

UNIVERSIDAD TECNICA DE AMBATO



**FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E
INDUSTRIAL**

CENTRO DE ESTUDIOS DE POSGRADO

MAESTRÍA EN GESTIÓN DE BASES DE DATOS

Tema:

**EL DATA MINING Y SU INCIDENCIA EN LA TOMA DE DECISIONES
DEL CATASTRO DE ESTABLECIMIENTOS Y LA EMISIÓN DE LOS
PERMISOS DE FUNCIONAMIENTO POR PARTE DE LA DIRECCION
PROVINCIAL DE SALUD DE COTOPAXI.**

Trabajo de Investigación

Previa a la obtención del Grado Académico de Magister en Gestión de Bases de
Datos

Autor: Ing. Hernán Ignacio Carvajal Parra.

Director: Ing. Mg. Franklin Oswaldo Mayorga Mayorga.

Ambato – Ecuador
2012.

Al Consejo de Posgrado de la UTA

El tribunal receptor de la defensa del trabajo de investigación con el tema: **“EL DATA MINING Y SU INCIDENCIA EN LA TOMA DE DECISIONES DEL CATASTRO DE ESTABLECIMIENTOS Y LA EMISIÓN DE LOS PERMISOS DE FUNCIONAMIENTO POR PARTE DE LA DIRECCION PROVINCIAL DE SALUD DE COTOPAXI.”**, presentado por: el Ing. Hernán Ignacio Carvajal Parra y conformado por: Ing. Mg. Elsa Pilar Urrutia Urrutia, Ing. Mg. Marco Antonio Jurado Lozada, Ing. Mg. Víctor Manuel Pérez Rodríguez, Miembros del Tribunal, Ing. Mg. Franklin Oswaldo Mayorga Mayorga, Director del trabajo de investigación y presidido por: Ing. M.Sc. Oswaldo Paredes O. Presidente del Tribunal; Ing. Mg. Juan Garcés Chávez Director del CEPOS – UTA, una vez escuchada la defensa oral el Tribunal aprueba y remite el trabajo de investigación para uso y custodia en las Bibliotecas de la UTA.

Ing. Msc. Oswaldo Paredes O.
Presidente del Tribunal de Defensa

Ing. Mg. Juan Garcés Chávez
DIRECTOR CEPOS

Ing. Mg. Franklin Oswaldo Mayorga Mayorga
Director del Trabajo de Investigación

Ing. Mg. Elsa Pilar Urrutia Urrutia
Miembro del Tribunal

Ing. Mg. Marco Antonio Jurado Lozada
Miembro del Tribunal

Ing. Mg. Víctor Manuel Pérez Rodríguez
Miembro del Tribunal

AUTORÍA DE LA INVESTIGACIÓN

La responsabilidad de las opiniones, comentarios y críticas emitidas en el trabajo de investigación con el tema: “EL DATA MINING Y SU INCIDENCIA EN LA TOMA DE DECISIONES DEL CATASTRO DE ESTABLECIMIENTOS Y LA EMISIÓN DE LOS PERMISOS DE FUNCIONAMIENTO POR PARTE DE LA DIRECCION PROVINCIAL DE SALUD DE COTOPAXI.” nos corresponde exclusivamente a el Ing. Hernán Ignacio Carvajal Parra, Autor y de Ing. Mg. Franklin Oswaldo Mayorga Mayorga, Director del trabajo de investigación; y el patrimonio intelectual del mismo a la Universidad Técnica de Ambato

Ing. Hernán Ignacio Carvajal Parra
C.I. 0502256225

Ing. Mg. Franklin Mayorga Mayorga
C.I.1802503993

DERECHOS DEL AUTOR

Autorizo a la Universidad Técnica de Ambato, para que haga de este trabajo de investigación o parte de él un documento disponible para su lectura, consulta y procesos de investigación, según las normas de la Institución.

Cedo los Derechos de mi trabajo de investigación, con fines de difusión pública, además apruebo la reproducción de ésta, dentro de las regulaciones de la Universidad.

Ing. Hernán Ignacio Carvajal Parra

DEDICATORIA

Para mi esposa e hijo quienes día tras día han sabido brindarme el apoyo y comprensión necesarios para continuar estudiando y alcanzar nuevas metas.

Hernán Ignacio Carvajal Parra

AGRADECIMIENTO

Dedico este proyecto de tesis a Dios y a mis padres, porque han sido mi fortaleza para vencer los limitantes a lo largo del camino

Para aquellas personas que de una u otra manera estuvieron presentes con su conocimiento en el desarrollo de esta investigación.

Mi sincero agradecimiento al Ing. Mg. Franklin Mayorga, ya que con su acertada dirección se llevó a cabo el desarrollo de este trabajo de investigación.

Mi gratitud a la FISEI que impulsó la ejecución de este programa de Maestría en Gestión de Bases de Datos, del cual tuve el privilegio de ser parte de ella.

A todas las personas que de una u otra manera me han ayudado a cumplir esta meta.

EL AUTOR

ÍNDICE GENERAL

PORTADA.....	i
AUTORÍA DE LA INVESTIGACIÓN.....	ii
DERECHOS DEL AUTOR.....	iii
DEDICATORIA	iv
AGRADECIMIENTO	v
ÍNDICE GENERAL.....	vi
ÍNDICE DE CUADROS.....	ix
ÍNDICE DE GRÁFICOS	x
ÍNDICE DE ANEXOS.....	xii
RESUMEN.....	xiii
ABSTRACT.....	xiv
INTRODUCCIÓN	xv
CAPÍTULO I.....	1
EL PROBLEMA	1
Planteamiento del Problema.....	1
Contextualización.....	1
Árbol de Problemas.....	4
Análisis Crítico	5
Prognosis.....	5
Formulación del Problema	6
Interrogantes de la Investigación	6
Delimitación de la Investigación.....	6
Delimitación del contenido	6
Delimitación Espacial	7
Delimitación Temporal	7
Unidades de Observación.....	7
Justificación.....	7
Objetivos	8
Objetivo General	8
Objetivos Específicos:.....	9

CAPITULO II	10
MARCO TEÓRICO.....	10
Antecedentes de Investigación.....	10
Fundamentaciones.....	11
Fundamentación filosófica.....	11
Fundamentación Tecnológica.....	12
Fundamentación Legal.....	12
Categorías Fundamentales.....	14
Categorías de la Variable Independiente.....	15
Categorías de la Variable Dependiente.....	37
Hipótesis o Pregunta Directriz.....	44
Señalamiento de Variables.....	44
CAPITULO III.....	45
METODOLOGÍA.....	45
Enfoque.....	45
Modalidad de Investigación.....	45
Niveles o Tipos.....	46
Población y Muestra.....	47
Operacionalización de Variables.....	48
Variable Independiente:.....	48
Variable Dependiente:.....	49
Técnicas e Instrumentos.....	49
Plan para Recolección de la Información.....	49
Plan para el Procesamiento de la Información.....	51
CAPITULO IV.....	53
ANÁLISIS E INTERPRETACIÓN DE RESULTADOS.....	53
Verificación de la Hipótesis.....	61
CAPITULO V.....	65
CONCLUSIONES Y RECOMENDACIONES.....	65
Conclusiones.....	65
Recomendaciones.....	66
CAPITULO VI.....	67

LA PROPUESTA	67
Datos Informativos.....	67
Antecedentes de la Propuesta.....	67
Justificación	68
Objetivos	69
Objetivo General	69
Objetivos Específicos.....	69
Análisis de Factibilidad.....	70
Fundamentación	71
Metodología	71
Conclusiones y Recomendaciones	113
BIBLIOGRAFÍA	114
GLOSARIO DE TÉRMINOS.....	116
ANEXOS	119

INDICE DE CUADROS

Cuadro N. 1: Población y Muestra	48
Cuadro N. 2: Operacionalización de la Variable Independiente.....	49
Cuadro N. 3: Operacionalización de la Variable Dependiente	50
Cuadro N. 4: Plan de Recolección de Información.....	51
Cuadro N. 5:Análisis de resultados Pregunta 1.....	54
Cuadro N. 6:Análisis de resultados Pregunta 2.....	55
Cuadro N. 7:Análisis de resultados Pregunta 3.....	56
Cuadro N. 8:Análisis de resultados Pregunta 4.....	57
Cuadro N. 9:Análisis de resultados Pregunta 5.....	58
Cuadro N. 10:Análisis de resultados Pregunta 6.....	59
Cuadro N. 11:Análisis de resultados Pregunta 7.....	60
Cuadro N. 12:Análisis de resultados Pregunta 8.....	61
Cuadro N. 13:Frecuencias Observadas y esperadas.....	64
Cuadro N. 15:Calculo de Ji Cuadrado	64

ÍNDICE DE GRÁFICOS

Gráfico N. 1: Relación causa-efecto	6
Gráfico N. 2: Categorías fundamentales	16
Gráfico N.3: Componentes de una Base de Datos	18
Gráfico N.4: Subcategorías de la Variable Independiente	43
Gráfico N.5: Subcategorías de la Variable Dependiente	44
Gráfico N.6: Análisis de resultados Pregunta 1	54
Gráfico N.7: Análisis de resultados Pregunta 2	55
Gráfico N.8: Análisis de resultados Pregunta 3	56
Gráfico N.9 Análisis de resultados Pregunta 4	57
Gráfico N.10: Análisis de resultados Pregunta 5	58
Gráfico N.11: Análisis de resultados Pregunta 6	59
Gráfico N.12: Análisis de resultados Pregunta 7	60
Gráfico N.13: Análisis de resultados Pregunta 8	61
Gráfico N.14.: Curva de Ji Cuadrado para comprobación de hipótesis	65
Gráfico N.15: Analysis Services	79
Gráfico N.16: Orígenes de Datos	80
Gráfico N.17: Origen de Datos	80
Gráfico N.18: Destino de Datos	81
Gráfico N.19: Extracción de los datos	81
Gráfico N.20: Tarea ejecutar Sql	82
Gráfico N.21: Flujo de datos	82
Gráfico N.22: Origen OLE DB:	83
Gráfico N.23: Columna Derivada	83

Gráfico N.24: Destino OLE DB.....	84
Gráfico N.25: transformación de los datos	85
Gráfico N.26: Flujo de datos de transformación	85
Gráfico N.27: Origen OLE DB de transformación	86
Gráfico N.28: Dimensión de variación lenta.....	86
Gráfico N.29: Atributos variable.....	87
Gráfico N.30: Atributos y tipos de variables	87
Gráfico N.31: compatibilidad con miembros deducidos.....	88
Gráfico N.32: Finalizar asistente de variación lenta	88
Gráfico N.33: Complemento minería de datos	94
Gráfico N.34: Desde SQL server	94
Gráfico N.35: Asistente conexión de datos.....	95
Gráfico N.36: Conexión instancia del servidor.....	95
Gráfico N.37: Selección de la vista o tabla	96
Gráfico N.38: Guardar archivos de conexión de datos	96
Gráfico N.39: Importar datos a Excel	97
Gráfico N.40: Datos en Excel	97
Gráfico N.41: Clasificar	98
Gráfico N.42: Asistente clasificación de datos	98
Gráfico N.43: Selección Datos de origen.....	99
Gráfico N.44: Selección columnas de predicción:	99
Gráfico N.45: Finalizar asistente de clasificación.....	100
Gráfico N.46: Procesando información de predicción.....	100
Gráfico N.47: Escenario 1.....	101
Gráfico N.48: Escenario 1.1	101
Gráfico N.49: Escenario 1.1.1	102
Gráfico N.50: Escenario Total	102
Gráfico N.51: Selección campos Escenario 2	103
Gráfico N.52: Escenario 2 Total	103

INDICE DE ANEXOS

ANEXO 1: instrumentos para la encuesta	110
ANEXO 2: Distribución Ji cuadrado χ^2	111
ANEXO 3: Manual Operativo	115

UNIVERSIDAD TÉCNICA DE AMBATO
FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E
INDUSTRIAL
MAESTRÍA EN GESTIÓN DE BASES DE DATOS

EL DATA MINING Y SU INCIDENCIA EN LA TOMA DE DECISIONES DEL
CATASTRO DE ESTABLECIMIENTOS Y LA EMISIÓN DE LOS PERMISOS
DE FUNCIONAMIENTO POR PARTE DE LA DIRECCION PROVINCIAL DE
SALUD DE COTOPAXI.

Autor: Ing. Hernán Ignacio Carvajal Parra.

Director: Ing. Mg. Franklin Oswaldo Mayorga Mayorga.

Fecha: 20 Septiembre del 2012

RESUMEN

La investigación sobre “El Data MINING y su incidencia en la toma de decisiones del Catastro de establecimientos sujetos a Vigilancia Sanitaria y los Permisos de Funcionamiento emitidos por la Dirección Provincial de Salud de Cotopaxi.”, tiene como objetivo general Implementar el Data MINING para la toma de decisiones del Catastro de establecimientos y la emisión de los Permisos de Funcionamiento por parte de la Dirección Provincial de Salud de Cotopaxi.

El problema radica en la dificultad de la entrega oportuna y a tiempo de información gerencial del catastro de establecimientos para la correcta toma de decisiones, esto se da debido a que los tiempos de respuesta para obtener información son muy largos por el gran volumen de datos que se maneja, ya que los mismos se encuentran dispersos y en base de datos independientes y así eventualmente predecir procesos para realizar una buena toma de decisiones.

Con esta investigación se plantea una solución a estos problemas dando respuestas y prioridades a la emisión de permisos de funcionamiento, en tiempo real a cada uno de los usuarios de la Dirección Provincial de Salud de Cotopaxi

Descriptores: Almacén de datos, Datawarehouse, patrones de datos, comportamiento de datos, ETLs, Data Mining, Toma de decisiones.

TECHNICAL UNIVERSITY OF AMBATO
FACULTY OF ENGINEERING SYSTEMS, ELECTRONICS
AND INDUSTRIAL
MASTER OF DATABASE MANAGEMENT

THE DATAMINING AND ITS INCIDENCE TO TAKE DECISIONS OF REGISTER OF ESTABLISHMENTS AND THE EMISION OF PERFORMANCE PERMISSIONS PROVIDED BY THE HEALTH PROVINCIAL LEADERSHIP OF COTOPAXI

Author: Engineer Hernán Ignacio Carvajal Parra.

Director: Engineer Mg. Franklin Oswaldo Mayorga Mayorga.

Date: September 20, 2012

ABSTRACT

The research of “Data MINING and its incidence to take decisions of register of establishments under sanitary vigilance and the performance permissions provided by the Health Provincial Leadership of Cotopaxi, it has as general objective To implement the Data MINING to take decisions in the registration of establishments and the emission of performance permission by the Health Provincial Leadership of Cotopaxi.

The biggest difficult is the delivery on time of the general information of the establishments registration to take the best decisions, it is produced because the answer time to get information is very long due to the large volume of data are handled, these ones are scattered and according the independent data and predict processes to get good decisions.

With this investigating work a solution is proposed to these troubles giving answers and priorities to the emission of performance permission, in a real time for each user of Health Provincial Leadership of Cotopaxi

Descriptors: Datastore, Datawarehouse, Data patterns, Behavior of data, ETLs, Data Mining, Decision Making.

INTRODUCCION

Hoy en día, en un mundo globalizado, de alta incertidumbre y competitivo, la gestión de la información se convierte en una forma de marcar la diferencia y hacer ventaja competitiva. En este sentido, simples formatos y registros son calificados como herramientas básicas de recopilación de información, en especial de necesidades de clientes, de quejas, reclamos e incluso de nuevos servicios solicitados. Esto ayuda también a la incorporación de factores de innovación en las empresas

En el marco actual de las instituciones y empresas, la confluencia de nuevas infraestructuras de comunicación con potentes y flexibles herramientas de tratamiento de información (bases de datos, Data Warehouse, Data Mining, etc) mejoran la calidad, cantidad y eficiencia de los datos, así como el análisis, procesamiento y comunicación de los mismos. En otras palabras, pueden aportar a las instituciones la base tecnológica necesaria para afrontar los nuevos retos de la situación actual y las perspectivas de futuro de la gestión.

El trabajo de investigación está estructurado por capítulos de la siguiente manera:

El capítulo I denominado El Problema contiene: La Contextualización donde se hace un exhaustivo análisis de las causas principales que originan el problema, árbol de problemas, análisis crítico, prognosis, formulación del problema, interrogantes de la investigación, delimitación del objetivo de investigación, justificación, objetivo general, objetivos específicos.

El capítulo II llamado Marco Teórico se estructura con los antecedentes investigativos, las fundamentaciones filosófica, tecnológica, legal, red de

inclusiones contextuales, constelaciones de ideas, hipótesis / pregunta directriz con sus respectivas variables

El capítulo III llamado Metodología contiene: Modalidades de investigación, niveles o tipos de investigación, población y muestra, Operacionalización de variables, Plan de recolección de la información, Las formas, métodos, técnicas y medios que han sido utilizados para obtener la información necesaria acerca de los problemas de la entrega de permisos y catastros

El capítulo IV llamado análisis e interpretación de resultados contiene: El análisis e interpretación de los resultados de las encuestas realizadas a los ingenieros administradores de los enlaces de datos así como también a los administradores de las aplicaciones propias de cada empresa. También se presenta la demostración de la hipótesis basándose en cálculos estadísticos.

El capítulo V llamado conclusiones y recomendaciones contiene: Las conclusiones y recomendaciones de la investigación acerca del problema planteado.

El capítulo VI llamado propuesta contiene: Toda la información relacionada al Data Mining de emisión de permisos y catastros y así minimizar el tiempo de respuesta de información a nivel gerencial. Incluye además información de su funcionamiento y pruebas de operación del sistema. Finalmente se encuentra el glosario de términos, la Bibliografía y los respectivos Anexos

CAPITULO I

EL PROBLEMA DE INVESTIGACION

1.1.- Tema:

El Data MINING y su incidencia en la toma de decisiones del Catastro de establecimientos sujetos a Vigilancia Sanitaria y los Permisos de Funcionamiento emitidos por la Dirección Provincial de Salud de Cotopaxi.

1.2.- Planteamiento del problema

1.2.1.-Contextualización

Al convertirse la información en un elemento esencial los métodos de control y recuperación están cambiando y facilitando el acceso a ella como consecuencia de las innovaciones tecnológicas.

Los avances más espectaculares se están produciendo en el campo de las ciencias aplicadas, el sistema educativo, administración de datos en empresas a nivel público y privado y en consecuencia ha ido cambiando todo este proceso ante un mundo que exige competitividad y eficiencia.

Anteriormente era todo un dilema controlar y administrar de manera óptima la información de las empresas y las que lograban hacerlo era gracias a las horas extra de trabajo que realizaba su personal.

Las bases de datos, hoy en día, ocupan un lugar determinante en cualquier área del quehacer humano, comercial, y tecnológico. No sólo las personas involucradas en el área de Informática, sino todas las personas administrativas, técnicas y con

mayor razón los profesionales de cualquier carrera, deben tener los conocimientos necesarios para poder utilizar las bases de datos.

El actual gobierno del Ecuador ante el acelerado avance e influencia de tecnología ha creado la Subsecretaria de Tecnologías de la Información con la misión de “Mejorar la gestión del gobierno mediante la estandarización, regulación, control, integración y ejecución de los proyectos informáticos de las entidades del gobierno central, y coordinar acciones en este campo en las demás instituciones del sector público”.

El Ecuador no es un país que se ha excluido de esta revolución tecnológica, y es por ello que cada Institución ha sentido la necesidad de capacitar a su personal, responsable de la administración de la información, con el claro objetivo de optimizarla y agilitarla para la toma correcta de decisiones.

Actualmente la mayor parte de las organizaciones (empresas e instituciones gubernamentales) implementan sistemas gestores de bases de datos, gracias a su relativa facilidad para llevar el control y al mismo tiempo mejorar la seguridad de la información de millones de personas, así como de operaciones que requieren grandes niveles de seguridad.

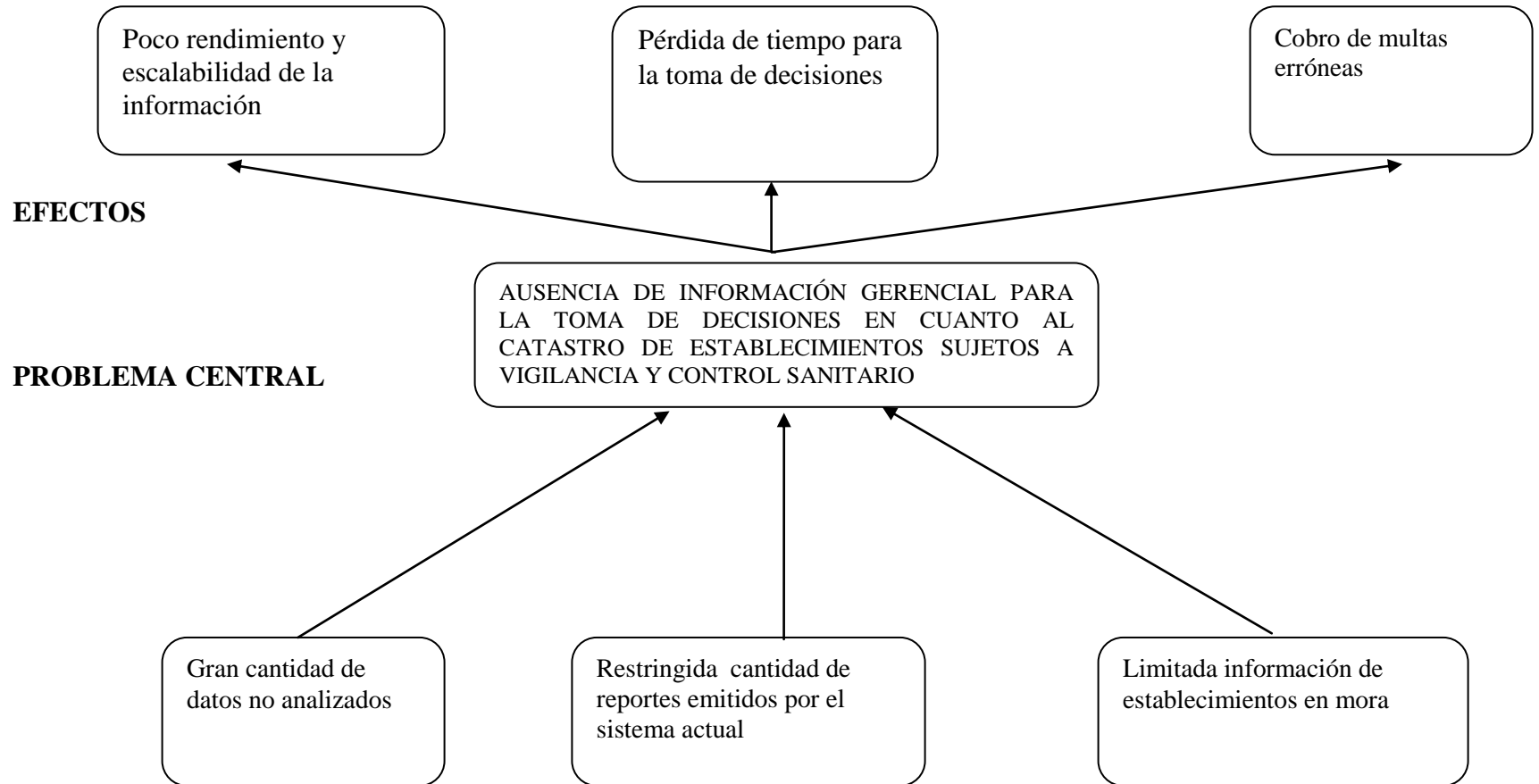
Se puede decir que muchas empresas almacenan un cierto tipo de datos al que se denomina dato-escritura, ya que sólo se guarda (o escribe) en el disco duro, pero nunca se hace uso de él. Generalmente, todas las empresas usan un dato llamado dato-escritura-lectura, que utilizan para hacer consultas dirigidas. Un nuevo tipo de dato al cual se lo denomina dato-escritura-lectura-análisis es el que proporciona en conjunto un verdadero conocimiento y apoya en las tomas de decisiones. Es necesario contar con tecnologías como es el llamado Data Mining que ayudará a explotar el potencial de este tipo de datos.

Por todos estos aspectos señalados surgió la necesidad de implementar la tecnología Data Mining en el Departamento de Vigilancia Sanitaria de la

Dirección Provincial de Salud de Cotopaxi donde se cuenta con una aplicación que lleva el catastro de establecimientos y el proceso de emisión de permisos de funcionamiento.

La capacidad de este departamento para generar y almacenar datos ha crecido en los últimos años a velocidades considerables. En contrapartida, su capacidad para procesarlos y utilizarlos no ha ido a la par. Por este motivo, el *Data Mining* se presenta como una tecnología de apoyo para explorar, analizar, comprender y aplicar el conocimiento obtenido usando grandes volúmenes de datos.

Árbol de Problemas



CAUSAS

Gráfico N 1.....: Relación causa-efecto
Elaborado por: Investigador

1.2.2.- Análisis crítico:

El presente trabajo investigativo parte de un análisis profundo de las debilidades existentes dentro del Departamento de Vigilancia Sanitaria de la Dirección Provincial de Salud de Cotopaxi, ya que se viene aplicando un software que posee un número limitado de reportes lo cual evidencia el poco tratamiento que se da a los datos almacenados en esta aplicación.

Un pilar fundamental en este proceso de administración de información es el rol que debe cumplir tanto el director y el coordinador de esta institución ya que no se realizó un correcto análisis para la implementación del software y los resultados que se obtendrían, una vez que el sistema cuente con la información real.

Otra problemática a resaltar es la falta de presupuesto en la institución para la implementación de un proyecto Data Mining, lo que ocasiona que existan datos que no son analizados de manera óptima para generar información que sea valedera en la toma de decisiones.

En cuanto al no contar con la información adecuada produciría graves problemas legales con los propietarios de los establecimientos en mora que serían sancionados a través de multas y/o clausura de sus establecimientos.

Además el personal que maneja el programa no posee un adecuado conocimiento en informática, siendo esta una gran falencia que retrasa el adelanto institucional.

Resultado de ello urge una solución inmediata a este problema que concierne a todos los involucrados.

1.2.3.- Prognosis:

Al no dar solución a lo anteriormente mencionado se continuaría observando la falta de información útil reflejada en la ausencia de reportes gerenciales para prevenir problemas y disconformidad de sus usuarios.

De no contar con la información apropiada las autoridades institucionales no podrán emitir informes confiables a los diferentes medios de comunicación, además se podría perjudicar tanto a usuarios como a la institución.

1.2.4.- Formulación del Problema:

¿Cómo incide el Data Mining en la toma de decisiones del Catastro de establecimientos sujetos a Vigilancia Sanitaria y los Permisos de Funcionamiento emitidos por la Dirección Provincial de Salud de Cotopaxi?

1.2.5.- Interrogantes de la Investigación:

- ¿Cómo se realiza actualmente la toma de decisiones del Catastro de establecimientos sujetos a Vigilancia Sanitaria y los Permisos de Funcionamiento emitidos por la Dirección Provincial de Salud de Cotopaxi?
- ¿La implementación de un Data MINING es factible y adecuado en Dirección Provincial de Salud de Cotopaxi.?
- ¿Existen alternativas de solución al problema de ausencia de información gerencial para la toma de decisiones en cuanto al catastro de establecimientos sujetos a vigilancia y control sanitario?

1.2.6.-Delimitación del Objeto de Investigación:

Un problema debidamente delimitado, facilita la viabilidad y profundización del mismo. Operativamente, se plantea:

Delimitación de contenido:

Campo: Sistemas
Área: Gestión de Base de Datos
Aspecto: Data MINING

Delimitación Espacial:

La investigación se desarrolló en los espacios de la Dirección Provincial de Salud de Cotopaxi.

Delimitación Temporal:

El presente trabajo de investigación se realizó durante el período comprendido entre Diciembre del 2011 y Julio del 2012.

Unidades de Observación:

- Coordinador de Vigilancia Sanitaria
- Personal de Vigilancia Sanitaria
- Inspectores Sanitarios
- Personal de Sistemas

1.3.- Justificación

La Revolución Tecnológica actual que el mundo globalizado y competente atraviesa obliga a que todos los seres humanos nos involucremos en este proceso de cambio a través de soluciones inmediatas a problemas de índole tecnológica, social, médica etc.

La presente investigación es **importante** porque provee a la Dirección Provincial de Salud de Cotopaxi una apertura hacia la implementación de Data MINING, y la aplicación de la informática en la toma de decisiones del Catastro de establecimientos sujetos a Vigilancia Sanitaria y los Permisos de Funcionamiento.

Es **útil** porque la toma de decisiones se la realiza bajo los lineamientos de la aplicación a proponerse, la misma que está normada de acuerdo a la base legal del Ministerio de Salud Pública y tiene un efecto inmediato para la obtención de datos, por cuanto la información obtenida será de relevante importancia.

El presente trabajo investigativo **beneficia** a la Dirección Provincial de Salud de Cotopaxi y a los propietarios de establecimientos calificados porque tienen la

confianza que la información de sus datos y transacciones realizadas son confiables. Por otro lado las autoridades de salud tienen a la mano una herramienta muy útil, versátil, moderna que facilitará la recopilación y análisis de la información para la toma de decisiones.

Causa alto **impacto**, porque da una profunda transformación en los procedimientos de manejo de datos e información en el Departamento de Vigilancia Sanitaria de la Dirección Provincial de Salud de Cotopaxi y sus Áreas a nivel de la provincia, esto a su vez puede ser tomado como referencia para otras direcciones provinciales de salud del país.

Esta investigación es **factible** por cuanto existe el interés de los involucrados en el proceso administrativo como son: el Director de salud, Coordinador y personal del Departamento de Vigilancia Sanitaria de la Dirección de Salud de Cotopaxi.

Finalmente el presente estudio proporcionará y orientará en la toma de decisiones del proceso, de manera que se garantice el mejoramiento de la emisión de reportes y administración de datos, y en consecuencia de la comunidad en general.

1.4.- Objetivos

1.4.1.- Objetivo General

- Diagnosticar las herramientas del Data Mining para mejorar la toma de decisiones del Catastro de establecimientos y la emisión de los Permisos de Funcionamiento por parte de la Dirección Provincial de Salud de Cotopaxi.

1.4.2.- Objetivos Específicos

- Analizar y establecer cómo se realiza actualmente la toma de decisiones del Catastro de establecimientos y la emisión de los Permisos de Funcionamiento.
- Estudiar las herramientas, tecnologías que engloban la creación de un Data Mining, estudiando los elementos que la conforman.
- Diseñar un Data Mining como alternativa de solución al problema de ausencia de información gerencial para la toma de decisiones en cuanto al catastro de establecimientos sujetos a vigilancia y control sanitario

CAPITULO II

MARCO TEORICO

2.1.- Antecedentes Investigativo

Realizado un recorrido por las principales bibliotecas de las universidades que ofertan las carreras de ingeniería en sistemas se encuentra que: en la Universidad Católica del Perú, existe una tesis cuyo tema es “Análisis, Diseño e Implementación de un Data Mining de Soporte de Decisiones para un Hospital del Sistema de Salud Público”, realizado por Álvaro Villanueva Ojeda (2008).

Su principal conclusión es: El trabajo de tesis presenta una solución que los hospitales pueden implementar para satisfacer sus necesidades de gestión, análisis y toma de decisiones. Otorga un panorama de lo que está sucediendo en el hospital y presenta esta información en línea.

Cabe resaltar que se investigó en la biblioteca de la facultad de Ingeniería en Sistemas, Electrónica e Industrial de la Universidad Técnica de Ambato y hasta la presente fecha, existe una tesis cuyos temas es: “Business Intelligence para manejo y recuperación de la empresa FERRIARCOS” , realizado por Carlos Nuñez (2012).

Cuya conclusión principal es que una solución efectiva a los problemas de recuperación de cartera es un sistema de inteligencia del negocio; el cual posee todas las características necesarias para administrar y mejorar los procesos de la

empresa. Un sistema de Datamining reducirá el tiempo en la toma de decisiones a nivel gerencial, para una mejor forma de recuperación de cartera y así tomar decisiones de manera oportuna.

Además existe otra tesis con el tema: “Estudio comparativo de herramientas BI para su aplicación en las Estadísticas del Censo Nacional Económico 2010 del Instituto Nacional de Estadísticas y Censos”, realizado por Juan Córdova Manjarrez (2012). Cuya conclusión principal es que un Business Objects EDGE31 de SAP en la fase de reporte BI, es la herramienta más adecuada para su aplicación en las estadísticas del Censo Nacional Económico 2010 del Instituto Nacional de Estadísticas y Censos

Fundamentaciones

2.2.- Fundamentación Filosófica

Para realizar el presente trabajo el investigador se ubica en el paradigma filosófico crítico - propositivo porque cuestiona la manera de hacer investigación y por el contrario plantea una propuesta de solución al problema investigado basado en la existencia de múltiples realidades socialmente construidas.

La palabra base de datos puede usarse para describir, contar, procesar, revelar y almacenar un conjunto de informaciones con los cuales se relaciona el hombre en su quehacer diario, entre ellas, una lista telefónica, un conjunto de reportes sobre los libros que se usan en una biblioteca escolar, los datos de profesores y estudiantes que se encuentran en una determinada escuela y el expediente de un paciente hospitalizado.

Sin embargo, aunque todos los seres humanos pudieran estar de acuerdo en que se necesita del control automático de los datos en todos los órdenes de la sociedad, probablemente la gran mayoría pudiera ignorar lo que es en realidad una base de datos.

El diccionario de la Real Academia Española indica que es "un conjunto de datos organizado de tal manera que permita obtener con rapidez diversos tipos de información" (1)

Matos Rosa María (2001) la define como **"conjunto de datos interrelacionados entre sí, almacenados con carácter más o menos permanente en la computadora. O sea, que una Base de Datos puede considerarse una colección de datos variables en el tiempo"**

2.3 Fundamentación Tecnológica

Tecnología es el conjunto de conocimientos técnicos, ordenados científicamente, que permiten construir objetos y máquinas para adaptar el medio y satisfacer las necesidades de las personas. Es una palabra de origen griego, τεχνολογία, formada por téchnē (τέχνη, "arte, técnica u oficio") y logía (λογία), el estudio de algo.

Aunque hay muchas tecnologías muy diferentes entre sí, es frecuente usar el término en singular para referirse a una de ellas o al conjunto de todas. Cuando se lo escribe con mayúscula, tecnología puede referirse tanto a la disciplina teórica que estudia los saberes comunes a todas las tecnologías, como a educación tecnológica, la disciplina escolar abocada a la familiarización con las tecnologías más importantes.

2.4.- Fundamentación Legal

La investigación tendrá su base legal en el Reglamento para otorgar permisos de funcionamiento a los establecimientos sujetos a vigilancia y control sanitario:

Según Ley Orgánica de Salud: "Art. 7.- A las Direcciones Provinciales de Salud les corresponde otorgar el permiso de funcionamiento anual, para lo cual el interesado deberá presentar una solicitud dirigida al Director Provincial de Salud de la jurisdicción a la que pertenece el domicilio del establecimiento..."

