



FACULTAD DE INGENIERÍA

Carrera de Ingeniería Informática y de Sistemas

**“APLICACIÓN DE MINERÍA DE DATOS PARA
DETERMINAR PATRONES DE CONSUMO FUTURO
EN CLIENTES DE UNA DISTRIBUIDORA DE
SUPLEMENTOS NUTRICIONALES”**

**Tesis para optar el Título Profesional de Ingeniero
Informático y de sistemas**

MIGUEL ANGEL GRÁNDEZ MÁRQUEZ

Asesores:

Ing. Carlos Takano Abratani

Ing. Jonathan Chavez Espinoza

Lima- Perú 2017

Dedicatoria

Dedico esta tesis a mis Padres, quienes son siempre apoyo en todas las etapas de mi vida y constantemente me impulsan en los diferentes objetivos trazados.

Agradecimientos

Mis sinceros agradecimientos a quien en vida fue, Ing. Carlos Takano Abratani ,mi primer Asesor de Tesis, así como también quiero agradecer al Ing. Jonathan Chávez Espinoza ,quien accedió en asesorarme luego de este hecho fortuito.

Dar las gracias a ambos por los consejos brindados y por la dirección en la conclusión de la presente tesis.

Resumen

El presente proyecto de tesis busca encontrar reglas que determinen el patrón de consumo de clientes en una distribuidora de suplementos nutricionales, se aplica técnicas de Minería de datos a través del software Visual Studio 2015 que contiene la extensión Service Analysis y como motor de Base de datos hacemos uso de SQL server 2014.

Abstract

The present thesis project seeks to find rules that determine the pattern of consumption on customers in a nutritional supplements distributor, in this project we apply data mining techniques through Visual Studio 2015 software that contains the Service Analysis extension and as a database engine we use SQL Server 2014.

INDICE DE CONTENIDOS

PROBLEMA DE INVESTIGACIÓN	7
PLANTEAMIENTO DEL PROBLEMA	8
FORMULACIÓN DEL PROBLEMA	8
JUSTIFICACIÓN DE LA INVESTIGACIÓN	8
MARCO REFERENCIAL	11
ANTECEDENTES	11
MARCO TEÓRICO	12
METODOLOGÍAS EN LA APLICACIÓN DE MINERÍA DE DATOS	12
EL PROCESO KDD	12
EL PROCESO SEMMA	13
EL PROCESO CRISP-DM	13
COMPARATIVO ENTRE METODOLOGÍAS	15
KDD Y SEMMA	15
METODOLOGÍA DE MINERÍA DE DATOS EMPLEADA	18
ALGORITMO DE ASOCIACIÓN DE MICROSOFT	25
ALGORITMO DE CLÚSTERES DE MICROSOFT	26
ALGORITMO DE RED NEURONAL DE MICROSOFT	26
COSTOS DEL PROYECTO DE MINERÍA DE DATOS	28
BENEFICIOS	32
OBJETIVOS E HIPÓTESIS	36
MÉTODO	37
TIPO Y DISEÑO DE INVESTIGACIÓN	37
TIPO DE INVESTIGACIÓN	37
DISEÑO DE INVESTIGACIÓN	37
VARIABLES	38
MUESTRA	38
INSTRUMENTOS DE INVESTIGACIÓN	40
PROCEDIMIENTOS DE RECOLECCIÓN DE DATOS	40
PROCEDIMIENTO EN EL DESARROLLO DE LA TESIS	41

<i>RESULTADOS</i>	51
<i>DISCUSIÓN</i>	74
<i>CONCLUSIONES</i>	76
<i>RECOMENDACIONES</i>	78
<i>REFERENCIAS</i>	79
<i>ANEXOS</i>	82

INDICE DE TABLAS

TABLA 1 RESULTADOS ESTUDIO ESPAÑA	5
TABLA 2 EDAD, ALTURA, PESO Y FRECUENCIA DE PRÁCTICA DEPORTIVA EN GIMNASIOS DE LA MUESTRA	6
TABLA 3 EVALUACION DE METODOLOGÍA (ELABORACIÓN PROPIA)	17
TABLA 4 LEYENDA PUNTAJE SW	19
TABLA 5 DESCRIPCION DE HERRAMIENTAS COMPARADAS ²³	19
TABLA 6 IBM SPSS MODELER	22
TABLA 7 ORACLE DATA MINING ODM	22
TABLA 8 MICROSOFT SQL SERVER INTEGRATION SERVICES	23
TABLA 9 LIONOSO	23
TABLA 10ADVANCEDMINER	24
TABLA 11PRECIOS DE LICENCIAMIENTO	24
TABLA 12CARACTERÍSTICAS ADICIONALES	25
ETAPA 1: TABLA 13 COSTOS COMPRENSIÓN DEL NEGOCIO.	28
ETAPA 2. TABLA 14 COMPRENSIÓN DE LOS DATOS	29
ETAPA 3. TABLA 15 PREPARACIÓN DE DATOS	29
ETAPA 4. TABLA 16 MODELADO	29
ETAPA 5. TABLA 17 EVALUACIÓN	30
ETAPA 6. TABLA 18 IMPLEMENTACIÓN	30
TABLA 19 COSTOS PLANEADOS EN SERVICIOS	30
TABLA 20 COSTOS PLANEADOS EN LICENCIAS	31
TABLA 21 COSTOS TOTALES.	31
TABLA 22 BENEFICIOS.	32
TABLA 23 PATRÓN DE CONSUMIDORES DE HARBINGER FITNESS® CORREA BIG	51
TABLA 24 PATRÓN DE CONSUMIDORES DE SOLGAR® GLUCOSAMINE	51
TABLA 25 PATRÓN DE CONSUMIDORES DE NUTREX® LIPO-6 CLA	54
TABLA 26 PATRÓN DE CONSUMIDORES DE SYNTRAX® NECTAR MEDICAL	55
TABLA 27 PATRÓN DE CONSUMIDORES DE CELLUCOR® WHEY COR PERFORMANCE	56
TABLA 28 PATRÓN DE CONSUMIDORES DE SMART SHAKE® SMART SHAKE	57
TABLA 29 PATRÓN DE CONSUMIDORES DE MUSCLETECH® PLATINUM 100% WHEY	57
TABLA 30 PATRÓN DE CONSUMIDORES DE MUSCLETECH® CLEAR MUSCLE	57
TABLA 31 PATRÓN DE CONSUMIDORES DE MET RX® LIQUID L-CARNITINE 1500	58
TABLA 32 PATRÓN DE CONSUMIDORES DE HARBINGER FITNESS® GUANTES POWER	58
TABLA 33 PATRÓN DE CONSUMIDORES DE MET RX® ULTRAMYOSYN WHEY	58
TABLA 34 RESULTADOS MODELO NEURONAL SYNTRAX® MATRIX	58
TABLA 35 RESULTADOS MODELO NEURONAL MUSCLETECH® HYDROXYCUT PRO	59
TABLA 36 RESULTADOS MODELO NEURONAL MUSCLETECH® PLATINUM 100% WHEY	60
TABLA 37 RESULTADOS MODELO NEURONAL NUTREX® LIPO-6 BLACK	60
TABLA 38 RESULTADOS MODELO CLUSTER	62

INDICE DE ILUSTRACIONES

<i>Ilustración 1 Comparativo Perú y Colombia</i>	3
<i>Ilustración 2 Comparativo Perú y España</i>	3
<i>Ilustración 3 Etapas en el procedimiento KDD</i>	12
<i>Ilustración 4 Esquema CRISP</i>	14
<i>Ilustración 5 Preferencia de metodología</i>	17
<i>Ilustración 6 Arquitectura Data Warehouse en 3 capas</i>	27
<i>Ilustración 7Arquitectura de un sistema típico de minería de datos</i>	28
<i>Ilustración 8 Calculadora Estadística</i>	39
<i>Ilustración 9 SQL Server Management Studio</i>	41
<i>Ilustración 10 Tiendas</i>	42
<i>Ilustración 11: Clientes</i>	42
<i>Ilustración 12 Ventas</i>	43
<i>Ilustración 13: Productos</i>	43
<i>Ilustración 14:Categoría</i>	44
<i>Ilustración 15: Actividad Física</i>	45
<i>Ilustración 16: Objetivos</i>	45
<i>Ilustración 17: Visual Studio</i>	45
<i>Ilustración 18: Conexión a DB</i>	46
<i>Ilustración 19: Modelo Minería de Datos</i>	46
<i>Ilustración 20: Algoritmos de la Herramienta.</i>	47
<i>Ilustración 21:Algoritmo Asociación</i>	47
<i>Ilustración 22:ITEM Sets</i>	48
<i>Ilustración 23: Red Neuronal</i>	48
<i>Ilustración 24: Diagrama de Cluster</i>	49
<i>Ilustración 25:Perfil de cluster</i>	49
<i>Ilustración 26: Características de Clúster</i>	50
<i>Ilustración 27 Resultados Algoritmo Clúster</i>	61
<i>Ilustración 28 Distribución de productos comercializados</i>	62
<i>Ilustración 29 Producto Syntrax Nectar</i>	64
<i>Ilustración 30Producto Mutant BCAA</i>	64
<i>Ilustración 31Producto Solgar Vitamin C</i>	64
<i>Ilustración32ProductoGakicV02</i>	64
<i>Ilustración 33 Nutrexliipo-6 cla</i>	65
<i>Ilustración 34 Muscletech Platinum 100% iso Whey</i>	65
<i>Ilustración 35 Syntrax Essence</i>	65
<i>Ilustración 36 Muscletech Phase 8</i>	65
<i>Ilustración 37 Solgar Glucosamine Chondroitin Complex</i>	66
<i>Ilustración 38Correa wrist support</i>	66
<i>ilustración 39 Correa lifting hook</i>	66
<i>ilustración 40 Correa wrist support</i>	66
<i>Ilustración 41 Guantes</i>	67
<i>Ilustración 42 Disney ® gummies disney princess</i>	68
<i>ilustración 43 Solgar ® calcium magnesium plus zinc</i>	68
<i>Ilustración 44 Fitmiss Burn</i>	69
<i>Ilustración 45 Guantes pro women</i>	69
<i>Ilustración 46 Met rx active women</i>	69

<i>ilustración 47 Guantes power women</i>	69
<i>ilustración 48 Hydroxycut hardcore elite caps</i>	70
<i>ilustración 49 Hydroxycut sx-7</i>	70
<i>ilustración 50 Hydroxycut hardcore next gen</i>	70
<i>ilustración 51 Gummies disney princess</i>	70
<i>Ilustración 52Kangavites chewable tablets ilustración 53 gummies disney car</i>	70
<i>ilustración 54 Arnold iron pack</i>	71
<i>ilustración 55 Armor-v</i>	71
<i>ilustración 56 Platinum multi vitamin</i>	72
<i>ilustración 57 Hydroxycut hardcore elite caps</i>	72
<i>ilustración 58 Hydroxycut sx-7</i>	72
<i>ilustración 59 Hydroxycut hardcore next gen</i>	73
<i>ilustración 60 Mutant BCAA</i>	73
<i>ilustración 61 Modern BCAA</i>	73
<i>Ilustración 62 Amino build next gen</i>	73
<i>Ilustración 63Diagrama de Base de Datos</i>	82

INTRODUCCIÓN

Granular la información nos permite conocer aspectos no analizados que pueden incidir en el desempeño de una organización.

Explorar los registros contenidos en la base de datos de la empresa con técnicas de minería de datos, nos facilita comprender reglas que determinan patrones de consumo y tendencias que siguen los datos, lo que a su vez genera un conocimiento; el cual nos permitirá asociar productos que tengan mayor rotación con aquellos que no la tienen, generando de esta forma venta cruzada que resulte en el incremento de ingresos.

La presente tesis se centra en aplicar Minería de Datos en un negocio de venta de suplementos nutricionales (SN), dado el crecimiento explosivo en la distribución de estos productos en Lima y que según investigaciones¹ en el año 2011 cada persona destinó 180 soles mensuales en suplementos y que para el 2016 se destinó hasta 270 soles mensuales en su consumo.

La distribuidora de suplementos nutricionales en la cual se enfoca la tesis es Lab Nutrition y nos centramos en la tienda Pardo Miraflores ,ya que es una de las tiendas que mejor desempeño presenta, se localiza en Av. José Pardo 370 Miraflores Lima- Perú , en esta tienda encontramos los datos enriquecidos ,ya que, en dicho local se realizan con regularidad (2 veces por mes) consultas nutricionales a cargo de profesionales en el área, quienes al momento de conocer las necesidades del cliente emiten su recomendación ; como resultado de la consulta se logra conocer aspectos adicionales a los recabados en la venta ,como por ejemplo datos antropométricos principales: Talla, peso así como también datos ocupacionales del cliente tales como: la profesión u ocupación, por lo cual la data ,correspondiente a los clientes seleccionados para el estudio, es de gran valor para el análisis.

La tienda comercializa 137 productos entre SN y accesorios (Guantes, Sogas ,Shakers), dichos productos se encuentran clasificados en categorías y sub categorías.

La tienda desea conocer reglas, las cuales determinen el patrón de consumo de sus clientes, además se busca generar paquetes de productos que puedan venderse de forma conjunta. (De acuerdo a las reglas formadas, y la función nutricional y/o características del producto), aumentando de esta forma su accionar en Cross selling; es en este escenario donde el análisis de la Base de Datos es preciso para lograr conocer mejor a sus clientes y sobre los resultados poder aplicar políticas adecuadas.

Se considera política adecuada todo accionar que logre incrementar el nivel de ventas de la tienda.

La alta dirección busca incrementar el nivel de ventas que presenta en la actualidad, puesto que ha observado que existen productos que tienen una rotación más alta que otros, por lo cual, busca incentivar el consumo de los productos menos demandados creando paquetes de consumo, la herramienta a aplicar es minería de datos: para analizar las reglas que determinan el consumo y de esta forma conocer el comportamiento de compra de sus clientes, con lo cual se pueda realizar ofrecimientos personalizados y crear paquetes de productos que conlleven a elevar el nivel de ventas.

El encargado de marketing, además espera conocer la información para sobre la base de los resultados, poder aplicar campañas dirigidas a clientes que sigan el mismo comportamiento; creando de esta forma acciones más eficaces.

Diariamente se pueden llegar a vender, en promedio, 100 productos de las diferentes categorías disponibles, tanto a clientes antiguos, como a clientes nuevos, esto hace crecer el nivel de la información de forma considerable, la empresa cuenta con un sistema de ventas y la tienda administra su información en una base de datos relacional.

Para poder conocer de forma adecuada las características de nuestros clientes, usamos como referencia estudios de este sector, de esta forma tenemos más datos de nuestro público objetivo, es decir, los grupos que conforman nuestra población de consumo.

Se ha tomado en consideración dos estudios: uno realizado en ²Colombia, país latinoamericano, hispanoparlante, con economía similar a la del Perú y un segundo estudio realizado en ³España, país hispanoparlante con el cual nos une la historia, que complementa con datos estadísticos los rasgos del consumidor de SN.

Perú			
País	Subcontinente	Continente	Mundo
			
• Capital: Lima • Población: 31.148.000 • Superficie: 1.285.220 km2 • Moneda: Soles (1 EUR=3,6642 PEN) • Pertenece a: Alianza del Pacífico, CAN, FMI, OEA, ONU, UNASUR			

Colombia			
País	Subcontinente	Continente	Mundo
			
• Capital: Bogotá • Población: 48.203.000 • Superficie: 1.141.749 km2 • Moneda: Pesos colombianos (1 EUR=3.206,5827 COP) • Pertenece a: Alianza del Pacífico, CAN, FMI, OEA, ONU, UNASUR			

Ilustración 1 Comparativo Perú y Colombia

Tomado el 25/10/16 de: <http://www.datosmacro.com/paises/comparar/peru/colombia>

Perú			
País	Subcontinente	Continente	Mundo
			
• Capital: Lima • Población: 31.148.000 • Superficie: 1.285.220 km2 • Moneda: Soles (1 EUR=3,6642 PEN) • Pertenece a: Alianza del Pacífico, CAN, FMI, OEA, ONU, UNASUR			

España			
País	Subcontinente	Continente	Mundo
			
• Capital: Madrid • Población: 46.524.943 • Superficie: 505.940 km2 • Moneda: Euros • Pertenece a: CoE, Zona Euro, EEE, FMI, OTAN, OCDE, ONU, OSCE, UE			

Ilustración 2 Comparativo Perú y España

Tomado el 25/10/16 de: <http://www.datosmacro.com/paises/comparar/peru/espana>

En el estudio Colombiano ² se encuentra lo siguiente:

De una muestra de 321 usuarios de entre 13 y 78 años pertenecientes a sectores económicos alto (55,5%), nivel medio (39%) y bajo (5%). Cuya distribución entre hombres y mujeres es 43,3% y 56,7% respectivamente.

Se conoce que el consumo se ve determinado por las siguientes variables:

- Edad.
- El sexo.
- La raza.
- El nivel de educación.
- El nivel de actividad física.
- El nivel económico de los usuarios.

Por ocupación del consumidor:

- 37,4% está conformado por estudiantes.
- 36,8% Empleados.
- 18,1% actividad económica independiente.
- 7,7% amas de casa o desempleados.

Las razones más escogidas para justificar el consumo de Suplementos Nutricionales fueron:

- Mejorar el aspecto físico.
- Cuidar la salud.
- Aumentar el rendimiento deportivo.
- Paliar algún déficit de la dieta.

En cuanto a la influencia para iniciar el consumo sabemos que las principales fuentes de influencia son:

- Familia.
- Amigos.
- Auto prescripción.
- Personal de salud.

El lugar donde más acuden los encuestados a comprar los Suplementos Nutricionales son:

- El gimnasio.
- Las tiendas especializadas.
- Las herboristerías.

En un segundo estudio realizado en **España**³ donde participaron 415 usuarios (260 varones y 155 mujeres) de 4 gimnasios de Sevilla se consiguieron los siguientes resultados:

Tabla 1 RESULTADOS ESTUDIO ESPAÑA

Resultados Obtenidos (Sexo, Ocupación , Nivel de estudios de la muestra)			
Frecuencia			%
Sexo	Varón	260	62,7
	Mujer	155	37,3
Ocupación	Estudiando	236	56,9
	Trabajando	140	33,7
	Ninguna	39	9,4
Nivel de Estudios	Graduado	63	15,2
	Bachiller	129	31,1
	Medios	127	30,6
	Superiores	96	23,1

El 48% de los encuestados acude al gimnasio por las siguientes razones:

- Estar en forma. (69.9%)
- Hacer ejercicio (19%).
- Salud (11,1%)

En varones:

- Por estar en forma
- Por hacer ejercicio
- Para sentirse bien

Mujeres:

- Estar en forma
- Por salud
- Por hacer Ejercicio.

Tabla 2 Edad, altura, peso y frecuencia de práctica deportiva en gimnasios de la muestra

		Total	Varones	Mujeres
Edad(años)	Media	31,63	29,95	34,43
	DS	11,15	10,22	12,06
	Mínimo	17	18	17
	Máximo	68	68	65
Altura (m)	Media	1,72	1,77	1,64
	DS	0,08	0,06	0,06
	Mínimo	1,50	1,58	1,50
	Máximo	1,96	1,96	1,78
Peso (Kg)	Media	73,38	79,15	63,69
	DS	11,26	7,41	9,90
	Mínimo	46,00	60	46
	Máximo	112,00	112	110
Años de Práctica	Media	3,01	3,13	2,82
	DS	2,89	2,82	3,01
	Mínimo	0,08	0,1	0,8
	Máximo	16,00	16	16
Días de Práctica Semanales	Media	3,64	3,78	3,41
	DS	1,02	1,02	0,97
	Mínimo	1	1	2
	Máximo	6	6	6
Tiempo al día de practica(Min)	Media	87,41	92,17	79,42
	DS	27,14	25,02	28,71
	Mínimo	30	45	17
	Máximo	240	210	240

Suplementos consumidos por varones:

- Proteína (42,7%),
- Creatina (26,9%)
- Bebida Deportiva (22,7%).

Suplementos consumidos por Mujeres:

- Diuréticos (29,7%)
- Complejos Vitamínicos (18,7%)
- Chitosan (14,2%).

Las razones más escogidas para justificar el consumo de SN fueron:

- Mejorar el aspecto físico.
- Cuidar la salud.
- Aumentar el rendimiento deportivo.
- Paliar algún déficit de la dieta.

Los estudios anteriormente citados nos sirven para entender mejor a nuestros clientes, entender qué razones persiguen los consumidores, así como conocer preferencias de SN tanto en varones como en mujeres y de esta forma realizar ofrecimientos sobre esta base. Por otro lado nos es de utilidad para seleccionar nuestra muestra, ya que, conocemos el grupo etario de los consumidores, datos antropométricos principales y las características del cliente para usarlas como variables de análisis (Edad, Sexo, Ocupación).

PROBLEMA DE INVESTIGACIÓN.

- No se han analizado los datos para de esta forma poder medir tendencias de compra y predecir qué características de los clientes determinan la compra de un producto.
- Se desconoce qué características tienen los consumidores de un determinado producto con una determinada probabilidad.
- No se tiene establecido campañas a grupos de clientes que siguen comportamientos similares.
- Se desconoce que variables del cliente determinan la compra de suplementos.

PLANTEAMIENTO DEL PROBLEMA.

El no analizar la data de los clientes de una distribuidora de suplementos nutricionales le representa costos a la empresa (costos de oportunidad), al no poder ofrecer productos complementarios a los usuarios de acuerdo a su histórico de compra , impide realizar la tarea de forma efectiva , ya que, no se puede conocer que reglas determinan el consumo de los clientes , clasificar a los clientes por SN consumido , aprovisionarse de manera adecuada y establecer campañas dirigidas a grupos que siguen comportamientos similares.

Aplicar minería de datos en esta industria que mueve miles de soles mensuales es una herramienta muy importante para evitar perder ventas y maximizar ingresos, al saber puntualmente las necesidades de cada cliente y al conocer qué aspectos del plan nutricional son más importantes para ellos tendremos un horizonte más claro.

Sabemos que cada uno de los clientes presenta necesidades diferentes y es muy importante ofrecer promociones sectorizadas a grupos de clientes que siguen las mismas reglas de consumo y por lo tanto presentan patrones de comportamiento y consumo similares, si lo hacemos de esta forma, las promociones serán apreciadas de mejor forma y los esfuerzos realizados generaran resultados favorables plasmados en el incremento de ventas.

FORMULACIÓN DEL PROBLEMA

¿Cuáles son las reglas que sigue el consumo de SN que determinan los patrones de consumo en clientes de la tienda de suplementos nutricionales?

JUSTIFICACIÓN DE LA INVESTIGACIÓN

El incremento de consumo de suplementos nutricionales se ha elevado en nuestro país de forma considerable, cada vez vemos más locales de venta de estos productos en diferentes zonas de la capital, y encontramos diferentes marcas tanto nacionales como extranjeras.

Las empresas que comercializan estos productos concentran sus actividades en los diferentes distritos de Lima y ya han iniciado su expansión a las principales provincias de nuestro país, es así que podemos encontrar tiendas tales como : ⁴**Lab Nutrition** (14 tiendas en Lima , 1 en Trujillo) ,⁵**Nutri Point** (6 tiendas en Lima ,1 en Cusco,1 en Chimbote,1 Trujillo, 1 Chiclayo ,1 Piura) ⁶**Nutriforma**(3 tiendas en Lima) ⁷**Tech Nutrition**

Perú que distribuye a través de ⁸**Nutrishop**(4 tiendas en Lima) ⁹**Suplement House Perú**(2 tiendas en Lima,1 en Chiclayo y 1 en Piura).¹⁰**Nutri Sport** (2 Locales en Lima) , ¹¹**Vitaminas Mason Perú** con distribución en Boticas y Farmacias como **Inka Farma** , **Mifarma** , **Farmacia Universal** entre otras, así como a través de tiendas por departamento : **Metro Y Wong** y finalmente ¹²**Santa Natura** (21 tiendas en Lima,11 tiendas en provincia,19 distribuidores en Lima,12 distribuidores en Provincia).

