

# Predictive model based on machine learning for the prevention of overstock in a footwear company

Karina Mercedes Aguirre Méndez, University Student<sup>1</sup>, Alfonso Lorenzo Moreno Torres, University Student<sup>2</sup>,  
Christian Ovalle<sup>3</sup>

<sup>1</sup>Universidad Privada del Norte, Perú, N00247996@upn.pe

<sup>2</sup>Universidad Privada del Norte, Perú, N00253071@upn.pe

<sup>3</sup>Universidad Privada del Norte, Perú, denis.ovalle@upn.pe

*Abstract - Sales forecasting is an essential process for companies as it allows them to plan and make appropriate decisions about their workforce, cash flow and resources. The objective of sales forecasting is to predict future sales from the previous period's sales information. For the retail industry, it is very important because accurate forecasts can help companies maximize their investments, reduce inventory costs, increase sales and profits, and reduce risks. The most recent and effective method for forecasting future data is Machine Learning. Likewise, in the present work, the logistic regression algorithm and decision tree have been applied to determine the best-selling products and categories. The logistic regression algorithm had an accuracy of 97%, with a confusion matrix of 98.1% and 94.4% of true positives and true negatives respectively. The accuracy metric was 97% and the completeness metric was 96%. The decision tree algorithm had an accuracy of 85%, with a confusion matrix of 86.6% and 83% of true positives and true negatives respectively. The accuracy metric was 87% and the completeness metric was 84%. It could be determined that the balerina category is the most sold category with 84.3%, and that the spring and summer seasons are the most sold.*

*Keywords— Predictive model, machine learning, algorithm, overstock, Crisp-DM.*

**Digital Object Identifier:** (only for full papers, inserted by LACCEI).  
**ISSN, ISBN:** (to be inserted by LACCEI).  
**DO NOT REMOVE**

# Modelo predictivo basado en machine learning para la prevención de sobrestock en una empresa de calzado

Karina Mercedes Aguirre Méndez, Estudiante<sup>1</sup>, Alfonso Lorenzo Moreno Torres, Estudiante<sup>2</sup>,  
Christian Ovalle<sup>3</sup>

<sup>1</sup>Universidad Privada del Norte, Perú, N00247996@upn.pe

<sup>2</sup>Universidad Privada del Norte, Perú, N00253071@upn.pe

<sup>3</sup>Universidad Privada del Norte, Perú, denis.ovalle@upn.pe

**Resumen** - La previsión de ventas es un proceso esencial para las empresas ya que les permite planificar y tomar decisiones adecuadas sobre su fuerza laboral, flujo de efectivo y recursos. El objetivo de la previsión de ventas es predecir las ventas futuras a partir de la información de ventas del período anterior. Para el sector minoritario, es muy importante ya que los pronósticos precisos pueden ayudar a las empresas a maximizar sus inversiones, reducir los costos de inventario, aumentar las ventas y las ganancias y reducir los riesgos. El método más reciente y efectivo para pronosticar datos futuros es el machine learning. Así mismo, en el presente trabajo se ha aplicado el algoritmo de regresión logística y árbol de decisión para determinar los productos y categorías más vendidos. El algoritmo de regresión logística tuvo una precisión del 97%, con una matriz de confusión de 98.1% y 94.4% de verdaderos positivos y verdaderos negativos respectivamente. La métrica de exactitud fue del 97% y la de exhaustividad del 96%. El algoritmo de árbol de decisión tuvo una precisión del 85%, con una matriz de confusión de 86.6% y 83% de verdaderos positivos y verdaderos negativos respectivamente. La métrica de precisión fue del 87% y la de exhaustividad del 84%. Se pudo determinar que la categoría balerina es la más vendida con un 84.3%, y que las estaciones de primavera y verano son las que más se venden.

**Palabras claves**— Modelo predictivo, machine learning, algoritmo, sobrestock, Crisp-DM.

## I. INTRODUCCIÓN

El pronóstico de ventas proporciona una comprensión de cómo una organización debe administrar su fuerza laboral, flujo de efectivo y recursos. Para que la empresa planifique y tome decisiones adecuadamente, este es un requisito previo esencial que da a las empresas la capacidad de planificar eficazmente sus estrategias comerciales [1].

El propósito de la previsión de ventas es predecir las ventas futuras en función de la información de ventas del período anterior. Para el sector empresarial minoritario, la previsión de las ventas de bienes es muy importante. Los pronósticos precisos pueden ayudar a las empresas a maximizar sus inversiones, reducir los costos de inventario, aumentar las ventas y las ganancias y reducir los riesgos. En la actualidad las tecnologías emergentes permiten registrar información sobre el proceso de venta de mercancías. El desarrollo de algoritmos puede ayudar a los investigadores a obtener conocimiento valioso a partir de cantidades masivas de datos, que luego

pueden usar para hacer predicciones precisas y dirigir la creación y venta de productos [2].

En la era moderna, cualquier pequeña o mediana empresa necesita una estrategia de gestión de existencias impecable para satisfacer las necesidades de los clientes. Requiere invertir mano de obra calificada y mucho dinero para su realización. Para mantener su stock de acuerdo con las demandas automáticas generadas por sus clientes a lo largo del tiempo, los gigantes del comercio electrónico utilizan técnicas de aprendizaje. Para las pequeñas o medianas empresas, esto significa que la gestión de almacenes se puede ofrecer como un servicio para pronosticar la demanda de una variedad de productos y aumentar sus ventas. Al considerar el pronóstico de demanda, un componente fundamental de todas las operaciones comerciales [3].

