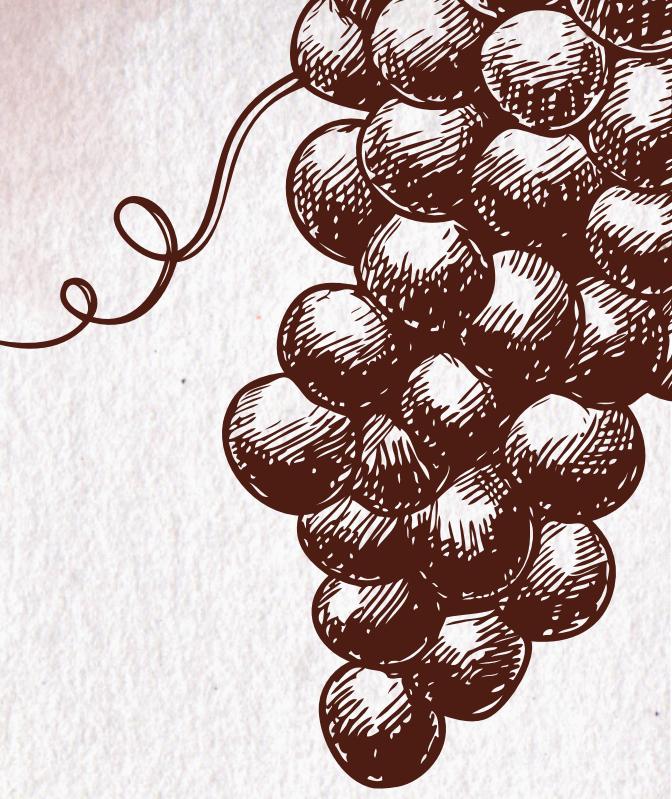


WINe



Íñigo Pascual Aguirre



"Vini, Vidi, Vino"

ÍNDICE

1 Introducción

2 Datos y características

3 Primeras visualizaciones

4 Correlación

5 Modelos Supervisados

6 Modelo No Supervisado

7 Métricas

8 Conclusiones



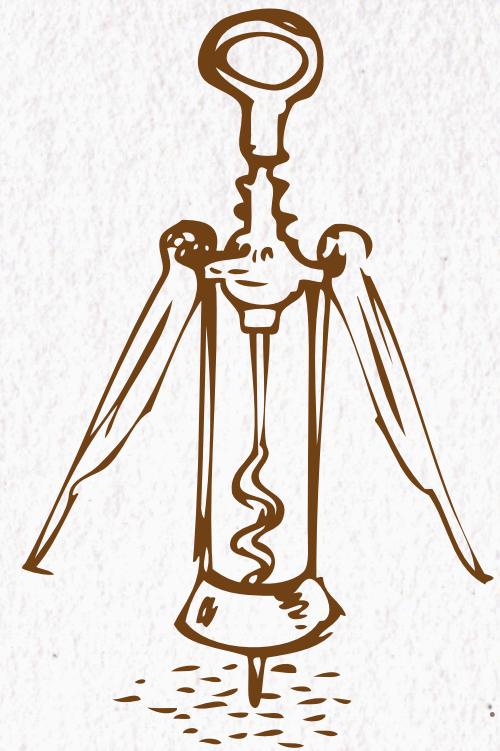


INTRODUCCIÓN

Este proyecto nace del amor por el vino y la búsqueda de la excelencia en su obtención

DATOS Y CARÁCTERÍSTICAS

- ONCE ATRIBUTOS Y UN TARGET
- IDENTIFICACIÓN DE OUTLIERS Y DESBALANCE EN EL TARGET
- ALGORITMO SUPERVISADO: CLASIFICACIÓN (MULTICLASE)





Alcohol

Mide el porcentaje de alcohol en volumen presente en el vino. El alcohol es un factor primario que influye en el sabor, cuerpo y complejidad del vino. Niveles más altos de alcohol generalmente resultan en vinos de cuerpo más completo, con aromas y sabores más intensos.

Sulfatos

Mide la concentración de sales de sulfato, principalmente sulfato de potasio, en el vino. Los sulfatos están presentes de forma natural en las uvas y también pueden añadirse durante los procesos de vinificación. Niveles excesivos de sulfatos pueden contribuir a un sabor amargo o metálico en el vino.

