



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

**VICERRECTORADO DE INVESTIGACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA
DIRECCIÓN DE POSTGRADOS**

**TRABAJO DE TITULACIÓN, PREVIO A LA OBTENCIÓN DEL
TÍTULO DE MAGISTER EN GESTIÓN DE SISTEMAS DE
INFORMACIÓN E INTELIGENCIA DE NEGOCIOS**

**TEMA: “ANÁLISIS DE TENDENCIAS Y DESCUBRIMIENTO
DE PATRONES DE COMPORTAMIENTO DE VENTAS PARA
LA EMPRESA GARWATER UTILIZANDO ALGORITMOS Y
TÉCNICAS DE MINERÍA DE DATOS”**

AUTOR:

ING. DELGADO CALERO, LUIS RICARDO

DIRECTOR:

ING. ECHEVERRÍA BRIONES PEDRO FABRICIO MSG

SANGOLQUÍ - ECUADOR

2017



DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
CARRERA DE MAESTRÍA EN SISTEMAS DE GESTIÓN DE LA
INFORMACIÓN E INTELIGENCIA DE NEGOCIOS

CERTIFICACIÓN

Certifico que el trabajo de titulación, “ANÁLISIS DE TENDENCIAS Y DESCUBRIMIENTO DE PATRONES DE COMPORTAMIENTO DE VENTAS PARA LA EMPRESA GARWATER UTILIZANDO ALGORITMOS Y TÉCNICAS DE MINERÍA DE DATOS” realizado por el señor LUIS RICARDO DELGADO CALERO, ha sido revisado en su totalidad y analizado por el software anti-plagio, el mismo cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, por lo tanto me permito acreditarlo y autorizar al señor LUIS RICARDO DELGADO CALERO para que lo sustente públicamente.

Sangolquí, 24 de marzo del 2017

MSG. PEDRO FABRICIO ECHEVERRÍA BRÍONES

DIRECTOR



**DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
CARRERA DE MAestrÍA EN SISTEMAS DE GESTIÓN DE LA
INFORMACIÓN E INTELIGENCIA DE NEGOCIOS**

AUTORÍA DE RESPONSABILIDAD

Yo, **LUIS RICARDO DELGADO CALERO**, con cédula de identidad N° 131289285-2 declaro que este trabajo de titulación “**ANÁLISIS DE TENDENCIAS Y DESCUBRIMIENTO DE PATRONES DE COMPORTAMIENTO DE VENTAS PARA LA EMPRESA GARWATER UTILIZANDO ALGORITMOS Y TÉCNICAS DE MINERÍA DE DATOS**” ha sido desarrollado considerando los métodos de investigación existentes, así como también se ha respetado los derechos intelectuales de terceros considerándose en las citas bibliográficas.

Consecuentemente declaro que este trabajo es de mi autoría, en virtud de ello me declaro responsable del contenido, veracidad y alcance de la investigación mencionada.

Sangolquí, 24 de marzo del 2017

LUIS RICARDO DELGADO CALERO

C.C. 1312892852



**DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
CARRERA DE MAESTRÍA EN SISTEMAS DE GESTIÓN DE LA
INFORMACIÓN E INTELIGENCIA DE NEGOCIOS**

AUTORIZACIÓN

Yo, **LUIS RICARDO DELGADO CALERO**, No autorizo a la Universidad de las Fuerzas Armadas ESPE publicar en la biblioteca virtual de la institución el presente trabajo de titulación “**ANÁLISIS DE TENDENCIAS Y DESCUBRIMIENTO DE PATRONES DE COMPORTAMIENTO DE VENTAS PARA LA EMPRESA GARWATER UTILIZANDO ALGORITMOS Y TÉCNICAS DE MINERÍA DE DATOS**” cuyo contenido, ideas y criterios de mi autoría y responsabilidad.

Sangolquí, 24 de marzo del 2017

LUIS RICARDO DELGADO CALERO

C.C. 1312892852

DEDICATORIA

“Los que dicen que es imposible no deberían molestar a los que lo están haciendo”.

Albert Einstein.

Con mi mayor gratitud dedico el presente trabajo de Tesis:

A mi familia por brindarme su apoyo y buenos deseos durante la maestría, fueron pieza clave para aventurarme a este capítulo de mi vida.

A mis maestros por transmitirme sus experiencias y disponer de su valioso tiempo, comprendiendo el sacrificio y el deber que conlleva ser un educador.

A mis compañeras y compañeros, con los que compartí buenos momentos e intercambios de experiencias y saberes, sin duda trabajar en equipo junto a ustedes fue una grata labor.

Luis Ricardo Delgado Calero

AGRADECIMIENTO

“El agradecimiento es la memoria del corazón”

J.B. Massieu.

Se les agradece de corazón:

Al director de tesis y al Tribunal de Revisión y Evaluación por el apoyo brindado al permitir el desarrollo del presente proyecto.

Al Gerente de la empresa de agua embotellada GARWATER por facilitarme los medios e información requerida para el estudio y desarrollo de la tesis.

Luis Ricardo Delgado Calero

ÍNDICE DE CONTENIDO

ÍNDICE

CERTIFICACIÓN	ii
AUTORÍA DE RESPONSABILIDAD	iii
AUTORIZACIÓN	iv
DEDICATORIA	v
AGRADECIMIENTO	vi
ÍNDICE	vii
ÍNDICE DE TABLAS	x
ÍNDICE DE FIGURAS.....	xi
RESÚMEN.....	xiii
ABSTRACT.....	xiv
CAPÍTULO I.....	1
ASPECTOS GENERALES	1
1. GARWATER S.A.....	1
1.1. Introducción histórica.....	1
1.2. Perfil de la empresa.....	1
1.3. Actualidad de GARWATER S.A.....	2
1.4. Planteamiento del problema.....	3
1.5. Formulación del problema a resolver.....	4
1.5.1. Objetivo General	5
1.5.2. Objetivos Específicos.....	5
1.6. Hipótesis.....	6

CAPÍTULO II	7
MARCO TEÓRICO.....	7
2. Cadena de valor.....	7
2.1. Infraestructura de la empresa.	9
2.2. Gestión de Talento Humano.....	9
2.3. Desarrollo Tecnológico.....	10
2.4. Aprovisionamiento.....	10
2.5. Logística interior / Actividades.....	11
2.6. Lógica Exterior.....	13
2.7. Marketing.....	13
2.8. Servicios.....	14
3. Modelo entidad relación.....	14
4. Modelo multidimensional.	19
5. Modelos de minería de datos.....	22
CAPÍTULO III.....	25
MARCO CONCEPTUAL.....	25
6. Comprensión del negocio.....	25
6.1. Objetivos del negocio.....	25
6.2. Situación actual de GARWATER.....	26
6.3. Objetivos de la Minería de Datos.....	27
6.4. Plan del proyecto de minería.....	27
7. Comprensión de los datos.	30
7.1. Recolección de datos iniciales.	30

7.2. Descripción de los datos.....	31
7.3. Exploración de los datos.	35
7.4. Verificación de los datos.....	37
8. Preparación de los datos.....	38
8.1. Selección de los datos.	38
8.2. Limpieza de los datos.....	38
8.3. Estructura de los datos.	40
8.4. Integración de los datos.....	40
8.5. Formateo de los datos.	40
9. Modelamiento de los datos.....	41
9.1. Selección de técnicas.....	41
CAPÍTULO IV.....	43
MODELO DE INVESTIGACIÓN	43
10. Deducción – inducción.....	43
11. Observación.....	44
12. Análisis.....	45
13. Fuentes de datos primarias.	49
14. Fuentes de datos secundarias.	49
15. Detección de modelos por comparación de las fuentes de datos.	50
CAPÍTULO V.....	54
ANÁLISIS DE RESULTADOS	54
16. Patrones obtenidos.	54
17. Demostración de la hipótesis.	62

18. Aplicación de resultados.....	63
CAPÍTULO VI.....	65
CONCLUSIONES Y RECOMENDACIONES.....	65
19. Conclusiones.....	65
20. Recomendaciones.....	66
ANEXOS.....	68
BIBLIOGRAFÍA.....	73

ÍNDICE DE TABLAS

Tabla 1. Catálogo de entidades del software GENESIS.....	16
Tabla 2. Plan de proyecto de minería de datos.....	28
Tabla 3. Versiones y licencias de PENTAHO utilizadas en el proyecto.....	30
Tabla 4. Técnicas de minería de datos.....	42
Tabla 5. Volumen de ventas desde el 2011 hasta el 2015.....	46
Tabla 6. Mejores meses en ventas desde el 2011 hasta el 2015.....	46
Tabla 7. Recaudación en ventas por producto desde el 2011 hasta el 2015.....	47
Tabla 8. Top 10 de Vendedores.....	47
Tabla 9. Detalle de la gestión en ventas.....	48
Tabla 10. TOP 10 de mejores Clientes.....	48
Tabla 11. Origen / Punto Matriz De Venta.....	48

ÍNDICE DE FIGURAS

Figura 1. Cadena de valor de Porter.....	7
Figura 2. Esquema de producción de GARWATER.	12
Figura 3. Componentes de un Diagrama Entidad / Relación.	15
Figura 4. Diagrama E/R de facturación del software GENESIS.	18
Figura 5. Esquema de un modelo multidimensional de ventas.....	19
Figura 6. Diagrama multidimensional de facturación de GARWATER.	21
Figura 7. Arquitectura de un modelo de minería de datos.	22
Figura 8. Cuadrante de Gartner.....	29
Figura 9. Estructura de tabla ge_empresa.	31
Figura 10. Consulta de la tabla ge_empresa.....	32
Figura 11. Estructura de la tabla in_cabecera.	33
Figura 12. Estructura de la tabla in_movimiento.....	34
Figura 13. Estructura de la tabla cxc_auxiliar.....	34
Figura 14. Estructura de la tabla in_cliente.....	35
Figura 15. Estructura de la tabla in_item.	35
Figura 16. Estadísticas de la Base de Datos del Software GENESIS - Ventas.....	36
Figura 17. Registros almacenados en tablas principales (2011 - 2015).....	36
Figura 18. Triggers, SPs, Views.	37

Figura 19. Flujo de Transformación y Limpieza de Datos en PENTAHO de la tabla IN_CABECERA.	39
Figura 20. Representación del método deductivo en la investigación.	44
Figura 21. Algoritmo de regresión lineal WEKA.	51
Figura 22. Gráfica del Algoritmo Forecasting.	51
Figura 23. Código generado por WEKA – FORECASTING.	53
Figura 24. Análisis del algoritmo Forecasting de WEKA.	54
Figura 25. Contraste de ventas y su mejor mes en el periodo 2011-2015.	55
Figura 26. Mejores compradores en el periodo 2011 – 2015.	56
Figura 27. Mejores vendedores en el periodo 2011 – 2015.	56
Figura 28. Productos de mayor influencia en las ventas de GARWATER.	57
Figura 29. Ponderación de ventas en las plantas de distribución con muestra de mejores clientes.	58
Figura 30. Distribución de ventas desde PLANTA PRINCIPAL.	58
Figura 31. Distribución de ventas desde PLANTA MANTA.	59
Figura 32. Comportamiento de ventas en consumo de productos GARWATER en los cantones de Manabí.	60
Figura 33. Ventas realizadas por clientes en GUAYAS y PICHINCHA.	61
Figura 34. Comparación de artículos más vendidos en función de los mejores vendedores.	62

RESÚMEN

GARWATER S.A. es una empresa embotelladora de agua, considerada por sus clientes como exitosa en la provincia de Manabí. En este contexto la empresa mantiene y genera mes a mes un gran volumen información, la cual fue levemente atendida y valorada desde que la planta inició formalmente sus operaciones, siendo este el caso la administración actual decidió estudiar esta información con el fin de obtener conocimiento de las ventas. Sin embargo, la tarea fue complicada al no existir un área dedicada a la inteligencia de negocios. Esta investigación comprende un desarrollo donde se aplicaron técnicas, herramientas y algoritmos de minería de datos sobre la información de las ventas de la empresa empleando la metodología CRISP – DM. Como solución a la problemática la actual dirigencia permitió realizar un estudio tomando como referencia las ventas comprendidas en el periodo 2011 – 2015, con la finalidad de encontrar conocimiento que permitiera optimizar y fortalecer la toma de decisiones. Los resultados del estudio permitieron cumplir los objetivos de GARWATER S.A., presentando una serie de reportes mostrando la situación actual de la empresa a los directivos, que si bien no mostraba pérdidas económicas si preocupaba la mala gestión de la empresa. Motivo por el cual la aplicación de las técnicas, herramientas y algoritmos de minería de datos fue limitada, gracias a este descubrimiento se plantearon nuevas metas y objetivos en la empresa que junto a las recomendaciones otorgadas por este estudio se permitirá corregir errores y mejorar las ventas en la empresa en futuros años.

Palabras Claves:

- **GAR WATER**
- **DATAMINING**
- **CRISP-DM**
- **FORECASTING**
- **PENTAHO**

ABSTRACT

GARWATER S.A. Is a water bottling company, considered by its clients to be successful in the province of Manabí. In this context, the company maintains and generates a large volume of information month by month, which was slightly attended and valued since the plant formally began its operations, being this the case the current administration decided to study this information in order to obtain knowledge of the sales. However, the task was complicated by the absence of an area dedicated to business intelligence. This research includes a development where techniques, tools and data mining algorithms were applied to the company's sales information using the CRISP - DM methodology. As a solution to the problem, the current leadership allowed a study to be carried out, taking as reference the sales comprised in the period 2011 - 2015, in order to find knowledge that would optimize and strengthen decision making. The results of the study allowed GARWATER S.A.'s objectives to be met, presenting a series of reports showing the current situation of the company to the executives, who did not show economic losses if he was concerned about the company's mismanagement. Reason why the application of techniques, tools and algorithms of data mining was limited, thanks to this discovery were set new goals and objectives in the company that along with the recommendations given by this study will be allowed to correct errors and improve sales In the company in future years.

Keywords:

- **GAR WATER**
- **DATAMINING**
- **CRISP-DM**
- **FORECASTING**
- **PENTAHO**

CAPÍTULO I

ASPECTOS GENERALES

1. GARWATER S.A.

1.1. Introducción histórica.

GARWATER S.A. salió al público un día 14 de junio de 2002, con el objetivo de brindar agua envasada purificada, de gran calidad, con la más alta tecnología, y poder contribuir a mejorar el estándar de vida de la comunidad, conociendo de las muchas falencias y calidad de agua distribuida de la empresa pública como privada. GARWATER S.A. es mucho más que una planta envasadora de agua, es una empresa que contribuye y aporta social y económicamente con la sociedad en general, ya que son los clientes con quienes la empresa se compromete a brindar un servicio de calidad.

1.2. Perfil de la empresa.

El Gerente de GARWATER es Julio Enrique Rivas Beltrón y la Administración está a cargo de la Ing. Bielka Rodríguez. La empresa está consolidada en el mercado manabita y distribuye sus productos en todos sus cantones. La compañía tiene sus oficinas principales y planta en, la parroquia Colón del cantón Portoviejo junto al INIAP en la vía Portoviejo – Santa Ana y junto sus bodegas y sucursales ubicadas en el centro de Portoviejo y en Manta emplean a casi 70 personas. Su misión es ofrecer agua purificada, envasada con los más altos estándares de calidad, cuidando la salud de los consumidores, su visión es mantenerse como primeros abastecedores de agua purificada, conservando calidad e innovando en tecnología para seguir siendo los mejores.

Su compromiso es sinónimo de responsabilidad del presente para alcanzar las metas en el futuro. GARWATER, se compromete con los socios, empleados, proveedores, clientes con toda la comunidad y al mundo al que la empresa sirve, de

actuar con el convencimiento de que la materia prima que no tiene precio en los productos, es la verdad, el honor, la dignidad y la integridad de quien los elabora. La empresa se compromete a tratar a al personal con el respeto, la compensación justa, el tratamiento honesto y equitativo. GARWATER fomentara siempre una cultura empresarial que aliente la excelencia, el liderazgo, la innovación. Los valores corporativos de la empresa son los siguientes:

- Ampliar las redes de distribución y comercialización del producto
- Responsabilidad social y honradez
- Calidad de nuestro producto
- Transparencia y ética profesional
- Independencia
- Respeto mutuo
- Eficacia, trabajo en equipo
- Solidaridad
- Innovación constante

Sus políticas son:

- Toda persona que labore para la entidad debe cuidar su imagen.
- El producto a elaborar será de calidad, contribuyendo al beneficio de la familia, y contando con todas las formas de higiene establecidos.
- Se garantizara la eficiencia y eficacia de nuestros empleados mediante una constante capacitación.
- Se realizaran estudios constantes de las actividades de los empleados
- Toda nueva bebida que la empresa desee lanzar al mercado debe ser elaborada con los más altos estándares de higiene y calidad.

1.3. Actualidad de GARWATER S.A.

GARWATER se enfrenta en la actualidad a los siguientes problemas:

- Mercado ecuatoriano saturado y surgimiento de nuevas marcas bajo el estandarte y apoyo de compañías multinacionales.
- Dificultades en el abastecimiento de algunas materias primas desde el capital.
- Emplea sistemas semi automáticos de producción.
- Poca inversión en recursos informáticos en contraste a las nuevas tendencias y uso de herramientas administrativas utilizando software para gestión de información.

1.4. Planteamiento del problema.

En la actualidad el desarrollo tecnológico y científico en el marco de la información, es palpable en toda su magnitud, desde el nacimiento hasta la muerte de una persona se deja una huella en los archivos digitales de una base de datos; ya sea en una factura o en un registro dental. Suministramos información a empresas, instituciones y gobiernos que de la mejor manera en la mayoría de los casos custodia celosamente nuestros datos para evitar su mal uso. Vivimos en una época donde el país más fuerte dejó de ser el más grande en extensión territorial y el que más conocimiento tenga concentrará más poder. Es así que los gobiernos y las empresas abanderan su poderío innovando y generando conocimiento, el mismo que se nutre segundo a segundo de la información que proporcionamos directa e indirectamente en nuestro diario vivir.

Esta dependencia a la información necesita de grandes almacenes de datos, que contengan volúmenes incalculables de información, mismos que irán incrementando sus registros considerablemente con el pasar del tiempo. Sin embargo ante tal necesidad de almacenar información se han mejorado los motores de bases de datos, se ha optimizado el rendimiento del hardware, se ha incrementado la capacidad de almacenamiento y se ha innovado la forma de como guardamos la información con su sincronización en red.

