

---

# Aplicación de minería de datos en la empresa Innovate Consulting para predecir valores de ventas

## *Application of data mining in the company Innovate Consulting to predict sales values*

Diego Geovanny Falconí Punguil<sup>1</sup>; Yomara Elizabeth Tello Oña<sup>2</sup>; Martha Patricia Calvopiña Oña<sup>3</sup>

<sup>1</sup>Universidad Técnica de Cotopaxi, Ecuador

<sup>2</sup>Fenix Corp, Latacunga, Ecuador

<sup>3</sup>Centro de Salud "A" IESS Amaguaña, Ecuador

### Resumen

El presente artículo presenta un estudio sobre la aplicación de minería de datos en *Innovate Consulting* para predecir valores de ventas. El objetivo fue comparar la capacidad predictiva de tres modelos principales: Random Forest, Deep Learning y Naive Bayes. Para ello, se implementó un método que incluyó el procesamiento de datos históricos de ventas, la configuración y entrenamiento de los modelos de Machine Learning, y la evaluación de su precisión mediante una matriz comparativa. Los resultados destacaron que Random Forest mostró la mayor precisión, seguido por Deep Learning y Naive Bayes. Estos hallazgos subrayan la importancia de seleccionar el modelo adecuado según las características específicas del conjunto de datos, proporcionando una base sólida para mejorar la planificación estratégica y la toma de decisiones.

**Palabras clave:** Random Forest, Deep Learning, Naive Bayes, Minería de Datos

Recibido: 20 de mayo de 2024 – revisión aceptada: 29 de julio de 2024

<sup>1</sup>Correspondiente al autor: diego.falconi4@utc.edu.ec

## Abstract

This article presents a study on the application of data mining at Innovate Consulting to predict sales values. The objective was to compare the predictive capability of three main models: Random Forest, Deep Learning, and Naive Bayes. To achieve this, a method was implemented involving the processing of historical sales data, configuration and training of machine learning models, and evaluation of their accuracy using a comparative matrix. Results highlighted that Random Forest exhibited the highest precision, followed by Deep Learning and Naive Bayes. These findings underscore the importance of selecting the appropriate model based on specific dataset characteristics, providing a robust foundation to enhance strategic planning and data-driven decision-making.

**Key words:** Random Forest, Deep Learning, Naive Bayes, Data Mining

---

## Introducción

La capacidad de predecir valores de ventas es crucial para las empresas, ya que permite una mejor planificación y toma de decisiones estratégicas. En la era de la información, la minería de datos se presenta como una herramienta poderosa para analizar grandes volúmenes de datos y extraer patrones útiles para la predicción. Las técnicas de minería de datos, como la regresión lineal, los árboles de decisión y las redes neuronales, han demostrado ser efectivas en la identificación de tendencias y patrones en los datos históricos de ventas (Han et al, 2012). La implementación de estas técnicas permite a las empresas no solo prever el comportamiento futuro del mercado, sino también ajustar sus estrategias comerciales de manera proactiva para maximizar sus ingresos y minimizar riesgos.

Innovate Consulting, una empresa dedicada a proporcionar soluciones innovadoras,

busca optimizar su proceso de predicción de ventas mediante la aplicación de estas técnicas avanzadas de minería de datos. Este estudio se centra en la implementación de diversos algoritmos de minería de datos para analizar datos históricos de ventas de Innovate Consulting, con el objetivo de evaluar la precisión y eficacia de los modelos desarrollados. La capacidad de predecir las ventas con precisión es esencial para la gestión de inventarios, la planificación de la producción y la formulación de estrategias de marketing, aspectos que son fundamentales para el éxito empresarial en un entorno competitivo.

Diversas investigaciones han explorado la aplicación de la minería de datos en la predicción de ventas, destacando su potencial para mejorar la toma de decisiones empresariales. Por ejemplo, un estudio reciente demostró que el uso de modelos predictivos basados en redes neuronales puede mejorar significativamente la precisión de las predicciones de ventas en

comparación con los métodos tradicionales (Chen et al, 2016). Además, las técnicas de minería de datos no solo permiten predecir los volúmenes de ventas, sino también identificar factores influyentes y patrones subyacentes en los datos de ventas. Otro estudio encontró que los árboles de decisión y las técnicas de agrupamiento pueden proporcionar insights valiosos sobre el comportamiento de los clientes y las tendencias de ventas (Murphy, 2016). Estos hallazgos subrayan la importancia de aplicar un enfoque sistemático y basado en datos para la predicción de ventas.