KEY FEATURES

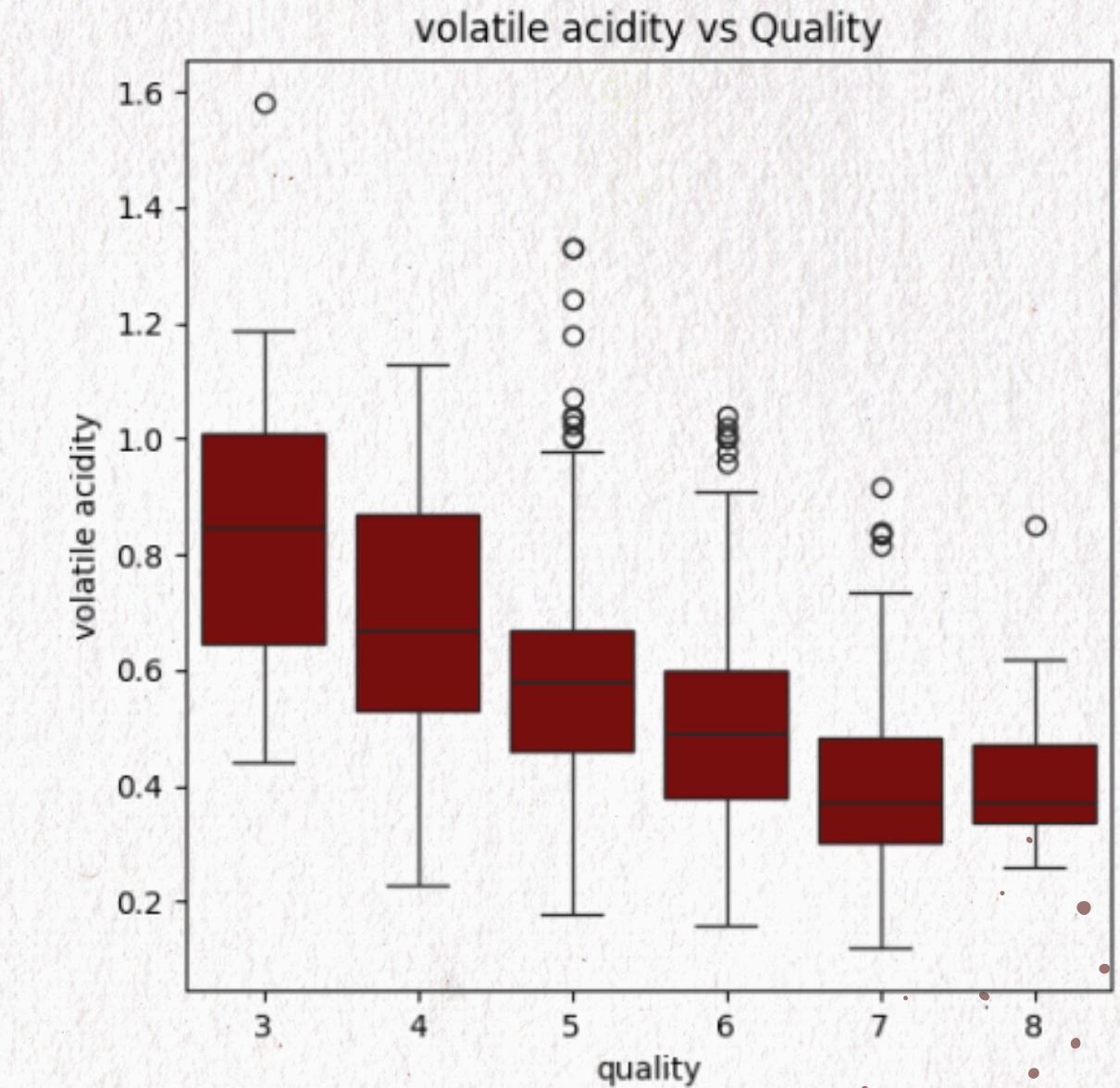
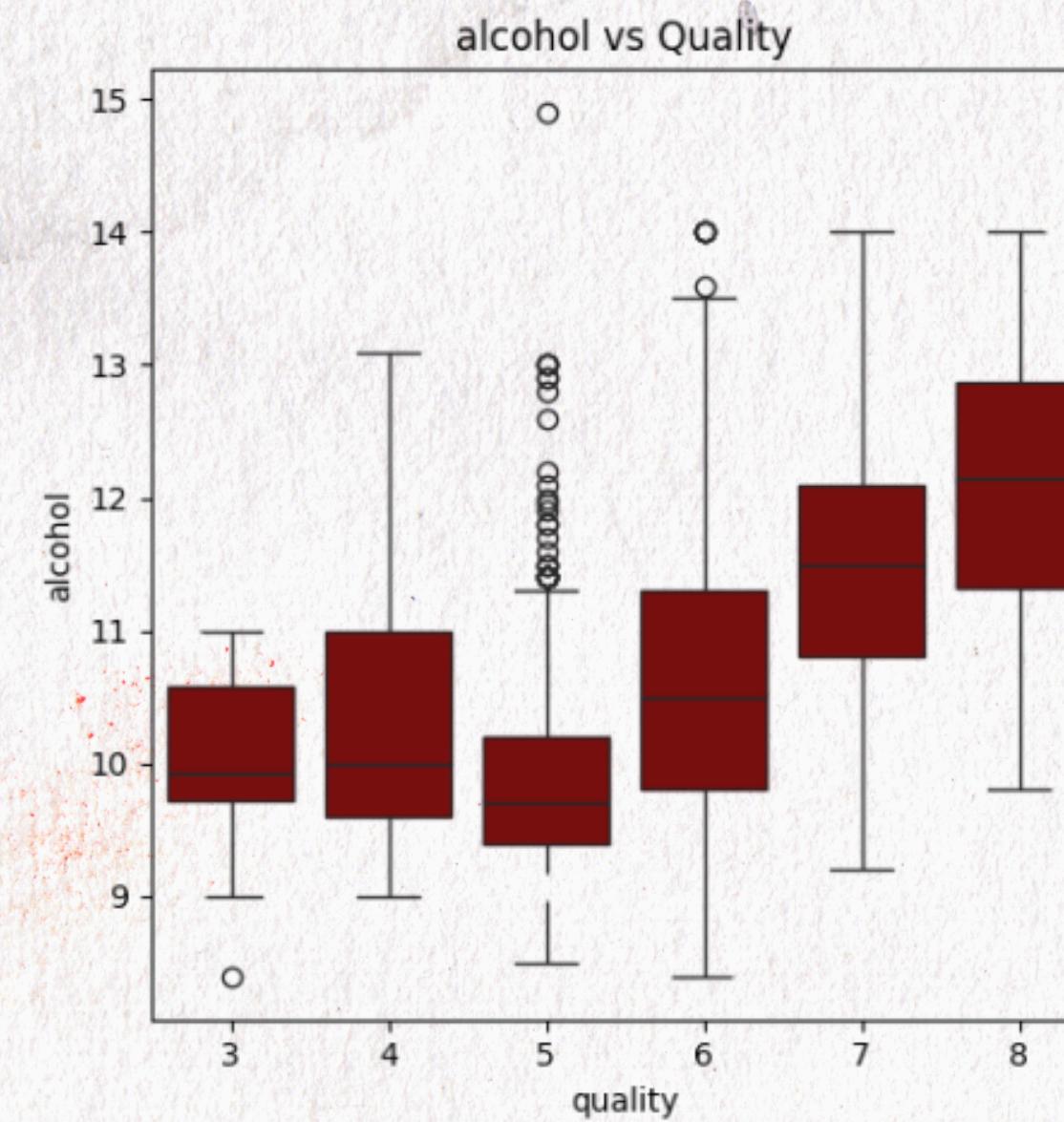
