

Roles, Tareas y Distribución de Puntos

La siguiente tabla resume la distribución de roles, tareas, tiempo dedicado y puntos asignados a cada integrante del grupo:

Integrante	Rol	Tarea Realizada	Tiempo Dedicado (horas)	Algoritmo Trabajado	Puntos Asignados
Juan Manuel Ramírez	Líder de Analítica	Implementación y optimización de XGBoost, validación final del mejor modelo, documentación de métricas.	6	XGBoost	33.33
Juan Diego Osorio	Líder de Proyecto	Implementación y análisis de Regresión Logística, coordinación de entregables, estructuración de la presentación.	6	Regresión Logística	33.33

Daniel Gómez	Líder de Datos	Implementación de Random Forest, limpieza de datos, vectorización con TF-IDF y preprocesamiento.	6	Random Forest	33.33
--------------	----------------	--	---	---------------	-------

Retos enfrentados en el proyecto:

- Uno de los principales retos que enfrentamos fue la coordinación del trabajo en equipo. Cada integrante tenía responsabilidades específicas, pero al tratarse de un proyecto que requería el manejo de datos, desarrollo de modelos analíticos y la comprensión del negocio, en algunos momentos la comunicación se volvió un desafío. Esto se sumó al reto de integrar nuestras diferentes áreas de trabajo, ya que mientras algunos tenían más experiencia en la parte técnica, otros destacaban en la comprensión del negocio. Otro desafío importante fue la selección del mejor algoritmo. Tanto Random Forest como XGBoost presentaron métricas muy similares, lo que hizo difícil decidir cuál utilizar. A pesar de que ambos tenían resultados comparables, fue crucial analizar en detalle sus características y su comportamiento en distintos escenarios para tomar la mejor decisión.

Formas Planteadas para Resolver los Retos

- Para abordar los problemas de coordinación y comunicación, decidimos aumentar la frecuencia de nuestras reuniones de sincronización y definir de forma más clara las responsabilidades de cada miembro en cada fase del proyecto. Esto nos permitió mantenernos alineados y asegurar que todos entendían el progreso y los próximos pasos. En cuanto a la selección del mejor modelo, se optó por XGBoost debido a su mejor desempeño en la detección de noticias falsas. Aunque los otros modelos presentaban resultados similares, XGBoost mostró mayor estabilidad en sus métricas y un mejor equilibrio entre precisión y recall, lo que indicaba una mayor adaptabilidad y precisión en las predicciones.

Uso de ChatGPT en el Proyecto

- Redacción de documentación para estructurar reportes y presentar los resultados de manera clara y organizada.
- Apoyo en preprocesamiento de datos

Reflexión sobre la Repartición de Puntos

En cuanto a la repartición de puntos, consideramos que la contribución de cada miembro del equipo fue equivalente y esencial para proyecto. Aunque cada integrante tuvo roles diferentes y abordó tareas específicas, todos aportaron de manera significativa, complementando las habilidades y conocimientos de los demás. El enfoque del equipo se basó en la colaboración continua. Todos participaron activamente en las discusiones y decisiones clave, y cada miembro aportó desde su área para fortalecer el trabajo. Esto permitió que el proyecto avanzara de manera fluida y que cada etapa se completara con el nivel de detalle requerido.

Referencia al Notebook

En esta parte del documento se hace referencia al notebook principal donde se encuentran detalladas las decisiones tomadas en el preprocesamiento, vectorización y modelado de datos igualmente en la wiki se puede apreciar esto.

Ubicación de cada sección en el Notebook:

1. **Sección 2: Entendimiento y Preparación de los Datos** → Sección: *Carga y Exploración de Datos*.
 - Detalles sobre la limpieza de datos y manejo de valores nulos.
 - Explicación sobre la eliminación de duplicados y conversión de fechas.
2. **Preprocesamiento y Vectorización** → Sección: *Preprocesamiento de Texto*.
 - Proceso de tokenización, lematización y eliminación de stopwords.
 - Creación de la matriz TF-IDF y su justificación.
3. **Sección 3: Modelado y Evaluación** → Sección: *Entrenamiento y Evaluación de Modelos*.
 - Implementación de Regresión Logística, Random Forest y XGBoost.
 - Comparación de métricas de desempeño y optimización de hiperparámetros.

4. **Sección 4: Resultados** → Sección: *Análisis de Resultados y Métricas de Desempeño*.

- Comparación de los modelos en términos de precisión, recall y F1-score.
- Justificación de la priorización en la detección de noticias falsas.

5. **Sección 5: Trabajo en Equipo** → Sección: *Gestión del Proyecto y División de Roles*.

- Descripción de roles y responsabilidades de cada integrante.
- Reflexión sobre la colaboración y distribución del esfuerzo.