



Reto | Mercadotecnia telefónica con aprendizaje supervisado

Entregable final del reto

Documento editable por el participante

Nombre del participante: Bernardo Lozano Wise

Indicaciones:

A continuación, se describen los pasos a realizar en la presente actividad, en lo que llamaremos el Plan del Reto:

1. Deberás utilizar el archivo llamado **bank_marketing.csv** con los datos de problema.
2. Utilizar el archivo **bank-names.txt** para obtener información de cada una de las variables.
3. Crear un proyecto tipo Jupyter Notebook en Google-Colab llamado **Solucion_Reto_SC_63_<nombre_y_apellido_del_estudiante>.ipynb**.
4. Incluye las librerías que consideres adecuadas y carga los datos del archivo en una variable llamada "data".
5. Obtener la información de dicha base de datos que incluya el número de registros, el total de variables, el tipo de cada variable, la cantidad de datos perdidos de cada variable en caso de que existan.
6. Transforma las variables categóricas de manera que puedan ser tratadas numéricamente. Justifica si utilizas **LabelEncoder o OneHotEncoder**.
7. Transforma las variables numéricas en los casos que se tenga algún tipo de sesgo.
8. Considera la variable "y" como la variable de salida y el resto de las variables como las variables de entrada.
9. Particiona los datos en los conjuntos de entrenamiento, validación y prueba en **60%, 20% y 20%**, respectivamente.
10. Aplica el modelo **Regresión Logística** en el conjunto de entrenamiento. Valida el modelo con las predicciones del conjunto de validación y su matriz de confusión. Ajusta los parámetros del modelo hasta obtener tu mejor resultado.
11. Aplica el modelo **Red Neuronal** en el conjunto de entrenamiento. Valida el modelo con las predicciones del conjunto de validación y su matriz de confusión. Ajusta los parámetros del modelo hasta obtener tu mejor modelo, entre ellos el número de neuronas y capas ocultas.
12. Selecciona **el mejor modelo encontrado** en los incisos anteriores y utiliza el conjunto de prueba para obtener el desempeño final del modelo y su matriz de confusión.

13. Incluye tus conclusiones del problema, en particular, ¿qué puedes decir acerca del uso de técnicas de inteligencia artificial en problemas de mercadotecnia?
14. Descarga tu script (archivo con extensión .ipynb) y guárdalo siguiendo la nomenclatura que se te indica en Formato de entrega de actividad.
15. Sube el archivo a GitHub.
16. Copia y pega en un archivo de edición de texto la liga de tu archivo.

6.a Justifica el uso de **LabelEncoder** o **OneHotEncoder**.

Se diseñó un plan de preprocesamiento a medida para cada tipo de variable. Se identificaron tres subtipos de variables categóricas, y cada uno recibió un tratamiento diferente:

- **Variables Nominales (job, marital, contact):** Estas variables representan categorías que no tienen un orden o jerarquía intrínseca. Por ejemplo, "management" no es inherentemente "mayor" o "menor" que "technician". OneHotEncoder es la técnica correcta aquí porque crea una nueva columna binaria para cada categoría, permitiendo que el modelo trate cada una como una característica independiente sin asumir una relación ordinal falsa.
- **Variables Ordinales (education, month, poutcome):** A diferencia de las nominales, estas variables tienen un orden lógico claro (ej. primary < secondary < tertiary). OrdinalEncoder fue elegido para mapear estas categorías a valores numéricos secuenciales (ej. 0, 1, 2, 3), preservando así esta valiosa información jerárquica para el modelo.
- **Variables Binarias (default, housing, loan):** Si bien OneHotEncoder funcionaría, es más eficiente y directo mapear estas variables de "sí/no" a valores de 1 y 0. Esto crea una sola columna numérica en lugar de dos, simplificando el dataset sin pérdida de información.

LabelEncoder no se utilizó porque asignaría un número entero arbitrario a cada categoría (ej. job='management' -> 5, job='technician' -> 10). Un algoritmo de ML interpretaría



erróneamente estos números como si tuvieran un orden y una magnitud, aprendiendo una relación falsa (ej. "technician" es el doble de importante que "management"). Esto introduciría un sesgo incorrecto en el modelo. OneHotEncoder evita este problema por completo.

Las variables numéricas se estandarizaron usando StandardScaler. Variables como age y balance operan en escalas muy diferentes. Los algoritmos que dependen de distancias o gradientes (como Regresión Logística, SVM, o la optimización en XGBoost) pueden verse indebidamente influenciados por las variables con magnitudes más grandes. La estandarización transforma todas las variables numéricas para que tengan una media de 0 y una desviación estándar de 1, asegurando que todas contribuyan de manera equitativa al proceso de aprendizaje.

13.a Incluye tus conclusiones de este problema en particular, ¿qué puedes decir acerca del uso de técnicas de inteligencia artificial en problemas de mercadotecnia?

Se concluye que la aplicación de técnicas de inteligencia artificial, como los modelos de clasificación desarrollados, representa una ventaja estratégica fundamental en problemas de mercadotecnia. Al ser capaz de predecir con un 82% de accuracy qué clientes son propensos a suscribir un depósito a plazo, el modelo permite a la institución financiera trascender el marketing masivo para adoptar un enfoque de alta precisión. En lugar de contactar a toda la base de datos, los esfuerzos se pueden concentrar en el segmento de clientes con mayor probabilidad de conversión, lo que se traduce directamente en una optimización de recursos, una reducción significativa de los costos operativos y, en última instancia, en un aumento sustancial del retorno de inversión (ROI) de la campaña. Por lo tanto, el Machine Learning se consolida como una herramienta esencial para transformar datos históricos de clientes en estrategias de mercado más inteligentes, eficientes y rentables.

