



**SEP**  
SECRETARÍA DE  
EDUCACIÓN PÚBLICA



TECNOLÓGICO NACIONAL DE MÉXICO

Instituto Tecnológico de Ciudad Guzmán



**INSTITUTO TECNOLÓGICO DE CD. GUZMÁN**

**PROGRAMA DE MAESTRÍA EN CIENCIAS  
DE LA COMPUTACIÓN**

**TESIS**

**TEMA:  
PREDICCIÓN DE VENTAS PARA UN MODELO DE  
NEGOCIOS.**

**QUE PARA OBTENER EL GRADO DE:  
MAESTRA EN CIENCIAS DE LA COMPUTACIÓN**

**PRESENTA:  
ING. TERESITA MORENO RAMÍREZ**

**ASESOR(A) :  
Dra. Rosa María Michel Nava**

**CODIRECTOR(A) :  
Dra. María Guadalupe Sánchez Cervantes**

**CD. GUZMÁN JALISCO, MÉXICO, AGOSTO DE 2019**



Instituto Tecnológico de Ciudad Guzmán

"2019, Año del Caudillo del Sur, Emiliano Zapata"

Cd. Guzmán, Jal. a **13/Agosto/2019**

Oficio No. DEPI/50/19

ASUNTO : AUTORIZACIÓN DE IMPRESIÓN

**C. TERESITA MORENO RAMÍREZ**  
N.C. M17290010

En cumplimiento con el documento normativo de las disposiciones para la operación de estudios de posgrado del Tecnológico Nacional de México y con base en la aprobación del Comité Tutorial comisionado para su revisión; la División de Estudios de Posgrado e Investigación le otorga la autorización de impresión de su trabajo de tesis intitulado:

**"PREDICCIÓN DE VENTAS PARA UN MODELO DE NEGOCIOS"**

dirigido por el **Dra. Rosa Maria Michel Nava**, desarrollado como requisito parcial para la obtención del grado de Maestro en Ciencias de la Computación, de acuerdo al plan de estudios MCOM-2011-05.

Sin otro asunto en particular, quedo de usted.

**ATENTAMENTE**

  
**DR. HUMBERTO BRACAMONTES DEL TORO**  
JEFE DE LA DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

C.p. Archivo



**S.E.P. TecNM**  
**INSTITUTO TECNOLÓGICO**  
**DE CD. GUZMAN**  
**DIVISION DE ESTUDIOS**  
**DE POSGRADO E**  
**INVESTIGACION**



## Índice general

Índice de figuras .....	iii
Índice de tablas .....	iv
Agradecimientos.....	1
Resumen .....	2
Abstract.....	3
1. Capítulo I. Introducción .....	5
1.1 Descripción del trabajo de investigación .....	6
1.2 Definición del problema .....	7
1.3 Justificación .....	8
1.4 Objetivos .....	9
1.5 Hipótesis .....	10
1.6 Motivación .....	11
2. Capítulo II. Fundamento teórico .....	12
2.1 Estado del arte.....	13
2.2 Marco teórico.....	20
2.3 Marco conceptual.....	21
2.4 Marco contextual .....	31
3. Capítulo III. Marco Metodológico .....	32
3.1 Tipo de Investigación.....	33
3.2 Universo, población o unidades de análisis .....	33
3.3 Criterios de inclusión/exclusión.....	33
3.4 Muestreo .....	33
3.5 Muestra .....	33
3.6 Instrumentos.....	34
3.7 Aparatos .....	34
3.8 Procedimientos.....	34
4. Capítulo IV. Resultados .....	38
4.1 Pruebas realizadas y resultados obtenidos .....	39
4.1.1 Regresión lineal.....	39
4.1.2 Promedio móvil.....	46
4.1.3 Suavización exponencial .....	57

4.1.4 Suavización exponencial ajustada a la tendencia.....	68
4.1.5 Árboles aleatorios.....	80
4.2 Porcentajes de errores .....	81
5. Capítulo V. Conclusiones.....	83
5.1 Conclusiones.....	84
5.2 Recomendaciones .....	84
5.3 Trabajo futuro .....	85
6. Fuentes de información.....	86
7. Glosario .....	90

## Índice de figuras

Figura 2.1 Software Minitab (fuente: <a href="http://www.minitab.com/es-mx/products/minitab/">http://www.minitab.com/es-mx/products/minitab/</a> ) .....	18
Figura 2.2 Software SAS (fuente: <a href="https://www.sas.com/es_mx/home.html">https://www.sas.com/es_mx/home.html</a> ).....	18
Figura 2.3 Software EViews 10 (fuente: <a href="http://www.eviews.com/home.html">http://www.eviews.com/home.html</a> ) .....	19
Figura 2.4 Software IBM SPSS (fuente: <a href="https://www.ibm.com/analytics/mx/es/technology/spss/">https://www.ibm.com/analytics/mx/es/technology/spss/</a> ) .....	20
Figura 2.5 Descripción general de los pasos que constituyen el proceso KDD (Fuente: <a href="https://www.kdnuggets.com/2016/03/data-science-process-rediscovered.html/2">https://www.kdnuggets.com/2016/03/data-science-process-rediscovered.html/2</a> ) .....	22
Figura 4.1 Gráfica de regresión lineal del producto 1 (Fuente: elaboración propia).....	41
Figura 4.2 Gráfica de regresión lineal del producto 2 (Fuente: elaboración propia).....	42
Figura 4.3 Gráfica de regresión lineal del producto 3 (Fuente: elaboración propia).....	43
Figura 4.4 Gráfica de regresión lineal del producto 4 (Fuente: elaboración propia).....	44
Figura 4.5 Gráfica de regresión lineal del producto 5 (Fuente: elaboración propia).....	45
Figura 4.6 Gráfica de regresión lineal del producto 6 (Fuente: elaboración propia).....	46
Figura 4.7 Gráfica de promedio móvil del producto 1 (Fuente: elaboración propia).....	48
Figura 4.8 Gráfica de promedio móvil del producto 2 (Fuente: elaboración propia).....	50
Figura 4.9 Gráfica de promedio móvil del producto 3 (Fuente: elaboración propia).....	51
Figura 4.10 Gráfica de promedio móvil del producto 4 (Fuente: elaboración propia).....	53
Figura 4.11 Gráfica de promedio móvil del producto 5 (Fuente: elaboración propia).....	55
Figura 4.12 Gráfica de promedio móvil del producto 6 (Fuente: elaboración propia).....	56
Figura 4.13 Gráfica de suavización exponencial del producto 1 (Fuente: elaboración propia)	59
Figura 4.14 Gráfica de suavización exponencial del producto 2 (Fuente: elaboración propia)	61
Figura 4.15 Gráfica de suavización exponencial del producto 3 (Fuente: elaboración propia)	62
Figura 4.16 Gráfica de suavización exponencial del producto 4 (Fuente: elaboración propia)	64
Figura 4.17 Gráfica de suavización exponencial del producto 5 (Fuente: propia).....	66
Figura 4.18 Gráfica de suavización exponencial del producto 6 (Fuente: elaboración propia)	67
Figura 4.19 Gráfica de suavización exponencial ajustada a la tendencia del producto 1 (Fuente: elaboración propia).....	70
Figura 4.20 Gráfica de suavización exponencial ajustada la tendencia del producto 2 (Fuente: elaboración propia).....	72
Figura 4.21 Gráfica de suavización exponencial ajustada a la tendencia del producto 3 (Fuente: elaboración propia).....	74
Figura 4.22 Gráfica de suavización exponencial ajustada a la tendencia del producto 4 (Fuente: elaboración propia).....	76
Figura 4.23 Gráfica de suavización exponencial ajustada a la tendencia del producto 5 (Fuente: elaboración propia).....	78
Figura 4.24 Gráfica de suavización exponencial ajustada a la tendencia del producto 6 (Fuente: elaboración propia).....	80

## Índice de tablas

Tabla 3.1 Pasos para el modelo de predicción (Fuente: elaboración propia) .....	37
Tabla 4.1 Información de ventas correspondiente a cuatro años, la media aritmética y la predicción con regresión lineal del producto 1 (Fuente: elaboración propia) .....	40
Tabla 4.2 Información de ventas correspondiente a cuatro años, la media aritmética y la predicción con regresión lineal del producto 2 (Fuente: elaboración propia) .....	41
Tabla 4.3 Información de ventas correspondiente a cuatro años, la media aritmética y la predicción con regresión lineal del producto 3 (Fuente: elaboración propia) .....	42
Tabla 4.4 Información de ventas correspondiente a cuatro años, la media aritmética y la predicción con regresión lineal del producto 4 (Fuente: elaboración propia) .....	43
Tabla 4.5 Información de ventas correspondiente a cuatro años, la media aritmética y la predicción con regresión lineal del producto 5 (Fuente: elaboración propia) .....	44
Tabla 4.6 Información de ventas correspondiente a cuatro años, la media aritmética y la predicción con regresión lineal del producto 6 (Fuente: elaboración propia) .....	45
Tabla 4.7 Información de ventas correspondiente por mes, año y la predicción generada con promedio móvil del producto 1 (Fuente: elaboración propia) .....	48
Tabla 4.8 Información de ventas correspondiente por mes, año y la predicción generada con promedio móvil del producto 2 (Fuente: elaboración propia) .....	49
Tabla 4.9 Información de ventas correspondiente por mes, año y la predicción generada con promedio móvil del producto 3 (Fuente: elaboración propia) .....	51
Tabla 4.10 Información de ventas correspondiente por mes, año y la predicción generada con promedio móvil del producto 4 (Fuente: elaboración propia) .....	53
Tabla 4.11 Información de ventas correspondiente por mes, año y la predicción generada con promedio móvil del producto 5 (Fuente: elaboración propia) .....	54
Tabla 4.12 Información de ventas correspondiente por mes, año y la predicción generada con promedio móvil del producto 6 (Fuente: elaboración propia) .....	56
Tabla 4.13 Información de ventas correspondiente por mes, año y predicción con suavización exponencial del producto 1 (Fuente: elaboración propia) .....	59
Tabla 4.14 Información de ventas correspondiente por mes, año y predicción con suavización exponencial del producto 2 (Fuente: elaboración propia) .....	60
Tabla 4.15 Información de ventas correspondiente por mes, año y predicción con suavización exponencial del producto 3 (Fuente: elaboración propia) .....	62
Tabla 4.16 Información de ventas correspondiente por mes, año y predicción con suavización exponencial del producto 4 (Fuente: elaboración propia) .....	64
Tabla 4.17 Información de ventas correspondiente por mes, año y predicción con suavización exponencial del producto 5 (Fuente: elaboración propia) .....	65
Tabla 4.18 Información de ventas correspondiente por mes, año y predicción con suavización exponencial del producto 6 (Fuente: elaboración propia) .....	67
Tabla 4.19 Información de ventas correspondiente por mes, año y predicción con suavización exponencial ajustada a la tendencia del producto 1 (Fuente: elaboración propia) .....	70

Tabla 4.20 Información de ventas correspondiente por mes, año y predicción con suavización exponencial ajustada a la tendencia del producto 2 (Fuente: elaboración propia) .....	71
Tabla 4.21 Información de ventas correspondiente por mes, año y predicción con suavización exponencial ajustada a la tendencia del producto 3 (Fuente: elaboración propia) .....	73
Tabla 4.22 Información de ventas correspondiente por mes, año y predicción con suavización exponencial ajustada a la tendencia del producto 4 (Fuente: elaboración propia) .....	75
Tabla 4.23 Información de ventas correspondiente por mes, año y predicción con suavización exponencial ajustada a la tendencia del producto 5 (Fuente: elaboración propia) .....	77
Tabla 4.24 Información de ventas correspondiente por mes, año y predicción con suavización exponencial ajustada a la tendencia del producto 6 (Fuente: elaboración propia) .....	79
Tabla 4.25 Tabla de porcentajes de certeza con árboles aleatorios de cada producto (Fuente: elaboración propia) .....	81
Tabla 4.26 Porcentajes de error calculado con MAPE de cada producto aplicando los diferentes modelos (Fuente: elaboración propia) .....	81

## **Agradecimientos**

En primera estancia agradezco a mis padres, por todas las enseñanzas y el apoyo brindado durante mi tiempo de formación, por darme las herramientas para estar en esta vida y el carácter de defender mis ideales, gracias a mi familia.

Sencillo no ha sido este proceso, pero gracias a la dedicación de mi querida asesora la Doctora Rosa María Michel Nava, al transmitirme su conocimiento, aconsejarme en cada uno de mis pruebas en este camino, escucharme y tomar en cuenta mi opinión, así como todo el tiempo dedicado y la paciencia que tuvo en cada sesión, he logrado este objetivo que es el culminar mi tesis con éxito, gracias por todo su apoyo ya que sin él esto no podría haber sido posible, más que considerarla sólo mi asesora, una amiga.

Gracias a la maestra María Eugenia Puga Nathal por proporcionarme información que necesité para el proyecto, al tomarse un momento para atender mis dudas.

Agradezco el tiempo invertido del Doctor Felipe Santoyo Telles al asesorarme sobre estadística y darme consejos útiles de cómo abordar el proyecto.

También he de mencionar a personas que han estado presentes en mi vida, gracias a José por brindarme su apoyo cuando lo he necesitado. Gracias a mis amigos por ser una segunda familia.

Gracias a la Teacher María José Lares y al Teacher Jesús Arellano por su apoyo en mi formación.

## Resumen

En la actualidad es de gran importancia que las empresas prevengan la cantidad de productos que deben tener en almacenes para poder cubrir las ventas en ciertos meses durante el año, así como medir qué comportamiento ha tenido en el paso del tiempo, con el fin de poder planificar el rumbo de su negocio, reducir pérdidas e incrementar las ventas. El objetivo de este proyecto consiste en tomar una muestra de la información de las ventas de seis productos, de la línea de artículos de mayor movimiento en la empresa, realizando un análisis de su comportamiento en un periodo de tiempo. Tomando esto como referencia se definieron dos variables de tipo cuantitativo, las ventas y el tiempo, para aplicarlas a un modelo de predicción.

Para poder generar una estimación de las ventas se analizaron diferentes técnicas de minería de datos para aplicar a las dos variables definidas. Entre las técnicas que se seleccionaron se encuentran series de tiempo, y dentro de ella, promedios móviles, suavización exponencial y suavización exponencial ajustada a la tendencia; también se seleccionó regresión lineal y árboles aleatorios. Se crearon modelos de predicción con cada una de las técnicas, aplicando las ecuaciones correspondientes y la información extraída de la base de datos, con el fin de generar predicciones de ventas con mayor precisión, para posteriormente comparar los diferentes resultados y medir el porcentaje de error, llegando a seleccionar un modelo.

La técnica seleccionada para medir el error fue MAPE (Error Porcentual Absoluto Medio) ya que, en ocasiones, es más útil calcular los errores de predicción en términos de porcentajes en lugar de cantidades.

Las pruebas con cuatro de las técnicas fueron generadas en el programa Excel. En lo que respecta a la técnica de árboles aleatorios se probó en un programa desarrollado en el lenguaje de programación Python.

Los resultados obtenidos en este trabajo van orientados a beneficiar a las empresas que manejen una abarrotera, de tal manera que puedan llevar un mejor control en su inventario y en sus ventas. Sin embargo, es necesario tener presente que para obtener un modelo confiable, es necesario contar más información y todas las variables que intervienen en las ventas.

## **Abstract**

Nowadays it is of great importance that companies report the quantity of products that they must have in stores to be able to cover sales in certain months during the year, as well as measure what behavior they have had over time, in order to be able to plan the direction of their business, reduce losses and increase sales. The objective of this project is to take a sample of the information of the sales of six products, from the line of articles of greater movement in the company and performing an analysis of their behavior over a period of time. Taking this as a reference, two quantitative variables, sales and time, were defined to apply them to a prediction model.

In order to generate a sales estimate, different data mining techniques were selected to be applied to two defined variables. Among the techniques that were selected was the time series, and within these were moving averages, exponential smoothing y trend adjusted exponential smoothing. Linear regression and random forests were also selected. Prediction models were created with each of these techniques, applying the corresponding equations based on information gathered from the database, to generate sales predictions with greater precision, so we may later compare the different results and measure the percentage of error, before selecting a prediction model.

The selected technique to measure the error was MAPE (Mean Absolute Percentage Error). Sometimes, it is more useful to calculate predictions errors in terms of percentage rather than quantities.

The tests with four of the techniques were generated in the Excel program. Regarding the random forest technique, it was tested in a program developed in the Python Programming Language.

The results obtained in this work are aimed at benefiting companies that manage a grocery store, so that they can keep better control their in inventory and sales. However, we must keep in mind that to obtain a reliable model, it is necessary to have all the information and all the variables involved in the total sales.

# **PREDICCIÓN DE VENTAS PARA UN MODELO DE NEGOCIOS**

## **1. Capítulo I. Introducción**

En el presente capítulo se describe el problema a resolver con el proyecto, así como sus respectivos objetivos, justificación e hipótesis.

## **1.1 Descripción del trabajo de investigación**

Mamá Coneja S.A de C.V es una abarrotera ubicada en Ciudad Guzmán, Jalisco, la cual maneja diferentes líneas de productos, para clasificar los artículos. Dentro de esas clasificaciones cabe mencionar, que existe una línea de vital importancia, llamada cereales y semillas, la cual es el giro principal de la empresa.

El presente trabajo de investigación consistió en proporcionar una herramienta computacional que sirva como apoyo en el análisis de las ventas de los artículos de la línea de productos antes mencionada, mediante la cual se pueda predecir en un periodo determinado qué cantidad de productos se estima vender, para tener el surtido necesario que cubra los requerimientos de los clientes sin tener productos en exceso, pero tampoco sin existencia.

El principal objetivo fue generar un modelo predictivo aplicable a los artículos contenidos en la línea de productos cereales y semillas, realizando el respectivo análisis de todos los artículos contenidos en dicha línea, para identificar qué datos introducir en el modelo, así como las salidas esperadas.

Como resultado de dicho análisis se determinó hacer las pruebas pertinentes con una muestra de seis artículos (que por cuestiones de confidencialidad se manejarán con una clave).

La investigación comenzó tomando la base de datos de la empresa, analizando cada una de las tablas, así como sus relaciones, hasta llegar a identificar las que se necesitarían en específico para las ventas de los productos, extrayendo la información que ingresaría al modelo predictivo (cabe mencionar que en este proyecto se seleccionaron técnicas de minería de datos para diseñar dicho modelo) para su procesamiento. Una vez hecho esto se presenta la predicción generada por el modelo, mostrando la cantidad estimada de ventas de cada producto en un periodo de tiempo. Los resultados obtenidos serán almacenados a su vez, en una base de datos diseñada con el fin de guardar un histórico de las predicciones generadas en diferentes periodos para poder consultarlas cada vez que sea requerido, sin necesidad de volver a procesar.

## 1.2 Definición del problema

Sarmiento (2008) describe cómo los empresarios se enfrentan a la problemática de predecir la actividad económica. Es importante identificar las variables principales que se tienen en el negocio para poder conocer la evolución y necesidades en el mismo, de tal manera que los dueños de pequeñas y medianas empresas puedan tomar decisiones de una forma intuitiva.

Se puede reducir el grado de incertidumbre causado por los cambios en el entorno, las empresas deben respaldar sus decisiones en algo más que las experiencias o la intuición, deben respaldarlo elaborando pronósticos precisos y confiables para satisfacer las necesidades en la planeación del negocio (Juárez, Zúñiga, Flores y Partida, 2016).

En la empresa de Mamá Coneja, se toman decisiones basadas en la experiencia que se tenía en el negocio, considerando los productos de mayor movimiento en ventas, lo cual no siempre resultaba de forma acertada, ya que se podía dar la situación de comprar producto de más, teniendo pérdidas al expirar en bodega, o por el contrario, no tener el surtido necesario para cubrir la demanda.

Así mismo, Sarmiento (2008) menciona la importancia de conocer la evolución y tendencia de las variables principales de los negocios, tomando ese análisis en ocasiones es necesario traducir sus métodos en modelos estadísticos, utilizando una técnica de minería de datos, para que los resultados obtenidos permitan ser utilizados como un instrumento para la toma de decisiones en situaciones futuras.

Con la presente investigación se dio respuesta a la siguiente interrogante: ¿usando una técnica de minería de datos se puede obtener un modelo predictivo de ventas en el negocio de Mamá Coneja, de Ciudad Guzmán?

### **1.3 Justificación**

El negocio de Mamá Coneja toma decisiones empíricamente para el análisis de ventas de los productos que tienen más movimiento en la empresa, teniendo incertidumbre de la cantidad de productos que debían tener en existencia para cubrir la demanda de los clientes, sin tener excedentes de mercancía en almacén o faltantes que no pudieran cubrir las cantidades solicitadas, lo que ocasionaba pérdidas monetarias.

Razón por la cual, se pensó en un modelo que mediante una técnica de análisis de minería de datos se procesara la información sobre las ventas, interpretándola mediante un modelo predictivo, de forma que ayudara a obtener información precisa sobre la cantidad que debería existir de un producto en un periodo de tiempo determinado.

Se considera que el proyecto es viable, ya que se cuentan con los recursos humanos, económicos y materiales para realizarlo.

Por una parte, el desarrollo del modelo de predicción tiene importancia en la sociedad al generarlo para un negocio en la región, haciendo mejoras específicas para dicha empresa.

Económicamente se verá reflejado en la posibilidad de disminuir la pérdida financiera que el negocio estaba teniendo.

Ambientalmente impacta al obtener predicciones que tomadas en cuenta permitan tener en existencia sólo la cantidad de producto necesaria, sin que se compre en exceso, haciendo que algunos productos caduquen, provocando contaminación ambiental.

Tecnológicamente se tendrá un impacto en el negocio, dado que ahora las decisiones estarán basadas en la predicción de ventas que el modelo genere.

