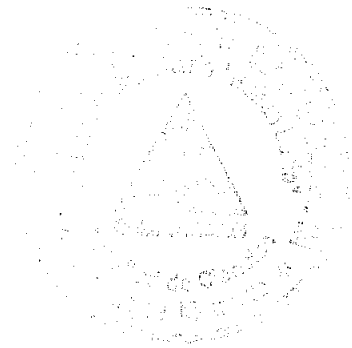


Universidad Nacional Autónoma de Nicaragua
Recinto Universitario Rubén Darío
Facultad de Ciencias e Ingenierías
Departamento de Computación
UNAN Managua

Seminario de Graduación.



Tema: Minería de datos.

Subtema:

Aplicación del algoritmo NaïveBayes de minería de datos en el área de mercadeo de la empresa MULTICOMERCIAL S.A. CECA en Managua Nicaragua en el II Semestre del 2015.

Br. Tito Alejandro Fariñas Osorio

Tutor: MSc. Luis Miguel Martínez Olivera

Managua, Nicaragua 2015

Instituto Central Salomón de la Selva

UNAN-Managua

Fecha de Impreso: 24/12/16

Compu: Des. XI. de Computación

Unidad C.

36984

Unidad No.

SM
COMP
378.242
Far
2015

AGRADECIMIENTOS

Le agradezco a Dios por haberme acompañado y guiado a lo largo de mi carrera, por ser mi fortaleza en los momentos de debilidad y por brindarme una vida llena de aprendizajes, experiencias y sobre todo felicidad.

Le doy gracias a mis padres Tito y Marta por apoyarme en todo momento, por los valores que me han inculcado, y por haberme dado la oportunidad de tener una excelente educación en el transcurso de mi vida. Sobre todo por ser un excelente ejemplo de vida a seguir.

A mi hermano por ser parte importante de mi vida y representar la unidad familiar.

A mi esposa por llenar mi vida de alegrías y amor cuando más lo he necesitado.

Agradezco también a mi tutor de proyecto de investigación, Luis Miguel Martínez Olivera, por su apoyo incondicional, entusiasmo, guía, paciencia y perseverancia, durante este año de investigación. Le agradezco por ser un gran educador y principalmente, una excelente persona. Así mismo quiero expresar mi reconocimiento a todos mis profesores por sus enseñanzas y experiencias, apoyo y dedicación en estos años de estudio.

Tito Fariñas

DEDICATORIA

Al tener la oportunidad de presentar este humilde trabajo,
me gustaría dedicarlo a:

Mis padres...

A mi hermano...

A mi amada esposa...

A nuestro gran personal docente...

Y todos mis amigos...

Tito Fariñas

INDICE DE CONTENIDOS

AGRADECIMIENTOS	2
DEDICATORIA	3
INDICE DE CONTENIDOS	4
ÍNDICE DE FIGURAS	7
ÍNDICE DE TABLAS	9
I. RESUMEN	10
II. INTRODUCCIÓN	11
III. ANTECEDENTES	12
IV. PLANTEAMIENTO DEL PROBLEMA.	14
4.1 CARACTERIZACIÓN DEL PROBLEMA.	14
4.2 DELIMITACIÓN DEL PROBLEMA.	14
4.3 FORMULACIÓN EL PROBLEMA.	14
4.4 SISTEMATIZACIÓN DEL PROBLEMA.	15
V. JUSTIFICACIÓN	16
VI. OBJETIVOS	17
6.1 OBJETIVO GENERAL	17
6.2 OBJETIVOS ESPECÍFICOS	17
VII. MARCO TEÓRICO	18
7.1 DATOS	18
7.2 INFORMACIÓN	19
7.3 CONOCIMIENTO	20
7.4 SISTEMAS DE INFORMACIÓN	21
7.4.1 Clasificación de los sistemas de información	22
7.5 BASE DE DATOS	23
7.6 SISTEMA DE ADMINISTRACIÓN DE BASES DE DATOS	24
7.7 BASE DE DATOS RELACIONAL	24
7.8 INTELIGENCIA DE NEGOCIOS	25
7.8.1 Business Intelligence	25

7.8.2	¿Qué es Business Intelligence?	25
7.8.3	Componentes de Business Intelligence	26
7.9	MINERÍA DE DATOS	27
7.9.1	Propósitos de minería de datos.	29
7.9.2	Aplicaciones de la minería de datos	29
7.9.3	Características y Objetivos de la Minería de Datos	30
7.9.4	Técnicas de Minería de Datos	32
7.9.5	Fases de La Minería de Datos.	32
7.9.6	¿Qué es un Algoritmo?	34
7.9.7	Categorías de los Algoritmos en minería de datos	35
7.9.8	Algoritmos de minería de datos	36
7.9.9	Explicación del Algoritmo NaiveBayes.	36
7.9.10	Herramientas de Minería de Datos.	43
7.10	MODELO DE CALIDAD ISO-9126	45
7.10.1	Estandar ISO 9126	45
7.10.2	Modelo de Calidad (ISO/IEC 9126-1)	46
7.10.3	Modelo de calidad para Calidad Externa e interna	48
7.10.4	Modelo de Calidad para Calidad en Uso	55
7.11	RESEÑA HISTÓRICA MULTICOMERCIAL CECA S.A.	57
VIII.	HIPÓTESIS	58
IX.	DISEÑO METODOLOGICO	59
9.1	TIPO DE ESTUDIO.	59
9.2	ÁREA DE ESTUDIO.	60
9.3	UNIVERSO Y MUESTRA.	60
9.3.1	Universo.	60
9.3.2	Muestra.	60
9.4	MÉTODOS DE RECOLECCIÓN DE DATOS.	61
9.4.1	Instrumentos.	61
9.4.2	Instrumento de recolección de datos.	61
9.4.3	La observación	62
9.4.4	La entrevista	62
9.4.5	La encuesta	63
9.5	PROCESO DE MINERÍA DE DATOS	63
X.	PRESENTACIÓN Y ANÁLISIS DE RESULTADOS	65
10.1	ANÁLISIS DE LA SITUACIÓN DE LA EMPRESA CECA MULTICOMERCIAL S.A.	66

10.2	CREACIÓN DE DATAWAREHOUSE (QUERY)	71
10.3	DIAGRAMA BASE DE DATOS (DATAWAREHOUSE)	73
10.4	SQL SERVER INTEGRATION SERVICES (SSIS): INTEGRACIÓN Y TRANSFORMACIÓN DE LOS DATOS.	74
10.4.1	Flujo de datos - Dimensión clientes	74
10.4.2	Flujo de datos - Dimensión Vendedores	75
10.4.3	Flujo de datos - Dimensión Tiempo	75
10.4.4	Flujo de datos - Dimensión Productos	75
10.4.5	Flujo de datos - Fact Table Temporal	75
10.5	CREACIÓN DE CUBO	76
10.6	EXPLORACIÓN Y MODELADO ALGORITMO NAIVE BAYES	77
10.6.1	¿Cuáles son los posibles compradores de audio profesional?	77
10.6.2	¿Cuáles son los posibles compradores de tecnología?	79
10.7	EVALUACIÓN DE CALIDAD NORMA ISO/IEC 9126-1 -CARACTERÍSTICA DE USABILIDAD	81
10.7.1	Capacidad para ser entendido	81
10.7.2	Capacidad para ser aprendido	82
10.7.3	Capacidad para ser operado	83
10.7.4	Capacidad de atracción	84
10.7.5	Impacto evaluación ISO9126-1 característica de Usabilidad	85
XI.	CRONOGRAMA DEL PROYECTO	87
XII.	PRESUPUESTO	88
XIII.	CONCLUSIONES	89
XIV.	RECOMENDACIONES	90
XV.	ANEXOS	91
XVI.	BIBLIOGRAFÍA	92

ÍNDICE DE FIGURAS

Figura 1: Esquema de una Base de Datos Relacional.....	24
Figura 2: Etapas de la minería de datos.....	33
Figura 3: Resultados del algoritmo naive bayes de microsoft	38
Figura 4: Modelo de calidad externa e interna.....	48
Figura 5: Modelo de calidad para la calidad en uso.....	55
Figura 6: Organigrama de MULTICOMERCIAL S.A. CECA.....	67
Figura 7: Paso para la toma de decisión MULTICOMERCIAL S.A. CECA	67
Figura 8: Tipos de problemas que se presentan en MULTICOMERCIAL S.A. CECA	68
Figura 9: Método para la toma de decisión MULTICOMERCIAL S.A. CECA	68
Figura 10: Tipo de decisiones que enfrenta MULTICOMERCIAL S.A. CECA	69
Figura 11: Condiciones frecuentes del mercado para la toma de decisiones MULTICOMERCIAL S.A. CECA	69
Figura 12: Grafico término 'minería de datos' MULTICOMERCIAL S.A. CECA	70
Figura 13: Grafico soluciones de minería de datos MULTICOMERCIAL S.A. CECA	70
Figura 14: Diagrama Base de datos (DataWarehouse).....	73
Figura 15: SISS - Integración y transformación de los datos	74
Figura 16: Flujo de datos - Dimensión Clientes.....	74
Figura 17: Flujo de datos - Dimensión Vendedores.....	75
Figura 18: Flujo de datos - Dimensión Tiempo.....	75
Figura 19: Flujo de datos - Dimensión Productos.....	75
Figura 20: Flujo de datos - Fact Table Temporal.....	75
Figura 21: Cubo Analysis Services.....	76
Figura 22: red de dependencia - compradores audio profesional	77

Figura 23: Perfil de atributos - compradores audio profesional	77
Figura 24: red de dependencia - compradores audio profesional	77
Figura 25: Característica atributos - compradores audio profesional	78
Figura 26: red de dependencia - compradores tecnología.....	79
Figura 27: Perfil de atributos - compradores Tecnología.....	79
Figura 28: Característica atributos - compradores Tecnología	80
Figura 29: Grafico de Impacto ISO9126-1 Usabilidad.....	86
Figura 30: Cronograma del Proyecto de Investigación.....	87
Figura 31: Anexo - pantalla reportes de ventas MULTICOMERCIAL S.A. CECA	91

ÍNDICE DE TABLAS

Tabla 1: Universo y muestra.....	66
Tabla 2: Base de datos MULTICOMERCIAL S.A. CECA.....	66
Tabla 3: ISO9126-1 Usabilidad-capacidad para ser entendido.	81
Tabla 4: ISO9126-1 Usabilidad-capacidad para ser aprendido.	82
Tabla 5: ISO9126-1 Usabilidad-capacidad para ser operado...	83
Tabla 6: ISO9126-1 Usabilidad-capacidad de atracción.....	84
Tabla 7: Impacto evaluación ISO9126-1 (USABILIDAD).....	85
Tabla 8: Presupuesto de la Investigación.....	88

I. RESUMEN

La presente investigación realiza el análisis y evaluación del algoritmo Naive Bayes en minería de datos. Así mismo hace énfasis en la necesidad de contar con una estrategia de minería de datos a fin de alcanzar los objetivos de marketing dirigido y tener éxito en el mercado.

Se consideró el tipo de estudio de desarrollo tecnológico y aplicado que conlleva la tarea de implementación del algoritmo naive bayes de minería de datos.

Los resultados obtenidos de esta investigación ayudarán a la empresa CECA MULTICOMERCIAL S.A. a tomar decisiones estratégicas comerciales que permitirán un mayor flujo de inventario y un incremento en sus ventas.

Se evaluó la calidad de la solución a través de la ISO 9126-1 por medio de la característica de usabilidad para garantizar la satisfacción de la empresa MULTICOMERCIAL S.A. CECA.

Palabras clave: *Minería de datos, Algoritmo Naive Bayes, Decisiones estratégicas comerciales*

II. INTRODUCCIÓN

La presente investigación se refiere al tema de Minería de datos la cual se puede definir como la extracción de conocimientos desde grandes volúmenes de datos. Actualmente la minería de datos es una herramienta importantísima para las empresas.

La característica principal de la minería de datos es suministrar información valiosa a la empresa para tomar decisiones comerciales empresariales.

En esta investigación abordaremos el algoritmo Naive Bayes que según (Jaime MacLennan, 2009) este algoritmo proporciona una forma muy efectiva de explorar sus datos.

El modelo de algoritmo Naive Bayes será utilizado en el área de mercadeo de la empresa CECA MULTICOMERCIAL S.A.

Por lo general aunque Naive Bayes puede ser un predictor poderoso, mucha gente utiliza algoritmos más sofisticados, como los arboles de decisiones y redes neuronales para la predicción cuando estas herramientas están disponibles.

III. ANTECEDENTES

Una inquietud muy grande de las empresas es ¿Cómo conocer a sus clientes? ¿Cómo conocer el mercado? ¿Cómo conocer la competencia? Y siempre esta información está basada en investigación de mercado. Pero actualmente tenemos otras fuentes muy importantes como son las bases de datos y la minería de datos, que es esa extracción de conocimiento que podemos hacer de todas esas fuentes de información.

Las grandes redes sociales como Facebook, Twitter, Instagram, LinkedIn entre otras utilizan minería de datos para crear publicidad personalizada según los gustos, intercambios financieros, amistad, relaciones sexuales o destinos visitados.

Empresas como Google que es el motor más grande de búsquedas en Internet también utilizan la minería de datos. (Google, 2015) *"En Google, la mayor parte de nuestro trabajo en nuestros productos primarios como la búsqueda, social, y los anuncios se basa en la minería de datos a gran escala. Aplicaciones de minería de datos se presentan en varios proyectos específicos, incluyendo el procesamiento del habla, la traducción automática y el procesamiento visual"*

(Google, 2015) *Nos afirma "tenemos uno de los grupos más activos que trabajan en prácticamente todos los aspectos de la minería de datos"*

La publicidad personalizada actual se basa en unos pocos rasgos demográficos, como el sexo, la edad, el empleo o el

lugar de residencia. La propuesta de Google añade nuevos parámetros, como aficiones, webs que visitamos; grupos a los que pertenecemos; personas con quienes hablamos, compartimos intereses o que leen nuestros blogs, e incluso el análisis semántico de lo que decimos. Estos tomarían el papel de atributos en el proceso de la minería de datos. A partir de ellos se obtendrían predicciones sobre si el usuario comprará más un tipo de producto u otro.

En Nicaragua empresas como ULTRANIC y SINSA utilizan minería de datos para buscar clientes e implementar promociones especiales para clientes clasificados por las predicciones de las técnicas de minería de datos.

IV. PLANTEAMIENTO DEL PROBLEMA.

4.1 Caracterización del Problema.

La empresa MULTICOMERCIAL S.A. CECA cuenta con un sistema automatizado de ventas e inventarios de productos en donde les permite hacer varios procedimientos. Una de las grandes debilidades de MULTICOMERCIAL S.A. CECA es que no cuenta con reportes que le permitan conocer los atributos de aquellos clientes que puedan comprar cierto producto u otro.

4.2 Delimitación del Problema.

La empresa MULTICOMERCIAL S.A. CECA necesita información detallada de los productos que no han tenido movimientos y que se puedan recomendar a los clientes para poder tomar decisiones de marketing y brindar los ingresos esperados.

4.3 Formulación el problema.

¿De qué manera podría la Minería de datos favorecer a la toma de decisión en el área de mercadeo de la empresa MULTICOMERCIAL CECAS.A?

4.4 Sistematización del problema.

¿Cuál es la situación actual en la que se encuentra MULTICOMERCIAL S.A. CECA con respecto al inventario de sus productos y el área de mercadeo?

¿De qué forma se puede presentar resultados de análisis para la toma de decisión de la empresa respecto al inventario para el área de mercadeo?

¿Cuáles son los beneficios esperados con la implementación de una aplicación de minería de datos para el análisis de datos y generación de reportes en la empresa MULTICOMERCIAL CECA S.A.?