Las empresas de hoy en día se enfrentan con frecuencia a ciclos de vida de productos más cortos, lo que hace que sea más importante predecir con precisión la demanda de los consumidores de productos introducidos recientemente. Son capaces de apoyar decisiones operativas como compras y control de inventario gracias a estos pronósticos. Pronosticar la demanda de nuevos productos es un desafío porque los datos históricos de ventas no están disponibles como un predictor de ventas futuras [4].

El método más reciente y efectivo para pronosticar muestras de datos futuras basadas en muestras masivas del pasado es el que enfatiza el machine learning. Esto permite al usuario realizar un análisis cuantitativo de los datos utilizando un subconjunto de datos conocido como "características" que se basan principalmente en la salida. Estas funciones permiten al usuario implementar un conjunto de reglas que resultan en una colección de ecuaciones que finalmente producen la salida [5].

Es necesario crear un modelo adecuado basado en datos de ventas anteriores para la previsión de ventas y, a continuación, utilizar ese modelo para predecir las ventas futuras es el siguiente paso. Los algoritmos de machine learning pueden ayudar a la hora de elegir un modelo. Rápidamente identificamos las características del modelo para que pudiera adaptarse a las ventas anteriores [1].

En el presente trabajo de investigación se desarrollará un modelo de machine learning, el cual será utilizado para la predicción de las ventas con el fin de prevenir el sobrestock de productos en una empresa de calzados. Para el desarrollo del modelo de machine learning se utilizará la herramienta Orange la cual es fácil de manipular y ayuda al análisis de datos de una manera intuitiva. Así mismo, se utilizará el árbol de decisión para tener un mapa de posibles resultados de una serie de decisiones.

## II. ESTADO DEL ARTE

### A. *Análisis predictivo*

Un análisis predictivo más avanzado implica pronosticar resultados potenciales y hacer recomendaciones teniendo en cuenta las tendencias actuales y las posibilidades futuras. Esto va más allá de las consultas y los informes, e incluye técnicas más sofisticadas como estadísticas, análisis predictivo, aprendizaje automático, simulación de resultados y optimización que busca patrones y tendencias en los datos, que suelen ser una combinación de datos estructurados y no estructurados [6].

Así mismo, el análisis predictivo utiliza una variedad de técnicas estadísticas, en este caso, aprendizaje automático, para examinar datos históricos o actuales sobre tendencias o patrones para pronosticar eventos futuros [7].

### B. *Recolección de datos*

La tarea más crucial cuando se trabaja con aprendizaje profundo y automático es la recopilación de datos. Primero, debemos confirmar qué datos tendrán un efecto en la predicción. Luego determine quién puede proporcionarnos esos detalles y hacia dónde debemos dirigir nuestros esfuerzos. Después de recopilar los datos necesarios, debemos procesar los datos. Porque el procesamiento de datos es un factor importante en la predicción. El resultado de nuestro modelo de predicción podría cambiar como resultado del comportamiento incorrecto de un solo dato [8].

En 2020 K. Singh, P. M. Booma, U. Eaganathan [9]. Durante su investigación afirman que se debe reunir un conjunto de datos de negocios relacionados para identificar modelos comerciales distintivos y cómo se ven afectadas las ventas. Así también, el tamaño del conjunto de datos debe ser superior a 20 meses, ya que una cantidad menor no es suficiente para proporcionar predicciones precisas y comparaciones de tendencias.

### C. *Machinne Learning*

El aprendizaje automático les da a las computadoras la

capacidad de aprender sin la intervención humana de la programación relacionado con el análisis predictivo. Para implementar modelos y algoritmos complejos que tienden a producir predicciones, el aprendizaje automático tiene una amplia gama de aplicaciones. Estos modelos ayudan a los estadísticos, científicos de datos, ingenieros e investigadores a tomar decisiones precisas y confiables que proporcionen los resultados deseados [10].

En 2019 M. A. I. Arif, S. I. Sany, F. I. Nahin y A. S. A. Rabby [8]. Durante su investigación afirman que la construcción de modelos para el reconocimiento de usuarios en dominios que van desde la visión asistida por computadora hasta el reconocimiento de voz y la comprensión de texto hasta los juegos de inteligencia artificial ha mostrado un éxito significativo en el campo del aprendizaje automático. En ingeniería y ciencias, el aprendizaje automático, y en particular el aprendizaje profundo, se están volviendo cada vez más significativos y fructíferos. A pesar de su gran éxito, el aprendizaje automático puede tener sus límites si se trata de datos de entrenamiento inadecuados. Una posible solución es la integración adicional de conocimientos previos en el proceso de aprendizaje que da como resultado la noción de aprendizaje automático informado. La principal contribución es la creación de una taxonomía que permita la categorización enfocada y la identificación de las principales vías. Esto ayuda a los usuarios actuales y futuros del aprendizaje informado automatizado a reconocer los enfoques apropiados para aplicar el conocimiento previo.

### D. *Árbol de decisiones*

El árbol de decisión es una clasificación conocida como espacio instantáneo recursivo. Es un método sólido para realizar análisis de múltiples variables y una herramienta confiable de minería de datos. Sus aplicaciones se encuentran en una variedad de campos, y este enfoque representa los factores que intervienen en el logro de un objetivo predeterminado, así como los factores necesarios para hacerlo y los métodos utilizados para lograrlo [1].