## **1.4 Objetivos**

### **Objetivo general**

Generar un modelo de predicción de ventas utilizando una técnica de minería de datos, para el modelo de negocios de Mamá Coneja en Cd. Guzmán, Jalisco.

### **Objetivos particulares o específicos**

Para lograr el objetivo general, fue necesario:

- Realizar investigación sobre artículos relacionados al proyecto.
- Comparar trabajos anteriores con la propuesta.
- Realizar investigación sobre técnicas de minería de datos para predicción aplicables al modelo de negocios del proyecto.
- Seleccionar la técnica minería de datos para predicción más conveniente para el proyecto.
- Diseñar el modelo de predicción con la técnica seleccionada.
- Mostrar los resultados de los valores de las predicciones por cada producto en la línea seleccionada.
- Contrastar los resultados de la predicción contra los valores reales.
- Procesar los resultados obtenidos.
- Dar a conocer los resultados.

## **1.5 Hipótesis**

La hipótesis que se pretendió comprobar en este proyecto es la siguiente:

La predicción de la cantidad de ventas de productos es más precisa aplicando una técnica de minería de datos.

## **1.6 Motivación**

Dentro de las razones que motivaron para la realización de este proyecto, fue el pensar involucrarse en el área de interés como lo son las bases de datos, big data, minería de datos, sin embargo, fue un horizonte totalmente diferente y muy extenso, el tratar de adaptarse a temas que son totalmente diferente a lo que se trabajó alguna vez, por el objetivo planteado, así como las técnicas que se aplicaron.

## **2. Capítulo II. Fundamento teórico**

La información recabada en el presente capítulo surge de trabajos encontrados relacionados con el tema de tesis, antecedentes y conceptos importantes para el proyecto.

## 2.1 Estado del arte

A continuación, se describen una serie de investigaciones o proyectos relacionados con el que se aborda en esta tesis.

El trabajo encontrado que se desarrolló para obtener el grado en ingeniería informática se relaciona con los antecedentes por tener conceptos importantes para esta investigación y técnicas utilizadas para la predicción.

### **Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso**

Este artículo muestra una investigación realizada sobre los conceptos de análisis predictivo, así como las técnicas más relevantes para el problema que se aborda. Se examinaron las principales herramientas de análisis predictivo disponible como SPSS, MATLAB o SAP, sin embargo, este sector tomó como objeto de estudio de este trabajo, de las que se tenga licencia open Source, centrándose en R y Weka. Se hizo una comparativa entre ambas, teniendo resultados favorecedores hacia el programa R, ya que a pesar de que Weka permite usar sus algoritmos en cualquier programa escrito en Java, la herramienta no llega a alcanzar el potencial de R. Weka se puede considerar más como un programa orientado a la formación, mientras que R tiene un uso más amplio entre los usuarios en el sector profesional (Espino, 2017, p. 6).

En este documento de tesis se probó una herramienta de open Source llamada Python, que es un lenguaje de código abierto, con librerías muy completas, utilizando análisis de datos y seleccionando técnicas de regresión lineal para generar predicciones.

El siguiente trabajo de tesis para obtener el grado de máster universitario en ingeniería computacional y sistemas inteligentes, aborda conceptos importantes relacionados con este documento.

### **Construcción de un modelo predictivo de tipo *Gradient boosting* para ventas online**

Los objetivos más concretos en este trabajo son el pre-procesar los datos, la construcción del modelo predictivo, el análisis del resultado y el cálculo del *score* del modelo.

Para este proyecto se utilizó la distribución de Python llamada Anaconda, creando un modelo predictivo de tipo Gradient boosting, se hicieron los cálculos necesarios para predecir el número de ventas online de unos productos X, en donde los resultados sólo muestran el número de identificación y el número de ventas en los diferentes doce meses del año, sin embargo, debido a que no se obtuvo el tipo de producto, sólo se pudo observar que en algunos meses existen ventas más elevadas sobre ciertos productos, llegando así a la conclusión que son ventas por temporadas altas (Basauri, 2015).

En el sistema que se desarrolló, en los productos de la línea seleccionada, se obtienen predicciones, pudiendo conocer la clave del producto, así como su descripción y total de ventas generadas; además de que se usó Python como herramienta de software.

A continuación, se muestra información importante de un artículo de investigación donde se dan a conocer modelos estadísticos para predecir ventas por catálogo.

### **Modelo estadístico de regresión múltiple, columna vertebral para predecir en empresas multinacionales con estilo de venta por catálogo**

Boada (2013) realizó una investigación donde se expone un modelaje sistemático de las ventas en una compañía con estilo de venta directa. Con este modelaje fue posible crear un esquema general estadístico, estandarizable a todos los productos de la compañía, con la finalidad de lograr simulaciones y predicciones con alto grado de confianza sobre su posible evolución en dimensión temporal establecida. Se hizo uso de productos con historial relevante, clasificados por códigos de producción y almacenamiento, agrupados por sus características físicas similares, valorando el impacto de diversas variables de mercadotecnia, aplicando ecuaciones de regresión lineal, con residuos de comportamiento aleatorio y valor esperado cero, tomando variable tanto cualitativas como cuantitativas.

En el presente trabajo de investigación se aplicaron técnicas de regresión lineal para evaluar la precisión de las predicciones, se tomaron en cuenta las fórmulas que se utilizaron en este artículo.

El artículo que se describe a continuación desarrolla un modelo para la predicción, donde se muestra la importancia de conocer la evolución y tendencia de las variables en los negocios con apoyo de los modelos estadísticos.

### **Predicción con series de tiempo y regresión**

Sarmiento (2008) elaboró un documento donde describe la importancia de conocer las variables principales de un negocio, para que con base en éstas se realice el análisis necesario y correcto para generar un modelo estadístico como apoyo para la toma de decisiones. Hacer predicciones sobre el rumbo a donde se dirige la empresa es fundamental, por tal motivo es necesario realizar un análisis de los modelos para establecer predicciones, tomando en cuenta el tipo de variable que se quiere analizar. Para las pruebas se utilizó una herramienta gratuita para demostrar que se puede generar un análisis del mismo nivel del software especializado que es de compra, con el fin de que se tomen en cuenta de forma imparcial, siempre y cuando se tenga información con la que se cuenta sea confiable.

El método que se seleccionó será para generar un modelo predictivo para información estructurada de una base de datos, a fin de crear un software de fácil acceso y especializado para el tipo de variable identificada.

El siguiente trabajo presenta un estudio sobre técnicas de series de tiempo y cómo evaluar algunos métodos.

### **Análisis de series de tiempo en el pronóstico de la demanda de almacenamiento de productos perecederos**

En este artículo se hace una comparación de algunas de las técnicas de series de tiempo para mejorar la toma de decisiones para la demanda en almacenamiento, tomando en cuenta que dicha empresa presentaba una tendencia de crecimiento. Hacer pronóstico de los productos perecederos fue de mucha importancia al tomar en cuenta que, si se adquiere producto de más puede llegar a caducarse, siendo así una pérdida significativa para el negocio, así como también es de dicha importancia tener en almacén los productos necesarios para tener buenas ventas.

Se muestra una descripción de las técnicas y cómo se aplicaron con los productos de la muestra, así mismo da una tabla de resultados con los errores en cada método y la conclusión del porqué se llega a escoger un modelo en específico (Juárez, Zuñiga, Flores y Partida, 2016).

En el presente documento se tomaron en cuenta las técnicas de series de tiempo que aparecen en este artículo como referencia para las pruebas al estudio de la posible técnica que se seleccionaría, analizando cuáles de éstas se aplicarían para el tipo de variable y el resultado que se esperaba obtener.

A continuación, se describe un artículo científico donde se muestra información de cómo crear un modelo de predicción.

### **Estrategias para la elaboración de modelos estadísticos de regresión**

El objetivo es crear un modelo de predicción siguiendo una serie de pasos plasmados en un algoritmo, de tal forma que se determina un objetivo principal para identificar los criterios de valoración para posteriormente seleccionar el método estadístico apropiado en función del resultado que se espera y el tipo de predicción. Así mismo se crea el modelo adecuado para evaluar el modelo final y presentar los resultados (Núñez, Steyerberg y Núñez, 2011).

Para este proyecto se enlistaron los pasos de un algoritmo para crear un modelo predictivo aplicable a dos variables de tipo cuantitativo, como es el tiempo y las ventas, evaluando la técnica de regresión lineal, suavización exponencial, método de Holt también llamado como suavización exponencial ajustada a la tendencia y por último promedios móviles.

En el siguiente artículo se hace una breve descripción de la teoría y las pruebas generadas con la técnica de arboles aleatorios.

### **Bosques aleatorios como extensión de los árboles de clasificación con los programas de R y Python**

Medina y Ñique (2017) desarrollaron un documento en el cual explican las diferencias entre los árboles de clasificación con los bosques aleatorios, haciendo énfasis en que si se tiene una cantidad de datos considerable y un buen número de variables, se puede llegar a hacer un análisis para generar predicción con un porcentaje de error bajo. Después de hacer las pruebas necesarias

se pudieron identificar las variables que tenían un peso más significativo, lo cual es de gran ayuda para poder hacer una posterior depuración de variables y un análisis más preciso. Cabe mencionar que al someter los datos a árboles aleatorios se obtuvo una buena tasa de error y queda como conclusión que si se ingresan nuevos casos asociados a las características del conjunto de datos, el error no incrementaría del porcentaje que resultó en la pruebas.

En este trabajo se tomó en cuenta el presente artículo ya que contiene conceptos de árboles aleatorios, así como su funcionamiento, la manera en que se debe aplicar y sus ventajas.

### **Software encontrado**

La siguiente información se refiere a software encontrado relacionado con el proyecto desarrollado.

#### Software Minitab

Minitab es un programa diseñado para ejecutar funciones estadísticas básicas y avanzadas. Explorando los datos desde una hoja de cálculo. Entre las muchas funciones que desempeña, se encuentran las siguientes:

Series de tiempo y pronósticos, que incluye:

- Gráficas de series de tiempo.
- Análisis de tendencia.
- Descomposición.
- Promedio móvil.
- Suavización exponencial.
- Método de Winters.
- Funciones de auto correlación, auto correlación parcial e intercorrelación.
- ARIMA.

En la figura 2.1 se muestra un ejemplo de lo que ofrece el software Minitab.

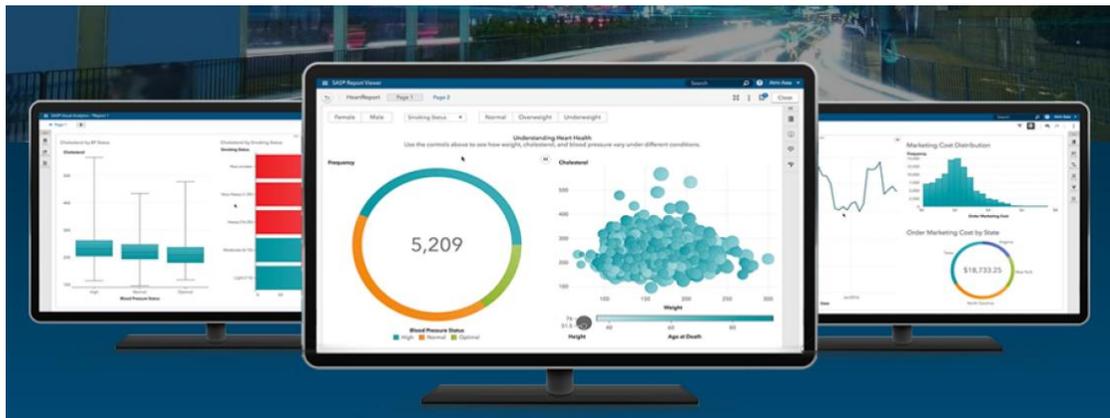


**Figura 2.1** Software Minitab (fuente: <http://www.minitab.com/es-mx/products/minitab/>)

## Software SAS

SAS (Statistical Analysis Software/Software de Análisis Estadístico) es un paquete de software desarrollado por SAS Institute para análisis avanzado, inteligencias empresariales, gestión de datos y análisis predictivo.

En la figura 2.2 se muestran algunos resultados sobre el software SAS.



**Figura 2.2** Software SAS (fuente: [https://www.sas.com/es\\_mx/home.html](https://www.sas.com/es_mx/home.html))

## Software EViews

EViews ofrece a investigadores académicos, corporaciones, agencias gubernamentales y estudiantes, acceso a herramientas estadísticas, de pronóstico y de modelado a través de una interfaz orientada a objetos innovadora y fácil de usar.

A continuación, se muestra una breve descripción de algunas de las áreas en las que el software EViews tiene relación.

#### Gobierno

EViews es el principal conjunto de herramientas de análisis y pronóstico macroeconómico utilizado por los bancos centrales, los bancos nacionales y las agencias gubernamentales de todo el mundo.

#### Académico

Los estudiantes pueden aprender rápidamente cómo importar datos, ejecutar regresiones y ver resultados gráficos, lo que permite a los profesores más tiempo para enseñar estadísticas y econometría en lugar de cómo usar el software.

#### Comercial

EViews ha sido una herramienta clave de análisis y pronóstico para empresas.

En la figura 2.3 se muestra la vista principal del software.



**Figura 2.3** Software EViews 10 (fuente: <http://www.eviews.com/home.html>)

#### Software IBM SPSS

Es un software de análisis predictivo que ofrece técnicas de recolección de datos, así como *insights* profundos y significativos, en un paquete fácil de usar que ayuda a encontrar nuevas oportunidades, mejorar la eficiencia y minimizar el riesgo.

Entre las opciones que contiene el software está el análisis estadístico y presentación, colección de datos, análisis, presentación de informes e implementación; modelado de predicción y minería de datos; administración e implementación de decisiones.

En la figura 2.4 se muestra una imagen representativa del software.



**Figura 2.4** Software IBM SPSS (fuente: <https://www.ibm.com/analytics/mx/es/technology/spss/>)

Una vez descritas las características del software encontrado, se puede mencionar que, en el software contemplado para el proyecto en cuestión, se consideraron dos variables específicas, que son el tiempo (meses) y la cantidad de ventas de cada producto. Se analizaron diferentes técnicas de estadística y minerías de datos para generar las predicciones, realizando diferentes pruebas. Como lenguaje de programación se pensó en utilizar Python, dado que es *Open Source* e incluye librerías para el manejo de dichas técnicas.

## 2.2 Marco teórico

Sarmiento (2008) plantea que las empresas que no son de un tamaño considerable tienden a tomar decisiones con base en su experiencia, sin tomar en cuenta algún método o modelo estadístico que los guíe por un camino más confiable, tomando así riesgos para sus negocios.

Por lo tanto, para obtener la predicción de ventas para el modelo de negocios de la empresa Mamá Coneja, se aplicó la teoría fundamental de minería de datos, que se define como el proceso que consiste en extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos. En ese sentido, la tarea

fundamental de la minería de datos es encontrar modelos inteligibles a partir de los datos recogidos (Vergaray, 2016).

La minería de datos se podría abstraer como la construcción de un modelo que ajustado a unos datos proporciona un conocimiento (Riquelme, Ruiz y Gilbert, 2006).

Para construir un modelo ajustado, existen diferentes técnicas dentro de la minería de datos, una de ellas es: series de tiempo, las cuales predicen valores futuros para la variable de interés basándose exclusivamente en el patrón histórico de esta variable, suponiendo que ese patrón histórico continuará. Existen diferentes técnicas, de las cuales sólo se aplicaron promedio móvil simple, suavización exponencial y suavización exponencial ajustada a la tendencia (Juárez, Zuñiga, Flores y Partida, 2016), cuyo funcionamiento se describirá en el apartado 2.3 Marco conceptual.

Existen otras técnicas para predicción como los son los arboles aleatorios, como se menciona en la página IBM Knowledge Center (s. f.), este método utiliza la partición reiterada para dividir los registros de entrenamiento en segmentos con valores de salida similares, esta técnica también se especifica en el siguiente subtema.

## **2.3 Marco conceptual**

Los siguientes conceptos están relacionados con el tema desarrollado en el proyecto.

### **Modelo de negocios**

De acuerdo con Barrientos y Rumiany (2005), un modelo de negocio representa una empresa. Se trata de un conjunto de variables que permiten simplificar, generando una abstracción con propósitos útiles. Cada una de las variables seleccionadas, define cada parte del negocio y el comportamiento de éste, puede ser interpretado a través del seguimiento de cada variable. El principal objetivo es describir los elementos centrales.

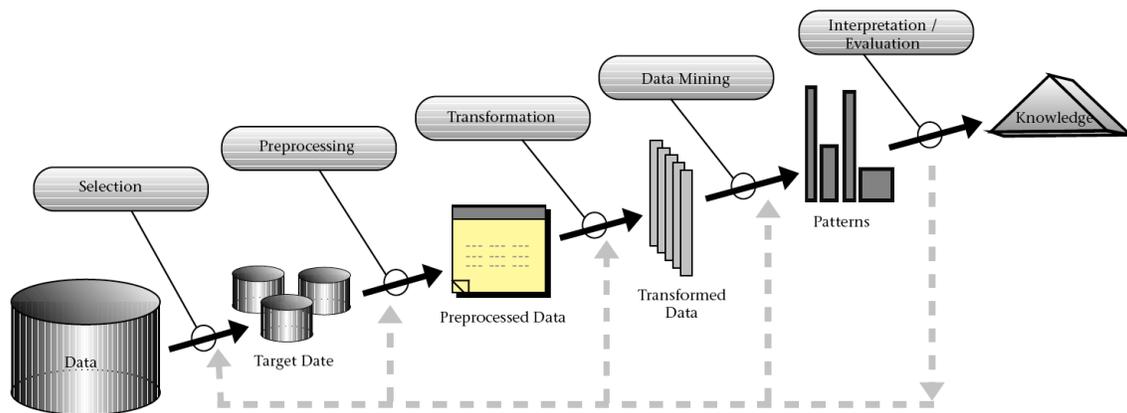
El objetivo de construir cualquier modelo es la expresión de una idea en términos más formales.

Otro concepto planteado por Preciado y Oliva (2011) se define de la siguiente manera: un modelo de negocio, se relaciona con la forma en la que un negocio hace dinero, entrega valor a

los clientes; hace referencia a la lógica de la compañía, es estático, no reflexivo, no es una descripción de la receta para el cambio y se centra en la oportunidad.

## Metodología KDD

Descrito por Vergaray (2016), las etapas para la realización de la minería de datos siempre son las mismas independientemente de la técnica específica a usar. El conjunto de las partes de este proceso, también conocido por sus siglas en inglés KDD (Knowledge Discovery in Databases o descubrimiento de conocimiento en bases de datos), se muestra a continuación en la figura 2.5.



**Figura 2.5** Descripción general de los pasos que constituyen el proceso KDD (Fuente: <https://www.kdnuggets.com/2016/03/data-science-process-rediscovered.html/2>)

Continuando con lo que se plantea por Vergaray (2016), el proceso de KDD es interactivo e iterativo que implica numerosos pasos que se resumen así:

1. Aprendizaje del dominio de la aplicación: este paso incluye adquirir el conocimiento previo relevante y el planteo de los objetivos de la aplicación. Para este punto se tomó en cuenta el siguiente concepto descrito por Grupo IGN (2017).

### Análisis predictivo

Es usar una variedad de técnicas para analizar datos y decidir qué acción es mejor tomar.

Existen numerosas técnicas para el análisis predictivo, como data mining, estadística clásica, machine learning o deep learning, a veces, el modelado predictivo y la analítica predictiva se consideran lo mismo, pero no lo son. Los modelos predictivos se utilizan sólo para crear una representación de conclusiones e información, mientras que los análisis predictivos son lo que crean esa conclusión.

Para los puntos 2 al 4 Vergaray (2016) plantea lo siguiente:

2. Creación de un conjunto de datos de destino: se escoge el conjunto de datos o se elige el subconjunto de variables o muestras de datos en los cuales el descubrimiento se va a realizar.
3. Limpieza de datos y pre-procesamiento: en éste se dan las operaciones básicas como la eliminación de ruido, el obtener la información necesaria para modelar, la determinación de estrategias para el manejo de los campos de datos.
4. Reducción de datos y de proyección: en este punto se lleva a cabo la búsqueda de las características útiles para representar los datos, dependiendo del objetivo y de la tarea.

En el proyecto se aplicaron varias técnicas de minería de datos para generar un modelo de predicción. Se hizo el análisis de la información, existente en un respaldo de la base de datos siendo ésta estructurada. Algunos conceptos importantes relacionados a este punto son los siguientes:

### **Técnicas de minería de datos**

Dentro de las técnicas usadas en la minería de datos, se encuentra una o más de las siguientes funciones, clasificación, regresión, clustering, resumen, recuperación de imágenes, extracción de reglas, etc., utilizadas para predicción, mientras que las técnicas de agrupamiento, la asociación y el descubrimiento de secuencias entre otras son más apropiados para describir relaciones existentes en los datos (Riquelme, Ruiz, y Gilbert, 2006).

## **Tareas predictivas y descriptivas**

De acuerdo con Riquelme, Ruiz, y Gilbert (2006), las tareas que se llevan a cabo con la minería de datos se pueden clasificar como predictivas y descriptivas. En las tareas predictivas cada elemento de la base de datos se caracteriza por tener unos parámetros de entrada y un parámetro de salida. El objetivo es predecir el valor del parámetro de salida utilizando la información proporcionada por los parámetros de entrada.

Dentro de las tareas predictivas existen dos tipos:

- Clasificación. El objetivo que se persigue es maximizar el poder de predicción de la clasificación de nuevos elementos para los cuales la clase es desconocida.
- Regresión. El valor asociado a cada elemento es un número real. El objetivo, maximizar la capacidad de predicción de este valor para un elemento nuevo a través de una función real que se debe aprender.