Este artículo tiene como objetivo describir la implementación de técnicas de minería de datos en Innovate Consulting, evaluar la precisión de los modelos predictivos desarrollados y discutir las implicaciones de los resultados para la estrategia comercial de la empresa. A través de este estudio, se espera contribuir al conocimiento existente sobre el uso de la minería de datos para la predicción de ventas y proporcionar recomendaciones prácticas para las empresas que buscan mejorar su proceso de toma de decisiones mediante el análisis de datos. En particular, se analizarán las metodologías empleadas, los desafíos enfrentados durante la implementación y las estrategias adoptadas para superar estos desafíos, con el fin de ofrecer una guía comprensiva para otras organizaciones interesadas en aprovechar la minería de datos para optimizar sus operaciones comerciales.

### **Minería de datos:**

La minería de datos es un proceso interdisciplinario que involucra técnicas de

estadística, aprendizaje automático y bases de datos para extraer información útil y patrones significativos a partir de grandes volúmenes de datos (Han et al, 2012). Este campo ha ganado prominencia debido a la creciente disponibilidad de datos y la necesidad de convertir estos datos en conocimientos accionables. Las técnicas de minería de datos se utilizan en diversas aplicaciones, como análisis de mercado, detección de fraudes, y, especialmente, en la predicción de ventas (Chen et al, 2016). Entre los algoritmos más utilizados están:

- A. Regresión Lineal
- B. Árboles de Decisión
- C. Redes Neuronales
- D. Algoritmos de Agrupamiento
- E. Algoritmos de Ensemble

### **Regresión Lineal:**

La regresión lineal es una técnica estadística que modela la relación entre una variable dependiente y una o más variables independientes. En el contexto de la predicción de ventas, se utiliza para predecir valores continuos basados en datos históricos. Su simplicidad y facilidad de interpretación la hacen una opción popular, aunque puede no capturar relaciones no lineales complejas presentes en los datos (Murphy, 2016).

### **Árboles de Decisión:**

Los árboles de decisión segmentan los datos en subgrupos homogéneos mediante una serie de decisiones binarias basadas en los valores de las variables predictoras. Esta

técnica es especialmente útil para identificar variables clave que influyen en las ventas y para construir modelos fácilmente interpretables. Sin embargo, pueden ser propensos al sobreajuste si no se controlan adecuadamente (Goodfellow et al, 2016). Los árboles de decisión se han utilizado con éxito en la segmentación de clientes y en la identificación de factores críticos que afectan las ventas (James et al, 2021).

### **Redes Neuronales:**

Inspiradas en el funcionamiento del cerebro humano, las redes neuronales son capaces de capturar relaciones no lineales y complejas en los datos. Estas técnicas son particularmente efectivas cuando los patrones en los datos son difíciles de identificar mediante métodos tradicionales. Las redes neuronales, especialmente las de tipo profundo (deep learning), han demostrado un rendimiento superior en tareas de predicción debido a su capacidad para aprender representaciones jerárquicas de los datos (Rokach & Maimon, 2014). Las redes neuronales profundas pueden manejar grandes volúmenes de datos y detectar patrones sutiles que otras técnicas podrían pasar por alto. Además, el desarrollo de técnicas avanzadas como las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN) ha ampliado las aplicaciones de deep learning a campos como la visión por computadora y el procesamiento del lenguaje natural (LeCun et al, 2015).

### **Algoritmos de Agrupamiento:**

Los algoritmos de agrupamiento, como el k-means, se utilizan para identificar grupos naturales dentro de los datos sin necesidad de

variables dependientes. Estos algoritmos son útiles para segmentar mercados y entender diferentes comportamientos de los clientes, lo que puede informar estrategias de ventas y marketing. La agrupación puede revelar segmentos de clientes con características similares que responden de manera similar a ciertas estrategias de marketing o productos (Jain, 2020).

### **Algoritmos de Ensemble:**

Los métodos de ensemble, como Random Forest y XGBoost, combinan múltiples modelos para mejorar la precisión y robustez de las predicciones. XGBoost, en particular, ha demostrado ser extremadamente eficaz debido a su capacidad para manejar grandes volúmenes de datos y capturar relaciones complejas. Este algoritmo utiliza un enfoque de boosting en árboles de decisión, lo que mejora la precisión y reduce el riesgo de sobreajuste (Raschka, 2017).