¿De qué manera se puede validar la calidad de la implementación del algoritmo naive bayes de minería de datos en la empresa MULTICOMERCIAL CECA S.A.?

V. JUSTIFICACIÓN

El incremento de la competitividad, resultado del marketing actual de los nuevos canales de distribución que disponemos como: redes sociales, aplicaciones móviles, telecomunicaciones; Las empresas se enfrentan a una economía globalizada y su éxito depende de la conservación de los clientes actuales y de conseguir nuevos. La necesidad de implementar la minería de datos la cual contiene tecnologías que permiten a las empresas analizar los factores que influyen decisivamente en estos temas.

Se analizó la necesidad de generar reportes específicos a través de la minería de datos para permitir una mejor implementación de estrategias de marketing con los datos adquiridos del análisis.

Esto ayudará a impulsar las ventas de la empresa MULTICOMERCIAL CECA S.A.

Entre los beneficios que proporcionaría la aplicación de una solución de Minería de datos serian:

- Mayor control de los datos de análisis del área de inventario y ventas de la empresa.
- Contribuir a la toma de decisiones tácticas y estratégicas para identificar información clave.
- Permitir a la empresa MULTICOMERCIAL S.A. CECA ser más competitiva.

VI. OBJETIVOS

6.1 Objetivo General

Aplicar el algoritmo Naïve Bayes de minería de datos en el área de mercadeo de la empresa MULTICOMERCIAL S.A. CECA en Managua Nicaragua en el II Semestre del 2015.

6.2 Objetivos Específicos

- Analizar la situación actual de la empresa MULTICOMERCIAL S.A. CECA con respecto al proceso de toma de decisión de mercadeo en la rotación del inventario.
- Aplicar técnicas de minería de datos utilizando el algoritmo de NaïveBayes para la obtención de datos de análisis.
- Presentar resultados de análisis mediante reportes con una solución de minería de datos para la toma de decisión respecto al inventario para el área de mercadeo.
- Valorar la calidad de la solución a través de la norma ISO 9126-1 con la implementación de la métrica de usabilidad.

VII. MARCO TEÓRICO

7.1 Datos

La palabra dato tiene su origen etimológico en el término latino «Datum» que significa "lo dado". Sin embargo, en sentido estricto, en el ámbito de la investigación científica, como señala (FLORES, 1994) "*La mayoría de los autores asumen que el investigador desempeña un papel activo respecto de los datos: el dato es el resultado de un proceso de elaboración, es decir, el dato hay que construirlo*" (p.15)

En el Diccionario de la Real Academia Española, se encuentran, entre varios significados, los siguientes: (Española, s.f.)

(Del lat. *datum*, lo que se da).

1. m. Antecedente necesario para llegar al conocimiento exacto de algo o para deducir las consecuencias legítimas de un hecho.
2. m. Documento, testimonio, fundamento.
3. m. *Inform.* Información dispuesta de manera adecuada para su tratamiento por un ordenador.

7.2 Información

El autor, (TOFFLER, 2006), en su libro «La revolución de la riqueza» nos brinda un concepto que diferencia entre lo que es **dato** y lo que es **información**: "Los datos suelen ser descritos como elementos discretos, huérfanos de contexto: por ejemplo, «300 acciones». Cuando los datos son contextualizados, se convierten en información: por ejemplo, «tenemos 300 acciones de la empresa farmacéutica X»"(p.154)

El autor (CHIAVENATO, 2006), afirma que la información "es un conjunto de datos conteniendo un significado" (p.546)

En el Diccionario de la Real Academia Española, se encuentran, entre varios significados, los siguientes: (Española, s.f.)

(Del lat. *informatiō*, -ōnis).

1. f. Acción y efecto de informar.
2. f. Oficina donde se informa sobre algo.
3. f. Averiguación jurídica y legal de un hecho o delito.
4. f. Pruebas que se hacen de la calidad y circunstancias necesarias en una persona para un empleo u honor. U. m. en pl.
5. f. Comunicación o adquisición de conocimientos que permiten ampliar o precisar los que se poseen sobre una materia determinada.
6. f. Conocimientos así comunicados o adquiridos.

7.3 Conocimiento

(O.C. FERRELL), se refiere al **conocimiento** como "comprender los datos mediante su estudio o por la experiencia" (p.121)

Este concepto es apoyado por el autor (CHIAVENATO, 2006), que indica que el conocimiento "es el recurso más importante de la era de la información. Constituye un activo intangible. Representa la familiaridad o comprensión de ganancias por medio del estudio y de la experiencia."(p.546)

Por otro lado en el Diccionario de la Real Academia Española, se encuentran, entre varios significados, los siguientes: (Española, s.f.)

1. m. Acción y efecto de conocer.
2. m. Entendimiento, inteligencia, razón natural

7.4 Sistemas de Información

Una organización es un sistema, la cual tiene funciones agrupadas en componentes que trabajan juntos para crear utilidades que beneficien tanto a la empresa como a las personas.

(SENN, 1992), indica que todo sistema organizacional depende, en mayor o menor medida, de una entidad abstracta denominada "**Sistema de información**". Este sistema es el medio por el cual los datos fluyen de una persona o departamento hacia otros y puede ser cualquier cosa, desde la comunicación interna entre los diferentes componentes de la organización y líneas telefónicas hasta sistemas de cómputo que generan reportes periódicos para varios usuarios. Los sistemas de información proporcionan servicio a todos los demás sistemas de una organización y enlazan todos sus componentes en forma tal que estos trabajen con eficiencia para alcanzar el mismo objetivo.

(Mannino, 2007), afirma que un sistema es un conjunto de componentes relacionados entre sí que trabajan para alcanzar un objetivo.

Un sistema de información es similar a un sistema físico (tal como el sistema circulatorio), con la excepción de que un **sistema de información** maneja datos en lugar de objetos físicos, como la sangre. (Mannino, 2007)

Un sistema de información acepta datos provenientes de su entorno, los procesa y genera datos de salida para la toma de decisiones. (Mannino, 2007)

Las bases de datos son componentes esenciales de muchos sistemas de información. El papel de una base de datos es proporcionar una memoria de largo plazo para un sistema de información. La memoria de largo plazo contiene entidades y relaciones. (Mannino, 2007)

Las bases de datos no son los únicos componentes de los sistemas de información; también contienen personas, procedimientos, datos de entrada, datos de salida, software y hardware. (Mannino, 2007)

Un sistema de información se puede definir técnicamente como un conjunto de componentes relacionados que recolectan (o Recuperan), Procesan, Almacenan y distribuyen información para apoyar la toma de decisiones y el control en una organización. (Itson, 2012)

7.4.1 Clasificación de los sistemas de información

Los sistemas de información se clasifican según el uso que la empresa le vaya a dar.

Respecto a esto (KENDALL K. E., 2011), nos dice: "Los sistemas de información se desarrollan para distintos fines, dependiendo de las necesidades de los usuarios humanos y la empresa" (p.2)

Cualquiera sea la necesidad de la empresa y con la ayuda del analista de sistema es que se clasifica el tipo de sistema a implementar, dentro de los cuales los autores (KENDALL K. E., 2011) destacan:

- Sistemas de información administrativa (MIS)
- Sistemas de soporte de decisiones (DSS)
- Inteligencia artificial y sistemas expertos (AI)
- Sistemas de soporte de decisiones en grupo (GDSS)
- Sistemas de soporte para ejecutivos (ESS)

7.5 Base de datos

(TSAI, 1990), argumenta que una **base de datos** es una colección de archivos interrelacionados creados con un DBMS. El contenido de una base de datos se obtiene combinando datos de todas la diferentes fuentes en una organización, de tal manera que los datos estén disponibles para todos los usuarios, y los **datos redundantes** puedan eliminarse o al menos minimizarse.

Otros autores como (KENDALL K. E., 2011) afirman que una base de datos es una fuente central de datos con el fin de que varios usuarios la compartan para su uso en varias aplicaciones, el corazón de una base de datos es el sistema de administración de base de datos (DBMS) el cual permite crear, modificar y actualizar las base de datos, la recuperación de los datos y la generación de informes y pantallas.

7.6 Sistema de Administración de bases de datos

Un sistema de administración de bases de datos (DBMS, por sus siglas en Inglés) es un conjunto de componentes que soportan la creación, el uso y el mantenimiento de bases de datos. Inicialmente, los DBMS proporcionaban un eficiente almacenamiento y recuperación de datos. (Mannino, 2007)

7.7 Base de datos relacional

Las bases de datos se pueden organizar en varias formas. El tipo más común es la base de datos relacional. Una base de datos relacional está organizada en tablas representativas, lo cual minimiza la repetición de los datos, lo cual a su vez reduce los errores y el espacio de almacenamiento. (KENDALL K. E., 2011)

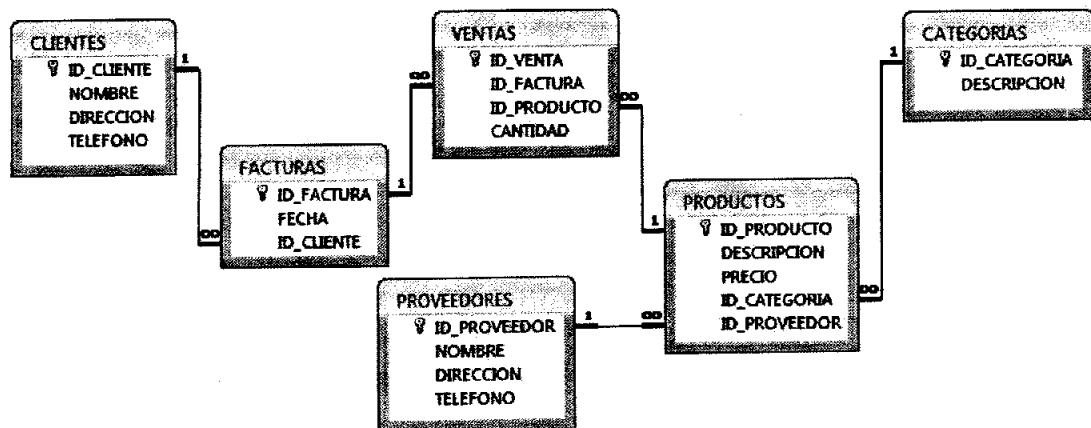


Figura 1: Esquema de una Base de Datos Relacional

Se basa en una representación lógica de los datos, que respeta el modelo relacional. Las bases de datos relacionales se han sabido imponer, ya que todas utilizan un mismo lenguaje estandarizado y normalizado de acceso a datos: SQL (Gabillau)

7.8 Inteligencia de Negocios

7.8.1 Business Intelligence

A como nos afirma (J.Vallejos, 2006) *"Algo peor que no tener información disponible es tener mucha información y no saber qué hacer con ella."*

La inteligencia de negocios o Business Intelligence (BI) es la solución a ese problema, pues por medio de dicha información puede generar escenarios, pronósticos y reportes que apoyen a la toma de decisiones, lo que se traduce en una ventaja competitiva.

En este nuevo mundo, la información reina afirma Geoffrey A. Moore, Director de Chasm Group. Vivimos en una época en que la información es la clave para obtener una ventaja competitiva en el mundo de los negocios.

7.8.2 ¿Qué es Business Intelligence?

Según (J.Vallejos, 2006) Business Intelligence (BI) se puede definir como el proceso de analizar los bienes o datos acumulados en la empresa y extraer una cierta inteligencia o conocimiento de ellos.

7.8.3 Componentes de Business Intelligence

Multidimensionalidad: La Información multidimensional se puede encontrar en hojas de cálculo, base de datos, etc. Una herramienta de BI debe de ser capaz de reunir información dispersa en toda la empresa e incluso en diferentes fuentes para así proporcionar a los departamentos la accesibilidad, poder y flexibilidad que necesitan para analizar la información.

Data Mining: Las empresas suelen generar grandes cantidades de información sobre sus procesos productivos, desempeño operacional, mercados y clientes. Pero el éxito de los negocios depende por lo general de la habilidad para ver nuevas tendencias o cambios en las tendencias. Las aplicaciones de data mining pueden identificar tendencias y comportamientos, no sólo para extraer información, sino también para descubrir las relaciones en base de datos que pueden identificar comportamientos que no son muy evidentes.

Agentes: Los agentes son programas que piensan. Ellos pueden realizar tareas a un nivel muy básico sin necesidad de intervención humana.

Data Warahouse: Es la respuesta de la tecnología de información a la descentralización en la toma de decisiones. Coloca información de todas las áreas funcionales de la organización en manos de quien toma las decisiones. También proporciona herramientas para búsqueda y análisis.

7.9 Minería de datos

Según (KENDALL K. E., 2011) la minería de datos se conoce también como el descubrimiento de conocimiento en base de datos (KDD). Algunos piensan que KDD es distinto a la minería de datos, debido a que su objetivo es ayudar a los encargados de tomar decisiones a encontrar patrones, en vez de ceder el control a un algoritmo para que los encuentre.

(KENDALL K. E., 2011) Afirma que la minería de datos recopila información personal sobre los clientes, en un esfuerzo para ser más específicos al interpretar y anticiparse a sus preferencias. (KENDALL K. E., 2011)

(Paola Garcia) Nos afirma que *"Para un experto, o para el responsable de un sistema, normalmente no son los datos en sí lo más relevante, sino la información que se encierra en sus relaciones, fluctuaciones y dependencias"*. (p.2)

La minería de datos puede identificar patrones que un humano no puede. O el encargado de tomar decisiones no puede ver un patrón, o tal vez no puede pensar en preguntar si existe ese patrón. Los algoritmos de minería de datos buscan patrones en los almacenes corporativos de datos mediante el uso de algoritmos. (KENDALL K. E., 2011)

Según (Microsoft, Conceptos de minería de datos, 2015) estos patrones y tendencias se pueden recopilar y definir como un modelo de minería de datos. Los modelos de minería de datos se pueden aplicar en escenarios como los siguientes:

- **Previsión:** calcular las ventas y predecir las cargas de servidor o el tiempo de inactividad del servidor.
- **Riesgo y probabilidad:** elegir los mejores clientes para la distribución de correo directo, determinar el punto de equilibrio probable para los escenarios de riesgo, y asignar probabilidades a diagnósticos u otros resultados.
- **Recomendaciones:** determinar los productos que se pueden vender juntos y generar recomendaciones.
- **Buscar secuencias:** analizar los artículos que los clientes han introducido en el carrito de compra y predecir los posibles eventos.
- **Agrupación:** separar los clientes o los eventos en clústeres de elementos relacionados, y analizar y predecir afinidades.

Según (M.Ricardo, 2009) La minería de datos (Data Mining) significa descubrir nueva información a partir de un conjunto de datos muy grandes. Por lo general, el conocimiento descubierto está en la forma de patrones o reglas. Además de la tecnología de base de datos, el minado de datos usa técnicas de los campos de estadísticas e inteligencia artificial, de manera especial del aprendizaje de máquinas. Puesto que involucra grandes cantidades de datos es necesario tener una base de datos grande o un almacén de datos.

7.9.1 Propósitos de minería de datos.

(M.Ricardo, 2009) Dice que para la mayoría de las empresas, el propósito final de la minería de datos es proporcionar conocimiento que dará a la compañía una ventaja competitiva, lo que permite ganar un mayor beneficio. Las empresas usan la minería de datos con la esperanza de que podrán lograr lo siguiente:

- Predecir el comportamiento futuro de los atributos.
- Clasificar ítems al colocarlos en las categorías correctas.
- Identificar la existencia de una actividad o un evento.
- Optimizar el uso de los recursos de la organización.

