

1. Introducción

Este informe presenta un análisis completo sobre la **predicción de la calidad del vino tinto** utilizando algoritmos de aprendizaje supervisado. El conjunto de datos contiene características fisicoquímicas del vino (como acidez, azúcares, sulfatos, alcohol, entre otros) y una variable objetivo multiclase que representa su calidad, medida en una escala de 3 a 8.

2. Técnicas Utilizadas

Se aplicaron las siguientes técnicas y herramientas en el análisis:

- **Análisis exploratorio de datos (EDA)**
 - **Preprocesamiento:**
 - Eliminación de columnas irrelevantes
 - Tratamiento de duplicados
 - Escalamiento de variables numéricas
 - **División del dataset en conjunto de entrenamiento (80%) y prueba (20%)**
 - **Modelado supervisado con:**
 - K-Nearest Neighbors (KNN)
 - Random Forest (RF)
 - Regresión Logística (RL)
 - **Evaluación de modelos mediante:**
 - Precisión (Accuracy)
 - Matrices de confusión
-

3. Resultados Obtenidos

A continuación, se presentan las precisiones obtenidas por cada modelo:

- **KNN (n=10):** ~0.58
- **Random Forest (n=50):** ~0.64
- **Regresión Logística:** ~0.53

El modelo que mejor desempeño logró fue **Random Forest**, que superó al resto tanto en precisión como en estabilidad. Le siguió KNN, mientras que la Regresión Logística fue el modelo con menor rendimiento, posiblemente por no ajustarse bien a la naturaleza multiclase y no lineal del problema.

4. Análisis Comparativo

- **Random Forest** demostró mayor robustez frente a la variabilidad del dataset, gracias a su enfoque de ensamble.
- **KNN** se comportó de forma aceptable, aunque dependiente del número de vecinos y muy sensible a la escala de los datos.
- **Regresión Logística** no logró modelar adecuadamente la complejidad del problema.
- Las **matrices de confusión** indicaron que los errores más frecuentes se daban entre clases cercanas (por ejemplo, predecir un 6 cuando el verdadero valor era 5 o 7), lo que sugiere que hay cierta continuidad en la variable objetivo.

5. Conclusiones

- **Random Forest** es el modelo más apropiado para este problema, ofreciendo una buena combinación entre precisión y estabilidad.
- El problema podría explorarse también como una **clasificación ordinal** o incluso una **regresión**, dado que las clases representan niveles de calidad que siguen una escala natural.
- Se recomienda para futuras versiones del proyecto:
 - Incorporar **modelos más avanzados** como XGBoost o LightGBM.
 - Realizar análisis de **importancia de variables** para interpretar el modelo.
 - Aplicar técnicas de **balanceo de clases**, si se detecta desbalance.
 - Generar una API para integrar este modelo en una aplicación de evaluación automática de calidad de vino.