Modelo de clasificación de clientes para la UEN - Previsión de una empresa funeraria basado en algoritmos de ML

Seminario de Analítica y Ciencia de Datos

# Resumen Descriptivo del Proyecto

Este proyecto presenta un modelo predictivo en el sector funerario, centrado en el perfilamiento de los clientes para identificar sus preferencias. A través de técnicas de asociación y clasificación, se facilitarán recomendaciones de productos que promuevan la venta cruzada y la fidelización de los clientes en las estrategias de marketing.

Marco Teórico

## Revisión de Literatura.

Hoy en día, la analítica de datos, se ha consolidado como un campo interdisciplinario que integra métodos estadísticos, informáticos y matemáticos para extraer conocimiento útil de grandes volúmenes de datos[1]. Así mismo, se ha convertido en una herramienta que apoya la toma de decisiones informada y que genera confianza para las organizaciones.

Considerando la descripción del proyecto y los fundamentos correspondientes desde la ciencia de datos, se plantea a continuación la consolidación de las diferentes teorías y/o metodologías que permiten abordar desde el campo de estudio la problemática planteada con respecto a la clasificación y asociación de los clientes. De este modo se intenta describir lo que se ha trabajado sobre el tema, además de aportes para el desarrollo de conocimiento en la materia.

Con el propósito de tener claridad y construir un esquema conceptual se describen a continuación las diferentes etapas que enmarcan el desarrollo de un proyecto de clasificación en conjunto con asociación :

1. **Fundamentos de la clasificación en la analítica de datos**

La **clasificación** es una técnica supervisada donde el objetivo es asignar una etiqueta o clase a un nuevo conjunto de datos, basado en características aprendidas durante el entrenamiento, basadas en observaciones previas. Es uno de los tipos más comunes de problemas abordados en machine learning y se utiliza en aplicaciones como el reconocimiento de imágenes, la detección de fraudes, tipificación de clientes, diagnósticos médicos, entre otros [2].

La clasificación puede ser:

* **Binaria**, cuando solo hay dos clases (por ejemplo, apto/no apto).
* **Multiclase**, cuando hay más de dos clases (por ejemplo, clasificación de tipos de clientes: AAA, AA, B, C).
* **Multietiqueta**, cuando a una instancia se le pueden asignar múltiples etiquetas.

Así mismo, es de interés para el desarrollo del proyecto, una vez se realice la clasificación, para efectos de la generación de acciones de mercadeo, realizar una asociación de los clientes ya clasificados. De ahí que también se debe describir el conocimiento acerca de las características de las reglas de asociación.

1. **Fundamentos de la minería de asociaciones**

Dentro de la analítica de datos, **la minería de asociaciones** o descubrimiento de reglas de asociación se ha establecido como una técnica clave para identificar relaciones ocultas entre variables de un conjunto de datos [1]. Esta actividad tiene como objetivo descubrir dependencias significativas, patrones frecuentes o correlaciones entre atributos, lo cual puede ser útil para mejorar estrategias de negocio, optimizar procesos y personalizar servicios.

La minería de asociaciones se popularizó con el desarrollo del algoritmo **Apriori**, propuesto por Agrawal et al. en 1993. Este algoritmo busca identificar **reglas de asociación**, que son relaciones del tipo "si-entonces" entre variables categóricas.

Dos medidas comunes para evaluar la calidad de una regla de asociación son, Soporte y Confianza[2].

Si bien el algoritmo Apriori ha sido ampliamente utilizado, tiene limitaciones en cuanto a la escalabilidad, especialmente con grandes bases de datos, lo que llevó al desarrollo de algoritmos más eficientes como FP-Growth (Frequent Pattern Growth) [3].

1. **Aplicación de metodologías de Machine Learning**

El machine learning, se enfoca en el desarrollo de algoritmos que pueden aprender patrones a partir de datos y realizar predicciones o decisiones sin estar explícitamente programados para cada tarea en procesos de clasificación [5]. Y así mismo para procesos de asociación, los algoritmos son capaces de identificar relaciones no lineales y patrones más complejos, permitiendo además la combinación con técnicas de reducción de dimensionalidad, detección de anomalías y análisis predictivos.