En cuanto a las tareas descriptivas, los elementos de la base de datos sólo tienen atributos de entrada y el objetivo es el de agruparlos maximizando la similitud entre los elementos de un mismo grupo y minimizando dicha similitud entre los diferentes grupos (Basauri, 2015).

5. Elección de la función de minería de datos: a través de esto se decide el propósito del modelo derivado por medio del algoritmo de minería de datos (Vergaray, 2016). Para seleccionar el tipo de técnica, se hizo un análisis teórico de los componentes que contienen estos algoritmos.

## **Minería de datos**

Se refiere al proceso de extraer conocimientos de bases de datos. Su objetivo es descubrir situaciones anómalas y/o interesantes, tendencias, patrones y secuencias en los datos. Minería de datos es una etapa dentro del proceso completo del descubrimiento del conocimiento, éste intenta obtener patrones o modelos a partir de los datos recopilados. Los algoritmos de minería de datos suelen tener tres componentes:

- El modelo, que contiene parámetros que han de fijarse a partir de los datos de entrada.
- El criterio de preferencia, que sirve para comparar modelos alternativos.
- El algoritmo de búsqueda, que viene a ser como cualquier otro programa de Inteligencia Artificial (IA).

Las aplicaciones de técnicas de minería de datos en grandes bases de datos persiguen los siguientes resultados: clasificación, regresión, agrupamiento (clustering), resumen, modelado de dependencias y análisis de secuencias (Valcárcel, 2004).

6. Elección del algoritmo de minería de datos: en esta parte se eligen los métodos que van a ser utilizados para la búsqueda de patrones en el conjunto de los datos, tales como decidir qué modelos y parámetros pueden ser apropiados y establecer con qué método de minería estos pueden ser aplicados (Vergaray, 2016). A continuación, las técnicas descritas, son algunas de las que se seleccionaron para las pruebas por el tipo de variables.

### **Técnica de regresión lineal**

Los modelos de regresión son el pilar de la analítica predictiva. El enfoque se basa en el establecimiento de una ecuación matemática como modelo para representar las interacciones entre las diferentes variables en consideración. El modelo de regresión lineal analiza la relación existente entre: la variable dependiente o de respuesta y un conjunto de variables independientes o predictores (Espino, 2017).

De acuerdo con Sarmiento (2008) la ecuación 1 representa la fórmula de regresión lineal, la cual está dada por:

$$\hat{Y} = \beta_0 + \beta_1 * X + \varepsilon. \quad (1)$$

Donde:

$\hat{Y}$  : variable dependiente que se pretende estimar.

$\beta_0$ : parámetro de regresión correspondiente al valor estimado. El punto donde se alarga la recta que intersecta con el eje y.

$\beta_1$ : parámetro de regresión correspondiente al crecimiento o disminución del valor. También representa la inclinación de la recta con el eje horizontal.

$X$ : variable independiente.

$\varepsilon$ : componente de error aleatorio.

### Series de tiempo

Los modelos de series de tiempo predicen valores futuros para la variable de interés basándose exclusivamente en el patrón histórico de ésta, suponiendo que ese patrón histórico continuara. Los métodos de series de tiempo, consisten en el uso de métodos analíticos, para poder determinar las tendencias y las variaciones estacionales (Juárez, Zúñiga, Flores y Partida, 2016).

Algunas de las técnicas de series de tiempo como lo mencionan Juárez, Zúñiga, Flores y Partida (2016) se enlistan a continuación:

- Promedio móvil simple. Es una técnica adecuada cuando se han utilizado las fuerzas que generan las series a pronosticar y cuando por lo general, no cambia el entorno donde existe la serie. En éste se utilizan datos recientes para reducir el efecto de las fluctuaciones aleatorias y responder al cambio en el proceso de una manera más rápida. Así el promedio móvil está dado por la suma de los últimos  $N$  datos, como se muestra en la ecuación 2.

$$M_T = \frac{1}{N}(d_{T-N} + d_{T-N+2} + \dots + d_T) = \frac{1}{N}. \quad (2)$$

Donde  $M_T$  representa el valor final de la predicción;  $N$  se refiere al número de periodos que se quiere considerar en el promedio móvil;  $d_t$  es la demanda histórica en el periodo  $T$ , y  $T$  es el proceso en el que se encuentra el periodo.

- Suavización exponencial. Está basada en el análisis de los errores de los pronósticos, en el que se asigna un mayor peso al último valor en la serie de tiempo y después, progresivamente, se asignan pesos más pequeños a los valores más antiguos como se muestra en la ecuación 3.

$$\hat{X}_t = \hat{X}_{t-1} + (\alpha \cdot (X_{t-1} - \hat{X}_{t-1})) \quad (3)$$

Donde  $\hat{X}_t$  es la variable a la cual se le asigna el resultado en el periodo t; donde  $\hat{X}_{t-1}$  se refiere al pronóstico de ventas del periodo t-1;  $X_{t-1}$  son las ventas reales en el periodo t-1, y  $\alpha$  es el coeficiente de suavización (su valor debe estar entre 0 y 1).

- Suavización exponencial ajustada a la tendencia. Usa valores recientes en la serie de tiempo para estimar cualquier tendencia ascendente o descendente, adicionalmente, tiende a seguir un número considerable pero definido de periodos, en lo que la tendencia se define como el cambio promedio del siguiente valor en una serie de tiempo (ecuación 4 a la 6).

Pronostico del periodo t,

$$\hat{X}_t = \hat{Z}_t + T_t \quad (4)$$

Serie de suavización exponencial,

$$\hat{Z}_t = \alpha(\hat{X}_{t-1}) + [(1 - \alpha)(\hat{Z}_{t-1} + T_{t-1})] \quad (5)$$

Estimado de la tendencia,

$$T_t = \beta(\hat{Z}_t - \hat{Z}_{t-1}) + [(1 - \beta)(T_{t-1})] \quad (6)$$

Donde  $\hat{X}_t$  es el pronóstico del periodo t;  $\hat{Z}_t$  representa el valor de la suavización de la serie;  $\hat{Z}_{t-1}$  es el valor de la suavización de la serie, pero del periodo t-1;  $T_t$  corresponde al resultado obtenido de calcular el estimado de la tendencia del periodo t;  $\hat{X}_{t-1}$  es el pronóstico del periodo t-1;  $T_{t-1}$  corresponde a la tendencia

del periodo  $t-1$ ;  $\alpha$  es el coeficiente de suavización (su valor debe estar entre 0.0 y 1.0), y  $\beta$  es el coeficiente de suavización de la tendencia (entre 0.0 y 1.0).

### **Arboles aleatorios**

Planteado por Alvaer (2018), los árboles aleatorios se construyen de la siguiente manera:

- El número de casos en el conjunto de entrenamiento es  $N$ . Una muestra de esos  $N$  casos se toma aleatoriamente con reemplazo, significa que un tercio de los datos no se usan para el entrenamiento y pueden ser usados para test. Esta muestra es el conjunto de entrenamiento para construir el árbol  $i$ .
- Si existen  $M$  variables de entrada, un número  $m < N$  se especifica tal que, para cada nodo,  $m$  variables se seleccionan aleatoriamente de  $M$ . La mejor división de estos  $m$  atributos es usado para ramificar el árbol. El valor de  $M$  se mantiene constante durante la generación de todo el bosque.
- Cada árbol crece hasta su máxima extensión posible y no hay proceso de poda.
- Nuevas instancias se predicen a partir de la agregación de las predicciones de los  $x$  árboles (es decir, mayoría de votos para clasificación, promedio para regresión).

De acuerdo con Vergaray (2016), plantea la etapa número 7 de la siguiente manera:

7. Realización de minería de datos: aquí se da la búsqueda de patrones de interés dentro de una forma de representación particular o el conjunto de tales representaciones. Siguiendo los pasos del procesamiento de minería, como se enlistan a continuación, se evalúa los resultados del modelo.

### **Procesamiento de la minería de datos**

En general el proceso de la minería de Datos itera a través de cinco pasos básicos descritos por Echeverri, Retamoza, De La Rosa, Barros, Álvarez y Guerrero, (2013) como se muestra a continuación:

- a) Selección de datos: consiste en buscar el objetivo y la herramienta del proceso de minería, identificando los datos a ser extraídos, buscando los atributos apropiados de entrada y la información de salida para representar la tarea.
- b) Transformación de datos: incluye organizar los datos en la forma deseada, convirtiendo un tipo de datos en otro definiendo nuevos atributos, reduciendo la dimensión de los datos, removiendo ruidos, “*outliers*”, normalizando.
- c) Minería de datos: los datos transformados son analizados o minados, utilizando una o más técnicas para extraer patrones de interés.
- d) Interpretación de resultados y validación: para comprender el significado del conocimiento extraído y su rango de validez, la aplicación de minería de datos prueba su robustez, utilizando métodos de validación establecidos y probándolo con datos diferentes a los utilizados para crear el modelo.
- e) Incorporación del conocimiento: presentación de los resultados del modelo para poder comprobar o resolver conflictos con creencia o resultados anteriores y aplicar el nuevo modelo.

Los pasos 8 y 9 describen las últimas dos etapas de la metodología KDD descrita por Vergaray (2016):

- 8. Interpretación: en este paso se realiza la interpretación de los patrones detectados y, en algunos casos, también se vuelve a cualquiera de los momentos anteriores. Luego se hace la visualización de los posibles patrones extraídos, la eliminación de los patrones redundantes o irrelevantes y la traducción de los patrones útiles a términos que sean comprensibles para los usuarios. Una forma de visualizar los resultados es medir el error que se tuvo en cada uno de los modelos, con esto se puede apreciar de una forma más entendible los resultados finales.

### **Error Porcentual Absoluto Medio MAPE**

De acuerdo con Silva (2013), menciona que, en ocasiones, es más útil calcular los errores de pronósticos en términos de porcentajes en lugar de cantidades. MAPE, se calcula al

encontrar el error absoluto en cada periodo, dividiéndolo entre el valor real observado para ese periodo y luego promediando los errores porcentuales absolutos. También puede ser utilizada para comparar la precisión de las mismas o diferentes técnicas en dos series totalmente distintas, a continuación, se presenta la ecuación (7).

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t}. \quad (7)$$

Donde la diferencia de la venta real menos la predicción está representada por  $|Y_t - \hat{Y}_t|$ , que se divide entre la demanda real dada por  $Y_t$ ; siendo  $t$  la representación del periodo al que pertenecen; así mismo realizando la sumatoria desde  $t-1$  hasta  $n$  que es el número de periodos pronosticados.

9. Utilización del conocimiento descubierto: este último paso implica la incorporación de los conocimientos extraídos en pro del rendimiento del sistema, tomando acciones basadas en dicho conocimiento o simplemente documentándolo e informándolo a las partes interesadas. Para la presentación del conocimiento se puede mostrar el resultado programando en lenguajes de programación como el que se describe a continuación:

### **Python**

Planteado por Challenger, Díaz y Becerra (2014) los lenguajes de programación son la herramienta básica de construcción de programas. Python ha ido ganando adeptos en comunidades como la de software libre, científica y educacional, por su sencillez y posibilidad de concentrarse en los problemas actuales.

Python cuenta con facilidades para la Programación Orientada a Objetos, imperativa y funcional, por lo que se considera un lenguaje multi-paradigmas. Fue basado en el *lenguaje ABC* y se dice que fue influenciado por otros como: *C*, *Algol 60*, *Modula-3* e *Icon* según su propio autor.

Es un lenguaje de alto nivel ya que contiene implícitas algunas estructuras de datos como listas, diccionarios, conjuntos y tuplas, que permiten realizar algunas tareas complejas en pocas líneas de código y de manera legible.

## **2.4 Marco contextual**

El proyecto está enfocado a las abarroteras del estado de Jalisco, específicamente en las abarroteras de la línea Mamá Coneja S.A de C.V., ubicadas en Cd. Guzmán, Jalisco, seleccionando la ubicada por la calle Primero de Mayo.

### **3. Capítulo III. Marco Metodológico**

Este capítulo presenta la metodología que se utilizó en el proyecto, el tipo de investigación, la muestra, el muestro, el universo en el que se estudió, así como los instrumentos que se utilizaron.

### **3.1 Tipo de Investigación**

El tipo de investigación que se utilizó fue aplicada por que se centra en encontrar las técnicas para lograr un objetivo en particular. El ámbito al que se aplico es específico y delimitado, se abordó un problema concreto.

Se aplicó también la investigación experimental ya que se manipularon más de una variable, que al tomarlas en cuenta para la predicción tuvo un efecto determinado.

### **3.2 Universo, población o unidades de análisis**

La población de estudio que se consideró para el proyecto son las abarroteras de Cd. Guzmán, específicamente las de la empresa Mamá Coneja.

### **3.3 Criterios de inclusión/exclusión**

De las empresas de Mamá Coneja que se encuentran en México se seleccionaron las que están ubicadas en Cd. Guzmán, realizando un estudio sobre las ventas de cada uno de los negocios. Se tomó como muestra la tienda que presentó mayor volumen en sus ventas de acuerdo al análisis realizado. Para la selección de los productos, se hizo una consulta de los 20 productos más vendidos para identificar qué línea de artículos es la de mayor movimiento. Optando por la línea de cereales y semillas de la cual se tomaron seis productos como muestra.

### **3.4 Muestreo**

Se utilizó el muestreo no probabilístico, porque la muestra fue dirigida a ese modelo de negocios, ya que se tiene un convenio con la empresa.

### **3.5 Muestra**

Se tomó como muestra la empresa de Mamá Coneja, S.A. de C.V., específicamente la que se encuentra ubicada por la calle Primero de Mayo, ya que el proyecto fue desarrollado a petición de la misma. La muestra sobre los productos no es aleatoria, ya que no se seleccionaron al azar,

se hizo una consulta de los productos más vendidos y con base en esta se hizo la debida selección de la línea de productos cereales y semillas.

### **3.6 Instrumentos**

Los tipos de instrumentos que se necesitaron para el desarrollo del proyecto, por una parte, fue la observación de los procesos y toma de nota de los mismos, y por la otra, el uso de formatos con los que se lleva a cabo el control de las ventas.

### **3.7 Aparatos**

Fue necesario el uso de una computadora, así como dispositivos de almacenamiento secundario para realizar respaldos de la base de datos de la empresa.

### **3.8 Procedimientos**

El procedimiento que se llevó a cabo fue el siguiente:

- 1) Investigar el estado del arte relacionado con este proyecto. Donde se investigaron proyectos acordes al que se desarrolló, así como también, se analizaron las diferentes metodologías y técnicas aplicables al proyecto, lo cual puede observarse con más detalle en el capítulo II.
- 2) Basado en la investigación que se realizó previamente y al resultado objetivo que se planteó, se determinó seguir la metodología KDD, donde se analizó la información que se tuvo, en este caso una base de datos equivalente a un respaldo de cuatro años; de ahí se seleccionó la información que se necesitó para las variables, se hizo un pre-procesamiento para posteriormente transformarla y aplicarla a una técnica de minería de datos.

De acuerdo con Vergaray (2016), las etapas de esta metodología se nombran de la siguiente manera:

- Aprendizaje del dominio de la aplicación.

- Creación de un conjunto de datos de destino.
- Limpieza de datos y pre procesamiento.
- Reducción de datos y de proyección.
- Elección de la función de minería de datos.
- Elección del algoritmo de minería de datos.
- Realización de minería de datos.
- Interpretación.
- Utilización del conocimiento descubierto.

A continuación, se describen las actividades realizadas en cada una de las etapas del proyecto:

- a) Aprendizaje del dominio de la aplicación. Primero se realizó la investigación sobre la aplicación en la cual está desarrollada la base de datos, así mismo, el manejador de la misma, con el fin de poder interpretar de una forma más fácil la información.
- b) Creación de un conjunto de datos de destino. Se especificaron dos variables cuantitativas con las que se trabajó y con base en éstas se extrajo la muestra descrita en el apartado 3.5 Muestra.
- c) Limpieza de datos y pre-procesamiento. Se especificó cuál sería la variable explicativa que representó al periodo y la variable dependiente que fue el total de ventas, donde se estableció predecir el total de ventas en función de un periodo seleccionado. Se instalaron los programas necesarios para el manejo de la base de datos, la cual está desarrollada en *Firebird*, analizando toda la estructura, se identificaron las tablas necesarias con sus respectivas relaciones para la consulta de la información, las consultas se realizaron en el manejador de base de datos *IBexpert*.

- d) Reducción de datos de proyección. En IBexpert se desarrollaron las consultas específicas, donde se extrajo la información dentro de un periodo específico de las ventas. Cabe mencionar que la información consultada fue de un respaldo de cuatro años que proporcionó la empresa. Teniendo la información de los productos más vendidos, se seleccionaron cinco dentro los diez más vendidos y uno del nivel 160 en la lista de las ventas, para ver el comportamiento y la adaptabilidad de las pruebas de los modelos que se plantearon, tanto en los productos de mayor movimiento, como los que se encuentran en una parte más inferior.
- e) Elección de la función de minería de datos. Se investigaron técnicas de minería de datos relacionadas con predicción, con el fin de que fueran aplicables al problema que se planteó.
- f) Elección del algoritmo de minería de datos. Para evaluar las técnicas de minería de datos se desarrolló un algoritmo, el cual se siguió para poder valorar qué modelo fue el apropiado, se propuso regresión lineal, promedio móvil, suavización exponencial, suavización ajustada a la tendencia y arboles aleatorios descritas en el marco conceptual. La tabla 3.1 muestra los pasos que se siguieron para llegar al modelo de predicción:

<b>Pasos para el establecimiento del modelo</b>
<b>Determinación del objetivo del modelo</b>
Modelo de predicción
<b>Elección del método estadístico apropiado en función del resultado y el tipo de predicción</b>
Regresión lineal
Promedio móvil
Suavización exponencial
Suavización ajustada a la tendencia
Arboles aleatorios
<b>Creación del modelo adecuado, incluida la validación interna</b>

Seleccionar las variables correctas
Acotar las variables al método estadístico seleccionado
<b>Evaluar el rendimiento del modelo</b>
Parámetros a evaluar
Validación interna: parámetros a evaluar con la información del respaldo de la base de datos
Validación externa: los mismos parámetros, pero con datos externos
<b>Presentación de los resultados</b>

**Tabla 3.1** Pasos para el modelo de predicción (Fuente: elaboración propia)

- g) Realización de minería de datos. Al aplicar la información de las dos variables seleccionadas a cada uno de las técnicas para la búsqueda de patrones, se ve reflejado cada resultado en gráficas generadas en cada modelo en el programa Excel, haciendo una comparación con las ventas reales, dichos resultados se pueden observar en la sección de resultados.
- h) Interpretación. Después de analizar las gráficas en cada uno de los modelos con la información procesada, se intentó encontrar algún tipo de patrón que se pudiera seguir en el transcurso del tiempo, dando un resultado no favorable para ninguno de los modelos planteados, tanto de regresión lineal, así como de series temporales o arboles aleatorios.
- i) Utilización del conocimiento descubierto. Se programó una conexión desde la base de datos de Firebird con el lenguaje de Python, esto con el fin de desarrollarlo en este ambiente, pero no se continuo con esta parte ya que no se encontró un modelo que se adaptara al objetivo planteado.
- 3) Difundir los resultados obtenidos. Se hizo un artículo titulado “Análisis del comportamiento de la predicción de ventas con diferentes técnicas de minería de datos”, enviado a la Revista Iberoamericana de Ciencias en donde se describe como se abordó el proyecto, así como los resultados obtenidos y recomendaciones de los factores que se deben tomar en cuenta para la predicción de ventas en una abarrotera.

#### **4. Capítulo IV. Resultados**

La información del siguiente capítulo muestra paso a paso las pruebas realizadas a los datos de entrada en cada una de las técnicas implementadas en el presente proyecto, así como las gráficas de los resultados obtenidos.

## 4.1 Pruebas realizadas y resultados obtenidos

En este apartado se describirán las pruebas que se realizaron con cada uno de los métodos, así como los resultados obtenidos.

### 4.1.1 Regresión lineal

En esta técnica, primero se calcularon tanto la media aritmética como la media geométrica, con base en la información de las ventas que se extrajo de la base de datos del respaldo antes mencionado, acomodándola por año y mes. Teniendo los resultados de cada una de las medias, se pudo observar cuál de las dos tuvo un índice de error más bajo, seleccionando la media aritmética. Para realizar el análisis de regresión lineal, se utilizó la herramienta de Microsoft Excel, a continuación, se enlistan cada uno de los pasos:

1. Se seleccionó de la cinta de opciones “**Datos**”.
2. En el apartado de herramientas de datos se seleccionó “**Análisis de datos**”, abriendo una ventana emergente, con las funciones de análisis disponibles.
3. Desplazándose por las diferentes opciones hasta llegar a la titulada como “**Regresión**”, dando clic en aceptar.
4. Ahí se seleccionó la información en cada una de las variables, tanto la variable dependiente que representa el tiempo, así como la variable independiente que fueron las ventas.
5. Se activaron las casillas de los residuales que se quieren mostrar en los resultados, en este caso fueron, “**Residuos**”, “**Residuos estándares**”, “**Gráficos residuales**”, “**Curva de regresión ajustada**”, así como el lugar en donde se quiere mostrar todos los cálculos de la regresión.
6. Se mostraron todos los resultados de la regresión en una hoja Excel nueva.

A continuación, en la tabla 4.1 a 4.6 se pueden observar los datos que se tomaron de cada producto como entrada para aplicar el método descrito, donde se puede apreciar que en la

primera columna se encuentran los meses representados por número, en la siguiente columna los meses por su respectivo nombre, posteriormente cada uno de los años con las ventas que se tuvieron en cada uno de los meses. En las figuras 4.1 a 4.6 se muestran gráficamente los resultados de cada uno de ellos.

