



**FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS
COMPUTACIONALES**

TEMA:

**Elaboración de un Prototipo Predictivo para la Toma de Decisiones
de la Gestión de Compras del Área de Importaciones con la
Generación de un Datamart y Técnicas de Datamining para la
Empresa Frutera Don Marcos S.A. (FRUDONMAR)**

AUTOR:

GARRIDO GARCÍA, FAUSTO JAVIER

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL
TÍTULO DE INGENIERA EN SISTEMAS COMPUTACIONALES**

TUTORA:

ING. ZURITA FABRE, ADELA, Ph.D.

**Guayaquil, Ecuador
Marzo de 2018**



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS
COMPUTACIONALES**

CERTIFICACIÓN

Certificamos que el presente trabajo de titulación **Elaboración de un prototipo predictivo para la toma de decisiones de la gestión de compras del área de importaciones con la generación de un Datamart y técnicas de Datamining para la empresa Frutera Don Marcos S.A. (FRUDONMAR)**, fue realizado en su totalidad por **Garrido García Fausto Javier**, como requerimiento para la obtención del título de **Ingeniero en Sistemas Computacionales**.

TUTORA

Ing. Zurita Fabre Adela, Ph.D.

DIRECTORA DE LA CARRERA

Ing. Guerrero Yépez Beatriz del Pilar, Mgs.

Guayaquil, a los cinco días del mes de marzo del año 2018



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS
COMPUTACIONALES**

DECLARACIÓN DE RESPONSABILIDAD

Yo, Garrido García, Fausto Javier

DECLARO QUE:

El Trabajo de Titulación **Elaboración de un prototipo predictivo para la toma de decisiones de la gestión de compras del área de importaciones con la generación de un Datamart y técnicas de Datamining para la empresa Frutera Don Marcos S.A. (FRUDONMAR)**, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias bibliográficas. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

Guayaquil, a los cinco días del mes de marzo del año 2018

EL AUTOR

Garrido García, Fausto Javier



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS
COMPUTACIONALES**

AUTORIZACIÓN

Yo, Garrido García, Fausto Javier

Autorizo a la Universidad Católica de Santiago de Guayaquil a la **publicación** en la biblioteca de la institución del Trabajo de Titulación **Elaboración de un prototipo predictivo para la toma de decisiones de la gestión de compras del área de importaciones con la generación de un Datamart y técnicas de Datamining para la empresa Frutera Don Marcos S.A. (FRUDONMAR)**, cuyo contenido, ideas y criterios son de mi exclusiva responsabilidad y total autoría.

Guayaquil, a los cinco días del mes de marzo del año 2018

EL AUTOR

Garrido García, Fausto Javier



UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL

FACULTAD DE INGENIERÍA CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

REPORTE SISTEMA ANTIPLAGIO

URKUND

Documento: [GARRIDO GARCIA FAUSTO-febrero 26 2018.docx](#) (D35967010)

Presentado: 2018-02-27 09:14 (-05:00)

Presentado por: adela.zurita.fabre (adela.zurita@cu.ucsg.edu.ec)

Recibido: adela.zurita.ucsg@analysis.urkund.com

Mensaje: [Mostrar el mensaje completo](#)

0% de estas 31 páginas, se componen de texto presente en 0 fuentes.

Categoría	Enlace/nombre de archivo
	GARRIDO GARCIA FAUSTO-febrero 26 2018.docx
	GARRIDO GARCIA FAUSTO-febrero 2018.docx
Fuentes alternativas	
	GARRIDO GARCIA-avance noviembre 22.docx
	VERA LETAMENDI_FINAL.docx
	http://culturacrm.com/data-mining/cinco-herramientas-data-mining/

100% #1 Activo Archivo de registro Urkund: Universidad Católica de Santiago de Guayaquil / GARRIDO GARCIA FAUST... 100%

FACULTAD DE INGENIERÍA CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

TEMA: Elaboración de un Prototipo Predictivo para la Toma de Decisiones de la Gestión de Compras del Área de Importaciones con la Generación de un Datamart y Técnicas de Datamind para la Empresa Frutera Don Marcos S.A. (FRUDONMAR)

AUTOR: GARRIDO GARCÍA, FAUSTO JAVIER

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERA EN SISTEMAS COMPUTACIONALES

TUTORA: ING.

ZURITA FABRE, ADELA, Ph.D.

Guayaquil, Ecuador Marzo de 2018

UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL

FACULTAD DE INGENIERÍA CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

CERTIFICACIÓN

Certificamos que el presente trabajo de titulación

AGRADECIMIENTO

A la empresa Frutera Don Marcos (Frudonmar) S.A. y mi tutora Ing. Adela Zurita por habernos brindado su apoyo y guía en el desarrollo de este trabajo de titulación

Fausto Javier Garrido García

DEDICATORIA

Este trabajo de titulación va dedicado a mis padres y hermano que han sido el apoyo principal en mi formación académica, que gracias a su apoyo y consejos me han ayudado a seguir adelante hasta la obtención de mi objetivo deseado.

Fausto Javier Garrido García



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS
COMPUTACIONALES**

TRIBUNAL DE SUSTENTACIÓN

Ing. Guerrero Yépez Beatriz del Pilar, Mgs.
DIRECTORA DE LA CARRERA

Ing. Yanza Montalván Ángela Olivia, Mgs.
COORDINADORA DEL ÁREA

Ing. Salazar Tovar César Adriano, Mgs.
OPONENTE

Índice General

RESUMEN.....	xviii
INTRODUCCIÓN	1
CAPÍTULO I.....	3
EL PROBLEMA	3
PLANTEAMIENTO DEL PROBLEMA.....	4
Delimitación del Problema	4
Evaluación del Problema	5
OBJETIVOS.....	5
General.....	5
Específicos.....	6
ALCANCE DE LA INVESTIGACIÓN	6
JUSTIFICACIÓN E IMPORTANCIA	6
DOMINIO Y LÍNEA DE INVESTIGACIÓN.....	7
CAPÍTULO II	8
MARCO TEÓRICO, CONCEPTUAL Y LEGAL	8
FUNDAMENTACIÓN TEÓRICA Y CONCEPTUAL.....	8
La gestión de inventarios, como base para la toma de decisiones.....	9
Modelos para la gestión de inventarios	10
Datos e información.....	11
Datawarehouse.....	12
Datamart	13
Datamart OLAP	15
Datamart OLTP	16
Bases de datos para la creación del modelo Datamart.....	16
Oracle	17
MySQL	18
SQLServer.....	18
Datamining	19
Redes neuronales artificiales	22
Reglas de inducción.....	22
Algoritmos genéticos.....	23

RapidMiner (formerly known as YALE).....	23
WEKA.....	24
Orange.....	24
KNIME	24
NLTK.....	25
Árbol de decisiones.....	25
Árbol de decisiones para la minería de datos.....	25
Reglas de Decisiones	27
FUNDAMENTACIÓN LEGAL	28
CAPÍTULO III.....	29
METODOLOGÍA DE INVESTIGACIÓN Y RESULTADOS.....	29
INVESTIGACIÓN Y CAMPO DE ESTUDIO	29
DESARROLLO DEL SOFTWARE Y CARACTERÍSTICAS.....	29
Pregunta Científica a Responder	30
Etapas de Metodología del Proyecto	30
Estudio y Comprensión de los Datos.....	30
Periodo de Evaluación.....	31
CRITERIOS DE ACEPTACIÓN DEL PROYECTO.....	31
CAPÍTULO IV.....	33
PROPUESTA TECNOLÓGICA	33
ANÁLISIS DE FACTIBILIDAD	33
Factibilidad Operacional.....	34
Factibilidad Técnica.....	34
Recurso de Software.....	35
Recursos de Hardware	38
Factibilidad Económica	39
CRONOGRAMA DE ACTIVIDADES.....	50
ANÁLISIS DE LOS DATOS Y SELECCIÓN DE CARACTERÍSTICAS.....	51
LAS VARIABLES DE ENTRADA O ANTECEDENTE:	51
RANGO DE VALORES DE LAS VARIABLES DE ENTRADA	52
ANÁLISIS DE LA INFORMACIÓN CON LA HERRAMIENTA DE MINERÍA DE DATOS PARA GENERAR LA REGLA DE DECISIONES	54
Primera Corrida de Datos con Algoritmo Modlem – Rt.....	54
Segunda Corrida de Datos con Algoritmo Furia	57

Tercera Corrida de Datos con Algoritmo Jrip	59
EVALUACIÓN DE LAS REGLAS	61
Primera Corrida de Datos con Algoritmo Modlem – Rt 1.....	61
Segunda Corrida de Datos con Algoritmo Furia	62
Tercera Corrida de Datos con Algoritmo Jrip	63
ESTRUCTURA DE LAS REGLAS DE DECISIONES PARA EL DISEÑO DEL PROTOTIPO	64
Despliegue	67
Diseño del Prototipo	67
Formas del proyecto	68
Menú transacciones.....	68
CORRIDA DE VALORES CON EL PROTOTIPO	70
CONCLUSIONES Y RECOMENDACIONES.....	74
CONCLUSIONES.....	74
RECOMENDACIONES	74
BIBLIOGRAFÍA.....	75

Índice de Tablas

Tabla 1. Diferencias entre Datawarehouse y Datamart	14
Tabla 2. Criterio para selección de algoritmos	36
Tabla 3. Recurso Humano para el desarrollo del prototipo	39
Tabla 4. Recurso tecnológico para el desarrollo del prototipo	39

Índice de Figuras

Figura 1: Datawarehouse.....	12
Figura 2: Datamart	14
Figura 3. Datamart híbrido.....	15
Figura 4. Datamart OLAP	16
Figura 5. Proceso de Datamining.....	20
Figura 6. Técnicas de Datamining	21

RESUMEN

El presente trabajo de titulación está basado en el desarrollo de un prototipo para la toma de decisiones del área de importaciones con técnicas de Datamining. Luego del análisis correspondiente de las diferentes bases de datos y sus bondades en cuanto a pertinencia y utilidad, se decidió que para el análisis de la información se utilice Weka, una herramienta de software libre, que permitirá realizar la generación de reglas de decisiones, utilizando algoritmos de selección con tratamiento de reglas difusas. Aplicadas las pruebas necesarias, se determinó que, en un mercado tan competitivo como el actual, contar con herramientas que provean información correctamente y en el momento oportuno, facilita las gestiones para tomar la mejor decisión. Este proyecto de titulación está enfocado en cubrir esa necesidad en la empresa objeto de este estudio, brindando un análisis histórico con grandes cantidades de información, permitiendo elaborar y desarrollar un prototipo que pueda cumplir con esta demanda.

Palabras clave: PROTOTIPO TOMA DE DECISIONES, HERRAMIENTA WEKA, ALGORITMOS DE CLASIFICACIÓN, DATAWAREHOUSE, DATAMART, DATAMINING

INTRODUCCIÓN

Gracias a la demanda de productos de exportación “Frutos y secos” para satisfacer un mercado en expansión la empresa Frutera Don Marcos S.A. (FRUDONMAR) se ve en la necesidad de abastecerse con la cantidad adecuada de inventario para cubrir la constante demanda de estos productos, es ahí donde una correcta toma de decisión en la importación es de vital importancia.

El área de importaciones es el motor de la empresa que se encarga de abastecer el inventario necesario y oportuno para cubrir la demanda de los clientes, es ahí donde se realiza un análisis de que producto será rentable para la empresa.

El análisis de la información es primordial para la toma de decisiones puesto que si esta no se encuentra bien estructurada y definida la información obtenida será errónea y provocará errores en la toma de decisiones más aún si no se cuenta con una herramienta que nos permita realizar este proceso de forma automática.

El surgimiento de Datawarehouse y Datamart son herramientas que facilitan la integración de información de diferentes repositorios en un solo esquema permitiéndonos realizar un análisis de gran cantidad de información desde una sola fuente.

El departamento de importaciones en la actualidad tiene el problema de no contar con la información oportuna y precisa cuando es requerida, para solventar este problema utilizan los reportes del sistema que ofrece esta información pero al estar segregada en diferentes módulos, realizan la unión de toda esta información en un documento en Excel detallando solo la información que es relevante para ellos, este proceso es manual y lleva un tiempo de horas incluso días, esto ocasiona que la toma de decisiones se la haga con información no tan confiable puesto que no se verifica los resultados agrupados en el documento.,

El presente documento propone la creación de un datamart que realice la consolidación de la información de diferentes tablas transaccionales para luego con un proceso de Datamining genere reglas de decisiones las cuales serán evaluadas con

un gran número de registros históricos que permita realizar la elaboración de un prototipo para la toma de decisiones.

La herramienta permitirá al departamento de importaciones realizar una correcta y oportuna toma de decisiones.

Este proyecto se basa en la elaboración de un prototipo de toma de decisiones departamental que analiza el comportamiento de la variabilidad de la venta de los productos generando reglas de decisiones que nos permitirá llevar a cabo estrategias de importaciones para el abastecimiento de productos y de esta forma satisfacer la demanda de los clientes obteniendo un correcto margen de utilidad.

El contenido de la investigación realizada se presenta en este documento, con la siguiente estructura: en el capítulo I se incluye la problemática en estudio, objetivos, justificación y alcance del estudio; en el capítulo II se presenta el marco teórico, conceptual y legal que enmarca el tema de la investigación; el capítulo III contempla la metodología de investigación, con sus correspondientes resultados; el capítulo IV contiene la propuesta tecnológica, para finalmente, presentar las conclusiones y recomendaciones relativas al estudio realizado.

CAPÍTULO I

EL PROBLEMA

En el ambiente económico y competitivo actual las metodologías tradicionales de toma de decisiones se han vuelto relativamente insuficientes ya que los responsables de realizar la dirección de las labores cotidianas de la empresa o institución en la que laboran se enfrentan a escenarios complejos y cambiantes que requieren de soluciones eficaces y prácticas que sean sustentados en procesos cuantitativos.

La toma de decisiones es fundamental en el desarrollo de la empresa, siendo fundamental que los resultados sean óptimos y coherentes para obtener el mayor beneficio para la empresa.

Las dificultades que se presentan al momento de realizar la toma de decisiones ha ocasionado que las empresas e instituciones busquen alternativas que les ayuden a efectuar de forma eficaz la toma de decisiones, dichos recurso tales como herramientas informáticas o técnicas de análisis de información son algunas alternativas que pueden ser utilizadas en este escenario.

Las herramientas informáticas tales como algoritmos de para generación de reglas de decisiones junto con Big Data para el procesamiento y análisis de datos han sido utilizadas para una amplia gama de situaciones en la toma de decisiones dentro de los distintos departamentos por parte de la gerencia. Se ha incrementado el uso de algoritmos para la generación de reglas de clasificación y estadísticos enfocados en la minería de datos para controlar, interpretar y predecir los resultados en la toma de decisiones gerenciales.

Dichas técnicas son utilizadas para la toma de decisiones en muchos ámbitos de negocios como son: abastecimiento y rotación de inventarios, pronósticos para las ventas, predicción de impacto del negocio y estrategias de inversión por mencionar algunos.

La necesidad de cada empresa es lograr posicionarse en la cima de su negocio con técnicas eficaces que logren encontrar el punto de equilibrio entre los objetivos esperados y los obtenidos, para de esta forma lograr su meta.

PLANTEAMIENTO DEL PROBLEMA

La empresa tiene La necesidad de mantener el inventario necesario para cumplir y abarcar el mercado, con productos que no están disponible todo el año, periodos altos y bajos de venta, precios que fluctúan no solamente por producto sino también dentro de un producto según el periodo de venta, el cliente y un sin número de variables que afectan la utilidad.

Por ello abastecerse de los productos adecuados con la cantidad precisa permitirá tener un margen de rendimiento alto, crecimiento de la empresa adecuado y copar el mercado atrayendo a los clientes en busca de productos variables.

Es necesaria una herramienta que permita tomar las decisiones adecuadas en el momento oportuno permitiendo minimizar costos, abastecerse adecuadamente puesto que al ser un producto perecible y con un periodo de vida no mayor de uno a dos meses aprovisionarse de una cantidad excesiva de productos dará como resultado baja de inventarios ocasionando perdida monetaria y cliente insatisfechos.

Delimitación del Problema

El área de importaciones realiza el abastecimiento de inventario en la empresa es por ello que para poder decidir qué cantidad de producto se debe adquirir y cuál será su margen de utilidad se ve en la necesidad de recurrir a una serie de reportes que se encuentran en el sistema de la compañía, para ello debe unir todos estos reportes en un documento en Excel filtrando la información relevante para el análisis, luego de ello junto con el área financiera se procede a realizar una validación y análisis de la información por medio de indicadores financieros que le darán una idea de cómo se comportará la venta y las utilidades que se podría tener por la adquisición de cierta cantidad de productos.

Esto toma un tiempo realmente largo, entre uno o dos días en consolidar la información con los datos relevantes, un día más en el análisis de dicha información

para luego de ello poder realizar la toma de decisión que permita al departamento de importaciones realizar la compra.

Es por esto que se ve en la necesidad de realizar una herramienta que pueda solventar este problema, el desarrollo de un prototipo que arroje los resultados que la empresa espera mediante un análisis previo de la información histórica y la obtención de reglas de decisiones que son integradas al prototipo permitirá ahorrar tiempo eliminado el proceso manual que hasta este momento se lleva a cabo.

Evaluación del Problema

Delimitado: Será enfocado al área de importaciones con indicadores financieros que nos permitirán establecer variables endógenas y exógenas para un periodo de tiempo desde el 2010 al 2017.

Claro: Se necesita de un prototipo para la toma de decisiones oportuna

Evidente: la falta de un proceso que permita realizar una toma de decisiones oportuna se hace evidente la necesidad de una herramienta informática para solventar este punto.

Concreto: Se necesita que el prototipo desarrollado retorne información concreta y precisa para la toma de decisiones.

Factible: se posee la información y las herramientas necesarias para proceder con el desarrollo y análisis del proyecto para de esta manera solventar el problema que se presenta en el departamento de importaciones.

OBJETIVOS

General

Desarrollo de un prototipo predictivo para la toma de decisiones de la gestión de compras del departamento de importaciones con la generación de un Datamart y técnicas de Datamining para la empresa FRUDONMAR.

Específicos

- Identificar variables de entrada y salida para los datos a evaluar
- Procesar la información histórica que la empresa nos proporciona con herramientas de Datamining para encontrar las reglas de decisiones
- Analizar y evaluar las reglas de decisiones halladas para filtrar las más optimas
- Elaborar un prototipo con las reglas de decisiones flitradas que nos ayude en la toma de decisiones

ALCANCE DE LA INVESTIGACIÓN

La empresa FRUDONMAR es una de las más grandes importadoras, ubicada en el Cantón duran, posee una gran infraestructura en cuanto a cámaras frigoríficas para almacenar los productos para la venta, todo proceso es controlado por un sistema ERP con módulos Administrativos, Contables, Financieros.

Para el presente trabajo de tesis se utilizará las herramientas apache y php para tener el mismo ambiente informático que la empresa utiliza en su ambiente de producción, y como gestor de base de datos MySQL para la elaboración del Datamart y como herramienta de análisis de dato Datamining Weka.

El objetivo será analizar la información provista por la empresa para encontrar reglas de decisiones con diferentes algoritmos de validación para luego con un mayor volumen de información validar las reglas seleccionando las que cumplan con las diferentes variabilidades de los datos, con las reglas de decisiones finales se procederá a elaborar el prototipo para la toma de decisiones del área de importaciones de la empresa FRUDONMAR.

JUSTIFICACIÓN E IMPORTANCIA

Con la ayuda de herramientas de Datamining se podrá encontrar patrones dispersos del comportamiento de las ventas, puesto que la variabilidad del precio del producto y los márgenes de utilidad junto con los costos tan cambiantes no permiten

encontrar dichos patrones con la información pequeña y limitada que se obtiene de los múltiples reportes del sistema informático.

Por medio del Datamining se optimizarán los procesos que en este momento se llevan manuales facilitando la toma de decisiones del área de Importaciones y optimar el proceso de compra de mercadería para el abastecimiento del inventario.

Adicionalmente la elaboración del prototipo para la toma de decisiones permitirá al usuario verificar información que le ayudará en el proceso de compra de mercadería en el departamento de importaciones.

El resultado de esta investigación puede dar pauta para que otras empresas, con características similares sigan esta buena práctica y la apliquen a lo interno de sus procesos.

DOMINIO Y LÍNEA DE INVESTIGACIÓN

El presente trabajo de titulación se enmarca en uno de los dominios determinados por la Universidad Católica de Santiago de Guayaquil (UCSG) denominado *tecnología y sistemas productivos* que responde a la Clasificación Internacional Normalizada de la Educación (CINE) *Tecnologías de la Información y la Comunicación* establecida por la UNESCO, además, responde a una de las líneas de investigación determinadas por la carrera Ingeniería en Sistemas Computacionales: (1) *Investigación y desarrollo de nuevos servicios o productos* y (2) *Utilización de software libre*.