Los métodos de clasificación y asociación en machine learning pueden dividirse en varias categorías. Cabe destacar si bien algunos de los algoritmos pueden ser aplicados en ambas metodologías o inclusive aplicar metodologías para el clustering, los resultados y/o conclusiones dependen específicamente del caso de estudio o problemática:

* 1. **Algoritmos de clasificación basados en modelos estadísticos**
     + Regresión Logística: Es uno de los métodos más simples para clasificación binaria. Modela la probabilidad de una clase como una función logística de las características de entrada [4, 6].
     + Análisis Discriminante Lineal (LDA): Utiliza una combinación lineal de características para separar las clases de manera óptima, asumiendo distribuciones gaussianas [7]**.**
  2. **Métodos de clasificación y asociación basados en Árboles de Decisión**
     + Árboles de Decisión: Construyen un árbol de decisiones donde cada nodo representa una característica y cada rama representa un resultado de la característica [8].
     + Random Forest: Un ensamble de árboles de decisión que reduce el sobreajuste y mejora la precisión combinando múltiples árboles en un solo modelo [9].
     + Gradient Boosting Machines (GBM): Mejoran el rendimiento ajustando iterativamente árboles de decisión que corrigen errores de los modelos anteriores [10].
  3. **Métodos de clasificación y asociación Basados en Redes Neuronales**
     + Redes Neuronales Artificiales (ANN): Son modelos inspirados en el cerebro humano que pueden aprender patrones complejos. Utilizan capas de neuronas conectadas entre sí para clasificar datos [11].
     + Redes Neuronales Profundas (Deep Learning): Son ANN con muchas capas ocultas, capaces de aprender representaciones abstractas de alto nivel en datos complejos como imágenes o texto [12].
     + Las **redes neuronales profundas (deep learning)** han comenzado a aplicarse a problemas de asociación complejos, especialmente en casos donde las interacciones entre ítems no son evidentes con los métodos tradicionales. Modelos como las **autoencoders** se utilizan para aprender representaciones latentes de los datos, lo que facilita el descubrimiento de asociaciones entre conjuntos de datos de alta dimensionalidad [13].
  4. **Máquinas de Soporte Vectorial (SVM) para modelos de clasificación**

Las SVM buscan un hiperplano que divida los datos en clases de manera que se maximice la distancia entre las clases. Son especialmente útiles cuando los datos no son linealmente separables, ya que pueden utilizar funciones núcleo (kernels) para proyectar los datos en un espacio de mayor dimensionalidad [14].

* 1. **Vecinos Más Cercanos (KNN) para modelos de clasificación**

Un enfoque simple basado en la idea de que las instancias similares deben tener etiquetas similares. Clasifica un nuevo dato en función de las clases de sus 𝑘 vecinos más cercanos en el espacio de características [15].

* 1. **Métodos basados en Agrupamiento**

Otra técnica utilizada para descubrir asociaciones es el clustering o agrupamiento. Algoritmos como K-Means y DBSCAN agrupan instancias similares, lo que permite identificar asociaciones entre los atributos que definen dichos grupos [16].

1. **Preprocesamiento de los datos** 
   1. **Normalización y estandarización**: Los algoritmos pueden verse afectados por escalas de características no homogéneas, por lo que es necesario transformar los datos para que todas las características tengan una escala comparable [16].
   2. **Imputación de valores faltantes**: Los datos incompletos pueden afectar el rendimiento de los modelos, por lo que es importante manejar los valores faltantes mediante técnicas de imputación o eliminación [17].
   3. **Selección de características**: No todas las características disponibles en un conjunto de datos son igualmente relevantes. Métodos como la selección basada en importancia de características o el análisis de componentes principales (PCA) pueden ayudar a reducir la dimensionalidad y mejorar el rendimiento del modelo [19].
2. **Evaluación del rendimiento de los modelos de clasificación**

La evaluación del rendimiento de un modelo de clasificación es crucial para entender su capacidad de generalización. Algunas de las métricas más comunes para la evaluación son:

* 1. Precisión: Proporción de predicciones correctas sobre el total de predicciones realizadas [19].
  2. Recall (Sensibilidad): Proporción de verdaderos positivos sobre el total de verdaderos positivos y falsos negativos [20].
  3. F1-Score: Combina la precisión y el recall en una sola métrica [21].
  4. Matriz de confusión: Representa visualmente el rendimiento del modelo mostrando los aciertos y errores en cada clase [22].
  5. AUC-ROC: Evalúa la capacidad del modelo para discriminar entre clases en términos de sensibilidad y especificidad [23].

