

Wine Classification Simulator

Machine Learning + Exploratory
Structural Analysis

Lautaro Silvestri

02/03/2026



Problem Statement

Dataset: Wine (UCI)

3 clases de vino (1, 2 y 3)

13 compuestos químicos

Objetivo: Clasificación y comprendimiento de diferencias estructurales entre vinos tintos

Localización: Piamonte, Italia



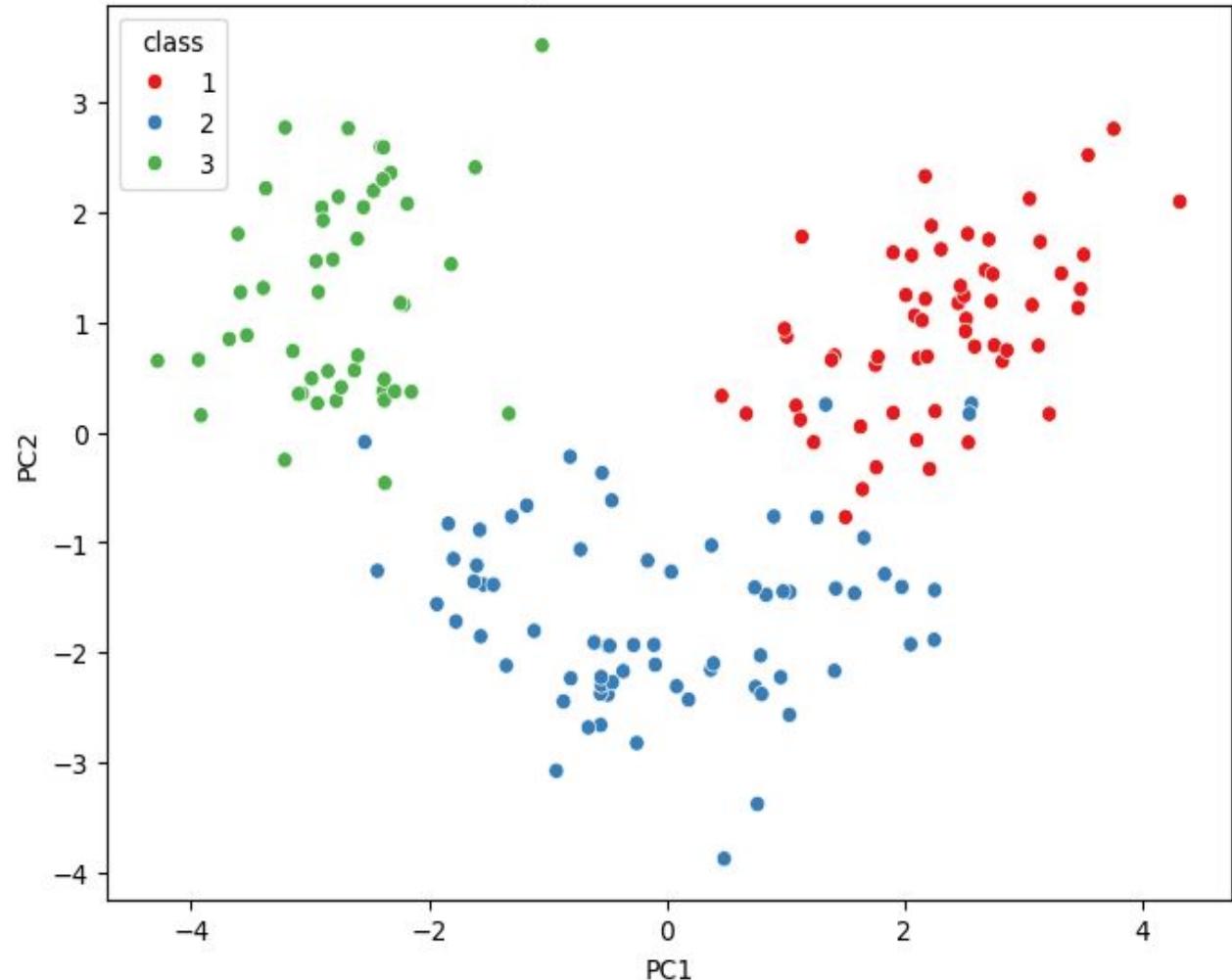
Análisis Estructural

Se creó un perfil promedio por clase

Discriminación de cada una de las variables → F_score y ANOVA Analysis

