



Tipo de artículo: Artículos originales
Temática: Inteligencia Artificial
Recibido: 28/12/2022 | Aceptado: 10/02/2023 | Publicado: 30/03/2023

Identificadores persistentes:
ARK: ark:/42411/s11/a84
PURL: 42411/s11/a84

Predicción del éxito del telemarketing bancario mediante el uso de árboles de decisión

Predicting the success of banking telemarketing through the use of decision trees

Rony Tito Ventura Ramos ^{1[0000-0002-2170-9217]*}, Andrew Pold Jacobo Castillo ^{2[0000-0003-0949-2139]}, Jesus Begazo Ticona ^{3[0000-0002-8259-4601]}, Brian Jhosep Gomez Velasco ^{4 [0000-0003-0623-4353]}

¹ Universidad Nacional de San Agustín. Arequipa, Perú. 04008. rventurar@unsa.edu.pe

² Universidad Nacional de San Agustín. Arequipa, Perú. 04000. ajacoboc@unsa.edu.pe

³ Universidad Nacional de San Agustín. Arequipa, Perú. 04146. jbegazoti@unsa.edu.pe

⁴ Universidad Nacional de San Agustín. Arequipa, Perú. 04018. bgomezv@unsa.edu.pe

* Autor para correspondencia: rventurar@unsa.edu.pe

Resumen

El telemercadeo es una técnica interactiva de mercadeo directo en la que un agente de telemercadeo solicita clientes potenciales a través del teléfono para realizar una venta de mercadería o servicio. Uno de los grandes problemas del telemarketing es especificar la lista de clientes que presentan una mayor probabilidad de comprar el producto que se ofrece. En este artículo proponemos un sistema de apoyo en la toma de decisiones personalizado que puede predecir automáticamente la decisión del público objetivo luego de realizar una llamada de telemarketing, con el fin de aumentar la efectividad de las campañas publicitarias directas y en consecuencia reducir el costo y tiempo de la campaña. El método de inteligencia artificial utilizado en este trabajo es el árbol de decisión evaluado con las métricas de precisión, exactitud y exhaustividad. Luego de aplicar el método de inteligencia artificial obtenemos una exactitud, precisión y exhaustividad mayor al 80%. Las conclusiones a las que el equipo llegó son que para mejorar el modelo de árbol de decisión es importante realizar un análisis previo de los datos mediante técnicas estadísticas o diagramas, para obtener referencia de los datos y aplicar técnicas de balanceo para obtener el mejor modelo posible.

Palabras clave: *Telemarketing, Árboles de decisión, Inteligencia artificial.*

Abstract

Telemarketing is an interactive direct marketing technique in which a telemarketing agent solicits potential customers over the phone to make a sale of merchandise or a service. One of the great problems of telemarketing is to specify the

list of clients that presents a greater probability of buying the product that is offered. In this article, we propose a personalized decision support system that can automatically predict the decision of the target audience after making a telemarketing call, in order to increase the effectiveness of direct advertising campaigns and consequently reduce the cost and cost. campaign time. The artificial intelligence method used in this work is the decision tree evaluated with the metrics of precision, accuracy and completeness. After applying the artificial intelligence method we obtain an accuracy, precision and completeness greater than 80%. The conclusions reached by the team are that in order to improve the decision tree model it is important to carry out a prior analysis of the data using statistical techniques or diagrams, to obtain a reference to the data and apply balancing techniques to obtain the best possible model.

Keywords: Telemarketing, Decision trees, Artificial Intelligence.

Introducción

Las campañas publicitarias constituyen una estrategia tradicional que las organizaciones utilizan para aumentar el número de clientes, es decir es el proceso mediante el cual se gestiona responsablemente las necesidades del cliente con el fin de entregar e intercambiar ofertas de valor con el cliente, socios o público en general [1]. Actualmente el marketing está cada vez más relacionado a los datos, automatización e inteligencia, los avances tecnológicos han producido cambios significativos en el marketing de modo que puede trabajar con la inteligencia artificial y generar oportunidades para la empresa [2]. El telemercado es una técnica interactiva de mercadeo directo en la que un agente de telemercado solicita clientes potenciales a través del teléfono para realizar una venta de mercadería o servicio [3]. Esta tecnología permite repensar el marketing enfocándose en maximizar el valor de por vida del cliente a través de la evaluación de la información disponible y métricas del cliente [4]. Existen 2 tipos de metodologías por las cuales las organizaciones promueven sus productos: a) centradas en la población general y b) campañas directas. La primera solo genera menos del 1% de reacciones positivas en toda la población, sin embargo, las campañas directas que se concentran en un grupo pequeño de personas que se cree que tienen una mayor probabilidad de sentirse interesados en el producto, son mucho más productivas para la empresa financiera [5, 7]. Uno de los grandes problemas del telemarketing es especificar la lista de clientes que presentan una mayor probabilidad de comprar el producto que se ofrece [6]. Para ello, los sistemas de soporte de decisiones que utilizan modelos predictivos pueden proporcionar una toma de decisiones mejor informada [6]. Incluso en la actualidad, muchos bancos han optado por la minería de datos para predecir los datos del cliente para clasificar