1. **Evaluación del rendimiento de los modelos de asociación**

Así mismo, dado que punto anterior, la evaluación del rendimiento de los algoritmos de asociación es clave para determinar la calidad de las reglas definidas. Algunas métricas y enfoques incluyen:

* 1. Medidas de soporte, confianza y lift: Aunque son las métricas tradicionales, estas medidas pueden ser insuficientes cuando los conjuntos de datos son altamente desbalanceados o cuando hay una gran cantidad de reglas generadas [24].
  2. Validación cruzada: Al igual que en otros problemas de machine learning, la validación cruzada puede ayudar a evaluar la capacidad de generalización de los modelos que descubren asociaciones [25].
  3. Métricas derivadas de los sistemas de recomendación, como precisión, recall y F1-Score, que son útiles cuando los métodos de asociación se aplican a la recomendación personalizada de ítems [14].

Ahora teniendo un contexto estructurado, acerca de las diferentes metodologías para el desarrollo del proyecto, cabe comprender la pertinencia de su aporte en la problemática de estudio, y por qué tiene valor para ser considerada dentro de este análisis.

1. **Hastie, Tibshirani y Friedman (2009)** [5], en *The Elements of Statistical Learning*, es una de las fuentes más citadas en temas de machine learning. Este libro proporciona un equilibrio entre teoría y práctica, lo que permite una comprensión profunda de los algoritmos estadísticos, incluidos los métodos de clasificación.
2. **Silver et al. (2016)** [10] y **Friedman (2001)** [7], son esenciales para abordar los avances más recientes en técnicas de clasificación, como el gradient boosting o el uso de redes neuronales profundas. Estos enfoques son altamente efectivos para el manejo de grandes volúmenes de datos y problemas complejos.
3. La obra de **Breiman (2001)** sobre **Random Forests** [7] es un hito en el campo de los algoritmos de clasificación. Breiman ofrece una explicación detallada sobre cómo los bosques aleatorios mejoran el rendimiento y reducen el sobreajuste, un problema común en los árboles de decisión individuales.
4. La referencia a **Cover y Hart (1967)** sobre el algoritmo de **Vecinos Más Cercanos (KNN)** [13] es interesante ya que introduce una técnica simple pero efectiva en ciertos escenarios de clasificación.

## **Modelos y Métodos ya existentes.**

Ya con el contexto evidenciado de la pertinencia para el uso de métodos de clasificación y asociación en el abordaje y propuesta de solución para el proyecto planteado, se analiza a continuación la implementación de dichas metodologías en problemáticas similares a la descrita.

Para este caso en particular, si bien el core del negocio de Los Olivos, son homenajes fúnebres, la UEN - Previsión tiene un comportamiento o estructura similar a una compañía de seguros. De ahí que se hace revisión de trabajos que además de aplicar técnicas de Machine Learning para la clasificación de clientes y posterior asociación de los mismos considerando la clasificación realizada; también incluyan dentro de su interés o consideración este tipo de negocio.

Así, de lo señalado anteriormente, se infiere que:

1. el uso de metodologías cómo **la Regresión logística y el Análisis Discriminante Lineal (ADL),** que son los algoritmos de machine learning más simples y más utilizados para la clasificación de dos clases. [4] Si bien en la práctica son de fácil implementación, son útiles como línea de base para cualquier problema de clasificación binaria; describen y estiman la relación entre una variable binaria dependiente y las variables independientes. De ahí que para efectos de este estudio no sería conveniente, puesto que la complejidad de los datos no tendría aplicabilidad para la clasificación de los clientes.
2. Modelos cómo **Random Forest,** que es un algoritmo de clasificación basado en la construcción de múltiples árboles de decisión independientes, combinando sus resultados para mejorar la precisión de las predicciones y tiene optimizaciones que mejoran la estabilidad y precisión del modelo.[6] Además de su aplicación en diferentes campos, incluída la clasificación de clientes según sus características. Lo perfila cómo un buen modelo a implementar cómo una aproximación viable a la problemática.
3. Así mismo, algoritmos cómo **XGBoost (Extreme Gradient Boosting)**

Que utiliza árboles de decisión de manera secuencial. Cada nuevo árbol corrige los errores cometidos por el árbol anterior, optimizando la capacidad predictiva del modelo. Permitiría obtener resultados más ajustados y con mayor nivel de precisión en la predicción.