7.9.2 Aplicaciones de la minería de datos

En el libro de (M.Ricardo, 2009) nos dice que la minería de datos se usa en una amplia variedad de aplicaciones. Entre las más exitosas están las siguientes:

- Venta al por menor
 - ✓ Administración de la relación con clientes.
 - ✓ Administración de campañas de publicidad.
- Banca y Finanzas
 - ✓ Calificaciones de crédito.
 - ✓ Detección y Prevención de Fraudes.

➤ Fabricación

- ✓ Optimización del uso de recursos.
- ✓ Optimización del proceso de fabricación.
- ✓ Diseño de Productos.

➤ Medicina

- ✓ Determinación de la efectividad de tratamientos
- ✓ Análisis del efecto de medicamentos.
- ✓ Descubrimiento de Relaciones.

Según (J.Vallejos, 2006) La minería de datos es la etapa de descubrimiento en el proceso KDD: este paso consiste en el uso de algoritmos concretos que generan una enumeración de patrones a partir de los datos procesados. Aunque se suelen usar indistintamente los términos KDD y Minería de Datos.

7.9.3 Características y Objetivos de la Minería de Datos

(J.Vallejos, 2006)

1. Explorar los datos que se encuentran en las profundidades de las bases de datos, como los almacenes de datos, que algunas veces contienen información almacenada durante varios años.
2. En algunos casos, los datos se consolidan en un almacén de datos y en mercados de datos; en otros, se mantienen en servidores de Internet e Intranet.
3. El entorno de minería de datos suele tener una arquitectura cliente servidor.

4. Las herramientas de minería de datos ayudan a extraer el mineral de la información enterrada en archivos corporativos o en registros públicos, archivados.
5. El minero, es muchas veces el usuario final con poca o ninguna habilidad de programación, facultado por barrenadoras de datos y otras poderosas herramientas indagatorias para efectuar preguntas adhoc y obtener rápidamente respuestas.
6. Hugar y sacudir a menudo implica el descubrimiento de resultados valiosos e inesperados.
7. Las herramientas de la minería de datos se combinan fácilmente y pueden analizarse y procesarse rápidamente.
8. Debido a la gran cantidad de datos, algunas veces resulta necesario usar procesamiento en paralelo para la minería de datos.

➤ La minería de datos produce cinco tipos de información:

- Asociaciones
- Secuencias
- Clasificaciones
- Agrupamientos
- Pronósticos.

7.9.4 Técnicas de Minería de Datos

Las técnicas de la minería de datos son el resultado de un largo proceso de investigación y desarrollo de productos. Esta evolución comenzó cuando los datos de negocios fueron almacenados por primera vez en computadoras, y continuo con mejoras en el acceso a los datos, y más recientes con tecnologías generadas para permitir a los usuarios navegar a través de los datos en tiempo real. La minería de datos toma este proceso de evolución más allá del acceso y navegación retrospectiva de los datos, hacia la entrega de información prospectiva y proactiva. La minería de datos esta soportado por tres tecnologías:

- Recolección Masiva de Datos
- Potentes computadoras con multiprocesadores.
- Algoritmos de Data Mining.

7.9.5 Fases de La Minería de Datos.

Los pasos a seguir para la realización de un proyecto de minería de datos según (J.Vallejos, 2006) son:

Siempre los mismos, independientemente de la técnica específica de extracción. De conocimiento usada.

El proceso de minería de datos pasa por las siguientes fases:

- Filtrado de Datos
- Selección de Variables
- Extracción de conocimiento
- Interpretación y Evaluación

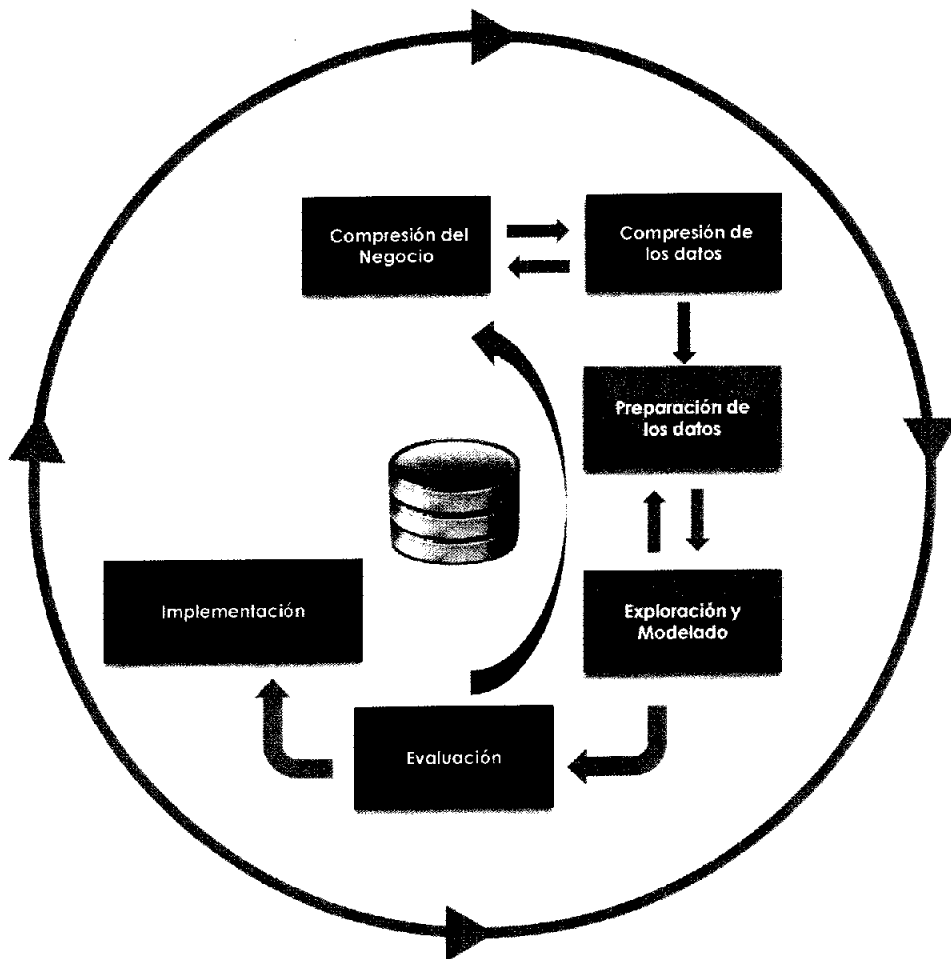
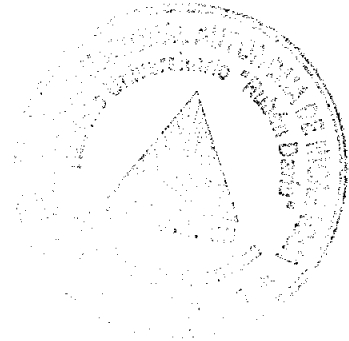


Figura 2: Etapas de la minería de datos

7.9.6 ¿Qué es un Algoritmo?

(Zahonero, 2007) Relata que algoritmo es una especificación concisa del método para resolver un problema con indicación de las acciones a realizar. Un Algoritmo es un conjunto finito de reglas que dan una secuencia de operaciones para resolver un determinado problema. Es por consiguiente, un método para resolver un problema que tiene en general una entrada y una salida. Las características fundamentales que debe cumplir todo algoritmo son:

- Un algoritmo debe ser preciso e indicar el orden de realización de cada paso.
- Un algoritmo debe estar bien definido. Si se sigue un algoritmo dos veces, se debe obtener el mismo resultado cada vez.
- Un algoritmo debe ser finito. Si se sigue un algoritmo, se debe terminar en su momento; es decir; debe tener un número finito de pasos.

También (Alfred, 1988) dice que un Algoritmo es una secuencia finita de instrucciones, cada una de las cuales tiene un significado preciso y puede ejecutarse con una cantidad finita de esfuerzo en un tiempo finito. Las instrucciones de un algoritmo pueden ejecutarse cualquier número de veces, siempre que ellas mismas indiquen la repetición.

No obstante, se exige que un algoritmo termine después de ejecutar un número finito de instrucciones, sin importar cuales fueron los valores de entrada. (Alfred, 1988) Así, un programa es un algoritmo mientras no entre en un ciclo infinito con ninguna entrada.

Es necesario aclarar un aspecto de esta definición de algoritmo. (Alfred, 1988) Ha dicho que cada instrucción de un algoritmo debe tener un significado preciso y debe ser ejecutable con una cantidad finita de esfuerzo; pero lo que está claro pero lo que está claro para una persona, puede no estarlo para otra.

En el centro de la minería de datos hay un conjunto de algoritmos que SSAS aplica a sus datos. (Hotek, 2009) Dice que un algoritmo de minería de dato es una ecuación matemática que cuando se aplica a sus datos le puede decir la probabilidad del resultado deseado.

7.9.7 Categorías de los Algoritmos en minería de datos

(Hotek, 2009) Dice que los algoritmos de minería de datos se pueden dividir en seis categorías básicas:

- Clasificación
- Asociación
- Regresión
- Pronóstico
- Análisis de secuencia
- Análisis de desviaciones

7.9.8 Algoritmos de minería de datos

En el libro de (Hotek, 2009) Encontramos que "SQL Server viene con siete Algoritmos de minería de datos que abarcan todas las categorías de la minería, excepto el análisis de desviación". Los algoritmos de minería de datos que están disponibles son:

- Naive Bayes
- Arboles de decisión de Microsoft
- Microsoft Clustering
- Clusters de Secuencia
- Reglas de Asociación
- Redes Neuronales
- Series de Tiempo

7.9.9 Explicación del Algoritmo NaïveBayes.

El algoritmo BayesNaive permite crear rápidamente modelos con habilidades de proporcionar un nuevo método de exploración y la compresión de sus datos nos dice (Jaime MacLennan, 2009). Los recuentos tomados por el algoritmo BayesNaive pueden ser utilizados para explorar las relaciones entre los diversos atributos en su modelo.

(Jaime MacLennan, 2009) Nos dice que este Algoritmo proporciona una forma muy efectiva de explorar sus datos. Por lo general aunque NaiveBayes puede ser un predictor poderoso, mucha gente utiliza algoritmos más sofisticados, como los arboles de decisiones y redes neuronales para la predicción cuando estas herramientas están disponibles.

El algoritmo Bayes naive de Microsoft es un algoritmo de clasificación basado en los teoremas de Bayes y que Microsoft SQL Server Analysis Services proporciona para el modelado de predicción. La palabra naïve (ingenuo en inglés) del término Bayes naive proviene del hecho que el algoritmo utiliza técnicas Bayesianas pero no tiene en cuenta las dependencias que puedan existir. (Microsoft, Algoritmo Naive Bayes , 2015)

7.9.9.1 Ejemplo de aplicación del algoritmo Naive Bayes

Como parte de su estrategia promocional, el departamento de comercialización de la empresa Adventure Works Cycles ha decidido atraer a posibles clientes realizando un envío por correo de folletos. Para reducir costos, desean enviar los folletos solo a los clientes de los que esperan recibir respuesta. La empresa almacena información en una base de datos sobre datos demográficos y respuestas a envíos de correo anteriores. Desean utilizar estos datos para ver el modo en que los datos demográficos como la edad o la ciudad pueden ayudarles a predecir la respuesta a una promoción, comparando los clientes potenciales con los que tienen características similares y con los que han adquirido productos de la empresa en el pasado. En concreto, lo que desean es ver las diferencias entre los clientes que adquirieron una bicicleta y los que no lo hicieron.

Mediante el algoritmo Bayes naive de Microsoft, el departamento de comercialización puede predecir rápidamente un resultado de un perfil de cliente concreto y, por tanto, puede determinar qué clientes responderán a los folletos con más probabilidad. Con el Visor Bayes naive de Microsoft de SQL Server Data Tools (SSDT), también pueden investigar visualmente qué columnas de entrada específicas contribuyen a conseguir respuestas positivas a los folletos.

7.9.9.2 Funcionamiento del algoritmo

El algoritmo Bayes naive de Microsoft calcula la probabilidad de cada estado de cada columna de entrada, dado cada posible estado de la columna de predicción. Para comprender cómo funciona, utilice el Visor Bayes naive de Microsoft de SQL Server Data Tools (SSDT) (como se muestra en el siguiente gráfico) para consultar una representación visual del modo en que el algoritmo distribuye los estados.

Atributos	Estados	Población... Tamaño: 18484	0 Tamaño: 9352	1 Tamaño: 9132	ausente Tamaño: 0
Age	<ul style="list-style-type: none"> ● 38 - 43 ● 29 - 34 ● 43 - 48 ● Other 	■	■	■	■
Commute Distance	<ul style="list-style-type: none"> ● 0-1 Miles ● 2-5 Miles ● 1-2 Miles ● Other 	■	■	■	■
Education	<ul style="list-style-type: none"> ● Bachelors ● Partial College ● High School ● Other 	■	■	■	■
Marital Status	<ul style="list-style-type: none"> ● M ● S ● Missing 	■	■	■	■
Number Cars Owned	<ul style="list-style-type: none"> ● 2 ● 1 ● 0 ● Other 	■	■	■	■
Number Children At Home	<ul style="list-style-type: none"> ● 0 ● 1 ● 2 ● Other 	■	■	■	■
Occupation	<ul style="list-style-type: none"> ● Professional ● Skilled Manual ● Management 	■	■	■	■

Figura 3: Resultados del algoritmo naive bayes de microsoft

Aquí, el Visor Bayes naive de Microsoft muestra cada columna de entrada del conjunto de datos e indica cómo se distribuyen los estados de cada columna, dado cada estado de la columna de predicción.

Esta vista del modelo se utilizaría para identificar las columnas de entrada que son importantes para diferenciar los distintos estados de la columna de predicción.

Por ejemplo, en la fila Commute Distance que se muestra aquí, la distribución de valores de entrada es visiblemente diferente para los compradores en comparación con los no compradores. Esto indica que la entrada, Commute Distance = 0-1 miles, es un factor de predicción potencial.

El visor también proporciona valores para las distribuciones, de modo que pueda ver que para los clientes que viajan entre una y dos millas para ir a trabajar, la probabilidad de que compren una bicicleta es de 0,387, y la probabilidad de que no la compren es de 0,287. En este ejemplo, el algoritmo utiliza la información numérica, derivada de un dato de cliente (como la distancia entre el domicilio y el lugar de trabajo), para predecir si un cliente compraría una bicicleta.

7.9.9.3 Datos requeridos para los modelos Naive Bayes

(Microsoft, Algoritmo Naive Bayes , 2015) Nos indica que al preparar los datos para su uso en un modelo de entrenamiento Naive bayes, conviene comprender qué requisitos son imprescindibles para el algoritmo, incluidos el volumen de datos necesario y la forma en que estos datos se utilizan.

Los requisitos para un modelo Naive bayes son los siguientes:

- **Una columna de una sola clave:** cada modelo debe contener una columna numérica o de texto que identifique cada registro de manera única. No están permitidas las claves compuestas.
- **Columnas de entrada:** en un modelo Naive bayes, todas las columnas deben ser discretas o discretizadas.

En un modelo Naive bayes, también es importante asegurarse de que los atributos de entrada sean independientes unos de otros. Esto es particularmente importante al utilizar el modelo para la predicción.

El motivo es que, si utiliza dos columnas de datos que ya están estrechamente relacionadas, el efecto sería multiplicar la influencia de esas columnas, lo que puede ocultar otros factores que influyen en el resultado.

- Al contrario, la capacidad del algoritmo de identificar las correlaciones entre las variables es útil cuando está explorando un modelo o conjunto de datos, para identificar las relaciones entre las entradas.
- **Al menos una columna de predicción:** el atributo de predicción debe contener valores discretos o discretizados.

Los valores de la columna predecible se pueden tratar como entradas. Este ejercicio puede ser útil si explora un nuevo conjunto de datos, para encontrar relaciones entre las columnas.