-Reglamento de control y funcionamiento de los establecimientos farmacéuticos:
“Art. 5.- La Dirección Provincial de Salud concederá permiso para la instalación de farmacias privadas cuando se justifique su necesidad, por el incremento de la densidad poblacional en un sector determinado y el número de establecimientos similares que hubieren al momento de la solicitud del permiso”

-Ley de propiedad intelectual del Ecuador.

“Art. 1.-El Estado reconoce, regula y garantiza la propiedad intelectual adquirida de conformidad con la ley, las decisiones de la Comisión de la Comunidad Andina y los convenios internacionales vigentes en el Ecuador”

Además el presente trabajo de grado se fundamenta en el Reglamento del Centro de Estudios de Posgrado (Cepos), el mismo que en sus Artículos esenciales indica:

“Art. 75 Para optar por el Título de Diploma Superior, Especialista o el grado académico de Magister, los estudiantes tendrán que realizar un trabajo de investigación para cada caso y sustentarlo ante un tribunal”.

2.5.- Categorías fundamentales

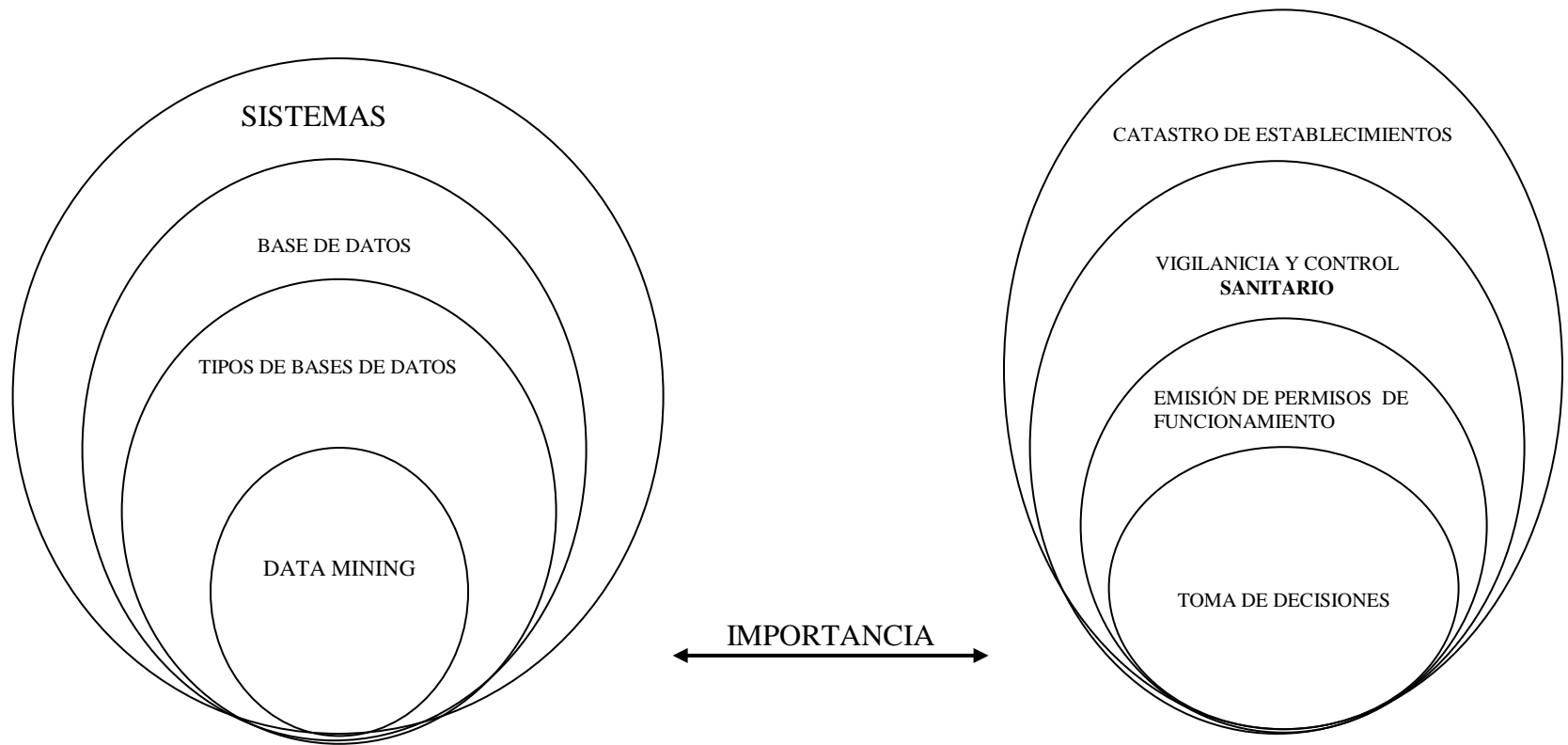


Gráfico N. 2: Categorías Fundamentales

Elaborado por: Investigador

Categorías de la Variable Independiente

Sistemas

Un **sistema**, es el conjunto de partes interrelacionadas, hardware, software y de recurso humano (humanware) que permite almacenar y procesar información. El hardware incluye computadoras o cualquier tipo de dispositivo electrónico inteligente, que consisten en procesadores, memoria, sistemas de almacenamiento externo, etc. El software incluye al sistema operativo, firmware y aplicaciones, siendo especialmente importante los sistemas de gestión de bases de datos. Por último el soporte humano incluye al personal técnico que crean y mantienen el sistema (analistas, programadores, operarios, etc.) y a los usuarios que lo utilizan.

Clasificación de los Sistemas Informáticos

Los Sistemas Informáticos pueden clasificarse en base a varios criterios como son:

Por su uso pueden ser:

- De uso general.
- De uso específico.

Por el paralelismo de los procesadores, que puede ser:

- SISD: Single Instruction Single Data
- SIMD: Single Instruction Multiple Data
- MIMD: Multiple Instruction Multiple Data

Por el tipo de ordenador utilizado en el sistema

- Estaciones de trabajo (Workstations)
- Terminales ligeros (Thin clients)
- Microordenadores (por ejemplo ordenadores personales)
- Miniordenadores (servidores pequeños)
- Macroordenadores (servidores de gran capacidad)
- Superordenadores

Por la arquitectura

- Sistema aislado
- Arquitectura cliente-servidor
- Arquitectura de 3 capas
- Arquitectura de n capas
- Servidor de aplicaciones
- Monitor de teleproceso o servidor de transacciones

BASE DE DATOS

La Base de Datos es un sistema que almacena datos que están relacionados. Es un repositorio en donde guardamos información integrada que podemos almacenar y recuperar.

- Un conjunto de información almacenada en memoria auxiliar que permite acceso directo y un conjunto de programas que manipulan esos datos

Componentes de una Base de Datos:

Los componentes básicos de una Base de Datos son:

- *Hardware*: constituido por dispositivo de almacenamiento como discos, tambores, cintas, etc.
- *Software*: que es el DBMS o Sistema Administrador de Base de Datos.
- *Datos*: los cuales están almacenados de acuerdo a la estructura externa y van a ser procesados para convertirse en información.

Estos componentes de la base de datos se pueden visualizar en el gráfico No 3

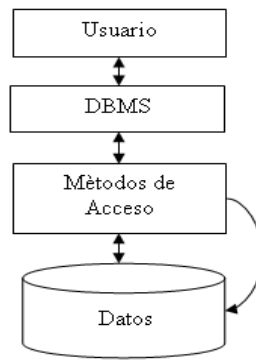


Gráfico N. 3: Componentes de una Base de Datos

Fuente: Introducción a los Sistemas de Base de Datos.

Elaborado por: Investigador tomando como referencia el documento Introducción a los Sistemas de Base de Datos – Marlon Ruiz

Tipos de Usuarios en Base de Datos

Los tipos de usuarios en Base de Datos son los siguientes:

- *Usuario Final:* es la persona que utiliza los datos, esta persona ve datos convertidos en información:
- *Desarrollador de Aplicaciones:* es la persona que desarrolla los sistemas que interactúan con la Base de Datos.
- *DBA:* es la persona que asegura integridad, consistencia, redundancia, seguridad este es el Administrador de Base de Datos quien se encarga de realizar el mantenimiento diario o periódico de los datos.

Las personas que tienen acceso a los Sistemas de Gestión de Bases de Datos (DBMS siglas en inglés) se clasifican de la siguiente manera:

USUARIOS INGENUOS. – Son aquellos que interactúan con el sistema por medio de aplicaciones permanentes.

USUARIOS SOFISTICADOS.– son aquellos con la capacidad de acceder a la información por medios de lenguajes de consulta.

PROGRAMADORES DE APLICACIÓN.- son aquellos con un amplio dominio del DML capaces de generar nuevos módulos o utilerías capaces de manejar nuevos datos en el sistema.

USUARIOS ESPECIALIZADOS.- son aquellos que desarrollan módulos que no se refieren precisamente al manejo de los datos, si no a aplicaciones avanzadas como sistemas expertos, reconocimientos de imágenes, procesamiento de audio y demás.

Conceptos Básicos de Base de datos

Algunos de los conceptos más utilizados en Base de Datos son:

- *Archivo:* son conjuntos de registros.
- *Registros:* son conjuntos de campos.
- *Campos:* es la mínima unidad de referencia.

Tipos de base de datos

Las bases de datos pueden clasificarse de varias maneras, de acuerdo al contexto que se esté manejando, la utilidad de las mismas o las necesidades que satisfagan.

Según la variabilidad de los datos almacenados

- Bases de datos estáticas

Son bases de datos de sólo lectura, utilizadas primordialmente para almacenar datos históricos que posteriormente se pueden utilizar para estudiar el comportamiento de un conjunto de datos a través del tiempo, realizar proyecciones, tomar decisiones y realizar análisis de datos para inteligencia empresarial.

- Bases de datos dinámicas

Éstas son bases de datos donde la información almacenada se modifica con el tiempo, permitiendo operaciones como actualización, borrado y adición de datos, además de las operaciones fundamentales de consulta. Un ejemplo de esto puede ser la base de datos utilizada en un sistema de información de un supermercado, una farmacia, un videoclub o una empresa.

Según el contenido

- Bases de datos bibliográficas

Sólo contienen un subrogante (representante) de la fuente primaria, que permite localizarla. Un registro típico de una base de datos bibliográfica contiene información sobre el autor, fecha de publicación, editorial, título, edición, de una determinada publicación, etc. Puede contener un resumen o extracto de la publicación original, pero nunca el texto completo, porque si no, estaríamos en presencia de una base de datos a texto completo (o de fuentes primarias).

- Bases de datos de texto completo

Almacenan las fuentes primarias, como por ejemplo, todo el contenido de todas las ediciones de una colección de revistas científicas.

- Directorios

Un ejemplo son las guías telefónicas en formato electrónico.

- Bases de datos o "bibliotecas" de información química o biológica

Son bases de datos que almacenan diferentes tipos de información proveniente de la química, las ciencias de la vida o médicas. Se pueden considerar en varios subtipos:

- Las que almacenan secuencias de nucleótidos o proteínas.
- Las bases de datos de rutas metabólicas.

- Bases de datos de estructura, comprende los registros de datos experimentales sobre estructuras 3D de biomoléculas
- Bases de datos clínicas.
- Bases de datos bibliográficas (biológicas, químicas, médicas y de otros campos): PubChem, Medline, EBSCOhost.

Modelos de bases de datos

Además de la clasificación por la función de las bases de datos, éstas también se pueden clasificar de acuerdo a su modelo de administración de datos.

Un modelo de datos es básicamente una "descripción" de algo conocido como *contenedor de datos* (algo en donde se guarda la información), así como de los métodos para almacenar y recuperar información de esos contenedores. Los modelos de datos no son cosas físicas: son abstracciones que permiten la implementación de un sistema eficiente de *base de datos*; por lo general se refieren a algoritmos, y conceptos matemáticos.

Algunos modelos con frecuencia utilizados en las bases de datos:

Bases de datos jerárquicas

En este modelo los datos se organizan en una forma similar a un árbol (visto al revés), en donde un *nodo padre* de información puede tener varios *hijos*. El nodo que no tiene padres es llamado *raíz*, y a los nodos que no tienen hijos se los conoce como *hojas*.

Las bases de datos jerárquicas son especialmente útiles en el caso de aplicaciones que manejan un gran volumen de información y datos muy compartidos permitiendo crear estructuras estables y de gran rendimiento.

Una de las principales limitaciones de este modelo es su incapacidad de representar eficientemente la redundancia de datos.

Base de datos de red

Éste es un modelo ligeramente distinto del jerárquico; su diferencia fundamental es la modificación del concepto de *nodo*: se permite que un mismo nodo tenga varios padres (posibilidad no permitida en el modelo jerárquico).

Fue una gran mejora con respecto al modelo jerárquico, ya que ofrecía una solución eficiente al problema de redundancia de datos; pero, aun así, la dificultad que significa administrar la información en una base de datos de red ha significado que sea un modelo utilizado en su mayoría por programadores más que por usuarios finales.

Bases de datos transaccionales

Son bases de datos cuyo único fin es el envío y recepción de datos a grandes velocidades, estas bases son muy poco comunes y están dirigidas por lo general al entorno de análisis de calidad, datos de producción e industrial, es importante entender que su fin único es recolectar y recuperar los datos a la mayor velocidad posible, por lo tanto la redundancia y duplicación de información no es un problema como con las demás bases de datos, por lo general para poderlas aprovechar al máximo permiten algún tipo de conectividad a bases de datos relacionales.

Un ejemplo habitual de transacción es el traspaso de una cantidad de dinero entre cuentas bancarias. Normalmente se realiza mediante dos operaciones distintas, una en la que se decrementa el saldo de la cuenta origen y otra en la que incrementamos el saldo de la cuenta destino. Para garantizar la atomicidad del sistema (es decir, para que no aparezca o desaparezca dinero), las dos operaciones deben ser atómicas, es decir, el sistema debe garantizar que, bajo cualquier circunstancia (incluso una caída del sistema), el resultado final es que, o bien se han realizado las dos operaciones, o bien no se ha realizado ninguna.

Bases de datos relacionales

Éste es el modelo utilizado en la actualidad para modelar problemas reales y administrar datos dinámicamente. Tras ser postulados sus fundamentos en 1970 por Edgar Frank Codd, de los laboratorios IBM en San José (California), no tardó en consolidarse como un nuevo paradigma en los modelos de base de datos. Su idea fundamental es el uso de "relaciones". Estas relaciones podrían considerarse en forma lógica como conjuntos de datos llamados "tuplas". Pese a que ésta es la teoría de las bases de datos relacionales creadas por Codd, la mayoría de las veces se conceptualiza de una manera más fácil de imaginar. Esto es pensando en cada relación como si fuese una tabla que está compuesta por *registros* (las filas de una tabla), que representarían las tuplas, y *campos* (las columnas de una tabla).

En este modelo, el lugar y la forma en que se almacenen los datos no tienen relevancia (a diferencia de otros modelos como el jerárquico y el de red). Esto tiene la considerable ventaja de que es más fácil de entender y de utilizar para un usuario esporádico de la base de datos. La información puede ser recuperada o almacenada mediante "consultas" que ofrecen una amplia flexibilidad y poder para administrar la información.

El lenguaje más habitual para construir las consultas a bases de datos relacionales es SQL, *Structured Query Language* o *Lenguaje Estructurado de Consultas*, un estándar implementado por los principales motores o sistemas de gestión de bases de datos relacionales.

Durante su diseño, una base de datos relacional pasa por un proceso al que se le conoce como normalización de una base de datos.

Durante los años 80 la aparición de dBASE produjo una revolución en los lenguajes de programación y sistemas de administración de datos. Aunque nunca debe olvidarse que dBase no utilizaba SQL como lenguaje base para su gestión.

Bases de datos multidimensionales

Son bases de datos ideadas para desarrollar aplicaciones muy concretas, como creación de **Cubos OLAP**. Básicamente no se diferencian demasiado de las bases de datos relacionales (una tabla en una base de datos relacional podría serlo también en una base de datos multidimensional), la diferencia está más bien a nivel conceptual; en las bases de datos multidimensionales los campos o atributos de una tabla pueden ser de dos tipos, o bien representan dimensiones de la tabla, o bien representan métricas que se desean estudiar.

Bases de datos orientadas a objetos

Es un modelo, bastante reciente, y propio de los modelos informáticos orientados a objetos, el cual trata de almacenar en la base de datos los *objetos* completos (estado y comportamiento).

Una base de datos orientada a objetos es una base de datos que incorpora todos los conceptos importantes del paradigma de objetos como son:

- Encapsulación - Propiedad que permite ocultar la información al resto de los objetos, impidiendo así accesos incorrectos o conflictos.
- Herencia - Propiedad a través de la cual los objetos heredan comportamiento dentro de una jerarquía de clases.
- Polimorfismo - Propiedad de una operación mediante la cual puede ser aplicada a distintos tipos de objetos.

En bases de datos orientadas a objetos, los usuarios pueden definir operaciones sobre los datos como parte de la definición de la base de datos. Una operación (llamada función) se especifica en dos partes. La interfaz (o *signatura*) de una operación incluye el nombre de la operación y los tipos de datos de sus argumentos (o *parámetros*). La implementación (o *método*) de la operación se especifica separadamente y puede modificarse sin afectar la interfaz. Los programas de aplicación de los usuarios pueden operar sobre los datos invocando a dichas operaciones a través de sus nombres y argumentos, sea cual sea la forma

en la que se han implementado. Esto podría denominarse independencia entre programas y operaciones.

SQL: 2003, es el estándar de SQL92 ampliado, soporta los conceptos orientados a objetos y mantiene la compatibilidad con SQL92.

Bases de datos documentales

Permiten la indexación a texto completo, y en líneas generales realizar búsquedas más potentes. Tesaurus es un sistema de índices optimizado para este tipo de bases de datos.

Bases de datos deductivas

Un sistema de base de datos deductiva, es un sistema de base de datos pero con la diferencia de que permite hacer deducciones a través de inferencias. Se basa principalmente en reglas y hechos que son almacenados en la base de datos. Las bases de datos deductivas son también llamadas bases de datos lógicas, a raíz de que se basa en lógica matemática. Este tipo de base de datos surge debido a las limitaciones de la Base de Datos Relacional de responder a consultas recursivas y de deducir relaciones indirectas de los datos almacenados en la base de datos.

Gestión de bases de datos distribuida (SGBD)

La base de datos y el software SGBD pueden estar distribuidos en múltiples sitios conectados por una red. Existen dos tipos:

1. Distribuidos homogéneos: utilizan el mismo SGBD en múltiples sitios.
2. Distribuidos heterogéneos: Da lugar a los SGBD federados o sistemas multibase de datos en los que los SGBD participantes tienen cierto grado de autonomía local y tienen acceso a varias bases de datos autónomas preexistentes almacenados en los SGBD, muchos de estos emplean una arquitectura cliente-servidor.

Estas surgen debido a la existencia física de organismos descentralizados. Esto les da la capacidad de unir las bases de datos de cada localidad y acceder así a distintas universidades, sucursales de tiendas, etcétera.

Modelos de Datos

Varias técnicas son usadas para modelar la estructura de datos. La mayor parte de sistemas de base de datos son construidos en torno a un modelo de datos particular, aunque sea cada vez más común para productos ofrecer el apoyo a más de un modelo. Ya que cualquier varia puesta en práctica lógica modela física puede ser posible, y la mayor parte de productos ofrecerán al usuario algún nivel de control en la sintonía de la puesta en práctica física, desde las opciones que son hechas tienen un efecto significativo sobre el funcionamiento. Un ejemplo de esto es el modelo emparentado: todas las puestas en práctica serias del modelo emparentado permiten la creación de índices que proporcionan rápido acceso a filas en una tabla si conocen los valores de ciertas columnas.

Modelo de tabla

El modelo de tabla consiste en una serie única, bidimensional de elementos de datos, donde todos los miembros de una columna dada son asumidos para ser valores similares, y todos los miembros de una fila son asumidos para ser relacionados el uno con el otro. Por ejemplo, columnas para el nombre y la contraseña que podría ser usada como una parte de una base de datos de seguridad de sistema. Cada fila tendría la contraseña específica asociada con un usuario individual. Las columnas de la tabla a menudo tienen un tipo asociado con ellos, definiéndolos como datos de carácter, fecha o la información de tiempo, números enteros, o números de punto flotante.

Modelo jerárquico

En un modelo jerárquico, los datos son organizados en una estructura parecida a un árbol, implicando un eslabón solo ascendente en cada registro para describir anidar, y un campo de clase para guardar los registros en un orden particular en

cada lista de mismo-nivel. Las estructuras jerárquicas fueron usadas extensamente en los primeros sistemas de gestión de datos de unidad central, como el Sistema de Dirección de Información (IMS) por la IBM, y ahora describen la estructura de documentos XML. Esta estructura permite un 1:N en una relación entre dos tipos de datos. Esta estructura es muy eficiente para describir muchas relaciones en el verdadero real; recetas, índice, ordenamiento de párrafos/versos, alguno anidó y clasificó la información. Sin embargo, la estructura jerárquica es ineficaz para ciertas operaciones de base de datos cuando un camino lleno (a diferencia del eslabón ascendente y el campo de clase) también no es incluido para cada registro. Una limitación del modelo jerárquico es su inhabilidad de representar manera eficiente la redundancia en datos. Los modelos de base de datos " el valor de atributo de entidad " como Caboodle por Swink están basados en esta estructura. En la relación Padre-hijo: El hijo sólo puede tener un padre pero un padre puede tener múltiples hijos. Los padres e hijos son atados juntos por eslabones "indicadores" llamados. Un padre tendrá una lista de indicadores de cada uno de sus hijos.

Modelo de red

El modelo de red (definido por la especificación CODASYL) organiza datos que usan dos fundamental construcciones, registros llamados y conjuntos. Los registros contienen campos (que puede ser organizado jerárquicamente, como en el lenguaje COBOL de lenguaje de programación). Los conjuntos (para no ser confundido con conjuntos matemáticos) definen de una a varias relaciones entre registros: un propietario, muchos miembros. Un registro puede ser un propietario en cualquier número de conjuntos, y un miembro en cualquier número de conjuntos.

El modelo de red es una variación sobre el modelo jerárquico, al grado que es construido sobre el concepto de múltiples ramas (estructuras de nivel inferior) emanando de uno o varios nodos (estructuras de nivel alto), mientras el modelo se diferencia del modelo jerárquico en esto las ramas pueden estar unidas a múltiples

nodos. El modelo de red es capaz de representar la redundancia en datos de una manera más eficiente que en el modelo jerárquico.

Las operaciones del modelo de red son de navegación en el estilo: un programa mantiene una posición corriente, y navega de un registro al otro por siguiente las relaciones en las cuales el registro participa. Los registros también pueden ser localizados por suministrando valores claves.

Aunque esto no sea un rasgo esencial del modelo, las bases de datos de red generalmente ponen en práctica las relaciones de juego mediante indicadores que directamente dirigen la ubicación de un registro sobre el disco. Esto da el funcionamiento de recuperación excelente, a cargo de operaciones como la carga de base de datos y la reorganización.

La mayor parte de bases de datos de objeto usan el concepto de navegación para proporcionar la navegación rápida a través de las redes de objetos, generalmente usando identificadores de objeto como indicadores "inteligentes" de objetos relacionados. Objectivity/DB, por ejemplo, los instrumentos llamados 1:1, 1:muchos, muchos:1 y muchos:muchos, llamados relaciones que pueden cruzar bases de datos. Muchas bases de datos de objeto también apoyan SQL, combinando las fuerzas de ambos modelos.

El modelo de red (definido por la especificación CODASYL) organiza datos que usan dos fundamentales construcciones, registros llamados y conjuntos. Los registros contienen campos (que puede ser organizado jerárquicamente, como en el lenguaje COBOL de lenguaje de programación). Los conjuntos (para no ser confundido con conjuntos matemáticos) definen de uno a varios relaciones entre registros: un propietario, muchos miembros. Un registro puede ser un propietario en cualquier número de conjuntos, y un miembro en cualquier número de conjuntos. El modelo de red es una variación sobre el modelo jerárquico, al grado que es construido sobre el concepto de múltiples ramas (estructuras de nivel inferior) emanando de uno o varios nodos (estructuras de nivel alto), mientras el

modelo se diferencia del modelo jerárquico en esto las ramas pueden estar unidas a múltiples nodos. El modelo de red es capaz de representar la redundancia en datos de una manera más eficiente que en el modelo jerárquico definen de uno a varias relaciones entre registros: un propietario, muchos miembros. Un registro puede ser un propietario en cualquier número Las operaciones del modelo de red son de navegación en el estilo: un programa mantiene una posición corriente, y navega de un registro al otro por siguiente las relaciones en las cuales el registro participa. Los registros también pueden ser localizados por suministrando valores claves. Aunque esto no sea un rasgo esencial del modelo, las bases de datos de red generalmente ponen en práctica las relaciones de juego mediante indicadores que directamente dirigen la ubicación de un registro sobre el disco. Esto da el funcionamiento de recuperación excelente, a cargo de operaciones como la carga de base de datos y la reorganización.

Modelo Dimensional

El modelo dimensional es una adaptación especializada del modelo relacional, solía representar datos en depósitos de datos, en un camino que los datos fácilmente pueden ser resumidos usando consultas OLAP. En el modelo dimensional, una base de datos consiste en una sola tabla grande de hechos que son descritos usando dimensiones y medidas.

Una dimensión proporciona el contexto de un hecho (como quien participó, cuando y donde pasó, y su tipo). Las dimensiones se toman en cuenta en la formulación de las consultas para agrupar hechos que están relacionados. Las dimensiones tienden a ser discretas y son a menudo jerárquicas; por ejemplo, la ubicación podría incluir el edificio, el estado, y el país.

Una medida es una cantidad que describe el hecho, tales como los ingresos. Es importante que las medidas puedan ser agregados significativamente - por ejemplo, los ingresos provenientes de diferentes lugares pueden sumarse.

En una consulta (OLAP), las dimensiones son escogidas y los hechos son agrupados y añadidos juntos para crear un reporte.

El modelo dimensional a menudo es puesto en práctica sobre la cima del modelo emparentado que usa un esquema de estrella, consistiendo en una mesa que contiene los hechos y mesas circundantes que contienen las dimensiones. Dimensiones en particular complicadas podrían ser representadas usando múltiples mesas, causando un esquema de copo de nieve.

Modelo de objeto

En años recientes, el paradigma mediante objetos ha sido aplicado a la tecnología de base de datos, creando un nuevo modelo de programa sabido (conocido) como bases de datos de objeto. Estas bases de datos intentan traer el mundo de base de datos y el uso que programa el mundo más cerca juntos, en particular por asegurando que la base de datos usa el mismo sistema de tipo que el programa de uso. Esto apunta para evitar el elevado (a veces mencionaba el desajuste de impedancia) de convertir la información entre su representación en la base de datos (por ejemplo como filas en mesas) y su representación en el programa de uso (típicamente como objetos). Al mismo tiempo, las bases de datos de objeto intentan introducir las ideas claves de programa de objeto, como encapsulación y polimorfismo, en el mundo de bases de datos.

Una variedad de estas formas ha sido aspirada almacenando objetos en una base de datos. Algunos productos se han acercado al problema del uso que programa el final, por haciendo los objetos manipulados según el programa persistente. Esto también típicamente requiere la adición de una especie de lengua de pregunta, ya que lenguajes de programación convencionales no tienen la capacidad de encontrar objetos basados en su contenido de la información. Los otros han atacado el problema a partir del final de base de datos, por definiendo un modelo de datos mediante objetos para la base de datos, y definiendo un lenguaje de

programación de base de datos que permite a capacidades de programa llenas así como instalaciones de pregunta tradicionales.

Las bases de datos de objeto han sufrido debido a la carencia de estandarización: aunque las normas fueran definidas por ODMG, nunca fueron puestas en práctica lo bastante bien para asegurar la interoperabilidad entre productos. Sin embargo, las bases de datos de objeto han sido usadas satisfactoriamente en muchos usos: Usualmente aplicaciones especializadas como bases de datos de ingeniería, base de datos biológica molecular, más bien que proceso de datos establecido comercial. Sin embargo, las ideas de base de datos de objeto fueron recogidas por los vendedores emparentados y extensiones influidas hechas a estos productos y de verdad a la lengua SQL.

Data Mining

La minería de datos o Datamining (es la etapa de análisis de "Knowledge Discovery in Databases" o KDD), es un campo de las ciencias de la computación, es el proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos. Utiliza los métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos. El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos transformándola en una estructura comprensible para su uso posterior. Además de la etapa de análisis en bruto, que involucra aspectos de bases de datos y gestión de datos, procesamiento de datos, el modelo y las consideraciones de inferencia, métricas de intereses, consideraciones de la Teoría de la complejidad computacional, post-procesamiento de las estructuras descubiertas, la visualización y actualización en línea.

El término es una palabra de moda, y es frecuentemente mal utilizado para referirse a cualquier forma de datos a gran escala o procesamiento de la información (recolección, extracción, almacenamiento, análisis y estadísticas), pero también se ha generalizado a cualquier tipo de sistema de apoyo informático decisión, incluyendo la inteligencia artificial, aprendizaje automático y la

inteligencia empresarial. En el uso de la palabra, el término clave es el descubrimiento, comúnmente se define como "la detección de algo nuevo". Incluso el popular libro "La minería de datos: sistema de prácticas herramientas de aprendizaje y técnicas con Java" (que cubre todo el material de aprendizaje automático) originalmente iba a ser llamado simplemente "la máquina de aprendizaje práctico", y el término "minería de datos" no se añadió por razones de marketing. A menudo, los términos más generales "(gran escala) el análisis de datos", o "análisis" -. o cuando se refiere a los métodos actuales, la inteligencia artificial y aprendizaje automático, son más apropiados.

La tarea de minería de datos real es el análisis automático o semi-automático de grandes cantidades de datos para extraer patrones interesantes hasta ahora desconocidos, como los grupos de registros de datos (análisis clúster), registros poco usuales (la detección de anomalías) y dependencias (Asociación Minera regla). Esto generalmente implica el uso de técnicas de bases de datos como los índices espaciales. Estos patrones pueden entonces ser visto como una especie de resumen de los datos de entrada, y puede ser utilizado en el análisis adicional o, por ejemplo, en la máquina de aprendizaje y análisis predictivo. Por ejemplo, el paso de minería de datos podría identificar varios grupos en los datos, que luego pueden ser utilizados para obtener resultados más precisos de predicción por un sistema de soporte de decisiones. Ni la recolección de datos, preparación de datos, ni la interpretación de los resultados y la información son parte de la etapa de minería de datos, pero que pertenecen a todo el proceso KDD como pasos adicionales.

Los términos relacionados con el dragado de datos, la pesca de datos y espionaje de los datos se refieren a la utilización de métodos de minería de datos a las partes de la muestra de un conjunto de datos de población más grandes establecidas que son (o pueden ser) demasiado pequeño para las inferencias estadísticas fiables que se hizo acerca de la validez de cualquier patrones descubiertos. Estos métodos pueden, sin embargo, ser utilizado en la creación de nuevas hipótesis que se prueba contra las poblaciones de datos más grandes.

Proceso de Minería de Datos

Un proceso típico de minería de datos consta de los siguientes pasos generales:

1. **Selección del conjunto de datos.-** Tanto en lo que se refiere a las variables objetivo (aquellas que se quiere predecir, calcular o inferir), como a las variables independientes (las que sirven para hacer el cálculo o proceso), como posiblemente al muestreo de los registros disponibles.
2. **Análisis de las propiedades de los datos.-** En especial los histogramas, diagramas de dispersión, presencia de valores atípicos y ausencia de datos (valores nulos).
3. **Transformación del conjunto de datos de entrada.-** Se realizará de diversas formas en función del análisis previo, con el objetivo de prepararlo para aplicar la técnica de minería de datos que mejor se adapte a los datos y al problema, a este paso también se le conoce como **pre procesamiento** de los datos.
4. **Seleccionar y aplicar la técnica de minería de datos.-** Se construye el modelo predictivo, de clasificación o segmentación.
5. **Extracción de conocimiento.-** Mediante una técnica de minería de datos, se obtiene un modelo de conocimiento, que representa patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables. También pueden usarse varias técnicas a la vez para generar distintos modelos, aunque generalmente cada técnica obliga a un pre-procesado diferente de los datos.
6. **Interpretación y evaluación de datos.-** Una vez obtenido el modelo, se debe proceder a su validación comprobando que las conclusiones que arroja son válidas y suficientemente satisfactorias. En el caso de haber obtenido varios modelos mediante el uso de distintas técnicas, se deben comparar los modelos en busca de aquel que se ajuste mejor al problema. Si ninguno de los modelos alcanza los resultados esperados, debe alterarse alguno de los pasos anteriores para generar nuevos modelos.

Si el modelo final no superara esta evaluación el proceso se podría repetir desde el principio o, si el *experto* lo considera oportuno, a partir de cualquiera de los pasos anteriores. Esta retroalimentación se podrá repetir cuantas veces se considere necesario hasta obtener un modelo válido.

Una vez validado el modelo, si resulta ser aceptable (proporciona salidas adecuadas y/o con márgenes de error admisibles) éste ya está listo para su explotación. Los modelos obtenidos por técnicas de minería de datos se aplican incorporándolos en los sistemas de análisis de información de las organizaciones, e incluso, en los sistemas transaccionales. En este sentido cabe destacar los esfuerzos del Data Mining Group, que está estandarizando el lenguaje **PMML** (Predictive Model Markup Language), de manera que los modelos de minería de datos sean interoperables en distintas plataformas, con independencia del sistema con el que han sido construidos. Los principales fabricantes de sistemas de bases de datos y programas de análisis de la información hacen uso de este estándar.

Tradicionalmente, las técnicas de minería de datos se aplicaban sobre información contenida en almacenes de datos. De hecho, muchas grandes empresas e instituciones han creado y alimentan bases de datos especialmente diseñadas para proyectos de minería de datos en las que centralizan información potencialmente útil de todas sus áreas de negocio. No obstante, actualmente está cobrando una importancia cada vez mayor la minería de datos desestructurados como información contenida en ficheros de texto, en Internet, etc.

Protocolo de un proyecto de minería de datos

Un proyecto de minería de datos tiene **varias fases necesarias** que son, esencialmente:

- Comprensión del **negocio** y del problema que se quiere resolver.
- Determinación, obtención y limpieza de los **datos** necesarios.
- Creación de **modelos** matemáticos.
- Validación, comunicación, etc. de los **resultados** obtenidos.

- **Integración**, si procede, de los resultados en un sistema transaccional o similar.

La relación entre todas estas fases sólo es lineal sobre el papel. En realidad, es mucho más compleja y esconde toda una jerarquía de sub-fases. A través de la experiencia acumulada en proyectos de minería de datos se han ido desarrollando *metodologías* que permiten gestionar esta complejidad de una manera más o menos uniforme.

Técnicas de minería de datos

Como ya se ha comentado, las técnicas de la minería de datos provienen de la Inteligencia artificial y de la estadística, dichas técnicas, no son más que algoritmos, más o menos sofisticados que se aplican sobre un conjunto de datos para obtener unos resultados.

