

# APLICACIÓN DE *MACHINE LEARNING* PARA CAMPAÑAS DE *MARKETING* EN LA BANCA COMERCIAL

GANÍMEDES T. ROSALES REYES

[grosalesr@unmsm.edu.pe](mailto:grosalesr@unmsm.edu.pe)

<https://orcid.org/0000-0001-6415-6109>

Universidad Nacional Mayor de San Marcos, Perú

XAVIER ALBERTO GUTIERREZ CORAL

[xavier.gutierrez@unmsm.edu.pe](mailto:xavier.gutierrez@unmsm.edu.pe)

<https://orcid.org/0000-0002-0250-0864>

Universidad Nacional Mayor de San Marcos, Perú

AUGUSTO ENRIQUE HAYASHIDA MARCHINARES

[augusto.hayashida@unmsm.edu.pe](mailto:augusto.hayashida@unmsm.edu.pe)

<https://orcid.org/0000-0002-4278-5008>

Universidad Nacional Mayor de San Marcos, Perú

**RESUMEN.** Los bancos usan el *telemarketing* para contactar directamente con los clientes potenciales para sus productos. Este canal de venta es complejo, pues requiere de grandes bases de datos de posibles prospectos y está sujeto a restricciones de tiempo y personal. Este artículo tiene tres objetivos: comparar cinco modelos de predicción basados en algoritmos de aprendizaje automático para encontrar el que ofrezca la mejor precisión predictiva; desplegar un piloto de este modelo; y recomendar una hoja de ruta para la futura arquitectura que lo soporte. Se encontró que el algoritmo seleccionado mejora considerablemente la eficacia de la identificación de clientes que aceptan el producto, que pasó de 11 % a 94 %, por lo que su implementación puede contribuir a la competitividad de estas organizaciones.

**PALABRAS CLAVE:** banca, *marketing*, depósitos a plazo fijo, aprendizaje automático, algoritmos de clasificación

## MACHINE LEARNING APPLICATION FOR CAMPAIGNS MARKETING IN COMMERCIAL BANKING

**ABSTRACT.** Banks use telemarketing to contact potential customers for their products directly. This sales channel is complex, requiring large databases of possible prospects, and is subject to time and personnel restrictions. This article has three objectives: to compare five prediction models based on machine learning algorithms to find the one that offers the best predictive accuracy, deploy a pilot of this model, and recommend a roadmap for the future architecture that supports it. The comparison results show that the selected algorithm considerably improves the identification of customers who accept the product, which went from 11 % to 94 %, so its implementation can contribute to the competitiveness of these organizations.

**KEYWORDS:** banking, marketing, fixed-term deposits, machine learning, classification algorithms

## 1. INTRODUCCIÓN

Una parte importante del negocio bancario son las operaciones pasivas, como la apertura de cuentas dirigidas a los clientes minoristas, en sus diferentes formas, ya sean cuentas de ahorros, cuentas corrientes o cuentas a plazo fijo. Sin embargo, uno de los desafíos que enfrenta la banca comercial en este ámbito es el hecho de tener que contactar a una cantidad importante de clientes, aun con los escasos recursos, tanto de tiempo como de materiales, con los que se cuenta. Por ello, para lograrlo los bancos han adoptado la ejecución de campañas comerciales en modalidad masiva o en modalidad directa, es decir, dirigidas a segmentos específicos de clientes (Dutta et al., 2021).

No obstante, es sabido que los resultados de este tipo de campañas no tienen ratios de éxito altos, por lo que es necesario el uso de estrategias cada vez más elaboradas a fin de mejorar su efectividad. En este esfuerzo, desde hace unos años, muchos bancos han ido incorporando de manera progresiva herramientas basadas en *machine learning* y minería de datos con el objetivo de incrementar el nivel de éxito de sus campañas, identificando para ello los principales factores que pueden conducir a él (Moro et al., 2011).