Producto 1							
No. de mes	Meses	2012	2013	2014	2015	Media aritmética	Predicción
1	Enero	655.200	3,864.809	6,703.468	3,138.025	3590.376	3021.195
2	Febrero	383.540	3,521.028	6,792.435	4,367.105	3766.027	3076.943
3	Marzo	315.100	3,936.770	4,259.319	4,063.337	3143.631	3132.691
4	Abril	454.930	4,438.960	2,847.660	5,161.090	3225.660	3188.439
5	Mayo	727.342	3,581.985	2,147.960	4,696.040	2788.332	3244.187
6	Junio	838.636	3,176.277	1,337.390	4,469.114	2455.354	3299.935
7	Julio	910.672	2,241.982	1,637.455	5,208.010	2499.530	3355.683
8	Agosto	2,231.026	2,684.181	2,364.480	6,334.435	3403.531	3411.431
9	Septiembre	2,246.059	2,653.173	2,906.850	4,831.320	3159.351	3467.179
10	Octubre	2,269.087	3,406.773	3,341.440	4,390.220	3351.880	3522.927
11	Noviembre	3,776.791	4,458.848	3,092.945	4,139.664	3867.062	3578.675
12	Diciembre	4,694.942	6,065.552	4,204.550	3,766.883	4682.982	3634.423

**Tabla 4.1** Información de ventas correspondiente a cuatro años, la media aritmética y la predicción con regresión lineal del producto 1 (Fuente: elaboración propia)

Como se puede apreciar en la figura 4.1 existe una notable diferencia entre la venta real con respecto a la predicción generada, puede haber una diferencia desde 10 productos hasta 1000 aproximadamente, lo que demuestra que la predicción no se asemeja a los valores reales y el modelo no es preciso. En las siguientes figuras de la predicción de regresión lineal se puede apreciar algo similar.

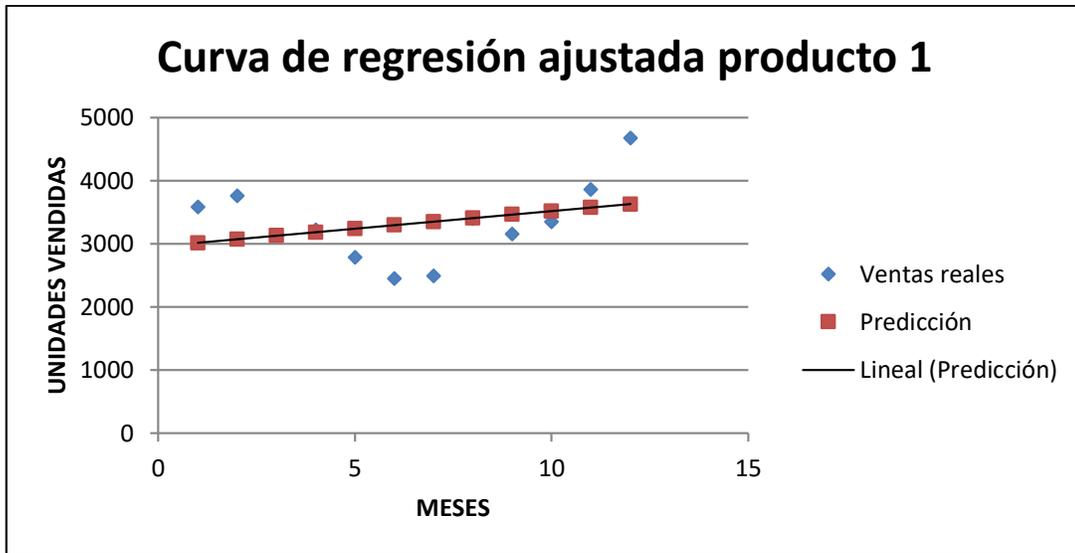
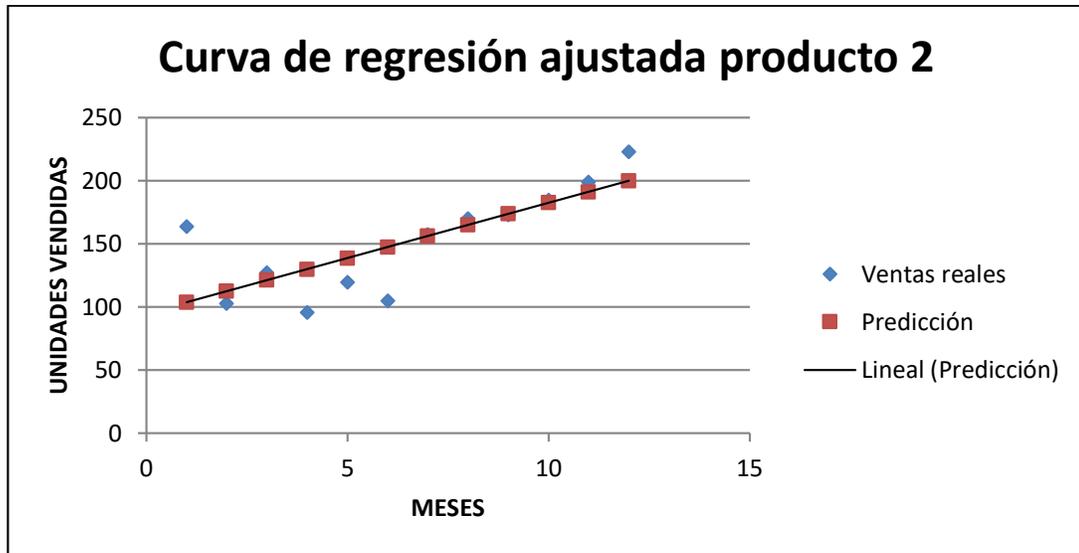


Figura 4.1 Gráfica de regresión lineal del producto 1 (Fuente: elaboración propia)

Producto 2							
No. de mes	Meses	2012	2013	2014	2015	Media aritmética	Predicción
1	Enero	35.810	165.865	129.055	325.045	163.944	103.785
2	Febrero	92.265	73.335	89.470	156.990	103.015	112.541
3	Marzo	74.890	93.475	86.200	254.980	127.386	121.297
4	Abril	37.430	75.810	87.045	184.135	96.105	130.054
5	Mayo	76.015	110.465	98.665	194.140	119.821	138.810
6	Junio	76.575	122.620	100.775	120.350	105.080	147.566
7	Julio	79.592	142.500	169.055	239.595	157.686	156.322
8	Agosto	102.500	109.360	250.020	219.545	170.356	165.078
9	Septiembre	121.345	144.910	119.325	307.322	173.226	173.834
10	Octubre	113.675	107.200	256.805	261.349	184.757	182.591
11	Noviembre	156.245	193.955	260.645	185.349	199.049	191.347
12	Diciembre	168.725	136.170	339.260	247.460	222.904	200.103

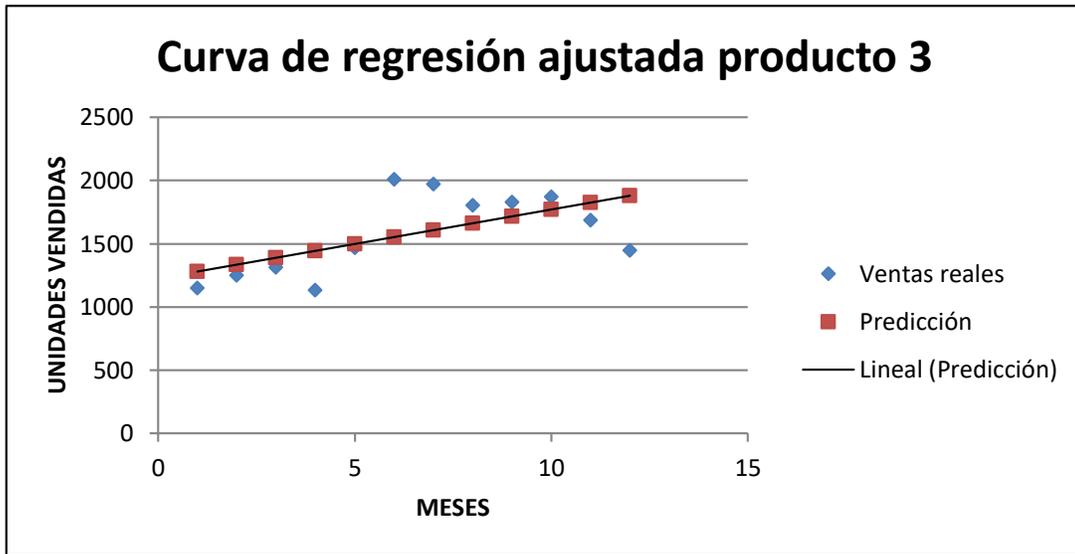
Tabla 4.2 Información de ventas correspondiente a cuatro años, la media aritmética y la predicción con regresión lineal del producto 2 (Fuente: elaboración propia)



**Figura 4.2** Gráfica de regresión lineal del producto 2 (Fuente: elaboración propia)

Producto 3							
No. de mes	Meses	2012	2013	2014	2015	Media aritmética	Predicción
1	Enero	877.475	711.224	924.346	2,079.590	1148.159	1278.063
2	Febrero	981.960	574.512	1,535.168	1,903.340	1248.745	1332.720
3	Marzo	821.910	464.607	2,198.707	1,771.185	1314.102	1387.378
4	Abril	703.585	496.100	1,807.395	1,526.770	1133.463	1442.036
5	Mayo	802.495	662.535	2,358.810	2,051.745	1468.896	1496.693
6	Junio	825.535	613.902	3,834.730	2,772.625	2011.698	1551.351
7	Julio	973.321	597.290	4,045.170	2,272.645	1972.107	1606.008
8	Agosto	955.902	469.020	3,975.075	1,826.265	1806.566	1660.666
9	Septiembre	782.464	547.000	3,994.445	1,995.079	1829.747	1715.323
10	Octubre	795.816	589.110	4,662.375	1,447.137	1873.610	1769.981
11	Noviembre	653.422	615.016	4,187.920	1,300.761	1689.280	1824.638
12	Diciembre	712.996	1,086.922	2,499.915	1,491.294	1447.782	1879.296

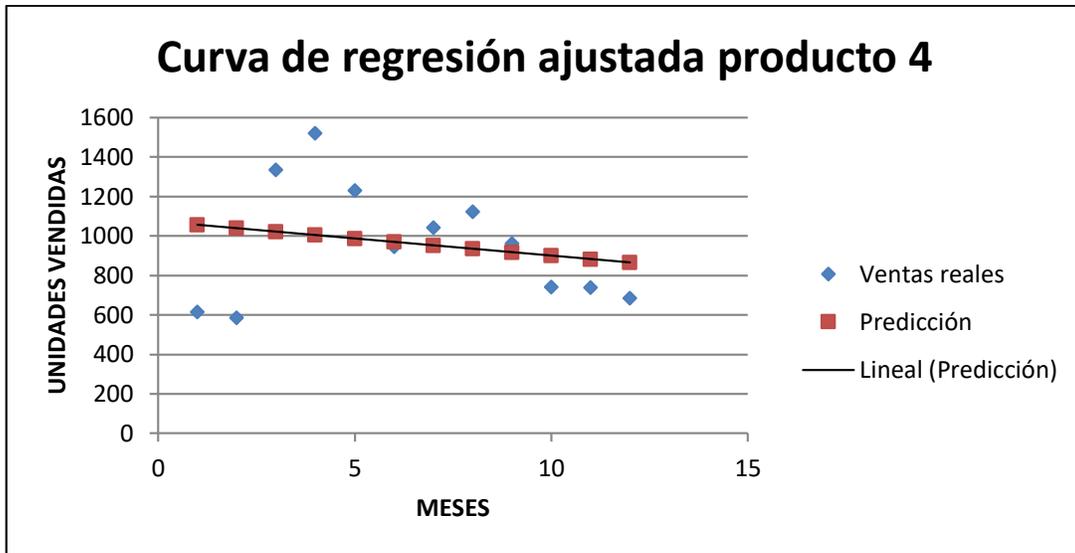
**Tabla 4.3** Información de ventas correspondiente a cuatro años, la media aritmética y la predicción con regresión lineal del producto 3 (Fuente: elaboración propia)



**Figura 4.3** Gráfica de regresión lineal del producto 3 (Fuente: elaboración propia)

Producto 4							
No. de mes	Meses	2012	2013	2014	2015	Media aritmética	Predicción
1	Enero	1,015.855	189.018	442.812	813.820	615.376	1057.799
2	Febrero	675.845	221.767	589.451	861.970	587.258	1040.299
3	Marzo	1,025.660	281.487	2,850.007	1,187.765	1336.230	1022.798
4	Abril	1,107.880	367.225	3,411.831	1,198.985	1521.480	1005.298
5	Mayo	517.559	291.810	3,348.401	765.925	1230.924	987.797
6	Junio	550.378	376.840	2,323.670	538.045	947.233	970.297
7	Julio	473.712	442.105	2,729.050	528.710	1043.394	952.796
8	Agosto	524.433	475.015	2,917.445	580.110	1124.251	935.296
9	Septiembre	685.955	309.190	2,375.210	478.127	962.121	917.795
10	Octubre	345.096	300.345	1,889.256	432.906	741.901	900.295
11	Noviembre	158.481	371.610	2,034.390	399.501	740.996	882.794
12	Diciembre	152.195	480.570	1,637.900	478.908	687.393	865.294

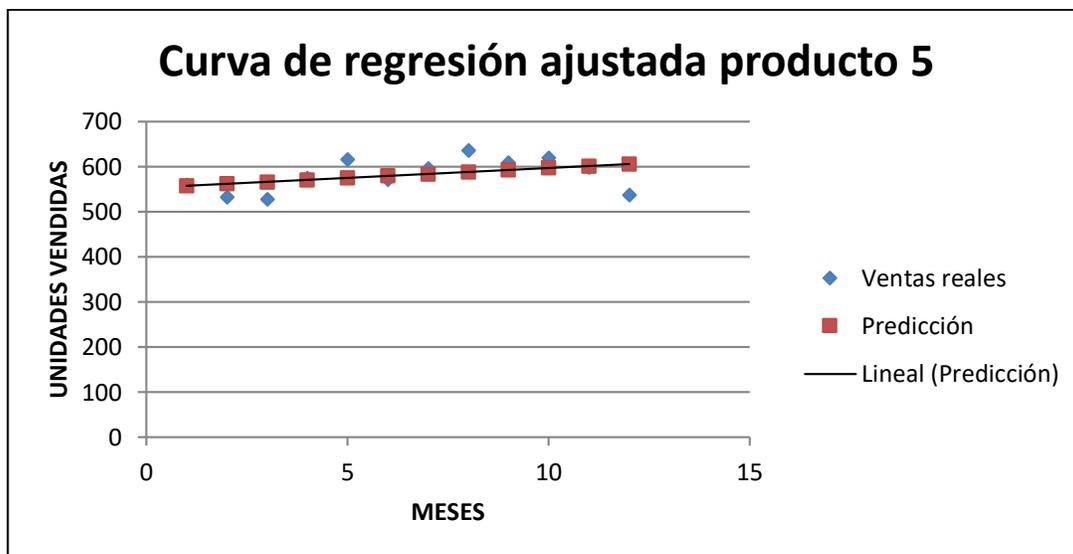
**Tabla 4.4** Información de ventas correspondiente a cuatro años, la media aritmética y la predicción con regresión lineal del producto 4 (Fuente: elaboración propia)



**Figura 4.4** Gráfica de regresión lineal del producto 4 (Fuente: elaboración propia)

Producto 5							
No. de mes	Meses	2012	2013	2014	2015	Media aritmética	Predicción
1	Enero	520.327	528.709	521.180	667.653	559.467	557.612
2	Febrero	517.950	488.980	554.805	568.808	532.636	562.015
3	Marzo	553.030	424.148	629.295	504.435	527.727	566.418
4	Abril	546.576	639.684	556.010	557.506	574.944	570.821
5	Mayo	538.926	642.712	654.020	631.671	616.832	575.224
6	Junio	540.110	553.660	603.010	590.556	571.834	579.628
7	Julio	530.333	576.675	651.705	626.914	596.407	584.031
8	Agosto	516.063	568.300	842.682	617.714	636.190	588.434
9	Septiembre	503.779	499.965	714.665	716.994	608.851	592.837
10	Octubre	626.752	517.660	712.700	623.224	620.084	597.240
11	Noviembre	493.036	504.365	769.611	629.552	599.141	601.643
12	Diciembre	497.439	492.080	590.047	571.779	537.836	606.046

**Tabla 4.5** Información de ventas correspondiente a cuatro años, la media aritmética y la predicción con regresión lineal del producto 5 (Fuente: elaboración propia)



**Figura 4.5** Gráfica de regresión lineal del producto 5 (Fuente: elaboración propia)

Producto 6							
No. de mes	Meses	2012	2013	2014	2015	Media aritmética	Predicción
1	Enero	44.000	106.000	96.000	116.000	90.500	100.054
2	Febrero	99.000	58.000	86.000	123.000	91.500	99.014
3	Marzo	105.000	61.000	128.000	147.000	110.250	97.974
4	Abril	95.000	94.000	108.000	106.000	100.750	96.934
5	Mayo	115.000	81.000	89.000	64.000	87.250	95.894
6	Junio	71.000	86.000	110.000	161.000	107.000	94.853
7	Julio	60.000	51.000	105.000	145.000	90.250	93.813
8	Agosto	129.000	26.000	56.000	118.000	82.250	92.773
9	Septiembre	92.000	64.000	126.000	148.000	107.500	91.733
10	Octubre	117.000	61.000	119.000	151.000	112.000	90.693
11	Noviembre	70.000	25.000	86.000	165.000	86.500	89.652
12	Diciembre	0.000	46.000	92.000	127.000	66.250	88.612

**Tabla 4.6** Información de ventas correspondiente a cuatro años, la media aritmética y la predicción con regresión lineal del producto 6 (Fuente: elaboración propia)

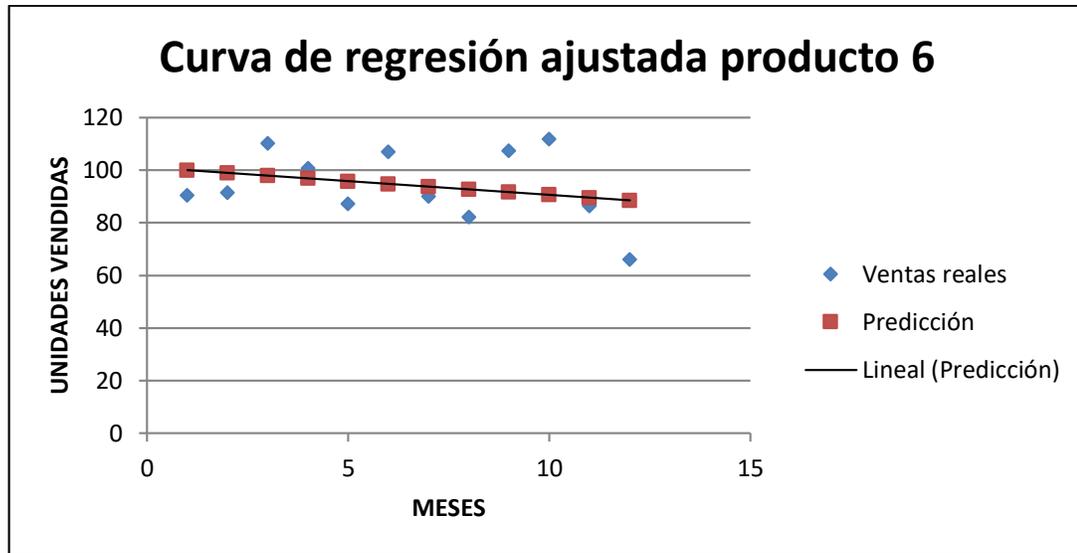


Figura 4.6 Gráfica de regresión lineal del producto 6 (Fuente: elaboración propia)

#### 4.1.2 Promedio móvil

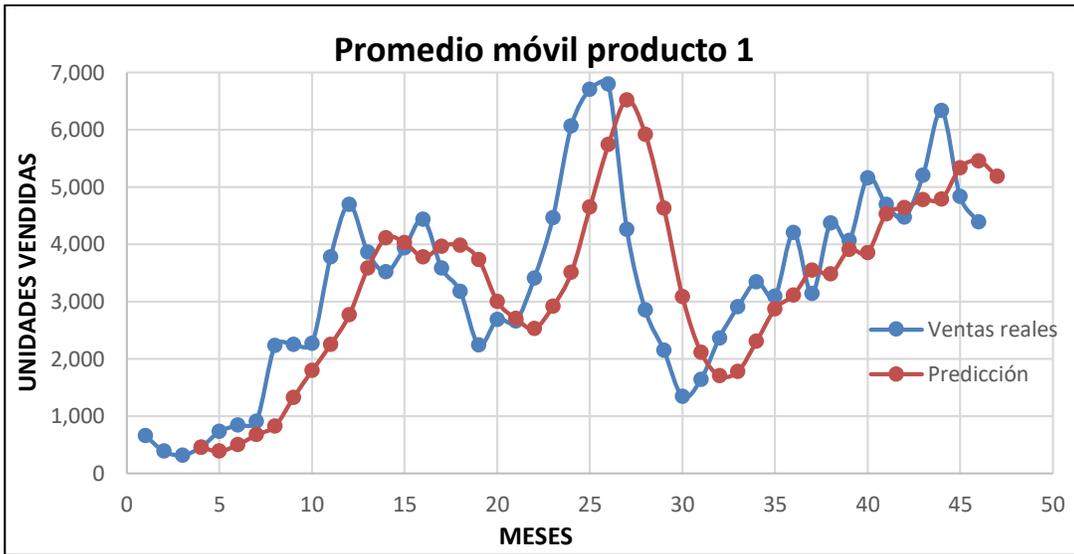
Para aplicar esta técnica se comenzó por definir un número que representaría el periodo para considerar el promedio móvil, siendo la constante que es representada con la letra **N** en la ecuación (2), se le asignó un valor de tres, siguiendo la fórmula se aplicó a la información de cada uno de los meses en los cuatro años que se tenían de respaldo, tomando como referencia el cuarto mes para comenzar a aplicar dicha ecuación para generar la predicción, por el valor especificado en el número de periodo para el promedio móvil. Así como se especificó en la ecuación (2),  $M_T$  será el resultado final de la predicción,  $d_T$  es la demanda del mes anterior y **T** es el número en el que se encuentra el periodo, la herramienta con la que se realizaron las pruebas fue Excel.

