**Reto | Mercadotecnia telefónica con aprendizaje supervisado**

*Entregable final del reto*

*Documento editable por el participante*

**Nombre del participante: Hugo Bonifacio Castillo Flores**

**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 OneHotEcoder**.
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 OneHotEcoder**.

|  |
| --- |
| **Transformación de variables categóricas**  Justificación de OneHotEncoder vs LabelEncoder:   * LabelEncoder: se utiliza cuando hay una relación entre los valores de la variable categórica, es decir, cuando los valores representan un orden. * OneHotEncoder: se utiliza cuando no hay relación de orden entre los valores de la variable. |

**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?

|  |
| --- |
| Para este tipo de problemas de mercadotecnia, el uso de técnicas de inteligencia artificial como la regresión logística y las redes neuronales puede ofrecer ventajas para predecir el comportamiento del cliente, permitiendo una toma de decisiones más informada sobre a quién contactar para mejorar la efectividad de las campañas de marketing, sin embargo, siempre es importante tener la opinión de los expertos para determinar que estrategia es la mejor para cada caso.  Ya ajustando los hiperparámetros:   **Regresión logística**: Score:0.816 con el conjunto de validación   * **TN (892)**: 892 clientes fueron correctamente clasificados como "no suscribirán". * **FP (166)**: 166 clientes fueron incorrectamente clasificados como "suscribirán" cuando en realidad no lo hicieron. * **FN (165)**: 165 clientes fueron incorrectamente clasificados como "no suscribirán" cuando en realidad sí lo hicieron. * **TP (577)**: 577 clientes fueron correctamente clasificados como "suscribirán".    **Red neuronal**: Score:0.83 con el conjunto de validación   * **TN (906)**: 906 clientes fueron correctamente clasificados como "no suscribirán". * **FP (152)**: 152 clientes fueron incorrectamente clasificados como "suscribirán". * **FN (154)**: 154 clientes fueron incorrectamente clasificados como "no suscribirán". * **TP (588)**: 577 clientes fueron correctamente clasificados como "suscribirán". |
| Debido al score de validación y que los resultados en la matriz fueron mejores, se utilizó el modelo de red neuronal con los siguientes resultados:   * **Score:** 0.833 con el conjunto de prueba * **TN (867)**: 867 clientes fueron correctamente clasificados como "no suscribirán". * **FP (138)**: 138 clientes fueron incorrectamente clasificados como "suscribirán". * **FN (163)**: 163 clientes fueron incorrectamente clasificados como "no suscribirán".   **TP (632)**: 632 clientes fueron correctamente clasificados como "suscribirán". |