

UNIVERSIDAD NACIONAL AUTÓNOMA DE NICARAGUA, MANAGUA
UNAN-MANAGUA

Recinto Universitario "Rubén Darío"
Facultad de Ciencias e Ingenierías
Departamento de Computación

Biblioteca Central "Salomón de la Selva"
UNAN-Managua
Fecha de Ingreso: 23/01/17
Comprado: Don. Dpt. Comp.
Precio: C\$ U\$ 46223
Registro No.



Minería de datos

Desarrollo de un modelo basado en minería de datos aplicando el algoritmo de Clúster en el área de Mercadeo y Ventas de la empresa DISPROCEL S.A (Distribuidora de Productos móviles) en el II semestre del año 2015.

Autores:

- ✚ Br. Marbell Del Socorro Acuña Gaitán
- ✚ Br. Leila Tatiana Palma.
- ✚ Br. Julio Cesar Huete

Tutor

MSc. Luis Miguel Martínez Olivera.

19 de Noviembre del 2015

4-20
COMP
378.242
ACU
2015

AGRADECIMIENTOS

Agradecemos:

A Dios por permitirnos concluir esta etapa de formación profesional, ya que sin su bondad y ayuda no podríamos cumplir nuestras metas.

A nuestros amigos que siempre nos motivaron a seguir adelante y nos brindaron su apoyo en el momento que más lo necesitábamos.

El agradecimiento es una de las cualidades más ilustres de la persona, con tal razón, agradecemos a nuestro tutor el MSc. Luis Miguel Martínez por brindarnos su paciencia, apoyo y conocimiento como también a todos los profesores del departamento de computación que fueron parte de nuestra formación académica lo cual nos han otorgado una gran ayuda en muchos momentos difíciles de la vida.

Autores

DEDICATORIA

A Dios por permitirme culminar esta fase de aprendizaje y no abandonarme en ningún momento.

A mis padres Esperanza del Socorro Gaitán Cano y Terencio de Jesús Acuña López por ser fuente de mi inspiración y apoyo constante en cada etapa de mi vida, quienes con su amor, paciencia y comprensión me han ayudado para poder cumplir con esta meta.

Que este esfuerzo sirva de ejemplo para las demás generaciones que vendrán después de mí, para que ellos cumplan con cabalidad los objetivos que se propongan.

Br. Marbell del Socorro Acuña Gaitán

A Dios por permitirme llegar a este momento, por darme la fuerza y la fortaleza cada día para poder culminar esta meta.

A mis padres Cándida Vásquez y Francisco Palma por su amor y apoyo incondicional en cada etapa de mi vida y en todos estos años de estudios.

A mi esposo por ser mi ayuda idónea en cada momento que lo necesitaba.

A mis hermanos y también amigos que me brindaron su ayuda y me motivaron a seguir adelante y lograr mis objetivos.

A mi pastor Julio Cesar Soza por ser mi guía espiritual y brindarme sus consejos cuando más los necesitaba.

Br. Leila Tatiana Palma Vásquez.

*“La paciencia es un árbol de raíces amargas,
pero de frutos muy dulces”*

Proverbio persa.

Se la dedico a mi familia por quienes soy lo que soy, por su apoyo, comprensión, consejos y ayuda en los momentos cumbres y por brindarme su solidaridad y recursos necesarios para estudiar.

A mi madre, a quien debo mis valores, mis principios, mi carácter, mi empeño, mi perseverancia y mi coraje para conseguir mis objetivos en la vida.

A todas aquellas personas que han estado presentes durante mis años de estudios y han sabido guiarme por el camino del éxito y darme las fuerzas necesarias para no ceder ante los problemas que se presentaban, enseñándome a encarar las adversidades sin desfallecer en el intento.

Br. Julio Cesar Huete.

INDICE

I. RESUMEN	5
II. INTRODUCCION.....	6
III. ANTECEDENTES.....	7
IV. PLANTEAMIENTO DEL PROBLEMA	8
1. DELIMITACION DEL PROBLEMA	8
2. FORMULACION DEL PROBLEMA.....	8
3. SISTEMATIZACION DEL PROBLEMA.....	8
V. JUSTIFICACION.....	9
VI. OBJETIVOS.....	10
OBJETIVO GENERAL.....	10
OBJETIVOS ESPECIFICOS	10
VII. MARCO TEORICO.....	11
1. DATOS	11
2. INFORMACION	11
3. CONOCIMIENTO.....	12
4. SISTEMAS DE INFORMACION.....	12
4.1 SISTEMAS.....	12
4.2 CARACTERISTICAS DE LOS SISTEMAS DE INFORMACION	14
4.4 BASES DE DATOS RELACIONALES	14
5.5 MODELO RELACIONAL	15
5. MINERIA DE DATOS.....	16
5.1 TECNICA DE MINERIA DE DATOS	16
5.2 CARACTERISTICAS DE LA MINERIA DE DATOS.....	18
5.3 ALMACEN DE DATOS.....	18
5.4 ESTRUCTURA DE LA MINERIA DE DATOS	20
5.5 MODELO DE MINERIA DE DATOS	22
5.6 PROPIEDADES DEL MODELO DE MINERIA DE DATOS	26
6. HERRAMIENTAS DE LA MINERIA DE DATOS.....	26
6.1 HERRAMIENTAS DE SQL SERVER.....	27

6.2 ALGORITMO DE MINERIA DE DATOS	28
6.3 EMPRESA DISPROCEL S.A	32
VIII. HIPOTESIS.....	37
IX. DISEÑO METODOLOGICO	38
1.1 TIPO DE ESTUDIO	38
1.2 UNIVERSO Y MUESTRA.....	38
1.3 METODO DE RECOLECCION DE LA INFORMACION	38
1.4 ESTUDIO DE FACTIBILIDAD	39
X. FASES DEL DESARROLLO DE MINERIA DE DATOS.....	40
XI. HERRAMIENTAS QUE UTILIZARON PARA REALIZAR EL MODELO.....	41
XII. ANALISIS DE RESULTADO	42
XIII. PRESUPUESTO	75
XIV. CRONOGRAMA DE ACTIVIDADES	76
XV. CONCLUSIONES	77
XVI. RECOMENDACIONES.....	78
XVII. BIBLIOGRAFIA.....	79
XVIII. COMPENDIOS	80

INDICE DE FIGURA

FIGURA 1. SISTEMAS DE INFORMACIÓN	13
FIGURA 2. ESTRUCTURA DE UNA BASE DE DATOS	15
FIGURA 3. CLASIFICACIÓN MINERÍAS DE DATOS	17
FIGURA 4. MINERIAS DE DATOS	23
FIGURA 5. ALGORITMO DE CLUSTER	28
FIGURA 6. DIAGRAMA DE LA EMPRESA DISPROCEL	36
FIGURA 7. PROCESO DE MINERIA	41
FIGURA 8. PROCESO DE EXTRACCION DE DATOS	43
FIGURA 9. BASE DE DATOS TRANSACCIONAL	44
FIGURA 10. DATA WAREHOUSE	45
FIGURA 11. FLUJO DE DATOS	50
FIGURA 12. PROCESO FINALIZADO DIMCLIENTE	50
FIGURA 13. FLUJO DE DATOS	50
FIGURA 14. PROCESO FINALIZADO DIMEMPLEADO	50
FIGURA 15. FLUJO DE DATOS	51
FIGURA 16. PROCESO FINALIZADO DIMPRODUCTO	51
FIGURA 17. FLUJO DE DATOS	51
FIGURA 18. PROCESO FINALIZADO DIMTIEMPO	51
FIGURA 19. FLUJO DE DATOS	52
FIGURA 20. PROCESO FINALIZADO DIMITIPONEGOCIO	52
FIGURA 21. FLUJO DE DATOS	52
FIGURA 22. PROCESO FINALIZADO DIMZONA	52
FIGURA 23. FLUJO DE DATOS	53
FIGURA 24. PROCESO FINALIZADO FTVENTAS	53
FIGURA 25. PROCESO ETL COMPLETO	54
FIGURA 26. DIAGRAMA ESTRUCTURA DE MINERIA DE DATOS	55
FIGURA 27. DIAGRAMA DE CLUSTER	57
FIGURA 28. PERFILES DEL CLUSTER	58
FIGURA 29. CARACTERISTICAS PARA POBLACION	58
FIGURA 30. VARIABLES DE PRODUCTO1	60
FIGURA 31. VARIABLES DE PRODUCTO2	60
FIGURA 32. VARIABLES DE PRODUCTO3	60
FIGURA 33. CLUSTER 2	61
FIGURA 34. CLUSTER 7	61
FIGURA 35. CLUSTER PULPERIAS	63
FIGURA 36. CLUSTER TIENDA	63
FIGURA 37. CLUSTER FARMACIA	64
FIGURA 38. CLUSTER MISCELANEA	65
FIGURA 39. VENTAS PARTICULARES	66
FIGURA 40. VENTAS	67
FIGURA 41. OTROS TIPOS DE NEGOCIOS	67
FIGURA 42. MEJORES CLIENTES	71
FIGURA 43. GRÁFICO DE EVALUACIÓN FINAL DEL USUARIO	74



INDICE DE TABLA

TABLA 1. CANTIDAD PRODUCTOS VENDIDOS.....	66
TABLA 2. PRODUCTOS BENDIDOS SEGUN TIPO NEGOCIO	67
TABLA 3. EFECTIVIDAD	70
TABLA 4.PRODUCTIVIDAD	73
TABLA 5. SASTIFACCION	73
TABLA 6. PRESUPUESTO	73

I. RESUMEN

En el presente trabajo se muestra el análisis e implementación de una solución de minería de datos para la empresa DISPROCEL S.A.

Primeramente se analizaron los procesos que realiza la empresa mediante los métodos de recolección de la información, identificando las necesidades y problemas presentes con respecto al flujo de la información, lo que sirvió como base para la creación de la solución.

Después del análisis realizado anteriormente se planteó una justificación para en donde se determinó el propósito de este proyecto y que áreas vendría a solucionar con la implementación de esta solución de minería. Una vez ya planteada la justificación para este proyecto se detallan los objetivos planteados para este proyecto.

Se detallan una serie de conceptos, características y herramientas utilizadas o con el objetivo de obtener más conocimiento sobre el tema a implementarse en el proyecto.

Se describen los métodos de recolección de información utilizados así como también la toma de muestra que se tomó al momento de la entrevista.

Luego se muestran los resultados del análisis y el desarrollo de la solución de minería de datos, en donde se presentó la solución a cada objetivo anteriormente planteado. Como primer resultado se presenta el análisis sobre los procesos de las áreas de venta y mercadeo así como la base de datos transaccional que utiliza la empresa, en nuestro segundo resultado se presenta los diferentes proceso que se realizaron para llegar a la implementación del algoritmo de clúster y como nuestro tercer resultado se muestra la evaluación por parte del usuario final.

Finalmente se realizó una conclusión sobre el proyecto y se brindaron una serie de recomendaciones para el buen manejo de la solución.

II. INTRODUCCION

En los últimos años debido al desarrollo tecnológico a niveles exponenciales tanto en el área de cómputo como en la de transmisión de datos, ha sido posible que se gestionen de una mejor manera el manejo y almacenamiento de la información. Sin duda existen cuatro factores importantes como: El incremento de las velocidades de cómputo en los procesadores, las mejoras en la confiabilidad y aumento de la velocidad en la transmisión de datos, el desarrollo de sistemas administradores de bases de datos más poderosos.

Actualmente todas estas ventajas nos han llevado a abusar del almacenamiento de la información en las bases de datos. Podemos decir que las empresas almacenan datos pero no todas hacen uso de ellos. En cambio, si estos datos son leídos y analizados pueden proporcionar en conjunto un verdadero conocimiento que ayude en la toma de decisiones.

La minería de datos surge como una tecnología que intenta ayudar a comprender el contenido de una base de datos, de forma general, los datos son la materia prima bruta.

En este trabajo se aplicó minería de datos a la empresa DISPROCEL en donde se describen los pasos del proceso de extracción de conocimiento en bases de datos y como se pueden utilizar la explotación de minería de los datos, los patrones del comportamiento entre usuarios y personal de la empresa además comprender los patrones de uso de los recursos de información en la empresa.

DISPROCEL como una empresa que está al tanto de la tecnología reconoce que para brindar un servicio eficiente a los clientes se necesita un manejo oportuno de la información y así de esta manera proporcionaría un servicio de calidad efectivo. Por tanto, se ha propuesto modernizar su modelo extracción de la información aplicando el algoritmo de Clúster a dicha información.

III. ANTECEDENTES DEL PROGRAMA

Proyecto SKYCAT de la NASA.

El sistema SkyImageCataloguing and AnalysisTool (SKYCAT) se basa en técnicas de agrupación (clustering) y árboles de decisión para poder clasificar los objetos en estrellas, planetas, sistemas, galaxias, etc. con una alta confiabilidad. Los resultados han ayudado a los astrónomos a descubrir dieciséis nuevos quásar con corrimiento hacia el rojo que los incluye entre los objetos más lejanos del universo y, por consiguiente, más antiguos. Estos quásares son difíciles de encontrar y permiten saber más acerca de los orígenes del universo.

Proyecto en NBA

El Advanced Scout es un software que emplea técnicas de data mining y que han desarrollado investigadores de IBM para detectar patrones estadísticos y eventos raros. Tiene una interfaz gráfica muy amigable orientada a un objetivo muy específico: analizar el juego de los equipos de la National Basketball Association (NBA). El software utiliza todos los registros guardados de cada evento en cada juego: pases, encestes, rebotes y doble marcaje (doublé team) a un jugador por el equipo contrario, entre otros. El objetivo es ayudar a los entrenadores a aislar eventos que no detectan cuando observan el juego en vivo o en película.

Se realizó una investigación con el fin de encontrar en Nicaragua alguna empresa que haya implementado el algoritmo de clúster. La búsqueda fue infructuosa.

Actualmente con este proyecto se utilizara se utilizará la solución de minería de datos aplicando el algoritmo de Clúster en la empresa DISPROCEL en el área de Mercadeo y Ventas.

IV. PLANTEAMIENTO DEL PROBLEMA

DISPROCEL S.A es una empresa que brinda los servicios de venta de dispositivos móviles recarga electrónica y Simcards. Actualmente posee un sistema de recolección de información digital sobre las ventas que se generan a diario, sin embargo este sistema no tiene un mecanismo eficiente de análisis del gran volumen de información que se recolecta de manera digital.

1. DELIMITACION DEL PROBLEMA

En la empresa DISPROCEL S.A se necesita aplicar una solución de minería de datos que transforme los datos en información útil de manera que el área de ventas de la empresa pueda utilizar la información , brindándoles un soporte en el cual puedan respaldar la toma de decisiones estratégicas.

2. FORMULACION DEL PROBLEMA

Tomando en cuenta lo antes expuesto, se puede definir la siguiente interrogante:

¿Cómo podría la minería de datos contribuir a mejorar el análisis en el área de mercadeo y ventas de la empresa DISPROCEL S.A?

3. SISTEMATIZACION DEL PROBLEMA

¿Cuál es la situación actual en que se encuentra la empresa con respecto al análisis de datos?

¿Cómo se podría mejorar el análisis de datos y generación de reportes para toma de decisiones estratégicas en la empresa?

¿Cómo evaluar la solución de minería de datos que se implementara en el área de Mercadeo y ventas de la empresa?

V. JUSTIFICACION

En el área de ventas de DISPROCEL se maneja un gran volumen de información digital almacenada en su base de datos. Los reportes generados actualmente no logran extraer de manera eficiente la información que la empresa considera relevante y necesaria. Por lo tanto, se constata la necesidad de implementar una solución de minería de datos para generar reportes con información relevante para la empresa de tal manera que se pueda aprovechar los beneficios de la base de datos.

Entre los beneficios que nos proporcionaría la solución serían:

- Proporcionar el apoyo necesario para la toma de decisiones, hacia el cumplimiento de sus proyectos de manera exitosa.
- Identificar a los clientes con mayor volumen de compra de productos, para brindarles una mejor atención en el tiempo requerido.
- Descubrir patrones escondidos dentro de la enorme cantidad de datos de las bases de datos y así lograr mejorar la interacción con el cliente y optimizar las ventas.
- Simplificar los esfuerzos del personal involucrado.

VI. OBJETIVOS

OBJETIVO GENERAL

Desarrollar un modelo basado en minería de datos aplicando el algoritmo de **Clúster** en el área de mercadeo y ventas de la empresa DISPROCEL S.A en Managua en el II semestre 2015.

OBJETIVOS ESPECIFICOS

- Analizar los principales procesos que se realizan en el área de Mercadeo y Ventas de DISPROCEL S.A
- Desarrollar una solución de minería de datos que permita el análisis de datos y generación de reportes en el área de mercadeo y ventas de manera que contribuya a la toma de decisiones estratégicas en la empresa.
- Evaluar la solución de minería de datos por parte de los usuarios finales de la Empresa DISPROCEL S.A, siguiendo el proceso definido en la norma ISO 9126-1.

VII. MARCO TEORICO

1. DATOS

De acuerdo a (Forouzan, 2006) los datos son un conjunto de caracteres con algún significado el cual puede ser numérico, alfabético, o alfanumérico y que se puede representar por medio de texto, números, imagen, audio o video. También se puede definir los datos como un conjunto de señales o signos con un significado particular y que dentro de un contexto empresarial, el concepto de dato es definido como un registro de transacciones. Un dato no dice nada sobre las cosas, y por sí mismo tiene poca o ninguna relevancia o propósito. Los datos describen únicamente una parte de lo que pasa en la realidad, no proporcionan juicios de valor o interpretaciones, y por lo tanto no son orientativos para la acción. La toma de decisiones utilizará los datos, pero éstos nunca dirán lo que hacer, no dicen nada acerca de lo que es importante o no. A pesar de ello, éstos son indispensables para las organizaciones, pues son la base para la creación de información(Senn., 1990)

2. INFORMACION

Según (Idalberto, 2006), información es un conjunto de datos con un significado 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. La información es un conjunto de datos acerca de algún suceso, hecho o fenómeno, que organizados en un contexto determinado tienen su significado, cuyo propósito puede ser el de reducir la incertidumbre o incrementar el conocimiento acerca de algo(Senn., 1990).