La presente investigación tiene tres objetivos. El primero de ellos es mostrar la evaluación de cuatro algoritmos entre los más comunes en el campo de los sistemas cognitivos a fin de demostrar las ventajas de este tipo de herramientas para la gestión comercial en la banca personal; los algoritmos evaluados fueron (i) *decision tree*, (ii) KNN, (iii) ANN (RPROP), (iv) *naive Bayes* y (v) XGBoost. Para dicha evaluación, se tomaron los datos de una campaña comercial llevada a cabo por un banco europeo con una muestra de 41 188 registros, y se utilizó la estrategia de particiones, es decir, se subdividió en dos grupos ese conjunto de datos, de los cuales el 70 % servirá para que el algoritmo evaluado aprenda y el otro 30 % para que dicho algoritmo sea probado. El segundo objetivo es mostrar como demostración una aplicación web que implementa el algoritmo que mejor desempeño tuvo. Por último, el tercer objetivo consiste en proponer una arquitectura empresarial futura con visión integral que permita construir una solución integral al proceso de campañas de *marketing*.

## 2. ESTADO DEL ARTE

A fin de lograr los tres objetivos planteados, se describe a continuación el estado del arte para los algoritmos de *machine learning*, así como para el concepto de arquitectura empresarial.

Respecto a los algoritmos de *machine learning*, se puede afirmar que los bancos están usando diferentes técnicas en sus implementaciones para descubrir y luego

utilizar patrones a partir de sus bases de datos. Entre los principales se encuentran los siguientes:

### 2.1 *Decision tree*

Los árboles de decisión se han convertido en uno de los modelos más potentes y populares en la ciencia de datos, como ciencia y tecnología de exploración de grandes y complejos conjuntos de datos, donde ayuda a descubrir patrones útiles. El objetivo del árbol de decisión es clasificar una instancia en un conjunto predefinido de clases en función de los valores de sus atributos (Rokach, 2016). Asimismo, Zhang et al. (2019) indican que, a diferencia de la regresión logística, el árbol de decisión se construye de arriba hacia abajo en una secuencia; esto es, todo el conjunto de datos se divide en particiones más pequeñas hasta que no se pueden realizar más particiones.

### 2.2 *K-nearest neighbors (KNN)*

Es un modelo simple y eficaz que no requiere parámetros. El proceso de clasificación de KNN consiste en realizar el cálculo de la similitud entre un objeto objetivo y los  $k$  vecinos más cercanos y similares en el conjunto de muestra de entrenamiento. La distancia de similitud de KNN normalmente se mide por la distancia euclidiana (Zhang et al., 2019).

#### Figura 1

*Fórmula para el cálculo de la distancia que utiliza el método KNN*

$$d(x, x_i) = \sqrt{\sum_{i=1}^n (x - x_i)^2}$$

*Nota.* Donde  $x$  es el objetivo y  $x_i$  es el  $i$ -ésimo vecino similar más cercano. Luego, al estar  $x$  más cerca de sus vecinos, el destino se asignará a la clase más común entre sus  $k$  vecinos más cercanos.

### 2.3 *Redes neuronales ANN (RPROP)*

La ANN está inspirada biológicamente en el cerebro humano. Las neuronas están interconectadas en el cerebro humano de manera similar a como los nodos están interconectados en la red neuronal artificial (Asha & Kumar, 2021). De acuerdo con Riedmiller y Braun (1993), es un algoritmo de aprendizaje para redes de alimentación directa multicapa, que fue propuesto para superar las desventajas inherentes del descenso-gradiente puro. RPROP realiza una adaptación local de las actualizaciones

ponderadas según el comportamiento de la función de error. A diferencia de otros modelos adaptativos, el efecto del proceso de adaptación de RPROP no se ve empañado por la influencia imprevisible del tamaño de la derivada, sino que solo depende del comportamiento temporal de su signo. A fin de mostrar la lógica del algoritmo, en la Figura 1 se presenta un fragmento de pseudocódigo que muestra el núcleo del proceso de adaptación y aprendizaje de RPROP.