De alguna forma el cuidado personal y de la salud se ha incrementado en Lima en esta última década, provocada quizás por influencia de los medios de comunicación y redes sociales en los que se promueve la alimentación adecuada, así como el consumo de suplementos alimenticios para tener una adecuada ingesta de nutrientes que requiere el cuerpo para su adecuado desenvolvimiento diario.

Esto desencadena en un importante negocio que va creciendo día a día, el vender productos clasificados dentro de este sector es sin duda muy atractivo.

Cuando se realiza una venta de un SN tenemos el registro de un cliente con diferentes atributos que lo identifican (**Nombres, Apellidos, Dirección**), estos datos además se ven enriquecidos por campañas nutricionales en las que se registran más características del cliente; el analizar el comportamiento de compra de ese consumidor, de forma apropiada, basados en sus características; puede simbolizar un incremento exponencial en las ventas de productos alternos.

El artículo ¹³ know your customer indica que el cross-selling es una estrategia esencial para mejorar la lealtad de marca, aumentar los ingresos, y maximizar el valor de relaciones con los clientes. La base para el éxito del cross-selling es **el conocimiento del cliente**. Un profundo conocimiento de las necesidades personales y profesionales de sus clientes le pone en una posición de recomendar soluciones adicionales.

La tesis busca conocer reglas que determinen el patrón de consumo de los clientes de la distribuidora de SN, para sobre esta base asociar productos que pueden comercializarse de forma conjunta; esperando aumentar el nivel de ventas. Las reglas que determinan el patrón de consumo las conoceremos luego de procesar los registros seleccionados del histórico de ventas del primer semestre del año 2016 haciendo uso de un software que aplica algoritmos especializados conocidos como: **Reglas de Asociación, Clustering, Redes Neuronales**, se espera identificar cuál de los modelos se ajusta mejor a los datos almacenados y conocer reglas que determinen el consumo de suplementos con una determinada probabilidad.

Las reglas que encontremos nos permitirán conocer el comportamiento de compra del cliente sobre la base de su histórico de adquisiciones , con esta información se busca establecer patrones de consumo que nos servirán para realizar campañas dirigidas a grupos de clientes que tienen características y comportamientos similares , así como mejorar el aprovisionamiento de existencia de acuerdo a demanda de productos. Esperamos con la aplicación de este conjunto de técnicas poder elevar el nivel de ventas anuales en un 25%, igualando de esta forma al crecimiento que viene experimentando el sector en ¹⁴ territorio brasileño.

La minería de datos permite que una organización se centre en la información más importante de la base de datos, lo que facilita a los gerentes tomar decisiones más informadas al predecir futuras tendencias y comportamientos (Chopoorian et al., 2001). ¹⁵

El marketing es una de las áreas más importantes donde se puede aplicar técnicas de minería de datos, ya que, permite a una organización, a partir de enormes cantidades de datos de clientes, orientarse hacia los más adecuados. Esto es de vital importancia para el departamento de marketing de cualquier organización. Se puede ahorrar importante cantidad de tiempo y dinero si una organización sabe quiénes son sus clientes y son capaces de predecir cuáles serán sus patrones de gasto. ¹⁵

La minería de datos ayuda a las personas en el departamento de marketing a entender lo que impulsa a un determinado comportamiento del cliente, el desempeño de un producto, o evento, y por lo tanto permite a la organización ser proactivo en lugar de reactivo a las condiciones cambiantes en el entorno empresarial .¹⁵

¹⁵ El comercio al por menor es el campo de aplicación principal de la minería de datos, porque este tipo de actividad (por ejemplo, supermercados, tiendas) recopila una gran cantidad de datos acerca de los clientes y sus compras, consumo y/o servicio recibido, lo cual proporciona a la minería de datos una gran cantidad de recursos, y debemos de considerar que "el comportamiento pasado de los consumidores es la mejor explicación de su propensión futura de consumo. "

¹⁶ A través de la recolección y el procesamiento de grandes cantidades de información de los clientes se llega a conocer el comportamiento del consumidor , grupos de consumo específico , intereses específicos de individuos, hábitos de gasto, tendencias de consumo para luego poder inferir el siguiente paso en el comportamiento del consumidor y basados en esto realizar campañas de marketing específicas.

Por todo lo citado con anterioridad se infiere la importancia de la aplicación de minería de datos en Retails y su importancia en el marketing, pues nos permite tener la información apropiada para establecer políticas y lineamientos que promuevan el buen desempeño organizacional.

MARCO REFERENCIAL

ANTECEDENTES

Existe un estudio previo referido al comportamiento del consumidor basados en técnicas de minería de datos titulado: ¹⁷ DATA MINING TECHNIQUES: A SOURCE FOR CONSUMER BEHAVIOR ANALYSIS fue realizado por el departamento de Ciencias de la computación de Vivekanand College, en la India.

En él se explica que para aplicar minería de datos es necesario escoger el modelo adecuado para poder obtener los resultados más certeros.

Mencionan además que el conocimiento generado con técnicas de minería de datos es muy útil en las decisiones del negocio y que su aplicación es bastante importante para poder asociar clientes con la compra de determinados productos y de esta forma saber el cambio hacia determinada marca de producto que logre satisfacer su necesidad ya que sus hábitos se encuentran estudiados previamente.

Se llega a concluir que la minería de datos es útil para estudiar el comportamiento de compra de los clientes en retails, también mencionan que hay ciertos hábitos de compra de los clientes que pueden ser utilizados para ofrecer mejor alternativas al cliente y poder retenerlo.

En el documento Know Your Buyer: A predictive approach to understand online buyers' behavior By Sandip Pal ¹⁸

Se estudia al cliente en el comercio electrónico y se llega a la conclusión que Millones de consumidores visitan sitios de comercio electrónico, pero sólo unos pocos llegan a comprar los productos, por lo tanto, se aplica minería de datos para lograr tener una mayor tasa de conversión de visitantes a compradores.

El objetivo del estudio es identificar a los compradores potenciales basándose en sus características demográficas, historial de transacciones, de acuerdo a los clics que hace, modelo de navegación, con lo cual se aplican políticas para mejorar la experiencia del cliente y de esta forma mejorar la tasa de conversión.

MARCO TEÓRICO

METODOLOGÍAS EN LA APLICACIÓN DE MINERÍA DE DATOS:

El Proceso KDD ¹⁹

El proceso KDD, tal como se presenta en (Fayyad et al, 1996), es el proceso de usar métodos de Data Mining (DM) para extraer conocimiento, utilizando una base de datos junto con cualquier pre procesamiento, sub- muestreo y transformación requeridos de la base de datos. Se consideran cinco etapas:

- 1. Selección** - Esta etapa consiste en crear un conjunto de datos objetivo, o enfocarse en un subconjunto de variables o muestras de datos, en las que se debe realizar el descubrimiento.
- 2. Procesamiento previo** - Esta etapa consiste en la limpieza y pre-procesamiento de datos de destino para obtener datos consistentes.
- 3. Transformación** - Esta etapa consiste en la transformación de los datos usando métodos de reducción de dimensionalidad o transformación.
- 4. Minería de Datos** - Esta etapa consiste en la búsqueda de patrones de interés en una forma representacional particular, dependiendo del objetivo de minería de datos (por lo general, la predicción).
- 5. Interpretación / Evaluación** - Esta etapa consiste en la interpretación y evaluación de las patrones.

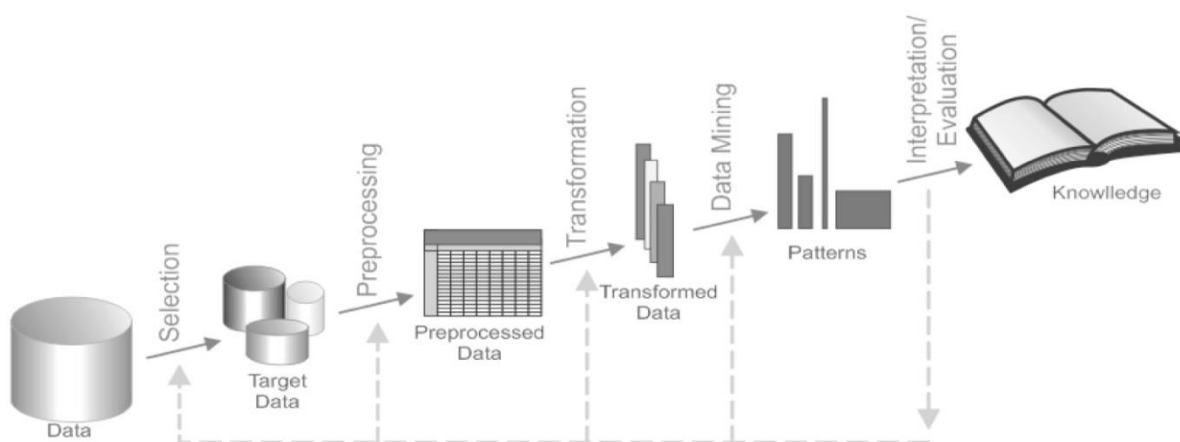


Ilustración 3 Etapas en el procedimiento KDD Stages in the KDD process. Fuente: Fayyad et al. (1995).

El proceso SEMMA ²⁰

El proceso SEMMA fue desarrollado por el Instituto SAS. El acrónimo SEMMA significa (Sample, Explore, Modify, Model and Assess).

Se considera un ciclo con 5 etapas para el proceso:

1. Muestra - Esta etapa consiste en sacar una muestra de un conjunto de datos; lo suficientemente grande como para contener la información significativa, pero lo suficientemente pequeño para manipular rápidamente. Esta etapa se señala como opcional.

2. Explorar - Esta etapa consiste en la exploración de los datos mediante la búsqueda de tendencias imprevistas y anomalías con el fin de obtener la comprensión y las ideas.

3. Modificar - Esta etapa consiste en la modificación de los datos creando, seleccionando y transformando las variables para enfocar el proceso de selección del modelo.

4. Modelo - Esta etapa consiste en modelar los datos permitiendo que el software busque automáticamente una combinación de datos que predice de manera fiable un resultado deseado.

5. Evaluar - Esta etapa consiste en evaluar los datos mediante la evaluación de la utilidad y fiabilidad de los resultados del proceso de minería de datos y estimar cuán bien se realiza.

El proceso SEMMA está vinculado al software SAS Enterprise Miner y pretende guiar al usuario en las implementaciones de DM.

SEMMA ofrece un proceso de fácil comprensión, permitiendo un desarrollo y mantenimiento adecuado y organizado de los proyectos de DM.

El proceso CRISP-DM ²⁰

El proceso CRISP-DM se desarrolló por medio del esfuerzo de un consorcio inicialmente compuesto por Daimler Chrysler, SPSS y NCR. CRISP-DM (CROSS-Industry Standard Process for Data Mining).

Consiste en un ciclo que consta de seis etapas:

1. Comprensión del negocio - Esta fase inicial se centra en la comprensión de los objetivos del proyecto y los requisitos desde una perspectiva de negocio, a continuación, convertir este conocimiento en una definición de problema de minería de datos y un plan preliminar diseñado para lograr los objetivos.

2. Comprensión de los datos - La fase de comprensión de los datos comienza con una recopilación inicial de datos y prosigue con las actividades para familiarizarse con los datos, identificar problemas de calidad de los datos, descubrir los primeros datos o detectar subconjuntos interesantes para formar hipótesis para Información oculta.

3. Preparación de datos - La fase de preparación de datos abarca todas las actividades para construir el conjunto de datos final a partir de los datos iniciales sin procesar.

4. Modelado - En esta fase, se seleccionan y aplican diversas técnicas de modelado y se calibran sus parámetros a valores óptimos.

5. Evaluación - En esta etapa el modelo o modelos obtenidos se evalúan más a fondo y se revisan los pasos ejecutados para construir el modelo para asegurarse de que alcanza adecuadamente los objetivos de negocio.

6. Implementación

- La creación del modelo generalmente no es el final del proyecto. Incluso si el propósito del modelo es aumentar el conocimiento de los datos, el conocimiento adquirido deberá ser organizado y presentado de manera que el cliente pueda utilizarlo.

Ilustración 4 Esquema CRISP



Imagen extraída de:

https://www.ibm.com/support/knowledgecenter/SS3RA7_15.0.0/com.ibm.spss.crispdm.help/crispoverview.htm?pos=2

COMPARATIVO ENTRE METODOLOGÍAS ²¹

Si comparamos KDD y SEMMA podemos afirmar que son equivalentes:

Podemos decir que las cinco etapas del proceso SEMMA pueden ser vistas como una implementación práctica de las cinco etapas del proceso KDD, ya que está directamente vinculada al software SAS Enterprise Miner.

KDD	SEMMA
Muestra	Selección.
Explorar	Pre-procesamiento
Modificar	Transformación
Modelo	Data Mining
Evaluación	Interpretación / Evaluación

Si comparamos KDD y CRISP-DM podemos observar que la metodología CRISP-DM incorpora los pasos que deben preceder y seguir el proceso KDD:

KDD	CRISP-DM
La fase de Entendimiento de Negocios	Puede identificarse con el desarrollo de una comprensión del dominio de la aplicación, el conocimiento previo relevante y los objetivos del usuario final.
La fase de implementación	Puede identificarse con la consolidación incorporando este conocimiento en el sistema.
La fase de Entendimiento de Datos	Puede ser identificada como la combinación de Selección y Pre procesamiento.
La fase de preparación	Se puede identificar con Transformación.
La fase de modelado	Se puede identificar con Data Mining
La fase de Evaluación	Puede ser identificada con Interpretación/ Evaluación.

Comparación CRISP –DM Y SEMMA

SEMA	CRIP-DM
Muestra y Exploración	Fase de Entendimiento de Datos
Modificar	Fase de preparación de datos
Modelo	Fase de modelado,
Evaluar paralelos	Fase de evaluación
Ambos modelos pretenden ser algo cíclicos en lugar de lineales.	

Diferencias entre CRISP-DM y SEMMA. ²²

SEMMA fue desarrollado para un paquete específico de software: SAS Enterprise Miner y, pone menos énfasis en las fases de planificación inicial cubiertas en las fases de CRISP-DM (Business Understanding y Data Understanding) y omite totalmente la fase de implementación.

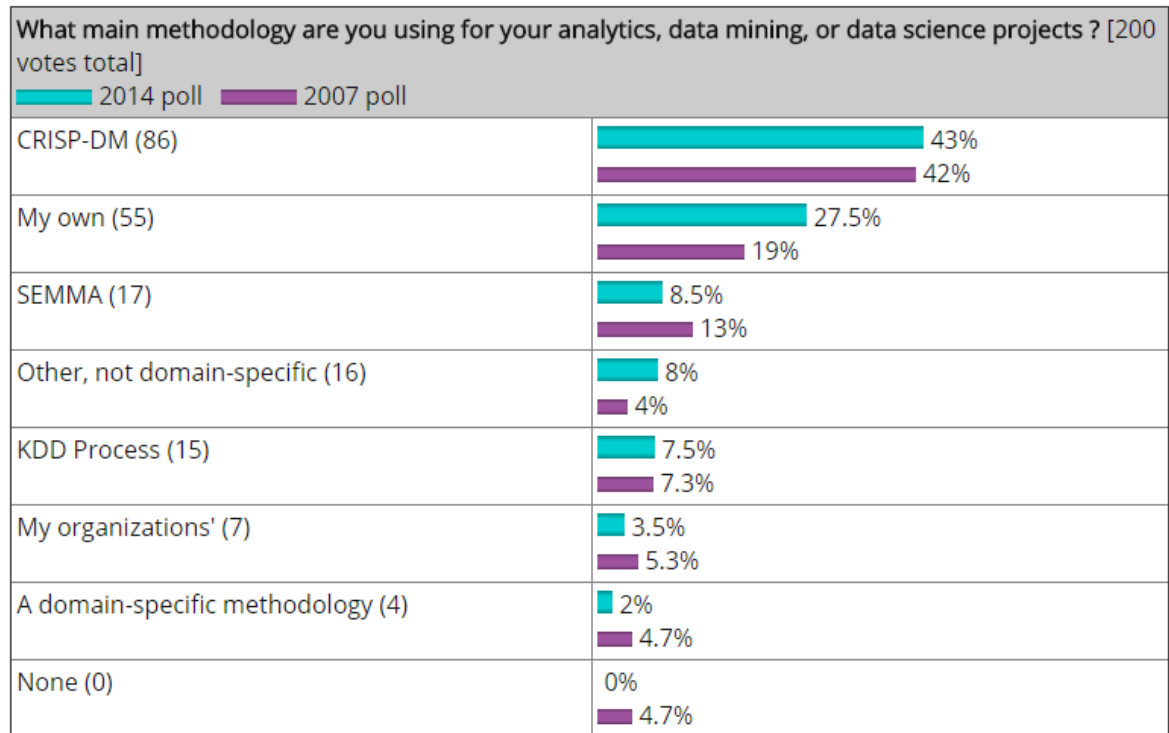
ELECCIÓN DE LA METODOLOGÍA

Para elegir la metodología a implementar realizamos un análisis sobre la base de características deseadas, luego de lo cual se elige implementar la metodología que satisface de forma completa los criterios de selección.

Deseamos una metodología que sea de amplio uso entre profesionales por lo cual recurrimos a un análisis de kdnuggets ²⁴ en el que se puede ver la preferencia de 200 usuarios en el uso de una metodología de minería de datos.

Vemos que en primer lugar de preferencia se encuentra la metodología CRISP –DM con un 43%, en segundo lugar SEMMA con un 8.5% y en tercer lugar KDD con 7.5%.

Ilustración 5 PREFERENCIA DE METODOLOGÍA ²⁴



Además de ser una metodología predominante en su uso, también deseamos otros aspectos

Tabla 3 EVALUACION DE METODOLOGÍA (Elaboración propia)

CRITERIOS/METODOLOGÍAS	CRISP-DM	SEMMA	KDD
Metodología Estructurada	✓	✓	✓
Metodología Independiente	✓	X	✓
Ampliamente Usada	✓	X	X
Mejora la calidad de resultados en proyectos de Data Mining.	✓	✓	✓
Herramientas y técnicas independientes	✓	✓	✓
Finalidad diversa (Ej. Ampliamente estable en la resolución de problemas variados).	✓	✓	✓
Fácil de implementar	✓	✓	✓

METODOLOGÍA DE MINERÍA DE DATOS EMPLEADA EN EL PRESENTE PROYECTO

Luego del análisis realizado, decidimos seleccionar e implementar la metodología CRISP-DM en el presente proyecto de tesis, ya que, es una metodología neutral, fácil de implementar, mejora la calidad de resultados en los proyectos, es una metodología estructurada y ha sido ampliamente adoptada en análisis de minería de datos.

De acuerdo a Duncan Ross, especialista en análisis avanzado de CRIPM-DM, copatrocinador de Teradata, la tecnología de minería de datos ha madurado mucho desde que se lanzó en 1999.

CRISP-DM ²⁰ fue financiado inicialmente por una subvención de la Unión Europea a mediados de los años noventa y, a diferencia de los enfoques elaborados por grupos de servicios profesionales de proveedores, fue diseñado para ser neutral en cuanto a herramientas y proveedores.

De acuerdo a Ross: Las empresas que participaron de la creación de la metodología buscan ayudar a la gente a entender cómo hacer que la minería de datos sea una actividad empresarial principal en las empresas.

Benchmark Analysis: Softwares de Minería de Datos.

En el mercado existen muchas herramientas que nos permiten realizar análisis de minería de datos, algunas de estas son más reconocidas que otras y cuentan con mayor presencia en el mercado por pertenecer a empresas líderes en tecnología y brindar mayores funcionalidades. Se han seleccionado 5 de ellas y se las ha sometido a evaluación de usuarios, conforme a la experiencia de uso experimentada. Considerando 6 aspectos:

- Facilidad de uso.
- Características y funcionamiento.
- Características avanzadas.
- Integración.
- Soporte para el cliente.
- Desempeño.

Para cada SW se consideran 10 opiniones expresadas con un puntaje en el rango de 1-10 puntos, basados en la experiencia del usuario al usar la herramienta.

Los puntajes fueron asignados por personas que se desempeñan en el área de análisis de datos y tiene contacto con la herramienta por:

- Desempeño laboral
- Herramienta investigativa

- Herramienta educativa.

La puntuación se realiza bajo el sgte lineamiento:

Tabla 4 LEYENDA PUNTAJE SW

FACILIDAD DE USO	▪ 1 Difícil de utilizar. ▪ 10 Fácil de utilizar.
CARACTERÍSTICAS Y FUNCIONAMIENTO	▪ 1 cuenta con características y funcionamiento limitado. ▪ 10 cuenta con muchas características y amplio funcionamiento.
CARACTERÍSTICAS AVANZADAS	▪ 1 cuenta con reducidas características avanzadas. ▪ 10 cuenta con muchas características avanzadas.
INTEGRACIÓN	▪ 1 se integra difícilmente con motores de base de datos o datos externos a la herramienta ▪ 10: se integra fácilmente con motores de base de datos y recursos externos.
SOPORTE PARA EL CLIENTE	▪ 1 es muy escaso el soporte brindado. ▪ 10 es muy fácil de solicitar soporte
DESEMPEÑO	▪ 1 el desempeño no satisface mis necesidades. ▪ 10 estoy satisfecho con el resultado obtenido.

A manera de resumen, se entrega al encuestado un resumen de las herramientas en mención.

Tabla 5 DESCRIPCION DE HERRAMIENTAS COMPARADAS ²³

LIONOSO	Es una máquina de aprendizaje completo y herramienta de optimización inteligente desarrollado por un equipo de expertos, ofrece diferentes soluciones dirigidas a la salud y las empresas. • Herramienta de optimización y aprendizaje automático. • Herramienta ideal para rápido análisis. • Está clasificada en el ámbito académico y organización sin fines de lucro. • El software promueve la investigación y el desarrollo industrial. • Investigadores pueden hacer su trabajo con esta herramienta. • El software puede conectar e integrarse con base de datos. • Brinda soluciones para bioinformática, áreas de salud y negocios. • Permite aprender de los datos.
ADVANCEDMINER	Es una herramienta analítica integrada para el procesamiento, análisis y modelado de datos. ▪ Tiene una interfaz gráfica. ▪ Ofrece un ambiente completo y amigable para la exploración de datos. ▪ Permite segmentación y perfilado de usuarios. ▪ Analiza los datos y prepara modelo de datos.