CAPÍTULO II

MARCO TEÓRICO, CONCEPTUAL Y LEGAL

La matriz de la empresa FRUDONMAR está ubicada en el cantón Durán, de la provincia del Guayas; cuenta con una sucursal en Tambillo; es una empresa que se dedica a la importación de frutas y con el paso del tiempo ha tenido la necesidad de realizar un análisis del comportamiento de sus productos principales: manzanas, peras, uvas, naranjas, para poder realizar un correcto abastecimiento de su inventario en el periodo en que estos productos no están disponibles.

La gran cantidad de información que se genera ha sido muy difícil de procesar y analizar, es por esto que se propuso el diseño de un modelo para la toma de decisiones, mediante el uso de Datamart como repositorio de los indicadores financieros y técnicas de Datamining para la creación del prototipo para la toma de decisiones, con el fin de identificar qué se debe comprar para abastecerse en el periodo de no disponibilidad de algunas de las frutas mencionadas.

Para diseñar el modelo que se propone en este trabajo de titulación, se ha revisado algunas teorías y principios relativos al manejo de inventarios e instrumentos que facilitan estas opciones; también se incluye algunas conceptualizaciones de términos mayormente técnicos que tienen relación con el desarrollo del producto para, finalmente, hacer referencia a algunos elementos legales que sustentan el quehacer de este tipo de empresas.

FUNDAMENTACIÓN TEÓRICA Y CONCEPTUAL

El ámbito de este estudio está relacionado con la administración de inventarios de un bien, por ello, se hace mención de algunas opiniones provenientes de investigadores sobre el tema, complementando con conceptos de elementos y términos que se relacionan con un sistema de control y predicción de inventarios, que permiten convertir competitiva a una empresa.

La gestión de inventarios, como base para la toma de decisiones

El manejo de inventarios en una empresa, independientemente de la actividad a la que se dedique, es crucial si quiere mantenerse en el mercado. De acuerdo con Aguilar Santamaría (2012, p. 153) “el inventario es considerado como un "buffer" o amortiguador entre dos sistemas, uno de oferta (producción o abastecimiento) y otro de demanda (clientes o distribuidores)”.

La gestión de inventarios, independientemente del modelo que se aplique, está orientado a garantizar la disponibilidad, tanto en tiempos como en costos así como en el lugar adecuado, de todos los insumos requeridos para el funcionamiento de la empresa, tales como materia prima, productos terminados, equipos, repuestos, otros insumos (Salas-Navarro, Manguel-Mejía, & Acevedo-Chedid, 2017).

La disposición de los inventarios depende también de la cadena de suministros, que a su vez está influenciada por patrones que se representan en una mayor o menor incertidumbre, según el comportamiento del mercado; todo ello debe estar considerado cuando las empresas realizan su planificación, adoptan procesos de manejo de inventario y modelos para la toma de decisiones (Aguilar Santamaría, 2012).

Para mantenerse en el mercado y ser competitivas en el campo de su funcionamiento, las empresas deben saber tomar decisiones oportunas a todo lo largo de su cadena de procesos, lo que involucra también al manejo de sus inventarios.

Los inventarios en proceso y los inventarios de producto final constituyen un aspecto de gran importancia para la organización y son un punto de partida para la toma de decisiones estratégicas de la empresa; en este sentido, la gestión de inventarios para la eficiente comercialización de bienes y servicios se convierte en una herramienta para registrar las cantidades que poseen la empresa, las cuales juegan un papel fundamental en la etapa de abastecimiento y en el desarrollo de la demanda, dando como resultado estados confiables en el control de materiales y productos. (Causado Rodríguez, 2015, p. 165).

El manejo adecuado del mapa de procesos y la cadena de suministros, dentro de la gestión de inventarios, permiten una “transmisión adecuada y oportuna de los planes de demanda a los diferentes niveles de la cadena, favoreciendo procesos productivos, niveles de inventarios de seguridad más real y mejora en tiempos de respuesta al cliente final”(Salas-Navarro et al., 2017, p. 328).

En general, la gestión de los inventarios representa uno de los temas de mayor complejidad en la administración de una empresa, ya que si no se adoptan mecanismos adecuados para su efectiva ejecución puede caerse en errores tales como contar con demasiadas unidades de lo que no se vende y poco o casi nada de los productos de lo que sí se vende; ello implica realizar una apropiada proyección de compra, tomando en consideración aquello que requiere el cliente para su satisfacción; esto es, considerar las variables más significativas de ¿cuánto importar? y ¿cuándo pedir? (Pérez-Vergara, Cifuentes-Laguna, Vásquez-García, & Marcela-Ocampo, 2013)

Modelos para la gestión de inventarios

Entre diversos modelos, se puede mencionar algunos modelos matemáticos que pueden predecir el comportamiento de la demanda, tales como la planificación de requerimiento de material o MRP (Material Requirements Planning), justo a tiempo o JIT (de su nombre en inglés Just in Time), sistema de planificación de recursos empresariales o ERP (Enterprise Resource Planning) (Bofill Placeres, Sablón Cossío, & Florido García, 2017).

La planificación jerárquica, es el método de planificación que más se encuentra en la literatura para planificar la cadena de suministros (*supply chain*) sin embargo, al reunir una gran cantidad de variables en un sólo modelo centralizado, se vuelve rígido y la velocidad de respuesta se hace más lento (Ramos-Maldonado, Maness, & Salinas-Sandoval, 2015, p. 614).

En los tiempos actuales en que las tecnologías de la información y la comunicación (TIC) se han convertido en la base para todo tipo de gestión, debido a la facilidad y agilidad con la que se transfieren, procesan y sistematizan los datos y la información, las empresas han diseñado y/o implementado modelos y sistemas de

gestión para los diferentes procesos que en ellas se ejecutan; por supuesto, con las debidas seguridades, dado que es necesario garantizar la disponibilidad, confidencialidad e integridad de la información (Miranda Cairo, Valdés Puga, Pérez Mallea, Portelles Cobas, & Sánchez Zequeira, 2016), con el fin de evitar cualquier pérdida o modificación de la información.

El uso de la tecnología no es solamente para reaccionar rápidamente ante una situación extrema, sino que también está encauzada a generar opciones para una adecuada toma de decisiones y cumplir con los objetivos de producción planteados (Ramos-Maldonado et al., 2015).

Algunos elementos que han sido mencionados en los párrafos anteriores se presentan a continuación con su correspondiente definición.

Datos e información

Se ha hecho referencia a que las TIC sirven para transmitir, procesar y sistematizar datos e información, cabe hacer una diferencia entre dato e información. Por un lado, el *dato* es un término que se refiere a hechos o transacciones que han ocurrido o registrados y la *información* resulta del procesamiento de dichos datos facilitando su interpretación. Los datos pueden ser alfabéticos, numéricos, simbólicos (/,\, \$,, @,-, &,+,...)

Los datos son números, letras o símbolos que describen objetos, condiciones o situaciones. Son el conjunto básico de hechos referentes a una persona, cosa o transacción de interés para distintos objetivos, entre los cuales se encuentra la toma de decisiones. (Dip, 2008, párr. 1)

La información se puede definir como un conjunto de datos procesados y que tienen un significado (relevancia, propósito y contexto), y que por lo tanto son de utilidad para quién debe tomar decisiones, al disminuir su incertidumbre. Los datos se pueden transforman en información añadiéndoles valor. (Sinnexus, 2016d, párr. 6)

Datawarehouse

Aparece por los años 93 y 94 de la pasada década y nace por la necesidad de unir grandes volúmenes de información segregada en diferentes repositorios de base de datos permitiendo un análisis completo de la administración y la toma de decisiones de la empresa.

El concepto informático de Datawarehouse hace alusión al proceso mediante el cual una organización o empresa particular almacena todos aquellos datos e información necesarios para el propio desempeño de la misma. Se presupone que este tipo de material se organiza de manera tal de facilitar el análisis y la realización de reportes en las situaciones en que sea necesario hacerlo. Contar con un fácil acceso a los datos de importancia tendrá directa relación con la efectividad de las diversas tareas de la empresa. (Bembibre, 2009, párr. 1).

Contiene la información de todas las bases de datos y sistemas que la empresa posee por medio de una estructura multidimensional.

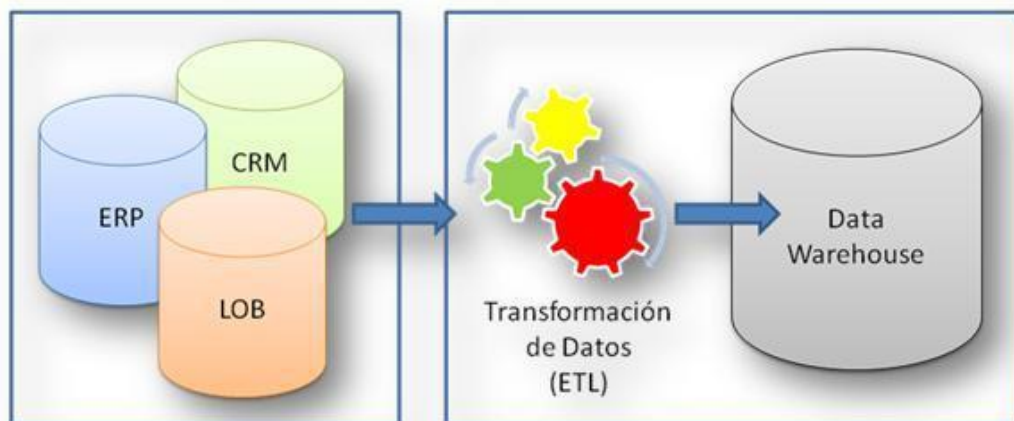


Figura 1: **Datawarehouse**. Tomado de Navarro Arango (2015)

La principal ventaja que Datawarehouse da es el fácil acceso a los datos desde un repositorio completo y estructurado con información completa y relevante para los usuarios y sus requerimientos.

Toda la información dispersa en la empresa es almacenada en ella, lo que permite realizar accesos rápidos y completos para un mejor y mayor análisis corporativo.

La ventaja principal de este tipo de bases de datos radica en las estructuras en las que se almacena la información (modelos de tablas en estrella, en copo de nieve, cubos relacionales... etc.). Este tipo de persistencia de la información es homogénea y fiable, y permite la consulta y el tratamiento jerarquizado de la misma (siempre en un entorno diferente a los sistemas operacionales). (Sinnexus, 2016c, párr. 2)

Una característica del Datawarehouse es que contiene metadatos que son datos sobre datos, estos permiten identificar de donde se extrajo esta información permitiendo así su refresco de información y priorizar lo relevante.

Entre las principales ventajas del Datawarehouse se puede mencionar las siguientes:

- Facilita la toma de decisiones en cualquier área de la empresa por la integración de la información en un solo repositorio.
- Facilita un análisis de la información mediante técnicas estadísticas permitiéndonos encontrar resultados ocultos por la segregación de la información.
- Permite aprender de datos pasados y poder proveer información futura en diferentes escenarios.

Datamart

Un Datamart puede proceder de un Datawarehouse o estar integrada por diferentes fuentes de información.

Es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. (Sinnexus, 2016b, párr. 1)

El Datamart se enfoca a un área específica (para nuestro estudio el departamento de importaciones) y elaborada por información relevante para cada área donde el administrador podrá generar y manejar dicha información

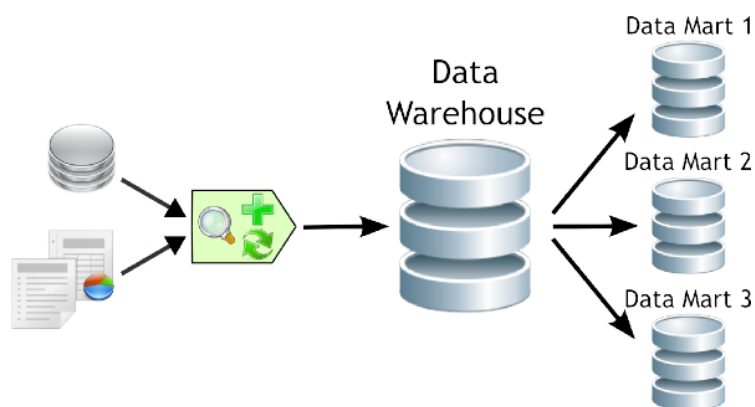


Figura 2: **Datamart.** Tomado de Riquelme (2013)

Cabe ahora hacer una relación entre el Datawarehouse y el Datamart. El primero surge de la necesidad integración de la información de la empresa en un solo repositorio para un mejor análisis de la información, el problema se presenta cuando crece y se forma más complejo su análisis es ahí cuando el Datamining entra en funcionamiento simplificando y consolidando la información segregándola en áreas del negocio alimentado por el Datawarehouse.

Tabla 1. **Diferencias entre Datawarehouse y Datamart**

Aspectos de análisis	Datawarehouse	Datamart
Alcance	Construido para satisfacer las necesidades de información de toda la organización	Construido para satisfacer las necesidades de información de toda la empresa
Objetivo	Diseñado para optimizar la integración y la administración de los datos fuente	Diseñado para optimizar la entrega de información de soporte a decisiones
Características de los datos	Administra grandes cantidades de datos históricos a nivel atómico	Se concentra en administrar resúmenes y/o datos totalizados
Pertenencia	Pertenece a toda la organización	Pertenece al área de negocios al cual está orientada
Administración	Administrado por la unidad de sistemas de la organización	Administrado por personal de sistemas de la unidad propietaria del datamart

Nota: Tomado de Abc del datamart (2016b)

Volviendo al Datamart, puede ser clasificado como dependiente, independiente o híbrido.

- **Dependiente.** Recibe la información desde un Datawarehouse; en este tipo el origen de información es única.
- **Independiente:** Toma su información directamente de los programas operacionales y/o transaccionales y no depende de un Datawarehouse (ésta es la que será utilizada para este estudio).
- **Híbrido:** Se alimentan de los dos tipos de orígenes, Datawarehouse o sistemas operaciones y/o transaccionales.

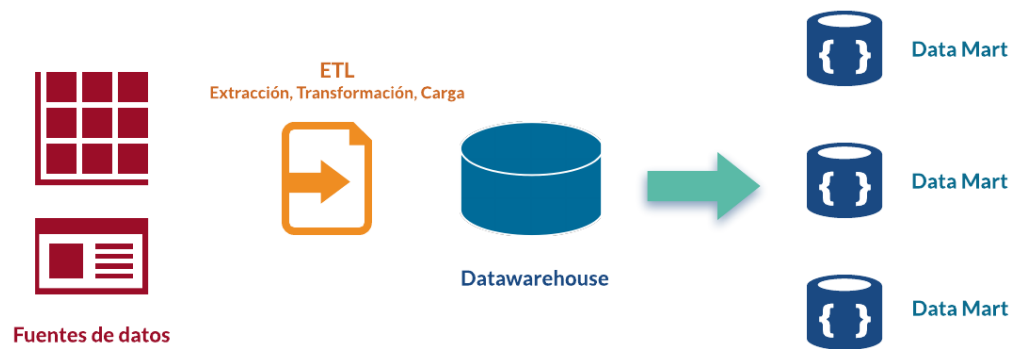


Figura 3. **Datamart híbrido.** Tomado de Abc del Datamart (Evaluando Software, 2016b)

Datamart OLAP

Se derivan de los cubos OLAP, que se generan en base a los requisitos de cada área o departamento de la empresa, definiendo las dimensiones y los indicadores necesarios de cada cubo. El modo de creación, uso y mantenimiento de los cubos OLAP es muy variado en su uso, según la herramienta final que se utilice.

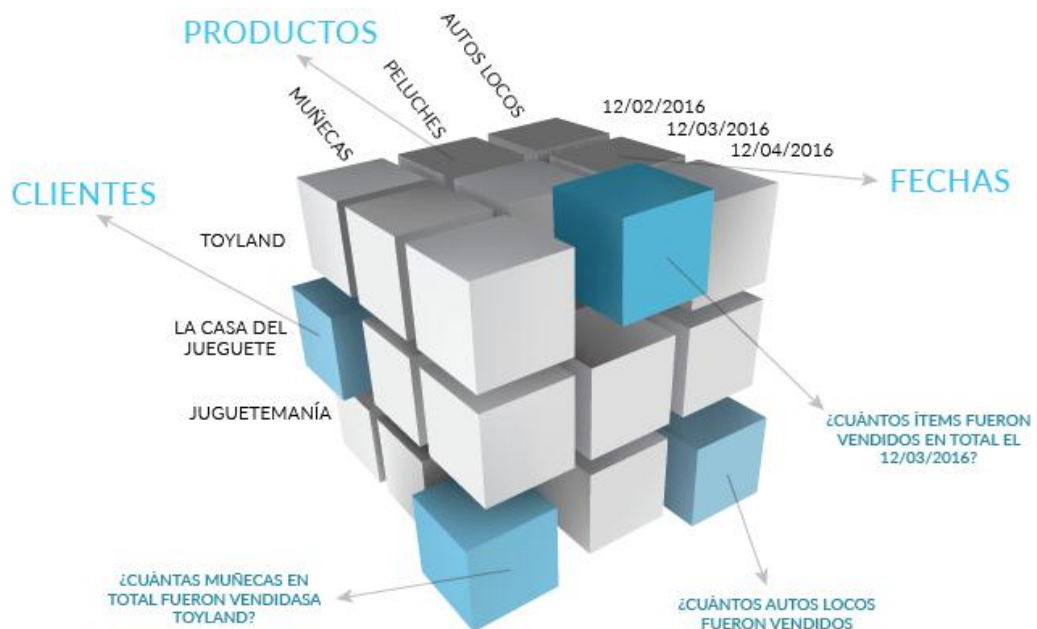


Figura 4. **Datamart OLAP.** Tomado de Giménez (2017)

Datamart OLTP

Puede generarse de un extracto simple del Datawarehouse, mejorando su rendimiento para ello se dispone de filtrados y agregaciones siendo las operaciones más usuales, con ello se pueden aprovechar las características particulares de cada departamento de la empresa.

Este tipo de Estructura está dotada de las siguientes características (Sinnexus, 2016a):

- Poco volumen de datos.
- Mayor rapidez de consulta.
- Validación directa de la información.
- Facilidad para la historia de los datos.

Bases de datos para la creación del modelo Datamart

En el mercado existe un sinnúmero de bases de datos que no puede ayudar a la elaboración de nuestro modelo, entre ellas detallamos las más utilizadas entre las empresas por su facilidad de uso y herramientas administrativas: Oracle, MySQL y SqlServer

Oracle

Oracle Database es la base de datos más utilizados por empresas medianas y grandes; se trata de un sistema de gestión de base de datos objeto-relacional (ORDBMS, por sus iniciales en inglés de Object-Relational Database Management System), desarrollado por Oracle Corporation; esta base de datos es considerada como uno de los más completos, ya que incluye soporte de transacciones, estabilidad, escalabilidad, y soporte multiplataforma; tiene suficiente capacidad para soportar todos los datos relacionales estándares, así como también datos oriundos como XLM, XQuery y WebDAV, JDBC, SQLJ, ODBC.Net, OLE.Net, ODP.Net (Formatalent, 2017).

Hay versiones de Oracle, en especial las últimas que tuvieron que ser certificadas para poder trabajar bajo Linux y GNU. Como contraparte, recientemente ha surgido competencia de Microsoft SQL Server y de la oferta de otros RDBMS con licencia libre como PostgreSQL, MySQL o Firebird.

Oracle nace en 1977 bajo el nombre SDL (por su nombre en inglés Software Development Laboratories). En 1979 SDL cambia su nombre a RSI (por sus siglas de Relational Software, Inc), cuya función estuvo orientada principalmente por un estudio sobre los Sistemas Gestores de Base de Datos (SGBD) de George Koch.

En la actualidad, Oracle todavía encabeza la lista, cuya tecnología de Oracle está posicionada en casi todas las industrias alrededor del mundo, principalmente en 98 empresas de las 100 mencionadas por la revista Fortune 100.

La base de datos de Oracle 10G Standar Edition es compatible con medianas industrias, rápida instalación sin contratiempos tanto en un único servidor como en un ambiente de clúster. Esto incluye Real Application Clúster, para crear protección en contra de fallos de hardware. Es muy sencillo de instalar y configurar, y viene con su propio software de clustering; utiliza gestiones de concurrencia similares a la de Oracle Enterprise Edition, asegurando así el máximo rendimiento para todas las cargas de trabajo. Los procedimientos almacenados pueden ser escritos en java PL SQL o utilizando .Net CLR Support en Oracle Database 10g Release Dos.