a los clientes antes de ofrecer servicios especiales, estas técnicas se centran en hacer coincidir los atributos del cliente y otras características a diferentes resultados [5]. Sin embargo, un factor importante que afecta el rendimiento de la predicción es el número de características de entrada. Especialmente, la información del cliente del banco es que normalmente tiene muchas características, por lo que hace que disminuya el rendimiento de la predicción [3]. Otra de los problemas presentamos es manejar la distribución de conjuntos de datos desequilibrados de manera confiable; los enfoques comúnmente usados imponen una sobrecarga de procesamiento o conducen a la pérdida de información. Las Redes Neuronales Artificiales (RNN) como se indica en [8][9] en algunos casos son muy competitivas, ya que pueden manejar cualquier tipo de distribución de datos desequilibrada, pero si se crean modelos sensibles al costo tenemos un fuerte candidato para dar una solución eficiente.

En este artículo proponemos un sistema de apoyo en la toma de decisiones personalizado que puede predecir automáticamente la decisión del público objetivo luego de realizar una llamada de telemarketing, con el fin de aumentar la efectividad de las campañas publicitarias directas y en consecuencia reducir el costo y tiempo de la campaña, para ello nos enfocamos en un rango limitado de características de entrada, principalmente indicadores socioeconómicos como: edad, ocupación, educación, moroso, etc. Los registros son analizados y se les aplica una técnica de Inteligencia artificial para obtener el sistema de apoyo. El artículo está organizado de la siguiente manera en la sección de metodología se explica la aplicación de una técnica de inteligencia artificial y el procedimiento que se realizó para obtener el sistema soporte de decisiones en la siguiente sección presentamos la discusión de los resultados y finalmente las conclusiones.

Trabajos relacionados

El trabajo realizado por Albrecht, Tobias, Rausch aborda los métodos y prácticas de implementación de inteligencia artificial en las llamadas que ingresan a un call center, este trabajo investiga las capacidades de los modelos de machine learning para pronosticar las llamadas que ingresan diariamente en un call center y medir su precisión y practicidad para ello los investigadores hacen uso de dos datasets de dos empresas retail, apoyo al cliente y reclamos del cliente, como primer paso realizan el análisis preliminar, luego diseñan la investigación, las variables predicativas, finalmente obtienen el resultado con diferentes algoritmos, random

forest, gradient boosting machine, K-nearest neighbor y vector regression, los resultados que los investigadores obtuvieron muestran que el algoritmo de random forest es el modelo que tuvo mejor rendimiento en las predicciones [10]. El artículo sirve de perspectiva para el desarrollo de esta investigación, sin embargo, difiere en el algoritmo usado y el enfoque de la investigación, ya que el artículo presenta un punto de partida para el pronóstico de llamadas de un call center mediante los algoritmos de inteligencia artificial.

El trabajo realizado por Zeinulla, Bekbayeva y Yazici [11] tiene como objetivo evaluar varios modelos de clasificación para la predicción de resultados de campañas de telemarketing bancario en cuanto a la probabilidad de suscripción del cliente al depósito. Para este propósito, la eficacia de los algoritmos ha sido evaluados mediante el análisis de la curva de características del operador de receptor (ROC) y el perfil de precisión acumulada (CAP), la precisión del algoritmo y la varianza de las predicciones. Según los resultados de la investigación, los mejores modelos para la predicción de la efectividad del telemarketing bancario son Random Forest con una precisión de 0.90 por ROC y Redes Neuronales Artificiales Profundas (DANN) con una precisión de 0.86. El artículo sirve de perspectiva para el desarrollo de esta investigación, sin embargo, difiere del algoritmo usado y la metodología de análisis.