En 2021 A. Shehadeh, O. Alshboul, R. E. Al Mamlook y O. Hamedat [11]. Se investigó el uso del aumento de gradiente para disminuir la varianza y aumentar la precisión del modelo. Además, los árboles de decisión se investigaron utilizando "Gini" y "entropía" para obtener información, y Ganancia se utilizó para determinar qué característica o atributo puede proporcionar la mayor cantidad de información sobre una clase. Los mejores resultados se obtuvieron utilizando bagging con un bin máximo de 100 y una tasa de aprendizaje promedio de 0.04, mientras que el número de hojas requeridas fue de 32.

### E. KDD

Knowledge Discovery Databases (KDD) es el proceso de extraer conocimiento oculto de las bases de datos. KDD requiere conocimientos previos pertinentes, así como una comprensión rápida del dominio y los objetivos de la aplicación. La naturalidad iterativa e interactiva caracteriza el modelo de proceso KDD [12].

### F. CRISP-DM

La metodología CRISP-DM, también conocida como Modelo de Proceso Estándar para Minería de Datos, divide el ciclo de vida de un proyecto de exploración de datos en seis fases, cada una de las cuales interactúa iterativamente con las demás a lo largo del desarrollo del proyecto [13].

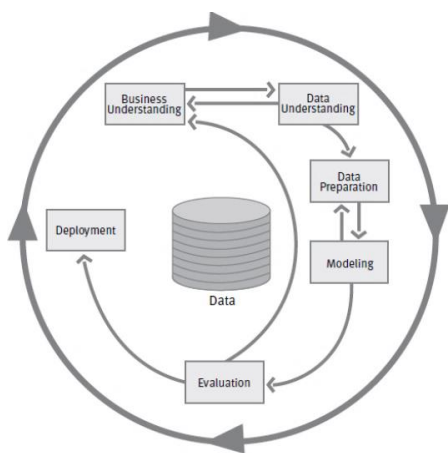


Fig. 1 Fases de la metodología CRISP-DM.

### G. SEMMA

Un método de minería de datos creado por el Instituto SAS se llama estándar SEMMA (Sample, Explore, Modify, Model, and Assess). Ofrece y permite la comprensión, organización, desarrollo y mantenimiento de proyectos de minería de datos. ayuda en el suministro de respuestas a los problemas y metas del negocio. SAS Enterprise Miner está conectado a SEMMA, que es principalmente una organización lógica para herramientas de trabajo para proyectos de minería de datos. Tiene un ciclo de cinco etapas [14].

## III. METODOLOGÍA

El presente trabajo de investigación acorde a su propósito se sitúa en el tipo de investigación aplicada, ya que, busca producir conocimiento que sea directamente aplicable a los problemas de la sociedad o del sector

industrial, centrándose en el vínculo entre la teoría y el resultado [15]. Así también, conforme al objetivo principal de la investigación se puede determinar que el nivel de investigación es predictivo, debido a que, este se basa en el conocimiento adquirido que se puede aplicar en el futuro [16].

Por otro lado, debido a que se manipulara la variable independiente el diseño de la investigación es experimental. El diseño experimental, es una situación de control cuando una o más variables independientes se cambian intencionalmente para estudiar los efectos de tal manipulación en una o más variables dependientes [17]. Así mismo, se puede afirmar que el tipo de diseño de la investigación es pre-experimental, ya que su grado de control es mínimo y se aplican a un solo grupo. Por último, se puede determinar que la presente investigación es de tipo cuantitativa, ya que, permitirá estimar las magnitudes u ocurrencia de los fenómenos y probar hipótesis [18].

### A. Análisis comparativo de metodologías de minería de datos contempladas

FASE	CARACTERÍSTICAS	METODOLOGÍAS		
		CRISP-DM	KDD	SEMMA
ANÁLISIS DEL PROBLEMA	Evaluación de la organización	SI		
	Identificación de stakeholders	SI	SI	SI
	Definición del problema			SI
	Evaluación de fuentes de datos		SI	SI
	Análisis de soluciones potenciales	SI		
	Definición de objetivos del proyecto	SI		
	Determinación de criterio de éxito	SI	SI	
	Análisis de técnicas de DM	SI	SI	
SELECCIÓN Y PREPARACION DE DATOS	Especificación de documentos(entregables)			
	Análisis exploratorio	SI	SI	
	Limpieza de datos	SI	SI	SI
	Transformación de variables y creación de atributos derivados	SI	SI	SI
	Análisis descriptivo de datos depurados			SI
MODELADO	Revisión del conjunto de datos final con el usuario		SI	
	Selección de técnicas de minería de datos	SI	SI	
	Evaluación de resultados	SI	SI	SI
	Evaluación de nuevos modelos	SI		SI
EVALUACION	Directivas para el descubrimiento de patrones		SI	
	Interpretación de modelos	SI	SI	SI
	Comparación y ponderación de modelos	SI	SI	
	Revisión del proceso	SI	SI	SI
IMPLEMENTACION	Directivas para modelos no viables	SI	SI	
	Implementación de nuevos modelos	SI	SI	SI
	Programa de mantenimiento	SI	SI	
	Resumen del proyecto	SI		
Total de características cumplidas:		19	17	11
% Cumplimiento		76%	68%	44%

Fig. 2 Actividades específicas por metodología.

Realizado el análisis comparativo entre las metodologías de minería de datos KDD, SEMMA y CRISP-DM, se determinó que la metodología utilizada para el desarrollo será CRISP-DM, esto en base a los resultados obtenidos en la Figura 2.