	▪ Permite la exploración de datos. ▪ Permite la detección de fraudes y el análisis de riesgo. ▪ Ofrece un entorno completo y fácil de usar para la exploración de datos. ▪ Permite el procesamiento de datos. ▪ Ofrece funciones para extraer y guardar datos de diferentes sistemas y archivos de bases de datos. ▪ Realiza varias funciones de análisis. ▪ Puede preparar y evaluar varios tipos de modelos de datos. ▪ Presenta informes de calidad después de evaluaciones modales. ▪ Tiene algunas características avanzadas, con estas se puede realizar la integración con el paquete JAVA y R. ▪ En funciones avanzadas, el software también puede soportar lenguaje SQL. ▪ El software lleva a cabo análisis de riesgo, calificación de crédito, segmentación de clientes, perfil de clientes y análisis basado en el mercado, análisis de tasa de churn, optimización de campañas de marketing , detección de fraudes, venta cruzada y up-selling.
IBM SPSS MODELER	Es una plataforma analítica de IBM, que brinda inteligencia predictiva a problemas de negocios del día a día, ▪ Provee un rango de análisis avanzado que incluye análisis de texto, análisis de redes sociales, modelado automático, preparación de datos, gestión y optimización de decisiones. ▪ Se conecta con datos almacenados en data warehouse, base de datos, Hadoop, o archivos planos al margen de si están estructurado por edad, precio, producto, locación o sin estructurar como ejemplo en un texto, email o medios sociales.
MICROSOFT SQL SERVER INTEGRATION SERVICES	Microsoft Integration Services es una plataforma para crear soluciones de integración de datos y transformaciones de datos a nivel de empresa. ▪ Se utiliza para resolver problemas empresariales complejos al copiar o descargar archivos, enviar mensajes de correo electrónico en respuesta a eventos, actualizar almacenes de datos, limpiar y extraer datos y administrar objetos y datos de SQL Server. ▪ Los paquetes pueden trabajar solos o en conjunto con otros paquetes para resolver las complejas necesidades del negocio. ▪ Puede extraer y transformar datos de una amplia variedad de fuentes, como archivos de datos XML, archivos planos y fuentes de datos relacionales, y luego cargar los datos en uno o más destinos. ▪ Incluye un rico conjunto de tareas y transformaciones integradas: Herramientas para la construcción y administración de paquetes. ▪ Puede utilizar las herramientas gráficas para crear soluciones sin escribir una sola línea de código. ▪ Puede crear paquetes mediante programación y codificar tareas personalizadas y otros objetos de paquete.

ORACLE DATA MINING ODM	Oracle Data Mining ODM proporciona una potente funcionalidad de minería de datos y permite a los usuarios descubrir nuevos conocimientos sobre datos ocultos. ▪ Oracle Data Mining ODM tiene varios algoritmos de minería de datos y análisis de datos y forma parte de Oracle Relational Database Management System Enterprise Edition. ▪ Estos algoritmos proporcionan medios para la creación, manipulación, aplicación, prueba y despliegue de modelos. ▪ Se utilizan para la clasificación, predicción, regresión, asociaciones, selección de características, detección de anomalías, extracción de características y análisis especializados. ▪ Los modelos se implementan en el kernel de la base de datos Oracle y se almacenan como objetos de la base de datos. ▪ Los modelos se pueden crear utilizando una interfaz gráfica de usuario. ▪ Los analistas de negocios pueden hacer análisis predictivos utilizando la hoja de cálculo de Oracle para análisis predictivo. ▪ La mayoría de los enfoques de aprendizaje de máquina conocidos como árboles de decisión, bayes ingenuos, máquinas vectoriales de apoyo, modelo lineal generalizado para minería predictiva, reglas de asociación, K-medios y partición ortogonal, agrupación y factorización matricial no negativa están disponibles. ▪ La minería de texto también está disponible en Oracle Data Mining ODM.
--------------------------------------	---

PUNTAJES POR HERRAMAMIENTA OBTENIDOS:

Tabla 6 IBM SPSS MODELER

IBM SPSS MODELER

CRITERIOS CONSIDERADOS	PROFESIONAL QUE USA HERRAMIENTAS DE DATA MINING.										
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	$\bar{X}$
FACILIDAD DE USO	7	8	8	8,5	8,5	8	7,5	7	7,5	8	7,80
CARACTERÍSTICAS Y FUNCIONAMIENTO	7,5	6	7,3	7,6	7,7	7	7,4	7,5	7,2	7,8	7,30
CARACTERÍSTICAS AVANZADAS	7,5	9	8	7	7,5	8	7,5	7	8	7,5	7,70
INTEGRACIÓN	7,8	8	7,4	7,5	7,5	7,5	7,6	7,5	7,3	7,8	7,59
SOPORTE PARA EL CLIENTE	7,5	9	9	8	8	7	8	7	8	8	7,95
DESEMPEÑO	7,5	7	8	7,5	7,5	7	7,6	7,7	8	7,8	7,56
PROMEDIO GENERAL											7,65

Tabla 7 ORACLE DATA MINING ODM

CRITERIOS CONSIDERADOS	PROFESIONAL QUE USA HERRAMIENTAS DE DATA MINING,										
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	$\bar{X}$
FACILIDAD DE USO	7,6	7,7	7,4	8	7,5	7,5	7,6	9	7,3	7,8	7,74
CARACTERÍSTICAS Y FUNCIONAMIENTO	7	9	9	7,6	7,7	7,5	7,4	7,3	7,2	7,8	7,75
CARACTERÍSTICAS AVANZADAS	7,5	9	7,4	8	7,5	7,4	7,6	8	7,5	7,4	7,73
INTEGRACIÓN	7,8	7,7	8	7,5	7,5	8	7,6	7,5	9	7,8	7,84
SOPORTE PARA EL CLIENTE	9	8	7,4	7,5	7,5	7,5	7,6	7,8	7,3	8	7,76
DESEMPEÑO	7,5	7,7	9	7,5	8	8	7,6	7,7	9	7,8	7,98
PROMEDIO GENERAL											7,80

Tabla 8 MICROSOFT SQL SERVER INTEGRATION SERVICES

CRITERIOS CONSIDERADOS	PROFESIONAL QUE USA HERRAMIENTAS DE DATA MINING,										
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	$\bar{X}$
FACILIDAD DE USO	7,6	7,7	7,5	7,5	7,5	7,4	7,6	7,7	7,5	7,8	7,58
CARACTERÍSTICAS Y FUNCIONAMIENTO	7,5	8	7,5	7,6	7,7	8	7,4	8	7,5	7,8	7,70
CARACTERÍSTICAS AVANZADAS	7,5	7,7	7,5	7,5	7,5	7,5	7,6	7,7	8	7,8	7,63
INTEGRACIÓN	7,8	7	8	8	7,5	9	7,6	9	7,5	7,8	7,92
SOPORTE PARA EL CLIENTE	8	8	7,5	7,5	8	7,5	7,6	8	8	7,8	7,79
DESEMPEÑO	9	8	7,4	8	9	7,5	7,6	8	8	9	8,15
PROMEDIO GENERAL											7,80

Tabla 9 LIONOSO

CRITERIOS CONSIDERADOS	PROFESIONAL QUE USA HERRAMIENTAS DE DATA MINING,										
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	$\bar{X}$
FACILIDAD DE USO	7	7	8	8	7	7	6	7	7	7	7,10
CARACTERÍSTICAS Y FUNCIONAMIENTO	5	7	7,5	7,6	8	8	7,5	7	6	7,5	7,11
CARACTERÍSTICAS AVANZADAS	5	7	7	7	7,5	7	7,5	7	7	7,5	6,95
INTEGRACIÓN	7	7	7	6	7	8	7	6	7	7,5	6,95
SOPORTE PARA EL CLIENTE	4	5	5	7	4	4	5	5	4	6	4,90
DESEMPEÑO	7	6	6	7	8	7	7	7	7	7	6,90
PROMEDIO GENERAL											6,65

Tabla 10 ADVANCEDMINER

CRITERIOS CONSIDERADOS	PROFESIONAL QUE USA HERRAMIENTAS DE DATA MINING,										
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	$\bar{X}$
FACILIDAD DE USO	7	6	7	6	6	7	6	6	7	7	6,50
CARACTERÍSTICAS Y FUNCIONAMIENTO	7,5	7	7	7,6	7	7,5	7	7	8	7	7,26
CARACTERÍSTICAS AVANZADAS	7,5	7	7	8	7	7	9	7,7	7,3	7,5	7,50
INTEGRACIÓN	6	7	7,5	7,5	7	7	6	6	6	5	6,50
SOPORTE PARA EL CLIENTE	4	3	6	5	4	4	4	4	5	5	4,40
DESEMPEÑO	7,5	7,7	7	7,5	7,5	7,4	7,6	7	7,3	7,8	7,43
PROMEDIO GENERAL											6,60

Oracle Data Mining ODM	Es parte de Oracle RDBMS Enterprise Edition.
Microsoft SQL Server integration services	Soportado con todas las funcionalidades en SQL Server Enterprise, Developer y parcialmente en Standard, Express.

Tabla 11 PRECIOS DE LICENCIAMIENTO²⁴

ESPECIFICACIONES	MICROSOFT SQL SERVER 2008 R2	MICROSOFT SQL SERVER 2012	ORACLE
4CPU,4 NUCLEOS POR CPU	\$113,916	\$113,916	\$380,000
4CPU,8 NUCLEOS POR CPU	\$113,916	\$227,832	\$1,520,000
8CPU,4CORES POR CPU	\$113,916	\$227,832	\$1,520,000
16CPU,8 CORES POR CPU	\$455664	\$911,328	\$3,040,000

Tabla 12 CARACTERÍSTICAS ADICIONALES

ESPECIFICACIONES	PARTICIONAMIENTO DE TABLA	COMPRESIÓN DE DATOS	OLAP	COSTO ADICIONAL TOTAL
4CPU,4 NUCLEOS POR CPU	SQL server \$0 Oracle \$92000	SQL server \$0 Oracle \$92000	SQL server \$0 Oracle \$92000	SQL server \$0 Oracle \$368,000
4CPU,8 NUCLEOS POR CPU	SQL server \$0 Oracle \$184,000	SQL server \$0 Oracle \$184,000	SQL server \$0 Oracle \$184,000	SQL server \$0 Oracle \$736,000
8CPU,4CORES POR CPU	SQL server \$0 Oracle \$184,000	SQL server \$0 Oracle \$184,000	SQL server \$0 Oracle \$184,000	SQL server \$0 Oracle \$736,000
16CPU,8 CORES POR CPU	SQL server \$0 Oracle \$736,000	SQL server \$0 Oracle \$736,000	SQL server \$0 Oracle\$1,472,000	SQL server \$0 Oracle \$2,944,000

ELECCIÓN DE LA HERRAMIENTA DE MINERÍA DE DATOS.

Luego del análisis de la herramienta hemos encontrado un puntaje similar en dos de las herramientas seleccionadas:

MICROSOFT SQL SERVER INTEGRATION SERVICES y ORACLE DATA MINING ODM

Con un puntaje promedio de 7,8 por lo tanto, procedemos a los análisis económicos de estos últimos para lograr un desempate y poder inclinarnos por uno de los SW.

Si nos fijamos en los precios de licenciamiento, el servidor de gama alta de Oracle se encuentra en \$ 6,000,000 mientras que el costo de Microsoft SQL Server 2012 es inferior a \$ 1,000,000.

Por los resultados observados, elegimos la herramienta de Microsoft debido a que cuenta con uno de los mayores puntajes y el licenciamiento es más económico.

ALGORITMO DE ASOCIACIÓN DE MICROSOFT ²⁵

Los modelos de asociación se generan basándose en conjuntos de datos que contienen identificadores para casos individuales y para los elementos que contienen los casos. Un grupo de elementos de un caso se denomina un conjunto de elementos. Un modelo de asociación se compone de una serie de conjuntos de elementos y de las reglas que describen cómo estos elementos se agrupan dentro de los casos.

El algoritmo de asociación de Microsoft puede encontrar potencialmente muchas reglas dentro de un conjunto de datos. El algoritmo usa dos parámetros, compatibilidad y probabilidad, para describir los conjuntos de elementos y las reglas que genera.

CÓMO FUNCIONA EL ALGORITMO

El algoritmo de asociación de Microsoft recorre un conjunto de datos para hallar elementos que aparezcan juntos en un caso. Después, agrupa en conjuntos de elementos todos los elementos asociados que aparecen, como mínimo, en el número de casos especificado en el parámetro MINIMUM_SUPPORT.

ALGORITMO DE CLÚSTERES DE MICROSOFT ²⁶

Es un algoritmo de segmentación o clústeres que itera en los casos de un conjunto de datos para agruparlos en clústeres que contengan características similares. Estas agrupaciones son útiles para la exploración de datos, la identificación de anomalías en los datos y la creación de predicciones.

CÓMO FUNCIONA EL ALGORITMO

El algoritmo de clústeres de Microsoft identifica primero las relaciones de un conjunto de datos y genera una serie de clústeres basándose en ellas.

ALGORITMO DE RED NEURONAL DE MICROSOFT ²⁷

El algoritmo de red neuronal de Microsoft es una implementación popular y adaptable para el aprendizaje automático. El algoritmo funciona mediante la prueba de cada estado posible del atributo de entrada en cada estado posible del atributo de predicción y calcula las probabilidades para cada combinación basándose en los datos de entrenamiento.

Puede usar estas probabilidades para tareas de clasificación o regresión para predecir un resultado en función de algunos atributos de entrada. Una red neuronal también puede utilizarse para el análisis de asociación.

CÓMO FUNCIONA EL ALGORITMO

El algoritmo de red neuronal de Microsoft crea una red que se compone de tres niveles de nodos (a veces denominado neuronas). Estas capas son la capa de entrada, la capa oculta, y la capa de salida.

Nivel de entrada: los nodos de entrada definen todos los valores de atributo de entrada para el modelo de minería de datos y sus probabilidades.

Nivel oculto: El nivel oculto es donde se asignan pesos a las distintas probabilidades de las entradas. Un peso describe la relevancia o importancia de una entrada determinada para el nodo oculto. Cuanto mayor sea el peso asignado a una entrada, más importante será el valor de dicha entrada. Los pesos pueden ser negativos, lo que significa que la entrada puede desactivar, en lugar de activar, un resultado concreto.

Nivel de salida: nodos de salida representan valores de atributo de predicción para el modelo de minería de datos.

ARQUITECTURA DATA WAREHOUSE EN 3 CAPAS ²⁸

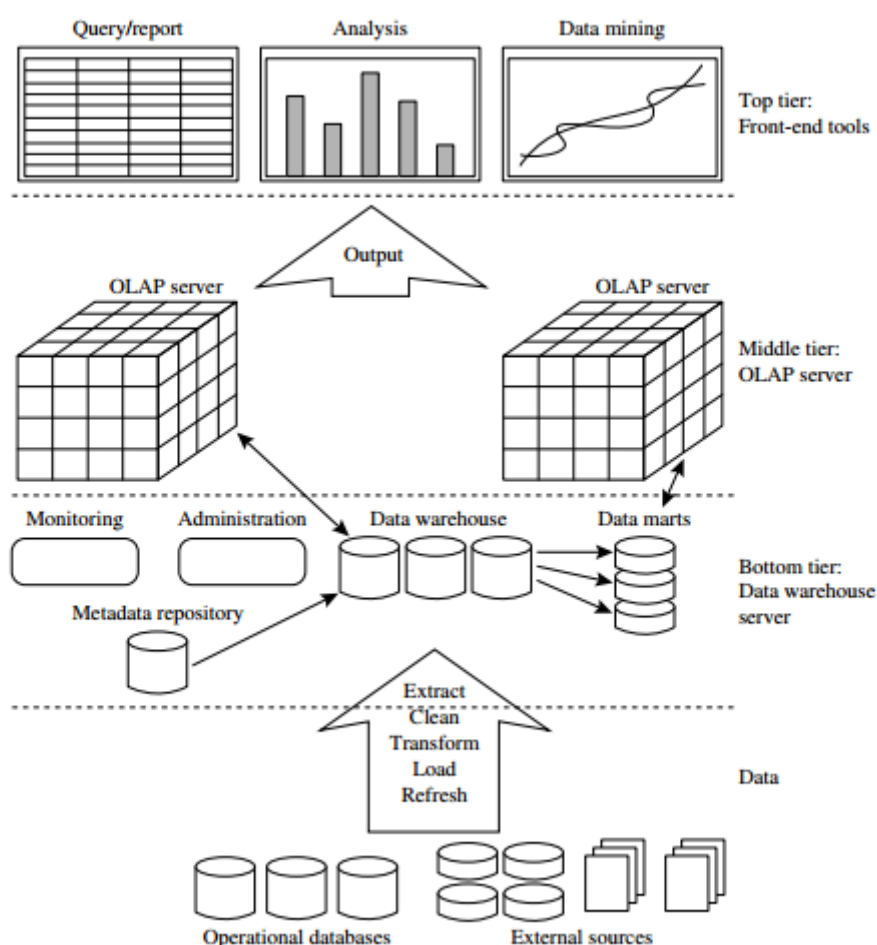


Ilustración 6 ARQUITECTURA DATA WAREHOUSE EN 3 CAPAS

ARQUITECTURA DE UN SISTEMA TÍPICO DE MINERÍA DE DATOS ²⁹

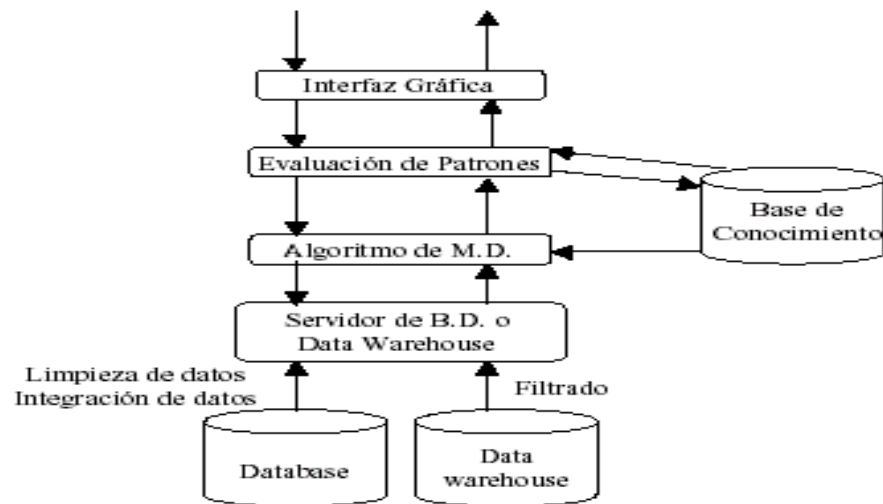


Ilustración 7 ARQUITECTURA DE UN SISTEMA TÍPICO DE MINERÍA DE DATOS

COSTOS DEL PROYECTO DE MINERÍA DE DATOS

Tomando como base las 6 etapas de la metodología CRISP –DM, costos en servicios varios y licencias requeridas para el desarrollo del proyecto.

Se costean principales roles empleados en el proyecto, haciendo una estimación de la cantidad de horas por cada recurso de acuerdo a la Etapa.

Costos planeados en sueldos. (Soles)

ETAPA 1: TABLA 13 COSTOS COMPRENSIÓN DEL NEGOCIO.

Roles	Remuneración /Hora(S/)	Horas semanales	Horas totales	Costo
Administrador de la Tienda.	20	3	12	240
Gerente del proyecto.	50	5	20	1000
Analista de minería de datos.	25	10	40	1000
Total Etapa 1				S/ 2,240

ETAPA 2. TABLA 14 COMPRENSIÓN DE LOS DATOS

Roles	Remuneración /Hora(S/)	Horas Semanales	Horas Totales	Costo
Gerente del proyecto	50	8	30	1500
Analista de minería de datos	25	10	40	1000
Analista de base de datos	25	5	30	750
Total Etapa 2				S/ 3,250

ETAPA 3. TABLA 15 PREPARACIÓN DE DATOS

Roles	Remuneración /Hora(S/)	Horas Semanales	Horas Totales	Costo
Gerente del proyecto	50	3	10	500
Analista de minería de datos	25	4	50	1250
Analista de base de datos	25	3	40	1000
Total Etapa 3				S/ 2,750

ETAPA 4. TABLA 16 MODELADO

Roles	Remuneración /Hora(S/)	Horas semanales	Horas totales	Costo
Gerente del proyecto	50	3	10	500
Analista de minería de datos	25	5	20	500
Total Etapa 4				S/ 1000

ETAPA 5. TABLA 17 EVALUACIÓN

Roles	Remuneración /Hora(S/)	Horas semanales	Horas totales	costo
Gerente del proyecto	50	3	10	500
Analista de minería de datos	25	5	15	375
Total Etapa 5				S/ 875

ETAPA 6. TABLA 18 IMPLEMENTACIÓN

Roles	Remuneración /Hora(S/)	Horas Semanales	Horas Totales	Costo
Gerente del proyecto	50	3	10	500
Analista de minería de datos	25	5	10	250
Total Etapa 6				S/750

TABLA 19 COSTOS PLANEADOS EN SERVICIOS

Ítems	Costo por semana(S/)	Costo por 20 semanas	Costo Total
Luz	20	400	400
Internet	20	400	400
Telefonía	10	200	200
Material de Oficina	10	200	200
Otros Gastos	30	600	600
Cantidad			
Laptop	1	1100	1100
Costos en Servicios			S/ 2900

TABLA 20 COSTOS PLANEADOS EN LICENCIAS

Licencia	Cant	Precio por unidad(S/)	Duración Licencia Años.	Costo Total	Costo por Mes	Costo por 5 meses(S/)
Microsoft office	1	900	2	900	37.5	187.5
SQL Server 2014 Standard - 2 Core License - Open Business	1	11028	Perpetuo	11028	919	4595
Visual Studio Professional 2015(Stand Alone License)	1	1697	Perpetua	1697	141.42	701.1
Costos en Licencias						S/ 5483.6

Tabla 21 COSTOS TOTALES.

Total sueldos	S/ 10,865
Total de servicios	S/ 2,900
Total de Licencias	S/ 5,483.6
Total del proyecto	S/ 19,248.6

En el presente proyecto no se incurrió en costos de datos, estos costos se deben considerar si como parte del proyecto es necesaria la compra de los mismos, para adquirirlos podemos recurrir a una investigadora de mercado, en Perú existen investigadores de mercado, cito una de ellas, Arellano Marketing puede facilitarnos acceso a los datos que se requieren por una suscripción anual.

El costo anual en el cual se incurre por el acceso a su base de datos es de 6000 soles anuales, lo cual nos da derecho a una base de datos dinámica que se va actualizando o de lo contrario podemos acceder al precio por contacto “a la carta” con un precio de 1.5-1.8 soles por registro con información estática.

BENEFICIOS.

- Conocer las reglas de consumo nos permite realizar ofrecimientos de paquetes que incrementen los ingresos a la tienda.
- Crear campañas orientadas.
- Conocer mejor a nuestros clientes.
- La tienda puede reorganizar los productos en la tienda dando más visualización a los productos que más se consumen y asociando productos a características de usuarios para mejorar la experiencia del cliente.

Tabla 22 BENEFICIOS.

PARA	Administradores, Gerentes, Supervisores.
QUIENES	Necesitan un análisis que les permita saber las reglas que determina el consumo, con la finalidad de poder realizar campañas de Marketing, tomar decisiones basadas en información, aumentar el nivel de ventas.
EL(NOMBRE DEL ANALISIS)	Análisis de minería de datos para determinar patrones de consumo futuro.
QUE	Permitirá conocer las reglas formadas por características de grupos de usuario y las cuales determinan el consumo de un producto en particular.
DIFERENTE A	Mantener la información sin análisis.
NUESTRO ANALISIS	Permite conocer reglas que determinan la compra de productos, el cual nos permite ofrecer paquetes de productos que pueden venderse de forma conjunta.

³⁰ **DATA MINING: A COMPETITIVE WEAPON FOR BANKING AND RETAIL INDUSTRIES**

En los Retailers, la minería de datos se puede utilizar para proporcionar información sobre las tendencias de ventas de productos, los hábitos y preferencias de compra del cliente, los plazos de entrega de proveedores y el rendimiento de entrega, variaciones estacionales, los períodos de mayor tráfico de clientes y datos predictivos similares para tomar decisiones proactivas.