Las técnicas más representativas son:

- **Redes neuronales.**- Las Redes neuronales son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida. Algunos ejemplos de red neuronal son:
 - El Perceptrón.
 - El Perceptrón multicapa.
 - Los Mapas Auto organizados, también conocidos como redes de Kohonen.
- **Regresión lineal.**- Es la más utilizada para formar relaciones entre datos. Rápida y eficaz pero insuficiente en espacios multidimensionales donde puedan relacionarse más de 2 variables.
- **Árboles de decisión.**- Un árbol de decisión es un modelo de predicción utilizado en el ámbito de la inteligencia artificial, dada una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a

los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema. Ejemplos:

- Algoritmo ID3.
- Algoritmo C4.5.

- **Modelos estadísticos.**- Es una expresión simbólica en forma de igualdad o ecuación que se emplea en todos los diseños experimentales y en la regresión para indicar los diferentes factores que modifican la variable de respuesta.

- **Agrupamiento o *Clustering*.**- Es un procedimiento de agrupación de una serie de vectores según criterios habitualmente de distancia; se tratará de disponer los vectores de entrada de forma que estén más cercanos aquellos que tengan características comunes. Ejemplos:
 - Algoritmo K-means.
 - Algoritmo K-medoids.

- **Reglas de asociación.**- Se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos.

Según el objetivo del análisis de los datos (Weiss y Indurkha, 1998), los algoritmos utilizados se clasifican en:

- Algoritmos supervisados (o predictivos): predicen un dato (o un conjunto de ellos) desconocido a priori, a partir de otros conocidos.
- Algoritmos no supervisados (o del descubrimiento del conocimiento): se descubren patrones y tendencias en los datos.

Minería de datos basada en teoría de la información

Todas las herramientas tradicionales de minería de datos *asumen* que los datos que usarán para construir los modelos contienen la información necesaria para

lograr el propósito buscado: obtener suficiente conocimiento que pueda ser aplicado al *negocio* (o problema) para obtener un beneficio (o solución).

El inconveniente es que esto no es necesariamente cierto. Además, existe otro problema mayor aún. Una vez construido el modelo no es posible conocer si el mismo ha capturado toda la información disponible en los datos. Por esta razón la práctica común es realizar varios modelos con distintos parámetros para ver si alguno logra mejores resultados.

Un enfoque relativamente nuevo al análisis de datos soluciona estos problemas haciendo que la práctica de la minería de datos se parezca más a una ciencia que a un arte.

En 1948 Claude Shannon publicó un trabajo llamado “Una Teoría Matemática de la Comunicación”. Posteriormente esta pasó a llamarse Teoría de la información y sentó las bases de la comunicación y la codificación de la información. Shannon propuso una manera de medir la cantidad de información a ser expresada en bits.

En 1999 Dorian Pyle publicó un libro llamado “Data Preparation for Data Mining” en el que propone una manera de usar la Teoría de la Información para analizar datos. En este nuevo enfoque, una base de datos es un canal que transmite información. Por un lado está el mundo real que captura datos generados por el negocio. Por el otro están todas las situaciones y problemas importantes del negocio. Y la información fluye desde el mundo real y a través de los datos, hasta la problemática del negocio.

Con esta perspectiva y usando la Teoría de la información, es posible medir la cantidad de información disponible en los datos y qué porción de la misma podrá utilizarse para resolver la problemática del negocio. Como un ejemplo práctico, podría encontrarse que los datos contienen un 65% de la información necesaria para predecir qué cliente rescindirán sus contratos. De esta manera, si el modelo final es capaz de hacer predicciones con un 60% de acierto, se puede asegurar que la herramienta que generó el modelo hizo un buen trabajo capturando la información disponible. Ahora, si el modelo hubiese tenido un porcentaje de

aciertos de solo el 10%, por ejemplo, entonces intentar otros modelos o incluso con otras herramientas podría valer la pena.

La capacidad de medir información contenida en los datos tiene otras ventajas importantes.

Al analizar los datos desde esta nueva perspectiva se genera un mapa de información que hace innecesario la preparación previa de los datos, una tarea absolutamente imprescindible si se desea buenos resultados, pero que lleva enorme cantidad de tiempo.

Es posible seleccionar un grupo de variables óptimo que contenga la información necesaria para realizar un modelo de predicción.

Una vez que las variables son procesadas con el fin de crear el mapa de información y luego seleccionadas aquellas que aportan la mayor información, la elección de la herramienta que se usará para crear el modelo deja de tener importancia, ya que el mayor trabajo fue realizado en los pasos previos.

Categorías de la Variable Dependiente

Catastro de establecimientos

El término **Catastro** es aquel que se usa para designar al censo o registro que tiene el Estado de los diferentes tipos de propiedades privadas y establecimientos existentes en su territorio. El principal objetivo del catastro es el conocimiento de tales espacios a fin de poder aplicar de manera adecuada y proporcional los correspondientes impuestos. El catastro puede servir además para el control y para la organización geográfica del espacio ya que permite conocer qué espacios están libres, cuáles están ocupados, cuáles deben ser renovados, etc. La palabra catastro proviene del griego *katastikhon*, que significa registro o censo.

El proceso de censo catastral se aplica desde tiempos remotos ya que siempre los Estados han necesitado contribuciones de los habitantes del territorio en forma de

impuestos. Así, el registro catastral se desarrollaba para poder obtener ingresos y organizar o planear el territorio de la manera más adecuada posible.

En lo que respecta al Ministerio de Salud Pública a nivel nacional, el catastro se aplica al registro de los establecimientos públicos y privados que son sujetos a control y vigilancia sanitaria, tales como: Hospitales, Clínicas, Restaurantes, Centros de esparcimiento, Plantas procesadoras de alimentos, entre otros. El catastro constituye un registro actualizado de establecimientos activos, cerrados, clausurados, en mora por falta del pago de impuestos.

La Dirección Provincial de Salud de Cotopaxi no es exenta de llevar un catastro de establecimientos a nivel provincial y cantonal, ya que tiene la obligación de verificar información valedera y oportuna de los establecimientos que se postulan a obtener un permiso de funcionamiento, los mismos que deben reunir los requisitos mínimos de acuerdo a su categoría para una adecuada atención al público.

Vigilancia y control sanitario

La **Vigilancia y Control Sanitario** es el conjunto de acciones capaz de eliminar, disminuir o prevenir riesgos a la Salud, intervenir en los problemas sanitarios provenientes del medio Ambiente, de la producción y circulación de bienes, así como de la prestación de servicios de interés de la Salud.

Es responsabilidad del Departamento de Vigilancia y Control sanitaria lo siguiente:

- Elaborar y ejecutar programas de educación sanitaria y de higiene en el expendio de alimentos.
- Control de bienes de consumo que, directa o indirectamente, se relacionen con la Salud, en todas las etapas y procedimientos de la producción de consumo.

- Control de la prestación de servicios que se relacionen, directa o indirectamente, con la Salud.
- Velar por el cumplimiento de las Leyes, Ordenanzas y Normas sobre la materia;
- Vigilar que los servicios de salud dirigidos a beneficiar a la colectividad, cuenten con los medios necesarios para dar una mejor atención a los habitantes del Cantón.
- Elaborar normas de higiene y salud para el manejo y expendio de productos alimenticios, a fin de precautelar la salud de la población.
- Participar en la elaboración de un sistema de control de las formas de faenamiento, transporte y comercialización de carnes y productos cárnicos.
- Vigilar que los locales donde se expenden alimentos tengan las condiciones higiénico-sanitarias adecuadas.
- Efectuar el control de calidad y contaminación microbiológica en alimentos, refrescos y similares que se expenden en el Cantón, mediante el análisis bromatológico de éstos.
- Realizar un estricto control sobre la calidad del ganado que se faena en el Cantón.
- Establecer normas de control tendientes a optimizar las condiciones de salud e higiene de los mercados y camales del Cantón.
- Autorizar o solicitar la revocatoria de permisos para ocupación o uso del espacio público.
- Supervisar la organización de los sistemas operativos de los mercados y camales; y,
- Las demás funciones que le asigne el Director de Salud y Medio Ambiente.

Emisión de permisos de funcionamiento

El permiso de funcionamiento es el documento otorgado por la autoridad sanitaria nacional a los establecimientos sujetos a control y vigilancia sanitaria que cumplen con todos los requisitos para su funcionamiento.

A las Direcciones Provinciales de Salud les corresponde otorgar el permiso de funcionamiento anual, para lo cual el interesado deberá presentar una solicitud dirigida al Director Provincial de Salud de la jurisdicción a la que pertenece el domicilio del establecimiento, con los siguientes datos:

- Nombre del propietario o representante legal;
- Nombre o razón social o denominación del establecimiento;
- Número de Registro Único de Contribuyentes (RUC) y cédula de ciudadanía o identidad del propietario o representante legal del establecimiento;
- Actividad o actividades que se realizan en el establecimiento
- Ubicación del establecimiento: cantón, parroquia, sector, calle principal, número e intersecciones, teléfono, fax, correo electrónico si lo tuviere.

La Dirección Provincial de Salud de Cotopaxi a través del proceso de Vigilancia Sanitaria realizará la verificación de la documentación presentada y emitirá el informe de la evaluación documental y verificación de la ubicación del local y se realizará una inspección del establecimiento por parte de una comisión técnica conformada por profesionales técnicos del proceso de Vigilancia Sanitaria.

En caso de que el resultado de la inspección sea favorable el proceso de Vigilancia Sanitaria procederá a ingresar los datos en el sistema de cómputo y entregará al usuario la orden de pago del derecho por el servicio correspondiente; una vez cancelado el valor se emitirá el permiso de funcionamiento debidamente legalizados por el Director Provincial de Salud de Cotopaxi y Coordinador de Vigilancia Sanitaria Provincial con sus firmas y sellos respectivos.

Toma de decisiones

Previo a la toma de decisiones por parte de las autoridades de salud se cumple un proceso de vigilancia, inspección, calificación y categorización de

establecimientos, los mismos que deben cumplir con todos los requisitos solicitados por parte del departamento de Control Sanitario.

Las autoridades están en la obligación de analizar cada uno de los aspectos reportados por parte de este departamento con la finalidad de de tomar decisiones tales como:

- Otorgación de permisos de funcionamiento
- Aplicación de multas y/o sanciones
- Clausuras de establecimientos
- Acciones judiciales

Constelación de Ideas de la Variable Independiente

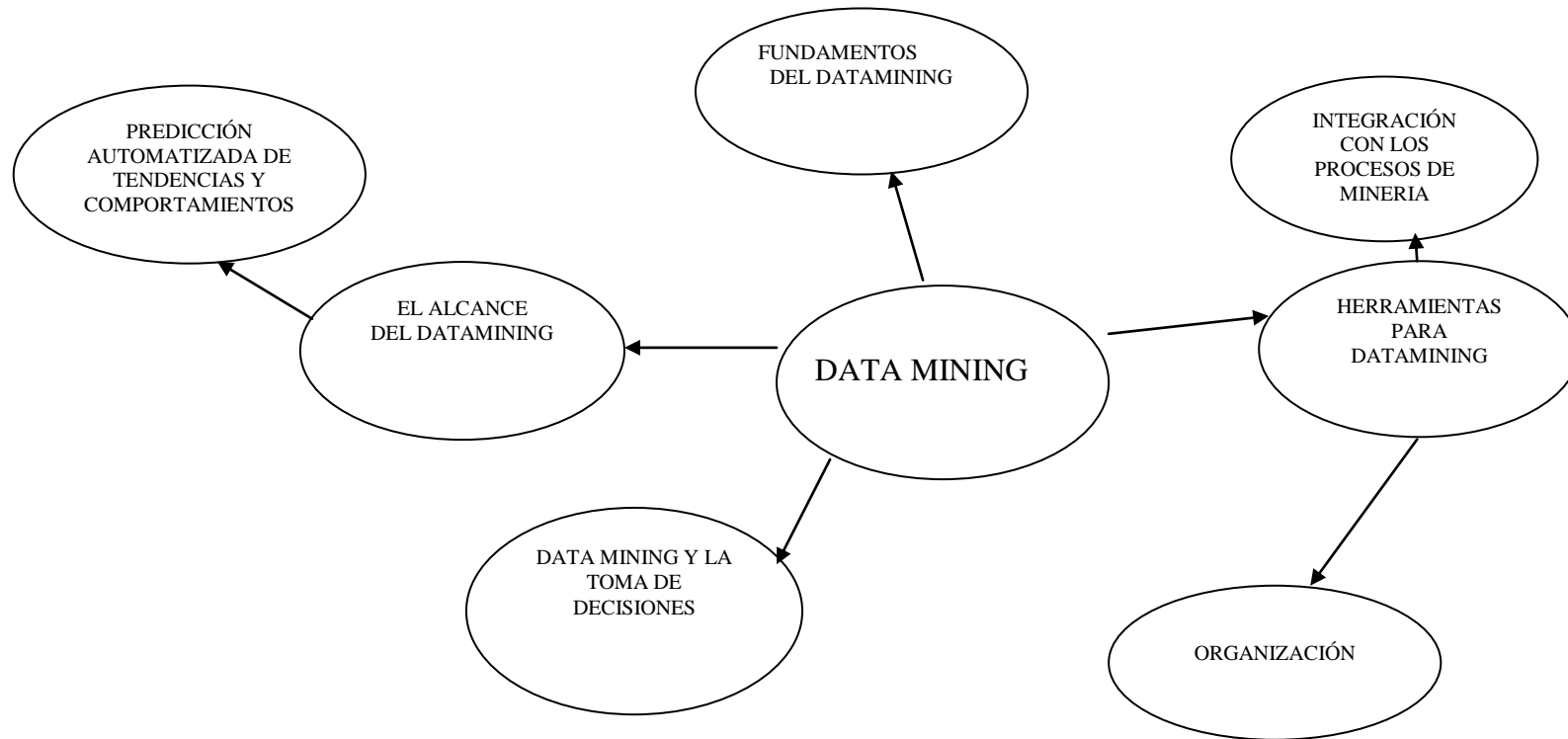


Gráfico N.4.: Subcategorías de la Variable Independiente
Elaborado por: Investigador

Constelación de Ideas de la Variable Dependiente

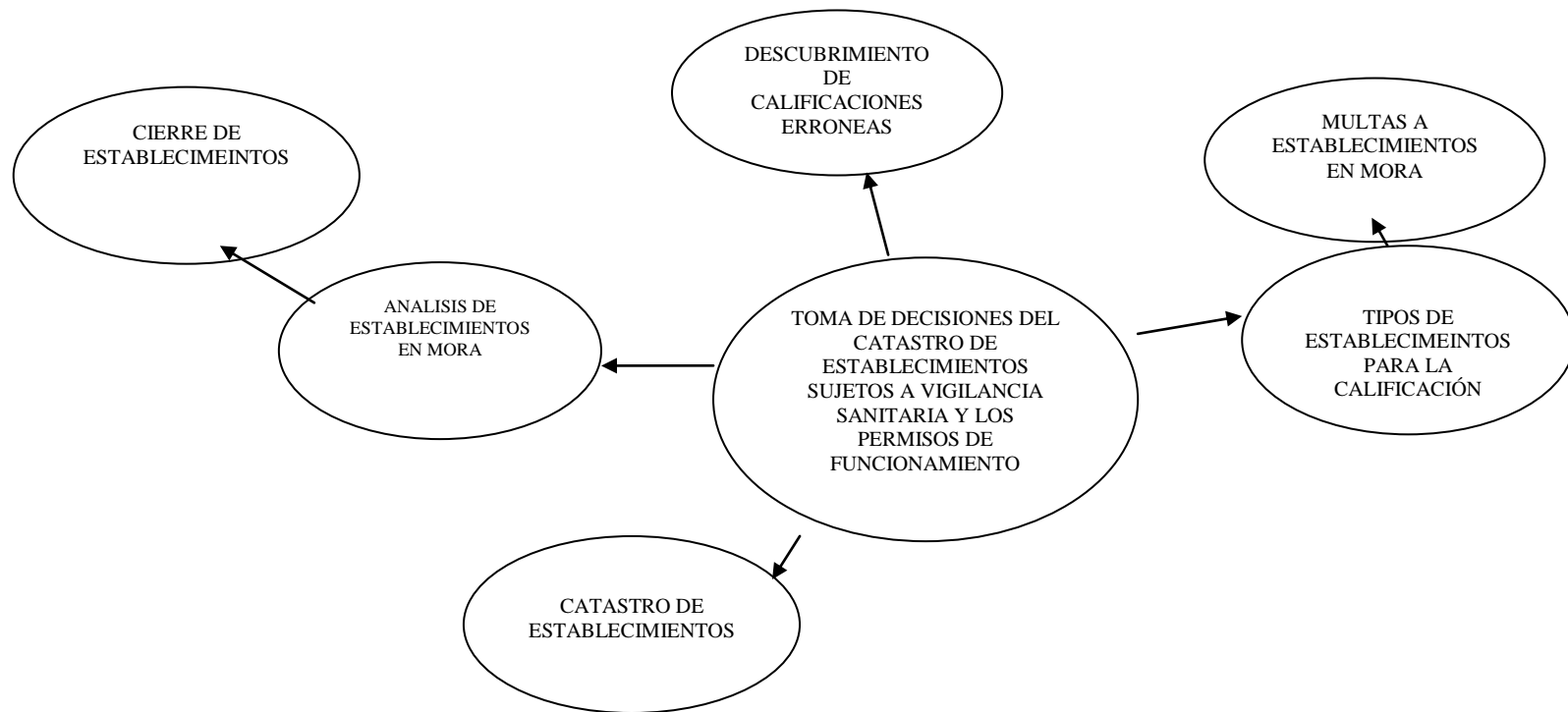


Gráfico N.5.: Subcategorías de la Variable Dependiente
Elaborado por: Investigador

2.5.-Hipótesis

¿La implementación del Data Mining incidirá positivamente en la toma de decisiones del Catastro de establecimientos y la emisión de los Permisos de Funcionamiento por parte de la Dirección Provincial de Salud de Cotopaxi.?

2.6.- Señalamiento de variables

Variable independiente

El Data MINING

Variable dependiente

Toma de decisiones del Catastro de establecimientos y la emisión de los Permisos de Funcionamiento

CAPITULO III

METODOLOGIA

Enfoque

La presente investigación tiene un enfoque de investigación de Ciencias Sociales ya que permite utilizar varios métodos predominantemente cualitativos, que se caracterizan por buscar la comprensión de los fenómenos sociales, a través de una observación naturalista cuyas conclusiones son contextualizadas, además está orientado al descubrimiento de las hipótesis para llegar a propuestas concretas, enfatiza en el proceso y asume una realidad dinámica.

3.1.- Modalidad de Investigación

Investigación de campo

La investigación tiene la modalidad de campo porque el investigador acudió al lugar en donde se producen los hechos para obtener información relacionada con los objetivos del trabajo de grado

Investigación Documental-bibliográfica

Con el propósito de fortalecer la investigación, se recurrió a obtener la información teórica de diferentes autores obtenidas en fuentes secundarias (libros, revistas especializadas, publicaciones, internet, etc.), y de fuentes de información primaria a través de documentos válidos y confiables.

El trabajo de grado responde a un proyecto factible de intervención social porque se plantean una propuesta viable de solución al problema investigado dentro de un contexto determinado.

Proyectos factibles de Intervención Social

El trabajo de grado responde a un proyecto factible de intervención social porque se planteara una propuesta viable de solución al problema investigado dentro de un contexto determinado.

Proyectos especiales

Con la finalidad de dar solución a los problemas de un contexto determinado la investigación tiene la modalidad de proyectos especiales porque con la utilización de la tecnología se desarrolló una herramienta orientada a proporcionar información gerencial referente a los establecimientos sujetos de control y vigilancia sanitaria.

3.2.- Tipos o Niveles de Investigación

Exploratorio

La investigación pasa por el nivel de investigación exploratorio porque sondea un problema poco investigado o desconocido en un contexto determinado, especialmente para definir el tema de investigación.

Descriptivo

Es descriptivo porque se busca informar los resultados obtenidos de la investigación entre la comparación de 2 variables tomando en cuenta criterios de coherencia interna y pertinencia.

Asociación de variables

Porque permite establecer y estudiar las tendencias de comportamiento entre las variables en un contexto determinado.

3.3.- Población y Muestra

La investigación se realiza en la Dirección Provincial de Salud de Cotopaxi para lo cual se toma en cuenta a las siguientes personas

POBLACIONES	FRECUENCIA	PORCENTAJE
Coordinador de Vigilancia Sanitaria	2	13,5%
Personal de Vigilancia Sanitaria	3	20%
Inspectores Sanitarios	8	53%
Personal de Sistemas	2	13,5%
TOTAL	15	100%

Cuadro N.1

Elaborado por: Investigador

En virtud de que ninguna de las poblaciones a ser investigada sobrepasa de 100 elementos se trabaja con la totalidad del universo sin que sea necesario sacar muestras representativas.

3.4.- Operacionalización de Variables

Operacionalización de la Variable Independiente

Variable: El Data MINING

CONCEPTUALIZACIÓN	DIMENSIONES	INDICADORES	ÍTEMS BASICOS	TÉCNICAS
El Data MINING, es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta	Base de Datos	Base de datos en: Negocios	¿Cómo se puede determinar la técnica más conveniente para validar el proceso de un modelo exitoso?	Encuesta Cuestionario
	Análisis	Modelos Estadísticos, descriptivos predicciones	¿Contribuye significativamente el Data MINING en múltiples aplicaciones de negocios?	Encuesta Entrevista Guía de la entrevista

Cuadro N.2

Elaborado por: Investigador

Operacionalización de la Variable Dependiente

Variable: Calificación de establecimientos sujetos a vigilancia y control sanitario

CONCEPTUALIZACIÓN	DIMENSIONES	INDICADORES	ÍTEMS BASICOS	TÉCNICAS
La toma de decisiones del Catastro de establecimientos sujetos a Vigilancia Sanitaria y los Permisos de Funcionamiento tiene el propósito mejorar el servicio a los usuarios , y velar por el cumplimiento de los reglamentos de salud	Servicio a los usuarios	Rapidez en la atención Evitar multas y sanciones Calidad de atención al usuario	¿Es posible pasar a un Data MINING lo más automatizada posible sobre el servicio a los usuarios?	Encuesta Cuestionario
	Cumplimiento de los reglamentos de salud	Presentación de requisitos previos a la obtención del permiso de funcionamiento Cobro por multas	¿Se pueden determinar los establecimientos con mora en los últimos años?	Encuesta Entrevista Guía de la entrevista

Cuadro N.3:

Elaborado por: Investigador

Técnicas e Instrumentos

Las técnicas e instrumentos que se utilizaron para la recolección de información en esta investigación fueron:

- Encuesta(Anexo 1)

3.5.- Plan para Recolección de la Información

Para la recolección de información se emplearon cuadernos de notas, fichas bibliográficas y se aplicaron encuestas a los Ingenieros y técnicos relacionados con el área de estudio.

Plan de Recolección de Información

Para concretar el plan de recolección de información, conviene contestar las siguientes preguntas:

PREGUNTAS BÁSICAS	EXPLICACIÓN
1. ¿Para qué?	Para alcanzar los objetivos de la investigación
2. ¿De qué personas u objetos?	Personal de Vigilancia Sanitaria, dueños de establecimientos
3. ¿Sobre qué aspectos?	Importancia del Data MINING para la toma de decisiones
4. ¿Quién, quiénes?	Investigador: Ing. Hernán Carvajal
5. ¿Cuándo?	Entre Diciembre del 2011 y Julio del 2012.
6. ¿Dónde?	Dirección Provincial de Salud de Cotopaxi
7. ¿Cuántas veces?	15 encuestados
8. ¿Qué técnicas de recolección?	Encuesta
9. ¿Con qué?	Cuestionario
10. ¿En qué situación?	Favorables porque existe la colaboración por parte de las autoridades

Cuadro N.4: Recolección de la Información

Elaborado por: Investigador

3.6.- Plan para el Procesamiento de la Información

Para el procesamiento de la información obtenida se realizaron las siguientes actividades:

- Definición de las preguntas para la encuesta.
- Se encuestó al Coordinador de Vigilancia Sanitaria, Personal de Vigilancia Sanitaria, Inspectores Sanitarios, Personal de Sistemas
- Se realizó la tabulación de los datos obtenidos en las encuestas y entrevista.
- Se estudió los datos críticamente para su correcta interpretación.
- Se realizó un análisis estadístico para la comprobación del problema.

Plan de Procesamiento de Información

Los datos recogidos se transforman siguiendo ciertos procedimientos.

- Revisión crítica de la información recogida; es decir, limpieza de la información defectuosa: contradictoria, incompleta, no pertinente, etc.
- Repetición de la recolección, en ciertos casos individuales, para corregir fallas de contestación.
- Tabulación o cuadros según variables de cada hipótesis: cuadros de una sola variable, cuadro con cruce de variables, etc.
- Manejo de información (reajuste de cuadros con casillas vacías o con datos tan reducidos cuantitativamente, que no influyen significativamente en los análisis).
- Estudio estadístico de datos para presentación de resultados.
- La presentación de datos puede hacerse siguiendo los siguientes procedimientos:
 - Representación escrita
 - Representación semi tabular
 - Representación tabular
 - Representación gráfica
- Análisis e interpretación de resultados

- Análisis de los resultados estadísticos, destacando tendencias o relaciones fundamentales de acuerdo con los objetivos e hipótesis.
- Interpretación de los resultados, con apoyo del marco teórico, en el aspecto pertinente.
- Comprobación de hipótesis Para la verificación estadística conviene seguir la asesoría de un especialista.
- Establecimiento de conclusiones y recomendaciones

CAPÍTULO IV

ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

4.1.- Análisis e interpretación de resultados.

Se aplicó una encuesta a Coordinadores, Personal, Inspectores de Vigilancia Sanitaria y Personal de Sistemas.

Pregunta N. 1:

¿Considera indispensable el uso de una herramienta informática (Datamining) que analice automática y rápidamente grandes cantidades de datos para obtener información valiosa del catastro de establecimientos?

Opciones	Frecuencia	Porcentaje
Totalmente	10	66.67%
Parcialmente	5	33.33%
Nada	0	-
TOTAL	15	100%

Cuadro N.5

Elaborado por: Investigador

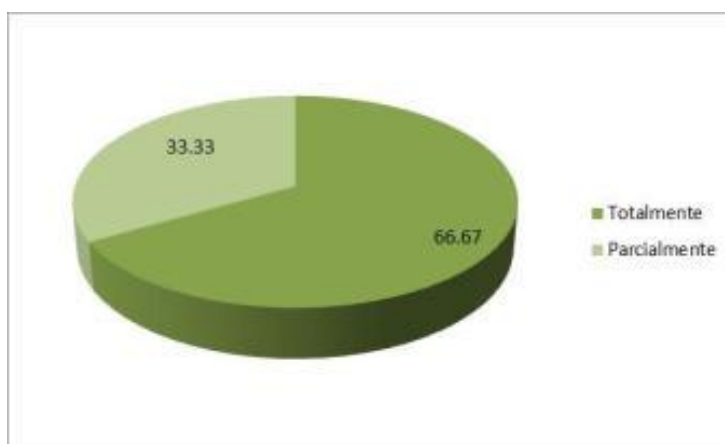


Gráfico N.6.:

Elaborado por: Investigador

ANÁLISIS E INTERPRETACIÓN

Se obtiene que el 66.67% responden que mejoraría TOTALMENTE es indispensable el uso de una herramienta informática (Datamining) que analice automática y rápidamente grandes cantidades de datos para obtener información valiosa del catastro de establecimientos, el 33.33% responden que mejoraría PARCIALMENTE, mientras que el 0% responde que no mejoraría NADA.

Pregunta N. 2:

¿Cree usted que la implementación del Datamining optimizaría el tiempo de trabajo de los inspectores de salud en el proceso de calificación de establecimientos?

Opciones	Frecuencia	Porcentaje
Totalmente	11	73,33%
Parcialmente	4	26,67%
Nada	0	-
TOTAL	15	100%

Cuadro N.6
Elaborado por: Investigador

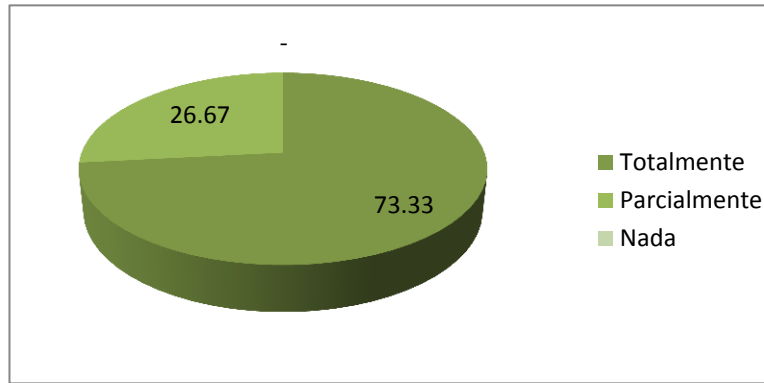


Gráfico N.7.:
Elaborado por: Investigador

ANÁLISIS E INTERPRETACIÓN

Se obtiene que el 73.33% responden que TOTALMENTE la implementación del Datamining optimizará el tiempo de trabajo de los inspectores de salud en el proceso de calificación de establecimientos, el 26.67% responden que PARCIALMENTE, mientras que el 0% responde que NADA.

Pregunta N. 3:

¿Considera usted útil contar con el uso de una herramienta informática que permita conocer los posibles establecimientos que incurrirán en mora?

Opciones	Frecuencia	Porcentaje
Totalmente	12	80%
Parcialmente	3	20 %
Nada	0	-
TOTAL	15	100%

Cuadro N.7
Elaborado por: Investigador

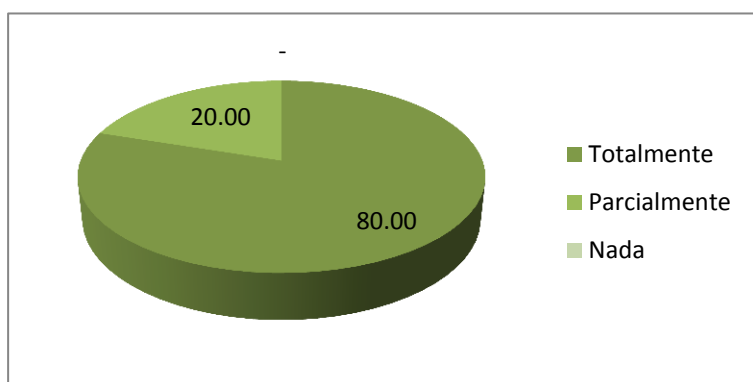


Gráfico N.8.:
Elaborado por: Investigador

ANÁLISIS E INTERPRETACIÓN

Se obtiene que el 80% responden que TOTALMENTE será útil contar con el uso de una herramienta informática que permita conocer los posibles establecimientos que incurrirán en mora, el 20% responden que PARCIALMENTE, mientras que el 0% responde que NADA.

Pregunta N. 4:

¿Cree usted que la información obtenida puede ayudar a prevenir el pago de multas por parte de los propietarios de establecimientos?

Opciones	Frecuencia	Porcentaje
Totalmente	12	80%
Parcialmente	3	20%
Nada	0	-
TOTAL	15	100%

Cuadro N.8
Elaborado por: Investigador

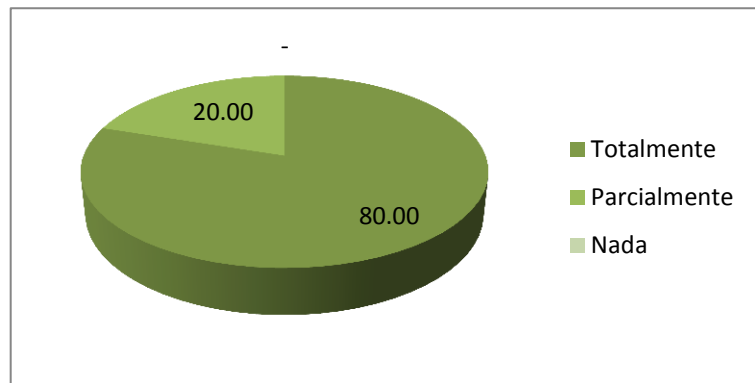


Gráfico N.9.:
Elaborado por: Investigador

ANÁLISIS E INTERPRETACIÓN

Se obtiene que el 80% responden que TOTALMENTE la información obtenida puede ayudar a prevenir el pago de multas por parte de los propietarios de establecimientos, el 20% responden que PARCIALMENTE, mientras que el 0% responde que NADA.

Pregunta N. 5

¿Cree usted que contar con información detallada de los establecimientos de salud ayudará a las autoridades a tomar decisiones previas a la aprobación para la apertura de nuevos establecimientos?

Opciones	Frecuencia	Porcentaje
Totalmente	12	80%
Parcialmente	3	20%
Nada	0	-
TOTAL	15	100%

Cuadro N.9
Elaborado por: Investigador

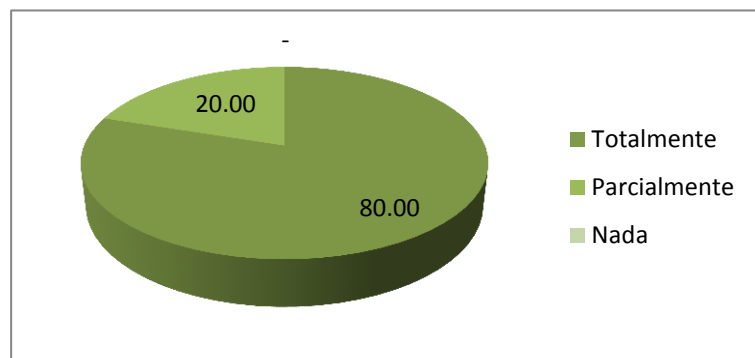


Gráfico N.10
Elaborado por: Investigador

ANÁLISIS E INTERPRETACIÓN

Se obtiene que el 80% responden que TOTALMENTE al contar con información detallada de los establecimientos de salud ayudará a las autoridades a tomar decisiones previas a la aprobación para la apertura de nuevos establecimientos, el 20% responden que PARCIALMENTE, mientras que el 0% responde que NADA.

Pregunta N. 6

¿Cree usted que mejoraría el desempeño en el trabajo al contar con estadísticas de control de calificaciones y emisión de permisos de funcionamiento?

Opciones	Frecuencia	Porcentaje
Totalmente	14	93%
Parcialmente	1	6%
Nada	0	-
TOTAL	15	100

Cuadro N.10
Elaborado por: Investigador

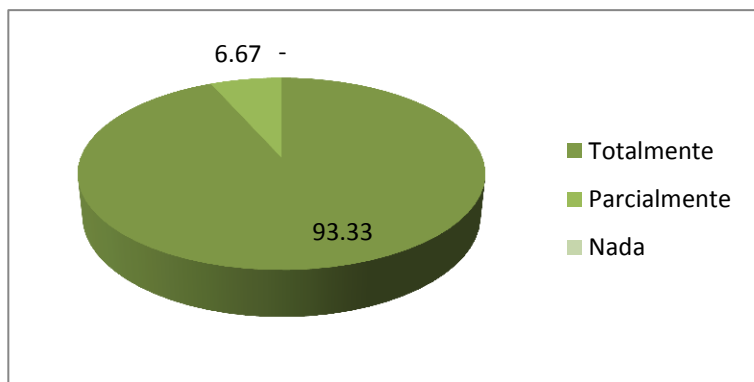


Gráfico N.11.:
Elaborado por: Investigador

ANÁLISIS E INTERPRETACIÓN

Se obtiene que el 93.33% responden que TOTALMENTE mejorará el desempeño en el trabajo al contar con estadísticas de control de calificaciones y emisión de permisos de funcionamiento, el 6.67% responden que PARCIALMENTE, mientras que el 0% responde que NADA.