## B. Modelo predictivo para reducción de sobrestock

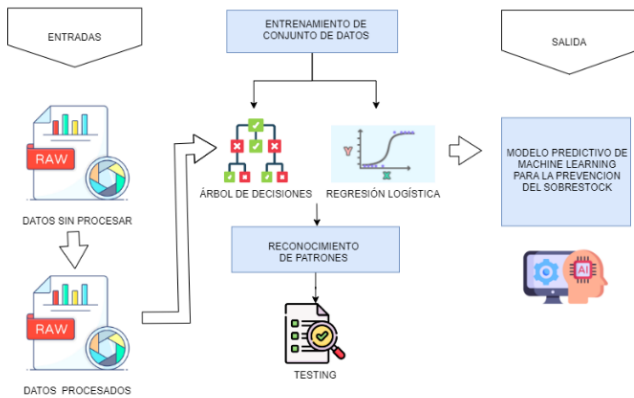


Fig. 3 Modelo predictivo para reducción de sobrestock.

En el modelo de Machine Learning propuesto los pasos a desarrollar son:

Paso 1: Recolección de datos, dada la problemática de la empresa se procederá a recopilar la información relacionada al histórico de ventas.

Paso 2: Preparación de datos con el fin de determinar la información relevante con respecto a los productos vendidos.

Paso 3: Se aplicarán algoritmos de predicción (árbol de decisiones y regresión logística) con la finalidad de tener una proyección de la demanda de productos.

Paso 4: Se aplicarán métricas de rendimiento que permitan evaluar la eficiencia de los algoritmos empleados.

## C. Desarrollo de la metodología CRISP-DM

### Fase 1: Comprender el negocio

La empresa de calzados comercializa desde hace 15 años diferentes marcas de calzados de dama, cuenta con productos nacionales provenientes de Arequipa, Trujillo y Lima.

En la actualidad la empresa quiere progresar y evitar, entre otras, un exceso de stock, para ello se deberá aplicar herramientas que ayuden a gestionar de manera eficiente el abastecimiento de productos. Los proveedores y clientes confiarán más en aquellas empresas capaces de anticiparse y de reaccionar ante las fluctuaciones constantes que se dan entre la oferta y la demanda.

Lo que se busca es ser capaz de detectar a tiempo el

comportamiento de compra de sus consumidores. Si en vez de intentar anticiparse y actuar a tiempo frente a imprevistos, se posee un inventario protegido, llevará a la empresa a costes ocultos como cuando los pedidos sean menores que el punto óptimo o la obsolescencia de los productos.

Así mismo, actualmente la información no es analizada, es importante considerar que no solo hay que almacenar los datos de las ventas, sino, que es necesario analizar esa información. Por otro lado, el análisis manual de información representa para las empresas un proceso ineficiente costoso y con riesgos de errores en la interpretación, esta falencia surge al no contar con información que permita tener una visión general de empresa. Los resultados que la empresa obtiene son de bajo rendimiento, ya que, la organización tiene que aceptar y acomodarse a la información disponible.

### Fase 2: Comprender los datos

En esta fase se recopila los datos, se describe, se explora y se verifica la calidad de los datos. Los datos que se emplearan para la construcción del modelo será el registro de venta de la empresa de calzados, así también la herramienta en la que se realizara el modelo predictivo será Orange Datamining.

En primer lugar, se seleccionó de la pestaña Data la herramienta file la cual nos permitirá cargar nuestra base de datos.

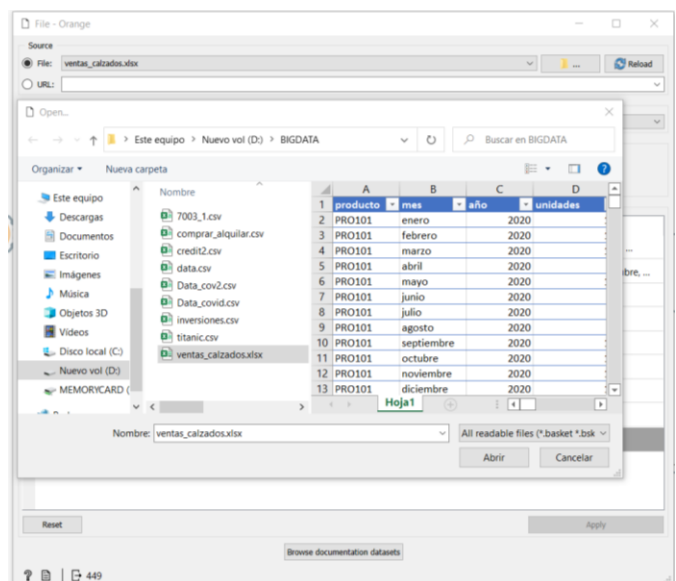


Fig. 4 Cargando data.

En segundo lugar, se seleccionó de la pestaña Data la herramienta data table la cual nos permita visualizar toda la

información en una tabla, la cual consta 9 variables y un total de 449 registros.

Fig. 5 Utilizando herramienta data table.

En tercer lugar, para realizar la exploración de los datos se utilizó de la pestaña Visualice las herramientas distributions y box plot.