De acuerdo con Groth (1998), los minoristas están interesados en la creación de modelos de minería de datos para responder a preguntas tales como:

¿Cuál es el gasto del Cliente en el Tiempo?

¿Cuál es la frecuencia de compra del cliente?

En el caso particular de la tesis aplicaremos minería de datos a la información almacenada por la distribuidora de suplementos nutricionales logrando entender el comportamiento de compra de los clientes, luego del análisis de su información histórica y de esta forma poder presentarle productos y promociones que simbolizen mayor valor para este.

³¹ **Data Mining Concepts and Techniques Second Edition**

Jiawei Han and Micheline Kamber

Data Mining for the Masses Dr. Matthew North ISBN: 0615684378

“La minería de datos como disciplina es en gran medida transparente para el mundo. La mayoría de las veces, ni siquiera notamos que está sucediendo. Pero cada vez que te registras para obtener una tarjeta de compras, vas a una tienda de comestibles, pagas con tarjeta de crédito, o navegas por la web, estamos creando datos. Estos datos se almacenan en grandes conjuntos de potentes ordenadores propiedad de las compañías.

Situados dentro de esos conjuntos de datos están los patrones de indicadores de nuestros intereses, nuestros hábitos y nuestros comportamientos. La minería de datos permite a los investigadores encontrar e interpretar esos patrones, lo que ayuda a tomar decisiones más informadas y servir mejor a sus clientes.”

El autor menciona en su libro técnicas y herramientas para la aplicación de minería de datos, la importancia de aplicar técnicas para conocer hábitos y comportamientos de los clientes y sobre los resultados poder tomar decisiones apropiadas; es justamente lo que deseamos plantear en el proyecto de tesis.

³² **MINERÍA DE DATOS A TRAVÉS DE EJEMPLOS**

María Pérez Marqués ISBN: 978-84-941801-4-9

“De un modo sencillo podemos definir la minería de datos como un conjunto de técnicas encaminadas al descubrimiento de la información contenida en grandes conjuntos de datos. Se trata de analizar comportamientos, patrones, tendencias, asociaciones y otras características del conocimiento inmerso en los datos.

Actualmente se dispone de grandes cantidades de datos y es más necesario que nunca poder analizarlos ordenadamente para extraer de un modo automatizado la inteligencia contenida en ellos utilizando técnicas especializadas apoyadas en herramientas informáticas. Estas técnicas constituyen la minería de datos.”

Indica que se pueden aplicar técnicas de minería de datos para analizar comportamientos, patrones y tendencias, esto se plantea realizar en el presente trabajo de tesis, descubrir reglas que determinen el comportamiento de compra futuro de clientes haciendo el uso de herramientas informáticas.

³³ **La autora Sofía Vallejos en su trabajo de investigación de minería de datos sostiene :**

“La Minería de Datos ha facilitado el proceso de extraer información de los grandes volúmenes de datos, revelando conocimiento innovador a las organizaciones, permitiendo conocer de una forma más detallada el comportamiento de variables sumamente importantes para las empresas, como lo son inventarios, ventas, el comportamiento de los clientes, entre otros factores.

Hace uso de técnicas, análisis de información así como diversas áreas de conocimiento como la estadística, inteligencia artificial, computación gráfica, bases de datos y el procesamiento masivo.”

El conocimiento oportuno de la información puede definir el éxito de una organización, en la distribuidora de suplementos tienen ese pensamiento que concuerda con las teorías de múltiples autores en el campo de la minería de datos.

³⁴**El autor (Chung et al., 2009). Menciona que** IT Ayuda a las organizaciones empresariales a mejorar su Competencia y sostener el crecimiento continuo de la empresa.

Definitivamente la época donde se delimitaba el accionar de IT solamente a la infraestructura de la empresa ya quedo de lado, actualmente a las empresas les interesa que áreas de tecnología creen valor en las diferentes áreas de la empresa, con la presente tesis vamos a hacer uso de la tecnología para incrementar la competencia de la empresa.

³⁵ **Según el autor Kwan et al.(2005), indica que:** “La tecnología puede apoyar la gestión del conocimiento en el negocio, a través del almacenamiento de datos, minería de datos, el Internet y los sistemas de gestión de documentos. En recientes décadas, la minería de datos se ha aplicado a una amplia gama de temas y áreas.

La mayoría de negocios y organizaciones lo utilizan para encontrar el área de problemas y permiten tomar decisiones estratégicas que facilitan a las organizaciones empresariales alcanzar el éxito.

Se aplica con el fin de comprender el razonamiento de los patrones de compra de los clientes, muchas empresas utilizan herramientas automatizadas para estudiar el comportamiento de sus clientes. Una vez que la información pertinente se ha obtenido, se puede utilizar de una manera que permitirá la organización para predecir el comportamiento de sus clientes.”

³⁶ **Aplicación de técnicas de minería de datos para la fidelización y retención de clientes**

Autor: Victor López Fandiño Business Intelligence Solutions Specialist IBM Software Group

Data Mining puede ser aplicada a cualquier iniciativa de modelización de los datos sobre clientes, el objetivo puede ser múltiple y siempre debe estar encaminada a soportar la toma de decisiones estratégicas.

Se puede aplicar para Analizar:

Compra de productos, Hábitos de comportamiento, Quejas y reclamaciones, Clasificaciones, Segmentación de clientes, Afinidad entre productos, Patrones secuenciales de compra, Rentabilidad de productos y clientes, Rendimiento de campañas, Tratamiento diferenciado a clientes. Esquemas de precio diferenciados, Optimización de la labor comercial, Venta cruzada de productos, Efectividad de promociones, Incremento de cuota de mercado, Retención de clientes Recuperación de clientes Captación de clientes.

En nuestro caso, la presente tesis analiza la información buscando conocer hábitos de comportamiento de los clientes, afinidad entre producto y cliente, patrones secuenciales de compra, Rentabilidad de productos y clientes y venta cruzada de productos.

OBJETIVOS E HIPÓTESIS

OBJETIVOS

Aplicación de minería de datos a través de un Software Informático para encontrar reglas de consumo que determinen patrones de consumo futuro en una distribuidora de suplementos nutricionales con el fin implementar políticas que incrementen el nivel de ventas.

Objetivos específicos.

- Identificar reglas de consumo en base a características del consumidor de un determinado suplemento.
- Formar paquetes de productos de acuerdo a reglas de Consumo.
- Identificar aquellos productos que prefieren los clientes y aplicar Cross-Selling.
- Mostrar la probabilidad y la importancia de las reglas que determinan el consumo.

HIPÓTESIS.

El comportamiento de compra del cliente de SN de Lab Nutrition –Local Pardo que realizó consultas con nutricionistas de la tienda puede ser representado a través de reglas de consumo con una probabilidad de ocurrencia del 100%.

MÉTODO

TIPO Y DISEÑO DE INVESTIGACIÓN

Tipo de investigación

Exploratorio

Nos basaremos en la información provista por la distribuidora de suplementos, estos datos nos permitirán aplicar técnicas y procedimientos de Data Mining para lograr identificar reglas de compra de los clientes.

Se busca plantear el problema a través del uso de minería de datos, logrando una visión global de tipo aproximativo, el tema planteado ha sido poco explorado aunque empieza poco a poco a tomar más importancia en nuestro país.

Podemos decir que es un tema que cada día va tomando relevancia entre los empresarios, debido a que los ingresos que generan las ventas son vitales para la subsistencia del negocio. Es una investigación que involucra técnicas informáticas para conocer y predecir el comportamiento futuro que adoptará el cliente.

Con la tesis se busca crear nuevo conocimiento para la distribuidora con lo cual pueda establecer nuevas políticas que generen el incremento de las ventas.

Diseño de investigación

El diseño es no experimental de tipo transeccional o transversal correlacional-causal ya que en el estudio no se realiza manipulación intencional y las variables se relacionan entre sí para conocerse las reglas de consumo.³⁶

Para esto procederemos de la siguiente manera:

- Se extrae información de la empresa perteneciente al último semestre de operaciones.(Enero –Junio 2016)
- Se interpreta la información.
- Se clasifica la información relevante.
- Identificamos el modelo de minería de datos que mejor se ajuste a los datos en estudio.
- Se identifica la mejor herramienta de minería de datos que permita procesar los datos.

- Se obtienen resultados luego del procesamiento.
- Se interpreta resultados
- Se aplica cross-selling para determinar e identificar que productos pueden venderse en forma conjunta.
- Concluimos

Variables (Cliente)

- Edad
- Genero
- Ocupación
- Actividad Física
- Peso
- Estatura
- Estado civil
- Ingreso mensual
- Número de hijos

MUESTRA

Datos de clientes de la distribuidora de suplementos nutricionales-local Miraflores.

Son clientes (varones y mujeres) de entre 17 a 50 años que generan compras recurrentes de suplementos nutricionales en el local de Miraflores (Lima), identificaremos en ellos reglas que determinan sus patrones de compra sobre la base de su información histórica.

Los clientes de este local pertenecen al sector socioeconómico A y B, con una frecuencia de compra quincenal, mensual. La frecuencia de compra varía debido a que la distribuidora ofrece una gran gama de productos deportivos, accesorios y suplementos alimenticios usados en la práctica de deportes.

La mayoría de ellos ejercita sus cuerpos con regularidad y requiere consumir los suplementos alimenticios para poder lograr objetivos varios planteados (objetivos similares a los de los estudios Colombia, España).

Tomamos una muestra de 611 clientes que realizaron compras entre Enero del 2016 a Junio del 2016, sobre ellos se aplicarán técnicas de minería de datos, el total de registros de compras que generaron es de 12154 y corresponden a 3714 transacciones, cabe precisar que los 611 clientes seleccionados realizaron consultas con los nutricionistas durante este periodo.

Si consideramos la media de ventas diarias, al mes tendríamos 3000 ventas y en los 6 meses 18000, entonces nuestro universo son 18000 transacciones por lo que la muestra seleccionada de 3714 representa el 20.63%.

Por otro lado, si consideramos en número de clientes; tendríamos de igual forma 18000 clientes que acuden a comprar durante los 6 meses de operaciones.

El tamaño de la muestra seleccionada es adecuado, ya que, si hacemos uso de una calculadora estadística los valores que obtenemos son:

Con un nivel de confianza del 95% la muestra que se nos sugiere es de 377 registros y con un nivel de 98% la muestra que se nos sugiere es de 526, por lo cual, con 611 tenemos más que suficiente.

18000	TAMAÑO DEL UNIVERSO Número de personas que componen la población a estudiar.
50	HETEROGENEIDAD % Es la diversidad del universo. Lo habitual es usar 50%, el peor caso.
5	MARGEN DE ERROR % Menor margen de error requiere mayor muestra.
98	NIVEL DE CONFIANZA % Mayor nivel de confianza requiere mayor muestra. Lo habitual es entre 95% y 99%
526	MUESTRA Personas a encuestar

CALCULAR

Ilustración 8 Calculadora Estadística

<http://www.netquest.com/es/panel/calculadora-muestras/calculadoras-estadisticas.html>

En la tesis hacemos uso de la metodología ²⁰CRISP-DM , pero eso no nos impide citar a la metodología SEMMA ²⁰ , también usada en Minería de datos ,que nos sugiere la muestra a utilizar y dice: debe de ser lo suficientemente grande para contener la información significativa, pero lo suficientemente pequeña para poder procesarla rápidamente.³⁷

³⁷Por otro lado, para elegir una muestra significativa e imparcial (lo cual es fundamental para el análisis estadístico), se hace uso del principio de aleatoriedad a través de la técnica de muestreo simple al azar, (Phillips 1996, página 91).

Si bien con el crecimiento de hardware en el presente, algunos podrían llegar a pensar que es posible y recomendable procesar la totalidad de data que almacenamos, esto resulta en un mito: **“Un montón de hardware le permite evitar el muestreo”**.

Si bien las mejoras tecnológicas han sido sustanciales en los últimos años. En particular, espacio en disco y la memoria, sin embargo, las restricciones tecnológicas siguen siendo las siguientes:

- Ancho de banda de la red
- Rendimiento
- Gestión de la memoria
- Equilibrio de carga de las CPU.

Desde una perspectiva de los recursos necesarios, la minería de datos es el proceso de selección de Datos (ancho de banda de red y rendimiento), exploración de datos (memoria) y datos de modelado (Gestión de memoria y CPU) para descubrir información previamente desconocida para lograr ventaja competitiva.

INSTRUMENTOS DE INVESTIGACIÓN

Entrevistas a administradores y vendedores de la tienda

Se realizan entrevistas con el administrador y los vendedores de la tienda para lograr comprender el negocio en su totalidad, estas entrevistas a través de un conjunto de preguntas, nos permiten conocer aspectos operacionales y del negocio que nos sirve para conocer el negocio y poder plantear, modelar y dar solución al problema.

Procedimientos de recolección de datos.

Los datos aquí contenidos corresponden a clientes que realizaron compras en Lab Nutrition y además participaron de campañas médico- informativas de la tienda.

Tenemos acceso a los registros de ventas, comprendidos entre el periodo Enero 2016-Junio 2016, estos datos se nos facilitan como parte de una consultoría, los datos se nos entregaron en un archivo plano, se realizara limpieza sobre ellos y se estructura en una DB relacional SQL server

Para recabar esta información se hará uso de un único recurso, el tesista, en disponibilidad total pues estamos enfocados 100% al proyecto de tesis.

Procedimiento en el desarrollo de la tesis:

Depuramos los datos de la distribuidora y los alojamos en un motor de base de datos SQL Server 2014.

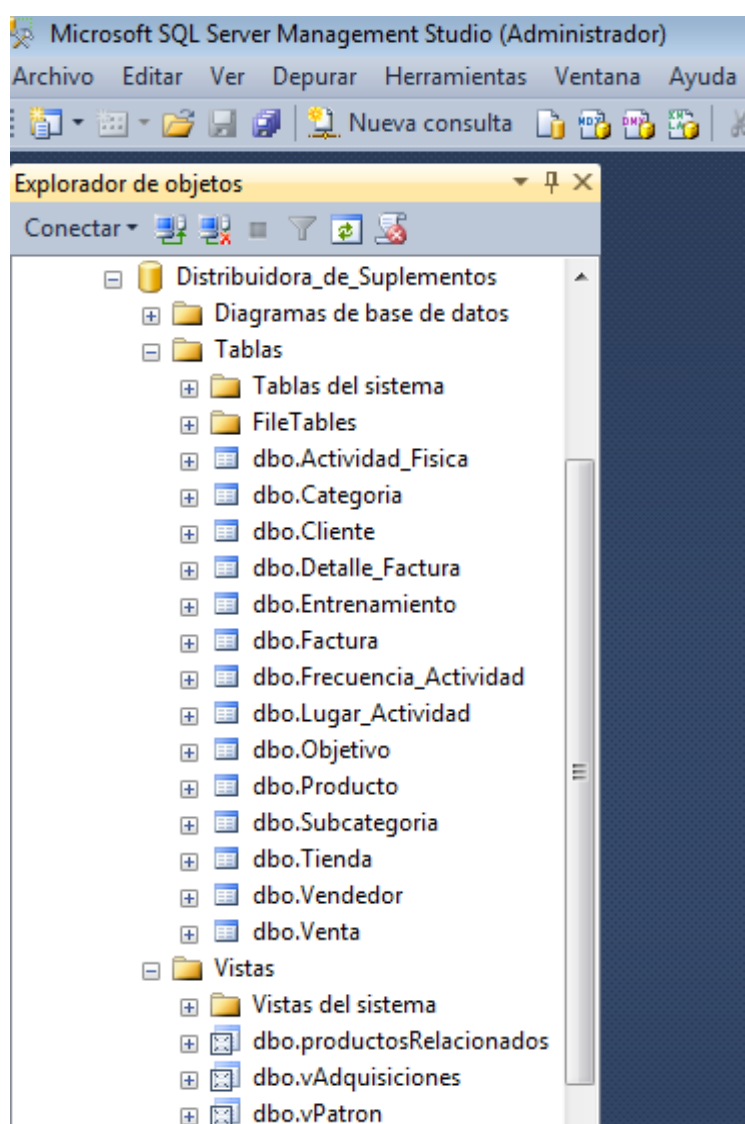


Ilustración 9 SQL Server Management Studio

Selección de registros :Tiendas

SQLQuery8.sql - LA...IGUEL\Mouses (62)) X LAPMIGUEL.Distrib...entos - dbo.Tienda SQLQuery7.sql - LA...IGUEL\Mouses (59))* SQLQuery6.sql - LA...IGUEL\Mouses (58))

/***** Script para el comando SelectTopNRows de SSMS *****/

100 %

Resultados Mensajes

	Tienda_Id	tienda_nombre	Direccion
1	1	TIENDA CHACARILLA (C.C. CAMINOS DEL INCA)	Calle Monterrey 258 - Tda. 104, Jirón Monterrey, Santiago de Surco 15038
2	2	TIENDA EL POLO	Av. La Encalada 1587 C.C. EL POLO / Dentro del Bodytech Lima, EL POLO Perú
3	3	TIENDA JOCKEY PLAZA	C.C. Jockey Plaza - Surco - Av. Javier Prado Este 4200 (2do. piso al costado de Ripley) Lima, Surco Perú
4	4	TIENDA MALL AVENTURA PLAZA - SANTA ANITA	C.C. Mall Aventura Plaza - Tda. 30038 (Av. Carretera Central 111) Lima, Santa Anita Perú
5	5	TIENDA ÓVALO GUTIERREZ (CINEPLANET ALCÁZAR)	AV SANTA CRUZ 814 TIENDA 04 SEGUNDO PISO - MIRAFLORES Lima, Miraflores Perú
6	6	TIENDA PARDO MIRAFLORES	Av. José Pardo 370 Lima, Miraflores Perú
7	7	TIENDA PLAZA LIMA SUR - CHORRILLOS	C.C. Plaza Lima Sur - Tda 211 (Av. Prol. Paseo de la República S/N) Lima, Chorrillos Perú
8	8	TIENDA PLAZA NORTE	C.C. Plaza Norte - Segundo Piso (Cruce Panamericana Norte - Tomás Valle - Túpac Amaru) Lima, Plaza Norte Perú
9	9	TIENDA SAN ANTONIO MIRAFLORES (C.C. PASO 28 DE J...	C.C. Jockey Plaza - Surco - Av. Javier Prado Este 4200 (2do. piso al costado de Ripley) Lima, Surco Perú
10	10	TIENDA SAN ANTONIO MIRAFLORES (C.C. PASO 28 DE J...	AV 28 DE JULIO Nº 1005, INTERIOR 205, URBANIZACION SAN ANTONIO - MIRAFLORES Lima, Miraflores Perú
11	11	TIENDA SAN BORJA	Av. Javier prado Este 1974 Lima, San Borja Perú
12	12	TIENDA SAN ISIDRO	Calle Choquehuanca 611 Lima, San Isidro 051 Perú
13	13	TIENDA SAN JUAN DE LURIGANCHO	Cuadra 1 Av. Gran Chimú, San Juan de Lurigancho Lima, San Juan de Lurigancho Perú
14	14	TIENDA SAN MIGUEL	C.C D Paso (Cruce Av. Universitaria con Av. La Mar) Lima, San Miguel Perú
15	15	TIENDA TRUJILLO	Mall Aventura Plaza (Av. América Oeste 750, Trujillo - Perú) Trujillo, La Libertad Perú

Ilustración 10 Tiendas

Selección de registros: Clientes

SQLQuery1.sql - LA...IGUEL\Mouses (53)) X

/***** Script para el comando SelectTopNRows de SSMS *****/

100 %

Resultados Mensajes

	Cliente_Id	Cliente_Nombre	Cliente_ApPaterno	Cliente_ApMaterno	Cliente_Direccion	Cliente_Ingreso_Mensual	Cliente_Estado_Civil	Cliente_Ocupacion
1	1	ABEL	MONTESINOS	MEDICO	Calle Ignacio Merino, 636	10000	C	Abogado
2	2	JIMENA MARIA	ZIMMERMANN	CAMPOS	JR Sucre 183	7000	S	Administradora
3	3	JORGE LUIS	MOTTA	PRENTICE	MA Amendariz 271	8000	C	Oficial
4	4	NOEMI	RODAS	FONSECA	Calle. Geman Gomez Sanchez 553	9500	S	Modelo
5	5	YOKO KATHERINE	KANNO	BARRANTES	Calle General Belisario Suarez 645	3500	S	Estudiante
6	6	SILVIA RUTH	CANO ALVA	TORRES	Calle Javier Fernandez 221	5500	C	Secretaria
7	7	LUCIO RAMIRO	CUEVA	MEJIA	AV Arica 320	5400	S	Asesor
8	8	CAROLINA	ALVAREZ	BISSO	CL General Miguel Iglesias 471	6500	S	Contador
9	9	JOHN ANTONNY	ABANTO	MALAYER	Av.2 De Mayo N° 1080	5900	S	Contador
10	10	VICTOR ARMANDO	AQUINO	SALCEDO	Jr General Mendiburu 740	5400	S	Veterinario
11	11	CATHERINE CAR...	ARRIETA	AVENDAÑO	AV Casimiro Ulloa 353	8400	S	Arquitecto
12	12	VIVIEN JOSEFINA	GELDRES	DE LASO	AV Jose Pardo 930	5400	S	Asesor
13	13	CHRISTIAN BRYAN	BJORK	OLIVARES	AV Reducto 1091	6400	S	Reportero
14	14	ALFONSO	BENAVIDES	CORREA	CL Santa Isabel 154 Urb. Miraflores	7200	C	Independiente
15	15	DULCE	VIVANCO	DEL CASTILLO ...	JR Tacna 276	6200	C	Arquitecta
16	16	ROXANA ZOILA	ANDRADE	MEDINA	CL Elias Aguirre 430	8200	C	Psicologa
17	17	JOSE ALBERTO	BACHMANN	TRISOLINI	AV Jose Larco 109	4800	S	Chef
18	18	TERESA ISABEL	OJEDA	FLORES DE BUT...	Jr Italia 248	8000	C	Independiente
19	19	GONZALO PASC...	PALACIOS	HERNANDEZ	Ca Geman Aparicio Gomez Sanc...	10000	D	Independiente
20	20	NANCI	MARCHENA	PALACIOS	AV Ricardo Palma 821 St. 201 Ur...	3000	S	Modelo
21	21	VICTOR DANIEL	NEGRON	PAZOS	Calle General Miguel Iglesias 482	5500	S	Arquitecto
22	22	ADOLFO	VATTUONE	CHAVEZ	Calle Jose Domingo Choquehuan...	2000	S	Estudiante

Ilustración 11: Clientes

Selección de registros : Ventas.