MySQL

Según Rouse (2015), MySQL es un sistema de gestión de base de datos relacional (RDBMS, por sus siglas en inglés de Relational Database Management System) de código abierto, basado en lenguaje de consulta estructurado (SQL) (TechTerms, 2017); fue concebido originalmente por la empresa sueca MySQL y adquirida por Oracle en 2008. Aunque los desarrolladores de este sistema pueden utilizarlo bajo licencia pública general, las empresas que deseen aplicarlo deben comprar la licencia comercial de Oracle.

Se ejecuta en Linux, UNIX y Windows, entre otras plataformas; se asocia más con las aplicaciones basadas en la web y la publicación en línea, aunque también se puede usar en varias opciones de aplicaciones, siendo un componente importante en el ámbito empresarial con código abierto denominado LAMP.

LAMP es una plataforma de desarrollo web que utiliza Linux como sistema operativo, Apache como servidor web, MySQL como sistema de gestión de base de datos relacional y PHP como lenguaje de programación orientado a objetos (a veces, Perl o Python se utiliza en lugar de PHP) (Rouse, 2015).

Los derivados de MySQL incluyen (Pratt, 2014):

- Drizzle - un sistema de gestión de base de datos ligero de código abierto en el desarrollo basado en MySQL 6.0.
- MariaDB - un reemplazo popular "drop-in" desarrollado en la comunidad para MySQL que utiliza las API y los comandos de MySQL.
- Percona Server con XtraDB - una versión mejorada de MySQL conocido por su escalabilidad horizontal.

SQLServer

Según Rouse (2015), SQL Server es también un RDBMS, diseñado para ser aplicado en el entorno empresarial SQL, convierte aplicaciones críticas en aplicaciones inteligentes con rendimiento in-memory y análisis avanzado integrado (Microsoft, 2016).

Se ejecuta en Transact-SQL (T-SQL) un conjunto de extensiones de programación de Sybase y Microsoft que añaden varias características a SQL estándar, incluyendo control de transacciones, excepción y manejo de errores, procesamiento fila, así como variables declaradas.

Bajo el nombre código Yukon en su etapa de desarrollo, SQL Server 2005 fue lanzado en noviembre de 2005. Se dice que el producto 2005 proporcionó una mayor flexibilidad, escalabilidad, confiabilidad y seguridad a las aplicaciones de base de datos, y permitió que fueran más fáciles de crear y desplegar, lo que reduce la complejidad y el tedio involucrado en la gestión de bases de datos. SQL Server 2005 también incluía más soporte administrativo.

El código original de SQL Server ha sido desarrollado por Sybase; a finales de 1980, Microsoft, Sybase y Ashton-Tate colaboraron para producir la primera versión del producto, SQL Server 4.2 para OS/2. Posteriormente, tanto Sybase como Microsoft ofrecieron productos de SQL Server. Sybase cambió después el nombre de su producto a Adaptive Server Enterprise.

Datamining

De acuerdo con lo expuesto por Pérez López (2007),

Las técnicas de minería de datos persiguen el descubrimiento automático del conocimiento contenido en la información almacenada en modo ordenado en grandes bases de datos. Estas técnicas tienen como objeto descubrir patrones, perfiles y tendencias a través del análisis de los datos utilizando tecnologías de reconocimiento de patrones, redes neuronales, lógica difusa, algoritmos genéticos y otras técnicas avanzadas de análisis de datos. (p. 1)

Con el Datamining tenemos técnicas que nos ayudan a explotar una gran cantidad de datos de una empresa con el objetivo de encontrar mediante procesos estadísticas o neuronales patrones, tendencias o reglas del comportamiento de los datos en un periodo determinado.

Aunque cada estudio de datos es independiente se suele tener cuatro etapas principales las cuales son:

- **Planteamiento de las necesidades.** Delimita las necesidades que el cliente desea bajo la supervisión del experto en Datamining.
- **Pre-procesamiento de los datos.** Se entiende a la selección, reducción y la transformación por medio del procesamiento de datos hacia un Datamart desde las bases de datos del ERP o Datawarehouse del cliente. Esta etapa consume un gran tiempo del proyecto de Datamining alrededor de un 70%.
- **Especificación del modelo.** Se inicia con un análisis estadístico de los datos, luego de ello se realiza una visualización gráfica de estos para así tener una primera aproximación. Dependiendo de las necesidades establecidas y los objetivos que deben llevarse a cabo, se suele utilizar algoritmos desarrollados en diferentes áreas de la Inteligencia Artificial.
- **Análisis de los resultados.** Constata que la información obtenida es coherente y los compara con los obtenidos por los análisis estadísticos y de visualización gráfica. En esta etapa el personal adecuado para esta revisión debe ser especializado en el área que se está evaluando para que pueda interpretar y verificar su contenido.

Luego de ello se puede realizar el despliegue del proyecto que consiste en la implementación de un proceso de Datamining repetible.



Proceso de Knowledge Discovery in Databases (KDD)

Figura 5. **Proceso de Datamining.** Tomado de Técnicas de Datamining (Evaluando Software, 2016a)

Dentro de la etapa de especificación del modelo existen dos tipos los cuales son (Berzal, 2006):

- **Métodos descriptivos.** Encuentran patrones que puedan ser interpretados para referir datos; pueden ser: clustering, descubrimiento de reglas de asociación y de patrones secuenciales. Este método describe cómo se

comportan los datos de manera que puedan ser interpretados por usuarios expertos.

- **Métodos predictivos.** Usan variables dependientes para predecir información futura o desconocida de variables independientes, estos pueden ser: regresión, clasificación y detección de la desviación (se usará este método en el proyecto, por ser de predicción y dependiente de variables).

A continuación se detalla Ejemplos de Modelos de Minería de Datos (Tan, Steinbach, & Kumar, 2006)

- **Reglas de asociación** (modelo descriptivo): Los compradores de colas también suelen comprar arroz.
- **Clustering** (modelo descriptivo): Clasificación de los clientes de un banco: cuenta ahorristas ocasiones que depositan gran cantidad de dinero, cuenta ahorristas habituales que depositan una cantidad de dinero media, cuenta ahorristas ocasionales que depositan una cantidad de dinero media
- **Clasificación** (modelo predictivo): Registros de un sistema de control de acceso no autorizado, historial de un cliente que solicita un préstamo bancario.



Figura 6. **Técnicas de Datamining.** Tomado de Guevara Vega (2015)

Para poder elaborar un Datamining existen diferentes técnicas que se centran en algoritmos detallados a continuación (Cultura CRM, 2016a):

Redes neuronales artificiales

Los métodos de Datamining con base en redes neuronales se utilizan para identificar patrones de comportamiento a partir de ciertos datos y proceder a clasificarlos, aunque los resultados que se obtienen son de difícil comprensión hacen que sea una de las técnicas menos aplicadas.

Son modelos predictivos no lineales, que recuerdan a una red de neuronas del sistema nervioso. La finalidad de las redes neuronales es construir un modelo que sea capaz de reproducir el método de aprendizaje del cerebro humano. Entre las empresas que han llevado a cabo el modelo de red neuronal artificial se encuentra Google, en su caso fue útil para reconocer los números de las calles en las imágenes que iban tomando con sus coches, consiguiendo un 96% de precisión en sus resultados. Por otro lado, se han aplicado en el control meteorológico, medición de audiencias de televisión, previsión de impacto en RRSS o en la predicción en los mercados financieros. (Cultura CRM, 2016a, párr. 3).

Reglas de inducción

Estas reglas facilitan la expresión de disyuntivas y opciones que pueden ser utilizadas más fácilmente que los árboles en Datamining y tienden a preferirse con respecto a los árboles por tender a representar “pedazos” de conocimiento relativamente independientes. Adicionalmente, consiguen que se pueda generar y contrastar patrones partiendo de los datos que se ingresaron inicialmente en el sistema; estos datos son conjuntos clasificados que están asociados a determinadas variables previamente definidas.

Por tanto, se basan en cálculos estadísticos que siguen un patrón clásico de “si-entonces”, es decir, como los clásicos juegos de rol, dependiendo el camino que elijas encontrarás una u otras aventuras ante ti. Básicamente, sirven para extraer conclusiones y hacer pronósticos. (Cultura CRM, 2016a, párr. 5).

Algoritmos genéticos

Este método es uno de los más utilizados en el Datamining. Aplican técnicas de optimización con base en conceptos provenientes de la combinación genética, la mutación y la selección natural para buscar conjuntos óptimos de parámetros que permitan realizar una descripción de predicción.

Esta técnica utiliza un proceso muy iterativo de selección, cruzado y de mutación de operadores, evolucionando las sucesivas generaciones de modelos. En cada iteración, cada modelo compite con todos los otros modelos por la herencia de los rasgos de los anteriores hasta que sólo el modelo más predictivo sobrevive. (Cultura CRM, 2016a, párr. 6)

Existen varios softwares de minería de datos para la evaluación de nuestro modelo que utilizan las técnicas descritas anteriormente los cuales se detallan a continuación, con sus correspondientes páginas de descarga (Tan et al., 2006):

- KNIME (<http://www.knime.org/>)
- RapidMiner (<http://rapidminer.com/>)
- Weka (<http://www.cs.waikato.ac.nz/ml/weka/>)
- R (<http://www.r-project.org/>)
- SPSS Modeler (<http://www.spss.com/software/modeler/>)
- SAS Enterprise Miner (<http://www.sas.com/>).

A continuación, se describe algunos detalles importantes de las herramientas más utilizadas para la minería de datos. (Cultura CRM, 2016b).

RapidMiner (formerly known as YALE)

Con un lenguaje de programación Java, RadipMiner permite realizar un análisis avanzado de los datos, a través de plantillas. Esta herramienta ofrece un servicio excelente, ocupando la primera posición entre las mejores herramientas de Datamining.

Además, dispone de la funcionalidad de pre-procesamiento y visualización de datos, análisis predictivo y modelos estadísticos, así como evaluación y despliegue de la información.

WEKA

La versión original de esta herramienta se destina al análisis de datos del sector agrícola. Pero fue evolucionando y desarrollo una herramienta superior basada en Java. Esta versión permite analizar datos y establecer modelos predictivos.

Igual que RapidMiner, Weka realiza trabajos de Datamining estándar, incluyendo pre-procesamiento de datos, clustering, clasificación, regresión, visualización y selección de características.

Orange

Esta herramienta es simple y fácil de usar, a la vez que potente. Orange dispone de una estupenda programación visual y de script. Además, cuenta con componentes de aprendizaje propio, complementos de informática y minería de texto.

KNIME

Capacidad de pre-procesamiento de datos con tres componentes principales: la extracción, la transformación y la carga de datos. Dispone de una interfaz gráfica que permite el montaje de nodos de procesamiento de datos. Se trata de una plataforma de análisis de datos de origen, con la posibilidad de presentar informes e integración abierta. KNIME también integra diversos componentes para el aprendizaje automático y minería de datos a través de su concepto de segmentación de datos modular, características que han llamado la atención del Business Intelligence y el análisis de datos financieros.

Por otro lado, destacar que esta herramienta está escrita en Java y basado en Eclipse. KNIME tiene la facilidad de añadir plugins y funcionalidades adicionales en cualquier momento.

NLTK

Cuando se trata de tareas de procesamiento del lenguaje, nada puede superar NLTK. NLTK proporciona un conjunto de herramientas de procesamiento del lenguaje, incluyendo la minería de datos, aprendizaje automático, raspado de datos, análisis de los sentimientos y otras tareas de procesamiento del lenguaje. Debido a que está escrito en Python, se pueden construir aplicaciones sobre sí misma y personalizarlo para tareas pequeñas.

Árbol de decisiones

Un árbol de decisiones es un modelo predictivo el cual nos ayuda a la toma de decisiones complejas, utilizando una serie de ramificaciones nos permite comparar posibles acciones entre sí.

Árbol de decisiones para la minería de datos

Según Lucidchart (2017) un árbol de decisión también puede usarse para ayudar a crear modelos predictivos automatizados, que puedan emplearse en el aprendizaje automático, la minería de datos y las estadísticas. Conocido como "el aprendizaje basado en árboles de decisión", este método toma en consideración las observaciones sobre un elemento para predecir su valor.

En estos árboles de decisión, los nodos representan datos en lugar de decisiones. Este tipo de árbol también se conoce como "árbol de clasificación". Cada ramificación contiene un conjunto de atributos o reglas de clasificación asociadas a una etiqueta de clase específica, que se halla al final de la ramificación.

Estas reglas, también conocidas como "reglas de decisión", se pueden expresar en una cláusula "Si... entonces...". Cada valor de datos o decisión forma una cláusula, de tal manera que, por ejemplo, "si las condiciones 1, 2 y 3 se cumplen, entonces el resultado X será el resultado definitivo con certeza Y".

Cada dato adicional ayuda a que el modelo prediga de forma más precisa a qué conjunto finito de valores pertenece el asunto en cuestión. Esa información se puede usar posteriormente como una entrada en un modelo más grande de toma de decisiones.

A veces la variable predicha será un número real, como un precio. Los árboles de decisión con resultados posibles, infinitos y continuos se llaman "árboles de regresión".

Para una mayor precisión, se suelen combinar múltiples árboles con métodos de ensamblado:

- **Bagging (Ensamblado)** crea múltiples árboles de decisión haciendo un remuestreo de los datos de origen, luego hace que los árboles voten para llegar a un consenso.
- Un **clasificador de selva** aleatoria consiste en múltiples árboles diseñados para aumentar la tasa de clasificación.
- **Los árboles ampliados** se pueden usar en árboles de clasificación y de regresión.
- Los árboles en una **Selva aleatoria (Rotation Forest)** se entrenan mediante el uso del análisis de componentes principales (ACP) en una porción aleatoria de los datos.

Un árbol de decisión se considera ideal cuando representa la mayor cantidad de datos con el menor número de niveles o preguntas. Los algoritmos diseñados para crear árboles de decisión optimizados incluyen CART, ASSISTANT, CLS y ID3/4/5. Un árbol de decisión también se puede generar mediante la creación de reglas de asociación, ubicando la variable objetivo a la derecha.

Cada método debe determinar cuál es la mejor forma de dividir los datos en cada nivel. Los métodos comunes para hacerlo incluyen la medición de la impureza de Gini, la obtención de información y la reducción de variaciones.

Emplear los árboles de decisión en el aprendizaje automático tiene numerosas ventajas:

- El costo del uso del árbol para predecir los datos disminuye con cada punto de datos adicional.
- Funciona para los datos numéricos o categóricos.
- Puede modelar problemas con múltiples resultados.
- Usa un modelo de caja blanca (lo que hace que los resultados sean fáciles de explicar).
- La fiabilidad de un árbol se puede cuantificar y poner a prueba.
- Tiende a ser preciso independientemente de si viola las suposiciones de los datos de origen.

Pero también tienen algunas desventajas:

- Cuando se presentan datos categóricos con múltiples niveles, la información obtenida se inclina a favor de los atributos con mayoría de niveles.
- Los cálculos pueden volverse complejos al lidiar con la falta de certezas y numerosos resultados relacionados.
- Las conjunciones entre nodos se limitan a AND, mientras que los gráficos de decisión admiten nodos relacionados mediante OR.

Reglas de Decisiones

Según (Universidad de Valencia, s. f.)

Se decide que las hipótesis están equivocadas y las poblaciones son diferentes a lo establecido en las Hipótesis Nulas si la evidencia observada es poco probable bajo el supuesto de que las hipótesis sean verdaderas. Generalmente se considera poco probable valores iguales o inferiores a 0.05 ó 0.01.

Hay dos maneras de aplicar la regla de decisión al contraste de hipótesis, una se basa en la probabilidad de observar valores del estadístico de contraste bajo el supuesto de que la Hipótesis Nula sea verdadera. La otra se basa en determinar si el valor observado del estadístico de contraste se sitúa en la región de rechazo (conjunto de valores cuya observación es poco probable si la Hipótesis Nula fuera verdadera).

La aplicación de la regla de decisión de la primera manera consiste en rechazar la Hipótesis Nula si la probabilidad de observar el estadístico de contraste es igual o menor a la probabilidad criterio (habitualmente 0.05 ó 0.01).

Ejemplo. Si se lanza diez veces la probabilidad de observar 6, 7, 8, 9, 10 ó 0, 1, 2, 3 ó 4 caras es igual a 0.754. En consecuencia, pensamos que este resultado no es raro (la probabilidad no es baja) bajo el supuesto de que la Hipótesis Nula sea verdadera, y por ello la aceptamos.

La aplicación de la regla de decisión en la segunda manera consiste en rechazar la Hipótesis Nula si el estadístico de contraste se sitúa en la región de rechazo (conjunto de valores cuya probabilidad de ocurrencia es muy baja bajo el supuesto de que la Hipótesis Nula sea verdadera). El resto de valores constituye la región de aceptación.

FUNDAMENTACIÓN LEGAL

De acuerdo con la LEY DE DERECHOS DE PROPIEDAD INTELECTUAL en el artículo 7 incluye algunos términos relacionados con el tema; en cuanto a *Programa de ordenador (software)* lo define como:

Toda secuencia de instrucciones o indicaciones destinadas a ser utilizadas, directa o indirectamente, en un dispositivo de lectura automatizada, ordenador, o aparato electrónico o similar con capacidad de procesar información, para la realización de una función o tarea, u obtención de un resultado determinado, cualquiera que fuere su forma de expresión o fijación. El programa de ordenador comprende también la documentación preparatoria, planes y diseños, la documentación técnica, y los manuales de uso. (Congreso Nacional, 2006, p. 5)

CAPÍTULO III

METODOLOGÍA DE INVESTIGACIÓN Y RESULTADOS

En este capítulo se presenta dos aspectos de la metodología a aplicar: la primera, la relacionada con la investigación propiamente dicha que facilita la obtención de información como base para el cumplimiento de este proyecto; la segunda, se refiere propiamente a la metodología utilizada para el desarrollo de software para uso de la empresa beneficiaria.

INVESTIGACIÓN Y CAMPO DE ESTUDIO

El enfoque aplicado para este estudio fue el cualitativo, descriptivo y documental, dado que estaba orientado a obtener información relacionada al manejo interno de los procesos que llevan a la importación de un bien en determinadas épocas, en función de su oferta y demanda; para ello se utilizó datos históricos de toma de decisiones de manera que sirvan de base para realizar proyecciones mediante diferentes herramientas informáticas. Esto es, la investigación no tuvo como propósito estudiar con métodos estadísticos posibles relaciones entre variables, más bien estuvo orientada a encontrar relaciones de comportamiento de algunos elementos que inciden en la toma de decisiones (Hernández, Fernández, & Baptista, 2014).

En reuniones de trabajo con personal de los departamentos de importaciones y financiero, se recogieron las observaciones presentadas y que sirvieron para establecer las características que debe cumplir el software a desarrollar.

Por ello, no se requirió aplicar algún tipo de instrumento dado que el punto de partida fue la documentación existente y las experiencias previas para luego pasar a aplicar el software para las proyecciones esperadas.

DESARROLLO DEL SOFTWARE Y CARACTERÍSTICAS

La metodología que se aplica para diseñar un software se constituye en un conjunto de técnicas que permiten estructurar, planear y controlar las etapas del proceso desarrollo de un sistema de información

Existen diferentes tipos de técnicas de investigación que no facilitan un correcto planteamiento y control del proyecto por ello se debe seleccionar uno que permita abarcar y realizar ajustes en cada una de las etapas sin necesidad de pérdida de tiempo y recursos, no olvidando que esta metodología deberá estar enfocada a la minería de datos que es el ámbito de este estudio.

Por lo antes expuesto la metodología utilizada es la denominada *Proceso de Construcción Cruzada Estándar de Minería de Datos* (CRISP-DM), que es una de las más usadas en el desarrollo de proyectos enfocados a la minería de datos.

Pregunta Científica a Responder

¿De qué forma se analiza la información cuando tiene un gran volumen de datos?

¿Cómo conseguir información procesada y detallada de la base de datos analizada que permitan tomar las mejores decisiones?

Etapas de Metodología del Proyecto

Para realizar el desarrollo del proyecto se realizó una reunión con representantes de la empresa FRUDONMAR para que faciliten el acceso a los datos necesarios para la elaboración de un Datamart con aplicación Datamining; se contó con la colaboración de personal del departamento de importaciones quienes detallaron sus necesidades y expectativas, que fueron la base para establecer el alcance del proyecto y, que finalmente les ayudará a realizar la toma de decisiones oportunas y adecuadas para la compra de productos y mantener abastecido el inventario en las temporadas bajas del producto en las cuales no está disponible para su compra.

