. necesario evaluar el comportamiento del conjunto de reglas al completo y la aportación de la regla individual al mismo, y
- El enfoque IRL (Iterative Rule Learning) en el que cada cromosoma representa una regla, pero la solución del Algoritmo Genético es el mejor individuo y la solución global está formada por los mejores individuos de una serie de ejecuciones sucesivas.([66]); ([67])

---

[66] M.J. DEL JESUS, P.GONZALEZ, F.HERRERA,M.MESONERO., **op.cit**.
[67] HERNANDEZ J, RAMIREZ J, FERRI C, **op.cit**, pp 390-391

La estrategia de elección del esquema de representación depende de la tarea a realizar por el algoritmo y por tanto, del tipo de regla a descubrir. El objetivo es encontrar un conjunto reducido de reglas en las que la calidad de la regla se evalúa de forma independiente del resto.

Una ***regla de asociación*** es una proposición probabilìstica sobre la ocurrencia de ciertos estados en una base de datos. En el algoritmo genético, los individuos "Reglas-Cromosoma" son hipótesis tomadas al azar para una muestra de casos de ejemplo particular de una población sobre la que se busca predecir su comportamiento. Una regla de asociación se convierte en ***regla de clasificación*** cuando el consecuente solo presenta una condición. El contenido del consecuente no tiene que ser un identificador de categoría o clase, sino puede tratarse de una condición de cualquier tipo (que es la que se pretende predecir, de ahí el nombre de ***regla de predicción***).

La codificación de una regla se puede describir de forma genérica como:

SI <condición 1 > Y < condición 2> Y... Y <condición N> ENTONCES <consecuente>

SI <antecedente> ENTONCES <consecuente>

El esquema de representación del algoritmo genético aplicado a clasificación mediante descubrimiento de reglas debe cumplir con ciertos requerimientos:

- Las condiciones del antecedente pueden implicar atributos nominales o numéricos. Si los atributos son numéricos, se puede utilizar un proceso de discretización que establezca una correspondencia entre valores e intervalos.
- El número de condiciones que aparece en el antecedente es variable.
- Si la regla es de clasificación, en el consecuente sólo aparecerá una condición que implica a un atributo, el atributo de clase o atributo objetivo, y éste no puede estar contenido en el antecedente.
- Si la regla es de asociación no tiene porqué existir necesariamente una diferenciación entre atributos predictores y objetivo, y también puede existir más de una condición en el consecuente.
- Ya que cada atributo puede tomar varios valores dentro del antecedente de la regla, se puede adoptar una esquema de codificación binario, en el que además

hay que agregar por cada atributo un bit , que tenga un valor 0 para determinar si el atributo no interviene en el antecedente y un valor 1 si interviene. De esta manera se tiene un solo esquema de codificación variable sobre el cual se pueden extraer un número ilimitado de reglas, cada una con sus atributos específicos y que cumplan múltiples condiciones.

Para el funcionamiento del algoritmo genético aplicado a clasificación mediante descubrimiento de reglas se deben tomar en cuenta los siguientes aspectos:

- Ejecutar el algoritmo tantas veces como clases distintas existan para el consecuente de la regla, es decir, se debe asociar primero todos los individuos de la población que tengan la misma clase y luego ejecutar el algoritmo.
- Evitar la convergencia de la población a un individuo en particular, ya que un sólo individuo no es una solución válida para el problema cuyo enfoque es el de "*Cromosoma = Regla*".
- Elegir, de forma determinística, la clase más adecuada para el antecedente de la regla. Lo cual se logra eliminando o añadiendo condiciones para el antecedente, previo a la ejecución del algoritmo.
- En clasificación no es adecuado representar en el consecuente más de una variable, esto se lo hace sólo si el objetivo es descubrir reglas para expresar dependencias funcionales, no así si el objetivo es obtener un atributo objetivo único (predicción).

- **La función de adaptación** generalmente es la media aritmética de la *completitud* (soporte, cobertura o prevalencia) y la *confianza* (precisión) de la regla representada en el cromosoma. ([68])

$$.\text{adaptación}(c) = \frac{\text{Completitud}(c) + \text{Confianza}(c)}{2}$$

Fórmula 2.3. Función de adaptación generalizada para A.G. descubrimiento de reglas

[HERNANDEZ J, RAMIREZ J, FERRI C]

[68] HERNANDEZ J, RAMIREZ J, FERRI C., **op cit.**, pág 396.

En ciertos casos se puede aplicar un factor de certeza a cada uno de estos parámetros para darle más evidencia a la función, como se muestra a continuación:

$$.\text{fitness} = \frac{W_1 * \text{Completitud (c)} + W_2 * \text{Confianza (c)}}{W_1 + W_2}$$

Fórmula 2.4. Función de adaptación ponderada para A.G. descubrimiento de reglas
[**M.J. DEL JESUS, P.GONZALEZ, F.HEERERA, M.MESONERO**]

Según GIRALDEZ Raúl ([69]), encontrar una función de evaluación apropiada encargada de medir la calidad de los individuos respecto a la clasificación que éstos realizan, no es una tarea trivial, ya que la mayoría de bases de datos incorporan ruido. Este autor propone una función de evaluación de la forma:

$$.f(r) = N – EC(r) + A(r) + \text{cobertura}(r)$$

Fórmula 2.5. Función de adaptación con reducción de error.
[GIRALDEZ Raúl, RIQUELME José C]

Donde N es el número de ejemplos del proceso, EC(r) es el error de clase, es decir, el número de ejemplos que pertenecen a la región definida por la regla pero que no comparten la misma clase que dicha regla; A(r) es el número de ejemplos totalmente clasificados; y la cobertura(r) es el volumen normalizado de la región cubierta por la regla.

No obstante, la función de adaptación definida al comienzo como la media aritmética de los parámetros de confianza y completitud, es la más indicada para aplicarla de manera general. Para aplicar la función, el método usado tradicionalmente por los algoritmos de aprendizaje evolutivos, consisten en realizar un recorrido lineal de los datos de entrenamiento para cada individuo de la población. Proceso que, aunque sencillo de aplicar, según opinión de

[69] GIRALDEZ Raúl, RIQUELME José C., **op cit**.

GIRALDES Raúl [13], puede resultar altamente costoso en términos de coste computacional.

- La ***confianza*** determina la precisión de la regla, ya que refleja el grado con que los ejemplos pertenecientes a la zona del espacio delimitado por el antecedente verifican la información indicada en el consecuente. Es el porcentaje de veces que la regla se cumple cuando se puede aplicar y se calcula como la suma del grado de pertenencia de los ejemplos de la clase en la zona determinada por el antecedente dividido entre la suma del grado de pertenencia de todos los ejemplos (independientemente de la clase a la que pertenecen). La fórmula varía de acuerdo a si las reglas son o no difusas (con múltiples grados de pertenencia de la clase a la zona de los ejemplos). Generalmente las reglas son no difusas (con solo dos grados 0 para no pertenece o 1 para pertenece).

- La ***completitud,*** es una medida del grado de cobertura que la regla ofrece a los ejemplos de la clase, también se define como el número de instancias que la regla predice correctamente (en porcentaje) . Se calcula como el cociente entre el número de nuevos ejemplos de la clase que cubre la regla y el número de ejemplos de la clase que quedaban por cubrir.

Estos parámetros se pueden explicar mejor de la siguiente manera: Teniendo a A como el conjunto de atributos-valor que aparecen en el lado derecho de la regla, definido como antecedente; y a B como el atributo-valor que aparece en el lado izquierdo definido como consecuente.

$$\text{Soporte}(A \Rightarrow B) = P(A \cup B)$$

$$\text{Confianza}(A \Rightarrow B) = P(B \mid A) = \frac{\text{soporte}(A \cup B)}{\text{soporte}(A)}$$

Fórmula 2.6. **Soporte y Confianza de evaluación de reglas de asociación.**

[MORALES MANZANARES, Eduardo, **op cit**]

La forma de hallar estos valores es la siguiente:

De n casos, el antecedente puede ser cierto en r_a casos y de éstos, r_c casos lo es también el consecuente. Por consiguiente,

Soporte, cobertura o completitud = $T_s = r_c / n$ = P(Cons ^ Ante)

(número de casos o porcentaje en los que se aplica satisfactoriamente r_c)

Confianza o certeza de la regla = $T_c = r_c / r_a$. = P(Cons / Ante)

El objetivo global de la función de adaptación es orientar la búsqueda hacia reglas que maximicen la precisión (confianza), minimizando el número de ejemplos negativos y no cubiertos.

El esquema completo de funcionamiento del algoritmo es el expresado por M.J. DEL JESUS, P.GONZALEZ, F.HEERERA, M.MESONERO ([70]), en su aplicación propuesta de algoritmo evolutivo, es descrito de la manera siguiente:

```
INICIO
 Cto_Reglas ← Ø
 REPETIR
   Ejecutar el AG obteniendo la regla R
   Búsqueda Local (R)
   CtoReglas ← CtoReglas + R
   Modificar el conjunto de ejemplos
 MIENTRAS confianza(R) >= confianza_min y
          R cubra ejemplos nuevos
FIN
```

Figura 2.4 Algoritmo genético aplicado al descubrimiento de reglas

[**M.J. DEL JESUS, P.GONZALEZ, F.HEERERA, M.MESONERO**]

El algoritmo genético permite obtener una regla para cada clase con un valor alto de confianza a un nivel de completitud adecuado.

El conjunto de reglas (Cto.Reglas) que se obtiene de aplicar el algoritmo genético, representan a una élite de individuos dentro de su generación, reglas con un alto nivel de aceptación entre toda la población. (**Figura 2.4**).

[70] M.J. DEL JESUS, P.GONZALEZ, F.HEERERA, M.MESONERO, **op.cit**.

Una variación de este mismo algoritmo maximiza el conjunto de reglas élite descubierto, de la forma siguiente:

Inicio

Crear la población P y el almacén élite E

Mientras (generación < MaxGeneraciones) hacer

Seleccionar los futuros padres a partir de los individuos de P y E

Aplicar operadores genéticos sobre los padres seleccionados

Evaluar los hijos obtenidos

Actualizar los conjuntos P y E

.generación = generación + 1

Fin Mientras

Devolver el conjunto E

Fin

Figura 2.5 Variación del a**lgoritmo genético aplicado al descubrimiento de reglas**
[ROMERO Cristóbal**, VENTURA S**]

En esta variación del algoritmo, se involucra en el proceso de selección un conjunto de individuos E, obtenido de la salida de cada generación intermedia. ([71])

En síntesis, los algoritmos genéticos de descubrimiento de reglas dan como resultado un conjunto de reglas élite, obtenido por optimización con alguna condición de convergencia o de parada, en la cual se aplica un límite de generaciones predefinido.

Para los ***operadores genéticos***, el algoritmo utiliza un modelo de reproducción de estado estacionario. La recombinación (cruce) se realiza a través del operador de cruce multipunto y un operador de mutación uniforme, donde el propósito es incrementar la generalidad de las reglas. ([72])

[71] ROMERO Cristóbal, **op cit**.
[72] M.J. DEL JESUS, P.GONZALEZ, F.HEERERA, M.MESONERO, **op.cit**.

2.3.1.2 Otras estrategias de selección de técnicas de inteligencia artificial para realizar pronósticos.

2.3.1.2.1 Técnica de Algoritmos genéticos para realizar pronósticos de transición de estados.

Se aplica en problemas de contexto económico, en los cuales el algoritmo genético sirve como herramienta de predicción de la composición de una población inicial, a partir de la cual se somete a los individuos a adaptarse a su medio socio-económico hasta que finalmente sobreviven los más aptos y se obtiene una población final resultado un proceso de transformación. Para el proceso de selección se recurre a una matriz de transición que establece las reglas probabilísticas mediante las cuales deben prevalecer los individuos más aptos hasta conformar la población futura. Esta técnica constituye un algoritmo genético híbrido, por cuanto sustituye los operadores genéticos de cruce y mutación y se sustenta en las probabilidades de transformación de las características de la población.([73]).

2.3.1.2.2 Técnica de árboles de decisión para realizar pronósticos.

En esta técnica se utilizan los árboles de decisión para clasificación y se denominan "**árboles de regresión**", ya que etiquetan en cada nodo de hoja del árbol a valores continuos de variables, en lugar de valores discretos. El árbol en conjunto representa a una función aprendida. Para encontrar la función se aplica un algoritmo de aprendizaje por inducción, que utiliza un conjunto de datos de entrenamiento. Al final del entrenamiento se construye el árbol con las hipótesis que mejor predicen las salidas para nuevos ejemplos. ([74]) ([75])

2.3.1.2.3 Técnica de Redes Neuronales Artificiales para realizar pronósticos.

Las redes neuronales artificiales para predicción son redes supervisadas que utilizan algoritmos de aprendizaje en los cuales se puede introducir con anterioridad las entradas y las salidas, luego se generaliza el modelo con los

---

[73] HERNANDEZ LOPEZ M, **op.cit.** pp 389-407.
[74] RUSSELL S, NORVIG P., **op.cit**. p.756-758
[75] HERNANDEZ J, RAMIREZ J, FERRI C., **op.cit**., p.293.

ejemplos aprendidos y al final se obtienen nuevas salidas con los nuevos datos de entradas. ([76]). Se utilizan algunos modelos de arquitectura de redes neuronales, de los cuales los más comunmente utilizados para predicción son el perceptrón, el adaline y el perceptrón multicapa. En este último modelo se aplica el algoritmo de aprendizaje por retropropagación o BP ("backpropagation"), en el cual se realiza un proceso de activación de neuronas entre la capa de entrada, las capas ocultas y la capa de salida, en ambas direcciones, hasta descubrir los pesos en cada neurona, que permitan predecir las salidas.

2.3.1.2.4 Técnica de Sistemas Expertos para realizar pronósticos.

Los sistemas expertos son ideales para resolver problemas de predicción y pronósticos ya que ofrecen respuestas inmediatas que parten de una base de conocimiento sobre la que se aplican métodos de inferencia. La arquitectura del sistema experto orientado a predicción debe considerar tres estados: Un estado inicial, en el cual se representan los datos de entrada; los estados intermedios, en los cuales se encuentra una interfaz de usuario capaz de dar mantenimiento a los hechos y parámetros de búsqueda y: los estados finales, los cuales representan las situaciones objetivo para alcanzar las soluciones..

2.3.1.2.5. Técnicas combinadas de inteligencia artificial para realizar pronósticos.

Los sistemas que combinan algoritmos genéticos y redes neuronales artificiales, prescinden del algoritmo de entrenamiento típico (retropropagación) y lo reemplazan por un algoritmo genético (AG). Con ello disminuyen sensiblemente la necesidad de fijar los parámetros de los cuales depende la operación de la red neuronal y su problema radica en realizar un diseño óptimo de la arquitectura de la red. [KURI MORALES A] ([77]). Otras técnicas combinadas utilizan a los algoritmos genéticos para tareas de búsqueda heurística, y para estimación de probabilidades. Otra alternativa es combinar métodos de vecindad con técnicas evolutivas.([78]). Esto consiste en sustituir la función de adaptación por la función de distancia del vecino más próximo u otra técnica de búsqueda hacia atrás.

---

[76] ZAVALA TOLEDO C.A., **op.cit.**
[77] KURI MORALES A., **op.cit**., pp 10-12.
[78] HERNANDEZ J, RAMIREZ J, FERRI C., **op.cit**. p 449.

2.3.2 PRINCIPALES VENTAJAS Y DESVENTAJAS DE LAS TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA REALIZAR PRONÓSTICOS.

2.3.2.1 Ventajas y Desventajas de Selección de Técnicas de Algoritmos Genéticos para realizar pronósticos frente a otras técnicas.

Las principales ***ventajas*** del uso de los algoritmos genéticos para realizar pronósticos, frente a otras técnicas son:

- Tratan el problema como un todo y permiten descubrir nuevas soluciones cada vez que se aplica el algoritmo, lo que los hace especialmente diseñados para problemas de optimización y adecuados para la extracción de conocimiento a través de reglas de asociación.
- Evitan dudar sobre la veracidad de los resultados, ya que estos pueden ser medibles y pueden ser analizados. Además, son el resultado de un proceso de selección entre las mejores soluciones.

Las principales ***desventajas*** del uso de algoritmos genéticos para realizar pronósticos son:

- Su eficacia depende de escoger una representación adecuada de los datos.. Una codificación errada obliga a estudiar de nuevo el genoma para saber si cumple las restricciones del problema y puede predecir resultados.
- La población inicial debe ser suficientemente grande para evitar que se produzcan individuos no deseados. En algunos casos es necesario aplicar una mutación beneficiosa para evitar que se repitan los mismos patrones en el entrenamiento y no se llegue a una rápida conversión.
- En el enfoque "*Cromosoma-Regla*" es difícil determinar un valor de aptitud para todas las reglas en conjunto, ya que cada regla se evalúa individualmente con su función de adaptabilidad.

2.3.2.2 Ventajas y Desventajas de Selección de Otras técnicas para realizar pronósticos.

Las principales **ventajas** de otras técnicas para realizar pronósticos son:

- En los árboles de decisión, a través del análisis de los datos, se obtiene por inducción la estructura del árbol. Se hace un buen trabajo de predicción de entrenamiento, siempre que el árbol encontrado sea pequeño y permita llegar a la solución en forma rápida.
- El aprendizaje adaptativo a través de ejemplos y su auto-organización, convierten a las redes neuronales artificiales en ideales para producir aproximaciones a la solución de problemas predictivos. Además sus cómputos pueden hacerse en paralelo y con miles de datos.
- Las técnicas del vecino más próximo son muy útiles para el aprendizaje supervisado, especialmente cuando se conocen todos los casos que se pueden presentar. Además las fórmulas de distancia son simples de usar.
- Los sistemas expertos resuelven problemas predictivos en los cuales se desconoce el modelo matemático cuya solución es compleja. Además preservan el conocimiento de expertos humanos en la solución.
- Las técnicas combinadas se complementan para elevar el nivel de confianza del pronóstico. Las técnicas de inteligencia artificial pueden utilizar tareas estadísticas para su evaluación. En otros casos simplifican tareas. Por ejemplo, las redes neuronales pueden seleccionar los mejores casos de entrenamiento a través de un algoritmo genético.

Las principales **desventajas** de otras técnicas de predicción son:

- El problema de los árboles de decisión en predicción es que requieren mucha memoria para representarlos, además memorizan las observaciones del conjunto de entrenamiento, y en ciertos casos es difícil extrapolar a ejemplos que no se parezcan a los proporcionados. ([79])

[79] RUSSELL S, NORVIG P., **op.cit**., p.746.

- El inconveniente de las redes neuronales artificiales es que obligan a utilizar el método de prueba y error, lo cual subordina el problema a la estructura de red elegida para el entrenamiento. Además, requiere de muchos parámetros y es difícil interpretar su funcionamiento, el cual se comporta como una caja negra.([80])
- Mientras que al aplicar los algoritmos genéticos se obtiene cada vez mejores soluciones y además se puede controlar el punto de parada, en las redes neuronales artificiales se llega a memorizar los errores del sobre-entrenamiento, y es difícil controlar su convergencia.
- El problema de la técnica de vecinos más próximos cuando utiliza métodos retardados es que generaliza en base a la selección de un conjunto de ejemplos cercanos al nuevo caso, y para otros casos que se presenten debe buscarse otro conjunto de ejemplos cercanos, lo que puede tomar mucho tiempo si existen muchas posibles soluciones para cada ejemplo.
- Los sistemas expertos requieren de actualizaciones periódicas a la base de conocimientos. Además el proceso de adquisición de conocimiento es lento y su conversión a símbolos y lenguaje natural es complejo.
- La combinación de técnicas puede surtir efecto para un problema de predicción específico, pero no son aceptables para generalizar a todos los casos. Por eso es importante conocer, por ejemplo, la relación que existe entre los métodos de clasificación y de regresión y si los unos se pueden adaptar a los otros. También es necesario diferenciar entre métodos determinísticos y probabilísticos.

[80] MARTIN DEL BRIO B.M., SANZ A., **op.cit**. p.227, p.72

CAPITULO 3. IMPLEMENTACIÓN DE UN SISTEMA DE PREDICCIÓN Y ANÁLISIS DE DATOS APLICADO AL COMPORTAMIENTO DE LOS CASOS DE CONSTRUCCIÓN DE VIVIENDAS EN EL PAÍS.

3.1 ANÁLISIS.

3.1.1 ESPECIFICACIÓN DE REQUERIMIENTOS DE SOFTWARE.

3.1.1.1 Introducción.

El Sistema de Predicción y Análisis de Datos aplicado al comportamiento de casos de construcción de viviendas del país (SPACV), es un software diseñado para realizar análisis exploratorio de datos históricos, y aplicar técnicas de inteligencia artificial para predicción acorde a la naturaleza del problema. El ciclo de vida básico del software comprende las fases especificadas en la metodología O.M.T (Object Modeling Technique) como análisis, diseño, desarrollo, implementación y pruebas. La especificación de los requerimientos de software utiliza el estándar recomendado por la IEEE ("Recomended Practice for Software Requirements Specification" ANSI / IEEE 830, 1998).

En la construcción del software, la metodología a usar es la Orientada a Objetos (OO), con la notación de Booch. Parte del análisis será reforzado con diagramas tomados de UML (Unified Modeling Language). La implementación usará una herramienta de programación visual de propósito general interactiva.

El propósito es llevar a cabo la predicción y análisis asistido por computadora de los casos de edificación para uso de viviendas residenciales y no residenciales en el Ecuador obtenidos a través de la publicación de resultados de la Encuesta de Edificaciones, investigada por el Instituto Nacional de Estadística y Censos (INEC)

en base a los permisos de construcción que entregan a lo largo del año los municipios del país. El sistema SPACV estará dirigido al público con énfasis en el sector de la construcción y en el área estadística, preferentemente. Se utilizará los casos de edificaciones en todo el país desde el año 2000 hasta el 2005. La técnica seleccionada es la de algoritmos genéticos para descubrimiento de reglas de asociación entre variables para pronosticar el comportamiento del fenómeno.

Definiciones, Acrónimos y Abreviaturas.

Definición	Descripción
SPACV	Sistema de Predicción y Análisis de Construcción de Viviendas.
Algoritmo	Conjunto de operaciones de almacenamiento, clasificación o cálculo que se las puede hacer por intermedio del computador y que permitan asistir al usuario en problemas de razonamiento, aprendizaje o predicción.
Inteligencia Artificial	"El estudio de cómo lograr que las computadoras realicen tareas que por el momento, los humanos las hacen mejor". [RICH, 1991]
Predicción	Conjunto de tareas, técnicas y herramientas para la toma de decisiones cuya función es la de encontrar las fronteras en las cuales se tienen valores futuros esperados como la continuación de patrones o tendencias del pasado.
Árboles de decisión	Estructura arbórea que representa un conjunto de decisiones, las cuales generan reglas para clasificación de un conjunto de datos.
Algoritmos genéticos	Técnicas de automatización que usan un proceso similar a la combinación genética, mutación y selección natural en un diseño basado sobre la teoría de la evolución de Charles Darwin.
Redes Neuronales Artificiales	Modelos predictivos no lineales que aprenden directamente del entrenamiento y simulan en su estructura a las redes neuronales biológicas.
Estadística	Rama de las matemáticas que se ocupa de reunir, organizar y analizar datos numéricos para ayudar a resolver problemas como el diseño de experimentos y la toma de decisiones.
Ingeniería	Conjunto de conocimientos y técnicas que permiten aplicar el saber científico a la resolución de problemas mediante la invención, construcción u otras realizaciones provechosas para el hombre.
Ingeniería de software	Estudio de los principios y metodologías para desarrollo y mantenimiento de sistemas de software.
Sistema	Conjunto de partes coordinadas y en interacción para alcanzar un conjunto de objetivos.
Objeto	Parte o componente de un sistema de variedad limitada y que puede ser parte del mundo físico (partículas, animales, artefactos) o abstracto (variables matemáticas, ecuaciones, reglas, software). Unidad básica de la programación orientada a objetos, que se construyen como instancias de una clase.
Clase	Definición de un objeto

Tabla 3.1. Análisis: Especificación de Requerimientos - Definiciones

Acrónimo	Descripción
BCE	Banco Central del Ecuador
INEC	Instituto Nacional de Estadística y Censos
AOO	Análisis Orientado a Objetos
DSS	Decision Support System. (Sistema de ayuda a la toma de decisiones.)
IEEE	"Institute for Electrical and Electronics Engineers" (Instituto para Ingenieros Eléctricos y Electrónicos).
IA	Inteligencia Artificial
IS	Ingeniería de Software
IC	Ingeniería del Conocimiento
SBC	Sistemas basados en el conocimiento
AG	Algoritmo genético
RNA	Red Neuronal Artificial
PIB	Producto Interno Bruto
IPCO	Indice de Precios a la Construcción
DLL	Data Linkage Library (Librería de encadenamiento de datos).
EXE	Extensión del archivo ejecutable de la aplicación.
IPC	Indice de Precios al Consumidor
OO	Object Oriented (Orientado a Objetos)

Tabla 3.2. Análisis: Especificación de Requerimientos - Acrónimos

Abreviatura	Descripción
.op.cit	Obra citada
p.ej.	Por ejemplo
Fig.	Figura.
Pb.9	Power Builder versión 9
SQL	Lenguaje de Consulta en Base de Datos
Ms-SQL Server 2000	Motor y Administrador de la Base de Datos de Microsoft SQL Server versión 2000.

Tabla 3.3. Análisis: Especificación de Requerimientos – Abreviaturas

Visión Global.

Se dispone de casos históricos de construcción de edificaciones para viviendas residenciales y no residenciales, a partir de los cuales se requiere hacer un análisis exploratorio de los datos sobre una determinada localidad geográfica hasta un año dado, para luego aplicar una técnica de inteligencia artificial que prediga su comportamiento oculto y tendencias futuras.

3.1.1.2 Descripciones.

Perspectivas del Producto.

El sistema SPACV estará compuesto de una sola aplicación a ser instalada y ejecutada en sistemas operativos Windows 2000 o Windows Xp. El sistema funcionará en un ambiente Cliente-Servidor con conexión a la base de datos Ms-SQL-Server 2000. La aplicación incluirá interfaz visual de usuario en la cual se llevarán a cabo las tareas de introducción y obtención de resultados de predicciones de datos sobre casos de construcción de viviendas en el país.

Funciones del Producto.

- Ingresar y actualizar los casos de construcción de viviendas en el país de años anteriores importándolos, así como también los datos complementarios para el análisis relacionados con el fenómeno de la construcción de viviendas.
- Ingresar y actualizar los parámetros predeterminados de entrada de datos para los algoritmos a ser aplicados.

- Introducir casos de predicción asistida por el sistema mediante la selección de opciones, tanto para delimitación de parámetros de entrada como para la generación de resultados.
- Aplicar el algoritmo de predicción a uno o varios casos de prueba.
- Procesar y presentar resultados de las predicciones realizadas sobre los casos introducidos y de acuerdo a los parámetros dados.
- Realizar el mantenimiento general de los datos administrativos (respaldos, recuperación, seguridades) y configuración de usuarios.

Características del Usuario.

Los usuarios del sistema SPACV serán aquellos que tengan algún conocimiento sobre la construcción y necesiten conocer y estimar el comportamiento de las viviendas en base a los datos históricos de casos que le ofrezca el sistema. Los usuarios requerirán de asistencia para la instalación de la base de datos y de la aplicación inicialmente, si es que no disponen de algún conocimiento o experiencia en instalación de programas. Si los usuarios están familiarizados con el sistema operativo Windows, deberán seguir los pasos de instalación del sistema y el software adicional de soporte.

El sistema contemplará dos perfiles básicos: El administrador del sistema, quien realiza tareas de mantenimiento de datos y la configuración de usuarios; y el usuario predictor, el cual introduce los casos de predicción y recibe las respuestas de los resultados de la predicción.

Restricciones generales.

El sistema SPACV permitirá la introducción de datos para actualización en la base de datos de casos históricos de construcción de viviendas y en las variables macroeconómicas relacionadas con el fenómeno de construcción de viviendas de por lo menos los últimos cinco años.

El sistema no permitirá realizar cambios en el funcionamiento de los algoritmos de predicción pero si podrá hacer modificación a los parámetros y restricciones tanto del algoritmo como de los valores de rangos en las variables de conformación de casos de construcción de viviendas.

El sistema no permitirá realizar otros tipos de análisis de datos que los relacionados con el algoritmo de predicción seleccionado. Los resultados intermedios producto de la ejecución serán almacenados temporalmente, sin embargo, se permitirá guardar los resultados finales.

3.1.1.3 Requerimientos específicos.

Requerimientos de Interfaces externas.

Para las interfaces de usuario se requieren las siguientes ventanas:

1) Ventana de presentación inicial del sistema.
2) Ventana de autenticación de usuario y de conexión al sistema.
3) Ventana opcional de ayuda de contenido e índice de opciones.
4) Ventana de navegación en el menú general de opciones del sistema.
5) Ventana de mantenimiento de tablas
6) Ventana de actualización de parámetros.
7) Iconos de acceso a las tareas más frecuentes.
8) Ventana de ingreso de casos de predicción y de ejecución de los procesos de análisis y predicción junto con sus resultados.
9) Ventana de mantenimiento a usuarios y cambio de contraseña.

Este sistema no requerirá de interfaces de hardware adicionales a las que ya son proporcionadas por el sistema operativo.

Este sistema no requerirá de interfaces de software adicional. Sin embargo, el usuario podrá realizar otras tareas del sistema operativo mientras se encuentre activada la ventana de opciones de la aplicación.

Requerimientos Funcionales.

Para llevar a cabo los procesos y funciones de predicción en el sistema, será necesario que se hayan instalado o ingresado y preparado los siguientes datos:

- Los datos históricos de casos de construcción de viviendas con las variables a utilizarse y demás datos complementarios relacionadas con el fenómeno.
- Los datos de parámetros aplicables a las técnicas a aplicarse.

El usuario estará en capacidad de introducir los casos de predicción, modificar los parámetros de entrada y ejecutar el procedimiento o ciclo de predicción.

El procedimiento o ciclo de predicción se iniciará con la introducción de un caso de predicción por tiempo y localidad. Luego se introducirán los parámetros del algoritmo. A continuación será ejecutado hasta su convergencia, momento en cual el sistema estará en capacidad de presentar los resultados finales

Requerimientos de Funcionamiento.

Hardware:

- Un computador PC Pentium IV o superior con espacio en disco que sustente la instalación de SQL Server 2000 y los ejecutables de la aplicación (aproximadamente 200 Mb)
- El equipo servidor, requerirá de 50 Mb. Adicionales para la base de datos histórica de casos de construcción de viviendas, series históricas y el resto de archivos de parámetros y restricciones.
- Es opcional es uso de impresora.

Software.

- Sistema Operativo Windows 2000, o Windows Xp.
- Ms-SQL-Server 2000.
- Power Builder 9.0

3.1.1.2 Restricciones del Sistema.

Metodología de Desarrollo.

El modelo de desarrollo será documentado acorde a la metodología OMT (Object Modeling Technique), ya que el modelo de objetos contiene una enorme riqueza semántica, ideal para el desarrollo de sistemas basados en la representación del conocimiento. El contenido es el siguiente:

Análisis
- Descripción del problema.
- Modelo de Objetos (Clases, atributos, asociaciones, operaciones)
- Modelo Dinámico.(escenarios, eventos entre objetos, transición de estados)
- Modelo Funcional (especificación de operaciones)

Diseño
- Diseño del sistema(arquitectura del sistema, subsistemas)
- Diseño de objetos (combinación de los modelos especificados en el análisis, diseño de los algoritmos, modelos de datos)

Desarrollo e Implementación
- Pseudocódigo de los métodos y funciones principales.
- Codificación principal con su descripción documentada.

Pruebas.
- De unidad o módulo.
- De Integración
- De Validación.
- Globales o del Sistema.

El flujo de datos del modelo funcional será complementado con diagramas de realización de casos de uso que se utilizan en el Lenguaje Unificado de Modelado de Objetos (UML). Se usará un diagrama de flujo de datos para el algoritmo de predicción genético, aparte de los diagramas establecidos por la metodología..

3.1.2 DESCRIPCIÓN DEL PROBLEMA.

La construcción de viviendas en el país obedece a un comportamiento poco conocido y que se polariza hacia los centros urbanos más poblados, particularmente hacia las ciudades ejes del desarrollo socio-económico y que son: Quito, Guayaquil, Cuenca y Machala, las que además representan las tendencias actuales de crecimiento del sector de la construcción en el Ecuador, según el estudio preliminar de dicho fenómeno abordado en el Capítulo 1.

A partir de las variables estudiadas en la encuesta de edificaciones que publica anualmente el INEC en base a los permisos de construcción otorgados por los municipios del país, se propone llevar a cabo la extracción de conocimiento de estos datos de manera inteligente para predecir sus reglas de comportamiento. Entre las variables investigadas están: los materiales predominantes en los cimientos, paredes, estructura y cubierta de las construcciones, el tipo de obra, usos de la edificación, el origen y destino del financiamiento.

La técnica a aplicarse según una estrategia presentada en el Capítulo 2, consiste en utilizar algoritmos genéticos para clasificación mediante extracción de reglas de asociación, escogida por sus amplias ventajas de exploración de espacios de búsqueda en forma óptima.

3.1.3 MODELO DE OBJETOS

3.1.3.1 Descripción del modelo de Objetos.

Edificación.- Representa a un caso de construcción. De acuerdo a su uso, éste puede ser residencial, no residencial o mixto. Las viviendas se encuentran en el grupo residencial o mixto. Estos datos se obtienen de los resultados de la encuesta de edificaciones (permisos de construcción), que se realiza anualmente y cuya fuente primaria proviene de los municipios del país.

Tiempo.- El tiempo es una de las dimensiones del fenómeno de la construcción de viviendas. Se establece en años generalmente.

Localidad geográfica.- Espacio físico en el cual se desarrolla el fenómeno de construcción de viviendas. Las edificaciones se localizan en el espacio geográfico del país, cuya división política administrativa está jerarquizada en regiones, provincias, cantones y parroquias.

Indicador económico.- Es una medida del crecimiento, actividad o comportamiento económico del país. Cada indicador influye directa o indirectamente en toda actividad económica inclusive la construcción de viviendas. Sus valores se representan en el tiempo y localidad geográfica. Son datos secundarios para complementar el proceso de análisis.

Parámetro.- Representa las reglas de juego del algoritmo genético básico aplicado al proceso de extracción de reglas de asociación. Se incluyen, porcentajes de cruce y mutación, el número máximo de corridas y el número de generaciones. Estos pueden ser modificados para mejorar la efectividad del algoritmo genético.

Restricción de edificación.- Describen las reglas de decisión aplicadas a los casos de edificaciones de viviendas, que se utilizan para determinar si la edificación es válida o no, en función de su composición interna. Estas restricciones son datos complementarios al fenómeno.

Caso de predicción.- Conjunto de especificaciones que determinan un caso de edificación, del cual se desconocen sus reglas de comportamiento. Está circunscrito a una localidad geográfica y a un tiempo establecido en años de recolección de información estadística.

Serie Indicador Económico.- Es la asociación entre indicador económico, tiempo y localidad geográfica. En esta asociación se especifican como atributos los valores de cada serie de indicador económico, tales como el PIB, la inflación, los precios de los materiales de construcción, etc. y que son datos secundarios usados para el análisis de casos de edificaciones y podrían ser usados para encontrar alguna relación remota de causalidad con el fenómeno.

Variable Caso Indicador.- Es la asociación entre Indicador Económico y Caso de Predicción. En esta asociación se especifican como atributos las series de indicadores que se relacionan con el caso de estudio. Se las puede utilizar como complemento secundario para el análisis exploratorio de los datos.

Variable Caso de Predicción.- Consiste en los valores de la distribución porcentual de las variables definidas para el caso de predicción y su finalidad es usarlas en el proceso exploratorio de análisis de forma independiente a las reglas de asociación extraídas para predicción .

Generación.- Comprende los valores aplicados al conjunto de individuos que conforman la población a la cual se aplica el algoritmo genético.

Individuo.- Representación genética de todas las variables posibles que intervienen en el antecedente de una regla de asociación a extraer. Un individuo es un "***Cromosoma-Regla***" cuyo genotipo se codifica como una sola secuencia

de bits, dividida en siete cadenas de ocho bits, cada una de las cuales representa al valor que puedan tomar cada una de las siete condiciones del antecedente de la regla, y para el cual se mantiene fijo un consecuente conocido y previamente declarado como condición objetivo.

Cromosoma individuo.- Define la composición genética de una de las características del individuo "***Cromosoma-Regla***" para una generación. Cada cromosoma individuo define a un atributo único para el individuo o regla de asociación. Una instancia de cromosoma individuo contiene la secuencia de ocho bits que define una característica propia del individuo sobre algún aspecto del mismo. Son características del individuo las variables de cimientos, estructura, pared, cimientos, usos de la edificación y origen del financiamiento.

Regla asociación.- Describe las reglas resultantes del proceso de extracción de conocimiento mediante el algoritmo genético. Incluye las condiciones del antecedente y la condición objetivo o consecuente. Contiene los parámetros de cobertura y confianza de la regla, usados para calcular la adaptabilidad de la regla durante la ejecución del algoritmo genético.

Tipo cromosoma.- Consiste en la agrupación de todos los valores que describen a los atributos de las variables discretas que componen el antecedente del “Cromosoma Regla” y cuyos valores sirven para conformar el genotipo del individuo. Se especifican los puntos de cruce y mutación dentro de las cadenas de bits que representan a cada característica genética de la población de reglas.

Cromosoma gen.- Permite tener una correspondencia binario-decimal entre los posibles valores que pueden tomar los tipos de cromosoma en el genotipo del “cromosoma-regla” o individuo de una generación. Para los valores que puedan tomar los tipos de cromosoma se han establecido rangos máximos y mínimos.

3.1.3.2 Diagrama del modelo de objetos.

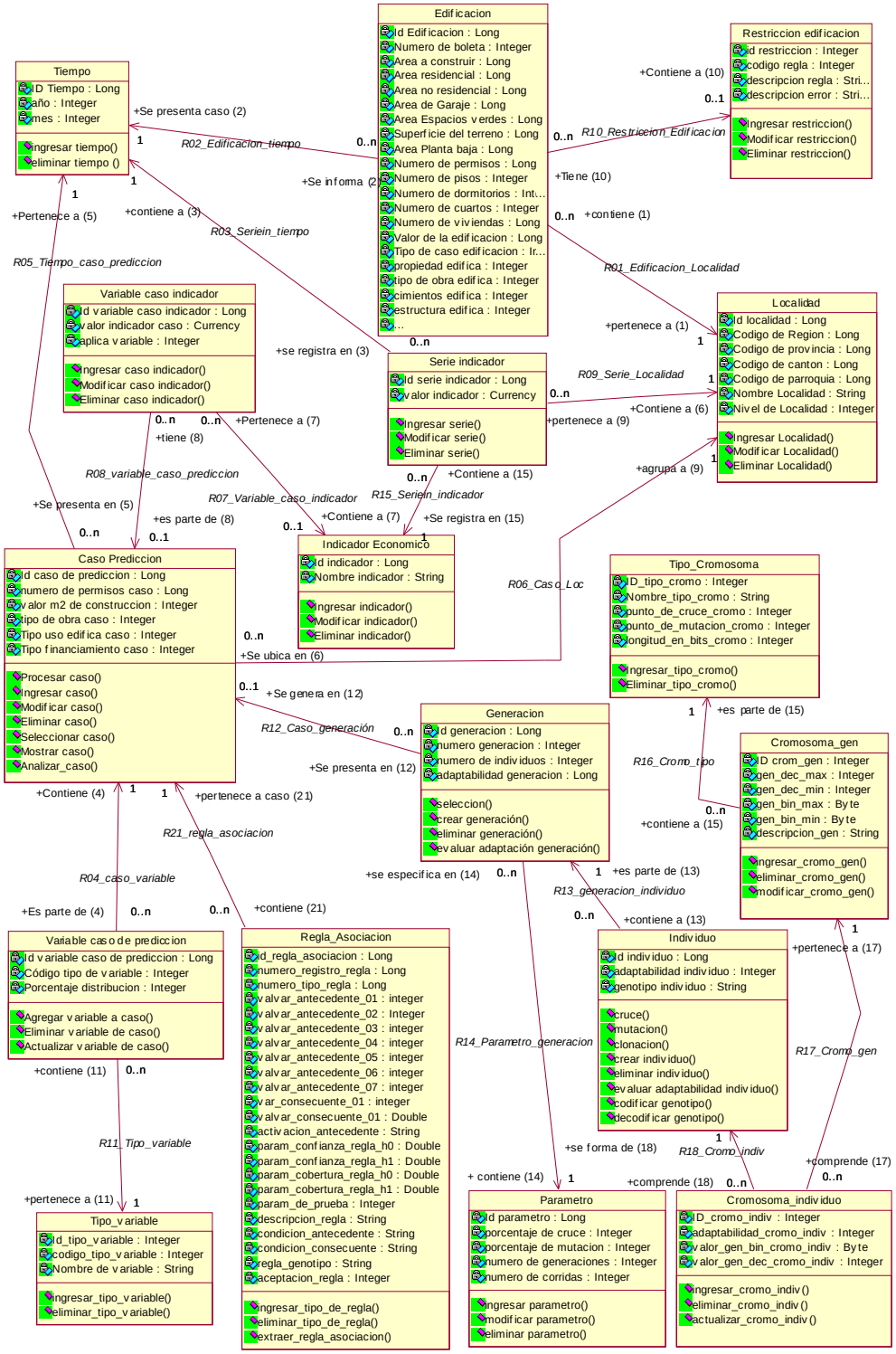


Figura 3.1. Diagrama del modelo de objetos.

Tipo variable.- Consiste en la codificación de todas las variables del fenómeno con todos sus niveles de desagregación posibles. Guarda relación con el tipo-cromosoma y cromosoma-gen creados para las variables del antecedente. Determina la selección del tipo de variable objetivo o consecuente y el valor discreto que puede tomar en la regla de asociación.

Usuario.- Identifica a los actores que intervienen en el sistema con sus roles específicos, tales como el administrador de la base de datos y la persona que realiza los procesos de análisis y predicción en el sistema.