En las tablas 4.7 a 4.12 se muestran los datos de cada producto que se usaron para aplicar el método especificado. La primera columna representa el año, la segunda columna el número que identifica a cada uno de los meses; la tercera indica el nombre del mes; la cuarta columna muestra las ventas reales y, por último, la predicción generada. Continuando en esta sección se presentan las figuras 4.7 a 4.12 que contienen las gráficas con los resultados al aplicar la técnica promedios móviles en cada uno de los productos de la muestra, donde se puede observar que existe un desfase de la predicción con la venta real.

<b>Producto 1</b>				
<b>Año</b>	<b>No. de mes</b>	<b>Mes</b>	<b>Ventas reales</b>	<b>Predicción</b>
2012	1	Enero	655.200	
	2	Febrero	383.540	
	3	Marzo	315.100	
	4	Abril	454.930	451.280
	5	Mayo	727.342	384.523
	6	Junio	838.636	499.124
	7	Julio	910.672	673.636
	8	Agosto	2,231.026	825.550
	9	Septiembre	2,246.059	1326.778
	10	Octubre	2,269.087	1795.919
	11	Noviembre	3,776.791	2248.724
	12	Diciembre	4,694.942	2763.979
2013	13	Enero	3,864.809	3580.273
	14	Febrero	3,521.028	4112.180
	15	Marzo	3,936.770	4026.926
	16	Abril	4,438.960	3774.202
	17	Mayo	3,581.985	3965.586
	18	Junio	3,176.277	3985.905
	19	Julio	2,241.982	3732.407
	20	Agosto	2,684.181	3000.081
	21	Septiembre	2,653.173	2700.813
	22	Octubre	3,406.773	2526.445
	23	Noviembre	4,458.848	2914.709
	24	Diciembre	6,065.552	3506.264
2014	25	Enero	6,703.468	4643.724
	26	Febrero	6,792.435	5742.622
	27	Marzo	4,259.319	6520.485
	28	Abril	2,847.660	5918.407
	29	Mayo	2,147.960	4633.138
	30	Junio	1,337.390	3084.979
	31	Julio	1,637.455	2111.003
	32	Agosto	2,364.480	1707.601
	33	Septiembre	2,906.850	1779.775
	34	Octubre	3,341.440	2302.928
	35	Noviembre	3,092.945	2870.923
	36	Diciembre	4,204.550	3113.745
2015	37	Enero	3,138.025	3546.311
	38	Febrero	4,367.105	3478.506
	39	Marzo	4,063.337	3903.226
	40	Abril	5,161.090	3856.155
	41	Mayo	4,696.040	4530.510

	42	Junio	4,469.114	4640.155
	43	Julio	5,208.010	4775.414
	44	Agosto	6,334.435	4791.054
	45	Septiembre	4,831.320	5337.186
	46	Octubre	4,390.220	5457.921
	47	Noviembre	4,139.664	5185.325
	48	Diciembre	3,766.883	4453.734

**Tabla 4.7** Información de ventas correspondiente por mes, año y la predicción generada con promedio móvil del producto 1 (Fuente: elaboración propia)



**Figura 4.7** Gráfica de promedio móvil del producto 1 (Fuente: elaboración propia)

Producto 2				
Año	No. de mes	Mes	Ventas reales	Predicción
2012	1	Enero	35.810	
	2	Febrero	92.265	
	3	Marzo	74.890	
	4	Abril	37.430	67.655
	5	Mayo	76.015	68.195
	6	Junio	76.575	62.778
	7	Julio	79.592	63.340
	8	Agosto	102.500	77.394
	9	Septiembre	121.345	86.222
	10	Octubre	113.675	101.145
	11	Noviembre	156.245	112.506
	12	Diciembre	168.725	130.421
2013	13	Enero	165.865	146.215

	14	Febrero	73.335	163.611
	15	Marzo	93.475	135.975
	16	Abril	75.810	110.891
	17	Mayo	110.465	80.873
	18	Junio	122.620	93.250
	19	Julio	142.500	102.965
	20	Agosto	109.360	125.195
	21	Septiembre	144.910	124.826
	22	Octubre	107.200	132.256
	23	Noviembre	193.955	120.490
	24	Diciembre	136.170	148.688
2014	25	Enero	129.055	145.775
	26	Febrero	89.470	153.060
	27	Marzo	86.200	118.231
	28	Abril	87.045	101.575
	29	Mayo	98.665	87.571
	30	Junio	100.775	90.636
	31	Julio	169.055	95.495
	32	Agosto	250.020	122.831
	33	Septiembre	119.325	173.283
	34	Octubre	256.805	179.466
	35	Noviembre	260.645	208.716
36	Diciembre	339.260	212.258	
2015	37	Enero	325.045	285.570
	38	Febrero	156.990	308.316
	39	Marzo	254.980	273.765
	40	Abril	184.135	245.671
	41	Mayo	194.140	198.701
	42	Junio	120.350	211.085
	43	Julio	239.595	166.208
	44	Agosto	219.545	184.695
	45	Septiembre	307.322	193.163
	46	Octubre	261.349	255.487
	47	Noviembre	185.349	262.738
48	Diciembre	247.460	251.34	

**Tabla 4.8** Información de ventas correspondiente por mes, año y la predicción generada con promedio móvil del producto 2 (Fuente: elaboración propia)

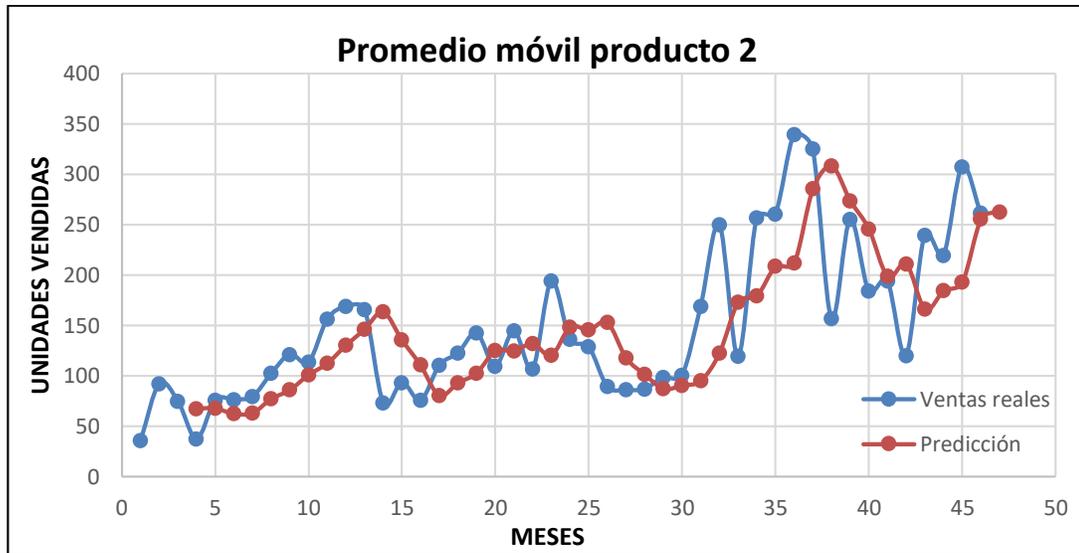
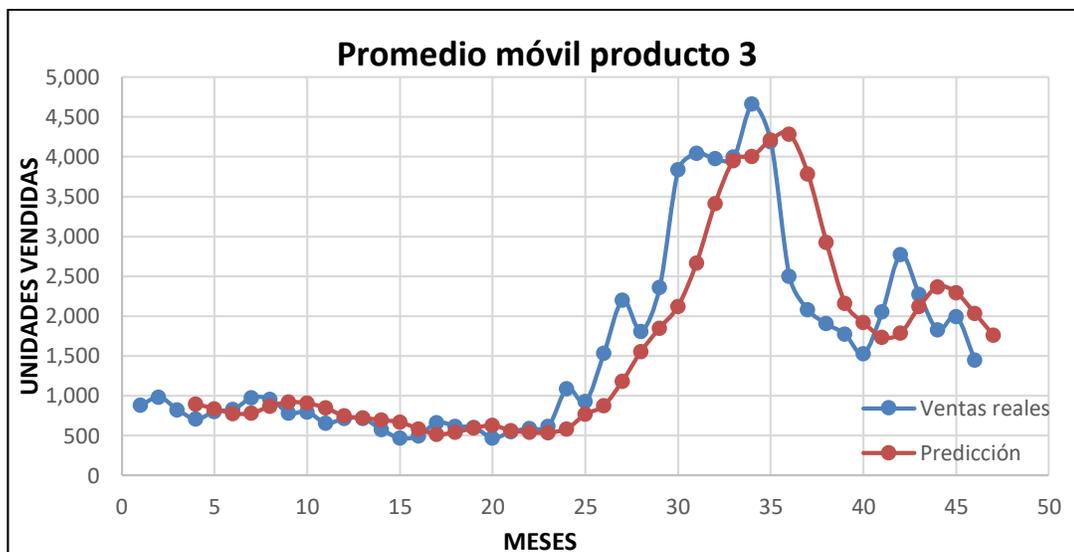


Figura 4.8 Gráfica de promedio móvil del producto 2 (Fuente: elaboración propia)

Producto 3				
Año	No. de mes	Mes	Ventas reales	Predicción
2012	1	Enero	877.475	
	2	Febrero	981.960	
	3	Marzo	821.910	
	4	Abril	703.585	893.781
	5	Mayo	802.495	835.818
	6	Junio	825.535	775.996
	7	Julio	973.321	777.205
	8	Agosto	955.902	867.117
	9	Septiembre	782.464	918.252
	10	Octubre	795.816	903.895
	11	Noviembre	653.422	844.727
	12	Diciembre	712.996	743.900
2013	13	Enero	711.224	720.744
	14	Febrero	574.512	692.547
	15	Marzo	464.607	666.244
	16	Abril	496.100	583.447
	17	Mayo	662.535	511.739
	18	Junio	613.902	541.080
	19	Julio	597.290	590.845
	20	Agosto	469.020	624.575
	21	Septiembre	547.000	560.070
	22	Octubre	589.110	537.770
	23	Noviembre	615.016	535.043
	24	Diciembre	1,086.922	583.708

2014	25	Enero	924.346	763.682
	26	Febrero	1,535.168	875.428
	27	Marzo	2,198.707	1182.145
	28	Abril	1,807.395	1552.740
	29	Mayo	2,358.810	1847.090
	30	Junio	3,834.730	2121.637
	31	Julio	4,045.170	2666.978
	32	Agosto	3,975.075	3412.903
	33	Septiembre	3,994.445	3951.658
	34	Octubre	4,662.375	4004.896
	35	Noviembre	4,187.920	4210.631
	36	Diciembre	2,499.915	4281.580
2015	37	Enero	2,079.590	3783.403
	38	Febrero	1,903.340	2922.475
	39	Marzo	1,771.185	2160.948
	40	Abril	1,526.770	1918.038
	41	Mayo	2,051.745	1733.765
	42	Junio	2,772.625	1783.233
	43	Julio	2,272.645	2117.046
	44	Agosto	1,826.265	2365.671
	45	Septiembre	1,995.079	2290.511
	46	Octubre	1,447.137	2031.329
	47	Noviembre	1,300.761	1756.160
	48	Diciembre	1,491.294	1580.992

**Tabla 4.9** Información de ventas correspondiente por mes, año y la predicción generada con promedio móvil del producto 3 (Fuente: elaboración propia)

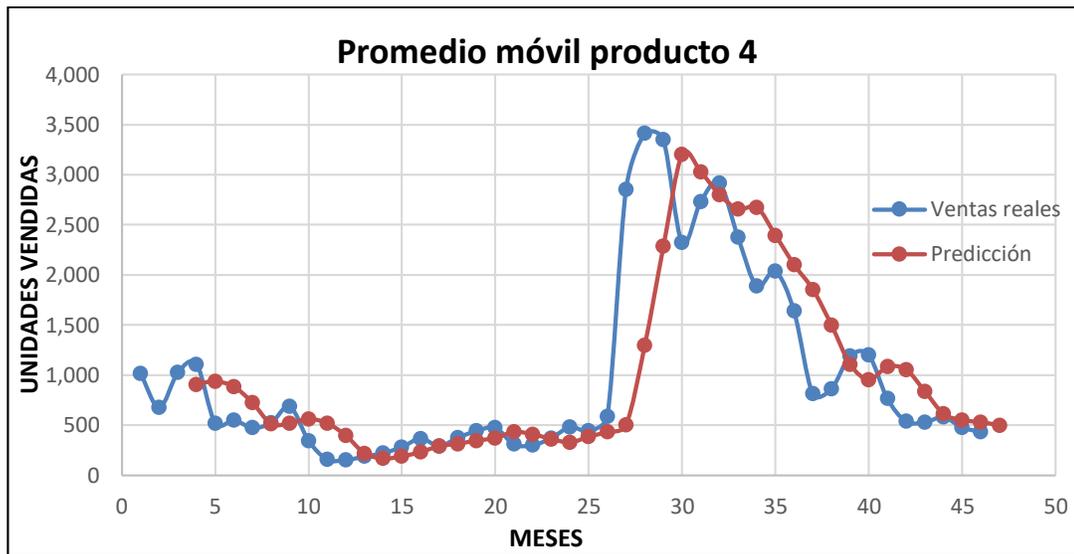


**Figura 4.9** Gráfica de promedio móvil del producto 3 (Fuente: elaboración propia)

<b>Producto 4</b>				
<b>Año</b>	<b>No. de mes</b>	<b>Mes</b>	<b>Ventas Reales</b>	<b>Predicción</b>
2012	1	Enero	1,015.855	
	2	Febrero	675.845	
	3	Marzo	1,025.660	
	4	Abril	1,107.880	905.786
	5	Mayo	517.559	936.461
	6	Junio	550.378	883.699
	7	Julio	473.712	725.272
	8	Agosto	524.433	513.883
	9	Septiembre	685.955	516.174
	10	Octubre	345.096	561.366
	11	Noviembre	158.481	518.494
	12	Diciembre	152.195	396.510
2013	13	Enero	189.018	218.590
	14	Febrero	221.767	166.564
	15	Marzo	281.487	187.660
	16	Abril	367.225	230.757
	17	Mayo	291.810	290.159
	18	Junio	376.840	313.507
	19	Julio	442.105	345.291
	20	Agosto	475.015	370.251
	21	Septiembre	309.190	431.320
	22	Octubre	300.345	408.770
	23	Noviembre	371.610	361.516
	24	Diciembre	480.570	327.048
2014	25	Enero	442.812	384.175
	26	Febrero	589.451	431.664
	27	Marzo	2,850.007	504.277
	28	Abril	3,411.831	1294.090
	29	Mayo	3,348.401	2283.763
	30	Junio	2,323.670	3203.413
	31	Julio	2,729.050	3027.967
	32	Agosto	2,917.445	2800.373
	33	Septiembre	2,375.210	2656.721
	34	Octubre	1,889.256	2673.901
	35	Noviembre	2,034.390	2393.970
	36	Diciembre	1,637.900	2099.618
2015	37	Enero	813.820	1853.848
	38	Febrero	861.970	1495.370
	39	Marzo	1,187.765	1104.563
	40	Abril	1,198.985	954.518
	41	Mayo	765.925	1082.906

	42	Junio	538.045	1050.891
	43	Julio	528.710	834.318
	44	Agosto	580.110	610.893
	45	Septiembre	478.127	548.955
	46	Octubre	432.906	528.982
	47	Noviembre	399.501	497.047
	48	Diciembre	478.908	436.844

**Tabla 4.10** Información de ventas correspondiente por mes, año y la predicción generada con promedio móvil del producto 4 (Fuente: elaboración propia)



**Figura 4.10** Gráfica de promedio móvil del producto 4 (Fuente: elaboración propia)

<b>Producto 5</b>				
<b>Año</b>	<b>No. de mes</b>	<b>Mes</b>	<b>Ventas reales</b>	<b>Predicción</b>
2012	1	Enero	520.327	
	2	Febrero	517.950	
	3	Marzo	553.030	
	4	Abril	546.576	530.435
	5	Mayo	538.926	539.185
	6	Junio	540.110	546.177
	7	Julio	530.333	541.870
	8	Agosto	516.063	536.456
	9	Septiembre	503.779	528.835
	10	Octubre	626.752	516.725
	11	Noviembre	493.036	548.864
	12	Diciembre	497.439	541.189
2013	13	Enero	528.709	539.075

	14	Febrero	488.980	506.394
	15	Marzo	424.148	505.042
	16	Abril	639.684	480.612
	17	Mayo	642.712	517.604
	18	Junio	553.660	568.848
	19	Julio	576.675	612.018
	20	Agosto	568.300	591.015
	21	Septiembre	499.965	566.211
	22	Octubre	517.660	548.313
	23	Noviembre	504.365	528.641
	24	Diciembre	492.080	507.330
2014	25	Enero	521.180	504.701
	26	Febrero	554.805	505.875
	27	Marzo	629.295	522.688
	28	Abril	556.010	568.426
	29	Mayo	654.020	580.036
	30	Junio	603.010	613.108
	31	Julio	651.705	604.346
	32	Agosto	842.682	636.245
	33	Septiembre	714.665	699.132
	34	Octubre	712.700	736.350
	35	Noviembre	769.611	756.682
36	Diciembre	590.047	732.325	
2015	37	Enero	667.653	690.786
	38	Febrero	568.808	675.770
	39	Marzo	504.435	608.836
	40	Abril	557.506	580.298
	41	Mayo	631.671	543.583
	42	Junio	590.556	564.537
	43	Julio	626.914	593.244
	44	Agosto	617.714	616.380
	45	Septiembre	716.994	611.728
	46	Octubre	623.224	653.874
	47	Noviembre	629.552	652.644
48	Diciembre	571.779	656.590	

**Tabla 4.11** Información de ventas correspondiente por mes, año y la predicción generada con promedio móvil del producto 5 (Fuente: elaboración propia)

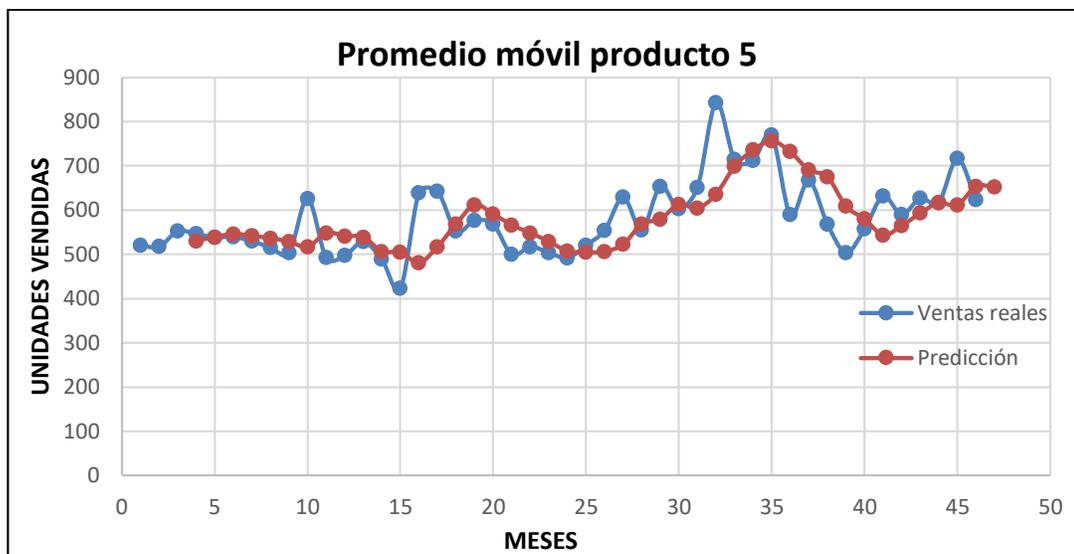
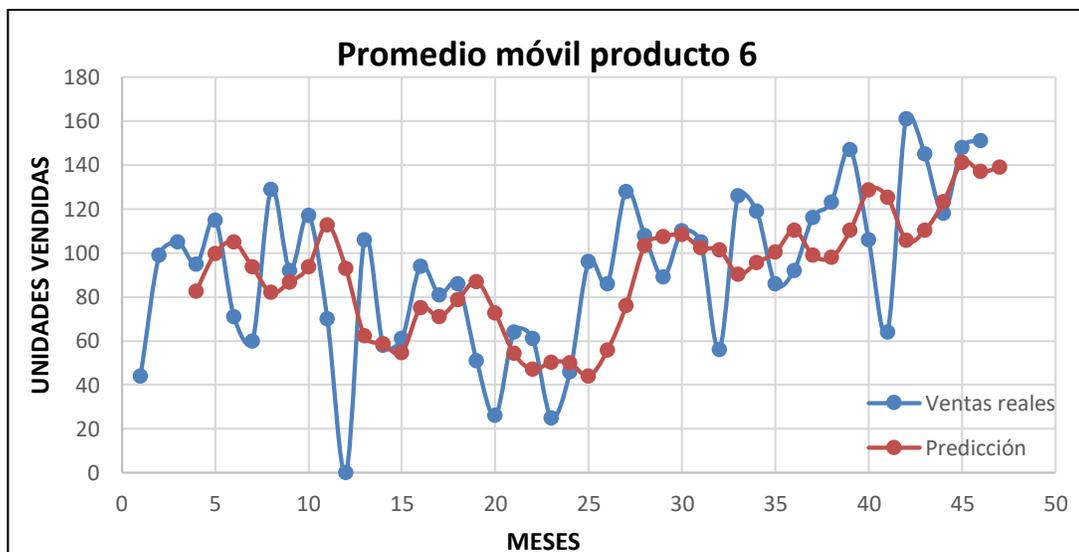


Figura 4.11 Gráfica de promedio móvil del producto 5 (Fuente: elaboración propia)

Producto 6				
Año	No. de mes	Mes	Ventas reales	Predicción
2012	1	Enero	44.000	
	2	Febrero	99.000	
	3	Marzo	105.000	
	4	Abril	95.000	82.666
	5	Mayo	115.000	99.666
	6	Junio	71.000	105.000
	7	Julio	60.000	93.666
	8	Agosto	129.000	82.000
	9	Septiembre	92.000	86.666
	10	Octubre	117.000	93.666
	11	Noviembre	70.000	112.666
	12	Diciembre	0.000	93.000
2013	13	Enero	106.000	62.333
	14	Febrero	58.000	58.666
	15	Marzo	61.000	54.666
	16	Abril	94.000	75.000
	17	Mayo	81.000	71.000
	18	Junio	86.000	78.666
	19	Julio	51.000	87.000
	20	Agosto	26.000	72.666
	21	Septiembre	64.000	54.333
	22	Octubre	61.000	47.000
	23	Noviembre	25.000	50.333
	24	Diciembre	46.000	50.000

2014	25	Enero	96.000	44.000
	26	Febrero	86.000	55.666
	27	Marzo	128.000	76.000
	28	Abril	108.000	103.333
	29	Mayo	89.000	107.333
	30	Junio	110.000	108.333
	31	Julio	105.000	102.333
	32	Agosto	56.000	101.333
	33	Septiembre	126.000	90.333
	34	Octubre	119.000	95.666
	35	Noviembre	86.000	100.333
	36	Diciembre	92.000	110.333
2015	37	Enero	116.000	99.000
	38	Febrero	123.000	98.000
	39	Marzo	147.000	110.333
	40	Abril	106.000	128.666
	41	Mayo	64.000	125.333
	42	Junio	161.000	105.666
	43	Julio	145.000	110.333
	44	Agosto	118.000	123.333
	45	Septiembre	148.000	141.333
	46	Octubre	151.000	137.000
	47	Noviembre	165.000	139.000
	48	Diciembre	127.000	154.666

**Tabla 4.12** Información de ventas correspondiente por mes, año y la predicción generada con promedio móvil del producto 6 (Fuente: elaboración propia)



**Figura 4.12** Gráfica de promedio móvil del producto 6 (Fuente: elaboración propia)

### 4.1.3 Suavización exponencial

Trabajando en la herramienta Excel, se comenzó por recopilar una serie de datos específicos para hacer la predicción con la fórmula de suavización exponencial.