Marco Teórico

Árbol de decisión

Un árbol de decisiones es una herramienta de apoyo con una estructura similar a un árbol que modela los resultados probables, el costo de los recursos, las utilidades y las posibles consecuencias. Los árboles de decisión proporcionan una forma de presentar algoritmos con declaraciones de control condicional, estos árboles incluyen ramas que representan pasos de toma de decisiones que pueden conducir a un resultado favorable.

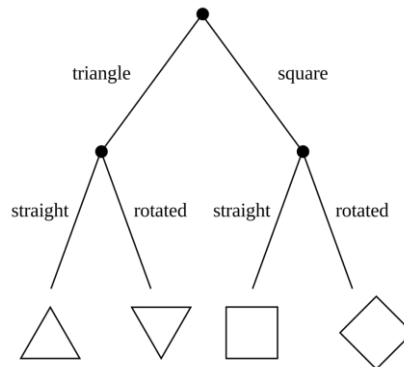


Figura 1. Árbol de decisión simple

La estructura del diagrama de flujo incluye nodos internos que representan pruebas o atributos en cada etapa y cada rama representa un resultado para los atributos, mientras que el camino desde la hoja hasta la raíz representa reglas para la clasificación. Los árboles de decisión son una de las mejores formas de algoritmos de aprendizaje basados en varios métodos de aprendizaje. Impulsan los modelos predictivos con precisión, facilidad de interpretación y estabilidad. Las herramientas también son efectivas para ajustar relaciones no lineales, ya que pueden resolver desafíos de ajuste de datos, como la regresión y las clasificaciones.

Python

El lenguaje de programación python es un lenguaje de alto nivel, las propiedades de este lenguaje es que es un lenguaje de alto nivel orientado a objetos, su semántica fue desarrollado por Guido van Rossum, Este lenguaje fue diseñado para ser fácil y entendible de programar. Python es un lenguaje de programación fácil de entender por los programadores principiantes por la sintaxis sencilla de entender y la gran cobertura de sus métodos. Python es utilizado para el desarrollo web, el desarrollo de programas de software, las matemáticas y la creación de secuencias de comandos del sistema y lenguaje de secuencias de comandos debido a sus estructuras de datos integradas de alto nivel. escritura dinámica y enlace dinámico. Las librerías creadas para Python facilitan los programas modulares y la reutilización del código. Python es un lenguaje comunitario de código abierto, por lo que numerosos programadores independientes crean continuamente bibliotecas y funcionalidades para él.

Algoritmo C4.5

El algoritmo C4.5 es una versión mejorada del algoritmo ID3 desarrollado por Ross Quinlan, este algoritmo utiliza el ratio de ganancia y la función de bondad para dividir el conjunto de datos, en cambio el algoritmo ID3 usa la ganancia de información. La función de ganancia de información prefiere las características con más categorías, los árboles de decisión que nos da C4.5 son usados mayormente para la clasificación. La ventaja de usar este algoritmo para el proyecto es que acepta variables categóricas y variables numéricas del tipo continuo y discretas [20].

Telemarketing bancario

El telemarketing es una técnica interactiva de mercadeo directo en la que un agente de telemarketing solicita clientes potenciales a través del teléfono para realizar una venta de mercadería o servicio [3]. Esta tecnología permite repensar el marketing enfocándose en maximizar el valor de por vida del cliente a través de la evaluación de la información disponible y métricas del cliente [4]. Existen 2 tipos de metodologías por las cuales las organizaciones promueven sus productos: a) centradas en la población general y b) campañas directas. La primera solo genera menos del 1% de reacciones positivas en toda la población, sin embargo, las campañas directas que se concentran en un grupo pequeño de personas que se cree que tienen una mayor probabilidad de sentirse interesados en el producto, son mucho más productivas para la empresa financiera [5, 7]. Uno de los grandes problemas del telemarketing es especificar la lista de clientes que presentan una mayor probabilidad de comprar el producto que se ofrece [6].