3.1.3.3 Diccionario de datos.

ATRIBUTO	TIPO DE DATO	CLASE O ASOCIACION	DESCRIPCION
Id edificación	Numérico10	Edificación	Identificación única de la edificación de vivienda.
Área espacios verdes	Numérico 7	Edificación	Área destinada a espacios verdes en m2
Área planta baja	Numérico 7	Edificación	Área de planta baja de la edificación en m2
Área a construir	Numérico 7	Edificación	Área a construir de la edificación en m2
Área de garaje	Numérico 7	Edificación	Área destinada a garaje en m2
Área no residencial	Numérico 7	Edificación	Área no residencial de la edificación
Área residencial	Numérico 7	Edificación	Área residencial de la edificación
Número de boleta	Numérico 5	Edificación	Número de la boleta o cuestionario
Número de cuartos	Numérico 3	Edificación	Número de cuartos en la vivienda
Número de dormitorios	Numérico 3	Edificación	Número de dormitorios en la vivienda
Número de permisos	Numérico 3	Edificación	Número de permisos de construcción
Número de pisos	Numérico 2	Edificación	Número de pisos en la construcción
Número de viviendas	Numérico 3	Edificación	Número de viviendas
Superficie del Terreno	Numérico 7	Edificación	Superficie total del terreno
Tipo de caso edificación	Numérico 1	Edificación	Tipo de caso de edificación: histórico o actual.
Valor de la edificación	Numérico 15	Edificación	Valor de la edificación en dólares americanos.
Cimientos edifica	Numérico 3	Edificación	Material predominante en cimientos
Código de área urbana	Numérico 1	Edificación	Código de área urbana
Cubierta edifica	Numérico 3	Edificación	Material predominante en cubierta
Tipo área a construir edifica	Numérico 1	Edificación	Tipo del área a construir de la edificación
Tipo permiso financiamiento	Numérico 1	Edificación	Tipo del permiso según financiamiento
Tipo superficie edifica	Numérico 1	Edificación	Tipo de la superficie de la edificación
Tipo uso edifica	Numérico 2	Edificación	Tipo de uso de la edificación
Tipo valor financiamiento	Numérico 2	Edificación	Tipo del valor del financiamiento
Estructura edifica	Numérico 3	Edificación	Material predominante en la estructura
Pared edifica	Numérico 3	Edificación	Material predominante en las paredes
Propiedad edifica	Numérico 1	Edificación	Tipo de propiedad de la edificación: privada o pública.
Tipo de obra edifica	Numérico 1	Edificación	Tipo de obra de edificación: nueva construcción, ampliación o reconstrucción.
Id restricción	Numérico 8	Restricción Edificación	Identificación de restricción edificación
Código regla	Numérico 3	Restricción Edificación	Código de regla de validación
Descripción regla	Caracter 249	Restricción Edificación	Descripción de la regla de validación
Descripción error	Caracter 249	Restricción Edificación	Descripción del error al aplicar la regla de validación.
Id Cromosoma Gen	Numérico 8	Cromosoma Gen	Identificación de un gen en el cromosoma
Número de cromosoma gen	Numérico 4	Cromosoma Gen	Número de gen en el cromosoma
Código de cromosoma gen	Numérico 4	Cromosoma Gen	Código del gen en el cromosoma
gen bin max	Caracter 8	Cromosoma Gen	Número máximo en binario para gen en el cromosoma
gen bin min	Caracter 8	Cromosoma Gen	Número mínimo en binario para gen en el cromosoma
gen dec max	Caracter 8	Cromosoma Gen	Número máximo en decimal para gen en el cromosoma
gen dec min	Caracter 8	Cromosoma Gen	Número mínimo en decimal para gen en el cromosoma
descripcion gen	Caracter 80	Cromosoma Gen	Descripción del gen en el cromosoma
adaptabilidad cromosoma	Numérico 15,2	Cromosoma Individuo	Valor de adaptabilidad de un cromosoma-regla
valor gen bin cromosoma	Numérico 8	Cromosoma Individuo	Secuencia de bits que representan al valor del cromosoma
valor gen dec cromosoma	Numérico 3	Cromosoma Individuo	Representación decimal del valor del cromosoma
Id parámetro	Numérico 8	Parámetro	Identificación de parámetro
Porcentaje de cruce	Numérico 6,2	Parámetro	Porcentaje de individuos para aplicar cruce
Porcentaje de mutación	Numérico 6,2	Parámetro	Porcentaje de individuos para aplicar mutación.
Número de generaciones	Numérico 5	Parámetro	Número de generaciones de individuos

Número de corridas	Numérico 5	Parámetro	Número de corridas del algoritmo genético.
adaptabilidad convergencia	Numérico 6,2	Parámetro	Índice de adaptabilidad de convergencia máximo
gap generacional	Numérico 6,2	Parámetro	Porcentaje de individuos que permanecen sin cambios
Id generación	Numérico 8	Generación	Identificación de generación
Número generación	Numérico 5	Generación	Número secuencial de generación
Número de individuos	Numérico 15	Generación	Número de individuos para la generación
adaptabilidad generación	Numérico 15,2	Generación	Valor de adaptabilidad de la generación
estado generación	Numérico 1	Generación	Estado de proceso de la generación en el algoritmo
Id individuo	Numérico 8	Individuo	Identificación de individuo
adaptabilidad individuo	Numérico 15,2	Individuo	Valor de adaptabilidad de individuo
probabilidad individuo	Numérico 6.2	Individuo	Probabilidad de aparición del individuo en la generación
genotipo individuo	Caracter 249	Individuo	Genotipo de individuo: código de características genéticas
id regla asociación	Numérico 8	Regla Asociación	Identificación de regla de asociación
Número registro regla	Numérico 5	Regla Asociación	Número de registro de regla de asociación
valvar antecedente_01	Numérico 3	Regla Asociación	Valor de la condición atributo 01 del antecedente
valvar antecedente_02	Numérico 3	Regla Asociación	Valor de la condición atributo 02 del antecedente
valvar antecedente_03	Numérico 3	Regla Asociación	Valor de la condición atributo 03 del antecedente
valvar antecedente_04	Numérico 3	Regla Asociación	Valor de la condición atributo 04 del antecedente
valvar antecedente_05	Numérico 3	Regla Asociación	Valor de la condición atributo 05 del antecedente
valvar antecedente_06	Numérico 3	Regla Asociación	Valor de la condición atributo 06 del antecedente
valvar antecedente_07	Numérico 3	Regla Asociación	Valor de la condición atributo 07 del antecedente
var consecuente_01	Numérico 3	Regla Asociación	Código de la variable atributo del consecuente
valvar consecuente_01	Numérico 15,2	Regla Asociación	Valor de la condición atributo del consecuente
activación antecedente	Caracter 7	Regla Asociación	Cadena de activación de los atributos del antecedente
param confianza regla_h0	Numérico 15,2	Regla Asociación	Parámetro de confianza de la regla de asociación calculada por el algoritmo genético para hipótesis nula.
param confianza regla_h1	Numérico 15,2	Regla Asociación	Parámetro de confianza de la regla de asociación de selección en ejemplos para contraste de hipótesis
param cobertura regla_ho	Numérico 15,2	Regla Asociación	Parámetro de cobertura de la regla de asociación calculada por el algoritmo genético para hipótesis nula
param cobertura regla_h1	Numérico 15,2	Regla Asociación	Parámetro de cobertura de la regla de asociación de selección de ejemplos para contraste de hipótesis
param de prueba	Numérico 2	Regla Asociación	Parámetro de Prueba seleccionado entre confianza y cobertura
descripción regla	Carácter 249	Regla Asociación	Descripción de la regla de asociación
condición antecedente	Carácter 249	Regla Asociación	Descripción del antecedente para hacer búsquedas
condición consecuente	Carácter 249	Regla Asociación	Descripción del consecuente para hacer búsquedas
regla genotipo	Carácter 249	Regla Asociación	Equivalente de la regla en código binario para la estructura del algoritmo genético
aceptación regla	Numérico 1	Regla Asociación	Indicador si la confianza de la regla es o no aceptada luego de la prueba de hipótesis.
Id caso predicción	Numérico 8	Caso Predicción	Identificación única del caso de predicción
Código de caso predicción	Numérico 8	Caso Predicción	Código secuencial del caso de predicción
Código de variable a predecir	Numérico 2	Caso Predicción	Código de la variable a predecir
Valor de la variable a predecir	Numérico 15,2	Caso Predicción	Valor de la variable a predecir
Descripción del caso	Carácter 249	Caso Predicción	Descripción del caso de predicción
Nivel de tiempo del caso	Numérico 1	Caso Predicción	Nivel de tiempo en años o meses de la predicción
Nivel de localidad del caso	Numérico 1	Caso Predicción	Nivel de localidad según división territorial para predicción
Id grupo restricción	Numérico 8	Grupo Restricción	Identificación única de grupo de restricciones
Código grupo restricción	Numérico 3	Grupo Restricción	Código de grupo de restricciones
Id Indicador	Numérico 4	Indicador económico	Identificación única de Indicador económico
Código indicador	Numérico 4	Indicador económico	Código de identificador económico
Nombre Indicador	Carácter 120	Indicador económico	Nombre de identificador económico
Nivel de tiempo Indicador	Numérico 1	Indicador económico	Nivel de tiempo del Indicador económico
Nivel de localidad Indicador	Numérico 1	Indicador económico	Nivel de localidad del Indicador económico
Unidad Indicador	Numérico 1	Indicador económico	Unidad del indicador económico
Detalle Indicador	Carácter 120	Indicador económico	Detalle del Indicador económico
Id serie Indicador	Numérico 8	Serie Indicador	Identificación de la serie del indicador económico
Valor Indicador	Numérico 15,2	Serie Indicador	Valor del indicador económico en la serie
Variación porcentual anual	Numérico 6,2	Serie Indicador	Variación porcentual anual del indicador en la serie
Variación porcentual mensual	Numérico 6,2	Serie Indicador	Variación porcentual mensual del indicador en la serie
Id Localidad	Carácter 10	Localidad	Identificación única de la localidad geográfica
Anio localidad	Numérico 4	Localidad	Año base de actualización de la localidad geográfica
Código de región	Carácter 1	Localidad	Código de región de la localidad según división territorial
Código de provincia	Carácter 2	Localidad	Código de provincia de localidad según división territorial
Código de cantón	Cáracter 2	Localidad	Código de cantón de localidad según división territorial
Código de parroquia	Carácter 2	Localidad	Código de parroquia de localidad según división territorial
Nombre de localidad	Carácter 2	Localidad	Nombre de la localidad geográfica.
Nivel de localidad	Numérico 1	Localidad	Nivel de localidad geográfica
Id Tiempo	Carácter 6	Tiempo	Identificación única de registro de tiempo
Año Tiempo	Numérico 4	Tiempo	Año del registro de tiempo
Mes Tiempo	Numérico 2	Tiempo	Mes del registro de tiempo
Nivel de Tiempo	Numérico 1	Tiempo	Nivel del registro de tiempo en años o meses
Id Tipo cromo	Numérico 8	Tipo Cromosoma	Identificación única del Tipo cromosoma
Cod Tipo Cromo	Numérico 4	Tipo Cromosoma	Código de Tipo cromosoma
Nombre Tipo Cromo	Carácter 80	Tipo Cromosoma	Nombre del Tipo Cromosoma
Puntos de cruce cromo	Carácter 8	Tipo Cromosoma	Puntos de cruce en el cromosoma
Puntos de mutación cromo	Carácter 8	Tipo Cromosoma	Puntos de mutación en el cromosoma
Logitud en bits cromo	Numérico 2	Tipo Cromosoma	Longitud en bits del cromosoma

Id tipo variable	Numérico 8	Tipo Variable	Identificación única de registro en tipo variable
Cod Tipo Variable	Numérico 3	Tipo Variable	Código secuencial de tipo variable
Cod Clase Variable	Numérico 2	Tipo Variable	Código de clase a la que pertenece la variable
Nombre de variable	Carácter 80	Tipo Variable	Nombre descriptivo de la variable
Nivel de variable	Numérico 1	Tipo Variable	Nivel de la variable, si es categoría o variable
Id Variable caso de predicción	Numérico 8	Variable Caso de Predicción	Identificación de la Variable principal en el caso de predicción
Porc var caso	Numérico 6,2	Variable Caso de Predicción	Porcentaje de variación de la variable para el caso de predicción
Valor var caso	Numérico 15,2	Variable Caso de Predicción	Valor de la variable para el caso de predicción
Porc distmax var caso	Numérico 6,2	Variable Caso de Predicción	Porcentaje máximo que toma la variable en la distribución porcentual
Porc distmin var caso	Numérico 6,2	Variable Caso de Predicción	Porcentaje mínimo que toma la variable en la distribución porcentual
Id variable caso indicador	Numérico 8	Variable Caso Indicador	Identificación de la Variable Indicador para el Caso de Predicción
Valor indicador caso	Numérico 15,2	Variable Caso Indicador	Valor de la variable indicador en el caso de predicción
Variación porc anual	Numérico 6,2	Variable Caso Indicador	Variación porcentual anual de la variable indicador en el caso de predicción
Variación porc mensual	Numérico 6,2	Variable Caso Indicador	Variación porcentual mensual de la variable indicador en el caso de predicción
Aplica variable	Numérico 1	Variable Caso Indicador	Aplicabilidad de la variable indicador al caso de predicción
Id nusu usu	Numérico 2	Usuario	Identificación única de usuario
scla usu	Carácter 15	Usuario	Clave de usuario
bestado usu	Numérico 1	Usuario	Estado de conexión del usuario, conectado o no conectado
spass usu	Carácter 15	Usuario	Contraseña de usuario
sroles usu	Carácter 20	Usuario	Rol al que pertenece el usuario

Tabla 3.4. Análisis: Modelo de Objetos - Descripción de Atributos

OPERACIÓN	CLASE O ASOCIACION	DESCRIPCION
Ingresar restricción	Restricción Edificación	Ingresar la descripción de una restricción para validar los datos de una edificación. Todo caso de edificación debe someterse a evaluación respecto a la consistencia de datos en su estructura.
Modificar restricción	Restricción Edificación	Modificar la descripción de una restricción para validar datos de una edificación
Eliminar restricción	Restricción Edificación	Elimina la descripción de una restricción para validar datos de una edificación.
Preprocesar caso	Caso de Predicción	Consiste en preparar datos para el caso de predicción, de modo que se limiten los datos históricos a una localidad y tiempo asignados.
Procesar caso	Caso de Predicción	Se aplica el algoritmo genético al caso de predicción que previamente ha sido seleccionado y preprocesado.
Crear generación	Generación	Creación al azar de individuos que representan nuevos casos de edificaciones para un caso de predicción, y que son sometidos a un proceso de adaptación.
Eliminar generación	Generación	Operación que a través de un proceso de selección, elimina una generación creada y la sustituye por nuevos individuos por los sobrevivientes típicos no repetidos de la generación anterior
Evaluar adaptación generación	Generación	Consiste en evaluar adaptabilidad de la generación de acuerdo a la utilidad de sus individuos y a un algoritmo de selección aplicado a toda la generación.
Crear Individuo	Individuo	Creación al azar de un individuo para una nueva generación sometida a un proceso de adaptación.
Eliminar Individuo	Individuo	Suprimir un individuo de la generación por no cumplir con los requerimientos mínimos de adaptabilidad respecto al ambiente.
Evaluar adaptabilidad individuo	Individuo	Consiste en evaluar la adaptabilidad de un solo individuo de acuerdo a la aptitud de su estructura genética y como resultado de aplicar un procedimiento de evaluación de su adaptabilidad.
Selección	Generación	Aplicar la selección natural del algoritmo genético en la cual prevalecen los individuos más aptos de la generación.
Cruce	Individuo	Operación genética aplicada a un individuo en la generación, en el cual se intercambian bits de dos cromosomas en forma aleatoria para obtener un nuevo individuo a ser seleccionado.
Mutación	Individuo	Operación genética aplicada a un individuo en la generación, en la cual se alteran bits del cromosoma, para producir un nuevo individuo con nuevas características a ser evaluadas.
Clonación	Individuo	Operación genética aplicada un individuo de la generación, para obtener una copia exacta del mismo que se integre a la nueva generación.
Codificar genotipo	Individuo	Transformar a binario un valor decimal en la estructura del genotipo del individuo.
Decodificar genotipo	Individuo	Transformar a decimal un valor binario de la estructura del genotipo del individuo
Ingresar caso de edificación	Edificación	Consiste en registrar al sistema un caso de edificación, el cual se presenta en una determinada localidad geográfica y en una fecha del año.
Ingresar serie de indicador	Serie indicador económico	Consiste en registrar al sistema una serie de valores para un indicador macroeconómico para un mes o año determinado.

-

OPERACIÓN	CLASE O ASOCIACION	DESCRIPCION
Ingresar caso	Caso de Predicción	Consiste en registrar al sistema un caso de predicción, el cual consiste en las especificaciones de lugar y tiempo en el cual se han declarado permisos de construcción de viviendas, y de cuyas variables se trata de obtener reglas de comportamiento de aplicabilidad futura.
Eliminar edificación	Edificación	Consiste en eliminar del sistema un caso de edificación registrado en el tiempo para una localidad geográfica.
Eliminar serie	Serie indicador económico	Consiste en eliminar del sistema una serie de indicador económica en el tiempo.
Eliminar caso	Caso de predicción	Consiste en eliminar del sistema un caso de predicción registrado.
Modificar edificación	Edificación	Consiste en modificar los valores de las variables atributos de un caso de edificación registrado en el sistema.
Modificar serie	Serie indicador económico	Consiste en modificar los datos de una serie de indicador macroeconómico registrados en el sistema para análisis.
Modificar caso	Caso de predicción	Consiste en modificar los valores de las variables atributos del caso de predicción registrado en el sistema.
Ingresar indicador	Indicador Económico	Consiste en crear un nuevo indicador económico con su descripción y agregarlo al sistema.
Eliminar Indicador	Indicador Económico	Consiste en eliminar un indicador económico registrado en el sistema con todas sus series históricas.
Agregar variable a caso	Variable caso de Predicción	Consiste en agregar al caso de predicción las variables que intervienen en el mismo para el resultado del análisis.
Eliminar variable de caso	Variable caso de Predicción	Consiste en eliminar del sistema a una variable del conjunto de variables resultado del análisis.
Ingresar parámetro	Parámetro	Consiste en crear un nuevo conjunto de valores de parámetros para el algoritmo genético y registrarlos en el sistema.
Ingresar tipo de regla	Regla Asociación	Consiste en definir la conformación de la regla de asociación con los atributos que forman parte del antecedente y la condición del consecuente
Eliminar tipo de regla	Regla Asociación	Consiste en eliminar un tipo de conformación de regla
Extraer regla asociación	Regla Asociación	Consiste en extraer los datos resultado de la regla, es decir, los valores de las condiciones de los atributos antecedentes, obtenidos mediante la aplicación del algoritmo genético.

Tabla 3.5. Análisis: Modelo de Objetos - Descripción de Operaciones

3.1.4 MODELO DINÁMICO.

3.1.4.1 Identificación de escenarios y eventos

En el Sistema SPACV los escenarios más relevantes se producen entre las clases edificación, serie indicador económico, parámetros de predicción, caso de predicción y reglas de asociación. Los actores de los eventos son: el sistema, el **analista experto** y el analista predictor. El analista experto se encarga de transmitir al sistema, el conocimiento histórico del fenómeno de la construcción de viviendas en el país, esto es, los casos de construcción de viviendas por localidad y tiempo, y demás datos secundarios relacionados con el fenómeno, como indicadores, restricciones y reglas del sistema. El **analista predictor** es el usuario que define los casos de predicción y los procesa en el sistema.

1) Introducir casos de predicción al sistema

Descripción: El analista predictor ingresa al sistema las características del caso de predicción. El sistema examina el caso y pide los parámetros. El usuario especifica los parámetros. El sistema queda en espera hasta procesar el caso.

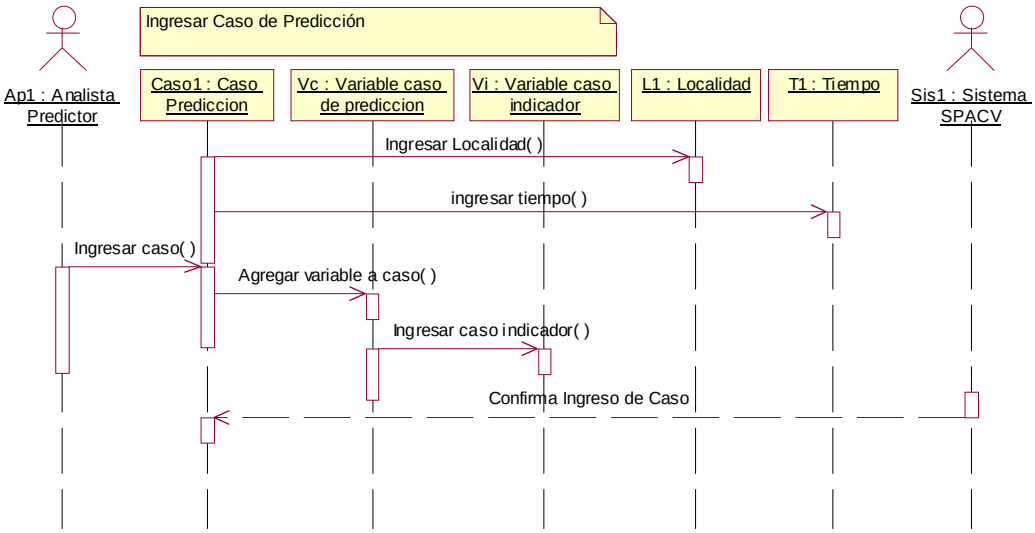


Figura 3.2. Escenario para introducir casos de predicción al sistema

Tabla de Escenario Introducir casos de predicción al sistema.			
Nombre	**Género**	**Agente**	**Descripción**
Ingresar caso	Operación	Analista Predictor	Ingresa caso de predicción
Agregar variable a caso	Operación	Analista Predictor	Agrega variable a caso de predicción
Ingresar caso indicador	Operación	Analista Predictor	Ingresa variables indicador al caso
Mensaje de confirmación de ingreso	Evento	Analista Predictor	Mensaje de confirmación de ingreso y registro del caso de predicción

Tabla 3.6. Análisis – Modelo Dinámico – Escenario para introducir casos de predicción.

2) Procesamiento de caso de predicción en el sistema

Descripción: El usuario predictor selecciona el caso analizado. También construye o define el tipo de regla de asociación a extraer y actualiza los parámetros del algoritmo genético. El sistema construye el genotipo de la regla y aplica el algoritmo genético para extraer las reglas de asociación de más alto valor de confianza, Luego se consulta los resultados de las reglas extraídas y

para cada una se aplica las pruebas de hipótesis respectivas sobre una muestra diferente o sobre toda la población para evaluar la persistencia de la regla.

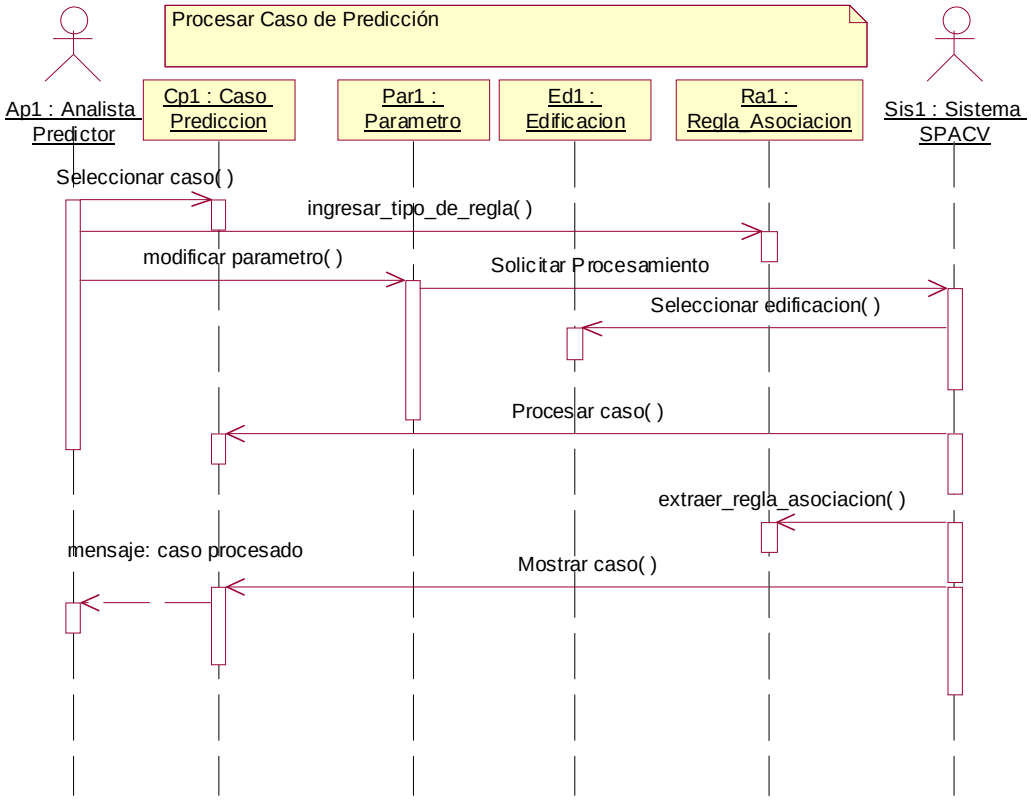


Figura 3.3. Escenario para Procesamiento de caso de predicción

Tabla de Escenario para procesamiento de caso de predicción.			
Nombre	**Género**	**Agente**	**Descripción**
Seleccionar caso	Operación	Analista Predictor	Selecciona el caso de predicción
Modificar parámetro	Operación	Analista Predictor	Modifica parámetros para el caso
Solicita procesamiento	Evento	Analista Predictor	Solicita procesamiento al Sistema
Seleccionar Edificación	Operación	Sistema	Recupera y selecciona casos de ejemplo de edificaciones anteriores
Procesar caso	Operación	Sistema	Procesa caso y aplica algoritmo
Mostrar caso	Operación	Sistema	Muestra reglas extraídas para el caso procesado
Mensaje de Fin de Proceso	Evento	Sistema	Mensaje de fin de proceso algoritmo

Tabla 3.7. Análisis – Modelo Dinámico – Escenario para procesamiento de caso de predicción.

3) Ejecución del algoritmo genético

Descripción: Se construye la primera generación de individuos, con su estructura binaria y decimal de cromosomas y genes que representan a los valores que

toman las variables del antecedente de las reglas de asociación. Se inicializan los valores de adaptación para cada cromosoma-regla y se somete la generación a los operadores genéticos para evaluar la adaptabilidad de los individuos. En el proceso de adaptación, se aplica la función de adaptabilidad calculado en base a los parámetros de cobertura y confianza que tiene cada regla en la muestra de ejemplos.

Tabla de Escenario para ejecución del algoritmo genético.			
Nombre	**Género**	**Agente**	**Descripción**
Seleccionar caso	Operación	Analista Predictor	Selecciona el caso de predicción
Modificar parámetro	Operación	Analista Predictor	Modifica parámetros para el caso
Solicita procesamiento	Evento	Analista Predictor	Solicita procesamiento al Sistema
Seleccionar Edificación	Operación	Sistema	Recupera y selecciona casos de ejemplo de edificaciones anteriores
Procesar caso	Operación	Sistema	Procesa caso y aplica algoritmo
Mostrar caso	Operación	Sistema	Muestra reglas extraídas para el caso procesado
Mensaje de Fin de Proceso	Evento	Sistema	Mensaje de fin de proceso algoritmo

Tabla 3.8. Análisis - Modelo Dinámico. Escenario para ejecución del algoritmo genético.

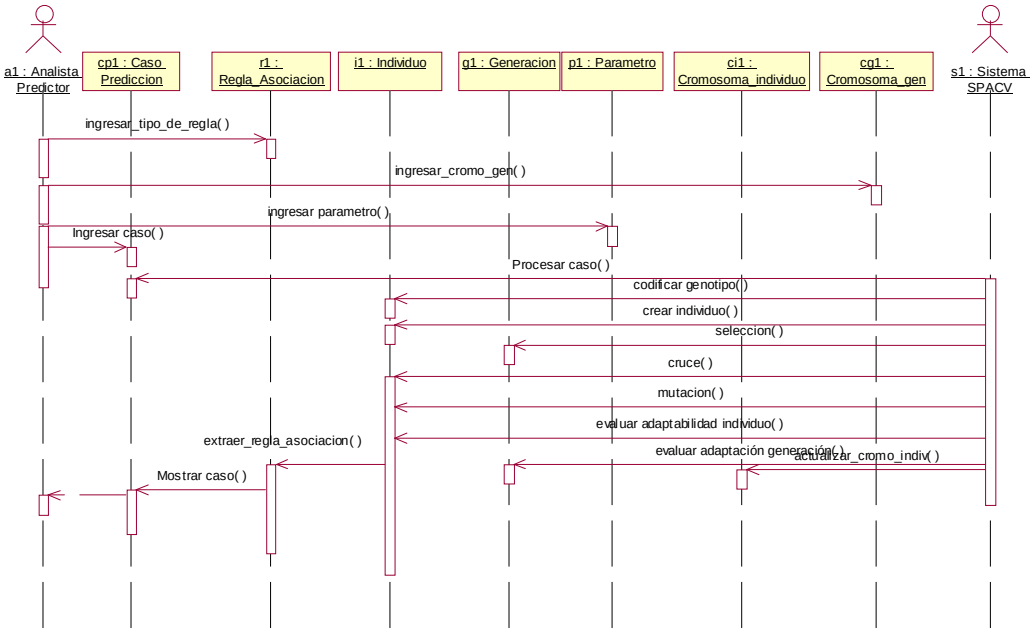


Figura 3.4. Escenario para ejecución del algoritmo genético.

En el proceso se asigna un puntaje de aceptación de los individuos de cada generación por la certeza de aparición que tuvieron en el pasado. Luego de un número preasignado de iteraciones, se evalúa al final la condición de parada

preestablecida en los parámetros. Los resultados se extraen de la última generación y son las reglas de asociación con sus valores para el antecedente.

3.1.4.2 Diagramas de transición de estados

1) <u>Diagrama de transición de estados de la clase caso de predicción</u>

Un caso de predicción es ingresado por el usuario con sus características de localidad y tiempo. Previamente deben haberse registrado los casos de ejemplo de edificaciones, la codificación de las variables, los parámetros del algoritmo genético (porcentaje de cruce, porcentaje de mutación, puntos de cruce, condición de parada), los tipos de reglas de asociación y las series de indicadores económicos complementarios para el análisis. Luego se aplica el algoritmo genético para extraer un cierto tipo de regla de asociación seleccionado.

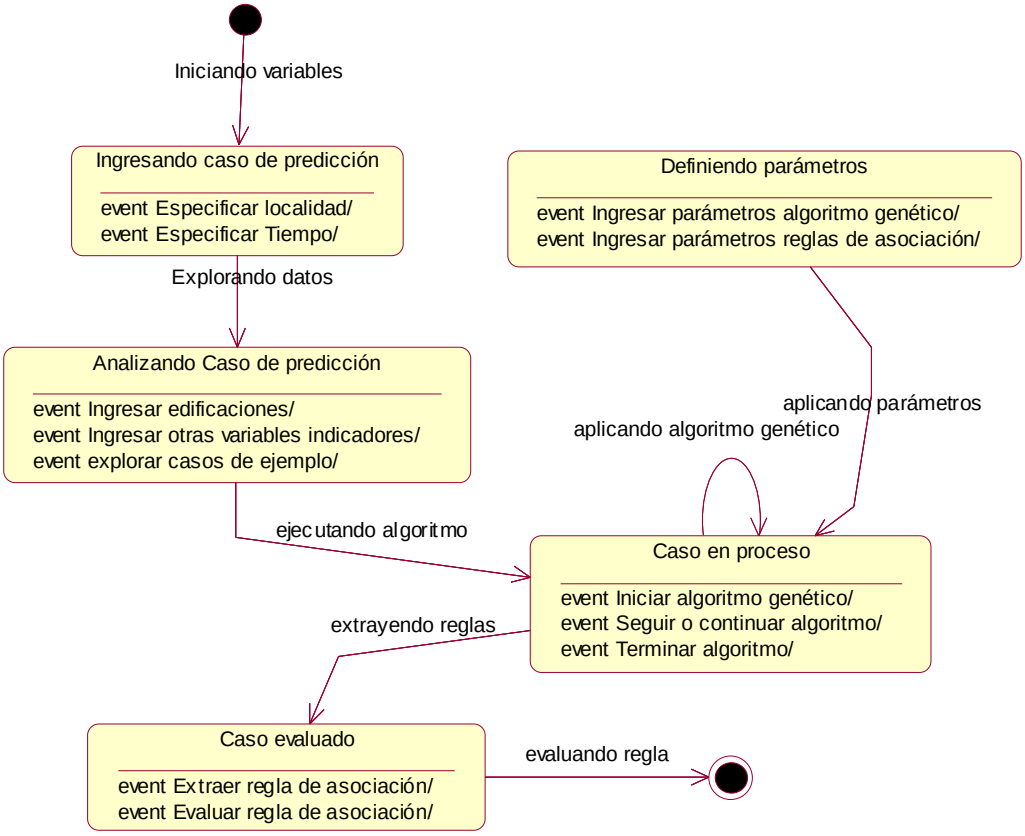


Figura 3.5. Diagrama de estados para clase caso de predicción.

2) Diagrama de transición de estados de la clase individuo

En "individuo" y "cromosoma_individuo" se realizan las operaciones del algoritmo genético y los resultados son extraídos hacia la clase "regla_asociación".

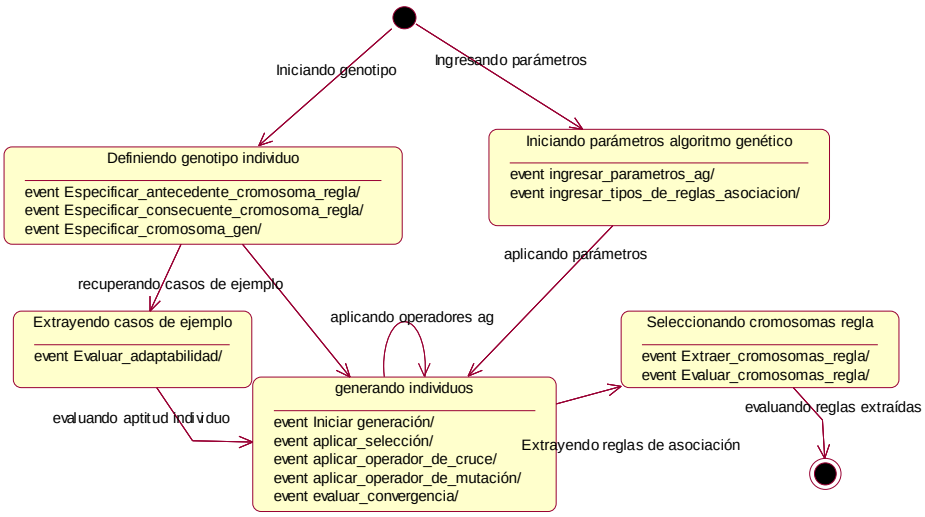


Figura 3.6. Diagrama de estados para clase individuo.

3.1.5 MODELO FUNCIONAL.

3.1.5.1 Diagrama de flujo de datos.

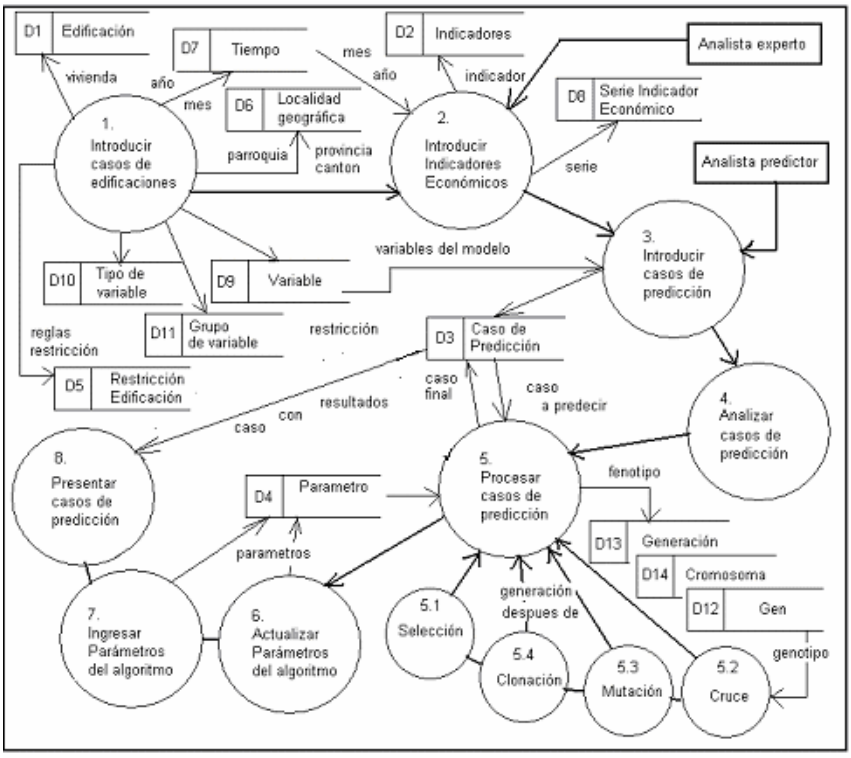


Figura 3.7. Diagrama de Flujo de datos.

Sistema: Sistema de Predicción y Análisis de construcción de viviendas (SPACV)

Entradas; Edificación, Caso de Predicción, Parámetro

Salidas: Generación, Caso de Predicción

Proceso; El Sistema SPACV se inicia con los procesos de introducción de datos en edificación y parámetros por medio del analista experto. Luego, el analista predictor ingresa el caso de predicción, realiza el preprocesamiento y procesamiento y obtiene resultados de la aplicación del algoritmo.

3.1.1.3 Especificación de operaciones y funciones.

Operación : **Proceso de Predicción.**

Entradas: Caso de Predicción; Restricción Edificaciones; Parámetros

Salidas: Caso de Predicción con resultado de la predicción

Funciones: Preprocesamiento; Procesamiento.

Descripción: El proceso de predicción se realiza a partir de un caso de predicción originado por una interfaz de usuario que selecciona la localidad geográfica y el tiempo expresado en años. Durante el proceso se extrae una muestra de ejemplos para evaluar los valores que tome el antecedente de una de las reglas de asociación escogida para evaluar su comportamiento. En la predicción se aplica el algoritmo genético para extracción de reglas hasta que se cumpla con una condición de convergencia o se supere el número de generaciones. El algoritmo se aplica para más de una regla, una a la vez y se entrega al final las reglas de asociación con su valoración de confianza y cobertura para ser evaluadas mediante prueba de contraste de hipótesis.

Función: **Análisis o Preprocesamiento**

Parámetros: variables del caso de predicción

Retorno: parámetros de valoración para el caso de predicción

Descripción: Con el caso de predicción delimitado en localidad geográfica y tiempo, se busca en la base histórica de datos de edificaciones casos de ejemplo que sirvan para hacer un análisis exploratorio de los datos, con estas variables. Este análisis es opcional y sirve para conocer la distribución de las variables.

Función: **Procesamiento**

Parámetros: Variables seleccionadas para el antecedente,
Variable objetivo para el consecuente y su valor constante
Parámetros del algoritmo genético.

Retorno: Reglas con valores en las condiciones del antecedente.

Descripción: Se genera la primera población de individuos con parámetros del algoritmo. Se aplica el algoritmo genético con los operadores de *selección, mutación y cruce* aplicados sobre estructuras genéticas de los cromosomas-regla cuyo contenido se crea a partir de secuencias de bits tomadas al azar entre 0-255 para cada atributo del antecedente de la regla. Tras sucesivas generaciones se aplica una condición de parada y como resultado se obtiene la población con las reglas de mayor adaptabilidad para sus valores en el antecedente, y que son el resultado de la predicción. Durante el proceso se extrae una muestra de ejemplos para evaluar la adaptabilidad de cada regla.

Función: **Algoritmo genético**.

Parámetros: caso de predicción con sus variables: parámetros.

Retorno: generación de reglas mejor adaptadas

Descripción: Se genera la primera población de individuos "cromosoma-reglas" creada al azar con un número de individuos dado por parámetro. Se aplican los operadores de selección, cruce y mutación y bajo una condición de parada, se retorna al final la distribución de población mejor adaptada.

Función: **Selección**.

Parámetros: cromosomas-regla.

Retorno: selección de reglas mejor adaptadas

Descripción: Se selecciona los individuos por torneo en la que se aplica una evaluación de la adaptabilidad de los cromosomas. La adaptabilidad se la calcula a partir del promedio entre los parámetros de confianza y completitud o cobertura. Para lo una muestra aleatoria de casos de ejemplo es extraída de la población de edificaciones en la localidad y tiempo del caso de predicción. Los individuos seleccionados son clonados a partir de sus padres mejor adaptados.

Función: **<u>Cruce</u>.**

Parámetros: cromosomas de cruce.

Retorno: Generación de individuos luego del cruce

Descripción: Elige al azar los individuos para el cruce. Se realiza el cruce entre los "cromosomas-regla" aplicando los puntos de corte, en un porcentaje de acuerdo a parámetros y sobre las cadenas de bits que representan a las variables que intervienen en el antecedente de las reglas.

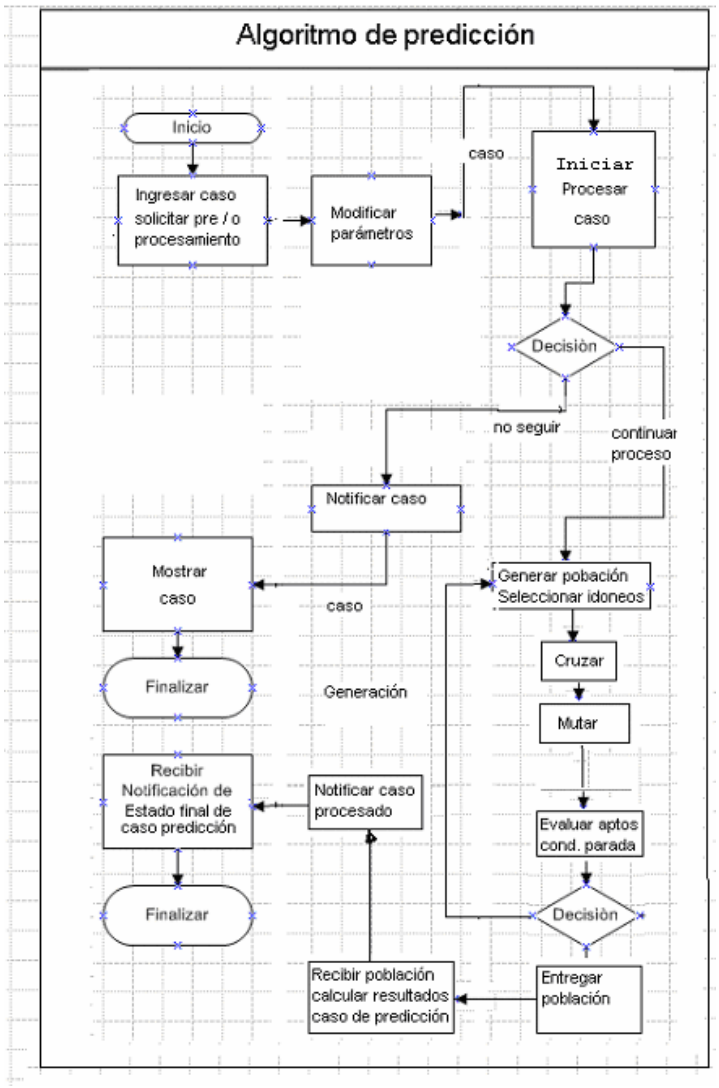


Figura 3.8. Flujo de Control de Algoritmo Genético

Función: **<u>Mutación.</u>**

Parámetros: cromosomas de mutación; Parámetros del algoritmo

Retorno: Generación con un porcentaje de individuos mutantes.

Descripción: Elige al azar individuos para aplicar mutación en porcentaje especificado por parámetros y sobre las cadenas de bits que representan a las variables que intervienen en el antecedente de las reglas. En un punto de corte al azar en los genes del cromosoma, se altera uno o varios bits por mutación.

3.1.5.2 Realización de casos de uso aplicados al modelo funcional.

El actor analista predictor proporciona los casos de predicción al sistema a través de la una interfaz para procesamiento. El sistema utiliza las clases generación, individuo, parámetro, variable de predicción, regla-asociación para obtener los resultados del caso de predicción. Se presenta al usuario las salidas desde la clase regla-asociación.

ACTOR	DESCRIPCION
Analista experto	Es aquel que introduce la información histórica de casos de edificaciones, series históricas de variables relacionadas con el fenómeno.
Analista predictor	Es aquel que presenta los casos de predicción al sistema, introduce los parámetros y espera las respuestas del sistema.
Sistema SPACV	Es el sistema automatizado, que realiza el proceso en base a las entradas recibidas del analista predictor, y el que entrega las respuestas.

Tabla 3.9. Análisis: Modelo Funcional - Identificación de actores

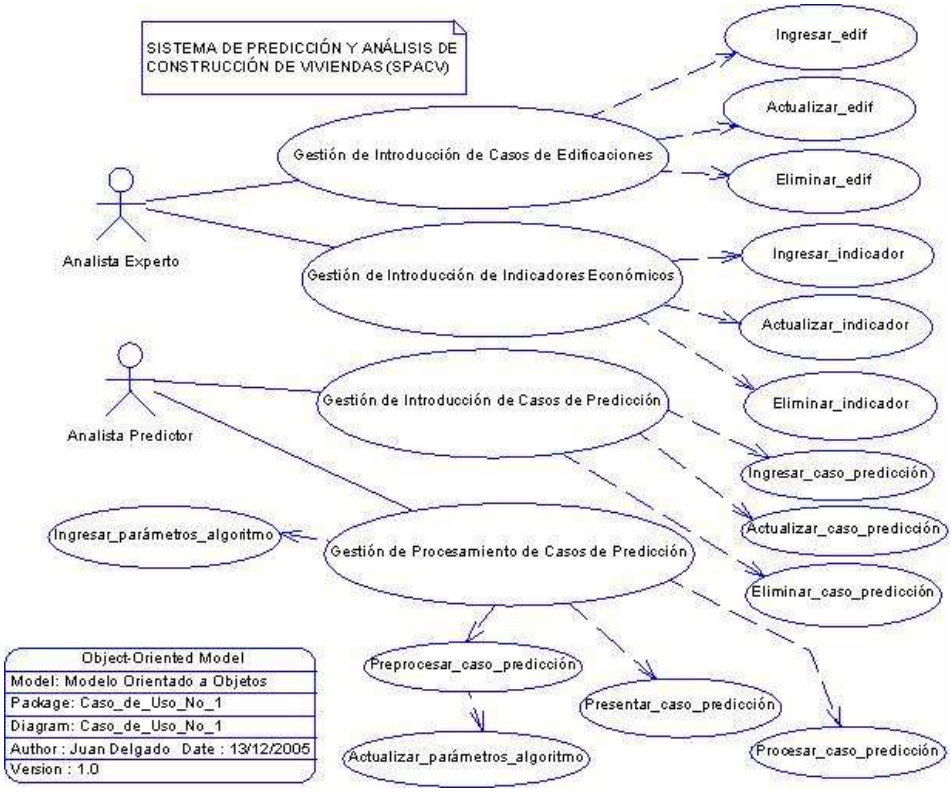


Figura 3.9. Caso de Uso del Sistema de Predicción y Análisis.