### **Deep Learning:**

El deep learning es una subcategoría del aprendizaje automático que utiliza redes neuronales profundas para modelar y entender patrones complejos en los datos. Estas redes están compuestas por múltiples capas de neuronas artificiales que permiten el aprendizaje jerárquico de características a diferentes niveles de abstracción (Al Jarrah, 2016). En el contexto de la predicción de ventas, el deep learning ha demostrado ser extremadamente eficaz debido a su capacidad para manejar grandes volúmenes de datos y aprender representaciones complejas que capturan tendencias y patrones sutiles en los datos históricos de ventas.

## Aplicaciones en la Predicción de Ventas:

La predicción de ventas es una aplicación crucial de la minería de datos que ayuda a las empresas a anticipar la demanda de sus productos y servicios, optimizando así la gestión de inventarios, la planificación de la producción y las estrategias de marketing. Estudios recientes han demostrado la efectividad de diversos algoritmos de minería de datos en la mejora de las predicciones de ventas.

Un estudio demostró que el uso de modelos predictivos basados en redes neuronales puede mejorar significativamente la precisión de las predicciones de ventas en comparación con los métodos tradicionales. Las redes neuronales fueron capaces de capturar patrones complejos en los datos históricos de ventas que otros métodos no lograron identificar (Chen et al, 2016). Las redes neuronales profundas, en particular, han revolucionado el campo de la predicción de ventas debido a su capacidad para manejar grandes volúmenes de datos y aprender representaciones jerárquicas complejas. Estas redes son especialmente útiles cuando los datos de ventas incluyen múltiples características y patrones no lineales que son difíciles de modelar con técnicas tradicionales (Al Jarrah, 2016).

Otro estudio encontró que los árboles de decisión y las técnicas de agrupamiento pueden proporcionar insights valiosos sobre el comportamiento de los clientes y las tendencias de ventas. Los árboles de decisión, en particular, permitieron a los investigadores identificar variables clave que influyen en las ventas, facilitando la interpretación y la toma de decisiones estratégicas (Murphy, 2016).

## Materiales y Métodos

La metodología empleada en este estudio se estructura en varias etapas cruciales para aplicar técnicas de minería de datos en la predicción de valores de ventas en Innovate Consulting. En primera instancia se lleva a cabo una exhaustiva recopilación de datos históricos de ventas de los clientes de la empresa, así como datos económicos y de mercado relevantes. Esta fase fue fundamental para asegurar la disponibilidad de datos completos y representativos que servirían de base para el análisis predictivo (Arthur, 2019).

Posteriormente, se realizará un riguroso proceso de preprocesamiento de datos, que incluye la limpieza para eliminar valores atípicos y datos faltantes, así como la normalización de variables numéricas y la codificación de variables categóricas. Estas acciones son esenciales para garantizar la calidad y consistencia de los datos utilizados en los modelos predictivos desarrollados (Jain, 2020).

Para comprender mejor la estructura de los datos y las relaciones entre las variables, se llevará a cabo un análisis exploratorio detallado. Mediante visualizaciones y estadísticas descriptivas, se identifican patrones significativos que guían la selección y transformación de características relevantes para la predicción de ventas (Murphy, 2016).

Las características más influyentes fueron seleccionadas utilizando métodos estadísticos y de aprendizaje automático, como el análisis de importancia de características y técnicas de reducción de dimensionalidad. Además, se aplicaron transformaciones adicionales para

mejorar la capacidad predictiva de los modelos construidos (Raschka, 2017). Se exploraron varios modelos de aprendizaje automático, tales como regresión lineal, árboles de decisión y redes neuronales, los cuales son ajustados y validados utilizando técnicas de validación cruzada. Esto permite seleccionar el modelo más adecuado para predecir con precisión los valores de ventas futuros en Innovate Consulting (Arthur, 2019).

El modelo seleccionado se implementó en un entorno de producción, integrando flujos de trabajo automatizados para asegurar la actualización continua de las predicciones de ventas. Se establecieron también procedimientos de monitoreo y retroalimentación para validar y mejorar iterativamente la precisión del modelo (Arthur, 2019).