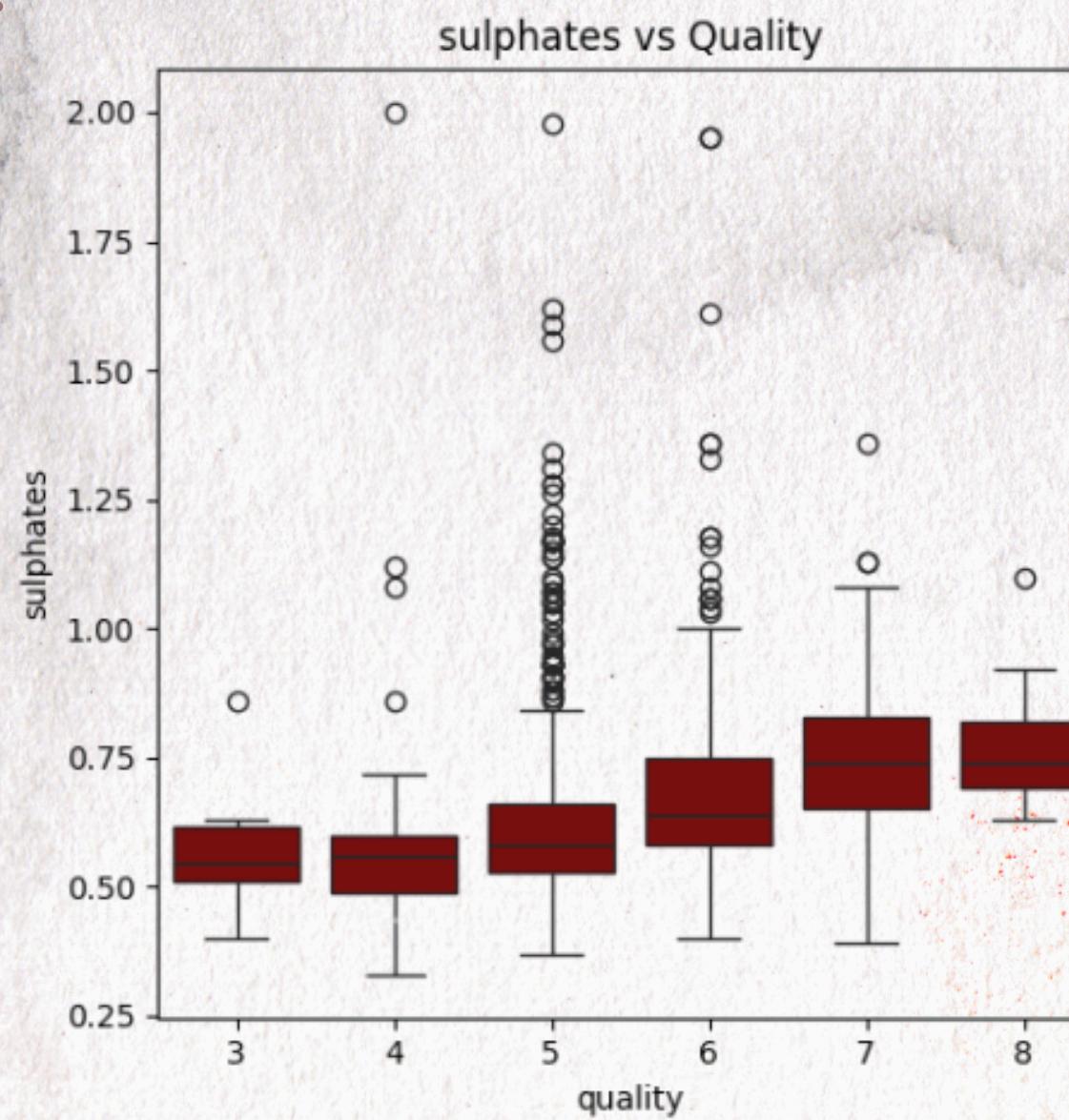
Acidez Volátil