Estudio y Comprensión de los Datos

Según el análisis financiero realizado por el departamento de Importaciones y Financiero de FRUDONMAR, los indicadores indispensables para realizar el proceso de Datamining son:

- **Indicador de precio de venta al público:** Medirá el precio de venta del producto en un determinado tiempo, siendo que el precio de venta es variable para el mismo producto esto implica que se debe determinar un precio relativo que nos permita realizar el análisis del valor de venta del producto para de esta manera determinar un margen de utilidad óptimo que para este estudio debería ser del 15-60% sobre el valor de compra, este análisis lo determinará el área Financiera para luego de su revisión proyectar con el área de importaciones las compras que se podrían realizar en el año para el abastecimiento de inventario.
- **Indicador de gasto de importación:** Medirá el gasto de las importaciones que se han realizado en un determinado tiempo, ya que la unión entre el producto y el exportador es de uno a uno no se necesitará diferenciar el proveedor para este proceso.
- **Indicador de producto vendido:** Medirá el producto vendido en un determinado periodo de tiempo, a estos valores de venta no se descontarán las devoluciones realizadas por parte del cliente.

Periodo de Evaluación

La información que será evaluada corresponde al periodo de 2010-2017.

CRITERIOS DE ACEPTACIÓN DEL PROYECTO

Este proyecto analiza y evalúa las necesidades iniciales del área de importaciones por medio de indicadores para lograr información eficiente y completa, que proporcione los datos básicos para la gestión interna.

La construcción de un prototipo junto con un Datamart constituye una herramienta válida para el análisis de los datos cuantitativos y cualitativos para la toma de decisiones del área de importaciones, que es el propósito planteado dentro del presente trabajo de titulación.

Se generó información confiable y relevante para su análisis; esta información es obtenida desde la estructura de la base de datos que el sistema informático de la

empresa maneja, relacionando la información de varias tablas transaccionales para obtener información consolidada y precisa.

En el transcurso de este proyecto se determinó una herramienta para el análisis de información como Weka que luego de un exhaustivo estudio se pudo verificar que es el mejor programa para obtener reglas de decisiones por medio de un conjunto de algoritmos, que permitió hacer un análisis comparativo de cada una de las reglas generadas para la elaboración del prototipo.

Se realizó un análisis de los equipos informáticos para la elaboración e instalación del prototipo constatando que se cuenta con una infraestructura adecuada y completa contando con paquetes informáticos como Apache, php y MySQL.

CAPÍTULO IV

PROPUESTA TECNOLÓGICA

La creación de un prototipo de un modelo predictivo para la toma de decisiones de la gestión de compras del área de importaciones con la generación de un Datamart y técnicas de Datamining para la empresa FRUDONMAR es factible ya que se cuenta con todo el acceso a la información relevante para el procesamiento de datos generando información coherente.

ANÁLISIS DE FACTIBILIDAD

Elaborar un prototipo con diseño de Datamart para el departamento de importaciones permitirá poseer información histórica permitiendo orientarse correctamente en los productos con mayor flujo de ventas en el presente y el tiempo.

EL principal beneficio de un Datamart para el departamento de importaciones es conseguir analizar la información de forma oportuna y ágil por medio de consultas rápidas y consolidadas puesto que por la variabilidad del producto y el precio de venta el volumen de información es extenso y segregado.

Esto permitirá a la empresa la toma de decisiones estratégicas para la compra de mercadería que pueda acaparar el mercado y proveedor a las bodegas de suficiente producto para el periodo del año en que es estos productos no están disponibles.

El proyecto poseerá las siguientes etapas:

- Análisis y planteamiento de la propuesta
- Fundamentación teórica
- Elaboración del prototipo
- Ejecución del Prototipo
- Evaluación de los resultados

Se utilizarán herramientas que permita la aplicación de Inteligencia de negocios, por lo cual podremos decir que el proyecto es factible.

Factibilidad Operacional

Es necesario poseer una herramienta que permita obtener información completa y consolidada la cual servirá para la toma de decisiones para el área de importaciones de FRUDONMAR. Con esta herramienta podrá obtener un grupo de datos que permitirá generar reportes estadísticos con toda la información histórica lo cual en estos momentos no se posee.

Se desarrollará un Datamart que contendrá toda la información histórica consolidada que permitirá ahorrar horas de trabajo repetitivo con la filtración de los reportes existentes y procesos estadísticos que el Datamining lo realizará.

Los usuarios del área de importaciones están de acuerdo en que se debe contar con una herramienta que les permita gestionar de mejor manera el abastecimiento del inventario con productos que puedan estar disponibles en todas las temporadas para los clientes por ello la desarrollo del Datamart es factible.

Este prototipo utilizará la Base de Datos Mysql para el Datamart y Weka para su análisis con ello tendrá las siguientes características:

- Son procesadas por el usuario final
- Se procesa en línea y con gran cantidad de datos
- Se extrajo toda la información que el usuario necesita para información relevante y concreta para la correcta toma de decisiones
- Los datos serán extraídos de las tablas transaccionales FRUDONMAR.

Factibilidad Técnica

Datamart tiene la característica de obtener información segregada de diferentes bases de datos para consolidarlas en un repositorio que nos permita obtener información filtrada y procesada de un departamento en específico, con ello podremos realizar consultas concretas permitiéndonos un control de la información óptimo.

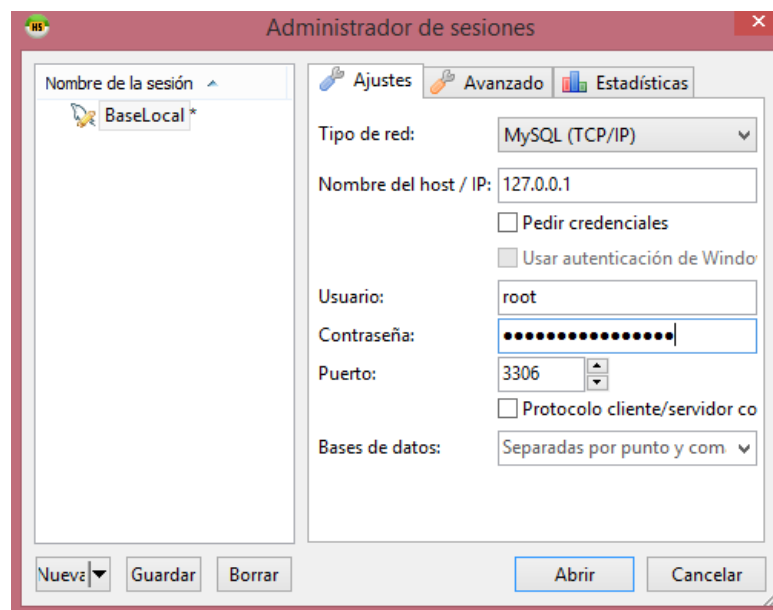
Esta tecnología es óptima y sencilla de aplicar ya que existe mucha información disponible que nos permite orientarnos al desarrollo y elaboración de un correcto diseño del Datamart y herramientas de análisis de datos.

Dispondrá de una interfaz que permita una correcta interacción con el usuario con ello facilitará la generación de reportes óptimos para la toma de decisiones (Los usuarios deberán ser capacitados por el uso del proceso).

La factibilidad para los componentes de hardware y software es viable ya que contamos en nuestro medio de herramientas y equipos que pueden soportar nuestro desarrollo.

Recurso de Software

Se utiliza como gestos de base de Datos Mysql para realizar el diseño de Datamart con HeidiSQL como herramienta que permitirá gestionar de manera rápida y sencilla las bases de datos, pero el punto más importante es que facilitará la comunicación y obtención de los datos que se va a utilizar en el prototipo, puesto que la empresa utiliza esta base de datos para el almacenamiento de la información de los sistemas modulares que posee.



Manejador de Base de Datos HeidiSQL

Para el análisis de la información se utilizará WEKA, que trabaja con algoritmos de clasificación y filtro que nos ayudan a realizar análisis de datos y

modelos predictivos, con una interfaz gráfica de usuario sencilla que facilita el acceso funcional, ideal para la minería de datos y procesamiento de grandes volúmenes de información que permite realizar conexiones a diferentes bases de datos incluido Mysql con sólo instalar un archivo de conexión, con ello se puede extraer directamente del Datamart la información, por ello es la herramienta ideal para este estudio.

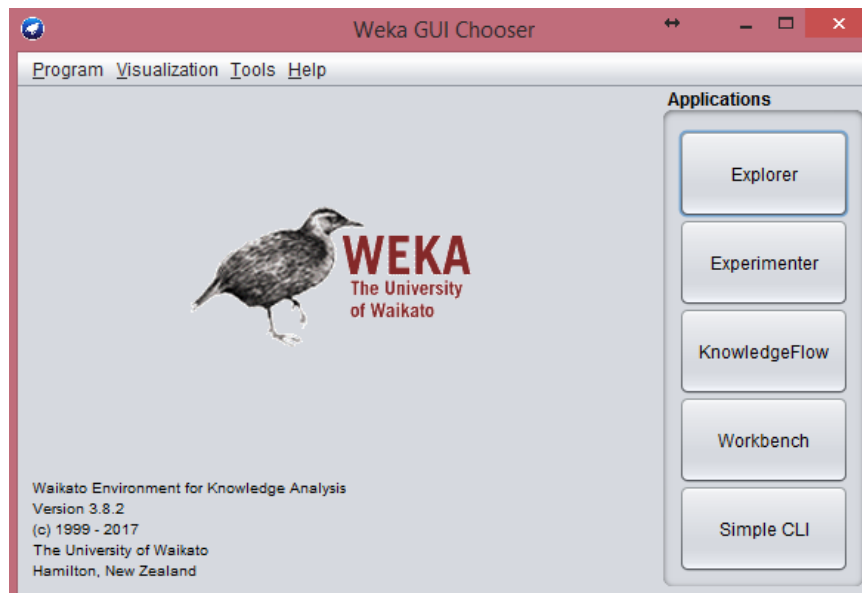
Utilizaremos algoritmos de reglas de clasificaciones que son basadas en reglas difusas permitiéndonos evaluar la información y encontrando reglas de decisiones aceptables para el procesamiento de los datos iniciales. Posteriormente con este grupo de reglas se realizará un análisis con un mayor número de datos y filtraremos las que cumplan con las características necesarias, evaluaremos nuestra muestra con los siguientes algoritmos:

Tabla 2. Criterio para selección de algoritmos

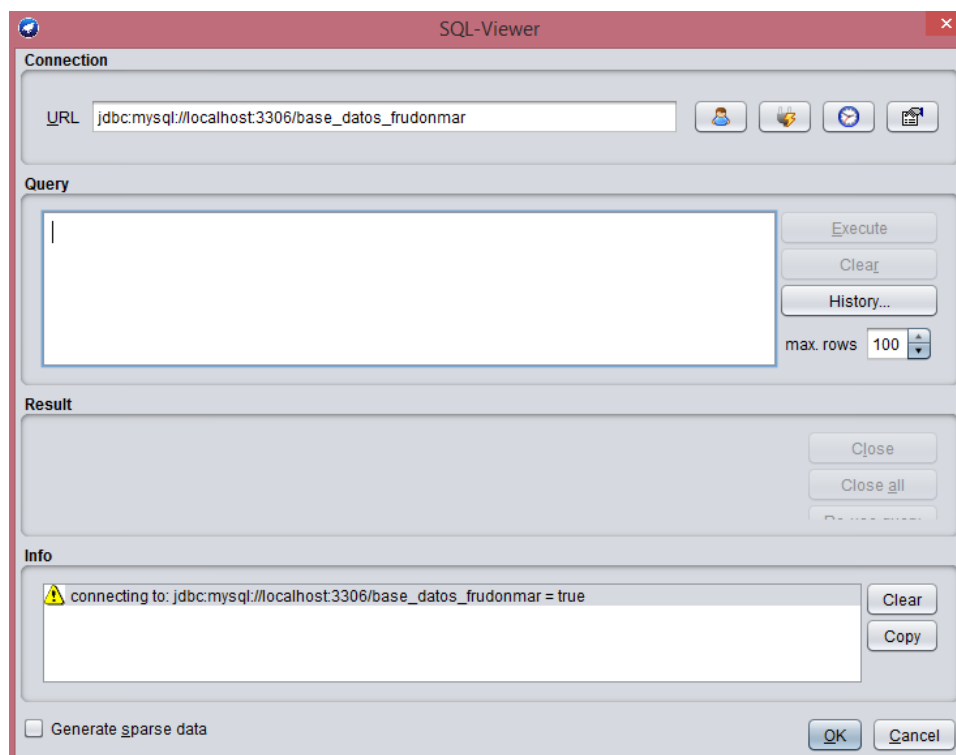
Algoritmo	Método de Clasificación	Lógica Difusa	Construcción de Reglas Predictivas
ZeroR	SI	NO	NO
Furia	SI	SI	SI
Jrip	SI	SI	SI
MODLEM	SI	SI	SI
OneR	SI	NO	SI

Según el criterio de selección de los algoritmos evaluados que cumplen con los parámetros establecidos para nuestro estudio son:

- Furia.- que es un algoritmo de inducción de reglas desordenadas “Difusas”.
- Jrip.- algoritmo RIPPER regla de inducción eficaz.
- Modlem.’ Algoritmo de clasificación que trata de encontrar la mínima cantidad de reglas no redundantes.



Programa Weka y sus funcionalidades



Conexión Weka con la base de datos MYSQL

Para la extracción de la información de la base de datos y los procesos de consolidación de la información para el Datamart se usará php con librerías javascript y CSS para la interfaz gráfica.

Para la instalación de las herramientas descritas se utilizará Appserv, que es un paquete de instalación OpenSource para Windows con Apache, MySQL y PHP

realizando la configuración de las aplicaciones de forma automática, lo que facilitará ejecutar un completo servidor web.

Este servidor web “Apache” junto con el gestor de páginas dinámicas “php” y el motor de base de datos “Mysql” son las herramientas elegidas para desarrollar el prototipo para FRUDONMAR, como plataforma de su sistema.

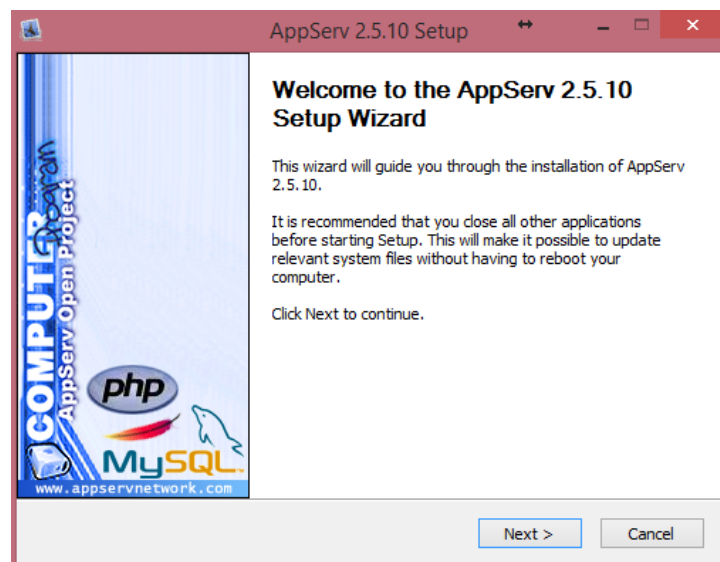
Appserv 2.5.10 (última versión estable)

Paquetes:

Apache 2.2.8

PHP 5.2.6

MySQL 5.0.51b



Paquete de instalación AppServ

Recursos de Hardware

Para el desarrollo del prototipo se necesitará un computador en el cual se realizará el desarrollo de la herramienta, un disco externo para guardar respaldos cotidianos de lo desarrollado y para levantar el servicio web un servidor en el cual se instalará el Appserv.

Factibilidad Económica

Se detalla el costo de la adquisición de los equipos solicitados y el personal para el desarrollo del proyecto.

Tabla 3. Recurso Humano para el desarrollo del prototipo

Actividades	Horas	Valor por hora	Total
Identificar los datos que serán procesados para ingresarlos en el Datamart	12	10.00	120.00
Análisis y diseño del Datamart	6	10.00	60.00
Creación del Datamart	5	10.00	50.00
Desarrollo de los proceso de consolidación de la información en el Datamart	15	10.00	150.00
Prueba de Validación de la información entregada por la empresa con la herramienta de Weka y obtención de las reglas de decisiones.	8	10.00	80.00
Creación de los procesos que validen las reglas de decisiones detalladas por Weka con la información entregada por la empresa	12	10.00	120.00
Desarrollo del prototipo con las reglas de decisiones	2	10.00	20.00
Pruebas y depuración del Prototipo	6	10.00	60.00
Total	34		620.00

Nota: montos expresados en USDólares

Tabla 4. Recurso tecnológico para el desarrollo del prototipo

Equipo	Cantidad	Valor	Total
Computador	1	800.00	800.00
Disco Duro Externo	1	80.00	80.00
Servidor	1	1,400.00	1,400.00
Total			2,280.00

Nota: montos expresados en USDólares

La FRUDONMAR facilitará un computador del área de sistemas el cual será utilizado para el desarrollo del proyecto junto con un disco externo para realizar los respaldos del desarrollo del prototipo. Posee un servidor en línea y con las herramientas instaladas como son: Apache, Php y MySQL en el cual se podrá cargar la herramienta desarrollada.

Para el factor humano el desarrollo de todo el proyecto se lo realizará como parte de este trabajo de titulación, por ello no tendrá costo alguno.

Por todo lo expuesto la elaboración del proyecto es factible económicamente.

Extracción de la información necesaria para el Datamart

Se detalla al área de Sistemas de la empresa FRUDONMAR entregue la información correspondiente al periodo establecido “2010-2017” para generar las reglas de decisiones:

Las herramientas con las que vamos a realizar el proyecto son las siguientes:

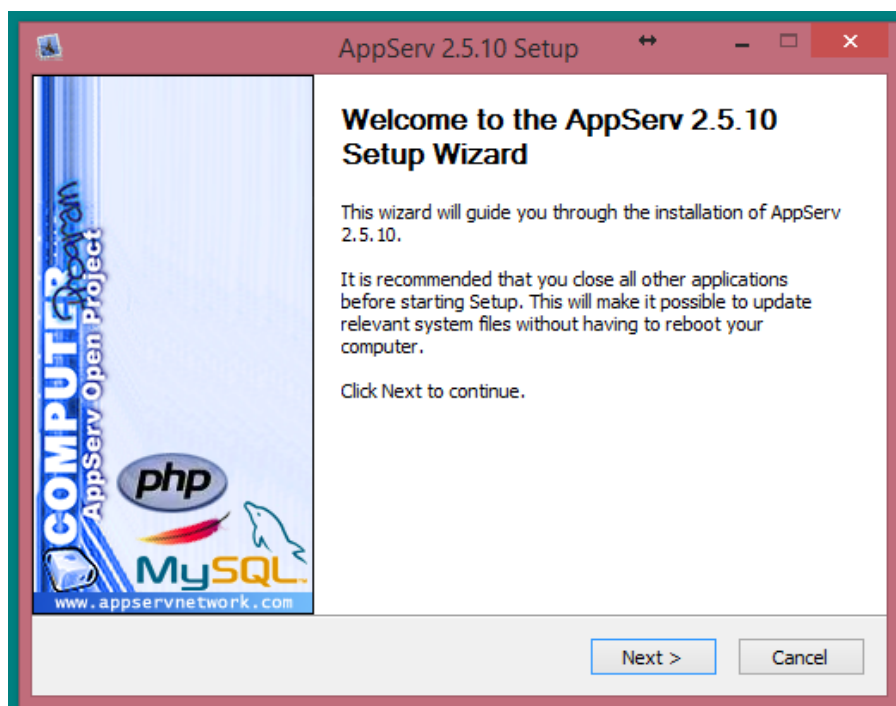
Appserv. Paquete de instalación con los siguientes programas:

Appserv 2.5.10 (última versión estable)