Sistema de soporte a la decisión

Un sistema de soporte a la decisión (DSS) es una herramienta que se encuentra dentro del paraguas de Inteligencia de negocios, enfocada al análisis de datos de una organización para ayudar en la toma de decisiones de una organización, un DSS usa las tecnologías de la información para apoyar la toma de decisiones, existen muchos subcampos en los cuales se puede aplicar, desde pequeños sistemas personales hasta grandes sistemas para organizaciones, estos sistemas usan técnicas de inteligencia artificial (IA) para apoyar la toma de decisiones, por ende es la columna vertebral detrás de estos sistemas [4]. Los DSS están compuestos por una interfaz de usuario que recibe y responde las consultas y un motor de inferencias, el cual

es capaz de usar todo el conocimiento acumulado, base de conocimiento, y dado unos parámetros de entrada es capaz de ejecutar una serie de reglas lógicas y responder preguntas de alto nivel de acuerdo a la consulta que se realizó. El administrador tiene acceso al resultado del análisis consolidado y toma decisiones en beneficio de la organización [13].

Herramientas Utilizadas

Sklearn

También conocido como scikit-learn es una librería hecha en Python que nos proporciona un amplio repertorio de algoritmos de aprendizaje automático además que son de última generación, dicha librería prioriza llevar el aprendizaje automático a los no especialistas, teniendo en cuenta que se realiza por medio de un lenguaje de alto nivel como lo es Python. Esta librería proporciona distintas ventajas como la facilidad de uso que tiene, rendimiento, documentación, coherencia de la API. Además que su dependencia llega a ser mínima, se distribuye bajo licencia BSD. Es bastante utilizado en sectores comerciales y también en los entornos académicos[18].

Google Colab

Colaboratory o “Colab”, es un producto de Google Research. Permite a cualquier usuario escribir y ejecutar código de Python en el navegador. Este software es a menudo utilizado para tareas de aprendizaje automático, análisis de datos y educación. La razón de esto es que Colab nos permite cambiar los ajustes del entorno para realizar ejecuciones más potentes que usualmente serían poco eficientes ejecutarlas en nuestro propio ordenador [12].

Pandas

Pandas es un paquete de Python que provee estructuras de datos rápidas, flexibles y expresivas, diseñado para hacer tareas con datos etiquetados fácil e intuitivos, pretende ser un fundamental bloque de alto nivel para realizar prácticas reales en análisis de datos en Python. Adicionalmente tiene como propósito fundamental llegar a ser la herramienta open source más poderosa y flexible para el análisis y manipulación de datos en

cualquier lenguaje de programación [15]. Pandas fue desarrollado a inicios del 2008 y fue publicado como open source a finales del siguiente año, algunas características fundamentales de esta herramienta son las siguientes, permite el fácil manejo de datos faltantes, las estructuras de datos que provee son fácilmente mutables en sus dimensiones, permite convertir datos, además de mapeos y mezclas entre otras funciones útiles [15, 16].

Resultados y discusión

Análisis del problema

Los datos están relacionados con campañas de marketing directo de una entidad bancaria portuguesa. Las campañas de marketing se basaron en llamadas telefónicas. El objetivo de la clasificación es predecir si el cliente suscribirá (si/no) un depósito a plazo (variable). Tenemos 9 variables categóricas de entrada que son: job, marital, education, default, month, housing, loan, contact, poutcome. Además 7 variables numéricas de entrada que son: age, balance, day, duration, campaign, pdays, previous. Finalmente una variable categórica de salida que es: y (yes, no)

Variable	Descripción
Age	Edad del objetivo, numérico
Job	Tipo de trabajo, categórico
Marital	Estado marital, categórico.
Education	Grado educativo, categórico.
Default	Tiene crédito en mora, binario.
Balance	Saldo medio anual en años, numérico.
Housing	Tiene crédito hipotecario, binario.
Loan	Tiene crédito personal, binario.
Contact	Tipo de comunicación, categórico.
Day	Último día de contacto del mes, numérico.
Duration	Tiempo de duración del último contacto en segundos, numérico.
Campaign	Número de contactos realizados durante la campaña actual, numérico.
Pdays	Número de días desde la última campaña, numérico.
Previous	Número de contactos realizados antes de esta campaña, numérico.
Poutcome	Resultado de la campaña anterior, categórico.
Month	Último mes de contacto en el año, categórico.
Salida	El cliente se suscribe o no (yes, no)

Tabla 1: Descripción de variables.

Análisis de los datos

Con la ayuda de Matplot se graficó todas las variables con el objetivo de localizar datos no relevantes (en el caso de las variables categóricas) y datos outliers (en el caso de las variables numéricas). Es importante aplicar un método de tratamiento sobre estas para que el modelo mejore, especialmente si se usa árboles de decisión. En nuestro caso elegimos eliminarlas.

A. Data Cleaning

En las variables categóricas hacemos uso de diagramas de barras para representar los datos con el fin de detectar anomalías en los datos.