7.9.9.4 Parámetros de BayesNaive

La implementación de BayesNaive es bastante sencilla y por lo tanto, no está fuertemente parametrizada nos dice (Jaime MacLennan, 2009). Este algoritmo considera todas las combinaciones de atributos por pares, el tiempo y la memoria para procesar los datos que se relacionan con el número total de valores de entrada multiplicado por el número total de valores de salida.

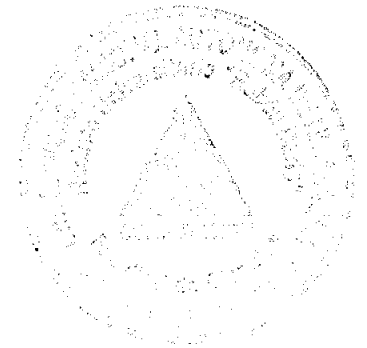
En general según (Jaime MacLennan, 2009) El algoritmo hace un buen trabajo de la elección que se consideran entradas y salidas cuando los parámetros se aplican, y cada uno de ellos se puede desactivar para obligar al algoritmo a considerarlas todas.

En el libro de (Jaime MacLennan, 2009) Encontramos los siguientes parámetros del algoritmo BayesNaive:

Atributos de entrada Máxima: Los parámetros de entrada máxima determinan el número de atributos que serán considerados como insumo para la formación. Si hay más de este número de entradas el algoritmo seleccionara las entradas más importantes e ignorar el resto.

Atributos de salida Máxima: El parámetro máximo de salida determina el número de atributos que serán considerados como salida para la formación. Si hay más de este número de salidas el algoritmo seleccionara las salidas más importantes.

Estado Máximos Unidos: Controla la consideración de muchos estados de un atributo. Si un atributo tiene más que este número de estado, solo los estados más populares se utilizaran.



7.9.9.5 Limitaciones del Algoritmo BayesNaive

El Algoritmo BayesNaive pertenece a una clase de algoritmos conocidos como lineal clasificadores. En el Espacio Multidimensional definido por las entradas, un clasificador lineal actúa como un hiperplano que separa los puntos de datos que pertenecen a diferentes clases nos dice (Jaime MacLennan, 2009)

7.9.10 Herramientas de Minería de Datos.

Actualmente las herramientas comerciales de minerías de datos utilizan varias técnicas comunes para la extracción del conocimiento según nos dice (ELMASRI, 2007). Entre estas se incluyen las reglas de asociación, el agrupamiento, las redes neuronales, la secuencia y el análisis estadístico. También se utilizan arboles de decisiones, que son una representación de las reglas utilizadas en la clasificación o en agrupamiento, y los análisis estadísticos, que pueden incluir la regresión y muchas otras técnicas.

(ELMASRI, 2007) Nos menciona que otros productos comerciales utilizan técnicas avanzadas como los algoritmos genéticos, los razonamientos basados en casos, las redes Bayesianas, la regresión no lineal, la optimización combinatoria, el emparejamiento de patrones y la lógica difusa.

(ELMASRI, 2007) Aclara que "la mayoría de las herramientas de minerías de datos utilizan la interfaz ODBC (Open DatabaseConnectivity, conectividad de bases de datos abierta). ODBC es un estándar de la industria que trabaja

con bases de datos; permite el acceso a los datos en la mayoría de los programas de bases de datos más conocidos como Access, dBASE, Informix, Oracle y SQL Server.” (p.845)

7.9.10.1 SQL Server

(Microsoft, Sql Server, 2015) Nos indica que SQL Server™ es un sistema de administración y análisis de bases de datos relacionales de Microsoft para soluciones de comercio electrónico, línea de negocio y almacenamiento de datos. En esta sección, encontrará información sobre varias versiones de SQL Server. También encontrará artículos sobre bases de datos y aplicaciones de diseño de bases de datos así como ejemplos de los usos de SQL Server.

7.9.10.2 Analysis Services

Analysis Services es un motor de datos analíticos en línea que se usa en soluciones de ayuda a la toma de decisiones y Business Intelligence (BI), y proporciona los datos analíticos para informes empresariales y aplicaciones cliente como Excel, informes de Reporting Services y otras herramientas de BI de terceros. Un flujo de trabajo típico para Analysis Services incluye la creación de un modelo de datos OLAP o tabular, la implementación del modelo como base de datos en una instancia de Analysis Services, el procesamiento de la base de datos para cargarla con datos y, a continuación, la asignación de permisos para permitir el acceso a datos. Cuando esté listo, se puede obtener acceso a este modelo de datos con varios fines desde cualquier

aplicación cliente que admita Analysis Services como origen de datos. (Microsoft, Analysis Services, 2015)

7.9.10.3 Microsoft Excel

La minería de datos con la conocida herramienta de Microsoft Excel. Simplicidad y facilidad de uso puede crear potentes modelos predictivos sin conocimiento de minería de datos de profundidad.

7.10 Modelo de Calidad ISO-9126

7.10.1 Estandar ISO 9126

La organización internacional para la estandarización en inglés (International Organization for Standardization) ISO y la Comisión Electrotécnica Internacional en inglés (International Electrotechnical Commission) IEC son organizaciones que permiten estandarizar o normar sistemas o directrices para la calidad, evaluación, seguridad, etc, para la industria del software y de las ciencias de computación a nivel mundial.

La Norma ISO/IEC 9126 estandariza la calidad del producto de software, esta consta de cuatro partes:

- Parte 1: Modelo de Calidad (ISO/IEC 9126-1)
- Parte 2: Métricas Externas (ISO/IEC 9126-2)
- Parte 3: Métricas Internas (ISO/IEC 9126-3)
- Parte 4: Métricas de Calidad en Uso (ISO/IEC 9126-4)

El hecho de que un producto de software, sea este una aplicación web, aplicación de escritorio, aplicación móvil, aplicación Smart client cumpla las directrices de la ISO 9126 nos da un grado de confianza de que ese producto de software tiene calidad.

Para nuestro caso de estudio, el hecho de que la aplicación del algoritmo Naive Bayes modelo estructural de minería de datos en la empresa CECA MULTICOMERCIAL S.A. cumpla la norma ISO 9126 garantizaría una calidad aceptable a nivel internacional y por ende facilitaría la comercialización de este producto en mercados similares.

7.10.2 Modelo de Calidad (ISO/IEC 9126-1)

En esta parte de la norma ISO/IEC 9126 se detalla el modelo a usar para la calidad del producto de software, que a su vez se divide en dos partes:

- Calidad interna y calidad externa
- Calidad en uso

La Calidad Interna y Calidad Externa del modelo describe a la calidad del software, basándose en seis características principales que a su vez se dividen en sus respectivas sub características.

La calidad en Uso del modelo se base en cuatro características primordiales para determinar la calidad de uso desde la perspectiva del usuario de un sistema.

El estándar ISO/IEC 9126 puede ser usado desde varias perspectivas como son: adquisición, desarrollo, uso, soporte, mantenimiento y auditoria de software.

7.10.2.1 Ejemplos de uso del Modelo de Calidad

- Validar la integridad de una definición de requisitos.
- Identificar requisitos del software
- Identificar objetivos para el diseño software
- Identificar requisitos para el aseguramiento de la calidad
- Identificar criterios de aceptación para un producto software en producción

7.10.3 Modelo de calidad para Calidad Externa e interna

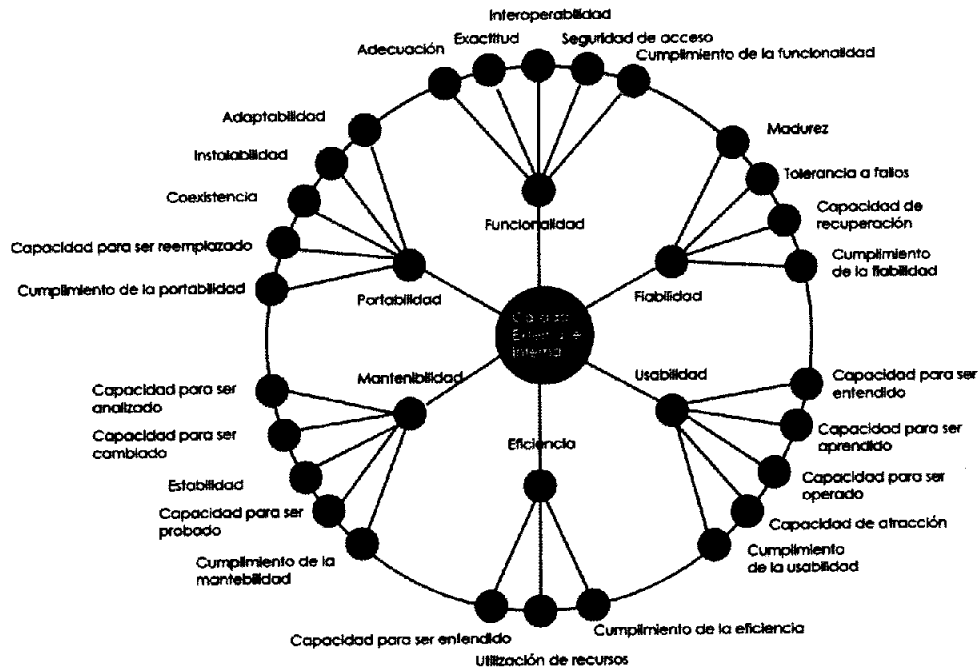


Figura 4: Modelo de calidad externa e interna

El modelo de calidad de la ISO 9126 se describe a partir de seis características generales (Funcionalidad, Fiabilidad, Usabilidad, Eficiencia, Mantenibilidad y Portabilidad) para la calidad interna y externa, cada una de ellas con sub características que pueden ser medidas por métricas internas o externas según corresponda.

Funcionalidad: es la capacidad del producto de software para proporcionar funciones que permitan satisfacer las necesidades básicas de funcionamiento cuando el software es usado en condiciones específicas.

Las subcaracterísticas de la funcionalidad son:

- **Adecuación:** capacidad del producto de software para proveer un conjunto apropiado de funciones para tareas y objetivos de usuario específicos.
- **Exactitud:** capacidad del producto de software para proveer los resultados o efectos correctos o acordados, con el grado necesario de precisión.
- **Interoperabilidad:** capacidad del producto de software para operar o interactuar con uno o más sistemas especificados.
- **Seguridad de acceso:** capacidad del producto de software para proveer una excelente protección de la información y datos que maneja el producto de software, de manera que las personas o sistemas ajenos a este, o no autorizados no puedan leerlos o modificarlos.

Es decir que con esta característica se dé el acceso a la información a usuarios autorizados y se deniegue el acceso a las personas o sistemas no autorizados.

- **Cumplimiento de la funcionalidad:** capacidad del producto de software para adherirse a estándares, normas, y buenas prácticas relacionadas con funcionalidad.

Fiabilidad: es la capacidad del producto de software para mantener un buen nivel aceptable de rendimiento cuando es usado bajo parámetros o condiciones específicas.

Las subcaracterísticas de la fiabilidad son:

- **Madurez:** capacidad del producto de software para evitar un fallo técnico del producto de software, no como resultado de alguna falla provocada por el usuario.
- **Tolerancia a fallos:** capacidad del producto para mantener un buen nivel aceptable de rendimiento en caso de fallos de software.
- **Capacidad de recuperación:** capacidad del producto de software para restablecer un nivel aceptable de rendimiento específico y de recuperar los datos involucrados después de algún fallo en el producto de software.
- **Cumplimiento de la fiabilidad:** capacidad del producto de software para adherirse a estándares, normas, convenciones, regulaciones, o buenas prácticas relacionadas con la fiabilidad.

Usabilidad: es la capacidad del producto de software para ser aprendido, facilidad de ser usado y que sea un producto de software considerado atractivo para el usuario bajo condiciones específicas. Para esta característica pueden incluirse perspectivas diferentes como: usuarios, operadores, usuarios finales y usuarios indirectos que tienen relación con el uso del software.

Las subcaracterísticas de la usabilidad son:

- **Capacidad para ser entendido:** capacidad del producto de software que permite a un determinado usuario entender si el software es adecuado para sus necesidades y cómo puede ser usado para determinadas tareas o condiciones de uso.
- **Capacidad para ser aprendido:** capacidad del producto de software que permite al usuario aprender el manejo del producto de software.
- **Capacidad para ser operado:** capacidad del producto de software que permite al usuario operarlo y controlarlo.
- **Capacidad de atracción:** capacidad del producto de software para ser considerado atractivo a un determinado usuario.
- **Cumplimiento de la usabilidad:** capacidad del producto de software para adherirse a estándares, normas, convenciones, guías de estilo, regulaciones o buenas prácticas relacionadas con la usabilidad.

Eficiencia: es la capacidad del producto de software para proporcionar un apropiado y básico rendimiento, relativo a la cantidad de recursos usados bajo parámetros y condiciones específicas.

Las subcaracterísticas de la eficiencia son:

- **Comportamiento temporal:** capacidad del producto de software para proporcionar tiempos de respuesta y tiempos de proceso apropiados, bajo condiciones determinadas.
- **Utilización de recursos:** capacidad del producto de software para usar adecuadamente los recursos adecuados cuando el producto de software está funcionando y operando bajo condiciones determinadas.
- **Cumplimiento de la eficiencia:** capacidad del producto de software para adherirse a estándares, normas, convención eso buenas prácticas relacionadas con la eficiencia.

Las características como la funcionalidad, fiabilidad, usabilidad y eficiencia pueden ser medidas externamente por la calidad en uso mediante diferentes perspectivas de usuarios que utilizan el producto de software.

Mantenibilidad: es la capacidad del producto de software para ser modificado al estar en producción, las modificaciones pueden incluir correcciones, mejoras, adaptaciones del software, cambios en el entorno de operación del software o sugerencias por parte de los usuarios.

Las sub características de la mantenibilidad son:

- **Capacidad para ser analizado:** es la capacidad del producto de software para diagnosticar deficiencias o causas de los fallos en el software, o para identificar las partes que van a tener que ser modificadas.
- **Capacidad para ser cambiado:** capacidad del producto de software que permite que una determinada modificación sea implementada sin afectar otras funcionalidades del producto de Software.
- **Estabilidad:** capacidad del producto de software para evitar efectos inesperados a causa de modificar el producto de software
- **Capacidad para ser probado:** capacidad del producto de software que permite que el software modificado sea validado y cumpla la funcionalidad por la cual se modificó.
- **Cumplimiento de la mantenibilidad:** capacidad del producto software para adherirse a estándares, normas, convenciones, buenas prácticas relacionadas con la mantenibilidad.

Portabilidad: es la capacidad del producto de software para ser trasladado de un ambiente determinado donde está funcionando correctamente hacia otro. El <ambiente puede ser una organización o entornos de hardware o software determinados.

Las subcaracterísticas de la portabilidad son:

- **Adaptabilidad:** capacidad del producto de software para ser adaptado a diferentes entornos o ambientes específicos, sin aplicar acciones o mecanismos diferentes de aquellos proporcionados inicialmente para el correcto funcionamiento del producto de software.
- **Instalabilidad:** capacidad del producto software para ser instalado en un entorno específico (Entorno de Hardware y Software).
- **Coexistencia:** capacidad del producto software para coexistir con otro software independiente a éste, en un ambiente o entorno común, compartiendo recursos específicos.
- **Capacidad para reemplazar:** capacidad del producto de software para ser usado en lugar de otro producto de software, para cumplir el mismo propósito, y en el mismo entorno de operación del software.

- **Cumplimiento de la portabilidad:** capacidad del producto software para adherirse a estándares, normas, convenciones, o buenas prácticas relacionadas con la portabilidad.

