

Reporte de trabajo en equipo

Reporte de trabajo en equipo

Miembros del equipo:

- - Juan Cantillo - Ingeniería de datos, Ciencia de datos, Análisis de negocio
- - Leonardo Beltrán - Análisis de datos, Tablero de datos, Despliegue y mantenimiento

1. Comprensión de negocio y definición de preguntas (Tarea 1)

Juan lideró la formulación de las dos preguntas de negocio orientadas a la Secretaría de Educación:

- - Caracterización de estudiantes con mayores puntajes globales.
- - Identificación de estudiantes con alta probabilidad de ser elegibles para becas.

Además, definió el enfoque general del proyecto y redactó el plan de acción, alineando los objetivos del análisis con las necesidades del usuario final.

2. Selección, limpieza y alistamiento de datos (Tarea 2)

Juan se encargó de la ingeniería de datos:

- - Descargó la base de datos del portal de Datos Abiertos del ICFES.
- - Configuró los buckets en AWS S3 y el crawler en Glue.
- - Ejecutó las consultas en Athena para extraer los datos del año 2017.
- - Cargó y transformó los datos en un entorno Python para su posterior análisis.
- - Realizó la imputación de datos faltantes, incluyendo reemplazos con moda y etiquetas como "No especificado", además de eliminar registros con fechas de nacimiento erradas.

3. Exploración de datos (Tarea 3)

Leonardo lideró esta etapa mediante el análisis descriptivo.

- - Calculó estadísticas generales: edad, estrato, puntaje global.
- - Generó visualizaciones como histogramas, boxplots y gráficos de barras que muestran diferencias según acceso a computador, estrato y edad.
- - Identificó patrones relevantes para orientar el modelado posterior y segmentar visualmente los datos por factores socioeconómicos.

4. Modelamiento (Tarea 4)

Juan lideró esta tarea desde el rol de Ciencia de Datos:

- - Seleccionó las variables relevantes e hizo la ingeniería de características.
- - Construyó y comparó cuatro arquitecturas de redes neuronales.
- - Evaluó cada una usando métricas como recall, para priorizar la correcta identificación de estudiantes elegibles.
- - Registró los experimentos con MLflow y seleccionó el modelo final de 3 capas (100, 50, 25 neuronas) por su mejor desempeño.

5. Diseño y desarrollo del tablero (Tarea 5)

Leonardo fue el encargado del diseño del wireframe del dashboard y su implementación con Dash.

- - Definió campos de entrada (edad, estrato, puntaje, etc.).
- - Integró el modelo de red neuronal para realizar predicciones desde el frontend.
- - Implementó visualizaciones adicionales como el mapa interactivo por departamentos.
- - Aseguró que el diseño fuera intuitivo y útil para el usuario final.

6. Evaluación del producto (Tarea 6)

Juan lideró la evaluación del proyecto, cruzando los insights del análisis descriptivo y los resultados del modelo con las preguntas de negocio.

- - Validó la coherencia del producto final, argumentando cómo cada parte del proyecto responde a las necesidades de la Secretaría de Educación.

7. Despliegue y mantenimiento (Tarea 7)

Leonardo fue el responsable de:

- - Serializar el modelo entrenado en formato .h5.
- - Integrar el modelo en el backend del dashboard.
- - Empaquetar el sistema completo en un contenedor Docker.
- - Desplegar exitosamente el tablero en AWS, asegurando su accesibilidad a través de la URL pública.

Este reporte refleja la colaboración equitativa y especializada de ambos integrantes, aprovechando sus fortalezas técnicas en roles complementarios para lograr un producto final funcional y alineado con los objetivos del curso.