**Figura 2**

*Núcleo del proceso de adaptación de RPROP*

```

For all weights and biases{
  if  $\left(\frac{\partial E}{\partial w_{ij}}(t-1) * \frac{\partial E}{\partial w_{ij}}(t) > 0\right)$  then {
     $\Delta_{ij}(t) = \text{minimum}(\Delta_{ij}(t-1) * \eta^+, \Delta_{\max})$ 
     $\Delta w_{ij}(t) = - \text{sign} \frac{\partial E}{\partial w_{ij}}(t) * \Delta_{ij}(t)$ 
     $w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t)$ 
  }
  else if  $\left(\frac{\partial E}{\partial w_{ij}}(t-1) * \frac{\partial E}{\partial w_{ij}}(t) < 0\right)$  then {
     $\Delta_{ij}(t) = \text{maximum}(\Delta_{ij}(t-1) * \eta^-, \Delta_{\min})$ 
     $w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t-1)$ 
     $\frac{\partial E}{\partial w_{ij}}(t) = 0$ 
  }
  else if  $\left(\frac{\partial E}{\partial w_{ij}}(t-1) * \frac{\partial E}{\partial w_{ij}}(t) = 0\right)$  then {
     $\Delta w_{ij}(t) = - \text{sign} \frac{\partial E}{\partial w_{ij}}(t) * \Delta_{ij}(t)$ 
     $w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t)$ 
  }
}

```

*Nota.* Reproducido de "A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm" (p. 588), por M. Riedmiller y H. Braun, 1993, *IEEE International Conference on Neural Networks*.

Se supone que el operador mínimo (máximo) debe entregar el mínimo (máximo) de dos números. El operador de signo devuelve +1 si el argumento es positivo; -1 si el argumento es negativo; y 0 en caso contrario.

## 2.4 Naive Bayes

Este modelo de clasificación permite estimar la probabilidad de una hipótesis a partir de la data experimental. Conforme se suma más data a la muestra, se va ajustando la probabilidad resultante. Su base matemática es el teorema de Bayes. En el modelo se asume que el valor de cada característica de la base de datos es independiente de las demás características, dada la variable de clase, y que cada una contribuye independientemente a la predicción final, sin establecerse correlaciones entre ellas; de allí la calificación de *naive* o ingenuo (Berrar, 2018).

Dado un vector  $x = (x_1, x_2, \dots, x_n)$  que representa una instancia por ser clasificada, formada por  $n$  variables independientes, el modelo considera la asignación de probabilidades para cada clase  $k$  (Murty & Devi, 2011).

$$p(C_k | x_1, x_2, \dots, x_n)$$

Según el teorema de Bayes, se puede descomponer la probabilidad condicional de la siguiente manera:

$$p(C_k | x) = (p(C_k) p(x | C_k)) / (p(x))$$

$p(C_k)$  es la probabilidad *a priori* de la hipótesis.

$p(x | C_k)$  es probabilidad de los datos dada la hipótesis o *likelihood*.

$p(x)$  es la probabilidad de observar los datos, independientemente de la hipótesis especificada o evidencia.

## 2.5 XGBoost

Es un algoritmo de aprendizaje automático basado en un árbol de decisiones y utiliza un marco de potenciación de gradientes (Data Science Team, 2019). Las redes neuronales artificiales superan en su mayoría a otros marcos o algoritmos cuando predicen problemas con texto, imágenes y otros datos no estructurados. XGBoost (*Extreme Gradient Boosting*) es una implementación avanzada del algoritmo de refuerzo de gradiente, y ha demostrado ser un algoritmo de ML muy eficaz, ampliamente utilizado en competiciones de aprendizaje automático y *hackathones*. Tiene un alto poder de predicción y es casi diez veces más rápido que las otras técnicas de *gradient boosting* (Singh, 2018). También incluye una variedad de regularización que reduce el sobreajuste y mejora el rendimiento general, de ahí que también se conozca como *técnica de refuerzo regularizado*.

El algoritmo XGBoost fue el resultado de un proyecto de investigación organizado en la Universidad de Washington. Carlos Guestrin y Tianqi Chen presentaron su trabajo en la conferencia SIGKDD del 2016. Desde entonces, XGBoost se ha convertido en un renombrado algoritmo que revolucionó el panorama del aprendizaje de las máquinas. Cabe señalar que permite utilizar una amplia gama de aplicaciones para resolver problemas de predicción, clasificación y regresión definidos por el investigador.