7.10.4 Modelo de Calidad para Calidad en Uso

El modelo describe a la calidad en uso del producto de software a partir de cuatro características generales (Efectividad, Productividad, Seguridad, Satisfacción)

Lograr la calidad en uso depende básicamente de lograr la calidad externa y esta depende de lograr la calidad interna del Producto de Software.

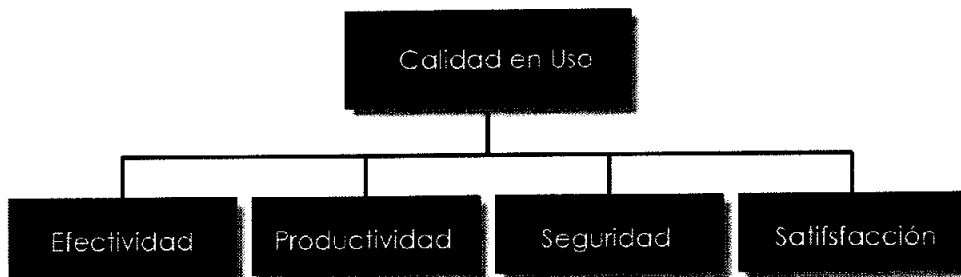


Figura 5: Modelo de calidad para la calidad en uso

Calidad en Uso: es la capacidad del producto de software de proveer características como: efectividad, productividad, seguridad y satisfacción al momento que el producto de software está en producción y desde las diferentes perspectivas de los usuarios que utilizan dicho producto.

Efectividad: capacidad del producto de software para alcanzar objetivos específicos con exactitud y completitud dependiendo las necesidades de cada uno de los usuarios que utilizan el producto de software dentro de un determinado uso específico.

Productividad: capacidad del producto de software que permite a los usuarios utilizar un porcentaje adecuado de los recursos con relación a la efectividad alcanzada al utilizar el producto de software dentro de un determinado uso específico.

Seguridad: capacidad del producto de software para alcanzar niveles mínimos y aceptables del riesgo de producir daño a personas, al negocio, al software, a la organización, a las propiedades o al medio ambiente dentro de un determinado uso específico del producto de software.

Satisfacción: capacidad del producto de software para satisfacer las necesidades mínimas que tienen los usuarios al utilizar el producto de software dentro de un determinado uso específico del producto de software.

7.11 Reseña Histórica MULTICOMERCIAL CECA S.A.

MULTICOMERCIAL, S.A. CECA fue fundada en 1971 en la ciudad de Managua, Nicaragua, como un negocio minorista de repuestos electrónicos y accesorios. Luego fueron incorporados sistemas de audio profesional, instrumentos musicales, iluminación profesional, equipos y accesorios para computadoras.

Con más de 40 años en el mercado, una ubicación estratégica en un área central de comercio de la capital, dos puntos principales de venta (CECA Electrónica y CECA Musical), un personal profesional, enfocado en proporcionar a nuestros clientes, servicios y soluciones integrales y productos de calidad a precios competitivos; nos ha convertido en un socio por excelencia en las áreas de Audio Profesional y Electrónica.



VIII. HIPÓTESIS

"Con la aplicación de minería de datos la empresa MULTICOMERCIAL S.A. CECA obtendría un incremento en la rotación de sus productos."



IX. DISEÑO METODOLOGICO

9.1 Tipo de Estudio.

En esta investigación analizaremos el proceso de toma de decisión en el área de mercadeo de la empresa con respecto a la rotación del inventario por lo que se considerará el tipo de estudio retrospectivo de desarrollo tecnológico y aplicado.

Consideramos que la investigación según el periodo y secuencia del estudio será del tipo transversal debido a que tomaremos un determinado periodo (primer semestre 2015) de tiempo para el desarrollo de nuestra investigación.

Según (Sampieri, 2006), *"los diseños de investigación transeccional o transversal recolectan datos en un solo momento, en un tiempo único. Su propósito es describir variables y analizar su incidencia e interrelación en un momento dado."* (p.208)

Además según su diseño metodológico es descriptivo estos según (Sampieri, 2006, pág. p.102) *"Buscan especificar las propiedades, las características y perfiles de las personas, grupos, comunidades, procesos, objetos o cualquier otro fenómeno que se someta a un análisis"*. Es decir, miden, evalúan o recolectan datos sobre diversos conceptos (variables), aspectos, dimensiones o componentes del fenómeno a investigar.

9.2 Área de Estudio.

El área de estudio será en la empresa MULTICOMERCIAL S.A. CECA ubicada en la Calle Principal de Altamira Managua N° 558 y 560 distribuida en dos edificios que ofrecen Instrumentos Musicales, Audio Profesional, Iluminación Profesional, Tecnología, Electrónica y artículos para el Hogar.

Debido a la gran gama de artículos que ofrece MULTICOMERCIAL S.A. CECA se seleccionaron las categorías de Tecnología, Equipos Electrónicos y Audio Profesional para llevar acabo nuestra investigación.

9.3 Universo y Muestra.

9.3.1 Universo.

El Universo en MULTICOMERCIAL S.A. CECA está constituida por la gerencia, gerencia comercial, gerencia financiera, departamento de ventas, departamento de compras, Contabilidad y el departamento de Informática; conformado por 48 trabajadores de los dos edificios que conforman esta entidad.

9.3.2 Muestra.

La muestra está conformada por 8 empleados del área de gerencia y jefes de departamentos de MULTICOMERCIAL S.A. CECA seleccionados por su papel en la toma de decisiones de la empresa.

9.4 Métodos de recolección de datos.

Existen dos métodos de recolección de datos como son: Cuantitativos y Cualitativos.

En nuestro caso de estudio utilizaremos el método cuantitativo.

9.4.1 Instrumentos.

En el presente trabajo investigativo se presentan diferentes tipos de instrumentos que me sirvieron de apoyo en el lapso del estudio. La recopilación de datos fue una de las etapas del proceso de investigación y nos permitió la utilización de diversos instrumentos de gran importancia que nos permitieron la extracción de datos del campo de estudio, como son la observación, las entrevistas y las encuestas estos fueron elaborados por un medio de escritos que nos sirvieron de guía en el momento de ser aplicados.

9.4.2 Instrumento de recolección de datos.

Una vez definido el tipo de investigación, fue necesario determinar los métodos de investigación a utilizar para la obtención de la información precisa de la situación en estudio. Al respecto (SEQUEIRA V. Y CRUZ, 1997) Expresa: *"En el desarrollo de las etapas del proceso investigativo, especialmente en la recolección de la información, se requiere el uso de métodos que permitan el contacto con el fenómeno de estudio y posibiliten el acceso a la información que demanda la investigación para el logro de los objetivos"*.

Como instrumento de recolección de datos utilicé la observación, la entrevista y la encuesta como análisis de la información.

9.4.3 La observación

(SEQUEIRA V. Y CRUZ, 1997) Afirma: "Ha sido considerada como un método fundamental en la búsqueda de la información cuando se estudia un determinado fenómeno".

El método permitió tener contacto más cercano con el fenómeno que se estudió y tener un conocimiento más objetivo de sus características. El método de observación directa se utilizó para tener contacto visual con la situación real del problema ya que ésta es la única forma de comprender con mayor profundidad el proceso de análisis de la información obtenida de la base de datos de parte de la gerencia de MULTICOMERCIAL S.A. CECA.

9.4.4 La entrevista

La entrevista es la comunicación establecida entre el investigador y el sujeto de estudio a fin de obtener respuestas verbales a las interrogantes planteadas sobre el problema propuesto.

Al respecto (SEQUEIRA V. Y CRUZ, 1997) Nos indica que "*La entrevista al igual que la encuesta y la observación, es un método que permite registrar la información del fenómeno que se estudia*". Esta técnica la utilizaremos con el propósito de recopilar información planteada por personas que

interactúan directamente con el desarrollo de estos procesos y funciones que se llevan a cabo en la organización.

En la presente investigación se realizó la entrevista a fin de obtener respuesta a las interrogantes planteadas sobre el caso de estudio lo que permitió analizar la situación actual en la que se encuentra la empresa MULTICOMERCIAL S.A. CECA al tomar una decisión comercial con respecto al proceso de mercadeo de sus productos.

9.4.5 La encuesta

Según (Bernal, 2010), "La encuesta o cuestionario es un conjunto de preguntas diseñadas para generar los datos necesarios, con el propósito de alcanzar los objetivos del proyecto de investigación. Se trata de un plan formal para recabar información de la unidad de análisis objeto de estudio y centro del problema de investigación." (p.250)

En esta investigación se utilizó la encuesta para obtener resultados de calidad de uso del estudio realizado.

9.5 Proceso de Minería de datos

El estándar utilizado para llevar a buen término el proceso de minería de datos es el modelo CRIPS-DM (Cross Industry Estándar Process for Data Mining) (Figura 2: Etapas de la minería de datos)

Que es un modelo basado en el entorno industrial y empresarial el cual da las pautas para realizar un buen proceso de minería de datos.

Porque no solamente es extraer la información sino también entender cómo funciona el negocio. Por supuesto entender también los datos. Una cosa es entender el negocio y otra muy diferente como almacenan los datos.

Preparar los datos: Es muy importante que los datos estén limpios de suciedad, estén limpios de datos inútiles, que los datos tengan calidad antes de pasar a la fase de modelado que ya es la fase de extracción del conocimiento. Y una vez extraída la información es muy importante la evaluación. Si la información es lo suficientemente precisa se pasa a la fase de desarrollo para que el cliente tomé las decisiones del negocio.

X. PRESENTACIÓN Y ANÁLISIS DE RESULTADOS

En este apartado se muestra los resultados obtenidos durante la implementación de minería de datos, a continuación se enumeran cada uno de los resultados.

- Análisis de la situación de la empresa CECA MULTICOMERCIAL S.A.
- Creación de Datawarehouse (query)
- Diagrama Base de datos (DataWarehouse)
- SQL Server **Integration Services** (SSIS): Integración y transformación de los datos.
 - o Flujo de datos - Dimensión Clientes
 - o Flujo de datos - Dimensión Vendedores
 - o Flujo de datos - Dimensión Tiempo
 - o Flujo de datos - Dimensión Productos
 - o Flujo de datos - Fact Table Temporal
- Creación de Cubo
- Exploración y Modelado Algoritmo Naive Bayes
- Evaluación de Calidad Norma ISO/IEC 9126-1 - característica de usabilidad

10.1 Análisis de la situación de la empresa CECA
MULTICOMERCIAL S.A.

Tabla 1: Universo y muestra

Universo	Muestra
48 empleados	8 empleados (Gerencia y jefes de áreas)

Tabla 2: Base de datos MULTICOMERCIAL S.A. CECA

Base de datos Relacional	Objetivo
EXACTUS	Producción
CECA	<ul style="list-style-type: none"> • Información Clientes • Reportes Gerenciales
EXATEST	Pruebas de ensayo

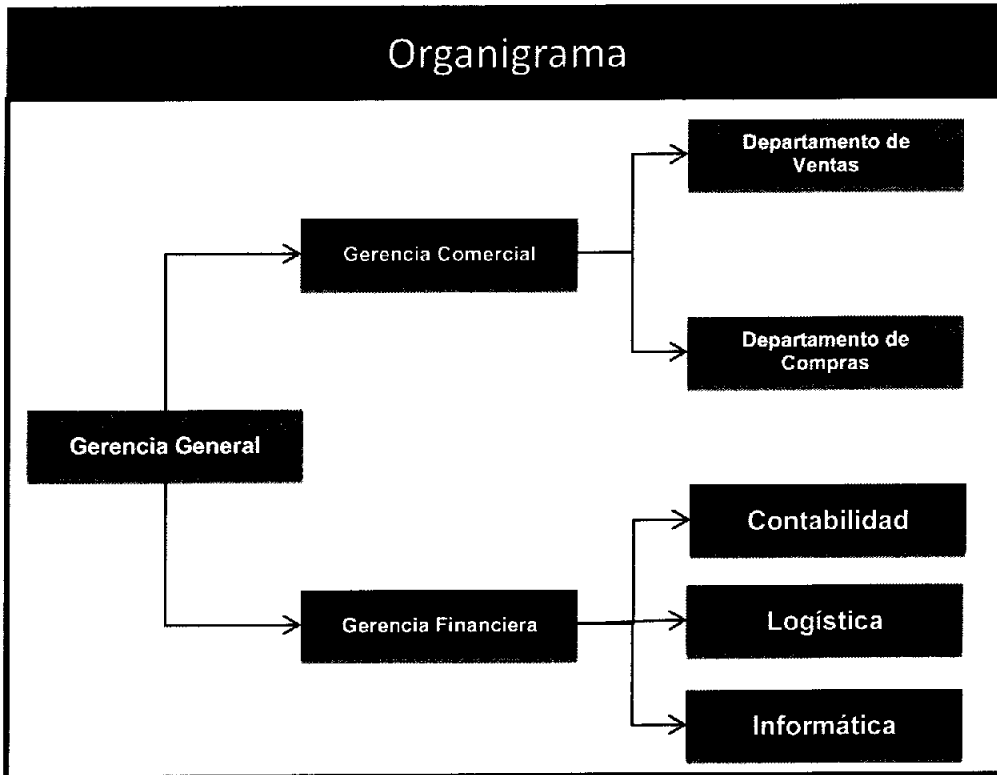
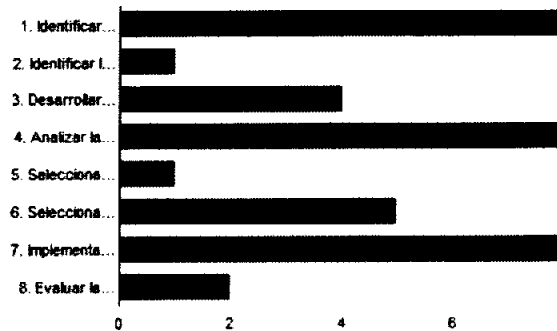


Figura 6: Organigrama de MULTICOMERCIAL S.A. CECA

¿Qué procedimiento o serie de pasos regularmente utiliza para tomar decisiones?



1. Identificar el problema	8	100%
2. Identificar los criterios de la decisión	1	12.5%
3. Desarrollar las alternativas	4	50%
4. Analizar las alternativas	8	100%
5. Seleccionar las alternativas	1	12.5%
6. Seleccionar una alternativa	5	62.5%
7. Implementar la decisión	8	100%
8. Evaluar la decisión	2	25%

Figura 7: Paso para la toma de decisión MULTICOMERCIAL S.A. CECA

¿Qué método utiliza para tomar decisiones?

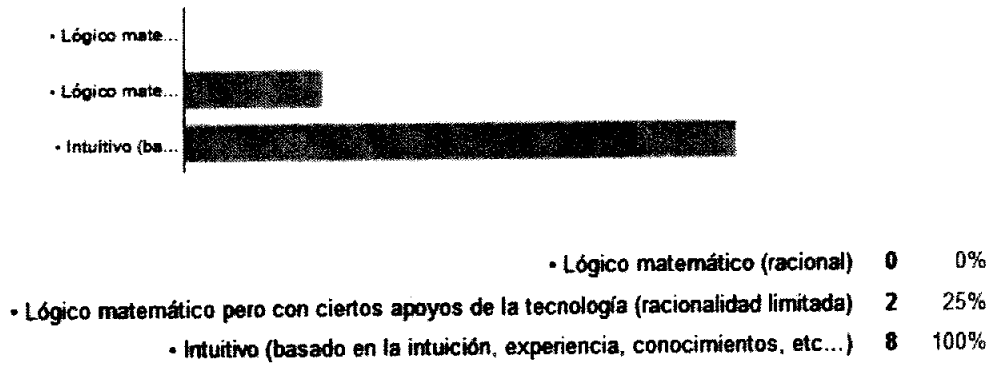


Figura 9: Método para la toma de decisión MULTICOMERCIAL S.A. CECA

¿Qué tipos de problemas se le presentan con más frecuencia?