3.2 DISEÑO.

3.2.1 DISEÑO DEL SISTEMA.

3.2.1.1 Descripción de Subsistemas.

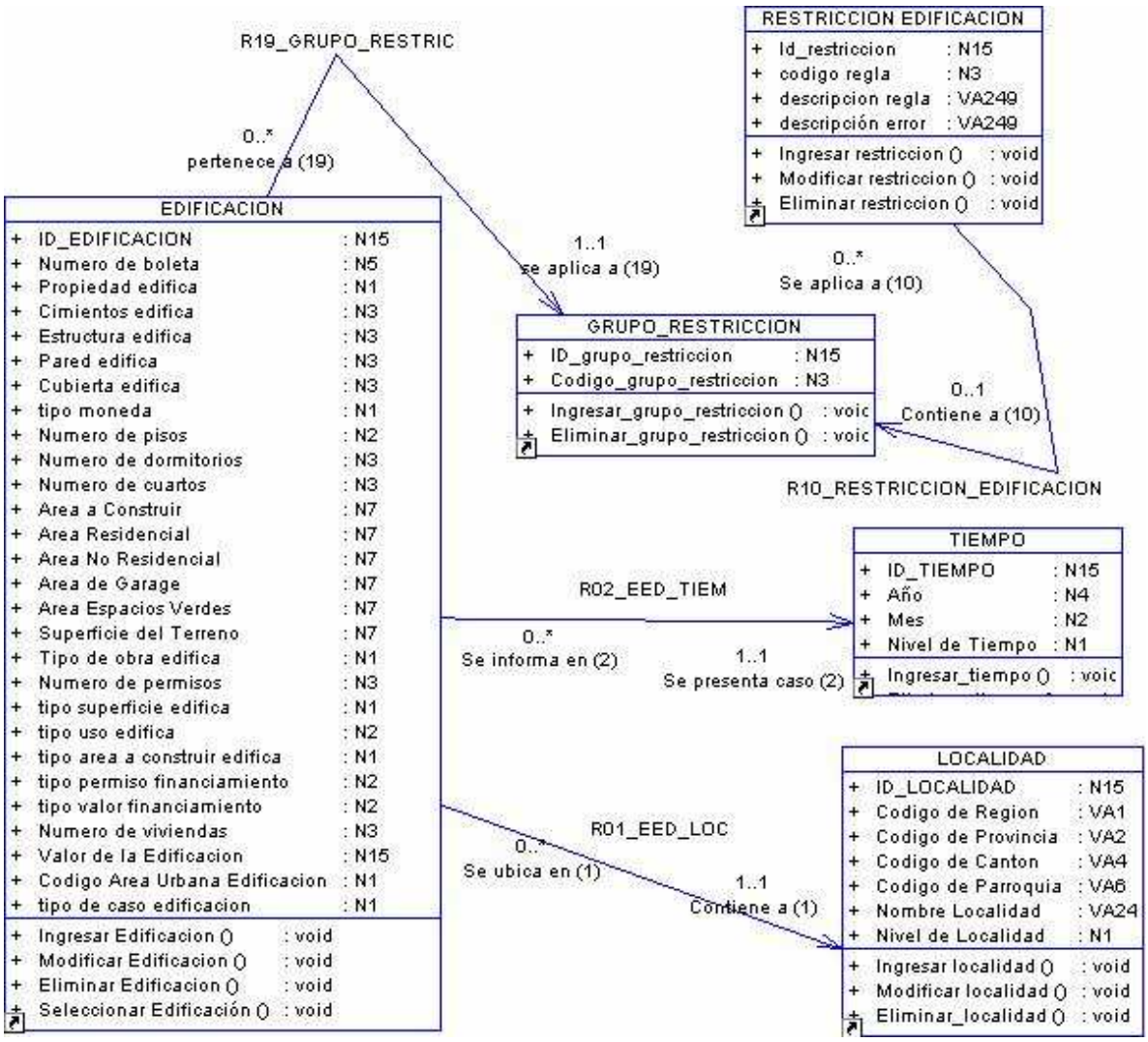


Figura 3.10. Diagrama de Clases del módulo Edificación

Subsistema Edificación.- En la clase edificación se registra los casos históricos de edificación de viviendas en tiempo y lugar. Las operaciones de estas clases son de mantenimiento (agregar, modificar, eliminar).

El nivel de localidad para los casos históricos es siempre el más bajo y se ubica en una parroquia, la cual pertenece a una región, provincia y cantón. El tiempo

en el que se registra una edificación es anual y mensual. Se almacenan descripciones de las restricciones aplicadas a los casos de edificación. También se definen en esta clase los mensajes de error al violar una restricción.

Subsistema Indicador Económico.- En tiempo y lugar, se registran las series de indicadores económicos representados por las clases Indicador Económico y Serie Indicador Económico.

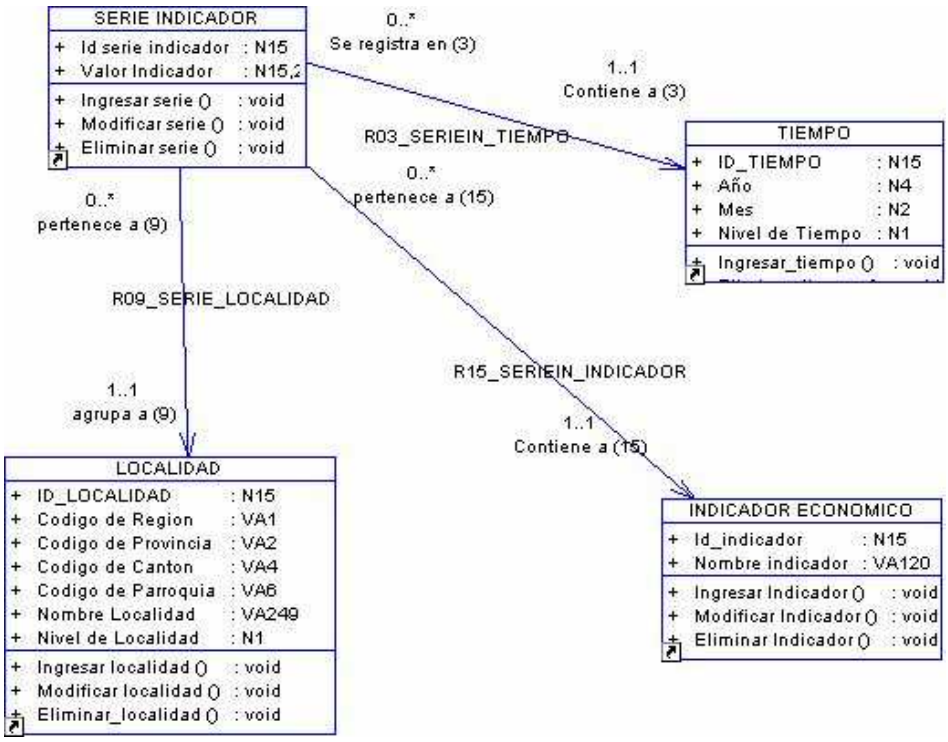


Figura 3.11. Diagrama de Clases del Módulo Indicador Económico.

En estas clases, se tiene operaciones de ingreso, eliminación y modificación.

Subsistema Predicción.- En la clase caso de predicción se realizan operaciones de procesamiento del algoritmo de predicción genético. Se enlaza al subsistema Algoritmo Genético a través de la clase generación. Un caso de predicción se representa por sus variables definidas en variable caso indicador y variable caso de predicción. En la clase variable caso indicador se definen los atributos para incorporar indicador económico a cada instancia de caso de predicción. En la clase variable caso de predicción tiene como atributos a las variables individuales que conforman las reglas de asociación, y que están presentes en tipo variable.

Los resultados de la predicción son las reglas descubiertas de asociación entre variables y se encuentran en la clase regla-asociación, y pueden ser una o más dependiendo del caso de predicción..

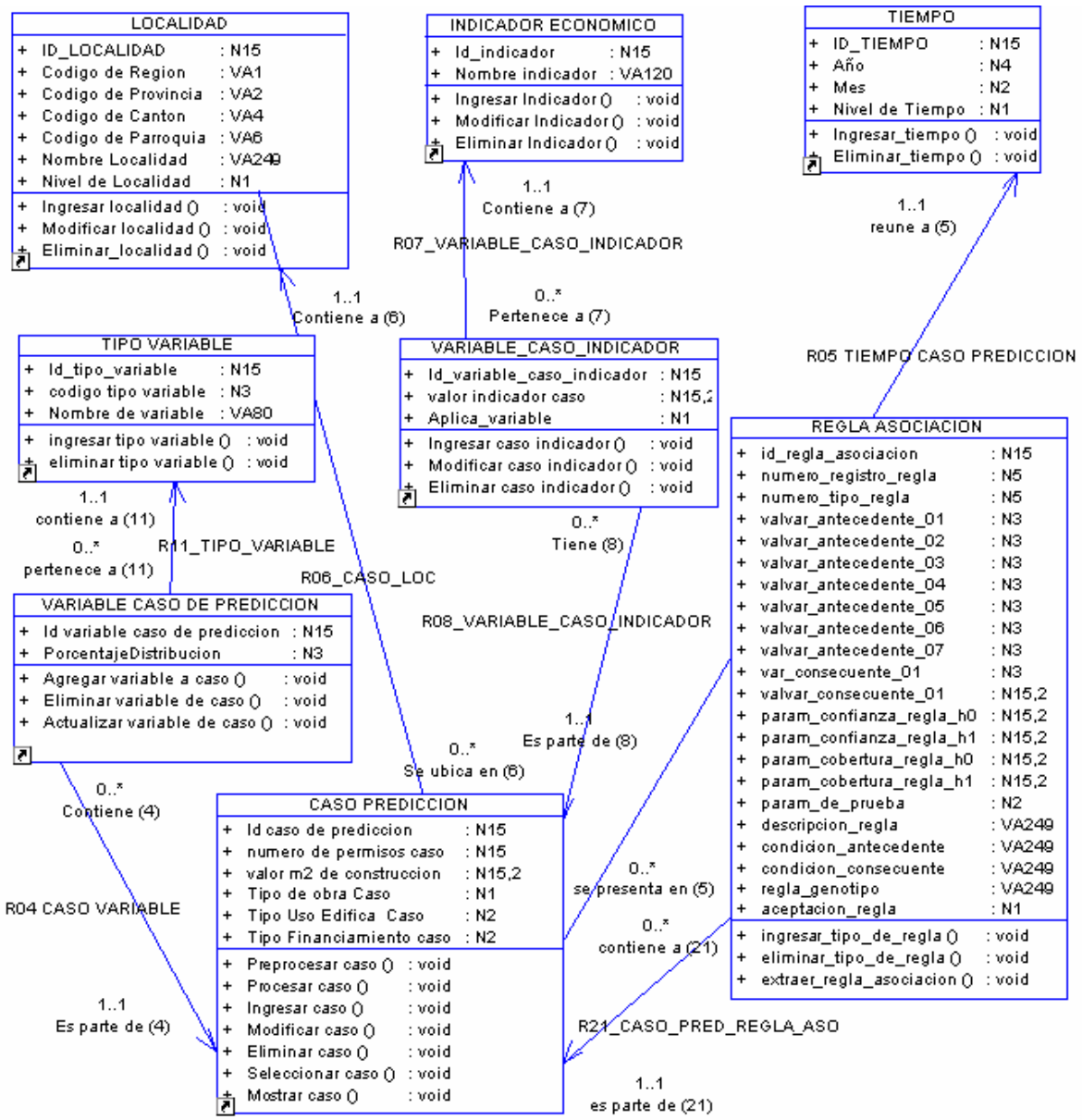


Figura 3.12. Diagrama de Clases del Módulo Predicción

Subsistema Algoritmo Genético.-

El subsistema Algoritmo de Predicción Genético está representado por las clases generación, parámetro, cromosoma-individuo, individuo, cromosoma-gen, tipo cromosoma.

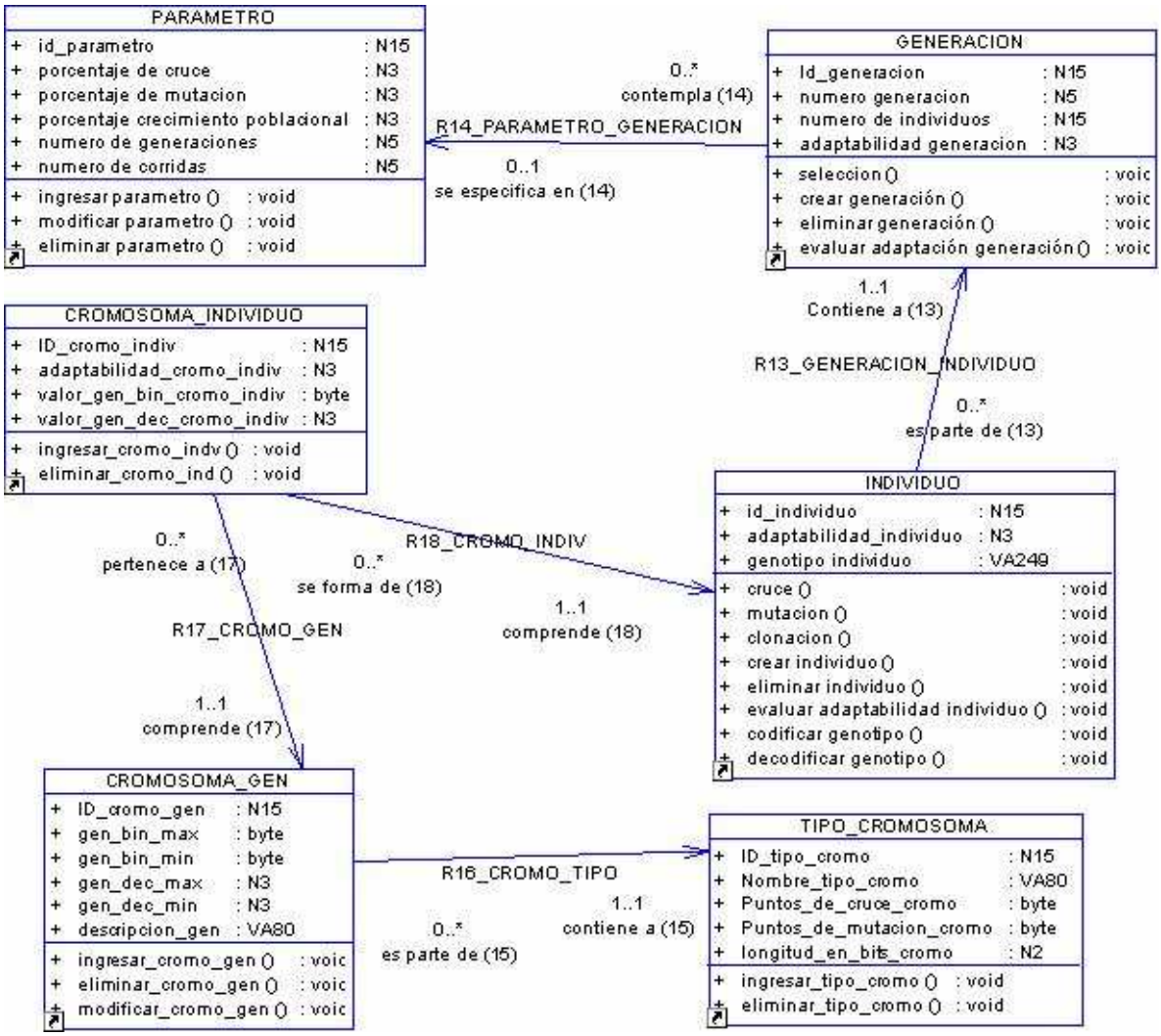


Figura 3.13. Diagrama de Clases del Módulo Algoritmo Genético.

Para cada caso de predicción se crean en forma sucesiva varias generaciones de individuos. Las transformaciones ocurren en las clases generación, individuo y cromosoma-individuo a través de los operadores genéticos de selección, cruce y mutación. Al final los resultados de las reglas descubiertas por el algoritmo quedan en la última generación y se extraen de cromosoma-individuo.

Diseño Estructural del Algoritmo Genético.

El algoritmo genético utiliza la clase parámetro para definir las especificaciones de ejecución. Cada individuo es un “cromosoma regla”, el cual es evaluado según el grado de adaptabilidad que soporta su estructura genética en función de muestras de casos de ejemplo extraídas al azar. Para el antecedente de la regla se registran las características del individuo como subcadenas de bits que representan a los atributos de cada condición del antecedente en cada cromosoma. La forma de representar los cromosomas de un individuo para un caso de predicción consiste en especificar en las clases Tipo-cromosoma y Cromosoma-Gen, tanto la definición, estructura y componentes de los cromosomas de la edificación como también sus atributos particulares.

CIMIENTOS	ESTRUCTURA	PARED	CUBIERTA	TIPO DE OBRA	TIPO DE USO	TIPO FINANCIAMIENTO
01111100	01111001	10001110	10000111	00000010	00000111	01111100

Figura 3.14. Representación genética del antecedente de un individuo Cromosoma-Regla

Se han considerado dos grupos de cromosomas en el antecedente: El primer grupo sirve para representar la composición física de los individuos en cuatro tipos de cromosomas vinculados a los materiales predominantes en la construcción. Estos son: cimientos, estructura, paredes, y cubierta o techo. El segundo grupo permite determinar el comportamiento de las edificaciones en el medio externo: tipo de obra, tipo de uso de la edificación y tipo de financiamiento.

La composición genética es única para cada cromosoma, expresada en cadenas de 56 bits, de las cuales se toman 8 bits para representar a cada una de las 7 variables que podrá formar parte del antecedente de la regla. Para saber si un atributo forma parte del antecedente de la regla se asigna a una variable del sistema una cadena de 7 bits, donde cada bit se utiliza para cada variable del antecedente y contiene un valor de 1 o 0 para indicar si está o no incluido en el antecedente. De esta manera se pueden registrar antecedentes de longitud variable, es decir, un antecedente puede estar formado por un número diferente de variables. En la figura 3.22 se aprecia una cadena de activación del

antecedente de una regla, en la cual los bits que tienen valor de 1 activan a los atributos del antecedente según la posición del bit en cadena.

0 1 1 1 1 1 0	CIMIENTOS= 0, (ESTRUCTURA, PAREDES, CUBIERTA, TIPO DE OBRA Y USO EDIFICACIÓN) = 1

Figura 3.15. Activación de atributos del antecedente de un individuo Cromosoma-Regla

La Tabla 3-16 muestra una vista de los cromosomas-individuo que definen la composición de los materiales predominantes de la vivienda y con los cuales se forma un cromosoma-regla.

COMPOSICION GENETICA EN CROMOSOMA INDIVIDUO PARA MATERIALES PREDOMINANTES EN LA CONSTRUCCIÓN					
NOMBRE TIPO CROMOSOMA	**DESCRIPCION GEN**	**GEN BIN MIN**	**GEN BIN MAX**	**GEN DEC MIN**	**GEN DEC MAX**
CIMIENTOS	Hormigón armado	00000000	00100100	0	36
CIMIENTOS	Hormigón ciclópeo	00100101	01001000	37	72
CIMIENTOS	Sobre pilotes de madera	01001001	01101100	73	108
CIMIENTOS	Sobre pilotes de hormigón	01101101	10010000	109	144
CIMIENTOS	Otros	10010001	10110100	145	180
CIMIENTOS	Piedra	10110101	11011000	181	216
CIMIENTOS	No se aplica, indeterminado	11011001	11111111	217	255
ESTRUCTURA	Hormigón armado	00000000	00110011	0	51
ESTRUCTURA	Metálica	00110100	01100110	52	102
ESTRUCTURA	Madera	01100111	10011001	103	153
ESTRUCTURA	Otros	10011010	11001100	154	204
ESTRUCTURA	No se aplica, indeterminado	11001101	11111111	205	255
PARED	Ladrillo	00000000	00100000	0	32
PARED	Bloque	00100001	01000000	33	64
PARED	Madera	01000001	01100000	65	96
PARED	Adobe o Tapia	01100001	10000000	97	128
PARED	Caña revestida o bareque	10000001	10100000	129	160
PARED	Prefabricadas	10100001	11000000	161	192
PARED	Otros	11000001	11100000	193	224
PARED	No se aplica, indeterminado	11100001	11111111	225	255
CUBIERTA	Hormigón Armado	00000000	00101011	0	43
CUBIERTA	Eternit, ardex, asbesto, cemento	00101100	01010110	44	86
CUBIERTA	Teja	01010111	10000001	87	129
CUBIERTA	Zinc	10000010	10101100	130	172
CUBIERTA	Otros	10101101	11010111	173	215
CUBIERTA	No se aplica, indeterminado	11011000	11111111	216	255

Tabla 3.10. Composición Genética Cromosomas de materiales predominantes de construcción.

Se tienen valores entre 0-255 para cada una de las 7 secciones de 8 bits que corresponden a cada atributo del antecedente del cromosoma-regla. La longitud máxima de un antecedente es de 56 bits, dependiendo de los atributos seleccionados. La notación decimal o fenotipo constituye el valor real del atributo específico y corresponde al código de la variable del fenómeno. La composición de un cromosoma-regla en binario es el genotipo del individuo. Las características representadas en los atributos del antecedente son: cimientos, estructura, paredes, cubierta, tipo de obra, uso de edificación, y origen del financiamiento de

la construcción. Pueden existir de 1 a n relaciones funcionales entre los atributos del antecedente que determinen el comportamiento de la construcción de la edificación de la vivienda.

Los parámetros del algoritmo con sus valores predeterminados se muestran en la Tabla 3.11. Uno de estos es el tamaño de la población inicial: Los porcentajes de cruce y mutación controlan las variaciones en la población y la reproducción de los individuos hasta obtener la población final. Los otros parámetros aseguran la convergencia del algoritmo luego de aplicar los operadores genéticos de cruce y mutación a través de alteraciones sucesivas a la composición de la población.

PARAMETRO	VALOR PREDETERMINADO
Tamaño población inicial	500
Porcentaje de cruce	40%
Porcentaje de mutación	0.1
Número de generaciones	10
Número de corridas	1
Índice de convergencia	75 %

Tabla 3.11. Parámetros algoritmo de predicción genético

Diseño Funcional del Algoritmo Genético.

La funcionalidad del algoritmo genético parte de su diseño estructural, en el cual, sobre cada cromosoma-regla se aplican operadores de cruce, mutación y clonación. A partir de evolución de algunas generaciones se obtienen nuevos individuos adaptados a las condiciones del ambiente representado por las particularidades de localidad y tiempo del caso de predicción. La herencia juega un papel importante en la adaptación, la constituyen las experiencias del pasado y es traducida a valores de confianza en los casos de ejemplo.

A los tipos de cromosoma que representan la composición física de la edificación, se asignan puntos de cruce y mutación al azar al momento de la operación. Para la operación de cruce se debe tomar los bits en sentido de izquierda a derecha partir del punto de cruce en el primer cromosoma, para remplazarlos en el

segundo cromosoma desde este punto hacia la derecha. Los cromosomas del cruce pertenecen a dos individuos diferentes que intercambian material genético y producen un nuevo individuo que debe ser sometido a adaptación en su ambiente. Para la operación de mutación se especifica el lugar que ocupa el bit en el cromosoma de izquierda a derecha.

La finalidad del cruce entre cromosomas de un mismo tipo es la de obtener siempre una composición diferente de material predominante en cada individuo en la generación, asegurando la aleatoriedad de aparecimiento de estas características de generación en generación.

La función de cruce puede ser expresada de la siguiente manera:

Resultado_cruce = función cruce(tipo_cromosoma_progenitor_1;
tipo_cromosoma_progenitor_2;
valor_gen_bin_cromo_progenitor_1;
valor_gen_bin_cromo_progenitor_2;
punto_de_cruce_cromo)

Fórmula 2.7. Resultado del cruce. A.G.

Los cromosomas progenitores son de un mismo tipo para diferentes individuos de la misma generación. El punto de cruce es tomado al azar al momento de realizar la operación del cruce. El corte se realiza en uno de los atributos del antecedente y se ubica en uno de los 8 bits de la sección correspondiente a la clase Tipo_Cromosoma. Como resultado se tiene un valor lógico que indica si se concretó o no la operación de cruce y se crea una nueva instancia de Cromosoma_Individuo con su estructura genética heredada de sus progenitores. El cruce se aplica únicamente para "cromosoma individuo" que representan a los atributos que fueron escogidos para formar parte del antecedente de la regla.

La finalidad de la mutación aplicada a la composición de un cromosoma, es hacer que se produzcan individuos con una característica poco frecuente cada cierto tiempo en la generación actual o siguiente.

De la mutación se obtiene un valor lógico que indica si se concretó la operación, y la instancia de cromosoma-individuo alterada en una cadena que guarda los puntos de mutación, elegidos al azar a momento de la operación.

La función de mutación puede ser expresada de la siguiente manera:

```
Resultado_mutación = función mutación (tipo_cromosoma_ mutante;
                                        valor_gen_bin_cromo_mutante;
                                         punto_de_mutacion_cromo
                                       )
```

Fórmula 3.1. Expresión del resultado del cruce. A.G.

Para una o varias generaciones a la vez, se asignan como parámetros, las probabilidades de cruce o mutación expresadas en un porcentaje de individuos que serán afectados por estas operaciones dentro del total. También se define el número de corridas que se hará hasta obtener la generación resultante.

La operación evaluar adaptabilidad de individuo toma los datos de la estructura genética del individuo y la somete a prueba según herencia y ambiente representados por su grado de pertenencia a la muestra de ejemplos extraída de los casos históricos. Este grado de pertenencia se mide en función de los valores de cobertura y confianza aplicables para esta técnica (detalles en el Capítulo 2 - Selección de Técnicas de Inteligencia Artificial). Luego se eligen los individuos con valores de adaptabilidad más elevados (cercanos) y se descartan los individuos con valores bajos (lejanos). De esta forma el algoritmo genético actúa como un proceso exhaustivo de búsqueda.

```
adaptabilidad_individuo = función evaluar_adaptabilidad_individuo (
                                        genotipo_individuo,
                                        regla seleccionada,
                                        casos de ejemplo,
                                        número de casos total,
                                        número de casos que cumple con el consecuente,
                                        adaptablilidad_cromosoma_regla,
                                        cobertura de la regla,
                                        )
```

Fórmula 3.2 Expresión de adaptabilidad del individuo. A.G.

```
Procedimiento Algoritmo Genético {
Inicio
        .nt = tamaño población
        . i = 1
        Lectura de caso de predicción y estructura de la regla de asociación.
        Lectura de valores de los parámetros
        .inicializar generación ( T i )
        .Procesar generación ( T i )
        .n =  seleccionar_idóneos ( T i )
        For j = 1  to n ( T I )
        .cruce ( j )
        .mutación ( j )
        .evaluar_adaptabilidad_individuo ( j )
        Next  j
        .seleccionar_sobrevivientes ( T i )
        . i  = i  +  1
        .evaluar_adaptabilidad_generación  ( T i )
        .obtener_generacion (T i + 1)
        .condicion_de_parada =  evaluar_condición_de_parada ( T i )
        Until  (condición_de_parada)  ó   ( i <= número de generaciones )
Fin
```

Figura 3.16. Pseudocódigo del Algoritmo Genético

El algoritmo se inicia con la identificación del caso de predicción y los parámetros. Luego se escoge o se crea una estructura de regla de asociación a descubrir. Se inicializa una población de individuos o generación T_i inicial. Los atributos a ser evaluados son los que se encuentran presentes en el antecedente de la regla, información predefinida por el usuario en los parámetros de las reglas. A través de la selección por torneo, van quedando parejas de cromosomas-regla que tienen mayor adaptabilidad, es decir se produce la clonación del progenitor más idóneo en la pareja. Sobre los idóneos se aplican los operadores de selección, cruce y mutación para transmitir características a la nueva generación en forma aleatoria. Cada cromosoma-regla se somete a evaluación en función de su adaptabilidad. Se puede decir que la función de adaptabilidad es un método retardado ya que al

momento de evaluación se extrae una muestra de ejemplos anteriores al año de predicción a través de la cual se mide la probabilidad de aparición de la regla en la población, expresada en los valores de confianza y cobertura. Esta probabilidad se entiende como la posibilidad de que sobrevivan los individuos en el espacio de búsqueda. De esta manera se produce una transición entre la generación (T_i) hacia la generación (T_{i+1}). Luego se aplica una condición de parada, la cual considera cual debe ser la generación final, esto es, cuando se haya alcanzado el objetivo en el cual un cierto número de individuos ha superado un nivel óptimo en su valor de adaptabilidad.

3.2.1.2 Arquitectura del Sistema.

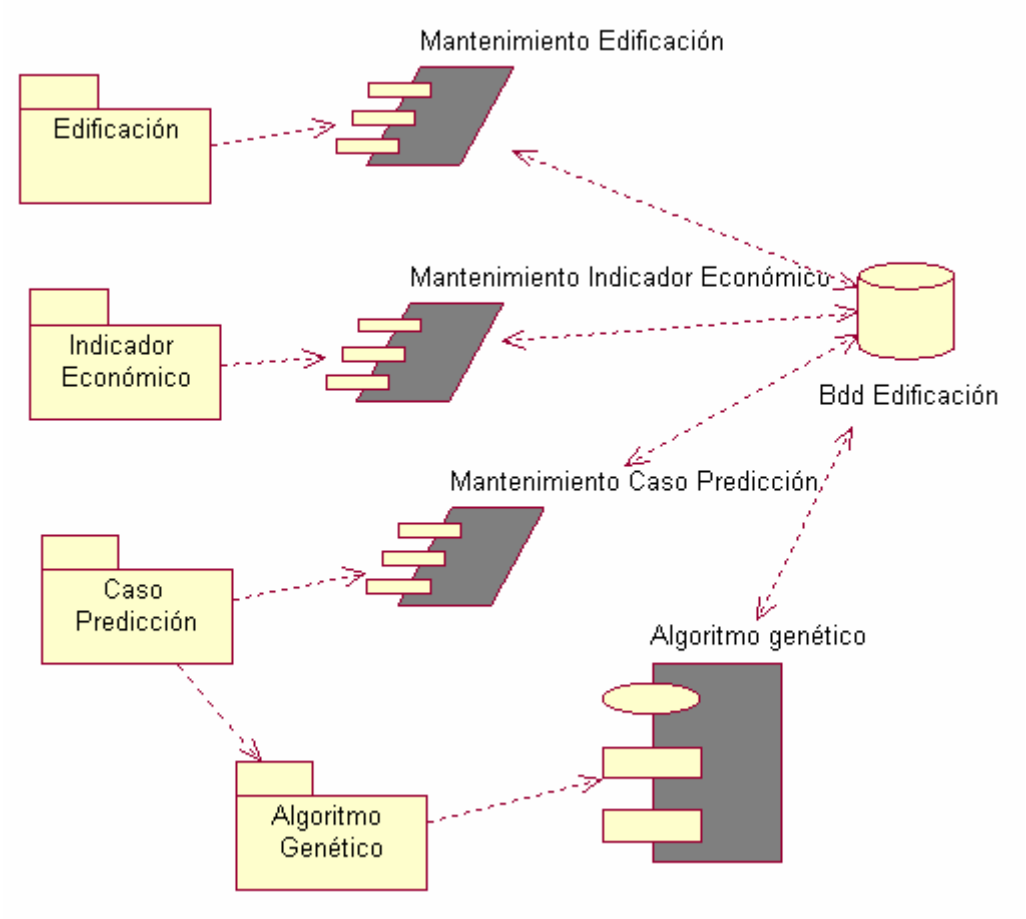


Figura 3.17. Diagrama de Arquitectura del Sistema SPACV.

La arquitectura del sistema SPACV consiste en la interrelación entre cuatro subsistemas o paquetes: Edificación, Indicador Económico, Caso Predicción, y Algoritmo genético. A cada uno de estos subsistemas se asocia un componente

de tareas de mantenimiento. El Subsistema Caso Predicción se asocia al Subsistema Algoritmo genético para llevar a cabo la aplicación del algoritmo sobre cada caso de predicción y para cuyos efectos utiliza los casos históricos de ejemplo desde la base de datos edificación. La estrategia de almacenamiento consiste en tener un solo repositorio con la información de los subsistemas, entre clases, objetos y asociaciones.

Mantenimiento Edificación.- Se encarga de realizar operaciones de ingreso, eliminación y modificación de objetos en las clases de datos históricos de construcción de viviendas.

Mantenimiento Indicador Económico .- Se encarga de realizar las operaciones de ingreso, eliminación y modificación de objetos instancia de las clases del subsistema Indicador Económico.

Mantenimiento Caso Predicción.- Se encarga de realizar las operaciones de ingreso, eliminación y modificación de objetos instancia de las clases del subsistema Caso Predicción. Se crean instancias de Variable Indicador económico y Variable Caso de Predicción y se llevan a cabo las operaciones de procesamiento del caso de predicción y la operación de extracción de reglas de asociación resultantes de la aplicación del algoritmo genético..

Algoritmo Genético.- Se encarga de aplicar el algoritmo genético a un caso de predicción, para el cual se inicia, se transforma y se termina el proceso en las clases generación, individuo y cromosoma-individuo del subsistema algoritmo genético y además se realizan operaciones de mantenimiento de parámetros

Otros componentes.- En la arquitectura del sistema hay otros servicios generales secundarios, como es el de control de usuarios y seguridades de acceso, y las operaciones de actualización de series de tiempo indicadores para procesos de análisis, entre otros.

3.2.2 DISEÑO DE OBJETOS.

3.2.2.1. Diccionario de objetos.

DICCIONARIO DE OBJETOS: ATRIBUTOS PRINCIPALES				
Nombre de atributo	**Código**	**Dominio**	**Tipo**	**Clase**
Id serie indicador	ID_SERIE_INDICADOR	IDENTIFICADOR	N15	SERIE_INDICADOR
Valor Indicador	VALOR_INDICADOR		N15,2	SERIE_INDICADOR
Id_tiempo	ID_TIEMPO	IDENTIFICADOR	N15	TIEMPO
Año	AÑO_TIEM		N4	TIEMPO
Mes	MES_TIEM		N2	TIEMPO
Nivel de tiempo	NIVEL_DE_TIEMPO		N1	TIEMPO
Id_Edificación	ID_EDIFICACION	IDENTIFICADOR	N15	EDIFICACION
Numero de boleta	NUBO		N5	EDIFICACION
Propiedad edifica	PROPIE_EED		N1	EDIFICACION
Cimientos edifica	CIMI_EED		N3	EDIFICACION
Estructura edifica	ESTRU_EED		N3	EDIFICACION
Pared edifica	PARED_EED		N3	EDIFICACION
Cubierta edifica	CUBI_EED		N3	EDIFICACION
Tipo moneda	MONEDA_EED		N1	EDIFICACION
Número de pisos	NUPIRES		N2	EDIFICACION
Número de dormitorios	NUDOCAL		N3	EDIFICACION
Número de cuartos	NUCUCAL		N3	EDIFICACION
Área a construir	CARCO		N7	EDIFICACION
Área Residencial	CARES		N7	EDIFICACION
Área No Residencial	CARNRES		N7	EDIFICACION
Área de Garage	CARDGAR		N7	EDIFICACION
Área Espacios Verdes	CAESV		N7	EDIFICACION
Superficie del Terreno	CSUTE		N7	EDIFICACION
Tipo de obra edifica	CTIPOBR_EED		N1	EDIFICACION
Número de permisos	CPERM		N3	EDIFICACION
Tipo superficie edifica	CDISPSUT_EED		N1	EDIFICACION
Tipo uso edifica	CDISPUSO_EED		N2	EDIFICACION
Tipo área a construir edifica	CDISPARC_EED		N1	EDIFICACION
Tipo permiso financiamiento	CDISPRFI_EED		N2	EDIFICACION
Tipo valor financiamiento	CDISPRFX_EED		N2	EDIFICACION
Número de viviendas	CNUVICAL		N3	EDIFICACION
Valor de la Edificación	CVAE		N15	EDIFICACION
Código Area Urbana Edifica	CAREAUR_EED		N1	EDIFICACION
Tipo de caso edificación	TIPOCASOED		N1	EDIFICACION
Id_indicador	ID_INDICADOR	IDENTIFICADOR	N15	INDICADOR ECONOMICO
Nombre indicador	NOMBRE_INDICADOR	CADENA LARGA	VA120	INDICADOR ECONOMICO
Id variable caso de predicción	ID_VARIABLE_CASO_DE_ PREDICCION	IDENTIFICADOR	N15	VARIABLE CASO DE PREDICCIÓN
Porcentaje Distribución	PORCDIST		N3	VARIABLE CASO DE PREDICCIÓN
Id caso de predicción	ID_CASO_PREDICCIÓN	IDENTIFICADOR	N15	CASO PREDICCIÓN
Numero de permisos caso	CPERM_CASO		N15	CASO PREDICCION
Valor m2 de construcción	VALOR_M2_C		N15,2	CASO PREDICCIÓN
Tipo de obra Caso	CTIPOBR_CASO		N1	CASO PREDICCIÓN
Tipo Uso Edifica Caso	CDISPUSO_CASO		N2	CASO PREDICCION
Tipo Financiamiento Caso	CDISPRFI_CASO		N2	CASO PREDICCION
Id localidad	IDENTIFICADOR	IDENTIFICADOR	N15	LOCALIDAD
Código de Región	CODIGO_DE_REGION		VA1	LOCALIDAD
Código de Provincia	CÓDIGO_DE_PROVINCIA		VA2	LOCALIDAD
Código de Cantón	CÓDIGO_DE_CANTON		VA4	LOCALIDAD
Código de Parroquia	CÓD_PARR_LOC		VA6	LOCALIDAD
Nombre Localidad	NOMBRE_LOCALIDAD	CADENA_SUPER	VA249	LOCALIDAD
Nivel de Localidad	NIVEL_DE_LOCALIDAD		N1	LOCALIDAD
Id_regla asociacion	ID_REGLA_ASOCIACION	IDENTIFICADOR	N15	REGLA_ASOCIACION
numero registro_regla	NUMERO_REGISTRO_REGLA		N5	REGLA_ASOCIACION
numero tipo_regla	NUMERO_TIPO_REGLA		N5	REGLA_ASOCIACION
condicion antecedente	CONDICION_ANTECEDENTE		VA249	REGLA_ASOCIACION
condicion consecuente	CONDICION_CONSECUENTE		VA249	REGLA_ASOCIACION
regla genotipo	REGLA_GENOTIPO		VA249	REGLA_ASOCIACION
param confianza_regla_h0	PARAM_CONFIANZA_REGLA_H0		N15,2	REGLA_ASOCIACION

Tabla 3.12. Diccionario de Objetos

3.2.2.2. Modelo Lógico basado en objetos: Modelo Entidad Relación

LISTA DE ENTIDADES	
Nombre de entidad	**Código**
SERIE INDICADOR	SERIE_INDICADOR
TIEMPO	TIEMPO
EDIFICACION	EDIFICACION
INDICADOR ECONOMICO	INDICADOR_ECONOMICO
VARIABLE CASO DE PREDICCION	VARIABLE_CASO_DE_PREDICCION
CASO PREDICCION	CASO_PREDICCION
LOCALIDAD	LOCALIDAD
VARIABLE CASO INDICADOR	VARIABLE_CASO_INDICADOR
GENERACION	GENERACION
INDIVIDUO	INDIVIDUO
PARAMETRO	PARAMETRO
RESTRICCION EDIFICACION	RESTRICCION_EDIFICACION
TIPO VARIABLE	TIPO_VARIABLE
CROMOSOMA GEN	CROMOSOMA_GEN
TIPO CROMOSOMA	TIPO_CROMOSOMA
CROMOSOMA INDIVIDUO	CROMOSOMA_INDIVIDUO
GRUPO RESTRICCION	GRUPO_RESTRICCION
REGLA ASOCIACION	REGLA_ASOCIACION

Tabla 3.13. Lista de Entidades

LISTA DE RELACIONES					
Relación	**Cod.**	**Entidad 2**	**Entidad 1**	**E2 → E1**	**E1 → E2**
R01_EED_LOC	R01	LOCALIDAD	EDIFICACION	1,1	0,n
R02_EED_TIEM	R02	TIEMPO	EDIFICACION	1,1	0,n
R03_SERIEIN_TIEMPO	R03	TIEMPO	SERIE_INDICADOR	1,1	0,n
R04_CASO_VARIABLE	R04	CASO_PREDICCION	VARIABLE_CASO_DE_PREDICCION	1,1	0,n
R05_TIEMPO_CASO_PREDICCION	R05	TIEMPO	CASO_PREDICCION	1,1	0,n
R06_CASO_LOC	R06	LOCALIDAD	CASO_PREDICCION	1,1	0,n
R07_VARIABLE_CASO_INDICADOR	R07	INDICADOR_ECONOMICO	VARIABLE_CASO_INDICADOR	1,1	0,n
R08_VARIABLE_CASO_PREDICCION	R08	CASO_PREDICION	VARIABLE_CASO_INDICADOR	1,1	0,n
R09_SERIE_LOCALIDAD	R09	LOCALIDAD	EDIFICACION	1,1	0,n
R10_RESTRICCION_EDIFICACION	R10	GRUPO_RESTRICCION	RESTRICCION_EDIFICACION	0,1	0,n
R11_TIPO_VARIABLE	R11	TIPO_VARIABLE	VARIABLE_CASO_DE_PREDICCION	1,1	0,n
R12_CASO_GENERACION	R12	CASO_PREDICCION	GENERACION	1,1	0,n
R13_GENERACION_INDIVIDUO	R13	GENERACION	INDIVIDUO	1,1	0,n
R14_PARAMETRO_GENERACION	R14	GENERACION	PARAMETRO	0,n	0,1
R15_SERIEIN_INDICADOR	R15	INDICADOR_ECONOMICO	SERIE_INDICADOR	1,1	0,n
R16_CROMO_TIPO	R16	TIPO_CROMOSOMA	CROMOSOMA_GEN	1,1	0,n
R17_CROMO_GEN	R17	CROMOSOMA_GEN	CROMOSOMA_INDIVIDUO	1,1	0,n
R18_CROMO_INDIV	R18	INDIVIDUO	CROMOSOMA_INDIVIDUO	1,1	0,n
R19_GRUPO_RESTRIC	R19	GRUPO_RESTRICCION	EDIFICACION	1,1	0,n
R20_PARAMETRO_CASO	R20	CASO_PREDICCION	PARAMETRO	0,n	1,1
R21_CASO_PRED_REGLA	R21	CASO_PREDICCION	REGLA_ASOCIACION	1,1	0,n

Tabla 3.14. Lista de Relaciones

3.2.2.3. Modelo Relacional basado en registros: Modelo Físico

LISTA DE TABLAS		
Nombre de Tabla	**Código**	**Clave primaria**
SERIE INDICADOR	SERIE_INDICADOR	ID_SERIE_INDICADOR
TIEMPO	TIEMPO	ID_TIEMPO
EDIFICACION	EDIFICACION	ID_EDIFICACION
INDICADOR ECONOMICO	INDICADOR_ECONOMICO	ID_INDICADOR_ECONOMICO
VARIABLE CASO DE PREDICCION	VARIABLE_CASO_DE_PREDICCION	ID_VARIABLE_CASO_DE_PRED
CASO PREDICCION	CASO_PREDICCION	ID_CASO_PREDICCION
LOCALIDAD	LOCALIDAD	ID_LOCALIDAD
VARIABLE CASO INDICADOR	VARIABLE_CASO_INDICADOR	ID_VARIABLE_CASO_INDICADOR
GENERACION	GENERACION	ID_GENERACION
INDIVIDUO	INDIVIDUO	ID_INDIVIDUO
PARAMETRO	PARAMETRO	ID_PARAMETRO
RESTRICCION EDIFICACION	RESTRICCION_EDIFICACION	ID_RESTRICCION_EDIFICACION
TIPO VARIABLE	TIPO_VARIABLE	ID_TIPO_VARIABLE
CROMOSOMA GEN	CROMOSOMA_GEN	ID_CROMOSOMA_GEN
TIPO CROMOSOMA	TIPO_CROMOSOMA	ID_TIPO_CROMOSOMA
CROMOSOMA INDIVIDUO	CROMOSOMA_INDIVIDUO	ID_CROMOSOMA_INDIVIDUO
GRUPO RESTRICCION	GRUPO_RESTRICCION	ID_GRUPO_RESTRICCION
REGLA_ASOCIACION	REGLA_ASOCIACION	ID_REGLA_ASOCIACION

Tabla 3.15. Lista de Tablas

LISTA DE REFERENCIAS				
Relación	**Código**	**Tabla Padre**	**Tabla Hija**	**Constraint**
R01_EED_LOC	R01	LOCALIDAD	EDIFICACION	FK_EDIFICAC_R01_LOCALIDA
R02_EED_TIEM	R02	TIEMPO	EDIFICACION	FK_EDIFICAC_R02_TIEMPO
R03_SERIEIN_TIEM	R03	TIEMPO	SERIE_INDICADOR	FK_SERIE_IN_R03_TIEMPO
R04_CASO_VARIABLE	R04	CASO_PREDICCION	VARIABLE_CASO_DE_ PREDICCION	FK_VARIABLE_R04_CASO_PRE
R05_TIEMPO_CASO_ PREDICCION	R05	TIEMPO	CASO_PREDICCION	FK_CASO_PRE_R05_TIEMPO
R06_CASO_LOC	R06	LOCALIDAD	CASO_PREDICCION	FK_CASO_PRE_R06_LOCALIDA
R07_VARIABLE_CASO_ INDICADOR	R07	INDICADOR_ ECONOMICO	VARIABLE_CASO_ INDICADOR	FK_VARIABLE_R07_INDICADO
R08_VARIABLE_CASO_ INDICADOR	R08	CASO_PREDICCION	VARIABLE_CASO_ INDICADOR	FK_VARIABLE_R08_CASO_PRE
R09_SERIE_LOCALIDAD	R09	LOCALIDAD	SERIE_INDICADOR	FK_SERIE_IN_R09_LOCALIDA
R10_RESTRICCION_ EDIFICACION	R10	GRUPO_RESTRICCION	VARIABLE_CASO_ INDICADOR	FK_RESTRICC_R10_GRUPO_RE
R11_TIPO_VARIABLE	R11	TIPO_VARIABLE	VARIABLE_CASO_ PREDICCION	FK_VARIABLE_R11_TIPO_VAR
R12_CASO_GENERACION	R12	CASO_PREDICCION	GENERACION	FK_GENERACI_R12_CASO_PRE
R13_GENERACION_ INDIVIDUO	R13	GENERACION	INDIVIDUO	FK_INDIVIDU_R13_GENERACI
R14_PARAMETRO_ GENERACION	R14	PARAMETRO	GENERACION	FK_GENERACI_R14_PARAMETR
R15_SERIEIN_INDICADOR	R15	INDICADOR_ ECONOMICO	SERIE_INDICADOR	FK_SERIE_IN_R15_INDICADO
R16_CROMO_TIPO	R16	TIPO_CROMOSOMA	CROMOSOMA_GEN	FK_CROMOSOM_R16_TIPO_CRO
R17_CROMO_GEN	R17	CROMOSOMA_GEN	CROMOSOMA_ INDIVIDUO	FK_CROMOSOM_R17_CROMOSOM
R18_CROMO_INDIV	R18	INDIVIDUO	CROMOSOMA_ INDIVIDUO	FK_CROMOSOM_R18_INDIVID
R19_GRUPO_RESTRIC	R19	GRUPO_RESTRICCION	EDIFICACION	FK_EDIFICAC_R19_GRUPO_RE
R20_PARAMETRO_CASO	R20	CASO_PREDICCION	PARAMETRO	FK_PARAM_R20_CASO_PRE
R21_CASO_PRED_REGLA	R21	CASO_PREDICCION	REGLA_ASOCIACION	FK_REGLA_ASO_R21_CASO_PRE

Tabla 3.16. Lista de Referencias

3.2.2.4. Diseño de Interfaz de usuario.

3.2.2.4.1. Componentes de Interfaz

Dispositivos de entrada y salida.

