

- Juan Miguel Delgado - 202314903
- Rodrigo García - 202212475
- Daniel Gómez – 202215189

## Sección 5. Trabajo en equipo

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 (horas)	Algoritmo trabajado	Puntos asignados
<b>Daniel Gómez</b>	Líder de Proyecto	Coordinación del equipo, definición de fechas y entregables, documentación de Sección 1 (Canvas), revisión final del notebook.	6 h	K-Nearest Neighbors (KNN)	33.3
<b>Juan Miguel Delgado</b>	Líder de Datos	Limpieza y preparación de datos, vectorización con TF-IDF, apoyo en perfilamiento y exportación de resultados a Excel.	6 h	Regresión Logística	33.3
<b>Rodrigo García</b>	Líder de Analítica	Implementación de Árbol de Decisión, evaluación de métricas, análisis de resultados y palabras clave por ODS.	6 h	Decision Tree	33.3

## Retos enfrentados en el proyecto

Durante el desarrollo del proyecto, Uno de los principales retos que enfrentamos fue la coordinación del trabajo en equipos, ya que cada uno tenía responsabilidades específicas y distintos horarios. Esto generó momentos en los que la comunicación se volvió complicada, especialmente al integrar las partes técnicas de preprocesamiento y modelado con la documentación. Otro desafío importante fue el desbalance de clases en los datos, lo que afectó la estabilidad de las métricas y exigió pruebas adicionales para no favorecer únicamente a la clase mayoritaria. Finalmente, la selección del mejor modelo también representó un reto, pues, aunque los algoritmos evaluados arrojaron métricas similares, fue necesario analizar en detalle el comportamiento de cada uno para tomar una decisión fundamentada.

- Juan Miguel Delgado - 202314903
- Rodrigo García - 202212475
- Daniel Gómez – 202215189

### **Formas planteadas para resolver los retos**

Para superar estas dificultades, el equipo decidió realizar reuniones cortas de seguimiento semanal, además de una reunión de planeación inicial y otra final de consolidación, lo cual permitió mejorar la comunicación y la coordinación de tareas. En cuanto al desbalance de clases, se aplicó validación cruzada estratificada y se revisaron los promedios macro de las métricas para asegurar que el modelo seleccionado tuviera un rendimiento equilibrado. Respecto a la elección del mejor algoritmo, se compararon los resultados de cada uno de manera detallada y se priorizó aquel que mostró mayor estabilidad en F1-macro (Logistic Regression).

### **Uso de ChatGPT en el proyecto**

ChatGPT se utilizó principalmente como herramienta de apoyo para estructurar la documentación y organizar las secciones del informe de manera clara y coherente. Además, sirvió como guía para mejorar el preprocesamiento de los datos textuales y estandarizar el flujo de trabajo en el notebook.

### **Reflexión sobre la repartición de puntos**

La repartición de puntos se realizó de manera equitativa, asignando 33.3 a cada integrante, ya que se consideró que todos aportaron de forma significativa al cumplimiento de los objetivos. Aunque cada miembro asumió un rol específico dentro del proyecto, el trabajo en equipo fue colaborativo y constante, lo que garantizó que ninguna tarea quedara desatendida. Todos los integrantes también participaron en la preparación y grabación del video de resultados, asegurando que la presentación final reflejara el esfuerzo conjunto. En conclusión, la división equitativa de puntos representa fielmente el compromiso y la calidad de las contribuciones de cada miembro.

### **Ubicación de cada sección en el Notebook**

**Sección 2: Entendimiento y Preparación de los Datos** → Sección: Carga y Exploración de Datos

- Perfilamiento inicial del dataset Datos\_proyecto.xlsx.
- Limpieza de caracteres especiales, normalización de mayúsculas/minúsculas.
- Manejo de valores nulos y análisis de distribución de clases.
- Preprocesamiento y Vectorización → Sección: Preprocesamiento de Texto

- Juan Miguel Delgado - 202314903
  - Rodrigo García - 202212475
  - Daniel Gómez – 202215189
- 
- Proceso de tokenización, lematización y eliminación de stopwords en español.
  - Representación de los textos mediante TF-IDF, con comparación frente a Bag of Words.
  - Análisis de palabras más frecuentes y proporción de stopwords.

### **Sección 3: Modelado y Evaluación** → Sección: Entrenamiento y Evaluación de Modelos

- Implementación de K-Nearest Neighbors, Decision Tree y Logistic Regression.
- Búsqueda de hiperparámetros con GridSearchCV y validación cruzada K-Fold.
- Comparación de métricas (Accuracy, Precision, Recall, F1-score) y matrices de confusión.

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

- Comparación de los modelos en términos de desempeño global y balance entre clases.
- Selección del modelo final (Logistic Regression) por estabilidad y mayor F1-macro.
- Análisis de palabras más representativas por ODS (1, 3, 4).
- Exportación del archivo final datos\_prueba\_clasificados.xlsx.

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

- Descripción de los roles y responsabilidades de cada integrante del grupo.
- Retos enfrentados, soluciones implementadas y reflexión sobre la distribución del esfuerzo.