

# Análisis del comportamiento de la predicción de ventas con diferentes técnicas de minería de datos

Teresita Moreno Ramírez<sup>1</sup> y Rosa María Michel Nava<sup>2</sup>

Maestría en ciencias de la computación<sup>1</sup>, Sistemas y Computación<sup>2</sup>  
Tecnológico Nacional de México/ Instituto Tecnológico de Cd. Guzmán  
Ciudad Guzmán, Jalisco, México

teresita.moreno@outlook.com, michel91\_3@hotmail.com (autor de correspondencia)

*Abstract*— 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, taking for the analysis two quantitative variables, sales and time, were defined to apply them to a prediction model. As part of the methodology, data mining techniques were applied: the time series (moving averages, exponential smoothing y trend adjusted exponential smoothing); linear regression and random forest. Prediction models were created with these techniques, for generate sales predictions with greater precision. The results are aimed at benefiting grocery companies, for control of their inventory and sales.

*Keyword*— Prediction, sales, business model, data mining.

*Resumen*— 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 una empresa, tomando para el análisis dos variables de tipo cuantitativo, las ventas y el tiempo, para aplicarlas a un modelo de predicción. Como parte de la metodología, se aplicaron técnicas de minería de datos: series de tiempo (promedios móviles, suavización exponencial y suavización exponencial ajustada a la tendencia); regresión lineal y arboles aleatorios. Se crearon modelos de predicción con las técnicas, para generar predicciones de ventas con mayor precisión. Los resultados van orientados a beneficiar a las empresas abarroteras, para el control en su inventario y en sus ventas.

*Palabras claves*—Predicción, ventas, modelo de negocios, minería de datos.

## I. INTRODUCCIÓN

En la actualidad las empresas abarroteras tienen un índice de venta que ha ido incrementando con el paso del tiempo, razón por la cual, los empresarios analizan la información de sus ventas, con el fin de poder identificar qué productos son los que representan un mayor movimiento en sus ganancias. 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.

Así mismo 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 abarrotera que se tomó para el estudio, las decisiones están basadas en la experiencia que se tiene en el negocio, considerando los productos de mayor movimiento en ventas, lo cual no siempre resulta de forma acertada, ya que se puede llegar a 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 base en esto, nace la propuesta para realizar un estudio sobre técnicas de minería de datos para la estimación de las ventas sobre la información de dicha empresa. Mediante el análisis de datos se quiere llegar a una conclusión de cómo se debe abordar la problemática de predicción de ventas, tomando en cuenta dos variables, ventas y el tiempo, para así evaluar los resultados con las ventas reales y hacer una comparativa, para demostrar si se pueden generar predicciones precisas seleccionando y aplicando una técnica de minería de datos, y que esto pueda servir de ayuda para la toma de decisiones.

Para esto se seleccionó una línea de productos denominada línea 1, que contiene una cantidad de 260 productos, de los cuales se tomaron cinco productos enlistados dentro de los diez más vendidos y un sexto producto de un nivel aleatorio. El principal objetivo fue el aplicar diferentes técnicas a los artículos seleccionados, para poder analizar si existe una técnica de minería de datos que pueda dar como resultado predicciones con porcentaje de error bajo con respecto a la venta real, tomando en cuenta sólo las variables planteadas.

Se usaron herramientas informáticas como minería de datos, análisis de datos, programas manejadores de bases de datos como IBExpert y Firebird, así como el lenguaje de programación Python y la herramienta de Microsoft Office Excel.

## II. TRABAJOS RELACIONADOS

A continuación, se hace la descripción de algunas investigaciones vitales para el desarrollo de este proyecto, basándose en su fundamento teórico.

### *A. 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.

### *B. 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).

### C. Estrategias para la elaboración de modelos estadísticos de regresión

El objetivo de este proyecto es el de 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).

### D. 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 el porcentaje que resultó en las pruebas.

## III. FUNDAMENTO 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 abarrotera con la que se trabajó el proyecto, 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).

Existen otras técnicas para predicción como los son los árboles 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.

Las ecuaciones utilizadas para la aplicación de los modelos, se describen a continuación:

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.

Algunas de las técnicas con las que se trabajaron de series de tiempo las 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.