Después de obtener y realizar la preparación de los datos, seleccionamos la técnica de minería de datos más apropiada para predecir las ventas en la empresa Innovate Consulting. Para llevar a cabo un correcto análisis se incluirá las siguientes técnicas:

- A. Árboles de decisión: que pueden generar modelos predictivos a partir de las variables de ventas, productos, clientes, etc.
- B. Deep Learning: capaces de aprender relaciones complejas en los datos de ventas y hacer predicciones precisas.
- C. Naive Bayes: es un clasificador probabilístico fundamentado en el teorema de Bayes y algunas hipótesis simplificadoras adicionales.

Bajo la misma línea, se consideró que es importante experimentar y cotejar el desempeño de los diferentes modelos para elegir aquel que minimice el error de predicción. Posteriormente de entrenar los diferentes modelos para estas predicciones, es trascendental valorar su desempeño y optar por el más apropiado, se debe calcular métricas de error para permitir la comparación de la precisión de las predicciones con los valores de ventas reales.

### **Materiales Esenciales**

Se usaron un conjunto de datos estructurado y de alta calidad que contenía información histórica sobre ventas, precios, promociones, condiciones económicas, datos demográficos de clientes y otras variables relevantes de la empresa Innovate Consulting, extraída directamente desde su base de datos de ventas.

Con dichos datos se realizará un análisis de datos, desarrollando un algoritmo en Python que pueda ser capaz de mostrar datos en base a las métricas anteriormente mencionadas. Además, se utilizó el software RapidMiner para realizar el proceso de minería de datos y entrenamiento de modelos en base a los diferentes algoritmos para evaluar cual de es el que mayor porcentaje de precisión se obtiene.

### **Método de Recolección de Datos**

El dataset fue recopilado de los registros históricos de Innovate Consulting y almacenado en un archivo Excel. Los campos del dataset incluyen:

- A. N° (Número secuencial)
- B. TipoSolicitud (Tipo de solicitud)

- C. Estado (Estado de la solicitud)
- D. Contenedor (Identificación del contenedor)
- E. Validez (Validez de la solicitud)
- F. ConRuc (Indicación de RUC),
- G. Nombres (Nombres del solicitante)
- H. Apellido1 (Primer apellido del solicitante)
- I. Apellido2 (Segundo apellido del solicitante)

### Preparación de Datos

La preparación de datos es crucial para garantizar la calidad y la usabilidad del dataset. Los siguientes pasos se realizaron en RapidMiner:

- A. Carga de Datos: El archivo Excel se cargó en RapidMiner utilizando el operador "Read Excel".
- B. Limpieza de Datos: Se eliminaron los registros duplicados y se manejaron los valores faltantes. Los campos no relevantes para la predicción se eliminaron.
- C. Transformación de Datos: Se normalizaron los datos y se codificaron

las variables categóricas utilizando el operador "Nominal to Numerical".

### Selección y Entrenamiento del Modelo

Se prueban varios algoritmos de minería de datos para identificar el modelo más adecuado. Los algoritmos seleccionados incluyen:

- A. Árboles de decisión
- B. Deep Learning
- C. Naive Bayes

### Implementación y Monitoreo

El modelo seleccionado se implementará para predecir los valores de ventas futuros. Se establece un sistema de monitoreo continuo para actualizar y refinar el modelo con nuevos datos. RapidMiner proporciona herramientas como "Apply Model" y "Performance Monitoring" para facilitar esta etapa.

### Resultados

Se ejecutó en un análisis en base al dataset proporcionado por la empresa Innovate Consulting el cual consta de 166 registros, en la figura 1 se observan parte de los datos en formato CSV:

Nº	TipoSolicitud	Estado	Contenedor	Validez	ConRuc	Nombres	Apellido1	Apellido2	Renovacion
1	PERSONA NATURAL	APROBADO	ARCHIVO	1	NO	KATERIN GISSELA	ZAPATA	CHASILUISA	SI
2	PERSONA NATURAL	APROBADO	ARCHIVO	1	NO	EVELYN MARITZA	CHANGO	HOLGUIN	SI
3	PERSONA NATURAL	APROBADO	ARCHIVO	1	SI	JESSICA CRISTINA	TITUAÑA	GALLARDO	SI
4	PERSONA NATURAL	APROBADO	ARCHIVO	2	SI	MARIA ELVIA	ALMACHI	GUAYTA	SI
5	PERSONA NATURAL	APROBADO	ARCHIVO	1	SI	MAGDALENA MERCEDES	ROMERO	CELA	SI
6	PERSONA NATURAL	APROBADO	ARCHIVO	5	SI	NOEMI	BONILLA	SANCHEZ	SI
7	PERSONA NATURAL	APROBADO	ARCHIVO	2	SI	AMPARITO ELIZABETH	BASTIDAS	BRAVO	SI
8	PERSONA NATURAL	APROBADO	ARCHIVO	5	SI	JULIO ROMULO	CURAY	TAIPE	SI
9	PERSONA NATURAL	APROBADO	ARCHIVO	1	SI	EDGAR RODRIGO	HERRERA	ZAPATA	SI
10	PERSONA NATURAL	APROBADO	ARCHIVO	1	SI	HECTOR FABIAN	HERRERA	ZAPATA	SI
11	PERSONA NATURAL	APROBADO	ARCHIVO	1	NO	OSMAR GUIDO	NANTIPIA	MITIAP	SI