1. **K-Nearest Neighbors (KNN),** por ejemplo, es un algoritmo de clasificación que basa su funcionamiento en la distancia entre puntos de datos.[10] Clasifica una nueva entrada según la mayoría de sus vecinos más cercanos en un espacio de características. Y es en general de valioso implementación dado que contratando lo que indican en los diferentes trabajos y varios autores, este método es potente y flexible y puede capturar eficazmente complejas dependencias de funciones no lineales, así como ofrecer resultados de alta calidad en términos de precisión, predicción y capacidad de generalización. Lo que lo perfila cómo un método para aplicar al modelo.
2. Por otro lado, también están los modelos de **Redes Neuronales,** inspirados en el funcionamiento del cerebro humano y son capaces de aprender patrones complejos a partir de datos no lineales. [12] Por otro lado, las redes bayesianas son modelos probabilísticos que representan dependencias entre variables y permiten hacer inferencias. Adicionalmente, tienen la capacidad de utilizar sus conexiones para aproximar funciones y dinámicas aprendiendo a partir de ejemplos; se han convertido en una alternativa muy popular para resolver una gran variedad de problemas. [23]

Con base en las anteriores consideraciones, la implementación de modelos predictivos en el sector funerario, centrados en el perfilamiento de clientes, puede beneficiarse de las técnicas de machine learning que han demostrado su eficacia en otros sectores, como se evidencia en los trabajos de C. Álvarez Florez y L. T. Molina Ruiz, así como de M. Sánchez Sardaña. Álvarez Florez y Molina Ruiz (2023) subrayan la importancia de la personalización en la atención al cliente, lo que resulta importante en el contexto funerario donde las decisiones son altamente influenciadas por factores emocionales.[28]

Por otro lado, Sánchez Sardaña (2022) compara algoritmos como Random Forest y XGBoost, que han mostrado un rendimiento superior en la predicción de comportamientos de consumo, lo cual es relevante para facilitar la venta cruzada y la fidelización de clientes [29].

Sin embargo, la aplicación de estos modelos en un sector como el funerario requiere una adaptación específica a pesar de tener similitudes con el sector de seguros, ya que las características del cliente en este sector pueden no alinearse perfectamente con los perfiles analizados en estudios de otras industrias. La falta de consideración de aspectos emocionales en la decisión de compra puede limitar la efectividad de las recomendaciones generadas por los modelos. Por lo tanto, es fundamental integrar enfoques que no solo se basen en datos cuantitativos, sino que también consideren las dinámicas psicológicas y culturales que influyen en el comportamiento del cliente en este contexto[30]

Asimismo, para el diseño de modelos predictivos de este proyecto se consideraron otros enfoques, aunque aplicados en sectores económicos distintos como las telecomunicaciones y la informática, ofrece valiosos aportes para el desarrollo de estrategias en el sector funerario. J. M. Alvarado Bustos (2011) destaca la utilidad de los modelos predictivos para detectar patrones de fuga en servicios, un enfoque que puede adaptarse para identificar patrones de comportamiento de los clientes en la industria funeraria[30] Al reconocer las señales que indican la posible desvinculación o desinterés de los clientes, es posible diseñar intervenciones más efectivas que promuevan la fidelización.