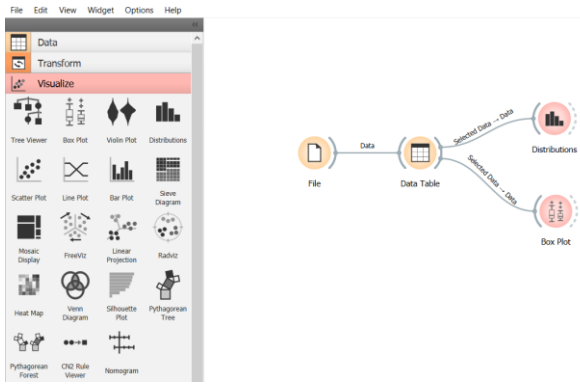


Fig. 6 Utilizando herramientas Distributions y Box plot.

Por último, para verificar la calidad de los datos se emplearon los siguientes gráficos:

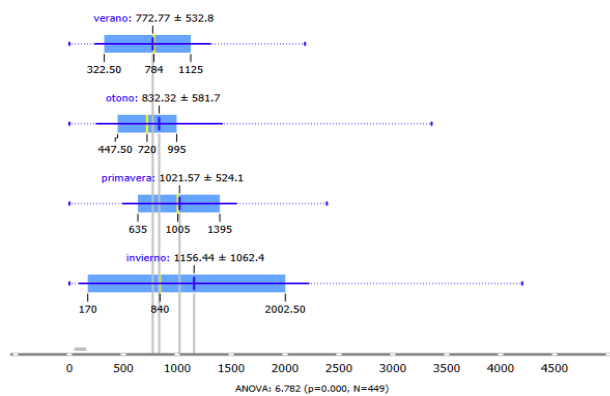


Fig. 7. Gráfico de cajas de monto total de ventas según la estación del año.

De la Fig.7, en el gráfico de caja de la estación de verano en relación con el monto total de las ventas podemos determinar que existe una media de 772.77 y una desviación estándar de 532.8, de la estación de otoño en relación con el monto total de las ventas podemos determinar que existe una media de 832.32 y una desviación estándar de 581.7, de la estación de primavera en relación con el monto total de las ventas podemos determinar que existe una media de 1021.57 y una desviación estándar de 524.1, por último de la estación de invierno en la relación con el monto total de las ventas podemos determinar que existe una media de 1156.44 y una desviación estándar de 1062.4.

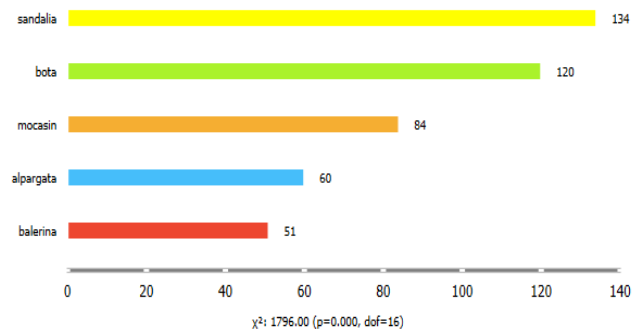


Fig. 8 Gráfico de cajas de categorías de productos.

De la Fig. 8, se puede determinar que de un total de 449 ítems 134 de ellos pertenecen a la categoría sandalia, 120 a la categoría bota, 84 a la categoría mocasin, 60 a la categoría alpargata y los 51 ítems restantes a la categoría balerina.

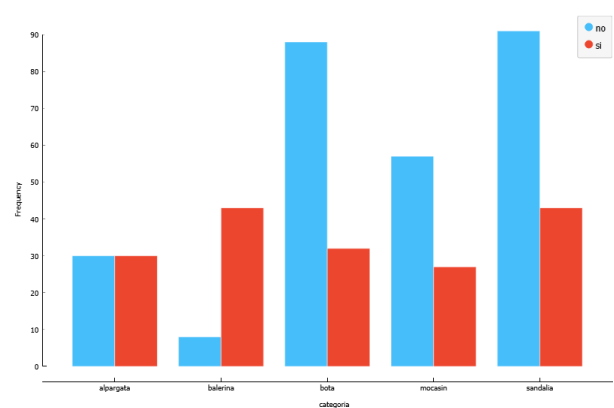


Fig. 9 Gráfico de distribución de categorías según el status.

De la Fig. 9, se puede observar la distribución por categoría en cuanto a los productos más vendidos y menos vendidos.



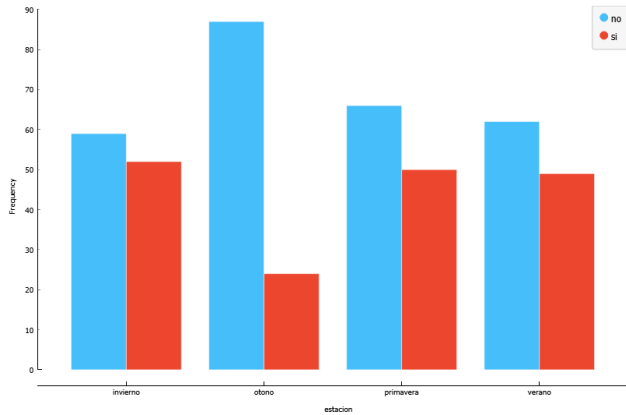


Fig. 10 Gráfico de distribución de las estaciones según el status.

De la Fig 10, se puede observar la distribución por estación en cuanto a los productos más vendidos y menos vendidos.

### Fase 3: Preparación de los datos

En esta fase se realizará la elección y limpieza de datos.

En primer lugar, de la pestaña Transform se utilizó la herramienta Select Columns la cual permitirá determinar las columnas que se trabajaran para la construcción del modelo.