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

Donde  $\dot{X}_t$  es la variable a la cual se le asigna el resultado en el periodo t; donde  $\dot{X}_{t-1}$  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,

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

Serie de suavización exponencial,

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

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).

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. 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).

#### 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.

## IV. METODOLOGÍA

Tomando en cuenta el tipo de proyecto que se abordó, el tipo de investigación que se utilizó fue aplicada porque se centra en encontrar las técnicas para lograr un objetivo en partículas. El ámbito al que se desarrollo es específico y delimitado, se aborda un problema concreto.

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

La población está conformada por una abarrotera, a la cuál está basada este proyecto. Se hicieron consultas para extraer la información necesaria en base a las variables planteadas, con el fin de tener

seleccionada solo la información requerida. La muestra que se planteó se extrajo de la abarrotera ubicada en Cd. Guzmán, Jalisco.

El procedimiento que se siguió para desarrollar la investigación, se basó en la metodología Descubrimiento de Conocimiento en Bases de Datos o KDD, se muestra a continuación en la figura 1.

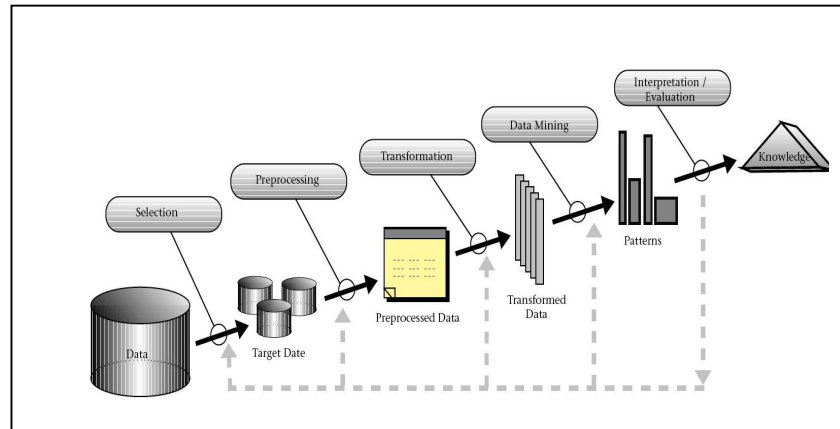


Fig. 1. Descripción general de los pasos que constituyen el proceso KDD.

A continuación, se enlistan los pasos de la metodología como lo plantea Vergaray (2016):

- Aprendizaje del dominio de la aplicación: este paso incluye adquirir el conocimiento previo relevante y el planteo de los objetivos de la aplicación.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.

- 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.

### V. RESULTADOS

A continuación, se pueden observar los resultados de las pruebas realizadas con las técnicas de regresión lineal, promedios móviles, suavización ajustada, suavización ajustada a la tendencia y arboles aleatorios:

En la figura 2 se muestra la gráfica correspondiente al modelo de regresión lineal, donde en el eje vertical se representan las unidades vendidas; en el eje horizontal se representan los meses correspondientes a cada venta. Para aplicar esta técnica primero se hicieron los cálculos necesarios para obtener la media aritmética y geométrica, para seleccionar la que presentara menor error, así mismo, seleccionando la media aritmética. Cabe mencionar que la información en la cual se realizó la extracción de los datos, corresponde a un respaldo de cuatro años. Se puede apreciar como la línea de predicción (puntos rojos) no está ajustada en comparación con los puntos azules que representan la venta real.

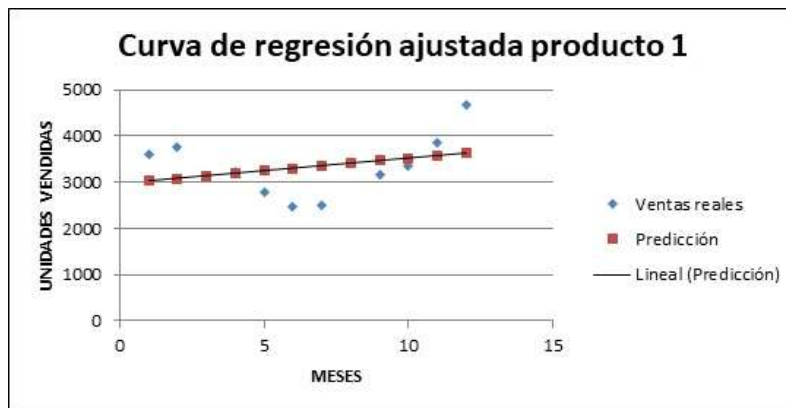


Fig. 2. Gráfica de regresión lineal del producto 1 (elaboración propia)