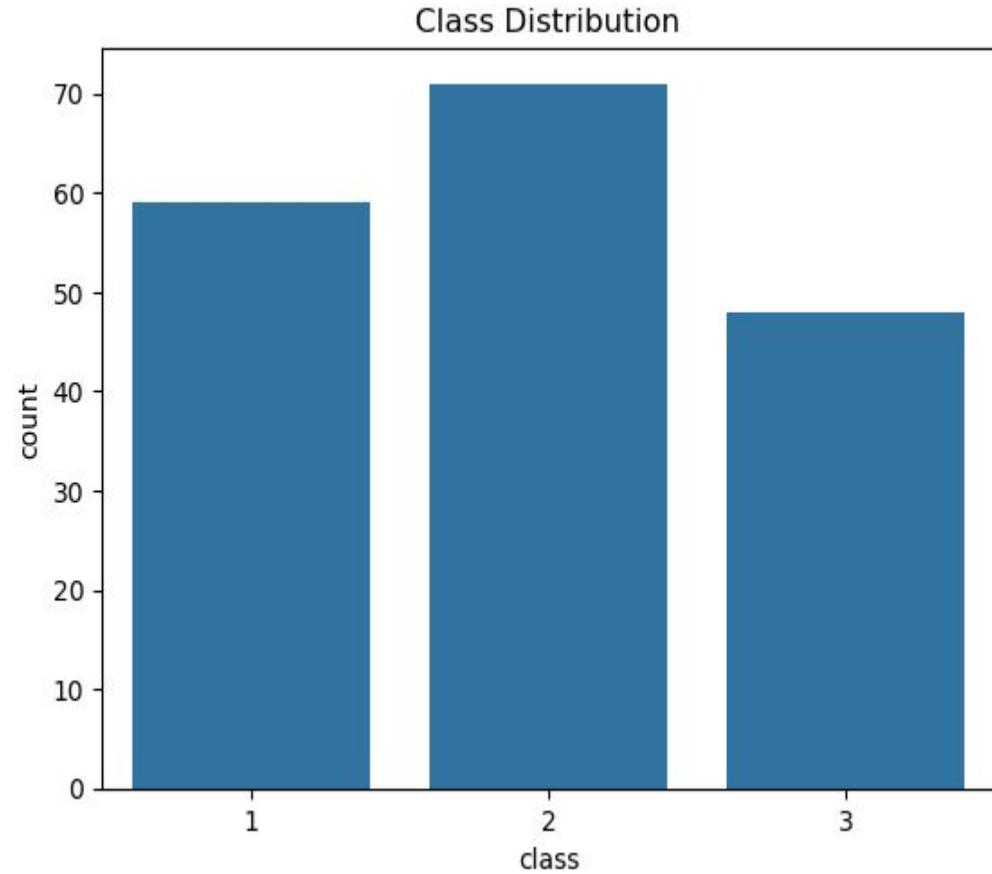
Proyección PCA para visualizar la separación entre clases

PCA - Proyección de las clases en 2D



Análisis de distribución de clases

- Clase 2: 39.88% de la muestra total
- Clase 1: 33.14%
- Clase 3: 26.98%



Variables clave

Top variables (según ANOVA):

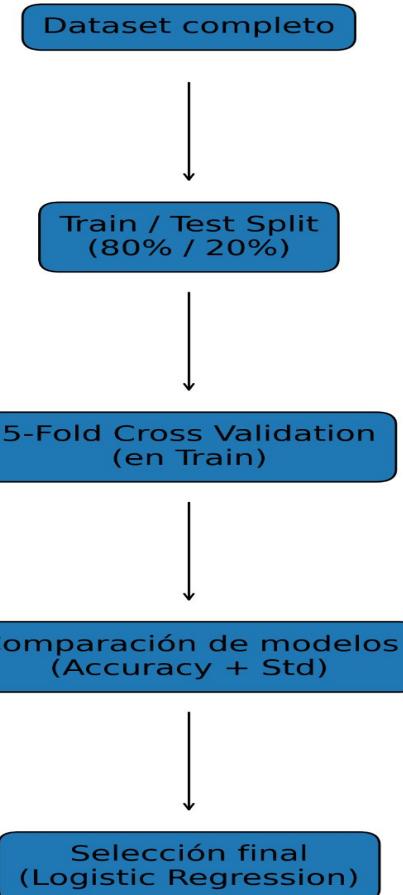
- Flavanoides (compuestos vegetales)
- Prolina (aminoácidos)
- OD280/OD315 (parámetro de luminiscencia)
- Alcohol (graduación alcohólica)
- Intensidad del color