Mide la concentración de ácidos volátiles, principalmente ácido acético, en el vino. Un exceso de acidez volátil, a menudo causado por la degradación bacteriana, puede producir un aroma y sabor desagradables similares al vinagre.

Ácido Cítrico

Mide la cantidad de ácido cítrico, otro componente natural del jugo de uva, presente en el vino. El ácido cítrico contribuye a la acidez y frescura del vino, influyendo en la percepción general de su sabor.

VISUALIZACIONES



- Alcohol (porcentaje en volumen): Mayor en vinos de alta calidad.

- Acidez volatil (g/L): Menor en vinos de alta calidad.

- Sulfatos (g/L): Mayor en vinos de alta calidad.

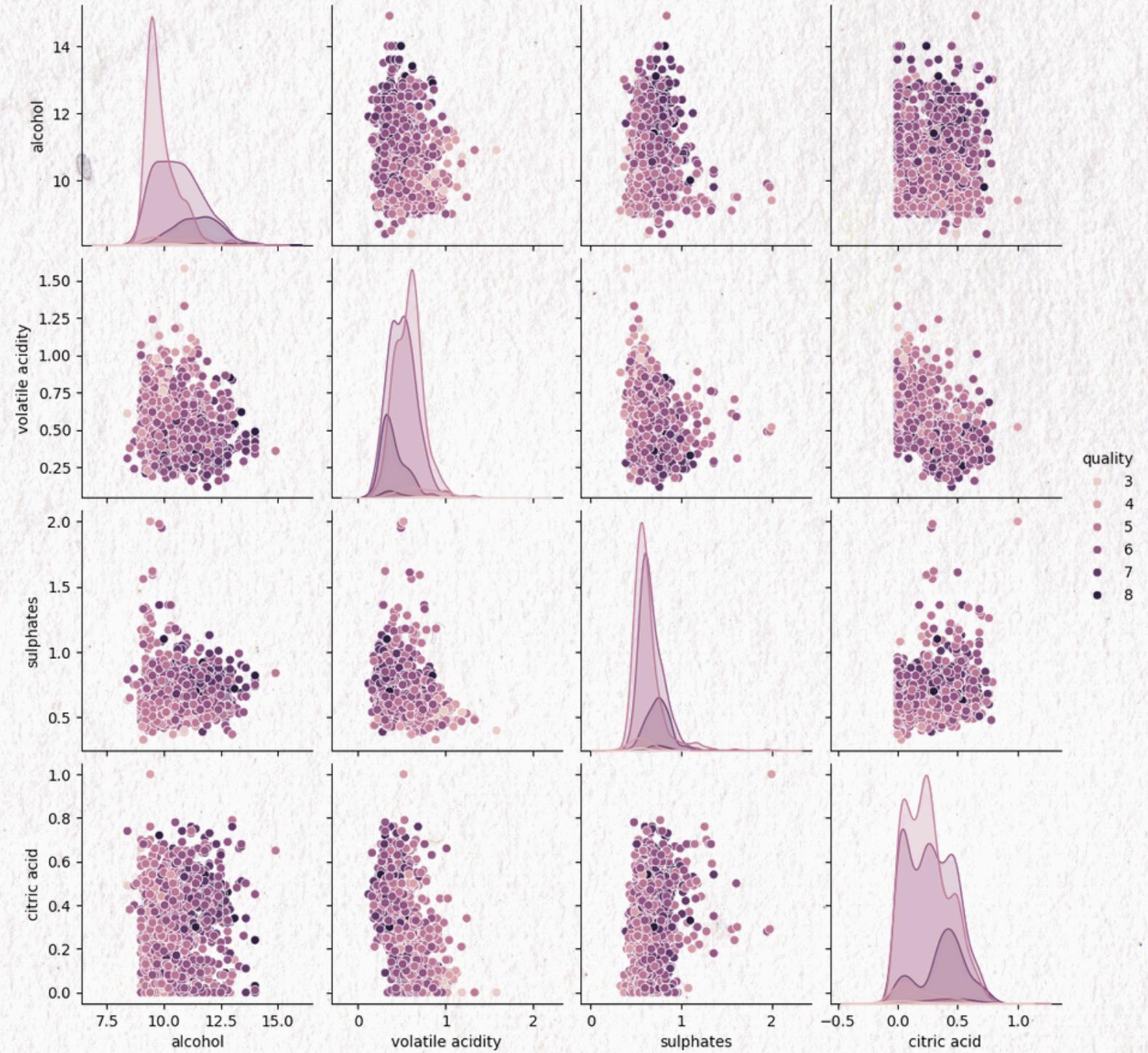
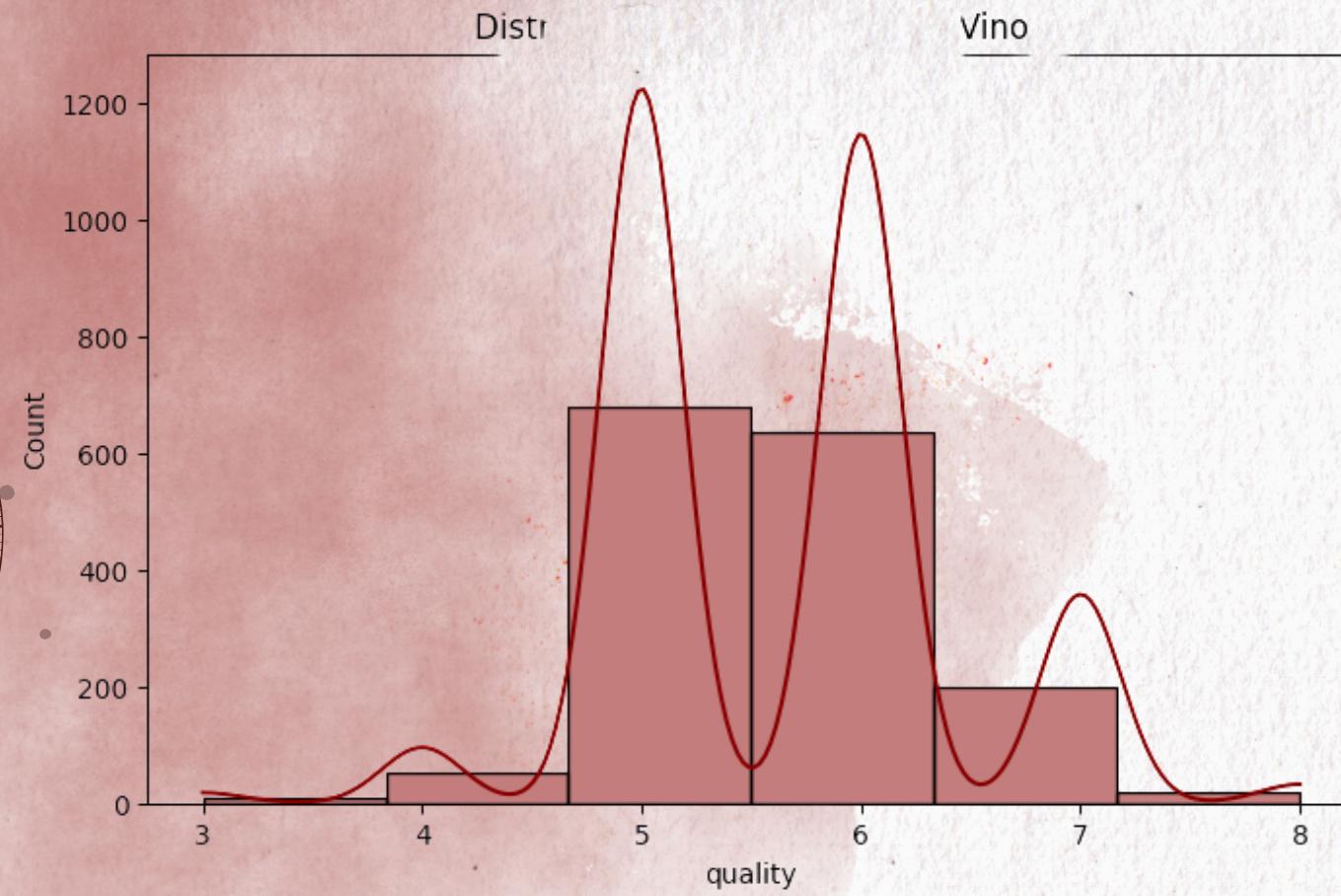
ACCIONES