El uso de la minería de datos combinada con una buena toma de decisiones y administración puede llevar a cualquier empresa a la cúspide del éxito. En Ecuador son muchas las empresas y organizaciones que emplean la minería de datos como una herramienta para la toma de decisiones y el éxito de sus operaciones es producto del buen uso de estas herramientas. Sin embargo existen muchas empresas en diferentes provincias que poseen herramientas informáticas para el desempeño de sus operaciones laborales y que generan grandes volúmenes de información pero no efectúan minería de datos, lo que ocasiona que se estanquen y adopten un estado conformista.

En la provincia de Manabí existe la empresa GARWATER que basa sus actividades comerciales en la producción de agua embotellada y la distribuye en toda la provincia, emplea un software contable desde el año 2009, el cual ha cumplido exitosamente sus funciones hasta la fecha. Sin embargo en los últimos años el volumen de información que posee la base de datos de este software en la empresa ha sido referente de análisis y reportes sin un éxito total, con el fin de mostrar tendencias en las ventas, volumen de entregas, pedidos, rutas de distribución, satisfacción del cliente, marketing, etc.

La empresa se encuentra en una etapa de crecimiento y sus ingresos por facturación superan los cien mil dólares, el análisis de los datos toma un papel crucial en el crecimiento y expansión de la empresa en el futuro, y esto solo se lograra implementando minería de datos para detectar tendencias y patrones de comportamientos que permitan tomar decisiones por el bienestar y prosperidad de la empresa y la provincia.

1.5. Formulación del problema a resolver.

Planteado el problema se procede a generar las siguientes interrogantes:

- ¿La empresa GARWATER necesita implementar minería de datos?

- ¿La metodología de minería de datos seleccionada para el proyecto cumplirá las expectativas de la empresa?
- ¿Las tendencias y patrones descubiertos servirán en la toma de futuras decisiones de la empresa?
- ¿Servirá el estudio de minería de datos para elaborar un plan de negocios que incremente las ventas de la empresa?

1.5. Objetivos

1.5.1. Objetivo General

- Aplicar técnicas y algoritmos de minería de datos para descubrir tendencias y patrones de comportamiento en las ventas de la empresa GARWATER.

1.5.2. Objetivos Específicos

- Formalizar con los directivos de la empresa GARWATER el estudio de su información, socializando la metodología CRISP – DM, las técnicas, herramientas y algoritmos de minería de datos que serán empleados.
- Conseguir permisos de acceso a las fuentes de información propias de las ventas de la empresa GARWATER.
- Crear un entorno OLTP con las ventas de GARWATER desde el año 2011 hasta el año 2015 empleando PENTAHO.
- Encontrar tendencias y patrones en las ventas construyendo dashboards, empleando PENTAHO.
- Socializar los resultados documentados en la investigación, con los directivos más influyentes de la empresa GARWATER.
- Participar en la elaboración de un plan de negocios que incremente las ventas en la empresa en los próximos años.

1.6. Hipótesis

- Es posible encontrar patrones y tendencias en las ventas de la empresa GARWATER, a través de la aplicación de técnicas de Minería de Datos, que permitan además predecir las ventas del año 2016.

CAPÍTULO II

MARCO TEÓRICO

2. Cadena de valor.

La cadena de valor es una herramienta que permite realizar un análisis interno de una empresa, clasificando cada una de las actividades que otorgan valor a la misma. Usando esta herramienta se consigue descubrir mejor las fortalezas y debilidades mediante: la identificación de las fuentes de valor, análisis de las ventajas competitivas y los costos asociados a cada actividad. (CreceNegocios, 2016)

Todas las empresas mantienen una cadena de valor, esta debe estar integrada por dinamismos que van desde diseñar un producto, hasta la distribución del mismo que oferte la empresa. Esta cadena plantea categorías para cada actividad que genere valor en las actividades de línea y las actividades de cuenta como apoyo ofreciendo el soporte necesario, en la figura 1 se observan dichas actividades. (CreceNegocios, 2016)



Figura 1. Cadena de valor de Porter.
Fuente: (CreceNegocios, 2016)

La empresa está consolidada en el mercado de la provincia de Manabí y distribuye sus productos en todos sus cantones. La compañía tiene sus oficinas principales y planta en, la parroquia Colón del cantón Portoviejo junto al INIAP en la vía Portoviejo – Santa Ana y junto sus bodegas y sucursales ubicadas en el centro de Portoviejo y en Manta emplean a casi 70 personas. (Delgado Calero & GARWATER, 2016)

La misión de esta empresa consiste en entregar agua pura, embotellada con los más altos estándares de calidad, protegiendo la salud de sus clientes. En cuanto a su visión la empresa espera ubicarse como líder de abastecedores de agua purificada en la región, manteniendo la calidad sin descuidar su precio accesible al público. Su compromiso representa la responsabilidad que mantiene con sus clientes y empleados en el presente para alcanzar las metas planeadas en el futuro extendiendo el compromiso a los socios, proveedores expandiéndose por mundo al que la empresa sirve.

La empresa actúa convencida de que la materia prima no tiene precio pues el agua es una fuente de vida la cual se debe cuidar y proteger para el futuro de la vida en el planeta. (Delgado Calero & GARWATER, 2016) La empresa está comprometida con el personal cuidando el respeto, la remuneración justa, honesta y equitativa, con el fin de fomentar siempre una cultura empresarial que fomente la excelencia, el liderazgo y la innovación. GARWATER incorpora los siguientes valores corporativos:

- Ampliar las redes de distribución y comercialización del producto
- Responsabilidad social y honradez
- Calidad de nuestro producto
- Transparencia y ética profesional
- Independencia
- Respeto mutuo
- Eficacia, trabajo en equipo

- Solidaridad
- Innovación constante

La empresa embotelladora de agua GARWATER S.A. tiene su cadena de valor que en el contexto de las diferentes actividades que generan valor, estas se especifican detalladamente a continuación:

2.1. Infraestructura de la empresa.

GARWATER utiliza una estructura distribuida en zonas en torno a la provincia de Manabí para llevar a cabo una estrategia de distribución en zonas. La distribución por zonas es un tipo de estrategia en la que las decisiones de la empresa se enfocan en las características poblacionales y demográficas de los sectores en cuestión. La empresa opera utilizando una estructura de dirección consistente en la distribución de sus productos en todos los cantones de la provincia de Manabí tomando como referencia y punto de abastecimiento sus dos bodegas localizadas en Portoviejo y Manta. (Delgado Calero & GARWATER, 2016)

Mediante esta estrategia GARWATER S.A. puede dirigir un negocio de expansión manteniendo un acercamiento multilocal. En el corazón de este acercamiento se sitúa la planta embotelladora, en la cual la empresa produce agua embotellada distribuida a los consumidores, clientes y vendedores independientes manteniendo siempre su margen de venta. (Delgado Calero & GARWATER, 2016)

2.2. Gestión de Talento Humano.

GARWATER tiene un compromiso continuado hacia la formación y la creación de un lugar de trabajo libre de discriminación y fatiga. De esta forma sus empleados pueden proporcionar al producto la calidad deseada. La empresa proporciona el entrenamiento y herramientas necesarias para hacer frente a un mercado cambiante y para satisfacer mejor las necesidades de los clientes. (Delgado Calero & GARWATER, 2016)

2.3. Desarrollo Tecnológico.

Con la pésima calidad del agua en los hogares en algunas ciudades de la provincia la empresa GARWATER vio la oportunidad de contribuir con la salud de los ciudadanos, ofreciéndoles un producto elaborado con los procesos más estrictos y así devolver a sus clientes un agua libre de impurezas y apta para el consumo humano, gracias a estas necesidades la empresa nace como emprendedora convirtiéndose en la primera empresa embotelladora de agua propia de la provincia.

Por las características únicas de sabor y de calidad del agua que embotellamos mejorando la presentación e innovando sus procesos de purificación para obtener un mejor producto de consumo a un bajo costo. Un ejemplo claro de investigación de nuevos productos es GARTWER con GAS, que se encuentra en fase mercadeo para detectar el recibimiento del producto en el mercado. Aunque GARWATER no desarrolla a gran escala nuevos productos, mantiene a sus clientes ofreciéndole calidad y una buena atención al cliente. (Delgado Calero & GARWATER, 2016)

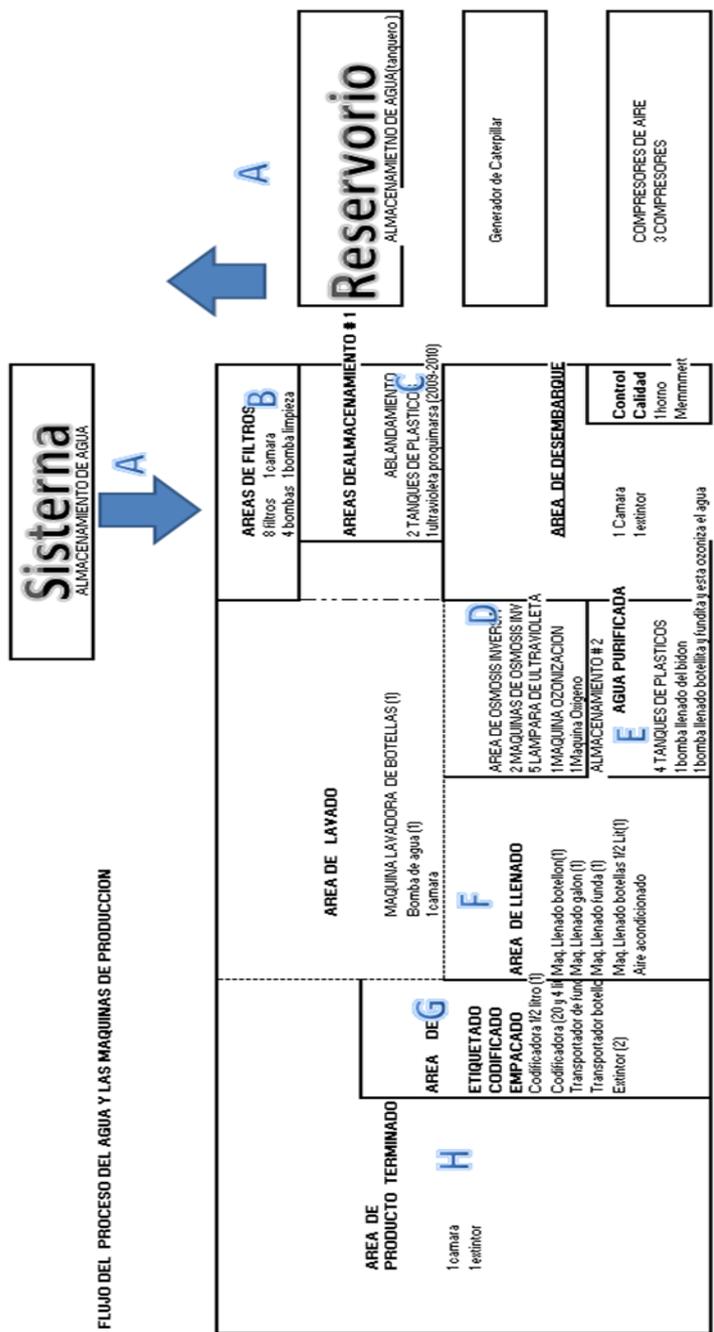
2.4. Aprovechamiento.

La empresa mantiene relaciones con los siguientes proveedores que son los encargados de abastecer de materia prima, equipos, plásticos y químicos utilizados en el los productos de GARWATER:

- MONGAR C.A.
- PROQUIMARSA S.A.
- TEMEIN S.A.
- ULTI S.A.
- PLASTICOS PANAMERICANOS PLAPASA
- EPMAFAP.
- FLEXOFAMA S.A.

2.5. Logística interior / Actividades.

En el corazón del acercamiento multilocal se encuentra el sistema de embotelladores utilizados por la empresa que produce agua embotellada para toda la provincia de Manabí. A continuación se detalla en la figura 2 el proceso de elaboración de los productos de GARWATER: (Delgado Calero & GARWATER, 2016)



2.6. Lógica Exterior.

Los productos son distribuidos por los embotelladores o, como ocurre en gran parte del país y el mundo, se utilizan camiones distribuidores. La empresa posee una flota de 9 camiones, los cuales efectúan 7 rutas diarias cortas y 2 rutas diarias largas. Estos transportes tienen la misión de transportar el producto a todos los rincones de la provincia, manteniendo las buenas normas planteadas en el manual de procedimientos de la empresa para la distribución de los productos. (Delgado Calero & GARWATER, 2016)

Lista de productos de GARWATER:

- Bidón de 20 litros.
- Galón de 5 litros.
- Galón 3 litros.
- Botella de 500 ml.
- Funda 300 ml.

2.7. Marketing.

La empresa GARWATER posee actualmente más de 2000 clientes, distribuidos en los cantones de la provincia de Manabí. La empresa en factor promedio factura alrededor de \$150.000 dólares a pesar que no emplea el uso de propagandas y marketing como fuerte, esto se debe a que la empresa tiene un tiempo considerable de funcionamiento y buena administración ofreciendo confianza y calidad a los clientes. (Delgado Calero & GARWATER, 2016)

Uno de los grandes recursos de la empresa es la marca. Y para mantenerla, GARWATER S.A. pregona su publicidad en los medios de comunicación locales, eventos masivos públicos y selectos, auspiciando eventos y realizando convenios de imagen con la marca con el equipo de futbol de la ciudad. Así se logra resumir los elementos diferenciadores en los siguientes:

- Acercamiento multilocal.
- Producto y marca GARWATER.
- Diseño de nuevos productos.
- Búsqueda de la calidad, a través de:
 - La dirección
 - Los empleados
 - La producción
 - El aprovisionamiento.
- Publicidad que potencia la imagen de marca y la calidad del producto.
- Facilidad de adquisición de los productos por parte del consumidor final.

2.8. Servicios.

GARWATER trabaja día a día para mejorar los servicios por el bien sus clientes, el medioambiente y la comunidad en general, es por tal motivo que la empresa en cada una de sus ventas brinda los siguientes servicios: (Delgado Calero & GARWATER, 2016)

- Atención al cliente
- Entregas inmediatas
- Uso de productos biodegradables
- Campañas de reciclaje
- Capacitaciones periódicas al personal

3. Modelo entidad relación.

Un diagrama Entidad Relación (E/R) representa a la realidad de una base de datos a través de un esquema gráfico, el mismo que emplea terminología y objetos que identifican la forma en que coexiste la información en una base de datos. Con el diagramado y se distinguen entidades las mismas que están compuestas por atributos los cuales deben de ser configurados según el tipo de información que se almacenaría permitiendo establecer relaciones entre entidades. (ECURED, 2016) Los diagramas

E/R nos ofrecen un lenguaje gráfico que muestra conceptos, informalmente estos diagramas están formados por simples dibujos o gráficos que muestran información referente un proceso o sistema, en la Figura 3 se logra observar los gráficos comúnmente empleados para la elaboración de estos diagramas:

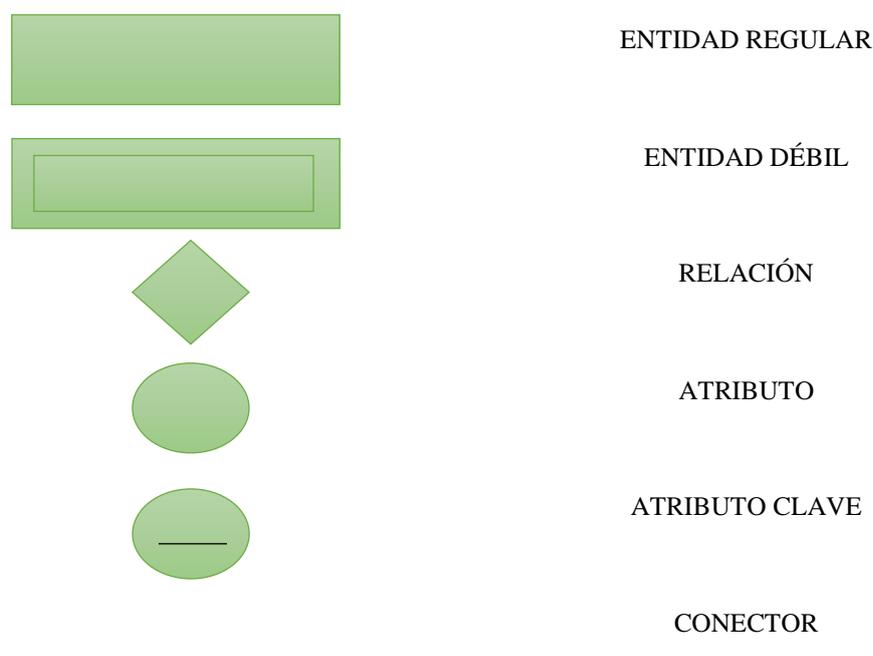


Figura 3. Componentes de un Diagrama Entidad / Relación.
Fuente: (ECURED, 2016)

GARWATER emplea un software que maneja la facturación y otras actividades propias de la empresa, GENESIS (VER ANEXO 10) es la encargada de brindar la instalación y soporte de este sistema que lleva su mismo nombre, desde el año 2009 esta empresa vela por los datos contables de GARWATER. Este aplicativo es de escritorio que utiliza clientes conectados a un servidor local ubicado en las instalaciones de la empresa. Como el estudio de la presente tesis está confinado a detectar patrones y tendencias en las ventas de GARWATER se establecieron los convenios necesarios de intercambio de información entre ambas empresa, para lo cual se solicitó formalmente el diagrama de Entidad Relación del software contable, pieza clave para el análisis de los datos descubriendo la forma en que se registran.

El software contable que emplea la empresa como asistente de facturación posee una base de datos en SYBASE de 244 tablas, siendo necesario extraer las más importantes para el caso de estudio. La selección fue enfocada a las ventas de GARWATER, para ello se seleccionaron las tablas con mayor vinculación a las ventas. En la figura 4 se han realizado ajustes para poder visualizar un esquema representativo del diagrama entidad relación de las ventas. En el proceso de extracción del diagrama entidad relación se obtuvieron 20 tablas, que con ayuda de la herramienta Open Source CA ERwin Community se logró estructurar y esquematizar para una mejor y práctica interpretación. En la tabla 1 se explican los tipos de entidades y documentos, así también las entidades principales donde se mantienen las facturas y los movimientos de estas.

Tabla 1
Catálogo de entidades del software GENESIS.