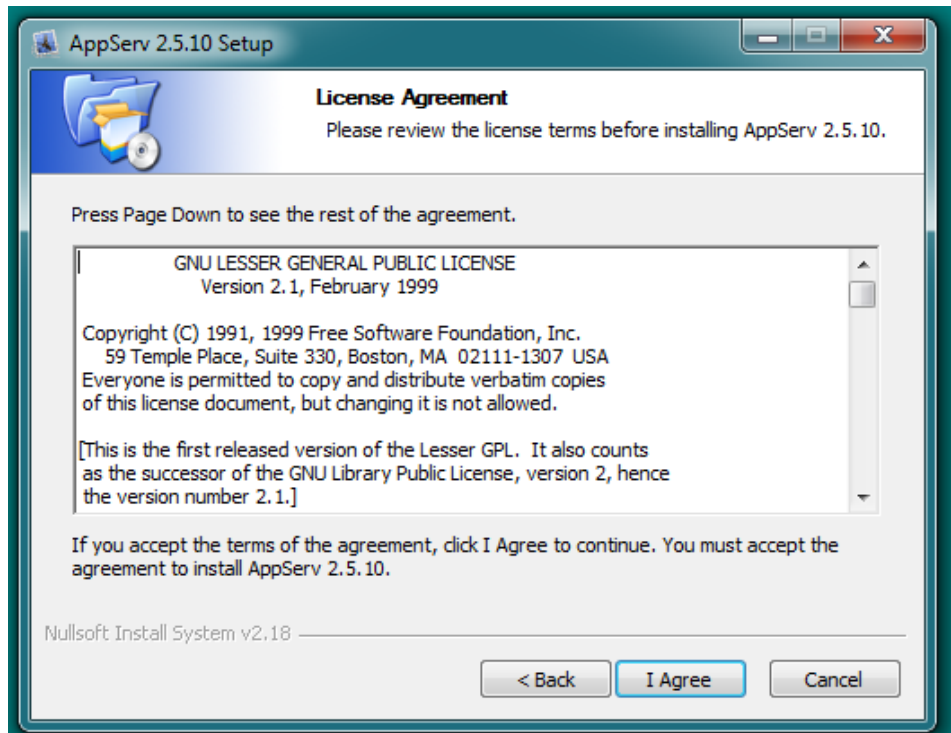
- Apache 2.2.8
- PHP 5.2.6
- MySQL 5.0.51b

Página de descarga: <https://www.appserv.org/download/>

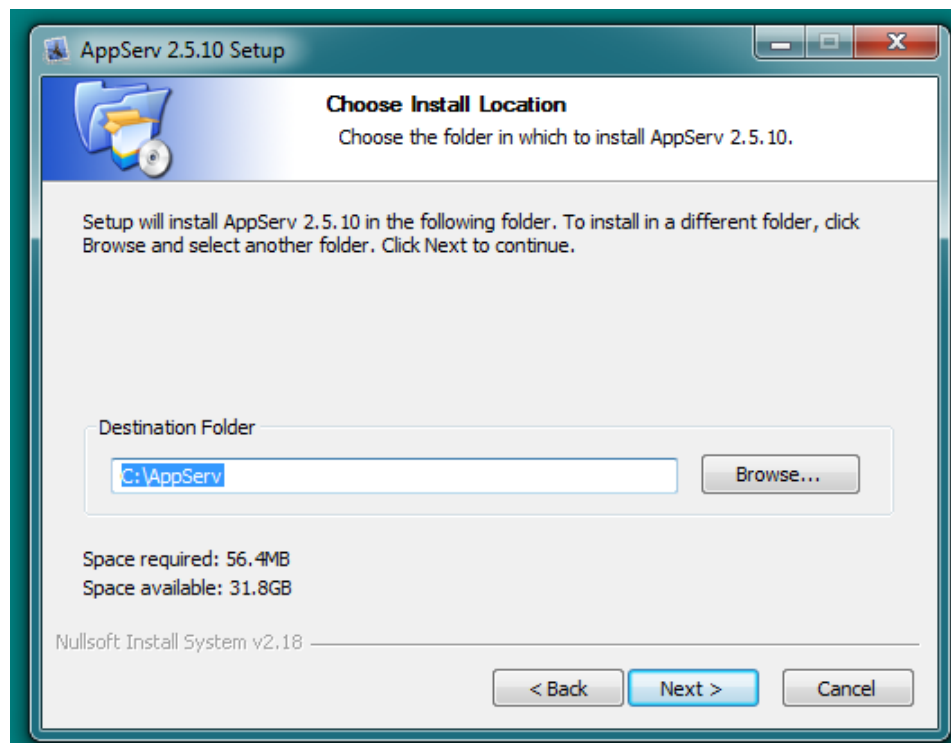
Secuencia de instalación del programa.



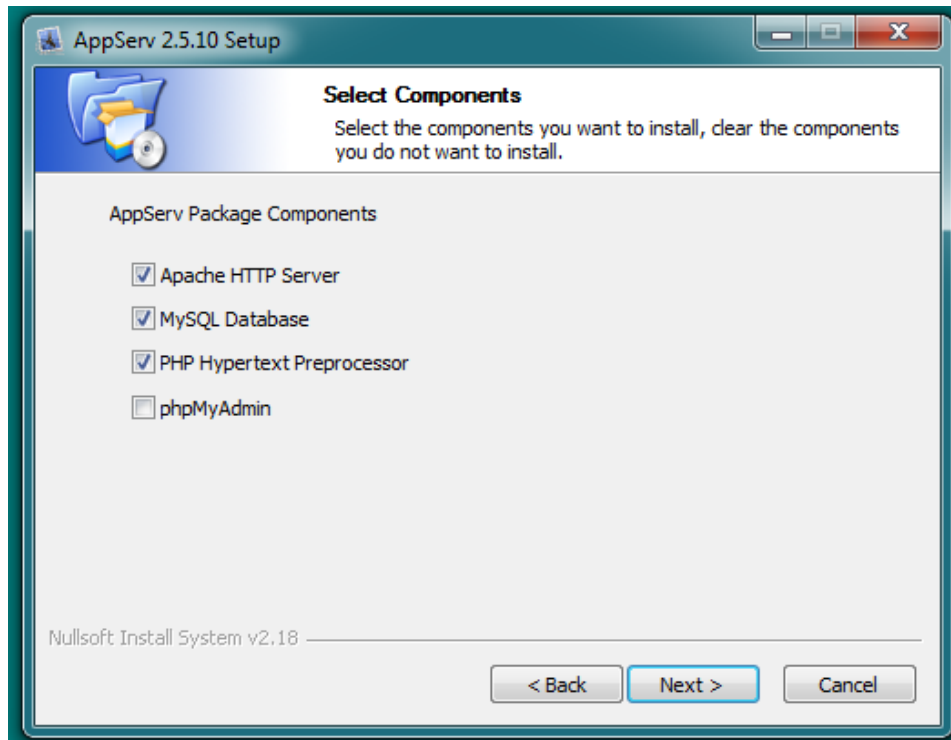
Pantalla de inicio del Wizard



Términos de licencia



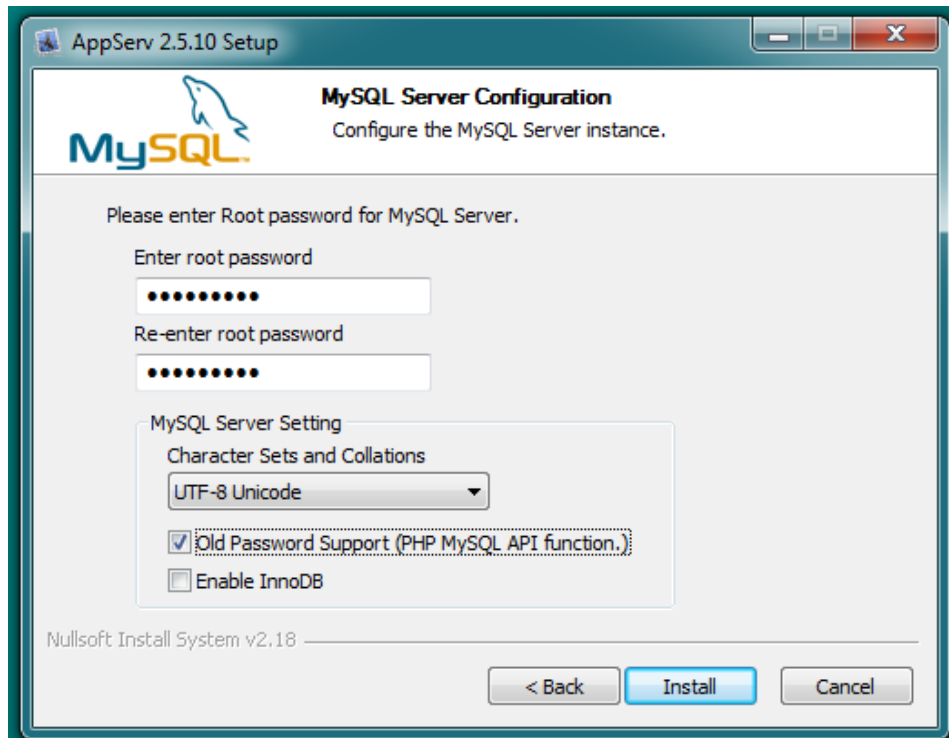
Ubicación de la instalación



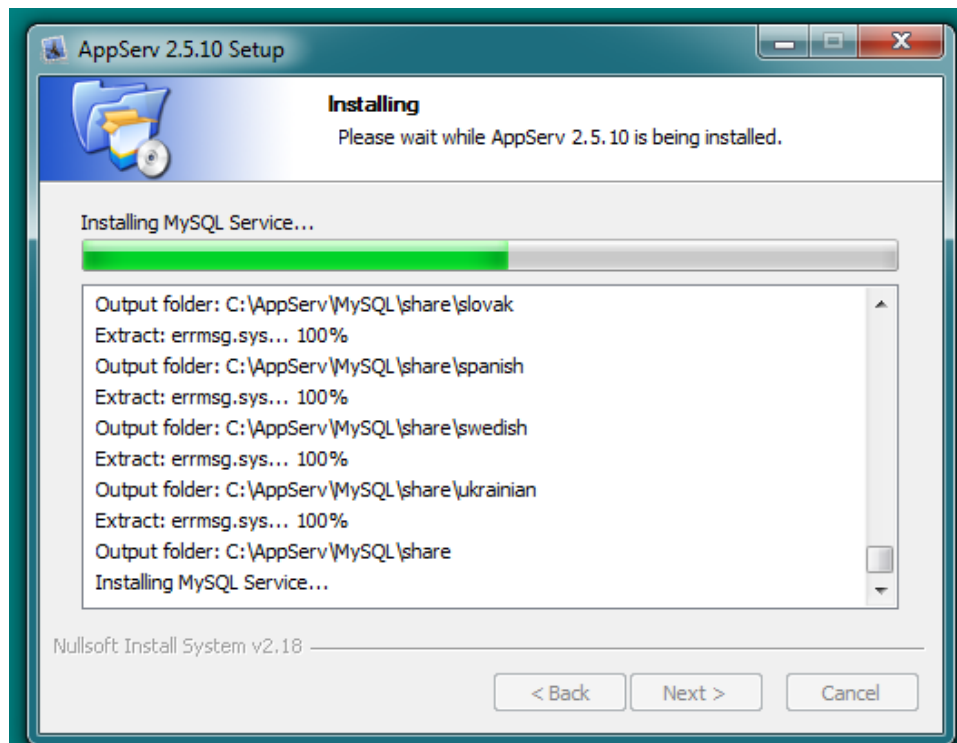
Programas que serán instalados



Nombre del servidor Web, Mail de Administración y puerto HTTP del Apache



Clave de Acceso a la Base de Datos MySQL con el usuario Administrador



Instalación de los programas seleccionados

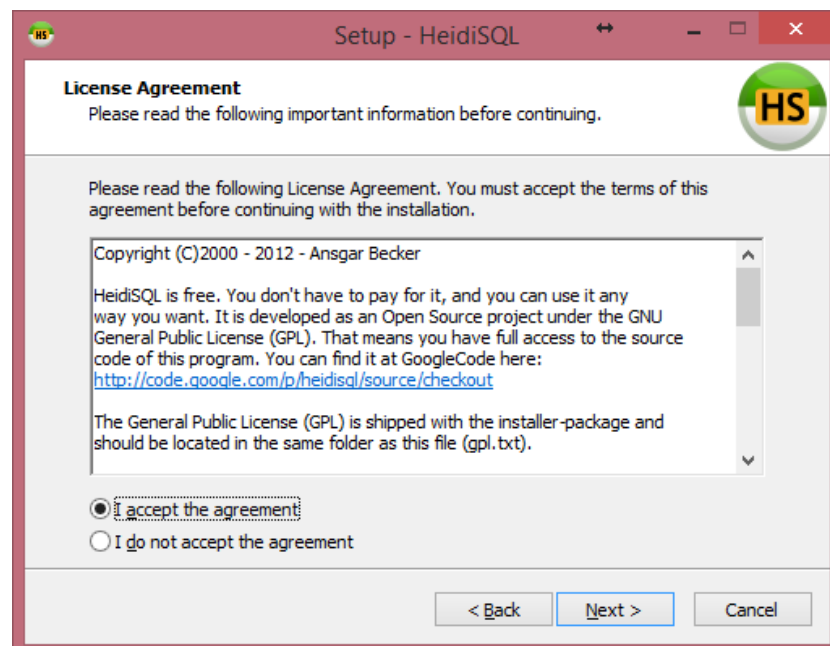
HeidiSQL.- gestor de base de datos

Página de descarga: <https://www.heidisql.com/download.php>

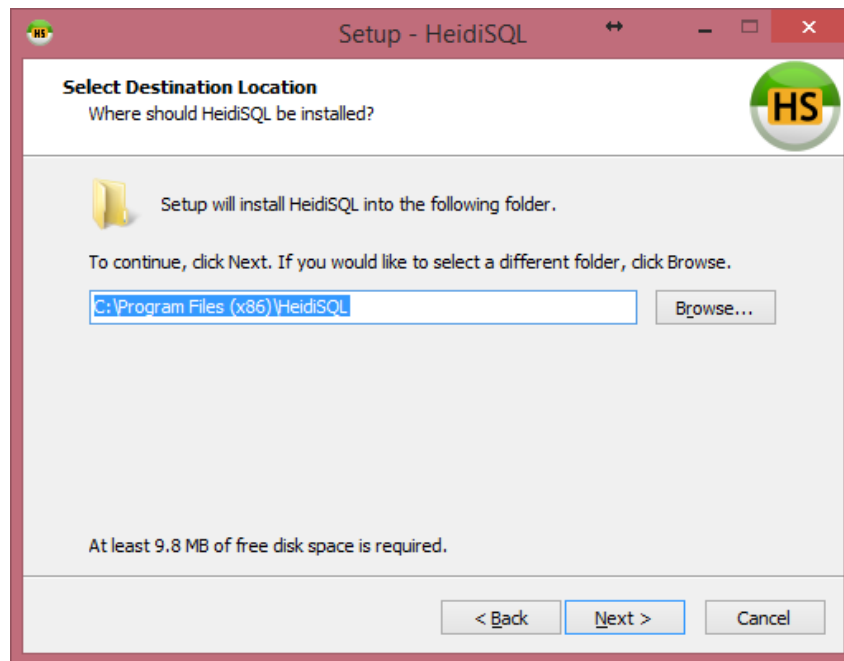
Secuencia de instalación del programa



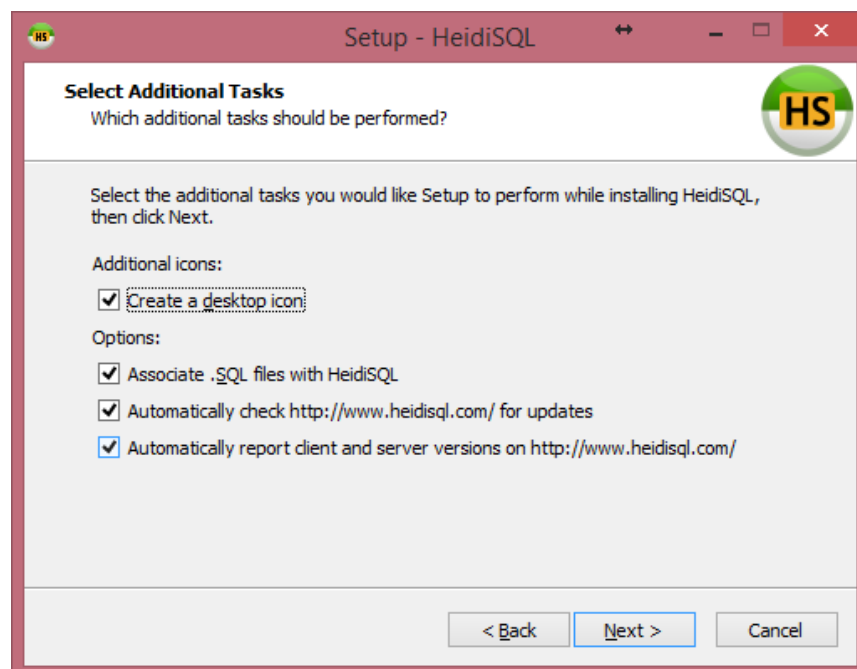
Pantalla inicial del wizard



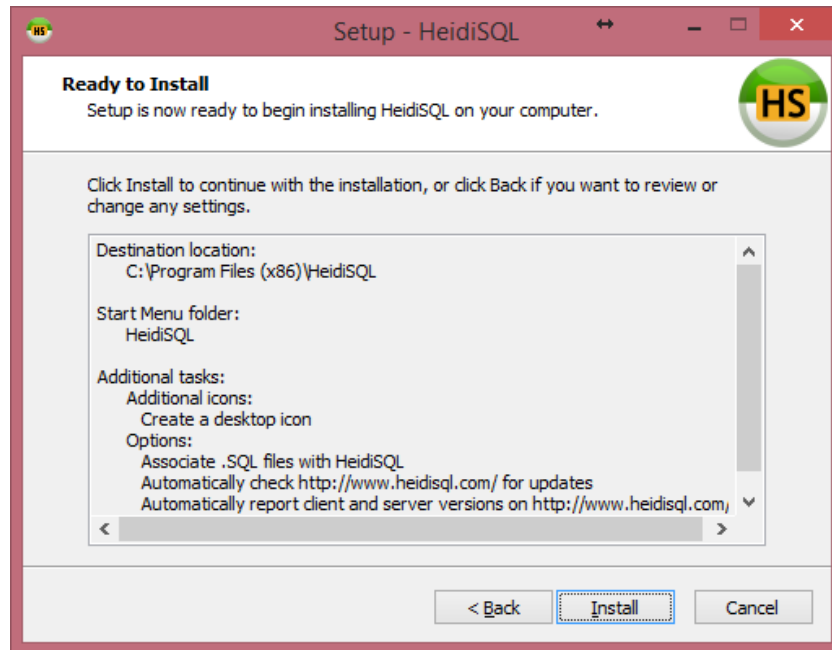
Términos de licencia



Ubicación de la instalación



Opciones que serán instaladas



Listo para la instalación



Instalación terminada

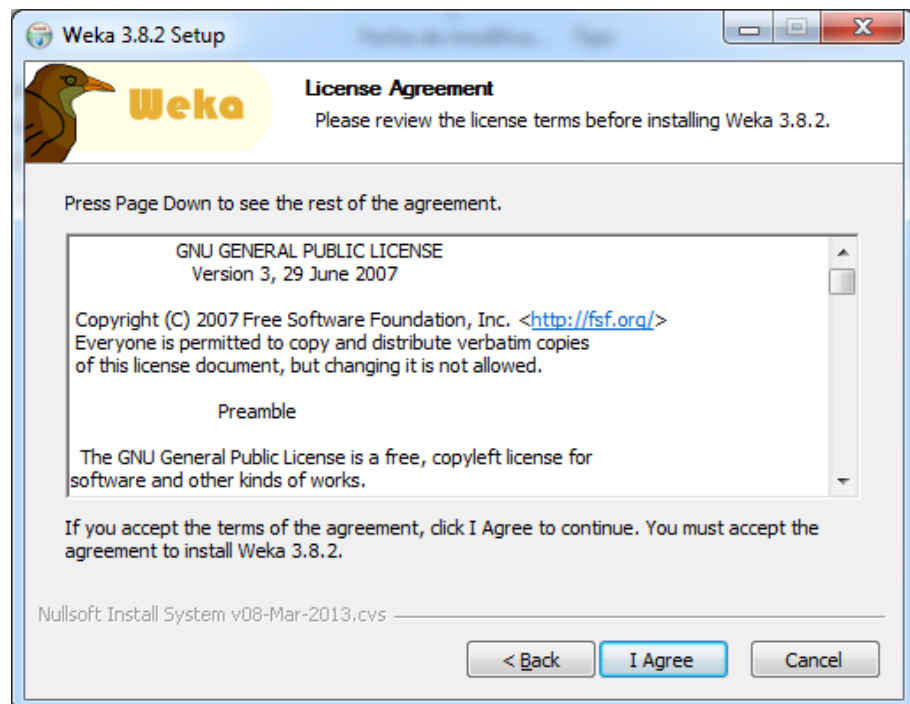
Weka.-Plataforma para el análisis de minería de datos.