En segundo lugar, seleccionamos las variables según el algoritmo a utilizar; para el algoritmo de árbol de decisión se seleccionaron las variables estación, categoría, estatus (target o variable objetivo) y producto (variable meta). Mientras que para el algoritmo de regresión logística se determinó emplear las 9 variables, las cuales son: producto, mes, año, unidades, precio unitario, total, estación, categoría y estatus (target variable objetivo).

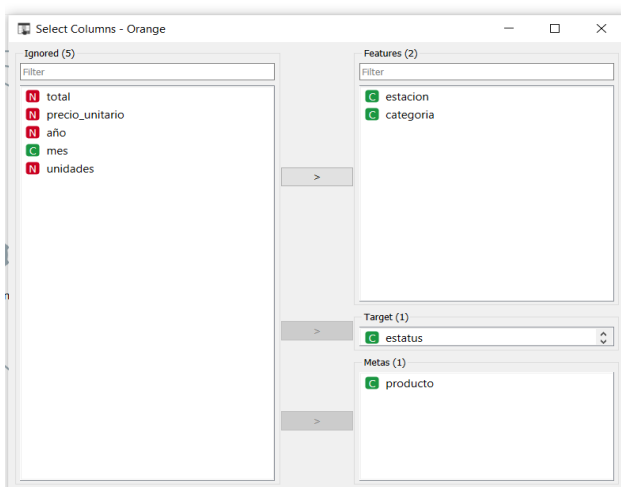


Fig. 11 Realizando selección de variables.

### Fase 4: Modelado

Durante esta fase se realizará la construcción del modelo utilizando los algoritmos previamente seleccionados, los cuales fueron: árbol de decisión y regresión logística. Para ello se utilizaron de la pestaña Model las herramientas Tree y Logistic Regression.

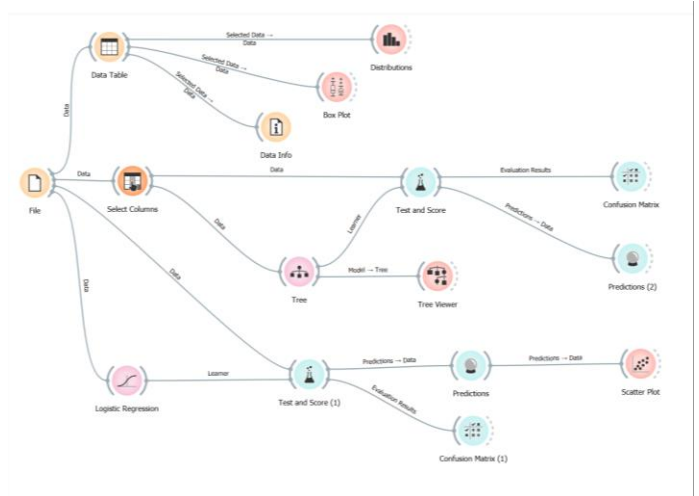


Fig. 12 Construcción del Modelo predictivo.

## IV. RESULTADOS

### A. Resultados obtenidos del algoritmo regresión logística

OE 1: Determinar cuáles son los productos más vendidos

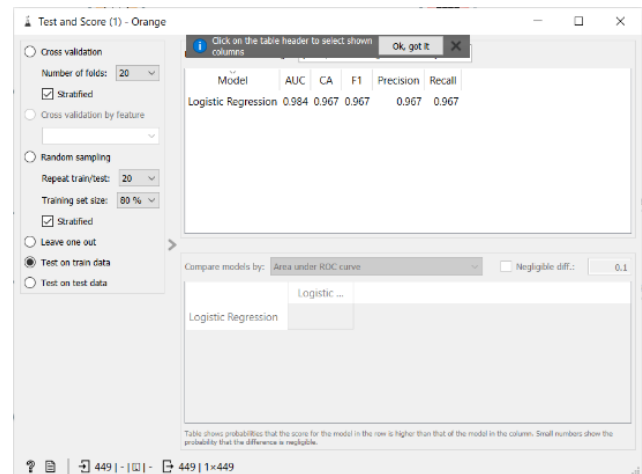


Fig. 13 Test and Score del algoritmo de regresión logística.

De la Fig. 13, se puede determinar que el nivel de precisión al aplicar el algoritmo de regresión logística es del 97% lo cual demuestra un nivel alto para realizar la predicción.



Fig. 14 Matriz de confusión del algoritmo regresión logística.

De la Fig. 14, se puede determinar que los valores de 98.1% y 94.4% corresponden a los valores estimados de forma correcta por el modelo tanto de verdaderos positivos como de verdaderos negativos. Mientras que los valores 1.9% y 5.6% corresponden a los casos en los que el modelo predijo de manera incorrecta.

#### MÉTRICA DE EXACTITUD

$$\text{Exactitud} = \frac{TP + TN}{TP + FN + FP + TN} = 0.97$$

Se obtuvo un nivel de exactitud del 97% lo cual demuestra que el número de elementos clasificados correctamente con relación al número total de elementos es alto.

#### MÉTRICA DE EXHAUSTIVIDAD

$$\text{Exhaustividad} = \frac{TP}{TP + FN} = 0.96$$

Se obtuvo un nivel de exhaustividad del 96% lo cual demuestra que la cantidad de verdaderos positivos que el modelo a clasificado en función del número total de valores positivos ha sido determinado de manera correcta.

#### B. Resultados obtenidos del algoritmo árbol de decisión

OE 2: Determinar cuáles son las categorías más vendidas durante las diferentes estaciones del año.