Posteriormente se aplicó la ecuación (3), en donde representada por  $\hat{X}_t$  es el resultado de la predicción de la venta en el periodo especificado  $t$ ; existen tres factores diferentes que intervienen, uno de ellos está representado por  $\hat{X}_{t-1}$  que representa la predicción de las ventas del periodo  $t-1$ , otro factor es  $X_{t-1}$  que representa la venta real del periodo  $t-1$  y por último el coeficiente de suavización representado por  $\alpha$ . A continuación, se muestran los pasos que fueron necesario para generar el valor de esta constate, mediante una opción de Excel:

1. Se seleccionó “**Datos**” de la cinta de opciones.
2. Ahí se localizó y seleccionó la opción “**Solver**” en el apartado de “**Análisis**”.
3. En primera estancia, se seleccionó la casilla “**Objetivo**” a la cual se quiera minimizar el rango de error, en este caso fue ME (que representa el promedio sobre los errores obtenidos). Se seleccionó la celda correspondiente y se calculó el mejor valor posible para poder reducir el promedio del error, llegando a la conclusión que el valor 1, fue el que mejor se adaptó para toda la muestra de productos.

En las tablas 4.13 a 4.18 se muestra la información de cada uno de los productos con la que se estuvo trabajando esta técnica, así como una constante a la que se le identificó con el nombre de  $\alpha$ , a la cual se le asigna un valor entre 0-1, con esto, se agregó una ponderación a los valores más cercanos en el tiempo, el valor de  $\alpha$  fue uno.

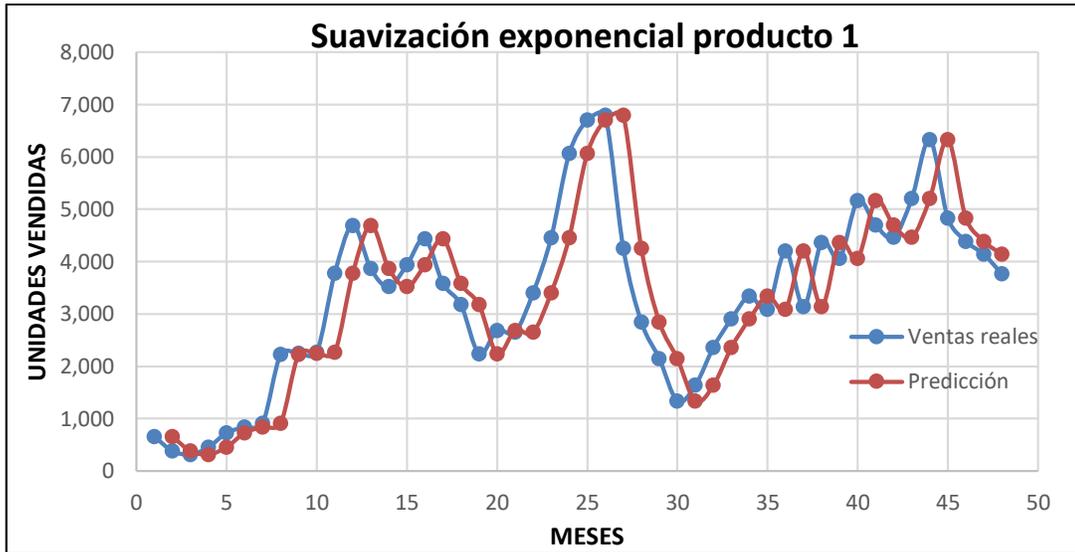
En las figuras 4.13 a 4.18 se pueden observar los resultados gráficamente, existiendo una diferencia notable entre la venta real y la predicción generada por el modelo, llegando a la conclusión que no tiene la precisión necesaria, así se puede observar el mismo comportamiento en las demás figuras.

Producto 1				
Año	No. de mes	Mes	Ventas reales	Predicción
2012	1	Enero	655.200	655.200
	2	Febrero	383.540	655.200
	3	Marzo	315.100	383.540

	4	Abril	454.930	315.100	
	5	Mayo	727.342	454.930	
	6	Junio	838.636	727.340	
	7	Julio	910.672	838.640	
	8	Agosto	2,231.026	910.670	
	9	Septiembre	2,246.059	2,231.030	
	10	Octubre	2,269.087	2,246.060	
	11	Noviembre	3,776.791	2,269.090	
	12	Diciembre	4,694.942	3,776.790	
	2013	13	Enero	3,864.809	4,694.940
		14	Febrero	3,521.028	3,864.810
		15	Marzo	3,936.770	3,521.030
16		Abril	4,438.960	3,936.770	
17		Mayo	3,581.985	4,438.960	
18		Junio	3,176.277	3,581.990	
19		Julio	2,241.982	3,176.280	
20		Agosto	2,684.181	2,241.980	
21		Septiembre	2,653.173	2,684.180	
22		Octubre	3,406.773	2,653.170	
23		Noviembre	4,458.848	3,406.770	
24		Diciembre	6,065.552	4,458.850	
2014	25	Enero	6,703.468	6,065.550	
	26	Febrero	6,792.435	6,703.470	
	27	Marzo	4,259.319	6,792.440	
	28	Abril	2,847.660	4,259.320	
	29	Mayo	2,147.960	2,847.660	
	30	Junio	1,337.390	2,147.960	
	31	Julio	1,637.455	1,337.390	
	32	Agosto	2,364.480	1,637.460	
	33	Septiembre	2,906.850	2,364.480	
	34	Octubre	3,341.440	2,906.850	
	35	Noviembre	3,092.945	3,341.440	
	36	Diciembre	4,204.550	3,092.950	
2015	37	Enero	3,138.025	4,204.550	
	38	Febrero	4,367.105	3,138.030	
	39	Marzo	4,063.337	4,367.110	
	40	Abril	5,161.090	4,063.340	
	41	Mayo	4,696.040	5,161.090	
	42	Junio	4,469.114	4,696.040	
	43	Julio	5,208.010	4,469.110	
	44	Agosto	6,334.435	5,208.010	
	45	Septiembre	4,831.320	6,334.440	
	46	Octubre	4,390.220	4,831.320	

	47	Noviembre	4,139.664	4,390.220
	48	Diciembre	3,766.883	4,139.660

**Tabla 4.13** Información de ventas correspondiente por mes, año y predicción con suavización exponencial del producto 1 (Fuente: elaboración propia)



**Figura 4.13** Gráfica de suavización exponencial del producto 1 (Fuente: elaboración propia)

<b>Producto 2</b>				
<b>Año</b>	<b>No. de mes</b>	<b>Mes</b>	<b>Ventas reales</b>	<b>Predicción</b>
2012	1	Enero	35.810	35.810
	2	Febrero	92.265	35.810
	3	Marzo	74.890	92.270
	4	Abril	37.430	74.890
	5	Mayo	76.015	37.430
	6	Junio	76.575	76.020
	7	Julio	79.592	76.580
	8	Agosto	102.500	79.590
	9	Septiembre	121.345	102.500
	10	Octubre	113.675	121.350
	11	Noviembre	156.245	113.680
	12	Diciembre	168.725	156.250
2013	13	Enero	165.865	168.730
	14	Febrero	73.335	165.870
	15	Marzo	93.475	73.340
	16	Abril	75.810	93.480
	17	Mayo	110.465	75.810
	18	Junio	122.620	110.470

	19	Julio	142.500	122.620
	20	Agosto	109.360	142.500
	21	Septiembre	144.910	109.360
	22	Octubre	107.200	144.910
	23	Noviembre	193.955	107.200
	24	Diciembre	136.170	193.960
2014	25	Enero	129.055	136.170
	26	Febrero	89.470	129.060
	27	Marzo	86.200	89.470
	28	Abril	87.045	86.200
	29	Mayo	98.665	87.050
	30	Junio	100.775	98.670
	31	Julio	169.055	100.780
	32	Agosto	250.020	169.060
	33	Septiembre	119.325	250.020
	34	Octubre	256.805	119.330
	35	Noviembre	260.645	256.810
	36	Diciembre	339.260	260.650
2015	37	Enero	325.045	339.260
	38	Febrero	156.990	325.050
	39	Marzo	254.980	156.990
	40	Abril	184.135	254.980
	41	Mayo	194.140	184.140
	42	Junio	120.350	194.140
	43	Julio	239.595	120.350
	44	Agosto	219.545	239.600
	45	Septiembre	307.322	219.550
	46	Octubre	261.349	307.320
	47	Noviembre	185.349	261.350
	48	Diciembre	247.460	185.350

**Tabla 4.14** Información de ventas correspondiente por mes, año y predicción con suavización exponencial del producto 2 (Fuente: elaboración propia)

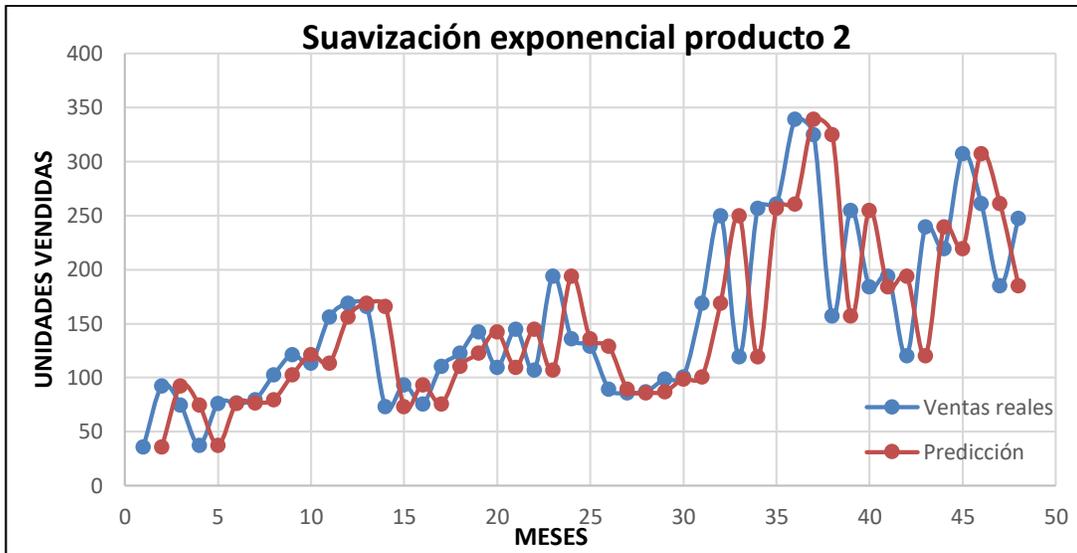
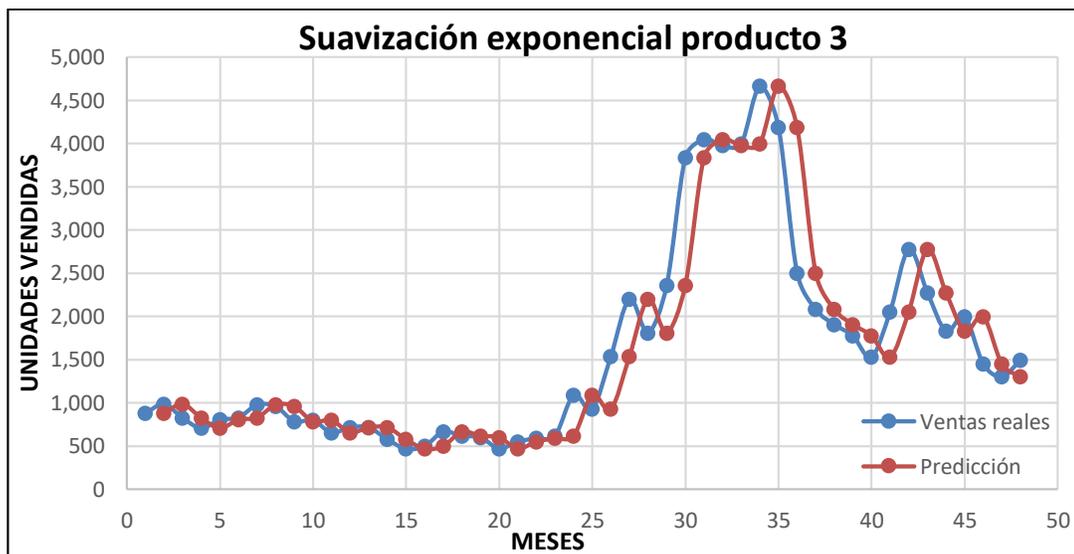


Figura 4.14 Gráfica de suavización exponencial del producto 2 (Fuente: elaboración propia)

Producto 3				
Año	No. de mes	Mes	Ventas reales	Predicción
2012	1	Enero	877.475	877.475
	2	Febrero	981.960	877.480
	3	Marzo	821.910	981.960
	4	Abril	703.585	821.910
	5	Mayo	802.495	703.590
	6	Junio	825.535	802.500
	7	Julio	973.321	825.540
	8	Agosto	955.902	973.320
	9	Septiembre	782.464	955.900
	10	Octubre	795.816	782.460
	11	Noviembre	653.422	795.820
	12	Diciembre	712.996	653.420
2013	13	Enero	711.224	713.000
	14	Febrero	574.512	711.220
	15	Marzo	464.607	574.510
	16	Abril	496.100	464.610
	17	Mayo	662.535	496.100
	18	Junio	613.902	662.540
	19	Julio	597.290	613.900
	20	Agosto	469.020	597.290
	21	Septiembre	547.000	469.020
	22	Octubre	589.110	547.000
	23	Noviembre	615.016	589.110
	24	Diciembre	1,086.922	615.020

2014	25	Enero	924.346	1,086.920
	26	Febrero	1,535.168	924.350
	27	Marzo	2,198.707	1,535.170
	28	Abril	1,807.395	2,198.710
	29	Mayo	2,358.810	1,807.400
	30	Junio	3,834.730	2,358.810
	31	Julio	4,045.170	3,834.730
	32	Agosto	3,975.075	4,045.170
	33	Septiembre	3,994.445	3,975.080
	34	Octubre	4,662.375	3,994.450
	35	Noviembre	4,187.920	4,662.380
	36	Diciembre	2,499.915	4,187.920
2015	37	Enero	2,079.590	2,499.920
	38	Febrero	1,903.340	2,079.590
	39	Marzo	1,771.185	1,903.340
	40	Abril	1,526.770	1,771.190
	41	Mayo	2,051.745	1,526.770
	42	Junio	2,772.625	2,051.750
	43	Julio	2,272.645	2,772.630
	44	Agosto	1,826.265	2,272.650
	45	Septiembre	1,995.079	1,826.270
	46	Octubre	1,447.137	1,995.080
	47	Noviembre	1,300.761	1,447.140
	48	Diciembre	1,491.294	1,300.760

**Tabla 4.15** Información de ventas correspondiente por mes, año y predicción con suavización exponencial del producto 3 (Fuente: elaboración propia)

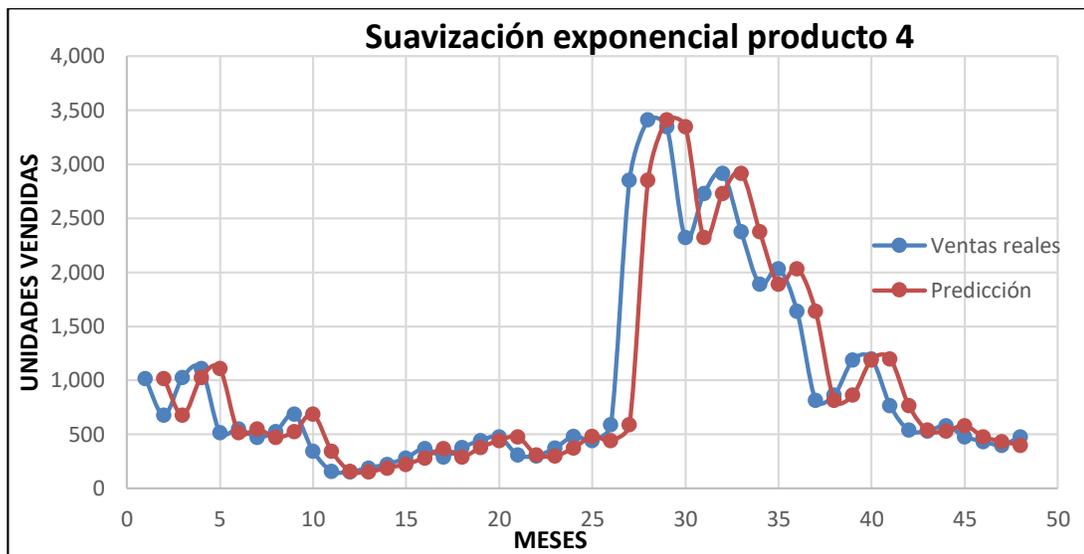


**Figura 4.15** Gráfica de suavización exponencial del producto 3 (Fuente: elaboración propia)

<b>Producto 4</b>				
<b>Año</b>	<b>No. de mes</b>	<b>Mes</b>	<b>Ventas reales</b>	<b>Predicción</b>
2012	1	Enero	1,015.855	1,015.855
	2	Febrero	675.845	1,015.860
	3	Marzo	1,025.660	675.850
	4	Abril	1,107.880	1,025.660
	5	Mayo	517.559	1,107.880
	6	Junio	550.378	517.560
	7	Julio	473.712	550.380
	8	Agosto	524.433	473.710
	9	Septiembre	685.955	524.430
	10	Octubre	345.096	685.960
	11	Noviembre	158.481	345.100
	12	Diciembre	152.195	158.480
2013	13	Enero	189.018	152.200
	14	Febrero	221.767	189.020
	15	Marzo	281.487	221.770
	16	Abril	367.225	281.490
	17	Mayo	291.810	367.230
	18	Junio	376.840	291.810
	19	Julio	442.105	376.840
	20	Agosto	475.015	442.110
	21	Septiembre	309.190	475.020
	22	Octubre	300.345	309.190
	23	Noviembre	371.610	300.350
	24	Diciembre	480.570	371.610
2014	25	Enero	442.812	480.570
	26	Febrero	589.451	442.810
	27	Marzo	2,850.007	589.450
	28	Abril	3,411.831	2,850.010
	29	Mayo	3,348.401	3,411.830
	30	Junio	2,323.670	3,348.400
	31	Julio	2,729.050	2,323.670
	32	Agosto	2,917.445	2,729.050
	33	Septiembre	2,375.210	2,917.450
	34	Octubre	1,889.256	2,375.210
	35	Noviembre	2,034.390	1,889.261
	36	Diciembre	1,637.900	2,034.390
2015	37	Enero	813.820	1,637.902
	38	Febrero	861.970	813.820
	39	Marzo	1,187.765	861.971
	40	Abril	1,198.985	1,187.770

	41	Mayo	765.925	1,198.990
	42	Junio	538.045	765.930
	43	Julio	528.710	538.050
	44	Agosto	580.110	528.710
	45	Septiembre	478.127	580.110
	46	Octubre	432.906	478.131
	47	Noviembre	399.501	432.911
	48	Diciembre	478.908	399.501