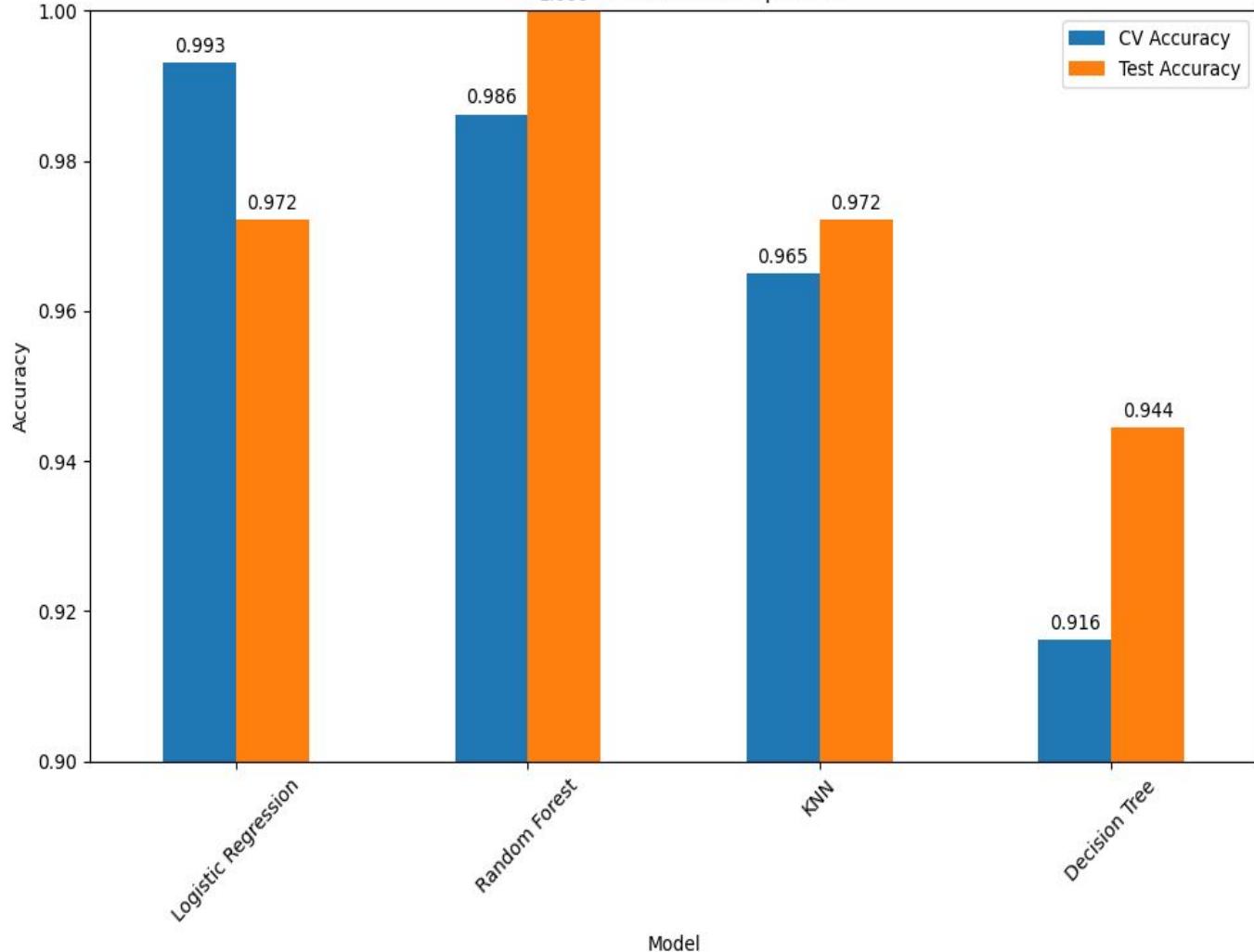
Evaluación de modelos

Proceso de validación:

- División Train/Test (80% / 20%)
- 5-Fold Cross Validation sobre el conjunto de entrenamiento
- Métrica principal: Accuracy
- Evaluación de:
 - Media de accuracy
 - Desviación estándar (estabilidad)
 - Complejidad del modelo
 - Interpretabilidad



Model Performance Comparison



Se selecciona Logistic Regression por ofrecer:

- Alto rendimiento
- Mayor estabilidad
- Menor complejidad
- Alta interpretabilidad

From Model to Interactive Tool

Creación de una app con las siguientes características:

- Gráficas para ajustar la proporción deseada de compuestos
- Botón de “Random realistic wine”
- Visualización de las probabilidades en función de la mezcla
- Interpretación de la clase:
 - Clase 1: Perfil equilibrado, buena estructura fenólica y balance general,
 - Clase 2: Perfil más intenso, mayor color intensity y carácter más marcado,
 - Clase 3: Perfil estructurado, alto contenido de flavanoides y prolina.