Tipos de tablas	
in	Inventario
in_fac	Inventario en factura
co	Contabilidad
te	Tesorería
Entidades principales	
in_cabecera	Cabecera de las facturas
in_movimiento	Movimiento de la factura
cxc_auxiliar	Pago de la factura
Tipos de datos en cabecera de facturas	
FC	Venta
CP	Compra
DV	Devolución en venta
DC	Devolución en compra

Las entidades poseen gran número de atributos, esto se debe a que GENESIS (VER ANEXO 10) es un software que da soporte a diferentes tipos de negocios. Por tal razón la forma en la que está construida la base de datos se basa en componentes genéricos que se utilizan dependiendo de la instalación que se le dé al software en la empresa. En la figura 4 se muestra el diagrama entidad relación de módulo de ventas

del software al cual se le han retirado campos que no son utilizados en las transacciones.

4. Modelo multidimensional.

El modelo multidimensional es considerado en el marco del diseño como una disciplina que se fundamenta con el modelo Entidad/Relación, mismo que depende de una ingeniería sobre los datos. Para realizar este modelado es necesario analizar las particularidades de los procesos que ocurren en una organización, dividiéndolos en mediciones y entorno. Las medidas son suelen ser numéricas, llamadas HECHOS. En los hechos existe un contexto que describe las condiciones y el momento cuando se registró este hecho. Aunque este entorno se ve como un todo, existen registros lógicos de diferentes características que dan descripción a un hecho, por ejemplo, si el hecho referido, es la venta de un producto, se podría dividir el entorno que rodea al hecho en la cantidad y el producto vendido, así también como el cliente que lo compró, la tienda y la fecha de realización de la venta. (Trujillo, 2006)

A estas divisiones se le denomina DIMENSIONES y a diferencia de los hechos que son numéricos, estos son de tipo texto. Los campos considerados como medidas, son registrados en las tablas de hechos, la combinación de las múltiples llaves foráneas que hacen referencia a las tablas de tipo dimensiones que componen los eventos ocurridos de este hecho, es decir, cada llaves extranjera que compone la tabla de hecho corresponden a la llave primaria de una dimensión, en la figura 5 se puede apreciar un modelo multidimensional del departamento de ventas de una tienda. (Trujillo, 2006)



Figura 5. Esquema de un modelo multidimensional de ventas.

Para la realización del modelo multidimensional a partir de la base de datos del software GENESIS (VER ANEXO 10), se necesitó de varias reuniones con los técnicos encargados de este aplicativo para comprender datos como: diseño de la base de datos transaccional, los diccionarios de datos, las relaciones y la configuración que la aplicación tenía al guardar los datos en los campos de las tablas. Gracias a este proceso investigativo realizado bajo la supervisión de los directivos de GARWATER se logró elaborar el diagrama entidad – relación, del cual se desprendería el diseño del diagrama multidimensional, dicho diagrama buscó elaborar un repositorio de datos enfocado en las ventas de la empresa, lo cual permitió la selección y creación de las dimensiones que establecerían las cualidades de la tabla de hechos.

En esta tabla de hechos se buscó obtener la mayor cantidad de datos cuantitativos, para lograr en las fases posteriores de la metodología de minería de datos en cuestión obtener una mayor cobertura de los datos, para analizarlos con las herramientas y técnicas de minería de datos empleadas en este estudio. Es importante conocer que el software contable emplea una lógica poco convencional al momento de almacenar los datos, sin embargo el aplicativo funciona correctamente y se acopla a las necesidades de GARWATER, si bien este aplicativo otorga ciertas libertades a la empresa al momento de dar las configuraciones iniciales y de parámetros particulares, esta no presenta dificultades al personal de ventas ya que lo manejan en su forma más cómoda.

Este factor implica que ciertos campos de algunas tablas en la base transaccional que fueron diseñados para almacenar nombres u códigos, sean utilizados para registrar aparentemente otro tipo de información. Un ejemplo claro es al registrar el nombre de un vendedor: en este caso por lógica se almacena una información de tipo personal es decir dos nombres y dos apellidos, sin embargo se utiliza para guardar el nombre del vendedor más el número del vehículo en el que se movilizan, esto se lo realiza con el fin de otorgar compensaciones e incentivos al

compromiso de venta que ellos tienen hacia la empresa. En la figura 6 se observa el modelo multidimensional de las ventas de GARWATER.

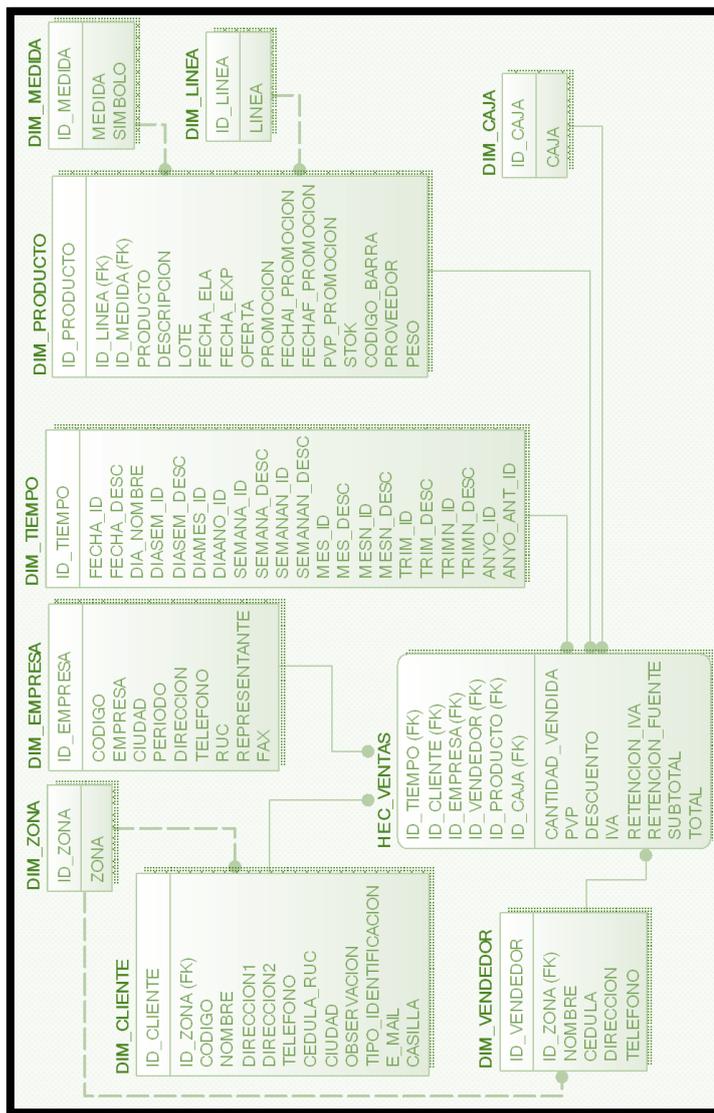


Figura 6. Diagrama multidimensional de facturación de GARWATER

5. Modelos de minería de datos.

Para definir un modelo de minería de datos es necesario la selección y aplicación de algoritmos sobre los datos, estos cumplen la función de generar información otorgando estadísticas y patrones. Mediante un modelo de minería de datos es posible predecir o descubrir información la misma que puesta en análisis del experto en minería de datos, es de mucha utilidad para la toma de decisiones en una empresa. Para entender el rol de un modelo de datos es necesario entender su arquitectura, a continuación se redacta un expreso pero concreto concepto que contempla la arquitectura de un modelo de minería de datos. (Microsoft, 2016)

5.1. Arquitectura de un modelo de datos

En un modelo de datos se reciben los datos referentes a la forma en que se representa la estructura de la minería de datos. En esta se describen los orígenes de datos, los algoritmos que se usaran, los objetos que se crearan y formaran parte de la minería, en síntesis la arquitectura representa a forma de metadatos el proceso de minería en un nivel lógico fácil de comprender y distribuir. (Microsoft, 2016) En este contexto el modelo de minería de datos tiene la función de guardar la información derivada del análisis estadístico de la información en forma de índices de los cuales se puede esquematizar el estudio. En un estado inicial el modelo de minería se encuentra vacío y únicamente se logrará llenar cuando se procese el modelo dando inicio al análisis estadístico de los datos. En un segundo paso se procesa conteniendo los metadatos, resultados de los análisis y los enlaces de la estructura de minería de datos. En la figura 7 se puede apreciar la arquitectura de un modelo de minería de datos. (Microsoft, 2016)



Figura 7. Arquitectura de un modelo de minería de datos.

Los metadatos otorgan el nombre y el servidor del modelo donde está alojado, dentro de estos se incluyen los campos de la estructura de minería de datos empleada para generarlo y los filtros que se aplican durante el procesamiento de datos describiendo el algoritmo utilizado. Estas opciones suelen ser muy influyentes en los resultados de los análisis. (Microsoft, 2016)

Dicho por la página de Microsoft detallando en la presentación de sus herramientas de minería, por lo general cada algoritmo estudia los datos de forma diferente, pues el contenido resultante de la ejecución del modelo de minería establece las estructuras de forma distinta, produciendo el mismo efecto durante el uso de datos de entrenamiento. Los autores dictaminan que los modelos cuyo entrenamiento se han realizado con una misma estructura de minería de datos, producen diferentes resultados al colocar filtros distintos. (Microsoft, 2016)

Por otro lado, es necesario detallar que bajo ningún motivo los datos se almacenan en los modelos; ya que solo contienen el sumario del estudio estadístico, en cuanto a lo datos reales estos se guardan en la estructura propia del modelo de minería. Tomando en cuenta lo descrito si se emplean filtros en los datos durante el entrenamiento del modelo, los conceptos del filtro anexaran como un objeto dentro del modelo. (Microsoft, 2016) Los modelos tienen un conjunto de enlaces, que señalan a los datos capturados en memoria. Estos enlaces otorgan acceso detallado a los resultados sin embargo, los datos reales se guardan en la memoria temporal de la estructura.(Microsoft, 2016)

5.2. Procesar un modelo de minería de datos

Los modelos de minería de datos son objetos vacíos y es solo hasta que se procesan cuando empiezan a tomar forma. Al procesar los datos la estructura se almacena en memoria caché y es aquí donde se pasan a través de un filtro donde solo pasa si se ha definido alguno dentro del modelo incluyendo al algoritmo que los analiza. Cada algoritmo es responsable llevar un conjunto de estadísticas a modo de

resumen que muestra los datos además de identificar las reglas y los patrones obtenidos. Cuando se procesa el algoritmo el modelo se llena con una gran cantidad de datos los mismos que considerados como patrones y se les atribuyen al éxito que mantuvo el análisis hecho por el modelo de minería. (Microsoft, 2016)

Para el desarrollo de la minería de datos en esta tesis, hubo un enfoque en la problemática de la empresa y fueron muchos los indicadores que se plasmaron en los objetivos a alcanzar, siendo este el caso se optó por seguir los modelos de minería de datos descriptivos y predictivos, esto debido a que GARWATER fue una empresa a la que no se ha incorporado este tipo de tecnología. Esto permitió que la empresa se convirtiera en un caso de estudio indicado para descubrir el conocimiento almacenado en su base de datos, en este caso el modelo descriptivo se utilizó con el fin de proporcionar información entre las relaciones de los datos y sus características, haciendo énfasis en los perfiles del dato que se quiere analizar.

Con este modelo de permitió descubrir las relaciones entre los clientes, los vendedores y las zonas de distribución, valiéndonos de análisis estadísticos; permitiéndonos utilizar técnicas de agrupamiento, segmentación y decisión. Del modelo predictivo se obtuvo conocimiento referente a la proyección en el tiempo de la información, correspondiente a los nuevos clientes, a las ventas y de la misma manera analizar las tendencias en las ventas. La puesta en función de estos modelos permitieron al área gerencial observar el conocimiento de su empresa de manera que fuera sencillo y comprensible, convirtiéndose en la oportunidad de tomar decisiones para el crecimiento de la empresa teniendo la certeza de que la información extraída fue sometida a un proceso que garantiza la veracidad de la información mostrada en los diferentes reportes y gráficos proporcionados por la herramienta de minería de datos.

CAPÍTULO III

MARCO CONCEPTUAL

6. Comprensión del negocio.

La comprensión del negocio es la primera fase de la metodología de minería de datos CRISP – DM, la cual se desarrolla realizando una serie de tareas esenciales para definir el marco en que se elaborará la minería estableciendo metas y objetivos que le permitan a la empresa trascender, las tareas que se realizan en esta fase son cinco las cuales se escriben a continuación. (Yoshibauco, 2011)

6.1. Objetivos del negocio

En esta tarea se determina cuál es el problema a resolver, seguido de las razones por la cuales se debe realizar minería de datos definiendo el propósito y los criterios para alcanzar el éxito. En GARWATER el problema principal radica en la escasa valoración de los datos que posee la empresa, si bien la empresa tiene un gran volumen de facturación y el software que cumple la tarea de asistir y registrar este proceso lo realiza exitosamente, se logró observar que en un nivel gerencial no se encuentra una herramienta que permita analizar y entender lo que sucede con las ventas, más allá de ver los valores contables siendo esto una limitación notable.

En este nivel gerencial los pedidos para establecer un análisis del comportamiento de los datos son muchos, pero el sistema con el que cuentan no logra entregar reportes estadísticos basados en algoritmos de minería de datos, lo cual es un problema para la empresa ya que se recurre a procesos arcaicos y reportes realizados en Excel y muchas veces llevados en papel. En GARWATER fue necesario realizar minería de datos y para ello se tomó la información correspondiente los cinco últimos años de la base de datos que emplea el software GENESIS (VER ANEXO 10) en la empresa. Los objetivos comerciales que la empresa persigue y servirse mediante el uso de minería de datos son los siguientes:

- Mejorar las ventas al realizar mejores recomendaciones.
- Otorgar un servicio más personalizado a los clientes para fomentar la lealtad.
- Optimizar las campañas de marketing.
- Optimizar las rutas de distribución de los productos.
- Expandir el territorio de ventas a otras provincias.

6.2. Situación actual de GARWATER.

GARWATER S.A. surgió en el mercado de agua embotellada el 14 de junio de 2002, con el único objetivo de entregar agua pura embotellada de calidad, empleando la más alta tecnología, para contribuir al cuidado de la salud de los ecuatorianos a partir de la detección de problemas en calidad de agua distribuida por la empresa pública como privada. La empresa tiene la filosofía de contribuir y aportar social y económicamente con la población ecuatoriana, ya que son fieles creyentes en que los clientes son la base en quienes se tiene entrega el compromiso entregar un producto de calidad. A la cabeza de la empresa se encuentra el Gerente General Ing. Julio Enrique Rivas Beltrón y la administradora Ing. Bielka Rodríguez.

GARWATER, se vincula con sus clientes mostrándose así misma al mundo como una empresa ejemplo en la manera de servir y actuar, muy convencida que materia prima es un recurso invaluable. La empresa se enfoca en la verdad constituyendo una institución que fomenta el honor y la integridad de quienes forjan los productos. GARWATER, está comprometida a tratar con respeto a sus empleados, dándoles una compensación justa y merecedora por su desempeño dentro y fuera de la empresa. Esta mantiene una cultura empresarial que alienta a la excelencia, el liderazgo y la innovación, sirviendo siempre a los clientes con la excelencia de los productos que se elaboran en las plantas de producción acompañado de la mejor atención y un buen trato.

La empresa se encamina a mejorar periódicamente y esto lo logra gracias al empeño y dedicación del personal que labora en ella, así también como la dirección

que las autoridades emplean para dotar a GARWATER con la mejor tecnología. Es así que la empresa planea lanzar campañas de expansión a otras provincias y para ello es necesario proyectar un sin número de factores que indicarían la factibilidad y el orden correcto de expansión. Se estima que la empresa factura un monto superior a los \$100.000 US mensual, indicador de la gran trascendencia en el mercado actual; en consecuencia en los repositorios digitales de información se mantienen datos valiosos que solo aplicando minería de datos se transformaría en información valiosa para el futuro productivo y lucrativo de la empresa.

6.3. Objetivos de la Minería de Datos.

Para alcanzar los objetivos comerciales de GARWATER fue necesario establecer los objetivos de la minería, estos permitieron el cumplimiento oportuno de los objetivos antes mencionados. A continuación se expone los objetivos de minería de datos:

- Realizar estudio de las tendencias en las ventas con la facturación de los últimos cinco años.
- Determinar el perfil de los clientes respecto a su frecuencia de compra.
- Realizar un análisis demográfico de los clientes para definir los medios y las formas en que las promociones y publicidad cumplan su función primordial.
- Descubrir patrones en la distribución y entrega de los productos.
- Proyectar las ventas asumiendo el comportamiento de las mismas en periodos futuros de tiempo.

6.4. Plan del proyecto de minería.

El plan del proyecto de minería se encuentra en la tabla 2, la cual resume las actividades desarrolladas en el estudio haciendo referencia al tiempo, recursos y los riesgos estimados, como efecto se logró proyectar las operaciones de cada fase del plan, socializando las actividades elaboradas con la presencia de las máximas autoridades de la empresa GARWATER. Con el desarrollo y seguimiento de esta

planificación se logró de manera exitosa la ejecución del proyecto, permitiendo conocer y valorar las herramientas, las técnicas y los algoritmos empleados en los bancos de información de la empresa, generando como resultado el descubrimiento de las tendencias, patrones y conocimientos que sirvieron como eje para la toma de decisiones otorgando valor a la empresa.

Tabla 2
Plan de proyecto de minería de datos.