3. CONOCIMIENTO

El conocimiento es el entendimiento de las causas y efectos que envuelven ideas y acciones que requieren el uso de habilidades de mayor orden o pensamiento crítico(Herrera, 2001). La búsqueda de los conocimientos es tan vieja como la historia de la humanidad refieren(Gomez, 1997). Cada día se producen más conocimientos y los conocimientos científicos y técnicos han ampliado los horizontes de la humanidad hasta límites insospechados. Su crecimiento es virtualmente irreversible, exponencial o casi y no existe previsiblemente límite.

4. SISTEMAS DE INFORMACION

4.1 SISTEMAS

Sistemas es un término que se usa comúnmente en todas las partes del mundo, cuando hablamos por ejemplo de sistemas informáticos, sistemas organizacionales, sistemas gubernamentales, sistemas financieros etc. Este término es bastante amplio y puede usarse en diferentes formas. "un sistema es una colección de componentes interrelacionados que trabajan conjuntamente para cumplir algún objetivo"(Sommerville, 2005)

Son parte de estos los **Sistemas de información**, que depende en mayor o menor medida de un sistema organizacional. Por medio de este fluyen los datos de una persona o de un departamento hacia otro y puede ser cualquier cosa, desde la comunicación interna entre los diferentes componentes de la organización hasta sistemas de cómputos que generen reportes periódicamente para varios usuarios.(Senn., 1990).

3. CONOCIMIENTO

El conocimiento es el entendimiento de las causas y efectos que envuelven ideas y acciones que requieren el uso de habilidades de mayor orden o pensamiento crítico(Herrera, 2001). La búsqueda de los conocimientos es tan vieja como la historia de la humanidad refieren(Gomez, 1997). Cada día se producen más conocimientos y los conocimientos científicos y técnicos han ampliado los horizontes de la humanidad hasta límites insospechados. Su crecimiento es virtualmente irreversible, exponencial o casi y no existe previsiblemente límite.

4. SISTEMAS DE INFORMACION

4.1 SISTEMAS

Sistemas es un término que se usa comúnmente en todas las partes del mundo, cuando hablamos por ejemplo de sistemas informáticos, sistemas organizacionales, sistemas gubernamentales, sistemas financieros etc. Este término es bastante amplio y puede usarse en diferentes formas. "un sistema es una colección de componentes interrelacionados que trabajan conjuntamente para cumplir algún objetivo"(Sommerville, 2005)

Son parte de estos los **Sistemas de información**, que depende en mayor o menor medida de un sistema organizacional. Por medio de este fluyen los datos de una persona o de un departamento hacia otro y puede ser cualquier cosa, desde la comunicación interna entre los diferentes componentes de la organización hasta sistemas de cómputos que generen reportes periódicamente para varios usuarios.(Senn., 1990).

Un sistema informático como todo sistema, es el conjunto de partes interrelacionadas, hardware, software y de recurso humano 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(Kendall, 2007).

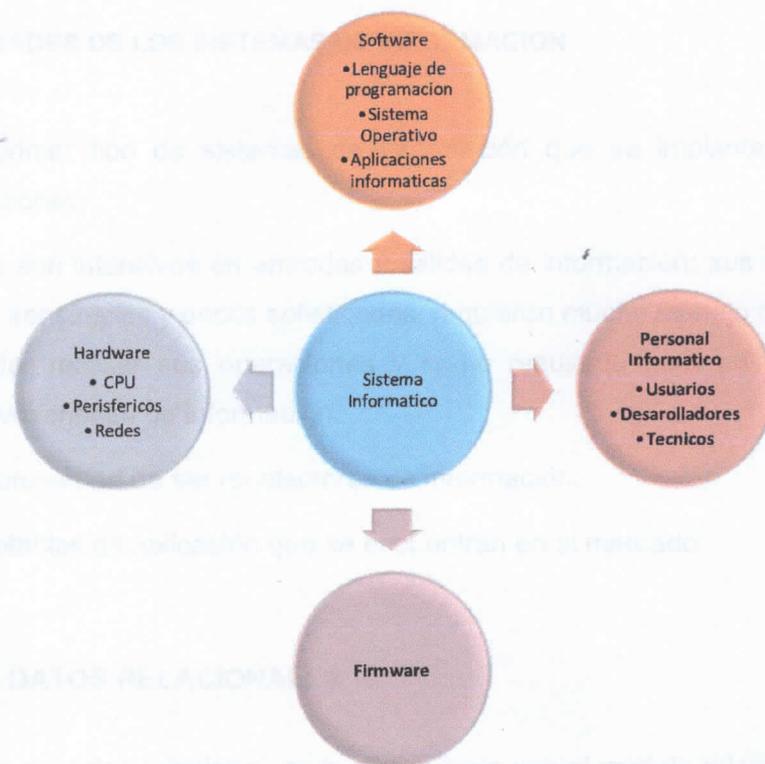


FIGURA1. SISTEMAS DE INFORMACIÓN

4.2 CARACTERISTICAS DE LOS SISTEMAS DE INFORMACION

Todo sistema de información es desarrollado para cumplir ciertos objetivos, para alcanzar estos objetivos este tiene que interactuar con su medio ambiente, que son todos los objetos que están fuera de la frontera de los sistemas, por lo que surgen dos tipos de sistema: Sistemas abiertos estos sistemas son los que interactúan con su medio ambiente, es decir que produce entradas y salidas, todos los sistemas actuales son abiertos(Senn., 1990).

4.3 GENERALIDADES DE LOS SISTEMAS DE INFORMACION

- Son el primer tipo de sistemas de información que se implanta en las organizaciones.
- Procesos son intensivos en entradas y salidas de información; sus cálculos y suelen ser simples y pocos sofisticados, requieren mucho manejo de datos para poder realizar sus operaciones y como resultado generan también grandes volúmenes de información.
- Tiene la propiedad de ser recolectores de información.
- Son adaptables de aplicación que se encuentran en el mercado.

4.4 BASES DE DATOS RELACIONALES

Una base de datos relacional es la que cumple con el modelo relacional, el cual es el modelo más utilizado en la actualidad para implementar bases de datos ya planificadas. Postuladas sus bases en 1970 por Edgar Frank Codd, de los laboratorios IBM, estas permiten establecer relaciones entre los datos, y así relacionar los datos de ambas tablas, de ahí proviene su nombre Modelo relacional. (Senn., 1990).

5.5 MODELO RELACIONAL

El modelo relacional propone un modelo basado en la teoría matemática de las relaciones, con el objetivo de mantener la independencia de la estructura lógica respecto al modo de almacenamiento y otras características de tipo físico. (Fernando, 2006).

Entre los primeros sistemas de gestión de bases de datos relacionales basados en microcomputadoras estaban dBase, R: baseyParadox. El Access de Microsoft, que usa el modelo relacional, es ahora el sistema de gestión de base de datos basado en microcomputadora más ampliamente utilizado. Oracle, DB2, Informix, Sybase y el SQL Server de Microsoft, todos los cuales usan el modelo relacional, en la actualidad son los sistemas de gestión de bases de datos empresariales más populares. (Fernando, 2006).

El modelo relacional introduce el concepto de relación o tablas como estructura básica del modelo. Todos los datos de una base de datos se representan en forma de relaciones cuyo contenido varía en el tiempo. El modelo relacional se basa en dos ramas de las matemáticas: la teoría de conjuntos y la lógica de predicados. Esto hace que sea un modelo seguro y robusto. (Senn., 1990).

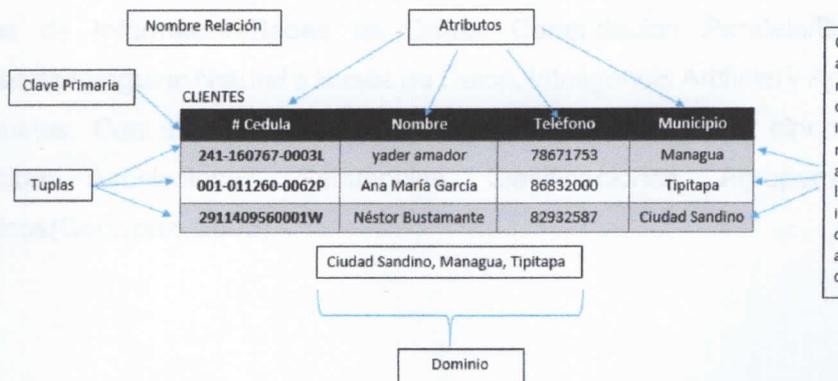


FIGURA 2.MODELO RELACIONAL

5. MINERIA DE DATOS

La minería de datos es definida por (Theodoulidis, 2003) como el proceso de descubrimiento del conocimiento sobre los repositorios de datos complejos mediante la extracción oculta y potencialmente útil en forma de patrones globales y relaciones estructurales implícitas entre datos. Otros como (Eibe, 2005) apuntan que la minería de datos como el proceso en que se extrae conocimiento útil y comprensible previamente desconocido y a partir de grandes conjuntos de datos almacenados en distintos formatos. (Hernandez, 2004) Aclara que existen muchos términos que se relacionan o utilizan como sinónimos de la minería de datos una de ellas es el KDD un proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y en última instancia comprensibles a partir de los datos, es un proceso más complejo que lleva no solo a obtención de modelos o patrones, que es el objetivo de la minería de datos si no que incluye además una evaluación y una posible interpretación de los mismos.

5.1 TECNICA DE MINERIA DE DATOS

De acuerdo Los elementos que hacen posible la minería de datos han estado bajo desarrollo por muchos años en áreas de investigación como Estadística, Sistemas de Información/Bases de Datos, Computación Paralela/Distribuida, Interfaces de Lenguaje Natural a Bases de Datos, Inteligencia Artificial y Aprendizaje de Máquinas. Con el apoyo de estas áreas, se puede producir cinco tipos de información: Asociaciones, Secuencias, Clasificaciones, Agrupamientos y Pronósticos (Gonzalez, 2005).

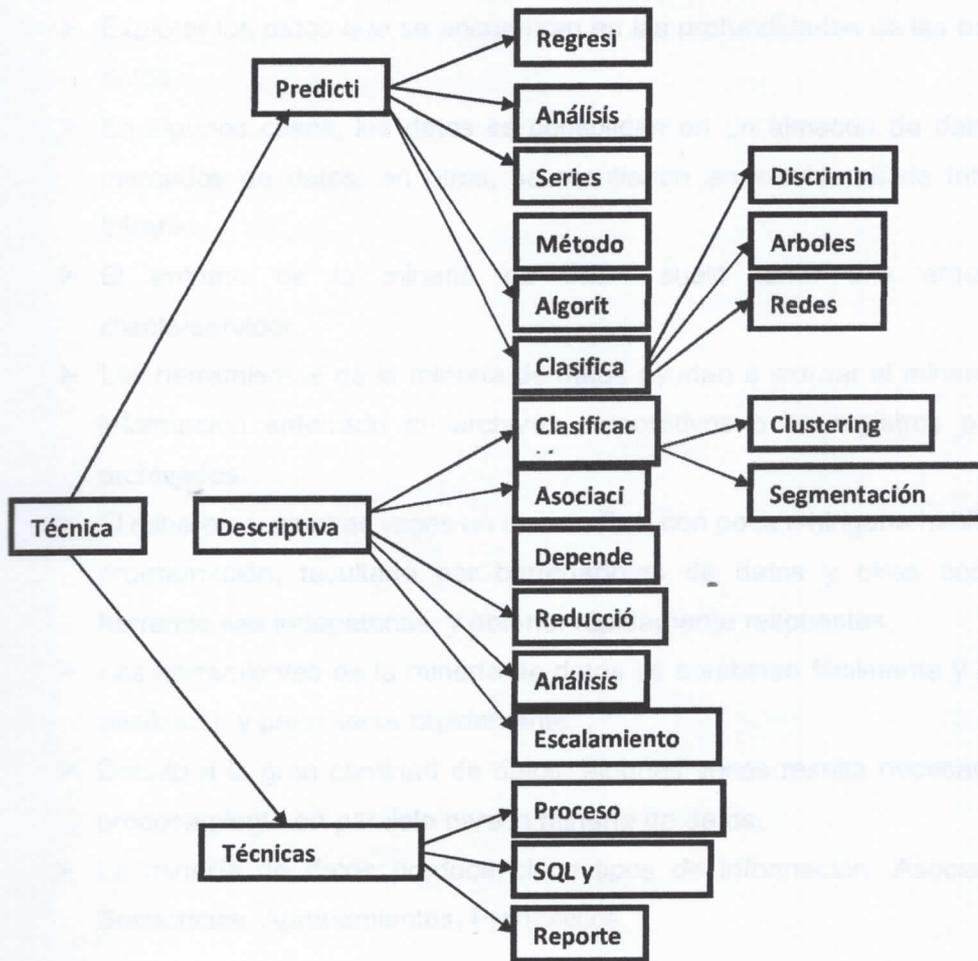


FIGURA 3. CLASIFICACIÓN MINERÍAS DE DATOS

5.2 CARACTERÍSTICAS DE LA MINERÍA DE DATOS

- Explorar los datos que se encuentran en las profundidades de las bases de datos.
- 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.
- El entorno de la minería de datos suele tener una arquitectura cliente/servidor.
- Las herramientas de la minería de datos ayudan a extraer el mineral de la información enterrado en archivos corporativos o en registros públicos, archivados.
- El minero es, muchas veces un usuario final con poca o ninguna habilidad de programación, facultado por barrenadoras de datos y otras poderosas herramientas indagatorias y obtener rápidamente respuestas.
- Las herramientas de la minería de datos se combinan fácilmente y pueden analizarse y procesarse rápidamente.
- 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, Agrupamientos, Pronósticos

5.3 ALMACEN DE DATOS

DW (Data Warehouse Almacén de Datos) es un sistema de base de datos que posee una colección y un software de gestión. Sin embargo a diferencia de un sistema de base de datos relacional está concebido para la toma de decisiones y es utilizado para la generación de reportes y análisis. Los datos almacenados en un DW son obtenidos de sistemas de base de datos transaccionales dichos datos pueden pasar por bases de datos intermedias para realizárseles operaciones adicionales necesarias para ser usados en el DW (Lewis William, 2001).

Típicamente un DW utiliza capas de preparación, integración y acceso para albergar sus funciones claves. La capa de preparación alberga los datos en bruto extraídos de cada sistema de fuente de datos. La capa de integración realiza la limpieza y homogenización de los datos provenientes de las diversas fuentes. Los datos integrados y limpiados son movidos a otra base de datos en la cual los datos son agrupados y ordenados en grupos jerárquicos y la capa de acceso está pensada para ayudar a los usuarios a recuperar los datos (Lewis William, 2001) .

OLAP (procesamiento analítico en línea) es un término utilizado para describir el análisis de datos complejos desde el almacén de datos. El modelo multidimensional es un modelo utilizado ampliamente para OLAP. En los modelos multidimensionales se rellenan matrices multidimensionales llamadas **cubos de datos** los cuales deben estar organizados en el cubo de manera tal que facilite la ejecución de la consulta no predeterminada de información agregada(Lewis William, 2001).

Para el almacenamiento multidimensional se necesitan dos tipos de tablas de dimensión y de hechos, las de dimensión constan de tuplas de atributos mientras que la de hechos son una agrupación de tuplas una por cada hecho registrado cada hecho contiene variables medidas y son asociadas mediante punteros con las tablas de dimensión que le dan un contexto(Lewis William, 2001).

El esquema de estrella es el esquema multidimensional más simple consiste en una tabla de hechos con una única tabla por cada dimensión. Este esquema es diseñado para optimizar la facilidad de uso por parte de los usuarios y el desempeño de la obtención de datos minimizando el número de tablas a combinar para llevar acabo las operaciones (Lewis William, 2001).

5.4 ESTRUCTURA DE LA MINERÍA DE DATOS

La estructura de minería de datos define los datos a partir de los cuales se generan los modelos de minería de datos, especifica la vista de datos de origen, el número y el tipo de columnas, y una partición opcional en conjuntos de entrenamiento y de pruebas. Una misma estructura de minería de datos puede admitir varios modelos de minería de datos que comparten el mismo dominio. (Hernandez, 2004)

La configuración de una estructura de minería de datos consta de los pasos siguientes.

- Definir un origen de datos.
- Seleccionar las columnas de estructura y definir una clave.
- Crear particiones de los datos de origen en un conjunto de entrenamiento y un conjunto de pruebas opcional.
- Procesar la estructura

Orígenes de datos para estructuras de minería de datos

Una vista del origen de datos permite combinar varios orígenes de datos y utilizarlos como un origen único en la estructura o el modelo de minería de datos que se ha creado. Los orígenes de datos originales no son visibles para las aplicaciones cliente.

Columnas de la estructura de minería de datos

Las columnas contienen información como el tipo de datos, el tipo de contenido y el modo en que se distribuyen los datos. La estructura de minería de datos no contiene información sobre el modo en que las columnas se utilizan para un modelo de minería de datos concreto, ni sobre el tipo de algoritmo que se utiliza para generarlo; esta información se define en el propio modelo de minería de datos.

Una estructura de minería de datos también puede contener tablas anidadas. Una tabla anidada representa una relación de uno a varios entre la entidad de un caso y sus atributos relacionados.(Hernandez, 2004)

Datos de aprendizaje y de pruebas

Cuando se definen los datos para la estructura de minería de datos, también es posible especificar que algunos de los datos se utilicen para aprendizaje y otros para pruebas. Por consiguiente, ya no es necesario crear particiones de los datos antes de crear una estructura de minería de datos. Puede especificar que un cierto porcentaje de los datos se reserven para pruebas y que el resto se utilice para aprendizaje, o puede especificar un cierto número de casos para que se utilicen como conjunto de datos de pruebas. La información de la partición se almacena en memoria caché con la estructura de minería de datos; por consiguiente, el mismo conjunto de pruebas puede utilizarse con todos los modelos que están basados en esa estructura.(Hernandez, 2004)

Procesar estructuras de minería de datos

Una estructura de minería de datos no es más que un contenedor de metadatos. Al procesar una estructura de minería de datos, AnalysisServices crea una caché que almacena estadísticas sobre los datos, información sobre cómo se discretizan los atributos continuos e información que usarán los modelos posteriormente. Por sí mismo, el modelo de minería de datos no almacena datos; hace referencia a la información de la caché. Por tanto, al procesar un modelo de minería de datos, la caché de la estructura debe estar disponible. Si no está disponible, se debe volver a procesar la estructura antes de que se genere el modelo. (López, 2007)

Si los datos en caché están disponibles, no será necesario volver a procesar la estructura de minería de datos cada vez que agregue a ella un nuevo modelo de minería de datos; puede procesar el modelo únicamente.