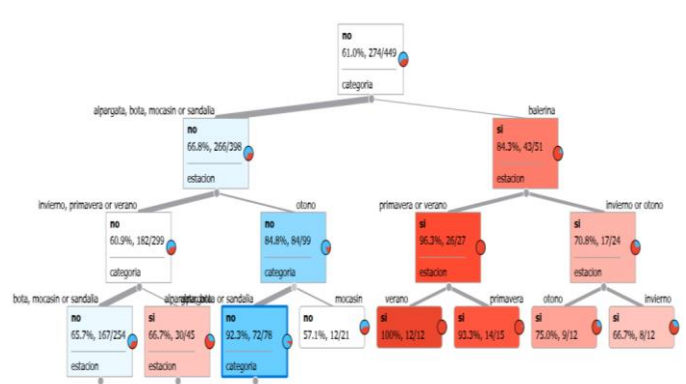


Fig. 15 Gráfico de árbol de decisión.

De la Fig. 15, se puede determinar en primera instancia que la categoría más vendida es balerina con un 84.3%, mientras que las categorías alpargata, bota, mocasin o sandalia tienen un 66.8% de productos vendidos. Así también, de la categoría balerina las estaciones con mayor índice de ventas son primavera o verano con un 96.3% de productos vendidos, mientras que de las categorías alpargata, bota, mocasin o sandalia la estación con menor índice de venta es otoño con un 84.8% de productos vendidos.

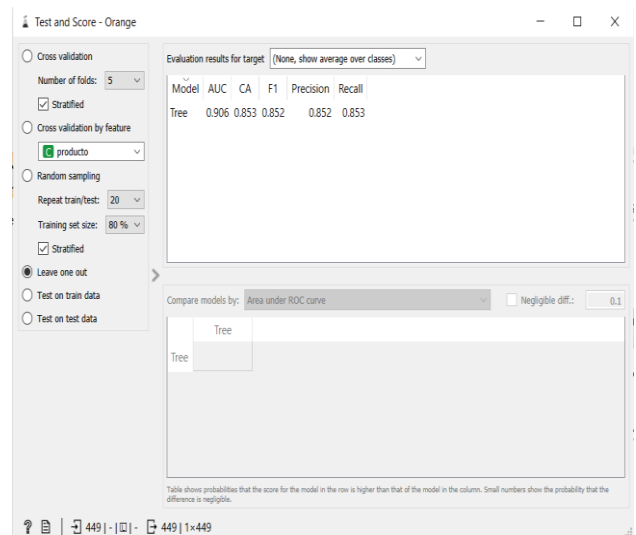


Fig. 16 Test and Score del algoritmo árbol de decisión.

De la Fig. 16, se puede determinar que el nivel de precisión al aplicar el algoritmo de árbol de decisión es del 85% lo cual demuestra un nivel alto para predecir las categorías más vendidas durante las estaciones del año.



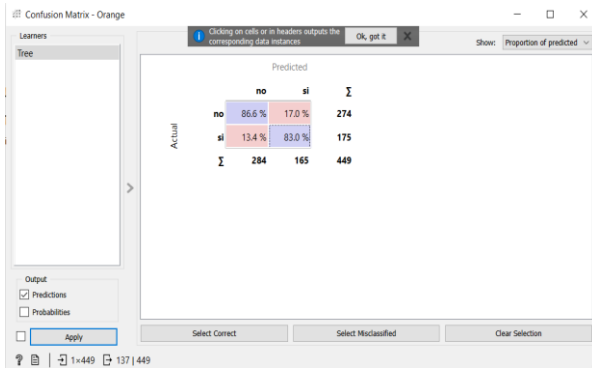


Fig. 17 Matriz de confusión del algoritmo árbol de decisión.

De la Fig. 17, se puede determinar que los valores de 86.6% y 83% corresponden a los valores estimados de forma correcta por el modelo tanto de verdaderos positivos como de verdaderos negativos. Mientras que los valores 17% y 13.4% corresponden a los casos en los que el modelo predijo de manera incorrecta.

#### MÉTRICA DE PRECISIÓN

$$\text{Precisión} = \frac{TP}{TP+FP} = \frac{246}{246+38} = 0.87$$

Se obtuvo un nivel de precisión del 87% lo cual demuestra que su nivel de predicción para determinar cuáles son los productos más vendidos durante las diferentes estaciones del año es alto.

#### V. CONCLUSIONES

Por lo expuesto durante la presente investigación se puede determinar que el desarrollo de un modelo predictivo basado en Machine Learning permite prevenir el sobrestock en la empresa de calzados. Así mismo, se obtuvo el histórico de ventas de la empresa de calzados, dicho archivo está compuesto de nueve variables y un total de 449 registros, cuya información fue vital en la construcción del modelo predictivo.

Se realizó la construcción del modelo predictivo basado en Machine Learning utilizando la herramienta Orange, así mismo, se realizó el análisis de los algoritmos predictivos donde se determinó que los algoritmos a aplicar serán regresión logística y árbol de decisión. De los resultados obtenidos al aplicar el algoritmo de regresión logística para predecir los productos más vendidos se obtuvo una precisión del 97% lo cual demuestra un nivel alto al realizar dicha predicción. Mientras que al aplicar el algoritmo de árbol de decisión para predecir cuales son las categorías más vendidas se obtuvo una precisión del 85% lo cual

demuestra un nivel alto al realizar dicha predicción.