Tal como, C. Espino Timón (2017) enfatiza el uso de herramientas de análisis predictivo de código abierto, que facilitan la implementación de modelos sin requerir grandes inversiones en software, lo que es crucial para las empresas del sector funerario, a menudo limitadas en recursos. [31] A pesar de las diferencias sectoriales, estos estudios subrayan la importancia de personalizar las aplicaciones de las técnicas de machine learning, permitiendo que los modelos se ajusten a las particularidades del cliente funerario, lo cual es esencial para ofrecer recomendaciones y servicios que realmente resuenen con las necesidades de los usuarios en momentos críticos. De esta manera, la integración de estas metodologías puede contribuir significativamente a mejorar la experiencia del cliente y optimizar las estrategias de marketing en el sector funerario [33,34]

Finalmente, haciendo un contraste entre las técnicas utilizadas y la aplicabilidad al modelo, se tienen suficientes metodologías para hacer el correspondiente diseño de un modelo que permita predecir la clasificación de los clientes del negocio para una posterior asociación de acuerdo a los productos y de ese modo, entregar resultados óptimos que permitan el desarrollo de estrategias asertivas.

En conclusión, el desarrollo de modelos predictivos y de clasificación, aunque aplicado en sectores diversos como la educación, el transporte y las finanzas, ofrece un marco valioso para el sector funerario. La investigación de Y. K. Amaya Torrado et al, que utiliza técnicas de minería de datos para prever la deserción estudiantil, destaca la importancia de identificar patrones de comportamiento aplicables al análisis de la lealtad del cliente [28]. Asimismo, el trabajo de M. M. Bru Díaz en la predicción de flujos de pasajeros demuestra cómo estos modelos pueden anticipar la demanda de servicios, lo que es crucial para la planificación en la industria funeraria [32]. Las metodologías presentadas por M. Carmona Mora y L. M. Londoño Morales en la detección de fraude financiero, así como el enfoque de P. A. Gómez Jaramillo en la clasificación de incidentes, subrayan la aplicabilidad de técnicas de machine learning para mejorar la atención al cliente. Estos estudios evidencian la versatilidad de las técnicas de análisis de datos y su potencial para transformar la atención al cliente en el sector funerario[33,34].

## Diferenciación del Proyecto

Desde la perspectiva del negocio y la compañía, la solución propuesta se considera innovadora en el sector de la previsión. Esto no solo introduce un enfoque nuevo, sino también mejora significativamente el planteamiento en las estrategias de marketing para la venta y fidelización de nuestros afiliados.

Tradicionalmente, este sector se ha centrado en la prestación de servicios fúnebres, así como en complementos para homenajes y previsión exequial. El público objetivo de la empresa está orientado hacia un segmento específico dentro del sector solidario, lo que había llevado a que la caracterización y el perfilamiento de los clientes no se perciban como una necesidad primordial.

No obstante, con la expansión de la unidad de negocio hacia un portafolio más integral, es esencial conocer al público actual. Este conocimiento permitirá profundizar en el portafolio de servicios y focalizar estrategias que faciliten la fidelización de los clientes, garantizando que permanezcan más tiempo con nosotros mediante la renovación de sus contratos.

La solución se centra en un modelo que permite realizar un perfilamiento de la base de clientes actual. A partir de esta clasificación, podemos identificar las características de nuevos clientes y, mediante técnicas de asociación, generar recomendaciones de productos para facilitar la venta cruzada. Esto no solo optimiza los recursos disponibles, sino que también ofrece una experiencia al cliente casi personalizada, anticipando sus necesidades.

Este proyecto representa un avance significativo en la forma en que se gestionan los servicios funerarios. Al integrar un enfoque analítico y centrado en el cliente, se establece un nuevo estándar en el sector de la previsión, que no solo beneficia a la empresa, sino que también contribuye a una experiencia satisfactoria

Este enfoque innovador no solo responde a las demandas actuales del mercado, sino que también se posiciona como un referente en la transformación del sector funerario, marcando el camino hacia un futuro más adaptado a las necesidades de los clientes.

## Referencias.

[1] M. M. Patil, “Data Analytics: Concepts, Techniques, and Applications,” *International Journal of Advanced Research in Computer Science*, vol. 9, no. 3, pp. 345–349, 2021.

[2] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Waltham: Morgan Kaufmann, 2011.

[3] A. Geron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow*, O'Reilly Media, 2017.

[4] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, 3rd ed., Wiley, 2013.

[5] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed., Springer, 2009.