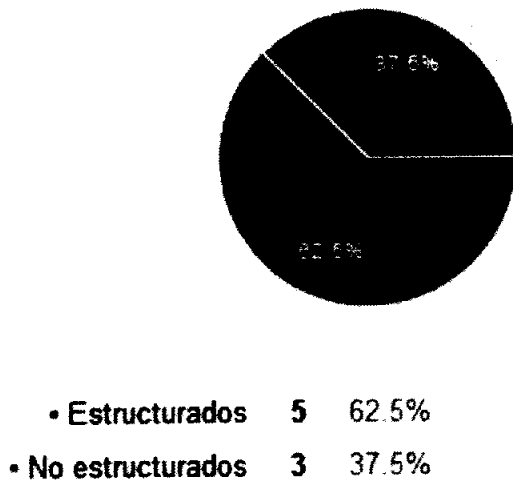
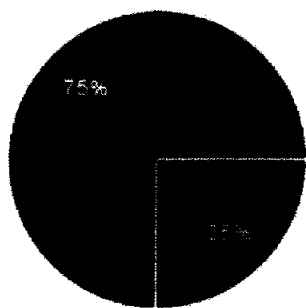


Figura 8: Tipos de problemas que se presentan en MULTICOMERCIAL S.A. CECA

¿Qué tipo de decisiones enfrenta con mayor frecuencia?

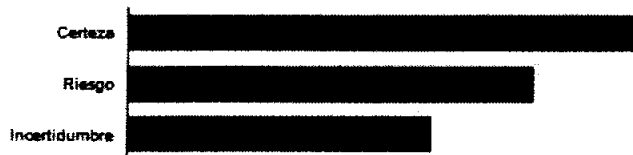


- Programadas 2 25%
- No programadas 6 75%



Figura 10: Tipo de decisiones que enfrenta MULTICOMERCIAL S.A. CECA

¿Cuáles son las condiciones más frecuentes del mercado en el que usted toma decisiones?



- Certeza 5 62.5%
- Riesgo 4 50%
- Incetidumbre 3 37.5%

Figura 11: Condiciones frecuentes del mercado para la toma de decisiones MULTICOMERCIAL S.A. CECA



¿Conoce el termino 'Minería de datos'?

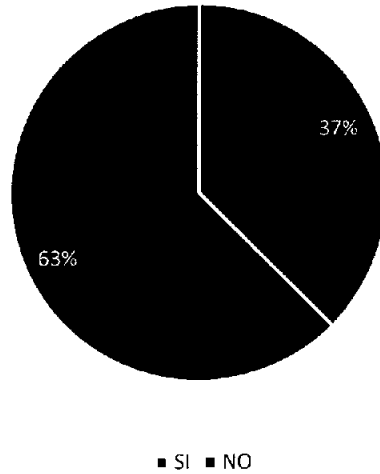


Figura 12: Grafico término 'minería de datos' MULTICOMERCIAL S.A. CECA

¿Conoce soluciones de minería de datos para tomar decisiones?

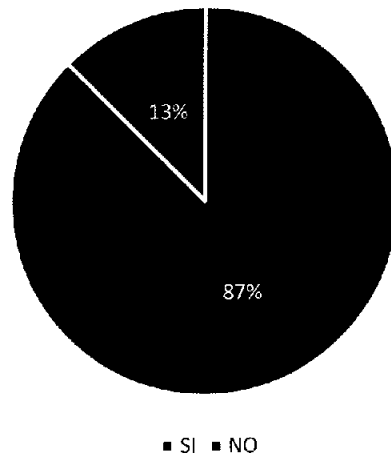


Figura 13: Grafico soluciones de minería de datos MULTICOMERCIAL S.A. CECA

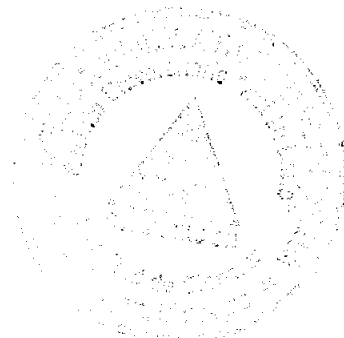
10.2 Creación de Datawarehouse (query)

```
CREATE TABLE Dim_Clientes(  
  Id_Cliente int IDENTITY(1,1) NOT NULL PRIMARY KEY,  
  IdAlt_Cliente nvarchar(50) NOT NULL,  
  Nombre_Cliente nvarchar(255) NULL,  
  Telefono1 nvarchar(10) NULL,  
  Telefono2 nvarchar(10) NULL,  
  Email nvarchar(255) NULL,  
  fechaIngreso date NULL,  
  addr nvarchar(255) NULL,  
  Fecha_ult_mov date NULL,  
  activo nvarchar(2) NULL,  
  SEXO nvarchar(2) NULL,  
  clasif_cliente nvarchar(50) NULL,  
  [ciudad] [nvarchar](50) NULL,  
  [stransporte] [nvarchar](1) NULL  
)
```

```
CREATE TABLE Dim_Productos(  
  Id_Producto int IDENTITY(1,1) NOT NULL PRIMARY KEY,  
  IdAlt_Producto nvarchar(50) NOT NULL,  
  Nombre_Producto nvarchar(255) NULL,  
  clasificacion nvarchar(255) NULL,  
  cantidad_disponible float NULL,  
  precio float NULL,  
  EstadoArticulo nvarchar(8) NULL  
)
```

```
CREATE TABLE Dim_Tiempo(  
  Id_Fecha int IDENTITY(1,1) NOT NULL PRIMARY KEY,  
  FechaFactura date NOT NULL,  
  AnoFactura nvarchar(255) NULL,  
  MesFactura nvarchar(255) NULL,  
  DiaFactura nvarchar(255) NULL,  
  SemanaFactura nvarchar(255) NULL,  
  TrimestreFactura nvarchar(255) NULL  
)
```

```
CREATE TABLE Dim_Vendedores(  
  Id_Vendedor int IDENTITY(1,1) NOT NULL PRIMARY KEY,  
  IdAlt_Vendedor nvarchar(10) NOT NULL,  
  Nombre_Vendedor nvarchar(255) NULL,  
  Nombre_tienda nvarchar(255) NULL  
)
```



```

CREATE TABLE Fact_Temp(
    Factura nvarchar(50) NOT NULL,
    FechaFactura date NOT NULL,
    Id_Vendedor nvarchar(10) NOT NULL,
    Id_Cliente nvarchar(50) NOT NULL,
    Id_Producto nvarchar(50) NOT NULL,
    Cantidad float NOT NULL,
    TotalMercaderiaDolar float NULL,
    TotalImpuesto1Dolar float NULL,
    TotalFacturaDolar float NULL,
    DescuentoTotalDolar float NOT NULL,
    TotalArticuloSinDscto float NULL,
    TotalArticuloConDscto float NULL
)

```

```

CREATE TABLE Fact_Ventas(
    [NoOrden] [int] IDENTITY(1,1) NOT NULL,
    [Id_Fecha] [int] NOT NULL,
    [Id_Vendedor] [int] NOT NULL,
    [Id_Cliente] [int] NOT NULL,
    [Id_Producto] [int] NOT NULL,
    [Factura] [nvarchar](50) NOT NULL,
    [Cantidad] [float] NOT NULL,
    [TotalMercaderiaDolar] [float] NULL,
    [TotalImpuesto1Dolar] [float] NULL,
    [TotalFacturaDolar] [float] NULL,
    [DescuentoTotalDolar] [float] NOT NULL,
    [TotalArticuloSinDscto] [float] NULL,
    [TotalArticuloConDscto] [float] NULL
)

```

```

ALTER TABLE Fact_Ventas
ADD CONSTRAINT pk_FV PRIMARY KEY
([NoOrden],[Id_Fecha],[Id_Vendedor],[Id_Cliente],[Id_Producto],[Factura])

```

```

ALTER TABLE Fact_Ventas
ADD CONSTRAINT fk_CF
FOREIGN KEY (Id_Cliente)
REFERENCES dim_clientes(Id_Cliente)

```

```

ALTER TABLE Fact_Ventas
ADD CONSTRAINT fk_VF
FOREIGN KEY ([id_Vendedor])
REFERENCES dim_vendedores([id_Vendedor])

```

```

ALTER TABLE Fact_Ventas
ADD CONSTRAINT fk_FF
FOREIGN KEY (Id_fecha)
REFERENCES dim_Tiempo(Id_fecha)

```

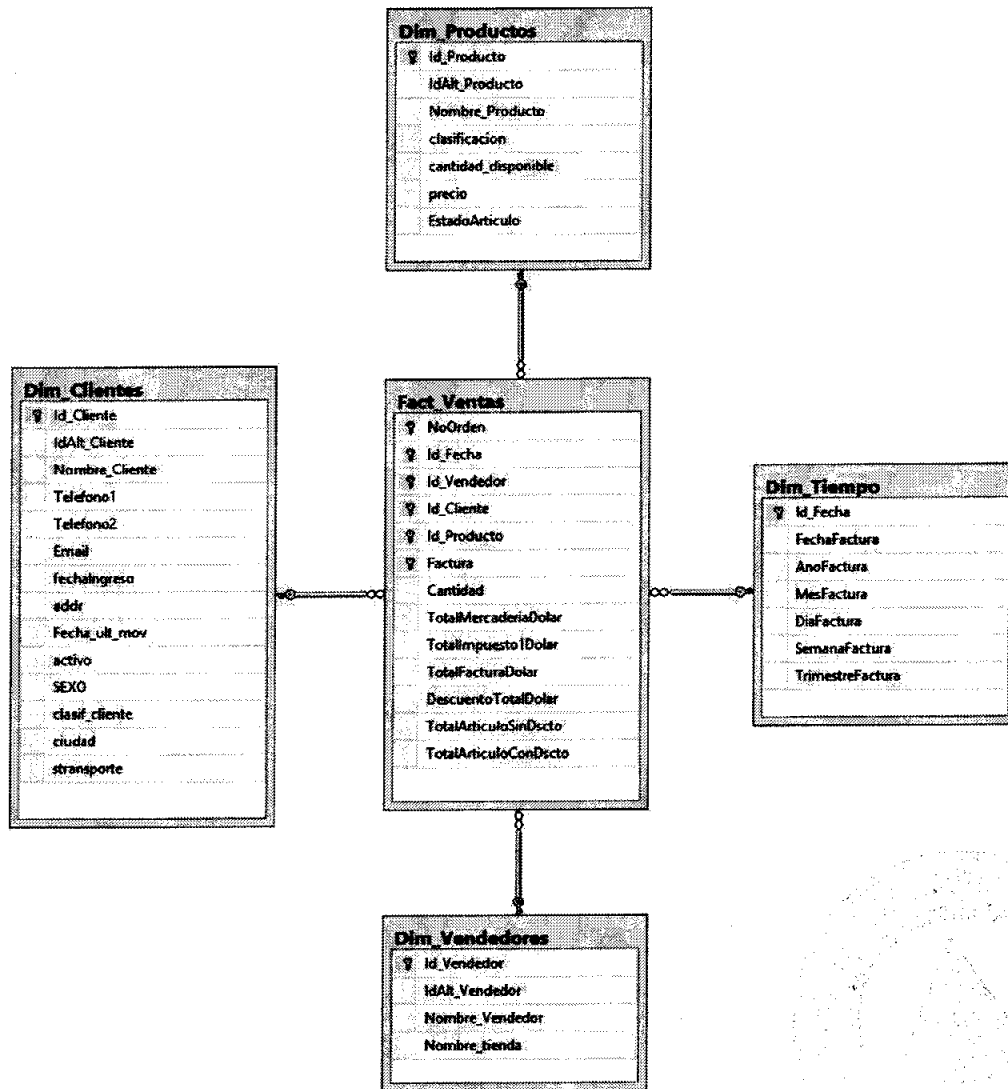
```

ALTER TABLE Fact_Ventas
ADD CONSTRAINT fk_PF
FOREIGN KEY (Id_Producto)
REFERENCES dim_productos(Id_Producto)

```


10.3 Diagrama Base de datos (DataWarehouse)

Figura 14: Diagrama Base de datos (DataWarehouse)