En la figura 3 se pueden observar los resultados obtenidos al aplicar la ecuación (2) descrita en el fundamento teórico, la cual corresponde al modelo de promedio móvil, que al analizar la predicción generada (línea roja) con respecto a la venta real que representa su comportamiento histórico (línea azul), muestra un desfase, así la cantidad de meses que se tomaron para promedio móvil fue de tres, ya que se hicieron pruebas que demostraron que esta cantidad fue la más idónea para continuar con la prueba de los demás productos.



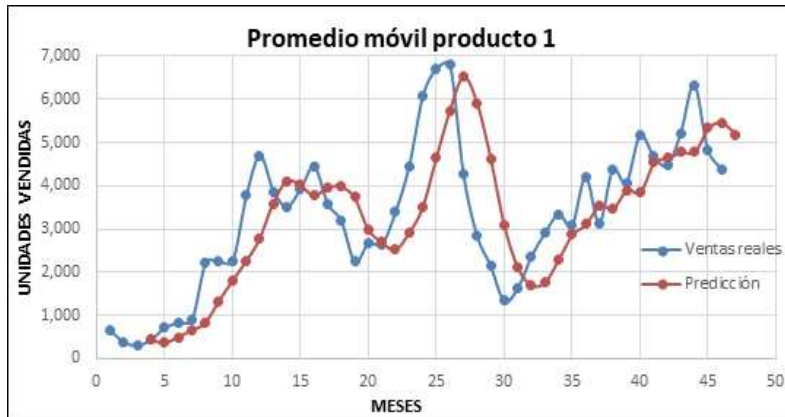


Fig. 3. Gráfica de promedio móvil del producto 1 (elaboración propia)

La figura 4 muestra el resultado obtenido gráficamente sobre el modelo de suavización exponencial en la que se aplicó la ecuación (3) (descrita en el apartado fundamento teórico), existiendo una diferencia notable entre la venta real (línea azul) y la estimación generada (línea roja).



Fig. 4. Grafica de suavización exponencial del producto 1 (elaboración propia)

A continuación en la figura 5 se pueden observar gráficamente los resultados obtenidos aplicando la técnica de suavización exponencial ajustada a la tendencia, 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 muy notoria, haciendo que este modelo no sea confiable.



Fig. 5. Gráfica de suavización exponencial ajustada a la tendencia del producto 1 (elaboración propia)



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 en 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.

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 óptima, 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 1 se pueden observar los porcentajes de certeza de los productos. Los resultados obtenidos de las pruebas sólo se muestran con las gráficas correspondientes al producto 1, a pesar de que se evaluaron seis, esto para evitar sobrecargar de gráficas el artículo.

Tabla I. Tabla de porcentajes de certeza con árboles aleatorios de cada producto.

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

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 II se puede apreciar el porcentaje de error de cada producto en cada modelo:

Tabla II. Porcentaje de error calculado con MAPE de cada producto aplicando los diferentes modelos.

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

Con respecto a la tabla II 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.

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.

## VI. 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 esta investigación se intentó elaborar una herramienta como apoyo a la toma de decisiones dentro de una abarrotera ubicada en Cd. Guzmán, tras aplicar técnicas de minería de datos, así como regresión lineal, se concluyó que al aplicarlas 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 no se cumplió, así fue, ya que se llegó hasta el final del desarrollo de cada uno de los modelos para posteriormente evaluarlos y analizar los riesgos que conllevaba un resultado con una diferencia significativa a la venta real.

Es necesario dar énfasis que a pesar que se probaron diferentes técnicas se identificó que el fallo principal 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 variables y otros factores, probablemente se hubieran podido generar resultados óptimos.

## REFERENCIAS

- Alvaer, J. O. (2018). *Árboles de decisión y Random Forest*. Recuperado el 20 de mayo de 2019 de <https://bookdown.org/content/2031/>
- 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\\_ddit\\_a/clementine/rf\\_general.html](https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddit_a/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>

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