

OCTOBER 2030

APLICACIÓN DE MACHINE LEARNING PARA CAMPANAS DE MARKETING EN LA BANCA COMERCIAL

NOMBRE:

*Valeria Liceth
Triviño Lucas*

tabla de contenido

01 introducción

02 Estado del arte

03 Metodología

04 Pruebas

05 Análisis de resultado

06 Futuro trabajo

07 Conclusiones



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.



Objetivo

- 1

E s 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

- 2

El segundo objetivo es mostrar como demostración una aplicación web que implementa el algoritmo que mejor desempeño tuvo.

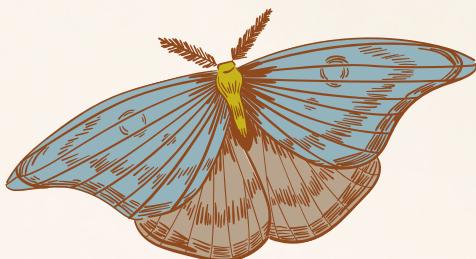
- 3

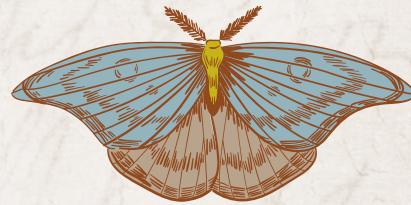
consiste en proponer una arquitectura empresarial futura con visión integral que permita construir una solución integral al proceso de campañas de marketing.



Estado del arte

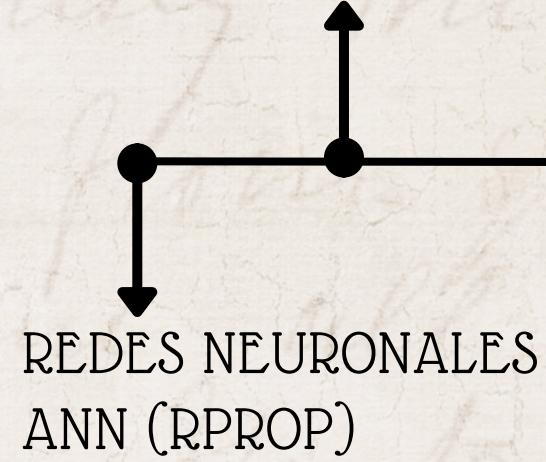
A fin de lograr los tres objetivos planteados, se describe a continuación el estado del arte para los algoritmos de machine learning, así como para el concepto de arquitectura empresarial.





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.

DECISION TREE



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

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.

K-NEAREST NEIGHBORS (KNN)

XGBOOST

Es un algoritmo de aprendizaje automático basado en un árbol de decisiones y utiliza un marco de potenciación de gradientes



metodología

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.



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



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

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

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.

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

		Algoritmo decision tree		Algoritmo KNN		Algoritmo naive Bayes		Algoritmo ANN (RPROP)		Algoritmo XGBoost	
		No	SI	No	SI	No	SI	No	SI	No	SI
No	No	9950	966	9112	1804	9656	1260	8963	1953	10 071	788
SI	SI	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 decision tree	Algoritmo KNN	Algoritmo naive Bayes	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

ANÁLISIS DE RESULTADOS

En este artículo se realizó una revisión completa de los algoritmos decision tree, KNN, naive Bayes, ANN (RPROP) y XGBoost

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,

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

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.



FUTUROS TRABAJOS

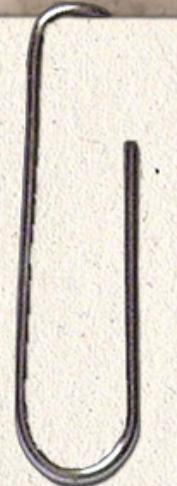
EN LA CAPA DE INFORMACIÓN Y DATOS se puede evaluar el incorporar la captura de variables relacionadas con el uso y preferencias de los usuarios actuales

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

EN LA CAPA DE INFRAESTRUCTURA Las opciones en nube permiten acceder a recursos para atender integralmente un proceso de campañas

WWW.REALLYGREATSITE.COM

Thank
you



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