Pregunta N. 7

¿Considera usted que la información procesada automática y ordenadamente de los establecimientos puede ayudar a realizar una mejor categorización?

Opciones	Frecuencia	Porcentaje
Totalmente	13	86%
Parcialmente	2	13%
Nada	0	-
TOTAL	15	100%

Cuadro N.11
Elaborado por: Investigador

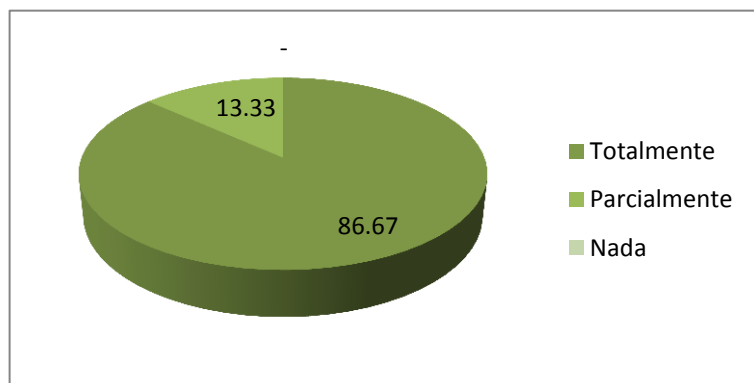


Gráfico N.12.:
Elaborado por: Investigador

ANÁLISIS E INTERPRETACIÓN

Se obtiene que el 86.67% responden que TOTALMENTE la información procesada automática y ordenadamente de los establecimientos puede ayudar a realizar una mejor categorización, el 13.3% responden que PARCIALMENTE, mientras que el 0% responde que NADA.

Pregunta N. 8

¿Cree usted que los resultados que arroje la minería de datos ayuden a determinar políticas de control en la emisión de permisos de funcionamiento?

Opciones	Frecuencia	Porcentaje
Totalmente	12	80%
Parcialmente	3	20%
Nada	0	-
TOTAL	15	100%

Cuadro N.12
Elaborado por: Investigador

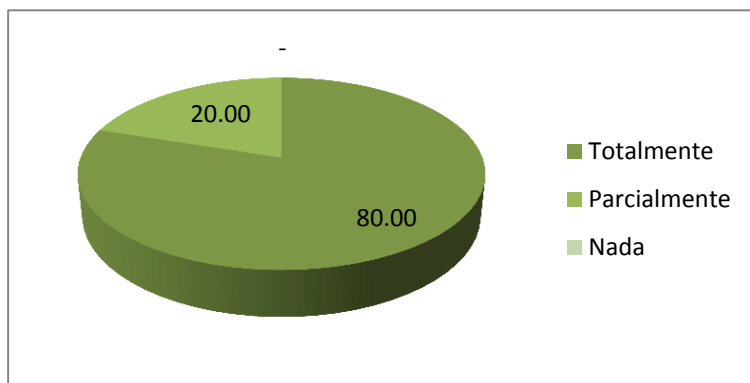


Gráfico N.13.:
Elaborado por: Investigador

ANÁLISIS E INTERPRETACIÓN

Se obtiene que el 80% responden que TOTALMENTE los resultados que arroje la minería de datos ayuden a determinar políticas de control en la emisión de permisos de funcionamiento, el 20% responden que PARCIALMENTE, mientras que el 0% responde que NADA.

4.3.- Verificación de la Hipótesis

La prueba de independencia Chi-cuadrado, nos permite determinar si existe una relación entre dos variables categóricas. Es necesario resaltar que esta prueba nos indica si existe o no una relación entre las variables, pero no indica el grado o el tipo de relación; es decir, no indica el porcentaje de influencia de una variable sobre la otra o la variable que causa la influencia.

Comprobación de la hipótesis con el Chi Cuadrado

Hipótesis Nula (H₀):

La implementación de un Data Mining *no incidirá en la* toma de decisiones del Catastro de establecimientos y la emisión de los Permisos de Funcionamiento por parte de la Dirección Provincial de Salud de Cotopaxi.

Hipótesis Alternativa H₁:

La implementación de un Data Mining *incidirá en la* toma de decisiones del Catastro de establecimientos y la emisión de los Permisos de Funcionamiento por parte de la Dirección Provincial de Salud de Cotopaxi.

Selección del nivel de significación

El nivel de significancia en la práctica, es frecuente la utilización del 0,05 o 0,01. Para nuestro cálculo utilizaremos el nivel de significancia del 0,05 (5%) al diseñar una regla de decisión, entonces hay unas cinco (5) oportunidades de entre cien (100) de rechazar la hipótesis cuando debiere ser aceptado; es decir, tenemos un 95% de confianza de que hemos adoptado por una decisión correcta. En tal caso decimos que la hipótesis ha sido rechazada al nivel de significancia de 0,05, lo que se interpreta que tal hipótesis una probabilidad de un 0,05 de ser falsa.

Especificación Estadística

Para la realización del presente trabajo se trabajará con una tabla consistente de dos filas por tres columnas proceso para la cual se utilizará la siguiente fórmula:

$$X^2 = \frac{\sum(O-E)^2}{E} \quad \text{Dónde:}$$

X^2 = Chi-cuadrado

Σ = Sumatoria

O = Frecuencia observada

E = Frecuencia esperada o teórica

Especificación de las regiones de Aceptación y Rechazo

Grado de libertad

Para el cálculo del grado de libertad se utilizará la siguiente fórmula:

$$G1 = (c-1) (h-1)$$

Dónde:

G1 = Grado de libertad

c = Columnas de la tabla

h = Filas de la tabla

Reemplazando tenemos:

$$G1 = (2-1) (3-1)$$

$$G1 = (1) (2)$$

$$G1 = 2$$

Recolección de datos y cálculos estadísticos

OPCIONES	Totalmente		Parcialmente		Nada		Total
	Fo	Fe	Fo	fe	Fo	Fe	
Pregunta N. 1: ¿Cuál es el grado de implementación de una solución de Minería de Datos en la Dirección Provincial de Salud de Cotopaxi que permita encontrar patrones faciliten el otorgamiento de permisos de funcionamiento de un establecimiento?	0	6	0	1,5	15	7,5	15
Pregunta N. 8: ¿Cree usted que los resultados que arroje la minería de datos ayuden a determinar políticas de control en la emisión de permisos de funcionamiento?	12	6	3	1,5	0	7,5	15
Total	12	12	3	3	15	15	30

Cuadro N.13 Frecuencia Observas y esperadas
Elaborado por: Investigador

Cálculo del chi cuadrado

FO	FE	Fo-Fe	(Fo-Fe) ²	(Fo-Fe) ² /Fe
0	6	-6	36	6
0	1,5	-1,5	2,25	1,5
15	7,5	7,5	56,25	7,5
12	6	6	36	6
3	1,5	1,5	2,25	1,5
0	7,5	-7,5	56,25	7,5
				30

Cuadro N.14 Cálculo del Chi Cuadrado
Elaborado por: Investigador

Verificación de la hipótesis

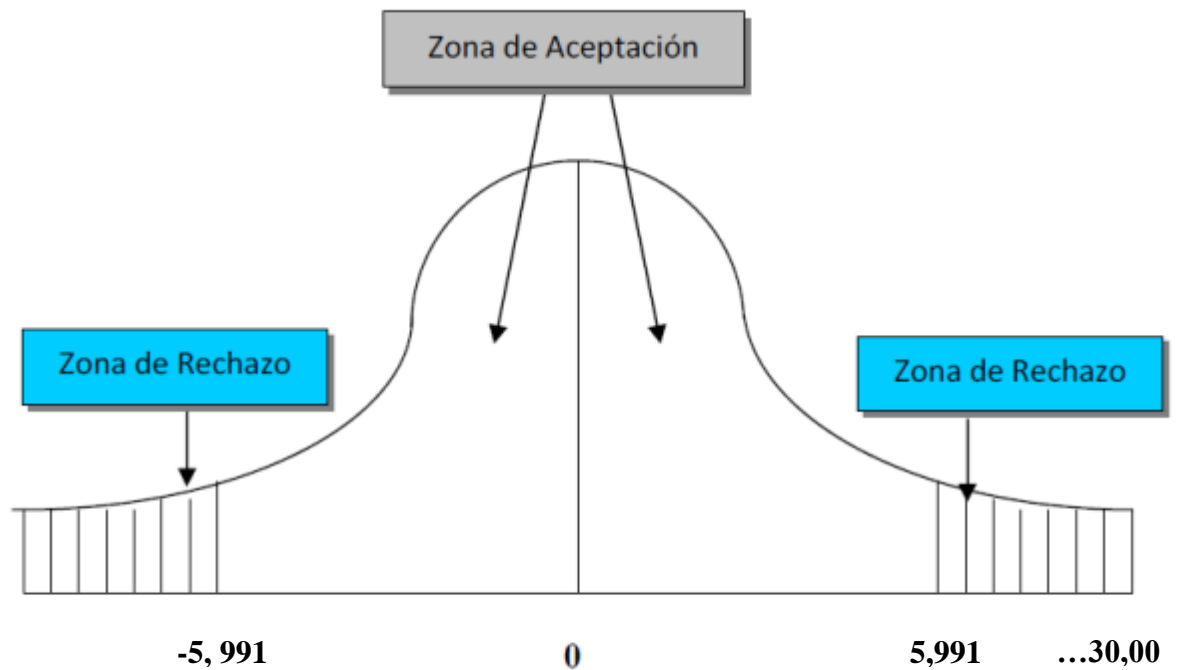


Gráfico N.14.: Chi Cuadrado
Elaborado por: Investigador

Análisis

Una vez realizado la prueba del Chi Cuadrado (X^2) se puede determinar que se rechaza la hipótesis nula (H_0) la misma que asevera que la implementación de un Data Mining *no incidirá en la* toma de decisiones del Catastro de establecimientos y la emisión de los Permisos de Funcionamiento por parte de la Dirección Provincial de Salud de Cotopaxi. Los valores obtenidos de X^2 según la tabla es de 30.00 que es mayor que el valor de los grados de libertad (5.991); ésta diferencia de valores nos indica que mientras más grande sea el valor de obtenido del X^2 con relación a los grados de libertad existe una gran y mayor asociatividad entre las variables dependiente e independiente, por lo tanto se acepta la hipótesis relativa la misma que expresa que la implementación de un Data Mining *incidirá en la* toma de decisiones del Catastro de establecimientos y la emisión de los Permisos de Funcionamiento por parte de la Dirección Provincial de Salud de Cotopaxi.

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

Luego de haber procesado toda la información obtenida a través de la encuesta y entrevistas realizadas a los funcionarios y a las autoridades de la Dirección Provincial de Salud de Cotopaxi, y en correlación con los objetivos específicos planteados se tienen las siguientes conclusiones:

- Se detectaron problemas como la carencia de datos en el catastro de establecimientos en tiempo real, limitante en la correcta toma de decisiones para la emisión de permisos de funcionamiento, lo anteriormente expuesto induce a la pérdida de tiempo y molestias a los usuarios.
- Existe gran cantidad de información que actualmente no esta siendo aprovechada en toda su dimensión, la misma que con un tratamiento y análisis adecuado ayudará a mejorar el proceso de toma de decisiones.
- Las nuevas tendencias tecnológicas apuntan a la utilización de sistemas complejos, con la finalidad de mejorar la calidad de servicios y la eficiencia en la entrega inmediata de la información.

Recomendaciones

1. Es recomendable manejar de manera adecuada el ingreso de la información al sistema de catastros para evitar datos inconsistentes e incompletos.
2. Es recomendable la realización y utilización de un Datamining para agilizar la recuperación y tratamiento de la información, con la finalidad de obtener patrones de datos y comportamientos de los mismos que ayuden a la toma de decisiones.
3. Se recomienda aplicar tecnología de punta en el manejo y procesamiento de información gerencial con el objetivo de mejorar la calidad de los servicios institucionales y atención a usuarios externos.

CAPÍTULO VI

LA PROPUESTA

6.1.- Datos Informativos

Nombre de la Institución:	Dirección Provincial de Salud de Cotopaxi
Provincia:	Cotopaxi
Parroquia:	La Matriz
Dirección:	Calle Dos de Mayo y Tarqui
Teléfono:	2813230 – 2813231
Beneficiarios:	Usuarios del Sistema Informático- usuarios externos
Tiempo Estimado:	Durante 1 año
Equipo Técnico Responsable:	Departamento de Sistemas.
Autor:	Ing. Hernán Carvajal Parra
Tutor:	Ing. Mg. Franklin Oswaldo Mayorga Mayorga.

6.2.- Antecedentes de la Propuesta

La Dirección Provincial de Salud es una institución que se dedica atender ramas de sanidad y asistencia social en salud a nivel nacional por lo que se ve en la necesidad de agilizar sus procesos y tomar predicciones, debido al gran volumen de información que se maneja

El principal problema que se ha encontrado, es la dificultad en la entrega oportuna y a tiempo de información y emitir reportes gerenciales del catastro para la toma de decisiones en la emisión de los permisos, esto se da debido a que los tiempos de respuesta para obtener información son muy largos por el gran volumen de datos que se maneja, ya que los mismos se encuentran dispersos, por lo que existe pérdida de tiempo y confusiones

Esta situación se presenta ya que los clientes no tienen un sistema que los ayude a presentar información y resultados reales a tiempo para una buena toma de decisiones por lo que los métodos actuales no son adecuados para minimizar o eliminar el problema.

Estado del arte

Luego de haber realizado la investigación acerca del problema se evidencia que la Dirección de Salud no posee un sistema que soporte el gran volumen de datos y a su vez pueda hacer predicciones para así tomar decisiones, esto se debe a que la información es llevada y manejada en un sistema obsoleto como es el FoxPro, por lo que las decisiones son tomadas empíricamente

6.3.-Justificación

Es **importante** realizar esta propuesta de creación de un **Data Mining** para impulsar el uso de nuevas tecnologías, con esto se obtiene correctas y acertadas decisiones en cuanto a la emisión de permisos de funcionamiento en la Dirección de Salud, con el propósito de optimizar procesos ofreciendo mejorar los servicios a los clientes y garantizando la agilidad de los procesos para satisfacción de sus usuarios.

El proponer la implementación de un Data Mining brindará la oportunidad de obtener información clara y oportuna para predecir y tomar acciones correctivas para el pago de permisos de funcionamiento, la cual será administrada por un encargado del área de sistemas, garantizando el manejo adecuado de los datos, tendrá las seguridades necesarias para que ningún usuario manipule los datos de manera inadecuada.

Es importante realizar esta propuesta ya que un Data Mining propondrá un conjunto de ideas, acciones organizativas y predictivas con los desarrollos tecnológicos para simplificar y normalizar los procedimientos administrativos facilitando trámites, ahorrando tiempo y dinero. Se debe tomar en cuenta que el

Data mining debe ser de fácil manejo y entendimiento, para todos sus usuarios, todas estas herramientas manejadas facilitarán los procesos y ayudarán a obtener resultados confiables y oportunos.

Los beneficios que aportará la implementación de un Data Mining favorecerá a los usuarios que arriban diariamente a la institución obteniendo información oportuna previo la obtención del permiso de funcionamiento, evitando de esta manera el aglutinamiento de trabajo y pérdida de tiempo, razón por la cual el beneficio directo se dará al personal que atiende en la institución, creando un ambiente amigable y facilitando las labores y procesos en los tramites diarios, elevando de esta manera el desarrollo del sector público, acreditando la rapidez en el manejo de datos considerando que lo más importante son las personas, más no la tecnología.

6.4.- Objetivos

Objetivo General

- Determinar e Implementar el Data Mining para la toma de decisiones del Catastro de establecimientos y la emisión de los Permisos de Funcionamiento por parte de la Dirección Provincial de Salud de Cotopaxi.

Objetivos Específicos

- Determinar el proceso para la toma de decisiones del Catastro de establecimientos y la emisión de los Permisos de Funcionamiento.
- Analizar las herramientas, tecnologías que engloban la creación de un Data Mining, estudiando los elementos que la conforman.
- Crear un Data Mining como alternativa de solución al problema de ausencia de información gerencial para la toma de decisiones en cuanto al catastro de establecimientos sujetos a vigilancia y control sanitario

6.5.- Análisis de Factibilidad

La aplicación de la propuesta establecida es factible porque se cuenta con el apoyo y colaboración de las autoridades que conforman la Dirección Provincial de Salud de Cotopaxi, quienes consideran que la propuesta es un aporte positivo para la entidad en virtud de que se podrá disponer de un instrumento orientador que permita posteriormente implementar la herramienta más óptima de minería de datos para contar con resultados importantes que servirán para la toma de decisiones que propendan a cumplir con metas y objetivos propuestos.

Factibilidad Económica: La Dirección Provincial de Salud de Cotopaxi, cuenta con una infraestructura informática apta para el análisis de las herramientas y su posterior implementación, es decir posee todos los equipos hardware y software por lo que la institución no tendrá que hacer inversión en la adquisición de recursos.

Factibilidad Técnica: La Dirección Provincial de Salud de Cotopaxi en la actualidad cuenta con la tecnología necesaria, la misma que permitirá viabilizar y hacer posible el desarrollo del estudio comparativo, ya que cuenta con:

- Sistema de emisión de permisos y catastros.
- Base de datos con licencias de Sql server 2005,
- Servidor, posee licencias de Windows 20003 server
- Capacitación al personal que estará a cargo de administrar el sistema
- Presupuesto necesario para la realización definitiva del data mining, y ampliar a más departamentos la nueva tecnología.

Factibilidad Operativa. Para llevar a efecto el desarrollo de la propuesta se cuenta con el apoyo incondicional del Director provincial de Salud, al igual que todo el recurso humano del departamento de Vigilancia Sanitaria.

Factibilidad Legal. Para el desarrollo del sistema no se ha encontrado ningún tipo de impedimento legal, ya que cuenta con la aprobación de las autoridades respectivas.

6.6.- Fundamentaciones

Filosófica

Para realizar el presente trabajo el investigador se ubica en el paradigma filosófico crítico propositivo porque cuestiona la manera de hacer investigación y por el contrario plantea una propuesta de solución al problema investigado basado en la existencia de múltiples realidades socialmente construidas

Además utilizaremos La metodología CRISP (CRoss-Industry Standard Processfor Data Mining) la cual es un proceso jerárquico formado por varias tareas que ofrece a las organizaciones la estructura necesaria para obtener mejores y más rápidos resultados en la minería de datos

Metodología de aplicación del Data Mining

En base a la experiencia adquirida en el análisis de los problemas se ha planteado la siguiente metodología para la implementación del sistema de emisión de permisos.

Esta metodología, diseñada para dar una guía a inexpertos en minería de datos, es un modelo genérico adaptable a las necesidades de una empresa o sector particular. El proceso de Data Mining se divide en las siguientes fases:

- Conocimiento del negocio.
- Conocimiento de los datos.
- Preparación de los datos.
- Modelado.
- Evaluación.
- Despliegue

Científico Técnico

Esta propuesta está basada en el manejo y estudio de un sistema de Data Mining para la emisión de permisos, para lo cual se utilizó la metodología de **CRISP** la cual se detalla a continuación:

Metodología de CRISP

La metodología CRISP-DM (Chapman, 1999) consta niveles de abstracción, organizados de forma jerárquica en tareas que van desde el nivel más general hasta los casos más específicos.

A nivel más general, el proceso está organizado en seis fases, estando cada fase a su vez estructurada en varias tareas generales de segundo nivel. Las tareas generales se proyectan a tareas específicas, donde se describen las acciones que deben ser desarrolladas para situaciones específicas. Así, si en el segundo nivel se tiene la tarea general “limpieza de datos”, en el tercer nivel se dicen las tareas que tienen que desarrollarse para un caso específico, como por ejemplo, “limpieza de datos numéricos”, o “limpieza de datos categóricos”. El cuarto nivel, recoge el conjunto de acciones, decisiones y resultados sobre el proyecto de Data Mining específico.

La metodología CRISP-DM proporciona dos documentos distintos como herramienta de ayuda en el desarrollo del proyecto de Data Mining: el modelo de referencia y la guía del usuario. El documento del modelo de referencia describe de forma general las fases, tareas generales y salidas de un proyecto de Data Mining en general. La guía del usuario proporciona información más detallada sobre la aplicación práctica del modelo de referencia a proyectos de Data Mining específicos, proporcionando consejos y listas de comprobación sobre las tareas correspondientes a cada fase.

La metodología CRISP-DM estructura el ciclo de vida de un proyecto de Data Mining en seis fases, que interactúan entre ellas de forma iterativa durante el desarrollo del proyecto.

1.- Fase de conocimiento del negocio o problema

La primera fase de la guía de referencia CRISP-DM, denominada fase de conocimiento del negocio o problema, es probablemente la más importante y aglutina las tareas de comprensión de los objetivos y requisitos del proyecto desde una perspectiva empresarial o institucional, con el fin de convertirlos en objetivos técnicos y en un plan de proyecto. Sin lograr comprender dichos objetivos, ningún algoritmo por muy sofisticado que sea, permitirá obtener resultados fiables. Para obtener el mejor provecho de Data Mining, es necesario entender de la manera más completa el problema que se desea resolver, esto permitirá recolectar los datos correctos e interpretar correctamente los resultados. En esta fase, es muy importante la capacidad de poder convertir el conocimiento adquirido del negocio, en un problema de Data Mining y en un plan preliminar cuya meta sea el alcanzar los objetivos del negocio. Una descripción de cada una de las principales tareas que componen esta fase es la siguiente:

Determinar los objetivos del negocio. Esta es la primera tarea a desarrollar y tiene como metas, determinar cuál es el problema que se desea resolver, por qué la necesidad de utilizar Data Mining y definir los criterios de éxito.

Identificar la falta y demora en el pago de la cuota de emisión de permisos, para el normal funcionamiento de los establecimientos.

Evaluación de la situación. En esta tarea se debe calificar el estado de la situación antes de iniciar el proceso de Data mining, considerando aspectos tales como: ¿cuál es el conocimiento previo disponible acerca del problema?, ¿se cuenta con la cantidad de datos requerida para resolver el problema?, etc. En esta fase se definen los requisitos del problema, tanto en términos de negocio como en términos de DataMining.

En la situación actual es imposible determinar patrones para justificar el comportamiento humano y saber por qué no pagan los permisos de funcionamiento

Se cuenta con los recursos financieros, tecnológicos y humanos para la realización de estos procesos, además se tiene datos históricos de la base de datos del sistema de catastros.

2.- Fase de conocimiento de los datos

Comprende la recolección inicial de datos, con el objetivo de establecer un primer contacto con el problema, familiarizándose con ellos, identificar su calidad y establecer las relaciones más evidentes que permitan definir las hipótesis. Esta fase junto a las próximas dos fases, son las que demandan el mayor esfuerzo y tiempo en un proyecto de Data Mining.

Las principales tareas a desarrollar en esta fase del proceso son:

Recolección de datos iniciales.

Los datos van hacer extraídos del sistema transaccional de catastros que se encuentra en visual FoxPro.

Los datos se obtuvieron del sistema de catastros la cual tenía una base de datos histórica, cabe recalcar que los datos tenían mucha información inconsistente como datos basura, datos blancos, en nulo, etc

Descripción de los datos.

Las tablas analizarse fueron: áreas, cantón, parroquia, tipo establecimiento, categorías, subcategoria2, subcategoria3, sexo, tipo de local, establecimiento y pagospf, esta última tabla es la transaccional y se tiene una cantidad de registros De todos los establecimientos a nivel de toda la provincia.

Exploración de datos.

A continuación se muestra el modelo de datos de acuerdo a la regla del negocio, que es identificar la demora en el pago de permisos de funcionamiento

Verificación de la calidad de los datos.

Los datos de la base transaccional de emisión de permisos se encuentran con inconsistencias, datos en blanco, datos nulos por lo que se ve la necesidad de crear procedimientos almacenados para verificar la calidad de los datos y poder tener buenos resultados para el data mining.

3.- Fase de preparación de los datos

En esta fase y una vez efectuada la recolección inicial de datos, se procede a su preparación para adaptarlos a las técnicas de Data Mining que se utilicen posteriormente, tales como técnicas de visualización de datos, de búsqueda de relaciones entre variables u otras medidas para exploración de los datos. La preparación de datos incluye las tareas generales de selección de datos a los que se va a aplicar una determinada técnica de modelado, limpieza de datos, generación de variables adicionales, integración de diferentes orígenes de datos y cambios de Formato.

4.- Fase de modelado

Se seleccionan las técnicas de modelado más apropiadas para el proyecto de Data Mining específico. Las técnicas a utilizar en esta fase se eligen en función de los siguientes criterios:

- Ser apropiada al problema.
- Disponer de datos adecuados.
- Cumplir los requisitos del problema.
- Tiempo adecuado para obtener un modelo.
- Conocimiento de la técnica.

Previamente al modelado de los datos, se debe determinar un método de evaluación de los modelos que permita establecer el grado de bondad de ellos. Después de concluir estas tareas genéricas, se procede a la generación y evaluación del modelo. Los parámetros utilizados en la generación del modelo dependen de las características de los datos y de las características de Precisión que se quieran lograr con el modelo.

En este caso se escogió el modelado de Arboles de decisiones, debido a que se centra en el objetivo de ver los datos más influyentes para el no pago de los permisos de funcionamiento de los establecimientos

5.- Fase de evaluación

En esta fase se evalúa el modelo, teniendo en cuenta el cumplimiento de los criterios de éxito del problema. Debe considerarse además, que la fiabilidad calculada para el modelo se aplica solamente para los datos sobre los que se realizó el análisis. Es preciso revisar el proceso, teniendo en cuenta los resultados obtenidos, para poder repetir algún paso anterior, en el que se haya posiblemente cometido algún error. Considerar que se pueden emplear múltiples herramientas para la interpretación de los resultados:

6.- Fase de implementación

Una vez que el modelo ha sido construido y validado, se transforma el conocimiento obtenido en acciones dentro del proceso de negocio, ya sea que el analista recomiende acciones basadas en la observación del modelo y sus resultados, ya sea aplicando el modelo a diferentes conjuntos de datos o como parte del proceso. Generalmente un proyecto de Data Mining no concluye en la implantación del modelo, pues se deben documentar y presentar los resultados de manera comprensible para el usuario, con el objetivo de lograr un incremento del conocimiento.

Para esto se seguirán los pasos que a continuación se detallan con las siguientes herramientas:

Selección de productos e instalación: se evalúa y selecciona cuales son los componentes necesarios específicos de la arquitectura (plataforma de hardware, motor de la BD, herramienta de ETL, etc.).

Se necesita para la realización del sistema Data mining lo siguiente:

Hardware: Un servidor con Windows 2003 server Sp2

Motor de base de datos :Microsoft Sql server 2005

Generación de ETL: Sql Server Business Intelligence Development Studio

Explotación de resultados Microsoft Excel con el complemento de minería de datos

Para la realización del data mining en el presente proyecto se escogió la Herramienta SQL Server 2005 porque tiene todas las herramientas necesarias unidas en el mismo paquete para realizar el proyecto planteado:

Herramienta *SQL SERVER* 2005

SQL Server 2005 es un sistema para la gestión de bases de datos producida por Microsoft basado en el modelo relacional. Microsoft SQL Server 2005 es una completa inteligencia de negocios que ofrece las características, herramientas y funcionalidades para crear tipos clásicos e innovadores de aplicaciones analíticas.

Las herramientas más utilizadas para el desarrollo y administración de SQL Server 2005 son:

- **Base de datos relacional:** Un motor de base de datos relacional más seguro, confiable, escalable y altamente disponible con el mejor rendimiento y compatible para datos estructurados y sin estructura (XML).
- **Servicios de replica:** Replica de datos para aplicaciones de procesamiento de datos distribuidas o móviles, alta disponibilidad de los sistemas, concurrencia escalable con almacenes de datos secundarios para soluciones de información empresarial e integración con sistemas heterogéneos, incluidas las bases de datos Oracle existentes.
- **NotificationServices:** Capacidades avanzadas de notificación para el desarrollo y el despliegue de aplicaciones escalables que pueden entregar actualizaciones de información personalizadas y oportunas a una diversidad de dispositivos conectados y móviles.

- **IntegrationServices:** Capacidades de extracción, transformación y carga (ELT) de datos para almacenamiento e integración de datos en toda la empresa.

Herramientas para extraer, transformar y cargar fuentes de datos.

- **Extracción:** Consiste en extraer los datos desde los sistemas de origen. Cada sistema separado puede usar una organización diferente de los datos o formatos distintos. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación.
- **Transformación:** Luego de que la información fue extraída, se pueden realizar diferentes pasos de transformación, como ser: limpieza de la información, botar a la basura lo que consideramos innecesario, seleccionar campos específicos que consideremos necesarios para el análisis del data mining, realizar combinaciones fuentes de datos, etc.
- **Carga:** Es el momento en el cual los datos de la fase anterior (transformar) son cargados en el sistema de destino. Dependiendo de los requerimientos de la organización, este proceso este proceso puede abarcar una amplia variedad de acciones diferentes. En algunas bases de datos se sobrescribe la información antigua con nuevos datos.

Herramientas de Sql Server

La plataforma de datos SQL Server incluye las siguientes herramientas:

- **Analysis Services:** Capacidades de procesamiento analítico en línea (OLAP) para el análisis rápido y sofisticado de conjunto de datos grandes y complejos, utilizando almacenamiento multidimensional.

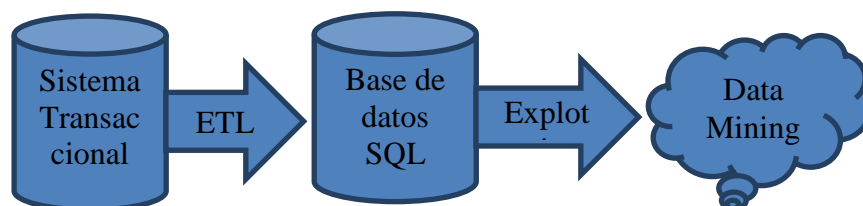


Gráfico N.15.: Analysis Services
Elaborado por: Investigador

- **Herramientas de administración:** incluye herramientas integradas de administración para administración y optimización avanzadas de bases de datos, así como también integración directa con otras herramientas. Los protocolos de acceso de datos estándar reducen drásticamente el tiempo que demanda integrar los datos en SQL Server con los sistemas existentes. Asimismo, el soporte de servicio Web nativo está incorporado en SQL Server para garantizar la interoperabilidad con otras aplicaciones y plataformas.
- **Herramientas de desarrollo:** SQL Server ofrece herramientas integrada de desarrollo para el motor de base de datos, extracción, transformación y carga de datos, minería de datos, OLAP e informes que están directamente integrados con Microsoft Visual Studio para ofrecer capacidades de desarrollo de aplicación de extremo a extremo. Cada subsistema principal en SQL Server se entrega con su propio modelo de objeto y conjunto de interfaces del programa de aplicación (API) para ampliar el sistema de datos en cualquier dirección que sea específica de su negocio.
- **Reporting service:** que es una solución global para crear, administrar y proporcionar tanto informes tradicionales orientados al papel como informes interactivos basados en la Web
- **Especificación de Aplicaciones para usuario finales:** se identifican los roles o perfiles de usuarios para los diferentes tipos de aplicaciones necesarias en base al alcance de los perfiles detectados.

Definición de ETL

ETL son las siglas en inglés de Extraer, Transformar y Cargar (Extract, Transform and Load).

Pasos para la creación de un ETL

Para la creación de un ETL se utiliza la herramienta Sql Server Business Intelligence Development Studio, para lo cual se detalla a continuación los siguientes pasos:

1.- Se crea un nuevo proyecto de Integration services.