Entrada: La introducción de datos en el sistema se lo hará a través de dispositivos como el ratón y el teclado. En ciertas ventanas se usará el botón derecho del ratón para desplegar un menú emergente cuando sea necesario.

Salida: La salida comprende mensajes del sistema al usuario, y cuadros de informe presentados por pantalla e impresora. Para la salida en impresora se presentará una ventana de vista previa antes de la impresión definitiva.

Dispositivos físicos y virtuales.

Físicos: El teclado de 101 teclas mínimo, el ratón y sus funciones, la impresora, el monitor con resolución de pantalla predeterminada a 1024 x 768 pixels.

Virtuales: pantalla principal, ventanas de mantenimiento y procesos, barra de herramientas, cuadros de texto, controles de desplazamiento de registro, botones de comandos, cuadros de selección, formularios de entrada de datos y de reporte.

Componentes virtuales de la interfaz de usuario.

Utilización de barra de herramientas para mantenimiento.

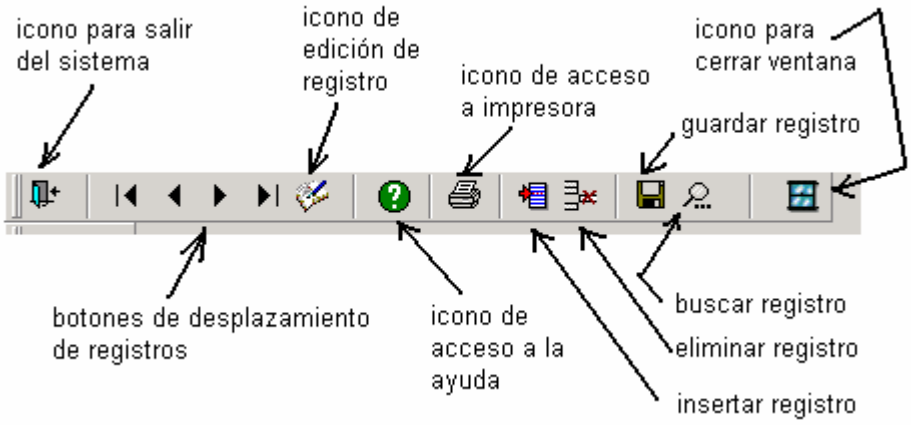


Figura 3.18. Barra de herramientas para mantenimiento

Estará disponible en el extremo superior de la ventana principal, una barra de herramientas para activarse conjuntamente con la ventana. Los iconos de insertar, eliminar, buscar y guardar registro, se aplican sobre la tabla que activa para mantenimiento. (Ver Figura 3.23).

Utilización de Iconos de acceso directo a opciones del sistema.

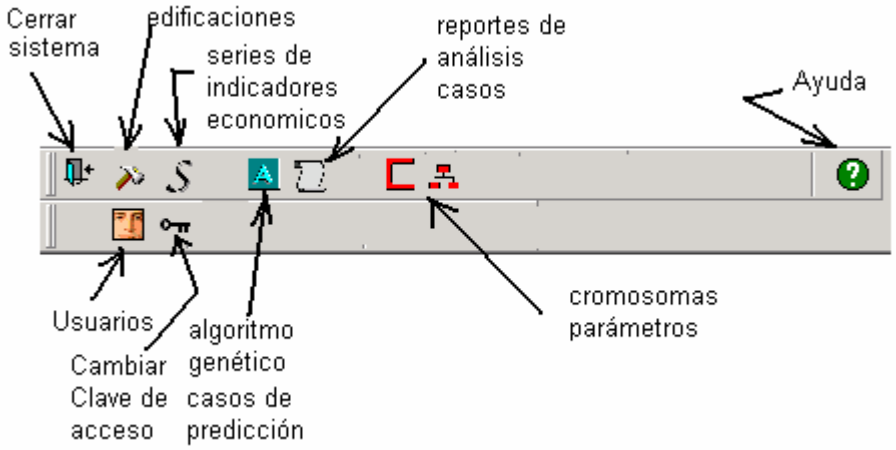


Figura 3.19. Iconos de acceso directo a opciones del sistema.

Para facilitar el acceso directo hacia opciones más frecuentes del sistema hay iconos visibles en la región superior de la pantalla principal. El icono de acceso a los casos de predicción y aplicación del algoritmo genético está en el medio. En la segunda fila van iconos de administración y cambio de clave.

Utilización de menús.

El menú principal de opciones se diseña en estilo horizontal, ubicado en la parte superior de la pantalla. (Ver Figura 3.25):

Figura 3.20. Menú principal de opciones del sistema.

Utilización de ventanas

Se diseña una ventana principal contenedora, la cual dispone de área de título, con botones estándar de Microsoft Windows para minimizar, maximizar y cerrar. Esta dispone de una línea de mensajes al extremo inferior, la cual podrá ser utilizada desde cualquier objeto del sistema al momento de su ejecución.

Formato de pantalla para una ventana base.

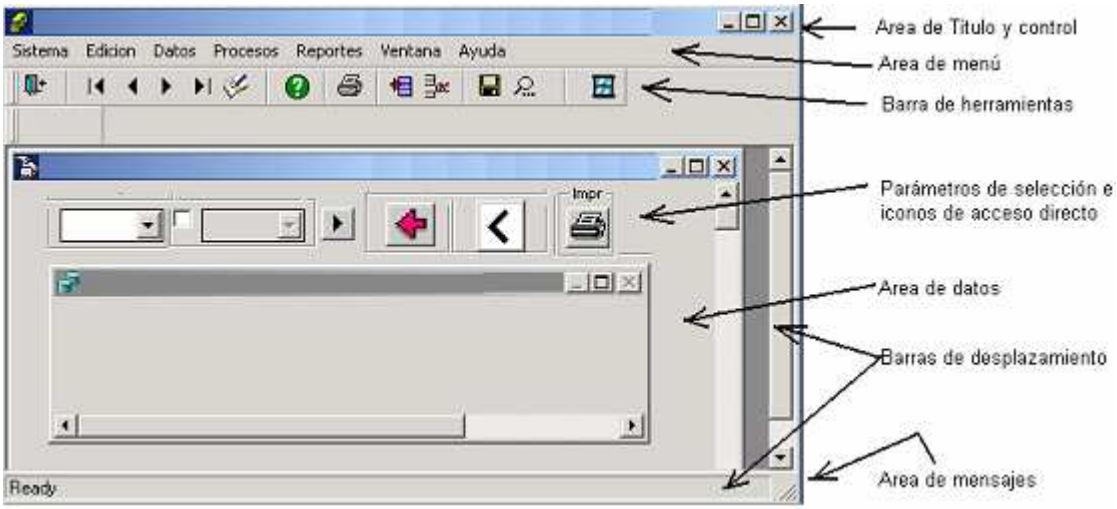


Figura 3.21. Formato de Pantalla

La pantalla se ha divido en cuatro áreas: El área de titulo y control contiene la referencia al sistema junto a los botones de maximización, minimización y cierre.

Se ha colocado la barra de herramientas bajo el área de menú para facilitar su acceso mediante íconos. El área de datos comprende un 70% de la pantalla total y se encuentra en el medio. Se divide en dos secciones. La primera es una franja superior en la cual se colocan los controles de parámetros de selección de datos e iconos de acceso directo a otras ventanas. La segunda consiste en una ventana interior, la cual puede ser de actualización o presentación de datos. Se ha diseñado barras de desplazamiento tanto horizontal como vertical. En el extremo inferior se ha dispuesto un área de línea de mensajes.

Diseño de Ventana de acceso (Login).

Es una ventana tipo respuesta y consta de dos cuadros de texto para la introducción de la identificación y contraseña del usuario. Tiene dos botones para aceptar o cancelar el proceso. Muestra un icono de identificación del sistema.

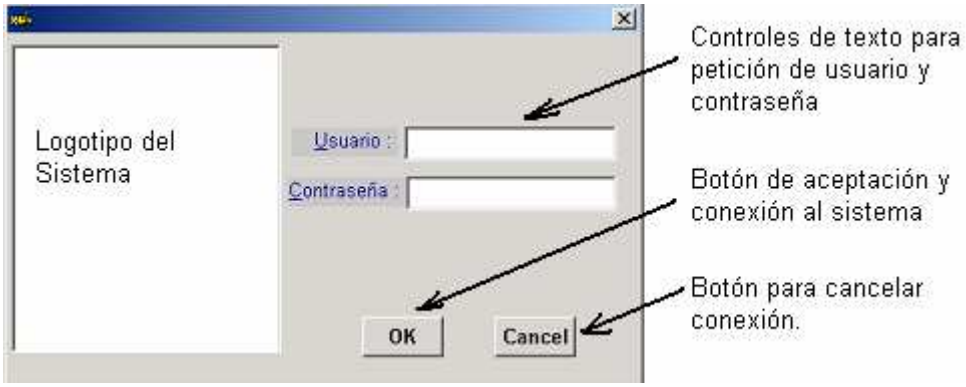


Figura 3.22. Diseño de Ventana Login

Diseño de Ventana de petición de parámetros de impresión.

Es una ventana tipo respuesta que contiene los objetos necesarios para introducir parámetros de impresión y para producir una impresión previa.

Diseño de Ventana de mantenimiento y presentación de datos.

La ventana mantenimiento es heredada de la ventana base y se ha diseñado para transacciones de ingreso, eliminación y actualización en los datos de una determinada tabla o relación del sistema. Contendrá un área de datos interna en la cual los datos se disponen en forma de cuadrilla, navegables de izquierda a derecha y de arriba hacia abajo, mediante una barra de desplazamiento..

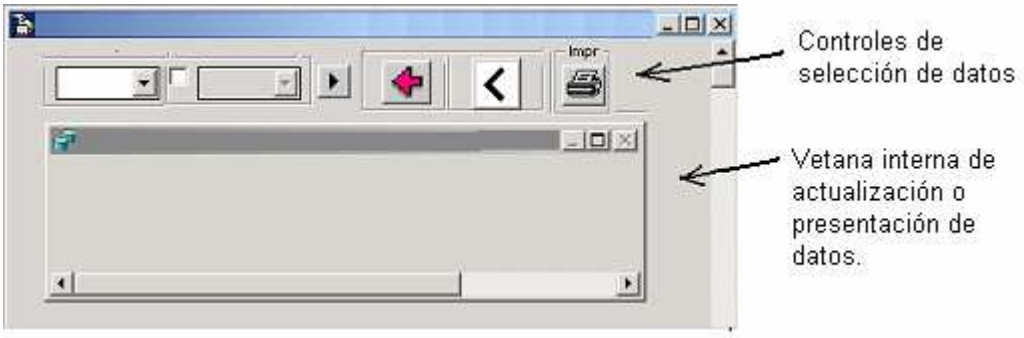


Figura 3.23. Diseño de Ventana Mantenimiento / Presentación de datos.

Diseño de Ventana Petición de Parámetros.

Esta ventana tipo respuesta permite la introducción de parámetros como por ejemplo los que se usan para la ejecución del algoritmo genético. Consiste en un conjunto de controles mediante los cuales se pueden modificar los parámetros.

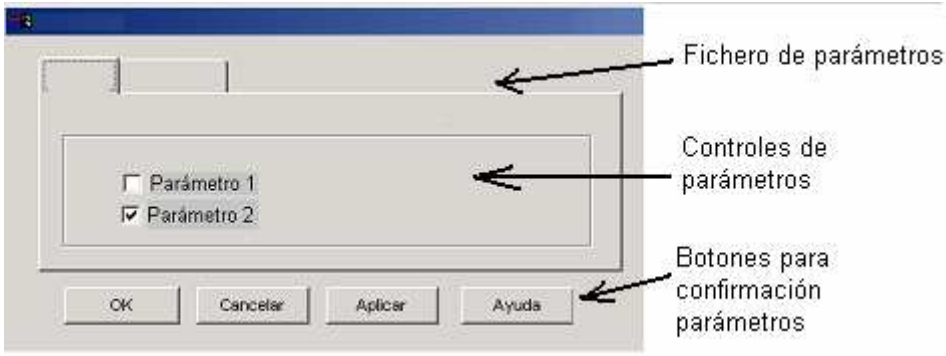


Figura 3.24. Diseño de Ventana Petición de parámetros.

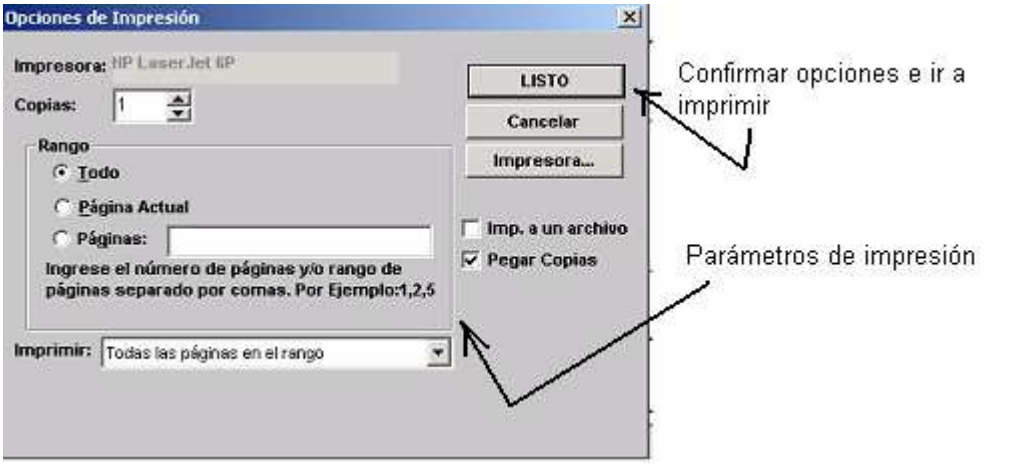


Figura 3.25. Diseño de Ventana de petición de parámetros de impresión.

Diseño de Ventana Interfaz Casos de Predicción

La ventana principal contendrá tres secciones: La primera servirá para introducir y seleccionar los casos de predicción. La segunda para exploración y análisis de casos y la tercera para aplicar técnicas de predicción y obtener resultados.

1. Ventana Interfaz Casos de Predicción - Sección Caso de Predicción

Esta sección contendrá los controles para introducir un caso de predicción a través de un icono de herramientas para escoger lugar y tiempo de la predicción.

Figura 3.26. Diseño de Ventana Interfaz Casos de Predicción: Casos de Predicción.

Se utiliza la barra de herramientas para insertar, eliminar o buscar otro caso.

2. Ventana Interfaz Casos de Predicción - Sección Análisis

Diseñada para realizar análisis exploratorio de datos de edificaciones del caso de predicción introducido por localidad y tiempo.

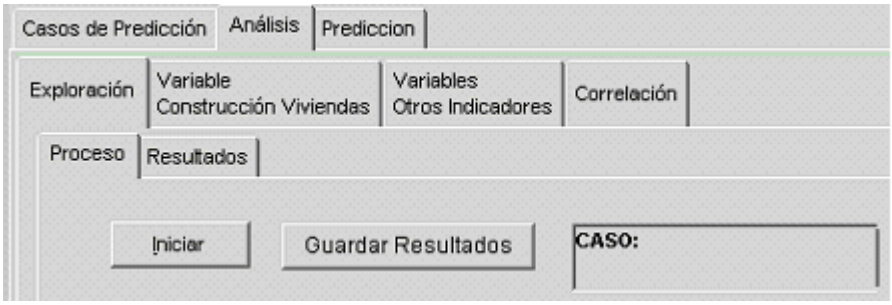


Figura 3.27. Diseño de Ventana Interfaz de Casos de Predicción: Análisis

También permitirá encontrar relaciones de causalidad con otras variables.

3. Ventana Casos de Predicción - Sección Predicción.

Esta sección se ha diseñado para contener las técnicas de inteligencia artificial a aplicar para el caso de predicción introducido. Para el caso de estudio del sistema será necesario incluir una sección para aplicar el algoritmo genético.

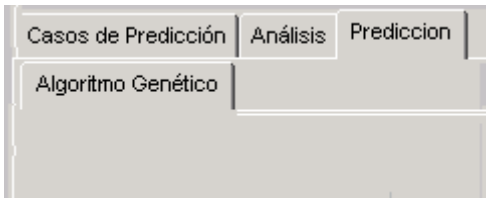


Figura 3.28. Diseño de Ventana Interfaz de Casos de Predicción: Predicción

La sección que incluye la aplicación del algoritmo genético contendrá tres subsecciones adicionales: Una subsección para introducir los parámetros del algoritmo, otra para realizar el proceso del algoritmo y una tercera para obtener y visualizar los resultados de las reglas extraídas y evaluarlas.

La sección de parámetros está diseñada para introducir parámetros que se aplican a toda una generación de individuos y además contiene las características genéticas en la composición de cromosomas y sus puntos de cruce y mutación.

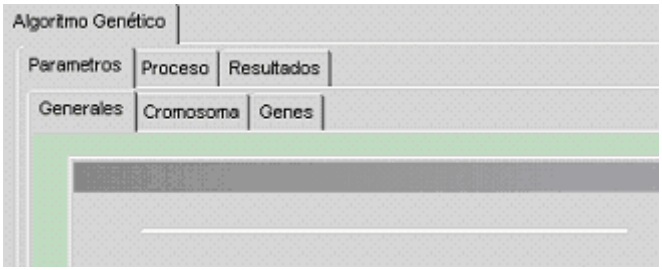


Figura 3.29. Diseño de Ventana Interfaz Sección Predicción AG: Parámetros.

En la subsección de proceso de la sección predicción se realizará la ejecución del algoritmo genético. Se inicia con una generación de individuos inicial.

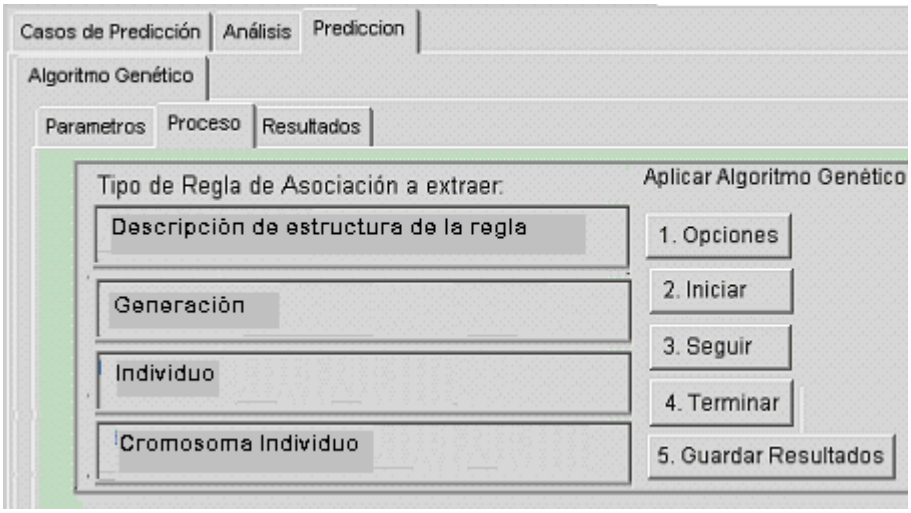


Figura 3.30. Diseño de Ventana Interfaz Sección Predicción AG: Proceso.

En la ventana hay botones para iniciar y continuar el proceso desde su estado anterior hasta que se cumplan las condiciones de parada del algoritmo. Se muestra la última generación y se ha creado adicionalmente un botón para guardar los resultados.

En la subsección de resultados de la sección predicción se ha diseñado un cuadro de ventana para presentar los resultados de extracción de las reglas mediante el algoritmo genético. Se permite introducir opciones antes de probar la regla.

Figura 3.31. Diseño de Ventana Interfaz Sección Predicción AG: Resultados.

3.2.2.4.2. Flujo de Ventanas. .

El flujo de ventanas en la ejecución del sistema sigue el siguiente ciclo:

1. Presentación de la ventana de acceso al sistema (Login)._ Se pide al usuario su identificación y contraseña. El usuario ingresa identificación y contraseña. El sistema reconoce al usuario y realiza la conexión con la base de datos.

2. Introducción a la sesión del sistema, al menú y selección de opción._ Se presenta la pantalla principal del sistema con menú e iconos de acceso directo a las opciones. Se muestra en el título de la ventana la identificación de usuario. El usuario selecciona una opción a través de un icono de acceso directo o un ítem del menú principal.

3. Presentación y ejecución de ventana desde el menú de opciones._El usuario activa una de las ventanas desde el menú de opciones, esta puede ser de mantenimiento de datos, o de introducción de casos, ejecución y presentación de resultados del proceso de predicción.

4. Cierre de ventana de opción y / o cierre de sesión del sistema.

3.2.2.4.3. Diseño de Opciones de Menú

Opción	Sub-opciones	Descripción
Sistema	Cerrar ventana	Cerrar la ventana activa en el área de datos de la pantalla.
	Cambiar contraseña	Cambiar la contraseña de usuario actual conectado.
	Cerrar Sistema	Cerrar la sesión de usuario en el sistema.
Edición	Primero; anterior; siguiente: último	desplazamiento de registros en la ventana de datos
	Editar; insertar; borrar	Transacciones de actualización, inserción o eliminación de registros en la ventana de datos
	Guardar; buscar; imprimir	Operaciones de guardar, buscar o imprimir los registros de la ventana de datos.
	Barra de herramientas	Modificar las propiedades de la barra de herramientas.
Datos	Edificaciones	Transacciones de actualización, inserción o eliminación de registros en edificaciones con sus restricciones y reglas
	Indicadores Económicos	Transacciones de actualización, inserción o eliminación de series de indicadores económicos
	Localidades Geográficas	Transacciones de actualización, inserción o eliminación.
	Tiempo	Transacciones de actualización, inserción o eliminación.
	Parámetros	Transacciones de actualización, inserción o eliminación.
	Usuarios	Transacciones de actualización, inserción o eliminación.
Procesos	Casos de Predicción	Introducción de un caso de predicción y ejecución del algoritmo de predicción genético.
Reportes	Análisis casos de predicción	Presentación de resultados del caso de predicción
Ventana	Barra de Herramientas	Cambiar propiedades de la barra de herramientas
Ayuda	Contenido	Muestra los contenidos de la ayuda general del sistema
	Resumen del Sistema	Muestra un resumen de las opciones del sistema
	Acerca de	Muestra la presentación del sistema con su autor.

Tabla 3.17. Opciones de menú del sistema

3.2.2.4.4. Diseño de Perfil de Usuarios

Rol Usuario	Descripción	Permisos de Acceso a Opciones
Administrador	Tiene un conocimiento general del sistema y de la base de datos. Tiene acceso a todas las transacciones y procesos del sistema. Puede crear nuevos usuarios y asignarles contraseña.	Todas las Opciones del Sistema
		Creación, eliminación, actualización de objetos en el sistema
		Seguridades y Respaldos
Analista experto	Realiza el mantenimiento de los datos de base antes de aplicarlos a un caso de predicción	Opción de Edición, Datos, Ayuda
		Inserción, eliminación y actualización de registros en las tablas de edificación, restricciones, series de indicadores económicos, localidad geográfica, tiempo, parámetros
Analista predictor	Realiza el proceso de predicción aplicado a un caso de edificación de vivienda. Puede acceder a la ventana de presentación de resultados.	Opción de Procesos, Reportes, Ayuda.
		Consulta a resultados de casos de predicción.
		Introduce casos de predicción, ejecuta el algoritmo de predicción genético.

Tabla 3.18.Perfil de Usuarios.

3.3 DESARROLLO Y PROGRAMACIÓN.

3.3.1 CONFIGURACIÓN DEL ENTORNO DE DESARROLLO.

3.3.1.1 Descripción del Entorno de desarrollo.

El entorno de desarrollo para el sistema es "**Power Builder**". En este entorno se crean los proyectos de desarrollo de aplicaciones a partir de los espacios de trabajo o **workspace**. En un workspace cabe más de un proyecto. A su vez, un proyecto puede tener varios destinatarios de objetos de aplicaciones o **targets**. De destinatarios "targets", el **Target powerscript** contiene una aplicación cliente servidor ejecutable o un componente del servidor. Adicionalmente se tienen en la ventana principal otros elementos como: el **System tree,** para colocar las clases y objetos de la aplicación; y otros objetos de botones para acceder a cada uno de los objetos de usuario incluidos librerías y bases de datos. Además botones **New** e **Inherit,** para crear y heredar objetos y el botón **Exit** para salir.

A estas tareas de los botones también pueden accederse desde las opciones **File, Run, y Tools** del menú principal de la pantalla del entorno de PowerBuilder.

3.3.1.2 Creación del espacio de trabajo.

Mediante el asistente de creación de espacios de trabajo y la opción del menú **File-New** se ha creado un nuevo espacio de trabajo ("WorkSpace"), el mismo que se encuentra en el árbol del sistema con el nombre de **SPACV_WKS**.

3.3.1.3 Creación del destinatario de la aplicación.

De la misma manera y mediante la opción del menú **File-New** se ha creado un destinatario para la aplicación de tipo **Target Application** y de nombre: **spacv**.
Este tipo de objeto hace que la aplicación sea Cliente-Servidor.

3.3.2 CREACIÓN DE LA BASE DE DATOS RELACIONAL

3.3.2.1 Creación de la Base de Datos Relacional.

Se ha elegido el motor de base de datos SQL Server 2000 por cuanto está diseñado para funcionar también como base de datos independiente directamente en el mismo equipo de una aplicación y por sus características de escalabilidad y facilidad de uso, que le permiten trabajar eficazmente en un único equipo sin consumir demasiados recursos y sin que sean necesarias tareas administrativas por parte del usuario independiente.

En SQL-Server 2000, la creación de la base de datos relacional se la hace a través de un código fuente escrito en el servidor o "script", el cual contiene las definiciones de localización y espacio asignado a los datos "data" y al registro de transacciones o "log. El "script" termina con la creación de tablas y demás objetos de la base de datos. Se incluye la creación de usuarios y claves de autenticación.

```
CREATE DATABASE [spacv]
ON (NAME = N'spacv_Data', FILENAME = N'C:\BASES\SPACV\data\spacv_Data.MDF' ,
SIZE = 76, FILEGROWTH = 10%) LOG ON (NAME = N'spacv_Log', FILENAME =
N'C:\BASES\SPACV\Log\spacv_Log.LDF' , SIZE = 284, FILEGROWTH = 10%)
COLLATE Modern_Spanish_CI_AS
```

Figura 3.32. Código de creación de la base de datos SPACV.

3.3.2.2 Configuración de los Objetos principales de la Base de Datos.

Antes de llevar a cabo la creación de usuarios, debe haberse configurado el inicio de sesión SQLServer 2000 a través del Administrador Corporativo

En el proceso de creación de objetos, los usuarios tienen su clave de autenticación y sus roles definidos para ingreso, eliminación y actualización en los datos. En este sistema se realiza la asignación de perfiles a través del cliente o programa de menú principal del sistema. Únicamente el acceso del administrador a la base de datos se configura desde SQL Server 2000.

```
CREATE TABLE [dbo].[CASO_PREDICCION] (
        [ID_CASO_PREDICCION] [IDENTIFICADOR] IDENTITY (1, 1) NOT NULL ,
        [COD_CASO_PREDICCION] [numeric](8, 0) NOT NULL ,
        [ID_TIEMPO] [IDENT_TIEMPO] NOT NULL ,
        [ID_LOCALIDAD] [IDENT_LOCALIDAD] NOT NULL ,
        [COD_VAR_A_PREDECIR] [numeric](2, 0) NOT NULL ,
        [VAL_VAR_A_PREDECIR] [numeric](15, 2) NULL ,
        [DESCRIPCION_CASO] [CADENA_SUPER] NULL ,
        [NIVEL_DE_TIEMPO_CASO] [numeric](1, 0) NOT NULL ,
        [NIVEL_DE_LOCALIDAD_CASO] [numeric](1, 0) NOT NULL
  ) ON [PRIMARY]
```

Figura 3.33. Código de creación de la tabla Caso_Predicción.

En la creación de tablas se usan los comandos CREATE TABLE para la definición de la tabla y ALTER TABLE para añadir otros objetos como asociaciones con otras tablas. Por medio del administrador corporativo de SQLServer 2000, se puede tanto crear como acceder a los objetos creados para la base de datos SPACV. En el siguiente código se agregan a la tabla "Caso_Predicción" sus respectivas claves foráneas de asociación con las tablas "Tiempo" y "Localidad":

```
ALTER TABLE [dbo].[CASO_PREDICCION] ADD
        CONSTRAINT [FK_CASO_PRE_R05_TIEMPO] FOREIGN KEY
        ([ID_TIEMPO]
        ) REFERENCES [dbo].[TIEMPO] (
                [ID_TIEMPO]
        ),
        CONSTRAINT [FK_CASO_PRE_R06_LOCALIDA] FOREIGN KEY
        ([ID_LOCALIDAD]
        ) REFERENCES [dbo].[LOCALIDAD] (
                [ID_LOCALIDAD]
)
```

Figura 3.34. Agregación de claves foráneas de tiempo y localidad a la tabla Caso_Predicción.

Se han creado dos grupos de usuarios de la base de datos SPACV: El primero tiene todos los privilegios y sus usuarios son: "Administrador", "dbo", "jdelgado".
Para el segundo grupo de usuarios se ha creado el usuario "**spacv**", que tiene privilegios de selección, inserción, eliminación, actualización y ejecución en las tablas del sistema. Los perfiles se asignan en menú del cliente.

Nombre	Nombre de inicio de sesión	Acceso a base de datos
administrador	jdelgado	Permiso
dbo	PCENTRAL\jdelgado	Permiso
jdelgado		Permiso
spacv	spacv	Permiso

Figura 3.35. Usuarios configurados en la base de datos SPACV a través del administrador..

3.3.3 DESARROLLO DE LA INTERFAZ DE USUARIO.

3.3.3.1 Creación de clases de objetos visuales.

Los objetos visuales se crean como instancias de ventanas definidas para el sistema y que son herencias de clases definidas por PowerBuilder.

Objeto	Nombre de clase	Abre-viatura	Descripción	Propiedades
none	Botón de comando (CommandButton)	Cmd_ Btn_	Ejecuta un proceso, función o método Permite el flujo hacia otra ventana. Cancela un proceso.	Color de fondo: Gris Color de texto : Negro Formato de texto: Arial 8 – 10
	Cuadro de comando (PictureButton)	Pbt_	Similar al botón de comando. Muestra un icono relacionado con el proceso o función al cual direcciona.	Color de fondo: Gris
none	Casilla de verificación (CheckBox)	Chk_	Marca un objeto. Verifica selección de algún dato.	Color de fondo: Gris Color de texto: Negro Color de casilla: blanco Formato de texto: Arial 8-10
none	Botón de opción (RadioButton)	rbt_	Elige una opción Activa un parámetro o alguna variable.	Color de fondo: Gris oscuro Color de texto: Negro Color de botón: blanco Formato de texto: Arial 8-10
none	Texto estático (Static Text)	st_	Texto informativo o mensaje estático	Color de fondo: Gris Color de texto: azul o negro
	Caja contenedora (GroupBox)	gb_	Contiene a otros controles	Color de línea: blanco Ancho de línea: 1,5 puntos
	Línea de edición (SingleLineEdit)	sle_	Permite editar un texto en pantalla	Color de fondo: Blanco Color de texto: Negro Formato de texto: Arial 8-10
.00	Máscara de edición (EditMax)	em_	Permite editar un campo de dato para tipos numéricos, fecha o texto. El color de fondo cambia a blanco en edición, y a gris en deshabilitado.	Color de fondo: Gris / Blanco Color de texto: Negro / Blanco Formato de texto: Arial 8-10

-

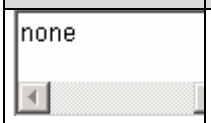
Objeto	Nombre de la clase	Abre-viatura	Descripción	Propiedades
	Edición en varias líneas (MultiLineEdit)	mle_	Permite escribir texto en varias líneas	Color de fondo: Blanco Color de texto: Negro Formato de texto: Arial 8 – 10
	Cuadro combinado de lista. (DropDownListBox)	ddlb_	Permite seleccionar un ítem de la lista	Color de fondo: Blanco Color de texto: Negro Formato de texto: Arial 8 – 10
	Caja de lista (ListBox)	lb_	Permite seleccionar un ítem de la lista, la cual aparece completamente visible.	Color de Fondo: Blanco Color de Texto: Negro Formato de texto: Arial 8 – 10
	Ventana de visualización y actualización de datos. (Data Window)	dw_	Ventana cuya función es la actualización o visualización de los datos en forma de tabla. LLeva barras de desplazamiento horizontales y verticales.	Color de fondo Título: Gris / Azul Color de texto Título: Blanco Color barra desplazamiento: Gris Color de fondo datos: Blanco Color de texto datos: Negro Formato de texto: Arial 8 - 10
	Ventana general (Window)	w_	Ventana general que sirve de control contenedor de otros objetos. Puede contener un conjunto de paneles que se activan a voluntad del usuario.	Color de fondo Título: Azul Color de texto Título: Blanco Color barra desplazamiento: Gris Color de fondo datos: Blanco Color de texto datos: Negro Formato de texto: Arial 8 – 10

Tabla 3.19. Objetos visuales básicos

3.3.3.2 Creación de clases de objetos no visuales

Objeto transacción.- Diseñado para enviar una petición de datos al sistema a través de un comando SQL. Busca una conexión con la base de datos, la utiliza, realiza una operación de inserción, eliminación, o actualización y devuelve un mensaje de error si no se realizó la transacción. Este objeto se ha creado con el nombre **n_tr_app** a partir de la clase Transacción y se lo utiliza en la aplicación SPACV a través de una instancia en la variable global **gtr_diad**.

Objeto conexión.- Diseñado como contenedor de las funciones principales de conexión a la base de datos. De acuerdo a los parámetros de conexión preestablecidos en el archivo de conexión **"spacv.ini"** al momento de ejecución del sistema, se realiza la conexión a la base de datos. Este objeto se ha creado

con el nombre de **n_cst_loginapp** y sus funciones se invocan desde los eventos open() y clicked() del botón cb_aceptar de la ventana **w_login.**

Nombre de clase ventana	Código	Descripción Funcional de Clase ventana
Ventana ancestro	w_anc	Ventana con el diseño original, en color gris de fondo y color negro para el texto. Contiene botones para minimizar, maximizar y salir.
Ventana base	w_base	Ventana con métodos necesarios para edición de registros de objetos activos de visualización y actualización de datos. Hereda las propiedades y métodos de la ventana ancestro.
Ventana de Interfaz múltiple de documentos (MDI - Multiple Document Interface)	w_frame	Ventana ancestro contenedora principal o MDI. Sirve de armazón al resto de ventanas que se activan en la pantalla principal a través de las opciones del menú del sistema.
Ventana de petición de respuestas.	w_response	Ventana que detiene el proceso de ejecución normal para pedir una respuesta de usuario a través de botones de comando. Estos botones pueden tener funciones de Aceptar, Cancelar o de Ayuda.
Ventana Logín (Ventana de autentificación de usuario y conexión al sistema).	w_login	Esta ventana es una herencia de la ventana de petición de respuestas y permite la introducción de identificación y contraseña de usuario. Una vez que los datos han sido reconocidos por el sistema, se ejecuta la ventana principal de opciones w_frame y se activa el menú del sistema.
Ventana de petición de parámetros de impresión	w_opcion_impresion	Ventana petición de respuestas diseñada con controles necesarios para pedir especificaciones de impresión al usuario, tales como número de páginas, tipo de impresora, orientación del papel, etc.

Tabla 3.20. Clases de ventana ancestro y contenedoras

3.3.3.3 Creación de clases de objetos de interfaz.

Las clases de objetos de interfaz se crean a partir de clases ventana "w_". La ventana "w_frame" se instancia con el objeto menú principal "m_marco_principal", y a través de estos objetos se hace el llamado al resto de ventanas de interfaz a voluntad del usuario.

Las ventanas contienen botones de comando "cb_" o "pb_", y a través de estos se hace el llamado a funciones y eventos. Las funciones globales se inician con "f_" y las funciones locales con "fl_". Estas últimas solo pueden ser llamadas desde el objeto ventana que las contiene.

En las ventanas de mantenimiento y de presentación se han incrustado objetos contenedores de datos que son instancian de la clase "uo_dw_base", creada a partir de la clase visual de PowerBuilder "Datawindow". La interfaz utiliza ciertos eventos de usuario creados y encapsulados en esta clase para llevar a cabo las operaciones de inserción, eliminación y actualización, y son nombrados como "ue_insertar", "ue_borrar", "ue_editar" respectivamente. El resto de eventos de usuario se reconocen por "ue_" seguido de su nombre descriptivo abreviado..

Estos eventos utilizan la instancia de la clase no visual **n_tr_app** a través de la variable global **gtr_diad** cada vez que requieren realizar las transacciones de inserción, eliminación y actualización.

Nombre de clase en el sistema SPACV	clase ancestro	Descripción Funcional
w_mant_password	w_response	Ventana de creación de contraseña. Realiza el mantenimiento de las contraseñas de usuario.
w_mant_edifica	w_base	Ventana de mantenimiento de casos de edificaciones. Realiza ingreso, eliminación e importación de casos de edificación de viviendas.
w_mant_serie_indicador	w_base	Ventana de mantenimiento de series de indicadores. Realiza el mantenimiento de las series indicadores para variables de análisis.
w_mant_usu	w_base	Ventana de mantenimiento de usuarios. Realiza la creación, eliminación y descripción de usuarios.
w_mant_localidad	w_base	Ventana de mantenimiento de localidad geográfica. Realiza el ingreso, eliminación, actualización e importación de localidades geográficas según división política administrativa país.
w_mant_tiempo	w_base	Ventana de mantenimiento de tiempo. Realiza ingreso, eliminación, actualización e importación de registros de tiempo en años y meses.
w_mant_parametro	w_base	Ventana de mantenimiento de parámetros. Realiza ingreso, actualización y eliminación de parámetros generales para el algoritmo genético.
w_mant_caso	w_base	Ventana de mantenimiento de casos de predicción. Introduce casos por localidad y tiempo y realiza todo el proceso de predicción y análisis incluido el llamado al algoritmo genético para descubrimiento de reglas.
w_proceso_ag_pasos_reglas	w_response	Ventana de ejecución del algoritmo genético. Permite seguir paso a paso todo el proceso de ejecución del algoritmo genético para descubrimiento de reglas.
w_pred_opciones_reglas	w_response	Ventana de introducción de opciones para extraer reglas. Permite ingresar nuevos tipos de reglas y sus opciones
w_pred_ag_ingresar_reglas	w_response	Ventana de ingreso de reglas. Permite ingresar una regla de asociación al sistema.
w_pred_opciones_hipotesis	w_response	Ventana de introducción de opciones para evaluar hipótesis. Permite ingresar las opciones para evaluar una regla de asociación mediante contraste de hipótesis dado una muestra de ejemplos
w_pred_opciones_prueba	w_response	Ventana de presentación de resultados de la prueba de hipótesis. Permite describir los resultados de la prueba de hipótesis de acuerdo a las opciones de contraste de hipótesis dadas.
w_pred_base_ayuda	w_response	Ventana de presentación de ayuda del sistema. Permite presentar la ayuda de navegación del sistema.
m_marco_principal	Menú	Menú principal de navegación del sistema. Menú principal del sistema con las opciones de navegación.
uo_dw_base	datawindow	Contenedor de datos para usarlo como objeto incrustado en las ventanas. Se instancia a partir de la clase genérica "datawindow" de PowerBuilder. Contiene funciones y eventos transaccionales de consulta, ingreso, eliminación y actualización de datos.

Tabla 3.21. Clases de objetos de interfaz

3.4. IMPLEMENTACIÓN.

3.4.1 IMPLEMENTACIÓN DEL SISTEMA SPACV.

3.4.1.1 Implementación de las Librerías.

La librería principal de la aplicación es **spacv.pbl,** y en ésta se encuentra el objeto aplicación "**spacv"** para acceso al sistema. Tanto ésta como el resto de librerías y sus objetos estarán en el directorio **\SPACV**, en el cliente.

Mediante el asistente de creación del destinatario o target de la aplicación se crea la ventana principal de la aplicación: **w_frame**, la cual es ventana contenedora del menú de opciones **m_marco_principal**, y para la cual se utilizan las clases windows MDI y menu de PowerBuilder respectivamente.

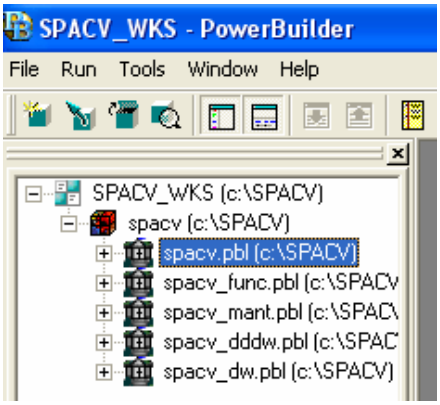


Figura 3.36 Desarrollo de la aplicación SPACV – Creación de librerías y objetos principales.

A continuación se crean las clases de interfaz, los objetos ancestro y sus herencias para luego irlos colocando en las librerías respectivas.

Spacv.pbl.- Librería principal de objetos, contiene la aplicación spacv, objetos de conexión a la base de datos, objetos y clases contenedoras, clases y objetos ancestro. Los objetos desde dónde se ejecuta y se cierra la aplicación como la ventana w_frame, el menú del sistema m_marco_principal, la ventana de acceso w_login, y objetos no visuales de conexión a la base de datos ntr_app y de ejecución de la aplicación: ntr_cst_loginapp

spacv_func.pbl .- Librería de funciones comunes.

spacv_mant.pbl.- Librería de ventanas y objetos de mantenimiento. Fue dividida en dos partes: **spacv_mant_01**.pbl y **spacv_mant_02.pbl** para optimizar velocidad.

spacv_dddw.pbl.- Librería de objetos tipo “dropdowndatawindows” usados para recuperar datos que se incorporan a las listas de selección en las interfaces.

spacv_dw.- Librería de objetos tipo “datawindow” que recuperan datos para consulta y actualización en el sistema y se los usa en las ventanas de mantenimiento o de proceso en el sistema.

3.4.1.2 Implementación de los Subsistemas.

Están implementados por medio de los módulos de interfaz, los cuales tienen su llamado en cada opción del menú principal del sistema, en el objeto "m_marco_principal". La opción "Datos" contiene el código de acceso a los subsistemas de mantenimiento de edificaciones e indicador económico. La opción "Procesos" contiene el llamado al módulo de ventana "w_caso_predicion", con la implementación de los subsistemas casos de predicción y algoritmo genético.

Sistema SPACV			
Implementación de los Subsistemas.			
Subsistema	**Clases que intervienen**	**Opción del menú**	**Módulos de Interfaz**
Mantenimiento de Edificaciones	edificación restricción_edificación grupo_restricción tiempo; localidad	Datos	w_mant_edifica w_mant_localidad w_mant_tiempo
Mantenimiento de Indicadores Económicos	serie_indicador tiempo; localidad indicador_económico	Datos	w_mant_serie_indicador w_mant_indicador
Predicción	tiempo; localidad caso_prediccion regla_asociación tipo_variable variable_caso_de_predicción variable_caso_indicador	Procesos	w_mant_caso
Algoritmo Genético	parámetro; generación cromosoma_individuo individuo cromosoma_gen tipo_cromosoma	Procesos	w_proceso_ag_pasos_reglas w_pred_ag_opciones_reglas w_pred_ag_ingresar_reglas w_ag_opciones_hipotesis w_ag_opciones_prueba w_mant_parametro

Tabla 3.22. Implementación de los Subsistemas de SPACV

3.4.1.3 Implementación de clases, operaciones y módulos.

La implementación de las operaciones de las clases Caso_predicción, Individuo, Generación, Cromosoma_Individuo, Regla_asociación, Parámetro y otras, incluidas las tareas de predicción y análisis, se lo hace a través de una clase de interfaz de tipo ventana contenedora de métodos y funciones, denominada "w_mant_caso". El propósito de ésta es aprovechar los métodos y eventos encapsulados presentes en los objetos ancestro "w_base" y "uo_dw_base". Esto se logra instanciando controles datawindow "dw_" , incrustados en sus secciones, los cuales realizan las transacciones a través de objetos datawindow "d_", a través de un código de recuperación escrito en instrucciones embebidas en SQL..