10.4 SQL Server Integration Services (SSIS): Integración y transformación de los datos.

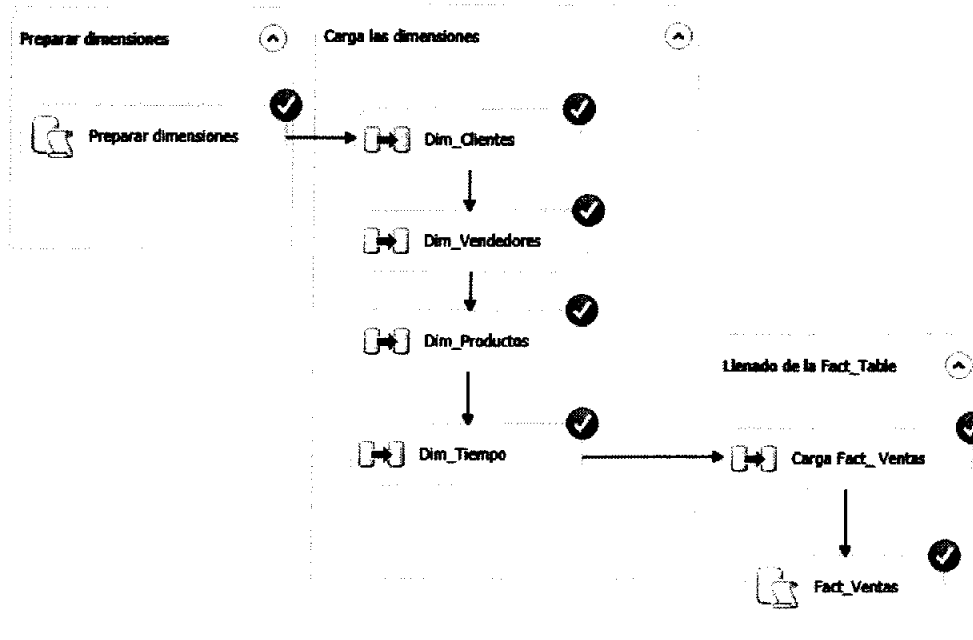


Figura 15: SISS - Integración y transformación de los datos

10.4.1 Flujo de datos - Dimensión clientes



Figura 16: Flujo de datos - Dimensión Clientes

10.4.2 Flujo de datos - Dimensión Vendedores

Dim_Vendedores



Figura 17: Flujo de datos - Dimensión Vendedores

10.4.3 Flujo de datos - Dimensión Tiempo

Dim_Tiempo



Figura 18: Flujo de datos - Dimensión Tiempo

10.4.4 Flujo de datos - Dimensión Productos

Dim_Productos



Figura 19: Flujo de datos - Dimensión Productos

10.4.5 Flujo de datos - Fact Table Temporal

Fact_Temp

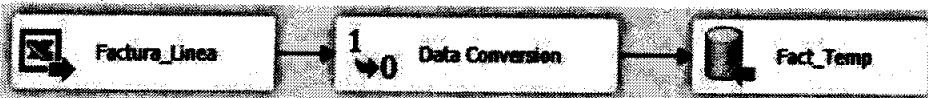


Figura 20: Flujo de datos - Fact Table Temporal

10.5 Creación de cubo

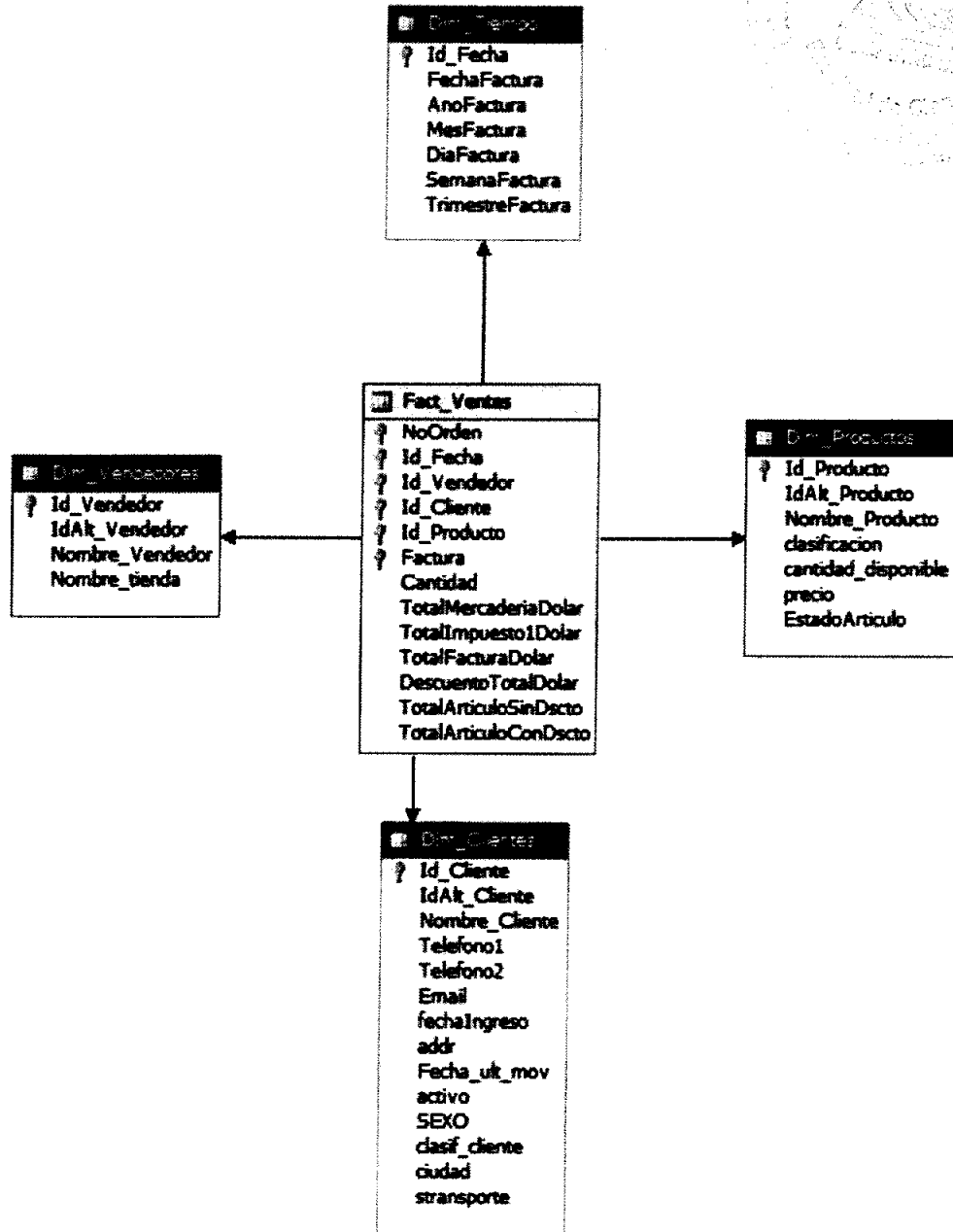


Figura 21: Cubo Analysis Services

10.6 Exploración y Modelado Algoritmo Naive Bayes

10.6.1 ¿Cuáles son los posibles compradores de audio profesional?

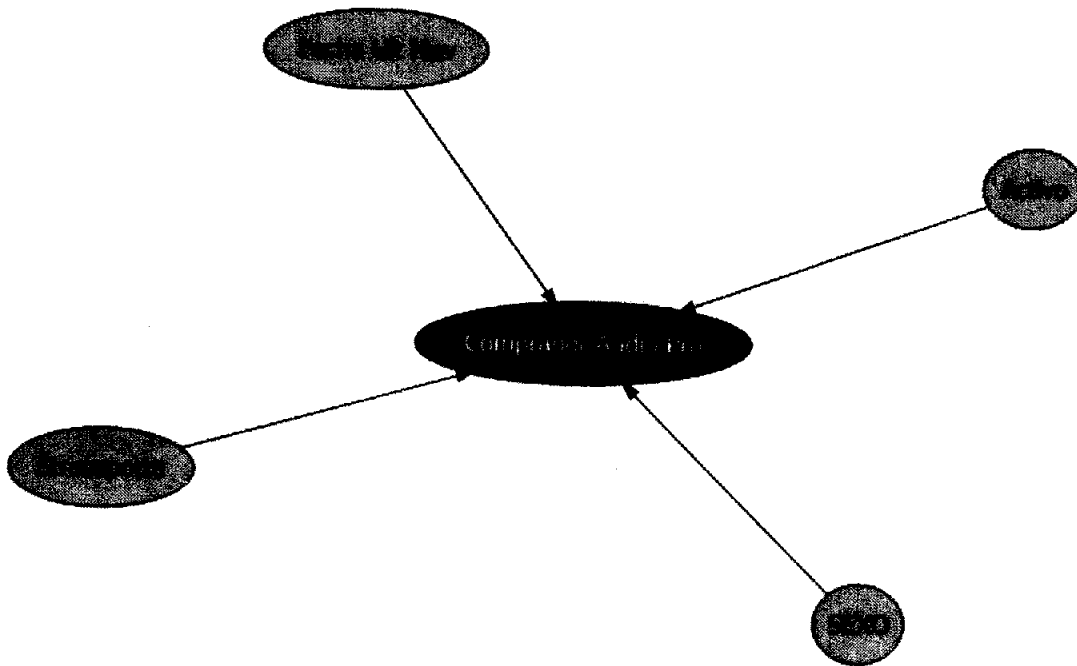


Figura 22: red de dependencia - compradores audio profesional

Attributes	States	Populatio... Size: 301	0 - 1 Size: 301	missing Size: 0
Activo	<ul style="list-style-type: none"> ● S ● Missing 			
Fecha Ult Mov	<ul style="list-style-type: none"> ● 14/03/2013 21: ● >= 01/07/2015 ● < 30/05/2008 6 ● 30/05/2008 6:0 ● Other ... 			
SEXO	<ul style="list-style-type: none"> ● NA ● M ● F ● Missing 			
Transporte	<ul style="list-style-type: none"> ● 1 ● 0 ● Missing 			

Figura 23: Perfil de atributos - compradores audio profesional

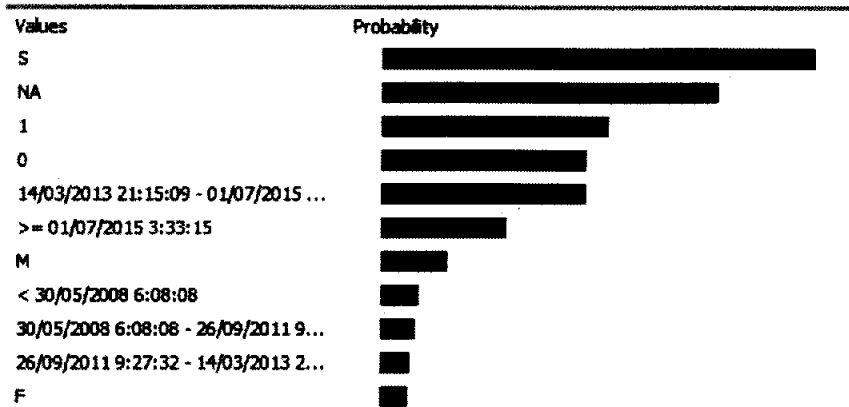


Figura 25: Característica atributos - compradores audio profesional

10.6.2 ¿Cuáles son los posibles compradores de tecnología?

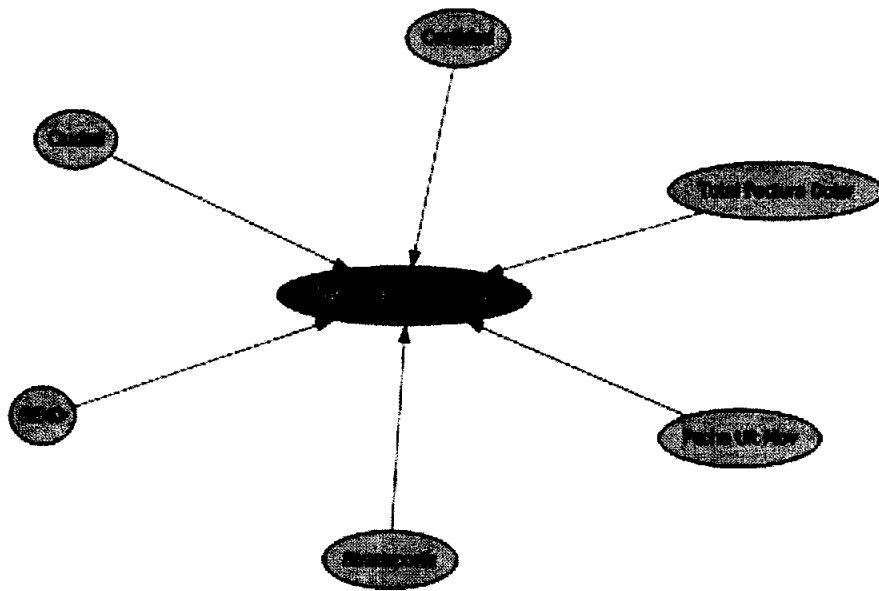


Figura 26: red de dependencia - compradores tecnología

Attributes	States	Populatio... Size: 39	0 - 1 Size: 39	missing Size: 0
Cantidad	● < 8,7537448752	■	■	■
	● >= 324,5	■	■	■
	● 106,753445952	■	■	■
	● Missing	■	■	■
	● Other	■	■	■
Ciudad	● Managua	■	■	■
	● Granada	■	■	■
	● Masaya	■	■	■
	● Boaco	■	■	■
	● Other	■	■	■
Fecha Ult Mov	● >= 07/10/2014	■	■	■
	● 06/01/2013 8:3	■	■	■
	● 12/12/2010 15:	■	■	■
	● < 28/10/2007 7	■	■	■
	● Other	■	■	■
SEXO	● NA	■	■	■
	● M	■	■	■
	● Missing	■	■	■
	● F	■	■	■
Transporte	● 0	■	■	■
	● 1	■	■	■
	● Missing	■	■	■



Figura 27: Perfil de atributos - compradores Tecnología

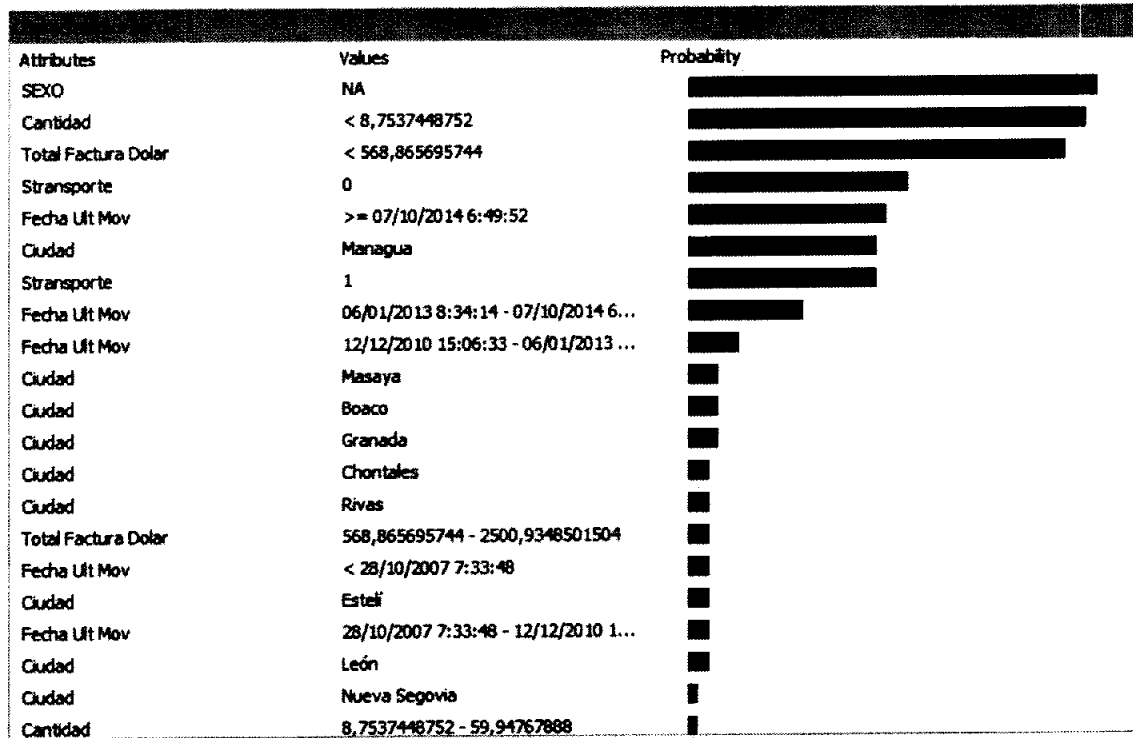


Figura 28: Característica atributos - compradores Tecnología

10.7 Evaluación de Calidad Norma ISO/IEC 9126-1 -
 característica de usabilidad

10.7.1 Capacidad para ser entendido

Tabla 3: ISO9126-1 Usabilidad-capacidad para ser entendido

USABILIDAD	Puntuación				
	1	2	3	4	5
Capacidad para ser entendido	Totamente en desacuerdo	En desacuerdo	Indiferente	De acuerdo	Totamente de acuerdo
¿El software utiliza una terminología adecuada con respecto a las técnicas de minería de datos?			2	6	
¿El lenguaje utilizado es accesible y fácil de leer?		1		7	
¿El software contiene algún módulo para la visualización gráfica de los resultados?				8	
¿Los mensajes emitidos son fáciles de entender?		3	2	3	
El mensaje de error propone una solución		7		1	
¿Los tipos y tamaños de letra son legibles y apropiados?				8	
El tamaño de las imágenes e iconos es adecuado para saber qué representan				8	
¿El lenguaje está más cercano al utilizado por el usuario que al informático o técnico?		2	1	3	
TOTALES	0	26	15	184	0

10.7.2 Capacidad para ser aprendido

Tabla 4: ISO9126-1 Usabilidad-capacidad para ser aprendido

USABILIDAD	Puntuación				
	1	2	3	4	5
Capacidad para ser aprendido	Totamente en desacuerdo	En desacuerdo	Indiferente	De acuerdo	Totamente de acuerdo
Facilidad para entender los cambios producidos en las operaciones		4	2	2	
Facilidad para entender que datos ingresar y facilidades proporcionados		1	1	0	
Desde cualquier punto un usuario puede salir de la herramienta				8	
En caso que el proceso requiera varios pasos es posible volver a los anteriores para modificar los datos		4	1	3	
La terminología es constante en toda la herramienta		2	1	5	
El mismo elemento aparece igual en toda la herramienta		1	1	0	
Se dan indicaciones para completar campos problemáticos		8			
Se muestran fácilmente las figuras, las tablas, los comandos claves y el tipo de acción que se debe realizar		1	2	4	
Si una tarea tiene opciones por defecto, están a disposición del usuario		3		5	
La herramienta no requiere volver a escribir la información solicitada anteriormente				8	
La herramienta soporta el aprendizaje colaborativo			1	7	
TOTALES	0	48	27	220	0