2.- Se realizan los orígenes de datos, un origen y un destino, como se demuestra en el gráfico N° 16



Gráfico N.16: Orígenes de datos

Fuente: Análisis del Investigador
Elaborado por: Investigador

3.- Se crea un origen el cual es la base de datos original de la que se va a extraer la información, es decir la base de datos transaccional en Visual Fox Pro, como se demuestra en el gráfico N° 17

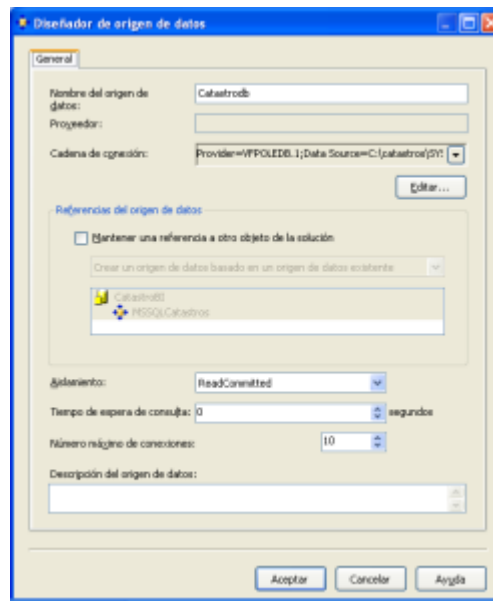


Gráfico N.17: Origen de Datos

Fuente: Análisis del Investigador
Elaborado por: Investigador

4.- Se procede a generar un destino, como se demuestra en el gráfico N° 18, que es en donde se pondrán los datos extraídos sin errores del paso anterior, esta es la base creada en sql server

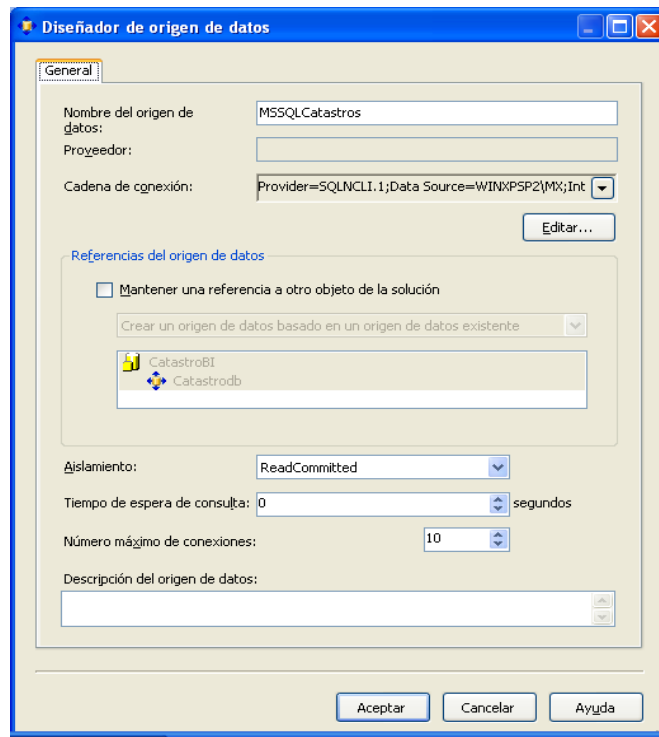


Gráfico N.18: Destino de los datos

Fuente: Análisis del Investigador
Elaborado por: Investigador

5.- Se crea paquetes SSIS para realizar las extracciones, transformaciones y carga:

De la barra de herramientas se escoge una tarea a ejecutar y una tarea de flujo de datos, como se demuestra en el gráfico N° 19

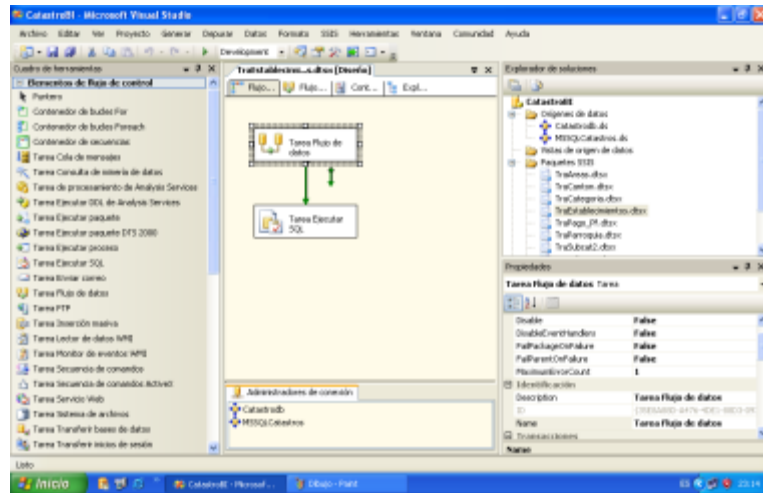


Gráfico N.19: Extracción de los datos

Fuente: Análisis del Investigador
 Elaborado por: Investigador

6.- En la tarea a ejecutar se realiza el llamado al procedimiento de campos adicionales, como se demuestra en el gráfico N° 20, con el fin de que la tabla donde se cargará la información se encuentre los mismos datos del origen

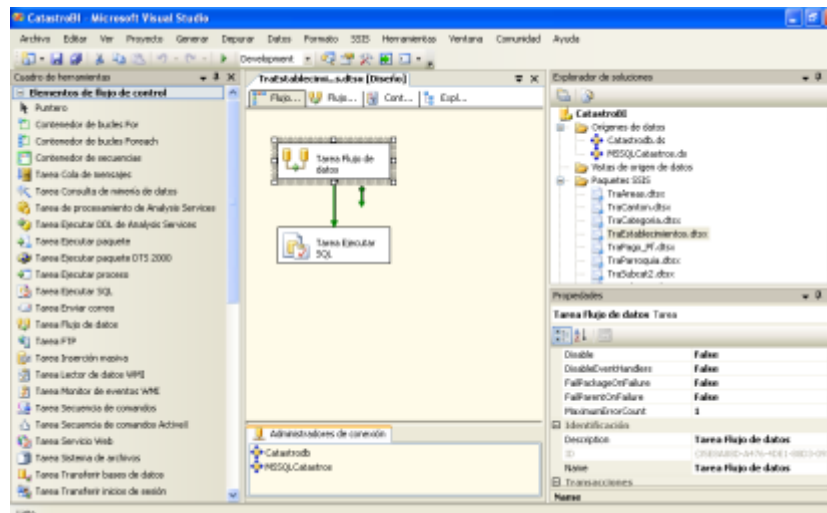


Gráfico N.20: Tarea ejecutar Sql

Fuente: Análisis del Investigador
 Elaborado por: Investigador

7.- En la tarea de flujo de datos se crea un origen de OLE DB, un destino de OLE DB y una columna derivada, como se demuestra en el gráfico N° 21:

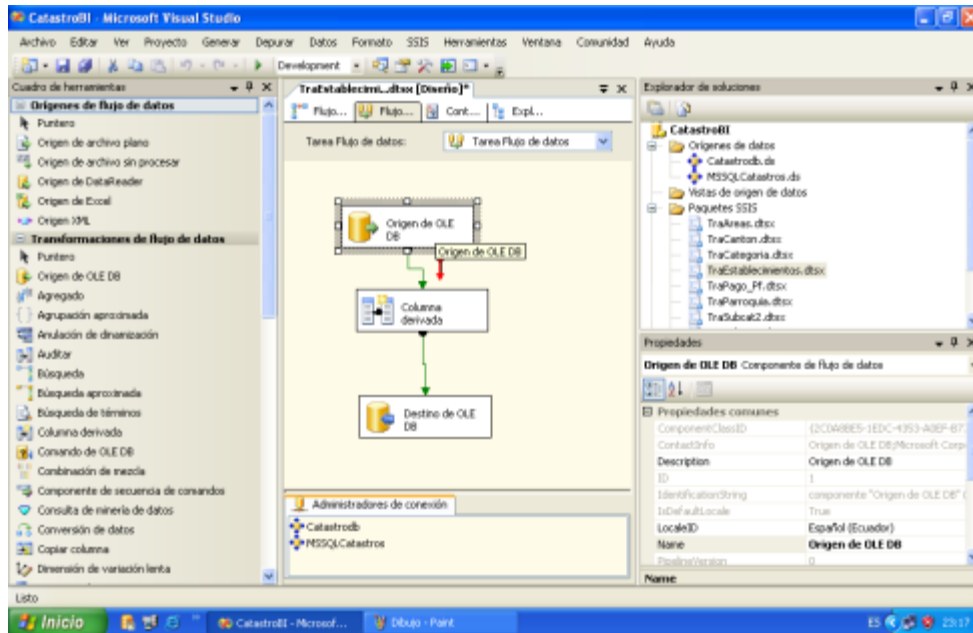


Gráfico N.21: Flujo de datos

Fuente: Análisis del Investigador
Elaborado por: Investigador

8.- En el origen de OLE DB se crea el comando Sql para extraer los datos de la base origen y pasar la información al destino comprobando que no existan datos nulos, blancos o incorrectos, como se demuestra en el gráfico N° 22 así:

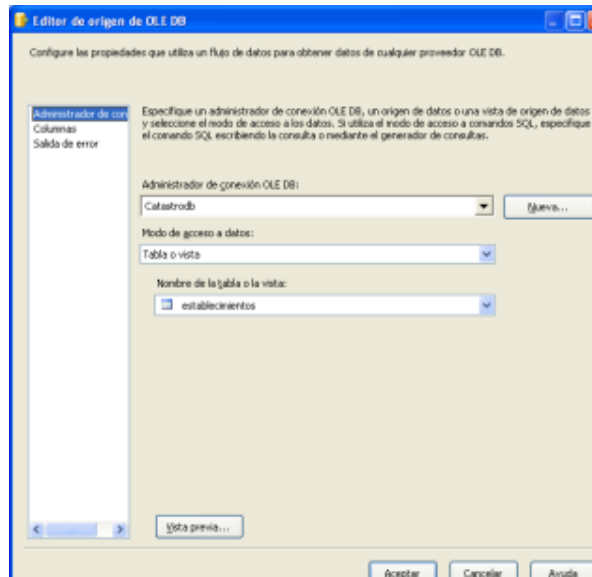


Gráfico N.22: Origen OLE DB

Fuente: Análisis del Investigador
 Elaborado por: Investigador

9.- En la columna deriva se pondrán los campos faltantes para poder copiar los datos del origen al destino y coincida la información del origen con el destino, como se demuestra en el gráfico N° 23

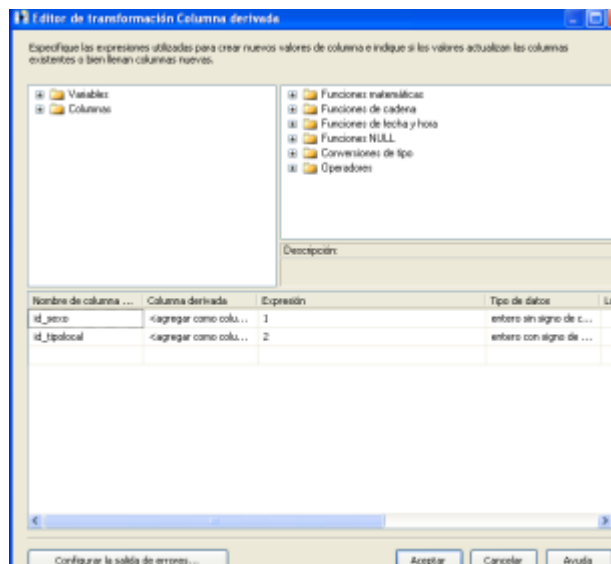


Gráfico N.23: Columna derivada

Fuente: Análisis del Investigador
 Elaborado por: Investigador

10.- En el destino se escoge la tabla en la cual se van a poner los datos extraídos en el paso anterior sin errores. Nótese que todos los campos del origen deben coincidir con su destino, como se demuestra en el gráfico N° 24:

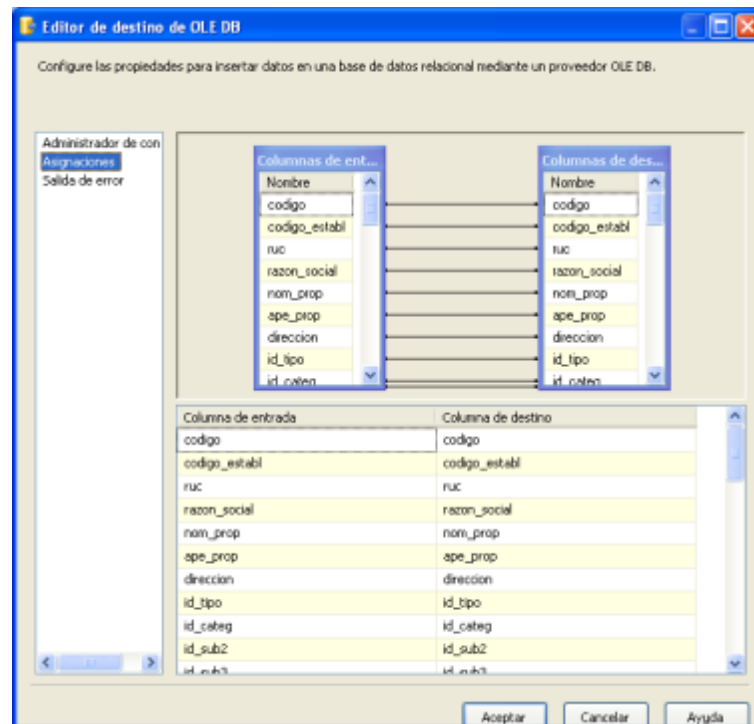


Gráfico N.24: Destino OLE DB

Fuente: Análisis del Investigador

Elaborado por: Investigador

Un parte intrínseca del proceso de extracción es la de analizar los datos extraídos, se realiza un chequeo, el cual verifica si los datos cumplen la pauta o estructura que se esperaba. Si no es así son rechazados.

11.- Para realizar la transformación se crea otro Paquete SSIS con el fin de mantener ordenada la información, para ello se escoge de la barra de herramientas una tarea de flujo de datos, como se demuestra en el gráfico N° 25:

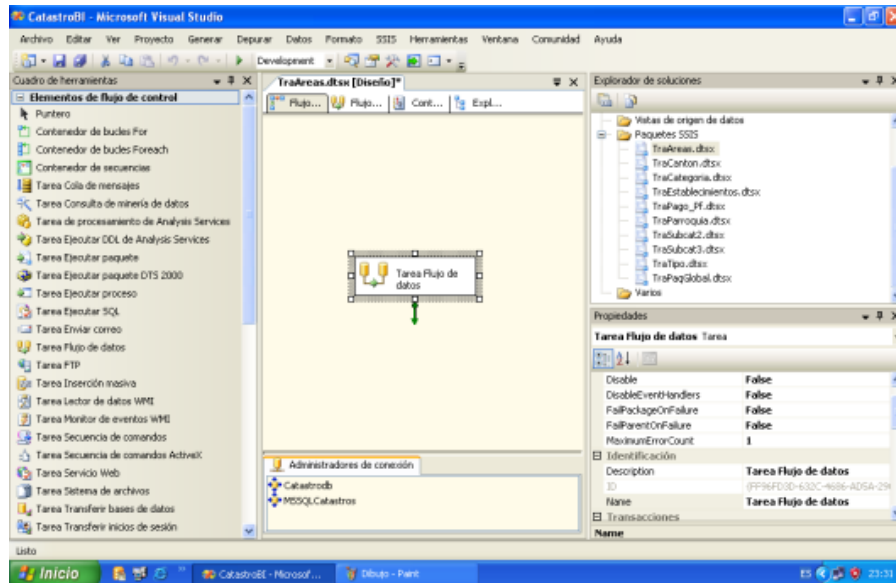


Gráfico N.25: Transformación de datos

Fuente: Análisis del Investigador
 Elaborado por: Investigador

12.- Dentro de la tarea de flujo de datos se pondrá un origen de OLEDB y una dimensión de variación lenta, como se demuestra en el gráfico N° 26:

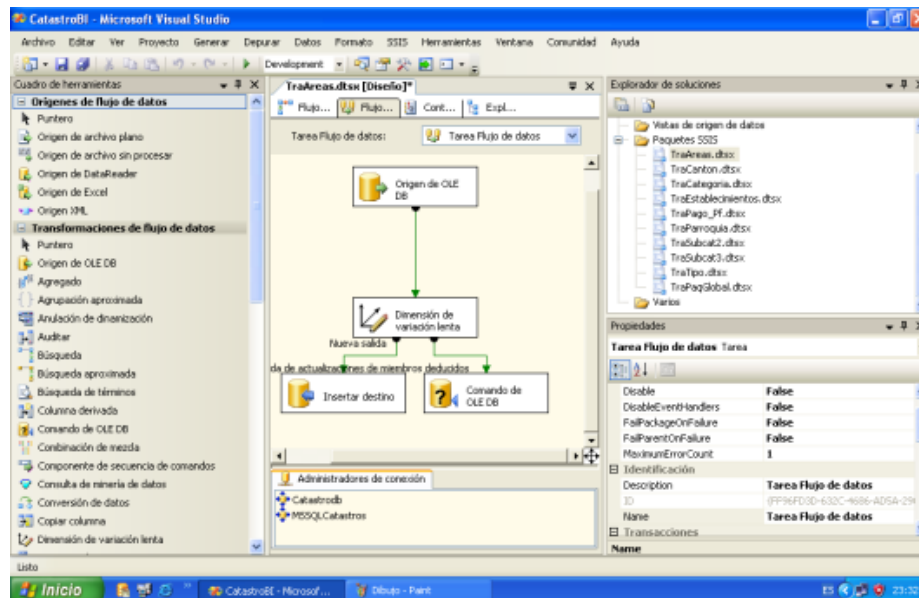


Gráfico N.26: Flujo de datos de transformación

Fuente: Análisis del Investigador
 Elaborado por: Investigador

13.- En el origen de OLEDB se seleccionará los campos del origen que se enviarán al destino, como se demuestra en el gráfico N° 27

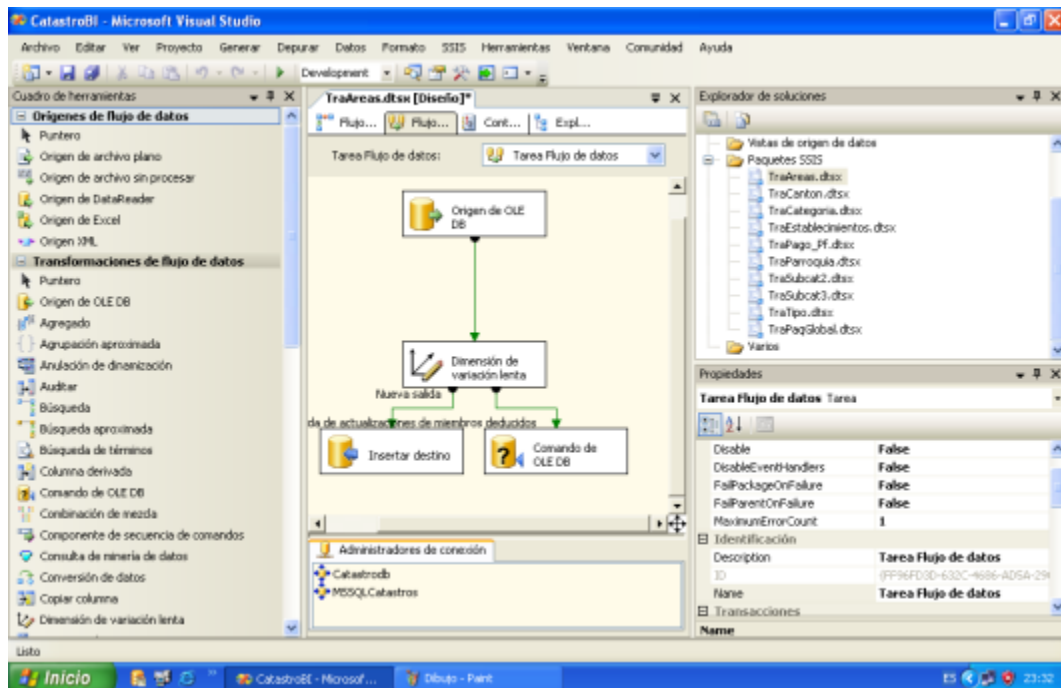


Gráfico N.27: Origen OLE DB de transformación

Fuente: Análisis del Investigador
Elaborado por: Investigador

14.- En la dimensión de variación lenta se seleccionará las columnas de entrada, las columnas de dimensión y el tipo de clave , como se demuestra en el gráfico N° 28

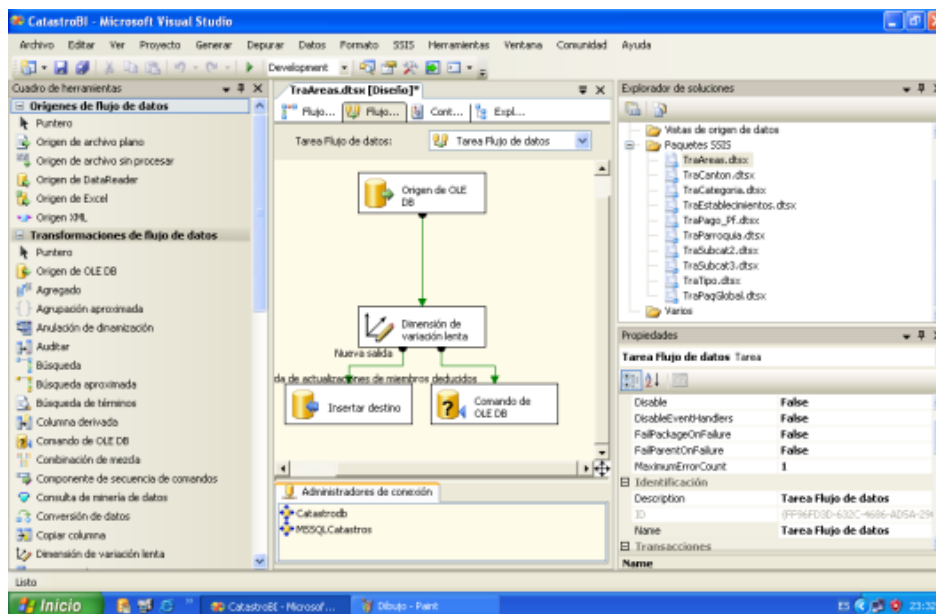


Gráfico N.28: Dimensión de variación lenta

Fuente: Análisis del Investigador
 Elaborado por: Investigador

15.- Se procede a dar Clic en Siguiente, se selecciona las columnas de dimensión que van hacer de atributos variables, como se demuestra en el gráfico N° 29

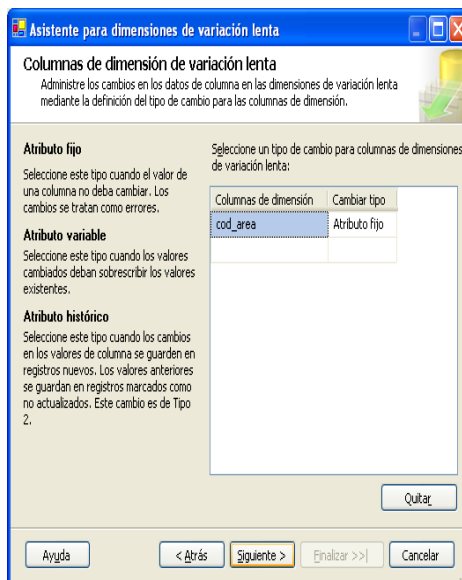


Gráfico N.29: Atributos variable

Fuente: Análisis del Investigador
 Elaborado por: Investigador

16.- Pulsar Clic en siguiente y seleccionar Cambiar todos los registros coincidentes, incluidos todos los registros no actualizados, cuando se detecten cambios en un atributo variable, como se demuestra en el gráfico N° 30:

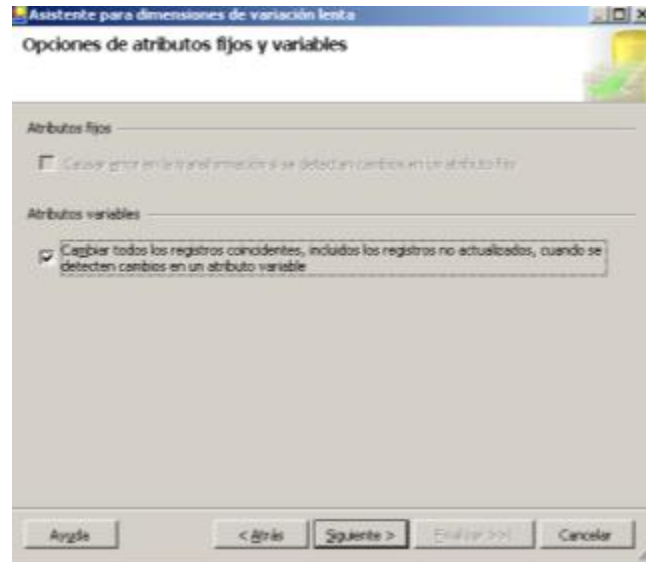


Gráfico N.30: Atributos y tipos de variables

Fuente: Análisis del Investigador
Elaborado por: Investigador

17.- Pulsar Clic en siguiente y se habilitará la compatibilidad con miembros deducidos, como se demuestra en el gráfico N° 31:

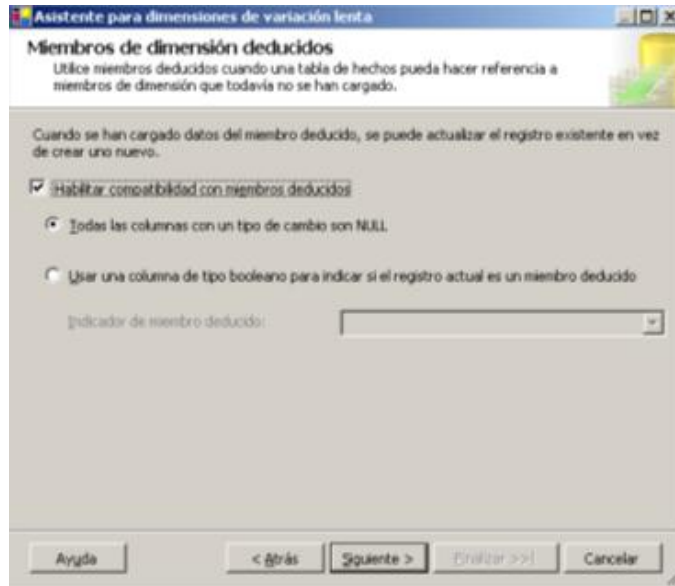


Gráfico N.31: compatibilidad con miembros deducidos

Fuente: Análisis del Investigador
 Elaborado por: Investigador

18.- Pulsar Clic en siguiente, como se demuestra en el gráfico N° 32:



Gráfico N.32: Finalizar asistente de variación lenta

Fuente: Análisis del Investigador
 Elaborado por: Investigador

19.- **Desarrollo de aplicaciones para usuario finales:** El desarrollo involucra configuraciones de los metadatos y construcción de reportes específicos.

Para realizar los reportes se utilizará la Herramienta del Sql Server 2005 denominada Análisis Service que es una solución global para crear, administrar y proporcionar informes de predicción, con la ayuda de Microsoft Excel para realizar una aplicación sencilla de fácil uso y entendimiento.

A continuación se describirá los principales pasos que se deben llevaron a cabo

Se crearon las siguientes vistas:

Se creará una vista de datos Pibotante de las tablas del sistema incluyendo la mayor cantidad de campos posibles de análisis de minería.

```
USE [catastros]
```

```
GO
```

```
SETANSI_NULLSON
```

```
GO
```

```
SETQUOTED_IDENTIFIERON
```

```
GO
```

```
CREATEVIEW [dbo].[VPivotanteMineria]
```

```
AS
```

```
SELECT areas.id_area,  
        areas.nombre as AreaNombre,  
        areas.cod_area,  
        areas.ruc,  
        areas.cuenta,  
        areas.nombre_cta,  
        areas.cuenta2,  
        areas.nombre_cta2,  
        canton.id_canton,
```

canton.nombre as CantonNombre,
parroquias.id_parro,
parroquias.nombre as ParroquiaNombre,
tipo.id_tipo,
tipo.nombre as TipoNombre,
tipo.codigo as TipoCodigo,
tipo.coeficient as TipoCoeficient,
tipo.derechos as TipoDerechos,
tipo.cabecera as TipoCabecera,
tipo.ambito as TipoAmbito,
categoria.id_categ,
categoria.nombre as CategNombre,
categoria.codigo as CategCodigo,
categoria.coeficient as CategCoeficient,
categoria.derechos as CategDerechos,
categoria.actividad as CategActividad,
categoria.ambito as CategAmbito,
subcat2.id_sub2,
subcat2.nombre as SubCat2Nombre,
subcat2.codigo as SubCat2Codigo,
subcat2.coeficient as SubCat2Coeficient,
subcat2.derechos as Subcat2Derechos,
subcat2.cabecera as Subcat2Cabecera,
subcat2.actividad as Subcat2Actividad,
subcat3.id_sub3,
subcat3.nombre as SubCat3Nombre,
subcat3.codigo as SubCat3Codigo,
subcat3.coeficient as Subcat3Coeficient,
subcat3.derechos as SubCat3Derechos,
subcat3.actividad as SubCat3Actividad,
establecimientos.codigo as EstablCodigo,
establecimientos.ruc as EstablRuc,

establecimientos.razon_social,
establecimientos.nom_prop,
establecimientos.ape_prop,
establecimientos.direccion,
establecimientos.actividad as EstablActividad,
establecimientos.telefono as EstablTelf,
establecimientos.transportar,
establecimientos.producir,
establecimientos.vender,
establecimientos.importar,
establecimientos.otros,
establecimientos.descrip_otros,
establecimientos.calificado,
establecimientos.exonerado,
establecimientos.estado,
establecimientos.nuevo,
sexo.descripcion as DescripSexo,
tipolocal.descriplocal

FROM areas,

canton,

parroquias,

tipo,

categoria,

subcat2,

subcat3,

sexo,

tipolocal,

establecimientos

Where areas.id_area = canton.id_area and

canton.id_canton = parroquias.id_canton and

tipo.id_tipo = categoria.id_tipo and

categoria.id_categ = subcat2.id_categ and

```

        subcat2.id_sub2 = subcat3.id_sub2 and
        areas.id_area = establecimientos.id_area and
        canton.id_canton = establecimientos.id_canton and
        parroquias.id_parro = establecimientos.id_parro and
        tipo.id_tipo = establecimientos.id_tipo and
        subcat2.id_sub2 = establecimientos.id_sub2 and
        subcat3.id_sub3 = establecimientos.id_sub3 and
        sexo.id_sexo = establecimientos.id_sexo and
        tipolocal.id_tipolocal = establecimientos.id_tipolocal

```

Crear una vista de datos que determine si un establecimiento a pagado o no.

```
USE [catastros]
```

```
GO
```

```
SETANSI_NULLSON
```

```
GO
```

```
SETQUOTED_IDENTIFIERON
```

```
GO
```

```
CREATEVIEW [dbo].[VPagoEstablecimeintos]
```

```
AS
```

```

        SELECT establecimientos.codigo,
               (Select(CaseWhencount(*)> 0 Then 1 Else 0 End) ExiPago
               From pago_pf
               Where establecimientos.codigo = pago_pf.codigo)as

```

```
ExiPagoEstabl
```

```
FROM establecimientos
```

Crear una vista de datos para el análisis de minería de datos, tratando de encontrar el o los factores más influyentes para que un establecimiento pague los permisos de funcionamiento o no.

```
USE [catastros]
```

```
GO
```

```
SETANSI_NULLSON
```

```
GO
```

```
SETQUOTED_IDENTIFIERON
```

```
GO
```

```
CREATEVIEW [dbo].[VFacInfPagoEstablEstado]
```

```
AS
```

```
SELECT    dbo.VPagoEstablecimientos.ExiPagoEstabl,  
          dbo.VPivotanteMineria.nuevo, dbo.VPivotanteMineria.calificado,  
          dbo.VPivotanteMineria.exonerado,  
          dbo.VPivotanteMineria.TipoNombre, dbo.VPivotanteMineria.CategNombre,  
          dbo.VPivotanteMineria.estado  
FROM      dbo.VPagoEstablecimientos INNERJOIN  
          dbo.VPivotanteMineria ON dbo.VPagoEstablecimientos.codigo =  
          dbo.VPivotanteMineria.EstablCodigo
```

Crear una vista de datos que determine si un establecimiento a pagado o no para el sistema transaccional.

```
USE [catastros]
```

```
GO
```

```
SETANSI_NULLSON
```

```
GO
```

```
SETQUOTED_IDENTIFIERON
```

```
GO
```

```
CREATEVIEW [dbo].[VFacInfPagoEstablTrans]
```

```
AS
```

```
SELECT    dbo.VPagoEstablecimientos.ExiPagoEstabl,  
          dbo.VPivotanteMineria.nuevo, dbo.VPivotanteMineria.calificado,  
          dbo.VPivotanteMineria.exonerado,  
          dbo.VPivotanteMineria.TipoNombre, dbo.VPivotanteMineria.CategNombre,  
          dbo.VPivotanteMineria.estado, dbo.VPivotanteMineria.transportar,
```



```

        dbo.VPivotanteMineria.producir, dbo.VPivotanteMineria.vender,
        dbo.VPivotanteMineria.importar, dbo.VPivotanteMineria.DescripSexo,
        dbo.VPivotanteMineria.descripload
FROM      dbo.VPagoEstablecimientos INNERJOIN
        dbo.VPivotanteMineria ON dbo.VPagoEstablecimientos.codigo =
        dbo.VPivotanteMineria.EstablCodigo

```

20 Luego se instalará el complemento SQLServer2005_DMAddinMineria.msi, que es una herramienta para la explotación del datamining para Excel.