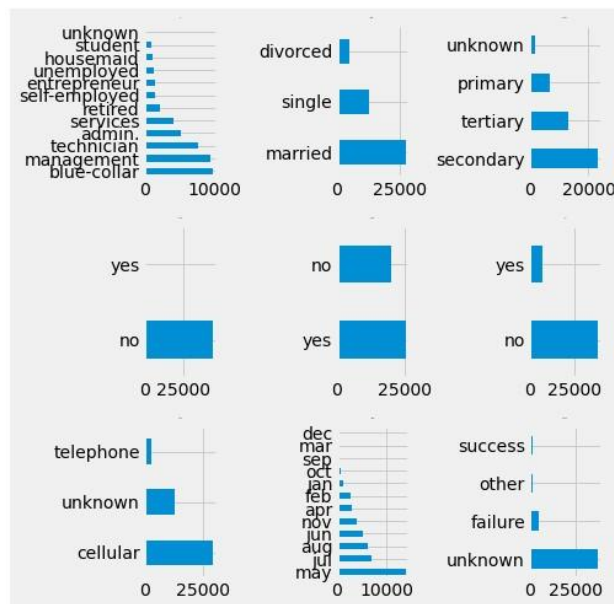


Figura 2: Diagrama de Barras para las variables categóricas

Primero eliminamos las variables irrelevantes que no aportan mucho a nuestro modelo que son: marital, month, day, contact, previous y default. En las variables que si consideramos vemos que hay valores “unknown”, estas se eliminan. En job, el valor “student” lo consideramos irrelevante al igual que el valor “other” de poutcome. Esto nos deja con 40716 filas. Para encontrar outliers en nuestras variables numéricas

nos hemos apoyado de diagramas de Caja y Bigote que nos permiten visualizar si estas tienen datos atípicos. Se ha utilizado la librería seaborn con la función plotbox().