3.4.1.3.1. Implementación del módulo de casos de predicción

La ventana "w_mant_caso" contiene en la sección "tab_ag" el control datawindow "dw_mant_caso" con el nombre abreviado "dw_c", a través del cual se realiza la introducción de casos de predicción, mediante la activación de los eventos ue_insertar(), ue_borrar(), ue_editar(), los cuales se ejecutan una vez que el usuario da clic en los iconos de inserción, eliminación y edición de la barra de herramientas. El evento ue_insertar hace el llamado a ue_antesdeinsertar() desde dónde se hace el llamado a la función fl_insertar_registro_caso(). Esta función realiza la apertura de la ventana w_pedir_tiempolocalidad_caso(), en la cual se introducen los parámetros de localidad y tiempo del caso de predicción.

Estos parámetros de localidad y tiempo obtenidos de la ventana se transmiten a una instancia de la variable de estructura "lstr_c" y son usados en la instrucción de inserción del nuevo registro en la base de datos

En las secciones de análisis y predicción de la ventana "w_caso_predicción" se encuentran los controles necesarios para realizar el análisis exploratorio de los datos y la aplicación de técnicas de inteligencia artificial para predicción, de las cuales se halla implementada la técnica de algoritmo genético para descubrimiento de reglas de asociación y predicción por clasificación.

3.4.1.3.2. Implementación del algoritmo genético en el módulo casos de predicción.

Los objetos de control "datawindow" incrustados en la subsección de parámetros de la sección predicción algoritmo genético, con los nombres de "dw_mant_parametro", "dw_mant_cromosoma", "dw_mant_genes", permiten la introducción de parámetros del algoritmo genético a través del polimorfismo de los eventos "ue_antesdeinsertar", "ue_insertar", "ue_borrar", "ue_editar", los cuales tienen un comportamiento diferente según el objeto y la tabla que alteran.

En "dw_mant_parametro" se actualizan los porcentajes de cruce, mutación, número máximo de corridas, de generaciones y el índice de convergencia máximo en porcentaje. En "dw_mant_cromosoma" se introducen los alelos del antecedente del cromosoma regla que representan a las variables de construcción de viviendas. Los puntos de cruce y sus puntos de mutación son descriptivos aquí, ya que estos se generan al azar en el proceso.

En "dw_mant_genes" se introducen tanto en binario como en decimal los rangos máximos y mínimos de cada cromosoma que pueden alcanzar los fenotipos y genotipos de las variables del antecedente y consecuente en las reglas de asociación ó individuos de la población, según el algoritmo genético.

Los controles "cb_parar", "cb_iniciar", "cb_seguir", "cb_terminar_algoritmo", "cb_guardar" incrustados en la subsección "proceso" de la sección "algoritmo genético" en la ventana "w_caso_predicción", contienen instrucciones para iniciar y realizar el ciclo del algoritmo genético hasta el final, conducido por el usuario.

El control "cb_parar" está disponible para alterar las opciones de ejecución del algoritmo genético cuando sea necesario. En el evento "clicked()" de este control se encuentra el llamado a ejecución de la ventana "w_pred_ag_opciones_reglas", en la cual se altera el tamaño de la muestra de ejemplos, se ingresa y se selecciona el tipo de regla de asociación a descubrir y se decide si el tipo de procesamiento se lo hará paso a paso o hasta finalizar.

El tipo de estructura para la regla de asociación a descubrir y elegida en opciones, permanece en una instancia de la variable global de estructura "str_ag_regla_asociacion". En ésta se halla la definición de activación de genes del antecedente en la cadena "activación_antecedente", la cual, por motivos de seguridad y respaldo se copia al archivo "reglas.ini" en el destino de la aplicación.

El control "cb_iniciar", en su evento "clicked()" hace el llamado a la función "fl_inicio_prediccion()", a través de la cual se crean los individuos "cromosoma_regla" para la primera generación. La asignación de alelos para el genotipo de la regla se la realiza al azar con valores permitidos de 1 a 255 para cada una de las siete secciones del cromosoma que se corresponden con los genes establecidos y que representan a las siete variables posibles para el antecedente de la regla, a saber: "cimientos"; "estructura"; "paredes", "tipo de obra", "uso de la edificación", y "origen del financiamiento".. Esta asignación inicial se hace a través del llamado a las funciones "f_inserta_reg_individuo()" y "f_insertar_reg_cromosoma_individuo".

En la función "fl_inicio_prediccion()" se llama a la función global de adaptación "f_adaptacion_individuo()" para cada individuo de la primera generación. Esta función evalúa la adaptación de los primeros "cromosoma_regla" creados, en función de sus valores de confianza y cobertura que tienen con relación a una muestra de ejemplos de casos de edificaciones, la cual es extraída al azar por medio de la función "fl_extrae_casos_ejemplo_regla()", que a su vez llama a la función global "f_enc_casos_edifica_regla()", la que busca los casos por lugar, tiempo y según el antecedente y consecuente del tipo de regla establecida. Los casos de ejemplo permanecen en un objeto "datastore" de almacenamiento temporal instanciado en la variable "ids_casos" mientras dura la sesión iniciada para el proceso de extracción de reglas del algoritmo genético.

El control "cb_seguir", en su evento "clicked()" hace el llamado a la función "fl_algoritmo_prediccion_reglas", la cual continúa el proceso del algoritmo genético hasta su convergencia, a través de la ejecución de la ventana "w_proceso_ag_pasos_reglas". Los controles y botones de comando de esta

ventana se activan sólo si en las opciones de ejecución se hubiere pedido que el proceso se realice paso a paso. No obstante, si en opciones se ha elegido procesar hasta el final, entonces desde el evento "open()" de la misma ventana se hace el llamado a la función "fl_procedimiento_algoritmo_genetico()", la cual aplica el algoritmo genético propiamente dicho sin intervención del operador hasta descubrir las reglas de asociación de mayor valor de aptitud, con la estructura de antecedente-consecuente que fuera predefinida en opciones..Desde esta función se hace el llamado a las funciones de selección cruce, mutación. En estas funciones se aplican los operadores genéticos a través de las funciones "fl_operacion_cruce()" y "fl_operacion_mutacion()".

Cuando se aplica cruce y mutación es necesario volver a evaluar la aptitud del "cromosoma-regla", a través de la función "f_adaptacion_individuo()" y registrar los cambios mediante las funciones "fl_actualiza_seleccion_por_torneo()", "fl_actualiza_cruce()" y "fl_actualiza_mutacion()".

El ciclo del algoritmo termina una vez que se ha superado el límite de convergencia evaluado en la función "fl_calc_indice_convergencia()", o se ha superado el número de generaciones permitido y extraído de los parámetros, validación que se realiza en la función "fl_validar_convergencia()".

Las reglas de asociación de mayor aptitud descubiertas pasan a formar parte de la última generación de individuos "cromosomas-regla" y se muestran en los controles "datawindow" "dw_indiv" y "dw_cromo_indiv" de "w_mant_caso", con su genotipo resultado y su estructura antecedente-consecuente, respectivamente.

El control "cb_terminar_algoritmo", se activa a través de su evento "clicked()" para agrupar y clasificar las reglas de asociación descubiertas, dejando las de mayor aptitud, y descartando las demás, desde la última generación de individuos.

Luego, si por acción del operador se activa el evento "clicked" del control "cb_guardar", entonces los resultados son almacenados permanentemente en la tabla "regla_asociacion" para el caso de predicción actual seleccionado.

Las reglas así descubiertas y almacenadas para el caso de predicción seleccionado se recuperan por acción del operador en el control datawindow "dw_reglas_aso", a través del evento "clicked()" del control "pb_reglas_aso" ubicado en la subsección "resultados" de la sección "prediccion-algoritmo genético" de la ventana "w_caso_prediccion".

En la subsección "resultados" indicada anteriormente, se encuentran los controles "cb_ociones_pruebas" y "cb_probar_hipotesis", y en la parte central está el control datawindow "dw_reglas_aso", necesarios para probar las reglas como hipótesis.

El evento "clicked()" de "cb_opciones_pruebas" activa la ventana "w_pred_ag_opciones_hipotesis" para predefinir entre otros parámetros, el nivel de significación de la prueba, el tamaño de la muestra de casos de ejemplo, la decisión de tomar o no todos los casos de ejemplo, y la variable parámetro que podría ser de cobertura o de confianza. Todas estas opciones se necesitan para aplicar contraste de hipótesis a la nueva regla extraída.

El evento "clicked()" del control "cb_probar_hipotesis" permite probar la regla seleccionada en el llamado a la función "fl_probar_regla_aso()". Esta, llama a la función global "f_evaluar_regla_en_casos_de_prueba()", donde se evalúa la regla, ya sea sobre una nueva muestra de casos de ejemplo de edificaciones extraída al azar, ó bien tomando todos los casos de ejemplo. En la evaluación, se califica el antecedente de la regla mediante la función "f_clasificar_casos_de_prueba_regla()", y se calcula la completitud y la confianza. Si la decisión es tomar todos los ejemplos del caso, entonces se llama a la función "fl_extrae_poblacion_casos_regla()", de lo contrario se llama a "fl_extrae_casos_ejemplo_regla()". Los resultados se presentan desde la ventana "w_pred_ag_opciones_prueba", desde donde se indica si se acepta o se rechaza la hipótesis nula para la regla de asociación extraída y seleccionada.

Resumen final de Clases, operaciones y módulos				
Sistema: SPACV; Subsistema: Algoritmo Genético				
Clase	**Operación**	**Módulo de Interfaz**	**Método o Función**	**Evento**
Pará-metro	ingresar_parametro	w_mant_caso. .dw_mant_parametro	fl_inserta_reg_param_caso	ue_antesdeinsertar()
	modificar_parametro			ue_editar()
	Eliminar_parametro			ue_antesdeborrar()
Gene-ración	Selección	w_proceso_ag_pasos_reglas	fl_seleccion_por_torneo()	pb_siguiente.clicked() open()
	crear_generacion	w_mant_caso	fl_inicio_prediccion() fl_preparar_datos_caso() fl_insertar_reg_generacion _caso()	cb_iniciar.clicked()
	Eliminar_generacion	w_mant_caso .dw_mant_caso	f_elimina_un_caso_prediccion()	dw_mant_caso.ue_antesdeborrar()
	evaluar_adaptacion_generacion	w_proceso_ag_pasos_reglas	fl_calc_indice_convergencia()	pb_actualizar_conv.clicked() open()
Cromo-soma Indivi-duo	ingresar_cromo_indiv	w_mant_caso	fl_inicio_prediccion() fl_insertar_reg_cromosoma _individuo()	cb_iniciar.clicked()
	Eliminar_cromo_indiv	w_mant_caso.dw_mant_caso	f_elimina_un_caso_prediccion()	dw_mant_caso_ue_antesdeborrar(); ue_borrar()
Indivi-duo	Cruce	w_proceso_ag_pasos_reglas	fl_cruce()	pb_siguiente.clicked() open()
	Mutación		fl_mutacion()	
	Clonacion		fl_actualiza_seleccion_por_torneo()	
	crear_individuo	w_mant_caso	fl_inicio_prediccion() fl_insertar_reg_individuo()	cb_iniciar_ckicked()
	Eliminar_individuo	w_mant_caso.dw_mant_caso	f_elimina_un_caso_prediccion()	dw_mant_caso_ue_antesdeborrar(); ue_borrar()
	evaluar_adaptabilidad_individuo	w_mant_caso	fl_algoritmo_prediccion_reglas() f_adaptacion_individuo()	dw_mant_caso.cb_seguir.clicked()
	codificar_genotipo	w_mant_caso	fl_inicio_prediccion() fl_insertar_reg_cromosoma _individuo f_genera_aleatorio() f_conver_decabin()	cb_iniciar.clicked()
	decodificar_genotipo	w_mant_caso	fl_probar_regla_aso() f_evaluar_regla_en_casos_de_prueba() f_clasificar_casos_de_prueba_regla() f_conver_binadec()	cb_probar_hipotesis.clicked()
Cromo-soma gen	ingresar_cromosoma_gen	w_mant_caso.dw_mant_genes	fl_insertar_genes()	ue_antesdeinsertar()
	modificar_cromosoma_gen			ue_editar()
	Eliminar_cromosoma_gen			ue_borrar()
Tipo Cromo-soma	ingresar_tipo_cromosoma	w_mant_caso.dw_mant_cromosoma	fl_insertar_registro_crom()	ue_insertar()
	Eliminar_tipo_cromosoma			ue_borrar()

Tabla 3.23. Resumen final de clases, operaciones y módulos - Subsistema Algoritmo Genético.

3.5. PRUEBAS.

3.5.1 PRUEBAS DE UNIDAD.

3.5.1.1 Plan de pruebas de unidad.

El proceso de pruebas de unidad a seguir para el sistema SPACV es el siguiente:

1. Preparar los casos de prueba para operaciones de clase, módulos, funciones y eventos críticos.
2. Revisar la funcionalidad del código
3. Separar los segmentos de código críticos.
4. Realizar pruebas de caja negra en los módulos de mayor prioridad, con el programa principal actuando como conductor.
5. Realizar pruebas de caja blanca en los segmentos de código de mayor complejidad y críticos.
6. Aprobar los módulos y calificar los casos de prueba

Los recursos de programación y depuración a usar serán los del entorno de programación de PowerBuilder. Servirán de conductores los llamados a funciones presente en los eventos de los botones de comando de las ventanas de interfaz de cada módulo. Se consideran ya probados los eventos de objetos ancestros.

3.5.1.2 Descripción de los casos de prueba de unidad.

Casos de prueba de unidad para el Sistema SPACV					
Subsistemas: Mantenimiento Edificaciones, Indicadores					
No. del Caso de Prueba	**Propósito de la Prueba**	**Módulos**	**Objeto.método**	**Tipo**	**Prioridad**
101	Asegurar ingreso, edición y eliminación de datos correctos en las tablas.	.w_mant_edifica .w_mant_serie_indicador .w_mant_indicador .w_mant_localidad .w_mant_tiempo .w_mant_parametro .w_opciones_de_mantenimiento	.uo_dw_base.insertar()	Evento	1
			.uo_dw_base.borrar()	Evento	
			.uo_dw_base.editar()	Evento	
102	Asegurar importación correcta de datos en las tablas		.w_mant_edifica.fl_importar_reg()	Función	2
			.w_mant.indicador.fl_importar_reg()	Función	
			.w_mant_localidad.fl_importar_reg()	función	
			.f_inserta_parametro()	función	

Tabla 3.24 Casos de prueba de unidad - Subsistema: Mantenimiento Edificaciones, Indicadores

Casos de prueba de unidad para el Sistema SPACV					
Subsistema: Procesamiento Casos de Predicción					
No. del Caso de Prueba - Propósito de la Prueba		**Módulo**	**Objeto.método**	**Tipo**	**Prioridad**
201	Ingreso y recuperación correcto de un caso de predicción	w_mant_caso	open()	Evento	1
			dw_mant_caso.ue_antesdeinsertar()	Evento	
			fl_insertar_registro_caso()	Función	
202	Análisis y exploración de casos históricos confiable	w_mant_caso	fl_inicio_rbc()	Función	4
			fl_guardar_resultados_rbc()	Función	
203	Inicialización correcta de variables para predicción	w_mant_caso	fl_prepara_datos_caso()	Función	2
204	Eliminación sin fallos de un caso de predicción	w_mant_caso	dw_mant_caso.ue_antesdeborrar()	Evento	3

Tabla 3.25. Casos de prueba de unidad - Subsistema: Casos de Predicción.

Casos de prueba de unidad para el Sistema SPACV					
Subsistema: Procesamiento Algoritmo Genético					
No. de Caso de Prueba Propósito de la prueba		**Módulo**	**Objeto.método**	**Tipo**	**Prior idad**
301	Ingreso y recuperación correcta de parámetros del algoritmo genético	.w_mant_caso	dw_mant_parametro.ue_antesdeinsertar()	evento	5
			dw_mant_parametro.ue_antesdeborrar()	evento	
			dw_mant_parametro.ue_editar()	evento	
			dw_mant_cromosoma.ue_antesdeinsertar()	evento	
			dw_mant_genes.ue_antesdeinsertar()	evento	
			dw_mant_genes.itemchanged()	evento	
302	Introducción de la estructura de una regla de asociación correcta	.w_mant_caso	w_pred_ag_opciones_regas.fl_agregar_nueva_regla()	función	3
			w_pred_ag_ingresar_reglas.fl_validar_regla()	función	
			w_pred_ag_ingresar_reglas.fl_agregar_nuevo()	función	
			w_pred_ag_opciones_reglas.fl_llenar_reglas()	función	
303	Ciclo de ejecución del algoritmo genético hasta su convergencia libre de fallos.	.w_mant_caso	fl_inicio_prediccion()	función	2
			fl_algoritmo_prediccion_reglas()	función	
			w_proceso_ag_pasos_reglas .fl_procedimiento_algoritmo_genetico()	función	
304	Ejecución correcta de la prueba de aceptación para una regla de asociación descubierta	.w_mant_caso	w_pred_ag_opciones_hipotesis.cb_aceptar.clicked()	evento	7
			fl_probar_regla_aso()	función	

Tabla 3.26. Casos de prueba de unidad - Subsistema: Algoritmo genético.

3.5.1.3 Ejecución de las pruebas de unidad y evaluación de resultados.

Resultados de las Pruebas de Unidad		Tipo de Prueba: Caja Negra		
Subsistema: Mantenimiento Edificaciones, indicadores				
Caso de Prueba	**Entradas**	**Módulo de prueba Objeto.método**	**Salidas**	**Aceptar**
101	.codigo_de_provincia: 17 .codigo_de_canton: 01 .código_de_parroquia: 50	.w_mant_localidad .dw_mant.uo_dw_base .ue_antesdeinsertar() .ue_insertar()	Salidas Esperadas: .nombre_localidad: Quito .nivel_de_localidad: 04 Salidas Obtenidas: Tabla: LOCALIDAD .nombre_localidad:QUITO .nivel_de_localidad: 04	Si
102	.archivo de lectura .EED2005W.TXT	.w_mant_edifica .fl_importar_reg()	Salidas Esperadas: Registros importados .desde 2005000001 hasta 2005018824 Salidas Obtenidas: Tabla: EDIFICA Reistros importados .del 2005000001 al 2005018824	Si

Tabla 3.27. Casos de prueba de unidad - Subsistema: Mantenimiento Edificaciones, Indicadores

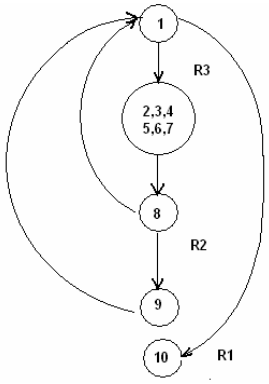
Resultados de las Pruebas de Unidad			Tipo de Prueba: Caja Negra	
Subsistema: Casos de Predicción y Análisis				
Caso de Prueba	**Entradas**	**Módulo de Prueba Objeto.método**	**Salidas**	**Aceptar**
201	.año: 2006 .mes: Todos .provincia: 03 .canton: 01 .parroquia: 50	.w_mant_caso .dw_mant_caso.antesdeinsertar() .fl_insertar_registro_caso()	Salidas Esperadas: .nuevo caso : tiempo: 2006-00 localidad: 03-01-50 Salidas Obtenidas: Tabla: CASO_DE_PREDICCION .Año: 2006 Localidad 030150 Azoguez. .caso: 00003; cod_caso: 2 .id_tiempo: 200600; .id_localidad = 2002030150	Si
202	Caso: 00003 cod_caso:2 .año 2006 .localidad: Azoguez	.w_mant_caso .fl_inicio_rbc() .fl_guardar_resultados_rbc()	.Salidas Esperadas: .valores promedio para el metro cuadrado de construcción .valores de análisis para otras variables de edificaciones. .Salidas Obtenidas: PANTALLA . VALOR DEL METRO CUADRADO DE CONSTRUCCIÓN Valor máximo alcanzado: $ 600 Valor mínimo alcanzado: $ 20 Valor de confianza máximo: $ 197 Valor de confianza mínimo: $ 47 promedio: $ 122 desviación estandar: 75	Si
203	Caso: 00003 cod_caso:2 .año 2006 .localidad: Azoguez	.w_mant_caso.open() .dw_mant_caso.rowfocuschanged() .fl_preparar_datos_caso()	.Salidas Esperadas: .variables del caso inicializadas .controles de pantalla activados. .Salidas Obtenidas: .año inicial: 2006 .Tablas: PARAMETRO; GENERACION .id_parametro: 03; id_caso: 02 .id_generacion: 02; id_caso: 03 .Controles: . cb_iniciar.enabled = true . resto de controles con valor = false	Si
204	Caso: 00003 cod_caso:2 .año 2006 .localidad: Azoguez	.w_mant_caso .dw_mant_caso.ue_antesdeborrar() .f_elimina_un_caso_prediccion()	Salidas Esperadas: . eliminación de caso 00003 de todas las tablas dependientes de CASO_PREDICCION Salidas Obtenidas: .registros de caso 00003 en: .Tablas: GENERACION: 0 PARAMETRO: 0 REGLA_ASOCIACION: 0 VARIABLE_CASO_PREDICCION: 0 VARIABLE_CASO_INDICADOR: 0 CASO_PREDICION: 00003 eliminado.	

Tabla 3.28. Resultados de las Pruebas de Unidad - Subsistema: Casos de Predicción y Análisis

Resultados de las Pruebas de Unidad			Tipo de prueba: Caja Negra	
Subsistema: Algoritmo Genético				
Caso	**Entradas**	**Módulo de Prueba Objeto.método**	**Salidas**	**Aceptar**
301	Caso: 00002 cod_caso:1 .año 2006 .localidad: Quito	.w_mant_caso .dw_mant_parametro .ue_antesdeinsertar() .dw_mant_cromosoma .ue_antesdeinsertar() .fl_inserta_reg_param_caso()	Salidas Esperadas: .caso 00002 con parámetros inicializados para procesar algoritmo genético. Salidas Obtenidas: .Tablas: .PARAMETRO caso 00002 . porc de cruce: 40% . porc de mutación: 10% . número máx. de generac.: 5 . número máx. de corridas: 1 . indice de convergencia: 75% .CROMOSOMA_GEN .TIPO_CROMOSOMA . cima, estru, paredes, cubi, tipo de obra, uso de edificación, origen de financiamiento.	Si
302	.Regla: .antecedente: .CIMI → No .ESTRU → No .PAREDES → Si .CUBI → Si .consecuente: . Dist_valor_m2 = . >= $100 y < $200	.w_mant_caso .w_pred_ag_opciones_regas. . fl_agregar_nueva_regla() . fl_validar_regla() .	Salidas Esperadas: . estructura de la regla de asociación. Salida Obtenida: . Si PAREDES = ? Y CUBI = ? . entonces . Dist_valor_m2 = (>= $100 y <= $200) . registro en archivo de parámetros REGLAS.INI:R08=001100033002	Si
303	Caso: 00002 cod_caso:1 .año 2006 .localidad: Quito .Estructura de regla: . Si PAREDES = ? Y CUBI = ? . entonces . ESTRU = Hormigón armado Parámetros: ·# individuos = 500 .tam.muestra = 100 .% cruce = 40 .% mutación = 10 .num max gen = 5 .num max corr = 1 .ind.converg= 75%	.w_mant_caso .fl_inicio_prediccion() .fl_algoritmo_prediccion_reglas() .w_proceso_ag_pasos_reglas .fl_procedimiento_algoritmo_genetico()	Salidas Esperadas: Valores para PAREDES y CUBI, .consecuente: ESTRU = Horm arm. Salidas Obtenidas: Si PAREDES = Ladrillo y CUBI = Teja, .entonces ESTRU = Horm arm. Si PAREDES = Bloque y CUBI = Teja, .entonces ESTRU = Horm arm. Si PAREDES = Bloque y CUBI = Eternit-ardex-asbesto, .entonces ESTRU = Horm arm. Si PAREDES = Ladrillo y CUBI = Horm arm., .entonces ESTRU = Horm arm. Si PAREDES = Bloque y CUBI = Horm arm., .entonces ESTRU = Horm arm.	Si
304	Caso: 00002 cod_caso:1 .año 2006 .localidad: Quito Regla descubierta: Si PAREDES = Bloque y CUBI = Horm.Arm., .entonces ESTRU = Horm.Arm	.w_mant_caso .w_pred_ag_opciones_hipotesis .cb_aceptar.clicked() .fl_probar_regla_aso()	Salidas Esperadas: .parametros de confianza y cobertura .resultados de prueba de hipótesis Salidas Obtenidas: .cobertura: = 67.94; confianza = 96.33 .	Si

Tabla 3.29 Resultados de las Pruebas de Unidad - Subsistema: Algoritmo genético. Caja Negra

Resultados de las Pruebas de Unidad		Tipo de prueba: Caja Blanca - Cobertura de Caminos		
Subsistema: Algoritmo Genético				
Caso	**Entradas Caminos**	**Módulo de Prueba Objeto.método**	**Salidas**	**Aceptar**

303	Caso: 00002
	cod_caso:1
	.año 2006
	.localidad: Quito
	
	.Estructura de regla:
	. Si PAREDES = ?
	Y CUBI = ?
	. entonces
	. ESTRU =
	Hormigón armado
	
	Parámetros:
	·# individuos = 500
	.tam.muestra = 100
	.% cruce = 40
	.% mutación = 10
	.num max gen = 5
	.num max corr = 1
	.ind.converg= 75%
		.fl_procedimiento_algoritmo_genetico()
		li_nmax_de_generaciones=istr_ag.ds_reg_parametro.Object.numero_de_generaciones[1]
		// Compara con numero de generaciones actual
		if (igeneracion >= li_nmax_de_generaciones) then
		RETURN 0
		end if
		// bucle del proceso completo del algoritmo genético
		lb_finalizar = false
		DO // **(1)**
		// Procedimientos secuenciales: lazos de inicialización,
		// lazos de selección, cruce y mutación
		// . . **.(2) (3) (4) (5) (6) (7)**
		// evalua convergencia
		ib_alcanzo_convergencia = false
		fl_llenar_dw_dist()
		idec_ind_conv = fl_calc_indice_convergencia()
		ib_alcanzo_convergencia = fl_validar_convergencia(..)
		//
		if ib_alcanzo_convergencia then // **(8)**
		lb_finalizar = true
		else // **(9)**
		if (igeneracion >= li_nmax_de_generaciones) then
		lb_finalizar = true
		else
		// continuar con nueva generación
		// hasta cumplir condiciones de parada
		igeneracion = igeneracion + 1
		end if
		end if
		LOOP UNTIL lb_finalizar
		// fin (10)
		Cobertura de caminos
		 
		
		Grafo de flujo de estructuras lógicas del algoritmo genético
		Complejidad Ciclomática V(G) = Aristas – Nodos + 2
		V(G) = 6-5 + 2 = 3
			"Cromosoma-Regla"
			en Tabla CROMOSOMA_INDIVIDUO.
			
			"Reglas "
			en Tabla REGLA_ASOCIACION
				Si

	num max gen = 1 ind_converg <= 35 %	Camino 1 : 1-10	reglas extraídas >= 0	Si
	num max gen > 0 ind_converg <= 35 %	Camino 2: 1-2-3-4-5-6-7-8-1-10	reglas extraidas >= 0	Si
	num max gen >= 5 ind_converg >= 75 %	Camino 3: 1-2-3-4-5-6-7-8-9-1-10	reglas extraidas >= 0	Si

Tabla 3.30. Resultados de las Pruebas de Unidad - Subsistema: Algotrimo genético.

3.5.2 PRUEBAS DE INTEGRACIÓN.

3.5.2.1 Plan y estrategias de integración

Se ha diseñado tres estrategias: La primera consiste en realizar una integración descendente y en anchura para todos los módulos de mantenimiento de tablas presentes en los subsistemas de edificaciones, indicadores, y parámetros.
La segunda estrategia tiene que ver con la integración del módulo de casos de predicción al sistema, la cual se la hará en forma descendente y en profundidad.
La tercera estrategia se aplicará al subsistema algoritmo genético y consistirá en integrar todas las funciones y eventos del algoritmo de forma ascendente hasta llegar a la ejecución completa del módulo principal de proceso de predicción.

3.5.2.2 Descripción de las fases de integración.

Se han considerado dos fases de integración:

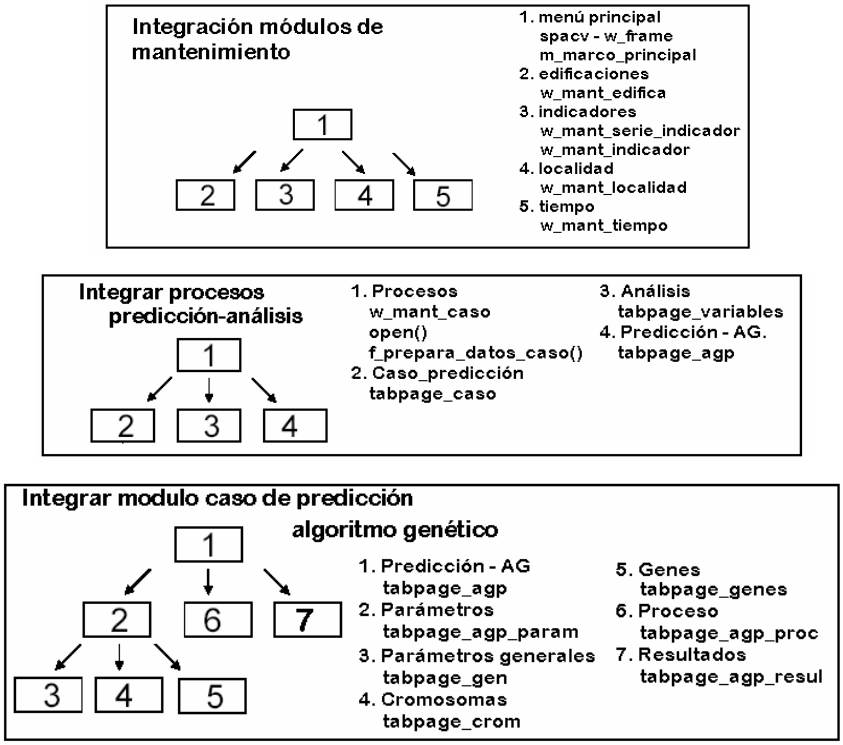


Figura 3.37. Fases de Integración de módulos.

Primera fase: Probar por separado y llevar a cabo la integración de los módulos de mantenimiento de los subsistemas de edificaciones, indicadores y parámetros.
Segunda fase: Probar en conjunto e integrar los módulos de los dos subsistemas restantes: casos de predicción y algoritmo genético. La integración se realiza entre las secciones nombradas como "tabpage_" de la ventana "w_mant_caso".

3.5.2.3 Ejecución de las pruebas de integración y evaluación de resultados

Resultados de las Pruebas de Integración				
Fase	**Subsistemas**	**Módulos a integrar**	**Ejecución y Resultados**	**Acep -tar**
1	Edificaciones	mantenimiento de edificaciones e indicadores económicos mantenimiento de localidades geográficas mantenimiento de tiempo	Se ejecuta el sistema y serealizan las siguientes tareas: 1. Importación de tablas de tiempo. 2. Importación de tablas de localidades geográficas 3. Importación de tablas de variables y de series históricas de indicadores. 4. Importación de tablas anuales de casos de edificaciones de viviendas.	Si
2	Caso de Predicción	mantenimiento de casos de predicción análisis de casos de predicción	Se ejecuta el sistema y se realizan las siguientes tareas: 1. Introcucción de un caso de predicción por localidad y tiempo 2. Análisis exploratorio de variables de edficaciones histórico hasta el año del caso de predicción	Si
2	Caso de Predicción	procesamiento de la técnica de inteligencia artificial (algoritmo genético) e integración al sistema	Se ejecuta el sistema y el algoritmo de predicción para el caso introducido. Se introducen los parámetros y se ejecuta el algoritmo hasta el final. Se obtienen los siguientes resultados: Reglas promedio extraídas para el caso de predicción: 5.	Si

Tabla 3.31. Resultados de Integración del Sistema.

.3.5.3 PRUEBAS DE VALIDACIÓN.

3.5.3.1 Plan de validación

El plan a seguir para las pruebas de validación es el siguiente:

1. Identificar los requerimientos de usuario prioritarios a validarse.
2. Identificar los criterios de aceptación del usuario final.
3. Ejecutar el sistema y probar los requerimientos de validación.

Las estrategias a seguir para aplicar este plan es la de aplicar caja-negra en los módulos afectados por los requerimientos a probar y realizar su seguimiento.

3.5.3.2 Desarrollo de las pruebas de aceptación

Identificación de requerimientos de usuario para las pruebas de aceptación.			
	Caso de Uso	**Requerimiento**	**Prioridad**
1	Gestión de Introducción de Casos de Edificaciones.	Asegurar la introducción correcta de datos históricos por localidades geográficas en los años de proceso.	1
2	Gestión de Introducción de Casos de Predicción. .	Asegurar la introducción correcta de parámetros de la técnica a aplicar y de parámetros de ejecución del algoritmo	2
3	Gestión de Procesamiento de Casos de Predicción.	Asegurar la introducción correcta de un caso de predicción, la ejecución de la técnica de predicción aplicada (algoritmo genético) y la obtención de los resultados (reglas de asociación extraídas).	3

Tabla 3.32. Identificación de los requerimientos de usuario para las pruebas de aceptación.

Identificación de Criterios de aceptación para las pruebas		
	Criterio	**Métrica aplicada**
1	Se cumplieron la mayor parte de los requerimientos ?	AMPLITUD = Número de requerimientos probados / numero de requerimientos totales
2	Se probaron al menos el 75% de los caminos en los requerimientos?	PROFUNDIDAD: % de caminos probados / total de caminos en las opciones de los requerimientos.
3	El número de fallos encontrados en las salidas es menor al número de fallos esperados ?	TOLERANCIA A FALLOS: número de fallos encontrados / total de fallos esperados.

Tabla 3-33. Identificación de los criterios de aceptación para las pruebas.

3.5.3.3 Ejecución de las pruebas de validación y evaluación de resultados

Resumen de pruebas de validación.			
	Criterio	**Métrica**	**% de aceptación.**
1	Se cumplieron la mayor parte de los requerimientos ?	AMPLITUD = (3/ 4) * 100	75 %
2	Se probaron al menos el 75% de los caminos de los requerimientos ?	PROFUNDIDAD = (13 / 15) * 100	86 %
3	El número de fallos encontrados en las salidas es menor al número de fallos esperados ?	TOLERANCIA A FALLOS = 0 / 13	100 %

Tabla 3.34. Resumen de pruebas de validación.

Resultados de las pruebas de aceptación .				
#	**Requerimiento - Entradas**	**Prueba alfa**	**Prueba beta**	**Acep-tación**
1	Asegurar la introducción correcta de datos históricos por localidades geográficas en los años de proceso. ENTRADAS: dpa_tiempo.txt dpa_localidad.txt variables.txt serie_indicador.txt indicador.txt EED2000W.TXT EED2001W.TXT EED2002W.TXT EED2003W.TXT EED2004W.TXT EED2005W.TXT	AUTENTICACIÓN: usuario: DBA contraseña: administrador Registros importados: TIEMPO. 208 . LOCALIDAD: 1262 TIPO_VARIABLE: 176 INDICADOR_EDONOMICO: 11 SERIE_INDICADOR: 467 EDIFICA: 104528 registros Resultados: Numero de caminos = 1 x 6 entradas caminos probados = 6 fallos encontrados = 0	AUTENTICACIÓN: usuario: spacv contraseña: spacv Registros importados: TIEMPO: 208. LOCALIDAD: 1262. TIPO_VARIABLE: 176. INDICADOR_ECONOMICO: 11 SERIE_INDICADOR: 467 EDIFICA: 104528 registros Resultados: Numero de caminos = 1 x 6 entradas caminos probados = 6 fallos encontrados = 0	Si
2	Asegurar la introducción correcta de parámetros de la técnica a aplicar y de parámetros de ejecución del algoritmo	Registros inicializados: PARAMETROS: 1 registro TIPO_CROMOSOMA: 7 registros CROMOSOMA_GEN: 58 registros REGLAS: 11 tipos de reglas Tipo de proceso: Hasta finalizar Número de individuos inicial: 500 Tamaño de muestra de ejemplos: 100 Resultados: Numero de caminos: 1 x 4 entradas caminos probados = 4 fallos encontrados: 0	Registros inicializados: PARAMETROS: 1 registro TIPO_CROMOSOMA: 7 registros CROMOSOMA_GEN: 58 registros REGLAS: 11 tipos de reglas Tipo de proceso: Hasta finalizar Número de individuos inicial: 500 Tamaño de la muestra de ejemplos: 100 Resultados: Numero de caminos: 1 x 4 entradas caminos probados = 4 fallos encontrados: 0	Si
3	Asegurar la introducción correcta de un caso de predicción, la ejecución de la técnica de predicción aplicada (algoritmo genético) y la obtención de los resultados (reglas de asociación extraidas).	Casos de predicción introducidos: 10 Promedio de reglas extraídas por cada caso: 5 Resultados: 1 entrada: introducción de parámetros 2 entrada: ejecución 3 entrada: aplicación del algoritmo Número de caminos: 1 x 3 entradas caminos probados: 3 fallos encontrados: 0	Caso de predicción introducidos: 10 Promedio de reglas extraidas por cada caso: 5 Resultados: 1 entrada: introducción de parámetros 2 entrada: ejecución 3 entrada: aplicación del algoritmo Número de caminos: 1 x 3 entradas caminos probados: 3 fallos encontrados: 0	Si

Tabla 3-35. Resultados de las pruebas de aceptación.

.5.4 PRUEBAS DEL SISTEMA.

3.5.4.1 Pruebas de recuperación.

Resultados de las pruebas de recuperación .				
	Detalle del Fallo e Interrupción del proceso.	**estados antes del fallo**	**estados después del fallo**	**Aceptación**
1	Se inicia importación de archivo EED2005W.txt al sistema mediante la opción de importar del módulo de mantenimiento en el subsistema EDIFICACIONES. Se produce la desconexión del sistema a la base de datos.	Datos origen: EED2005W.TXT Datos destino: EDIFICACIONES Antes de guardar.	Datos destino: inalterados Resultados: Se recuperaron los datos antes del fallo? No Se hizo necesario volver a introducir los datos manualmente ? Si Tiempo necesario para recuperar nuevamente los datos antes del fallo: 2 minutos Recuperación automática en el programa por medio de la instrucción ROLLBACK Mensaje al usuario: Fallo en la actualización de datos importados.	Si
2	Se introduce un caso de predicción general y se ejecuta el algoritmo genético para extracción de un tipo de regla. Durante la ejecución del algoritmo se interrumpe la ejecución y se cancela el programa.	Datos de entrada: caso de predicción: 1 tipos de reglas de asociación: reglas.ini Tabla parámetros.	caso de predicción: Inalterado reglas de asociación: Inalterado Tabla de parámetros inalterada. Resultados: Se recuperaron los datos antes del fallo ? Parcialmente. Se hizo necesario volver a introducir los datos manualmente ?: Unicamente los parámetros de ejecución. Tiempo necesario para recuperar los datos antes del fallo: 5 minutos (incluido reprocesamiento del algoritmo)	Si

Tabla 3.36. Resultados de las pruebas de recuperación.

3.5.4.2 Pruebas de seguridad

Resultados de las pruebas de seguridad .					
	Caso de prueba de seguridad	**Autenticación de usuario**	**Accesos esperados**	**Accesos permitidos**	**Acep-tación**
1	Acceso al sistema y ejecución de las opciones de mantenimiento de datos localidad, tiempo, edficaciones, series, usuarios	Usuario: DBA contraseña: Administrador	Todos lectura, escritura, ejecución	Todos	Si
		Usuario: PROC contraseña: proc	Procesos: Caso de predicción análisis predicción algoritmo lectura, escritura, ejecución	Procesos: Caso de predicción análisis predicción algoritmo lectura, escritura, ejecución	

		Usuario: MANT contraseña: mant	Datos Edificaciones,Indicadores, variables Localidad, tiempo Usuarios, Parámetros lectura, escritura, ejecución	Datos: Edificaciones, Indicadores, variables, Localidad, tiempo usuarios, Parámetros lectura, escritura, ejecución No permite acceso a procesos.	
2	Acceso al sistema, introducción de un caso de predicción y ejecución del algoritmo genético hasta la extracción de reglas de asociación.	Usuario: DBA contraseña: Administrador	Todos lectura, escritura, ejecución	Todos	
		Usuario: PROC contraseña: proc	a todos los casos de predicción para análisis y predicción lectura, escritura, ejecución	a todos los casos de predicción para análisis y predicción lectura, escritura, ejecución	
		para el resto de usuarios no es permitido el acceso,			

Tabla 3.37. Resultados de las pruebas de seguridad.

3.5.4.3 Pruebas de resistencia y rendimiento

Resultados de las pruebas de resistencia.					
	Caso de prueba de resistencia	**Condiciones Extremas:**	**Tiempo esperado**	**Tiempo empleado**	**Aceptación**
1	Introducción de un caso de predicción, un tipo de regla de asociación, parámetros y ejecución del algoritmo genético hasta la extracción de reglas de asociación	Caso de predicción: localidad: PAIS año: 2006 parametros: numero de generaciones max: 10 indice convergencia: 75% Número de variables en el antecedente: 7 tamaño de la muestra de ejemplos: 1000 Tamaño inicial de individuos: 1000	10 minutos	5 minutos	Si

Tabla 3.38. Resultados de las pruebas de resistencia

Resultados de las pruebas de rendimiento.						
	Caso de prueba de rendimiento	**espacio en disco**	**espacio en memoria**	**tiempo uso promedio de C.P.U**	**Tiempo Total empleado**	**Aceptación**
1	Ejecución del algoritmo genético desde inicio, selección, cruce, mutación hasta la extracción de la generación final.	reglas.ini: 1KB. base de datos: DATA: 77 Mb. LOG: 1 Mb.	236020K	70 %	5 min	Si

Tabla 3.39. Resultados de las pruebas de rendimiento.

CAPITULO 4.
APLICACIÓN DEL SISTEMA A UN CASO DE ESTUDIO.

4.1. SELECCIÓN DE EJEMPLOS DE CASOS DE PREDICCIÓN CON LAS VARIABLES INVESTIGADAS.

4.1.1. DELIMITACIÓN Y DESCRIPCIÓN DE LOS CASOS DE EJEMPLO.

4.1.1.1 Delimitación de los casos de ejemplo por localidad geográfica.

De acuerdo al estudio realizado en el primer capítulo, las ciudades de Quito, Guayaquil, Cuenca y Machala son ejes potenciales del desarrollo urbano y las más pobladas. Por consiguiente, el comportamiento de la construcción de viviendas en el país puede entenderse a través de casos de ejemplo del entorno urbano de estas cuatro ciudades, y para completar el estudio se ha aplicado una muestra de una de las provincias orientales amazónicas.

Provincia	**Cantón**	**Parroquia**	**Localidad Geográfica**
01 = Azuay	01= Cuenca	50 = Cuenca. Cabecera Cantonal	Cuenca. (Zona Urbana)
07 = El Oro	01= Machala	50 = Machala. Cabecera Cantonal	Machala. (Zona Urbana)
09 = Guayas	01 = Guayaquil	50 = Guayaquil. Cabecera Cantonal	Guayaquil (Zona Urbana)
17 = Pichincha	01 = Quito	50 = Quito. Distrito Metropolitano	Quito (Zona Urbana)
21 = Sucumbìos	01 = Lago Agrio	50 = Nueva Loja. Cabecera Cantonal	Nueva Loja (Zona Urbana)

Tabla 4.1. Delimitación de los casos de ejemplo por localidad geográfica.

4.1.1.2 Descripción de los casos de ejemplo por variables investigadas y reglas de asociación.

Los casos de ejemplo se extraen de los datos anuales de edificaciones. Las afirmaciones que se hacen sobre las variables investigadas pueden ser expresadas como resultado de reglas de asociación.

De acuerdo a la técnica de algoritmos genéticos para realizar pronósticos por clasificación mediante extracción de reglas de asociación, se tiene que para los casos de ejemplo sobre los cuales se va a aplicar el sistema de predicción y análisis se desconocen los valores de las variables condición del antecedente de

las reglas, y se tiene como atributo objetivo en el lado del consecuente una de las variables cuyo valor es conocido. Por consiguiente, para los casos de ejemplo a partir de los cuales se obtendrán reglas de asociación, se han seleccionado algunos atributos objetivos y variables para la condición del antecedente, mostrados en la Tabla 4-2.

Caso de ejemplo	Variables seleccionadas para la condición del antecedente de la regla	Atributo objetivo para el consecuente de la regla
1	CIMIENTOS, CUBIERTA	ESTRUCTURA
1	TIPO DE OBRA, USO DE LA EDIFICACION	NUMERO DE VIVIENDAS
1	TIPO DE OBRA, USO DE LA EDIFICACION	ORIGEN DEL FINANCIAMIENTO

Tabla 4.2. Descripción de variables seleccionadas para los casos de ejemplo.

4.2 INTRODUCCIÓN DE DATOS AL SISTEMA Y PROCESAMIENTO.

4.2.1 INTRODUCCIÓN DE DATOS DE EDIFICACIONES Y SERIES HISTÓRICAS.

4.2.1.1 Introducción de datos de Localidad Geográfica y Tiempo.

Los datos de localidades geográficas, años y meses de proceso se introducen primero al momento de la instalación del sistema, por medio de la opción de importación de datos. Las localidades geográficas se hallan codificadas según el año base de 2002, con relación a la división política administrativa del Ecuador según e último censo de población y vivienda del año 2001.