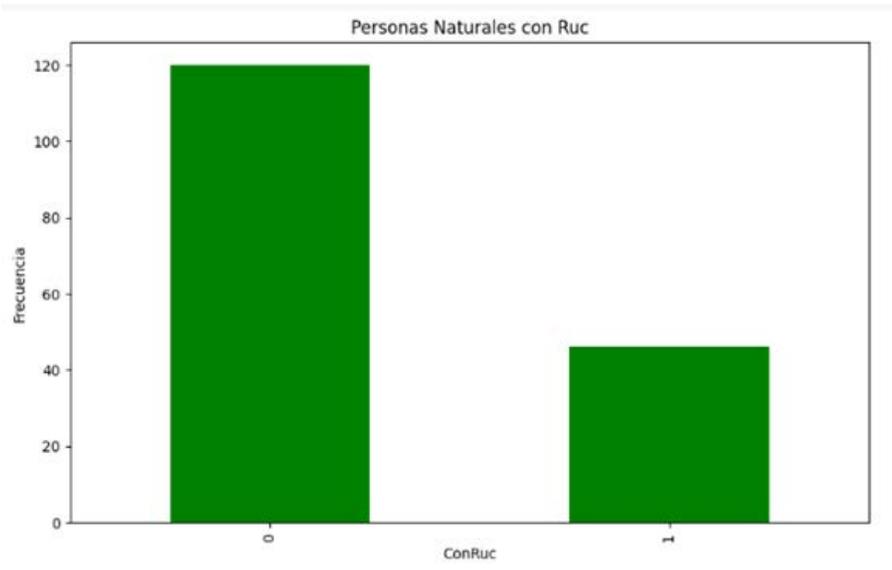
Figura 1. Dataset Innovate Consulting

Una vez recolectados los datos realizamos un algoritmo en Python como se muestra en la figura 2, para representar los datos recolectados indicando si la persona natural tiene o no ruc

como se lo puede apreciar en la figura 3. En la figura 3 se puede interpretar que en el dataset tiene un total de 120 personas que no tienen RUC y 46 que si tienen Ruc

```
import pandas as pd
import matplotlib.pyplot as plt
archivo_excel = '/content/MATRIZ_CLIENTES_FE_colab.xlsx'
df = pd.read_excel(archivo_excel)
print(df.head())
conruc_counts = df['ConRuc'].value_counts()
plt.figure(figsize=(10, 6))
colors = 'green'
conruc_counts.plot(kind='bar', color=colors)
plt.xlabel('ConRuc')
plt.ylabel('Frecuencia')
plt.title('Personas Naturales con Ruc')
plt.show()
```

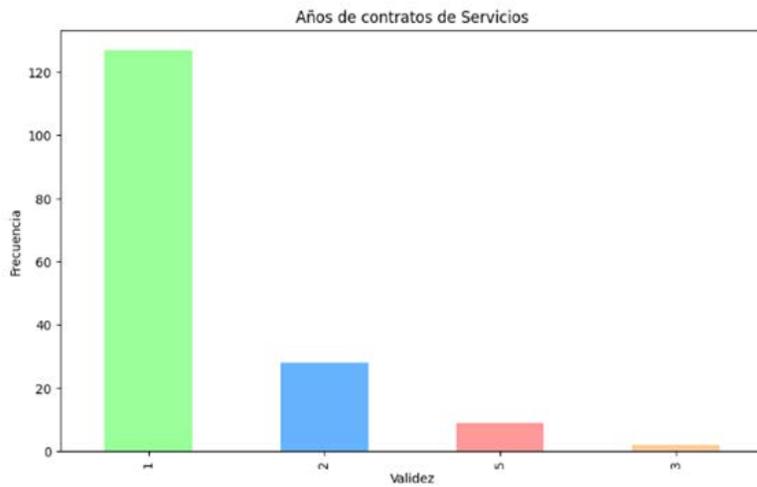
**Figura 2.** Algoritmo de gráfico de personas con ruc y sin ruc