FASE	TIEMPO	RECURSOS	RIESGO
Comprensión del negocio	2 semana	Investigador, Autoridades y empleados de GARWATER	Cambios económicos
Comprensión de los datos	4 semanas	Investigador, Soporte GENESIS, GARWATER	Problemas de datos, problemas tecnológicos
Preparación de los datos	4 semanas	Investigador, SYBASE, PENTAHO, GENESIS, POSTGRES.	Problemas de datos, problemas tecnológicos
Modelado	4 semanas	Investigador, WEKA PENTAHO,	Problemas de tecnología, incapacidad para encontrar un modelo adecuado
Evaluación	4 semana	Investigador, Autoridades y empleados de GARWATER	Cambio económico, incapacidad para implementar resultados
Distribución	2 semana	Investigador, Gerente GARWATER.	Cambio económico, incapacidad para implementar resultados

La valoración de la herramienta fue tema de discusión y clave para el desarrollo del proyecto, optando por una alternativa OPEN SOURCE, seleccionando

así el software PENTAHO para realizar la minería de datos, el argumento que permitió esta selección se obtuvo al observar el cuadrante de GARTNER actualizado hasta el mes de febrero del 2016 el cual posiciona esta herramienta como una buena alternativa. En la figura 7 se observa las diferentes herramientas que existen en el mercado para la realización de minería de datos y BI, lamentablemente la posición de GARWATER frente a esta tecnología genera una expectativa que empuja a los directivos a usar herramientas libres y no arriesgarse a comprar licencias de software ubicados en el primer cuadrante de Gartner. (Gartner, 2016)

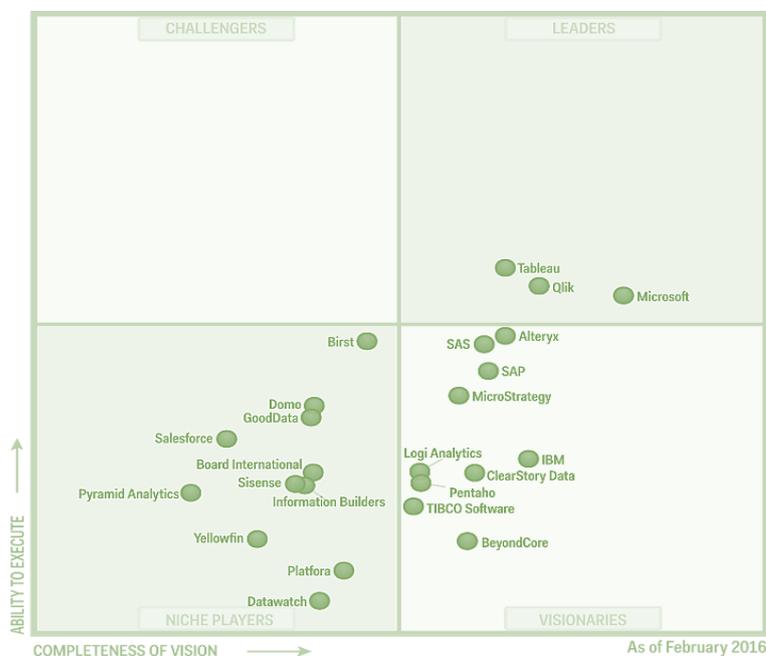


Figura 8. Cuadrante de Gartner.
Fuente: (Gartner, 2016)

En la tabla 3 se muestran las aplicaciones de PENTAHO junto a sus versiones y licencias para una mejor comprensión y detalle del software con sus aplicaciones para el desarrollo del proyecto, dichas aplicaciones fueron presentadas a la administración de la empresa para garantizar el proceso de selección bajo los argumentos presentados, otorgando confianza y seguridad en el proceso.

Tabla 3
Versiones y licencias de PENTAHO utilizadas en el proyecto.

APLICACIÓN	VERSIÓN	LICENCIA
Data Integration - Kettle	6.0	Apache License 2.0
Data Mining - Weka	3.7.13	GNU General Public License (GPL)
Reporting	6.0	GNU Lesser General Public License, version 2.1

7. Comprensión de los datos.

7.1. Recolección de datos iniciales.

Los datos necesarios para realizar la minería de datos se obtuvieron del servidor principal de la empresa GARWATER donde se aloja la base de datos del sistema GENESIS (VER ANEXO 10), esta base de datos está diseñada en Sybase 11. La empresa que da soporte al sistema contable, facilitó un técnico para la extracción y estudio de la bases de datos bajo el consentimiento del gerente de la empresa GARWATER, este servidor se encuentra en la planta principal ubicado en Portoviejo exactamente en la vía Portoviejo – Santa Ana a los lados del INIAP; que donde funciona el departamento administrativo de la empresa.

Es necesario marcar que se hizo un corte en la base de datos debido a una actualización del sistema contable generando dos bases de datos donde el primer respaldo contiene los registros del año 2009 hasta el 2014 y el actual respaldo contiene los registros del 2015 hasta la fecha. Para la recolección inicial de la información se emplearon las técnicas de la entrevista y la observación, las cuales fueron aplicadas a los directivos y los empleados vinculados directamente al área de ventas de la empresa GARWATER, de igual manera se aplicaron estas técnicas con el personal que le da soporte al software contable, para establecer los permisos y accesos a las fuentes de datos del aplicativo que realiza la facturación en la empresa.

7.2. Descripción de los datos.

El software GENESIS (VER ANEXO 10) interactúa con su base de datos que posee un total de 244 tablas, las cuales dependiendo la empresa y el negocio puede haber el caso de que no se utilicen todas. Sin embargo este software utiliza 150 tablas para almacenar los registros de la empresa GARWATER. Durante el proceso de implementación del sistema en la empresa se realizaron las capacitaciones correspondientes permitiendo que un usuario administrador pudiera configurar y crear ciertos parámetros ideales para el negocio, de tal manera algunos registros manejan una lógica poco estructurada esto se debe a las limitaciones del software y el uso de usuarios inexperto causando que en varios registros se muestren datos aparentemente incoherentes.

Como el caso de estudio se enfoca en las ventas de la empresa es necesario mostrar las tablas principales con sus respectivos campos, es importante señalar que la dinámica de almacenamiento en cada año se establece con la tabla **ge_empresa** que es la encargada de generar un código único, este código apunta a todas las tablas de manera que clasifica las actividades realizadas en cada año fiscal. En la figura 9 se observa la estructura de esta tabla.

Name	ID	Data Type
codigo	1	varchar(3)
empresa	2	varchar(35)
ciudad	3	varchar(20)
periodo	4	varchar(4)
direccion	5	varchar(50)
telefono	6	varchar(30)
ruc	7	varchar(13)
pais	8	varchar(20)
logo	9	long varchar
representante1	10	varchar(30)
representante2	11	varchar(30)
fax	12	varchar(30)
patronal	13	varchar(30)
rucContador	14	varchar(13)
idRepre	15	varchar(13)
tpldRepre	16	varchar(1)
email	17	varchar(60)
valor_v	18	decimal(15,2)

Figura 9. Estructura de tabla **ge_empresa**.

Al iniciar cada año se genera un nuevo registro que permitirá clasificar todas las actividades que se realicen en dicho año, actualmente la tabla contiene 14 registros lo que no concuerda con el número de años que tiene funcionando el aplicativo, esto se debe a que durante el año fiscal se efectúan cambios en cuentas o módulos del sistema por lo que es necesario crear un nuevo indicador. Cuando un usuario requiere la búsqueda de registros pertenecientes a periodos contables posteriores al actual, configura la aplicación para cargar todos los datos de dicho periodo, lo que internamente ocurre hace efecto al movimiento del índice de la tabla `ge_empresa` de la cual se relaciona la mayoría de las tablas vinculadas a las ventas y otras actividades del software en general. En la figura 10 se observa una consulta SQL donde se obtiene los registros de la tabla `ge_empresa` y se logra visualizar en especial el campo periodo donde se establece el año contable al que hace referencia el registro.

codigo	empresa	ciudad	periodo	direccion	telefono	ruc	pais	logo	representante1
1/001	GAR WATER	PORTOVIEJO	2009	CARRETERO COLON QUIIMIS VIA SANTA ANA	052420940	1391719078001	ECUADOR	(NULL)	SR JULIO RIVAS
2/002	GAR WATER_2010	PORTOVIEJO	2010	CARRETERO COLON QUIIMIS VIA SANTA ANA	052420940	1391719078001	ECUADOR	(NULL)	SR JULIO RIVAS
3/003	GAR WATER	PORTOVIEJO	2010	CARRETERO COLON QUIIMIS VIA SANTA ANA	052420940	1391719078001	ECUADOR	(NULL)	SR JULIO RIVAS
4/004	GAR WATER_2011	PORTOVIEJO	2011	CARRETERO COLON QUIIMIS VIA SANTA ANA	052420940	1391719078001	ECUADOR	(NULL)	SR JULIO RIVAS
5/005	GAR WATER_2010	PORTOVIEJO	2011	CARRETERO COLON QUIIMIS VIA SANTA ANA	052420940	1391719078001	ECUADOR	(NULL)	SR JULIO RIVAS
6/006	GAR WATER_2011	PORTOVIEJO	2012	CARRETERO COLON QUIIMIS VIA SANTA ANA	052420940	1391719078001	ECUADOR	(NULL)	SR JULIO RIVAS
7/007	GAR WATER_2012 -1	PORTOVIEJO	2012	CARRETERO COLON QUIIMIS VIA SANTA ANA	052420940	1391719078001	ECUADOR	(NULL)	SR JULIO RIVAS
8/008	GAR WATER_2012	PORTOVIEJO	2012	CARRETERO COLON QUIIMIS VIA SANTA ANA	052420940	1391719078001	ECUADOR	(NULL)	SR JULIO RIVAS
9/009	GAR WATER_2012 -1	PORTOVIEJO	2013	CARRETERO COLON QUIIMIS VIA SANTA ANA	052420940	1391719078001	ECUADOR	(NULL)	SR JULIO RIVAS
10/010	GAR WATER_2012	PORTOVIEJO	2013	CARRETERO COLON QUIIMIS VIA SANTA ANA	052420940	1391719078001	ECUADOR	(NULL)	SR JULIO RIVAS
11/011	GAR WATER 2013	PORTOVIEJO	2013	CARRETERO COLON QUIIMIS VIA SANTA ANA	052420940	1391719078001	ECUADOR	(NULL)	SR JULIO RIVAS

Figura 10. Consulta de la tabla `ge_empresa`.

Detallar en este documento todas las tablas necesarias para registrar las ventas en GARWATER es tarea extensa y en algunos casos repetitiva, es por tal razón que a continuación se explicaran las tablas más importantes, como lo son la cabecera de las facturas, los ítems de las facturas y la cancelación de la misma, acompañadas de tablas secundarias como productos y clientes que conforman el eje principal de las ventas en la empresa GARWATER. La tabla `in_cabecera` es la encargada de almacenar las cabeceras de las facturas que realiza GARWATER, para el uso del software contable esta tabla maneja un total de 52 campos, de los cuales solo se

llegan a usar un máximo de 22 campos, esto se debe a la configuración inicial y el acoplamiento del software a la empresa explicado en secciones anteriores, en la figura 11 se observa su estructura en la cual se determinan todos los campos acompañados de su tipo de dato.

Es necesario afirmar que el software contable mantiene una lógica genérica al momento de ser implementado en una empresa, esto se debe a que no es un software a medida y a eso se debe la gran cantidad de campos muchos de los cuales no son utilizados bajo ningún concepto por la empresa GARWATER. Para detallar la estructura de la tabla **in_cabecera** fue necesario unificar varias figuras las cuales fueron unificadas para una mejor representación visual dando paso a la figura 11.

tipo	1 varchar(2)	hora	33 time
documento	2 varchar(10)	estacion	34 varchar(5)
fecha	3 date	punto	35 varchar(5)
pro_cli	4 varchar(8)	nombre_usuario	36 varchar(40)
vendedor	5 varchar(8)	fecha_usuario	37 date
descuento	6 numeric(7,4)	hora_usuario	38 time
empresa	7 varchar(3)	habilitado	39 varchar(1)
ubicacion	8 varchar(3)	vendedor_usuario	40 varchar(6)
impuesto	9 numeric(7,4)	accion_usuario	41 varchar(37)
comentario	10 long varchar	documento2	42 varchar(20)
fechav	11 date	usuario	43 varchar(40)
marca_cxc	12 varchar(10)	pedido	44 varchar(10)
marca_co	13 varchar(10)	tr_anexos	45 varchar(10)
referencia	14 varchar(20)	oficial	46 varchar(8)
transporte	15 numeric(10,2)	propina	47 numeric(10,2)
seguro	16 numeric(12,2)	moneda	48 varchar(15)
orden	17 varchar(20)	guia_remision	49 varchar(20)
contabilidad	18 varchar(1)	ce_autoriza	50 varchar(50)
stado	19 varchar(1)	ce_clave_acceso	51 varchar(50)
stadoimp	20 varchar(20)	ce_rise	52 varchar(1)
fecha_retencion	21 date		
retencion_fuente	22 numeric(5,0)		
retencion_iva	23 numeric(5,0)		
contabiliza_ret	24 varchar(8)		
contabiliza_tran	25 varchar(1)		
nombre_trans	26 varchar(50)		
cedula_trans	27 varchar(13)		
direccion_dest	28 varchar(50)		
fecha_tras1	29 date		
fecha_tras2	30 date		
caja	31 varchar(8)		
contabiliza_iva	32 varchar(8)		

Figura 11. Estructura de la tabla in_cabecera.

De igual manera la en la figura 12 observamos la estructura de la tabla **in_movimiento**, encargada de almacenar los detalles de las facturas vinculada a la tabla **in_cabecera**. Esta tabla contiene 26 campos que bajo la lógica anteriormente explicada no se llegan a utilizar todos los campos, a continuación se muestra la estructura de la tabla.

tipo	1 varchar(2)	serie	17 varchar(80)
documento	2 varchar(10)	clase	18 varchar(10)
cantidad	3 numeric(12,2)	producto2	19 varchar(15)
valor	4 numeric(15,6)	formapago	20 char(2)
descuento	5 numeric(11,6)	transporte	21 numeric(15,4)
impuesto	6 numeric(7,2)	totalprome	22 numeric(15,4)
bonificacion	7 numeric(10,2)	cod_rf	23 numeric(15,4)
lista	8 varchar(3)	cod_ri	24 numeric(15,4)
producto	9 varchar(15)	codigo_concepto_re...	25 varchar(3)
empresa	10 varchar(3)	fecha_cad	26 date
secuencia	11 integer		
promedio	12 numeric(15,6)		
costo	13 numeric(15,6)		
valor1	14 numeric(15,6)		
ubicacion	15 varchar(3)		
numprecio	16 numeric(3,0)		

Figura 12. Estructura de la tabla in_movimiento.

Finalmente en la figura 13 se tiene la estructura de la tabla donde se registran los pagos de las facturas, acompañada de la estructura de las tablas de clientes denotada en la figura 14 y la tabla producto en la figura 15.

documento	1 varchar(10)	retencion	17 numeric(15,2)
tipo	2 varchar(2)	contabilidad	18 varchar(1)
forma_pago	3 numeric(10,0)	ret_iva	19 numeric(15,2)
fechae	4 date	recibo	20 numeric(10,0)
fechav	5 date	descuento	21 numeric(15,2)
numero	6 varchar(15)	abono	22 varchar(1)
valor	7 numeric(15,2)	ubicacion	23 varchar(15)
banco	8 numeric(15,0)	cxc_descuento	24 varchar(10)
referencia	9 varchar(100)	cxc_fuente	25 varchar(10)
empresa	10 varchar(3)	cxc_iva	26 varchar(10)
secuencia	11 numeric(15,0)	anticipo	27 numeric(15,2)
cxc_marca	12 varchar(10)	recibo2	28 varchar(10)
co_marca	13 varchar(10)	observacion	29 varchar(100)
cuenta	14 varchar(15)	documento2	30 varchar(10)
entidad	15 varchar(35)	retencion_trans	31 numeric(15,3)
numero_1	16 varchar(35)	ret_trans	32 numeric(15,0)
		ret_asumidas	33 numeric(15,0)

Figura 13. Estructura de la tabla cxc_auxiliar.

codigo	1	varchar(8)	e_mail	17	varchar(100)
nombre	2	varchar(35)	fax	18	varchar(30)
direccion1	3	varchar(35)	cta_contable	19	varchar(25)
direccion2	4	varchar(35)	estado	20	varchar(1)
telefono	5	varchar(30)	cta_contable_iva	21	varchar(25)
cedula_ruc	6	varchar(13)	cta_contable_retenci...	22	varchar(25)
cupo	7	numeric(14,2)	cta_contable_retenci...	23	varchar(25)
activo	8	varchar(1)	descuento	24	numeric(10,2)
ciudad	9	varchar(20)	zona	25	varchar(3)
observacion	10	varchar(100)	propietario	26	varchar(50)
contacto	11	varchar(30)	dias	27	integer
empresa	12	varchar(3)	institucion	28	varchar(3)
vendedor	13	varchar(8)	tipo	29	varchar(3)
cargo	14	varchar(30)	tipo_identificacion	30	varchar(2)
cobranza	15	varchar(30)			
casilla	16	varchar(10)			

Figura 14. Estructura de la tabla in_cliente.

codigo	1	varchar(15)	costo	16	numeric(15,6)
empresa	2	varchar(3)	pvp1	17	numeric(15,6)
descripcion1	3	varchar(35)	pvp2	18	numeric(15,6)
grupo	4	varchar(4)	pvp3	19	numeric(15,6)
linea	5	varchar(3)	por1	20	numeric(7,2)
medida	6	varchar(3)	por2	21	numeric(7,2)
codigo_barra	7	varchar(30)	por3	22	numeric(7,2)
maximo	8	numeric(12,2)	iva	23	varchar(1)
minimo	9	numeric(12,2)	promocion	24	varchar(30)
observacion	10	varchar(100)	consumo	25	varchar(1)
Marca	11	varchar(5)	dias	26	integer
stock	12	varchar(1)	promedio	27	numeric(15,6)
referencia	13	varchar(40)	imagen	28	varchar(100)
aplicacion	14	varchar(40)	calidad	29	varchar(1)
cuenta	15	varchar(20)	metatotal	30	numeric(15,4)
costo	16	numeric(15,6)	metacantidad	31	numeric(15,2)
			itemb	32	varchar(1)
			items	33	varchar(1)

Figura 15. Estructura de la tabla in_item.

7.3. Exploración de los datos.

La versión del software GENESIS (VER ANEXO 10) que emplea GARWATER es operativa para Windows desde la versión XP en adelante, utiliza una base de datos creada en Sybase Anyware 11 con un total de 244 tablas, de las cuales solo utiliza 50 para la empresa. De esta muestra se extrajeron 14 tablas para realizar el diagrama entidad relación enfocado a las ventas, en el cuadro 1 se logra apreciar dicha información donde solo se utilizara el 22% de las tablas del sistema.



Figura 16. Estadísticas de la Base de Datos del Software GENESIS - Ventas.

Como la muestra del estudio se obtuvo a partir de los datos generados en el periodo 2011 – 2015 el volumen de almacenamiento no supera los 2 GB, es decir que la tasa de crecimiento en cuanto al volumen físico de información es bajo. A continuación en el cuadro 2 se puede observar la cantidad de registros ingresados en cada periodo financiero detallando la participación de las tablas principales correspondientes a la cabecera de las facturas, el detalle de las facturas y la cancelación de las facturas. Como es de observar el crecimiento se viene dando hasta los actuales momentos en forma exponencial.