Año base localidad	Nivel de localidad	Códigos de Localidad	Tipo de Localidad Ingresada	Número de registros ingresados
2002	0	000000	PAIS	1
2002	1	000000	REGION	0
2002	2	010000- 220000; 900000	PROVINCIA	23
2002	3	010100- 220400;	CANTON	223
2002	4	010150- 220455;	PARROQUIA	1014

Tabla 4.3. Resumen de localidades geográficas ingresadas al sistema

Nivel de tiempo	Códigos de Tiempo	Tipo de registro de tiempo	Número de registros ingresados
1	199100 – 200600	ANUAL	16
2	199101 – 200612	MENSUAL	192

Tabla 4.4. Resumen de registros de tiempo ingresados al sistema.

4.2.1.2 Introducción de variables de edificaciones.

La descripción de las variables de edificaciones principales correspondientes a materiales predominantes en cimientos, estructura, paredes y cubierta, tipo de obra, uso de la edificación y origen del financiamiento con sus categorías, se introducen al sistema por medio de la opción de importar datos en la ventana de mantenimiento de edificaciones desde el archivo VARIABLES.TX. Se especifican rangos máximos y mínimos, como datos informativos.

Códigos de clase	Códigos de categoría	Nivel de la variable	Número de registros ingresados
1-33	0	0 Clase	33
1-33	1-999	1 Categoría	143

Tabla 4.5. Resumen de registros de variables de edificaciones ingresadas al sistema.

4.2.1.3 Introducción de datos históricos de edificaciones.

Los datos históricos de edificaciones para los casos de estudio, corresponden a los años 2000, 2001,2003, 2004 y 2005. Estos datos son resultado de la encuesta de edificaciones que realiza el Instituto Nacional de Estadística y Censos en base a los permisos de construcción otorgados por los municipios del país. Son construcciones y viviendas proyectadas para esos años. Se introducen al sistema a través de la opción de importar de ventana de mantenimiento de edificaciones.

La importación se la hace tomando como origen los archivos de texto mediante el botón de importar de la sección edificaciones de la misma ventana.

Archivo origen	Tiempo (año-meses)	Localidad (provincia)	Número de registros ingresados
EED2000W.TXT	2000 - (01-12)	01-22	17543
EED2001W.TXT	2001- (01-12)	01-22	17144
EED2002W.TXT	2002 - (01-12)	01-22	16943
EED2003W.TXT	2003 - (01-12)	01-22	16785
EED2004W.TXT	2004 - (01-12)	01-22	17289
EED2005W.TXT	2005 - (01-12)	01-22	18824

Tabla 4.6. Resumen de registros de edificaciones ingresados al sistema.

4.2.1.4 Introducción de series de indicadores económicos para análisis

Las series de indicadores económicos son datos secundarios y se introducen al sistema a través de la opción de importación en la ventana de mantenimiento de indicadores económicos. Antes de introducir las series deben importarse los códigos y nombres de las variables de indicadores en la misma ventana.
Estos datos sirven únicamente para análisis. Permiten encontrar alguna relación de causalidad entre la variable de número de permisos de construcción con otras variables externas del ambiente socioeconómico del país.

Archivo origen	Tiempo (años)	Localidad (provincia)	Número de registros ingresados
INDICADOR.TXT	2000 - 2005	01-22	11
SERIE_INDICADOR.TXT	2000 - 2005	01-22	467

Tabla 4.7. Resumen de registros de series de indicadores ingresados al sistema.

4.2.2 PROCESAMIENTO DE CASOS DE PREDICCIÓN.

4.2.2.1 Introducción de casos de predicción.

Los casos de predicción se introducen al sistema una vez concluida la importación de archivos. Haciendo uso de la ventana de proceso de casos de predicción en el sistema, se introduce cada caso por medio del icono insertar.

Caso	Descripción
1. Cuenca	CASO:[001] AÑO: [2006] LOCALIDAD: [010150] PROV:AZUAY; CANTON:CUENCA; CAB CANT: CUENCA.
2. Machala	CASO:[002] AÑO: [2006] LOCALIDAD: [070150] PROV:EL ORO; CANTON: MACHALA; CAB CANT: MACHALA
3. Guayaquil	CASO:[003] AÑO: [2006] LOCALIDAD: [090150] PROV:GUAYAS; CANTON: GUAYAQUIL; CAB CANT: GUAYAQUIL
4. Quito	CASO:[004] AÑO: [2006] LOCALIDAD: [170150] PROV: PICHINCHA; CANTON:QUITO: CAB. CANT: QUITO
5. Nueva Loja	CASO:[005] AÑO: [2006] LOCALIDAD [210150] PROV: SUCUMBIOS; CANTON: NUEVA LOJA; CAB.CANT: NUEVA LOJA

Tabla 4.8. Descripción de casos de prueba o predicción.

4.2.2.2 Introducción de parámetros del algoritmo genético.

En la ventana de proceso de casos de predicción, se abre la sección de parámetros del algoritmo genético. Los parámetros que se actualizan por cada caso son: porcentaje de cruce, porcentaje de mutación, número máximo de generaciones, número máximo de corridas e índice de convergencia. Por cada sección del “Cromosoma-regla”, se actualizan únicamente los puntos de corte

para el cruce y la mutación, ya que en el proceso de instalación se crean e inicializan las variables que formarán parte de las reglas.

No.	Parámetro	Aplicación
1	Porcentaje de cruce: 40 %	Todos los casos
2	Porcentaje de mutación: 0.15 %	Todos los casos
3	Número de generaciones: 7	Todos los casos
4	Número de corridas: 1	Todos los casos
5	Adaptabilidad de convergencia: 75 %	Todos los casos

Tabla 4.9. Parámetros del algoritmo genético ingresados para los casos de estudio.

De acuerdo a las pruebas realizadas, se descubrió que el número de generaciones necesario para alcanzar la convergencia del 75% es alrededor de cinco, en una corrida. Por otro lado, si se aumenta el porcentaje de convergencia, el número de reglas de asociación descubiertas se aproximará a uno, puesto que siempre se tiende a llegar a un sólo individuo más apto.

Los puntos de corte para el cruce y para la mutación se establecen al azar durante la aplicación del algoritmo.

4.2.2.3 Introducción de tipos de reglas de asociación y parámetros de ejecución.

La estructura de las reglas de asociación se ingresan al sistema junto a los parámetros de ejecución del algoritmo. Se eligen variables para el antecedente, de las cuales se ignoran sus valores y sólo se conocen los valores del consecuente. Este último permanece fijo durante la ejecución, según lo que establece la técnica del algoritmo descrita en el segundo capítulo. Unos pocos criterios para elegir las variables del antecedente y el consecuente se obtuvieron del análisis realizado en el capitulo primero.

Tipo Regla	Antecedente desconocido	Consecuente objetivo
1	**SI** CIMIENTOS = **?** Y CUBIERTA = **?**	**ENTONCES**, ESTRUCTURA = Hormigón Armado
2	**SI** TIPO DE OBRA = **?** Y USO DE LA EDIFICACION = **?**	**ENTONCES**, DISTRIBUCIÓN DE VIVIENDAS = 1 Vivienda
3	**SI** TIPO DE OBRA = **?** Y USO DE LA EDIFICACION = **?**	**ENTONCES**, ORIGEN DEL FINANCIAMIENTO= Recursos Propios

Tabla 4.10. Descripción de estructura de las reglas de asociación a descubrir para el caso de estudio.

Entre los parámetros de ejecución, se encuentran el número de individuos inicial del algoritmo y el tamaño de la muestra de ejemplos, necesaria para el cálculo de adaptabilidad de los “cromosoma-individuo” o reglas a descubrir.

No.	Parámetro	Valor
1	Número de individuos inicial	500 genotipos
2	Tamaño de la muestra de ejemplos	100 casos

Tabla 4.11. Parámetros de ejecución del algoritmo genético aplicado a los casos de prueba.

Para cada tipo de regla de la Tabla 4-10, se pueden llegar a descubrir más de una regla de asociación, dependiendo de los valores de adaptabilidad que se alcance para los atributos del antecedente, al final del proceso.

4.2.2.4 Proceso de ejecución del algoritmo genético para descubrimiento de reglas de asociación para casos de predicción.

El proceso de ejecución del algoritmo en el sistema se inicia con la selección del caso en la sección principal de la ventana de mantenimiento de casos, luego se siguen los pasos del algoritmo y que consisten en activar los botones de inicio, seguir, terminar y guardar resultados, consecuentemente.

Las reglas de asociación descubiertas por la aplicación del algoritmo genético para los casos de prueba ingresados se muestran a continuación.

Caso de estudio	Tipo Regla	Corrida	Generación alcanzada	Convergencia Alcanzada	Número de reglas descubiertas.
1-1. Cuenca	1	1	7	82.80	6
1-2. Machala	1	1	7	86.60	3
1-3. Guayaquil	1	1	7	89.00	5
1-4. Quito	1	1	7	86.60	4
1-5. Nueva Loja	1	1	7	89.00	7
2-1. Cuenca	2	1	7	83.60	3
2-2. Machala	2	1	8	90.80	3
2-3. Guayaquil	2	1	7	77.40	4
2-4. Quito	2	1	9	88.40	2
2-5. Nueva Loja	2	1	8	78.40	3
2-1. Cuenca	3	1	7	91.60	9
2-2. Machala	3	1	8	80.40	5
2-3. Guayaquil	3	1	7	94.40	3
2-4. Quito	3	1	7	83.60	5
2-5. Nueva Loja	3	1	7	86.20	5

Tabla 4.12 Resultados de la ejecución del algoritmo genético para los casos de predicción.

4.3 ANALISIS DE RESULTADOS.

4.3.1 PRUEBAS DE CONTRASTE DE HIPÓTESIS APLICADO A LAS REGLAS DE ASOCIACIÓN DESCUBIERTAS.

4.3.1.1 Descripción de la metodología de pruebas de contraste de hipótesis.

A partir de los datos contenidos en una muestra de ejemplos, se puede decidir si un parámetro de la población es aceptable o no. Para el caso de una regla de asociación descubierta, los parámetros de cobertura o completitud y la confianza que la regla obtuvo para la muestra de ejemplos durante la ejecución del algoritmo genético, pueden ser evaluados mediante contraste de hipótesis aplicado a proporciones.

> "Una hipótesis estadística es una afirmación respecto a alguna característica de una población. Contrastar una hipótesis es comparar las predicciones con la realidad observada. Las pruebas son paramétricas si se hace afirmaciones sobre valores de parámetros desconocidos. Para aceptar o rechazar la hipótesis se establece un margen de error o región de aceptación para los valores que pueda tomar un parámetro de la población" ([81]).

La hipótesis emitida se designa por $\mathbf{H_o}$ y se llama hipótesis nula, porque parte del supuesto que la diferencia entre el valor verdadero del parámetro y su valor hipotético es debido al azar, es decir no hay diferencia. La hipótesis contraria se designa por $\mathbf{H_1}$ y se llama hipótesis alternativa.

El contraste puede ser unilateral o bilateral, también llamado de una o dos colas, según se establezca la hipótesis. Se rechaza la hipótesis nula Ho, si el valor absoluto del estadístico entre proporciones calculado Z, tiene un valor: $|Z| >= Z_a$ para pruebas de una cola. y $|Z| >= Z_{a/2}$, para pruebas de dos colas, dónde 'a' es el nivel de significación de la prueba.

Los pasos para aplicar contraste unilateral de hipótesis acerca de la proporción en una población a partir de los datos extraídos de una muestra son:

---

[81] GARCÍA CEBRIAN, María José., **Contraste de Hipótesis**., **op.cit**.

1. Enunciar la hipótesis.

   Para comparar dos proporciones obtenidas en dos muestras diferentes, se expresa la hipótesis nula como Ho: $\Theta = \Theta_o$; y las hipótesis alternativas pueden ser H1: $\Theta < \Theta_o$; H1: $\Theta > \Theta_o$; ó bien, H1: $\Theta \neq \Theta_o$.

   Dónde Θ es la verdadera proporción de éxitos de una serie de ensayos.

   Para muestras grandes, la proporción muestral se define como P = X / n, dónde X representa el número de éxitos en n ensayos de Bernoulli, la cual tiende a seguir una distribución normal estándar (0,1).

   Se seleccionan dos muestras n1 y n2, en las cuales se definen estimadores independientes P1 y P2 para las proporciones Θ_1 y Θ_2 desconocidas.

   P1 = X1 / n1 : y P2 = X2 / n2. (**Fórmula 4.1** Estimador de Proporción)

2. Elegir un nivel de significación α y construir una zona de aceptación.

   El estadístico para comparación de proporciones es Z . y se expresa así:

$$Z = \frac{P_1 - P_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}}} \approx N(0,1)$$

   (**Fórmula 4.2** Estadístico Z)

   Para hallar Z se remplaza el promedio ponderado entre las proporciones:

$$\theta_1 = \theta_2 = \theta = \frac{x_1 + x_2}{n_1 + n_2}$$

   (**Fórmula 4.3** Promedio ponderado entre proporciones)

   Para un α existe un Z_α , tal que $P(|Z| < Z_\alpha) = 1 - \alpha$

   La zona de aceptación para este caso, se define como:

Figura 3.38 Zona de aceptación para contraste unilateral de hipótesis

3. Calcular Z, el cual es un indicador de comparación de las proporciones obtenidas en las dos muestras.
4. Decidir. Si el valor calculado cae dentro de la zona de aceptación se acepta la hipótesis y si no, se rechaza.

Hipótesis Nula	Hipótesis Alternativa	Rechaza sí:
$Ho: \theta_1 = \theta_2$	$H_1: \theta_1 \neq \theta_2$	$\lvert Z \rvert \geq Z_{\alpha/2}$
$Ho: \theta_1 = \theta_2$	$H_1: \theta_1 > \theta_2$	$Z \geq Z_{\alpha}$
$Ho: \theta_1 = \theta_2$	$H_1: \theta_1 < \theta_2$	$\lvert Z \rvert \geq Z_{\alpha}$

Tabla 4.13. **Aceptación de hipótesis para comparar dos proporciones.**

Fuera del intervalo sólo se encuentran los casos más raros o atípicos. La zona de rechazo se llama región crítica y su área es el nivel de significación. El contraste de hipótesis no establece la verdad de la hipótesis, sino un criterio que permite determinar si las muestras observadas difieren significativamente de los resultados esperados. El nivel de significación se fija de un 90 a 95%.

En el proceso se producen dos tipos de errores según sea la situación real y la decisión tomada: El error del tipo I es aquel que se comete al rechazar la hipótesis nula siendo cierta, y el error del tipo II es el que se comete al aceptar la hipótesis nula siendo falsa. El tamaño, potencia o nivel del contraste es la probabilidad de cometer el error de tipo I, y está dado por α = P(el). ([82])([83])

4.3.1.2 Aplicación de las pruebas de contraste de hipótesis a las reglas de asociación descubiertas.

El objetivo de aplicar pruebas de contraste de hipótesis a las reglas de asociación descubiertas por el algoritmo genético, es encontrar una región de aceptación para los valores que puedan tomar los parámetros de cobertura y confianza en una nueva muestra diferente, y así determinar un nivel de certeza o medida de aceptación de la regla para el futuro.

4.3.2 RESULTADOS DE LAS PRUEBAS DE CONTRASTE DE HIPÓTESIS.

4.3.2.1 Opciones de contraste de hipótesis aplicadas en el sistema.

Las opciones para aplicar contraste de hipótesis a las reglas descubiertas mediante el algoritmo genético tienen la finalidad de elevar el nivel de significación y tener la posibilidad de realizar la prueba con dos alternativas: bien sea con una

[82] CALDERON, C. Bernardo A., **Diferencia entre proporciones** ., **op.cit**.
[83] DE LA TORRE, Leticia., **Curso de Estadística., op.cit**

muestra variable de casos de ejemplo seleccionada al azar o con toda la población de casos de ejemplo.

Parámetro de prueba	Coeficiente de significación	Tamaño de la muestra de casos de ejemplo
.p = confianza	.a = 0.10 Z = 1.28	1000 casos
.p = confianza	.a = 0.05 Z = 1.645	1000 casos
.p = confianza	.a = 0.01 Z = 2.33	1000 casos
.p = confianza	.a = 0.005 Z = 2.58	1000 casos
.p = confianza	.a = 0.002 Z = 2.88	1000 casos

Tabla 4.14. Parámetros de contraste de hipótesis aplicados a las reglas descubiertas.

4.3.2.2 Obtención de resultados de las pruebas de contraste de hipótesis.

Pruebas de contraste de hipótesis aplicadas a reglas del tipo 1.

Para las reglas descubiertas del tipo 1:

"Si CIMI = ? Y CUBI = ?, ENTONCES ESTRU = Horm.Arm", se obtuvieron los siguientes resultados:

Para construcciones de estructura de hormigón armado, predomina el hormigón armado en cimientos, sin embargo la cubierta o techo utilizado varía según la localidad geográfica del caso de prueba. (Ver Tabla 4.15)

Las reglas de uso del hormigón armado en cubierta tienen una zona de aceptación bastante alta. En Quito se tiene mayor tendencia a construir con cimientos de hormigón ciclópeo (piedra). De las pruebas aplicadas, se deduce que hay una confianza promedio del 98 % con un 90% de significación a usar hormigón armado en cubierta, con cimientos y estructura del mismo material predominante, tanto en Quito como en Guayaquil.

Para el mismo tipo de regla anterior, se deduce que mientras en Quito, Guayaquil y Cuenca ya no se encuentran reglas de casos significativos con zinc en cubierta o techo como material predominante, en Machala y Nueva Loja la tendencia a utilizar el zinc en cubierta es mayoritaria, con 96.02% y 84.00 % de confianza, respectivamente. También se puede confirmar la hipótesis de que en Cuenca se tiene una aceptación de confianza mayor al 90% en el uso de la teja en cubierta con hormigón armado en estructura y cimientos.

Caso de prueba	Regla descubierta: Tipo de Regla: 1	Parámetro Confianza Algoritmo	Parametro Confianza Muestra	Aceptación (a = 0.005) | Z | < Za
1.Cuenca	SI CIMI = Horm. Ciclopeo Y CUBI = Eternit-Ardex o Asbesto ENTONCES ESTRU = Horm.Arm	94.76	95.78	SE ACEPTA 0.93 < 2.58
	SI CIMI = Horm. Arm. Y CUBI = Teja, ENTONCES ESTRU = Horm.Arm.	94.38	93.97	SE ACEPTA 0.16 < 2.58
	SI CIMI = Horm. Arm Y CUBI = Horm. Arm, ENTONCES ESTRU = Horm.Arm.	100.00	95.67	SE ACEPTA 1.12 < 2.58
	SI CIMI = Horm. Ciclópeo y CUBI = Teja, ENTONCES ESTRU = Horm.Arm	91.58	93.88	SE ACEPTA 0.88 < 2.58
	SI CIMI = Horm. Arm. Y CUBI = Eternit-Ardex o Asbesto ENTONCES ESTRU = Horm.Arm	88.39	90.25	SE ACEPTA 0.97 < 2.58
	SI CIMI = Horm. Ciclopeo y CUBI = Horm.Arm ENTONCES ESTRU = Horm.Arm.	93.55	96.55	SE ACEPTA 0.81 < 2.58
2.Machala	SI CIMI = Horm. Arm Y CUBI = Eternit-Ardex o Asbesto ENTONCES ESTRU = Horm.Arm.	98.20	97.94	SE ACEPTA 0.31 < 2.58
	SI CIMI = Horm. Arm Y CUBI = Zinc ENTONCES ESTRU = Horm.Arm	96.62	96.02	SE ACEPTA 0.43 < 2.58
	SI CIMI = Horm. Arm Y CUBI = Horm.Arm ENTONCES ESTRU = Horm.Arm	99.48	99.28	SE ACEPTA 0.27 < 2.58
3.Guayaquil	SI CIMI = Horm.Arm Y CUBI = Horm.Arm. ENTONCES ESTRU = Horm.Arm	100.00	99.67	SE ACEPTA 1.01 < 2.58
	SI CIMI = Horm.Arm. Y CUBI = Eternit-Ardex o AsbestoENTONCES ESTRU = Horm.Arm.	99.37	98.17	SE ACEPTA 1.57 < 2.58
	SI CIMI = Horm.Arm. Y CUBI = Teja ENTONCES ESTRU = Horm.Arm.	100.00	99.75	SE ACEPTA 0.26 < 2.58
4.Quito	SI CIMI = Horm.Arm Y CUBI = Horm.Arm ENTONCES ESTRU = Horm.Arm.	97.59	97.41	SE ACEPTA 0.27 < 2.58
	SI CIMI = Horm.Ciclópeo Y CUBI = Horm.Arm. ENTONCES ESTRU = Horm.Arm	97.73	98.47	SE ACEPTA 0.48 < 2.58
	SI CIMI = Horm.Arm. Y CUBI = Teja, ENTONCES ESTRU = Horm.Arm.	96.77	94.07	SE ACEPTA 0.60 < 2.58
	SI CIMI = Horm.Arm Y CUBI = Eternit-Ardex o Asbesto ENTONCES ESTRU = Horm.Arm.	61.54	59.34	SE ACEPTA 0.29 < 2.58
5.Nueva Loja	SI CIMI = Horm.Arm Y CUBI = Zinc, ENTONCES ESTRU = Horm.Arm	84.00	84.00	SE ACEPTA 0.01 < 2.58
	SI CIMI = Horm.Ciclopeo y CUBI = Eternit-Ardex o Asbesto ENTONCES ESTRU = Horm.Arm.	72.31	72.31	SE ACEPTA 0.03 < 2.58
	SI CIMI = Horm.Ciclópeo y CUBI = Horm.Arm ENTONCES ESTRU = Horm.Arm.	100.00	100.00	SE ACEPTA 0.00 < 2.58
	SI CIMI = Horm.Ciclopeo y CUBI = Zinc ENTONCES ESTRU = Horm.Arm.	100.00	100.00	SE ACEPTA 0.00 < 2.58

Tabla 4.15. Resultados de aplicación de contraste de hipótesis a reglas descubiertas del tipo 1.

Pruebas de contraste de hipótesis aplicadas a reglas del tipo 2.

Para las reglas descubiertas del tipo 2:

"Si TIPO DE OBRA = ? Y USO DE LA EDIFICACIÓN = ?, ENTONCES DISTRIBUICION DE VIVIENDAS = 1 ", se obtuvieron los siguientes resultados:

En Quito, las reglas del tipo 2 descubiertas son apenas dos, de las cuales la que tiene mayor aceptación es la que expresa que si se tiene nuevas construcciones para una familia, entonces hay una confianza mayor al 71.17 % de encontrar construcciones para una vivienda.

Caso de prueba	Regla descubierta:Tipo de Regla:2	Parámetro Confianza Algoritmo	Parametro Confianza Muestra	Aceptación (a = 0.005) | Z | < Za
1.Cuenca	SI TIPO DE OBRA = Nueva const. Y USO DE LA EDI F= 1 Familia,ENTONCES DIST.VIVIENDAS = 1	99.57	99.67	SE ACEPTA 0.39 < 2.58
	SI TIPO DE OBRA = Nueva const. Y USO DE LA EDIF= Mixta ENTONCES DIST.VIVIENDAS = 1	56.00	53.62	SE ACEPTA 0.23 < 2.58
	SI TIPO DE OBRA = Ampliación Y USO DE LA EDIF = 1 Familia ENTONCES DIST.VIVIENDAS = 1	100.00	100.00	SE ACEPTA 0.00 < 2.58
2.Machala	SI TIPO DE OBRA = Nueva const. Y USO DE LA EDI F= 1 Familia ENTONCES DIST.VIVIENDAS = 1	99.98	99.92	SE ACEPTA 0.29 < 2.58
	SI TIPO DE OBRA = Ampliación Y USO DE LA EDIF = 1 Familia ENTONCES DIST.VIVIENDAS = 1	100.00	100.00	SE ACEPTA 0.00 < 2.58
	SI TIPO DE OBRA = Reconstrucción Y USO DE LA EDIF= 1 Familia ENTONCES DIST.VIVIENDAS = 1	100.00	100.00	SE ACEPTA 0.00 < 2.58
3.Guayaquil	SI TIPO DE OBRA = Nueva const. Y USO DE LA EDI F= 1 Familia ENTONCES DIST.VIVIENDAS = 1	91.04	93.55	SE ACEPTA 1.81 < 2.58
	SI TIPO DE OBRA = Ampliación Y USO DE LA EDIF = 1 FamiliaENTONCES DIST.VIVIENDAS = 1	99.47	99.95	SE ACEPTA 2.09 < 2.58
	SI TIPO DE OBRA = Nueva const. Y USO DE LA EDIF= MixtaENTONCES DIST.VIVIENDAS = 1	57.63	52.54	SE ACEPTA 0.75 < 2.58
	SI TIPO DE OBRA = Reconstrucción Y USO DE LA EDIF = 1 Familia ENTONCES DIST.VIVIENDAS = 1	100.00	100.00	SE ACEPTA 0.00 < 2.58
4.Quito	SI TIPO DE OBRA = Ampliación Y USO DE LA EDIF = 1 Familia ENTONCES DIST.VIVIENDAS = 1	100.00	99.35	SE ACEPTA 0.54 < 2.58
	SI TIPO DE OBRA = Nueva const. Y USO DE LA EDI F= 1 Familia ENTONCES DIST.VIVIENDAS = 1	74.36	73.57	SE ACEPTA 0.28 < 2.58
5.Nueva Loja	SI TIPO DE OBRA = Ampliación Y USO DE LA EDIF = 1 Familia ENTONCES DIST.VIVIENDAS = 1	100	100.00	SE ACEPTA 0.00 < 2.58
	SI TIPO DE OBRA = Nueva const. Y USO DE LA EDIF= Mixta ENTONCES DIST.VIVIENDAS = 1	83.33	77.78	SE ACEPTA 0.29 < 2.58
	SI TIPO DE OBRA = Nueva const. Y USO DE LA EDI F= 1 Familia ENTONCES DIST.VIVIENDAS = 1	97.87	99.10	SE ACEPTA 0.72 < 2.58

Tabla 4.16 Resultados de aplicación de contraste de hipótesis a reglas descubiertas del tipo 2.

En Guayaquil, la regla de mayor aceptación corresponde a reconstrucción de viviendas para una familia. Esta regla alcanza una confianza del 100% en todas las pruebas. Lo que lleva a deducir que en esta ciudad tiende a tener más reconstrucciones que nuevas construcciones.

Por otra pare se observa que, en otras ciudades como Machala en la costa y Nueva Loja en la región amazónica, la regla de mayor aceptación es la ampliación de viviendas para una familia, seguida por nuevas construcciones de viviendas también para una familia.

Pruebas de contraste de hipótesis aplicadas a reglas del tipo 3.

En cuanto a las reglas descubiertas de mayor aceptación para tipo 3:

"Si TIPO DE OBRA= ? Y USO EDIFICACIÓN= ?, ENTONCES ORIGEN DEL FINANCIAMIENTO=Recursos Propios ", se obtuvieron los siguientes resultados:

Caso de prueba	Regla descubierta: Tipo de Regla: 3	Parámetro Confianza Algoritmo	Parametro Confianza Muestra	Aceptación (a = 0.005) | Z | < Za
1.Cuenca	SI TIPO DE OBRA = Reconstrucción Y USO DE LA EDI F= Una Familia, ENTONCES ORIGEN DEL FINANCIAMIENTO= Recursos Propios	100.00	93.48	SE ACEPTA 0.46 < 2.58
	SI TIPO DE OBRA = Ampliación Y USO DE LA EDIF= Una Familia ENTONCES ORIGEN DEL FINANCIAMIENTO= Recursos Propios	94.74	94.35	SE ACEPTA 0.12 < 2.58
	SI TIPO DE OBRA = Nueva Const. Y USO DE LA EDIF= 3 o más Familias ENTONCES ORIGEN DEL FINANCIAMIENTO= Recursos Propios	92.31	94.27	SE ACEPTA 0.50 < 2.58
2.Machala	SI TIPO DE OBRA = Ampliación Y USO DE LA EDIF= Dos Familias ENTONCES ORIGEN DEL FINANCIAMIENTO= Recursos Propios	75.00	84.62	SE ACEPTA 0.54 < 2.58
	SI TIPO DE OBRA = Nueva Const. Y USO DE LA EDIF= 2 Familias ENTONCES ORIGEN DEL FINANCIAMIENTO= Recursos Propio	97.30	95.24	SE ACEPTA 0.51 < 2.58
3.Guayaquil	SI TIPO DE OBRA = Nueva Const. Y USO DE LA EDIF= Una Familia ENTONCES ORIGEN DEL FINANCIAMIENTO= Recursos Propios	90.29	90.41	SE ACEPTA 0.07 < 2.58
	SI TIPO DE OBRA = Reconstrucción Y USO DE LA EDIF= Una Familia ENTONCES ORIGEN DEL FINANCIAMIENTO= Recursos Propios	96.55	99.68	SE ACEPTA 2.09 < 2.58
	SI TIPO DE OBRA = Ampliación Y USO DE LA EDIF= Una Familia ENTONCES ORIGEN DEL FINANCIAMIENTO= Recursos Propios	100	99.90	SE ACEPTA 0.44 < 2.58
4.Quito	SI TIPO DE OBRA = Ampliación Y USO DE LA EDIF= Comercial ENTONCES ORIGEN DEL FINANCIAMIENTO= Recursos Propios	100	90.32	SE ACEPTA 0.33 < 2.58
	SI TIPO DE OBRA = Nueva Const. Y USO DE LA EDIF= 3 o más Familias ENTONCES ORIGEN DEL FINANCIAMIENTO= Recursos Propios	78.57	78.57	SE ACEPTA 0.07 < 2.58
	SI TIPO DE OBRA = Nueva Const. Y USO DE LA EDIF= 2 Familias ENTONCES ORIGEN DEL FINANCIAMIENTO= Recursos Propio	78.42	79.96	SE ACEPTA 0.13 < 2.58
5.Nueva Loja	SI TIPO DE OBRA = Nueva Const. Y USO DE LA EDIF= Una Familia ENTONCES ORIGEN DEL FINANCIAMIENTO= Recursos Propios	85.11	86.43	SE ACEPTA 0.24 < 2.58
	SI TIPO DE OBRA = Nueva Const. Y USO DE LA EDIF= 3 o más Familias ENTONCES ORIGEN DEL FINANCIAMIENTO= Recursos Propio	80.00	83.33	SE ACEPTA 0.17 < 2.58

Tabla 4.17. Resultados de aplicación de contraste de hipótesis a reglas descubiertas del tipo 3.

La regla de mayor aceptación es la de financiamiento con recursos propios en ampliación para una y dos familias, seguida por nuevas construcciones para una familia. Solo en Guayaquil, la regla de reconstrucciones con financiamiento por recursos propios para una familia tiene una confianza superior al 99 %, por lo que se la podría considerar aceptable. De las otras ciudades, aparte de Quito y Guayaquil, se deduce que se prefiere a las nuevas construcciones de hasta tres o más familias antes que las ampliaciones y las reconstrucciones.

CAPÍTULO 5.
CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

En este trabajo se ha demostrado que la técnica de algoritmos genéticos puede usarse como método aleatorio de resolución de problemas de predicción y clasificación, a diferencia de otras técnicas de inteligencia artificial. Específicamente, para descubrimiento de reglas, el resultado que se obtiene al aplicar este método depende más del número de intentos y de la selección acertada de muestras de ejemplos que el alcanzar un óptimo, como se exige en otras aplicaciones del mismo algoritmo. El éxito del método es superar la convergencia mínima con un numero considerable de reglas descubiertas y cuyo parámetro de confianza se aproxime a valores óptimos.

En cuanto a la metodología de desarrollo de software utilizada: Orientación a Objetos (OMT / Booch), se pudo comprobar lo siguiente:

- Los diagramas del Modelo Unificado en UML usados, especialmente la representación del problema y los requerimientos en casos de uso, se aproximan bastante al lenguaje natural y permiten organizar acciones concretas en escenarios, que son guias para la construcción del software.

- La complejidad del modelo de datos no es un obstáculo, ya que la separación en paquetes y subsistemas permite ir refinándolo a medida que la aplicación crezca con nuevos requerimientos y cambios.

- El modelo funcional, para ciertos proyectos donde se aplican algoritmos de mayor complejidad, requieren de diagramas de flujo tradicionales para complementar la explicación de funcionalidad.

En cuanto a los resultados de aplicar algoritmos genéticos como técnica de inteligencia artificial para explicar y predecir el comportamiento de la construcción de viviendas en el país, se puede apreciar lo siguiente:

- Inicialmente se ejecutó el algoritmo para alcanzar la convergencia del 75% con valores de adaptabilidad máximos y con un máximo de 5 generaciones. Sin embargo, para la mayoría de casos se tuvo que aumentar el número de generaciones hasta 10 para superar la convergencia indicada. El número de individuos inicial para todos los casos se mantuvo en 100, y posibilitó discriminar adecuadamente casi en su totalidad las clases (reglas posibles).

- Para tipos de reglas a descubrir, en los cuales hay mayor número de características, siendo 7 el máximo número de variables para el antecedente de la regla, se tuvo que aumentar el porcentaje de mutación de 0.10% hasta más allá del 1% , de manera que en la siguientes generaciones aparecían nuevas reglas no detectadas en las generaciones de corridas anteriores.

- Al realizar las pruebas de contraste de hipótesis para validar el parámetro de confianza de las reglas descubiertas por el algoritmo, se necesitó aumentar el tamaño de la muestra de ejemplos de 100 a 1000 casos, para casos de estudio con un total de casos superior a 1500. De esta manera los rangos de aceptación de la hipótesis eran más estables y colocaban los valores de confianza más allá del 50%.

- Fue acertado el seleccionar muestras de ejemplos de acuerdo a la localidad geográfica, justamente en regiones que caracterizan al desarrollo urbano del país (según estudio del capítulo primero) y específicamente en ciudades de mayor crecimiento como Quito, Guayaquil, Cuenca, Machala y una ciudad típica de la amazonía como es Nueva Loja, ya que las reglas descubiertas variaban de una región a otra, incluso en ciudades tan cercanas como Guayaquil y Machala.

- Se pudo contar además con todo el conjunto de datos de casos de las encuestas de edificaciones disponible de los últimos cinco años, lo cual permite generalizar los resultados para la coyuntura actual que vive el país (posterior a la crisis económica de fines del siglo anterior y del fenómeno de la dolarización).

- Para medir la efectividad de la predicción, esto es, evaluar una regla de asociación descubierta por la técnica de algoritmos genéticos, es preciso comparar resultados estimados (aplicando muestras aleatorias) con resultados actuales (aplicando todos los ejemplos). Es así que, usando el parámetro de confianza de cada regla para el contraste de hipótesis, se llegó a determinar si existe evidencia significativa de que la proporción de confianza es o no la habitual, y por lo tanto puede o no predecir el comportamiento de las variables.

Entre otros aspectos, se puede concluir lo siguiente:

- Las técnicas de inteligencia artificial realmente permiten construir soluciones informáticas que integren en un mismo ambiente, representación de datos, predicción y evaluación de resultados para definir y explicar el comportamiento de un determinado dominio del problema.

- Los lenguajes de programación visuales y orientados a objetos (PowerBuilder) con un buen soporte de base de datos (SQL Server), en verdad proporcionan un ambiente de desarrollo para construir proyectos complejos donde se puede explorar y analizar grandes volúmenes de datos, se pueden implementar algoritmos para una o varias técnicas aprovechando el rendimiento del hardware al máximo y además evaluar resultados en forma gráfica, organizada, eficiente y amigable al usuario

- El estudio de una amplia variedad de técnicas de inteligencia artificial aplicadas a la solución de problemas de clasificación y predicción, ha

permitido identificar diversas formas de representación del problema, de las cuales, las que hacen una simulación con fenómenos presentes en la naturaleza tienen mejores posibilidades de aplicación práctica. Tanto las redes neuronales artificiales como la computación evolutiva exploran mejor los espacios de búsqueda que otras técnicas.

- Los experimentos hipotéticos permiten probar nuevas ideas, y también la efectividad de la aplicación de una determinada técnica. Las reglas descubiertas por algoritmos genéticos necesitan de una técnica capaz de determinar si éstas tienen validez para todos los casos presentes y probablemente futuros. Puesto que el parámetro de confianza de las reglas es una variable contínua, entonces es factible aplicar la estadística de estimación de parámetros, intervalos de hipótesis y pruebas de contraste para evaluar su aceptación presente y futura.

5.2 RECOMENDACIONES

- Uno de los factores que determinó el tiempo empleado en este trabajo fue estudiar en detalle el fenómeno, no solamente el comportamiento de las variables investigadas propuestas sino también todos los posibles factores externos que las explican. En éste como en futuros proyectos sería recomendable dar más importancia a la observación, ya que en metodología de la investigación es fundamental para ampliar y consolidar los planteamientos e hipótesis. Esto conduce a construir una representación clara y certera del conocimiento, y a través de un sistema inteligente asistido por una herramienta informática, se consigue llegar al logro de los objetivos y a la comprobación y generalización de los resultados.

- Se recomienda además, en lo posible, cambiar las técnicas y herramientas y la metodología que generalmente se emplean para la resolución de un problema, ya que esto permite también cambiar la forma de pensar sobre él. Usar algo diferente puede ocasionar que se descubran nuevas y

mejores soluciones. En el desarrollo del software, es recomendable no someter el problema a la herramienta, sino usar las herramientas tan creativamente como sea factible. También es recomendable tratar de evitar en lo posible el usar una misma herramienta para todos los problemas que se presenten.

- De acuerdo a las experiencias de este proyecto, en el proceso de construcción del software la creatividad e independencia de código pueden ser fundamentales para obtener un resultado acorde a los requerimientos. Mientras se tengan más clases, objetos y código especialmente diseñados para el problema y si se evita en lo posible el uso de componentes desarrollados para otras aplicaciones, se podrá adecuar mejor la herramienta a la solución y no lo contrario.

- Se recomienda que se piense en futuros proyectos informáticos similares, donde se estudien fenómenos socioeconómicos involucrados con la realidad nacional, ya que esto permite comprometer la ciencia y la tecnología con el progreso del país. Aunque dichos fenómenos parezcan a primera vista complejos y difíciles de comprender, siempre hay posibilidades y estrategias de búsqueda de soluciones, puesto que se tiene la enorme ventaja de contar con herramientas informáticas cada vez más avanzadas, especialmente aquellas que ofrecen soluciones inteligentes.

BIBLIOGRAFIA.

LIBROS.

1. ANDERSON D., SWEENEY D., ***"Estadística para Administración y Economía"***. , 7ª Ed., Thomson Editores, México 1999.

2. BETANZOS A., GUIJARRO B., LOZANO A.*,"* ***Ingeniería del conocimiento aspectos metodológicos"***., 1ra ed. , Edit. Pearson Educación, S.A., Madrid, 2004.

3. BOOCH, G., JACOBSON, I., RUMBAUGH, J., "***El Proceso Unificado de Desarrollo de Software"***., Edit. Pearson-Education, Madrid 2000.

4. CARRION Fernando., *"**Evolución del espacio urbano ecuatoriano; El Proceso de Urbanización en el Ecuador (del Siglo XVIII al Siglo XX)"*.** 1ra Ed., Edit. Concejo, Quito, 1986;

5. CARRION Fernando., *"**El proceso urbano en el Ecuador**"*, Edit. ILDIS., Quito, 1987.

6. DEL BRIO, M. SANZ MOLINA, A., "***Redes Neuronales y sistemas difusos"***, 2da ed. Edit.Alfaomega México D.F. 2002.

7. DOBSON Rick, "***Programación de Microsoft SQL Server 2000 con Microsoft Visual Basic.NET"***., 1ra. Ed., Edit Mc.Graw Hill Interamericana, Fareso, España 2002.

8. GOLDBERG, D.E., "***Messy genetic algorithms: Motivation, analysis, and first result., Complex System"***, vol 3. pp. 493-530, Michigan 1989.

9. HERNANDEZ J, RAMIREZ J, FERRI C. "***Introducción a la Minería de Datos".***, 1ra Ed. Pearson Educación S.A., Madrid 2004

10. MAKRIDAKIS - WHEELWRIGHT, **"*Métodos de Pronósticos*"**, Edit Limusa, México 2000.

11. PEREZ BELLO, C , **"*Pronóstico Empresarial*"**, 1ra Ed, Edit. Ecoe Ediciones, Bogotá, 2000.

12. PRESSMAN, R.*,"* ***Ingeniería del Software"***., 5ta. Ed., Edit. Mc-GrawHill, Madrid 2002.

13. RICH E, KNIGHT K., "***Inteligencia Artificial"***., 2da Ed., Mc-Graw Hill, Madrid1994.

14. RUMELHART D.E, McCLELLAND J.L, "***Parallel Distributed Processing"***, Vol 1: Foundations. Edit. MIT Press, Chicago 1986.

15. RUSSELL, S., NORVIG P., "***Inteligencia Artificial – Un enfoque moderno"*,** 2da Ed.., Edit Pearson – Prentice Hall., Madrid 2004.

16. SCHILDT, H., "***Utilización de C en Inteligencia Artificial"***., 1ra.Ed, Edit. McGraw Hill/Interamericana, S.A., Madrid, 1991

17. VASQUEZ. L, SALTOS, N., ***"Ecuador, su realidad 2005-2006"***, 13ed., Edit. Artes Gráficas Silva., Quito, 2005.

18. WEBSTER ALLEN L., "***Estadística aplicada a los negocios y la economía."***, 3ra Ed., Edit Mc.Graw-Hill, Bogotá, 2001.

19. WINCHESTER A.M., "***Genetics – A survey of the principles of heredity"***, 5ta Ed., Edit. Library of Congress, Denver 1972.

PUBLICACIONES.

1 INEC*., "**Encuesta Anual de Edificaciones***"*., Talleres Gráficos INEC, Quito, Publicaciones de años 1999-2005.

2 INEC., *"**Resultados del Censo de Población y Vivienda; Anuario Estadístico 2003"***., Talleres Gráficos INEC, Quito 2004.

3 INEC., *"**Encuesta de Ingresos y Gastos de los Hogares Urbanos (ENIGHU)"***, Talleres Gráficos INEC, Quito 2004

4 INEC., *"**Encuesta de Condiciones de Vida (ECV)"*,** Talleres Gráficos INEC, Quito 2004.

5 INEC., *"**Índice de Precios a la Construcción (IPCO)"*.** Publicaciones y Metodologías. , 1999-2003. , Talleres Gráficos INEC, Quito, 2004.

MANUALES

1. ANSI / IEEEm, "***Standard 829-1983 for Software Test ocumentation",*** IEEE Press 1983.

2. CRISTOFOLI, M., **"*Manual de Estadística con Excel*"**, 1ra Ed, Edit Omicron System.S.A., Buenos Aires, 2003.

3. MICROSOFT COORPORATION ©, ISEMGER David, "***SQL Server 2000 Reference Library***", 1era Ed Vol 6 msdn Library, Edit. Microsoft Press., Chicago 2000.

4. SYBASE © , *"**PowerBuilder Getting Started"***, Sybase Inc., Dublin 2003

SITIOS DE INTERNET.

1 ACOSTA, A, Ecuador: "***Deuda externa y migración, una relación incestuosa*"**. ,http://168.96.200.17/ar/libros/ecuador/idiuc/acosta.rtf; Octubre 2005.

2 BCE. , Banco Central del Ecuador, "***Cuentas Nacionales*"** (varios números); http://www.bce.fin.ec; Octubre 2005.

3 BID.,*"**Ecuador, Programa de Apoyo al Sector Vivienda, Documento del Banco Interamericano de Desarrollo"***. http://www.iadb.org/exr/doc98/apr/ec1416s.pdf; Octubre 2005.

4 CALDERON C, Bernardo A., Universidad de Antioquía, Colombia., ***"Prueba de Hipótesis sobre proporciones y diferencia de proporciones"***., http://bochica.udea.edu.co/~bcalderon/6_prop_difproporciones.html; Enero 2007.

5 CIDEIBER (Centro de Información y Documentación Empresarial sobre Ibero América."***Ecuador, Actividades del sector servicios, Construcción y Vivienda***.", http://www.cideiber.com/infopaises/Ecuador/Ecuador-06-07.html; Agosto 2005.

6 DÁVALOS SOTELO R,, RICALDE M., "***Elaboración de una guía práctica para el diseño y construcción de viviendas ligeras*"**: http://www.uv.mx/posgrado/memorias/3/3_6.htm, octubre 2006.

7 DE LA VEGA, Pablo. , Revista Aportes Andinos. , "***Análisis sobre Derecho al Desarrollo. , Deuda externa y migración*"**, Programa Andino de Derechos Humanos, Universidad Andina Simón Bolívar, Ecuador Agosto 2002., http://www.uasb.edu.ec/padh/revista3/articulos/pablodelavega.htm; Octubre 2005.

8 DE LA TORRE, Leticia., Instituto Tecnológico de Chihuahua, México., "***Curso de Estadística - Teoría del Muestreo - Diferencia de proporciones***". http://www.itch.edu.mx/academic/industrial/estadistica1/toc.html; enero 2007.

9 DOMÍNGUEZ I.., GABINO G., "***Aprendizaje Inductivo de Reglas basado en Algoritmos Genéticos"***, Monografía para la Asignatura de Inteligencia Artificial II en Ingeniería Informática de la Universidad de Valladolid, Junio 2003.; http://www.infor.uva.es/~calonso/TemasAlumnosIAII/AIRBAG.pdf, Enero 2005.