5.5 MODELO DE MINERIA DE DATOS

Un modelo recibe los datos de una estructura de minería de datos, los analiza utilizando un algoritmo, la estructura y el modelo son objetos independientes. Una vez procesado el modelo, contiene los metadatos, resultados y enlaces de la estructura de minería de datos. (Senn., 1992).

La generación de un modelo de minería de datos forma parte de un proceso mayor que incluye desde la formulación de preguntas acerca de los datos y la creación de un modelo para responderlas, hasta la implementación del modelo en un entorno de trabajo. Este proceso se puede definir mediante los seis pasos básicos siguientes:

FIGURA 4 MINERÍA DE DATOS

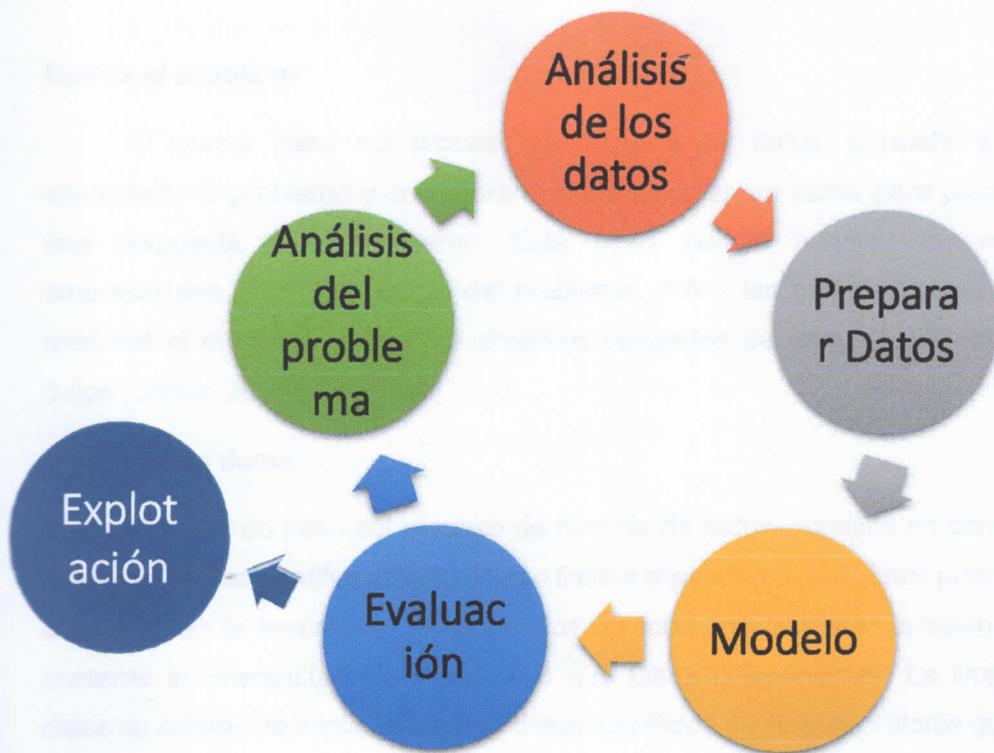


FIGURA 4. MINERIAS DE DATOS

Definir el problema:

El primer paso del proceso de minería de datos, consiste en definir claramente el problema y considerar formas de usar los datos para proporcionar una respuesta para el mismo. Este paso incluye analizar los requisitos empresariales, definir el ámbito del problema, definir las métricas por las que se evaluará el modelo y definir los objetivos concretos del proyecto de minería de datos. (López, 2007)

Preparar los datos:

El segundo paso del proceso de minería de datos, consiste en consolidar y limpiar los datos identificados en el paso Definir el problema. Los datos pueden estar dispersos en la empresa y almacenados en formatos distintos; también pueden contener incoherencias como entradas que faltan o incorrectas. La limpieza de datos no solamente implica quitar los datos no válidos o interpolar valores que faltan, sino también buscar las correlaciones ocultas en los datos, identificar los orígenes de datos que son más precisos y determinar qué columnas son las más adecuadas para el análisis.

Explorar los datos:

El tercer paso del proceso de minería de datos, consiste en explorar los datos preparados. Debe conocer los datos para tomar las decisiones adecuadas al crear los modelos de minería de datos. Entre las técnicas de exploración se incluyen calcular los valores mínimos y máximos, calcular la media y las desviaciones estándar, y examinar la distribución de los datos. Se puede usar herramientas como Master Data Services para sondear los orígenes de datos disponibles y determinar su disponibilidad para la minería de datos. También se Puede usar herramientas como SQL Server Data Quality Services, o el generador de perfiles de datos de Integration Services, para analizar la distribución de los datos (Hernandez, 2004).



Generar modelos:

El cuarto paso del proceso de minería de datos, consiste en generar el modelo o modelos de minería de datos. Se usaran los conocimientos adquiridos en el paso Explorar los datos para definir y crear los modelos. Antes de procesar la estructura y el modelo, un modelo de minería de datos simplemente es un contenedor que especifica las columnas que se usan para la entrada, el atributo que está prediciendo y parámetros que indican al algoritmo cómo procesar los datos. El procesamiento de un modelo a menudo se denomina entrenamiento. El entrenamiento hace referencia al proceso de aplicar un algoritmo matemático concreto a los datos de la estructura para extraer patrones (PEREZ., 2005).

Explorar y validar los modelos:

El quinto paso del proceso de minería de datos, consiste en explorar los modelos de minería de datos que ha generado y comprobar su eficacia. Analysis Services proporciona herramientas que ayudan a separar los datos en conjuntos de datos de entrenamiento y pruebas, para que pueda evaluar con precisión el rendimiento de todos los modelos en los mismos datos. El conjunto de datos de entrenamiento se utiliza para generar el modelo y el conjunto de datos de prueba para comprobar la precisión del modelo mediante la creación de consultas de predicción.

Implementar y actualizar los modelos:

El último paso del proceso de minería de datos, consiste en implementar los modelos que funcionan mejor en un entorno de producción. Una vez que los modelos de minería de datos se encuentran en el entorno de producción, puede llevar a cabo diferentes tareas. Utilizar Integration Services para crear un paquete en el que se utilice un modelo de minería de datos para dividir de forma inteligente los datos entrantes en varias tablas.

Actualizar dinámicamente los modelos, cuando entren más datos en la organización, y realizar modificaciones constantes para mejorar la efectividad de la solución.

5.6 PROPIEDADES DEL MODELO DE MINERÍA DE DATOS

Cada modelo de minería de datos tiene propiedades que definen el modelo y sus metadatos. Estos incluyen el nombre, la descripción, la fecha en que se procesó el modelo por última vez, los permisos del modelo y los filtros que se usan en los datos para el entrenamiento. Si alguna de las columnas que usa el modelo es una tabla anidada, también se le puede aplicar un filtro independiente. Además, cada modelo de minería de datos contiene dos propiedades especiales: Algoritmo y Usage.(Hernandez, 2004)

La propiedad Usage define cómo usa el modelo cada columna. Puede definir el uso de la columna como Input, Predict, PredictOnly o Key. La propiedad Usage se aplica a las columnas de modelo de minería de datos individuales y debe establecerse individualmente para cada columna que se incluye en un modelo. Si la estructura contiene una columna que no se usa en el modelo, el uso se establece en Ignore.(PEREZ., 2005)

6. HERRAMIENTAS DE LA MINERÍA DE DATOS

SQL (Structured Query Language) es un lenguaje de programación diseñado para el manejo de datos en sistemas de administración de datos relacionales. Para poder realizar la consulta, inserción eliminación y actualización de datos en varias bases de datos contempladas en la solución desarrollada en este proyecto es necesaria la correcta utilización de SQL ya que dichos datos van a estar almacenados en un sistema de administración de datos relacional más específicamente SQL Server 2012.

6.1 HERRAMIENTAS DE SQL SERVER

Integration Services:

Es un componente de SQL Server que puede ser utilizado para llevar a cabo un amplio rango de tareas de migración, integración y transformación de datos. Integration Services rápidas y flexibles para realizar dichas tareas. Dichas herramientas pueden ser usadas además para automatizar el mantenimiento de base de datos SQL Server y la utilización de datos multidimensionales de cubos. Proporciona una completa herramienta ETL (Extraction, Transformation, Loading) que nos permitirá captar datos de fuentes heterogéneas, convertirlos a datos compatibles con nuestras reglas de negocio e incorporarlos a nuestros almacenes de datos sean estos servidores SQL Server o no. (Knight, 2012).

Analysis Services:

Analysis Services (Servicios de Análisis de SQL Server) es una herramienta OLAP, de minería de datos y de reportes de Microsoft SQL Server. Es usada como una herramienta para analizar y dar sentido a información que podría estar repartida en múltiples bases de datos o en distintas tablas (Microsoft, 2012).

Microsoft SQL Server Business Intelligence Development Studio

Business Intelligence Development Studio (Estudio y desarrollo de inteligencia de negocios) Es un ambiente de desarrollo integrado de Windows y se usa para desarrollar análisis de datos e inteligencia de negocios utilizando servicios integrados. Está basado en el ambiente de desarrollo de Microsoft Visual Studio pero se agregaron proyectos específicos de SQL Server, así como tipos de proyectos incluyendo herramientas así como controles y proyectos para reportes y flujos de datos ETL, cubos de datos OLAP y estructuras de minería de datos. (Microsoft, 2012).

6.2 ALGORITMO DE MINERIA DE DATOS

Los Algoritmos de minería de datos son quienes forman un modelo una estructura, se analiza un conjunto de datos en los cuales se encuentran una serie de patrones y tendencias específicas es decir que encuentran todas las conexiones posibles que pueda haber en toda la información.

Algoritmo de clúster:



FIGURA 5. ALGORITMO DE CLUSTER

El análisis clúster es un conjunto de métodos o técnicas estadísticas que permiten describir y reconocer diferentes agrupaciones que subyacen en un conjunto de individuos o unidades de observación, es decir, permiten clasificar o dividir en grupos más o menos homogéneos, un conjunto de individuos que están definidos por diferentes variables. El análisis clúster como un método estadístico multivariante

de clasificación, trata a partir de una tabla de datos (individuos-variables), de situarlos en grupos homogéneos o conglomerados, de manera que los individuos que pueden ser considerados similares sean asignados a un mismo clúster.

El análisis clúster puede caracterizarse como descriptivo y no inferencial. No tiene bases estadísticas sobre las cuales deducir inferencias para una población a partir de una muestra, y se utiliza fundamentalmente como una técnica exploratoria. Las soluciones no son únicas, en la medida que la pertenencia al conglomerado para cualquier número de soluciones depende de muchos elementos del procedimiento y se pueden obtener muchas soluciones diferentes variando uno o más de estos elementos.

La selección de las variables puede hacerse con relación a consideraciones teóricas, conceptuales o prácticas. Se deben incluir sólo aquellas variables que caracterizan los objetos que se están agrupando, y que se refieren específicamente a los objetivos del análisis (Peña, 2002).

Medidas de similitud:

La similitud es una medida de correspondencia o semejanza entre los objetos que van a ser agrupados. La estrategia más común consiste en medir la equivalencia en términos de la distancia entre los pares de objetos. Los objetos con distancias reducidas entre ellos son más parecidos entre sí que aquellos que tienen distancias mayores y se agruparán, por tanto, dentro del mismo clúster (Peña, 2002).

El algoritmo de clústeres se diferencia de otros algoritmos de minería de datos, como el algoritmo de árboles de decisión de Microsoft, en que no se tiene que designar una columna de predicción para generar un modelo de agrupación en clústeres. El algoritmo de clústeres entrena el modelo de forma estricta a partir de las relaciones que existen en los datos y de los clústeres que identifica el algoritmo.

Cómo Funciona el algoritmo

El algoritmo de clústeres de Microsoft identifica primero las relaciones de un conjunto de datos y genera una serie de clústeres basándose en ellas. Un gráfico de dispersión es una forma útil de representar visualmente el modo en que el algoritmo agrupa los datos.

Después de definir los clústeres, el algoritmo calcula el grado de perfección con que los clústeres representan las agrupaciones de puntos y, a continuación, intenta volver a definir las agrupaciones para crear clústeres que representen mejor los datos. El algoritmo establece una iteración en este proceso hasta que ya no es posible mejorar los resultados mediante la redefinición de los clústeres.

Los requisitos para un modelo de agrupación en clústeres son los siguientes

Una columna key: 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 Cada modelo debe tener al menos una columna de entrada que contenga los valores que se utilizan para generar los clústeres. Puede tener tantas columnas de entrada como desee, pero dependiendo del número de valores existentes en cada columna, la adición de columnas adicionales podría aumentar el tiempo necesario para entrenar el modelo.

Una columna de predicción opcional El algoritmo no necesita una columna de predicción para generar el modelo, pero puede agregar una columna de predicción de casi cualquier tipo de datos. Los valores de la columna de predicción se pueden tratar como entradas del modelo de agrupación en clústeres, o se puede especificar que solo se utilicen para las predicciones. (Microsoft, Algoritmo de cluster, s.f.)

Un problema relacionado con el análisis de clúster es la selección de factores en tareas de clasificación, debido a que no todas las variables tienen la misma importancia a la hora de agrupar los objetos. Otro problema de gran importancia y que actualmente despierta un gran interés es la fusión de conocimiento, ya que existen múltiples fuentes de información sobre un mismo tema, los cuales no utilizan

una categorización homogénea de los objetos. Para poder solucionar estos inconvenientes es necesario fusionar la información a la hora de recopilar, comparar o resumir los datos (Microsoft, Algoritmo de cluster, s.f.).

NORMA ISO 9126-1

Modelo de Calidad (ISO/IEC 9126-1)

Según (L.Alonzo, 2006) 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.

Nuestro proyecto será evaluado a partir de la segunda característica calidad de uso

Modelo para Calidad en Uso

(L.Alonzo, 2006) La calidad en uso del modelo se basa en tres características (Efectividad, Productividad, Satisfacción) que son esenciales para determinar la calidad de uso desde la perspectiva del usuario fase se divide en cuatro partes.

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.

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.

6.3 EMPRESA DISPROCEL S.A

DISPROCEL S.A es una empresa nicaragüense fundada en el año 2010. La empresa se dedica a la venta de recargas electrónicas, dispositivos móviles y Simcard. Para esto, la empresa posee un contrato de exclusividad con la empresa Telefónica de Nicaragua (Movistar) para la distribución y venta de los productos que les abastece Telefónica de Nicaragua.

La empresa se encuentra ubicada en el residencial los robles de plaza el sol 2 cuadras al sur y 3 ½ al este. La empresa fue fundada por el Ing. Orlando Arguello socio único de la empresa. Actualmente posee un área de contabilidad y finanzas, un área de mercadeo y ventas, área de recursos humanos, un director general de la empresa.

El área de mercadeo y ventas está encargada de adquirir la mercadería, su ingreso a bodegas y distribución a cada uno de los ejecutivos de venta ubicados en cada uno de los departamentos del país.

Los ejecutivos de ventas son el último eslabón en el proceso de comercialización de los productos de la empresa, debido a que estos realizan el contacto directo con el cliente final.

Actualmente la empresa ha adquirido un sistema de monitoreo de las ventas del producto y visitas a los clientes. A través de este software los ejecutivos de venta registran la información de los clientes y las visitas y/o ventas realizadas.

Sin embargo, la empresa a pesar de contar con el software antes mencionado, aún carece de un método para el análisis de las grandes cantidades de información que se registran día a día por parte del grupo de vendedores.

Estructura de la empresa DISPROCEL

Gerencia General

El área está compuesta únicamente por el propietario fundador de la empresa, el Ing. Orlando Arguello.

Recursos Humanos

Esta área está conformada por un director (Lic. Nubia Pavón) y dos asistentes encargadas de las contrataciones de personal, elaboración de planilla y otras actividades propias a la situación laboral de cada uno de los empleados de la empresa.

Contabilidad y Finanzas

El área de contabilidad y finanzas se encarga de llevar la situación financiera de la empresa, tanto de los costos operacionales como de los ingresos generados por la actividad comercial de la empresa. El área está dirigida por el Lic. Wilfredo Velázquez y por una asistente contable en el caso de la Lic. Grace López

Dirección General

El director general es el encargado de coordinar la buena andanza de la actividad comercial de la empresa. Este se encuentra en constante comunicación con la empresa Telefónica de Nicaragua. El encargado del área es el Lic. Marlon Gonzales.

Mercadeo

Esta área en estrecha relación con el área de ventas es encargada de la comercialización de los productos y generan toda la actividad comercial de la empresa. Como encargada en esta área tenemos a la Lic. Blanca Mercado, quien es asistida por el Ing. York López.

Bodega

En cada uno de los departamentos del país existe un encargado de bodega, el cual debe velar y verificar el inventario que ingresa y ponerlo a disposición del área de ventas para su distribución al equipo de ventas.

Aparte de cada una de las bodegas departamentales, se posee una bodega central. A esta bodega central ingresa la mercadería proveniente de los pedidos realizados a Telefónica de Nicaragua y de ahí se redistribuye a cada una de las bodegas departamentales según sea demandado. José Luis Granja es el encargado de la bodega central y la bodega Managua.

Área de Ventas

El área de ventas no posee un director dedicado propiamente al área, esta viene a ser una extensión del área de mercadeo sin embargo se encuentra coordinada directamente por el director general de la empresa en conjunto con la directora del área de mercadeo.

En el área de ventas se encuentran un equipo de cuatro personas, tres asistentes liquidadoras (Karla Calderón, Cindy Laínez y Jennifer Sotelo) y un responsable del área (Paul Pérez) que coordina las actividades y orientaciones provenientes de las áreas superiores.