Desarrollado en Java

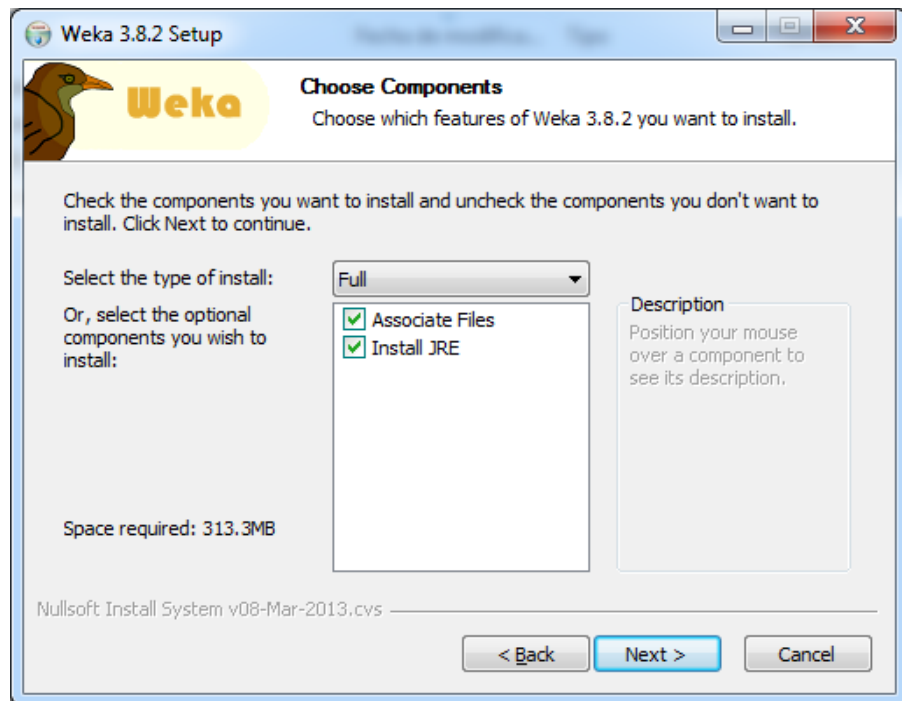
Página de descarga: <https://www.cs.waikato.ac.nz/ml/weka/downloading.html>



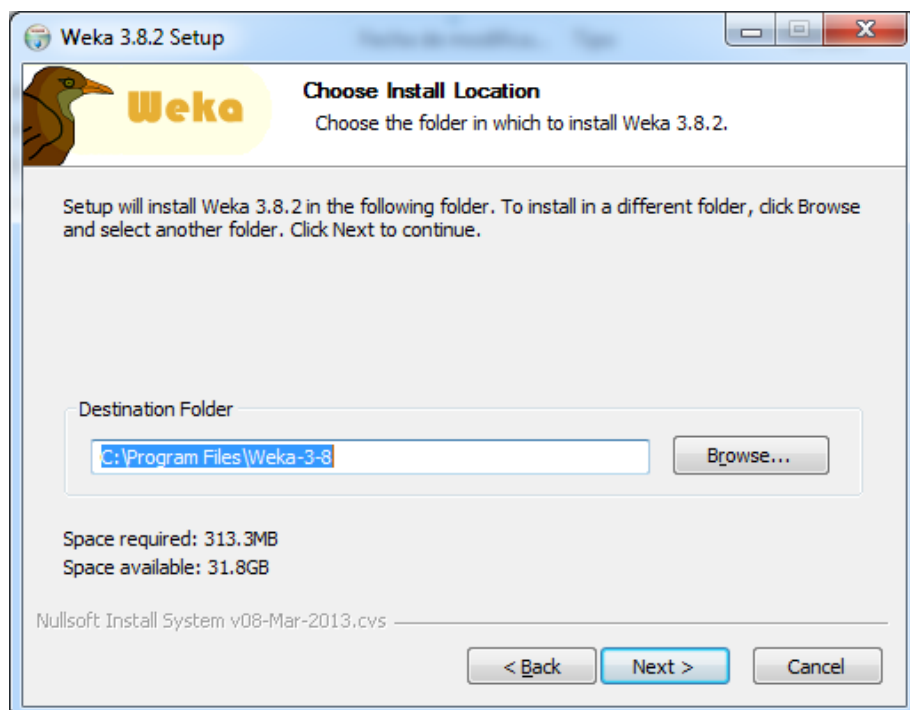
Página Principal del Wizard



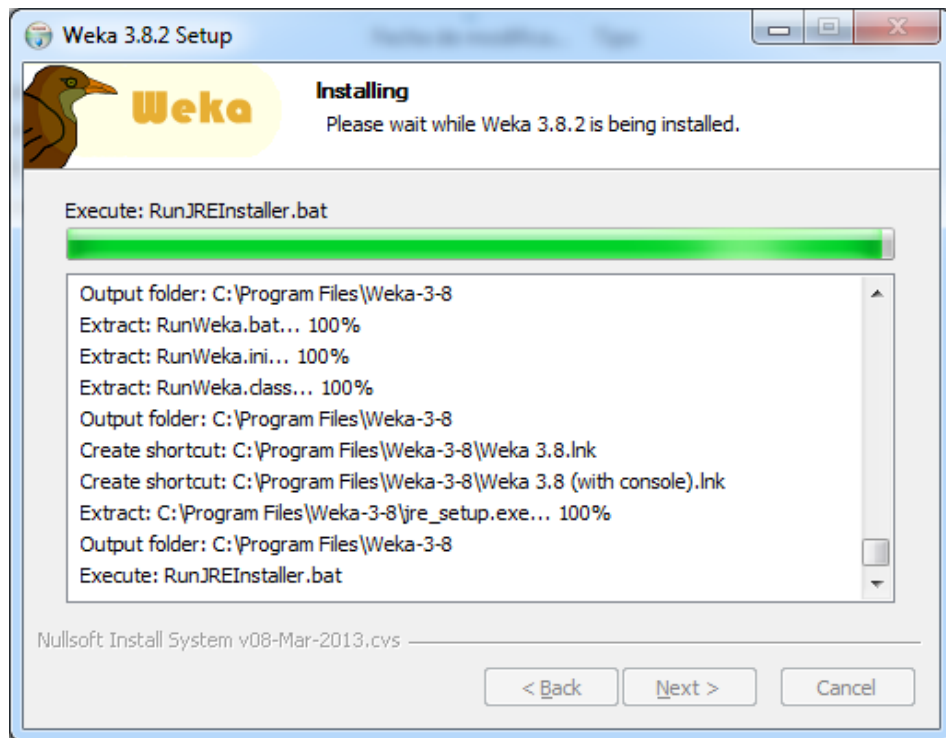
Términos de licencia



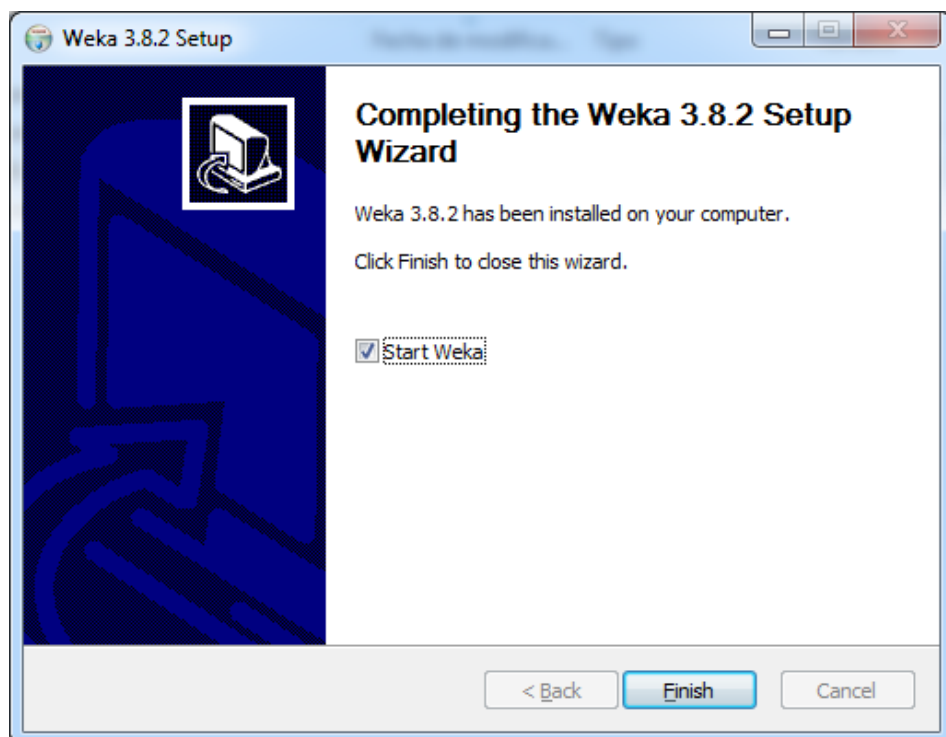
Programas que serán instalados



Ubicación de la instalación

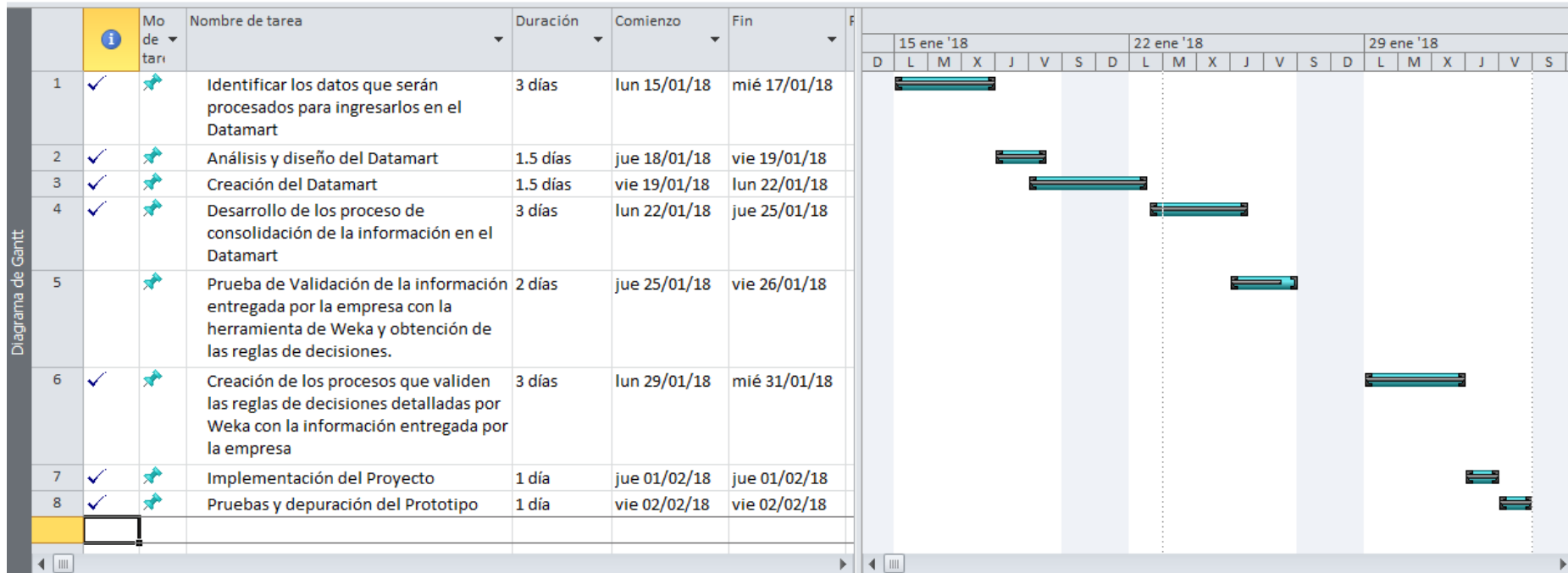


Proceso de Instalación



Final de la instalación

CRONOGRAMA DE ACTIVIDADES



Nota.

Los días de trabajo serán de 4 horas.

ANÁLISIS DE LOS DATOS Y SELECCIÓN DE CARACTERÍSTICAS

DATOS QUE SERÁN ANALIZADOS

Datos definidos y detallados por la empresa Frutera Don Marcos Frudonmar S.A. que serán analizados en la herramienta de minería de datos

Las variables de entrada o antecedente:

Cantidad.- índice de Cantidad de productos vendidos.

Gastos.- índice de Gasto de exportación por unidad del producto, en la cual contiene:

- Gastos de almacenaje
- Aranceles
- Flete

Marguen de Utilidad en venta.-Índice porcentual de utilidad de venta con relación al valor de compra “FOB”

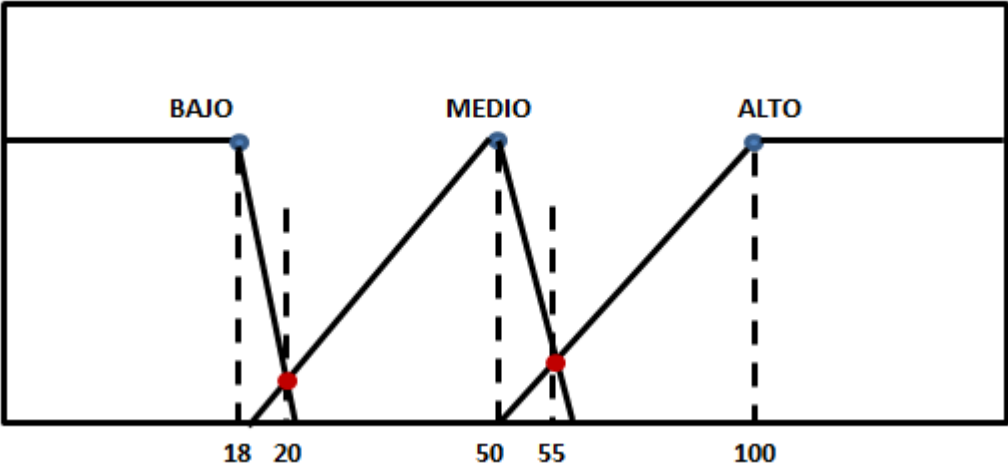
Variable de Salida o consecuente:

Utilidad total de la venta.- que corresponde a la ganancia total luego de todos los gastos que conlleva la venta.

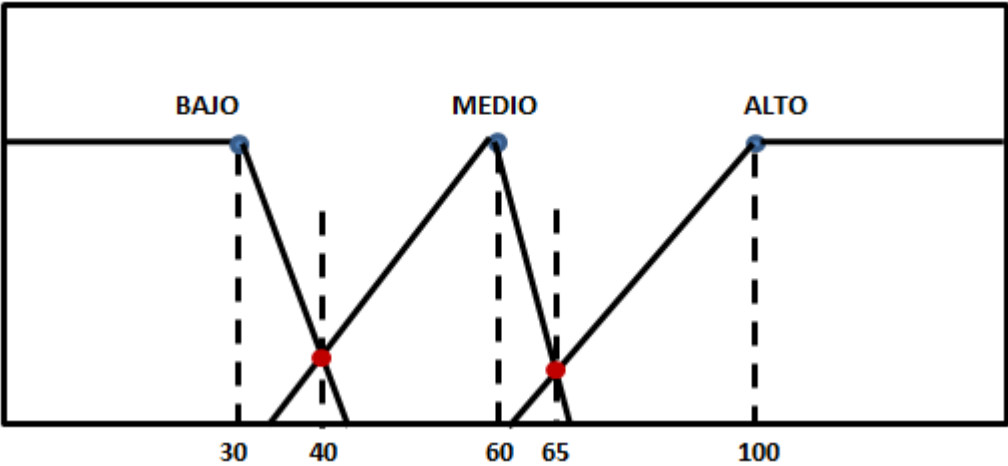
Cada variable es medida en tres posibles clases: BAJO, MEDIO, ALTO, cada uno de estos valores lingüísticos son definidos por un rango de valores según sea la variable.

RANGO DE VALORES DE LAS VARIABLES DE ENTRADA

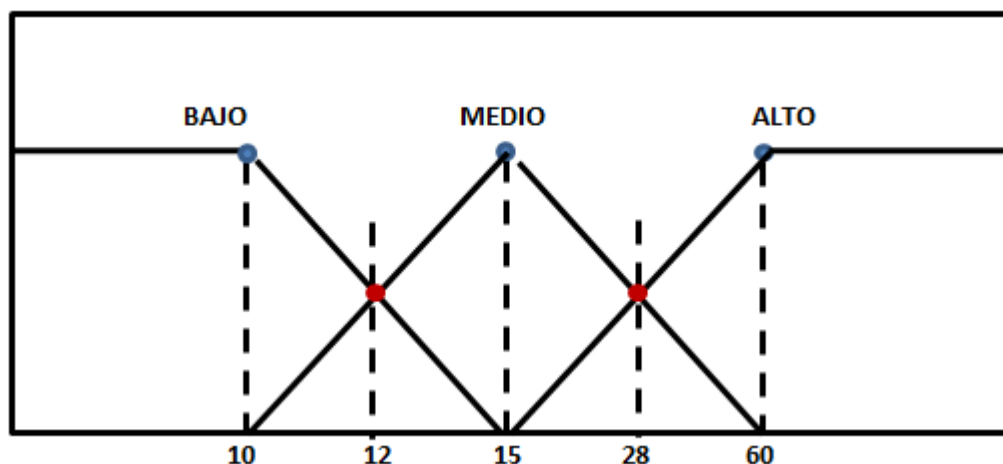
VARIABLE LINGÜÍSTICA CANTIDAD



VARIABLE LINGÜÍSTICA GASTOS

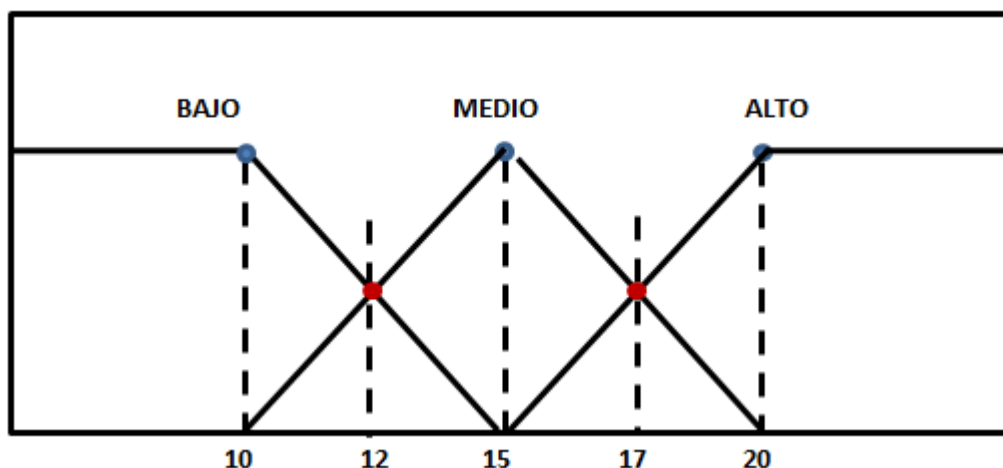


VARIABLE LINGÜÍSTICA MARGEN



RANGO DE VALOR DE LAS VARIABLE DE SALIDA

VARIABLE LINGÜÍSTICA UTILIDAD



Estos valores han sido definidos por el área financiera y de ventas de la empresa Frutera Don Marcos Frudonmar S.A. con los cuales han trabajado en el transcurso del tiempo y que para nuestro estudio serán los que definan que valor lingüístico se asignará a la variable según el valor que contenga.

Sistema de Inferencia Borroso.

Se utilizará para el modelo de sistema difuso el método Mamdani de la forma Min-Max, utilizando variables de entrada y salida de forma lingüística para nuestra

evaluación, reglas de formato (if-else) y una base de conocimiento de antecedente y consecuente.

Por medio de este modelo definiremos los valores numéricos de la información entregada por la empresa frutera Don Marcos Frudonmar S.A. en valores lingüísticos para realizar nuestro análisis.

Posteriormente el resultado que se arroja en formato lingüístico lo transformaremos en formato numérico para un mejor análisis del usuario.

ANÁLISIS DE LA INFORMACIÓN CON LA HERRAMIENTA DE MINERÍA DE DATOS PARA GENERAR LA REGLA DE DECISIONES

Procederemos a realizar la corrida de la información inicial para encontrar las reglas de decisiones que luego procederemos a verificar su validez con un mayor número de datos filtrando las reglas que tendrán un porcentaje bajo de aciertos para obtener las de mayor porcentaje de acierto y elaborar nuestro prototipo.

Para este análisis la información en formato numérico ha sido establecida en valor lingüístico con los tres posibles combinaciones “Alto, Medio y Bajo”.

Lo realizaremos con la Herramienta Weka con tres tipos de algoritmos que nos ayudarán a verificar la información mediante el modelo Mandani.