Microsoft SQL Server Management Studio (Administrador)

Herramientas Ventana Ayuda

SQLQuery2.sql - LA...IGUEL\Mouses (54)) x SQLQuery1.sql - LA...IGUEL\Mouses (53))

100 %

Resultados Mensajes

	Venta_Id	Venta_Transaccion	Venta_Fecha	Venta_Cliente_Id	Venta_Producto_Id	Venta_Vendedor_Id	Venta_Tienda_Id	Venta_Fact_Id
558	558	V00200	2016-01-02 19:45:12.0000000	189	10	448039345	6	1
559	559	V00200	2016-01-02 19:45:12.0000000	189	41	448039345	6	1
560	560	V00200	2016-01-02 19:45:12.0000000	189	75	448039345	6	1
561	561	V00201	2016-01-02 19:55:21.0000000	190	10	448039345	6	1
562	562	V00201	2016-01-02 19:55:21.0000000	190	41	448039345	6	1
563	563	V00202	2016-01-02 19:58:12.0000000	191	75	448039345	6	1
564	564	V00202	2016-01-02 19:58:12.0000000	191	10	448039345	6	1
565	565	V00202	2016-01-02 19:58:12.0000000	191	41	448039345	6	1
566	566	V00203	2016-01-02 19:59:54.0000000	192	75	448039345	6	1
567	567	V00204	2016-01-02 20:03:22.0000000	193	10	448039345	6	1
568	568	V00204	2016-01-02 20:03:22.0000000	193	41	448039345	6	1
569	569	V00205	2016-01-02 20:12:16.0000000	194	75	448039345	6	1
570	570	V00205	2016-01-02 20:12:16.0000000	194	10	448039345	6	1
571	571	V00206	2016-01-02 20:16:07.0000000	195	41	448039345	6	1
572	572	V00206	2016-01-02 20:16:07.0000000	195	75	448039345	6	1
573	573	V00206	2016-01-02 20:16:07.0000000	195	10	448039345	6	1
574	574	V00207	2016-01-02 20:25:22.0000000	196	41	448039345	6	1
575	575	V00207	2016-01-02 20:25:22.0000000	196	75	448039345	6	1

Ilustración 12 Ventas

Selección de registros: Productos

SQLQuery3.sql - LA...IGUEL\Mouses (55)) x SQLQuery2.sql - LA...IGUEL\Mouses (54)) SQLQuery1.sql - LA...IGUEL\Mouses (53))

100 %

Resultados Mensajes

	Producto_Id	Nombre_producto	Descripcion	Precio_unitario	Stock	Objetivo_Producto_ID	Categoria_Id
10	10	SYNTRAX ® NECTAR MEDICAL	La unica proteina de grado farmacéutico	110	1000	7	1
11	11	ON ® PLATINUM HYDROWHEY	La nueva evolución de la proteína platinum h...	275	1000	8	1
12	12	MET RX ® ULTRAMYOSYN WHEY ISOLATE	Evolución muscular asegurada	159	1000	1	1
13	13	SYNTRAX ® NECTAR GRAB N GO	La proteína más fina y dietética del mercado	60	1000	1	1
14	14	MUSCLETECH ® PLATINUM 100% ISO WHEY	La evolucion de la proteína aislada	189	1000	1	1
15	15	SYNTRAX ® NECTAR	La proteína más fina y dietética del mercado	189	1000	1	1
16	16	MUSCLETECH ® NITROTECH PERFORMANCE SERIES	Hecho para construir músculos	69	1000	1	1
17	17	MUSCLEMEDS ® CARNIVOR	La primera proteína aislada de carne	260	1000	1	1
18	18	MUSCLETECH ® PLATINUM 100% BEEF PROTEIN	La evolucion de la proteína de carne	249	1000	8	1
19	19	SYNTRAX ® MICELLAR	La caseína perfecta para bajar de peso	175	1000	1	2
20	20	MUSCLETECH ® PLATINUM 100% CASEIN	La evolucion de la caseína	199	1000	8	2
21	21	MUSCLETECH ® HALO PERFORMANCE	Crecimiento, fuerza y resistencia, todo en uno	189	1000	1	3
22	22	MUSCLEPHARM ® COMBAT	Alimenta tus músculos por 8 horas	199	1000	1	3
23	23	SYNTRAX ® ESSENCE	Mezcla perfecta para tu dieta	149	1000	1	3
24	24	MUSCLETECH ® PHASE 8	8 horas de alimentación para tus músculos	179	1000	8	3
25	25	SYNTRAX ® MATRIX	Equilibrio perfecto en proteínas antes, durant...	69	1000	1	3
26	26	MUSCLETECH ® HYDROXYCUT Hardcore NEXT GEN	Baja de peso de manera intensa	199	1000	2	4
27	27	MUSCLETECH ® HYDROXYCUT Hardcore ELITE POWDER	Fórmula pre-entreno ¡quema grasa a gran esc...	139	1000	2	4
28	28	MUSCLETECH ® HYDROXYCUT SX-7 BLACK ONYX	Obtén una experiencia sensorial potente	249	1000	8	4
29	29	MUSCLETECH ® HYDROXYCUT Hardcore ELITE CAPS	Fórmula n° 1 para bajar de peso	199	1000	2	4
30	30	MUSCLETECH ® HYDROXYCUT SX-7	El quemador de grasa de última generación	159	1000	2	4
31	31	MUSCLETECH ® HYDROXYCUT FAT BURNER	Quema la grasa y aumenta tu energía	89	1000	2	4

Ilustración 13: Productos

Selección de registros : Categoría.

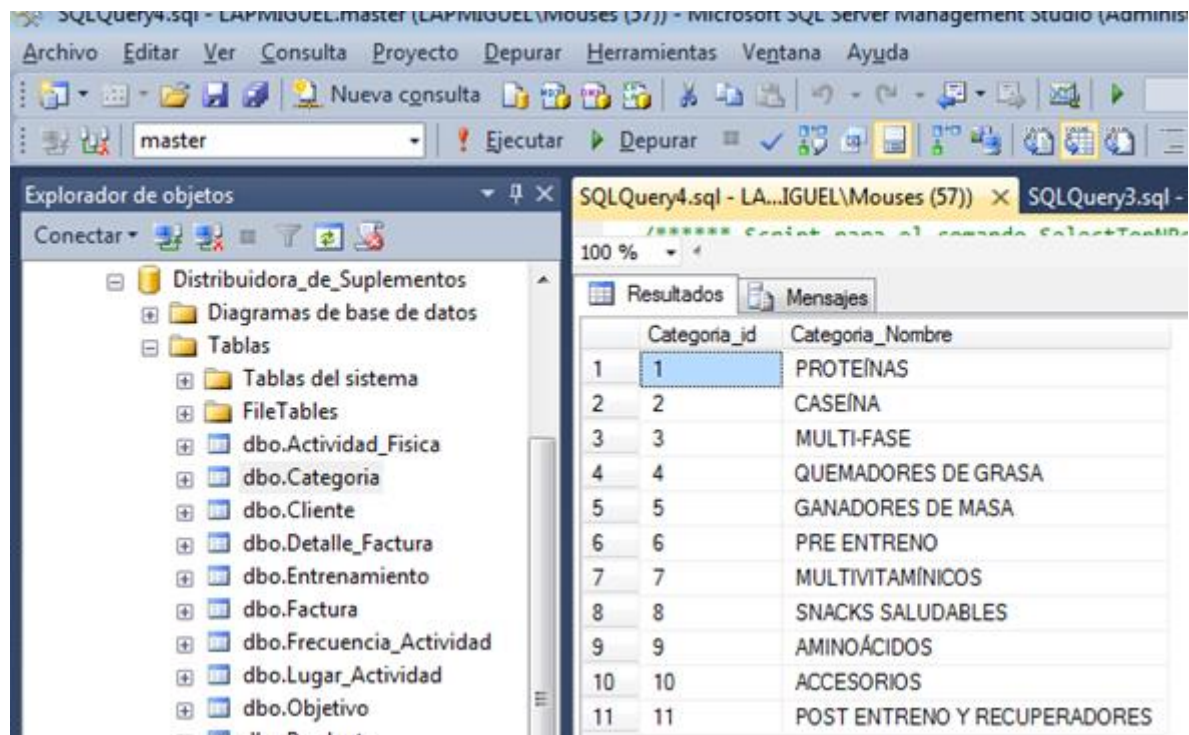
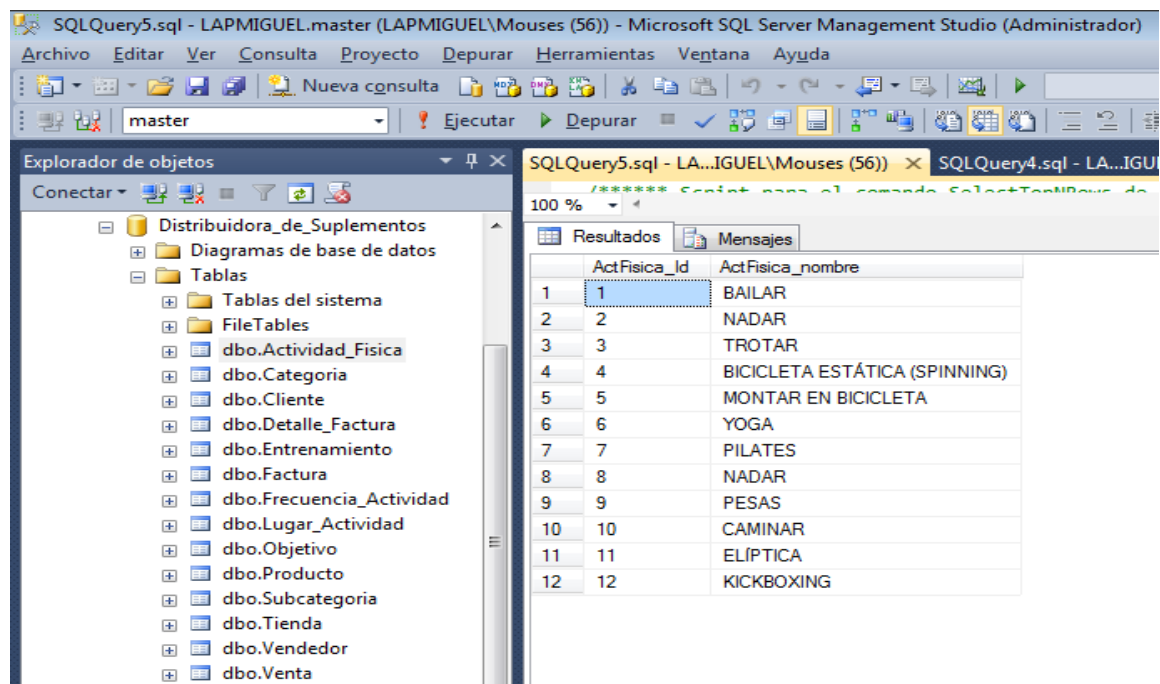


Ilustración 14: Categoría



Selección de Registros :Actividad_Física

Ilustración 15: Actividad Física

Selección de registros : Objetivo

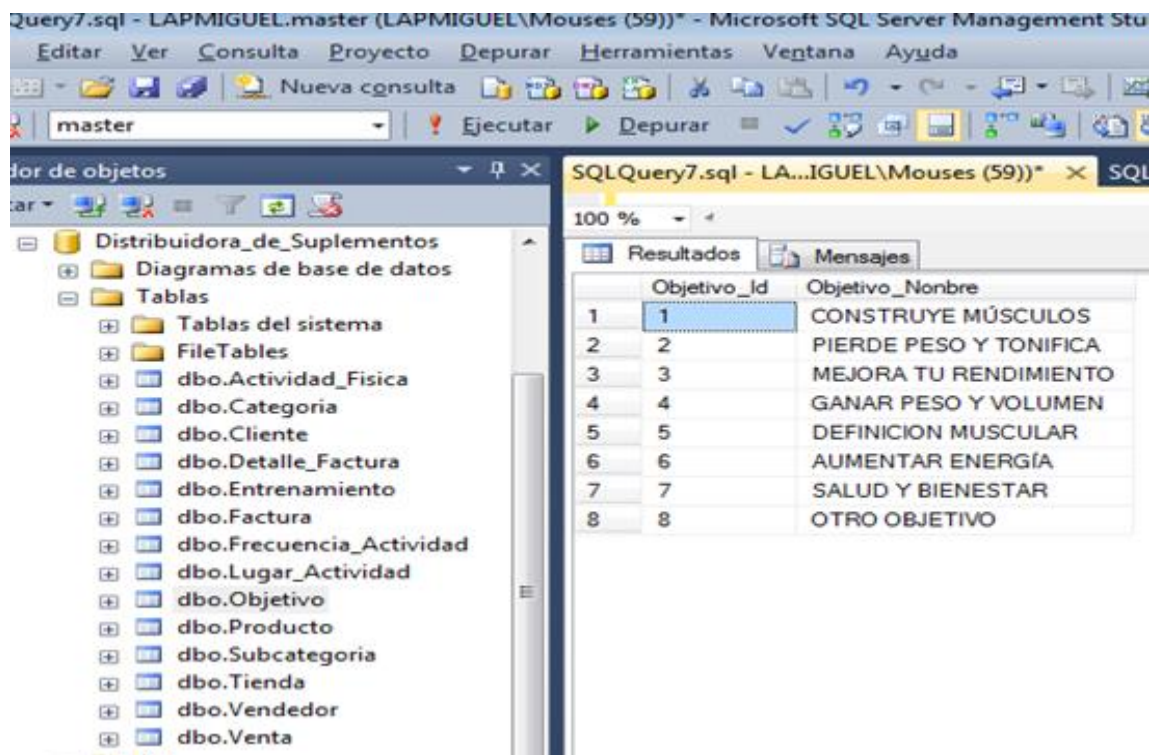


Ilustración 16: Objetivos

En Visual Studio : Creamos un proyecto de Analysis Services Multidimensional y Data Mining.

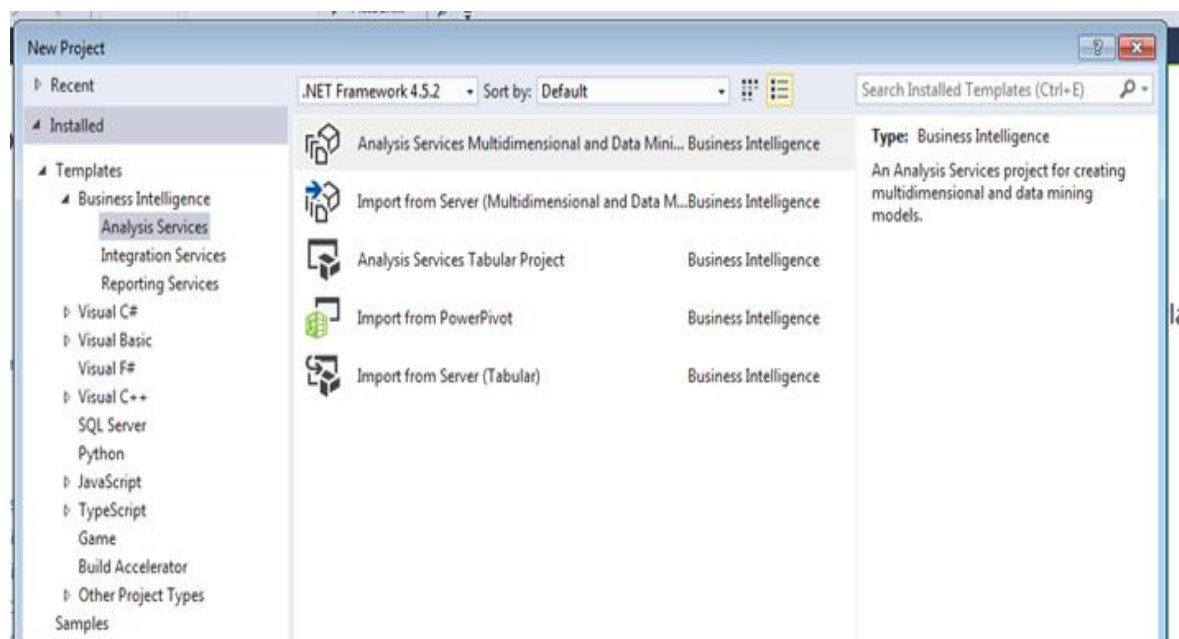


Ilustración 17: Visual Studio

Realizamos la conexión con la base de datos SQL server ,creamos una fuente de Datos “Data Source” , una vista de fuente de datos “ Data Source View” y una estructura de Minería de Datos “Mining Structure”.

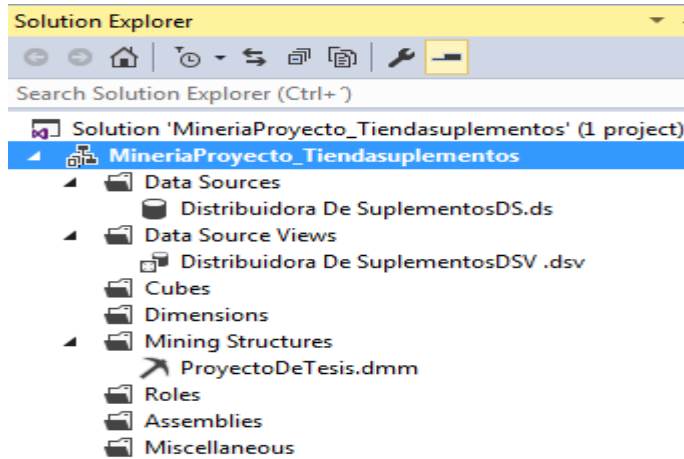


Ilustración 18: Conexión a DB

Creamos nuestro modelo de Minería de Datos.

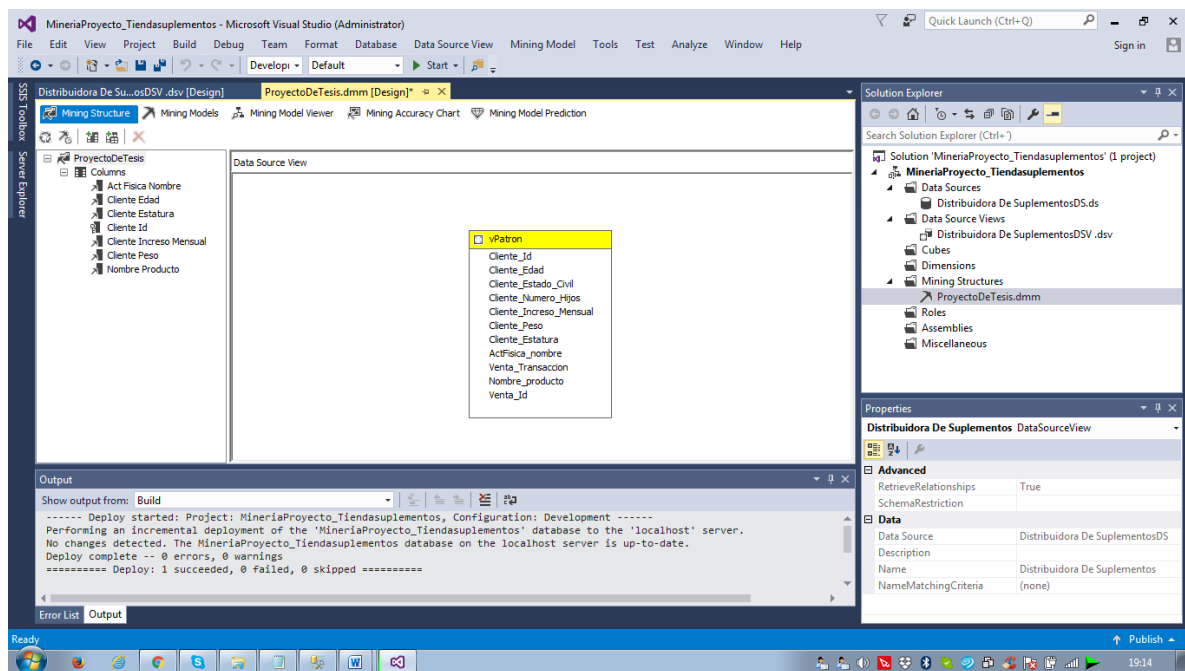


Ilustración 19: Modelo Minería de Datos

Hacemos uso de los algoritmos que nos provee la herramienta: Asociacion , Red Neuronal, Cluster .

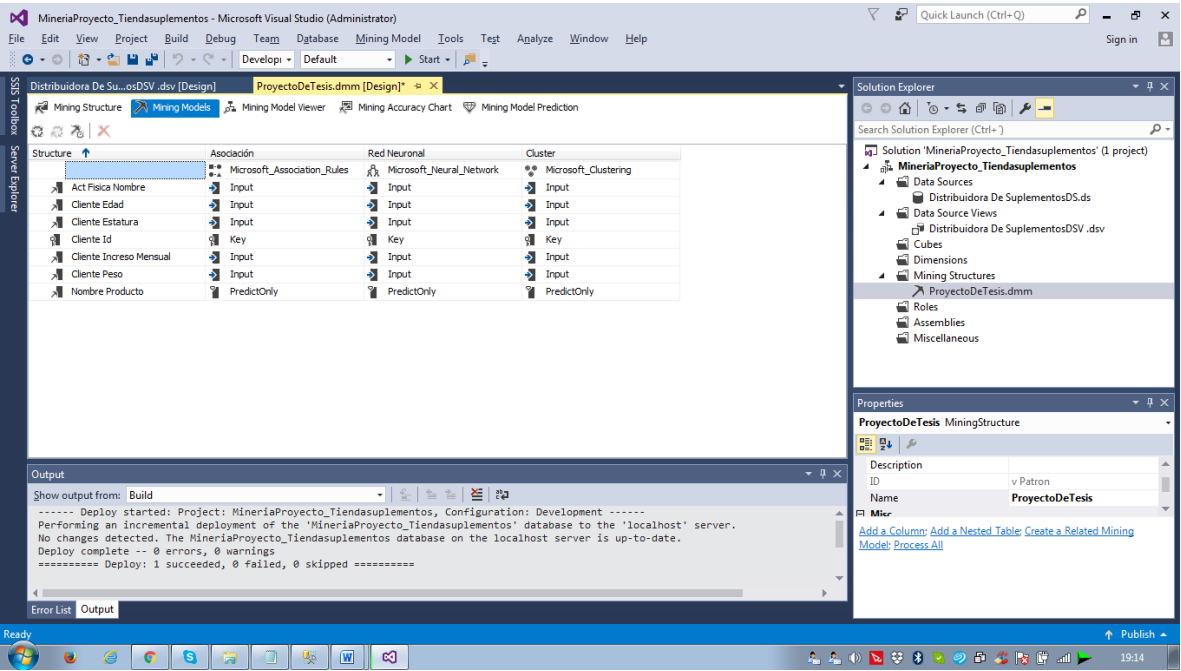


Ilustración 20: Algoritmos de la Herramienta.

Resultados del algoritmo de Asociacion .

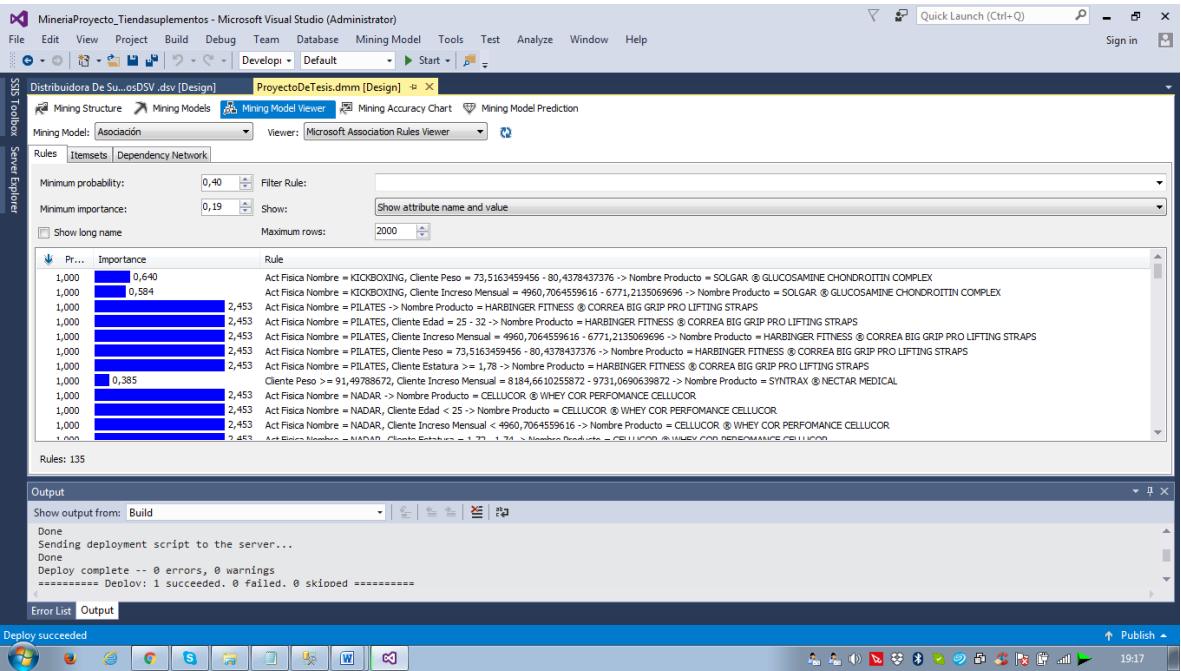


Ilustración 21: Algoritmo Asociación

ITEM SETS.