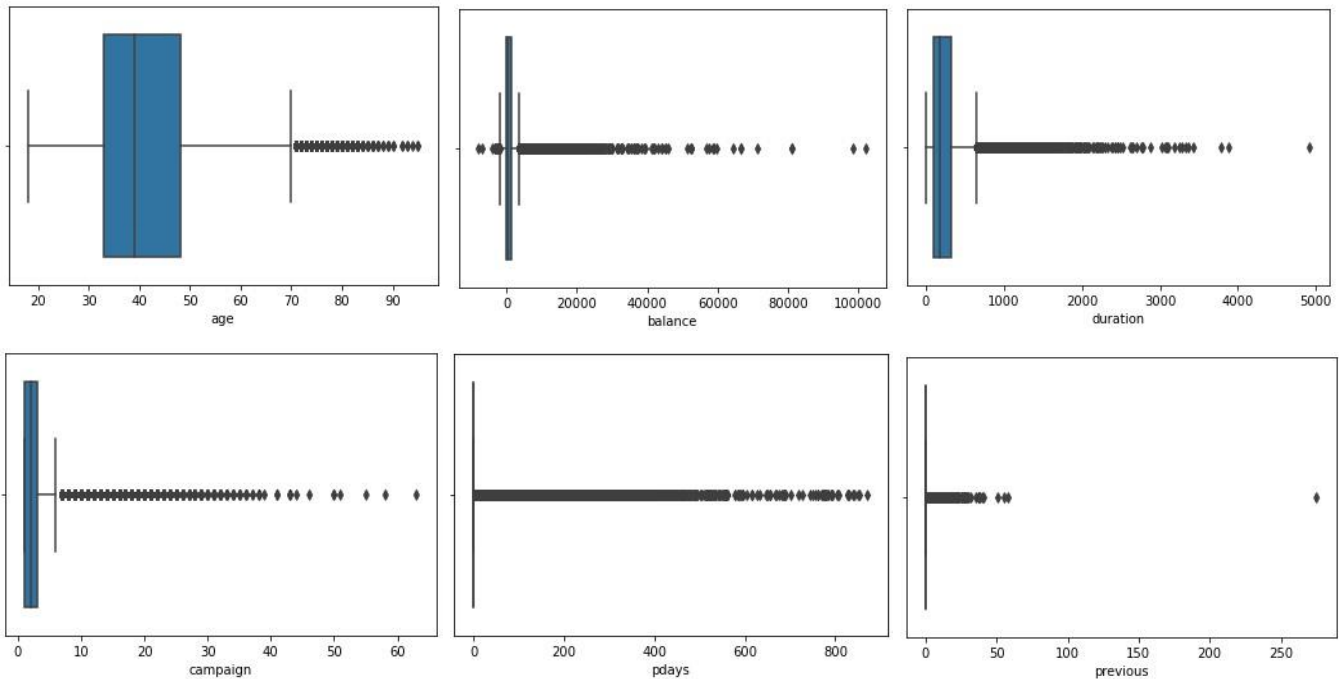


Figura 3: Diagramas de Caja y Bigote para las variables numéricas

Para eliminar este tipo de valores se ha empleado el puntaje estándar o también conocido como z-score, este valor estadístico no dará una idea de qué tan lejos de la media está un punto de datos. Para cumplir con nuestro objetivo hemos empleado la librería scipy con el módulo stats y su función `zscore()` que nos deja con un total de 39178 filas.

B. Transformación de los datos

El algoritmo C4.5 utilizado para el entrenamiento de nuestro árbol de decisión es una adaptación del ID3 que permite tanto variables numéricas (continuas y discretas) como categóricas. Sin embargo, sólo admite entradas numéricas lo cual lleva a realizar una transformación en los datos especialmente en las variables categóricas las cuales debemos asignar una representación numérica por cada clase diferente. Con la ayuda

de pandas y sus funciones map() podemos hacer el reemplazo de estas mismas y, además, eliminar valores vacíos como NaN o Null.

Variable	Valores
job	"admin.": 0, "unknown": 1, "unemployed": 2, "management": 3, "housemaid": 4, "entrepreneur": 5, "blue-collar": 6, "self-employed": 7, "retired": 8, "technician": 9, "services": 10
education	"secondary": 0, "primary": 1, "tertiary": 2
housing	"yes": 1, "no": 0
loan	"yes": 1, "no": 0
poutcome	"unknown": 0, "failure": 1, "success": 2
y	"yes": 1, "no": 0

Tabla 2: Transformación de las variables categóricas

C. Balanceo de datos

Algo que caracteriza a los datos usados para este trabajo es su gran desbalance con respecto a la variable de salida. En la literatura consultada se presentan diferentes métodos para resolver este tipo de problema. En una fase inicial se probó eliminando las filas que contengan “no” en la variable objetivo, sin embargo, al realizar este proceso provocó que las demás variables de entrada se desbalancen dando como resultado un árbol con bastante precisión en su modelo, pero mal entrenado. La solución más óptima encontrada fue aumentar las salidas de la clase “yes” para que de esta manera no tengamos tanta diferencia en el total de los resultados. Para ello se ha utilizado Smote que es una herramienta que se encarga de sintetizar nuevos ejemplos para la clase minoritaria. Al finalizar este proceso obtenemos 33280 filas de clase mayoritaria y 26624 de la clase minoritaria.

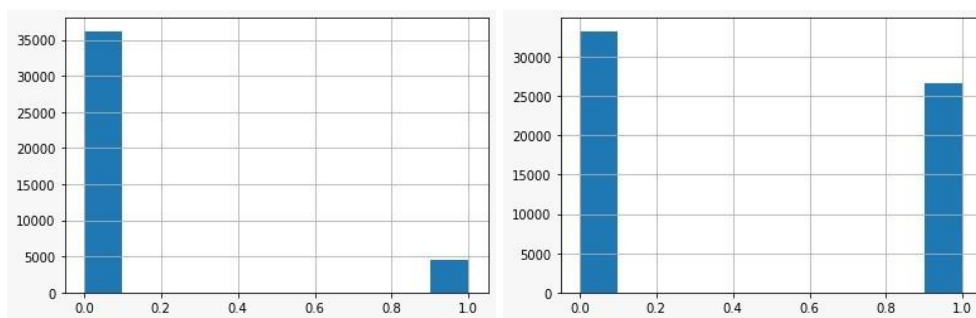


Figura 4: Balanceo de la clase minoritaria en la variable de salida

Entrenamiento

En esta sección describimos la manera de cómo se prepara el algoritmo, para el árbol de decisión, se separa los datos para el entrenamiento y otros para la prueba, en el caso propuesto se selecciona el 80% de los datos para el entrenamiento y el 20% para las pruebas. También se utiliza el criterio de “entropy” para el algoritmo, luego se determina el nivel máximo del árbol en este caso es 6, luego se realiza el entrenamiento con los datos de prueba y los parámetros establecidos.