**Tabla 4.16** Información de ventas correspondiente por mes, año y predicción con suavización exponencial del producto 4 (Fuente: elaboración propia)



**Figura 4.16** Gráfica de suavización exponencial del producto 4 (Fuente: elaboración propia)

<b>Producto 5</b>				
<b>Año</b>	<b>No. de mes</b>	<b>Mes</b>	<b>Ventas reales</b>	<b>Predicción</b>
2012	1	Enero	520.327	520.327
	2	Febrero	517.950	517.950
	3	Marzo	553.030	553.030
	4	Abril	546.576	546.576
	5	Mayo	538.926	538.926
	6	Junio	540.110	540.110
	7	Julio	530.333	530.333
	8	Agosto	516.063	516.063
	9	Septiembre	503.779	503.779
	10	Octubre	626.752	626.752
	11	Noviembre	493.036	493.036
	12	Diciembre	497.439	497.439

2013	13	Enero	528.709	528.709
	14	Febrero	488.980	488.980
	15	Marzo	424.148	424.148
	16	Abril	639.684	639.684
	17	Mayo	642.712	642.712
	18	Junio	553.660	553.660
	19	Julio	576.675	576.675
	20	Agosto	568.300	568.300
	21	Septiembre	499.965	499.965
	22	Octubre	517.660	517.660
	23	Noviembre	504.365	504.365
	24	Diciembre	492.080	492.080
2014	25	Enero	521.180	521.180
	26	Febrero	554.805	554.805
	27	Marzo	629.295	629.295
	28	Abril	556.010	556.010
	29	Mayo	654.020	654.020
	30	Junio	603.010	603.010
	31	Julio	651.705	651.705
	32	Agosto	842.682	842.682
	33	Septiembre	714.665	714.665
	34	Octubre	712.700	712.700
	35	Noviembre	769.611	769.611
	36	Diciembre	590.047	590.047
2015	37	Enero	667.653	667.653
	38	Febrero	568.808	568.808
	39	Marzo	504.435	504.435
	40	Abril	557.506	557.506
	41	Mayo	631.671	631.671
	42	Junio	590.556	590.556
	43	Julio	626.914	626.914
	44	Agosto	617.714	617.714
	45	Septiembre	716.994	716.994
	46	Octubre	623.224	623.224
	47	Noviembre	629.552	629.552
	48	Diciembre	571.779	520.327

**Tabla 4.17** Información de ventas correspondiente por mes, año y predicción con suavización exponencial del producto 5 (Fuente: elaboración propia)

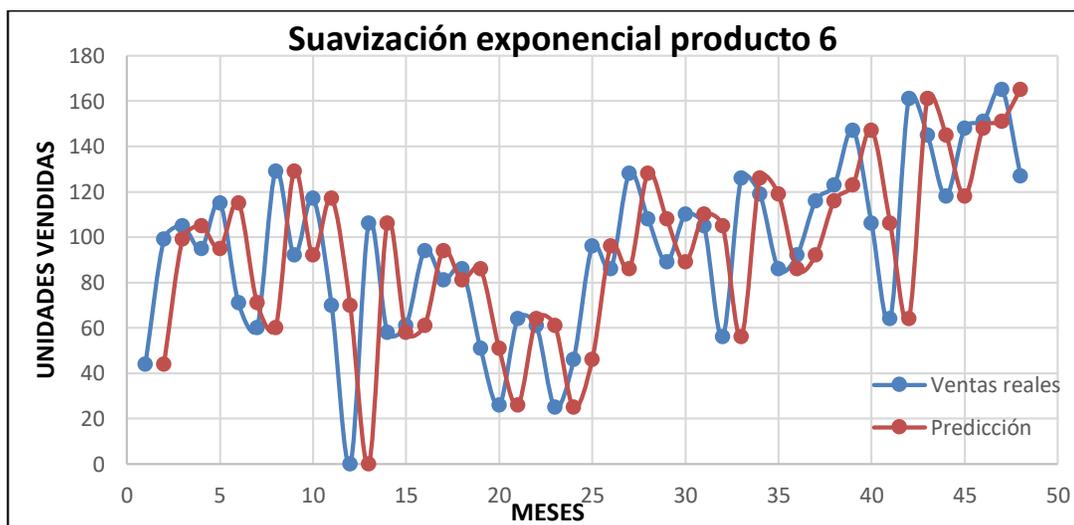


Figura 4.17 Gráfica de suavización exponencial del producto 5 (Fuente: propia)

Producto 6				
Año	No. de mes	Mes	Ventas reales	Predicción
2012	1	Enero	44.000	44.000
	2	Febrero	99.000	44.00
	3	Marzo	105.000	99.00
	4	Abril	95.000	105.00
	5	Mayo	115.000	95.00
	6	Junio	71.000	115.00
	7	Julio	60.000	71.00
	8	Agosto	129.000	60.00
	9	Septiembre	92.000	129.00
	10	Octubre	117.000	92.00
	11	Noviembre	70.000	117.00
	12	Diciembre	0.000	70.00
2013	13	Enero	106.000	0.00
	14	Febrero	58.000	106.00
	15	Marzo	61.000	58.00
	16	Abril	94.000	61.00
	17	Mayo	81.000	94.00
	18	Junio	86.000	81.00
	19	Julio	51.000	86.00
	20	Agosto	26.000	51.00
	21	Septiembre	64.000	26.00
	22	Octubre	61.000	64.00
	23	Noviembre	25.000	61.00

	24	Diciembre	46.000	25.00
2014	25	Enero	96.000	46.00
	26	Febrero	86.000	96.00
	27	Marzo	128.000	86.00
	28	Abril	108.000	128.00
	29	Mayo	89.000	108.00
	30	Junio	110.000	89.00
	31	Julio	105.000	110.00
	32	Agosto	56.000	105.00
	33	Septiembre	126.000	56.00
	34	Octubre	119.000	126.00
	35	Noviembre	86.000	119.00
	36	Diciembre	92.000	86.00
	2015	37	Enero	116.000
38		Febrero	123.000	116.00
39		Marzo	147.000	123.00
40		Abril	106.000	147.00
41		Mayo	64.000	106.00
42		Junio	161.000	64.00
43		Julio	145.000	161.00
44		Agosto	118.000	145.00
45		Septiembre	148.000	118.00
46		Octubre	151.000	148.00
47		Noviembre	165.000	151.00
48		Diciembre	127.000	165.00

**Tabla 4.18** Información de ventas correspondiente por mes, año y predicción con suavización exponencial del producto 6 (Fuente: elaboración propia)



**Figura 4.18** Gráfica de suavización exponencial del producto 6 (Fuente: elaboración propia)

#### 4.1.4 Suavización exponencial ajustada a la tendencia

Para esta técnica se trabajó en la herramienta de Excel, donde se aplicaron las ecuaciones (5), (6) y (7) respectivamente. Para llegar a obtener una predicción con dicho método, se comenzó por aplicar la ecuación (6), donde  $\hat{X}_{t-1}$  representa la predicción del periodo **t-1** o anterior al que se pretende calcular, a continuación se realizó una adición del coeficiente de suavización representado por  $\alpha$  que multiplica la venta real menos la predicción de las ventas, ambos siendo del mes anterior. En la ecuación (7), se multiplicó la constante  $\beta$  por la diferencia del resultado de la ecuación (6) del periodo **t** y **t-1**, posteriormente se le sumó la diferencia de uno y el valor de  $\beta$ , para terminar, multiplicándolo por el resultado de la ecuación (7) de periodo **t-1**. Para obtener el resultado final en la ecuación (5), se sumaron los resultados de la ecuación (6) y (7).

En la tabla 4.19 a 4.24 se puede apreciar el conjunto de datos que se extrajeron para aplicar la técnica, en la primera columna se puede apreciar el año de registro; la siguiente columna muestra la representación del mes en número; la tercera columna describe el nombre del mes respectivamente; la columna siguiente representa las unidades vendidas en cada uno de los meses. Por último, se encuentra el resultado de la predicción. Existen dos constantes, las cuales están identificadas como  $\alpha$  y  $\beta$ .  $\alpha$  tiene un valor asignado para agregar una ponderación, con respecto a los valores en el tiempo;  $\beta$  es un valor para suavización a la tendencia que se pueda presentar en el transcurso del tiempo. Estas dos constantes obtienen un valor entre 0-1.  $\alpha$  quedó con el valor de 0.70 y  $\beta$  con un valor de 0.54. De las figuras 4.19 a 4.24 se pueden observar gráficamente los resultados obtenidos aplicando esta técnica, en donde la predicción tiene un comportamiento aceptable mientras la venta se mantenga estable, cuando se presenta una variación muy marcada, la cantidad en la predicción tiene una diferencia razonable, haciendo este modelo no confiable. El mismo comportamiento se puede observar en todas las figuras de esta técnica.

Producto 1				
Año	No. de mes	Mes	Ventas reales	Predicción
2012	1	Enero	655.200	
	2	Febrero	383.540	657
	3	Marzo	315.100	362
	4	Abril	454.930	256

	5	Mayo	727.342	415
	6	Junio	838.636	747
	7	Julio	910.672	905
	8	Agosto	2,231.026	979
	9	Septiembre	2,246.059	2391
	10	Octubre	2,269.087	2539
	11	Noviembre	3,776.791	2423
	12	Diciembre	4,694.942	4133
2013	13	Enero	3,864.809	5206
	14	Febrero	3,521.028	4062
	15	Marzo	3,936.770	3453
	16	Abril	4,438.960	3927
	17	Mayo	3,581.985	4573
	18	Junio	3,176.277	3544
	19	Julio	2,241.982	2987
	20	Agosto	2,684.181	1917
	21	Septiembre	2,653.173	2507
	22	Octubre	3,406.773	2610
	23	Noviembre	4,458.848	3538
	24	Diciembre	6,065.552	4810
2014	25	Enero	6,703.468	6666
	26	Febrero	6,792.435	7298
	27	Marzo	4,259.319	7169
	28	Abril	2,847.660	3955
	29	Mayo	2,147.960	2134
	30	Junio	1,337.390	1488
	31	Julio	1,637.455	788
	32	Agosto	2,364.480	1352
	33	Septiembre	2,906.850	2409
	34	Octubre	3,341.440	3128
	35	Noviembre	3,092.945	3602
	36	Diciembre	4,204.550	3220
2015	37	Enero	3,138.025	4448
	38	Febrero	4,367.105	3168
	39	Marzo	4,063.337	4503
	40	Abril	5,161.090	4194
	41	Mayo	4,696.040	5410
	42	Junio	4,469.114	4843
	43	Julio	5,208.010	4450
	44	Agosto	6,334.435	5304
	45	Septiembre	4,831.320	6675
	46	Octubre	4,390.220	4840
	47	Noviembre	4,139.664	4146

	48	Diciembre	3,766.883	3905
--	----	-----------	-----------	------

**Tabla 4.19** Información de ventas correspondiente por mes, año y predicción con suavización exponencial ajustada a la tendencia del producto 1 (Fuente: elaboración propia)



**Figura 4.19** Gráfica de suavización exponencial ajustada a la tendencia del producto 1 (Fuente: elaboración propia)

Producto 2				
Año	No. de mes	Mes	Ventas reales	Predicción
2012	1	Enero	35.810	
	2	Febrero	92.265	38
	3	Marzo	74.890	98
	4	Abril	37.430	85
	5	Mayo	76.015	39
	6	Junio	76.575	74
	7	Julio	79.592	80
	8	Agosto	102.500	83
	9	Septiembre	121.345	107
	10	Octubre	113.675	129
	11	Noviembre	156.245	118
	12	Diciembre	168.725	166
2013	13	Enero	165.865	180
	14	Febrero	73.335	172
	15	Marzo	93.475	58
	16	Abril	75.810	79
	17	Mayo	110.465	67

	18	Junio	122.620	112
	19	Julio	142.500	130
	20	Agosto	109.360	151
	21	Septiembre	144.910	110
	22	Octubre	107.200	148
	23	Noviembre	193.955	105
	24	Diciembre	136.170	206
2014	25	Enero	129.055	140
	26	Febrero	89.470	124
	27	Marzo	86.200	78
	28	Abril	87.045	75
	29	Mayo	98.665	82
	30	Junio	100.775	99
	31	Julio	169.055	103
	32	Agosto	250.020	183
	33	Septiembre	119.325	280
	34	Octubre	256.805	118
	35	Noviembre	260.645	268
	36	Diciembre	339.260	281
2015	37	Enero	325.045	366
	38	Febrero	156.990	343
	39	Marzo	254.980	132
	40	Abril	184.135	242
	41	Mayo	194.140	174
	42	Junio	120.350	185
	43	Julio	239.595	102
	44	Agosto	219.545	246
	45	Septiembre	307.322	232
	46	Octubre	261.349	329
	47	Noviembre	185.349	272
	48	Diciembre	247.460	171

**Tabla 4.20** Información de ventas correspondiente por mes, año y predicción con suavización exponencial ajustada a la tendencia del producto 2 (Fuente: elaboración propia)



**Figura 4.20** Gráfica de suavización exponencial ajustada la tendencia del producto 2 (Fuente: elaboración propia)

<b>Producto 3</b>				
<b>Año</b>	<b>No. de mes</b>	<b>Mes</b>	<b>Ventas reales</b>	<b>Predicción</b>
2012	1	Enero	877.475	
	2	Febrero	981.960	879
	3	Marzo	821.910	991
	4	Abril	703.585	830
	5	Mayo	802.495	678
	6	Junio	825.535	773
	7	Julio	973.321	822
	8	Agosto	955.902	989
	9	Septiembre	782.464	988
	10	Octubre	795.816	788
	11	Noviembre	653.422	776
	12	Diciembre	712.996	615
2013	13	Enero	711.224	690
	14	Febrero	574.512	706
	15	Marzo	464.607	547
	16	Abril	496.100	416
	17	Mayo	662.535	465
	18	Junio	613.902	683
	19	Julio	597.290	635
	20	Agosto	469.020	600
	21	Septiembre	547.000	443

	22	Octubre	589.110	534
	23	Noviembre	615.016	599
	24	Diciembre	1,086.922	631
2014	25	Enero	924.346	1187
	26	Febrero	1,535.168	995
	27	Marzo	2,198.707	1670
	28	Abril	1,807.395	2454
	29	Mayo	2,358.810	1934
	30	Junio	3,834.730	2481
	31	Julio	4,045.170	4227
	32	Agosto	3,975.075	4442
	33	Septiembre	3,994.445	4184
	34	Octubre	4,662.375	4081
	35	Noviembre	4,187.920	4824
	36	Diciembre	2,499.915	4247
	2015	37	Enero	2,079.590
38		Febrero	1,903.340	1637
39		Marzo	1,771.185	1597
40		Abril	1,526.770	1586
41		Mayo	2,051.745	1385
42		Junio	2,772.625	2061
43		Julio	2,272.645	2975
44		Agosto	1,826.265	2363
45		Septiembre	1,995.079	1733
46		Octubre	1,447.137	1923
47		Noviembre	1,300.761	1323
48		Diciembre	1,491.294	1156

**Tabla 4.21** Información de ventas correspondiente por mes, año y predicción con suavización exponencial ajustada a la tendencia del producto 3 (Fuente: elaboración propia)

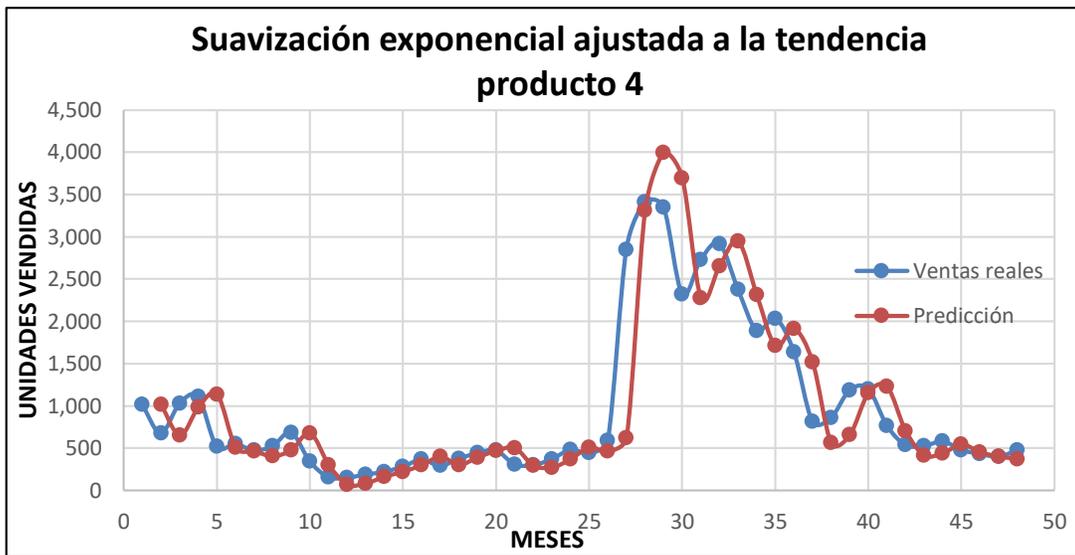


**Figura 4.21** Gráfica de suavización exponencial ajustada a la tendencia del producto 3 (Fuente: elaboración propia)

<b>Producto 4</b>				
<b>Año</b>	<b>No. de mes</b>	<b>Mes</b>	<b>Ventas reales</b>	<b>Predicción</b>
2012	1	Enero	1,015.855	
	2	Febrero	675.845	1018
	3	Marzo	1,025.660	650
	4	Abril	1,107.880	986
	5	Mayo	517.559	1136
	6	Junio	550.378	510
	7	Julio	473.712	462
	8	Agosto	524.433	406
	9	Septiembre	685.955	479
	10	Octubre	345.096	680
	11	Noviembre	158.481	303
	12	Diciembre	152.195	68
2013	13	Enero	189.018	83
	14	Febrero	221.767	162
	15	Marzo	281.487	222
	16	Abril	367.225	298
	17	Mayo	291.810	399
	18	Junio	376.840	303
	19	Julio	442.105	390
	20	Agosto	475.015	469
	21	Septiembre	309.190	502
	22	Octubre	300.345	294

	23	Noviembre	371.610	272
	24	Diciembre	480.570	369
2014	25	Enero	442.812	509
	26	Febrero	589.451	463
	27	Marzo	2,850.007	623
	28	Abril	3,411.831	3311
	29	Mayo	3,348.401	3996
	30	Junio	2,323.670	3696
	31	Julio	2,729.050	2279
	32	Agosto	2,917.445	2656
	33	Septiembre	2,375.210	2950
	34	Octubre	1,889.256	2315
	35	Noviembre	2,034.390	1709
	36	Diciembre	1,637.900	1916
	2015	37	Enero	813.820
38		Febrero	861.970	562
39		Marzo	1,187.765	656
40		Abril	1,198.985	1153
41		Mayo	765.925	1228
42		Junio	538.045	705
43		Julio	528.710	417
44		Agosto	580.110	439
45		Septiembre	478.127	546
46		Octubre	432.906	451
47		Noviembre	399.501	402
48		Diciembre	478.908	373

**Tabla 4.22** Información de ventas correspondiente por mes, año y predicción con suavización exponencial ajustada a la tendencia del producto 4 (Fuente: elaboración propia)



**Figura 4.22** Gráfica de suavización exponencial ajustada a la tendencia del producto 4 (Fuente: elaboración propia)

<b>Producto 5</b>				
<b>Año</b>	<b>No. de mes</b>	<b>Mes</b>	<b>Ventas reales</b>	<b>Predicción</b>
2012	1	Enero	520.327	
	2	Febrero	517.950	522
	3	Marzo	553.030	519
	4	Abril	546.576	556
	5	Mayo	538.926	553
	6	Junio	540.110	542
	7	Julio	530.333	540
	8	Agosto	516.063	530
	9	Septiembre	503.779	513
	10	Octubre	626.752	499
	11	Noviembre	493.036	646
	12	Diciembre	497.439	491
2013	13	Enero	528.709	483
	14	Febrero	488.980	527
	15	Marzo	424.148	484
	16	Abril	639.684	406
	17	Mayo	642.712	664
	18	Junio	553.660	679
	19	Julio	576.675	557
	20	Agosto	568.300	572
	21	Septiembre	499.965	565

	22	Octubre	517.660	485
	23	Noviembre	504.365	506
	24	Diciembre	492.080	498
2014	25	Enero	521.180	486
	26	Febrero	554.805	522
	27	Marzo	629.295	565
	28	Abril	556.010	653
	29	Mayo	654.020	562
	30	Junio	603.010	667
	31	Julio	651.705	610
	32	Agosto	842.682	659
	33	Septiembre	714.665	887
	34	Octubre	712.700	734
	35	Noviembre	769.611	708
	36	Diciembre	590.047	776
	2015	37	Enero	667.653
38		Febrero	568.808	651
39		Marzo	504.435	549
40		Abril	557.506	473
41		Mayo	631.671	545
42		Junio	590.556	646
43		Julio	626.914	599
44		Agosto	617.714	634
45		Septiembre	716.994	623
46		Octubre	623.224	737
47		Noviembre	629.552	626
48		Diciembre	571.779	622

**Tabla 4.23** Información de ventas correspondiente por mes, año y predicción con suavización exponencial ajustada a la tendencia del producto 5 (Fuente: elaboración propia)