El equipo de liquidadoras son las encargadas de asignar o distribuir los equipos y productos a cada uno de los ejecutivos de venta del departamento o zona correspondiente. Así mismo, llevan un control de las ventas diarias realizadas por cada uno de los ejecutivos que tienen a su cargo.

Equipo de venta

El equipo de venta está bajo la administración directa de la dirección general. Trabaja de manera estrecha con el área de ventas, debido a que es esta área la encargada de abastecer a cada uno de los miembros del equipo de venta. También es en esta área se da el proceso de liquidación o arqueo de las ventas realizadas y el inventario disponible por cada vendedor.

El equipo de ventas del departamento de Managua está compuesto por 15 ejecutivos que se movilizan en motos visitando clientes potenciales así como clientes que ya posee la empresa.

El equipo de ventas está compuesto además por un supervisor (Ing. Juan Incer). Entre las tareas que realiza el supervisor de los ejecutivos están: creación y asignación de rutas, asignación de rutas a vendedores, visitas de verificación a los clientes, transmitir orientaciones superiores al equipo de vendedores, asistirlos en caso de inconvenientes que surjan, etc.

Misión:

Somos una empresa de comercialización de productos y servicios en el área de telefonía móvil, confiable, eficiente; orientada a satisfacer las necesidades y aspiraciones de nuestros clientes. Somos un aporte positivo para la sociedad, generando empleo directo e indirecto dentro de un buen ambiente de trabajo y obteniendo un justo margen de utilidad.

Visión:

Situarnos como una empresa líder en el mercado, a través de nuestro producto, servicio, calidad e innovación. Teniendo como meta la satisfacción de nuestros clientes. Siempre guiados por una actitud ética y honesta.

Organigrama operación DISPROCEL

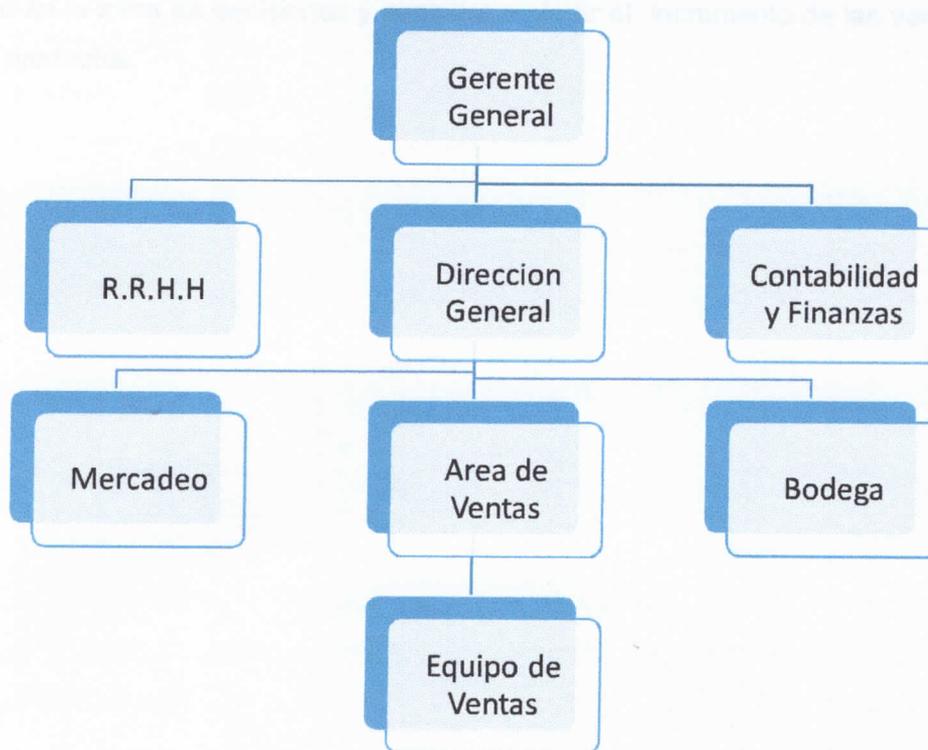


FIGURA 6. DIAGRAMA DE LA EMPRESA DISPROCEL

VIII. HIPOTESIS TOPOLOGICO

La solución de minería de datos en la empresa DISPROCEL fortalecerá el proceso en la toma de decisiones y permitirá mejorar el incremento de las ventas en sus productos.

1.2 UNIVERSO Y MUESTRA

Universo: Actualmente la empresa está compuesta por 7 áreas, el área de producción y ventas se encargará de realizar la aplicación de minería de datos por lo tanto confirmo nuestro universo.

Muestra: El área de producción está compuesta por 2 personas directas y asistencia y el área de ventas está compuesta por 4 personas que tienen el control de las ventas de las unidades, por lo tanto para la muestra se determinó utilizar el 100 % del mismo universo.

1.3 METODO DE RECOLECCION DE LA INFORMACION

Una vez definido el tipo de estudio se determinaron los métodos que serán usados para la recolección de información específica con respecto a cuánto se obtuvo que permitiera el contacto con el universo de estudio y permitir el acceso a la información que demanda la investigación sobre el universo de estudio.

IX. DISEÑO METODOLOGICO

1.1 TIPO DE ESTUDIO

En el campo de la investigación de minería de datos el presente estudio, es un desarrollo tecnológico y aplicado. De acuerdo a (PIURA, 2006) el tipo de estudio es descriptivo y analítico porque se tomaron datos los cuales se expusieron y se resumieron para luego analizar los resultados. Según el tiempo de ocurrencia de los hechos (Pineda, 2008) es de tipo prospectivo porque se está tomando información actual y según el periodo y secuencia es de tipo transversal porque según (Sampieri, 2010) se recolectan datos en un momento único con el su propósito de describir variables y analizar su incidencia e interrelación en un momento dado.

1.2 UNIVERSO Y MUESTRA

Universo: Actualmente la empresa está compuesta por 7 áreas, el área de mercadeo y ventas es donde se va realizar la solución de minería de datos por lo tanto conforma nuestro universo.

Muestra: El área de mercadeo está compuesto por 2 personas directora y asistente y el área de ventas está compuesto por 4 personas que llevan el control de las ventas de los productos, por lo tanto para la muestra se determinó utilizar el 100 % de nuestro universo.

1.3 METODO DE RECOLECCION DE LA INFORMACION

Una vez definido el tipo de estudio, se determinaran los métodos que serán usados para la recolección de información específica con respecto al objeto en estudio 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.

- Entrevista: Se realizó la entrevista al director de la empresa Lic. Marlon González, al encargado del área de mercadeo Lic. Blanca Mercado, en el área de ventas al equipo de liquidación que lo conforman 4 personas que son las que manejan la información que brindan de los vendedores a través de los dispositivos móviles.
- Correo electrónico: Se utilizó el correo electrónico para obtener información sobre algunas interrogantes que no fueron aclaradas en el momento de la entrevista.
- Cuestionario: se redactó un cuestionario con una serie de preguntas sobre información necesaria para llevar a cabo nuestro proyecto.

1.4 ESTUDIO DE FACTIBILIDAD

Estudio de factibilidad

Para la evaluación de la factibilidad del proyecto se ha tomado en cuenta el estudio de Factibilidad técnica, operativa y económica donde cada uno de ellos tiene su importancia en el desarrollo de una solución de minería de datos para mejorar la toma de decisiones en la empresa **DISPROCEL S.A (Distribuidora de Productos móviles)**.

Factibilidad Técnica

No es necesario adquirir nuevo equipo informático por que la empresa **DISPROCEL S.A (Distribuidora de Productos móviles)**. dispone con el equipo necesario para llevar a cabo la realización de la solución, por lo expuesto la solución es técnicamente factible.

Factibilidad operativa

Existen muchos procesos manuales en la elaboración de los informes para la toma de decisiones, ello conlleva a postergarlo además de generar ciertas dudas por el mismo hecho de hacerse manualmente.

Se toma la iniciativa de brindar una capacitación de como funcionaria la solución de minerías de datos que se va a poder explorar desde cualquier software, por lo tanto la solución si es operacionalmente factible.

Factibilidad económica

Se evalúa la factibilidad económica, dependiendo de la inversión, los costos operativos y los beneficios detallados en Presupuestos (Ver Presupuesto).

X. FASES DEL DESARROLLO DE MINERIA DE DATOS

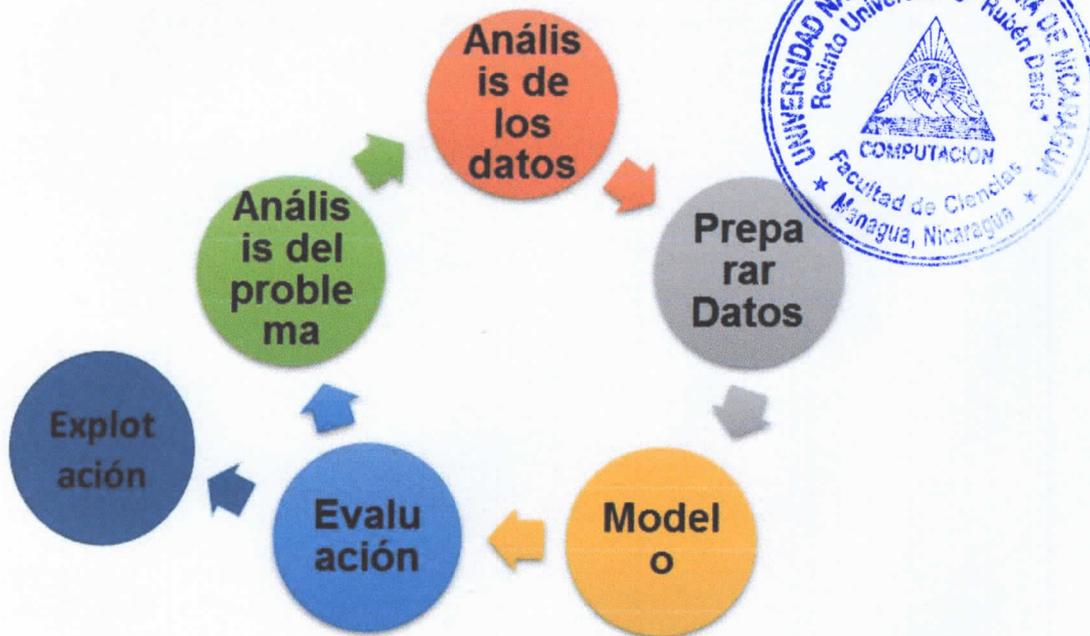


FIGURA 7. PROCESO DE MINERÍA DE DATOS

XI. HERRAMIENTAS QUE UTILIZARON PARA REALIZAR EL MODELO

- SQL Server 2012
- Embarcadero Studio
- Integration Services 2012
- Analysis Services 2012
- Microsoft Excel 2013
- Plugin de minería de datos para Excel

XII. ANALISIS DE RESULTADO

ESCENARIO:

DISPROCEL es una Empresa de comercialización de Equipos móviles. Actualmente desarrolla sus actividades en la ciudad de Managua.

Si bien su desarrollo es exitoso dentro de la ciudad, su mayor preocupación está en ampliar su mercado de comercialización en base a mejores precios y promociones.

NECESIDAD:

Por motivos de fidelidad por parte de los clientes hacia la empresa y agradecimiento a estos, se busca tener un listado de Clientes que más Compras hagan en un determinado Periodo y según esto poder darles descuentos en sus siguientes compras al igual ubicar las zonas donde menos se distribuyen los productos e impulsar decisiones estratégicas.

Posteriormente a esto se requiere información de los Vendedores que más ventas tengan en un lapso de un Año, para poder darles premios e incentivos. Se recomienda listar por tipos de ventas para el mayor ingreso de cada una de ellas.

Definir las Fuentes de Datos del Cubo.

A continuación se muestra la estructura de la base de datos que tiene la empresa DISPROCEL actualmente y es la base de datos transaccional que utilizamos para el proyecto es la base de datos que utiliza el sistema de la empresa para sus transacciones diarias, lo cual por medio de este, nos vamos a vasar para el desarrollo del diseño de diagrama de nuestro Data Warehouse que será indispensable para la creación del cubo.

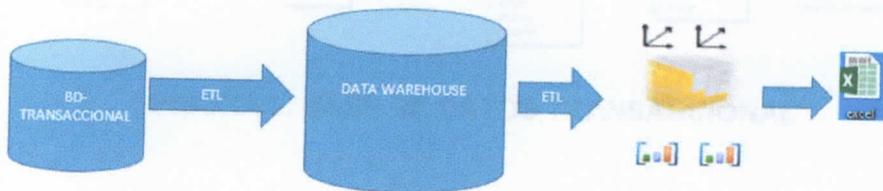


FIGURA 8.PROCESO DE EXTRACCION DE DATOS.

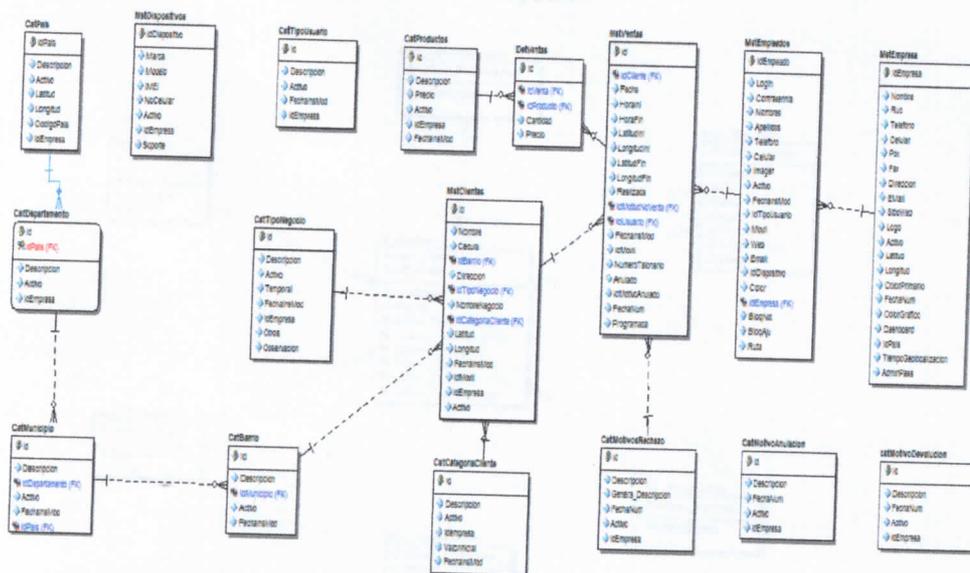


FIGURA 9. BASE DE DATOS TRANSACCIONAL

Definir la tabla de hechos y sus medidas

Los hechos están definidos por conjuntos de atributos (medidas), usualmente de tipo numérico que describe la actividad, las medidas para la tabla de hechos FT_Venta son: Cantidad de productos vendidos, precio de producto.

Las dimensiones están distribuidas en un grupo de atributos ordenadas jerárquicamente, así como la dimensión DimEmpleado está formado por un atributo nombre; la dimensión Destiempo está formado por siete atributos fecha, día, mes, cuatrimestre, trimestre, semestre, año; la dimensión DimProducto está formada por un atributo llamado descripción; la dimensión DimCliente está formado por un atributo llamado nombre; la dimensión DimTipoNegocio está formada por un atributo llamado descripción; la dimensión DimZona está formada por tres atributos departamento, municipio, barrio.

En la figura 8 se muestra nuestro Data Warehouse que se utilizó para realizar la solución de minería de datos, el cual se obtuvo mediante una reestructuración de la

base de datos inicial en la cual se tomaron las tablas que tienen la información principal y relevante para llevar a cabo este proyecto.

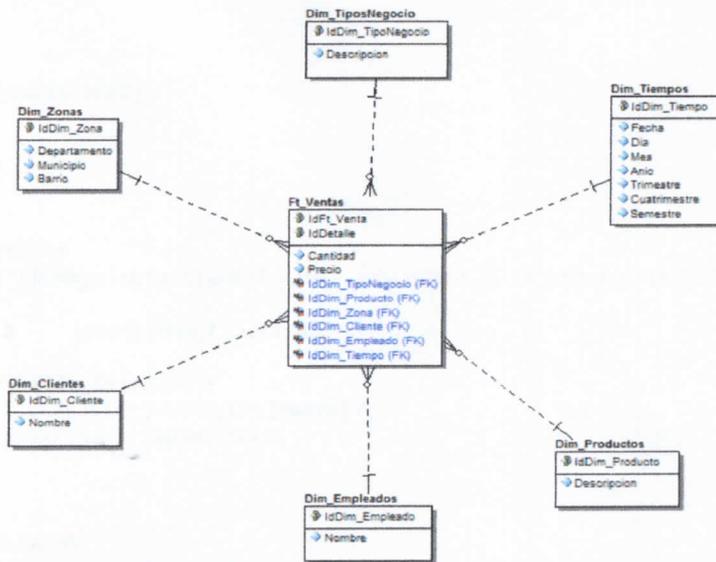


FIGURA 10. DATA WAREHOUSE

Creación de la Base de datos

Para el desarrollo de la base de datos utilizamos la herramienta SQL Server 2012 se creó una nueva base de datos que se llama DWDISPROCEL luego se crearon cada una de las tablas de dimensión que se diferencian por la palabra Dim y también se creó la de hechos.

En las tablas de dimensiones deben estar relacionadas con la tabla principal FT_Ventas la cual es la tabla de hechos que contiene las medidas del mismo. La tabla principal FT_Ventas tiene las claves foráneas de las claves principales de las tablas dimensiones las cuales tienen el mismo tipo de datos.