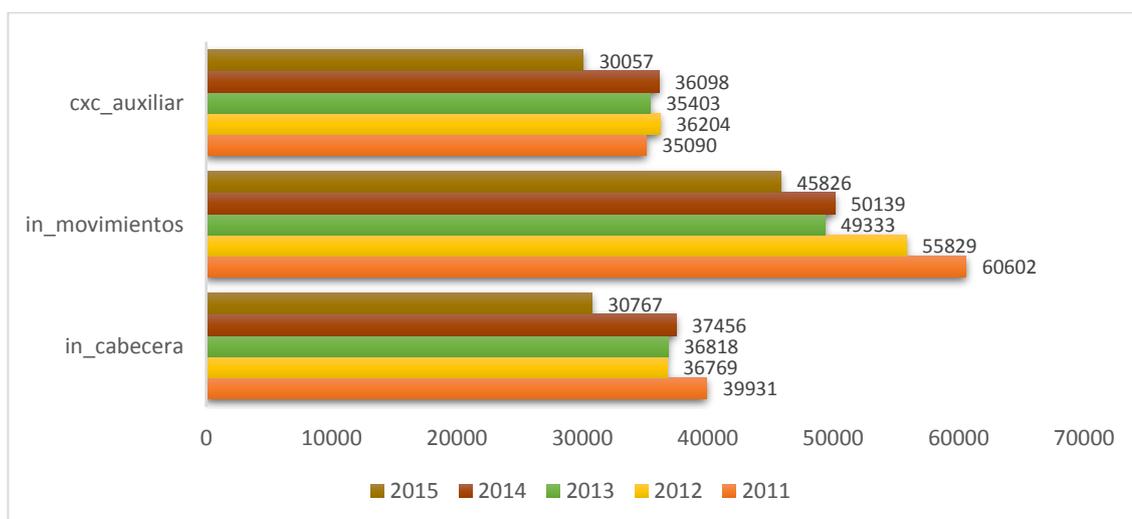


Figura 17. Registros almacenados en tablas principales (2011 - 2015).

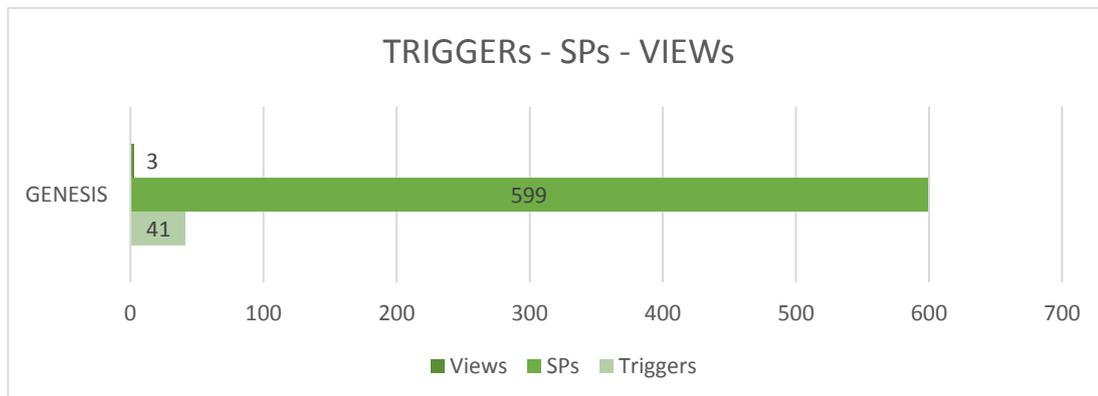


Figura 18. Triggers, SPs, Views.

7.4. Verificación de los datos.

Para la realización de esta tarea fue necesario el apoyo del personal técnico del software, debido a que en su base de datos precisamente en las tablas existe el inconveniente en el que no se utilizan la mayoría de sus campos cuando se guarda información de GARWATER, por lo que se utilizaron consultas SQL para verificar los campos que tomen parte en el análisis y los que no arrojaran el resultado esperado. Otro asunto de gran importancia fue entender que la empresa que da soporte al sistema, capacita a los empleados de la empresa, pero el personal a no poseer una cultura y preparación en el cuidado del ingreso de la información, se encontró con gran cantidad de registros basura u otros solo simples pruebas.

Detectar estos errores fue tarea fácil pero extensa por la gran cantidad de registros y la forma en que esta creada la base de datos, finalmente se validaron todos los datos y se registraron los rellenos que permitieron que la base de datos multidimensional funcione óptimamente reduciendo en un 100% los datos basura.

8. Preparación de los datos.

8.1. Selección de los datos.

La selección de los datos en la base de datos de GENESIS (VER ANEXO 10), se realizó al unir los registros de las tablas seleccionadas pertenecientes a las base 2014.db que posee información desde el año 2009 hasta el año 2014 y 2015.db que contienen la información del año 2015 hasta la fecha. En estas bases de datos se encuentra el historial de facturas y en efecto las ventas que ha realizado la empresa desde la implementación del software contable, es por tal motivo que todas las consultas SQL para la extracción de los datos, están filtradas para limitar los registros en el tiempo en el rango 2011 – 2015, permitiendo de esta manera obtener los datos necesarios para llenar el entorno OLAP para realizar la minería.

Es un hecho que la lógica de la base de datos del software contable en el módulo de ventas, gira en torno a la tabla IN_CABECERA que contiene los índices de las facturas, en la extracción del subconjunto de datos para el estudio al realizar las consultas SQL se obtuvieron un total de 180856 registros correspondiente al total de facturas realizadas en GARWATER en el periodo establecido para la realización de esta investigación.

8.2. Limpieza de los datos.

En la mayoría de las tablas de la base de datos de software contable se encontró con campos que no registraban información, esto se debe a la configuración inicial que se le da al software para su permanente funcionamiento, esto deja sin utilizar a tablas y campos que cumplen funciones especiales si la empresa cliente lo quisiera. El soporte técnico de la empresa GENESIS (VER ANEXO 10) en sus tareas de mantenimiento, deciden dejar estos campos; que en su mayoría se llenan con valores NULL, con 0 o en blanco, por tal motivo para limpiar los datos de esta fuente de información se emplearon los siguientes métodos de limpieza:

- **Análisis:** Se estudió en contenido de las tablas en el entorno transaccional, analizando conjuntamente con los técnicos de bases de datos de la empresa GENESIS (VER ANEXO 10) las características y funcionamiento de cada uno de los campos en las tablas vinculadas a las ventas.
- **Transformación de Datos:** Se utilizó este método para normalizar los tipos de datos, debido a las actualizaciones del software y base de datos, durante la unión de registros entre tablas ocurrían errores por el cambio de los tipos de datos y tamaño de los mismos gracias al incremento de tablas o versión del software.

La limpieza de datos se la realizo creando flujos de datos en PENTAHO, conocidos como transformaciones en la figura 16 se puede observar la limpieza de datos que se le realizó a la tabla IN_CABECERA, la misma que contenía campos inutilizados, registros en 0, null o en blanco, lo cual era un permanente problema para la realización de la minería de datos pudiendo generar errores o datos incoherentes en el estudio, de la misma manera se le realizó una exhaustiva limpieza de datos a todas las tablas transaccionales que funcionan en el módulo de ventas del software contable.

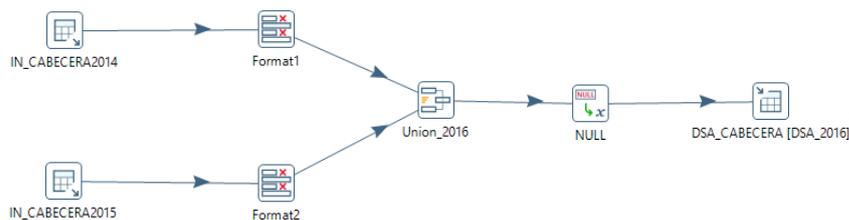


Figura 19. Flujo de Transformación y Limpieza de Datos en PENTAHO de la tabla IN_ CABECERA.

8.3. Estructura de los datos.

A la hora de proporcionar nuevos atributos a partir de ya existentes, integración de nuevos registros y la transformación de valores para atributos existentes, fue mínima debido a la estructura de la base de datos del software contable. Esta base de datos contenía una gran cantidad de campos calculados, unos a nivel de aplicación y otros valiéndose de triggers o procesos almacenados permitiendo ejecutar los mismos para generar los campos requeridos en el modelo multidimensional.

8.4. Integración de los datos.

Para la creación del repositorio de datos de prueba con el cual se obtuvo la información para generar la base de datos dimensional, fue necesario la unión de registros entre las bases de datos 2014.db y 2015.db, esta tarea fue compleja de realizar debido al incremento de los campos entre las tablas de ambas bases de datos y tipo de datos en especial los campos numéricos como los formatos de fecha. Sin embargo gracias a que cada registro en la base transaccional está identificado por el índice de la empresa que se crea al inicio de cada año, se pudo distinguir los segmentos de datos de ambas bases de datos logrando una comprobación de registros rápida y segura. Durante el proceso de relleno de campos en las dimensiones se crearon nuevos registros en las tablas de las bases de datos de origen para establecer sus respectivos índices de ser necesario en futuros estudios.

8.5. Formateo de los datos.

Durante el proceso de extracción de datos para la creación del repositorio de prueba realizado en PENTAHO, se pudo formatear los diferentes campos que generaba desbordamientos o causantes de errores en el momento de juntar las bases de datos de origen, es necesario destacar que el software PENTAHO en la mayoría de sus herramientas de transformación mantiene presente el formateo de los ingresos y salidas de datos entre flujos, lo que hace realmente práctico y oportuno al momento

de dar formatos a los datos para llenar las tablas de trabajo, en la figura 17 se observa el paso de formateo que posee la herramienta Pentaho.

9. Modelamiento de los datos.

9.1. Selección de técnicas.

En cuanto a las técnicas empleadas en el modelado se han mejorado desde sus inicios y la capacidad de almacenamiento información se han optimizado formidablemente. (Lantares, 2014) Se puede especificar que la potencia del cálculo en las técnicas de minería de datos determina la manera de automatizar el modelo, permitiendo así que el trabajo de los datos se realice cada vez en menos tiempo. La creación los modelos, aplicada a los negocios sin importar el escenario los siguientes objetivos:

- Extraer la mayor cantidad de conocimiento sobre los datos.
- Aumentar la valoración del conocimiento encontrado.
- Generar nuevas estrategias para las empresas. (Lantares, 2014)

En el caso de GARWATER bajo ningún concepto se le permitió obtener una extracción de conocimiento lo cual no dio paso a perfeccionar las estrategias en la empresa, por tal motivo se realizó este estudio empleando las siguientes técnicas de modelado, las cuales se detallan en la tabla 4 donde se puede apreciar el alcance y utilidad de estas técnicas dando razón del porque se fueron necesarias en la empresa, partiendo de un punto inicial donde el escenario muestra una empresa sólida en ventas con un gran volumen de clientes y una expansión territorial que cubre a la provincia de Manabí, que por factores ajenos al estudio no ha pasado por un proceso de extracción de conocimiento mucho menos un análisis de sus bases de datos.

Tabla 4
Técnicas de minería de datos.

CATEGORÍA	DESCRIPCIÓN	TÉCNICAS
Clasificación	Los métodos tienen la intención de aprender diferentes funciones que clasifiquen los datos dentro de un conjunto predefinido de clases. Dado un nuevo de clases predefinidas, un número de atributos y un conjunto de datos de aprendizaje o entrenamiento, los métodos de clasificación pueden automáticamente predecir la clase de los datos previamente no clasificados. Las claves más problemáticas relacionadas con la clasificación son la evaluación de los errores de clasificación y la potencia de predicción. (Milla, 2010)	*Árboles de decisión binarios. *Redes neuronales *Programación lineal y estadística.
Asociación	Mide la relación entre un ítem de una transacción y otro ítem en la misma transacción es utilizado para predecir patrones. (Milla, 2010)	*Reglas de asociación. *Arboles de decisión.
Predicción	La idea principal del análisis de predicción es descubrir las relaciones entre variables dependientes e independientes y las relaciones entre variables independientes. (Milla, 2010)	*Técnicas de regresión
Series temporales	Utiliza datos históricos junto con técnicas de regresión lineal o no lineal, produce curvas de regresión que se utilizaran para establecer predicciones a futuro. (Milla, 2010)	*Algoritmos de series de tiempo

Fuente: (Lantares, 2014)

La aplicación de estas técnicas permite alcanzar los objetivos empresariales de cualquier empresa, en el caso de GARWATER, cumpliendo con las expectativas de las autoridades de la empresa, se optó por asistir a modelos de predicción creando un nuevo punto de partida donde el escenario agrega un variante importante en tema de extracción de información para realizar el descubrimiento de nuevo conocimiento.

CAPÍTULO IV

MODELO DE INVESTIGACIÓN

10. Deducción – inducción.

Dada la problemática se trabajó en la investigación aplicando los métodos deductivo e inductivo para realizar el estudio de la información de GARWATER, partiendo de diferentes ámbitos se analizaron los datos aplicando el razonamiento lógico para dar lugar a observaciones particulares y leyes generales a favor de la investigación.

El razonamiento deductivo siempre ha constituido una de las principales características del proceso de enfoque cuantitativo de una investigación, tanto así que el físico Albert Einstein durante la elaboración de la Teoría de la Relatividad hizo uso de este método, pues él partió de una teoría, que imaginó dando a lugar un serie proposiciones o conceptos previos (Plata, 2008). Por otra parte el método de investigación inductivo que consiste en inferir de algunos casos particulares para posteriormente establecer una ley general que los enmarque, este razonamiento inductivo siempre constituirá uno de los pilares fundamentales del enfoque cualitativo de la investigación. (Plata, 2008)

Dentro de la metodología CRISP-DM, el método de investigación deductivo formo parte clave para el desarrollo de las fases de comprensión del negocio y de los datos, donde se sostuvo un análisis general de la situación de la empresa en estos últimos cinco años, dando lugar a marcos más específicos en los cuales se enfocó esta investigación logrando aplicar de forma asertiva las herramientas de minería de datos mencionados en este escrito. En la figura 18 se tiene la representación del método deductivo explicando a razón de bloques la transformación de la información partiendo de aspectos muy generales hacia conocimientos particulares, donde se obtuvieron las bases necesarias para la investigación recopilando información básica facilitando al investigador la comprensión de los datos y situación actual de la

empresa. Gracias a este método se logró implementar las herramientas de minería de datos necesarias, obteniendo resultados importantes para el futuro y dirección de GARWATER.

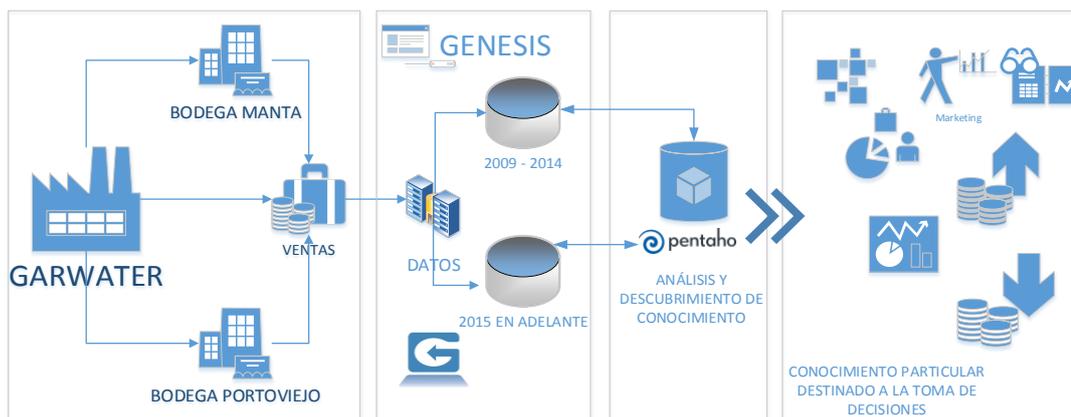


Figura 20. Representación del método deductivo en la investigación.

El conocimiento particular obtenido de los análisis constituye la fuente necesaria para construir las reglas y políticas que la empresa deberá establecer en años futuros y de esta forma con la aplicación del método inductivo de investigación, plasmar como un logro el establecimiento de los resultados presentados en este estudio, mismas que fueron puestas prácticas en la fase de “Despliegue” en la metodología CRISP – DM.

11. Observación.

La observación se realiza de manera empírica, siendo uno de los métodos más antiguos en el marco de la investigación. Al emplear la observación se establece una relación directa entre el investigador y los hechos expuestos por el entorno de donde se obtiene la información necesaria para realizar la investigación. (Fabbri, s.f.) Para la presente tesis el rol de la observación facilitó el desarrollo de la fase de Comprensión de Negocio en la metodología CRISP - DM, que inicialmente muestra a la empresa en un escenario situado en el cantón Portoviejo de la provincia de Manabí, donde la ciudad extiende su parque industrial dando cabida a la formación

de nuevas empresas, siendo el caso de GARWATER que toma el liderato local en la producción de agua embotellada para consumo humano entregando sus productos a mercados, tiendas, instituciones educativas, públicas, privadas y financieras.

Gracias a un organizado y estructurado sistema de producción y distribución, la empresa se ha mantenido de manera firme en el mercado desde sus inicios en el año 2009. Sin embargo la competencia y el desarrollo de nuevos modelos de negocios dejan a GARWATER en desventaja frente a otras empresas que han logrado actualizar sus métodos de operación con las tendencias más actuales del mercado, es así que observando el estado operativo y financiero actual de GARWATER fue oportuno proponer la realización del estudio de sus datos, para determinar con argumentos sólidos las fortalezas y falencias del negocio, para corregir errores muchos de los cuales provenientes de los departamentos financieros en complicidad con una mala operación del software que realiza el control de las ventas, para así lograr prevenir la ocurrencia de nuevos errores y reforzar de la actividad comercial en la que se desenvuelve la empresa.

12. Análisis.

GARWATER así como otras empresas solo se enfocaba en el almacenamiento bruto de información, con la finalidad de ser usada cuando las diferentes áreas de trabajo la necesitaran. Con la puesta en marcha de esta investigación se pudo realizar el estudio de su información realizando el análisis correspondiente, que valiéndonos de PENTAHO se realizó la oportuna extracción y carga de los datos mediante flujos de trabajos que transforman una extensa base de datos transaccional a una base de datos multidimensional. Todos estos procedimientos de almacenaron en un repositorio que permite la migración de los flujos así como las conexiones con las bases de datos de donde se extrajo la información, dicho procedimiento que se enmarca en la fase de modelado de la metodología CRISP - DM.