Primera Corrida de Datos con Algoritmo Modlem – Rt 1

==== Run information ====

Scheme: weka.classifiers.rules.MODLEM -RT 1 -CM 1 -CS 8 -AS 0

Relation DatosRegistros

Instances: 5955

Attributes: 4

cantidad

gastos

marguen

utilidad

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Rule 1. (marguen in {ALTO})&(cantidad in {MEDIO, BAJO}) => (utilidad = ALTO) (515/515, 99.04%)

Rule 2. (marguen in {BAJO})&(cantidad in {ALTO}) => (utilidad = ALTO) (5/5, 0.96%)

Rule 3. (gastos in {MEDIO, ALTO}) => (utilidad = BAJO) (2023/2023, 95.29%)

Rule 4. (marguen in {BAJO})&(cantidad in {BAJO}) => (utilidad = BAJO) (100/100, 4.71%)

Rule 5. (cantidad in {ALTA}) => (utilidad = MEDIO) (2/2, 100%)

Number of Rules: 5

Time taken to build model: 0.47 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	5125	86.0621 %
--------------------------------	------	-----------

Incorrectly Classified Instances	830	13.9379 %
Kappa statistic		0.7366
Mean absolute error		0.0929
Root mean squared error		0.3048
Relative absolute error		24.4749 %
Root relative squared error		69.9675 %
Total Number of Instances	5955	

=== Detailed Accuracy By Class ===

Area	Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC
	ALTO	1,000	0,281	0,783	1,000	0,878	0,751	0,860	0,783
	BAJO	0,865	0,000	1,000	0,865	0,928	0,889	0,933	0,921
	MEDIO	0,004	0,000	1,000	0,004	0,008	0,060	0,502	0,088
	Weighted Avg.	0,861	0,142	0,891	0,861	0,825	0,750	0,860	0,781

=== Confusion Matrix ===

a	b	c	<-- classified as
3000	0	0	a = ALTO
330	2123	0	b = BAJO
500	0	2	c = MEDIO

Segunda Corrida de Datos con Algoritmo Furia

=== Run information ===

Scheme: weka.classifiers.rules.FURIA -F 3 -N 2.0 -O 2 -S 1 -p 0 -s 0

Relation: DatosRegistros

Instances: 5955

Attributes: 4

cantidad

gastos

marguen

utilidad

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

FURIA rules:

=====

(marguen = ALTO) and (cantidad = MEDIO) => utilidad=ALTO (CF = 1.0)

(marguen = ALTO) and (gastos = BAJO) => utilidad=ALTO (CF = 0.97)

(cantidad = ALTO) and (gastos = BAJO) => utilidad=ALTO (CF = 0.97)

(marguen = BAJO) => utilidad=BAJO (CF = 1.0)

(marguen = MEDIO) => utilidad=MEDIO (CF = 0.67)

Number of Rules : 5

Time taken to build model: 0.13 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	5617	94.3241 %
Incorrectly Classified Instances	338	5.6759 %
Kappa statistic		0.9025
Mean absolute error		0.0376
Root mean squared error		0.1935
Relative absolute error		9.8943 %
Root relative squared error		44.4035 %
Total Number of Instances		5955

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,998	0,028	0,973	0,998	0,986	0,971	0,986	0,973	ALTO
0,864	0,001	0,998	0,864	0,926	0,886	0,932	0,920	BAJO
1,000	0,046	0,668	1,000	0,801	0,798	0,977	0,668	MEDIO
Weighted Avg.								
0,943	0,019	0,957	0,943	0,945	0,921	0,963	0,926	

=== Confusion Matrix ===

a	b	c	<-- classified as
2995	5	0	a = ALTO
83	2120	250	b = BAJO
0	0	502	c = MEDIO

Tercera Corrida de Datos con Algoritmo Jrip

=== Run information ===

Scheme: weka.classifiers.rules.JRip -F 3 -N 2.0 -O 2 -S 1

Relation: DatosRegistros

Instances: 5955

Attributes: 4

cantidad

gastos

marguen

utilidad

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

JRIP rules:

=====

(marguen = MEDIO) => utilidad=MEDIO (752.0/250.0)

(marguen = BAJO) and (cantidad = MEDIO) => utilidad=BAJO (2020.0/0.0)

(marguen = BAJO) and (cantidad = BAJO) => utilidad=BAJO (100.0/0.0)

(gastos = ALTO) => utilidad=BAJO (3.0/0.0)

(marguen = ALTO) => utilidad=ALTO (3080.0/80.0)

Number of Rules : 5

Time taken to build model: 0.07 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	5624	94.4416 %
Incorrectly Classified Instances	331	5.5584 %
Kappa statistic		0.9045
Mean absolute error		0.0551
Root mean squared error		0.166
Relative absolute error		14.5213 %
Root relative squared error		38.1076 %
Total Number of Instances		5955

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1,000	0,027	0,974	1,000	0,987	0,973	0,984	0,967	ALTO
0,865	0,000	1,000	0,865	0,928	0,889	0,971	0,972	BAJO
1,000	0,046	0,668	1,000	0,801	0,798	0,975	0,635	MEDIO

Weighted Avg.

0,944 0,018 0,959 0,944 0,947 0,924 0,978 0,941

==== Confusion Matrix ====

a	b	c	<-- classified as
2999	1	0	a = ALTO
80	2123	250	b = BAJO
0	0	502	c = MEDIO

EVALUACIÓN DE LAS REGLAS

Se procedió a verificar mediante la herramienta desarrollada las reglas obtenidas con la herramienta de minería de datos, con ello se obtuvo los siguientes resultados.

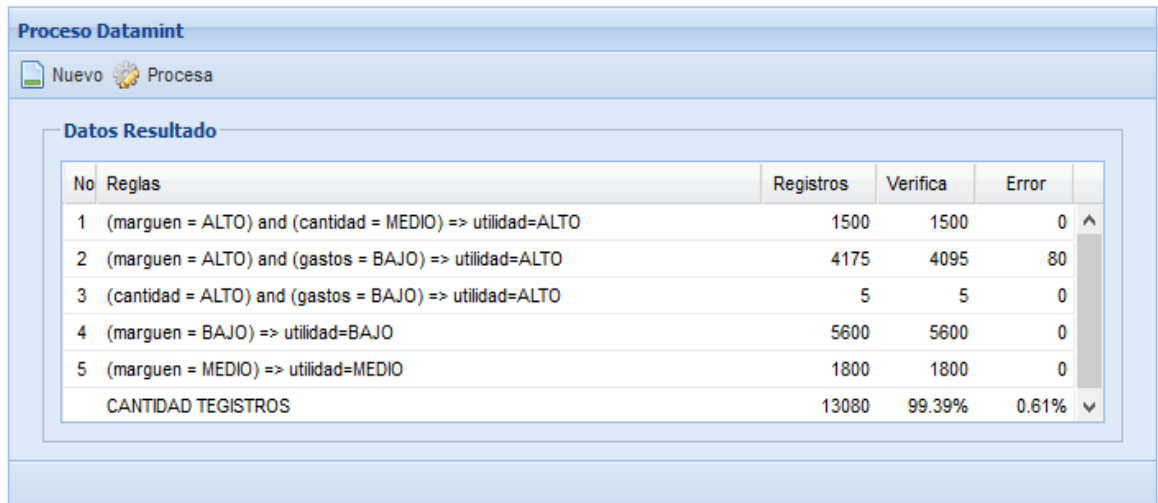
Primera Corrida de Datos con Algoritmo Modlem – Rt 1

No Reglas	Registros	Verifica	Error
1 marguen in {ALTO}&(cantidad in {MEDIO, BAJO}) => (utilidad = ALTO)	15	15	0
2 (gastos in {MEDIO, ALTO}) => (utilidad = BAJO)	4120	4120	0
3 (marguen in {BAJO}&(cantidad in {BAJO}) => (utilidad = BAJO)	1480	1480	0
4 (cantidad in {ALTA}) => (utilidad = MEDIO)	0	0	0
5 CANTIDAD REGISTROS	13080	42.93%	57.07%

Para el primer grupo de reglas de decisiones se obtuvo un porcentaje de acierto del 42% con una muestra de registros de 13080.

Con ello se puede deducir que las reglas analizadas no son confiables para la elaboración de prototipo, esto se debe a que no contempla todas las opciones que pueden tomar las variables independientes.

Segunda Corrida de Datos con Algoritmo Furia



No	Reglas	Registros	Verifica	Error
1	(marguen = ALTO) and (cantidad = MEDIO) => utilidad=ALTO	1500	1500	0
2	(marguen = ALTO) and (gastos = BAJO) => utilidad=ALTO	4175	4095	80
3	(cantidad = ALTO) and (gastos = BAJO) => utilidad=ALTO	5	5	0
4	(marguen = BAJO) => utilidad=BAJO	5600	5600	0
5	(marguen = MEDIO) => utilidad=MEDIO	1800	1800	0
CANTIDAD REGISTROS		13080	99.39%	0.61%

Para el segundo grupo de reglas obtenidas se logró un acierto del 99%, existiendo un error del 1%.

Esto representa 80 registros de los 13080 analizados que no cumplen con las reglas de decisiones.

Análisis de los registros con errores para el segundo grupo de las reglas de decisiones

Con esta cantidad de registros que no cumplen con las reglas de decisiones se pudo realizar la verificación con el área de sistemas de la empresa por qué no se aplican a estas reglas.

Luego de que ellos realizaran la revisión correspondiente se nos indicó que estos registros tienen valores de descuento en el precio de venta y por ello el margen de utilidad disminuyó, este proceso no afecta el registro principal manteniendo el valor inicial de venta y ya que en el proceso se analiza los datos de esta tabla se mostrará este desfase en las reglas.

Tercera Corrida de Datos con Algoritmo Jrip

No	Reglas	Registros	Verifica	Error
1	(marguen = MEDIO) => utilidad=MEDIO	1800	1800	0
2	(marguen = BAJO) and (cantidad = MEDIO) => utilidad=BAJO	4120	4120	0
3	(marguen = BAJO) and (cantidad = BAJO) => utilidad=BAJO	1480	1480	0
4	(gastos = ALTO) => utilidad=BAJO	0	0	0
5	(marguen = ALTO) => utilidad=ALTO	5675	5595	80
CANTIDAD REGISTROS		13080	99.35%	0.65%

Para el tercer grupo de reglas obtenidas se logró un acierto del 99%, existiendo un error del 1%.

Esto representa 80 registros de los 13080 analizados que no cumplen con las reglas de decisiones, que junto con la segunda corrida detallada en el anterior proceso podemos ir armando nuestras reglas de decisiones.

Análisis de los registros con errores para el tercer grupo de las reglas de decisiones

De la misma forma que se realizó con la segunda corrida los registros que no cumplen con las reglas de decisiones se realiza la verificación con el área de sistemas de la empresa por qué no se aplican a estas reglas.

Luego del análisis del área de Sistemas por parte de la empresa FRUDONMAR, se determinó que son los mismos registros de la segunda corrida, por lo cual los valores de descuento no se aplican en el precio de venta y por ello el margen de utilidad disminuyó, esto no afecta el registro principal por ello la utilidad del resultado no concuerda con las reglas.

El análisis de las pruebas se las pudo realizar ya que en la estructura del Datamart se almacena el código del registro que corresponde al dato extraído de la base de datos de la empresa.

Enviando este código al área de sistemas de la empresa pudo verificar dentro de sus registros la inconsistencia de la utilidad detallada por las reglas de decisiones.

ESTRUCTURA DE LAS REGLAS DE DECISIONES PARA EL DISEÑO DEL PROTOTIPO

Luego de la validación realizada por los procesos generados para los registros se pudo obtener que la segunda y tercera corrida arrojaran los mismos resultados, de esta manera podemos armar nuestro árbol de decisiones para implementarlo en el prototipo de Toma de decisiones.

Regla de Decisiones de la segunda corrida

1. (marguen = ALTO) and (cantidad = MEDIO) => utilidad=ALTO (CF = 1.0)
2. (marguen = ALTO) and (gastos = BAJO) => utilidad=ALTO (CF = 0.97)
3. (cantidad = ALTO) and (gastos = BAJO) => utilidad=ALTO (CF = 0.97)
4. (marguen = BAJO) => utilidad=BAJO (CF = 1.0)
5. (marguen = MEDIO) => utilidad=MEDIO (CF = 0.67)

Regla de Decisiones de la tercera corrida

1. (marguen = MEDIO) => utilidad=MEDIO (752.0/250.0)
2. (marguen = BAJO) and (cantidad = MEDIO) => utilidad=BAJO (2020.0/0.0)
3. (marguen = BAJO) and (cantidad = BAJO) => utilidad=BAJO (100.0/0.0)
4. (gastos = ALTO) => utilidad=BAJO (3.0/0.0)
5. (marguen = ALTO) => utilidad=ALTO (3080.0/80.0)

Unión de las reglas de decisiones

La primera regla para la tercera corrida y la quinta regla de la segunda corrida son similares por ello está dentro del árbol de decisiones a elaborar.

(marguen = MEDIO) => utilidad=MEDIO (752.0/250.0)

La segunda y tercera regla de la tercera corrida será unificadas a la cuarta regla de la segunda corrida ya que si el margen será bajo no importará las demás variables la utilidad será “BAJA”

Tercera Corrida

2. (marguen = BAJO) and (cantidad = MEDIO) => utilidad=BAJO (2020.0/0.0)

3. (marguen = BAJO) and (cantidad = BAJO) => utilidad=BAJO (100.0/0.0)

Segunda corrida

4. (marguen = BAJO) => utilidad=BAJO (CF = 1.0)

La primera y segunda regla de la segunda corrida es similar a la quinta regla de la tercera corrida ya que no importando las demás variables mientras el margen sea alto la utilidad será alta.

Segunda corrida

1. (marguen = ALTO) and (cantidad = MEDIO) => utilidad=ALTO (CF = 1.0)

2. (marguen = ALTO) and (gastos = BAJO) => utilidad=ALTO (CF = 0.97)

Tercera corrida

5. (marguen = ALTO) => utilidad=ALTO (3080.0/80.0)

Las reglas de decisiones restantes serán agregadas al prototipo en el orden de validación que se detallan en cada corrida.

3. (cantidad = ALTO) and (gastos = BAJO) => utilidad=ALTO (CF = 0.97)

4. (gastos = ALTO) => utilidad=BAJO (3.0/0.0)

Las reglas de decisiones que se estructuro son:

//(marguen = ALTO) and (cantidad = MEDIO) => utilidad=ALTO

```

if($marg_desc=="ALTO"&&$cant_desc=="MEDIO"){

    $util_desc="ALTO";

}else

//(marguen = ALTO) and (gastos = BAJO) => utilidad=ALTO
if($marg_desc=="ALTO"&&$gast_desc=="BAJO"){

    $util_desc="ALTO";

}else

//(cantidad = ALTO) and (gastos = BAJO) => utilidad=ALTO

if($cant_desc=="ALTO"&&$gast_desc=="BAJO"){

    $util_desc="ALTO";

}else

//(marguen = BAJO) => utilidad=BAJO

if($marg_desc=="BAJO"){

    $util_desc="BAJO";

}else

//(marguen = MEDIO) => utilidad=MEDIO

if($marg_desc=="MEDIO"){

    $util_desc="MEDIO";

}else

//(gastos = ALTO) => utilidad=BAJO

if($gast_desc=="ALTO"){

```

```
$util_desc="BAJO";

}else

//(marguen = ALTO) => utilidad=ALTO

if($marg_desc=="ALTO"){

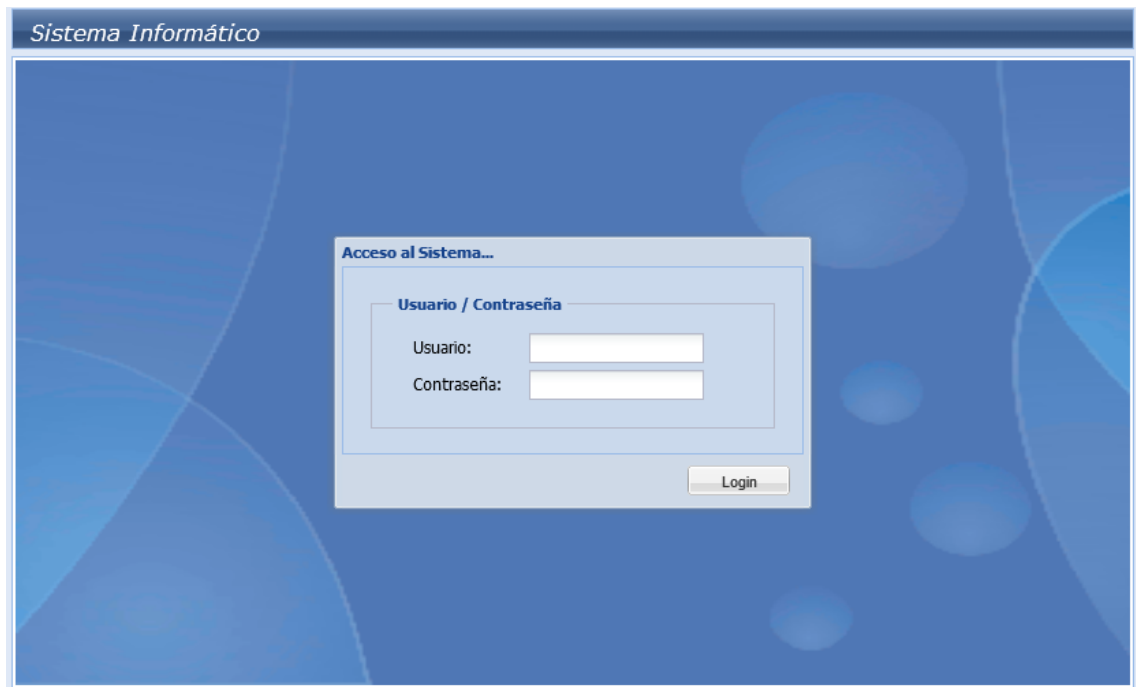
    $util_desc="ALTO";

}
```

Despliegue

Diseño del Prototipo

Desarrollado en html, CSS y php

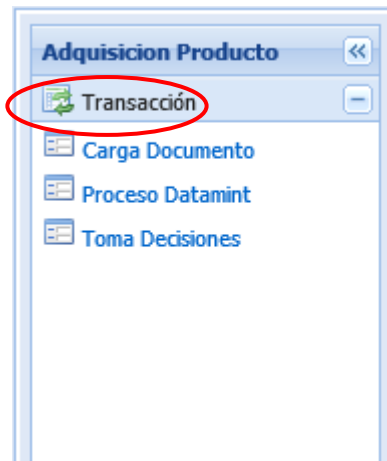


Ingreso al Sistema

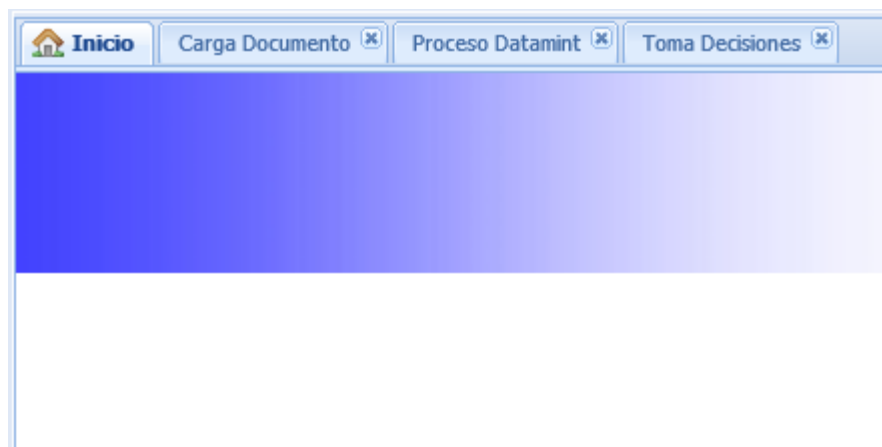
Una vez dentro del sistema, presentará una pantalla que contendrá el menú de opciones con las siguientes áreas:

- 1) **Superior.** Nombre de la aplicación, icono de la empresa.

- 2) **Lateral Izquierda.** Nombre del Módulo “Adquisición de Productos”, listado del menú, dividido en 3 paneles:
 - a. **Transacciones,** Realiza los procesos del módulo.



- 3) **Lateral derecha.** Pantalla principal, carga las formas seleccionadas en las opciones del menú, se solapará con Tabs que permitirán abrir múltiples ventanas a la vez.