Aparte de los algoritmos descritos, a fin de proveer a la solución que se va a implementar de un enfoque integrador y holístico, se plantea como marco de trabajo el concepto de arquitectura empresarial (AE). Para Goethals et al. (2006), la AE es un acercamiento holístico para el manejo y gestión de una organización, el cual adopta una vista integral que cubre sus procesos de negocio, los sistemas de información, los datos e información y la infraestructura tecnológica. De acuerdo con Lankhorst (2005), la AE es un conjunto coherente de principios, métodos y modelos que se utilizan en el diseño y la realización de la estructura organizacional de una empresa, los procesos de negocio, los sistemas de información y la infraestructura. Finalmente, Arango Serna et al. (2010) explican que el campo de conocimiento de la AE ha evolucionado con el objeto de hacer frente a dos problemas importantes: el primero es la capacidad de gestionar la creciente complejidad de los sistemas de información en las organizaciones, y el segundo se refiere al incremento en la dificultad en la generación de valor real por parte de los sistemas de información para las empresas.

### 3. METODOLOGÍA

#### 3.1 Descripción del conjunto de datos

- Número de observaciones: 41 188.
- Número de muestras realizadas para el presente estudio:
  - Entrenamiento: 28 831 observaciones, correspondientes al 70 % del conjunto de datos.
  - Prueba: 12 357 observaciones, correspondientes al 30 % del conjunto de datos.
  - Número de variables de entrada: 20.
  - Número de variables de salida: 1.

La Tabla 1 muestra las variables del conjunto de datos.

**Tabla 1**  
*Descripción de variables*

Tipo de datos	N.º	Variable	Escala	Tipo de variable
Datos del cliente	1	Edad del cliente	Cuantitativa discreta	Entrada
	2	Ocupación	Cualitativa politómica nominal	Entrada
	3	Estado civil	Cualitativa politómica nominal	Entrada
	4	Educación	Cualitativa politómica nominal	Entrada
	5	Mora bancaria	Cualitativa politómica nominal	Entrada
	6	Crédito hipotecario	Cualitativa politómica nominal	Entrada
	7	Crédito personal	Cualitativa politómica nominal	Entrada
Datos últimos de contacto de esta campaña	8	Medio de contacto	Cualitativa dicotómica nominal	Entrada
	9	Mes de último contacto	Cualitativa politómica nominal	Entrada
	10	Día de la semana del último contacto	Cualitativa politómica nominal	Entrada
	11	Duración del último contacto	Cuantitativa discreta	Entrada
Otros datos	12	Contactos en esta campaña	Cuantitativa discreta	Entrada
	13	Número de días desde que fue contactado la última vez	Cuantitativa discreta	Entrada
	14	Número de contactos antes de esta campaña	Cuantitativa discreta	Entrada
	15	Resultado de campañas previas	Cualitativa politómica nominal	Entrada
Datos del contexto social y económico	16	Tasa de variación del empleo	Cuantitativa continua	Entrada
	17	Índice de precios al consumo	Cuantitativa continua	Entrada
	18	Índice de confianza del consumidor	Cuantitativa continua	Entrada
	19	Índice euribor a 3 meses	Cuantitativa continua	Entrada
	20	Número de empleados	Cuantitativa discreta	Entrada
Clase	21	Cliente suscribió depósito en cuenta	Cualitativa dicotómica nominal	Salida

**3.2 Preparación de los datos**

Teniendo en cuenta que el número de clientes que aceptan campañas de *telemarketing* es mucho menor que aquellos que las rechazan, el *dataset* se encontraba desbalanceado. Por este motivo fue necesario balancearlo utilizando la técnica



estadística de sobremuestreo.

Debido a que varios de los modelos de clasificación utilizados requieren operar con datos numéricos en lugar de cadenas de caracteres, también se convirtieron los atributos de tipo cualitativo dicotómico o politómico a escala numérica. Se filtraron adicionalmente en el conjunto de datos final atributos invariables de contexto social y económico.

4. PRUEBAS

A partir del conjunto de datos definidos, se implementaron los modelos de *decision tree*, KNN, *naive Bayes*, ANN (RPROP) y XGBoost. Estos modelos fueron seleccionados con el objetivo de obtener el modelo predictivo de la mayor exactitud posible.

La Tabla 2 muestra la matriz de confusión de los cinco modelos desarrollados y sus respectivas gráficas en la Figura 2. Por su parte, en la Tabla 3 se aprecian las métricas de desempeño de los algoritmos evaluados. En ambos casos, la información se obtuvo luego de balancear el *dataset* de origen.