10 GARCÍA CEBRIAN, María José., "***Contraste de Hipótesis*"**., **http://descartes.cnice.mecd.es/Bach_HCS_2/inferencia_estadistica/contraste.htm**, Octubre 2006.

11 GALLEGO A.C**., "*Introducción a las redes neuronales artificiales*".,** Madrid 2004 http://www.gui.uva.es/login/13/redesn.html; Agosto 2005..

12 GILGERT, Karina., "***Técnicas híbridas de Inteligencia Artificial y Estadística para el descubrimiento de conocimiento y la minería de datos***"., Universidad de Cataluña, España., Enero 2007
http://www.lsi.us.es/redmidas/Capitulos/LMD12.pdf

13 GIRALDEZ Raúl, RIQUELME José C, AGUILAR Jesús., "***Aprendizaje evolutivo de Reglas: Mejoras en la codificación y evaluación de Individuos***", Departamento de Lenguajes y Sistemas Informáticos - Universidad de Sevilla., España, Enero 2007
http://www.lsi.us.es/redmidas/Capitulos/LMD27.pdf

14 GOMEZ RUIZ J.A***.,*** "***Un modelo para la Predicción de Recidiva de Pacientes Operados de Cáncer de Mama (CMO) basado en redes neuronales"***., Dpto. de Lenguajes y Ciencias de la Computación, Universidad de Malaga ;
http://tornado.dia.fi.upm.es/caepia/numeros/11/gomez.pdf; Octubre 2005.

15 GONZALEZ BOSQUE M., "***Conceptos Fundamentales de los Algoritmos Genéticos"***., Universidad de Buenos Aires 1998.,
http://www.geocities.com/SiliconValley/Byte/4713/ga/E-GA7.htm ; Septiembre 2005.

16 HERNANDEZ LOPEZ Monserrat., "***Predicción Económica con algoritmos genéticos: operadores genéticos versus matriz de transición"***., 1ra Ed. Talleres Gráficos - Facultad de Ciencias Económicas y Empresariales de la Universidad de la Laguna, España., Revista Estadística Española, Vol 46. Num 157, Madrid 2004.
http://www.ine.es/revistas/estaespa/157_1.pdf, ;Abril 2005.

17 HERNANDEZ ORALLO José., "***Introducción a la Prospección de Datos Masivos (Data Mining)***", Universidad Politécnica de Valencia - España., Enero 2007
http://www.dsic.upv.es/~jorallo/master/seminari.part.II.pdf

18 KURI MORALES A., "***Redes Neuronales y Algoritmos Genéticos"*.**;
http://cursos.itam.mx/akuri/Semestre2/AGS/19RNSyAGS.PDF; Noviembre 2005.

19 LOZANO L., FERNANDEZ, J; "***Razonamiento basado en casos: Una visión General"***., Universidad de Valladolid., Valladolid – España. 2006
http://www.infor.uva.es/~calonso/IAI/TrabajoAlumnos/Razonamiento%20basado%20en%20casos.pdf; Abril 2006.

20 M.J. DEL JESUS, P.GONZÁLEZ, F.HERRERA, M.MESONERO., "***Algoritmo Evolutivo de Extracción de Reglas de Asociación aplicado a un Problema de Márketing***.", Departamento de Informática., Universidad de Jaén, España.
http://sci2s.ugr.es/keel/pdf/keel/congreso/maeb04-reglas-jaen-granada.pdf: , Agosto 2006.

21 MERELO, J., "***Informática Evolutiva: Algorítmos Genéticos***.",
http://geneura.ugr.es/~jmerelo/ie/ags.htm ; Agosto 2005.;

22 MORALES MANZANAREZ Eduardo., "***Descubrimiento de Conocimiento en Bases de Datos***"., Enero 2007
http://ccc.inaoep.mx/~emorales/Cursos/KDD03/principal.html

23 NIKOS D., "***Inducción de Árboles de decisión"***, CBLU., University of Leeds.,
http://dns1.mor.itesm.mx/~emorales/Cursos/KDD01/node21.html; Octubre 2005.

24 OIT.,Oficina Internacional del Trabajo., "***Análisis de las Políticas y Programas Sociales en Ecuador –Las Dimensiones de la Pobreza en el Ecuador*"**., pp.35-41 http://www.oit.org.pe/ipec/documentos/politicas_sociales_ecuador.pdf: Septiembre 2005.

25 OLEAS, J, "***La construcción en 2003: buen año pero pudo ser mejor*"**., http://www.gestion.dinediciones.com/16/3.htm; Enero 2004.

26 OMS, Organización Mundial de la Salud. , Programa de Políticas Públicas y Salud.- División de Salud y Desarrollo Humano – "***Programa de Saneamiento Básico - Serie de Informes Técnicos – Capítulo ECUADOR*",** Washington, DC., ,http://www.oit.org.pe/ipec/documentos/politicas_sociales_ecuador.pdf; Octubre 2005

27 RED CIENTÍFICA, GAIA ., Ciencia, Tecnología y Pensamiento, "***Computación Evolutiva*"**., http://www.redcientifica.com/gaia/ce/ce_c.htm l; Septiembre 2006.

28 ROMERO Cristóbal, VENTURA S, Hervas C.,
[1] "***Descubrimiento de Reglas de Predicción en Sistemas de e-learning utilizando Programación Genética"***., Universidad de Córdova-España, Enero 2007 http://www.lsi.us.es/redmidas/Capitulos/LMD05.pdf
[2] "***Aplicación de Algoritmos Evolutivos como Técnica de Minería de Datos para la mejora de cursos hipermedia adaptativos basados en web***", Enero 2007., http://www.utpl.edu.ec/ried/index.php?option=com_content&task=view&id=23&Itemid=53

29 SERRANO C., "***La Inteligencia Artificial***", Universidad de Zaragoza, España ; http://ciberconta.unizar.es/LECCION/INTRODUC/490.HTM; Agosto 2005.

30 VERBEEK, E., "***Prejuicios y Realidades sobre Redes Neuronales*"**, Revista de ADEMO- Investigación y Marketting; http://www.aedemo.es/septiembre97/privado/redneuro.html; Octubre 2005.

31 VILLA A., SANCHEZ D., ESCOBAR L., "***Relaciones causales en reglas de asociación"*.,** XII Congreso Español sobre tecnologías y lógica fuzzy.,Enero 2006 http://decsai.ugr.es/~castro/docto-csi/LER/p44.pdf

32 VILLAVICENCIO, G., "***Globalización, Pobreza y Exclusión en las áreas urbanas del Ecuador*".**, CER-G; Conferencia Latinoamericana y Caribeña de Ciencias Sociales; Recife, Brasil; Noviembre 1999; http://www.fundaj.gov.br/clacso/paper08.doc, Octubre 2005.

33 ZAVALA TOLEDO C.A., "***Modelo de Red Neuronal en el Ensayo Híbrido en Línea en Subestructuras de Pórticos Planos Flexibles*"**. , Universidad Nacional de Ingeniería., Lima – Perú., 1998 Vol.8 No.02.; http://quipu.uni.edu.pe/OtrosWWW/webproof/public/revistas/tecnia/vol8n2/04art; Octubre 2005.

.

ANEXOS.

ANEXO 1.
MANUALES

1.1 MANUAL DE INSTALACIÓN.

1.1.1 REQUERIMIENTOS MÍNIMOS.

1.1.1.1 Requerimientos de Hardware.

1 Servidor de Datos

1 PC. Cliente Local para instalación de la aplicación y conexión al Servidor a través de red local.

Servidor de la red:

- Procesador P-IV de 1Ghz, con 128 MB de RAM
- Disco local de capacidad > 10 GB Y 100 MB disponibles
- Impresora HP LaserJet 8000 N PS o similar en Inyección a tinta.

Estación de trabajo (Usuario):

- Procesador P-IV de 1Ghz, con 128 MB de RAM
- Disco local de capacidad >= 10 GB Y 60 MB disponibles
- Tarjeta de Video resolución superior a 800 x 600 y memoria 8MB

Para la demostración del sistema se puede disponer de un sólo equípo con las características del harware del servidor.

1.1.1.2. Requermientos de Software.

SISTEMA OPERATIVO.

- Alternativas:

  Microsoft Windows 2000 Profesional.
  Microsoft Windows XP.

- Configuración Regional (Ecuador) con punto decimal y coma para separador de miles.

SOFTWARE INSTALADOS: CLIENTE

- Sistema de impresión HP LaserJet 8000
- WinZip
- Microsoft Office 2000 Professional
- SQL Server 2000

SOFTWARE INSTALADOS: SERVIDOR

- SQL Server 2000 (Administrador de servicios, Administrador Corporativo, Analizador de consultas, Herramientas de red, Ayuda-Libros en pantalla).

1.1.2. GUIA DE INSTALACIÓN.

1.1.2.1. Proceso de Instalación.

1.1.2.1.1. Tareas previas a la instalación.

a) Crear una instancia de SQL Server 2000 en el Servidor con características de tipo de instalación "Herramienta cliente y servidor". Luego, desde el Administrador Corporativo de SQL. Server:

   - Reconocer el nombre del servidor.
   - Revisar que en las propiedades registradas de SQL Server se encuentre seleccionado lo siguiente : Mostrar estado de SQL Server en la consola; Mostrar bases de datos y objetos del sistema, Iniciar automáticamente SQL Server al conectar y preferentemente esté activada autenticación de Windows y autenticación opcional SQL Server de confianza usuario "sa".

b) Revisar que la máquina PC CLIENTE cumpla con las características de requerimientos de hardware y software.

1.1.2.1.2 Instalación de la parte del Servidor.

Hay tres maneras de implantar la base de datos en el servidor:

1.- Generando y utilizando secuencias de comandos

2.- Separando y adjuntando la base de datos

3 - Con copia de seguridad y restauración

Se explicarán aquí las dos primeras, que son las que se usarán:

Pasos para generar y utilizar secuencia de comandos.

1. El script de creación de la base de datos se encuentra en el archivo "creacion_bdd_tablas.sql" del directorio \BASES del CD de Instalación.
   El script fue generado por la opción;
   [Todas las tareas] > [Generar Secuencia de comandos SQL]
   del Administrador Corportativo de SQL Server 2000, en desarrollo.

2. Para utilizar la secuencia de comandos o script de creación de la base de datos SPACV, debe abrirse el archivo "creacion_bdd_tablas.sql" indicado, desde el programa Analizador de Consultas de SQL Server 2000 (activándolo desde Inicio de programas de Windows).

   En el Analizador de consultas debe conectarse a la instancia de SQL Server del servidor local, y debe aparecer en uso la base "master".
   Usando la opción del menú [Archivo]>[Abrir], abrir el archivo "creacion_bdd_tablas.,sql". Antes de ejecutar el script, se debe crear en el disco duro del servidor, el directorio \BASES\ con su subdirectorio \SPACV y este a su vez con los subdirectorios \Data y \Log.
   C:\BASES\SPACV\Data
   C:\BASES\SPACV\Log

3. Ejecutar el archivo de script de creación "creacion_bdd_tablas.sql" desde el Analizador de Consultas SQL Server, como se indica en el paso anterior.
   Ver los mensajes en la ventan. Luego ir en el Administrador Corporativo, colocarse en el servidor de bases de datos con el cursor del mouse, y usar

el menú del botón derecho para activar la opción [Actualizar] Verificar luego que se encuentre creada y en servicio la base de datos SPACV.

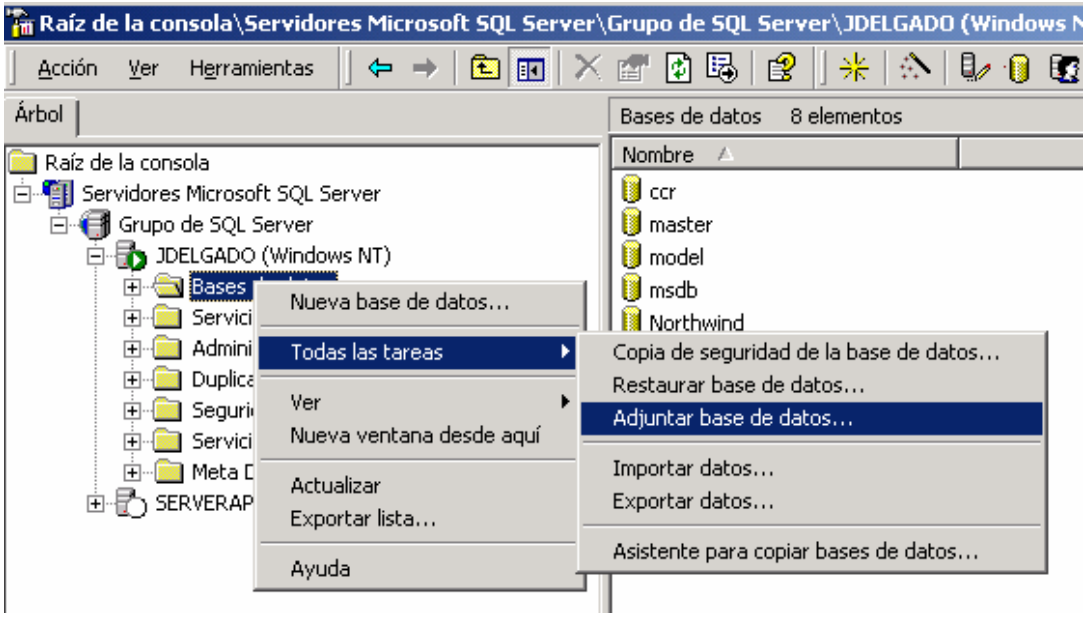
4. El contenido de los datos de prueba se carga a la base de datos a través del archivo de "scrip_insersion_01.sql" desde el Analizador de Consultas SQL Server de la misma manera que se hizo con el script de creación.
   Para cargar todos los datos, incluidos los datos históricos de edificaciones de los años 2000-2005, se debe ejecutar la aplicación SPACV, on clave de usuario "INST" y contraseña "mantenimiento", una vez que esta ya se haya instalado desde el Cliente.

Pasos para separar y adjuntar la base de datos.

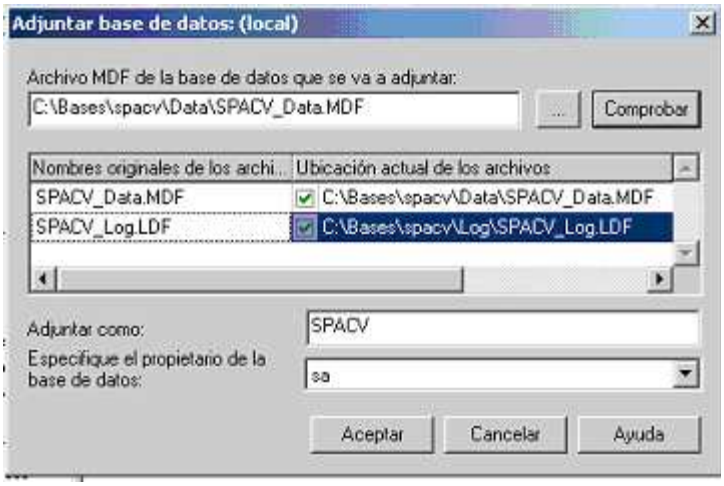
1. Separar y copiar los archivos de la base de datos SPACV hacia el directorio \BASES\SPACV y los subdirectorios \DATA y \LOG organizados en el servidor de producción como:
   C:\BASES\SPACV\DATA\SPACV_Data.MDF
   C:\BASES\SPACV\LOG\SPACV_Log.LDF

   Los datos fueron separados del servidor de desarrollo mediante la opción del Administrador Coorportativo de SQL Server: [Todas las Tareas] > [Separar base de datos] activada mediante el botón derecho del mouse una vez y apuntando el cursor en la base de datos a separar.
   Estos archivos se encontrarán en el directorio \BASES del CD de instalación.

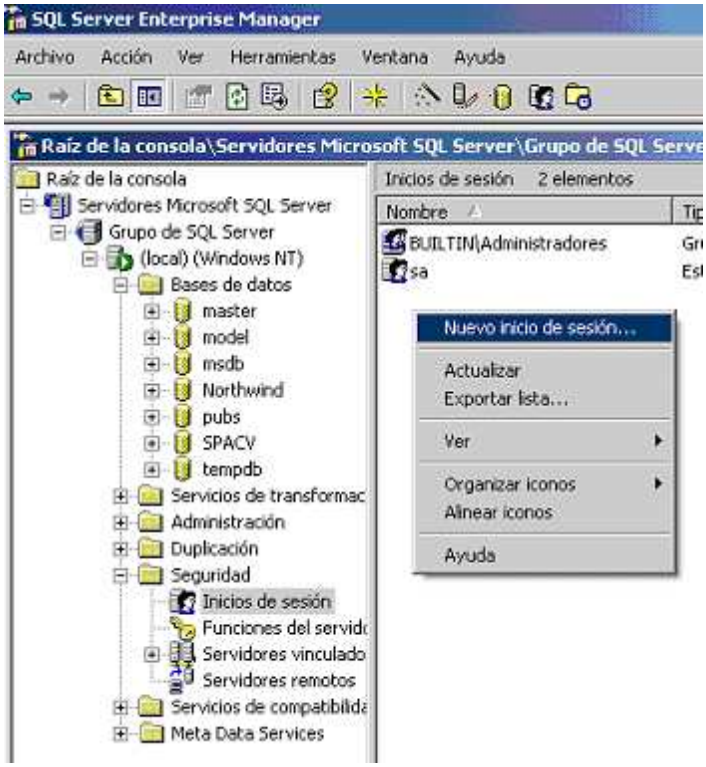
2. Ir a la opción Bases de datos [Todas las Tareas] > [Adjuntar Base de Datos] mediante el botón derecho del mouse apuntando el cursor en la base de datos a adjuntar, dentro del Administrador Coorporativo de SQL Server 2000 del servidor de producción.

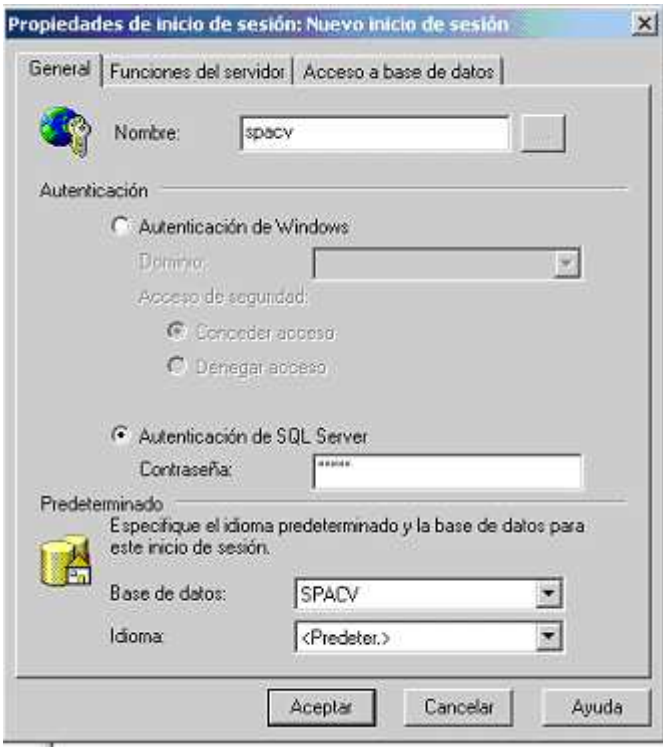


Indicar las ubicaciones de los archivos MDF y LDF en el Servidor.

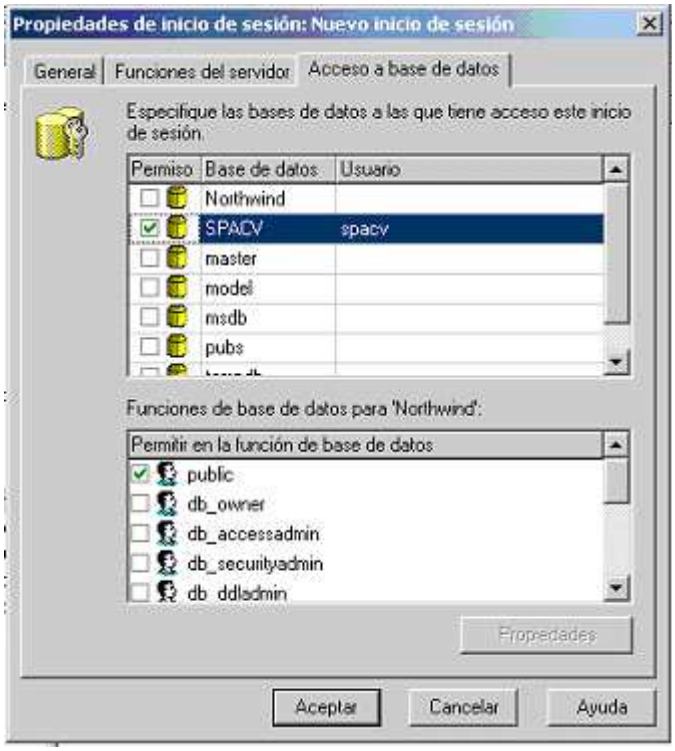


3. Crear un inicio de sesión de nombre "SPACV" hacia la base de datos "SPACV". (Antes se debe eliminar el usuario "SPACV", si este ya fue creado antes, para impedir que de error por duplicidad de nombres).





La autenticación debe ser de SQL Server con contraseña "spacv". La base de datos SPACV debe seleccionarse del cuadro de lista "Base de datos".



Los permisos de acceso deber ser tipo "public".

4. La base de datos SPACV debe aparecer luego en el Administrador Corporativo y estar presente en su carpeta "\Usuarios" como en su carpeta de "\Seguridad \ Inicios de sesión", el nombre "spacv".

1.1.2.1.2 Instalación de la parte del Cliente.

CREACIÓN DE LA CUENTA CLIENTE SQL SERVER 2000.

Seguir el proceso de instalación de SQL Server, con las opciones:

a) Crear una nueva instancia de SQL Server o instalar herramientas cliente.
b) Tipo de instalación: Solo herramientas cliente.

INSTALACIÓN DE LA APLICACIÓN

Para instalar la Aplicación SPACV se necesita:

a) El CD de Instalación etiquetado: SPACV 2006-Vers 1.0 Nov2006.
b) Este Manual.

c) Ejecutar el programa de instalación automático SETUP.EXE y seguir los PASOS DE INSTALACIÓN.

PASOS DE LA INSTALACIÓN DE LA APLICACIÓN DESDE EL PROGRAMA SETUP.EXE

- Copiar el contenido del directorio INSTALAR desde el CD hacia un directorio INSTALAR del disco duro del PC CLIENTE.
- Ejecutar desde el directorio INSTALAR creado en el disco duro, el programa SETUP.EXE.
- Pulsar el botón "Next" hasta concluir la Instalación.

  El directorio donde se instala la aplicación debe ser "C: \SPACV", preferentemente.
- Verificar que se haya creado el icono de escritorio. Sobre el mismo, presionar el botón derecho del mouse y a continuación dar clic en Propiedades. Luego escribir el nombre del directorio de la aplicación (C:\SPACV) en el cuadro de texto que aparece luego de "Iniciar en: ".
- Verificar que en los archivos de texto SPACV.INI y SPACV.DSN, conste nombre del servidor (local) con usuario "sa" y su contraseña. En SPACV.INI la ubicación de los nombres de directorios debe coincidir con la ubicación donde se encuentre la aplicación.

1.2. MANUAL DEL USUARIO.

1.2.1 INTRODUCCIÓN.

1.2.1.1 Acerca del Sistema.

El Sistema de Predicción y Análisis de Casos de Construcción de viviendas en el país utiliza como fuente de información el resultado de las Encuestas de Edificaciones que realiza el Instituto Nacional de Estadística y Censos (INEC) cada año en base a los permisos de construcción que informan los municipios de todo el país. Se extraen las variables de materiales predominantes en los cimientos, estructura, paredes y techo de las viviendas, el uso de la edificación y el origen del financiamiento. El sistema aplica la técnica de algoritmos genéticos para descubrir sus reglas de asociación y predice el comportamiento futuro de estas variables a través del grado de aceptación que lleguen a alcanzar las reglas descubiertas para una muestra aleatoria de ejemplos históricos de edificaciones extraida al azar.

1. 2.1.2 Objetivos y Alcance del Sistema.

Objetivos.

- Explicar y predecir el comportamiento de las variables de construcción de viviendas indicadas para casos de estudio en las ciudades del país donde el proceso de urbanización es más relevante.
- Utilizar el Sistema para hacer una análisis de las variables investigadas y aplicar la técnica de algoritmos genéticos para descubrimiento de reglas de asociación y predicción por clasificación.

Alcance.

- Aplicar el Sistema a casos de estudio en las ciudades de Quito, Guayaquil, Cuenca y Machala.

- Utilizar la información de edificaciones (permisos de construcción) de los años 2000,2001,2002,2003,2004 y 2005.

1.2.1.3 Terminología

1.2.1.3.1 Términos especiales

Término	Descripción
archivo DSN (file DSN)	Almacena información de conexión para una base de datos en un archivo que se guardó en el equipo. El archivo es de texto con la extensión .DSN. La información de conexión consta de parámetros y sus valores correspondientes que el Administrador de controladores ODBC utiliza para establecer una conexión.
archivo INI (file INI)	Almacena información de parámetros de propiedades de la conexión de datos y ubicación de los archivos de la aplicación incluidos el archivo DSN. También se guardan especificaciones de parámetros de ejecución del programa.
atributo	Una característica del fenómeno representado por una variable. Para el Sistema SPACV se trata de una característica de la construcción de viviendas que puede ser una de las variables investigadas como: tipo de obra, material predominante en cimientos, estructura, paredes y techo o cubierta, uso de la edificación y origen del financiamiento.
autenticación	Proceso por el que se valida si el usuario que está intentando conectarse a SQL Server está autorizado. También se utiliza el mismo término para indicar si el usuario se puede conectar al sistema SPACV, con su clave de acceso para conexión a la base de datos de SQL Server.
base de datos	Colección de información, tablas y otros objetos organizados y presentados para un propósito específico como, por ejemplo, búsqueda, clasificación y procesos de ejecución de algoritmos. Las bases de datos tienen dos archivos: el de datos y el de registro de transacciones.
base de datos de SQL Server	Base de datos basada en el lenguaje de consulta estructurado SQL y administradas por SQL Server
campo	área de una ventana o registro que guarda un sólo valor de datos
clase	Definición de un objeto. Guarda sus atributos y modo de comportamiento u operaciones.
clave	columna o grupo de columnas que identifican únicamente a una fila (PRIMARY KEY) y definen la relación entre dos tablas (FOREIGN KEY)
columna	En una tabla de SQL, el área en la fila que almacena el valor de los datos de algunos atributos del objeto modelado por la tabla.
conexión	Vínculo de comunicación entre procesos (IPC) establecido en una aplicación para SQL Server y una instancia de SQL Server 2000. La aplicación SPACV utiliza el vínculo IPC para enviar instrucciones Transact-SQL Server y para recibir resultados, errores y mensajes de SQL Server.
conexión de datos	Información necesaria para tener acceso a una base de datos específica, en este caso SPACV. La información incluye el nombre de origen de datos y la información de inicio de sesión, Esta información se guarda en el archivo SPACV.INI , y sus parámetros en el archivo SPACV.DSN
datos de ejemplo	Datos generados o presentados en lugar de los datos reales. Obtenidos de la extracción de información histórica almacenada en la base de datos, particularmente en la tabla edificación.
datos generados	Los datos producidos aleatoriamente de entre rangos máximos y mínimos para los valores de los atributos de las variables investigadas de la construcción de viviendas, se transforman a binario y se almacenan en un solo registro, el cual representa un genotipo.

diccionario de datos	Conjunto de tablas del sistema, almacenado en un catálogo, que incluye las definiciones de las estructuras de datos e información relacionada, como los permisos de construcción.
fila	En una tabla de SQL, colección de elementos que forman una línea horizontal en una tabla. Cada fila representa una sola aparición del objeto modelado por la tabla y almacena los valores de todos los atributos de dicho objeto.
Identificación de inicio de sesión	Identificación (ID) que un usuario debe utilizar para iniciar sesión en un servidor local.
instancia	Copia de SQL Server que se ejecuta en un equipo. En el equipo se pueden ejecutar varias instancias de SQL Server 2000.
interfaz	Conjunto definido de propiedades, métodos y colecciones que componen una agrupación lógica de datos y comportamientos. Una interfaz se puede implementar mediante distintas clases.
master, base de datos	Base de datos que controla el funcionamiento de cada instancia de SQL Server. Realiza un seguimiento de las cuentas de usuarios y contiene procedimientos almacenados para controlar otras bases de datos.
objeto	En bases de datos, uno de los componentes de una base de datos: una tabla, índice, desencadenador, vista, clave, etc. En Orientación a Objetos: Una instancia de una clase.
procedimiento almacenado	Colección precompilada de instrucciones Transac-SQL almacenadas bajo un nombre y procesadas como una unidad.
servidor local	En conexiones SQL Server 2000, instancia de SQL Server 2000 que se ejecuta en el mismo equipo que la aplicación.
tabla	Objeto de dos dimensiones que consta de filas y columnas y se utiliza para almacenar datos en una base de datos relacional.
usuario	En SQL Server: Cuenta o identificador de seguridad de SQL Server que representa un determinado usuario de una base de datos. En el Sistema SPACV: Cuenta de acceso a las opciones del sistema según su rol definido. Cada cuenta de usuario puede tener acceso al sistema en las opciones que tiene permiso trabajar.

1.2.1.3.2 Abreviaturas

Abreviatura	Descripción
ID	(Identificador). Utilizada para atributos clave o para nombrar a la identificación de usuario durante la autenticación.
DSN	(Data Source Name) Nombre de origen de datos utilizado en la conexión a la base de datos.
dw	(Data Window) Objeto de edición, presentación y consulta de datos en Power Builder
dddw	(DropDownDataWindow) Objeto de consulta de datos para ser usado en una lista de despliegue..
ho	hipótesis nula
AG	Algotrimo Genético

1.2.1.3.3 Siglas

Sigla	Descripción
SPACV	(Sistema de Predicción y Análisis de Construcción de Viviendas) Nombre del Sistema desarrollado para esta tesis.
ODBC	(Open Database Connectivity) Interfaz de programación de aplicaciones (API) de acceso a datos que admite el acceso a cualquier origen de datos para el que existe un controlador ODBC. ODBC cumple con las normas stándar de ANSI
ANSI	(American National Standards Institute) Instituto de Normas estándar americanas usadas en los Estados Unidos y su área de influencia, entre ellas Latinoamérica.
OSI	Organización Internacional de Estandarización
SQL	Lenguaje de Consulta estructurado

1.2.2 GENERALIDADES

1.2.2.1 Seguridad y Administración de Usuarios

La seguridad tanto para el Sistema SPACV como para el acceso a la base de datos se controla desde el Sistema Operativo (cuando se realiza la autenticación Windows), y luego continúa a través de SQL Server, con reconocimiento del usuario en el inicio de sesión. Y por último, el sistema SPACV reconoce al usuario en la tabla de usuarios de la base de datos.
La administración de las seguridades de la base de datos SPACV se la puede hacer a través del uso del Administrador Corporativo de SQL Server 2000.

1.2.2.1.1 Descripción de Usuarios.

Se han definido dos tipos de usuarios: El usuario de conexión a la base de datos SPACV y el usuario de conexión al sistema del mismo nombre:SPACV.

El primer tipo de usuario es quien realiza la conexión hacia la base de datos SQL Server y es el que se halla definido como objeto de la base de datos. Su ID y contraseña se guardan en los archivos SPACV.INI y SPACV.DSN

El segundo tipo de usuario es el que se registra en la tabla de usuarios de la base de datos SPACV. Los usuarios de este tipo pueden ser creados y modificados por el administrador desde la pantalla de administración de usuarios del sistema SPACV.

1.2.2.1.2 Cuentas de Seguridad.

La cuenta de seguridad usada para el Sistema SPACV es la misma cuenta de seguridad de inicio de sesión que se tiene para el acceso al servidor de base de datos SQL Server. El inicio de sesión se la hace luego de haber instalado una instancia de SQL Server. El sistema SPACV utiliza el inicio de sesión para usuario

Windows. De esta manera, SQL Server no tiene necesidad de volver a validar la identificación de usuario. La validación del inicio de sesión se realiza durante la conexión mediante lectura de los archivos SPACV.INI y SPACV.DSN

1.2.2.1.3 Autenticación

Para accesar al sistema SPACV, se requiere de identificación de usuario y contraseña. El proceso de autenticación es transparente y se la hace mediante lectura de parámetros del archivo SPACV.INI, durante la conexión. El acceso a la base de datos se la hace a través del inicio del inicio de sesión SQL Server.
Si se realiza algún cambio en los nombres de inicio de sesión y usuario de SQL Server para la base de datos SPACV, se requiere inmediatamente transmitir el cambio realizado hacia los parámetros de los archivos SPACV.INI y SPACV.DSN.
En SPACV.ini se especifica para la conexión, los parámetros:

```
Server=(local)          ( nombre del servidor)
UsrName="SPACV"         (Nuevo nombre de usuario)
UsrPass ="spacv"        (Contraseña del usuario)
```

En SPACV.DSN se especifica tanto nombre de servidor como identificador UID

```
[ODBC]
DRIVER=SQL Server
UID=sa
Trusted_Connection=Yes
DATABASE=spacv
WSID=
APP=Microsoft Open Database Connectivity
SERVER=(local)
Description =spacv
```

1.2.2.1.4. Roles y Permisos

Para esta aplicación se tienen dos tipos de roles: Los roles fijos definidos por SQL Server y los roles definidos por el usuario en el Sistema SPACV

Los roles definidos en SQL Server se dividen en roles de servidor y roles de bases de datos. Los roles de servidor llevan asociados permisos para realizar tareas de servidor, tales como creación, modificación, borrado de la base de datos o administración de inicios de sesión y la modificación de contraseñas. Los roles de bases de datos permiten asignar permisos sobre objetos de la base de datos como por ejemplo seleccionar finas de una tabla.

Para el sistema SPACV, estos permisos ya se encuentran predefinidos en la base de datos instalada, y no requiere de mayores complicaciones por parte del usuario administrador.

Las tareas más destacadas de estos roles son:

Nombre del rol	Tareas seleccionadas
Sysadmin	Puede realizar cualquier tarea y obtiene acceso sin restricción a todas las bases de datos
Serveradmin	Puede realizar operaciones sp_configure y SHUTDOWN
DBCreator	Puede realizar operaciones CREATE DATABASE, ALTER DATABASE y DROP DATABASE

Los roles definidos por el usuario en el Sistema SPACV, son fijos y sirven para crear nuevos usuarios desde la aplicación. Estos roles son los siguientes:

Nombre del rol	Tareas seleccionadas
ADM	Todas las opciones
CONS	Procesamiento y consulta a resultados
PROC	Procesamiento (Predicción y Análisis)
MANT	Solo opción de mantenimiento de tablas (instalación e importación de datos)

Los usuarios se crean en el sistema con un identificador, una contraseña y un rol específico, que se selecciona de los indicados en la tabla.

1.2.2.2 Breve descripción del Sistema.

El Sistema SPACV dispone de dos opciones principales en el menú general: La primera, nombrada como "Datos", permite realizar todas las tareas de administración, importación y mantenimiento de los datos históricos de edificaciones, variables, series de indicadores, tiempos y localidades geográficas, parámetros generales, y además la administración de usuarios, a través de ventanas de mantenimiento independientes entre sí. La segunda opción, denominada "Procesos" permite realizar tanto el análisis como la predicción de los casos de estudio a través de una sola ventana dividida en tres secciones. La primera sección de esta ventana sirve para la introducción y selección de casos de estudio. En la sección de introducción de casos de estudio se ingresa el año y la localidad geográfica. La segunda sección sirve para el análisis de datos correspondientes al caso de estudio. Y la tercera sección de la ventana es la que permite realizar todo el proceso de predicción.

La sección de predicción se divide a su ves en tres subsecciones. La primera sirve para introducir los parámetros del método aplicado. La segunda es una subsección de proceso con los pasos a seguir, los cuales corresponden a la técnica del algoritmo genético aplicada al descubrimiento de reglas de asociación, en dónde se introducen además parámetros de ejecución, de los cuales constan los tipos de reglas a descubrir. En la tercera subsección de este proceso se encuentran los resultados de las reglas extraídas y es en la cual, luego de concluir con el proceso, se puede seleccionar una determinada regla y evaluarla con un procedimiento de contraste de hipótesis. Los resultados pueden ser impresos desde pantalla.

El sistema dispone además de ayuda en línea que se activa con pulsar la tecla F1 desde cualquier parte que se encuentre. Hay también una pantalla de resumen de la ayuda y la ventana de acerca del sistema.

1.2.3 GUIA DE OPERACIÓN.

1.2.3.1 Uso de los dispositivos de Entrada.

Uso del teclado.

De preferencia debe utilizarse el teclado en español, el cual se debe seleccionar en la configuración del panel de control de Windows.
Para la introducción de datos numéricos se pueden utilizar las teclas de la sección derecha del teclado, para lo cual debe activarse la tecla "Bloq Num".
Los datos descriptivos se introducen mediante las teclas de letras desde la "a..z" e minúsculas y desde la "A..Z" en mayúsculas. Se pueden introducir indistintamente entre mayúsculas o minúsculas los nombres de localidades geográficas, y para los nombres específicos de variables preferentemente se debe introducirlos todos en mayúsculas. En estos nombres es conveniente no introducir caracteres especiales.
Las teclas pueden usarse de manera alternativa al mouse, si es que se conocen las combinaciones apropiadas y que se encuentran junto a las descripciones escritas frente a cada opción del menú principal.

Uso del mouse o ratón.

En el sistema SPACV el mouse o ratón se usa con la finalidad de activar eventos al dar "clic" izquierdo o principal, luego que se lo haya posicionado dentro de los controles de usuario de las pantallas, tales como botones de comando, listas de selección, opciones del menú principal, barra de herramientas y filas de los controles datawindow en las ventanas de mantenimiento y procesos.

Si se encuentra el cursor activo dentro de una ventana de mantenimiento, es posible dar un clic en el botón derecho del mouse para visualizar las opciones emergentes. En los casos de mantenimiento de localidades geográficas y de

tiempos es utilizado el botón derecho para acceder a la opción de importar registros a la tabla principal del datawindow activo.

1.2.3.2 Uso de los dispositivos de Salida.

El dispositivo de salida por excelencia usado en el sistema SPACV es la impresora. De acuerdo a las especificaciones del hardware, se debe disponer de una impresora laser o matricial que al momento de la ejecución del sistema debe encontrarse prendida. Los parámetros de impresión pueden ser modificados desde la ventana de impresión que aparece luego de dar un clic del botón principal del mouse en el icono de impresión. Estos parámetros tienen que ver con la orientación del papel, el número de copias y también la impresora predeterminada.

1.2.3.3. Acceso a la Aplicación.

El acceso a la aplicación se realiza desde la ventana del escritorio de Windows, que una vez que se haya instalado el sistema, aparece un icono distintivo del sistema SPACV desde el cual se tiene acceso a la pantalla LOGIN del sistema.

Login

En la pantalla LOGIN o de acceso al sistema se debe proporcionar la identificación de usuario y la contraseña. Luego se debe dar un <Enter> o un 'clic' del botón principal del mouse ubicado en el botón de comando 'Aceptar' de esta ventana.

El sistema realiza la autenticación del usuario, de forma transparente, y si es la clave correcta, entonces le muestra la ventana de opciones del menú general.

Pero si el usuario no es reconocido o la clave fue introducida de manera incorrecta, entonces se le muestra un mensaje de error y el usuario es abortado del sistema.

Independiente de cual sea el identificador de usuario y contraseña, el sistema intentará conectarse primero al servidor y base de datos SQL Server, leyendo la definición de parámetros de conexión se encuentran en los archivos SPACV.INI y SPACV.DSN. Por eso es preciso que el nombre del servidor sea el mismo que se encuentra declarado en el contenido de estos archivos.

1.2.3.4. Descripción de Módulos y Pantallas.

Una vez que el usuario se haya conectado con su clave de identificación y contraseña, entonces se presenta la ventana del menú principal. En esta se encuentra en primera línea horizontal las opciones a las cuales se accede ya sea a través del mouse o por medio de la combinación de teclas indicada al extremo derecho de la descripción de la opción.

SPACV Menú

Las opciones del menú principal son las siguientes:

Sistema: Permite acceder a la ventana de cambio de contraseña y contiene la una opción para salir del sistema.

Edición: Permite tener acceso a las opciones generales de edición de registros y a la barra de herramientas.

Datos: Por medio de esta opción se accede a las ventanas de mantenimiento de datos.

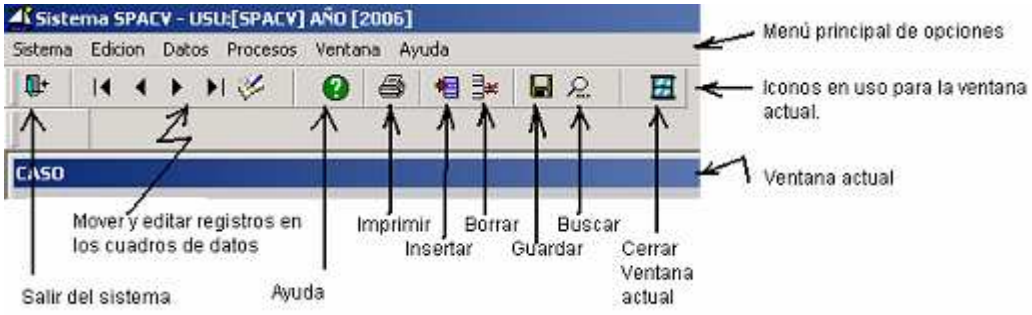
Procesos: Por medio de esta opción se accede a la ventana de predicción y análisis de casos de estudio.

Ventana: Permite hacer cambios en la presentación de las ventanas.

Ayuda: Presenta las ventanas de ayuda del sistema y la pantalla "Acerca de".

1.2.3.4.1. Módulo de Datos y Administración de Usuarios.

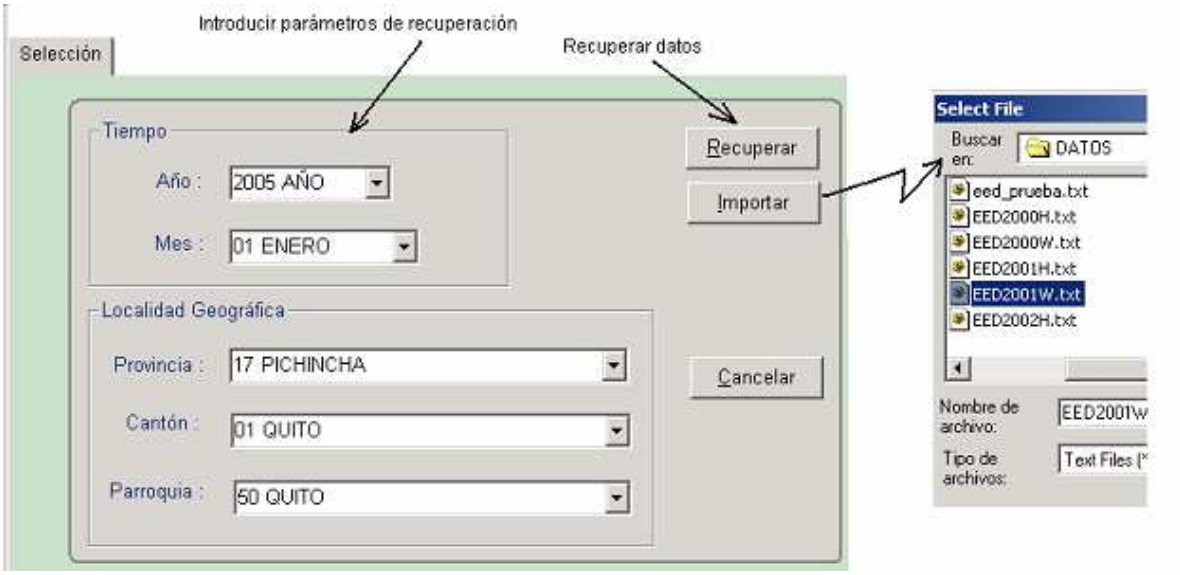
Durante la ejecución de estas ventanas se puede usar la barra de herramientas para edición de los registros y en algunos casos el botón derecho del mouse para tareas de importación de datos. Todas estas ventanas contienen un botón de "Cancelar" para abandonar la ventana. En la mayoría de los casos, los datos se guardan cada vez que se cierra la ventana, sin embargo es preciso utilizar el icono "guardar" de la barra de herramientas antes de abandonar la ventana. También para cerrar la ventana se dispone de icono de herramientas.



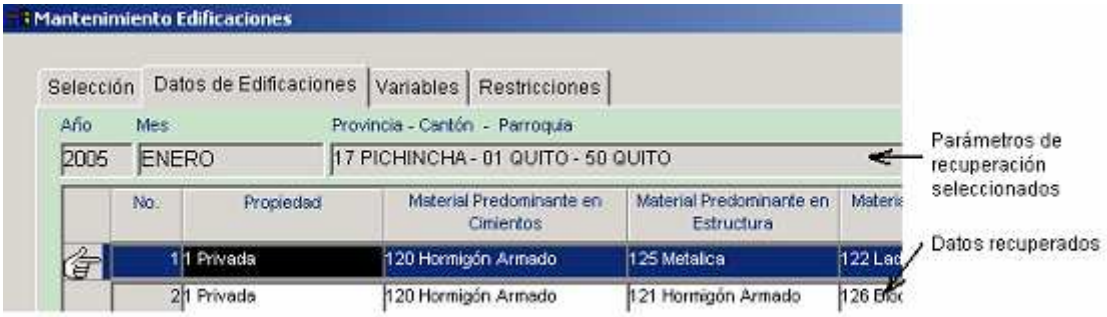
Pantalla de mantenimiento de Edificaciones.