Script de la Data warehouse

-- SCRIPT DATAWAREHOUSE: 18/10/2015 19:09:34

```
USEmaster
GO
```

```
CREATEDATABASE[DwDiprocel]
GO
```

```
USE[DwDiprocel]
GO
```

--DIMENSION CLIENTES

```
IFEXISTS(SELECT*FROMsys.objectsWHEREobject_id=OBJECT_ID(N'[dbo].[Dim_Clientes]')ANDt
ypein(N'U'))
    DROPTABLE [dbo].[Dim_Clientes]
```

```
CREATETABLE[dbo].[Dim_Clientes](
    [IdDim_Cliente][int]NOTNULLPRIMARYKEY,
    [Nombre][nvarchar](150)NOTNULL
)
GO
```

--DIMENSION EMPLEADOS

```
IFEXISTS(SELECT*FROMsys.objectsWHEREobject_id=OBJECT_ID(N'[dbo].[Dim_Empleados]')AND
typein(N'U'))
    DROPTABLE [dbo].[Dim_Empleados]
```

```
CREATETABLE[dbo].[Dim_Empleados](
    [IdDim_Empleado][bigint]NOTNULLPRIMARYKEY,
    [Nombre][nvarchar](80)NOTNULL
)
GO
```

--DIMENSION PRODUCTOS

```
IFNOTEXISTS(SELECT*FROMsys.objectsWHEREobject_id=OBJECT_ID(N'[dbo].[Dim_Productos]')
ANDtypein(N'U'))
    DROPTABLE [dbo].[Dim_Productos]
```

```
CREATETABLE[dbo].[Dim_Productos](
    [IdDim_Producto][bigint]NOTNULLPRIMARYKEY,
    [Descripcion][nvarchar](200)NOTNULL
)
GO
```

--DIMENSION TIEMPOS

```
IFNOTEXISTS(SELECT*FROMsys.objectsWHEREobject_id=OBJECT_ID(N'[dbo].[Dim_Tiempos]')AN
Dtypein(N'U'))
    DROPTABLE [dbo].[Dim_Tiempos]
```

```
CREATETABLE[dbo].[Dim_Tiempos](
    [IdDim_Tiempo][bigint]NOTNULLPRIMARYKEY,
    [Fecha][datetime]NOTNULL,
    [Dia][int]NOTNULL,
    [Mes][int]NOTNULL,
    [Anio][int]NOTNULL,
)
```

```

        [Trimestre][int]NOTNULL,
        [Cuatrimestre][int]NOTNULL,
        [Semestre][int]NOTNULL
    )
GO

--DIMENSION TIPOS DE NEGOCIOS
IFNOTEXISTS(SELECT*FROMsys.objectsWHEREobject_id=OBJECT_ID(N'[dbo].[Dim_TiposNegocio]')ANDtypein(N'U'))
        DROPTABLE [dbo].[Dim_TiposNegocio]

CREATETABLE[dbo].[Dim_TiposNegocio](
        [IdDim_TipoNegocio][int]NOTNULLPRIMARYKEY,
        [Descripcion][nvarchar](150)NOTNULL
)
GO

--DIMENSION ZONAS
IFNOTEXISTS(SELECT*FROMsys.objectsWHEREobject_id=OBJECT_ID(N'[dbo].[Dim_Zonas]')ANDt
ypein(N'U'))
        DROPTABLE [dbo].[Dim_Zonas]

CREATETABLE[dbo].[Dim_Zonas](
        [IdDim_Zona][int]NOTNULLPRIMARYKEY,
        [Departamento][nvarchar](150)NOTNULL,
        [Municipio][nvarchar](150)NOTNULL,
        [Barrio][nvarchar](150)NOTNULL
)
GO

--FACT TABLE VENTAS
IFNOTEXISTS(SELECT*FROMsys.objectsWHEREobject_id=OBJECT_ID(N'[dbo].[Ft_Ventas]')ANDt
ypein(N'U'))
        DROPTABLE [dbo].[Ft_Ventas]

CREATETABLE[dbo].[Ft_Ventas](
        [IdFt_Venta][bigint]NOTNULLPRIMARY KEY,
        [IdDetalle][bigint]NOTNULL,
        [Cantidad][int]NOTNULL,
        [Precio][float]NOTNULL,
        [IdDim_TipoNegocio][int]NOTNULL,
        [IdDim_Producto][bigint]NOTNULL,
        [IdDim_Zona][int]NOTNULL,
        [IdDim_Cliente][int]NOTNULL,
        [IdDim_Empleado][bigint]NOTNULL,
        [IdDim_Tiempo][bigint]NOTNULL
)

ALTERTABLE[dbo].[Ft_Ventas]WITHCHECKADDCONSTRAINT[FK_Ft_Ventas_Dim_Clientes]FOREIGNK
EY([IdDim_Cliente])
REFERENCES[dbo].[Dim_Clientes]([IdDim_Cliente])

ALTERTABLE[dbo].[Ft_Ventas]WITHCHECKADDCONSTRAINT[FK_Ft_Ventas_Dim_Empleados]FOREIGN
KEY([IdDim_Empleado])
REFERENCES[dbo].[Dim_Empleados]([IdDim_Empleado])

ALTERTABLE[dbo].[Ft_Ventas]WITHCHECKADDCONSTRAINT[FK_Ft_Ventas_Dim_Productos]FOREIGN
KEY([IdDim_Producto])

```

```
REFERENCES[dbo].[Dim_Productos]([IdDim_Producto])
```

```
ALTERTABLE[dbo].[Ft_Ventas]WITHCHECKADDCONSTRAINT[FK_Ft_Ventas_Dim_Tiempos]FOREIGNKEY([IdDim_Tiempo])REFERENCES[dbo].[Dim_Tiempos]([IdDim_Tiempo])
```

```
ALTERTABLE[dbo].[Ft_Ventas]WITHCHECKADDCONSTRAINT[FK_Ft_Ventas_Dim_TiposNegocio]FOREIGNKEY([IdDim_TipoNegocio])REFERENCES[dbo].[Dim_TiposNegocio]([IdDim_TipoNegocio])
```

```
ALTERTABLE[dbo].[Ft_Ventas]WITHCHECKADDCONSTRAINT[FK_Ft_Ventas_Dim_Zonas]FOREIGNKEY([IdDim_Zona])REFERENCES[dbo].[Dim_Zonas]([IdDim_Zona])GO
```

Migración de los datos desde IntegrationServices

Al integrar los datos encada una de las tablas DIM de nuestro data warehouse. Utilizamos el control flujo de datos para extraer datos, aplicar transformaciones de nivel de columna y cargar datos. Dentro del flujo de control se agrega el control origen de datos en este control realizamos la conexión al servidor en SQL y luego elegimos nuestra base de datos transaccional y realizamos una consulta para extraer los datos desde la base de datos destino a el data warehouse. Después de haber configurado el control de flujo de datos agregamos un nuevo control de conversión de datos este control lo que hace es convertir los datos en un tipo de datos de Integreion Service a los datos numéricos se les asigna un tipo de dato numérico, a los datos de cadena se les asigna un tipo de datos de caracteres. Posteriormente se le agrega otro control dimensión de variación lenta el cual coordina la actualización e inserción de registros en las tablas de dimensiones de almacenamiento de datos. Esta transformación admite cuatro tipos de cambios: atributo variable, atributo histórico, atributo fijo y miembro deducido. Se selecciona la conexión al servidor y se selecciona la conexión a la base de datos destino al data warehouse y seleccionamos la tabla demisión en donde vamos integrar los datos, al ejecutar el proyecto podemos observar si se han o no cargado los datos podemos comprobarlo haciendo una consulta en el data warehouse desde SQL y ver los datos que se ha insertado.

INTEGRANDO LOS DATOS DIM PRODUCTOS

 DIM_CLIENTES

FIGURA 11. FLUJO DE DATOS

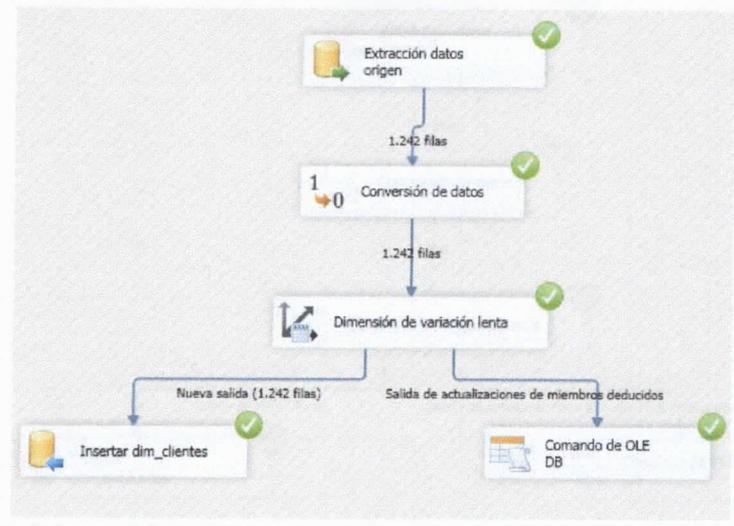


FIGURA 12. PROCESO FINALIZADO DIMCLIENTE

INTEGRANDO LOS DATOS DIMCLIENTE

 DIM_EMPLEADOS

FIGURA 13. FLUJO DE DATOS

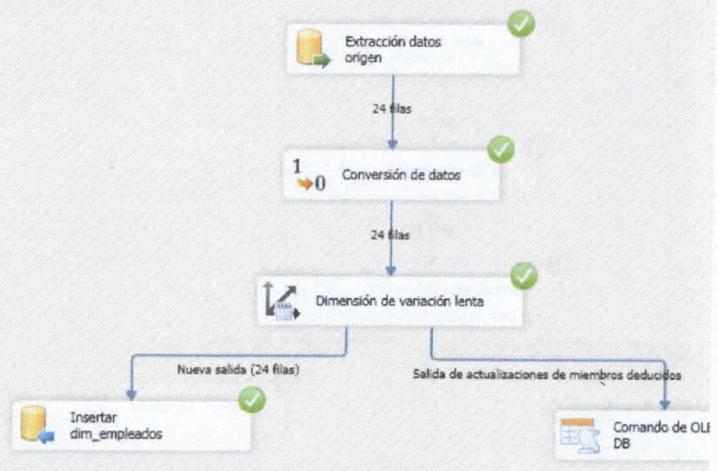


FIGURA 14. PROCESO FINALIZADO DIM EMPLEADO

INTEGRANDO LOS DATOS DIM EMPLEADOS

INTEGRANDO LOS DATOS DIM PRODUCTOS

 DIM_PRODUCTOS

FIGURA 15. FLUJO DE DATOS

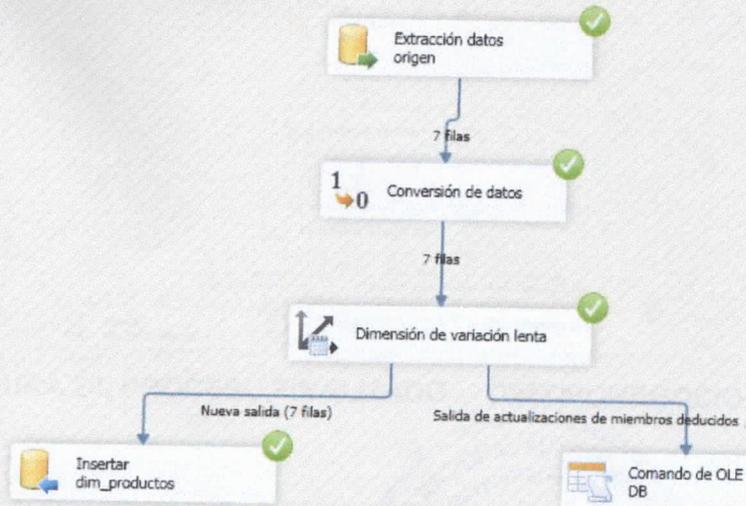


FIGURA 16. PROCESO FINALIZADO DIMPRODUCTO

INTEGRANDO LOS DATOS DIM TIEMPOS

 DIM_TIEMPOS

FIGURA 17. FLUJO DE DATOS

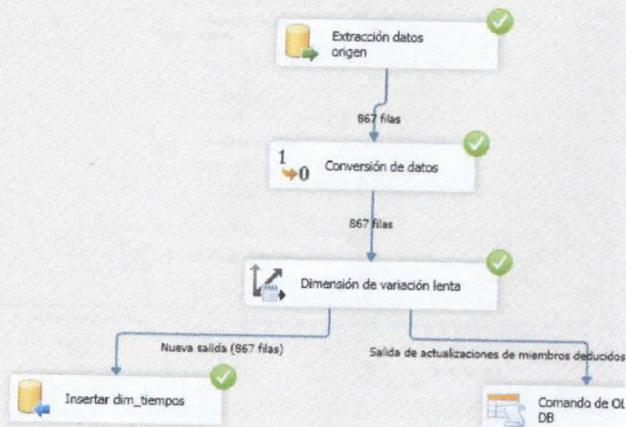


FIGURA 18. PROCESO FINALIZADO DIMTIEMPO

INTEGRANDO LOS DATOS TIPO_NEGOCIO

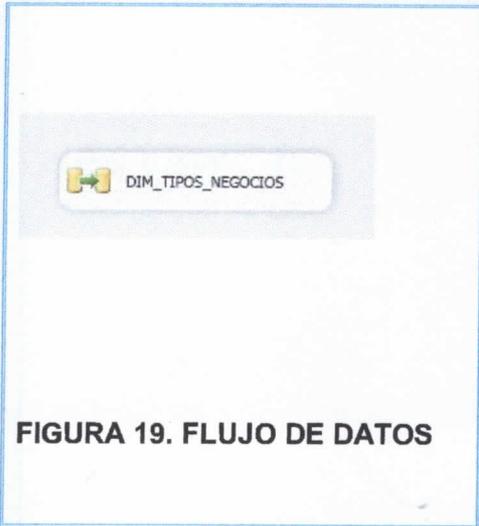


FIGURA 19. FLUJO DE DATOS



FIGURA 20. PROCESO FINALIZADO DIM TIPO NEGOCIO



INTEGRANDO DIM ZONAS

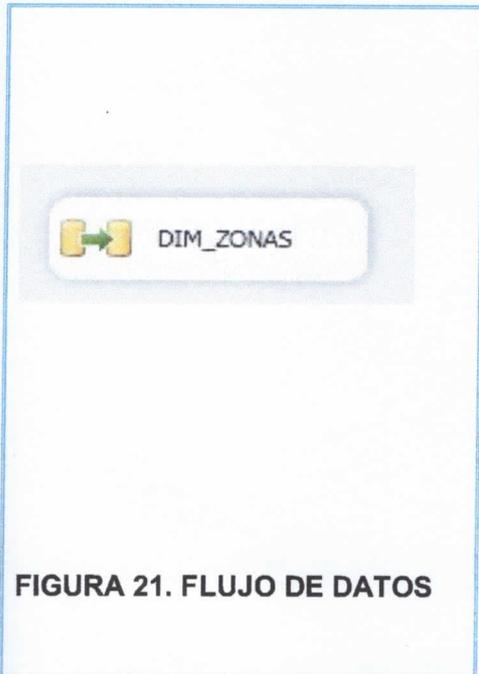


FIGURA 21. FLUJO DE DATOS

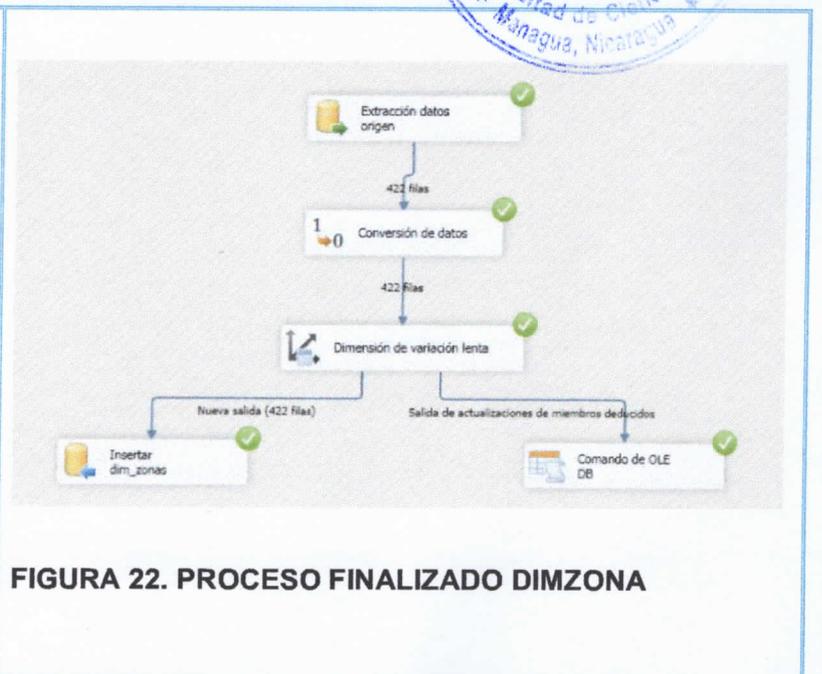


FIGURA 22. PROCESO FINALIZADO DIM ZONA

INTEGRANDO LOS DATOS DIM VENTA

FT_VENTAS

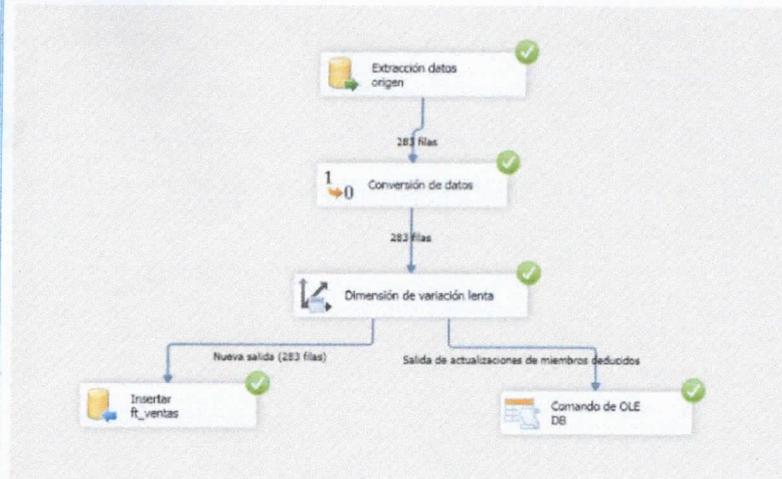


FIGURA 23. FLUJO DE DATOS

FIGURA 24. PROCESO FINALIZADO FTVENTAS

FIGURA 25. PROCESO ETL COMPLETO

PROCESO COMPLETO



FIGURA 25. PROCESO ETL COMPLETO

Proceso de análisis de datos con SSAS de SQL SERVER

Creación del modelo de minería de datos

Iniciamos el SQL Data Server Tools correspondiente con la versión del SQL Server de Microsoft. Una vez iniciado el programa, nos dirigimos al menú Archivo y luego al submenú Nuevo Proyecto, Una vez creada el proyecto iniciamos el desarrollo de la solución de minería de datos con la selección de un origen de datos.