**Figura 3.** Gráfico de personas con Ruc y sin Ruc

Posteriormente con el algoritmo de la figura 4 logramos representar un gráfico de barras de los años que se contrató los servicios a la empresa

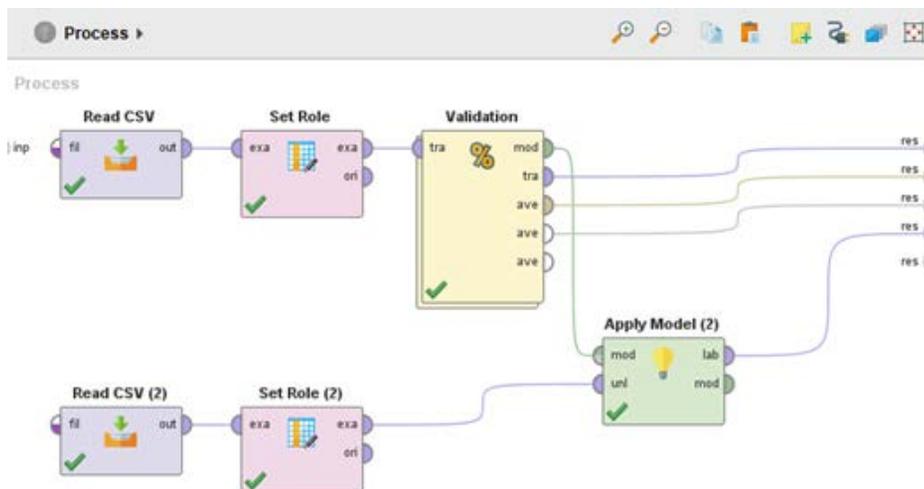
Innovate Consulting, en el mismo podemos concluir que alrededor de 120 personas naturales contratan el servicio para 1 año:



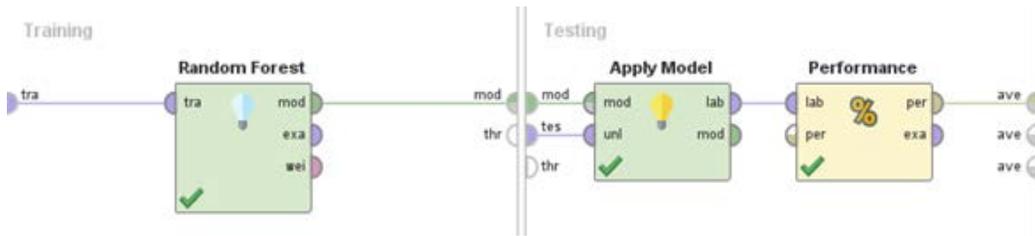
**Figura 4.** Gráfico años de servicio contratado.

En la figura 5 y figura 6 se observa el diagrama que se aplicó en la herramienta RapidMiner para obtener la predicción de renovación de servicios. La función “Read Csv” es se utiliza para realizar la carga del dataset.csv que se va a analizar, la función “Set Role” se utiliza para

setear el parámetro del cual se va a realizar la predicción, la función “Validation” es donde se configura el modelo que se va aplicar en este caso es “Random Forest”, las funciones “Apply Model” y “Performance” se utiliza para aplicar el modelo y ejecutarlo.



**Figura 5.** Diagrama RapidMiner



**Figura 6.** Aplicación de modelo Random Forest para obtener predicción

En la figura 7 se muestra que aplicando Random Forest se obtiene una precisión del 76% del modelo, lo que es algo favorable ya

que se estima que la mayoría de clientes van a volver a contratar servicios de la empresa Innovate Consulting.

accuracy: 76.00%

	true SI	true NO	class precision
pred. SI	27	8	77.14%
pred. NO	4	11	73.33%
class recall	87.10%	57.89%	

**Figura 7.** Precisión del Modelo Random Forest.

La figura 8 indica la predicción de renovación de servicios en la empresa de estudio, acorde