Está dividida en cuatro secciones:

Sección Selección: sirve para introducir los parámetros de lugar y tiempo de los datos históricos a recuperar. La recuperación se realiza a través del botón "Recuperar" del lado derecho. El botón "Importar" permite realizar la importación de los datos desde archivos de texto que existen para cada año.



Sección Datos de Edificaciones: Una vez que se ha introducido los parámetros y se ha dado clic principal del mouse en recuperar, el sistema le permite ir a esta sección para visualizar o editar los registros de datos de edificaciones. Los datos de edificaciones prácticamente no necesitan de edición, ya que se importan y son el resultado de la encuesta de permisos de construcción que realiza el INEC cada año y los publica.



Sección Variables: En esta ventana se muestran los nombres de las variables principales de la encuesta de edificaciones y usadas por el sistema SPACV. Estos nombres también fueron importados desde un archivo texto al momento de la creación e instalación de la base de datos, pero pueden ser editados haciendo uso de la barra de herramientas, al igual que los datos de rangos de algunas de estas variables.

Variables

Variables de casos de construcción y sus categorías

Código	Clase	Descripción / Especificación Categoría	Nivel de Variable
01	1 CIMENTOS	CIMENTOS	0 Clase
01	1 CIMENTOS	No Aplica	1 Categoría
120	1 CIMENTOS	Hormigón Armado	1 Categoría

Sección Restricciones: En esta ventana se muestra la información de restricciones de las variables de edificaciones.. Esta información también fue importada desde un archivo de texto al momento de la creación y es utilizada únicamente como meramente informativo para el sistema.

Pantallas de mantenimiento de Indicadores Económicos

La Pantalla Indicador Económico

Esta ventana muestra la información de nombres de las variables de indicadores económicas, usadas con fines de análisis. Cada registro muestra la información de un indicador. Los datos que aparecen fueron introducidos al sistema a través de la opción Importar del botón derecho al momento del proceso de instalación. Sin embargo, pueden agregarse nuevos indicadores económicos por medio de la opción insertar de la barra de herramientas.

Codigo	Nombre Indicador	Nivel de Tiempo	Nivel de Localidad	Unidad Indicador
1	01 PIB TOTAL (%)	1 ANUAL	0 PAIS	02 PORCENTUAL
2	02 INFLACION MENSUAL	2 MENSUAL	0 PAIS	02 PORCENTUAL
3	03 INFLACION ANUAL	2 MENSUAL	0 PAIS	02 PORCENTUAL
4	04 ACERO EN BARRAS	2 MENSUAL	0 PAIS	01 EN UNIDAD INDICE
5	05 CEMENTO PORTLAND TIP	2 MENSUAL	0 PAIS	01 EN UNIDAD INDICE
6	06 HORMIGON PREMEZCLAD	2 MENSUAL	0 PAIS	01 EN UNIDAD INDICE
7	07 COMBUSTIBLES (MEZCLA	2 MENSUAL	0 PAIS	01 EN UNIDAD INDICE
8	08 COMPONENTES VIVIENDA	2 MENSUAL	0 F	

Selección de opciones

Archivar Ver Editar Procesos

Guardar
Guardar como
Importar
Salir

Usar barra de herramientas para insertar, eliminar

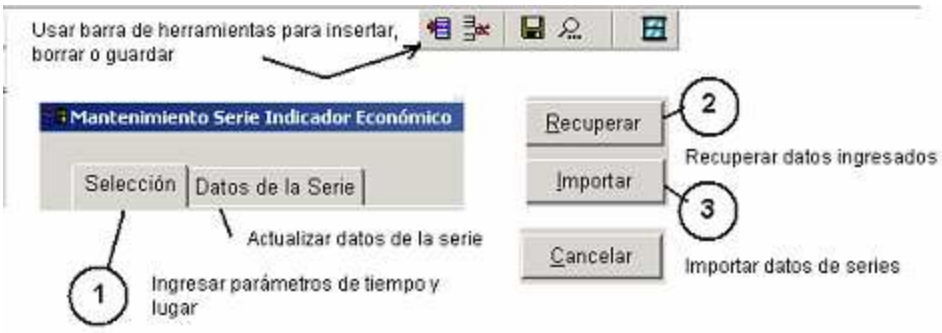
Usar boton derecho para importar registros.

mantenimiento series de indicadores económicos

Si ya existen datos de series económicas ingresadas para un indicador económico, el sistema no permite eliminar el registro de indicador económico. En este caso, si se quiere eliminar un registro de indicador, se debe eliminar primero las series de tiempo ingresadas para este indicador.

La Pantalla Series de Indicadores Económicos.

La pantalla de Series de Indicadores Económicos está dividida en dos secciones: "Selección" y "Datos de la Serie".



mantenimiento series de indicadores económicos

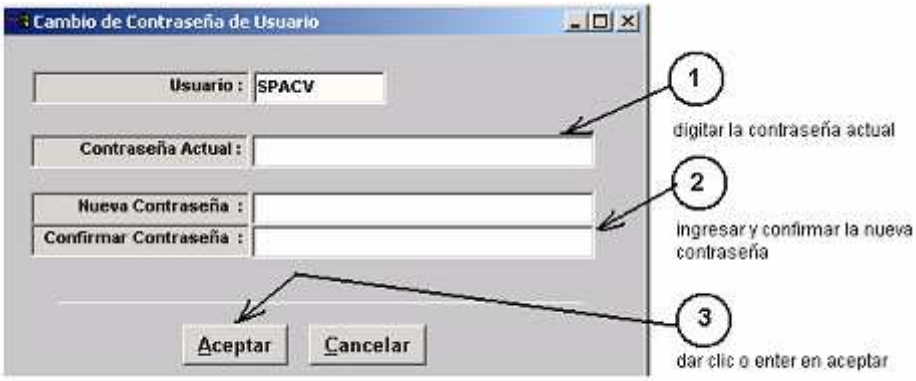
La primer sección "Selección" permite introducir los parámetros de tiempo y lugar necesarios para recuperar la información de indicadores. Para recuperar estos datos se tiene a la derecha el botón de comando "Recuperar".

El botón "Importar" permite ingresar los registros anuales o mensuales de los indicadores desde un archivo de texto. Esta operación se la hace al momento de la instalación y también puede hacérsela en cualquier momento.

La sección "Datos de la Serie" muestra los registros de tiempo y lugar ingresados para la serie de indicadores recuperada. Si se requiere introducir una nueva serie de indicador se puede usar el icono de "insertar" de la barra de herramientas.

Pantalla de Cambio de Contraseña.

Esta pantalla permite a cada usuario realizar su cambio de contraseña por seguridad. El sistema le pide que se ingrese la contraseña actual y que se confirme la nueva contraseña ingresada.

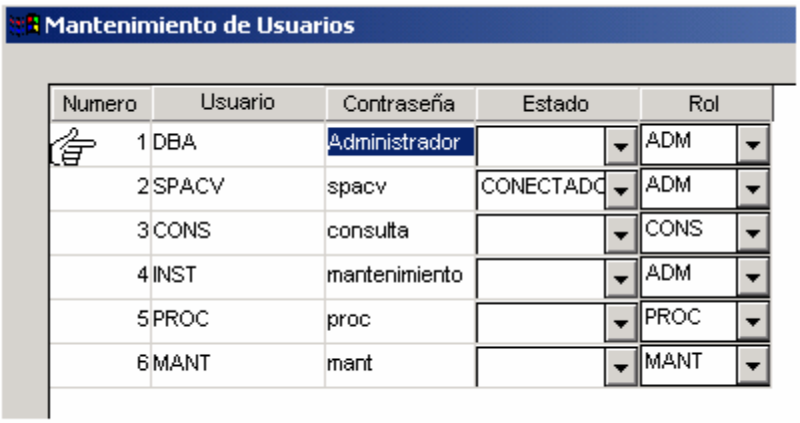


Cambiar contraseña

Después de hacer el cambio de contraseña, esta se hace efectiva en el momento en que se vuelve a ingresar al sistema. Si la contraseña es olvidada, únicamente el administrador puede conocerla mediante el acceso a la ventana de administrador de usuarios.

Pantalla de Administración de Usuarios.

El administrador de usuarios puede actualizar, insertar o eliminar un usuario del sistema. En cada registro puede consultar el estado en que se encuentra el usuario, es decir, puede verificar si un usuario se halla conectado o no al sistema. Puede también asignarle un "rol" en el sistema. El rol permite que el usuario tenga permiso de acceso a una o varias opciones del sistema.



administrar usuarios

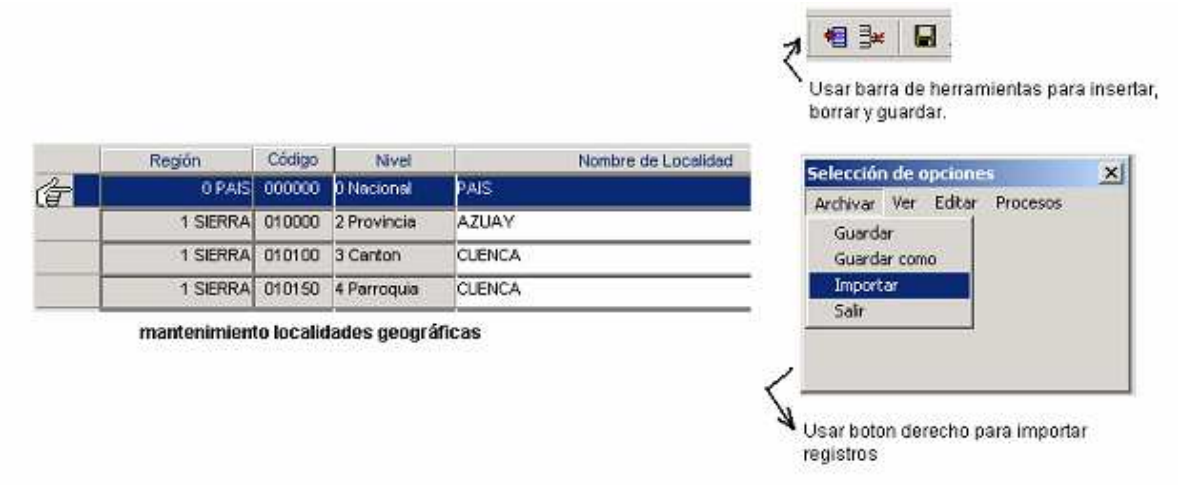
Los roles de usuarios básicamente son cuatro: Rol de Administrador: "ADM", el Rol de Proceso: "PROC", Consulta: "CONS", y de mantenimiento: "MANT".

El rol "ADM" tiene acceso a todas las opciones para consulta y modificación.

El rol "PROC" tiene acceso a la opción de proceso y otras, excepto la de datos.
El rol "CONS" tiene acceso a consultas en la opción de proceso.
El rol "MANT" tiene acceso a la ventana de mantenimiento del sistema.
A cada usuario se debe asignar un número de identificación y contraseña. Luego debe darse un rol escogido de la lista de selección. Todo esto se lo hace a través del icono de insertar de la barra de herramientas.

Pantalla de mantenimiento de Localidades Geográficas

En esta pantalla se ingresan los registros de localidades geográficas usadas para determinar los datos históricos de edificaciones, las series de indicadores y los casos de estudio o predicción.



La ventana muestra los registros de localidades ingresadas. Las localidades que aparecen fueron importadas por medio de la opción "Importar" del menú de despliegue de botón derecho al momento de la instalación. Corresponden a los registros de provincias, cantones y parroquias según la división territorial administrativa de la República del Ecuador, para el año base 2002. No requieren actualización, pero pueden editarse por medio de las herramientas.

Pantalla de mantenimiento de Tiempo (registro de años y meses)

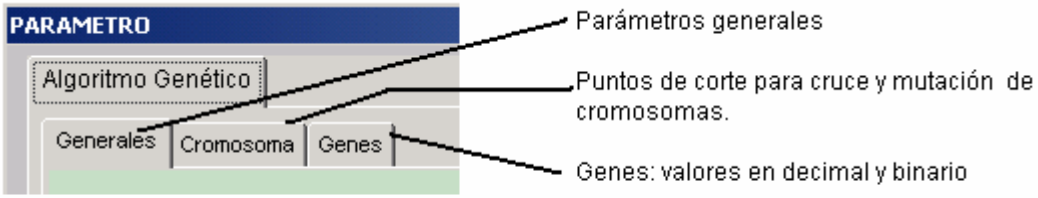
En esta pantalla se ingresan los registros de años y meses que van a servir para identificar a los datos históricos de edificaciones, series de indicadores y los casos de estudio o predicción.

mantenimiento de tiempos

La ventana muestra los registros de tiempos en años y meses ingresados. Los registros que aparecen fueron importados por medio de la opción "importar", del menú de despliegue del botón derecho al momento de la instalación. Estas identificaciones de tiempo corresponden a los años y meses a partir del año 1990. Estos datos no requieren modificación pero pueden ser administrados a través de los iconos de insertar, borrar y guardar de la barra de herramientas.

Pantalla de mantenimiento de Parámetros.

Esta pantalla fue creada para ingresar y actualizar los parámetros generales de las técnicas de inteligencia artificial empleadas. Para el caso de estudio del sistema se encuentran los parámetros generales del algoritmo genético para descubrimiento de reglas de asociación.

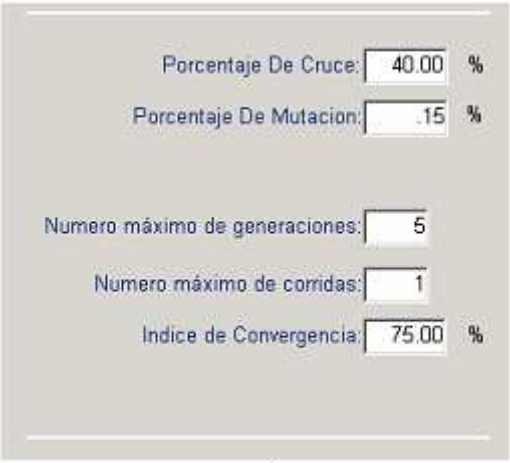


Actualizar parámetros

La ventana contiene una sola sección "Algoritmo Genético", la cual se divide en tres subsecciones: "Generales", "Cromosoma" y "Genes": Los valores que

aparecen son predeterminados y fueron ingresados al momento de la creación e instalación de la base de datos.

La subsección de "Parámetros Generales", permite ingresar o actualizar los parámetros de porcentaje de cruce, mutación, número de generación, de corridas y porcentaje de adaptabilidad de convergencia. Para activar la edición de los controles se requiere hacer uso del icono editar de la barra de herramientas, el cual debe ser pulsado dos veces por medio del botón principal del mouse. Luego de cada cambio debe darse un clic principal en el icono "guardar" de la barra de herramientas.



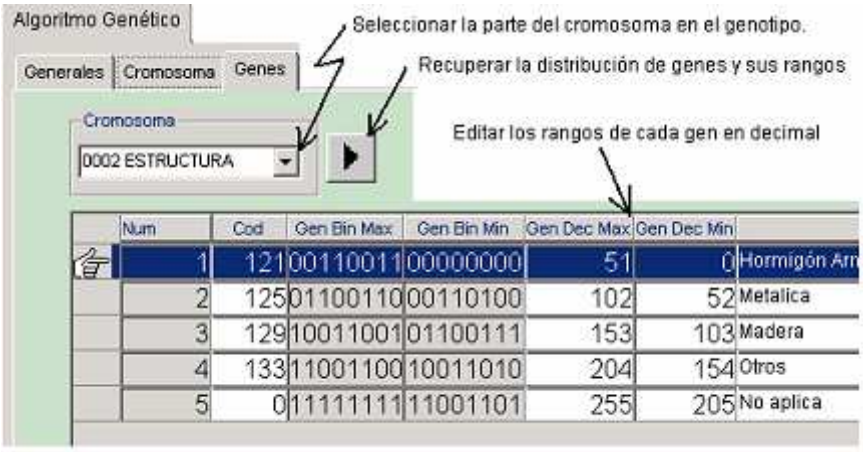
Actualizar parámetros caso

La subsección "cromosoma" muestra las partes en que se divide el antecedente de un "Cromosoma Regla" para la regla que será descubierta durante el proceso de predicción y al momento de aplicación del algoritmo genético al caso de estudio para extracción de reglas de asociación.

Cada registro se introduce al sistema al momento de la creación e instalación y permanecen fijos durante el uso del sistema.
Se muestran siete divisiones del cromosoma, cada uno relacionado con las variables investigadas: cimientos, estructura, paredes, cubierta o techo, tipo de obra, uso de la edificación y origen del financiamiento.

Las variables seleccionadas para las divisiones del cromosoma se ingresan y se actualizan al momento de la instalación, y es necesario hacer modificaciones.

La subsección genes permite dar nombres descriptivos a los rangos de valores representan a los genes para cada división del cromosoma. Las divisiones del cromosoma se escogen del cuadro de lista "cromosoma". Los rangos se asignan en decimal, y el sistema los interpreta en binario.

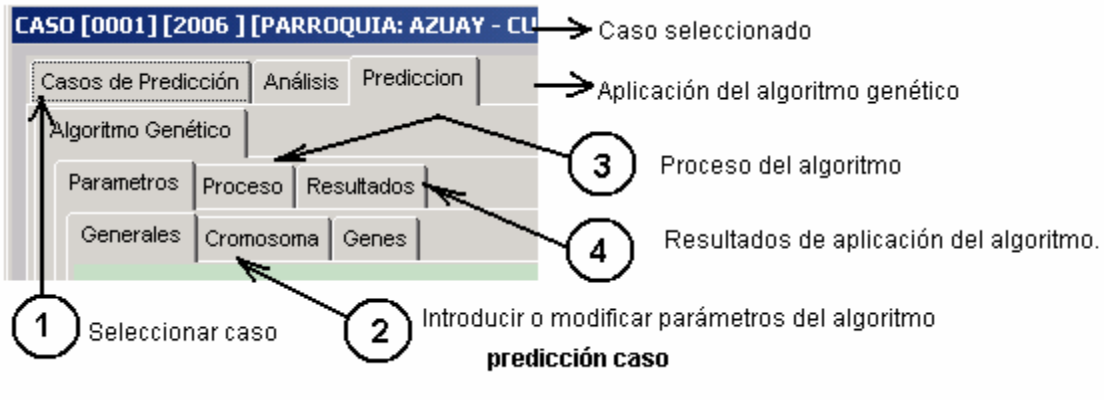


Cada rango permite reconocer a una característica de la variable investigada y puede ser definida en valores que van desde 0 hasta 255. De igual manera, se debe activar la edición de los registros mediante un clic principal del mouse en el icono de editar de la barra de herramientas. Debe volverse a presionar hasta que el registro se encuentre en edición. Para guardar los cambios se usa el icono de guardar de la barra de herramientas.

No se requiere insertar ni eliminar registros, ya que estos permanecen invariables durante la ejecución y se introducen una sola vez en el sistema.

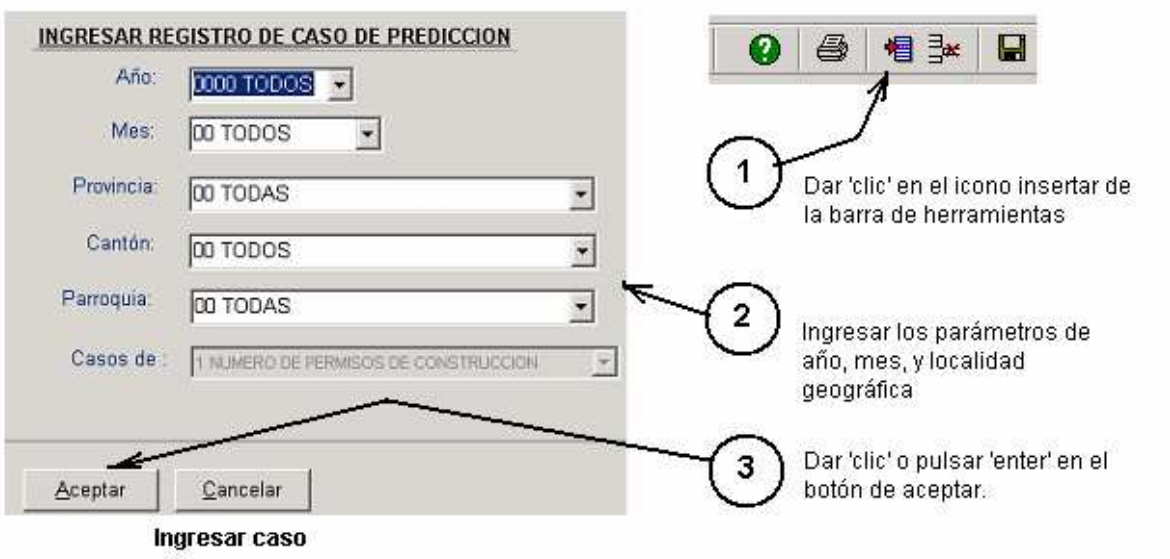
1.2.3.4.2. Módulo de Proceso.

La pantalla de proceso consiste en una ventana para introducir casos de estudio o predicción y para realizar procesos de análisis y predicción de estos casos desde sus secciones. La ventana de proceso se halla dividida en tres secciones: Casos de Predicción, Análisis y Predicción.



La sección Casos de Predicción.

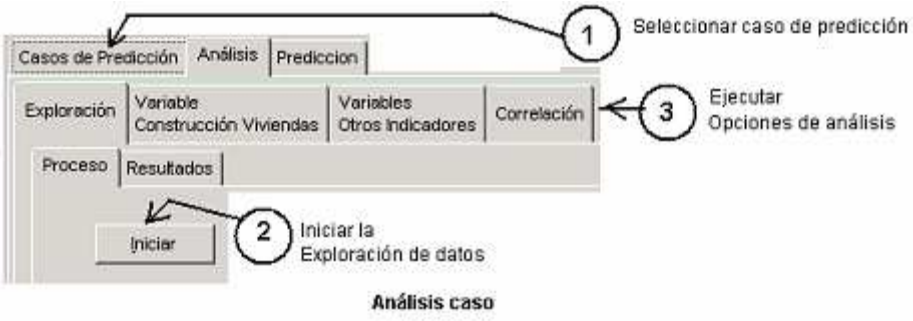
Esta sección permite introducir o seleccionar un caso de predicción al sistema. Se utilizan los iconos de insertar, borrar y guardar de la barra de herramientas. Para cada caso es preciso introducir el año, mes y localidad geográfica, en una ventana de introducción de datos que aparece luego de dar clic en el icono de insertar.



Cada registro del control de datos corresponde a un caso de predicción para una localidad y tiempo y debe ser seleccionado mediante un clic principal del mouse o mediante el uso de las teclas de subir y bajar.

La Sección Análisis

Una vez introducido un caso de predicción, se usa la ventana de análisis para hacer una exploración de los datos y un análisis causa y efecto entre la serie de número de permisos de construcción para el año de predicción y las variables de indicadores económicos. El procesamiento de esta ventana es opcional en el sistema y es meramente informativo.



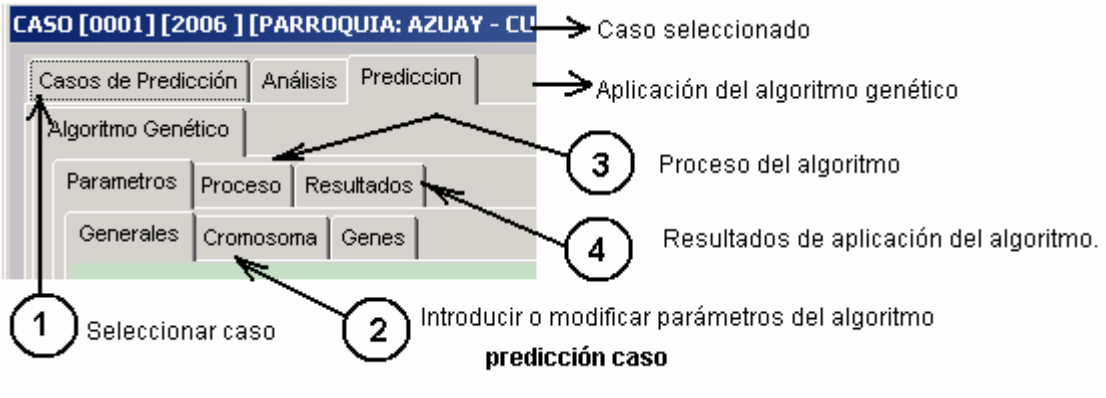
Esta sección se subdivide en otras secciones como: Exploración, Variables y Correlación. Para el proceso de análisis debe darse un clic principal en el botón "Iniciar" de la subsección "Exploración". Y en las otras ventanas se debe dar un clic en los botones de recuperación presentes en la parte superior del control de presentación central de datos, identificados con un triángulo negro.

Los datos para análisis son de visualización solamente.

La Sección Predicción.

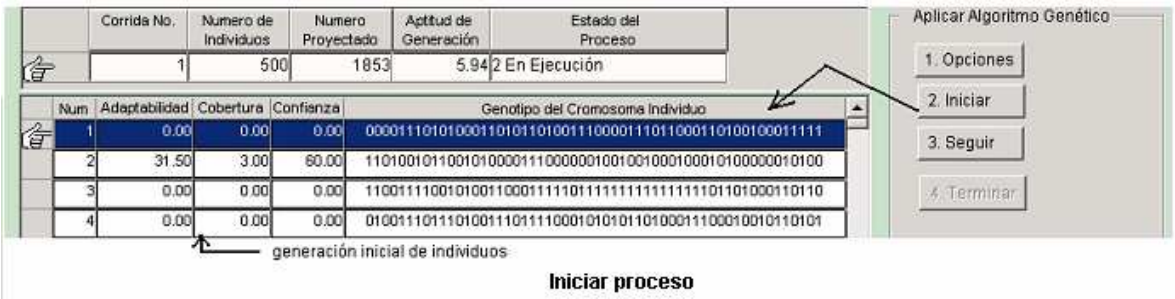
Esta sección es en la cual se realiza la aplicación del algoritmo genético por caso de estudio o predicción en la Sección "Casos de Predicción":

Esta sección contiene a su vez otra sección para aplicación de la técnica de algoritmo genético, la cual se subdivide en tres secciones principales: Parámetros. Proceso y Resultados.

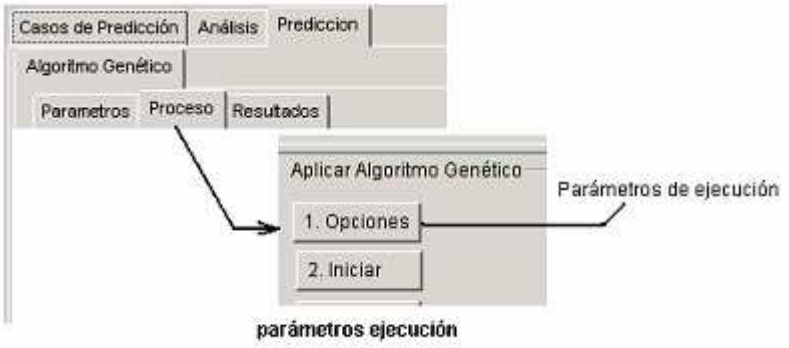


La primera sección es idéntica a la de Parámetros generales que fue explicada en el Módulo de datos y administración de usuarios. La segunda permite realizar el proceso y la tercera mostrar y evaluar los resultados En la subsección de proceso existen cinco botones: opciones, iniciar, seguir, terminar y guardar resultados.

El botón "iniciar" sirve para crear la primera generación de individuos del algoritmo genético y se activa con un clic principal del mouse. Al final se muestra un mensaje de concluido el proceso. Durante la ejecución se muestra una barra de avance en el centro de la ventana.



El botón "opciones" permite cambiar las opciones de ejecución del algoritmo elegir si se va a seguir paso a paso el proceso del algoritmo genético, y también permite introducir y seleccionar los tipos de reglas de asociación.



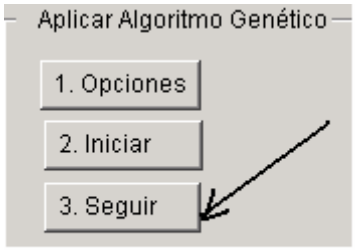
parámetros ejecución

Los tipos de reglas de selección aparecen en una ventana que se activa al momento de dar clic principal del mouse en el botón opciones. En el extremo derecho superior hay dos iconos para insertar y eliminar tipos de reglas.

parámetros reglas

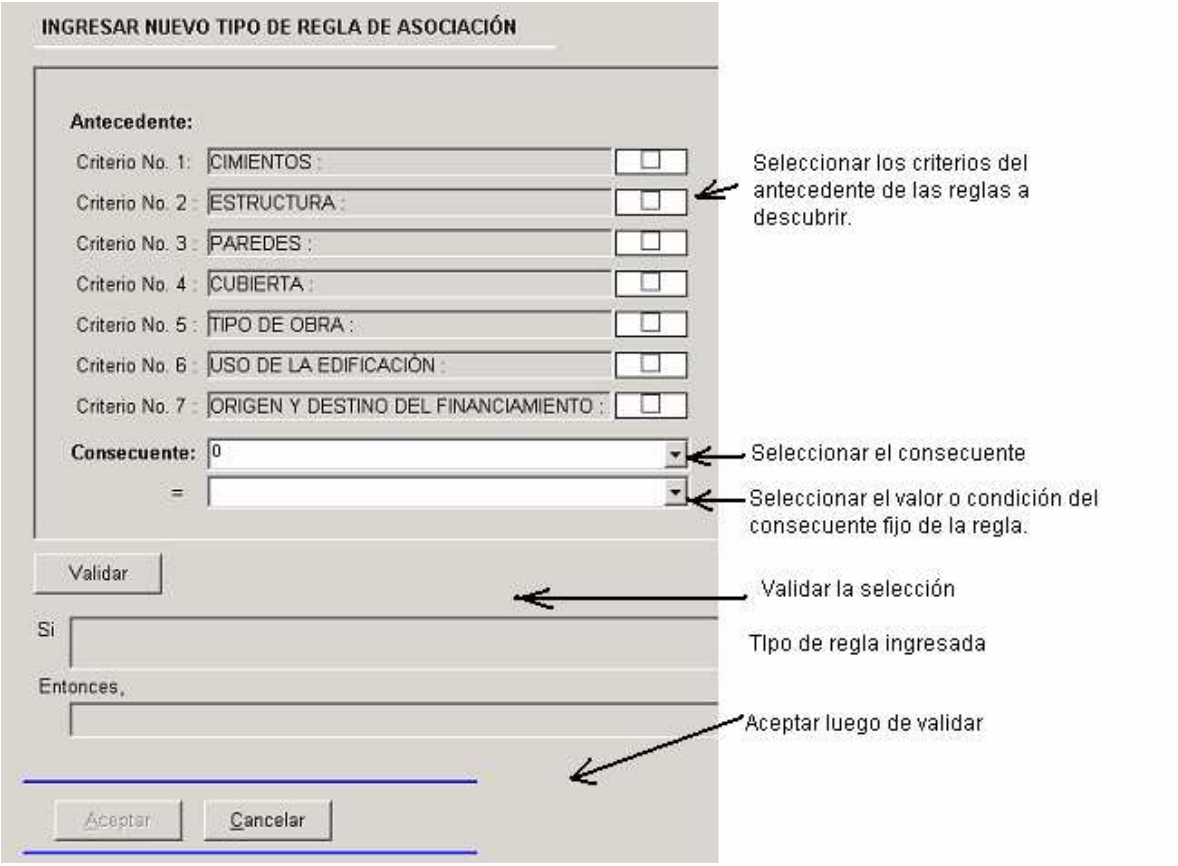
Al activar el icono insertar se muestra otra ventana con las opciones para introducir los antecedentes y consecuente de la regla y un botón aceptar para guardar y validar el tipo de regla introducido.

El botón "Seguir" permite continuar con el proceso o ir a la pantalla que muestra los resultados intermedios. Esta última ventana se activa siempre y cuando en la subsección de opciones se haya seleccionado Tipo de Proceso: Paso a Paso.

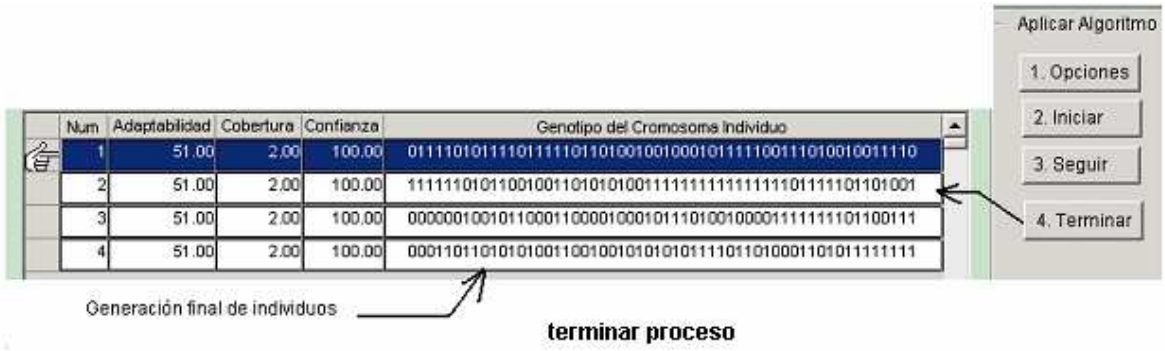


seguir proceso

Esta ventana tiene dos secciones: La primera tiene botones de flecha para continuar con la ejecución de un grupo de generaciones de individuos y la segunda sección permite evaluar la convergencia del algoritmo. Para salir debe darse clic principal del mouse en el botón: "Salir y Guardar".

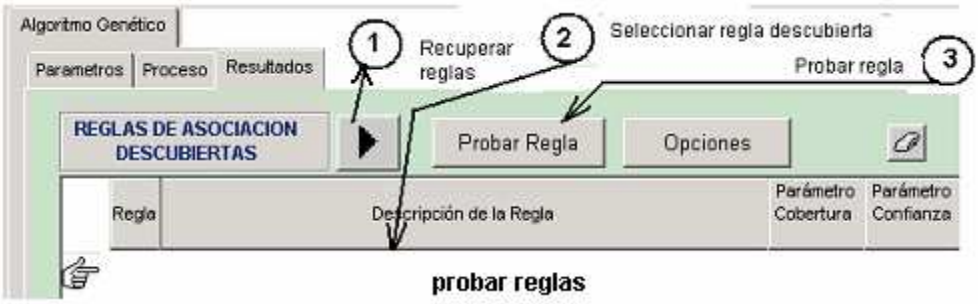


El botón "Terminar" se activa luego de haber concluido el proceso en el botón seguir. Debe presionarse inmediatamente mediante un clic principal del mouse para que el sistema termine el proceso.



El botón "Guardar resultados" es importante para que los datos resultantes de la última generación de individuos se conviertan en las reglas descubiertas y se guarden definitivamente para el caso de estudio seleccionado, y luego puedan ser recuperadas.

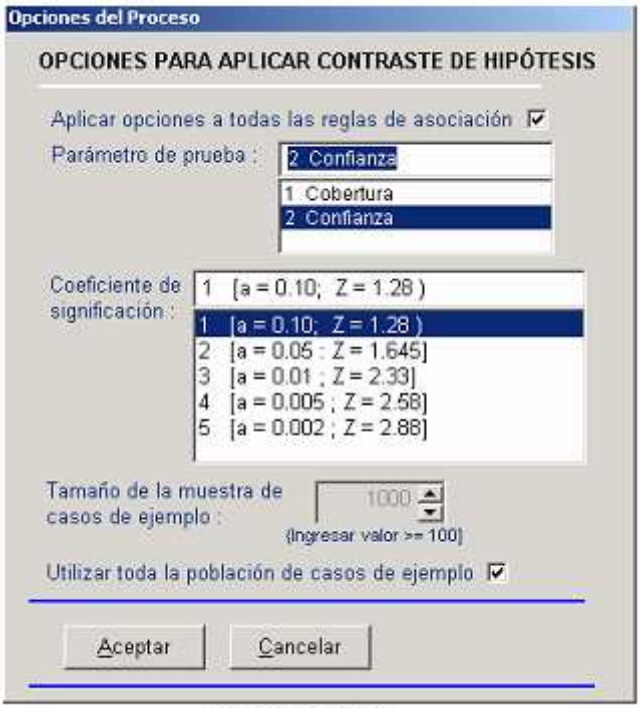
En la subsección Resultados, se encuentran dos botones y un cuadro de datos central con los registros de las reglas descubiertas para el tipo. El primer botón tiene el nombre de "Probar Regla" y permite probar la regla cuyo registro se ha seleccionado en el cuadro de datos. El segundo botón nombrado como "Opciones" permite introducir opciones para probar la regla mediante contraste de hipótesis.



Luego de activar el botón "probar regla", el sistema se demora unos instantes y luego le presenta un mensaje de resultados con la pregunta "Desea continuar": Si se pulsa el botón de "Si" por medio de un clic principal del mouse, entonces se presenta luego de otros instantes una segunda ventana de resultados con el mensaje de aceptación o rechazo de la hipótesis. Se debe dar clic principal en el botón "Aceptar" para volver a la ventana de resultados.

Para imprimir los resultados obtenidos de la aplicación de contraste de hipótesis a las reglas extraídas, se debe usar el icono de imprimir de la barra de herramientas. Al accionar este icono mediante un clic principal del mouse, se presenta otra ventana con la impresión previa del reporte y un botón superior para realizar la impresión. Luego de dar aceptación en este botón mediante otro clic del

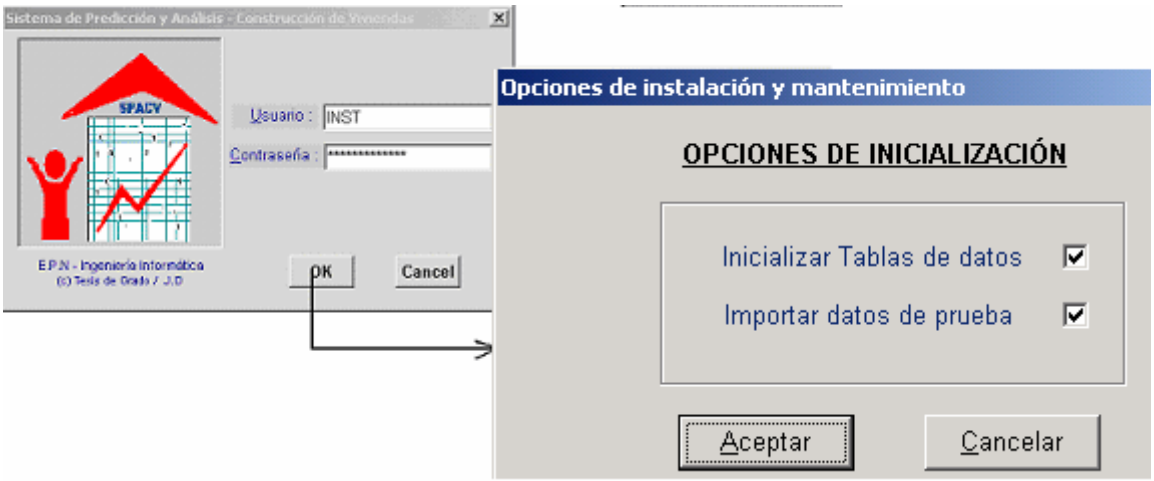
mouse, aparecerá la ventana de configuración de impresión, y en esta ventana se puede ir a imprimir directamente a impresora mediante el botón de "Ir".



opciones reglas

1.2.3.4.3. Acceso a la pantalla de instalación de datos de prueba y mantenimiento del sistema.

La ventana de mantenimiento del sistema sólo puede ser accedida mediante el usuario "INST" y su contraseña "mantenimiento".



Enseguida aparece la ventana de mantenimiento del sistema con sus controles. Esta opción soporta la posibilidad de crear nuevamente las tablas para volver a introducir los datos iniciales a través de la importación directa desde archivos de

extensión TXT, que se encuentran en el subdirectorio \DATOS del directorio del Sistema \SPACV.

La pantalla de mantenimiento solamente es útil para cargar los datos iniciales de prueba y cuando la base de datos se haya llegado a saturar con demasiada información, y se requiere preparar nuevamente el sistema para volver a comenzar desde el proceso de carga e importación.

Proceso de carga e importación de datos al sistema por primera vez luego de creación de la base de datos.

Si la base de datos contiene sus tablas vacías, ya sea por haber eliminado todos sus datos o por haber sido creada inicialmente, entonces se requiere un proceso de carga e importación de datos que debe seguir los siguientes pasos:

1. Importar los registros de tiempo desde el archivo "spacv\datos\dpa_tiempo.txt" mediante la opción "importar" del menú de despliegue del botón derecho del mouse en la ventana de mantenimiento de tiempo.
2. Importar los registros de localidades geográficas desde el archivo spacv\datos\dpa_localidad.txt mediante la opción "Importar" del menú de despliegue del botón derecho del mouse en la ventana de mantenimiento de localidades geográficas.
3. Importar los registros de variables de edificaciones desde el archivo spacv\datos\variables.txt mediante la opción "Importar" del menú de despliegue del botón derecho del mouse ubicado en el cuadro de datos de la sección "Variables" de la ventana de Edificaciones.
4. Importar los registros anuales de edificaciones desde los archivos spacv\datos\eed2000w.txt; eed2001w.txt; eed2002w.txt . . . eed2005w.txt. mediante el botón "Importar" de la sección "Selección" de la ventana "Edificaciones".
5. Importar los registros de Indicadores económicos desde el archivo de texto spacv\datos\indicador.txt mediante la opción "importar" del menú de

despliegue del botón derecho del mouse ubicado en el cuadro de datos de la ventana "Indicadores Económicos", activada por la opción Datos-Indicadores del menú principal.

6. Importar los registros de Series de Indicadores económicos desde el archivo de texto spacv\datos\serie_indicador.txt , mediante el botón "Importar" de la Sección "Selección" de la ventana "Series de Indicadores", activada por la opción Datos-Series del menú principal.

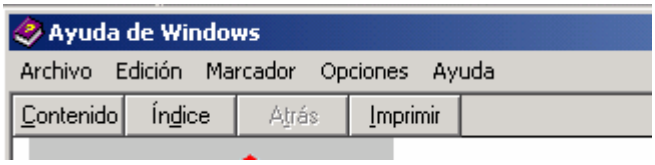
Para facilidad del proceso de carga e inicialización de datos, todos estos pasos se realizan automáticamente desde la ventana de instalación y mantenimiento del sistema que se ejecuta por acceso al sistema con el usuario "INST" y contraseña "mantenimiento".

1.2.3.4.4. Acceso a pantalla de ayuda.

Para acceder a la pantalla de ayuda en cualquier momento se puede presionar la tecla "F1", seleccionar la opción de "Ayuda" del menú principal o el icono de Signo de interrogación de la barra de herramientas del sistema.

La pantalla de ayuda contiene una sección de Índice y una sección de Búsqueda. La sección de índice contiene las frases y palabras claves que permiten acceder a la explicación de cada parte del sistema.

La ayuda se produce por efecto de lectura al archivo SPACV.HLP que se encuentra en el directorio de la aplicación. Existe además un resumen de ayuda del sistema que consiste en un árbol de opciones que aparecen en una ventana con un dos secciones, desde las cuales se puede navegar en el árbol conociendo las opciones del sistema. Esta ayuda lee los archivos de texto AyudaInd.txt y AyudaText.txt presentes en el directorio de la aplicación.



1.2.3.5. Ejecución de un caso de estudio.

Para ejecutarse un caso de estudio en el sistema debe tenerse presente los siguientes requerimientos:

1. Deben existir registros de tiempo y localidades geográficas, datos anuales de edificaciones y series de tiempo, variables de edificaciones.
2. Deben existir los registros de parámetros generales y sus valores.

Pasos para la ejecución de un caso de estudio.

1. Ingresar un nuevo caso de estudio o seleccionar uno existente a través de la ventana de Predicción de la opción "Procesos" del menú principal.
2. Seleccionar el caso ingresado e ir a la sección Proceso. Luego dar clic en el botón "Opciones" e ingresar tipo de regla de asociación desde esta ventana.
3. Volver y dar clic en el botón "Iniciar" y esperar hasta el mensaje del sistema.
4. Dar clic en el botón "Seguir" y esperar hasta que termine la ejecución.
5. Dar clic en el botón "Terminar" y esperar hasta que termine el proceso.
6. Dar clic en el botón "Guardar Resultados" y esperar el mensaje en la pantalla.
7. Ir a la sección resultados, y dar clic principal en el botón con flecha.
8. Dar clic en el botón de opciones desde la sección resultados y verificar o cambiar las opciones de prueba.
9. Seleccionar una regla descubierta por el algoritmo en el cuadro de datos central y luego dar clic en el botón de "Probar regla" desde la misma sección de resultados y esperar hasta que se muestre el mensaje con la pregunta "Desea continuar".
10. Dar clic en "Si" de la ventana con la pregunta y esperar hasta que se presente la ventana con el mensaje de aceptación o rechazo de la hipótesis para la regla seleccionada.
11. Luego de probar todas las reglas desde el paso 9 al 10, seleccionar el icono de impresión, dar clic principal del mouse y obtener el reporte final.