Tabla 2

Matriz de confusión de los algoritmos *decision tree*, KNN, *naive Bayes*, ANN (RPROP) y XGBoost

	Algoritmo <i>decision tree</i>		Algoritmo KNN		Algoritmo <i>naive Bayes</i>		Algoritmo ANN (RPROP)		Algoritmo XGBoost	
	No	Sí	No	Sí	No	Sí	No	Sí	No	Sí
No	9950	966	9112	1804	9656	1260	8963	1953	10 071	788
Sí	589	10 424	224	10 789	6477	4536	1557	9456	663	10 407

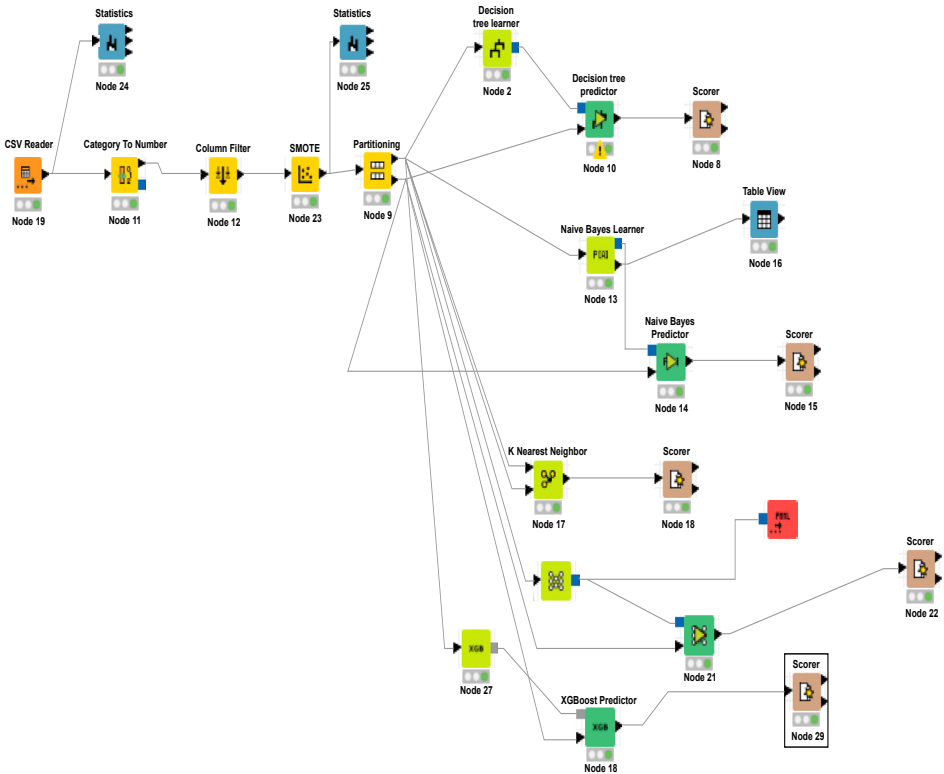
Tabla 3

Resultados de métricas de desempeño de los algoritmos *decision tree*, KNN, *naive Bayes*, ANN (RPROP) y XGBoost

	Algoritmo <i>decision tree</i>	Algoritmo KNN	Algoritmo <i>naive Bayes</i>	Algoritmo ANN (RPROP)	Algoritmo XGBoost
Accuracy	0,929	0,908	0,647	0,840	0,934
Recall	0,947	0,980	0,412	0,859	0,940
Precision	0,915	0,857	0,783	0,829	0,930
F-measure	0,931	0,914	0,540	0,843	0,935

Figura 3

Modelo del flujo de información analizada



5. ANÁLISIS DE RESULTADOS

En este artículo se realizó una revisión completa de los algoritmos *decision tree*, KNN, *naive Bayes*, ANN (RPROP) y XGBoost, así como su aplicación para la predicción de decisiones en la priorización de contactos con clientes en campañas comerciales para la venta del producto de depósitos a plazo. Se descubrió que los algoritmos de *machine learning* tienen diferentes aproximaciones teóricas y formas de aplicación. En términos generales, los modelos *decision tree*, KNN, *naive Bayes*, ANN (RPROP) y XGBoost son sensibles a los datos desequilibrados y son adecuados para muestras de gran tamaño.