Gracias a estos flujos de trabajo se obtuvo información sintetizada y confiable, la cual se convirtió en punto clave y estratégico para la toma de decisiones en la empresa. El estudio de esta información tuvo como objetivo generar ideas importantes tomando como origen las fuentes de información disponibles, formulando resultados sin ambigüedades, con el fin de guardar y recuperar información, la misma que se entregó a GARWATER para que efectúen sus reportes y tareas de una manera segura y con información confiable transmitida en un lenguaje sencillo ordenado, que resista a las críticas y la duda. Gracias al desarrollo de la fase de Preparación de los Datos y Modelado de la metodología CRISP - DM se logró generar reportes con contenidos coherentes, que difícilmente el software que gestionaba sus ventas podía generar, así se logró identificar el volumen de venta en sus últimos cinco años, los mejores meses, el producto más vendido el mejor vendedor, mejor cliente entre otros, en las siguientes tablas se observan los reportes anteriormente mencionados elaborados en PENTAHO.

Tabla 5
Volumen de ventas desde el 2011 hasta el 2015.

Años	Ventas Total
2011	1.452.725,48
2012	1.544.823,61
2013	1.200.808,04
2014	1.254.159,42
2015	1.311.620,08

Tabla 6
Mejores meses en ventas desde el 2011 hasta el 2015.

Años	Mejor Mes	Ventas \$[2011 – 2015]
2011	Marzo	171.737,76
	Diciembre	160.013,94
	Mayo	126.088,65
	Abril	122.823,95
2012	Mayo	178.796,17
	Marzo	167.147,01
	Abril	157.542,43
	Junio	131.217,54
2013	Diciembre	115.139,60

	Abril	109.147,32
	Mayo	105.141,22
	Marzo	101.876,65
2014	Diciembre	124.641,01
	Mayo	111.855,64
	Julio	110.311,34
	Febrero	109.709,56
2015	Diciembre	120.600,36
	Septiembre	117.718,04
	Junio	117.534,04
	Mayo	113.304,00

Tabla 7
Recaudación en ventas por producto desde el 2011 hasta el 2015.

Producto	Recaudado (\$)
Bidón de 20 Lt. (liq)	224.261,70
Botella de ½ Lt. (N° 3)	872.543,25
Botellón (Bidones para Venta)	2.707.096,16
Fundas de ½ Lt. (N° 2)	79.948,08
Galón de 4 Lt. (N° 4)	1.533.861,26
Galón de 5 Lt. (N° 5)	1.039.454,65

Tabla 8
Top 10 de Vendedores.

Vendedor GARWATER (TOP 10)	Vendido \$[2011 -2015]
Navas Arteaga Carlos Leonardo	678.839,69
Bernal Párraga Darwin Edilberto	588.799,97
García Murillo Ney Antero	557.798,44
Mera M. Cristian-veh 06	485.849,31
Zambrano Loor Colombo-veh09	418.649,38
Salazar Andrade C. Enrique-Veh14	401.280,54
Gar Water	347.834,03
Vélez Macías Manuel.Veh05	294.000,90
B. Manta	207.375,34
B. Portoviejo	193.930,60

Tabla 9
Detalle de la gestión en ventas.

Producto	PVP (ACT.)	COSTO (ACT.)	VENTAS \$ [2011 – 2015]	UNIDADES VENDIDAS
Bidón de 20 Lt. (liq)	1	0,400967	224.261,70	343495
Botella de ½ Lt. (N° 3)	0,1625	0,076516	872.543,25	1746311
Botellón (Bidones para Venta)	4,5	3,080357	2.707.096,16	2986500
Fundas de ½ Lt. (N° 2)	0,065	0,023094	79.948,08	1060386
Galón de 4 Lt. (N° 4)	0,75	0,414901	1.533.861,26	11113231
Galón de 5 Lt. (N° 5)	0,750026	0,414901	1.039.454,65	20466053

Tabla 10
TOP 10 de mejores Clientes

CLIENTES (TOP 10)	Vendido \$[2011 -2015]
García Zambrano Gloria Maribel	351,453,67
Mendoza Chávez Eudaldo Cayetano	344.792,68
Mendoza Cedeño Washington Nasael	310.765,60
Mieles Burgos Martha Kelia	220.575,71
Moreira Vásquez Juan Euclides	197.536,96
García Sabando Diego	193.573,02
Muriel Villegas Alfredo Enrique	189.512,00
Fernández Gómez Fernando Javier	166.479,80
Mejía Antón Marvin Daniel	161.974,59
CONSUMIDOR FINAL	160.006,56

Tabla 11
Origen / Punto Matriz De Venta

ORIGEN DE VENTA	Ventas \$[2011 – 2015]
B. 4	208.239,42
B. 5	165.777,75

B. MANTA	173.926,72
B. MANTA NUEVA	171.323,62
B. PLANTA	5.927.112,53
B. PORTOVIEJO	295.116,75
B. PORTOVIEJO NUEVA	174.747,18

13. Fuentes de datos primarias.

Las fuentes primarias permitieron encontrar información nueva y original, el proveedor principal de esta información fue la empresa GARWATER. Siendo una de las partes interesadas en la ejecución de esta investigación, la empresa fue la encargada de entregar la información necesaria y requerida sin excepciones, otorgando manuales de procedimiento, lista de empleados, clientes, acceso a oficinas e intercomunicador con la empresa de software GENESIS (VER ANEXO 10) que gracias a la disposición del gerente y dando paso a la investigación se facilitó un técnico de planta para el acceso a los datos del software contable en el servidor principal de la empresa embotelladora de agua. Para la manipulación del software PENTAHO y WEKA, se acudió a manuales y tutoriales de páginas oficiales, sin desmerecer la gran ayuda que proporcionaron los foros y la comunidad de operadores que usan estas herramientas para el estudio y análisis de sus datos.

14. Fuentes de datos secundarias.

Las fuentes secundarias toman índice a la estructura y formación de esta tesis, de la cual se sirvió la redacción de este documento para lo cual se revisaron artículos científicos, tesis, revistas y páginas web para estudiar las tendencias en inteligencia de negocios, el uso y la aplicación de las herramientas expuestas en este escrito. Dichas fuentes se encuentran inscritas en la bibliografía en la parte final del documento, ubicada con el objetivo de reconocer el trabajo de otros autores en materia de conocimiento del tema planteado en esta investigación.

15. Detección de modelos por comparación de las fuentes de datos.

El descubrimiento de un modelo de minería de datos estuvo en función de las pocas variables encontradas en el estudio inicial, esto ocurre debido a que la empresa posee un volumen extenso de registros en sus bases de datos, sin embargo la forma en la cual fueron ingresados manifestó la mala operación del software contable omitiendo datos de gran valor para los estudios, además de registrar información duplicada, incomprensible e incompleta, de tal manera que través de los años no se realizaron los correctivos necesarios permitiendo continuar con el mismo protocolo de ingreso de datos, definiendo una limitada detección del modelo de minería de datos a seguir.

Durante el desarrollo de la fase de Modelado en la metodología CRISP – DM es importante añadir que la inconsistencias en información demográfica, temporal y en los datos fue un limitante para el desarrollo libre de los modelos de minería de datos empleados en el estudio, más aun cuando se realizó la tarea de ajustar las variables significantes representadas en el entorno OLAP, desde aquí se analizaron los datos y de manera cierta se logró aplicar un modelo de minería de datos que atendió las necesidades de generar el nuevo conocimiento para la empresa. Con un panorama limitado de variables, fue difícil aplicar todo el potencial de la herramienta WEKA sobre los datos generados en el entorno multidimensional, dado este caso y sin alejarse de los objetivos planteados en esta investigación fue optimo seleccionar un modelo PREDICTIVO aplicando el algoritmo de regresión lineal de la herramienta WEKA 3.7.13 cuya representación se expone en la figura 19.

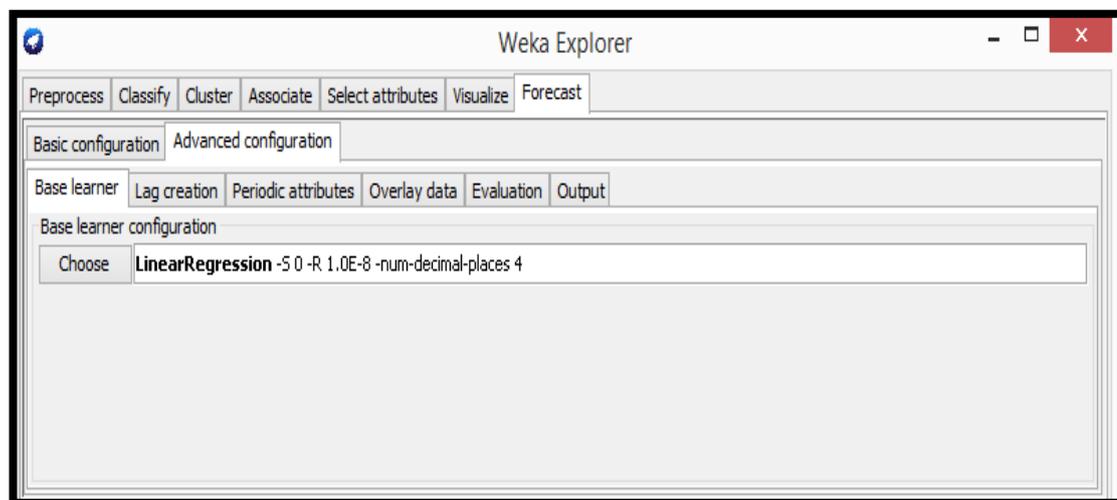


Figura 21. Algoritmo de regresión lineal WEKA.

Para el entrenamiento de modelo se realizó el diseño de un flujo de trabajo generado con PENTAHO recibiendo como entrada las ventas anuales de la empresa, obteniendo como salida un archivo con extensión “.arf” el cual contuvo el filtrado de las ventas desde el año 2011 hasta el año 2015, datos que fueron propuestos para el análisis de este estudio.

El objetivo del modelo entrenado es de predecir un nuevo periodo en las ventas de la empresa, para lo cual la herramienta muestra de forma gráfica los resultados del algoritmo Forecasting, mismo que se puede apreciar en la figura 20, esta figura está acompañada de un reporte textual de la configuración del algoritmo y datos estadísticos del modelo generado por el aplicativo.

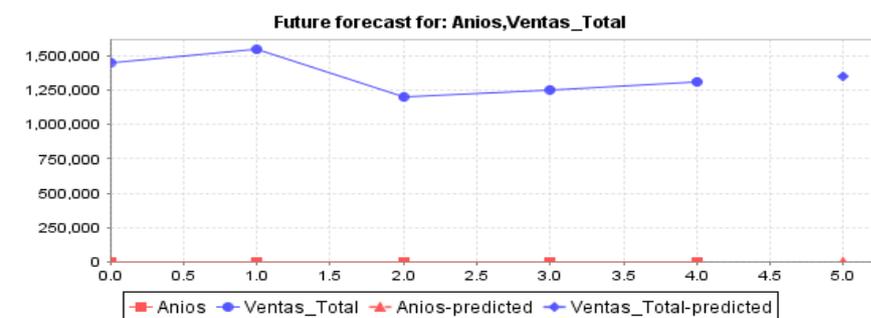


Figura 22. Gráfica del Algoritmo Forecasting.

=== Run information ===

Scheme:

LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4

Lagged and derived variable options:

-F "[Anios, Ventas_Total]" -L 1 -M 2 -G Anios

Relation: *C:\Users\CELI\Desktop\Reportes\forecasting.arff*

Instances: *5*

Attributes: *2*

Anios

Ventas_Total

Transformed training data:

Anios

Ventas_Total

Lag_Anios-1

Lag_Anios-2

Lag_Ventas_Total-1

Lag_Ventas_Total-2

Anios^2

Anios^3

*Anios*Lag_Anios-1*

*Anios*Lag_Anios-2*

*Anios*Lag_Ventas_Total-1*

*Anios*Lag_Ventas_Total-2*

Anios:

Linear Regression Model

Anios =

*-0.0233 * Lag_Anios-2 +
0 * Lag_Ventas_Total-2 +
0.0004 * Anios^2 +
-0 * Anios^3 +
0 * Anios*Lag_Anios-2 +
792.9197*

Ventas_Total:

Linear Regression Model

Ventas_Total =

```
+  
1352827.326  
  
=== Future predictions from end of training data ===  
inst#           Anios Ventas_Total  
2011             2011  1452725.48  
2012             2012  1544823.61  
2013             2013  1200808.04  
2014             2014  1254159.42  
2015             2015  1311620.08  
2016*           2016.0003 1352827.326
```

Figura 23. Código generado por WEKA – FORECASTING.

CAPÍTULO V

ANÁLISIS DE RESULTADOS

16. Patrones obtenidos.

Dentro de la Metodología CRISP – DM se procesan los modelos desarrollados enmarcados en la fase de Evaluación y Despliegue, continuando con esta metodología en este punto de la investigación encontramos que debido a la lógica del software que maneja el sistema de facturación y contabilidad de GARWARTER, fue limitada la obtención de variables que tuviesen participación en la aplicación de algoritmos de minería de datos. Sin embargo fue oportuno y necesario el uso un algoritmo de regresión lineal **Forecasting** proporcionado por WEKA, rescatando de las pocas variables obtenidas las más influyentes sobre la información de la empresa, gracias a los resultados de aplicar este algoritmo se reflejó el buen estado en el cual se mantiene la empresa en torno a sus ventas y como serán sus ganancias en el año 2016 en el caso de seguir manteniendo el ritmo de ventas presentado hasta el año 2015.

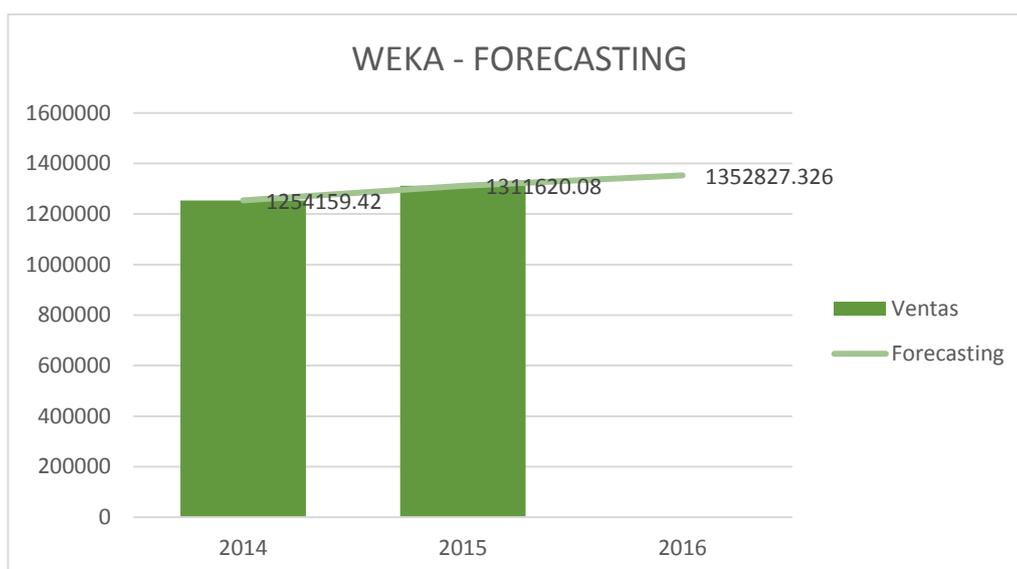


Figura 24. Análisis del algoritmo Forecasting de WEKA.

Sumando a este análisis se tiene el estudio realizado a los datos, producto de la extracción, limpieza y creación del entorno OLAP a partir de las bases de datos transaccionales del software GENESIS (VER ANEXO 10), se puede obtener reportes donde claramente se lograron observar patrones entorno a las ventas, los clientes y a sus vendedores, en la cuadro 5 se logra identificar los meses en donde las ventas se disparan. En los meses de mayo y diciembre, este patrón tiende a repetirse en los últimos 5 años lo que indica acrecentar la producción y optimizar la distribución de los productos durante estos meses para incrementar las ventas.

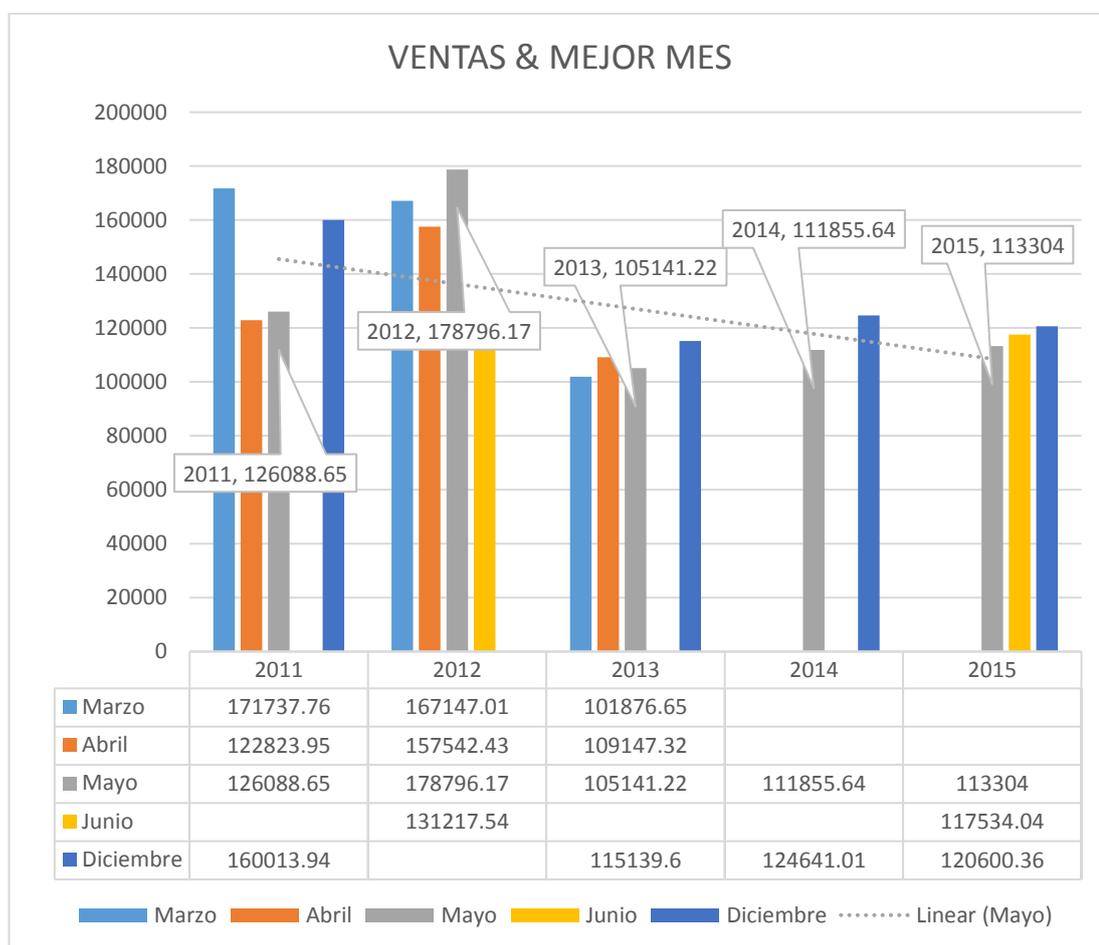


Figura 25. Contraste de ventas y su mejor mes en el periodo 2011-2015.