Mining Model: Asociación Viewer: Microsoft Association Rules Viewer

Rules Itemsets Dependency Network

Minimum support: 1 Filter Itemset:

Minimum itemset size: 0 Show: Show attribute name and value

Maximum rows: 2000 ☐ Show long name

Support	Size	Itemset
248	1	Act Fisica Nombre = BAILAR
230	1	Cliente Edad = 39 - 43
211	1	Cliente Estatura >= 1,78
211	1	Cliente Peso = 73,5163459456 - 80,4378437376
175	1	Cliente Ingreso Mensual = 6771,2135069696 - 8184,6610255872
142	2	Cliente Peso = 73,5163459456 - 80,4378437376, Cliente Estatura >= 1,78
142	2	Cliente Edad = 39 - 43, Act Fisica Nombre = BAILAR
129	2	Cliente Estatura >= 1,78, Act Fisica Nombre = BAILAR
128	2	Cliente Peso = 73,5163459456 - 80,4378437376, Cliente Edad = 39 - 43
125	2	Cliente Ingreso Mensual = 6771,2135069696 - 8184,6610255872, Cliente Edad = 39 - 43
124	2	Cliente Peso = 73,5163459456 - 80,4378437376, Act Fisica Nombre = BAILAR
120	2	Cliente Estatura >= 1,78, Cliente Edad = 39 - 43
117	1	Nombre Producto = SYNTRAX ® NECTAR MEDICAL
112	1	Cliente Peso = 65,3905212032 - 73,5163459456
105	2	Cliente Ingreso Mensual = 6771,2135069696 - 8184,6610255872, Act Fisica Nombre = BAILAR
102	2	Cliente Ingreso Mensual = 6771,2135069696 - 8184,6610255872, Cliente Estatura >= 1,78
100	1	Cliente Ingreso Mensual = 4960,7064559616 - 6771,2135069696
100	2	Cliente Ingreso Mensual = 6771,2135069696 - 8184,6610255872, Cliente Peso = 73,5163459456 - 80,4378437376

Ilustración 22: ITEM Sets

RESULTADO RED NEURONAL:

ProyectoDeTesis.dmm [Design] Mining Structure Mining Models Mining Model Viewer Mining Accuracy Chart Mining Model Prediction

Mining Model: Red Neuronal Viewer: Microsoft Neural Network Viewer

Input:

Attribute	Value
<All>	

Output:

Output Attribute: Nombre Producto

Value 1: SYNTRAX ® MATRIX

Value 2: MUSCLETECH ® HYDROXYCUT PRO CLINI

Variables:

Attribute	Value	Favors SYNTRAX ® MATRIX	Favors MUSCLETECH ® HYDROXYCU...
Act Fisica Nombre	PESAS		
Act Fisica Nombre	PILATES		
Cliente Edad	32 - 39		
Cliente Edad	< 25		
Cliente Ingreso Mensual	>= 9731,0690639872		
Cliente Ingreso Mensual	< 4960,7064559616		
Cliente Edad	25 - 32		
Act Fisica Nombre	KICKBOXING		
Act Fisica Nombre	BICICLETA ESTÁTICA (SPINNING)		
Cliente Peso	80,4378437376 - 91,49788672		
Act Fisica Nombre	BAILAR		
Cliente Estatura	1,72 - 1,74		
Cliente Edad	>= 43		
Cliente Estatura	< 1,72		

Ilustración 23: Red Neuronal

RESULTADO CLUSTER:

Diagrama de Clúster:

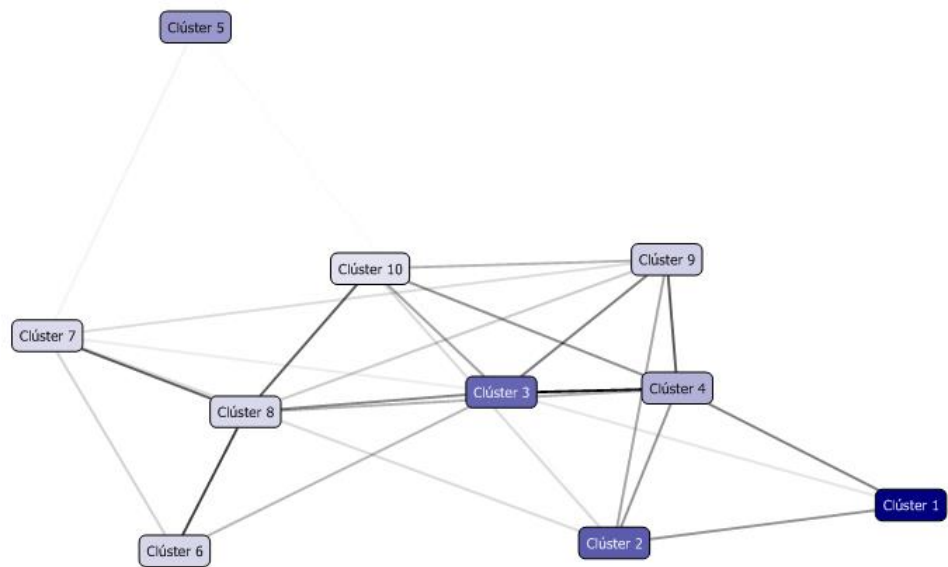


Ilustración 24: Diagrama de Clúster

Perfil de clúster creados:

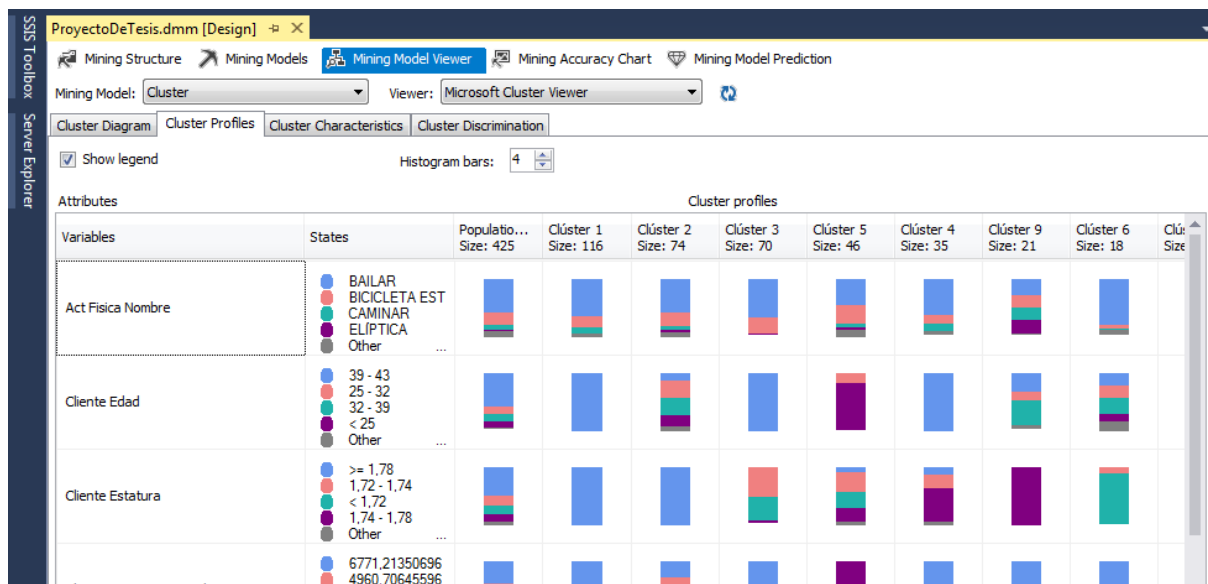


Ilustración 25: Perfil de clúster

Características del clúster

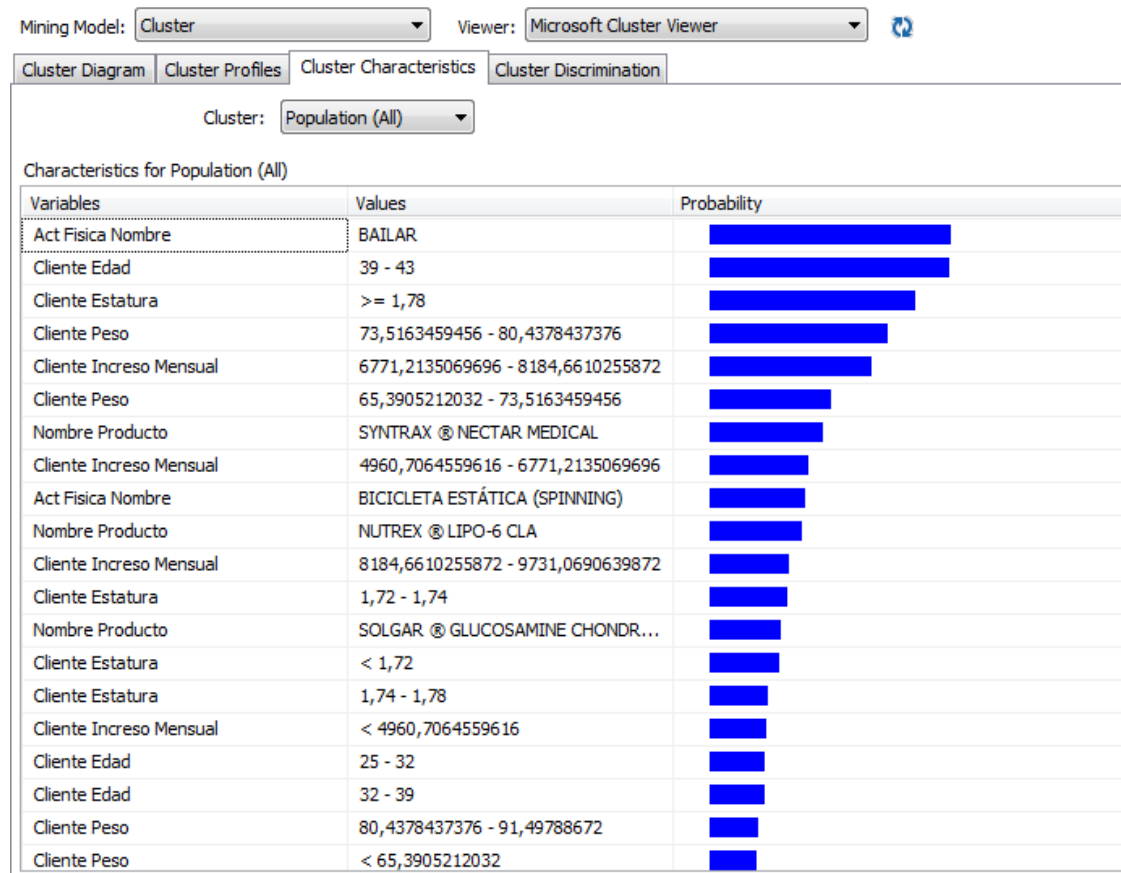


Ilustración 26: Características de Clúster

RESULTADOS

Presentación de resultados

TABLA 23 PATRÓN DE CONSUMIDORES DE HARBINGER FITNESS® CORREA BIG GRIP PRO LIFTING STRAPS

#	<u>Probabilidad</u>	<u>Importancia</u>	<u>Regla</u>
1	1,000	2,45	Act Física Nombre = PILATES, Cliente Edad = 25 - 32 -> Nombre Producto = HARBINGER FITNESS® CORREA BIG GRIP PRO LIFTING STRAPS
2	1,000	2,45	Act Física Nombre = PILATES, Cliente Ingreso Mensual = 4960,7-6771,2-> Nombre Producto = HARBINGER FITNESS® CORREA BIG GRIP PRO LIFTING STRAPS
3	1,000	2,45	Act Física Nombre = PILATES, Cliente Estatura >= 1,78 -> Nombre Producto = HARBINGER FITNESS® CORREA BIG GRIP PRO LIFTING STRAPS
4	1,000	2,45	Act Física Nombre = PILATES, Cliente Peso = 73,52 - 80,48 -> Nombre Producto = HARBINGER FITNESS® CORREA BIG GRIP PRO LIFTING STRAPS

TABLA 24 PATRÓN DE CONSUMIDORES DE SOLGAR® GLUCOSAMINE CHONDROITIN COMPLEX

Patrón de consumidores de SOLGAR® GLUCOSAMINE CHONDROITIN COMPLEX

#	<u>Probabilidad</u>	<u>Importancia</u>	<u>Regla</u>
1	1,000	0,66	Act Física Nombre = KICKBOXING, Cliente Peso = 73,52-80,44 -> Nombre Producto = SOLGAR® GLUCOSAMINE CHONDROITIN COMPLEX
2	1,000	0,62	Act Física Nombre = KICKBOXING, Cliente Estatura >= 1,78 -> Nombre Producto = SOLGAR® GLUCOSAMINE CHONDROITIN COMPLEX

3	1,000	0,62	Act Física Nombre = PESAS, Cliente Estatura >= 1,78 -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
4	1,000	0,62	Act Física Nombre = PESAS, Cliente Peso = 73,5163459456 - 80,4378437376 -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
5	1,000	0,62	Act Física Nombre = KICKBOXING, Cliente Numero Hijos < 2 - > Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
6	1,000	0,62	Act Física Nombre = PESAS, Cliente Edad = 39 - 43 -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
7	1,000	0,62	Act Física Nombre = PESAS, Cliente Estado Civil = C -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
8	1,000	0,57	Act Física Nombre = ELÍPTICA, Cliente Edad >= 43 -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
9	1,000	0,57	Act Física Nombre = KICKBOXING, Cliente Ingreso Mensual >= 9731,0690639872 -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
10	1,000	0,57	Act Física Nombre = PESAS, Cliente Ingreso Mensual = 8184,6610255872 - 9731,0690639872 -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
11	1,000	0,57	Act Física Nombre = PESAS, Cliente Ingreso Mensual = 6771,2135069696 - 8184,6610255872 -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
12	1,000	0,57	Act Física Nombre = KICKBOXING, Cliente Sexo = F -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX

13	1,000	0,57	Act Física Nombre = PESAS, Cliente Sexo = F -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
14	1,000	0,57	Act Física Nombre = PESAS, Cliente Numero Hijos = 2 - 3 -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
15	1,000	0,57	Act Física Nombre = KICKBOXING, Cliente Ingreso Mensual = 6771,2135069696 - 8184,6610255872 -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
16	1,000	0,57	Act Física Nombre = ELÍPTICA, Cliente Estatura < 1,72 -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
17	1,000	0,57	Act Física Nombre = KICKBOXING, Cliente Estatura = 1,74 - 1,78 -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
18	1,000	0,57	Act Física Nombre = KICKBOXING, Cliente Peso = 80,4378437376 - 91,49788672 -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
19	1,000	0,57	Act Física Nombre = KICKBOXING, Cliente Ingreso Mensual = 4960,7064559616 - 6771,2135069696 -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
20	1,000	0,57	Act Física Nombre = KICKBOXING, Cliente Edad = 32 - 39 -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
21	1,000	0,57	Act Física Nombre = ELÍPTICA, Cliente Peso >= 91,49788672 -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX

TABLA 25 PATRÓN DE CONSUMIDORES DE NUTREX ® LIPO-6 CLA

#	<u>Probabilidad</u>	<u>Importancia</u>	<u>Regla</u>
1	1,000	0,49	Act Física Nombre = ELÍPTICA, Cliente Sexo = F -> Nombre Producto = NUTREX ® LIPO-6 CLA
2	1,000	0,49	Cliente Estado Civil = D -> Nombre Producto = NUTREX ® LIPO-6 CLA
3	1,000	0,49	Cliente Estado Civil = D, Cliente Numero Hijos >= 3 -> Nombre Producto = NUTREX ® LIPO-6 CLA
4	1,000	0,49	Cliente Estado Civil = D, Cliente Edad = 32 - 39 -> Nombre Producto = NUTREX ® LIPO-6 CLA
5	1,000	0,49	Cliente Estado Civil = D, Cliente Ingreso Mensual = 6771,2-8184,7 -> Nombre Producto = NUTREX ® LIPO-6 CLA
6	1,000	0,49	Cliente Estado Civil = D, Cliente Estatura >= 1,78 -> Nombre Producto = NUTREX ® LIPO-6 CLA
7	1,000	0,49	Cliente Estado Civil = D, Cliente Peso = 73,5- 80,4-> Nombre Producto = NUTREX ® LIPO-6 CLA
8	1,000	0,49	Cliente Estado Civil = D, Cliente Sexo = M -> Nombre Producto = NUTREX ® LIPO-6 CLA
9	1,000	0,49	Cliente Estado Civil = D, Act Física Nombre = BAILAR -> Nombre Producto = NUTREX ® LIPO-6 CLA
10	1,000	0,49	Act Física Nombre = PILATES, Cliente Peso = 80,4 - 91,5-> Nombre Producto = NUTREX ® LIPO-6 CLA
11	1,000	0,49	Act Física Nombre = PILATES, Cliente Edad < 25 -> Nombre Producto = NUTREX ® LIPO-6 CLA

12	1,000	0,49	Act Física Nombre = PILATES, Cliente Ingreso Mensual < 4960,7- > Nombre Producto = NUTREX ® LIPO-6 CLA
13	1,000	0,49	Act Física Nombre = PILATES, Cliente Estatura = 1,74 - 1,78 -> Nombre Producto = NUTREX ® LIPO-6 CLA
14	1,000	0,49	Cliente Peso >= 91,5, Cliente Ingreso Mensual = 4960,7- 6771,2 - > Nombre Producto = NUTREX ® LIPO-6 CLA

TABLA 26 PATRÓN DE CONSUMIDORES DE SYNTRAX ® NECTAR MEDICAL

<u>#</u>	<u>Probabilidad</u>	<u>Importancia</u>	<u>Regla</u>
1	1,000	0,52	Cliente Numero Hijos >= 3, Act Física Nombre = BICICLETA ESTÁTICA (SPINNING) -> Nombre Producto = SYNTRAX ® NECTAR MEDICAL
2	1,000	0,48	Cliente Numero Hijos >= 3, Cliente Estatura < 1,72 -> Nombre Producto = SYNTRAX ® NECTAR MEDICAL
3	1,000	0,45	Cliente Numero Hijos >= 3, Cliente Peso = 80,4 - 91,5-> Nombre Producto = SYNTRAX ® NECTAR MEDICAL
4	1,000	0,39	Cliente Numero Hijos >= 3, Act Física Nombre = CAMINAR -> Nombre Producto = SYNTRAX ® NECTAR MEDICAL
5	1,000	0,39	Cliente Numero Hijos >= 3, Cliente Estado Civil = S -> Nombre Producto = SYNTRAX ® NECTAR MEDICAL
6	1,000	0,39	Cliente Peso >= 91,49788672, Cliente Sexo = F -> Nombre Producto = SYNTRAX ® NECTAR MEDICAL
7	1,000	0,39	Cliente Edad < 25, Cliente Ingreso Mensual = 6771,2- 8184,7 - > Nombre Producto = SYNTRAX ® NECTAR MEDICAL

**TABLA 27 PATRÓN DE CONSUMIDORES DE CELLUCOR ® WHEY COR
PERFOMANCE CELLUCOR**

#	<u>Probabilidad</u>	<u>Importancia</u>	<u>Regla</u>
1	1,000	2,45	Act Física Nombre = NADAR -> Nombre Producto = CELLUCOR ® WHEY COR PERFOMANCE CELLUCOR
2	1,000	2,45	Act Física Nombre = NADAR, Cliente Edad < 25 -> Nombre Producto = CELLUCOR ® WHEY COR PERFOMANCE CELLUCOR
3	1,000	2,45	Act Física Nombre = NADAR, Cliente Ingreso Mensual < 4960,7064559616 -> Nombre Producto = CELLUCOR ® WHEY COR PERFOMANCE CELLUCOR
4	1,000	2,45	Act Física Nombre = NADAR, Cliente Estatura = 1,72 - 1,74 -> Nombre Producto = CELLUCOR ® WHEY COR PERFOMANCE CELLUCOR
5	1,000	2,45	Act Física Nombre = NADAR, Cliente Peso = 65,4- 73,5-> Nombre Producto = CELLUCOR ® WHEY COR PERFOMANCE CELLUCOR
6	1,000	2,45	Act Física Nombre = NADAR, Cliente Estado Civil = S -> Nombre Producto = CELLUCOR ® WHEY COR PERFOMANCE CELLUCOR
7	1,000	2,45	Act Física Nombre = NADAR, Cliente Numero Hijos < 2 -> Nombre Producto = CELLUCOR ® WHEY COR PERFOMANCE CELLUCOR
8	1,000	2,45	Act Física Nombre = NADAR, Cliente Sexo = F -> Nombre Producto = CELLUCOR ® WHEY COR PERFOMANCE CELLUCOR

TABLA 28 PATRÓN DE CONSUMIDORES DE SMART SHAKE ® SMART SHAKE JAY CUTLER EDITION

#	<u>Probabilidad</u>	<u>Importancia</u>	<u>Regla</u>
1	1,000	2,15	Cliente Peso >= 91,5, Cliente Edad = 32 - 39 -> Nombre Producto = SMART SHAKE ® SMART SHAKE JAY CUTLER EDITION
2	1,000	2,15	Cliente Peso >= 91,5, Cliente Numero Hijos = 2 - 3 -> Nombre Producto = SMART SHAKE ® SMART SHAKE JAY CUTLER EDITION

TABLA 29 PATRÓN DE CONSUMIDORES DE MUSCLETECH ® PLATINUM 100% WHEY

#	<u>Probabilidad</u>	<u>Importancia</u>	<u>Regla</u>
1	1,000	1,31	Act Física Nombre = ELÍPTICA, Cliente Ingreso Mensual >= 9731 -> Nombre Producto = MUSCLETECH ® PLATINUM 100% WHEY
2	1,000	1,31	Cliente Peso >= 91,5, Act Física Nombre = CAMINAR -> Nombre Producto = MUSCLETECH ® PLATINUM 100% WHEY
3	1,000	1,31	Act Física Nombre = PESAS, Cliente Ingreso Mensual < 4960,7 -> Nombre Producto = MUSCLETECH ® PLATINUM 100% WHEY
4	1,000	1,31	Act Física Nombre = PESAS, Cliente Edad = 25 - 32 -> Nombre Producto = MUSCLETECH ® PLATINUM 100% WHEY
5	1,000	1,31	Act Física Nombre = PESAS, Cliente Estatura = 1,74 - 1,78 -> Nombre Producto = MUSCLETECH ® PLATINUM 100% WHEY
6	1,000	1,31	Act Física Nombre = PESAS, Cliente Peso = 65,3905212032 - 73,5 -> Nombre Producto = MUSCLETECH ® PLATINUM 100% WHEY
7	1,000	1,31	Act Física Nombre = PESAS, Cliente Estado Civil = S -> Nombre Producto = MUSCLETECH ® PLATINUM 100% WHEY

TABLA 30 PATRÓN DE CONSUMIDORES DE MUSCLETECH ® CLEAR MUSCLE

#	<u>Probabilidad</u>	<u>Importancia</u>	<u>Regla</u>
1	1,000	1,85	Act Física Nombre = KICKBOXING, Cliente Peso = 65,39- 73,52 -> Nombre Producto = MUSCLETECH ® CLEAR MUSCLE

TABLA 31 PATRÓN DE CONSUMIDORES DE MET RX ® LIQUID L-CARNITINE 1500

#	<u>Probabilidad</u>	<u>Importancia</u>	<u>Regla</u>
<u>1</u>	1,000	2,15	Cliente Ingreso Mensual >= 9731,1, Cliente Edad < 25 -> Nombre Producto = MET RX ® LIQUID L-CARNITINE 1500

TABLA 32 PATRÓN DE CONSUMIDORES DE HARBINGER FITNESS ® GUANTES POWER SERIES

#	<u>Probabilidad</u>	<u>Importancia</u>	<u>Regla</u>
<u>1</u>	1,000	1,98	Cliente Edad < 25, Cliente Ingreso Mensual = 8184,66- 9731,07- > Nombre Producto = HARBINGER FITNESS ® GUANTES POWER SERIES

TABLA 33 PATRÓN DE CONSUMIDORES DE MET RX ® ULTRAMYOSYN WHEY

#	<u>Probabilidad</u>	<u>Importancia</u>	<u>Regla</u>
<u>1</u>	1,000	2,15	Cliente Ingreso Mensual < 4960,71, Cliente Edad = 39 - 43 -> Nombre Producto = MET RX ® ULTRAMYOSYN WHEY

HACIENDO USO DEL MODELO NEURONAL

Con este modelo podemos descubrir el perfil del consumidor por cada uno de los productos comercializados a modo de ejemplo seleccionamos algunos de ellos.