ELIMINACIÓN DE
OUTLIERS

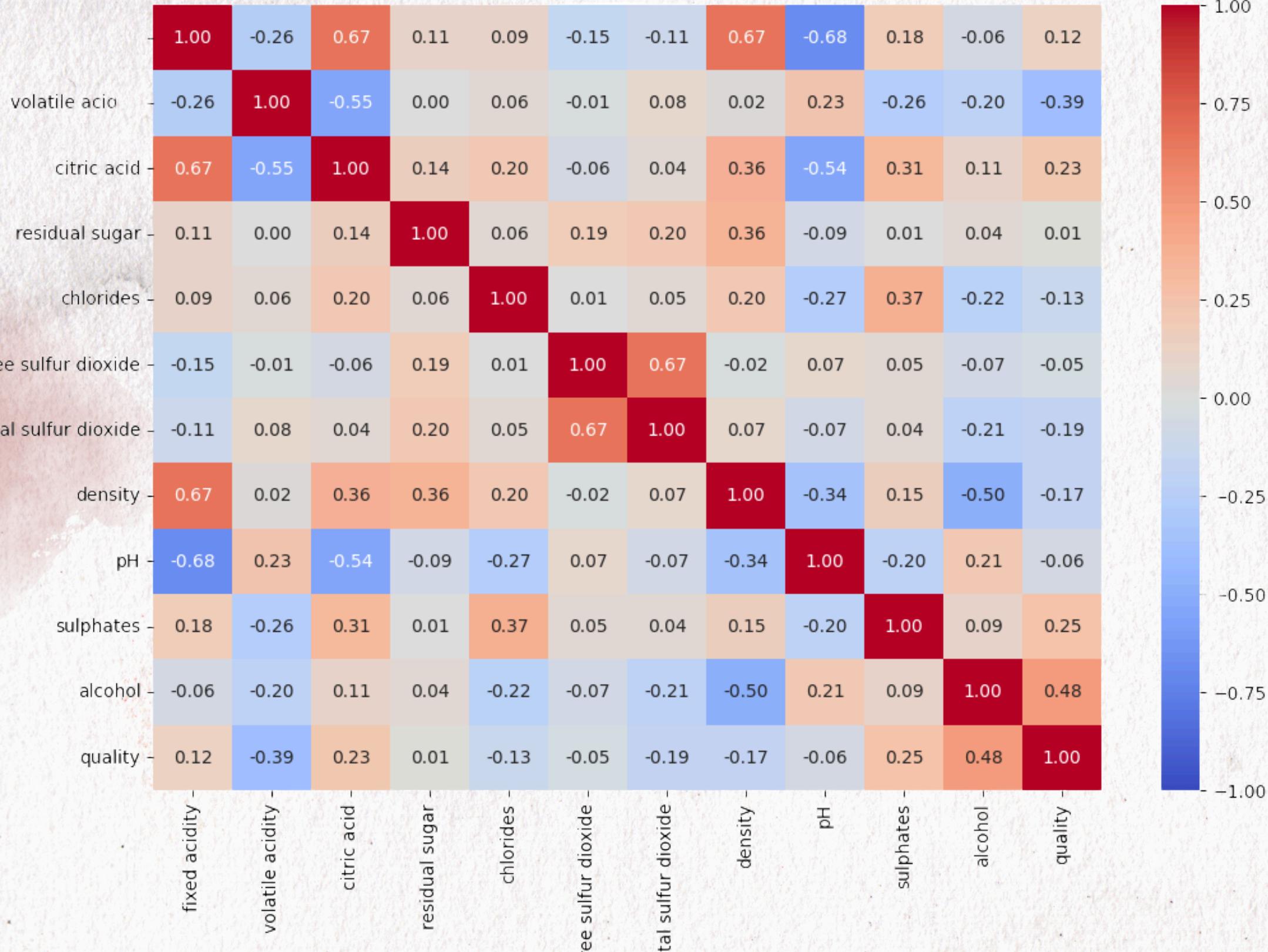
CATEGORIZACIÓN DE
QUALITY EN BAJA,
MEDIA Y ALTA