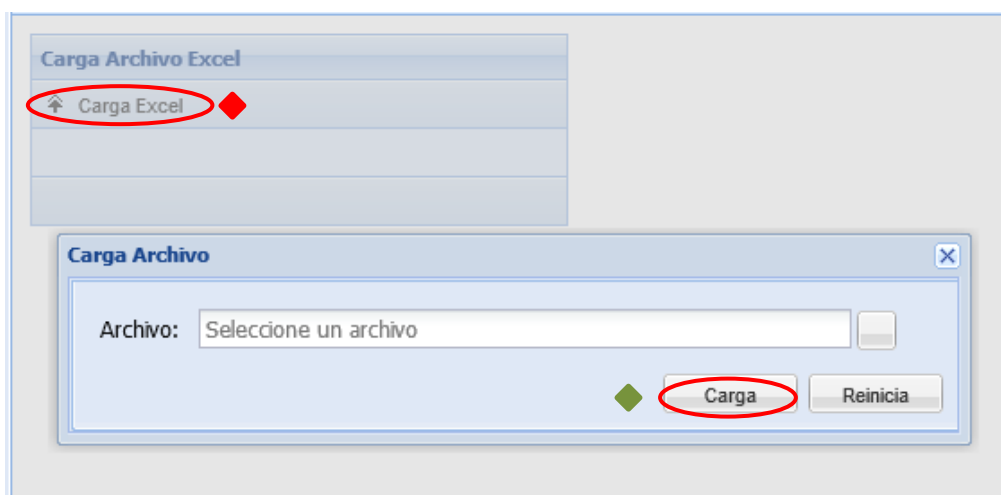


Formas del proyecto

Menú transacciones

- **Carga Archivo Excel:** Carga el archivo Excel que contiene información que la empresa Generó y que desea analizar con las reglas de decisiones

detalladas en la herramienta de minería de datos, esta información se almacena en el Datamining.



Nota: ♦ Botón para desplegar la opción de búsqueda del Archivo; ◆ Botón para Cargar el Archivo

- **Proceso Datamining.**-Verifica las reglas de decisiones que la herramienta de minería de datos detallo para así encontrar las que cumplan en un mayor porcentaje con los datos analizados y con ellas generar el prototipo de toma de decisiones.

Proceso Datamint

Nuevo **Procesa** ♦

Datos Resultado

No	Reglas	Registros	Verifica	Error
1	(marguen = ALTO) and (cantidad = MEDIO) => utilidad=ALTO	5675	5595	80
2	(marguen = ALTO) and (gastos = BAJO) => utilidad=ALTO	0	0	0
3	(cantidad = ALTO) and (gastos = BAJO) => utilidad=ALTO	5	5	0
4	(marguen = BAJO) => utilidad=BAJO	5600	5600	0
5	CANTIDAD REGISTROS	13080	99.39%	0.61%

◆

Nota: ♦ Procesa la información cargada en el Datamart para aplicarle las reglas de decisiones; ◆ Detalle del análisis de la información con las reglas de decisiones encontradas

- **Prototipo Toma de Decisiones:** Forma para analizar según las reglas de decisiones filtradas en los procesos anteriores cual será el resultado de la combinación las variables: cantidad de venta, gasto de almacenaje y porcentaje de utilidad. La utilidad que se tendrá con estas combinaciones. Estas variables tienen como formato:

Cantidad vendida en valor entero.

Gasto de almacenaje en valor flotante.

Porcentaje de Utilidad no mayor a 100%

Resultado Obtenido: mostrará un rango de utilidad en que se encontrará el valor esperado con la combinación de las variables de entrada

The screenshot shows a software interface titled "Prototipo Toma de Decisiones". At the top, there are two buttons: "Nuevo" and "Grabar". Below this is a section labeled "Datos" containing three input fields: "Cantidad Vendida:" with the placeholder "Cantidad Ven", "Gasto de Almacenaje:" with the placeholder "Gasto de Alm", and "Prctje de Utilidad:" with the placeholder "Margen de Ut". Below these fields is a larger text area labeled "Resultado:" which is currently empty.

CORRIDA DE VALORES CON EL PROTOTIPO

Primera corrida:

This screenshot shows the same interface as above, but with numerical values entered in the input fields. "Cantidad Vendida" is 21, "Gasto de Almacenaje" is 45, and "Prctje de Utilidad" is 50. The "Resultado" field now displays the text "Utilidad: ALTO, Rango: 17.00 - 80.00%".

Valores de Entrada

Cantidad Vendida: 21, valor Lingüístico "MEDIO".

Gato de Almacenaje: 45, valor Lingüístico "MEDIO".

Pctje Marguen: 50, valor Lingüístico "ALTO".

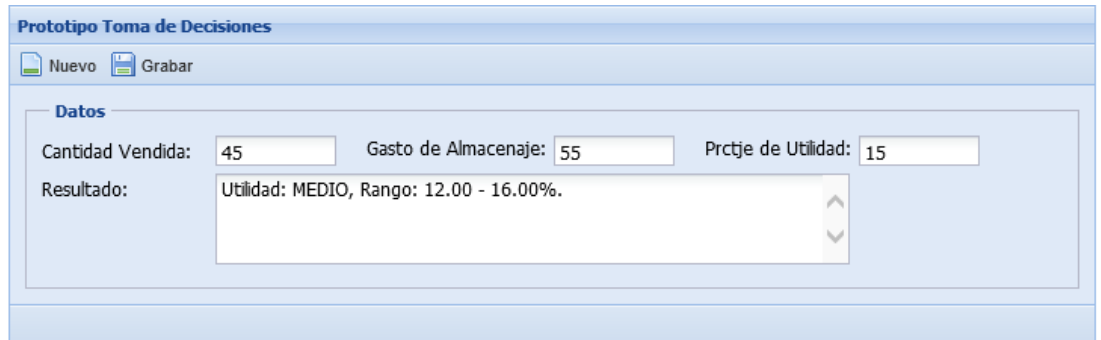
Resultado: ALTO, Rango 17.00-80.00%

Regla de decisiones

Los valores de entrada caen en la regla:

(marguen = ALTO) and (cantidad = MEDIO) => utilidad=ALTO

Segunda corrida:



The screenshot shows a software interface for decision support. At the top, it says 'Prototipo Toma de Decisiones'. Below that are two buttons: 'Nuevo' and 'Grabar'. Under the heading 'Datos', there are three input fields: 'Cantidad Vendida' with the value 45, 'Gasto de Almacenaje' with the value 55, and 'Prctje de Utilidad' with the value 15. Below these is a 'Resultado' field which contains the text 'Utilidad: MEDIO, Rango: 12.00 - 16.00%'.

Valores de Entrada

Cantidad Vendida: 45, valor Lingüístico “MEDIO”.

Gato de Almacenaje: 55, valor Lingüístico “MEDIO”.

Pctje Marguen: 15, valor Lingüístico “MEDIO”.

Resultado: MEDIO, Rango: 12.00 - 16.00%.

Regla de decisiones

Los valores de entrada caen en la regla:

(marguen = MEDIO) => utilidad=MEDIO

Tercera corrida:

Prototipo Toma de Decisiones

Nuevo Grabar

Datos

Cantidad Vendida: 55 Gasto de Almacenaje: 10 Prctje de Utilidad: 9

Resultado: Utilidad: ALTO, Rango: 17.00 - 80.00%.

Valores de Entrada

Cantidad Vendida: 55, valor Lingüístico “ALTO”.

Gato de Almacenaje: 10, valor Lingüístico “BAJO”.

Pctje Marguen: 15, valor Lingüístico “BAJO”.

Resultado: ALTO, Rango: 17.00 - 80.00%.

Regla de decisiones

Los valores de entrada caen en la regla:

(cantidad = ALTO) and (gastos = BAJO) => utilidad=ALTO

Cuarta corrida:

Prototipo Toma de Decisiones

Nuevo Grabar

Datos

Cantidad Vendida: 50 Gasto de Almacenaje: 12 Prctje de Utilidad: 8

Resultado: Utilidad: BAJO, Rango: 1.00 - 11.00%.

Valores de Entrada

Cantidad Vendida: 50, valor Lingüístico “MEDIO”.

Gato de Almacenaje: 12, valor Lingüístico “BAJO”.

Pctje Marguen: 8, valor Lingüístico “BAJO”.

Resultado: BAJO, Rango: 1.00 - 11.00%.

Regla de decisiones

Los valores de entrada caen en la regla:

(marguen = BAJO) => utilidad=BAJO (CF = 1.0)

CONCLUSIONES Y RECOMENDACIONES

CONCLUSIONES

- La información histórica de la empresa FRUDONMAR analizada en el presente proyecto es completa y adecuada y cumple con los indicadores financieros que la empresa necesita para obtener el correcto resultado para la toma de decisiones,
- Se establecieron posibles problemas que puedan impactar sobre la toma de decisiones por parte del área de importaciones.
- Los usuarios pueden evaluar resultados obtenidos por medio del prototipo determinando cada uno de los valores dentro de las variables establecidas para de esta manera tener una idea clara para la toma de decisiones.
- Por medio del uso de Indicadores financieros se pudo obtener la información que luego debió ser analizada en el desarrollo de este estudio.

RECOMENDACIONES

- El uso de datos históricos debe darse con un periodo mayor a 10 años ya que con esto se puede tener una muestra adecuada que permite analizar todo tipo de posibles eventos.
- Es necesario almacenar el resultado de la consolidación de los datos históricos en un Datamart para de esta forma realizar un correcto análisis para futuras decisiones.
- Con el uso del Prototipo propuesto luego del análisis de Datamart con técnicas Dataminig se logrará reducir el trabajo manual y el tiempo que éste toma en realizarse, para la obtención de información relevante para la toma de decisiones.
- Evaluar contantemente los indicadores financieros planteados en este proyecto para tomar decisiones oportunas.

BIBLIOGRAFÍA

- Aguilar Santamaría, P. A. (2012). Un modelo de clasificación de inventarios para incrementar el nivel de servicio al cliente y la rentabilidad de la empresa. *Pensamiento & Gestión*, (32), 142-164. Recuperado a partir de http://www.scielo.org.co/scielo.php?script=sci_abstract&pid=S1657-62762012000100007&lng=en&nrm=iso&tlng=es
- Bembibre, V. (2009, febrero 25). Definicion Datawarehouse. Recuperado 21 de noviembre de 2017, a partir de <https://www.definicionabc.com/tecnologia/datawarehouse.php>
- Berzal, F. (2006). Introduction to data mining. DECSAI Departamento de Ciencias de la Computación e I.A., Universidad de Granada. Recuperado a partir de <http://elvex.ugr.es/decsai/intelligent/slides/dm/D1%20Data%20Mining.pdf>
- Bofill Placeres, A., Sablón Cossío, N., & Florido García, R. (2017). Procedimiento para la gestión de inventario en el almacén central de una cadena comercial cubana. *Revista Universidad y Sociedad*, 9(1), 41-51. Recuperado a partir de http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2218-36202017000100006&lng=es&nrm=iso&tlng=es
- Causado Rodríguez, E. (2015). Inventory control model for economic order in food marketer. *Revista Ingenierías Universidad de Medellín*, 14(27), 163–177.
- Congreso Nacional. Ley de Propiedad Intelectual, Pub. L. No. Registro Oficial Suplemento 426, Codificación N° 2006-013 77 (2006). Recuperado a partir de http://www.wipo.int/wipolex/es/text.jsp?file_id=444010

Cultura CRM. (2016a, marzo). Técnicas de aplicación de Data Mining. Recuperado 22 de febrero de 2018, a partir de <http://culturacrm.com/data-mining/tecnicas-aplicacion-data-mining/>

Cultura CRM. (2016b, julio 6). Cinco herramientas de Data Mining o minería de datos. Recuperado 4 de diciembre de 2017, a partir de <http://culturacrm.com/data-mining/cinco-herramientas-data-mining/>

Dip, P. (2008, abril 20). Tecnología e Informática: DATOS Y DIGITALIZACIÓN. Recuperado 20 de febrero de 2018, a partir de <http://latecnologiavirtual.blogspot.com/2008/04/datos-y-digitalizacin.html>

Evaluando Software. (2016a, julio). Técnicas de Data Mining. Recuperado 22 de febrero de 2018, a partir de <http://www.evaluandosoftware.com/tecnicas-data-mining/>

Evaluando Software. (2016b, septiembre). ABC del data mart. Recuperado 22 de febrero de 2018, a partir de <http://www.evaluandosoftware.com/abc-del-data-mart/>

Formatalent. (2017, mayo). Formación en Oracle: Historia y Características. Recuperado 20 de febrero de 2018, a partir de <http://formatalent.com/formacion-en-oracle-historia-y-caracteristicas/>

Giménez, J. M. (2017, mayo). Datamart. Recuperado 22 de febrero de 2018, a partir de <https://cursoerpcrmblog.wordpress.com/2017/05/09/datamart/>

Guevara Vega, C. P. (2015). Desarrollo de una plataforma de business intelligence para facilitar el análisis de datos de las competencias generales de formación

aplicadas en el desempeño laboral de los egresados de la Universidad Técnica del Norte. Recuperado a partir de <http://repositorio.espe.edu.ec/jspui/handle/21000/10160>

Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación*. México, D.F.: McGraw-Hill Education.

Lucidchart. (2017, marzo 25). Qué es un diagrama de árbol de decisión. Recuperado 30 de enero de 2018, a partir de <https://www.lucidchart.com/pages/es/qu%C3%A9-es-un-diagrama-de-%C3%A1rbol-de-decisi%C3%B3n>

Microsoft. (2016). SQL Server 2016 | Microsoft. Recuperado 26 de febrero de 2018, a partir de <https://www.microsoft.com/es-es/sql-server/sql-server-2016>

Miranda Cairo, M., Valdés Puga, O., Pérez Mallea, I., Portelles Cobas, R., & Sánchez Zequeira, R. (2016). Metodología para la Implementación de la Gestión Automatizada de Controles de Seguridad Informática. *Revista Cubana de Ciencias Informáticas*, 10(2), 14-26. Recuperado a partir de http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2227-18992016000200002&lng=es&nrm=iso&tlng=es

Navarro Arango, R. (2015, agosto). Data Warehouse, Data mart y Data Mining. Recuperado 22 de febrero de 2018, a partir de <https://dabacodlab.wordpress.com/2015/08/20/data-warehouse-data-mart-y-data-mining/>

- Pérez López, C. (2007). *Minería de datos: técnicas y herramientas*. Editorial Paraninfo.
- Pérez-Vergara, I., Cifuentes-Laguna, A. M., Vásquez-García, C., & Marcela-Ocampo, D. (2013). Un modelo de gestión de inventarios para una empresa de productos alimenticios. *Ingeniería Industrial*, 34(2), 227-236. Recuperado a partir de http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S1815-59362013000200011&lng=es&nrm=iso&tlng=es
- Pratt, J. P. (2014). *Concepts of database management* (Eighth edition). Boston, MA: Cengage Learning.
- Ramos-Maldonado, M., Maness, T., & Salinas-Sandoval, D. (2015). Modelo de un sistema multi-agente para la optimización de la cadena de suministros en la industria de la madera de coníferas. *Maderas. Ciencia y tecnología*, 17(3), 613-624. <https://doi.org/10.4067/S0718-221X2015005000054>
- Riquelme, M. (2013, julio). ¿Que es un Data Mart? Recuperado 22 de febrero de 2018, a partir de <https://www.webyempresas.com/que-es-un-data-mart/>
- Rouse, M. (2015). ¿Qué es MySQL? - Definición en WhatIs.com. Recuperado 4 de diciembre de 2017, a partir de <http://searchdatacenter.techtarget.com/es/definicion/MySQL>
- Salas-Navarro, K., Maignel-Mejía, H., & Acevedo-Chedid, J. (2017). Metodología de Gestión de Inventarios para determinar los niveles de integración y colaboración en una cadena de suministro. *Ingeniare. Revista chilena de*

ingeniería, 25(2), 326-337. <https://doi.org/10.4067/S0718-33052017000200326>

Sinnexus. (2016a). Bases de datos OLTP y OLAP. Recuperado 22 de febrero de 2018, a partir de http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx

Sinnexus. (2016b). Datamart. Recuperado 21 de noviembre de 2017, a partir de http://www.sinnexus.com/business_intelligence/datamart.aspx

Sinnexus. (2016c). Datawarehouse. Recuperado 21 de noviembre de 2017, a partir de http://www.sinnexus.com/business_intelligence/datawarehouse.aspx

Sinnexus. (2016d). Datos, información, conocimiento. Recuperado 20 de febrero de 2018, a partir de http://www.sinnexus.com/business_intelligence/piramide_negocio.aspx

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining* (1st ed). Boston: Pearson Addison Wesley.

TechTerms. (2017, diciembre). RDBMS (Relational Database Management System) Definition. Recuperado 26 de febrero de 2018, a partir de <https://techterms.com/definition/rdbms>

Universidad de Valencia. (s. f.). 2.4 Regla de decisión. Recuperado 30 de enero de 2018, a partir de https://www.uv.es/webgid/Inferencial/24_regla_de_decisin.html



**Presidencia
de la República
del Ecuador**



**Plan Nacional
de Ciencia, Tecnología,
Innovación y Saberes**



SENESCYT

Secretaría Nacional de Educación Superior,
Ciencia, Tecnología e Innovación

DECLARACIÓN Y AUTORIZACIÓN

Yo, **Garrido García, Fausto Javier**, con CC # **0918748625**, autor del trabajo de titulación **Elaboración de un Prototipo Predictivo para la Toma de Decisiones de la Gestión de Compras del Área de Importaciones con la Generación de un Datamart y Técnicas de Datamining para la Empresa Frutera Don Marcos S.A. (FRUDONMAR)**, previo a la obtención del título de **Ingeniero en Sistemas Computacionales** en la Universidad Católica de Santiago de Guayaquil.

1.- Declaro tener pleno conocimiento de la obligación que tienen las instituciones de educación superior, de conformidad con el Artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de titulación para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.

2.- Autorizo a la SENESCYT a tener una copia del referido trabajo de titulación, con el propósito de generar un repositorio que democratice la información, respetando las políticas de propiedad intelectual vigentes.

Guayaquil, **05 de marzo de 2018**

Garrido García, Fausto Javier

C.C: 0918748625

REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA

FICHA DE REGISTRO DE TESIS/TRABAJO DE TITULACIÓN

TEMA Y SUBTEMA	Elaboración de un Prototipo Predictivo para la Toma de Decisiones de la Gestión de Compras del Área de Importaciones con la Generación de un Datamart y Técnicas de Datamining para la Empresa Frutera Don Marcos S.A. (FRUDONMAR)		
AUTOR	Fausto Javier Garrido García		
TUTORA	Adela Zurita Fabre		
INSTITUCIÓN	Universidad Católica de Santiago de Guayaquil		
FACULTAD	Facultad de Ingeniería		
CARRERA	Ingeniería en Sistemas Computacionales		
TITULO OBTENIDO	Ingeniero en Sistemas Computacionales		
FECHA DE PUBLICACIÓN	08 de marzo de 2018	No. DE PÁGINAS:	95
ÁREAS TEMÁTICAS	Bases de datos, minería de datos, sistemas de inferencia		
PALABRAS CLAVES/KEYWORDS	Prototipo toma de decisiones, herramienta Weka, algoritmos de clasificación, datawarehouse, datamart, datamining		
RESUMEN			
<p>El presente trabajo de titulación está basado en el desarrollo de un prototipo para la toma de decisiones del área de importaciones con técnicas de Datamining. Luego del análisis correspondiente de las diferentes bases de datos y sus bondades en cuanto a pertinencia y utilidad, se decidió que para el análisis de la información se utilice Weka, una herramienta de software libre, que permitirá realizar la generación de reglas de decisiones, utilizando algoritmos de selección con tratamiento de reglas difusas. Aplicadas las pruebas necesarias, se determinó que, en un mercado tan competitivo como el actual, contar con herramientas que provean información correctamente y en el momento oportuno, facilita las gestiones para tomar la mejor decisión. Este proyecto de titulación está enfocado en cubrir esa necesidad en la empresa objeto de este estudio, brindando un análisis histórico con grandes cantidades de información, permitiendo elaborar y desarrollar un prototipo que pueda cumplir con esta demanda.</p>			
ADJUNTO PDF	<input checked="" type="checkbox"/> SI	<input type="checkbox"/> NO	
CONTACTO CON AUTOR	Teléfono: +593-939799995	E-mail: xaviergarrido@hotmail.com	
CONTACTO CON LA INSTITUCIÓN (COORDINADOR DEL PROCESO UTE)	Nombre: Yanza Montalván, Ángela Olivia		
	Teléfono: +593-4-0983035702		
	E-mail: angela.yanza@cu.ucsg.edu.ec		
SECCIÓN PARA USO DE BIBLIOTECA			
Nº. DE REGISTRO (en base a datos)			
Nº. DE CLASIFICACIÓN			
DIRECCIÓN URL (tesis en la web)			