Los objetivos planteados en la presente investigación estuvieron enfocados en la prevención del sobrestock y una vez logrados brindaron una potencial solución al problema que motivo el desarrollo de este proyecto, ya que al contar con la información de los productos con mayor salida durante las diferentes estaciones del año se tomaran decisiones más acertadas en cuanto al abastecimiento de productos evitando así el sobrestock de los mismos.

#### REFERENCIAS

- [1] S. Cheriyan, S. Ibrahim, S. Mohanan, and S. Treesa, "Intelligent Sales Prediction Using Machine Learning Techniques," in 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), Southend, United Kingdom, Aug. 16-17, 2018, IEEE, 2018, <https://doi.org/10.1109/iccecome.2018.8659115>.
- [2] J. Wang, "A hybrid machine learning model for sales prediction", in 2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI), Sanya, China, 4–6 de diciembre de 2020. IEEE, 2020. Disponible: <https://doi.org/10.1109/ichci51889.2020.00083>
- [3] Preet, S. Sharma, D. Deepika and G. Singh, "Intelligent Warehouse Stocking Using Machine Learning", in 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNC), Tumkur, Karnataka, India, 3–4 de diciembre de 2021. IEEE, 2021. Disponible: <https://doi.org/10.1109/icmnc52512.2021.9688530>
- [4] R. M. van Steenberghe and M. R. K. Mes, "Forecasting demand profiles of new products", Decision Support Systems, vol. 139, p. 113401, diciembre de 2020. Disponible: <https://doi.org/10.1016/j.dss.2020.113401>
- [5] M. Al Maruf, A. Azim and S. Mukherjee, "Mushroom Demand Prediction Using Machine Learning Algorithms", in 2020 International Symposium on Networks, Computers and Communications (ISNCC), Montreal, QC, 20–22 de octubre de 2020. IEEE, 2020. Disponible: <https://doi.org/10.1109/isncc49221.2020.9297219>
- [6] J. Hlaváč and J. Štefanovič, "Machine Learning and Business Intelligence or from Descriptive Analytics to Predictive Analytics," in 2020 Cybernetics & Informatics (K&I), 2020, pp. 1-4, doi: 10.1109/KI48306.2020.9039874.
- [7] O. Haffner, E. Kučera and M. Moravčík, "Sales Prediction of Svijany Slovakia, Ltd. Using Microsoft Azure Machine Learning and ARIMA," 2020 Cybernetics & Informatics (K&I), 2020, pp. 1-9, doi: 10.1109/KI48306.2020.9039875.
- [8] M. A. I. Arif, S. I. Sany, F. I. Nahin and A. S. A. Rabby, "Comparison Study: Product Demand Forecasting with Machine Learning for Shop", in 2019 8th International

Conference System Modeling and Advancement in Research Trends (SMART), Moradabad, India, 22–23 de noviembre de 2019. IEEE, 2019. Disponible: <https://doi.org/10.1109/smart46866.2019.9117395>

[9] K. Singh, P. M. Booma and U. Eaganathan, "E-Commerce System for Sale Prediction Using Machine Learning Technique", Journal of Physics: Conference Series, vol. 1712, p. 012042, diciembre de 2020. Disponible: <https://doi.org/10.1088/1742-6596/1712/1/012042>

[10] R. R. Halde, "Application of Machine Learning algorithms for betterment in education system", in 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, India, 9–10 de septiembre de 2016. IEEE, 2016. Disponible: <https://doi.org/10.1109/icacdot.2016.7877759>

[11] A. Shehadeh, O. Alshboul, R. E. Al Mamlook and O. Hamedat, "Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, LightGBM, and XGBoost regression", Automation in Construction, vol. 129, p. 103827, septiembre de 2021. Disponible: <https://doi.org/10.1016/j.autcon.2021.103827>

[12] J. Benalcázar Tamayo, "Análisis comparativo de metodologías de minería de datos y su aplicabilidad a la industria de servicios," Tesis de Postgrado, Universidad de las Américas. Quito 2017.

[13] J. C. Carpio Ticona, "Modelo de Predicción de la Morosidad en el otorgamiento de Crédito Financiero Aplicando Metodología CRISP-DM," Tesis de grado, Universidad Andina Nestor Caceres Velasquez. Juliaca 2016.

[14] J. M. Moine, S. E. Gordillo y A. S. Haedo. "Análisis comparativo de metodologías para la gestión de proyectos de minería de datos". SEDICI - Repositorio de la Universidad Nacional de La Plata. <http://sedici.unlp.edu.ar/handle/10915/18749>

[15] J. Lozada, "Investigación Aplicada Definición, Propiedad Intelectual e Industria", CienciAmérica: Revista de divulgación científica de la Universidad Tecnológica Indoamérica, Vol. 3, n.º 1, pp. 47–50, 2014. Disponible: <https://dialnet.unirioja.es/servlet/articulo?codigo=6163749>

[16] E. Maya, Métodos y técnicas de investigación: Una propuesta ágil para la presentación de trabajos científicos en las áreas de arquitectura, urbanismo y disciplinas afines. Ciudad de Mexico: UNAM, 2014.

[17] R. Hernández, C. Fernández y M. d. P. Baptista, METODOLOGIA DE LA INVESTIGACION, 6a ed. Ciudad de México: McGraw-Hill Interamericana editores, 2014.

[18] R. Hernández-Sampieri y M. Christian, Metodología de la investigación: las rutas: cuantitativa, cualitativa y mixta. Mc Graw Hill education, 20