TÉCNICA SMOTE PARA
GENERAR MUESTRAS
SINTÉTICAS DE LAS CLASES
MINORITARIAS

ESTANDARIZACIÓN



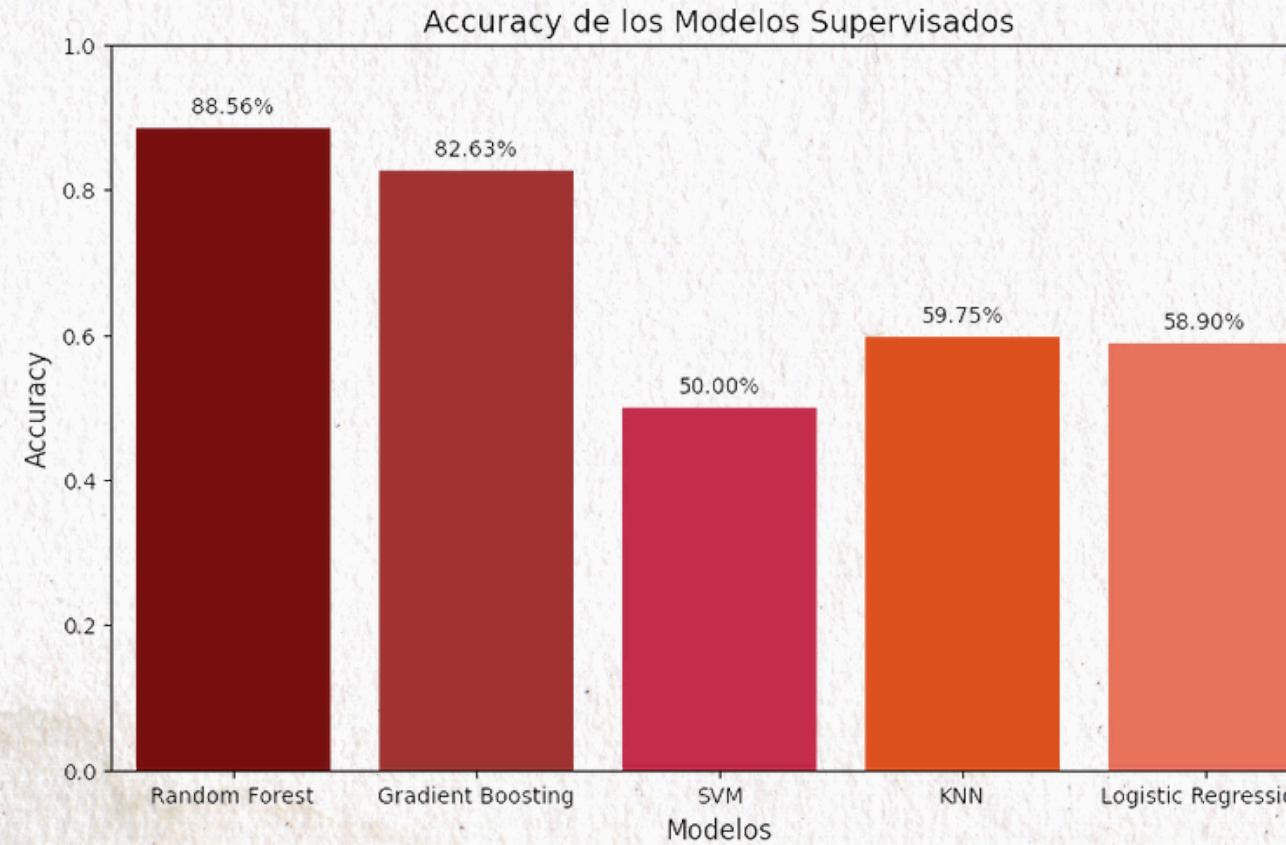
MATRIZ DE CORRELACIÓN