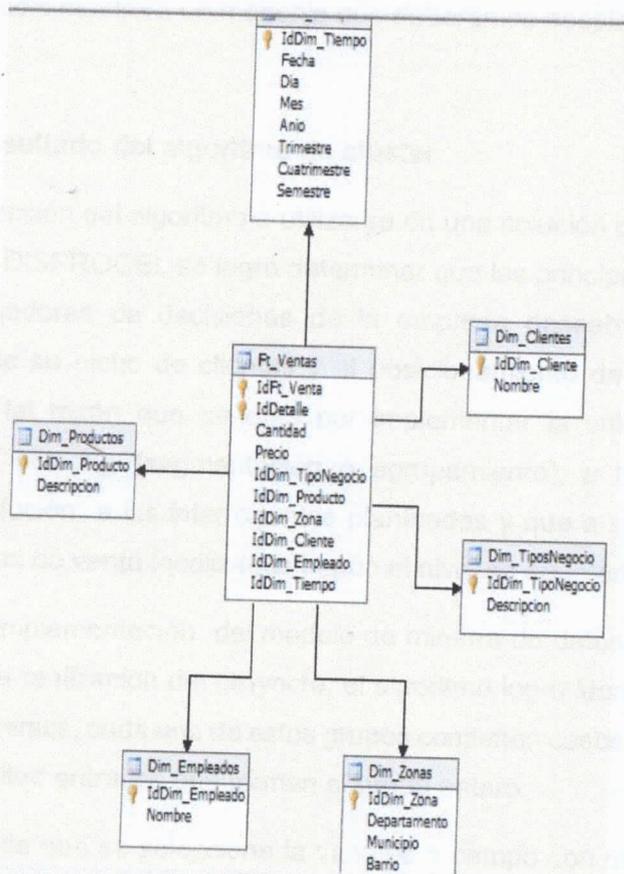


FIGURA 26. DIAGRAMA ESTRUCTURA DE MINERIA DE DATOS

Análisis del modelo.

Después de crear a la estructura de minería de datos, podemos empezar el análisis de los datos.

Entre las pestañas de la estructura de minería de datos seleccionamos visor de modelos de minería de datos. Si el modelo aún no ha sido implementado, el SQL Server Data Tools mostrara un mensaje que deberemos aceptar.

Análisis de resultado del algoritmo de clúster

Durante la selección del algoritmo a utilizarse en una solución de minería de datos en la empresa DISPROCEL se logró determinar que las principales incertidumbres que los manejadores de decisiones de la empresa deseaban aclarar eran la identificación de su nicho de clientes y el posicionamiento de su producto entre estos. Es por tal razón que se optó por implementar la solución utilizando el algoritmo de Clustering (segmentación o agrupamiento), a través del cual se planifico dar solución a las interrogantes planteadas y que a su vez contribuya a trazar estrategias de venta focalizadas según el nivel de segmentación alcanzada.

Después de la implementación del modelo de minería de datos sobre el datamart diseñado para la realización del proyecto, el algoritmo logro identificar diez grupos o segmentos diversos, cada uno de estos grupos contienen casos que poseen datos con alguna similitud entre los que forman el grupo entero.

El sistema permite que se seleccione la variable o campo con mayor densidad de ocurrencia dentro de los grupos. Al seleccionar el sistema establece el o los grupos que tienen una mayor densidad de ocurrencia de la variable seleccionada coloreándolos con un tono más intenso.

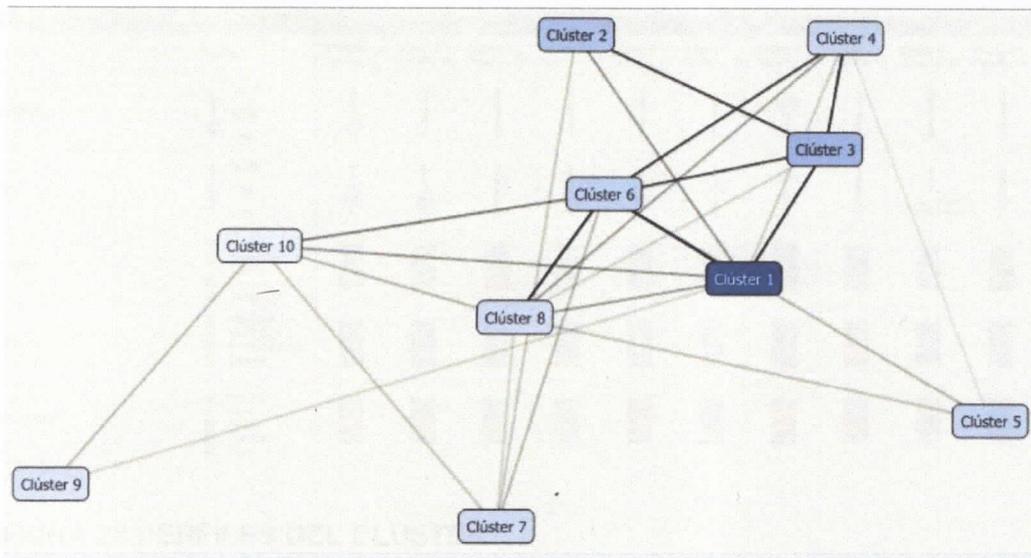


FIGURA 27. DIAGRAMA DE CLUSTER

Las variables que intervienen en el proceso de segmentación que se lograron identificar son: productos, tipo de negocios, municipio o zona y el día del mes. A continuación se realizara una interpretación o análisis de cada una de las variables para trata de encontrar patrones de comportamiento de estas dentro los diversos grupos obtenidos.

Luego de identificar el o los grupos que poseen la mayor densidad de la variable objeto del análisis a través del diagrama de clúster (FIGURA 27), procederemos a la visualizar los detalles que componen el grupo a través de la pestañas perfiles de clúster (FIGURA 28) y características del clúster (FIGURA 29)

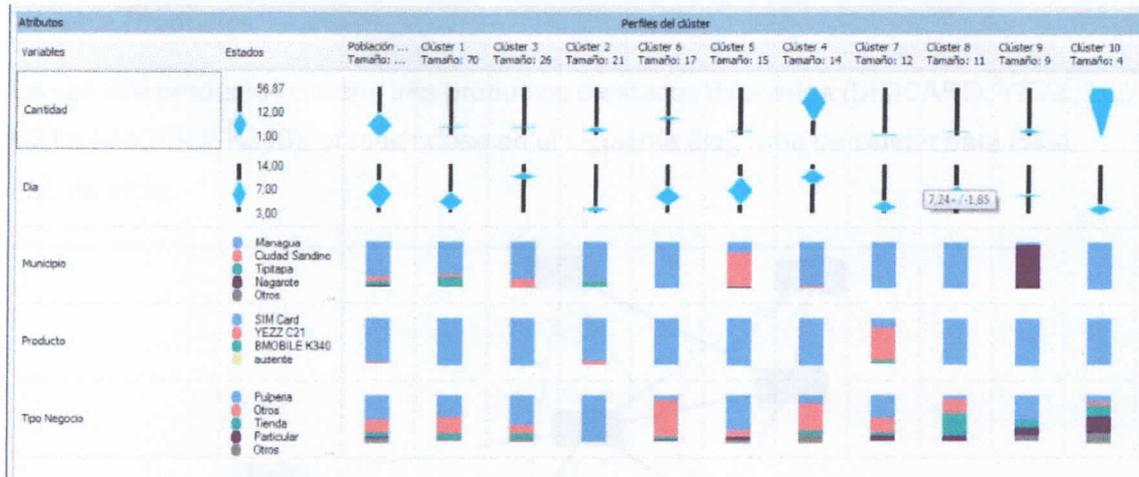


FIGURA 28. PERFILES DEL CLUSTER

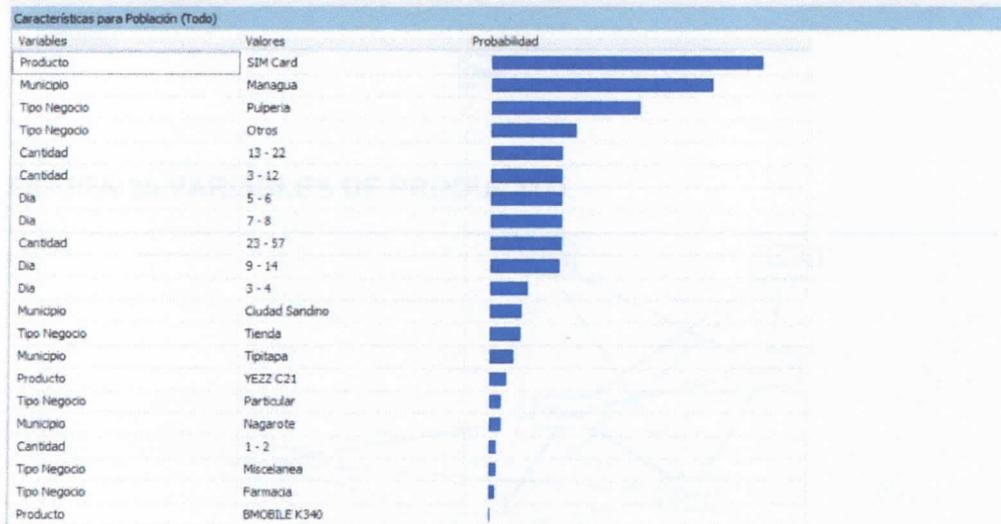


FIGURA 29. CARACTERÍSTICAS PARA POBLACION

Variable Productos

La variable producto contiene tres productos o estados diferentes (SIMCARD, YEZZ C21 y BMOBILE K340), obteniéndose en el siguiente diagrama de clúster para cada uno de ellos:

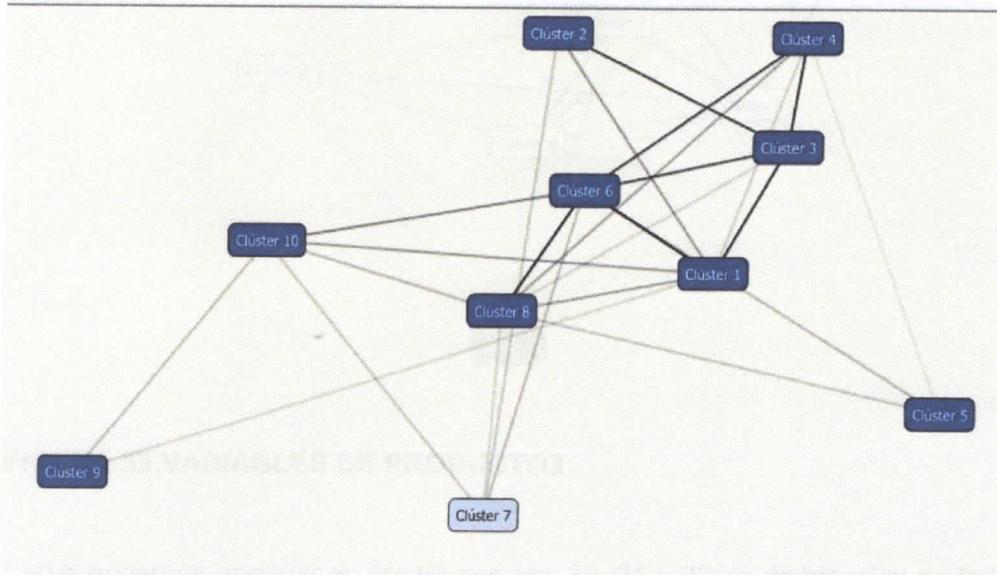


FIGURA 30.VARIABLES DE PRODUCTO1

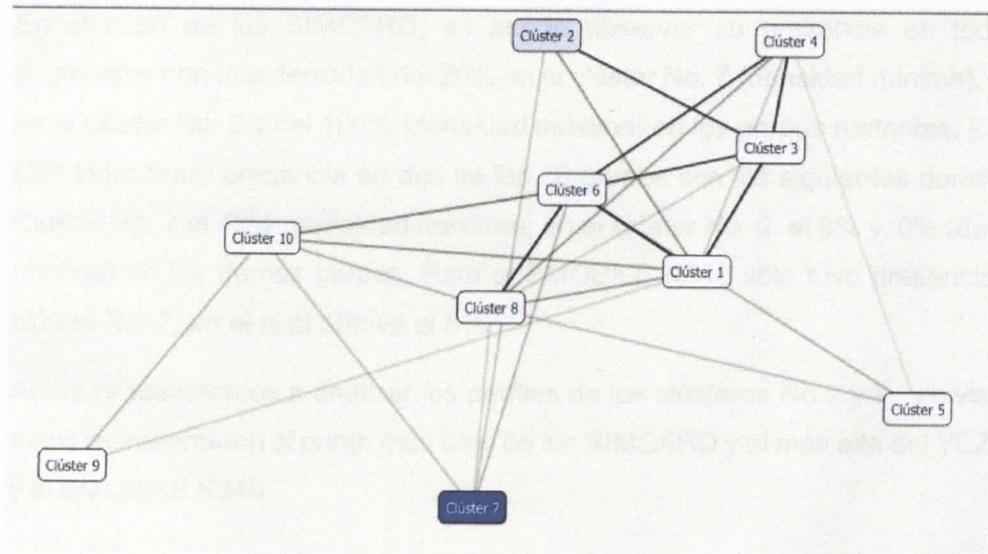


FIGURA 31.VARIABLES DE PRODUCTO2

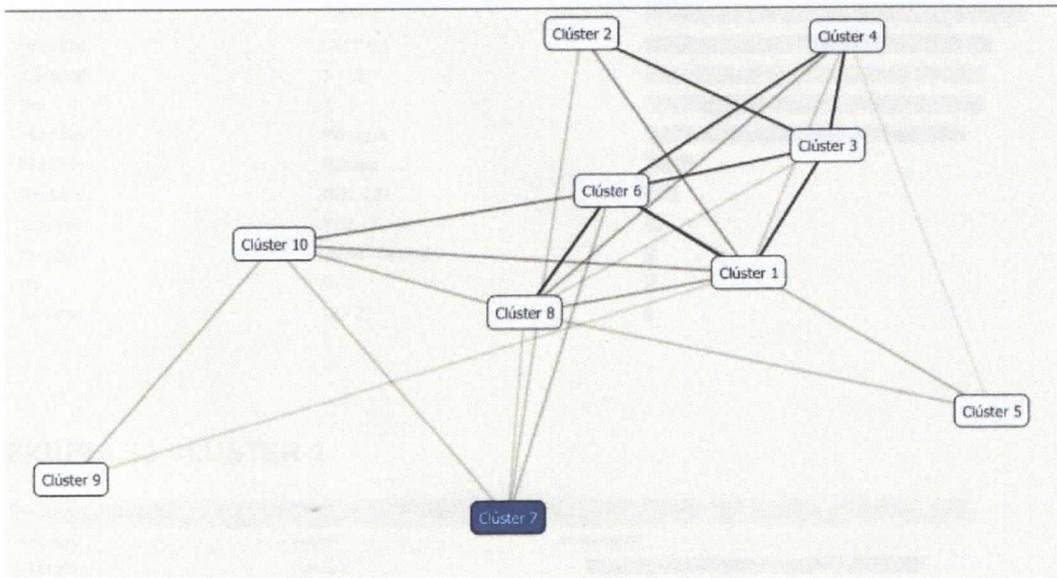


FIGURA 32.VARIABLES DE PRODUCTO3

Como podemos apreciar en las figuras No. 30, 31 y 32, la distribución de la venta de los productos es muy desigual.

En el caso de los SIMCARD, se puede observar su presencia en todos los segmentos con una densidad del 20% en el clúster No. 7 (densidad mínima), el 91% en el clúster No. 2 y del 100% (densidad máxima) en los grupos restantes. El YEZZ C21 logro tener presencia en dos de los 10 grupos con las siguientes densidades: Clúster No. 7 el 72% (densidad máxima), en el clúster No. 2 el 9% y 0% (densidad mínima) en los demás grupos. Para el BMOBILE K340 solo tuvo presencia en el clúster No. 7, en el cual obtuvo el 8%.

Ahora procederemos a analizar los perfiles de los clústeres No.2 y 7, en vista que estos representaron el punto más bajo de los SIMCARD y el más alto del YEZZ C21 y el BMOBILE K340.

1. Trazar campañas de negocios en busca de posicionar cada uno de los productos en el segmento de mercado en que se ubicaron
2. Desarrollar estrategias de mercado tales como promociones en pro de lograr la aceptación de los productos de menor demanda en los demás segmentos.
3. Retirar del mercado aquellos producto cuya demanda sea demasiada baja o poco aceptada por los clientes (es el caso del BMOBILE K340)

Variable Tipos de negocios

Con la inclusión de la variable tipos de negocios dentro del análisis, la empresa pretende reconocer a sus potenciales clientes en dependencia del tipo de negocio que estos posean, para de esta manera orientar a sus ejecutivos de ventas la visita o no de estos.

Actualmente, la empresa ha identificado siete diversos tipos de negocios (farmacias, tiendas o kioscos, distribuidora, miscelánea, pulpería, particular y otros), los cuales son visitados día a día por los ejecutivos de venta que se movilizan por las diversas zonas a bordo de sus motocicletas. Debido a la gran cantidad de valores que contiene esta variable, analizaremos una por una.

Las pulperías se lograron presenciar en todos los grupos (ver Figura 35), obteniendo valores que van desde el 8% (densidad mínima) para el clúster No. 8, hasta el 100% (densidad máxima) que se obtuvo en el clúster No. 2.