21.- Para comprobar si se instaló correctamente se abre Excel y aparecerá una pestaña que llamada Minería de Datos, como se demuestra en el gráfico N° 33



Gráfico N.33: Complemento minería de Datos

Fuente: Análisis del Investigador
 Elaborado por: Investigador

22.- Una vez instalado el complemento ubicarse en el menú datos y escoger la opción “De otras fuentes”, para luego seleccionar la opción “Desde SQL server”, como se demuestra en el gráfico N° 34:

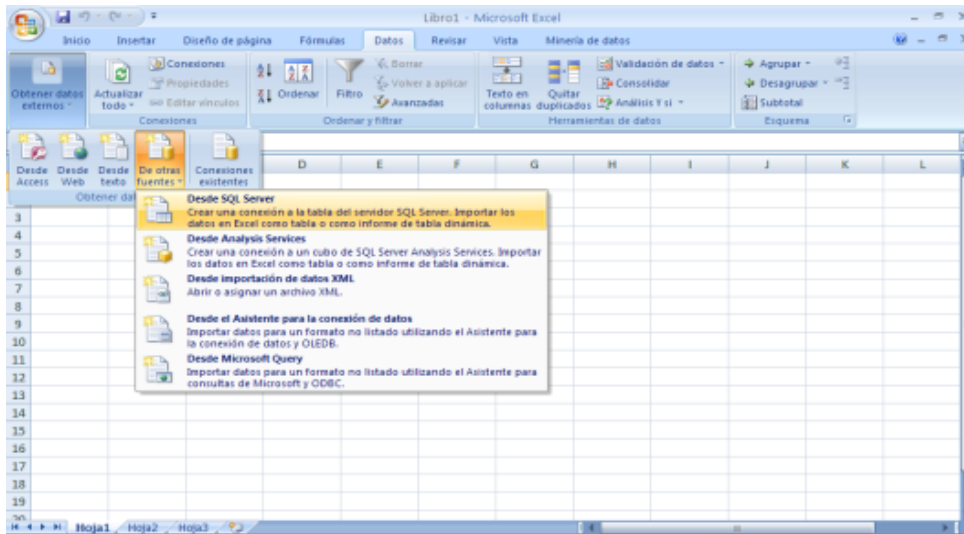


Gráfico N.34: Desde SQL Server

Fuente: Análisis del Investigador
 Elaborado por: Investigador

23.- Aparecerá el asistente de minería de conexión de datos, verificar que este seleccionado Microsoft SQL server y clic en siguiente, como se demuestra en el gráfico N° 35

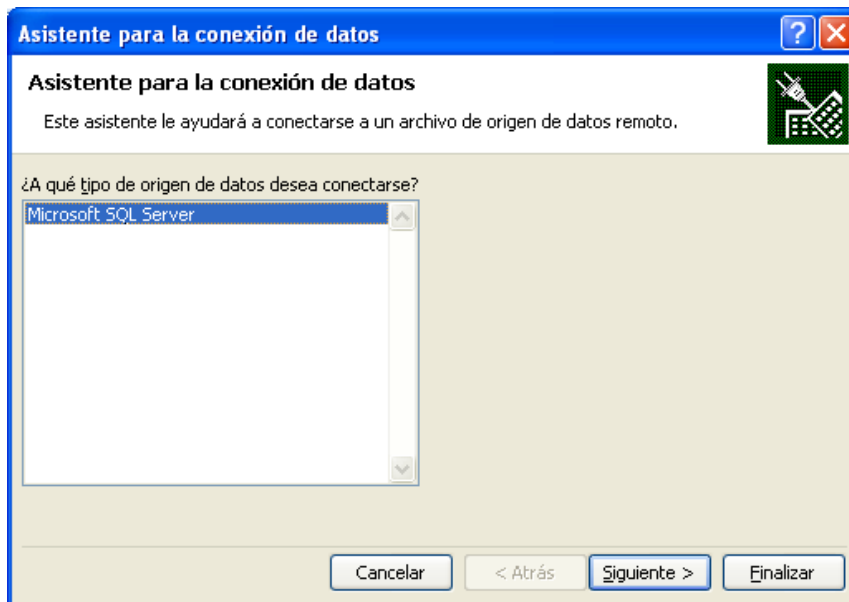


Gráfico N.35: Asistente conexión de datos

Fuente: Análisis del Investigador
 Elaborado por: Investigador

24.- Luego seleccionar la instancia de SQL server en la cual se creó la base y se cargaron los datos con los procesos ETL, clic en siguiente, como se demuestra en el gráfico N° 36

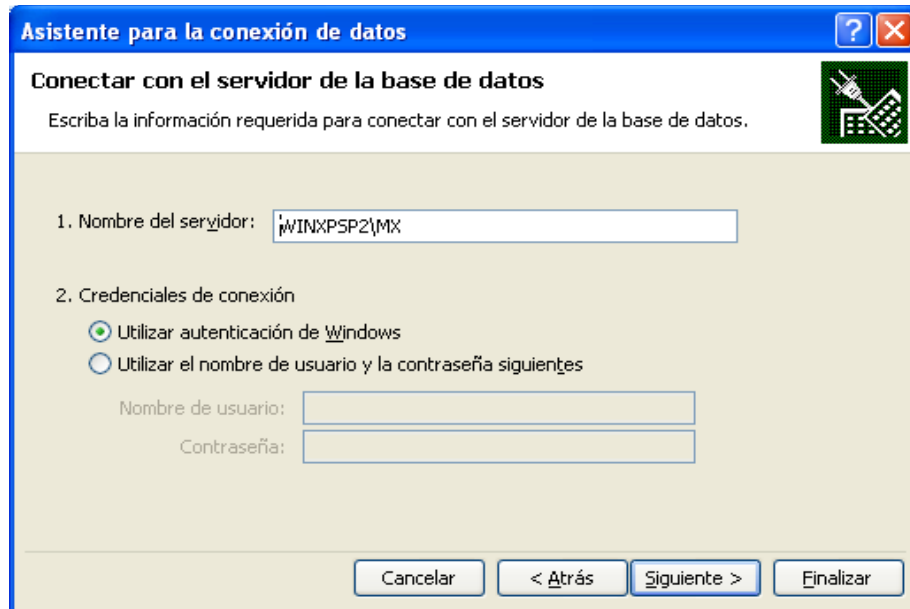


Gráfico N.36: Conexión instancia del servidor

Fuente: Análisis del Investigador
Elaborado por: Investigador

25.- Luego se selecciona la base de datos y la tabla o vista sobre la cual se va a trabajar, clic en siguiente, como se demuestra en el gráfico N° 37

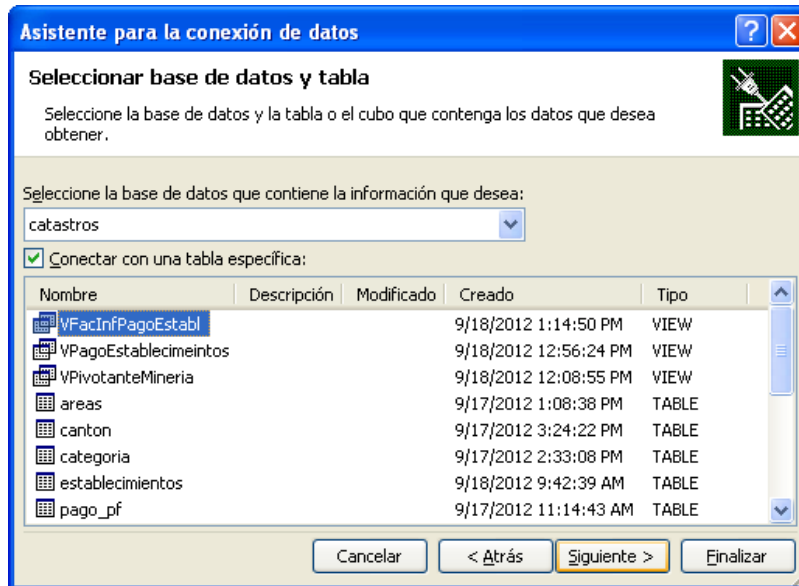


Gráfico N.37: Selección de la vista o tabla

Fuente: Análisis del Investigador
 Elaborado por: Investigador

26.- Se dará un nombre al reporte y clic en finalizar, como se demuestra en el gráfico N° 38

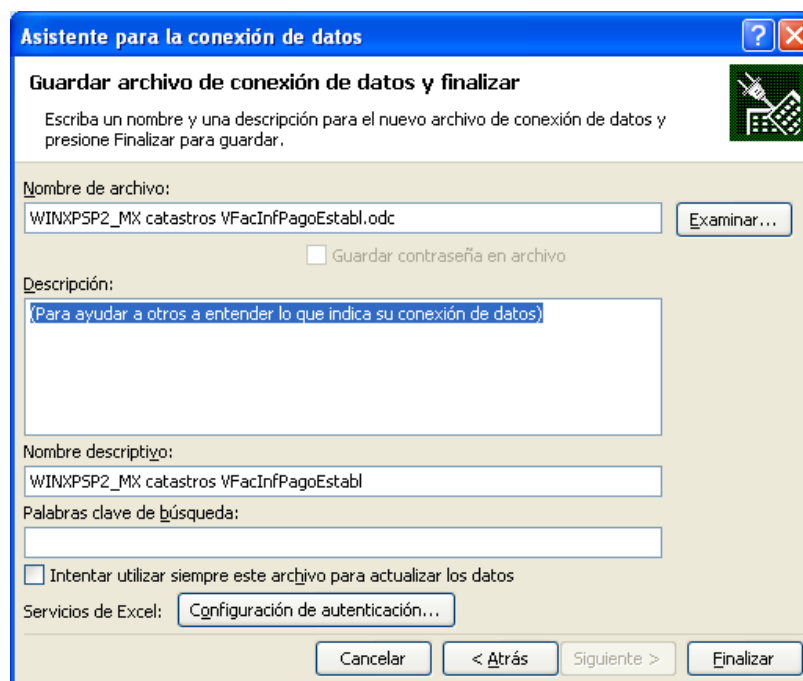


Gráfico N.38: Guardar archivos de conexión de datos

Fuente: Análisis del Investigador
 Elaborado por: Investigador

27.- El siguiente paso es importar los datos al Excel para lo cual se escoge la tabla, clic en aceptar, como se demuestra en el gráfico N° 39

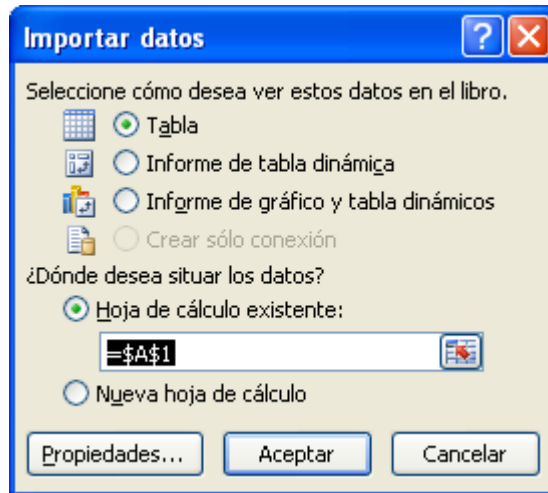


Gráfico N.39: Importar datos a Excel

Fuente: Análisis del Investigador
Elaborado por: Investigador

28.- Si todo esta correcto los datos se importarán al Excel de la forma como se demuestra en el gráfico N° 40

ExIPagoEstabl	AreaNombre	CantonNombre	ParroquiaNombre
1	Area 1 Latacunga, Centro de Salud	Latacunga	Ignacio Flores
2	Area 1 Latacunga, Centro de Salud	Latacunga	Ignacio Flores
3	Area 1 Latacunga, Centro de Salud	Latacunga	Ignacio Flores
4	Area 1 Latacunga, Centro de Salud	Latacunga	Ignacio Flores
5	Area 4 Saquisilí, Centro De Salud	Saquisilí	Saquisilí
6	Area 4 Saquisilí, Centro De Salud	Saquisilí	Saquisilí
7	Area 4 Saquisilí, Centro De Salud	Saquisilí	Saquisilí
8	Area 4 Saquisilí, Centro De Salud	Saquisilí	Saquisilí
9	Area 4 Saquisilí, Centro De Salud	Saquisilí	Saquisilí
10	Area 4 Saquisilí, Centro De Salud	Saquisilí	Saquisilí
11	Area 4 Saquisilí, Centro De Salud	Saquisilí	Saquisilí
12	Area 4 Saquisilí, Centro De Salud	Saquisilí	Saquisilí
13	Area 4 Saquisilí, Centro De Salud	Saquisilí	Saquisilí
14	Area 4 Saquisilí, Centro De Salud	Saquisilí	Saquisilí
15	Area 4 Saquisilí, Centro De Salud	Saquisilí	Saquisilí
16	Area 4 Saquisilí, Centro De Salud	Saquisilí	Saquisilí
17	Area 1 Latacunga, Centro de Salud	Latacunga	Ignacio Flores
18	Area 1 Latacunga, Centro de Salud	Latacunga	Ignacio Flores
19	Area 1 Latacunga, Centro de Salud	Latacunga	Ignacio Flores

Gráfico N.40: Datos en Excel

Fuente: Análisis del Investigador
 Elaborado por: Investigador

Árbol de Decisiones

29.- Para realizar la minería en el menú Dataming escoger la opción Clasificar, como se demuestra en el gráfico N° 41

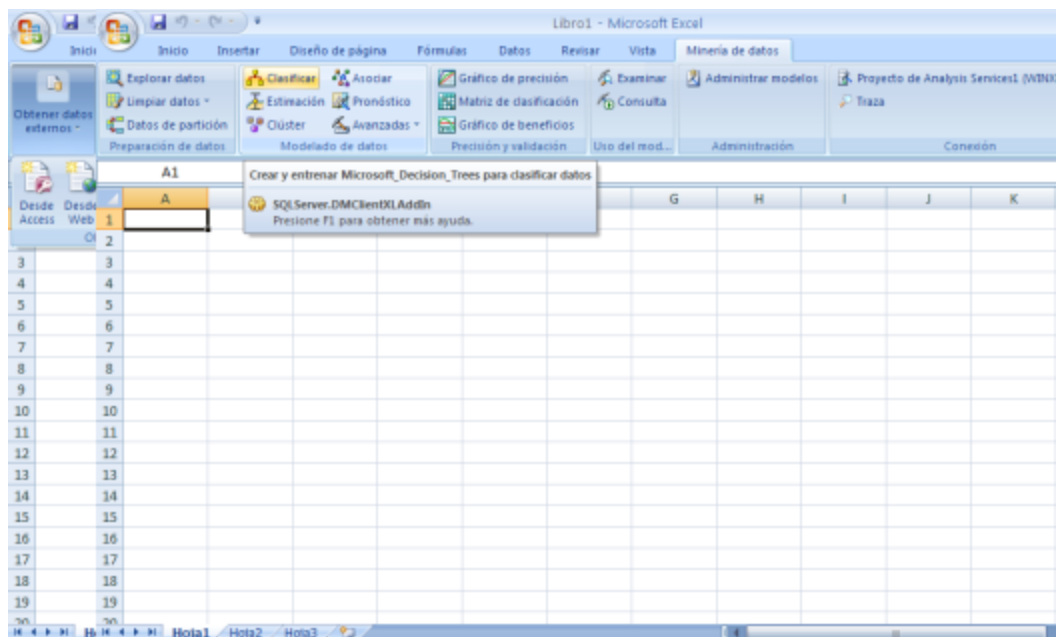


Gráfico N.41: Clasificar

Fuente: Análisis del Investigador
 Elaborado por: Investigador

30.- Aparecerá el asistente para clasificar, clic en siguiente, como se demuestra en el gráfico N° 42

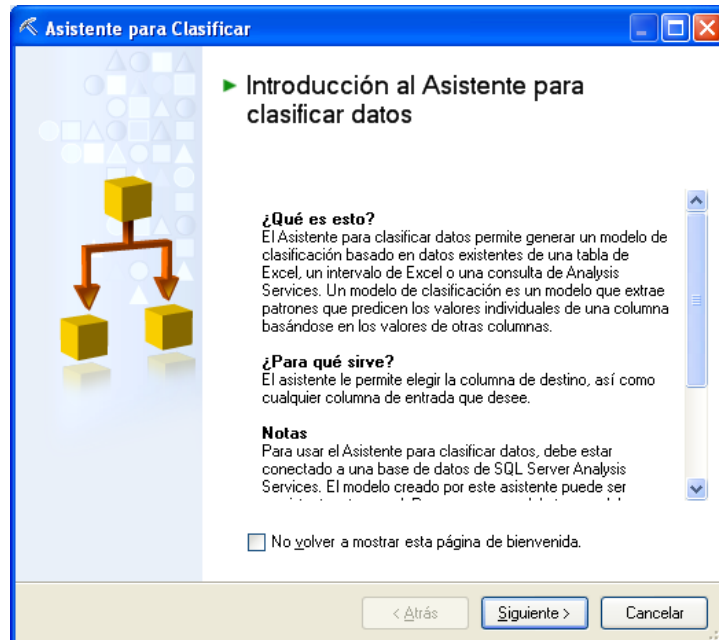


Gráfico N.42: Asiste de clasificación de datos

Fuente: Análisis del Investigador
Elaborado por: Investigador

31.- Seleccionar los datos del origen, estos se pondrán por defecto, clic en siguiente, como se demuestra en el gráfico N° 43

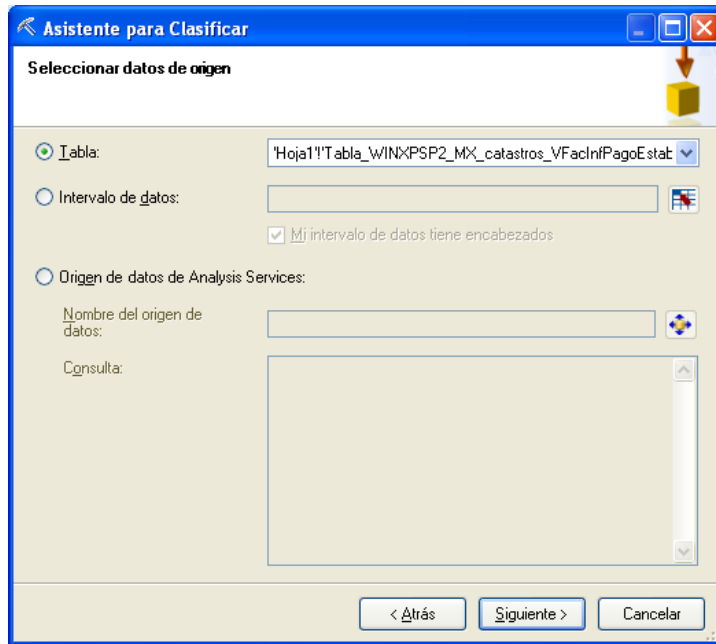


Gráfico N.43: Selección datos de origen

Fuente: Análisis del Investigador
 Elaborado por: Investigador

32.- Escoger los campos por los que se clasificará el data mining para que genere el árbol de decisión clic en siguiente, como se demuestra en el gráfico N° 44

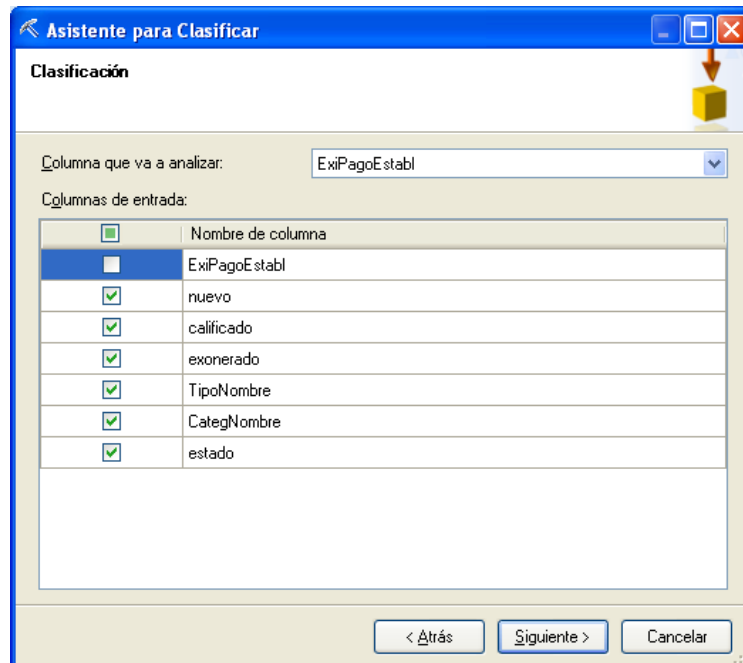


Gráfico N.44: Selección de columnas de predicción

Fuente: Análisis del Investigador
 Elaborado por: Investigador

33.- Se observará un resumen del árbol de decisión a crear, clic en finalizar, como se demuestra en el gráfico N° 45

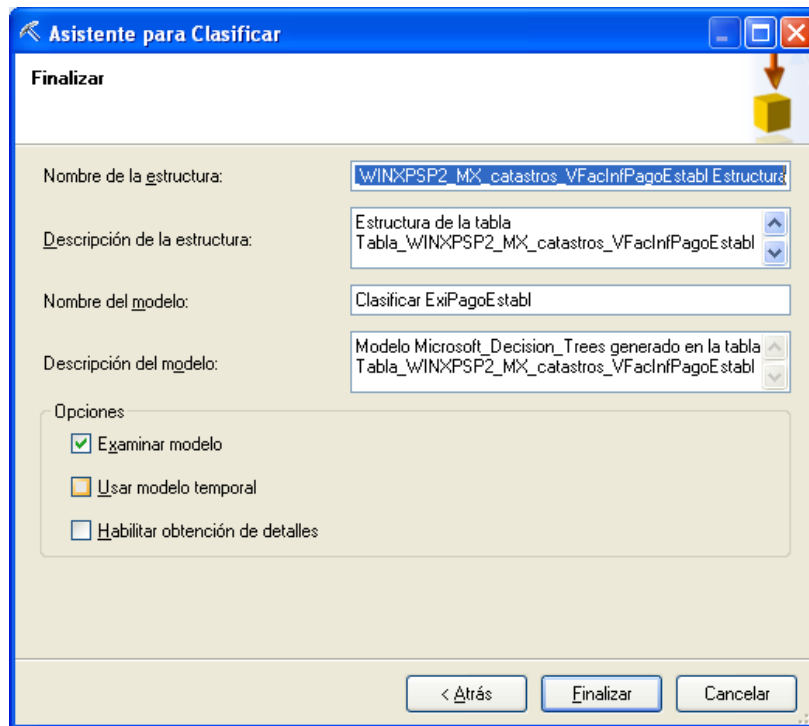


Gráfico N.45: Finalizar asistente de clasificación

Fuente: Análisis del Investigador
Elaborado por: Investigador

34.- Si todo esta correcto saldrá una pantalla como esta en donde se está generando el árbol de decisión, como se demuestra en el gráfico N° 46

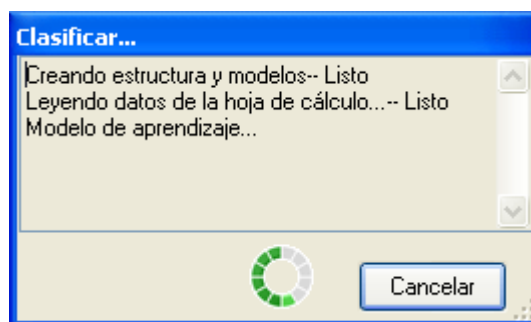


Gráfico N.46: procesando información de predicción

Fuente: Análisis del Investigador
Elaborado por: Investigador

Algunos de los escenarios creados son:

Escenario = Identificar la demora en el pago de permiso de funcionamiento.

Total de Escenarios = 56480, de los cuales 37824 han pagado la cuota y 18656 no lo han hecho.

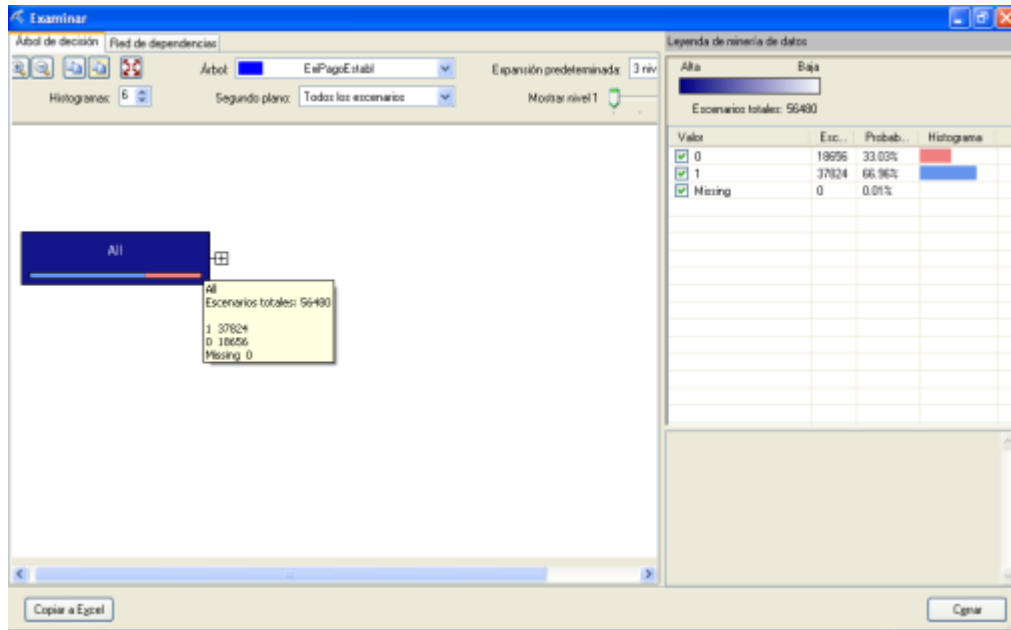


Gráfico N.47: Escenario 1

Fuente: Análisis del Investigador
Elaborado por: Investigador

Explorando más adentro identificamos que los que menos pagan, son los establecimientos cerrados, como se demuestra en el gráfico N° 48

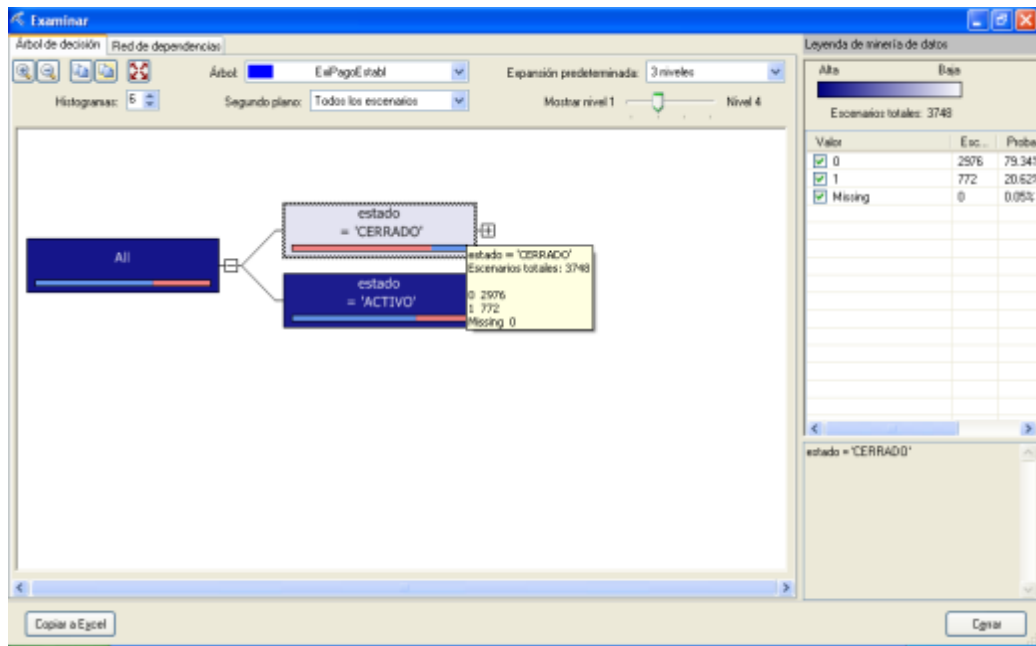


Gráfico N.48: Escenario 1.1

Fuente: Análisis del Investigador
Elaborado por: Investigador

De los establecimientos cerrados el factor más influyente para la falta de pago es que los establecimientos cerrados no son nuevos, sin embargo se puede ver que 772 establecimientos si han cancelado la cuota del permiso de funcionamiento, como se demuestra en el gráfico N° 49

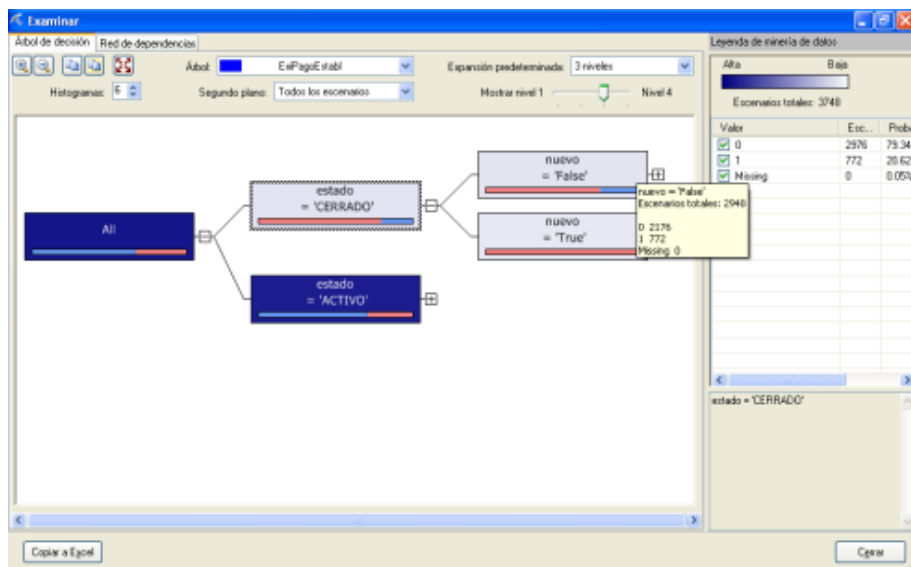


Gráfico N.49: Escenario 1.1.1

Fuente: Análisis del Investigador
Elaborado por: Investigador

Luego se determina que la mayor cantidad de establecimientos de falta de pago son empresas comerciales y de servicios, como se demuestra en el gráfico N° 50

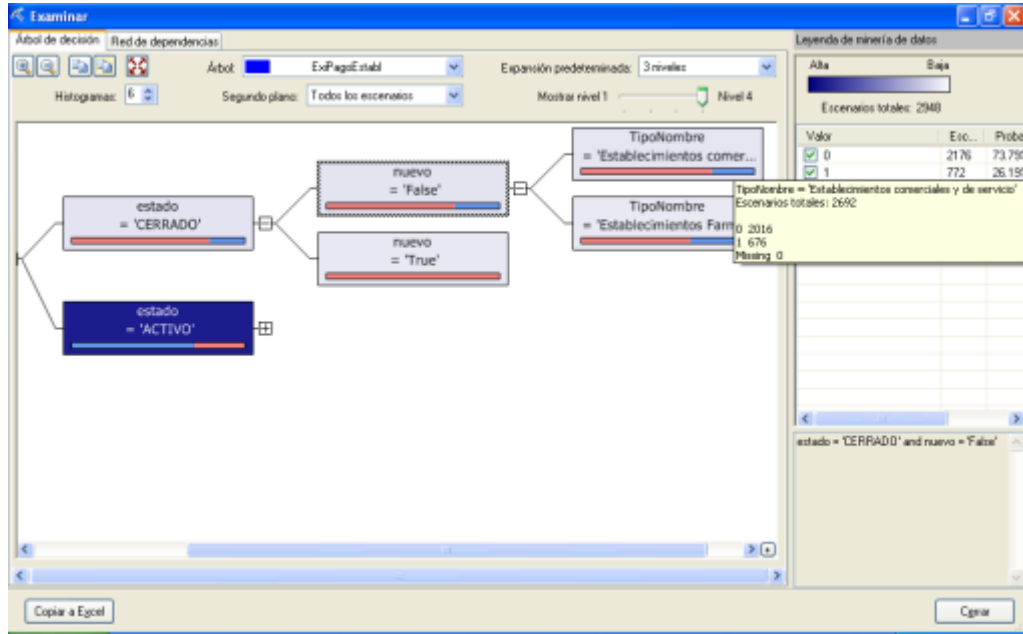


Gráfico N.50: Escenario Total

Fuente: Análisis del Investigador
Elaborado por: Investigador

Otro escenario:

Identificar el factor más influyente para la demora en el pago de los permisos de funcionamiento de los establecimientos por: Si el local es nuevo o no, si está calificado o no, si el local está activo o no, por sexo y si el local es propio o arrendado.

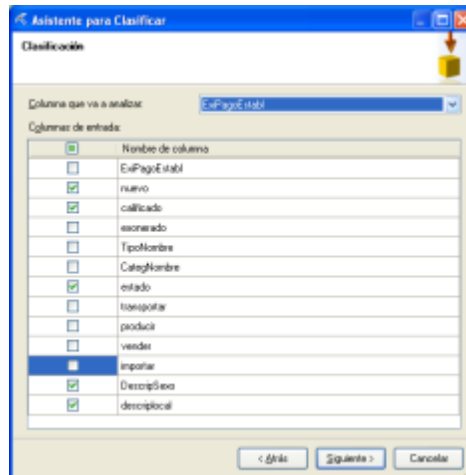


Gráfico N.51: Selección campos escenario 2

Fuente: Análisis del Investigador
Elaborado por: Investigador

Se concluye que las personas que tienen establecimientos propios y calificados y que pertenecen a dueños hombres son los que no mayormente no pagan la cuota para emisión de permisos de funcionamiento, como se demuestra en el gráfico N° 52

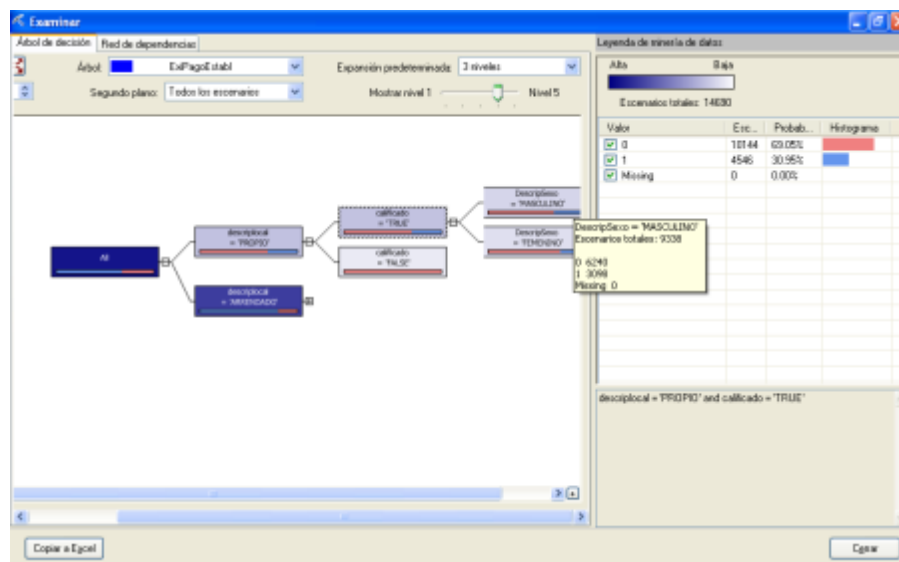


Gráfico N.52: Escenario Total 2

Fuente: Análisis del Investigador
Elaborado por: Investigador

Detección de Categorías

La herramienta Detectar categorías utiliza el algoritmo de clústeres de Microsoft para buscar automáticamente filas que contienen características similares en una tabla.

Cuando la herramienta finaliza, crea un informe que muestra las categorías que ha localizado y las características distintivas. De forma predeterminada, agrega una nueva columna a la tabla de datos con la categoría propuesta para cada fila de datos. Estas categorías se pueden revisar y se puede cambiar su nombre.