Como se puede visualizar en el gráfico durante estos últimos cinco años el mes que se mantiene en dentro del top 4 de mejores ventas es el mes de mayo

seguido por el mes de diciembre, que si bien no asoma en el año 2012 dentro de los 4 mejores meses se mantiene en los otros años. Sin duda este patrón se mantendrá en los próximos años de mantener el estado actual en el ejercicio de vender agua embotellada. Otro comportamiento detectado gira en torno a los clientes y los vendedores donde se resaltaron a los mejores en su ámbito, sin duda este patrón para la empresa genera un valor muy importante donde se pueden impulsar ofertas, descuentos, premios y promociones establecidas por GARWATER para recompensar la atención de sus clientes y empleados. En la figura 21 y 22 se puede observar a detalle lo descrito.



Figura 26. Mejores compradores en el periodo 2011 – 2015.

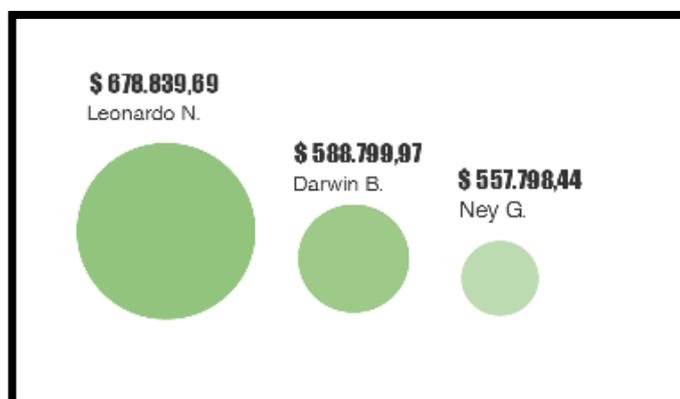


Figura 27. Mejores vendedores en el periodo 2011 – 2015.

Sin duda alguna la empresa conoce cuál es su producto estrella y en contraste con el análisis de los datos se corroboró que sus principales ventas van de la mano de los productos y artículos detallados en la figura 23.

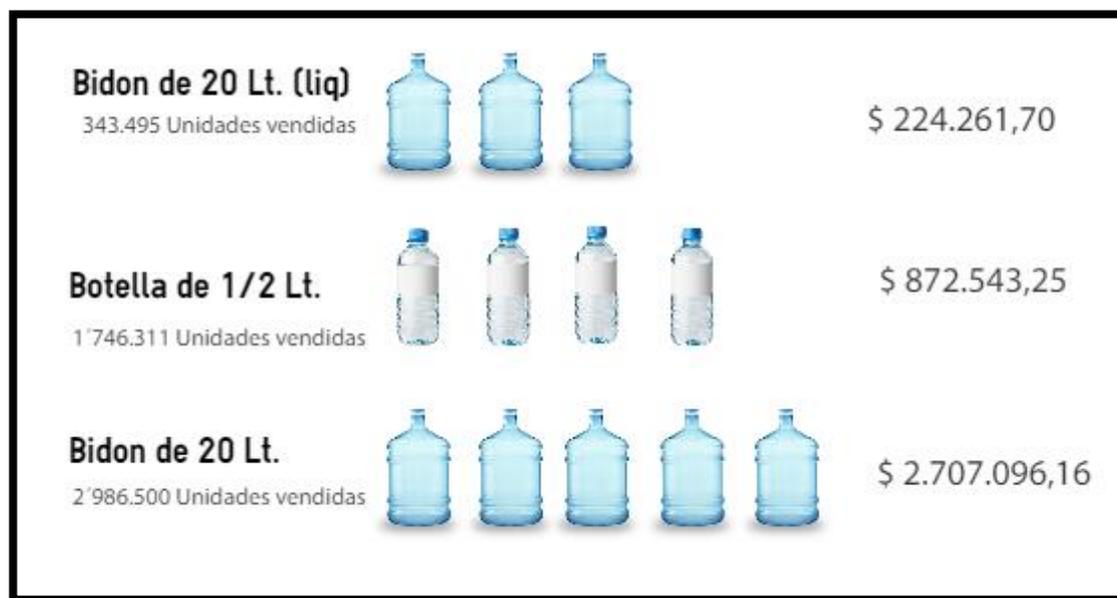


Figura 28. Productos de mayor influencia en las ventas de GARWATER.

Otro patrón importante encontrado se encuentra al observar los dos puntos más grandes de distribución en las ventas, como se puede observar en la Figura 24 que la ponderación se inclina ante la planta principal, en conclusiones simples parecería que son pocos los clientes en las cercanías de manta que consumen los productos de GARWATER, sin embargo la realidad es muy lejana pues la empresa se siente muy cómoda distribuyendo los productos desde Portoviejo a modo de estrella hacia las otras localidades de la provincia de Manabí.

Se puede pensar que GARWATER se acogió a la resistencia del cambio y tomo su zona de confort al no expandir nuevas plantas de producción que permitan ahorrar recursos y mejorar las entregas. Sin embargo el servicio ha permanecido con un control tan eficiente que no ha sido necesario la implementación de nuevos modelos de negocios en la empresa.

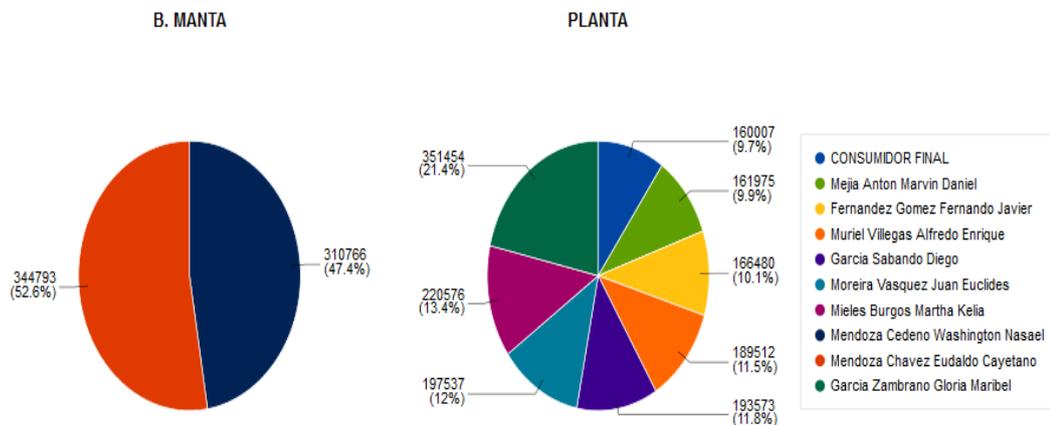


Figura 29. Ponderación de ventas en las plantas de distribución con muestra de mejores clientes.

Para lograr una mejor comprensión a lo anteriormente dicho se diseñó la siguiente Figura donde se puede observar el alcance de distribución que mantenía la planta principal, que siguiendo estándares propios logro mantener el mercado centralizando la producción

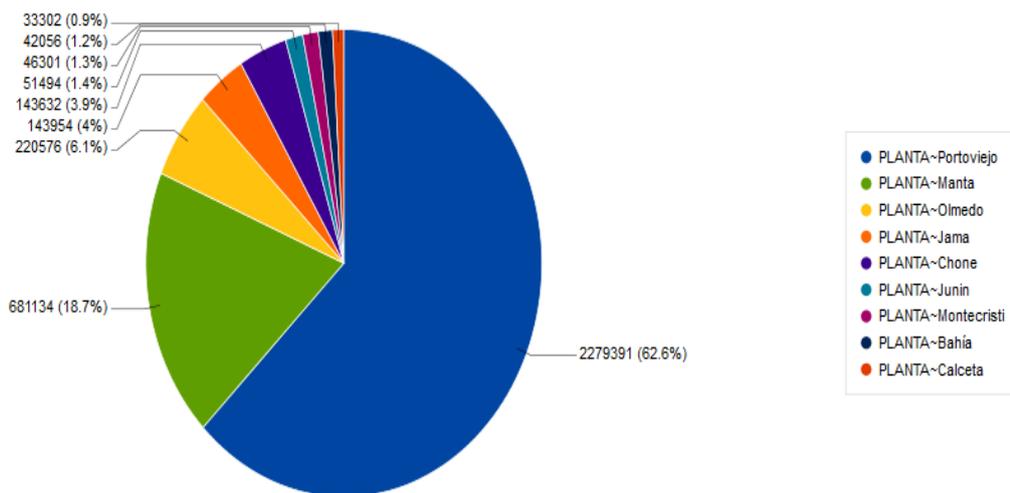


Figura 30. Distribución de ventas desde PLANTA PRINCIPAL.

Es apreciable la manera en que la forma centralizada aparentemente da sus frutos, sin estimar los gastos y riesgos a los que se expone esta forma de trabajo. En la Figura 26 se logra apreciar el comportamiento del punto de distribución de Manta y es claro notar que abarca menos cantones que a su efecto producen un número menor de ventas.

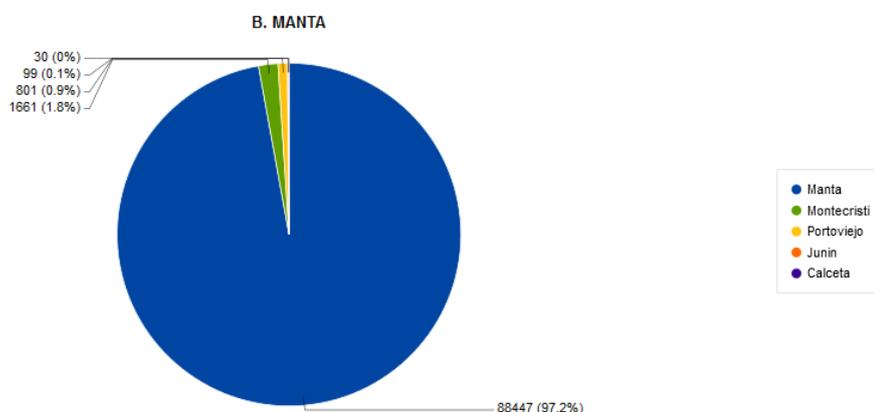


Figura 31. Distribución de ventas desde PLANTA MANTA.

Finalmente se consiguió comprender como se comportan las ventas y con la Figura 27 se puede observar como la necesidad de adquirir un producto de GARWATER es diferente en cada ciudad y gracias al estudio de los gráficos se pudo tener una percepción real conociendo a fondo el perfil de la población dando como resultados variaciones importantes y significativas, mismas que permitieron tomar decisiones gerenciales.

La distribución de la gráfica nos muestra los cantones de la provincia de Manabí donde consumen los productos de la empresa, se puede observar que cada cantón tiene un comportamiento distinto a la hora de consumir, se concluye que gracias a estos patrones se logrará establecer campañas optimizadas en función de las preferencias de los clientes ahorrando esfuerzos y recursos.

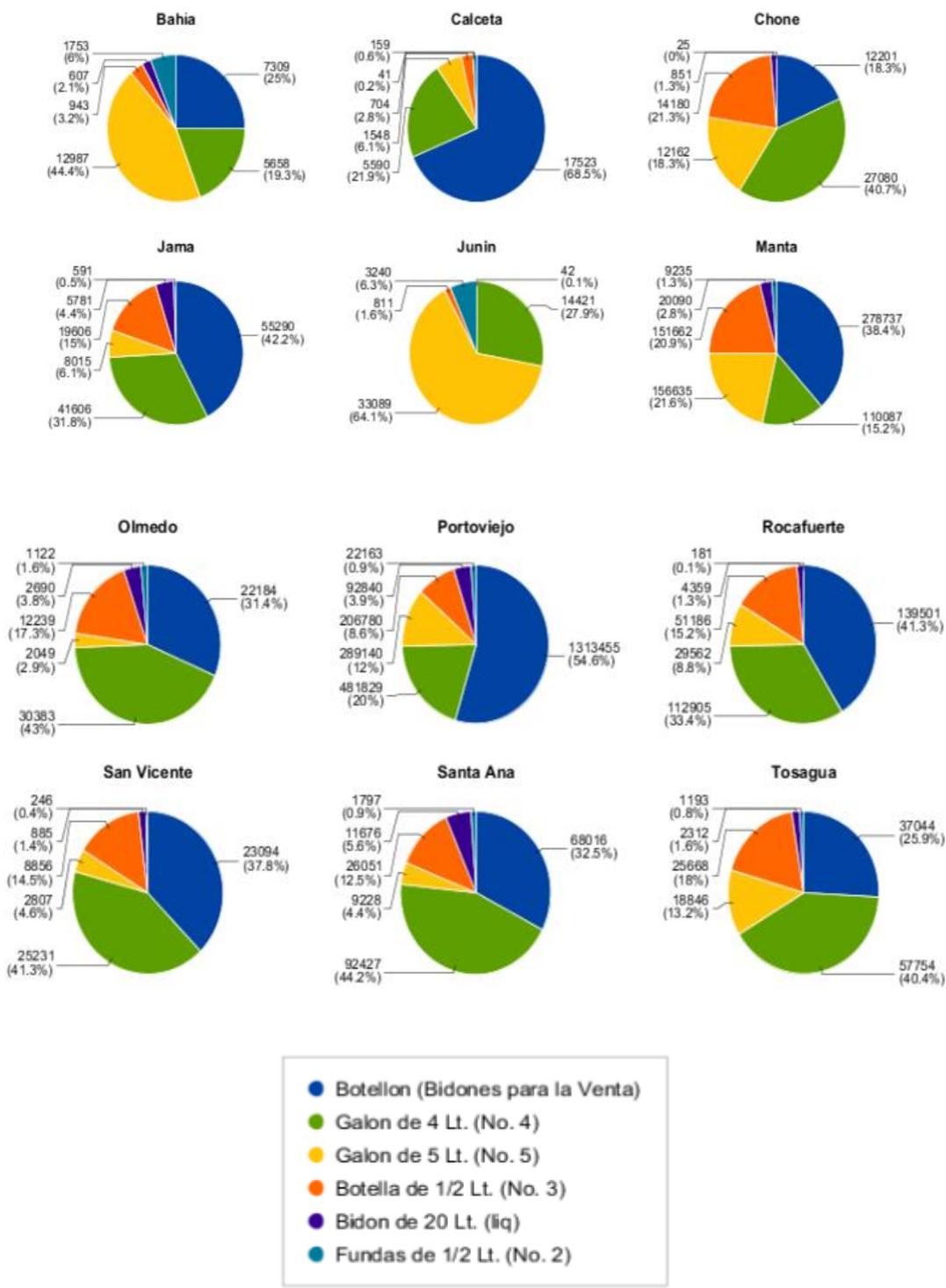


Figura 32. Comportamiento de ventas en consumo de productos GARWATER en los cantones de Manabí.

Otra tendencia se descubrió a partir de datos aparentemente olvidados pero que tienen la pieza clave para expandir la marca fuera de la provincia y llevarlas a otras como lo son Pichincha y Guayas, como se puede observar la demanda de los productos fue requerida en las fechas que especifica la Figura 28, donde claramente nos muestra que se realizaron compras a nombre de clientes con raíces en las provincias ya nombradas, sin embargo no se realizaron las operaciones correctas para sostener la constancia de estas ventas donde sin duda existirá la posibilidad de establecer nuevos negocios y generar nuevas rentas a la empresa sin olvidar las oportunidades que surgirán a raíz de enfocar esfuerzos para recuperar a estos clientes.

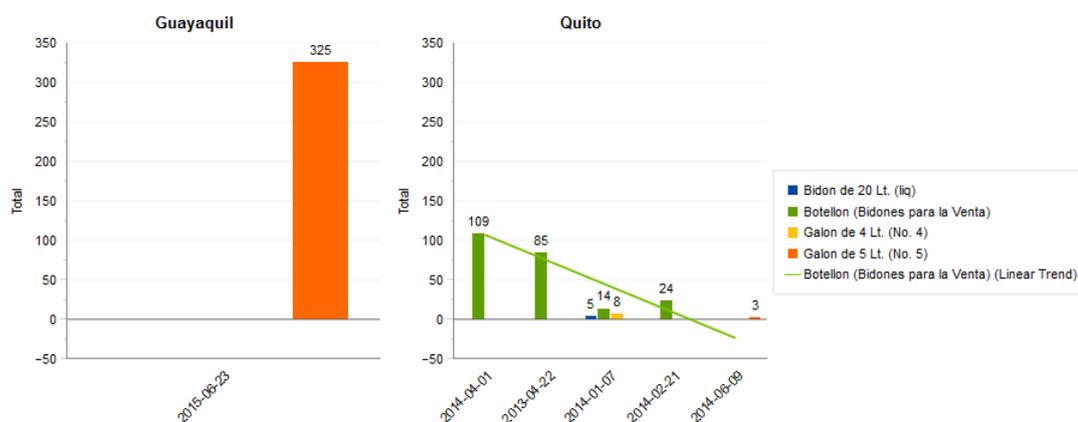


Figura 33. Ventas realizadas por clientes de la provincias del GUAYAS y PICHINCHA.

Es importante premiar el rol de los vendedores, sin embargo estudiar el comportamiento de estos en función de los productos que venden, cataloga cual es el artículo que tiene más éxito y es más fácil de vender, es así que tomando a los empleados más eficientes se obtuvo un tendencia a vender Bidones sin líquido, es una actividad que denota que se está llegando a más clientes, desde una tienda de barrio hasta los hogares en los diferentes cantones de la provincia de Manabí. A

pesar que existen empleados que venden exitosamente otros productos, las ganancias más fuertes tienden los Bidones en estos últimos 5 años, factor favorable para tomar en cuenta en futuras decisiones.