Para evaluar el desempeño de los cinco algoritmos en análisis, se seleccionaron las métricas de *accuracy*, *recall*, *precision* y la medida F1 para la clase objetivo, es decir, aquellos clientes de interés para el banco por ser potenciales ahorristas y

que estarían dispuestos a aceptar la campaña. Los resultados de esta evaluación se muestran en la Tabla 3.

Para la métrica *accuracy*, el algoritmo XGBoost es el que muestra un mayor valor con 0,934, seguido de *decision tree* con 0,929; luego KNN con 0,908; ANN (RPROP) con 0,840; y, finalmente, *naive Bayes* con 0,647. Para la métrica *recall*, el algoritmo KNN obtiene el mayor valor con 0,980, seguido por *decision tree* con 0,947; luego ANN (RPROP) con 0,859; y, finalmente, *naive Bayes* con 0,412. En caso de la métrica *precision*, el algoritmo XGBoost obtiene 0,930, seguido de *decision tree* con 0,915; luego KNN con 0,857; luego ANN (RPROP) con 0,829; y, finalmente, *naive Bayes* con 0,783. Para la métrica F1, el valor mayor es del algoritmo XGBoost con 0,935; luego *decision tree* con 0,931, seguido de KNN con 0,914; luego ANN (RPROP) con 0,843; y, finalmente, *naive Bayes* con 0,540. Asimismo, para llevar a la práctica el modelo, se implementó una aplicación web en la nube, la cual fue probada como piloto y dio los resultados esperados.

En la siguiente sección, se presentan recomendaciones para extender y masificar este modelo, integrándolo a los procesos y aplicativos de una institución financiera.

## 6. FUTUROS TRABAJOS

Uno de los objetivos de los sistemas de información es su alineamiento con las estrategias de negocios, y la arquitectura empresarial es la disciplina que permite a la organización que sus recursos de tecnologías de la información respondan adecuadamente a las fuerzas disruptivas de su entorno. La arquitectura empresarial define un marco de trabajo con una organización lógica en capas que describe los diferentes elementos de los procesos de negocio y tecnologías de la información (Saleem & Fakieh, 2020). En este sentido, consideramos que un análisis bajo el enfoque de la arquitectura empresarial es necesario para identificar los siguientes pasos en la evolución de soluciones de campañas de *marketing* en instituciones financieras, para extender el modelo más allá de los modelos algorítmicos de *machine learning*, a fin de integrar este componente con las demás aplicaciones y procesos de la institución que se relacionan con la comercialización de productos bancarios.

En la capa de información y datos, se puede evaluar el incorporar la captura de variables relacionadas con el uso y preferencias de los usuarios actuales que aceptaron la campaña y compraron el producto, de manera que se refinan los modelos para facilitar la venta recurrente a los mismos clientes, con campañas y atributos personalizados.

En la capa de aplicación, se puede analizar la manera de integrar la solución con otros canales de la institución y las aplicaciones que las soportan. De manera que el proceso de comercialización no esté solamente integrado a campañas de *telemarketing* telefónico, sino que, durante el uso de los canales transaccionales, y de manera contextual, los clientes puedan recibir propuestas comerciales que eventualmente sean calificadas y desembolsadas en línea.

Finalmente, en la capa de infraestructura, si bien las instituciones financieras usualmente han favorecido la utilización de recursos *on-premise*, las opciones en nube permiten acceder a recursos para atender integralmente un proceso de campañas. La habilitación de elementos de integración a las aplicaciones *core* financieras permitiría que una capa de campañas que resida en la nube pueda acceder a la información para identificar eventos que gatillen las ofertas a los clientes. La utilización de recursos como *data lakes* de *marketing* permitiría consolidar masivamente la información de manera que pueda ser aprovechada en múltiples usos posteriores, sin necesidad de una transformación previa (Miloslavskaya & Tolstoy, 2016).

## 7. CONCLUSIONES

La clasificación de datos mediante técnicas de *machine learning* se puede utilizar para mejorar la eficacia en la toma de decisiones de los responsables de las áreas comerciales en las entidades financieras, según las variables seleccionadas y sus ponderaciones.