Para realizar esta tarea se deben seguir los siguientes pasos:

- 1.- Una vez obtenidos los datos del origen o fuente, en la pestaña Analizar seleccionar la opción Detectar Categorías, como se demuestra en el gráfico N° 53

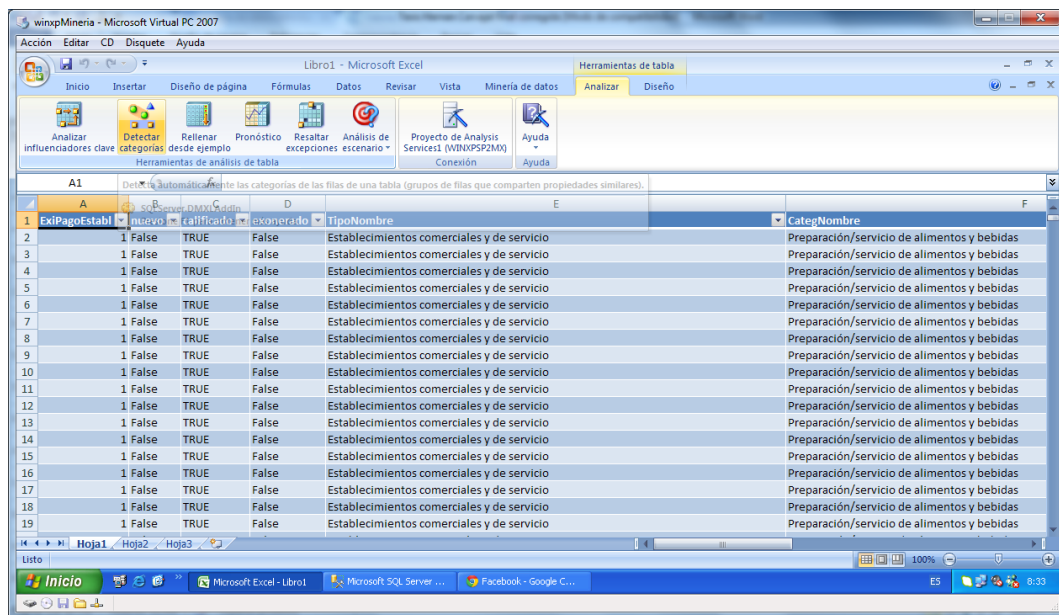


Gráfico N.53: Detectar Categorías

Fuente: Análisis del Investigador
Elaborado por: Investigador

2.- En el siguiente paso se debe seleccionar las columnas que se usarán para buscar las categorías posteriormente pulsar en Ejecutar, como se demuestra en el gráfico N° 54

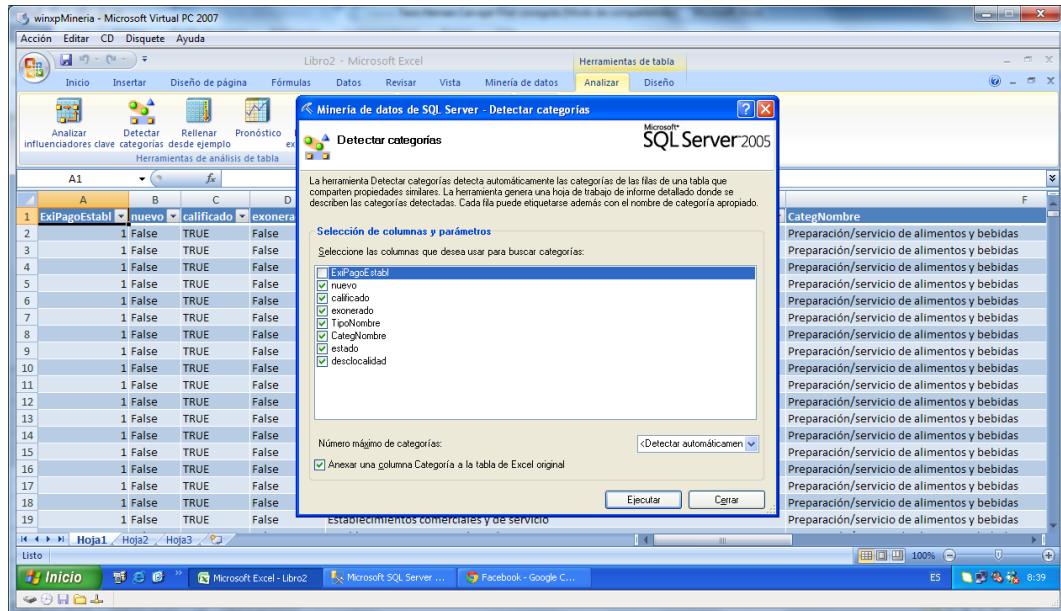


Gráfico N.54: Selección de columnas

Fuente: Análisis del Investigador
Elaborado por: Investigador

3.- Una vez que la herramienta termina de realizar el proceso para la detección de categorías, se obtiene como resultado una hoja de cálculo que contiene información referente a las categorías encontradas, en este caso de obtuvieron 3 Categorías, como se demuestra en el gráfico N° 55

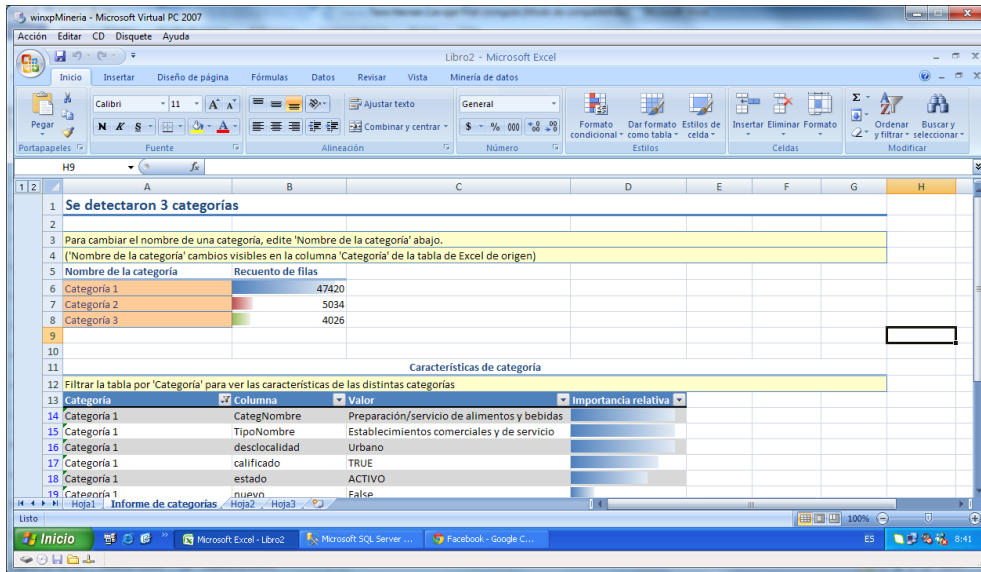


Gráfico N.55: Hoja de cálculo con información de las Categorías

Fuente: Análisis del Investigador
 Elaborado por: Investigador

4.- Se puede apreciar la tabla Características de categorías donde se muestran los factores que conforman cada categoría y su incidencia, como se demuestra en el gráfico N° 56

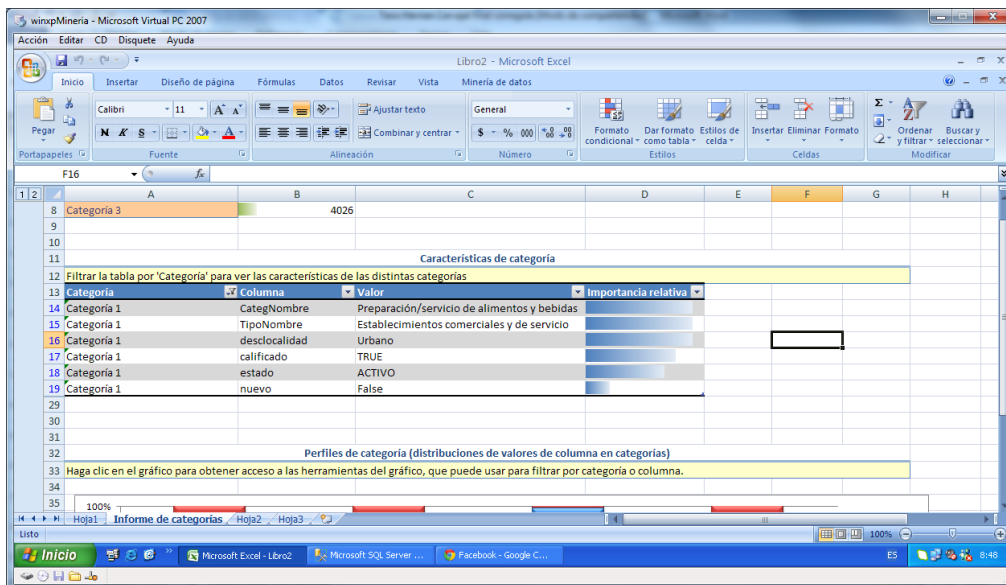


Gráfico N.56: Características de Categorías

Fuente: Análisis del Investigador
 Elaborado por: Investigador

5.- La herramienta también muestra un gráfico llamado Perfil de Categorías para el análisis en forma gráfica del comportamiento de los factores seleccionados en cada categoría, como se demuestra en el gráfico N° 57

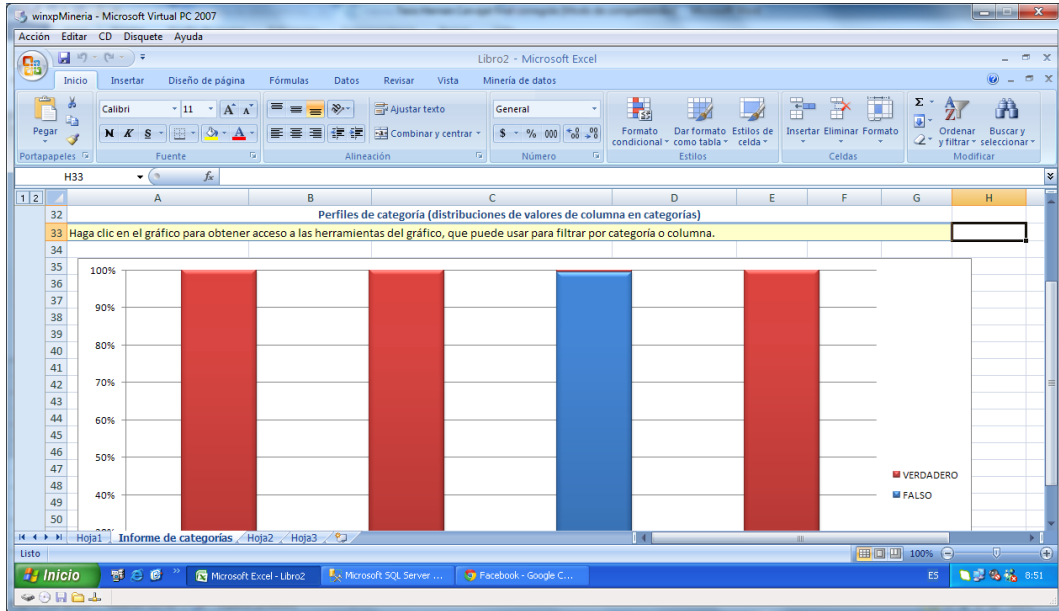


Gráfico N.57: Perfil de Categorías

Fuente: Análisis del Investigador
Elaborado por: Investigador

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

- La solución efectiva para detectar problemas en los pagos de emisión de permisos es un Data mining. El cual posee todas las características necesarias para administrar, predecir y mejorar los procesos.
- El paquete SQL server provee un ambiente amigable, de fácil uso, al mismo tiempo que cuenta con todas las herramientas necesarias para la realización de un Data mining de manera adecuada y confiable. Además la herramienta Microsoft Excel ayuda de manera fácil y sencilla a la explotación de la minería para generar los árboles de decisión y poderlos interpretar de mejor manera.

Recomendaciones

1. Para que la solución sea completamente efectiva es recomendable analizar el funcionamiento de la tecnología de minería de datos, para un mejor entendimiento de la misma y un manejo adecuado de la herramienta
2. Realizar una capacitación básica al personal que va a manejar la herramienta, para un mejor funcionamiento de la misma, para que los resultados sean confiables y exactos.
3. Tener conocimientos básicos de manejo del SQL server y Microsoft Excel, como también de manejo y utilización de base de datos.
4. Se recomienda impulsar la nueva tecnología de minería en toda la dirección de salud ya que esta ofrece información predictiva para solucionar problemas internos en el manejo de los procesos y así facilitar y agilizar los trabajos del personal brindando un mejor servicio a los clientes.

Materiales de referencia

1. BIBLIOGRAFÍA

José Hernández Orallo. (2004,.Introducción a la minería de datos. Edit. Pearson

MÉNDEZ, C. (1993). Metodología. Guía para elaborar Diseños de Investigación en Ciencias Económicas, contables, Administrativas. Edit. De Gasso, Barcelona.

Wilfredo Palma Claudio Palma.(2009.). Data Mining. El Arte De Anticipar. Edit. Ril Editores.

Silberschatz A. (2002). Fundamentos De Bases De Datos. Edit. Mcgraw-hill

Coekaerts J. (2006). Oracle Database 10g Administration Edit. Mcgraw-hill (1ª edición)

Michael Steinbach, Errata (March 25, **2006**). Introduction to Data Mining. Edit. **University** of Minnesota

Luis Aldana. (2010). DATA MINING, Principios y Aplicaciones. Edit.LafaBooks

HERNÁNDEZ, J.-LANDÁZURI, A. (.1999). Corrientes, Métodos y Técnicas de Investigación. EB-PRODEC. Quito.

HERNÁNDEZ, R. y otros. (2000). Metodología de la Investigación. McGRAW-HILL. 2ª. Edición. México D.F.

Lincografía

<http://www.microsoft.com/latam/technet/productos/servers/sql/2005/BI/default.mspx>

http://es.wikipedia.org/wiki/Miner%C3%ADa_de_datos

http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/data_mining.htm

<http://www.microsoft.com/en-us/download/details.aspx?id=8569>

<http://www.youtube.com/watch?v=Dd2KYLDxfSA>

<http://www.microsoft.com/spain/sql/technologies/dm/default.msp>

GLOSARIO DE TÉRMINOS

B

Base de datos: es un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso. En este sentido, una biblioteca puede considerarse una base de datos compuesta en su mayoría por documentos y textos impresos en papel e indexados para su consulta

Bases de datos multidimensionales: se utilizan principalmente para crear aplicaciones OLAP y pueden verse como bases de datos de una sola tabla, su peculiaridad es que por cada dimensión tienen un campo (o columna), y otro campo por cada métrica o hecho, es decir estas tablas almacenan registros cuyos campos son de la forma:

D

Data Mining: Es el proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos. Utiliza los métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos. El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos y la transformándola en una estructura comprensible para su uso posterior. Además de la etapa de análisis en bruto, que involucra aspectos de bases de datos y gestión de datos, procesamiento de datos, el modelo y las consideraciones de inferencia, métricas de Intereses, consideraciones de la Teoría de la complejidad computacional, post-procesamiento de las estructuras descubiertas, la visualización y actualización en línea.

I

Información: es un conjunto de datos con un significado, o sea, que reduce la incertidumbre o que aumenta el conocimiento de algo. En verdad, la información es un mensaje con significado en un determinado contexto, disponible para uso

inmediato y que proporciona orientación a las acciones por el hecho de reducir el margen de incertidumbre con respecto a nuestras decisiones

M

Microsoft SQL Server: es un sistema para la gestión de bases de datos producido por Microsoft basado en el modelo relacional. Sus lenguajes para consultas son T-SQL y ANSI SQL. **Microsoft SQL Server** constituye la alternativa de Microsoft a otros potentes sistemas gestores de bases de datos como son *Oracle*, *PostgreSQL* o *MySQL*.

O

OLAP: es el acrónimo en inglés de **procesamiento analítico en línea** (*On-Line Analytical Processing*). Es una solución utilizada en el campo de la llamada Inteligencia empresarial (o *Business Intelligence*), minería cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Para ello utiliza estructuras multidimensionales, que contienen datos resumidos de grandes Bases de datos o Sistemas Transaccionales **OLTP:** Se usa en informes de negocios de ventas, marketing, informes de dirección, minería de datos y áreas similares.

P

Predicción puede referirse tanto a la «acción y al efecto de predecir, como a «as palabras que manifiestan aquello que se predice; en este sentido, predecir algo es anunciar por revelación, ciencia o conjetura algo que ha de suceder

Procedimiento: es el modo de succionar determinadas relaciones que suelen realizarse de la misma forma, con una serie común de pasos claramente definidos, que permiten realizar una ocupación, trabajo, investigación, o estudio.

S

Sistema informático: es el conjunto de partes interrelacionadas, hardware, software y de recurso humano (humanware) que permite almacenar y procesar información

SQL: (por sus siglas en inglés *structured query language*) es un lenguaje declarativo de acceso a bases de datos relacionales que permite especificar diversos tipos de operaciones en estas. Una de sus características es el manejo del álgebra y el cálculo relacional permitiendo efectuar consultas con el fin de recuperar de una forma sencilla información de interés de una base de datos, así como también hacer cambios sobre ella.

ANEXOS

Anexo 1:

INSTRUMENTOS PARA LA ENCUESTA.

ENCUESTA DIRIGIDA AL PERSONAL DE LA DIRECCION DE SALUD DE COTOPAXI

OBJETIVO: Verificar la necesidad de la implementación de un sistema para el manejo de permisos de funcionamiento para los catastros.

Señores:

Estamos trabajando en un estudio que servirá para elaborar una tesis profesional acerca del grado de aceptación de desarrollo de la Gestión de Base de Datos en la Dirección de Salud de Cotopaxi. Sus respuestas serán confidenciales y anónimas.

DATOS GENERALES:

Fecha de la Encuesta.....

DATOS ESPECÍFICOS: Marque con X en el paréntesis de su elección

N.	PREGUNTAS	RESPUESTAS	COD.
1	¿Cuál es el grado de implementación de una solución de Minería de Datos en la Dirección Provincial de Salud de Cotopaxi que permita encontrar patrones faciliten el otorgamiento de permisos de funcionamiento de un establecimiento?	<ul style="list-style-type: none"> • Totalmente • Parcialmente • Nada 	1. () 2. () 3. ()
2	¿Cree usted que la implementación del Datamining optimizaría el tiempo de trabajo de los inspectores de salud en el proceso de calificación de establecimientos?	<ul style="list-style-type: none"> • Totalmente • Parcialmente • Nada 	1. () 2. () 3. ()
3	¿Considera usted útil contar con el uso de una herramienta informática que permita conocer los posibles establecimientos que incurrirán en mora?	<ul style="list-style-type: none"> • Totalmente • Parcialmente • Nada 	1. () 2. () 3. ()
4	¿Cree usted que la información obtenida puede ayudar a prevenir el pago de multas por parte de los propietarios de establecimientos?	<ul style="list-style-type: none"> • Totalmente • Parcialmente • Nada 	1. () 2. () 3. ()
5	¿Cree usted que contar con información detallada de los establecimientos de salud ayudará a las autoridades a tomar decisiones previas a la aprobación para la apertura de nuevos establecimientos?	<ul style="list-style-type: none"> • Totalmente • Parcialmente • Nada 	1. () 2. () 3. ()
6	¿Cree usted que mejoraría el desempeño en el trabajo al contar con estadísticas de control de calificaciones y emisión de permisos de funcionamiento?	<ul style="list-style-type: none"> • Totalmente • Parcialmente • Nada 	1. () 2. () 3. ()
7	¿Considera usted que la información procesada automática y ordenadamente de los establecimientos puede ayudar a realizar una mejor categorización?	<ul style="list-style-type: none"> • Totalmente • Parcialmente • Nada 	1. () 2. () 3. ()
8	¿Cree usted que los resultados que arroje la minería de datos ayuden a determinar políticas de control en la emisión de permisos de funcionamiento?	<ul style="list-style-type: none"> • Totalmente • Parcialmente • Nada 	1. () 2. () 3. ()

Gracias por su colaboración

Anexo 2

Distribución Ji cuadrado χ^2

P = Probabilidad de encontrar un valor mayor o igual que el Ji cuadrado tabulado, ν

ν/p	0,001	0,0025	0,005	0,01	0,025	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
1	10,8274	9,1404	7,8794	6,6349	5,0239	3,8415	2,7055	2,0722	1,6424	1,3233	1,0742	0,8735	0,7083	0,5707	0,4549
2	13,8150	11,9827	10,5965	9,2104	7,3778	5,9915	4,6052	3,7942	3,2189	2,7726	2,4079	2,0996	1,8326	1,5970	1,3863
3	16,2660	14,3202	12,8381	11,3449	9,3484	7,8147	6,2514	5,3170	4,6416	4,1083	3,6649	3,2831	2,9462	2,6430	2,3660
4	18,4662	16,4238	14,8602	13,2767	11,1433	9,4877	7,7794	6,7449	5,9886	5,3853	4,8784	4,4377	4,0446	3,6871	3,3567
5	20,5147	18,3854	16,7496	15,0863	12,8325	11,0705	9,2363	8,1152	7,2893	6,6257	6,0644	5,5731	5,1319	4,7278	4,3515
6	22,4575	20,2491	18,5475	16,8119	14,4494	12,5916	10,6446	9,4461	8,5581	7,8408	7,2311	6,6948	6,2108	5,7652	5,3481
7	24,3213	22,0402	20,2777	18,4753	16,0128	14,0671	12,0170	10,7479	9,8032	9,0371	8,3834	7,8061	7,2832	6,8000	6,3458
8	26,1239	23,7742	21,9549	20,0902	17,5345	15,5073	13,3616	12,0271	11,0301	10,2189	9,5245	8,9094	8,3505	7,8325	7,3441
9	27,8767	25,4625	23,5893	21,6660	19,0228	16,9190	14,6837	13,2880	12,2421	11,3887	10,6564	10,0060	9,4136	8,8632	8,3428
10	29,5879	27,1119	25,1881	23,2093	20,4832	18,3070	15,9872	14,5339	13,4420	12,5489	11,7807	11,0971	10,4732	9,8922	9,3418
11	31,2635	28,7291	26,7569	24,7250	21,9200	19,6752	17,2750	15,7671	14,6314	13,7007	12,8987	12,1836	11,5298	10,9199	10,3410
12	32,9092	30,3182	28,2997	26,2170	23,3367	21,0261	18,5493	16,9893	15,8120	14,8454	14,0111	13,2661	12,5838	11,9463	11,3403
13	34,5274	31,8830	29,8193	27,6882	24,7356	22,3620	19,8119	18,2020	16,9848	15,9839	15,1187	14,3451	13,6356	12,9717	12,3398
14	36,1239	33,4262	31,3194	29,1412	26,1189	23,6848	21,0641	19,4062	18,1508	17,1169	16,2221	15,4209	14,6853	13,9961	13,3393
15	37,6978	34,9494	32,8015	30,5780	27,4884	24,9958	22,3071	20,6030	19,3107	18,2451	17,3217	16,4940	15,7332	15,0197	14,3389
16	39,2518	36,4555	34,2671	31,9999	28,8453	26,2962	23,5418	21,7931	20,4651	19,3689	18,4179	17,5646	16,7795	16,0425	15,3385
17	40,7911	37,9462	35,7184	33,4087	30,1910	27,5871	24,7690	22,9770	21,6146	20,4887	19,5110	18,6330	17,8244	17,0646	16,3382
18	42,3119	39,4220	37,1564	34,8052	31,5264	28,8693	25,9894	24,1555	22,7595	21,6049	20,6014	19,6993	18,8679	18,0860	17,3379
19	43,8194	40,8847	38,5821	36,1908	32,8523	30,1435	27,2036	25,3289	23,9004	22,7178	21,6891	20,7638	19,9102	19,1069	18,3376
20	45,3142	42,3358	39,9969	37,5663	34,1696	31,4104	28,4120	26,4976	25,0375	23,8277	22,7745	21,8265	20,9514	20,1272	19,3374
21	46,7963	43,7749	41,4009	38,9322	35,4789	32,6706	29,6151	27,6620	26,1711	24,9348	23,8578	22,8876	21,9915	21,1470	20,3372
22	48,2676	45,2041	42,7957	40,2894	36,7807	33,9245	30,8133	28,8224	27,3015	26,0393	24,9390	23,9473	23,0307	22,1663	21,3370
23	49,7276	46,6231	44,1814	41,6383	38,0756	35,1725	32,0069	29,9792	28,4288	27,1413	26,0184	25,0055	24,0689	23,1852	22,3369
24	51,1790	48,0336	45,5584	42,9798	39,3641	36,4150	33,1962	31,1325	29,5533	28,2412	27,0960	26,0625	25,1064	24,2037	23,3367
25	52,6187	49,4351	46,9280	44,3140	40,6465	37,6525	34,3816	32,2825	30,6752	29,3388	28,1719	27,1183	26,1430	25,2218	24,3366
26	54,0511	50,8291	48,2898	45,6416	41,9231	38,8851	35,5632	33,4295	31,7946	30,4346	29,2463	28,1730	27,1789	26,2395	25,3365
27	55,4751	52,2152	49,6450	46,9628	43,1945	40,1133	36,7412	34,5736	32,9117	31,5284	30,3193	29,2266	28,2141	27,2569	26,3363
28	56,8918	53,5939	50,9936	48,2782	44,4608	41,3372	37,9159	35,7150	34,0266	32,6205	31,3909	30,2791	29,2486	28,2740	27,3362
29	58,3006	54,9662	52,3355	49,5878	45,7223	42,5569	39,0875	36,8538	35,1394	33,7109	32,4612	31,3308	30,2825	29,2908	28,3361

v/p	0,001	0,0025	0,005	0,01	0,025	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
30	59,7022	56,3325	53,6719	50,8922	46,9792	43,7730	40,2560	37,9902	36,2502	34,7997	33,5302	32,3815	31,3159	30,3073	29,3360
31	61,0980	57,6921	55,0025	52,1914	48,2319	44,9853	41,4217	39,1244	37,3591	35,8871	34,5981	33,4314	32,3486	31,3235	30,3359
32	62,4873	59,0461	56,3280	53,4857	49,4804	46,1942	42,5847	40,2563	38,4663	36,9730	35,6649	34,4804	33,3809	32,3394	31,3359
33	63,8694	60,3953	57,6483	54,7754	50,7251	47,3999	43,7452	41,3861	39,5718	38,0575	36,7307	35,5287	34,4126	33,3551	32,3358
34	65,2471	61,7382	58,9637	56,0609	51,9660	48,6024	44,9032	42,5140	40,6756	39,1408	37,7954	36,5763	35,4438	34,3706	33,3357
35	66,6192	63,0760	60,2746	57,3420	53,2033	49,8018	46,0588	43,6399	41,7780	40,2228	38,8591	37,6231	36,4746	35,3858	34,3356
36	67,9850	64,4097	61,5811	58,6192	54,4373	50,9985	47,2122	44,7641	42,8788	41,3036	39,9220	38,6693	37,5049	36,4008	35,3356
37	69,3476	65,7384	62,8832	59,8926	55,6680	52,1923	48,3634	45,8864	43,9782	42,3833	40,9839	39,7148	38,5348	37,4156	36,3355
38	70,7039	67,0628	64,1812	61,1620	56,8955	53,3835	49,5126	47,0072	45,0763	43,4619	42,0450	40,7597	39,5643	38,4302	37,3354
39	72,0550	68,3830	65,4753	62,4281	58,1201	54,5722	50,6598	48,1263	46,1730	44,5395	43,1053	41,8040	40,5935	39,4446	38,3354
40	73,4029	69,6987	66,7660	63,6908	59,3417	55,7585	51,8050	49,2438	47,2685	45,6160	44,1649	42,8477	41,6222	40,4589	39,3353
45	80,0776	76,2229	73,1660	69,9569	65,4101	61,6562	57,5053	54,8105	52,7288	50,9849	49,4517	48,0584	46,7607	45,5274	44,3351
50	86,6603	82,6637	79,4898	76,1538	71,4202	67,5048	63,1671	60,3460	58,1638	56,3336	54,7228	53,2576	51,8916	50,5923	49,3349
55	93,1671	89,0344	85,7491	82,2920	77,3804	73,3115	68,7962	65,8550	63,5772	61,6650	59,9804	58,4469	57,0160	55,6539	54,3348
60	99,6078	95,3443	91,9518	88,3794	83,2977	79,0820	74,3970	71,3411	68,9721	66,9815	65,2265	63,6277	62,1348	60,7128	59,3347
70	112,3167	107,8079	104,2148	100,4251	95,0231	90,5313	85,5270	82,2553	79,7147	77,5766	75,6893	73,9677	72,3583	70,8236	69,3345
80	124,8389	120,1018	116,3209	112,3288	106,6285	101,8795	96,5782	93,1058	90,4053	88,1303	86,1197	84,2840	82,5663	80,9266	79,3343
90	137,2082	132,2554	128,2987	124,1162	118,1359	113,1452	107,5650	103,9040	101,0537	98,6499	96,5238	94,5909	92,7614	91,0234	89,3342
100	149,4488	144,2925	140,1697	135,8069	129,5613	124,3421	118,4980	114,6588	111,6667	109,1412	106,9058	104,8615	102,9459	101,1149	99,3341
120	173,6184	168,0814	163,6485	158,9500	152,2113	146,5673	140,2326	136,0620	132,8063	130,0546	127,6159	125,3833	123,2890	121,2850	119,3340
140	197,4498	191,5653	186,8465	181,8405	174,6478	168,6130	161,8270	157,3517	153,8537	150,8941	148,2686	145,8629	143,6043	141,4413	139,3339
160	221,0197	214,8081	209,8238	204,5300	196,9152	190,5164	183,3106	178,5517	174,8283	171,6752	168,8759	166,3092	163,8977	161,5968	159,3338
180	244,3723	237,8548	232,6198	227,0563	219,0442	212,3039	204,7036	199,6786	195,7434	192,4086	189,4462	186,7282	184,1732	181,7234	179,3338
200	267,5388	260,7350	255,2638	249,4452	241,0578	233,9942	226,0210	220,7441	216,6088	213,1022	209,9854	207,1244	204,4337	201,8526	199,3337
250	324,8306	317,3609	311,3460	304,9393	295,6885	287,8815	279,0504	273,1944	268,5987	264,6970	261,2253	258,0355	255,0327	252,1497	249,3337
300	381,4239	373,3509	366,8439	359,9064	349,8745	341,3951	331,7885	325,4090	320,3971	316,1383	312,3460	308,8589	305,5741	302,4182	299,3336
500	603,4458	593,3580	585,2060	576,4931	563,8514	553,1269	540,9303	532,8028	526,4014	520,9505	516,0874	511,6081	507,3816	503,3147	499,3335
600	712,7726	701,8322	692,9809	683,5155	669,7690	658,0936	644,8004	635,9329	628,8157	622,9876	617,6713	612,7718	608,1468	603,6942	599,3335

v/p	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	0,975	0,99	0,995	0,9975	0,999
1	0,3573	0,2750	0,2059	0,1485	0,1015	0,0642	0,0358	0,0158	0,0039	0,0010	0,0002	0,0000	0,0000	0,0000
2	1,1957	1,0217	0,8616	0,7133	0,5754	0,4463	0,3250	0,2107	0,1026	0,0506	0,0201	0,0100	0,0050	0,0020
3	2,1095	1,8692	1,6416	1,4237	1,2125	1,0052	0,7978	0,5844	0,3518	0,2158	0,1148	0,0717	0,0449	0,0243
4	3,0469	2,7528	2,4701	2,1947	1,9226	1,6488	1,3665	1,0636	0,7107	0,4844	0,2971	0,2070	0,1449	0,0908
5	3,9959	3,6555	3,3251	2,9999	2,6746	2,3425	1,9938	1,6103	1,1455	0,8312	0,5543	0,4118	0,3075	0,2102
6	4,9519	4,5702	4,1973	3,8276	3,4546	3,0701	2,6613	2,2041	1,6354	1,2373	0,8721	0,6757	0,5266	0,3810
7	5,9125	5,4932	5,0816	4,6713	4,2549	3,8223	3,3583	2,8331	2,1673	1,6899	1,2390	0,9893	0,7945	0,5985
8	6,8766	6,4226	5,9753	5,5274	5,0706	4,5936	4,0782	3,4895	2,7326	2,1797	1,6465	1,3444	1,1042	0,8571
9	7,8434	7,3570	6,8763	6,3933	5,8988	5,3801	4,8165	4,1682	3,3251	2,7004	2,0879	1,7349	1,4501	1,1519
10	8,8124	8,2955	7,7832	7,2672	6,7372	6,1791	5,5701	4,8652	3,9403	3,2470	2,5582	2,1558	1,8274	1,4787
11	9,7831	9,2373	8,6952	8,1479	7,5841	6,9887	6,3364	5,5778	4,5748	3,8157	3,0535	2,6032	2,2321	1,8338
12	10,7553	10,1820	9,6115	9,0343	8,4384	7,8073	7,1138	6,3038	5,2260	4,4038	3,5706	3,0738	2,6612	2,2141
13	11,7288	11,1291	10,5315	9,9257	9,2991	8,6339	7,9008	7,0415	5,8919	5,0087	4,1069	3,5650	3,1118	2,6172
14	12,7034	12,0785	11,4548	10,8215	10,1653	9,4673	8,6963	7,7895	6,5706	5,6287	4,6604	4,0747	3,5820	3,0407
15	13,6790	13,0298	12,3809	11,7212	11,0365	10,3070	9,4993	8,5468	7,2609	6,2621	5,2294	4,6009	4,0697	3,4825
16	14,6555	13,9827	13,3096	12,6243	11,9122	11,1521	10,3090	9,3122	7,9616	6,9077	5,8122	5,1422	4,5734	3,9417
17	15,6328	14,9373	14,2406	13,5307	12,7919	12,0023	11,1249	10,0852	8,6718	7,5642	6,4077	5,6973	5,0916	4,4162
18	16,6108	15,8932	15,1738	14,4399	13,6753	12,8570	11,9462	10,8649	9,3904	8,2307	7,0149	6,2648	5,6234	4,9048
19	17,5894	16,8504	16,1089	15,3517	14,5620	13,7158	12,7727	11,6509	10,1170	8,9065	7,6327	6,8439	6,1673	5,4067
20	18,5687	17,8088	17,0458	16,2659	15,4518	14,5784	13,6039	12,4426	10,8508	9,5908	8,2604	7,4338	6,7228	5,9210
21	19,5485	18,7683	17,9843	17,1823	16,3444	15,4446	14,4393	13,2396	11,5913	10,2829	8,8972	8,0336	7,2889	6,4467
22	20,5288	19,7288	18,9243	18,1007	17,2396	16,3140	15,2787	14,0415	12,3380	10,9823	9,5425	8,6427	7,8648	6,9829
23	21,5095	20,6902	19,8657	19,0211	18,1373	17,1865	16,1219	14,8480	13,0905	11,6885	10,1957	9,2604	8,4503	7,5291
24	22,4908	21,6525	20,8084	19,9432	19,0373	18,0618	16,9686	15,6587	13,8484	12,4011	10,8563	9,8862	9,0441	8,0847
25	23,4724	22,6156	21,7524	20,8670	19,9393	18,9397	17,8184	16,4734	14,6114	13,1197	11,5240	10,5196	9,6462	8,6494
26	24,4544	23,5794	22,6975	21,7924	20,8434	19,8202	18,6714	17,2919	15,3792	13,8439	12,1982	11,1602	10,2561	9,2222
27	25,4367	24,5440	23,6437	22,7192	21,7494	20,7030	19,5272	18,1139	16,1514	14,5734	12,8785	11,8077	10,8733	9,8029
28	26,4195	25,5092	24,5909	23,6475	22,6572	21,5880	20,3857	18,9392	16,9279	15,3079	13,5647	12,4613	11,4973	10,3907
29	27,4025	26,4751	25,5391	24,5770	23,5666	22,4751	21,2468	19,7677	17,7084	16,0471	14,2564	13,1211	12,1278	10,9861

v/p	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	0,975	0,99	0,995	0,9975	0,999
30	28,3858	27,4416	26,4881	25,5078	24,4776	23,3641	22,1103	20,5992	18,4927	16,7908	14,9535	13,7867	12,7646	11,5876
31	29,3694	28,4087	27,4381	26,4397	25,3901	24,2551	22,9762	21,4336	19,2806	17,5387	15,6555	14,4577	13,4073	12,1961
32	30,3533	29,3763	28,3889	27,3728	26,3041	25,1478	23,8442	22,2706	20,0719	18,2908	16,3622	15,1340	14,0555	12,8104
33	31,3375	30,3444	29,3405	28,3069	27,2194	26,0422	24,7143	23,1102	20,8665	19,0467	17,0735	15,8152	14,7092	13,4312
34	32,3219	31,3130	30,2928	29,2421	28,1361	26,9383	25,5864	23,9522	21,6643	19,8062	17,7891	16,5013	15,3679	14,0568
35	33,3065	32,2821	31,2458	30,1782	29,0540	27,8359	26,4604	24,7966	22,4650	20,5694	18,5089	17,1917	16,0315	14,6881
36	34,2913	33,2517	32,1995	31,1152	29,9730	28,7350	27,3363	25,6433	23,2686	21,3359	19,2326	17,8868	16,7000	15,3243
37	35,2764	34,2216	33,1539	32,0532	30,8933	29,6355	28,2138	26,4921	24,0749	22,1056	19,9603	18,5859	17,3730	15,9652
38	36,2617	35,1920	34,1089	32,9919	31,8146	30,5373	29,0931	27,3430	24,8839	22,8785	20,6914	19,2888	18,0501	16,6109
39	37,2472	36,1628	35,0645	33,9315	32,7369	31,4405	29,9739	28,1958	25,6954	23,6543	21,4261	19,9958	18,7318	17,2612
40	38,2328	37,1340	36,0207	34,8719	33,6603	32,3449	30,8563	29,0505	26,5093	24,4331	22,1642	20,7066	19,4171	17,9166
45	43,1638	41,9950	40,8095	39,5847	38,2910	36,8844	35,2895	33,3504	30,6123	28,3662	25,9012	24,3110	22,8994	21,2509
50	48,0986	46,8638	45,6100	44,3133	42,9421	41,4492	39,7539	37,6886	34,7642	32,3574	29,7067	27,9908	26,4636	24,6736
55	53,0367	51,7391	50,4204	49,0554	47,6105	46,0356	44,2448	42,0596	38,9581	36,3981	33,5705	31,7349	30,0974	28,1731
60	57,9775	56,6200	55,2394	53,8091	52,2938	50,6406	48,7587	46,4589	43,1880	40,4817	37,4848	35,5344	33,7909	31,7381
70	67,8664	66,3961	64,8990	63,3460	61,6983	59,8978	57,8443	55,3289	51,7393	48,7575	45,4417	43,2753	41,3323	39,0358
80	77,7631	76,1879	74,5825	72,9153	71,1445	69,2070	66,9938	64,2778	60,3915	57,1532	53,5400	51,1719	49,0430	46,5197
90	87,6661	85,9925	84,2854	82,5111	80,6247	78,5584	76,1954	73,2911	69,1260	65,6466	61,7540	59,1963	56,8918	54,1559
100	97,5744	95,8078	94,0046	92,1290	90,1332	87,9453	85,4406	82,3581	77,9294	74,2219	70,0650	67,3275	64,8571	61,9182
120	117,4041	115,4646	113,4825	111,4186	109,2197	106,8056	104,0374	100,6236	95,7046	91,5726	86,9233	83,8517	81,0726	77,7555
140	137,2476	135,1491	133,0028	130,7657	128,3800	125,7580	122,7476	119,0293	113,6594	109,1368	104,0343	100,6547	97,5908	93,9253
160	157,1019	154,8555	152,5564	150,1583	147,5988	144,7834	141,5475	137,5457	131,7560	126,8700	121,3457	117,6791	114,3496	110,3592
180	176,9652	174,5799	172,1373	169,5879	166,8653	163,8682	160,4206	156,1526	149,9687	144,7413	138,8205	134,8843	131,3050	127,0114
200	196,8359	194,3193	191,7409	189,0486	186,1717	183,0028	179,3550	174,8353	168,2785	162,7280	156,4321	152,2408	148,4262	143,8420
250	246,5387	243,7202	240,8297	237,8085	234,5768	231,0128	226,9048	221,8059	214,3915	208,0978	200,9387	196,1604	191,8020	186,5537
300	296,2700	293,1786	290,0062	286,6878	283,1353	279,2143	274,6901	269,0679	260,8781	253,9122	245,9727	240,6631	235,8126	229,9620
500	495,3734	491,3709	487,2569	482,9462	478,3231	473,2099	467,2962	459,9261	449,1467	439,9360	429,3874	422,3034	415,8081	407,9458
600	594,9938	590,6057	586,0930	581,3623	576,2859	570,6681	564,1661	556,0560	544,1801	534,0185	522,3654	514,5285	507,3385	498,6219

Anexo 3

Manual Operativo

1.- En el equipo en el cual se correrá el aplicativo se deberá instalar el complemento SQLServer2005_DMAddinMineria.msi, que es una herramienta para la explotación del datamining para Microsoft Excel.