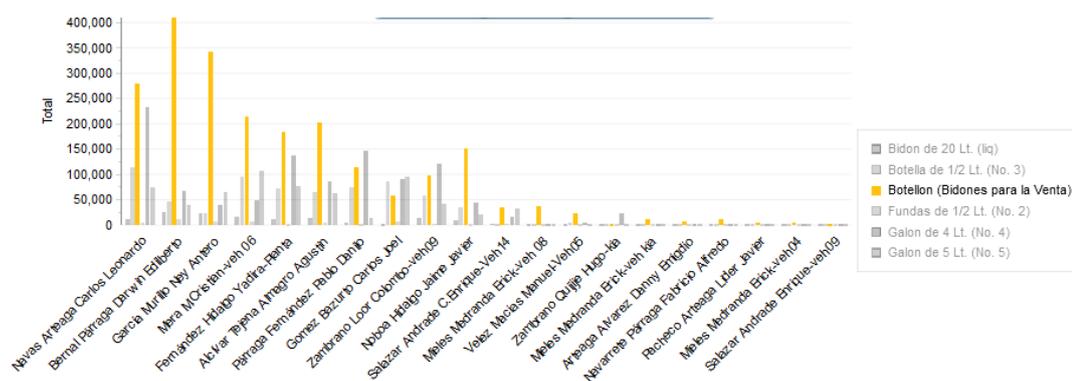


Figura 34. Comparación de artículos más vendidos en función de los mejores vendedores.

17. Demostración de la hipótesis.

Durante el desarrollo de esta investigación fue notorio observar que GARWATER nunca logró realizar un análisis total de la información referente a las ventas registradas desde el año 2011 hasta el año 2015. En vista de los antecedentes expuestos en la redacción de este documento, fue viable la aplicación de técnicas en minería de datos donde se recopiló y se estandarizó la información guardada en sus bases de datos, construyendo complejos flujos de procesos en PENTAHO para extraer registros válidos necesarios para el uso del algoritmo de regresión lineal Forecasting operado en WEKA.

En los resultados obtenidos por el algoritmo de regresión lineal se pudo evidenciar que para el 2016 la empresa tendría un crecimiento considerable en sus ventas, estableciendo una tendencia notable en las recaudaciones del nuevo periodo contable. Esta tendencia se justificó al construir dashboards entregables para la gerencia donde interactuó PENTAHO con el entorno OLTP, creado exclusivamente para GARWATER con el fin analizar el presente y el pasado de los datos vinculados

a las ventas, registrando los patrones descubiertos para su utilización en la toma de decisiones que beneficien al negocio.

Estos descubrimientos fueron incorporados como un activo clave para darle la dirección a la cual la empresa GARWATER debe apuntar si desea incrementar las ventas de agua embotellada, abanderando la marca con el propósito de extender las zonas de distribución, incrementar la clientela y conocer las necesidades de los consumidores actuales priorizando la calidad del producto para obtener una referencia coherente con las campañas de marketing que se realicen a futuro, gracias a estos resultados se logró afirmar que fue posible encontrar patrones y tendencias con la aplicación de técnicas de minería de datos consiguiendo predecir las ventas del año 2016 con la ejecución del algoritmo Forecasting.

18. Aplicación de resultados.

Los resultados obtenidos en esta investigación fueron presentados al gerente general, autoridades, área de ventas de la empresa y un asesor del software GENESIS (VER ANEXO 10), en un inicio se mantuvieron reuniones con las entidades involucradas en el proyecto puntualizando las actividades del plan de minería de datos que determinó la presentación de los descubrimientos de patrones y tendencias.

Terminado el estudio se expuso formalmente la investigación la cual puso a disposición toda la documentación recopilada sobre los resultados obtenidos, detallando los procesos, herramientas y técnicas de minería de datos utilizadas sirviendo como medio para presentar el comportamiento de las ventas desde año 2011 hasta el año 2015, prediciendo con un algoritmo de regresión lineal las ventas totales para el periodo 2016.

La directiva en general asumió los resultados como una oportunidad de innovar, mejorar y corregir los errores que permitan darle crecimiento a GARWATER, es así que se realizó un convenio con la empresa GENESIS (VER ANEXO 10) para capacitar al personal en buenas prácticas para el ingreso y el

mantenimiento de la información, se implementó un plan de negocios que involucra campañas de marketing en las zonas donde el producto tiene más demanda, cambiando la imagen en el etiquetado del producto embotellado.

Se realizó el contrato de una empresa de publicidad para crear y administrar los medios digitales como su página web y sus redes sociales, se contrató la construcción de un software que administre la producción registrando información para un futuro estudio en el comportamiento interno en los procesos de fabricación, las autoridades finalmente establecieron un enfoque nuevo para la empresa al considerar la premiación al cumplimiento, dedicación y fidelidad de los clientes y empleados que basados en los informes de esta investigación se encontró con personas que han dejado marcas en la historia de la empresa y que su ímpetu sea de ejemplo al recibir descuentos, premios y una notable vinculación con la empresa como un ente familiar de confianza, proyectando un buen mensaje a la población esperando crecer de la mano de las personas que formar GARWATER.

CAPÍTULO VI

CONCLUSIONES Y RECOMENDACIONES

19. Conclusiones

Para el término de esta investigación se establecieron las siguientes conclusiones:

- Se logró aplicar el algoritmo de regresión lineal Forecasting para predecir las ventas de GARWATER en el año 2016, partiendo de la información extraída y estructurada en un entorno OLAP que almacena las ventas del rango 2011 – 2015, pronosticando que existirá un incremento en las ventas de 3.14% equivalentes a 412017.24 dólares más en relación al año 2015.
- En el rango de tiempo establecido en la captura de datos para el análisis de las ventas de GARWATER, muestran que a pesar de no contar con un plan de negocios formal y de recurrir a operaciones categorizadas como empíricas, la empresa mantuvo un crecimiento considerable en los años 2011 y 2012, para luego descender en un 22.26% en el año 2013. La empresa realizó reajustes y en base a las pérdidas logra reponerse en un 4% en el año 2014 manteniendo para el 2015 el mismo incremento del 4%, en conclusión la empresa seguirá manteniendo una débil recuperación en relación al 2012 de no tomar las acciones acertadas que permitan recuperar a la empresa y generar más ventas.
- Con los flujos de trabajos realizados en PENTAHO sobre las bases de datos transaccionales se logró observar que los registros en su gran mayoría mantienen errores de escritura realizado por los usuarios, que combinados con malas configuraciones limitaron las variables para analizar y utilizar en la implementación de algoritmos complejos de minería de datos. Este descubrimiento permitió plantear nuevas políticas en la empresa para el uso de cualquier software que registre información que describa el comportamiento de la empresa en sus diferentes áreas operativas.

- Se descubrió que las ventas de los productos GARWATER cambian totalmente según las preferencias de la población tomando como referencia una distribución demográfica, obligando a la empresa a optimizar la entrega de productos a los diferentes cantones de la provincia de Manabí y efectuar las campañas publicitarias necesarias para fomentar el consumo de los productos menos vendidos.
- La empresa mantuvo ventas fuera de la provincia, sin embargo la dirigencia actual desconoció las actividades vinculadas a este proceso, fruto de este descubrimiento se realizaron investigaciones que determinaron que GARWATER vendía productos sin etiquetar a las ciudades de Quito y Guayaquil, razón valedera para incorporar a los objetivos de expansión vender los productos de la empresa a otras provincias, implicando la construcción de nuevas plantas.
- La implementación de esta tesis marco un hito en GARWATER, donde se replantearon las metas y los objetivos necesarios para expansión y crecimiento de la misma con la puesta en marcha de un plan de negocios basado y respaldado en las debilidades y fortalezas detectadas en este estudio. La labor de esta investigación no marca un fin lucrativo para el autor, en su contexto aporta lo mejor de su conocimiento adquirido en materia de minería de datos para el progreso y resurgimiento del cantón Portoviejo y la provincia de Manabí.

20. Recomendaciones.

Finalizando esta investigación se establecieron en consideración de la situación de la empresa, las siguientes recomendaciones:

- La empresa debe crear como política interna la implementación de capacitaciones periódicas, dedicado al personal de ventas sin importar la

función o el rol que cumplan en torno a la acción de vender, con fin de sincronizar la recopilación de información que gire alrededor del cliente y su compra, así de esta manera poder incluir datos que generen nuevos conocimientos.

- Se recomienda dialogar con la empresa desarrolladora del software GENESIS (VER ANEXO 10) para establecer convenios integrales, donde esta empresa se comprometa a efectuar actualizaciones exclusivas para GARWATER. Actualizaciones que cumplirán la tarea de satisfacer los nuevos requerimientos a nivel de interfaces, captura de información y base de datos, con el fin de facilitar el ingreso de nueva información, optimizando la base de datos transaccional de la cual se servirá el entorno OLAP dedicado a descubrir nuevo conocimiento con la ayuda las herramientas de análisis, descubrimiento de tendencias, patrones y el uso de nuevos algoritmos de minería de datos.
- Crear un departamento dedicado al control, monitoreo y mantenimiento de la infraestructura tecnológica así como los repositorios digitales de información que posee GARWATER.
- Adquirir los servicios de un especialista que asista a las necesidades que respondan a la inteligencia de negocios, con el suficiente conocimiento y experiencia en el área de la producción de bebidas embotelladas.
- Supervisar periódicamente que las actividades entorno a las ventas reflejen su situación real en los repositorios de información digital, considerando esta tarea como fundamental si se quiere controlar el correcto recibimiento de la información a las bases de datos de la empresa.

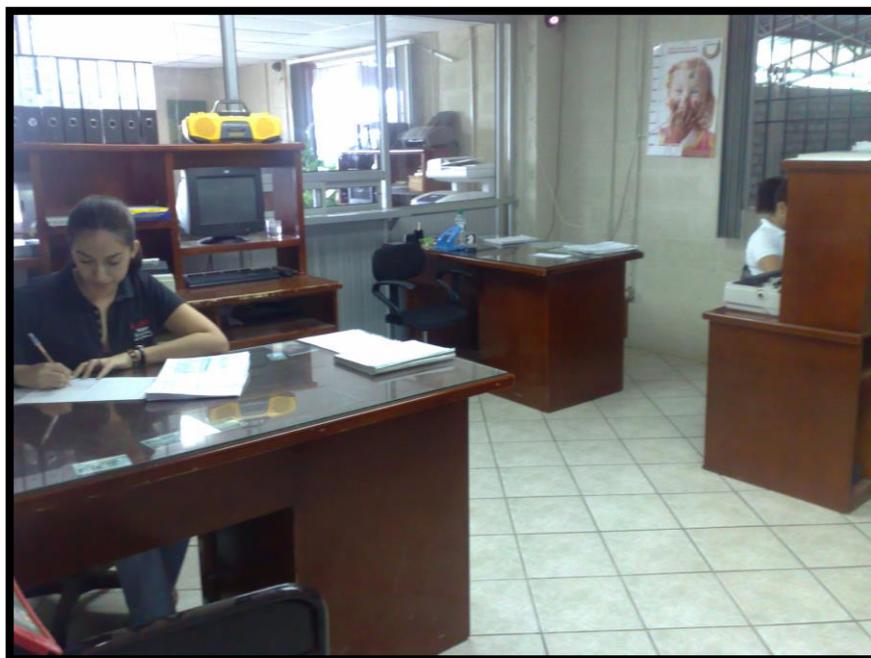
ANEXOS



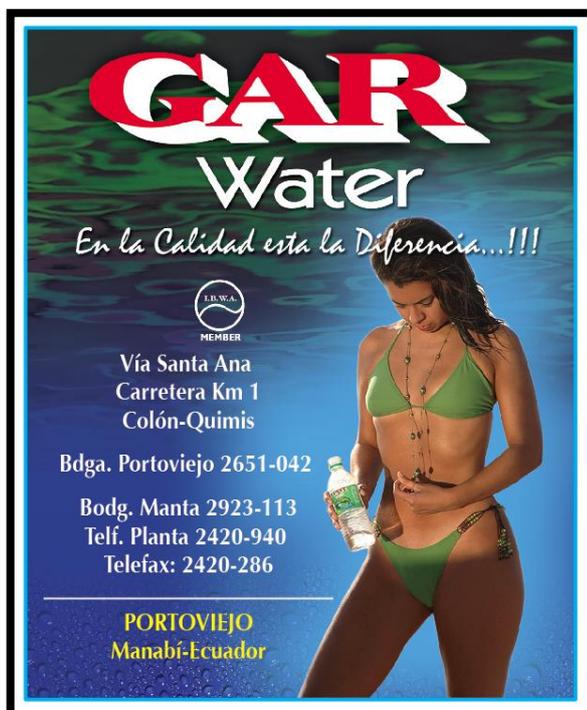
Anexo 1. Productos que ofrece GARWATER.



Anexo 2. Operador llenando de líquido los Bidones.



Anexo 3. Personal del Área de Ventas



Anexo 4. Publicidad de GARWATER



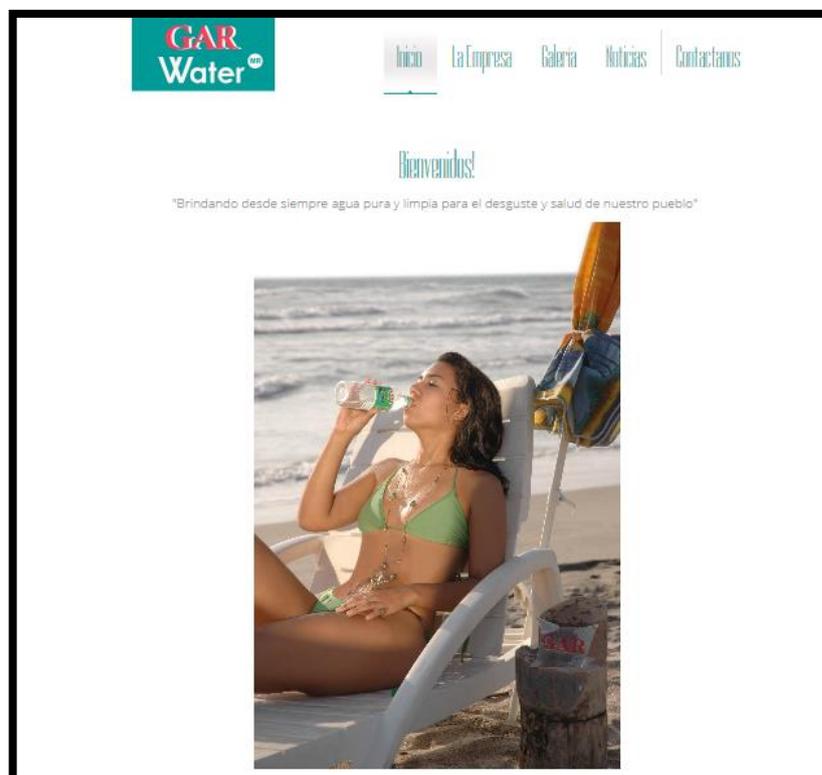
Anexo 5. Área de Llenado



Anexo 6. Entrevista con el Gerente de GARWATER



Anexo 7. Carga de productos a camiones de distribución.



Anexo 8. Página web de GARWATER (www.garwater.com).



Anexo 9. Facebook de GARWATER

Anexo 10. Sitio web de software GENESIS (VER ANEXO 10)

([http://sistemasGENESIS \(Ver anexo 10\).com.ec](http://sistemasGENESIS (Ver anexo 10).com.ec))

BIBLIOGRAFÍA

CreceNegocios. (7 de marzo de 2016). *CreceNegocios*. Obtenido de <http://www.crecenegocios.com/la-cadena-de-valor-de-porter/>

Delgado Calero, L. R., & GARWATER. (6 de enero de 2016). *GARWATER - Historia, Políticas, La empresa*. Portoviejo, Manabí, Ecuador.

Domínguez, Y. S. (2007). *El análisis de información y las investigaciones cuantitativa y cualitativa*. Obtenido de Rev Cubana Salud Pública: http://bvs.sld.cu/revistas/spu/vol33_3_07/spu20207.htm

ECURED. (6 de febrero de 2016). *DIAGRAMA ENTIDAD RELACIÓN*. Obtenido de http://www.ecured.cu/Diagrama_Entidad_Relaci%C3%B3n

Fabrizi, M. S. (s.f.). *Las técnicas de la investigación: La observa*. Obtenido de <http://www.fhumyar.unr.edu.ar/escuelas/3/materiales%20de%20catedras/trabajo%20de%20campo/solefabri1.htm>

Gartner. (13 de marzo de 2016). *Magic Quadrant for Business Intelligence and Analytics Platforms*. Obtenido de <https://www.gartner.com/doc/reprints?id=1-2XXET8P&ct=160204>

IBM. (13 de marzo de 2016). *IBM*. Obtenido de IBM Knowledge Center: http://www.ibm.com/support/knowledgecenter/SS3RA7_17.1.0/modeler_crispdm_ddita/modeler_crispdm_ddita-gentopic1.dita

Lantares. (2014). *Transformación Digital de los negocios, analizando el cambio*. Obtenido de <http://www.lantares.com/blog/modelos-de-data-mining-y-las-herramientas-mas-usadas>

Microsoft. (9 de marzo de 2016). *Microsoft Developer Network*. Obtenido de [https://msdn.microsoft.com/es-es/library/cc645779\(v=sql.120\).aspx](https://msdn.microsoft.com/es-es/library/cc645779(v=sql.120).aspx)

Milla, R. E. (2010). *El Rincon del BI: Descubriendo el Business Intelligence*. Obtenido de <https://churriwifi.wordpress.com>

MORENO A., F. J. (2009). Estado del arte de los modelos multidimensionales espacio temporales. *bdigital Universidad Nacional colombia*.

Plata, C. G. (15 de ABRIL de 2008). *TRABAJO INDIVIDUAL N° 01: METODO DEDUCTIVO Y METODO INDUCTIVO*. Obtenido de <http://colbertgarcia.blogspot.com/2008/04/metodo-deductivo-y-metodo-inductivo.html>

Rodriguez, O. (12 de marzo de 2016). *Metodología para el desarrollo de proyectos en minería de datos CRISP - DM*. Obtenido de Basado en la tesis: "Metodología para la definición de requisitos en proyectos de data mining ER - DM de Jose GallardoArancibia:
http://www.oldemarrodriguez.com/yahoo_site_admin/assets/docs/Documento_CRISP-DM.2385037.pdf

Trujillo, A. C. (2006). Modelo multidimensional. *Fundación Dialnet - Ingeniería Industrial Vol 27 N° 1, 4*.

Yoshibauco. (27 de abril de 2011). *Blog Wordpress Algoritmos TDIDT aplicado al Analisis de suelo*. Obtenido de <https://yoshibauco.wordpress.com/2011/04/27/empezando-con-las-etapas-de-crisp-dm/>