10.7.3 Capacidad para ser operado

Tabla 5: ISO9126-1 Usabilidad-capacidad para ser operado

USABILIDAD	Puntuación				
	1	2	3	4	5
	Totalmente de acuerdo	En desacuerdo	Indiferente	De acuerdo	Totalmente de acuerdo
¿El software utiliza las operaciones OLAP, para realizar las técnicas de minería de datos?			7	1	
¿El software tiene varias opciones para acceder a los datos (p.ej. archivos tipo texto, archivos de bases de datos Access, DBASE o...)?		3	3	2	
¿El usuario entiende fácilmente los mensajes enviados por el software?		1	2	5	
¿El software muestra el estado del proceso de minería de datos en uso?					8
¿Puede el usuario restablecer o corregir sus errores y recuperar la tarea de minería de datos?.			2	6	
¿Puede el usuario fácilmente completar una tarea de minería de datos satisfactoriamente?		1	3	4	
¿Puede el usuario retroceder en pasos, dentro de una tarea determinada y corregir algún error o dato para terminar la tarea?			1	7	
¿Existe una metodología establecida por el software para cada una de las técnicas de minería de datos que el software contiene?				8	
¿El software tiene varias opciones de técnicas de minería de datos?					8
La información proporcionada por las ayudas en línea, es útil en el momento que es requerida.		3	1	4	
¿La selección de opciones del software, se realiza por medio de la barra de herramientas o ratón (mouse)?					8
TOTALES	0	16	57	148	120

10.7.4 Capacidad de atracción

Tabla 6: ISO9126-1 Usabilidad-capacidad de atracción

USABILIDAD	Puntuación				
	1	2	3	4	5
Capacidad de atracción	Totamente en desacuerdo	En desacuerdo	Indiferente	De acuerdo	Totamente de acuerdo
Los títulos son descriptivos y distintivos				5	3
No despiega excesiva información en la interfaz de la herramienta		2	1	5	
Se utilizan las ventanas emergentes para mostrar información importante	2	6			
En los menús desplegables, los elementos de los listas son suficientemente explicativos y se visualizan todos a la vez			2	5	1
La apariencia de la herramienta es agradable y sencilla				7	1
TOTALES		16	9	88	25

10.7.5 Impacto evaluación ISO9126-1 característica de Usabilidad

Tabla 7: Impacto evaluación ISO9126-1 (USABILIDAD)

USABILIDAD	PUNTUACIÓN				
	1	2	3	4	5
	Totalmente en desacuerdo	En desacuerdo	Indiferente	De acuerdo	Totalmente de acuerdo
Capacidad para ser entendido	0	26	15	184	0
Capacidad para ser aprendido	0	48	27	220	0
Capacidad para ser operado	0	16	57	148	120
Capacidad de atracción	2	16	9	88	25

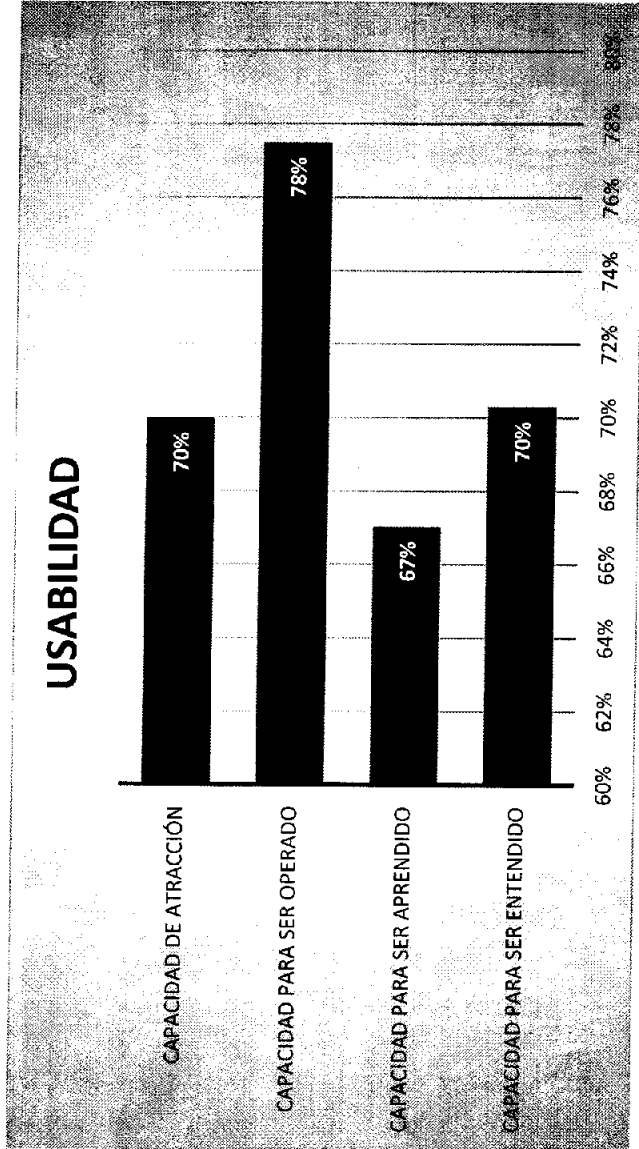


Figura 29: Grafico de Impacto de ISO9126-1 Usabilidad

XI. CRONOGRAMA DEL PROYECTO

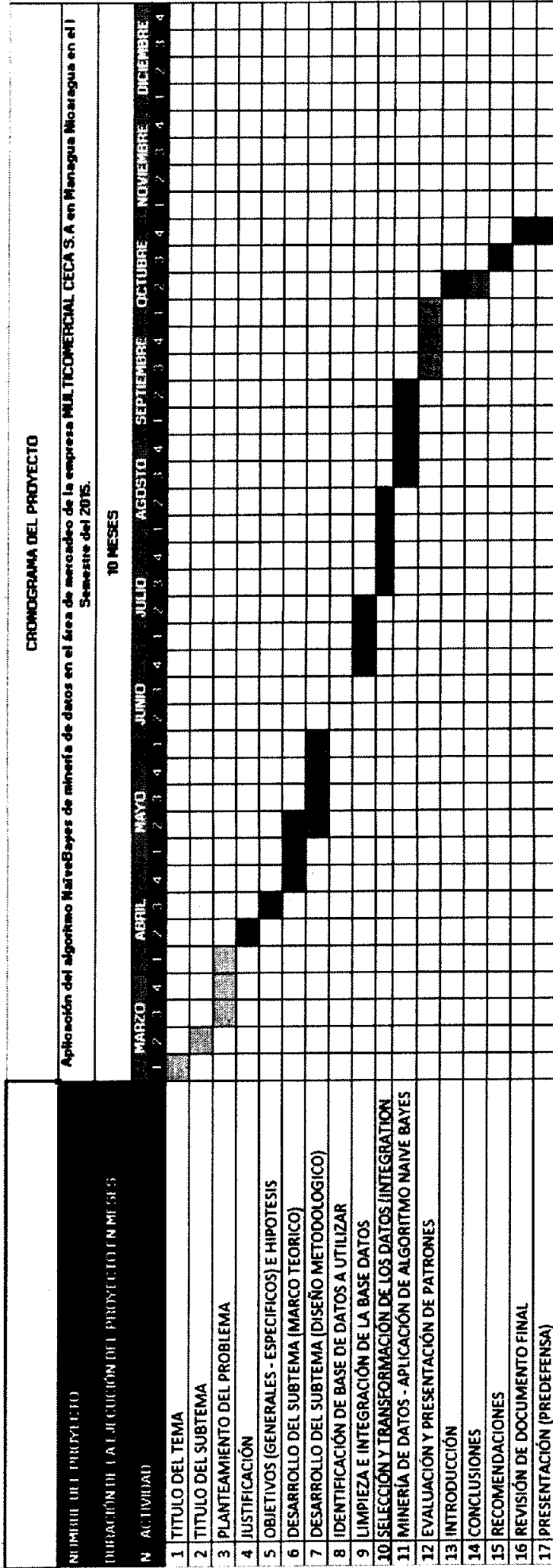


Figura 30: Cronograma del Proyecto de Investigación

XII. PRESUPUESTO

Tabla 8: Presupuesto de la Investigación

PRESUPUESTO INVESTIGACIÓN	UNIDADES	MENSUAL	TIEMPO	TOTAL
COSTO PERSONAL DE TRABAJO				
DESARROLLADOR (DATAMINER)	1	\$ 225.00	10 MESES	\$ 2,250.00
COSTOS OPERACIONALES				
MATERIALES	-	\$ 15.00	2 MESES	\$ 30.00
TRANSPORTE	-	\$ 24.00	10 MESES	\$ 240.00
COSTOS BÁSICOS FIJOS				
INTERNET	0	\$ 35.00	10 MESES	\$ 350.00
ELECTRICIDAD	0	\$ 25.00	10 MESES	\$ 250.00
TELÉFONO	1	\$ 36.00	10 MESES	\$ 360.00
AGUA	0	\$ 10.00	10 MESES	\$ 100.00
ARRIENDO DE OFICINA	1	\$ 100.00	10 MESES	\$ 1,000.00
COMPUTADORES				
DELL LATITUDE E5420	1	\$ 20.00	10 MESES	\$ 200.00
SAMSUNG RV54	1	\$ 20.00	10 MESES	\$ 200.00
COSTOS SOFTWARES				
SQL SERVER 2012	1	-	-	\$ 898.00
MICROSOFT OFFICE 2013	1	\$ 8.25	12 MESES	\$ 99.00
VALOR PROYECTO				
ETAPA DE ANALISIS	-	\$ 45.00	3 MESES	\$ 135.00
ETAPA DE DESARROLLO	-	\$ 50.00	4 MESES	\$ 200.00
ETAPA DE CONCLUSIONES Y RESULTADOS	-	\$ 30.00	3 MESES	\$ 90.00
TOTALES		\$ 643.25		\$ 6,402.00

XIII. CONCLUSIONES

- Como resultado de la investigación presentada, es posible resaltar la importancia que existe en la aplicación de minería de datos dentro de la empresa MULTICOMERCIAL S.A. CECA
- La implementación de minería de datos como una herramienta de decisión estratégica comercial va a permitir crear publicidad personalizada según los resultados obtenidos del algoritmo aplicado.
- La resultados de los reportes obtenidos por la implementación de minería de datos dentro del área de mercadeo del a empresa MULTICOMERCIAL S.A. CECA facilitarán la rápida toma de decisión.
- Se valoró la calidad de la solución de minería de datos por medio de la Norma ISO/IEC 9126-1 dando como resultado un valor positivo el cual permite ser entendido, aprendido y operado por los usuarios finales.

XIV. RECOMENDACIONES

- En este estudio se implementó el algoritmo Naive Bayes, en el futuro se podría implementar otro algoritmo, por ejemplo el algoritmo de Clustering ó el algoritmo de árboles de decisión.
- Tomar las acciones correspondientes al análisis obtenido de los datos para el mejoramiento en la rotación del inventario por parte de la gerencia de MULTICOMERCIAL S.A. CECA
- Establecer acciones o normas para que los jefes de cada área involucrada en la toma de decisiones se comprometan y responsabilicen a establecer estrategias relacionadas al mercadeo de la empresa.
- Continuar la evaluación de calidad de la soluciones de minería de datos que se vayan aplicar a futuro por medio de la norma ISO/IEC 9126-1.



XV. ANEXOS

CECA MULTICOMERCIAL, S.A.

Departamento de Informática

Reportes de Ventas

- > Reporte de Ventas por Clasificación
- > Reporte de Auditoría de Descuentos
- > Precios Vs Precio de Factura



NO ANDE DE LA CECA A LA MECA

Figura 31: Anexo - pantalla reportes de ventas MULTICOMERCIAL S.A. CECA



XVI. BIBLIOGRAFÍA

- Alfred, V. (1988). *Estructura de datos y Algoritmos*. Mexico: Addison-Wesley.
- Bernal, C. A. (2010). Metodología de la investigación. En C. A. Bernal, *Metodología de la investigación: administración, economía, humanidades y ciencias sociales* (Tercera edición ed., pág. 320). Colombia: PEARSON EDUCACIÓN.
- CHIAVENATO, I. (2006). *Introducción a la teoría general de la administración*. México: McGrawHill.
- ELMASRI, R. (2007). *Fundamentos de Sistemas de Bases de Datos*. Madrid: Pearson.
- Española, R. A. (s.f.). *Real Academia Española*. Recuperado el 26 de abril de 2015, de <http://www.rae.es/>
- FLORES, J. G. (1994). *Análisis de datos cualitativos, aplicación a la investigación educativa*. Barcelona: Primera edición.
- Gabillau, J. (s.f.). *SQL Server 2012*.
- Gillenson, M. L. (1988). *Introducción a la base de datos*. México: McGRAW-HILL.
- Google. (2015). *Data Mining*. Obtenido de Research at Google: <http://research.google.com/pubs/DataMining.html>
- Hotek, M. (2009). *Microsoft SQL Server 2008*. Estados Unidos de America.
- Itson, B. (2012). *Introduccion a Sistemas*. Mexico.

J.Vallejos, S. (2006). *Diseño y Administracion de datos.*
Argentinian.

Jaime MacLennan, Z. T. (2009). *Mineria de Datos con SQL
Server 2008.* Indianapolis.

KENDALL, K. E. (2011). *Análisis y Diseño de Sistemas* (Octava
Edición ed.). México: PEARSON EDUCACION.

KENDALL, K. Y. (2005). *Análisis y Diseño de Sistemas.* En K.
Y. KENDALL, *Análisis y Diseño de Sistemas.* México:
PEARSON EDUCACIÓN.

M.Ricardo, C. (2009). *Bases de Datos.* Mexico: Mcgraw-Hill
Interamericana Editores S.A.

Mannino, M. v. (2007). *Administracion de bases de Datos.*
Mexico: McGraw-Hill/Interamericana Editores, S.A.

Microsoft. (2015). *Algoritmo Naive Bayes* . Obtenido de
Microsoft-Developer Network:
<https://msdn.microsoft.com/es-es/library/ms174806.aspx>

Microsoft. (2015). *Analysis Services.* Obtenido de Microsoft-
Developer Network: [https://msdn.microsoft.com/es-
es/library/bb522607%28v=sql.120%29.aspx](https://msdn.microsoft.com/es-es/library/bb522607%28v=sql.120%29.aspx)

Microsoft. (2015). *Conceptos de minería de datos.* Obtenido de
Microsoft-Developer Network:
[https://msdn.microsoft.com/es-
es/library/ms174949%28v=sql.120%29.aspx](https://msdn.microsoft.com/es-es/library/ms174949%28v=sql.120%29.aspx)

Microsoft. (2015). *Sql Server.* Obtenido de Microsoft-
Developer Network: [https://msdn.microsoft.com/es-
es/library/bb545450.aspx](https://msdn.microsoft.com/es-es/library/bb545450.aspx)

O.C. FERRELL, G. A. (s.f.). *Introducción a los negocios en un mundo cambiante*. México: McGrawHill.

Paola Garcia, C. A. (s.f.). *Minería de Datos aplicada a las redes sociales*. Universidad Carlos III de Madrid, Madrid, España. Obtenido de <http://www.it.uc3m.es/jvillena/irc/practicass/08-09/08.pdf>

Sampieri, F. (2006). *Metodología de la investigación*. Mexico: McGraw-Hill Interamericana.

SENN, J. (1992). *Análisis y diseño de sistemas de información*. México: McGraw Hill.

SEQUEIRA V. Y CRUZ, A. (1997). *Investigar es fácil* (Segunda ed.). El Amanecer.

TOFFLER, A. Y. (2006). *La revolución de la riqueza*. México.

TSAI, A. Y. (1990). *Sistemas de Base de datos: Administración y uso*. Naucalan de Juárez, Edo. de México: PRENTICE-HALL HISPANOAMERICANA, S.A.

Zahonero, L. J. (2007). *Estructura de Datos en C++*. Madrid: Mc Graw Hill.