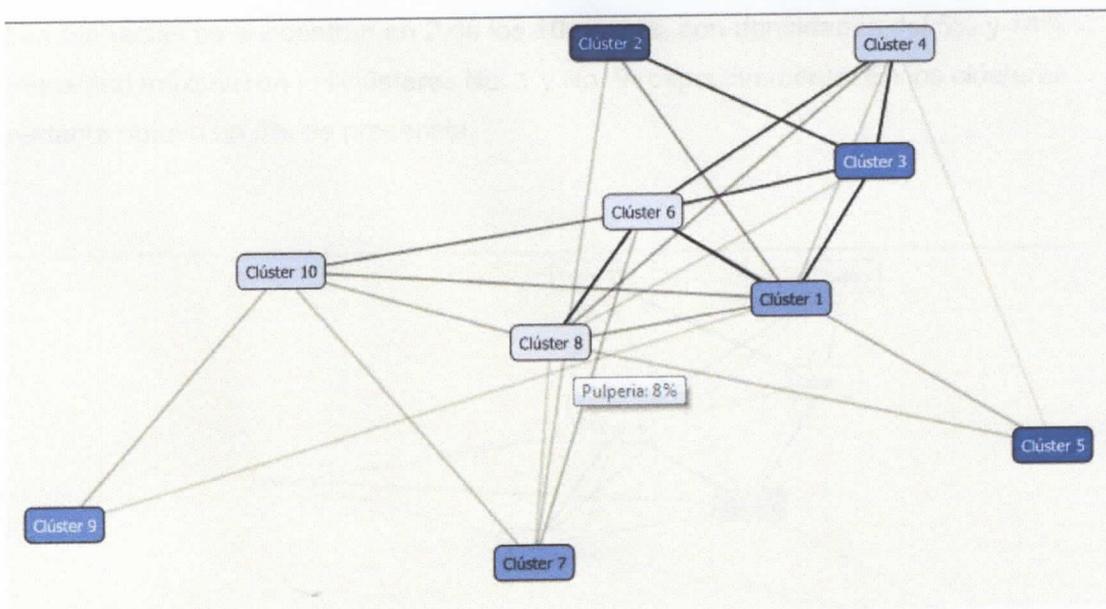


FIGURA 35. CLUSTER PULPERIAS.

En el caso de las tiendas o kioscos, se logró observar su presencia en casi la totalidad de los conglomerados a excepción del clúster No. 2, donde obtuvo un 0%. Aunque las tiendas se lograron ubicar en 9 de 10 grupos, solo logro alcanzar un máximo de densidad del 48% en el clúster No. 8 (ver figura 36. cluster tienda)

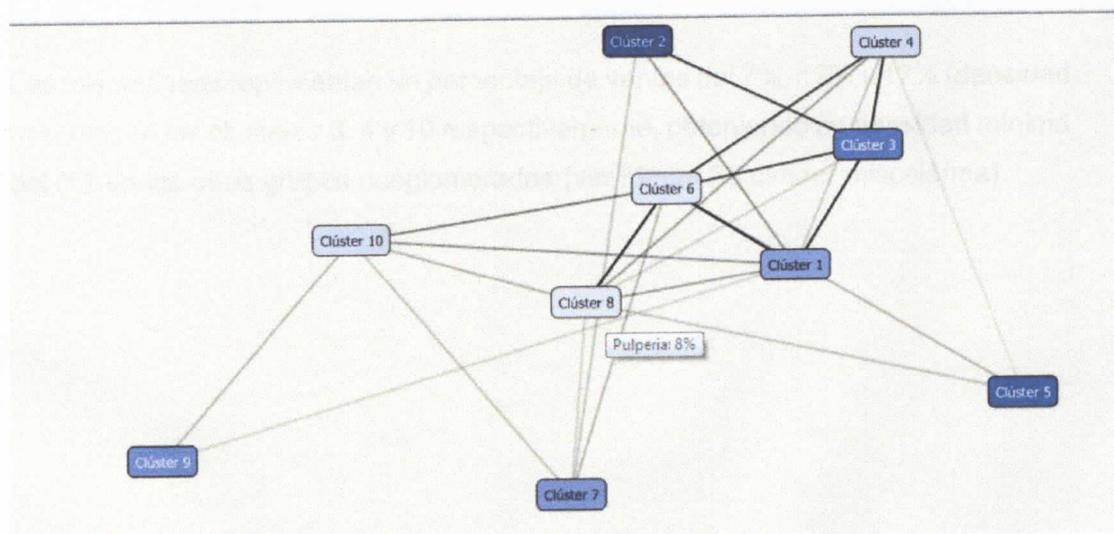


FIGURA 36. CLUSTER TIENDA

Las farmacias se encuentran en 2 de los 10 grupos, con densidades del 5% y 14% (densidad máxima) en los clústeres No. 1 y No. 9 respectivamente. En los clústeres restante obtuvo un 0% de presencia.

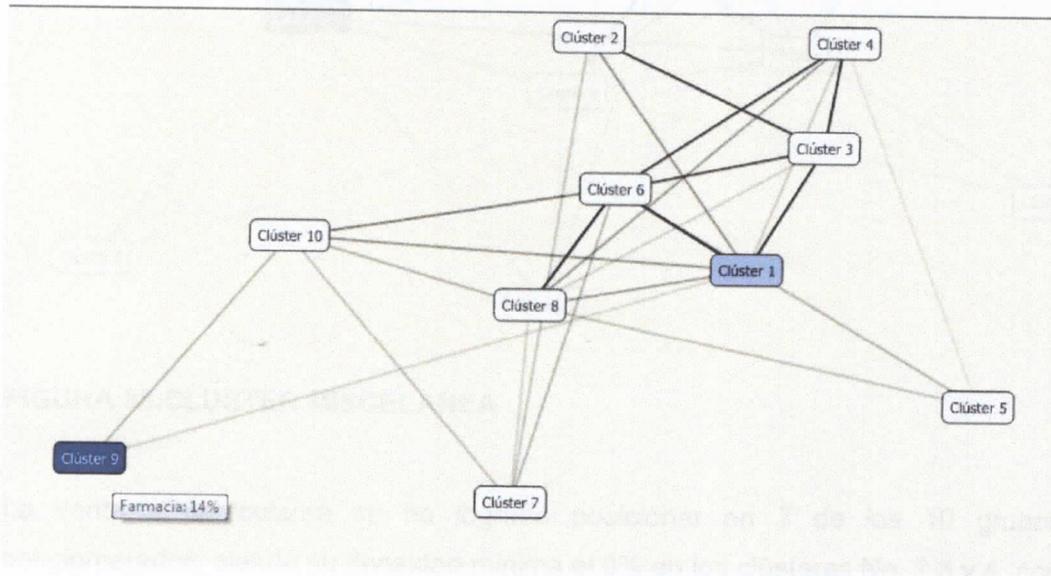


FIGURA 37. CLUSTER FARMACIA

Las misceláneas representan un porcentaje de ventas del 7%, 12% y 17% (densidad máxima) en los clústeres 3, 4 y 10 respectivamente, obteniendo su densidad mínima del 0% en los otros grupos conglomerados (ver Figura 38. cluster miscelánea).

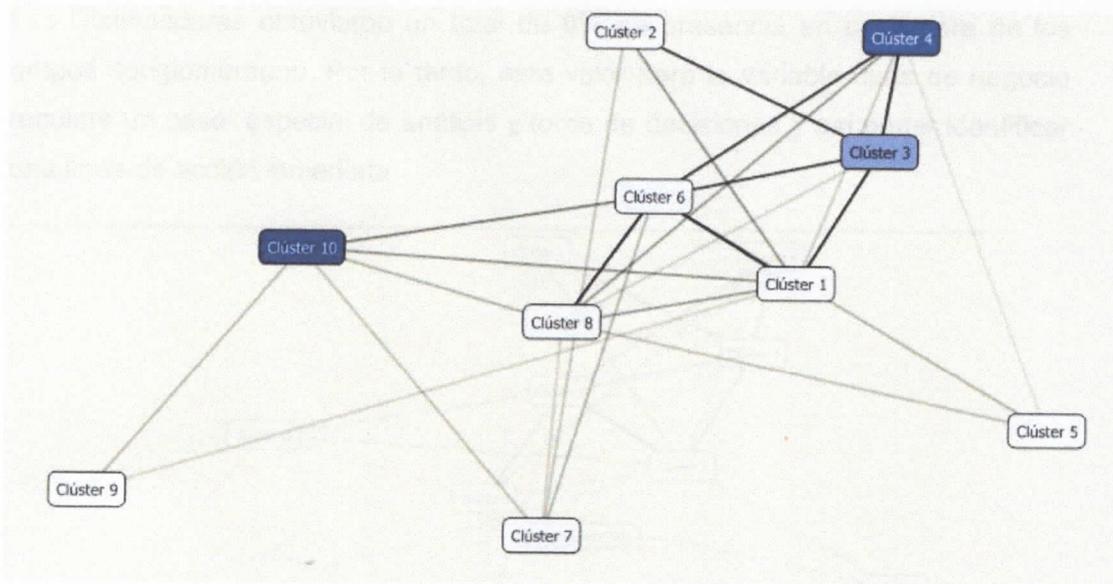


FIGURA 38. CLUSTER MISCELANEA

La venta a particulares se ha logrado posicionar en 7 de los 10 grupos conglomerados, siendo su densidad mínima el 0% en los clústeres No. 2,3 y 4, con una máxima densidad del 34% en el clúster No.10.

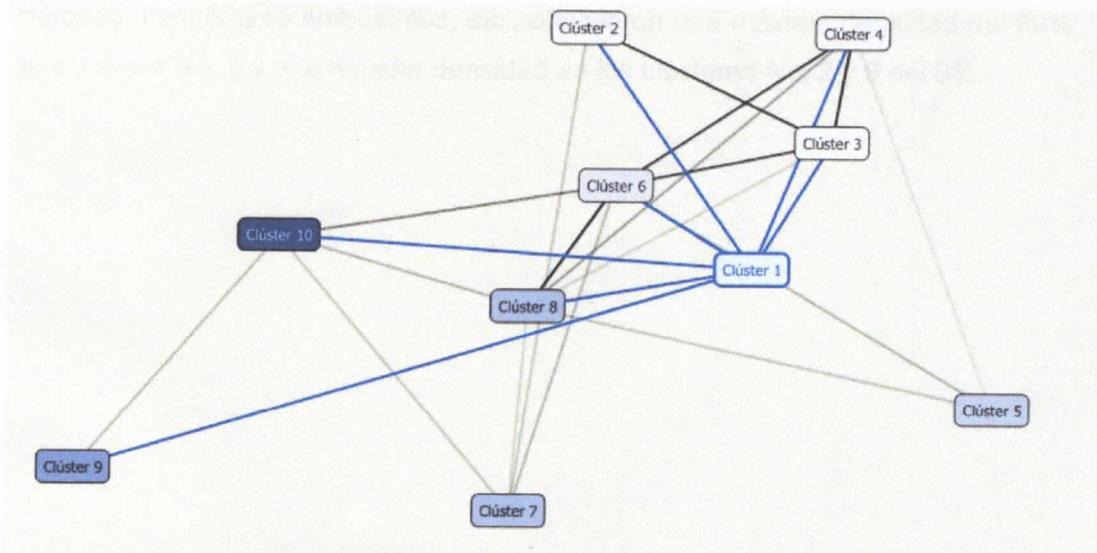


FIGURA 39. VENTAS PARTICULARES

Las Distribuidoras obtuvieron un total de 0% de presencia en cualquiera de los grupos conglomerados. Por lo tanto, este valor para la variable tipos de negocio requiere un caso especial de análisis y toma de decisiones y así poder identificar una línea de acción inmediata.

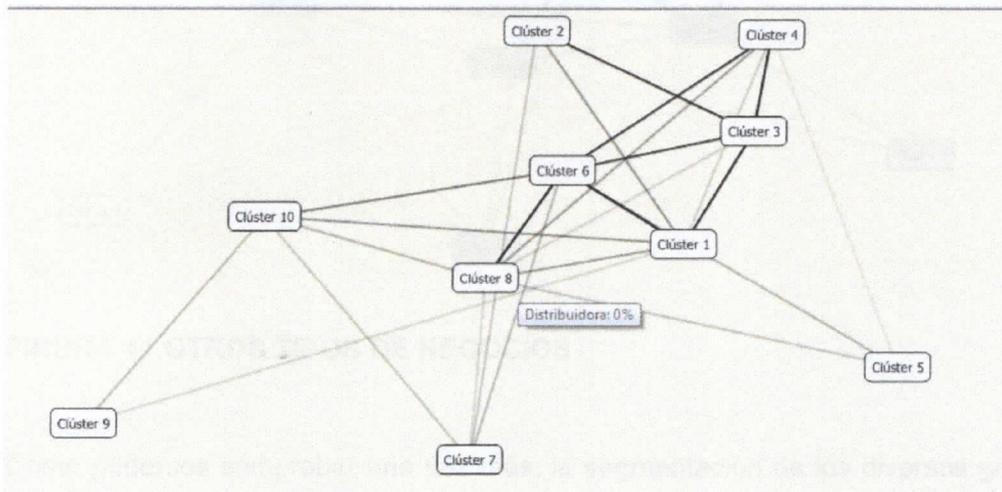


FIGURA 40. VENTAS

El valor Otros, que representa aquellos tipos de negocios que no estén incluidos en los catálogos suministrados por la empresa tales como: librerías, tramos de mercado, vendedores ambulantes, etc., obtuvieron una máxima densidad del 80% en el clúster No. 6 y una mínima densidad en los clústeres No. 2 y 9 del 0%.

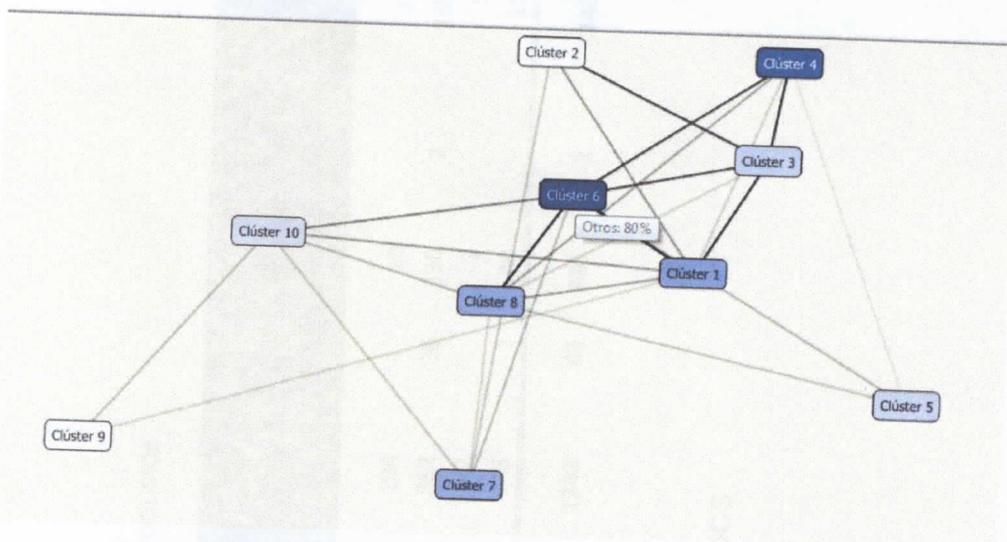


FIGURA 41. OTROS TIPOS DE NEGOCIOS

Como podemos comprobar una vez más, la segmentación de los diversos grupos de mercado ha contribuido a identificar problemas y bondades que ocurren durante el proceso de ventas y los cuales se encontraban ocultos dentro de nuestros almacenes de datos.

INFORMES EN EXCEL.

Salida de datos en Excel.

1. Cantidad de productos vendido en los municipios por tipo de negocios y productos.

Ciudad	Distribuidora		Farmacia		Miscelánea		Otros		Particular		Pulpería		Tienda		Total general	
	SIM Card	SIM Card	SIM Card	SIM Card	SIM Card	YEZZ C21	BMOBILE K340	SIM Card	SIM Card	YEZZ C21	YEZZ C21	SIM Card	YEZZ C21	YEZZ C21		
Sandino		5				45					194		10		254	
Managua	5	47	200	5	1	181					862		36	330	2	2890
Nagarote		5					8				42		3			58
Tipitapa						60		10			145		7	40		262
Total general	5	57	200	5	1	1326	5	199	1	1243	43	383	2	3464		

TABLA 1. CANTIDAD PRODUCTOS VENDIDOS

2. Cantidad de productos vendido por ejecutivo de venta por tipo de negocio y producto.

	Distribuidora		Farmacia		Miscelanea		Otros		Particular		Pulperia		Tienda		Total general
	SIM Card	SIM Card	SIM Card	SIM Card	SIM Card	SIM Card	YEZZ C21	BMOBILE K340	SIM Card	SIM Card	SIM Card	YEZZ C21	SIM Card	YEZZ C21	
Allan Jose												120		70	190
Andy Argenis				40	165	3						180			388
CristoferLerrom					10										10
David			15												
Eddy Ivan					35	2				31		195	2	100	2
Francisco Ramón				160	270							97	28		162
Jeffry										5					430
Jose Antonio			10		280							35		65	5
Jose de la cruz	5				36					140		89		60	390
Luis					60					10		145	7	40	330
Maria Luisa			10		55					5		61	6	35	262
Patricia			12		370				1			85			172
Walter			10		45					8		236		13	468
Total general	5	57	200	1326	5	1	199	1243	43	383	2	3464	2	3464	

TABLA 2. PRODUCTOS VENDIDOS SEGUN TIPO NEGOCIO

3. Mejores Clientes

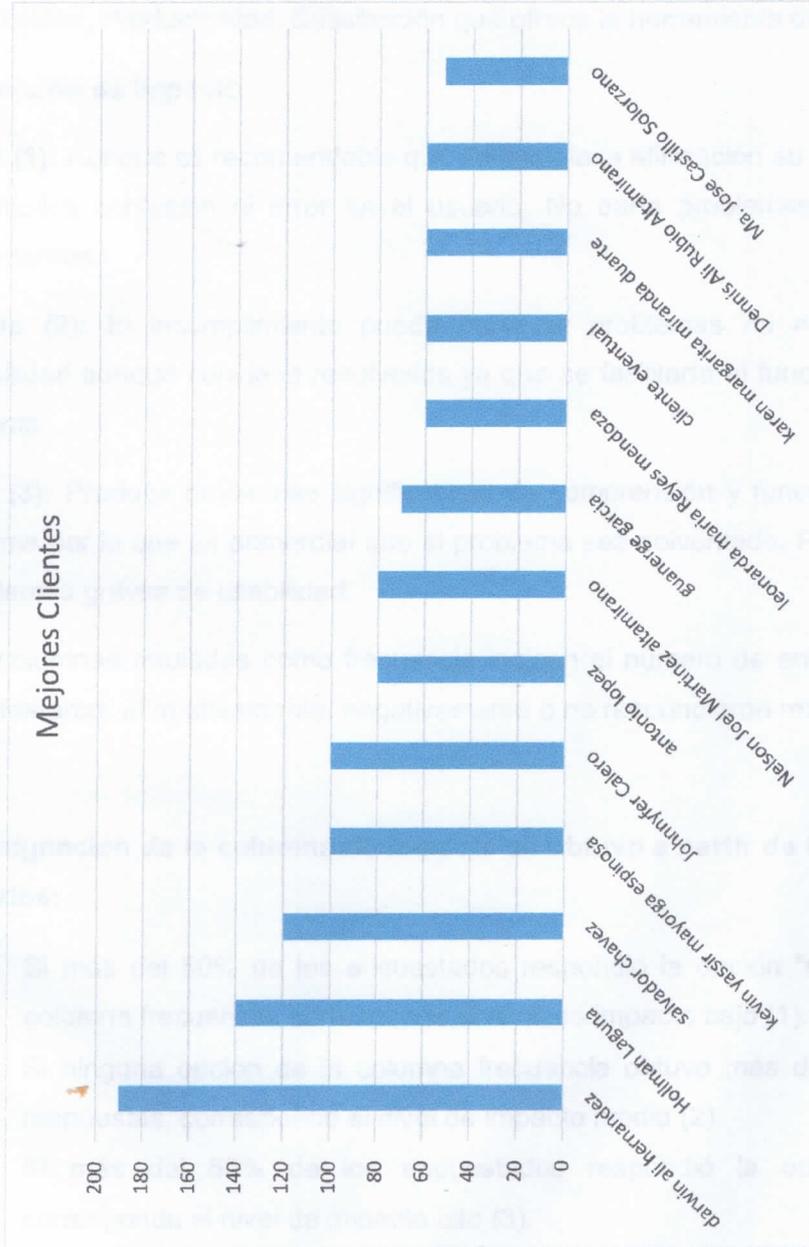


FIGURA 42. MEJORES CLIENTES



EVALUACION DE LA CALIDAD DE USO DE LA HERRAMIENTA DE INGENIERIA DE DATOS POR PARTE DEL USUARIO FINAL

Se dan a conocer los resultados obtenidos de la entrevista realizada al usuario final sobre la herramienta de minería de datos luego de una semana de la capacitación.

