

Práctica 2

En equipo

Sugerencia del Conjunto de datos:

El dataset a utilizar deberá extraerse de las cuentas personales de WhatsApp. Elija un grupo o persona para utilizar la conversación como conjunto de datos base. La unidad muestral a estudiar es el mensaje.

Diccionario de datos

Una vez realizada la extracción, cada registro de la tabla tendrá la siguiente estructura:

{Fecha}, {Hora} - {Autor}: {Mensaje}

Donde:

- Fecha: Indica la fecha en la que se envió el mensaje.
- Hora: Hora de envío del mensaje.
- Autor: Persona que escribió el mensaje.
- Mensaje: Contenido del mensaje.
- Nota: El formato puede variar en función del Sistema Operativo.

Para los siguientes puntos se asume que cada variable se encuentra correctamente separada en columnas.

- Para el conjunto de datos se requieren datos no estructurados (texto, fechas, imágenes, etc)(si puede haber variables continuas en los datos crudos, pero no deben ser mayores al 30%)

Solución Analítica

De acuerdo con el conjunto de datos previamente presentado:

1. Construya:
 - Variable objetivo discreta argumentando su utilidad
 - Variable objetivo continua argumentando su utilidad
2. Construya el mejor modelo para estimar la variable objetivo discreta.
3. Construya el mejor modelo para estimar la variable objetivo continua.
4. Realice la interpretación correspondiente en función del negocio.

Entregables:

El capitán de cada equipo enviará por correo en archivos sin comprimir:

To: dbhpresidente111@gmail.com

CC: a los miembros del equipo

[los demás integrantes del equipo]

Asunto: **CDD_G33_M2_P2_Equipo_# (ejemplo CDD_G33_M2_P2_Equipo_1)**

Cuerpo del correo:

- Integrantes del equipo (con su respectiva ponderación de participación en la práctica [debe sumar 1])
 - Ejemplo [Caso ideal]
 - Integrante 1 (.20) [+1 por participación]
 - Integrante 2 (.20)
 - Integrante 3 (.20)
 - Integrante 4 (.20)
 - Integrante 5 (.20)
 - Ejemplo [Caso extremo (no recomendable)]
 - Integrante 1 (.90)
 - Integrante 2 (.10)
 - Integrante 3 (0)
 - Integrante 4 (0)
 - Integrante 5 (0)

Breve resumen del proceso de tratamiento de datos y modelación; así como conclusión del problema basado en las métricas de ajuste [tanto en entrenamiento como en test].

Archivos adjuntos:

- Archivo **CDD_G33_M2_P2_Equipo_1.ipynb** Notebook del entrenamiento del modelo con su EDA y tratamiento de variables.
- Se debe entregar una presentación ejecutiva (maximo 3-5 slides) en formato PDF que contenga:

Concepto	Ponderación
Resumen de EDA (Principales Insights [más enfocados a negocio])	2
Justificación de: <ul style="list-style-type: none">● Variable objetivo discreta (1.5)● Variable objetivo continua (1.5)	3
Métricas del Modelo (Criterios de selección del mejor modelo) <ul style="list-style-type: none">● Variable objetivo discreta (1)● Variable objetivo continua (1)	2
Interpretación correspondiente en función del negocio. <ul style="list-style-type: none">● Variable objetivo discreta (2)● Variable objetivo continua (2)	4
Total	11

(los archivos sin comprimir)

Entrega (– ∞ , 21/01/2026 23:59:59]