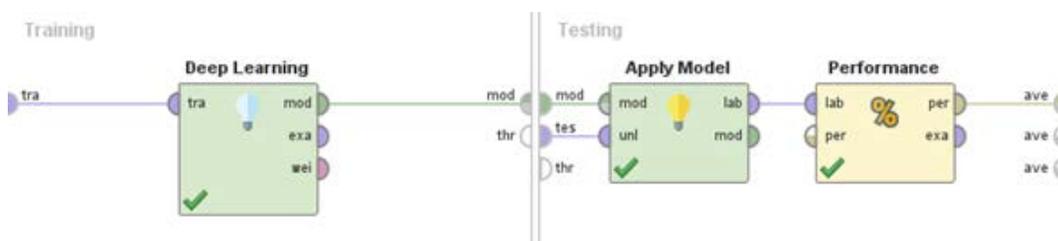
al Modelo Random Forest aplicado.

Row No.	Renovacion	prediction(R...	confidence(...	confidence(...	N°	TipoSolicitud	Estado	Contenedor	Valid
38	NO	NO	0.194	0.806	38	PERSONA N...	APROBADO	ARCHIVO	1
39	NO	NO	0.189	0.811	39	PERSONA N...	APROBADO	ARCHIVO	1
40	NO	NO	0.205	0.795	40	PERSONA N...	APROBADO	ARCHIVO	1
41	SI	SI	0.745	0.255	41	PERSONA N...	APROBADO	ARCHIVO	1
42	SI	SI	0.709	0.291	42	PERSONA N...	APROBADO	ARCHIVO	5
43	NO	NO	0.497	0.503	43	PERSONA N...	APROBADO	ARCHIVO	5
44	NO	NO	0.244	0.756	44	PERSONA N...	APROBADO	ARCHIVO	1
45	NO	NO	0.244	0.756	45	PERSONA N...	APROBADO	ARCHIVO	1
46	NO	NO	0.244	0.756	46	PERSONA N...	APROBADO	ARCHIVO	1
47	SI	NO	0.493	0.507	47	PERSONA N...	APROBADO	ARCHIVO	1
48	NO	NO	0.406	0.594	48	PERSONA N...	APROBADO	ARCHIVO	1
49	SI	SI	0.513	0.487	49	PERSONA N...	APROBADO	ARCHIVO	1
50	NO	NO	0.482	0.518	50	PERSONA N...	APROBADO	ARCHIVO	1
51	NO	NO	0.479	0.521	51	PERSONA N...	APROBADO	ARCHIVO	1

**Figura 8.** Predicción de renovación de servicios.

Se empleó la misma lógica usando el algoritmo de Deep Learning para verificar el porcentaje de precisión del mismo, en la figura 9 se muestra el

diagrama empleado en Rapid Miner y en la figura 10 se muestra que aplicando el algoritmo de Deep Learning se obtiene una precisión del 68%:



**Figura 9.** Aplicación de modelo Deep Learning para obtener predicción.

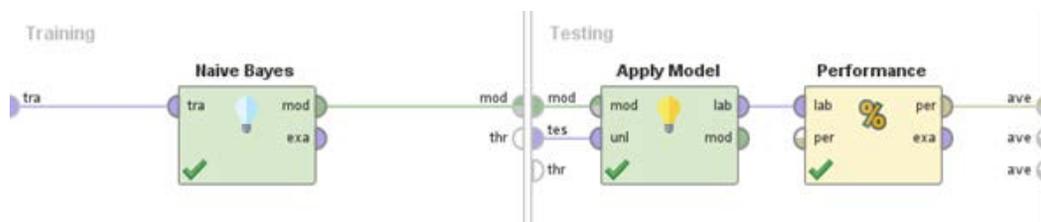
accuracy: 68.00%

	true SI	true NO	class precision
pred. SI	25	10	71.43%
pred. NO	6	9	60.00%
class recall	80.65%	47.37%	

**Figura 10.** Precisión del Modelo Deep Learning.

Finalmente, se realizó un último análisis usando el algoritmo de Naive Bayes para verificar el porcentaje de precisión del mismo, en la figura 9 se muestra el diagrama empleado

en Rapid Miner y en la figura 10 se muestra que aplicando el algoritmo de Naive Bayes se obtiene una precisión del 62%:



**Figura 11.** Aplicación de modelo Naive Bayes para obtener predicción.

accuracy: 62.00%

	true SI	true NO	class precision
pred. SI	17	5	77.27%
pred. NO	14	14	50.00%
class recall	54.84%	73.68%	

**Figura 12.** Precisión del Modelo Naive Bayes.

Luego de que se evaluaron los tres algoritmos diferentes en términos de su precisión en la predicción de valores de ventas para la empresa Innovate Consulting. Los algoritmos

considerados fueron Random Forest, Deep Learning y Naive Bayes. A continuación, se presenta un análisis detallado de los resultados obtenidos, como se muestra en la Tabla 1.