Tabla 34 RESULTADOS MODELO NEURONAL SYNTRAX ® MATRIX

ATRIBUTO	VALOR	Favors SYNTRAX ® MATRIX
CLIENTE EDAD	< 25	69,68
CLIENTE ESTADO CIVIL	S	64,93
CLIENTE NÚMERO HIJOS	2 – 3	43,02
CLIENTE PESO	80,4 - 91,4	41,36
ACT FÍSICA NOMBRE	ELÍPTICA	39,36
CLIENTE INGRESO MENSUAL	< 4960,7	36,99
ACT FÍSICA NOMBRE	KICKBOXING	36,21
ACT FÍSICA NOMBRE	BICICLETA ESTÁTICA (SPINNING)	23,95
ACT FÍSICA NOMBRE	PESAS	20,83
CLIENTE ESTATURA	>= 1,78	20,01
CLIENTE INGRESO MENSUAL	>= 9731	15,81

CLIENTE PESO	65,4 - 73,5	11,28
CLIENTE PESO	>= 91,5	9,27
CLIENTE ESTADO CIVIL	D	5,34
CLIENTE SEXO	M	1,64
CLIENTE ESTATURA	1,74 - 1,78	0,69

Tabla 35 RESULTADOS MODELO NEURONAL MUSCLETECH ® HYDROXYCUT PRO CLINICAL GUMMIES

ATRIBUTO	VALOR	MUSCLETECH ® HYDROXYCUT PRO CLINICAL GUMMIES
CLIENTE NUMERO HIJOS	>= 3	100
CLIENTE ESTATURA	< 1,72	65,71
ACT FÍSICA NOMBRE	BAILAR	64,49
CLIENTE EDAD	>= 43	55,22
CLIENTE NUMERO HIJOS	< 2	39,83
CLIENTE ESTADO CIVIL	C	35,08
ACT FÍSICA NOMBRE	NADAR	28,78
ACT FÍSICA NOMBRE	MONTAR EN BICICLETA	28,04
CLIENTE PESO	< 65	23,69
ACT FÍSICA NOMBRE	CAMINAR	22,94
CLIENTE EDAD	25 – 32	20,07
CLIENTE EDAD	32 – 39	19,96
ACT FÍSICA NOMBRE	PILATES	10,47
CLIENTE EDAD	39 – 43	8,76
CLIENTE INGRESO MENSUAL	4961- 6771	7,47
CLIENTE SEXO	F	3,62
CLIENTE ESTATURA	1,72 - 1,74	3,51
CLIENTE PESO	73,5 - 80,4	2,97
CLIENTE INGRESO MENSUAL	8184,7 - 9731,1	0,08
CLIENTE INGRESO MENSUAL	6771- 8185	0,01

Tabla 36 RESULTADOS MODELO NEURONAL MUSCLETECH ® PLATINUM 100% WHEY

ATRIBUTO	VALOR	FAVORS MUSCLETECH ® PLATINUM 100% WHEY
CLIENTE EDAD	25 – 32	100
CLIENTE ESTADO CIVIL	S	62,64
ACT FÍSICA NOMBRE	MONTAR EN BICICLETA	61,62
ACT FÍSICA NOMBRE	NADAR	54,83
CLIENTE ESTATURA	>= 1,78	50,39
CLIENTE ESTADO CIVIL	D	35,7
CLIENTE ESTATURA	1,72 - 1,74	30,61
CLIENTE INGRESO MENSUAL	>= 9731	26,36
ACT FÍSICA NOMBRE	BICICLETA ESTÁTICA (SPINNING)	21,96
CLIENTE INGRESO MENSUAL	8184,7 - 9731,1	19,82
CLIENTE PESO	73,5 - 80,4	15,7
CLIENTE INGRESO MENSUAL	6771 - 8184,6	14,69
CLIENTE NUMERO HIJOS	2 - 3	14,02
CLIENTE NUMERO HIJOS	< 2	13,89
CLIENTE PESO	80,4 - 91,5	13,36
CLIENTE PESO	65 - 73	10,53
ACT FÍSICA NOMBRE	BAILAR	8
ACT FÍSICA NOMBRE	CAMINAR	7,3
CLIENTE INGRESO MENSUAL	4961- 6771	6,55
CLIENTE EDAD	32 - 39	2,67

Tabla 37 RESULTADOS MODELO NEURONAL NUTREX ® LIPO-6 BLACK

ATRIBUTO	VALOR	FAVORS NUTREX ® LIPO-6 BLACK
ACT FÍSICA NOMBRE	PESAS	81,97
CLIENTE NUMERO HIJOS	>= 3	62,46
CLIENTE ESTATURA	< 1,72	56,59
CLIENTE ESTATURA	1,74 - 1,78	53,91
ACT FÍSICA NOMBRE	ELÍPTICA	51,13
ACT FÍSICA NOMBRE	KICKBOXING	48,77
CLIENTE INGRESO MENSUAL	< 4961	23,91
CLIENTE PESO	< 65	21,91
CLIENTE ESTADO CIVIL	C	17,77
CLIENTE EDAD	39 - 43	13,96
ACT FÍSICA NOMBRE	PILATES	10,12

CLIENTE EDAD	< 25	5,56
CLIENTE EDAD	>= 43	3,28
CLIENTE PESO	>= 91	1,23

ALGORITMO CLUSTER

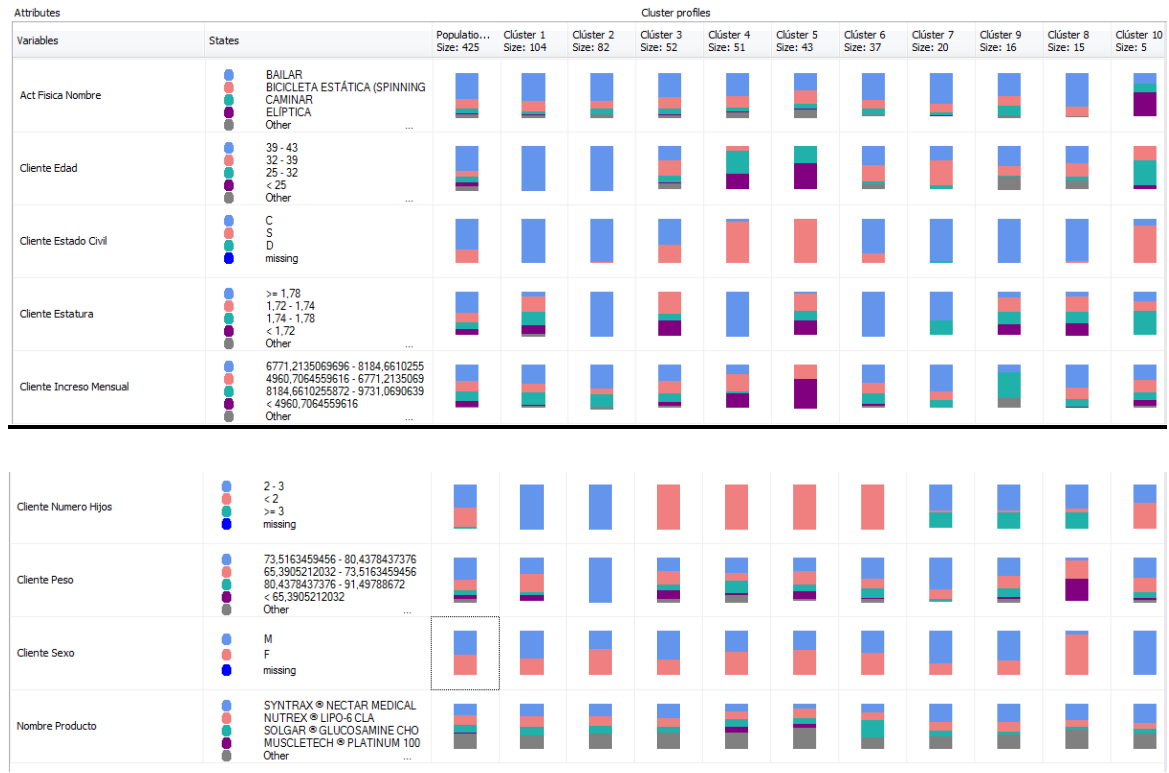


Ilustración 27 Resultados Algoritmo Clúster

PRODUCTOS COMERCIALIZADOS EN EL PRIMER SEMESTRE DEL 2016.

Color	Meaning	Distribution
	SYNTRAX ® NECTAR MEDICAL	0,271
	NUTREX ® LIPO-6 CLA	0,214
	SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX	0,181
	MUSCLETECH ® PLATINUM 100% WHEY	0,033
	FITMISS ® BURN	0,021
	MET RX ® ULTRAMYOSYN WHEY ISOLATE	0,021
	ON ® 100% WHEY GOLD STANDARD	0,019
	DISNEY ® GUMMIES DISNEY PRINCESS	0,014
	MET RX ® PROTEIN PLUS	0,012
	MET RX ® TONALIN CLA 1000	0,012
	ON ® PLATINUM HYDROWHEY	0,012
	CELLUCOR ® C4 50X	0,012
	MUSCLEPHARM ® COMBAT	0,012
	MUSCLETECH ® CLEAR MUSCLE	0,009
	NUTREX ® LIPO-6 BLACK	0,009
	HARBINGER FITNESS ® CUERDAS ANTI MICROBIAL SPEED ROPE	0,007
	HARBINGER FITNESS ® GUANTES POWER SERIES	0,007
	MUTANT ® MUTANT WHEY	0,007
	MUSCLETECH ® GAKIC VO2 MAX SX-7	0,007
	FITMISS ® CLEANSE	0,007
	MET RX ® BIG 100 COLOSSAL	0,005
	SYNTRAX ® MATRIX	0,005
	SOLGAR ® VITAMIN C 500 MG VEGETABLE	0,005
	SYNTRAX ® MICELLAR	0,005
	MUSCLETECH ® PLATINUM 100% BEEF PROTEIN	0,005

Ilustración 28 Distribución de productos comercializados

CARACTERÍSTICAS DE LOS CLUSTERS

Tabla 38 RESULTADOS MODELO CLUSTER

Variables	Valores	Probabilidad
CLIENTE ESTADO CIVIL	C	69,647%
ACT FÍSICA NOMBRE	BAILAR	57,647%
CLIENTE EDAD	39 - 43	56,215%
CLIENTE SEXO	M	53,412%
CLIENTE NUMERO HIJOS	2 - 3	51,314%
CLIENTE PESO	73,5 - 80,4	49,359%
CLIENTE ESTATURA	>= 1,78	47,604%
CLIENTE SEXO	F	46,588%
CLIENTE NUMERO HIJOS	< 2	44,066%
CLIENTE INGRESO MENSUAL	6771 - 8185	38,185%
CLIENTE ESTADO CIVIL	S	30,118%
CLIENTE INGRESO MENSUAL	4961 - 6771	23,572%
ACT FÍSICA NOMBRE	BICICLETA ESTÁTICA (SPINNING)	23,059%
CLIENTE ESTATURA	1,72 - 1,74	21,596%
CLIENTE INGRESO MENSUAL	8185 - 9731	21,327%

CLIENTE ESTATURA	1,74 - 1,78	16,287%
CLIENTE ESTATURA	< 1,72	14,513%
CLIENTE EDAD	32 - 39	14,222%
CLIENTE EDAD	25 - 32	14,141%
CLIENTE INGRESO MENSUAL	< 4961	13,260%
CLIENTE PESO	80- 91	11,876%
ACT FÍSICA NOMBRE	CAMINAR	11,059%
CLIENTE EDAD	< 25	10,619%
CLIENTE PESO	< 65	10,580%
CLIENTE EDAD	>= 43	4,802%
CLIENTE NUMERO HIJOS	>= 3	4,620%
CLIENTE INGRESO MENSUAL	>= 9731	3,657%
ACT FÍSICA NOMBRE	ELÍPTICA	3,294%
CLIENTE PESO	>= 91	2,757%
ACT FÍSICA NOMBRE	MONTAR EN BICICLETA	2,353%
ACT FÍSICA NOMBRE	KICKBOXING	1,176%
ACT FÍSICA NOMBRE	PESAS	0,706%

PAQUETES DE CONSUMO DEFINIDOS (CROSS-SELLING)

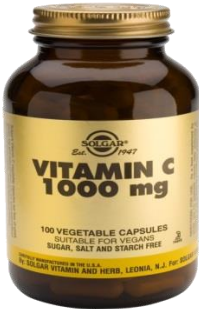
Basándose en el nivel de venta de cada producto obtenido y la función nutricional del producto procedemos a formar los siguientes paquetes de consumo.

- SYNTRAX MEDICAL NÉCTAR



OBJETIVO :CONSTRUYE MUSCULO,DEFINICION MUSCULAR
 Producto indispensable para programas de reducción de peso y grasa.
 Mejora el rendimiento físico y la recuperación
 Tonifica los músculos.

Ilustración 29 Producto Syntrex Nectar

NECTAR + MUTANT BCAA	NECTAR + SOLGAR VITAMIN C 1000 MG	NECTAR + GAKIC VO2 MAX SX-7
 OBJETIVO :CONSTRUYE MUSCULO,MEJORA TU RENDIMIENTO ,AUMENTA ENERGIA Mantiene la maso muscular evitando el catabolismo. Estimula la síntesis proteica y el desarrollo muscular. Incrementa la fuerza y potencia muscular Ilustración 30Producto Mutant BCAA	 OBJETIVO :SALUD Y BIENESTAR Conserva la fortaleza de las paredes de los vasos sanguíneos. Esencial para el metabolismo de los glóbulos blancos. Promueve la absorción del hierro Ilustración 31Producto Solgar Vitamin C	 OBJETIVO :AUMENTA ENERGIA Ilustración32ProductoGakic V02

NUTREX LIPO- 6 CLA

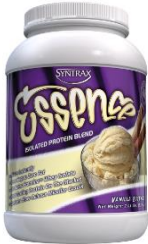


NUTREX[®]
LIPO 6 CLA

OBJETIVO:
PIERDE PESO Y TONIFICA

- Ayuda a formar un cuerpo magro
- Fomenta la actividad metabólica
- Sin estimulantes

Ilustración 33 NUTREXLIPO-6 CLA

NUTREX LIPO 6 + MUSCLETECH PLATINUM 100% ISO WHEY	NUTREX LIPO 6 + SYNTRAX ESSENCE	NUTREX LIPO 6 + MUSCLETECH PHASE 8
 MUSCLETECH[®] PLATINUM 100% ISO WHEY OBJETIVO: CONSTRUYE MÚSCULO, DEFINICIÓN MUSCULAR • Fórmula mejorada con enzimas digestivas • Se mezcla instantáneamente • Avalada por estudios científicos. Libre de prohibidas	 SYNTRAX[®] ESSENCE OBJETIVO: CONSTRUYE MÚSCULO • Alimenta los músculos de manera sana • Disminuye y controla el apetito • Incrementa el metabolismo	 MUSCLETECH[®] PHASE 8 OBJETIVO: CONSTRUYE MÚSCULO • Fórmula perfecta para construir músculos • Ideal para deportistas de alto rendimiento • Contiene 6 fuentes de proteína de la más alta calidad
Ilustración 34 Muscletech Platinum 100% iso Whey	Ilustración 35 Syntrax Essence	Ilustración 36 Muscletech Phase 8

- SOLGAR GLUCOSAMINE CHONDOITRIN COMPLEX



SOLGAR® GLUCOSAMINE CHONDROITIN COMPLEX

Ayuda en la formación del cartilago articular, tendones y ligamentos

Ilustración 37 Solgar Glucosamine Chondroitin Complex

GLUCOSAMINE CHONDROITIN COMPLEX + CORREA WRIST SUPPORT	GLUCOSAMINE CHONDROITIN COMPLEX + CORREA LIFTING HOOK	GLUCOSAMINE CHONDROITIN COMPLEX + CORREA WRIST SUPPORT
 HARBINGER FITNESS® CORREA WRIST SUPPORT • Permite mantener las manos y las muñecas en posición correcta • Potente estabilizador que da soporte para las • Diseño anti estiramiento Ilustración 38 CORREA WRIST SUPPORT	 HARBINGER FITNESS® CORREA LIFTING HOOK • Gancho amplio compatible con todas las barras • Da Soporte y es totalmente ajustable a la muñe • Ganchos que sostienen el peso cuando el agarri Ilustración 39 CORREA LIFTING HOOK	 HARBINGER FITNESS® VENDA KETTLEBELL ARM GUARD • Protege los antebrazos de magulladuras y heridas • Absorbe el sudor durante el ejercicio • Se adapta al antebrazo para un ajuste cómodo y seguro Ilustración 40 CORREA WRIST SUPPORT

POSIBLES CAMPAÑAS

Basándose en los resultados:

1	1,000	2,45	Act Física Nombre = PILATES, Cliente Edad = 25 - 32 -> Nombre Producto = HARBINGER FITNESS ® CORREA BIG GRIP PRO LIFTING STRAPS
---	-------	------	---

Como es 100% probable que nuestros clientes de entre 25-32 años, que realizan como actividad física Pilates compren el producto HARBINGER FITNESS ® CORREA BIG GRIP PRO LIFTING STRAPS

Podemos ofrecerle además nuestra línea de GUANTES:



Ilustración 41 Guantes

<u>2</u>	1,000	0,62	Act Física Nombre = KICKBOXING, Cliente Numero Hijos < 2 -> Nombre Producto = SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
----------	-------	------	---

Como es 100% probable que nuestros clientes que realizan Kickboxing y tienen menos de 2 hijos compren el producto SOLGAR ® GLUCOSAMINE CHONDROITIN COMPLEX
Podemos ofrecerle además:

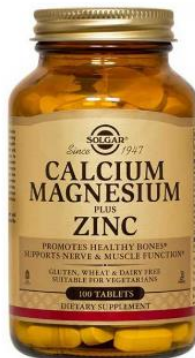


**DISNEY ®
GUMMIES DISNEY PRINCESS**

60 gr

Suplemento de vitaminas y minerales

Ilustración 42 DISNEY ® GUMMIES DISNEY PRINCESS



**SOLGAR ®
CALCIUM MAGNESIUM PLUS ZINC**

OBJETIVO:
SALUD Y BIENESTAR

- Ayuda al Sistema óseo
- Provee minerales esenciales para la salud
- Ayuda al funcionamiento muscular y nervioso

Ilustración 43 SOLGAR ® CALCIUM MAGNESIUM PLUS ZINC

3	1,000	0,49	Act Física Nombre = ELÍPTICA, Cliente Sexo = F -> Nombre Producto = NUTREX ® LIPO-6 CLA
---	-------	------	---

Como es 100% probable que nuestras clientes Mujeres que hacen ejercicios con Elíptica compren el producto NUTREX ® LIPO-6 CLA.

Podemos ofrecerle además:

Ilustración 44 Fitmiss Burn



**FITMISS ®
BURN**

El quemador ideal para mujeres

Ilustración 45 GUANTES PRO WOMEN

Ilustración 46 MET RX ACTIVE WOMEN

Ilustración 47 GUANTES POWER WOMEN



**HARBINGER FITNESS ®
GUANTES PRO WOMEN**

Los mejores accesorios deportivos del mercado



**MET RX ®
ACTIVE WOMEN**
Para la mujer activa



**HARBINGER FITNESS ®
GUANTES POWER WOMEN**

Los mejores accesorios deportivos del mercado

4	1.000	0,45	Cliente Numero Hijos >= 3, Cliente Peso = 80,4 - 91,5-> Nombre Producto = SYNTRAX ® NECTAR MEDICAL
---	-------	------	---

Como es 100% probable que nuestros clientes que tienen >= 3 hijos y tienen un peso entre 80,4 - 91,5 compren el producto SYNTRAX ® NECTAR MEDICAL
Podemos ofrecerle además productos que los ayuden a perder peso:



**MUSCLETECH ®
HYDROXYCUT HARDCORE ELITE CAPS**

100 caps
Fórmula n° 1 para bajar de peso



**MUSCLETECH ®
HYDROXYCUT SX-7**

70 caps | 140 caps
El quemador de grasa de última generación



**MUSCLETECH ®
HYDROXYCUT HARDCORE NEXT GEN**

100 caps | 180 caps
Baja de peso de manera intensa

Ilustración 48 HYDROXYCUT HARDCORE ELITE CAPS

Ilustración 49 HYDROXYCUT SX-7

Ilustración 50 HYDROXYCUT HARDCORE NEXT GEN

Y productos para sus hijos:



**DISNEY ®
GUMMIES DISNEY PRINCESS**

Suplemento de vitaminas y minerales



**SOLGAR ®
KANGAVITES CHEWABLE TABLETS**

Fórmula multivitamínica y multimineral
hecha a base de frutas y hortalizas



**DISNEY ®
GUMMIES DISNEY CAR**

Suplemento de vitaminas y minerales

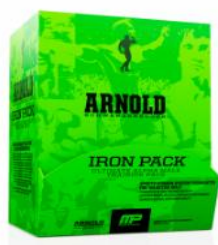
Ilustración 51 GUMMIES DISNEY PRINCESS

Ilustración 52 KANGAVITES CHEWABLE TABLETS Ilustración 53 GUMMIES DISNEY CAR

5	1,000	2,45	Act Física Nombre = NADAR, Cliente Estado Civil = S -> Nombre Producto = CELLUCOR ® WHEY COR PERFORMANCE CELLUCOR
---	-------	------	---

Como es 100% probable que nuestros clientes que NADAN y son solteros compren el producto CELLUCOR ® WHEY COR PERFORMANCE CELLUCOR

Podemos ofrecerle además productos multivitamínicos.