La estructura de la entrevista consta de tres partes divididas en tres tablas de manera que permitiera conocer el grado de satisfacción del usuario en cuanto a la Efectividad, Productividad, Satisfacción que ofrece la herramienta de minería.

Definición de impacto

Bajo (1): Aunque es recomendable que se cumpla la afirmación su incumplimiento no implica confusión ni error en el usuario. No daría problemas de usabilidad importantes.

Medio (2): El incumplimiento puede provocar problemas no muy graves de usabilidad aunque conviene resolverlos ya que se facilitaría el funcionamiento del sistema.

Alto (3): Produce problemas significativos de comprensión y funcionalidad en el sistema por lo que es primordial que el problema sea solventado. Puede provocar problemas graves de usabilidad.

Las columnas rotuladas como frecuencia indican el número de encuestados que manifestaron: afirmativamente, negativamente o no respondieron respectivamente.

La asignación de la columna de impacto se obtuvo a partir de los siguientes criterios:

- Si más del 50% de los encuestados respondió la opción "siempre" en la columna frecuencia, corresponde el nivel de impacto bajo (1).
- Si ninguna opción de la columna frecuencia obtuvo más del 50% de las respuestas, corresponde el nivel de impacto medio (2).
- Si más del 50% de los encuestados respondió la opción "nunca", corresponde el nivel de impacto alto (3).

Selección de los usuarios

La selección de usuarios es un elemento fundamental en el proceso de evaluación, la muestra se compuso de 6 participantes.

Efectividad:

Tabla 3.EFECTIVIDAD

Criterios de uso	Frecuencia			
	Siempre	Nunca	No respondió	Impacto
Se identifican fácilmente las figuras, las tablas, las zonas activas y el tipo de acción que se debe ejecutar	3	3	-	2
Los mensajes emitidos son fáciles de entender	1	5	-	3
El mensaje de error propone una solución	-	6	-	3
Los tipos y tamaños de letra son legibles y distinguibles	5	1	-	1
El lenguaje está más cercano al utilizado por el usuario que al informático o técnico.	4	2	-	1
La herramienta ayuda a recaudar información relevante	5	1	-	1

Productividad:

Tabla 4.PRODUCTIVIDAD

Criterios de uso	Frecuencia			
	Siempre	Nunca	No Respondió	Impacto
Facilidad de Aprendizaje				
No se necesita recurrir al desarrollador por ayuda en caso de no recordar el uso de la herramienta	5	1	-	1
Es fácil de recordar como ejecutar una tarea en la herramienta	3	3	-	2
La herramienta está diseñada para necesitar el mínimo de ayuda y de instrucciones.	3	3	-	2
Desde cualquier punto un usuario puede salir de la herramienta	5	-	-	1

Satisfacción:

Tabla 5. SASTIFACCION

Criterios de Uso	Frecuencia			
	Siempre	Nunca	No Respondió	Impacto
Facilidad de Aprendizaje				
La herramienta posee diseño agradable	5	-	-	1
La herramienta cumple todas las tareas que fueron planteadas	5	-	1	1
Es amigable con el usuario (fácil uso)	6	-	-	1
Ahorra tiempo en el proceso de creación de reportes	4	-	2	1

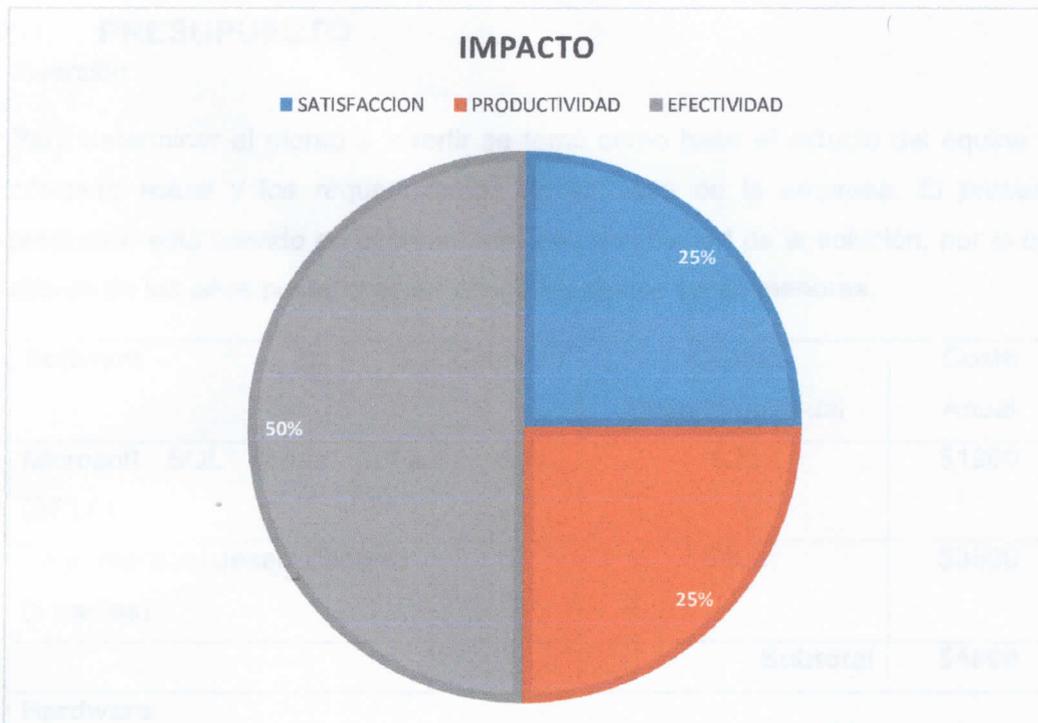


FIGURA 43. GRÁFICO DE EVALUACIÓN FINAL DEL USUARIO

TABLA 3. PRESUPUESTO

XIII. PRESUPUESTO

Inversión

Para determinar el monto a invertir se tomó como base el estudio del equipo de cómputo actual y los requerimientos de software de la empresa. El presente presupuesto está basado en el primer año de operatividad de la solución, por lo que diferirá de los años posteriores en donde los costos serán menores.

Software	Cantidad	Costos Unitario/mensual	Costo Anual
Microsoft SQL server 2012 (SPLA)	5	\$20	\$1200
Pago mensual desarrolladores (6 meses)	3	\$300	\$3600
		Subtotal	\$4800
Hardware			
máquinas de escritorio	5	\$550	\$2750
escritorios para pc	5	\$70	\$350
		Subtotal	\$3100
Costos fijos mensuales			
Infraestructura (agua, luz, etc.)	N/A	\$100	\$1200
Comunicación(internet, telefonía fija y móvil)	N/A	\$400	\$4800
Suministros varios	N/A	\$250	\$3000
Documentación	N/A	\$100	\$1200
		Subtotal	\$10200
Costo total (1er año)			\$18100

TABLA 6. PRESUPUESTO

XIV. CRONOGRAMA DE ACTIVIDADES

Nombre de tarea	Duración	Comienzo	Fin	Encargado Actividad
Gestion del proyecto	10 días	vie 08/05/15	jue 21/05/15	
Primer contacto DISPROCEL (propuesta)	1 día	vie 08/05/15	vie 08/05/15	Julio César Huete
Visita a las oficinas DISPROCEL	1 día	sáb 16/05/15	sáb 16/05/15	Julio César Huete
Visita a las oficinas DISPROCEL	1 día	jue 21/05/15	jue 21/05/15	Julio César Huete
Análisis	30 días	vie 22/05/15	jue 02/07/15	
Recopilación de requerimientos	15 días	vie 22/05/15	jue 11/06/15	Julio César Huete, Leyla Palma
Revisión cuestionarios y datos recopilados	5 días	jue 12/06/15	jue 18/06/15	Leyla Palma, Marbell Acuña
Verificación procesos operativos DISPROCEL	1 día	vie 19/06/15	vie 19/06/15	Julio César Huete, Marbell Acuña
Validación requerimientos funcionales	10 días	sáb 20/06/15	jue 02/07/15	Julio Cesar Huete, Leyla Palma, Marbell Acuña
Diseño	15 días	vie 03/07/15	jue 23/07/15	
Análisis Algoritmo minería de datos a utilizar	5 días	vie 03/07/15	jue 09/07/15	Leyla Palma, Marbell Acuña
Análisis BD transaccional	5 días	vie 10/07/15	jue 16/07/15	Julio Cesar Huete
Análisis DataMart DISPROCEL	3 días	vie 17/07/15	mar 21/07/15	Leyla Palma, Marbell Acuña
Análisis herramientas de minería de datos a utilizar	2 días	mié 22/07/15	jue 23/07/15	Leyla Palma, Marbell Acuña
Desarrollo del proyecto	30 días	sáb 03/08/15	jue 10/09/15	
Generación acceso a datos BD transaccional	1 día	lun 24/08/15	lun 24/08/15	Julio César Huete
Proceso ETL (Integration Services SQL SERVER)	10 días	mar 25/08/15	lun 07/09/15	Julio César Huete, Leyla Palma, Marbell Acuña
Creación de vista de datos y dimensiones (Inicio proyecto Analysis Services)	7 días	sáb 08/08/15	lun 17/08/15	Leyla Palma, Marbell Acuña
Generación del cubo y primeras pruebas con los datos	5 días	mar 18/08/15	lun 24/08/15	Julio César Huete, Leyla Palma, Marbell Acuña
Aplicación algoritmo Clustering Microsoft	5 días	mar 25/08/15	lun 31/08/15	Julio César Huete, Leyla Palma, Marbell Acuña
Pruebas y Documentación	15 días	mar 01/09/15	lun 21/09/15	
Análisis resultados	2 días	mar 01/09/15	mié 02/09/15	Julio César Huete, Leyla Palma, Marbell Acuña
Validación resultados mediante pruebas con datos seleccionados	2 días	jue 03/09/15	vie 04/09/15	Julio César Huete

XV. CONCLUSIONES

La información obtenida a través de los métodos de recolección empleados permitió analizar de manera eficaz el proceso principal que se realiza en las áreas de mercadeo y ventas de DISPROCEL dando como resultado una descripción detallada de los requerimientos.

Comprendidos los requerimientos se diseñó una solución de minería aplicando el algoritmo de clúster que permite al usuario obtener una serie de información, que es de suma importancia para la empresa ya que se pueden tomar decisiones estratégicas de manera acertada logrando obtener más clientes y por lo tanto un aumento en venta de productos.

Una vez finalizada la solución de minería se evaluó con el usuario final en donde se valoró la usabilidad de la misma a sí mismo la efectividad, productividad y el grado de satisfacción del usuario, lo que nos permitió constatar que la solución cumple con los requerimientos y ha ayudado en los procesos de toma de decisiones de la empresa.

VII. BIBLIOGRAFIA

XVI. RECOMENDACIONES

- Los usuarios que sean asignados a operar la solución deberán tener conocimiento sobre minería de datos.
- Brindar capacitaciones a los usuarios encargados sobre la aplicación del modelo de clúster.
- Realizar el proceso de limpieza de datos cada mes durante los periodos de corte de la empresa o bien cuando se crea necesario según solicitud de las áreas involucradas.

XVII. BIBLIOGRAFIA

- Eibe, L. H. (2005). *Data mining*. San Francisco: Morgan Kaufmann.
- Fernando, M. R. (2006). *Sistemas gestores de bases de datos*.
- Forouzan, B. (2006). *Introducción a la ciencia de la computación*. Mexico: International Thomson Editores.
- Gomez, A. (1997). *Ingeniería del Conocimiento*. Centro de Estudios Ramon Areces SA.
- Gonzalez, G. (2005). *Procesamiento de base datos masivas*.
- Hernandez, J. (2004). *Introducción a la minería de datos*. Madrid: Pearsons.
- Herrera, R. (2001). *La Gestión del Conocimiento y la Tecnología*. Mexico: SOMECE.
- Idalberto, C. (2006). *Introducción a la teoría General de la Administración*.
- INMON, W. H. (1996). *Building the Data Warehouse*.
- Kendall, K. E. (2007). *Análisis y diseño de sistemas*. Mexico: Pearson educación.
- Knight, B. (2012). *SQL SERVER 2012--INTEGRATION SERVICES*.
- L.Alonzo, P. (2006). *Ciencias Exactas y Glimensuras*.
- Lewis William, J. (2001). *Data Warehouse and E/Comerce*. Prentice Hall.
- López, P. (2007). *Pérez López & Santín González*.
- Microsoft. (2012). <http://msdn.microsoft.com>. Obtenido de Sql Server Analysis Services.
- Peña, D. (2002). *Análisis de Datos Multivariantes*. Mc Grawhill.
- PEREZ., S. Y. (2005). *Minería de datos Técnicas y herramientas*. España.
- Pineda, E. B. (2008). *Metología d la Investigación*. Washington.
- PIURA, D. J. (2006). *Metodología de la Investigación*.
- Sampieri, R. H. (2010). *Metodología de la Investigación*. MCGRAW-HILL.
- Sánchez S, A. M. (2005). *Minerías de datos y Características*.
- Senn., J. A. (1990). *Sistemas de Información para la Administración*. Grupo Editorial Iberoamericana.
- Sommerville, I. (2005). *Ingeniería del software*. Madrid: Pearson educación s.a.
- Theodoulidis, K. &. (2003). *Visual data mining modeling techniques*.

XVIII. COMPENDIOS

Cuestionario para el análisis y la implementación de un modelo de minería de datos

Nombre y Apellidos: _____

Cargo: _____ Fecha: _____ Hora: _____

El presente cuestionario tiene como objetivo recopilar información operativa de la empresa DISPROCEL con el fin de realizar un análisis acerca de la viabilidad de implementar un modelo de minería de datos. Sus respuestas deben ser lo más precisas posibles.

1. Se encuentra en constante contacto con la toma de decisiones de la empresa.
 - a. Si
 - b. No
 - c. No Aplica

2. Seleccione los problemas que se presentan en su departamento o área de trabajo.
 - a. Falta de acceso a la información
 - b. Periodos largos de espera por la información solicitada.
 - c. La información recibida es confusa y poco entendible
 - d. Es necesario volver a procesar la información recibida Para obtener los datos que se solicitan.
 - e. Los datos obtenidos no apoyan la toma de decisiones A nivel de su departamento o área de trabajo.

3. ¿Con que frecuencia se presentan estos inconvenientes?
- a. Todos los días
 - b. Semanal
 - c. Quincenal
 - d. Mensual
 - e. Raras veces
4. ¿Existe alguna preclasificación o categoría para las siguientes entidades del negocio?
- a. Clientes
 - b. Productos
 - c. Negocios
 - d. Vendedores
 - e. Otros
 - f. Ninguno de los anteriores
5. ¿De los siguientes conceptos, cuales le son familiares?
- a. Datamining
 - b. Datamart
 - c. Datawarehouse
 - d. Minería de datos
 - e. Limpieza de datos
 - f. Integración de datos
 - g. Segmentación/clustering/ de datos

6. ¿En caso de estar familiarizado con el concepto de minería de datos, piensa usted es necesario la implementación de un modelo de minería de datos en la empresa?

- a- Si
- b- No
- c- No sabe

7. ¿En caso de estar familiarizado con el concepto de minería de datos, piensa que existen las condiciones básicas necesarias para llevarlo a cabo?

- a. Si
- b. No
- c. No sabe

8. ¿En caso que el proyecto se realizara, estaría de acuerdo con el mismo?

- a. Si
- b. No

9. Enumere 5 beneficios que espera sean cubiertos por la solución de minería de datos que se pretende realizar.

- a. _____
- b. _____
- c. _____
- d. _____
- e. _____