[6] L. Breiman, “Classification and Regression Trees,” *Machine Learning*, vol. 45, pp. 5–32, 2001.

[7] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.

[8] J. Friedman, “Greedy function approximation: A gradient boosting machine,” *Annals of Statistics*, vol. 29, pp. 1189–1232, 2001.

[9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[10] D. Silver et al., "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484-489, 2016.

[11] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, pp. 273–297, 1995.

[12] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[13] T. Cover and P. Hart, “Nearest neighbor pattern classification,” *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[14] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, Springer, 1998.

[15] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer, 2009.

[16] A. Rodriguez, M. Esteban, and C. Ruiz, "Clustering methods and association rule mining for itemset discovery," *IEEE Access*, vol. 7, pp. 11270–11281, 2019.

[17] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[18] J. Davis and M. Goadrich, “The relationship between Precision-Recall and ROC curves,” *Proceedings of the 23rd International Conference on Machine Learning*, 2006.

[19] C. Drummond and R. Holte, “Cost curves: An improved method for visualizing classifier performance,” *Machine Learning*, vol. 65, pp. 95-130, 2006.

[20] P. Rousseeuw, “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,” *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.

[21] I. Witten, E. Frank, and M. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed., Morgan Kaufmann, 2011.

[22] J. A. Swets, “Measuring the accuracy of diagnostic systems,” *Science*, vol. 240, no. 4857, pp. 1285-1293, 1988.

[23] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012

[24] N. Kriegeskorte y T. Golan, «Neural network models and deep learning» Current Biology, vol. 29, pp. R231-R236, 2019.

[25] H. Mannila, "Randomization methods for mining associations," *Data Mining and Knowledge Discovery*, vol. 12, no. 3, pp. 307–328, 2006.

[26] I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed., Morgan Kaufmann, 2011.

[27] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Pearson, 2019.

[28] M. Sánchez Sardaña, "Modelo predictivo de venta cruzada en productos de Vida y Salud: Random Forest vs XGBoost," Trabajo Fin de Máster, Máster en Ciencias Actuariales y Financieras, Universidad Carlos III de Madrid, Madrid, España, junio 2022.

[29] C. Álvarez Florez y L. T. Molina Ruiz, "Modelos de aprendizaje automático para la predicción de la preferencia en el uso de canales de atención para un fondo de pensiones y cesantías," Tesis, Universidad de Antioquia, Medellín, Colombia, 2023

[30] J. M. Alvarado Bustos, "Diseño e implementación de un modelo predictivo para detectar patrones de fuga en los servicios de Telefónica del Sur," Tesis, Ingeniero Civil en Informática, Valdivia, Chile, 2011.

[31] C. Espino Timón, "Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso," Trabajo final de Grado en Ingeniería Informática, Universidad, 2017

[32] Y. K. Amaya Torrado, E. Barrientos Avendaño, y D. J. Heredia Vizcaíno, "Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos," *Universidad Francisco de Paula Santander*, Ocaña, Colombia, 2024.

[33] M. M. Bru Díaz, "Modelos de aprendizaje supervisado para predecir la cantidad de pasajeros que saldrán de la Terminal de Transporte Norte de Medellín a otras regiones del país," Monografía, Universidad de Antioquia, Facultad de Ingeniería, Especialización en Analítica y Ciencia de Datos, Medellín, Antioquia, Colombia, 2021.

[34] M. Carmona Mora y L. M. Londoño Morales, "Modelos de machine learning para la detección de fraude financiero," Tesis, Universidad de Antioquia, Facultad de Ingeniería, Medellín, Colombia, 2021.

[35] P. A. Gómez Jaramillo, "Modelo de clasificación de incidentes tecnológicos para una empresa aseguradora desde un enfoque machine learning," Tesis, Universidad de Antioquia, Facultad de Ingeniería, Maestría en Ingeniería, Medellín, Antioquia, Colombia, 2022.

[36] L. M. Martínez Arias y S. Marulanda Vélez, "Modelo de clasificación multiclases para la predicción de apuestas deportivas," Monografía, Universidad de Antioquia, Facultad de Ingeniería, Especialización en Analítica y Ciencia de Datos, Medellín, Antioquia, Colombia, 2023.