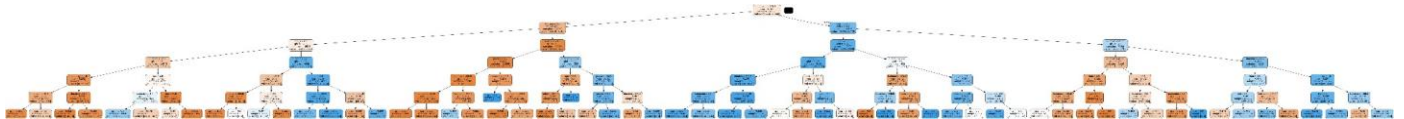


Figura 5: Árbol de Decisión.

Métricas

A. Precisión

La precisión es la capacidad del modelo de no clasificar como positiva una nuestra que es negativa, se calcula mediante la razón $tp / (tp + fp)$, donde tp es el número de verdaderos positivos y fp el número de falsos positivos. En nuestro caso de estudio usamos la librería sklearn para obtener la precisión del modelo y obtenemos 0.81 lo cual indica que el modelo es aceptable pero no óptimo.

```
precision = precision_score(y_test, y_pred)
print('Precisión del modelo:')
print(precision)

Precisión del modelo:
0.8139244458320325
```

Figura 6: Precisión del modelo.

B. Exactitud

La exactitud mide el porcentaje de casos que el modelo ha acertado, este modelo no funciona bien cuando los datos están desbalanceados, es mucho mejor el uso de métricas como precisión, exhaustividad, F1.

```
accuracy_score = accuracy_score(y_test, y_pred)
print('Exactitud del modelo:')
print(accuracy_score)

Exactitud del modelo:
0.8473192616018239
```

Figura 7: Exactitud del modelo

C. Exhaustividad

La exhaustividad es intuitivamente la capacidad que tiene el modelo para encontrar todas las muestras positivas, se calcula mediante, $tp / (tp + fn)$, donde tp son los verdaderos positivos y fn los falsos negativos.

```
recall_score = recall_score(y_test, y_pred)
print('Exhaustividad del modelo:')
print(recall_score)

Exhaustividad del modelo:
0.8550344375204986
```

Figura 8: Exhaustividad del modelo

D. Validación

Los valores obtenidos en las métricas de precisión, exactitud y exhaustividad son aceptables, se encuentran por encima de ochenta puntos, sin embargo no son valores óptimos, para comprobar que el modelo es capaz de predecir correctamente se tomó un valor aleatorio dentro del conjunto de datos y observar el resultado.

```
-----Valor X-----
age          35
job           4
education    3
balance     1047
housing      0
loan         0
duration    528
campaign     2
pdays      -1
Name: 50431, dtype: int64
-----Valor Y-----
1
***** Resultado *****
[1]
```

Figura 9: Validación del modelo

Conclusión

Dentro de la industria del marketing es importante optimizar el público objetivo para el telemercadeo, ya que escoger el público correcto reduce los costos y aumenta la probabilidad de ganancias. En este estudio nosotros presentamos la implementación de un árbol de decisión como modelo de solución, para saber si un cliente

acepta o no un crédito bancario. El conjunto de datos obtenidos muestra un desbalance excesivo, por ello se aplicó técnicas para el balanceo como, smote, eliminación de valores desconocidos y variables irrelevantes con el objetivo de mejorar el modelo de árbol de decisión. Adicionalmente se realizó la transformación de los datos mapeando el dominio de los datos, además de asignar valores a los datos desconocidos. Para comprobar si el modelo es útil se aplicó las siguientes métricas: precisión, exactitud y exhaustividad, obteniendo como resultado valores aceptables que están por encima del 80%, es importante realizar un análisis previo de los datos, mediante técnicas estadísticas o gráficas como los diagramas de barras, caja y bigote según el tipo de dato y su dominio.

Trabajos Futuros

En trabajos futuros, planteamos realizar estudios sobre el mismo conjunto de datos aplicando otras técnicas de inteligencia artificial como redes neuronales, regresión logística. Luego de aplicar las técnicas anteriormente mencionadas, realizar estudios comparativos para obtener la técnica y el objeto que mejor se ajuste a los datos.

