

# Informe Final: Predicción y Generación de Insights sobre Calidad de Vinos

## Proyecto Integrador de MLOps

Universidad del Magdalena – Noviembre 2025

Autores: Juan Mosquera, Anderson Bornachera

---

### 1. Objetivos

El presente proyecto tiene como objetivo general desarrollar un pipeline completo de MLOps que integre:

- Experimentación reproducible con MLflow Tracking.
- Registro y versionamiento de modelos mediante MLflow Model Registry.
- Una interfaz interactiva con Gradio para usuarios finales.
- Generación automática de explicaciones interpretativas usando IA generativa (Gen AI) con un modelo local (Ollama + Llama3).

Los objetivos específicos incluyen:

- Realizar al menos tres ejecuciones de entrenamiento con distintos hiperparámetros.
  - Seleccionar el mejor modelo basado en métricas (RMSE,  $R^2$ ) y promoverlo a producción.
  - Demostrar cómo la IA generativa puede complementar la comunicación de resultados de modelos de ML.
  - Reflexionar críticamente sobre los riesgos éticos del uso de LLMs en la interpretación de predicciones.
- 

### 2. Metodología

#### Dataset y preprocesamiento

Se utilizó el dataset público *Wine Quality White* de la UCI Machine Learning Repository, compuesto por 4,898 vinos blancos descritos por 11 atributos químicos (ej. acidez, alcohol, sulfatos) y una variable objetivo numérica: calidad (escala 0–10). No se realizaron transformaciones complejas; se usó el conjunto de datos tal como se proporciona, con división estratificada en entrenamiento (80%) y prueba (20%).

#### Modelo de machine learning

Se implementó un Random Forest Regressor por su capacidad para capturar relaciones no lineales y su robustez frente a colinealidades. Se probaron tres configuraciones de hiperparámetros:

1. `n_estimators=50, max_depth=5`
2. `n_estimators=100, max_depth=10`
3. `n_estimators=200, max_depth=None`

El entrenamiento, evaluación y registro se realizaron con MLflow, registrando parámetros, métricas (RMSE, MAE,  $R^2$ ), artefactos (importancia de features) y el modelo serializado.

#### Registro y producción


El mejor modelo (configuración #2: RMSE = 0.62,  $R^2$  = 0.61) fue registrado en MLflow Model Registry bajo el nombre WineQualityModel, primero en etapa *Staging* y luego promovido a *Production*.

#### Interfaz y Gen AI


Se construyó una interfaz con Gradio que carga el modelo en producción, permite ingresar valores o subir un CSV, y muestra la predicción. Además, al hacer clic en “Generar explicación”, se invoca a Ollama (modelo llama3) para crear una justificación textual, la cual se registra como artefacto en MLflow.

---

### 3. Resultados

- Se completaron 3 experimentos con MLflow Tracking, comparables en la UI.
- El mejor modelo alcanzó RMSE = 0.62 y  $R^2$  = 0.61, suficiente para una estimación realista de calidad.
- El modelo fue registrado y promovido exitosamente a Production en MLflow Model Registry.
- La interfaz de Gradio funciona localmente y permite predicciones interactivas.
-  Las explicaciones generadas por IA son coherentes y técnicamente plausibles, ej.:

*"Este vino tiene alta calidad porque presenta un alto contenido de alcohol (10.5%) y baja acidez volátil (0.27 g/dm<sup>3</sup>), lo que sugiere un buen equilibrio y estabilidad."*

-  Cada explicación se almacena como artefacto en MLflow (genai\_explanation.txt).

---

### 4. Reflexiones Éticas sobre el Uso de Gen AI

El uso de modelos de lenguaje grandes (LLMs) para explicar predicciones introduce beneficios y riesgos éticos:

#### Beneficios

- Mejora la interpretabilidad para usuarios no técnicos (ej. enólogos, productores).

- Humaniza la salida del modelo, facilitando la toma de decisiones.

#### Riesgos y sesgos

1. Alucinación: Los LLMs pueden inventar relaciones causales inexistentes (ej. atribuir calidad a "origen geográfico" aunque no esté en los datos).
2. Sobreconfianza: Los usuarios pueden confiar ciegamente en explicaciones que suenan razonables pero carecen de base estadística.
3. Falta de trazabilidad: A diferencia de SHAP o LIME (métodos post-hoc clásicos), las explicaciones de LLMs no son reproducibles ni auditables de forma determinista.
4. Refuerzo de sesgos: Si el LLM fue entrenado con datos sesgados (ej. "vino europeo = mejor"), podría reflejar estereotipos no presentes en el modelo original.

#### Conclusión ética

La Gen AI debe usarse como complemento, no como sustituto, de métodos de explicabilidad tradicionales. Es esencial transparencia: informar al usuario que la explicación es generada por IA y no proviene directamente del modelo de ML. Además, siempre se debe validar que las explicaciones sean coherentes con la lógica del dominio.