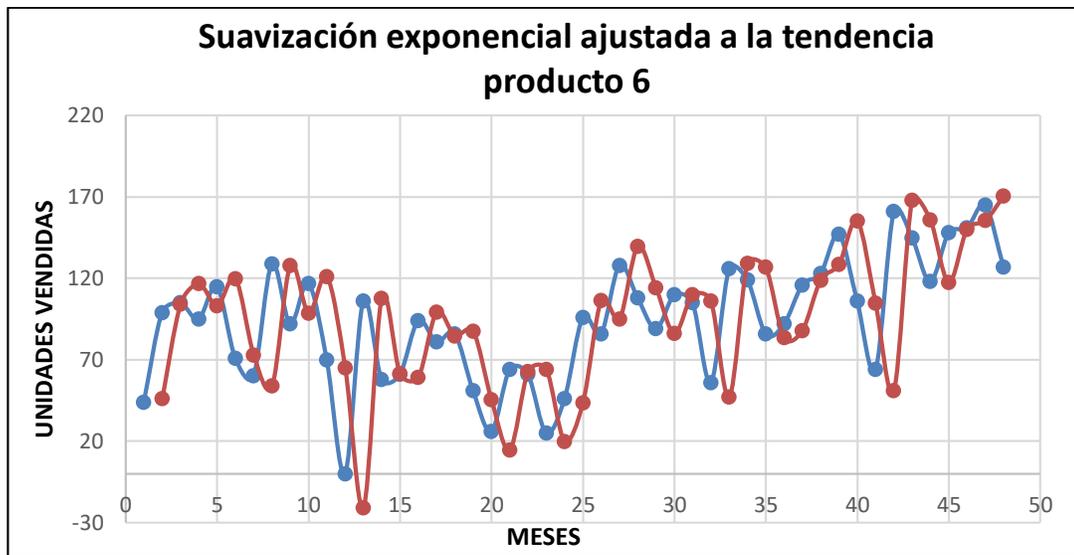


**Figura 4.23** Gráfica de suavización exponencial ajustada a la tendencia del producto 5 (Fuente: elaboración propia)

<b>Producto 6</b>				
<b>Año</b>	<b>No. de mes</b>	<b>Mes</b>	<b>Ventas reales</b>	<b>Predicción</b>
2012	1	Enero	44.000	
	2	Febrero	99.000	46
	3	Marzo	105.000	104
	4	Abril	95.000	117
	5	Mayo	115.000	103
	6	Junio	71.000	120
	7	Julio	60.000	73
	8	Agosto	129.000	54
	9	Septiembre	92.000	128
	10	Octubre	117.000	99
	11	Noviembre	70.000	121
	12	Diciembre	0.000	65
2013	13	Enero	106.000	-21
	14	Febrero	58.000	108
	15	Marzo	61.000	61
	16	Abril	94.000	59
	17	Mayo	81.000	99
	18	Junio	86.000	85
	19	Julio	51.000	88
	20	Agosto	26.000	45
	21	Septiembre	64.000	15
	22	Octubre	61.000	63

	23	Noviembre	25.000	64
	24	Diciembre	46.000	20
2014	25	Enero	96.000	43
	26	Febrero	86.000	106
	27	Marzo	128.000	95
	28	Abril	108.000	140
	29	Mayo	89.000	114
	30	Junio	110.000	86
	31	Julio	105.000	110
	32	Agosto	56.000	106
	33	Septiembre	126.000	47
	34	Octubre	119.000	129
	35	Noviembre	86.000	127
	36	Diciembre	92.000	83
	2015	37	Enero	116.000
38		Febrero	123.000	119
39		Marzo	147.000	129
40		Abril	106.000	155
41		Mayo	64.000	105
42		Junio	161.000	51
43		Julio	145.000	168
44		Agosto	118.000	156
45		Septiembre	148.000	117
46		Octubre	151.000	150
47		Noviembre	165.000	155
48		Diciembre	127.000	170

**Tabla 4.24** Información de ventas correspondiente por mes, año y predicción con suavización exponencial ajustada a la tendencia del producto 6 (Fuente: elaboración propia)



**Figura 4.24** Gráfica de suavización exponencial ajustada a la tendencia del producto 6 (Fuente: elaboración propia)

#### 4.1.5 Árboles aleatorios

Para la técnica de árboles aleatorios se comenzó proponiendo las diferentes variables que se necesitarían para ejecutar en el algoritmo. El código está desarrollado en Python e interviene una serie de parámetros que se extraen desde un archivo de Excel, guardado con la extensión “csv”. Tales parámetros fueron: ventas reales, día, mes, año, venta del día anterior y una estimación de la predicción que se esperaba obtener, dada por un número cercano a la venta real. Se ejecutó el programa para cada producto, con su respectivo archivo, generando 3000 árboles para tratar de obtener un mejor resultado con las iteraciones. Se evaluó el porcentaje de certeza más alto de cada producto, llegando a la conclusión de que sólo el producto 2 contaba con un porcentaje de confianza. Cabe mencionar que el respaldo de la base de datos que proporcionó la empresa no estuvo completo, se puede suponer que no se tuvieron ventas en algunos días o por causa de privacidad se modificaron algunos datos, dando como resultado un faltante de la información de las ventas en los productos, esto presentándose de forma aleatoria sin patrón identificado. El número de registros pertenecientes a los productos es: producto 1: 1,438, producto 2: 998, producto 3: 1,440, producto 4: 1,407, producto 5: 1,442 y producto 6: 1,157.

Al observar el porcentaje de certeza de la muestra, se realizaron pruebas con registros de ventas, para poder generar una predicción en unidades del producto 1, enviando como parámetro: venta

real, día, mes, año y la venta del día anterior, arrojando una predicción optima, siempre y cuando la venta fuera estable, al momento de que la venta bajara por mitad o más, el resultado difería de la venta real. A continuación, en la tabla 4.25 se puede observar los porcentajes de certeza de los productos.

Productos	Porcentaje de certeza
Producto 1	62.48%
Producto 2	99.10%
Producto 3	50.64%
Producto 4	27.43%
Producto 5	56.97%
Producto 6	15.03%

**Tabla 4.25** Tabla de porcentajes de certeza con árboles aleatorios de cada producto (Fuente: elaboración propia)

## 4.2 Porcentajes de errores

La técnica de Error Porcentual Absoluto Medio “MAPE” se seleccionó para evaluar la predicción generada de cada modelo. Aplicando la ecuación (7) se obtuvieron los porcentajes de error de cada producto; para las primeras cuatro técnicas el cálculo fue generado con Excel y para la técnica de árboles aleatorios se hizo en el código de Python, en la tabla 4.26 se puede apreciar el porcentaje de error de cada producto en cada modelo:

		Modelos				
		Regresión lineal	Promedios móviles	Suavización exponencial	Suavización exponencial ajustada a la tendencia	Arboles aleatorios
Productos	Producto 1	15.46%	28.87%	21.80%	21.71%	37.52%
	Producto 2	20.01%	29.56%	30.81%	33.70%	0.90%
	Producto 3	15.50%	22.05%	16.92%	18.19%	49.36%
	Producto 4	34.91%	38.95%	27.23%	30.83%	72.57%
	Producto 5	5.62%	7.72%	9.06%	9.86%	43.03%
	Producto 6	16.59%	28.49%	34.81%	37.57%	84.97%

**Tabla 4.26** Porcentajes de error calculado con MAPE de cada producto aplicando los diferentes modelos (Fuente: elaboración propia)

Con respecto a la tabla 4.26 se puede observar que el producto 1 tiene un porcentaje más bajo con el modelo de regresión lineal; el producto 2 tiene un porcentaje de 0.90% de error en la técnica de árboles aleatorios, siendo la mejor para éste. El producto 3 muestra similitud con el

producto 1 ya que el porcentaje más bajo es con la técnica de regresión lineal; para el producto 4 se puede notar una diferencia en el modelo con el porcentaje de error más bajo en la técnica de suavización exponencial. Por último, para los productos 5 y 6 se mostraron porcentajes más bajos con el modelo de regresión lineal. Cabe mencionar que el porcentaje más bajo de cada uno de los productos es en comparación con las otras técnicas.

A pesar de que el modelo de regresión lineal muestra en varios productos un porcentaje de error más bajo no se puede confirmar que sea una técnica confiable para generar predicciones, dado que, no resultó idónea para todos los productos.

## **5. Capítulo V. Conclusiones**

En este capítulo se describen comentarios finales sobre el resultado del proyecto, las cuales pueden ayudar como recomendaciones para tomar en cuenta en trabajo futuro.

## **5.1 Conclusiones**

La predicción de ventas es algo vital hoy en día y el saber qué rumbo tendrá un negocio puede ser de gran ayuda para los empresarios. Al desarrollar este proyecto se intentó elaborar una herramienta como apoyo a la toma de decisiones dentro de una abarrotera de Mamá Coneja ubicada en Cd. Guzmán, tras aplicar técnicas de minería de datos, así como regresión lineal, se concluyó que la hipótesis “La predicción de la cantidad de ventas de productos es más precisa aplicando una técnica de minería de datos” se ha rechazado dado que al aplicar diferentes técnicas no se pudo obtener un resultado óptimo en comparación con las ventas reales de la empresa.

Se hizo el respectivo análisis de la información, se generaron cinco modelos de predicción y con los resultados obtenidos se midió el error, arrojando un porcentaje alto para poder confiar en el resultado. Con esto se podría llevar a la pregunta si se hizo lo necesario para poder comprobar que la hipótesis no se cumplió, así fue, ya que a pesar de que hubo inconvenientes, con la información, las técnicas propuestas, se llegó hasta el final del desarrollo de cada una de ellas para así poder medir el error y analizar los riesgos que conllevaba un resultado con una diferencia significativa a la venta real.

Es necesario dar énfasis en que el fallo principal del proyecto fue el no considerar el total o la mayoría de las variables que intervienen en la venta en una tienda de este tipo, ya que, si se hubieran considerado más detalles, probablemente se hubiera podido cumplir la hipótesis.

Otro factor importante que se debe considerar para el éxito del desarrollo de un proyecto es, que cuando se involucra con una empresa es necesario tener una comunicación constante, por todo lo que eso conlleva, tanto solicitar los datos que se necesitan y que se los proporcionen en la brevedad posible, así como colaborar con las personas directamente involucradas a lo que se va necesitando.

## **5.2 Recomendaciones**

Se recomienda a futuros tesista que al abordar un tema de investigación se analice de forma profunda para poder plantearlo de la mejor manera, cabe mencionar que durante el desarrollo

de cada proyecto siempre existen cambios, como puede ser que se agreguen más puntos, o que los quiten, pero si se tiene bien formada la idea desde el inicio será una forma más fácil para poder sobrellevarlos.

### **5.3 Trabajo futuro**

Si existe la posibilidad de que alguien retome el proyecto para apoyo en algo similar, se debe tomar en cuenta que se tienen que replantear las estipulaciones, comenzando por plantear todas las variables que intervienen en una venta, hablar con las personas que llevan una relación sobre las ventas, su comportamiento, cómo se maneja la empresa y sobre todo, que ellos tengan un modelo de negocios bien definido, todo esto para facilitar el trabajo y evitar todos los contratiempos que otros investigadores han tenido.

## 6. Fuentes de información

### Libros

Geurts, L., Meertens, L. & Pemperton S. (2005). *ABC Programmer's Handbook*. United Kingdom: Editorial BOSKO BOOKS.

Vergaray, A. D. (2016). *Data mining: Minería de datos*. Lima, Perú: Editorial Macro.

### Digitales

Alvaer, J. O. (2018). *Árboles de decisión y Random Forest*. Recuperado el 20 de mayo de 2019 de <https://bookdown.org/content/2031/>

Basauri, O. J. (2015). *Construcción de un modelo predictivo de tipo Gradient boosting para ventas online*. Recuperado el 3 de septiembre de 2018 de <http://hdl.handle.net/10810/15852>

Barrientos, J., & Rumiany, D. (2005). *Modelo de negocios*. Publicaciones de Cátedra. Facultad de Ciencias Económicas. Universidad de Buenos Aires, Buenos Aires. Recuperado el 14 de marzo 2018 de <http://www.aximia.com/blog/wp-content/uploads/file/Modelo%20de%20negocios.pdf>

Boada, A. J., & de Vasconcelos, D. (2013). *Modelo estadístico de regresión múltiple, columna vertebral para predecir en empresas multinacionales con estilo de venta por catálogo*. Revista Lasallista de Investigación, 10 (1). Recuperado el 28 de septiembre de 2018 de <http://repository.lasallista.edu.co:8080/ojs/index.php/rldi/article/view/428/201>

Cardelli, L., Donahue, J., Glassman, L., Jordan, M., Kalsow, B. & Nelson, G. (1989). *Modular-3 Report*. Palo Alto, California. Recuperado el 10 de junio de 2019 de <https://www.hpl.hp.com/techreports/Compaq-DEC/SRC-RR-52.pdf>

Challenger, P. I., Díaz, R. Y. & Becerra, G. R. A. (2014). *El lenguaje de programación Python*. Ciencias Holguín, 20 (2), 1-12. Recuperado el 10 de mayo de 2018 de <http://www.ciencias.holguin.cu/index.php/cienciasholguin/article/view/826>

- Echeverri, L. A., Retamoza, A. M. P., De La Rosa, M. O., Barros, I. V., Álvarez, D. D. O., & Guerrero, E. C. (2013). *Minería de Datos como herramienta para el desarrollo de estrategias de mercadeo B2B en sectores productivos, afines a los colombianos: una revisión de casos*. *Sotavento mba*, (22), 126-136. Recuperado el 19 de diciembre de 2018 de [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2432274](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2432274)
- Espino T. C. (2017). *Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo-herramientas Open Source que permiten su uso*. Recuperado el 22 de octubre de 2018 de <http://hdl.handle.net/10609/59565>
- Griswold, R. E. & Griswold, M. T. (2002). *The icon programming language*. Recuperado el 1 junio de 2019 de <https://www2.cs.arizona.edu/icon/ftp/doc/lb1up.pdf>
- Grupo IGN. (2017). *Análisis predictivo: la información y el conocimiento es poder. IGN - Soluciones de gestión para pymes*. Recuperado 12 noviembre de 2017 de <https://ignsl.es/analisis-predictivo/>
- IBExpert documentation. (s. f.). Recuperado el 10 de junio de 2019 de <http://www.ibexpert.net/ibe/uploads/Doc/all.html>
- IBM Knowledge Center. *Nodo Arboles aleatorios*. Recuperado 30 de mayo de 2019 de [https://www.ibm.com/support/knowledgecenter/es/SS3RA7\\_sub/modeler\\_mainhelp\\_client\\_ddita/clementine/rf\\_general.html](https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddita/clementine/rf_general.html)
- Juárez, A. C., Zuñiga, C. A., Flores, J. L. M., & Partida, D. S. (2016). *Análisis de series de tiempo en el pronóstico de la demanda de almacenamiento de productos perecederos*. *Estudios Gerenciales*, 32 (141), 387-396. Recuperado 10 de abril de 2018 de <https://www.sciencedirect.com/science/article/pii/S0123592316300754>
- Kernighan, B. W. & Ritchie, D. M. (1991). *El lenguaje de programación en C*. Edo. De México: Editorial Pearson Education. Recuperado el 10 de junio de 2019 de [http://www.academicos.ccadet.unam.mx/mario.gonzalez/cursos\\_archivos/mn/Kernighan.pdf](http://www.academicos.ccadet.unam.mx/mario.gonzalez/cursos_archivos/mn/Kernighan.pdf)

- Medina, M. R. F. & Ñique, Ch. C. I. (2017). *Bosques aleatorios como extensión de los árboles de clasificación con los programas R y Python*. *Interfases*, (010), 165-189. Recuperado el 18 de abril de 2019 de <http://revistas.ulima.edu.pe/index.php/Interfases/article/view/1775>
- Núñez, E., Steyerberg, E. W., & Núñez, J. (2011). *Estrategias para la elaboración de modelos estadísticos de regresión*. *Revista española de cardiología*, 64 (6), 501-507. Recuperado 15 de Agosto 2018 de <https://www.revespcardiol.org/es-estrategias-elaboracion-modelos-estadisticos-regresion-articulo-S0300893211003502>
- Preciado, M. P. & Oliva, E. J. D. (2011). *Modelos de negocio: propuesta de un marco conceptual para centros de productividad*. *Administración & desarrollo*, 39 (53), 23-34. Recuperado el 22 de enero de 2019 de <http://www.bdigital.unal.edu.co/5152/1/940794.2011.pdf>
- Randell, B. & Russel, L. J. (1964). *Algol 60 implmentation, the translation and use of algol 60 programs on a computer*. New York, New York. Edition published by Academic Press inc. Recuperado el 5 de junio de [http://www.softwarepreservation.org/projects/ALGOL/book/Randell\\_ALGOL\\_60\\_Implementation\\_1964.pdf](http://www.softwarepreservation.org/projects/ALGOL/book/Randell_ALGOL_60_Implementation_1964.pdf)
- Revista UNAM, (s. f.). *Conceptos básicos de clustering*. Recuperado el 1 junio 2019 de <http://www.revista.unam.mx/vol.4/num2/art3/cluster.htm>
- Riquelme S., J. C., Ruiz R., & Gilbert, K. (2006). *Minería de Datos: Conceptos y tendencias*. *Inteligencia artificial: Revista Iberoamericana de Inteligencia Artificial*, 10(29), 11-18. Recuperado el 28 de diciembre de 2018 de <https://idus.us.es/xmlui/handle/11441/43290>
- Sarmiento, E. M. (2008). *Predicción con series de tiempo y regresión*. *Panorama*, 2 (4), 2. Recuperado el 18 de diciembre de 2018 de <https://dialnet.unirioja.es/servlet/articulo?codigo=4780125>

- Silva, R. M. A. (2013). *Modelo de pronóstico para la estimación de la utilización y confiabilidad de equipos dinámicos. Caso: Equipo de compresión de la RPMNE*. Recuperado el 10 de junio de 2019 de Bibliografía latinoamericana en revistas de investigación científica y social sitio web: <https://biblat.unam.mx/hevila/Ingenieriapetrolera/2013/vol53/no5/3.pdf>
- Tejera, V. T. (s. f.). *Firebird: lenguaje de definición de datos*. Recuperado el 10 de junio de 2019 de [https://firebird.com.mx/descargas/documentos/tema\\_3-ddl.pdf](https://firebird.com.mx/descargas/documentos/tema_3-ddl.pdf)
- The American Heritage® Dictionary of the English Language (1). (s. f.). *Insight*. Recuperado el 11 de junio 2019 de <https://ahdictionary.com/word/search.html?q=insight>
- The American Heritage® Dictionary of the English Language (2). (s. f.). *Open source*. Recuperado el 11 de junio 2019 de <https://ahdictionary.com/word/search.html?q=open+source>
- The American Heritage® Dictionary of the English Language (3). (s. f.). *Score*. Recuperado el 24 de junio de 2019 de <https://ahdictionary.com/word/search.html?q=score>
- Valcárcel, A. V. (2004). *Data Mining y el descubrimiento del conocimiento. Industrial Data*, 7(2). Recuperado el 11 de mayo de 2018 de <http://www.redalyc.org/html/816/81670213/>

## 7. Glosario

**Algol 60.** Es un lenguaje, basado en notación matemática normal, que ha sido ampliamente utilizado para la publicación de los detalles de los procesos computacionales, en forma de 'algoritmos', y que es lo suficientemente preciso para ser utilizado para la programación de ordenadores digitales automáticos (Randell & Russel, 1964).

**Clustering.** El concepto *clustering* se refiere a una técnica que permite combinar múltiples sistemas para que trabajen en paralelo y se comporten como un recurso informático unificado para: servir a un grupo de tareas, proporcionar tolerancia a fallos y tener disponibilidad continua (Revista UNAM, s. f.).

**Firebird.** Firebird es un sistema gestor de bases de datos relacional. Como tal, está diseñado para soportar la creación y mantenimiento de estructuras de datos abstractas, no sólo almacenar datos sino también mantener las relaciones y optimizar la velocidad y consistencia cuando los datos pedidos son enviados a los clientes (Tejera, s. f.).

**IBExpert.** IBExpert es un Entorno de Desarrollo Integrado (IDE) profesional para el desarrollo y la administración de las bases de datos Firebird (IBExpert documentation, s. f.).

**Icon.** Es uno de los lenguajes de programación de alto nivel de propósito general, que contiene una amplia variedad de características para procesar y presentar datos simbólicos, cadenas de caracteres y estructuras tanto en texto como en imágenes gráficas (Griswold, R. E, Griswold, M. T., 2002).

**Insights.** El acto o resultado del entendimiento de la interna u oculta naturaleza de las cosas o de percibirlo de una manera intuitiva (The American Heritage® Dictionary of the English Language (1), s. f.).

**Lenguaje ABC.** Es un lenguaje que fue diseñado con el objetivo de ofrecer un lenguaje fácil de usar, ideal para la enseñanza y el aprendizaje de la programación. ABC es más legible que otros idiomas y no tiene comandos oscuros (Geurts, Meertens & Pemperton, 2005).

**Lenguaje C.** Es un lenguaje de programación de propósito general que ha sido estrechamente asociado con el sistema UNIX en donde fue desarrollado puesto que tanto el sistema como los

programas que corren en él están escritos en lenguaje C. Sin embargo, este lenguaje no está ligado a ningún sistema operativo ni a ninguna máquina, y aunque se le llama “lenguaje de programación de sistemas” debido a su utilidad para escribir compiladores y sistemas operativos, se utiliza con igual eficacia para escribir importantes programas en diversas disciplinas (Kernighan & Ritchie, 1991).

**Modula-3.** Es uno de los lenguajes de programación que utiliza la equivalencia estructural de modula 2. El programa especifica un cálculo que actúa sobre una secuencia de componentes digitales llamados ubicaciones. Una variable es un conjunto de ubicaciones que representa un valor matemático de acuerdo con una convención determinada por el tipo de la variable (Cardelli, Donahue, Glassman, Jordan, Kalsow & Nelson, 1989).

**Open Source.** Abierto a la participación pública, como en el intercambio de información o el desarrollo de software: un sistema operativo de código abierto (The American Heritage® Dictionary of the English Language (2), s. f.).

**Score.** Un resultado, generalmente expresado numéricamente, de una prueba o examen (The American Heritage® Dictionary of the English Language (3), s. f.).