Referencias

- [1] ESAN, “El Marketing y sus definiciones | Conexión ESAN,” 2016. <https://www.esan.edu.pe/conexion-esan/el-marketing-y-sus-definiciones> (accessed Jun. 22, 2022).
- [2] S. Chintalapati and S. K. Pandey, “Artificial intelligence in marketing: A systematic literature review,” *Int. J. Mark. Res.*, vol. 64, no. 1, pp. 38–68, 2022, doi: 10.1177/14707853211018428.
- [3] Vajiramedhin, C., & Suebsing, A. (2014). Feature selection with data balancing for prediction of bank telemarketing. *Applied Mathematical Sciences*, 8(114), 5667-5672.
- [4] Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, 62, 22-31.

- [5] Asare-Frempong, J., & Jayabalan, M. (2017, September). Predicting customer response to bank direct telemarketing campaign. In *2017 International Conference on Engineering Technology and Technopreneurship (ICE2T)* (pp. 1-4). IEEE
- [6] Moro, S., Cortez, P., & Rita, P. (2018). A divide-and-conquer strategy using feature relevance and expert knowledge for enhancing a data mining approach to bank telemarketing. *Expert Systems*, 35(3), e12253.
- [7] Lau, Kn., Chow, H. & Liu, C. A database approach to cross selling in the banking industry: Practices, strategies and challenges. *J Database Mark Cust Strategy Manag* 11, 216–234 (2004).
- [8] Ghatasheh, N., Faris, H., AlTaharwa, I., Harb, Y., & Harb, A. (2020). Business analytics in telemarketing: cost-sensitive analysis of bank campaigns using artificial neural networks. *Applied Sciences*, 10(7), 2581.
- [9] Kim, K. H., Lee, C. S., Jo, S. M., & Cho, S. B. (2015, November). Predicting the success of bank telemarketing using deep convolutional neural network. In *2015 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR)* (pp. 314-317). IEEE.
- [10] Albrecht, Tobias; Rausch, Theresa Maria; Derra, Nicholas Daniel (2021). Call me maybe: Methods and practical implementation of artificial intelligence in call center arrivals forecasting. *Journal of Business Research*, 123(), 267–278, doi:10.1016/j.jbusres.2020.09.033
- [11] E. Zeinulla, K. Bekbayeva and A. Yazici, "Comparative study of the classification models for prediction of bank telemarketing," 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT), 2018, pp. 1-5, doi: 10.1109/ICAICT.2018.8747086.

- [12] "Google Colab". Google Research. <https://research.google.com/colaboratory/intl/es/faq.html#:~:text=Colaboratory,%20o%20Colab%20para,análisis%20de%20datos%20y%20educación>. (accedido el 17 de agosto de 2022).
- [13] G. Stalidis, D. Karapistolis, and A. Vafeiadis, "Marketing Decision Support Using Artificial Intelligence and Knowledge Modeling: Application to Tourist Destination Management," *Procedia - Soc. Behav. Sci.*, vol. 175, pp. 106–113, 2015, doi: 10.1016/j.sbspro.2015.01.1180.
- [15] *pandas - Python Data Analysis Library*. (n.d.). Retrieved August 14, 2022, from <https://pandas.pydata.org/>
- [16] *pandas · PyPI*. (n.d.). Retrieved August 14, 2022, from <https://pypi.org/project/pandas/>
- [17] MARTÍNEZ, Guillermo Roberto Solarte; MEJÍA, José A. Soto. Árboles de decisiones en el diagnóstico de enfermedades cardiovasculares. *Scientia et technica*, 2011, vol. 16, no 49, p. 104-109.
- [18] GARCÍA PICHARDO, Víctor Hugo, et al. Algoritmo ID3 en la detección de ataques en aplicaciones Web.
- [19] PEDREGOSA, Fabian, et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 2011, vol. 12, p. 2825-2830.
- [20] Larranaga, P., Inza, I., & Moujahid, A. Tema 10: árboles de clasificación. 2020 from <http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t10arboles.pdf>.

Roles de Autoría

Rony Tito Ventura Ramos: Conceptualización, Curación de datos, Análisis formal, Investigación, Metodología, Software, Validación, Redacción - borrador original. **Andrew Pold Jacobo Castillo:** Conceptualización, Curación de datos, Análisis formal, Investigación, Metodología, Software, Validación, Redacción - borrador original. **Jesus Begazo Ticona:** Conceptualización, Curación de datos, Análisis formal, Investigación, Metodología, Software, Validación, Redacción - borrador original. **Brian Jhosep Gomez Velasco:** Conceptualización, Curación de datos, Análisis formal, Investigación, Metodología, Software, Validación, Redacción - borrador original.