**ARNOLD ®
IRON PACK**

30 packs

Más que un multivitamínico, es tu aliado



**MUSCLEPHARM ®
ARMOR-V**

120 caps

Fortalece tu cuerpo desde adentro



**MUSCLETECH ®
PLATINUM MULTI VITAMIN**

90 caps

Máximo poder, vitaminas y minerales

Ilustración 54 ARNOLD IRON PACK

Ilustración 55 ARMOR-V

Ilustración 56 PLATINUM MULTI VITAMIN

6	1,000	2,15	Cliente Peso >= 91,5, Cliente Edad = 32 - 39 -> Nombre Producto = SMART SHAKE ® SMART SHAKE JAY CUTLER EDITION
---	-------	------	--

Como es 100% probable que los clientes con un peso igual o mayor a 91.5 Kg y además se encuentran en el rango etario 32-39 comprarán SMART SHAKE ® SMART SHAKE JAY UTLER EDITION.

Podemos ofrecerle productos que los ayuden a perder el exceso de peso presentado:



**MUSCLETECH ®
HYDROXYCUT HARDCORE ELITE CAPS**

100 caps

Fórmula n° 1 para bajar de peso



**MUSCLETECH ®
HYDROXYCUT SX-7**

70 caps | 140 caps

El quemador de grasa de última generación



**MUSCLETECH ®
HYDROXYCUT HARDCORE NEXT GEN**

100 caps | 180 caps

Baja de peso de manera intensa

Ilustración 57 HYDROXYCUT HARDCORE ELITE CAPS

Ilustración 58 HYDROXYCUT SX-7

Ilustración 59 HYDROXYCUT HARDCORE NEXT GEN

7	1,000	1,31	Act Física Nombre = PESAS, Cliente Estado Civil = S -> Nombre Producto = MUSCLETECH ® PLATINUM 100% WHEY
---	-------	------	--

Como es 100% probable que nuestros clientes que levantan Pesas y son solteros compren el producto MUSCLETECH ® PLATINUM 100% WHEY
Podemos ofrecerle además productos BCAAs que construyen músculos.



**MUTANT ®
MUTANT BCAA**

Potente dosis de 9,7 gramos de mezcla de bcaas.



**USPLABS ®
MODERN BCAA**



**MUSCLETECH ®
AMINO BUILD NEXT GEN**

El mejor aminoácido que te da energía

Ilustración 60 MUTANT BCAA

Ilustración 61 MODERN BCAA

Ilustración 62 AMINO BUILD NEXT GEN

DISCUSIÓN

Si nos basamos en la definición de patrón de consumo encontramos lo siguiente: es el conjunto de factores que determinan una forma particular de uso que incluye tipo de sustancia, en nuestro caso, por los estudios citados, hemos visto que el consumo de suplementos se ve determinado por el conjunto de características del cliente.

El análisis de la información ha sido realizado con Visual Studio 2015 que contiene la extensión Service Analysis y la base de datos se localiza en SQL server 2014.

Los algoritmos seleccionados para el análisis fueron:

- El modelo de Asociación de Microsoft.
- El modelo de Clúster de Microsoft.
- El modelo de Red Neuronal de Microsoft.

Estos algoritmos forman parte de la herramienta seleccionado y son los modelos que tienen mayor precisión para el análisis requerido.

En los resultados podemos encontrar una o varias reglas que se crean al relacionar diferentes atributos del cliente, éstas contienen probabilidades de ocurrencia y una importancia asociada para cada producto, a su vez determinan los patrones de consumo.

Las reglas identificadas permitirán a la distribuidora de SN conocer los productos demandados de acuerdo a las características del cliente, lo cual por un lado sirve para incluir dichos productos en paquetes que se comercialicen de forma conjunta y por otro lado, con el análisis de la información se llega a conocer aspectos desconocidos por la distribuidora que serán aprovechados y con lo que se espera incrementar las ventas futuras.

Los resultados encontrados pueden ser de utilidad en locales de la distribuidora que se encuentran en distritos de similar nivel socioeconómico.

Después de realizar el procedimientos con los tres algoritmos: Asociación, Red Neuronal y Clúster se ha escogido el algoritmo de Asociación de Microsoft como el más preciso ,ya que es el que se acerca al logro de los objetivos planteados, estos modelos de asociación se generan basándose en conjuntos de datos que contienen identificadores

para casos individuales. Un grupo de elementos de un caso se denomina un conjunto de elementos. Un modelo de asociación se compone de una serie de conjuntos de elementos y de las reglas que describen cómo estos elementos se agrupan dentro de los casos. Las reglas que el algoritmo identifica pueden utilizarse para predecir las probables compras de un cliente en el futuro.

Por la ilustración N° 25 podemos decir que los clientes tienden a consumir los siguientes productos con mayor frecuencia:

- SYNTRAX MEDICAL NÉCTAR
- NUTREX LIPO- 6 CLA
- SOLGAR CLUCOSAMINE CHONDOITRIN COMPLEX

Se observa además que el consumo de los mismos se ve determinado por las siguientes características:

- Estado Civil del Cliente
- Edad del Cliente.
- El sexo del Cliente
- El número de Hijos
- El peso del Cliente
- La estatura del cliente.
- El Ingreso mensual.
- La actividad Física.

CONCLUSIONES

Luego de aplicar técnicas de minería de datos a los registros de la distribuidora de suplementos nutricionales llegamos a las siguientes conclusiones:

Cuando realizamos minería de datos, es muy importante la metodología que seleccionemos, pues la forma en la que realizamos el análisis de la información y el procesamiento son de mucha importancia; debemos elegir una metodología de amplia difusión, que cuente con el respaldo de instituciones líderes del mercado, por otro lado ;escoger las herramientas tecnológicas precisas también es de vital importancia , las mismas deben de contar con basta documentación ,recursos en línea, foros activos que permitan y faciliten el intercambio de ideas.

Seleccionar algoritmos adecuados, que cubran las necesidades del caso de estudio, es importante para conseguir resultados.

El tamaño de la muestra: debe ser relevante y de tamaño adecuado y representativo.

Se considera que el modelo que se acerca más a lo buscado es el algoritmo de asociación de Microsoft, ya que, nos permite conocer las reglas que determinen el consumo de los productos.

Encontramos que el producto más demandado por los clientes, tanto hombres como mujeres son las proteínas, representado por el producto NECTAR ISOLATE, coincidiendo con el estudio español.

Es 100% probable que nuestros clientes de entre 25-32 años, que realizan como actividad física Pilates compren el producto Harbinger fitness ® correa big grip pro lifting straps.

Es 100% probable que nuestros clientes que realizan Kickboxing y tienen menos de 2 hijos compren el producto Solgar ® glucosamine chondroitin complex.

Es 100% probable que nuestras clientes Mujeres que hacen ejercicios con Elíptica compren el producto Nutrex ® lipo-6 cla.

Es 100% probable que nuestros clientes que tienen 3 o más hijos y tienen un peso entre 80,4 - 91,5 compren el producto Syntrex ® nectar medical.

Es 100% probable que nuestros clientes que NADAN y son solteros compren el producto Cellucor ® whey cor performance cellucor.

Es 100% probable que nuestros clientes que levantan Pesas y son solteros compren el producto Muscletech® platinum 100% whey.

El consumo de un tipo de SN se ve determinado por las características del consumidor, conocer las características apropiadas de acuerdo al producto comercializado es de vital importancia para construir un modelo de Data Mining adecuado; en la tesis encontramos, de la misma forma que en el estudio colombiano la importancia de las variables: edad, sexo y nivel económico son determinantes para el consumo de SN.

La minería de datos es un área de conocimiento que nos ayuda a tomar decisiones en base a información del propio negocio.

En el algoritmo de asociación podemos ver que se forman reglas con cierta probabilidad de ocurrencia, así como también con una importancia, estas reglas lo que hacen es encontrar patrones sólidos dentro de la base de datos que relacione las características de los consumidores con los productos que suelen consumir, mientras tanto en el algoritmo neuronal se indica que tan determinante es una variable en la compra de un determinado producto y en el algoritmo Clúster se agrupa las variables de acuerdo al valor que puede tomar y asigna a cada una un nivel de ocurrencia.

La base de datos debe de ser analizada y refinada antes de aplicar técnicas de minería de datos o mejor aún antes de iniciar un proyecto de Data Mining se debe realizar un planeamiento en la recolección de datos.

Es importante emplear el número adecuado de registros que sean significativos, ya que los algoritmos que utiliza minería de datos deben de detectar tendencias para mostrar resultados.

El consumo de SN no se distribuye de forma uniforme, existe una preferencia marcada por productos específicos, de la gran cantidad de productos ofrecidos, solo algunos son demandados por el cliente. Siendo los 3 productos más demandados. (Ilustración 25)

- SYNTRAX MEDICAL NÉCTAR
- NUTREX LIPO- 6 CLA
- SOLGAR CLUCOSAMINE CHONDOITRIN COMPLEX

Los resultados obtenidos pueden ser aplicados a otros locales de la distribuidora que cumplan similares características socioeconómicas con la tienda en estudio.

Ofrecer paquetes donde se incluya un producto altamente demandado puede incrementar los ingresos del comercio.

RECOMENDACIONES

Se recomienda instaurar los cambios en la tienda de Pardo, ofreciendo los diversos paquetes formados en el presente estudio, de tal manera que sirva de piloto para el resto de tiendas que guarden similares características socioeconómicas.

Se recomienda realizar mailing clasificado de acuerdo a reglas identificadas, tanto en mailing por correo electrónico, como mailing físico: al enviar publicidad impresa se evitará publicidad innecesaria y se reducen costos excesivos.

Se recomienda realizar estudios de minería de datos en los demás procesos claves de la empresa, permitirán conocer aspectos adicionales y poder tomar medidas correctivas.

Tener datos almacenados sin procesamiento no genera valor al negocio se recomienda procesarlos con técnicas de minería de datos para obtener información relevante y con la cual tomaremos decisiones acertadas e informadas.

Se recomienda continuar realizando estudios en minería de datos ya que es un campo muy importante de estudio que va tomando mucha importancia. El análisis de Los algoritmos utilizados en minería de datos, resultan muy interesantes. El presenta año 2017 ha sido catalogado por la organización mundial de aduanas (WCO- World Customs Organization) ³⁸ como el año del análisis de datos. Por otro lado de acuerdo a un artículo de IBM Big Data & analytics Hub ³⁹ señala que el Big Data y el análisis será la tendencia este año 2017.

Sería conveniente hacer investigaciones de Data Mining en el campo de la salud ya que nos ayudaría a descubrir nuevas formas de detección de enfermedades sobre la base de síntomas.

REFERENCIAS

- ¹ Andina del Perú para el Mundo. Recuperado de <http://www.andina.com.pe/agencia/noticia-familias-peruanas-sectores-a-y-b-gastan-s-180-mensuales-promedio-suplementos-nutricionales-359916.aspx>
Fecha última consulta: 11 Setiembre del 2016
- ² Perfil del consumidor de suplementos dietéticos en usuarios habituales de centros de acondicionamiento físico Martínez GM, Martínez LM, Cuesta DP, Carrillo DK, Salazar G, Castillo M
- ³ Estudio estadístico del consumo de, suplementos nutricionales, & dietéticos, e. g. (2008). Estudio estadístico del consumo de suplementos nutricionales y dietéticos en gimnasios. Archivos Latinoamericanos De Nutrición, 58(3), 221. Retrieved from <http://search.proquest.com/docview/1685935880?accountid=43847>
- ⁴ Lab Nutrition Recuperado de <https://www.labnutrition.com/storelocator/>
Fecha última consulta: 25/10/16
- ⁵ Nutri Point Recuperado de <http://nutripoint.com.pe/tiendas/>
Fecha última consulta: 25/10/16
- ⁶ Nutri Forma Recuperado de <http://www.nutriforma.pe/locales/>
Fecha última consulta: 25/10/16
- ⁷ Tech Nutrition Recuperado de <http://www.technutritionperu.com/comprar.php>
Fecha última consulta: 25/10/16
- ⁸ Nutri Shop Recuperado de <http://www.nutrishop.pe/>
Fecha última consulta: 25/10/16
- ⁹ Suplement House Recuperado de <http://www.supplementhouseperu.com/>
Fecha última consulta: 25/10/16
- ¹⁰ Tiendeo Recuperado de <http://www.tiendeo.pe/lima/mapa/nutrisport>
Fecha última consulta: 25/10/16
- ¹¹ Mason Vitamins Recuperado de http://vitaminasmason.pe/?page_id=3027
Fecha última consulta: 25/10/16
- ¹² Santa Natura Recuperado de <http://www.santanatura.com.pe/tiendas-santa-natura/>
Fecha última consulta: 25/10/16
- ¹³ Addis, Scott, C.P.C.U., C.R.A. (2014). KNOW YOUR CUSTOMER. Rough Notes, 157(6), 42-42, 44. Retrieved from <http://search.proquest.com/docview/1627047722?accountid=43847> Fecha última consulta: el 01 /10/2016.
- ¹⁴ Expo Nutrition. DATOS DEL SECTOR. Recuperado de <http://exponutrition.com.br/es/datos-del-sector/>
Fecha última consulta: el 01 /10/2016.
- ¹⁵ DATA MINING: A COMPETITIVE WEAPON FOR BANKING AND RETAIL INDUSTRIES Hormazi, A. M., & Giles, S. (2004). DATA MINING: A COMPETITIVE WEAPON FOR BANKING AND RETAIL INDUSTRIES. Information Systems Management, 21(2), 62-71. Retrieved from <http://search.proquest.com/docview/214122842?accountid=43847> Fecha última consulta: el 05 /10/2016.
- ¹⁶ Meng, J. L., Yang, Y. Y., & Niu, W. H. (2011). The research of commonly used technologies and application fields on data mining. Applied Mechanics and Materials, 130-134, 282. Doi : Recuperado de <http://dx.doi.org/10.4028/www.scientific.net/AMM.130-134.282> Fecha última consulta: el 01 /10/2016.
- ¹⁷ DATA MINING TECHNIQUES: A SOURCE FOR CONSUMER BEHAVIOR ANALYSIS. Recuperado de <https://arxiv.org/ftp/arxiv/papers/1109/1109.1202.pdf>
Fecha última consulta: el 15/10/2016.
- ¹⁸ KNOW YOUR BUYER: A PREDICTIVE APPROACH TO UNDERSTAND ONLINE BUYERS' BEHAVIOR

Recuperado de <http://www.happiestminds.com/whitepapers/a-predictive-approach-to-understand-online-buyers-behavior.pdf> Fecha última consulta: el 15 /10/ 2016.

¹⁹ KDD, SEMMA AND CRISP-DM: A PARALLEL OVERVIEW Recuperado de <http://recipp.ipp.pt/bitstream/10400.22/136/1/KDD-CRISP-SEMMA.pdf> Fecha de última consulta 25/01/2017

²⁰ Grimes, S. (2006). In BI deployments, methodology does matter. Intelligent Enterprise, 9(11), 9. Retrieved from <http://search.proquest.com/docview/200664901?accountid=43847> Fecha de última consulta 14/12/2016

²¹ TOP 33 DATA MINING SOFTWARE Recuperado de <http://www.predictiveanalyticstoday.com/top-data-mining-software/> Fecha de última consulta 29/03/2017

²² Oracle vs. SQL Server face-off: Microsoft SQL cheaper, simpler than Oracle DB Recuperado de <http://searchsqlserver.techtarget.com/tip/Oracle-vs-SQL-Server-faceoff-Microsoft-SQL-cheaper-simpler-than-Oracle-DB> Fecha de última consulta 24/01/2017

²³ SEMMA AND CRISP-DM: DATA MINING METHODOLOGIES Recuperado de <http://jesshampton.com/2011/02/16/semma-and-crisp-dm-data-mining-methodologies/> Fecha de última consulta 25/01/2017

²⁴ NIVELES DE USO DE METODOLOGIA EN DATA MINING Recuperado de <http://www.kdnuggets.com/polls/2014/analytics-data-mining-data-science-methodology.html> Fecha de última consulta 29/03/2017

²⁵ Algoritmo de asociación de Microsoft
<https://msdn.microsoft.com/es-es/library/ms174916.aspx>
Fecha última consulta: el 05 diciembre 2016.

²⁶ Algoritmo de clústeres de Microsoft
<https://msdn.microsoft.com/es-es/library/ms174879.aspx>
Fecha última consulta: el 05 diciembre 2016.

²⁷ Algoritmo de red neuronal de Microsoft
<https://msdn.microsoft.com/es-es/library/ms174941.aspx>
Fecha última consulta: el 05 diciembre 2016.

²⁸ Data Mining Concepts and techniques tercera edicion Jiawei Han [Micheline Kamber] Jian Pei
Retrieved from <http://romisatriawahono.net/lecture/dm/book/Han%20-%20Data%20Mining%20Concepts%20and%20Techniques%203rd%20Edition%20-%202012.pdf>
Fecha última consulta: el 05 diciembre 2016
29 Data Mining: Concepts and Techniques, 2nd ed. Jiawei Han and Micheline Kamber March 2006. ISBN 1-55860-901-6 Fecha última consulta: el 05 diciembre 2016

³⁰ DATA MINING: A COMPETITIVE WEAPON FOR BANKING AND RETAIL INDUSTRIES
Recuperado de http://cjou.im.tku.edu.tw/dm200707/dm_application.pdf
Fecha última consulta: el 15 Octubre 2016.

³¹ Data Mining Concepts and Techniques Second Edition Jiawei Han and Micheline Kamber Data Mining for the Masses Dr. Matthew North ISBN: 0615684378 Fecha última consulta: el 05 diciembre 2016

³² Minería de datos a través de ejemplos: María Pérez Marqués ISBN: 978-84-941801-4-9 Fecha última consulta: el 05 diciembre 2016

³³ Vallejos, Sofía. (2006). Minería de Datos.
http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/Mineria_Datos_Vallejos.pdf
Fecha última consulta: el 17 Octubre 2016.

³⁴ Advances in Knowledge Discovery and Data Mining, Part II: 14th Pacific-Asia conference, Pakdd 2010 hyderabad, India, June 2010. Fecha última consulta: 17 /10/ 2016.

³⁵Marketing and Consumer Behavior: Concepts, Methodologies, Tools, and Applications retrieved from [https://books.google.com.pe/books?id=r6GXBgAAQBAJ&pg=PA896&dq=Kwan+et+al.\(2005\)+data+mining&hl=es&sa=X&ved=0ahUKEwiuvfPc_YjRAhVr5YMKHa-ABHQQ6AEIlzAA#v=onepage&q=Kwan%20et%20al.\(2005\)%20data%20mining&f=false](https://books.google.com.pe/books?id=r6GXBgAAQBAJ&pg=PA896&dq=Kwan+et+al.(2005)+data+mining&hl=es&sa=X&ved=0ahUKEwiuvfPc_YjRAhVr5YMKHa-ABHQQ6AEIlzAA#v=onepage&q=Kwan%20et%20al.(2005)%20data%20mining&f=false) Fecha última consulta: 17 /10/ 2016.

³⁶Aplicación de técnicas de minería de datos para la fidelización y retención de clientes Autor: Víctor López Fandiño Business Intelligence Solutions Specialist IBM Software Group Fecha última consulta: 17 /10/ 2016.

³⁷Diseño de investigación no experimental extraído de <http://es.slideshare.net/conejo920/diseo-de-investigacion-no-experimental>. Fecha de última consulta 25 Enero 2017

³⁸ Data mining and case od sampling Extraído de http://sce.uhcl.edu/boetticher/ML_DataMining/SAS-SEMMA.pdf Fecha última consulta: 10 /11/ 2016.

³⁹World Customs Organization declares 2017 to be the year of Data Analysis Extraído de <http://www.wcoomd.org/en/media/newsroom/2016/november/world-customs-organization-declares-2017-to-be-the-year-of-data-analysis.aspx> Fecha última consulta: 25 /01/ 2016.

⁴⁰ Big data and analytics trends in 2017: James Kobielus's predictions Extraído de <http://www.ibmbigdatahub.com/blog/big-data-and-analytics-trends-2017-james-kobielus-s-predictions> Fecha última consulta: 25 /01/ 2016.

Anexos

DIAGRAMA DE BASE DE DATOS

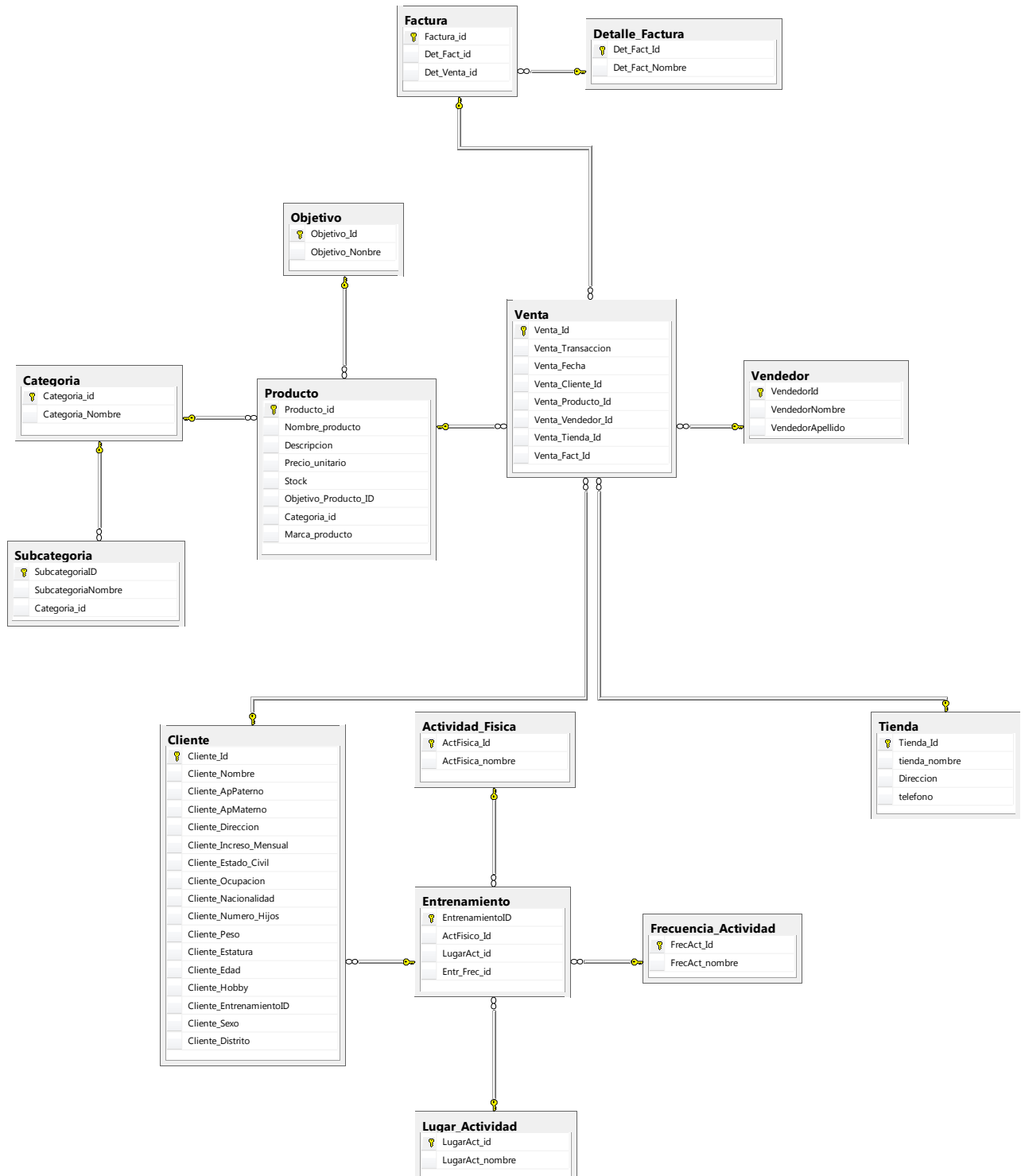


Ilustración 63 Diagrama de Base de Datos