2.- Una vez instalado el complemento se procede a abrir Excel y se comprobará que en la barra de herramientas se ha añadido la pestaña Minería de Datos , como se demuestra en el gráfico N° 53:



Gráfico N.53: Complemento minería de Datos

Fuente: Análisis del Investigador
Elaborado por: Investigador

3.- Ubicarse en la pestaña Datos y escoger la opción De otras fuentes y seleccionar Desde SQL Server, como se demuestra en el gráfico N° 54:

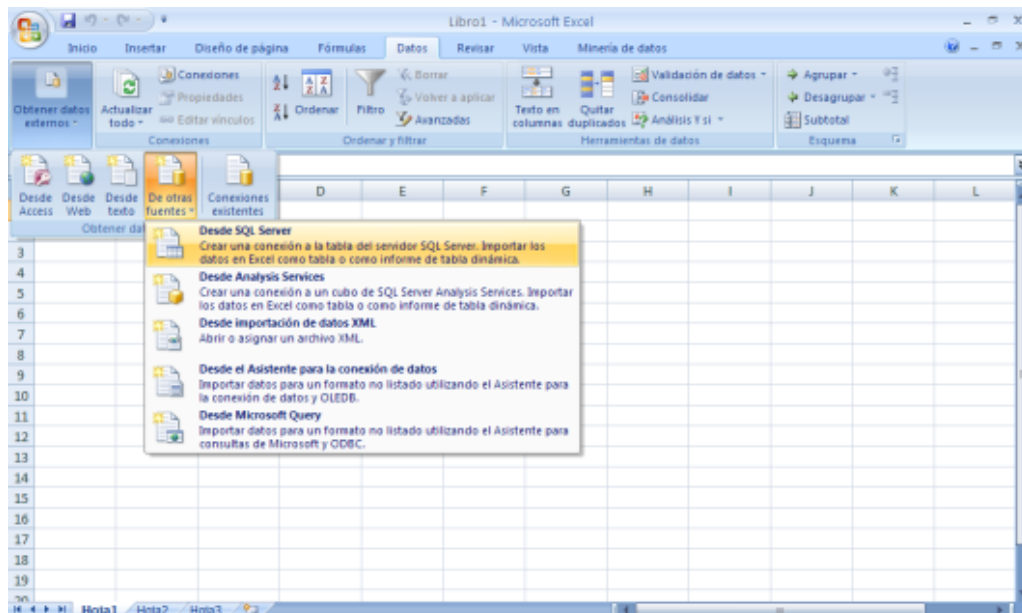


Gráfico N.54: Desde SQL Server

Fuente: Análisis del Investigador
Elaborado por: Investigador

4.- Aparecerá el asistente de minería para la conexión de datos, verificar que este seleccionado Microsoft SQL server y clic en siguiente, como se demuestra en el gráfico N° 55

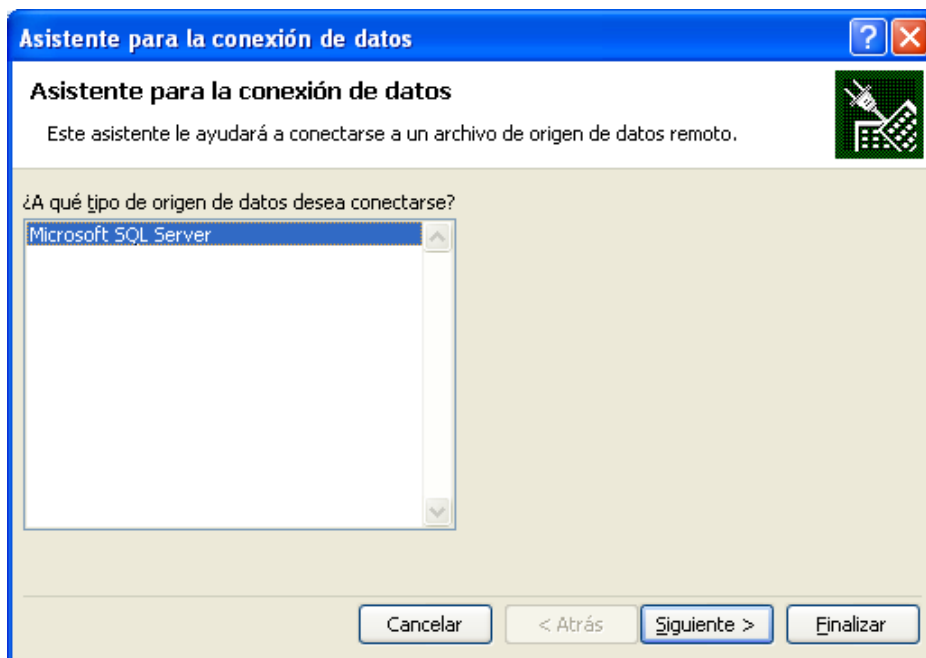


Gráfico N.55: Asistente conexión de datos

Fuente: Análisis del Investigador
Elaborado por: Investigador

5.- En nombre del servidor digitar el nombre del servidor de SQL server y la instancia en la cual se creó la base y se cargaron los datos con los proceso ETL, clic en siguiente, como se demuestra en el gráfico N° 56

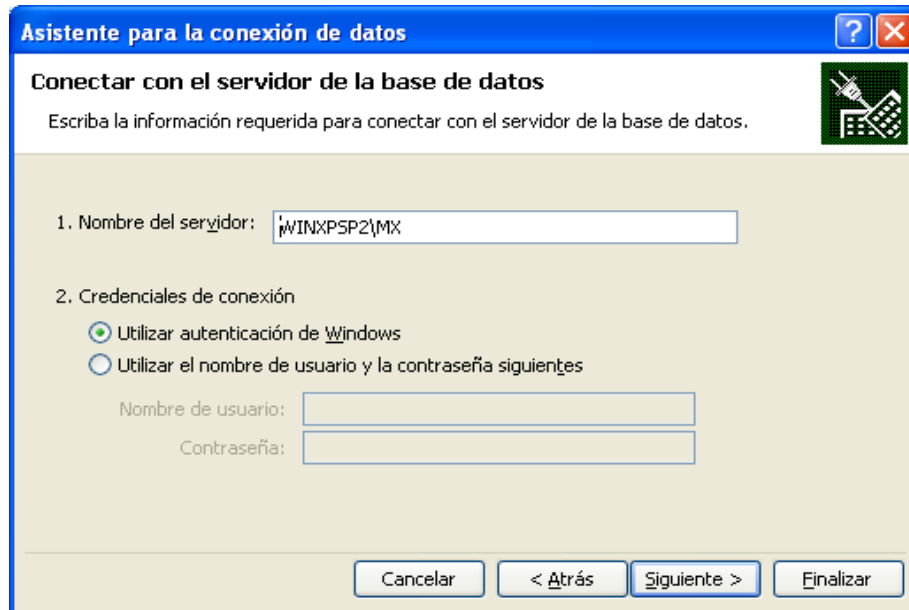


Gráfico N.56: Conexión instancia del servidor

Fuente: Análisis del Investigador
 Elaborado por: Investigador

6.- Seleccionar la base de datos y la tabla o vista sobre la cual se va a trabajar, clic en siguiente, como se demuestra en el gráfico N° 57

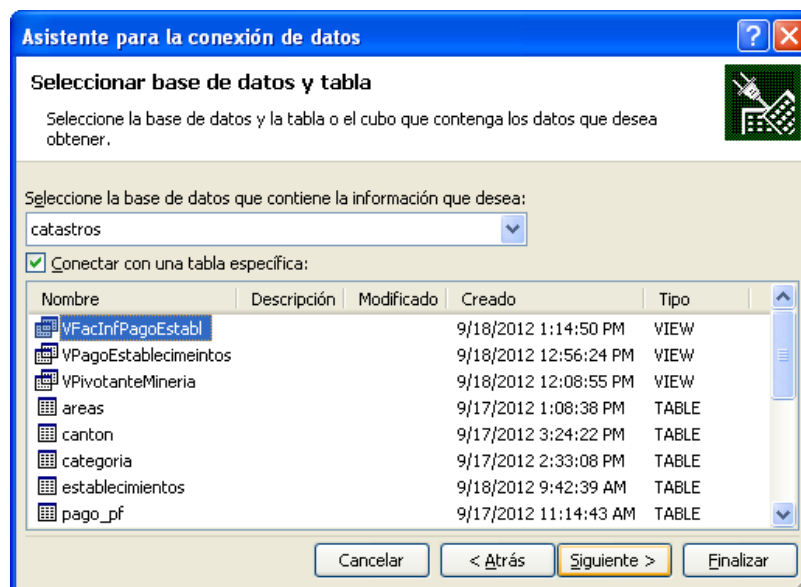


Gráfico N.57: Selección de la vista o tabla

Fuente: Análisis del Investigador
 Elaborado por: Investigador

7.- Se dará un nombre al reporte, clic en finalizar, como se demuestra en el gráfico N° 58

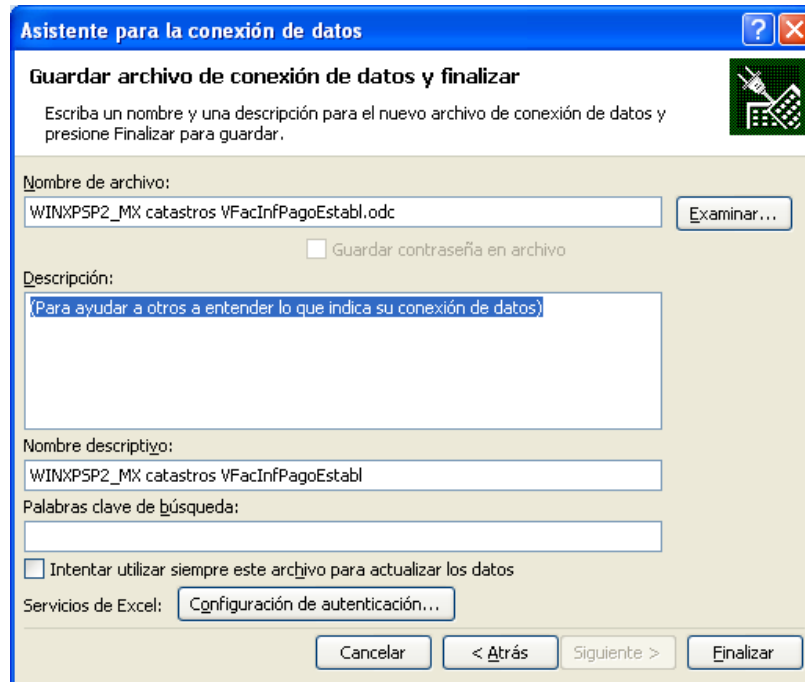


Gráfico N.58: Guardar archivos de conexión de datos

Fuente: Análisis del Investigador
Elaborado por: Investigador

8.- El siguiente paso es importar los datos a Excel para lo cual se escoge tabla y clic en aceptar, como se demuestra en el gráfico N° 59

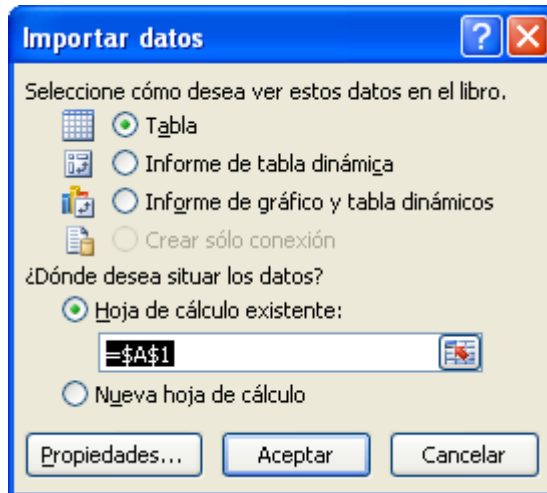


Gráfico N.59: Importar datos a Excel

Fuente: Análisis del Investigador
 Elaborado por: Investigador

9.- Si todo esta correcto los datos se importarán a Excel de la forma, como se demuestra en el gráfico N° 60:

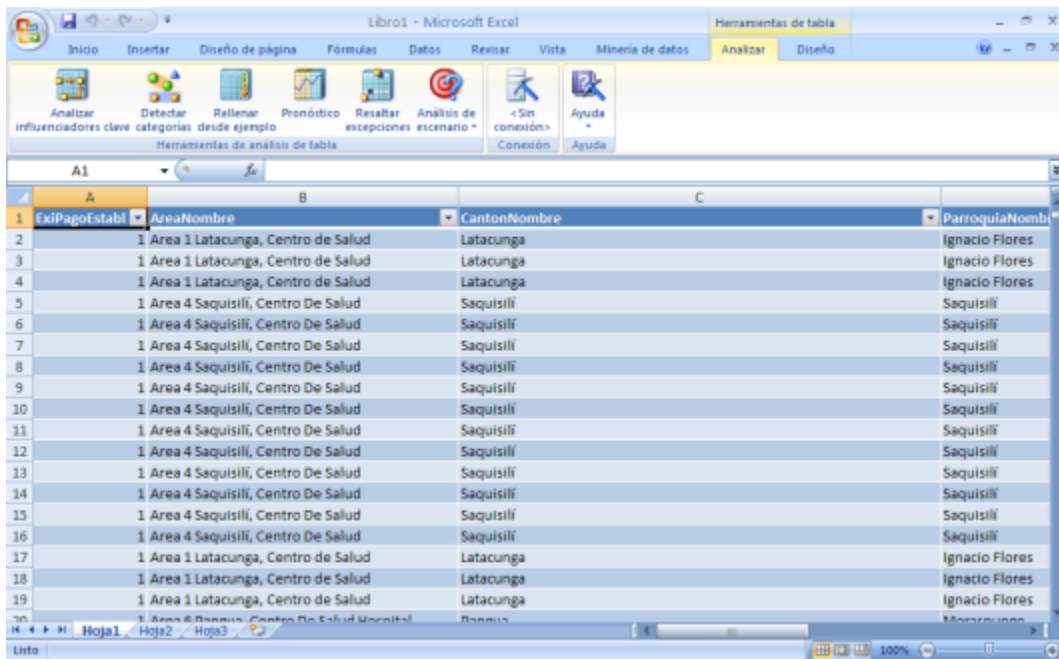


Gráfico N.60: Datos en Excel

Fuente: Análisis del Investigador
 Elaborado por: Investigador

10.- Para realizar la minería en el menú Dataming escoger clasificar, como se demuestra en el gráfico N° 61

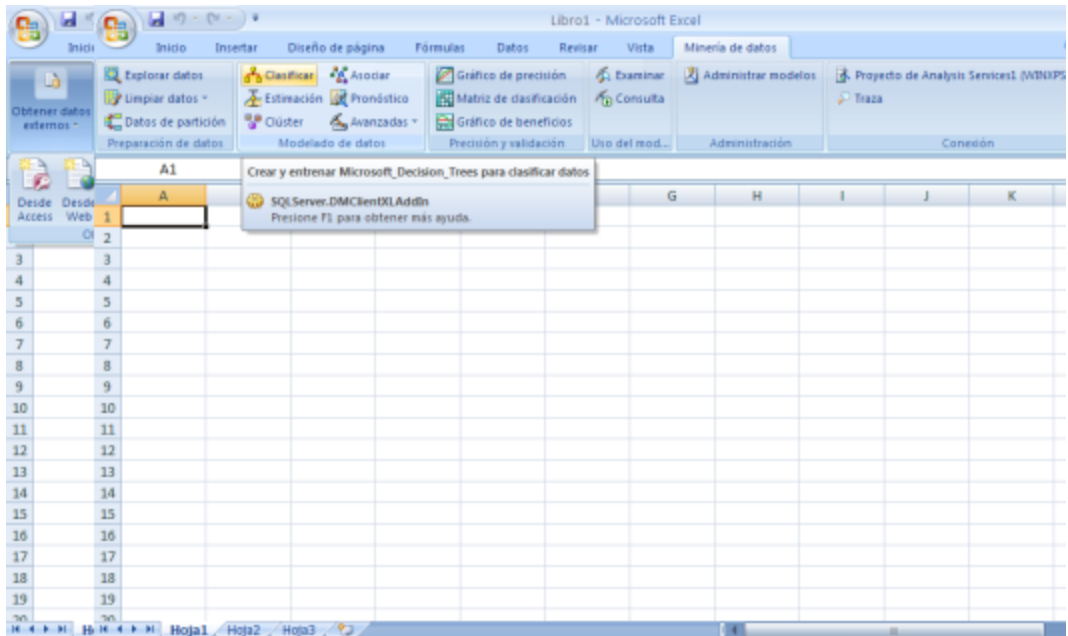


Gráfico N.61: Clasificar

Fuente: Análisis del Investigador
Elaborado por: Investigador

11.- Aparecerá el asistente para clasificar, clic en siguiente, como se demuestra en el gráfico N° 62



Gráfico N.62: Asiste de clasificación de datos

Fuente: Análisis del Investigador
Elaborado por: Investigador

12.- Seleccionar los datos del origen, estos se pondrán por defecto, clic en siguiente, como se demuestra en el gráfico N° 63

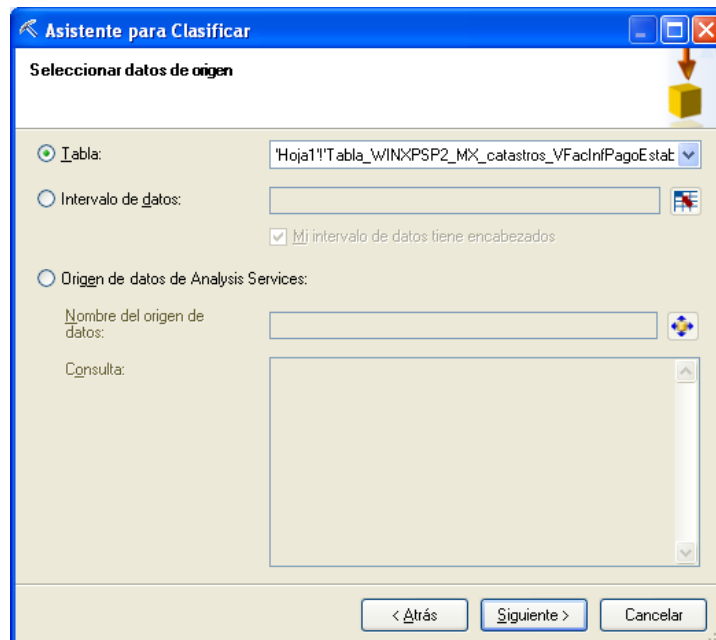


Gráfico N.63: Selección datos de origen

Fuente: Análisis del Investigador
Elaborado por: Investigador

13.- Escoger los campos por los que se clasificará el data mining , clic en siguiente, como se demuestra en el gráfico N° 64

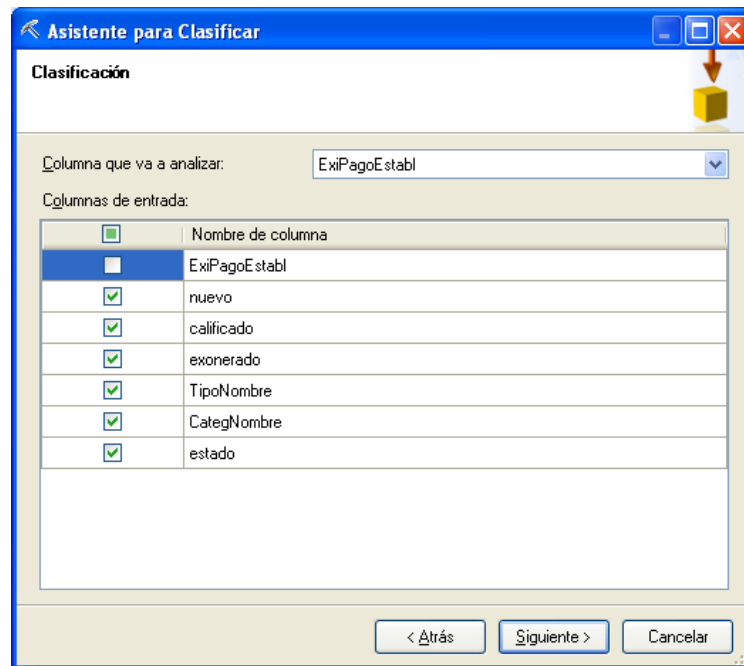


Gráfico N.64: Selección de columnas de predicción

Fuente: Análisis del Investigador
Elaborado por: Investigador

14.- Se observará un resumen del árbol de decisión a crear, clic en finalizar, como se demuestra en el gráfico N° 65

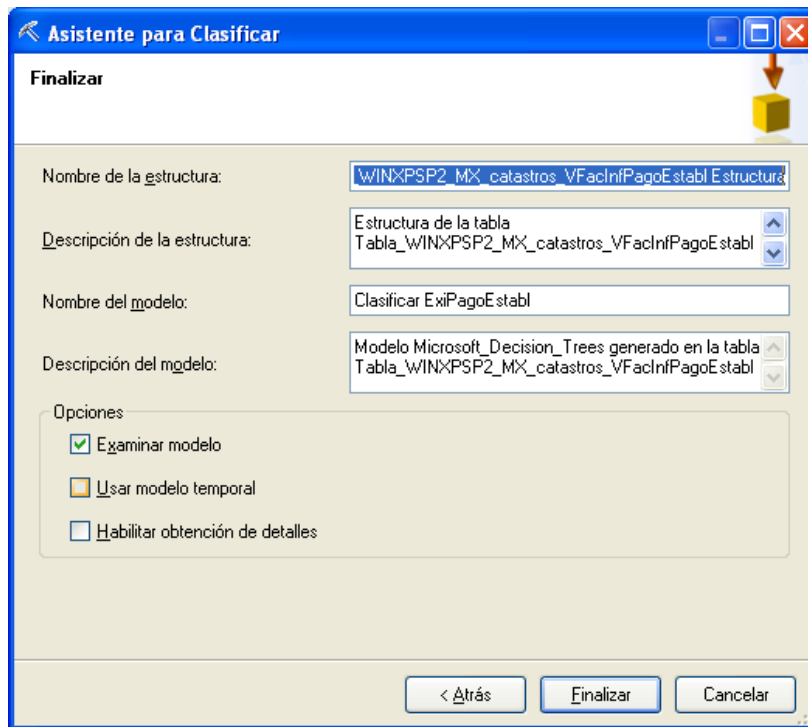


Gráfico N.65: Finalizar asistente de clasificación

Fuente: Análisis del Investigador
 Elaborado por: Investigador

15.- Si todo esta correcto saldrá una pantalla como ésta en donde se está generando el árbol de decisión, como se demuestra en el gráfico N° 66

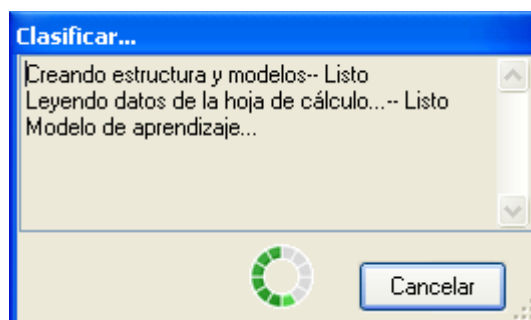


Gráfico N.66: procesando información de predicción

Fuente: Análisis del Investigador
 Elaborado por: Investigador

Tendremos ya los resultados que la minería arroja en la cual podemos ver algunos de los escenarios resultantes, como ejemplo tenemos:

Escenario = Identificar la demora en el pago de permiso de funcionamiento.

Total de Escenarios = 56480, de los cuales 37824 han pagado la cuota y 18656 no lo han hecho, como se demuestra en el gráfico N° 67

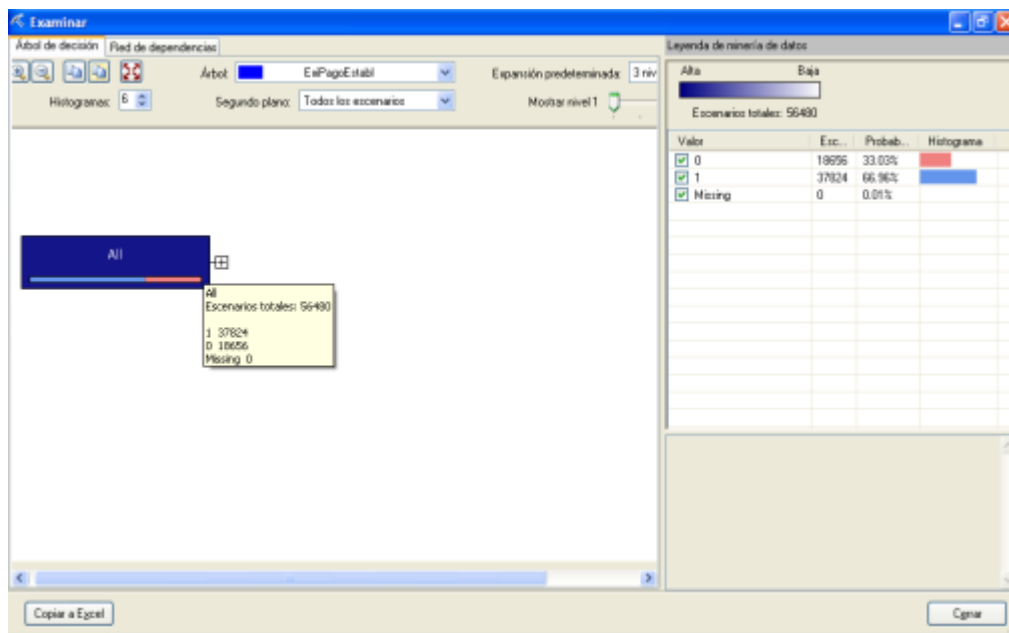


Gráfico N.67: Escenario 1

Fuente: Análisis del Investigador
Elaborado por: Investigador

Explorando más adentro identificamos que los que menos pagan, son los establecimientos cerrados, como se demuestra en el gráfico N° 68

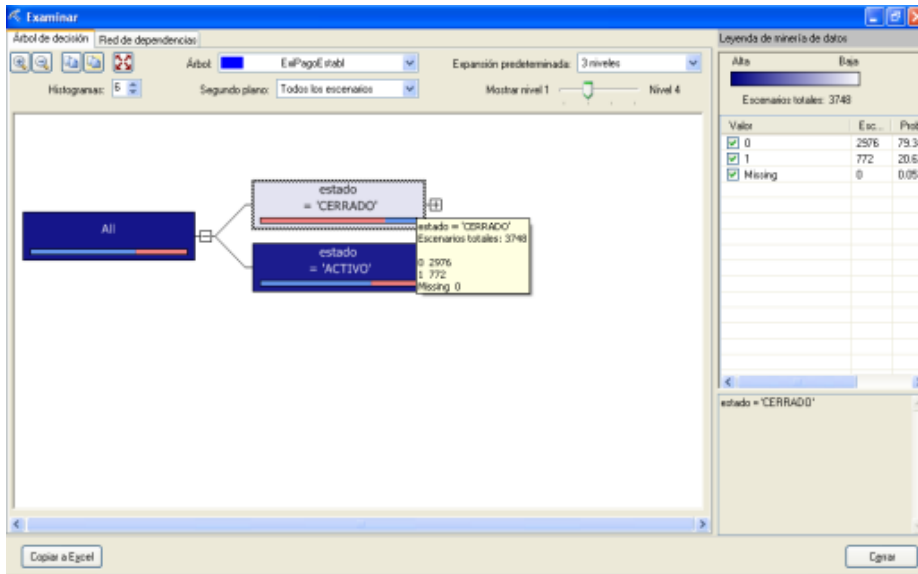


Gráfico N.68: Escenario 1.1

Fuente: Análisis del Investigador
Elaborado por: Investigador

De los establecimientos cerrados el factor más influyente para la falta de pago es que los establecimientos cerrados no son nuevos, sin embargo podemos ver que 772 establecimientos si han cancelado la cuota del permiso de funcionamiento, como se demuestra en el gráfico N° 69

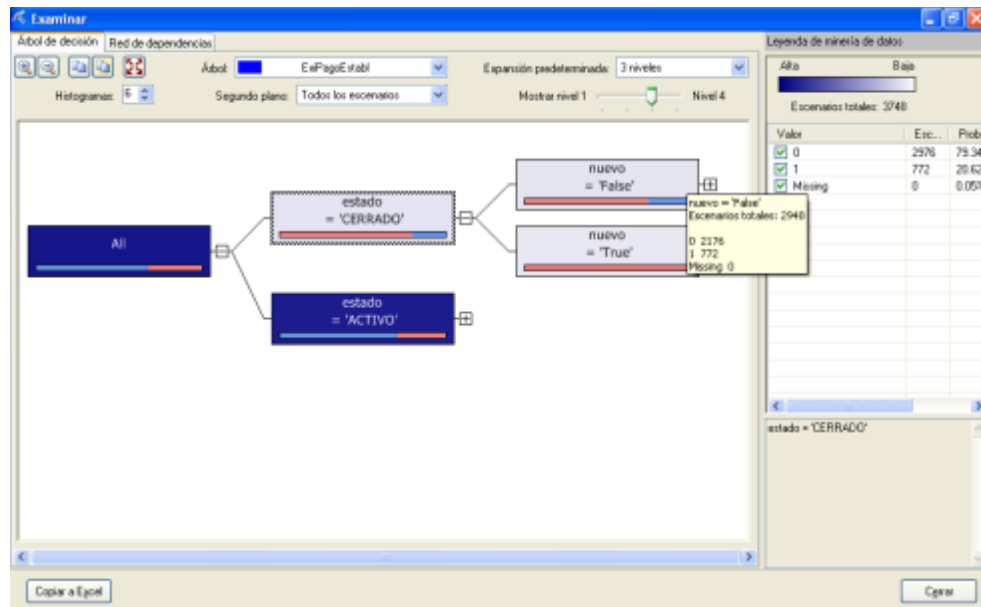


Gráfico N.69: Escenario 1.1.1

Fuente: Análisis del Investigador
Elaborado por: Investigador

Luego se determina que la mayor cantidad de establecimientos de falta de pago son empresas comerciales y de servicios, como se demuestra en el gráfico N° 70

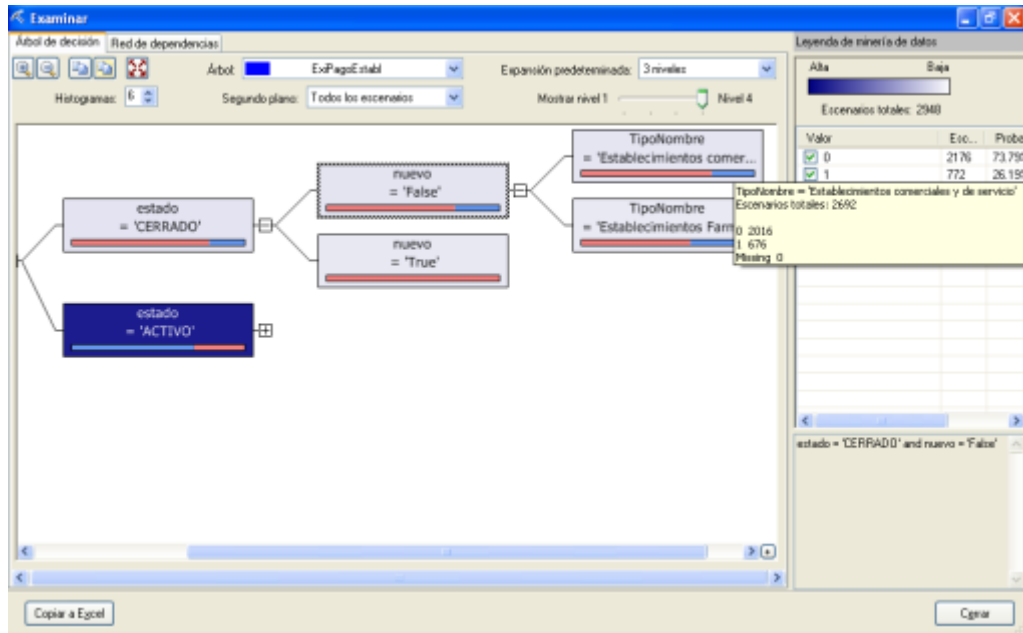


Gráfico N.70: Escenario Total

Fuente: Análisis del Investigador
 Elaborado por: Investigador