Dado que el conjunto de datos de entrada es desbalanceado, por el menor ratio de clientes que acepta la campaña comercial respecto a los que la rechazan, fue necesario realizar un balanceo de los datos utilizando la técnica SMOTE (Martínez Heras, 2020).

Sobre la base del desempeño de la métrica *accuracy*, se recomienda el algoritmo XGBoost para las predicciones en este tipo de campañas comerciales.

Teniendo en cuenta que el ratio de éxito actual de las campañas que realiza el banco es del 11 % (4640 clientes de la muestra total), el modelo recomendado incrementa este ratio al 94 % de acuerdo con la métrica *recall*, que mide la capacidad de identificar a aquellos clientes que realmente están interesados en los productos ofertados.

## REFERENCIAS

- Arango Serna, M. D., Londoño Salazar, J. E., & Zapata Cortés, J. A. (2010). Arquitectura empresarial: una visión general. *Revista Ingenierías Universidad de Medellín*, 9(16), 101-111.
- Asha, R. B., & Kumar, K. R. (2021). Credit card fraud detection using artificial neural network. *Global Transitions Proceedings*, 2(1), 35-41. <https://doi.org/10.1016/j.gltp.2021.01.006>
- Berrar, D. (2018). Bayes' theorem and naive Bayes classifier. En *Encyclopedia of bioinformatics and computational biology: ABC of bioinformatics* (vol. 1, pp. 403-412). Elsevier. <https://doi.org/10.1016/B978-0-12-809633-8.20473-1>
- Data Science Team. (2019). *XGBoost. ¿Qué es?* <https://Datascience.Eu/Es/Programacion/Xgboost-4/>
- Dutta, S., Bose, P., Goyal, V., & Bandyopadhyay, S. K. (2021). Applying convolutional-GRU for term deposit likelihood prediction. *International Journal of Engineering and Management Research*, 11(3), 265-272. <https://doi.org/10.20944/preprints202007.0101.v1>
- Goethals, F. G., Snoeck, M., Lemahieu, W., & Vandenbulcke, J. (2006). Management and enterprise architecture click: The FAD(E)E framework. *Information Systems Frontiers*, 8(2), 67-79. <https://doi.org/10.1007/s10796-006-7971-1>
- Martínez Heras, J. (2020, 9 de octubre). *Precision, recall, F1, accuracy en clasificación*. IArtificial.net. <https://www.iartificial.net/precision-recall-f1-accuracy-en-clasificacion/>
- Lankhorst, M. (2005). *Enterprise architecture at work. Modelling, communication and analysis*. Springer.
- Miloslavskaya, N., & Tolstoy, A. (2016). Big data, fast data and data lake concepts. *Procedia Computer Science*, 88, 300-305. <https://doi.org/10.1016/j.procs.2016.07.439>
- Moro, S., Laureano, R., & Cortez, P. (2011). Using data mining for bank direct marketing: An application of the CRISP-DM methodology. En P. Novais, J. Machado, C. Analide & A. Abelha (Ed.), *ESM'2011. The 2011 European Simulation and Modelling Conference* (pp. 117-121). EUROISIS.
- Murty, M. N., & Devi, V. S. (2011). *Pattern recognition an algorithmic approach*. Springer.
- Riedmiller, M., & Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The RPROP algorithm. En *IEEE International Conference on Neural Networks* (vol. 1, pp. 586-591). DOI: 10.1109/ICNN.1993.298623

- Rokach, L. (2016). Decision forest: Twenty years of research. *Information Fusion*, 27, 111-125. <https://doi.org/10.1016/j.inffus.2015.06.005>
- Saleem, F., & Fakieh, B. (2020). Enterprise architecture and organizational benefits: A case study. *Sustainability*, 12(19), 8237. <https://doi.org/10.3390/su12198237>
- Singh, A. (2018, 18 de junio). *A comprehensive guide to ensemble learning*. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/>.
- Zhang, C., Zhang, H., & Hu, X. (2019). A contrastive study of machine learning on funding evaluation prediction. *IEEE Access*, 7, 106307-106315. <https://doi.org/10.1109/ACCESS.2019.2927517>