**Tabla 1.** Resultados de precisión de los algoritmos

ALGORITMO	PRECISIÓN
Random Forest	76%
Deep Learning	68%
Naive Bayes	62%

El algoritmo Random Forest logró una precisión del 76%. Este resultado indica que el modelo basado en Random Forest es capaz de predecir los valores de ventas con una alta exactitud en comparación con los otros algoritmos evaluados. La alta precisión obtenida sugiere que este algoritmo es eficaz para capturar las relaciones complejas y no lineales en los datos de ventas, proporcionando predicciones confiables.

El algoritmo de Deep Learning obtuvo una precisión del 68%. Aunque este valor es inferior al alcanzado por el Random Forest, sigue siendo considerablemente alto y destaca la capacidad de las redes neuronales profundas para manejar grandes volúmenes de datos y aprender representaciones complejas. Las redes neuronales profundas son especialmente efectivas cuando los datos contienen múltiples características y patrones no lineales, lo que sugiere que, con ajustes adicionales en la arquitectura del modelo o el uso de técnicas de optimización, la precisión de Deep Learning podría mejorarse aún más. Además, las

capacidades de generalización de los modelos de Deep Learning son una ventaja significativa para la predicción de ventas a largo plazo.

El algoritmo Naive Bayes logró una precisión del 62%, siendo el algoritmo con el rendimiento más bajo entre los tres evaluados. La menor precisión de Naive Bayes en este caso sugiere que las relaciones entre las variables predictoras en los datos de ventas no son independientes, y, por lo tanto, este algoritmo no es el más adecuado para este tipo de predicciones.

## Conclusiones

La evaluación de tres algoritmos diferentes para la predicción de valores de ventas en *Innovate Consulting* reveló variaciones significativas en cuanto a su precisión. Según los resultados de la tabla de precisiones obtenidas, se observa que Random Forest lidera con una precisión del 76%, seguido por Deep Learning con un 68%, y Naive Bayes con un 62%. Estas diferencias destacan la importancia de seleccionar el algoritmo más adecuado según

las características específicas de los datos y los objetivos del análisis.

Random Forest mostró consistentemente la precisión más alta entre los tres algoritmos evaluados. Su capacidad para gestionar conjuntos de datos complejos y capturar relaciones no lineales lo posiciona como una opción robusta para la predicción de ventas. Por otro lado, aunque Deep Learning ofrece una precisión respetable, su rendimiento ligeramente inferior respecto a Random Forest sugiere que podría requerir ajustes adicionales en la configuración del modelo o una mayor cantidad de datos para competir efectivamente en términos de precisión. Por último, Naive Bayes, aunque el menos preciso en este contexto, sigue siendo una opción viable en escenarios donde la simplicidad y la velocidad de implementación son prioritarias sobre la precisión absoluta. La baja precisión observada puede atribuirse a la suposición simplificada de independencia entre características, que no siempre se cumple en conjuntos de datos complejos como los de ventas.

## Literatura citada

- Al Jarrah, R. (2016). Efficient Machine Learning for *Big Data: A Review*. *Big Data Research*, 87-93.
- Arthur, A. (2019). Data collection techniques and quality: Essentials for data-driven decision making. *IEEE Transactions on Data and Knowledge Engineering*, 442-455.
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Special Interest Group on Knowledge Discovery*

and *Data Mining (SIGKDD)*, 785-794.

- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. Waltham, 3.
- Jain, A. (2020). Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 651-666.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning with Applications in R*. Springer.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 436-444.
- Murphy, K. (2016). *Machine Learning: A Probabilistic Perspective*. Cambridge.
- Raschka, S. (2017). *Python Machine Learning*. Packt Publishing.
- Rokach, L., & Maimon, O. (2014). *Data Mining with Decision Trees: Theory and Applications*. Singapore: World Scientific.