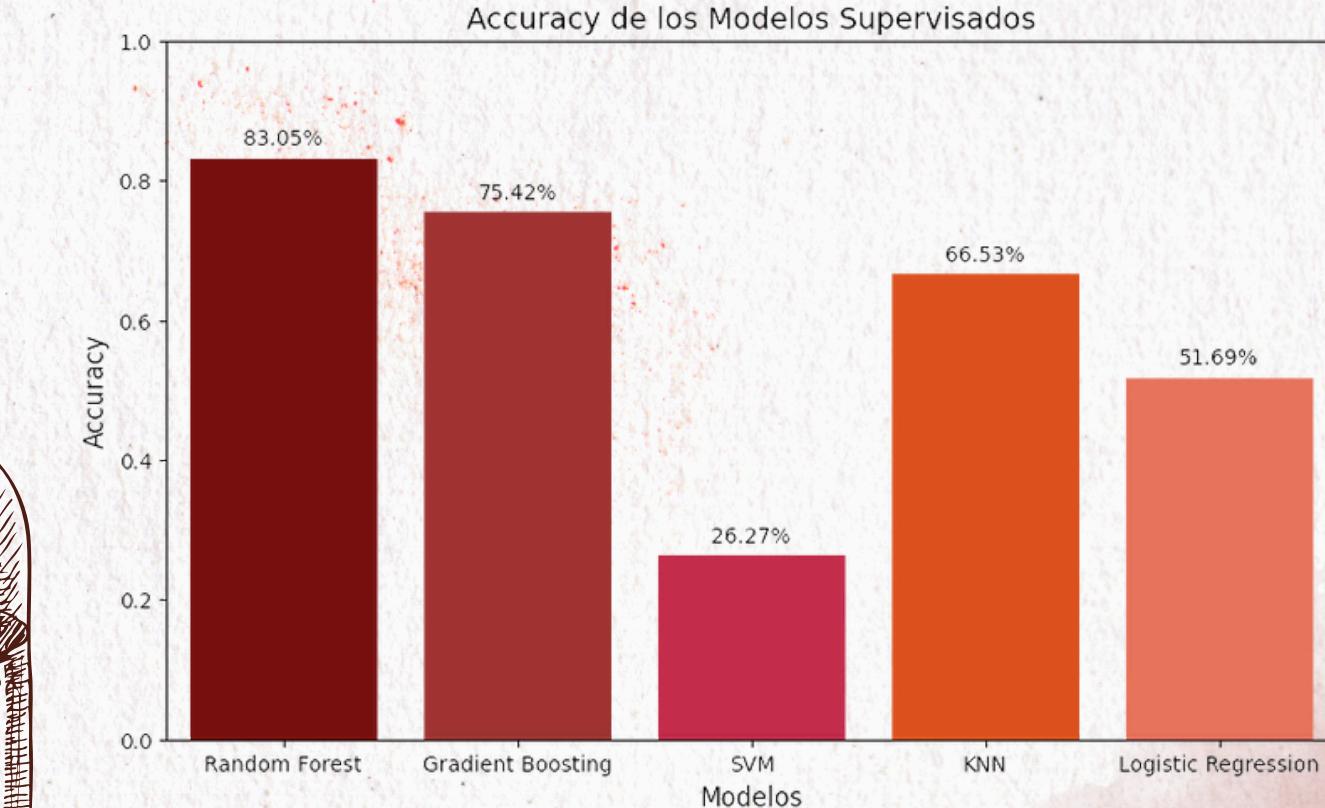
MODELOS SUPERVISADOS



ALL FEATURES

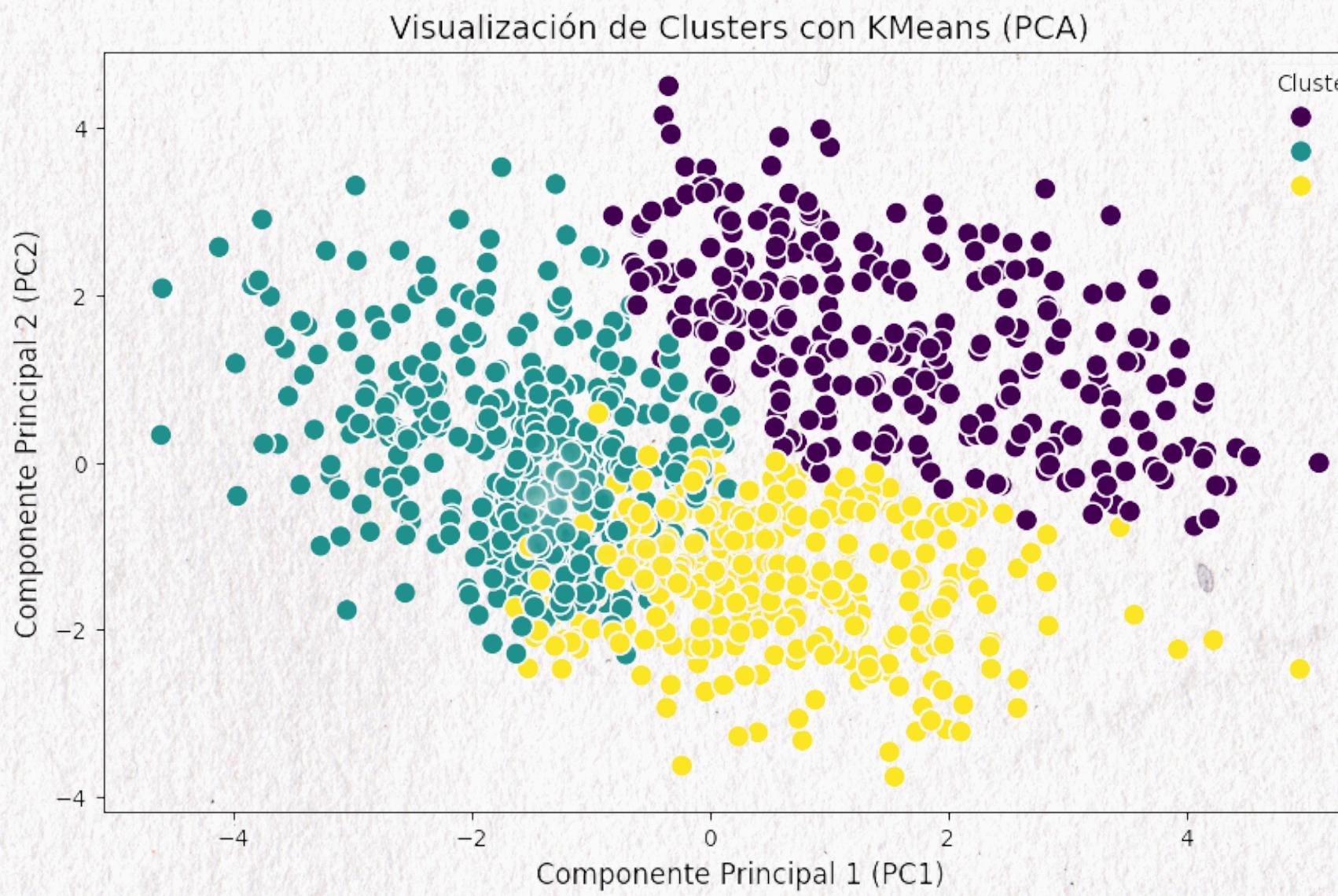


KEY FEATURES



MODELO NO SUPERVISADO

KMEANS



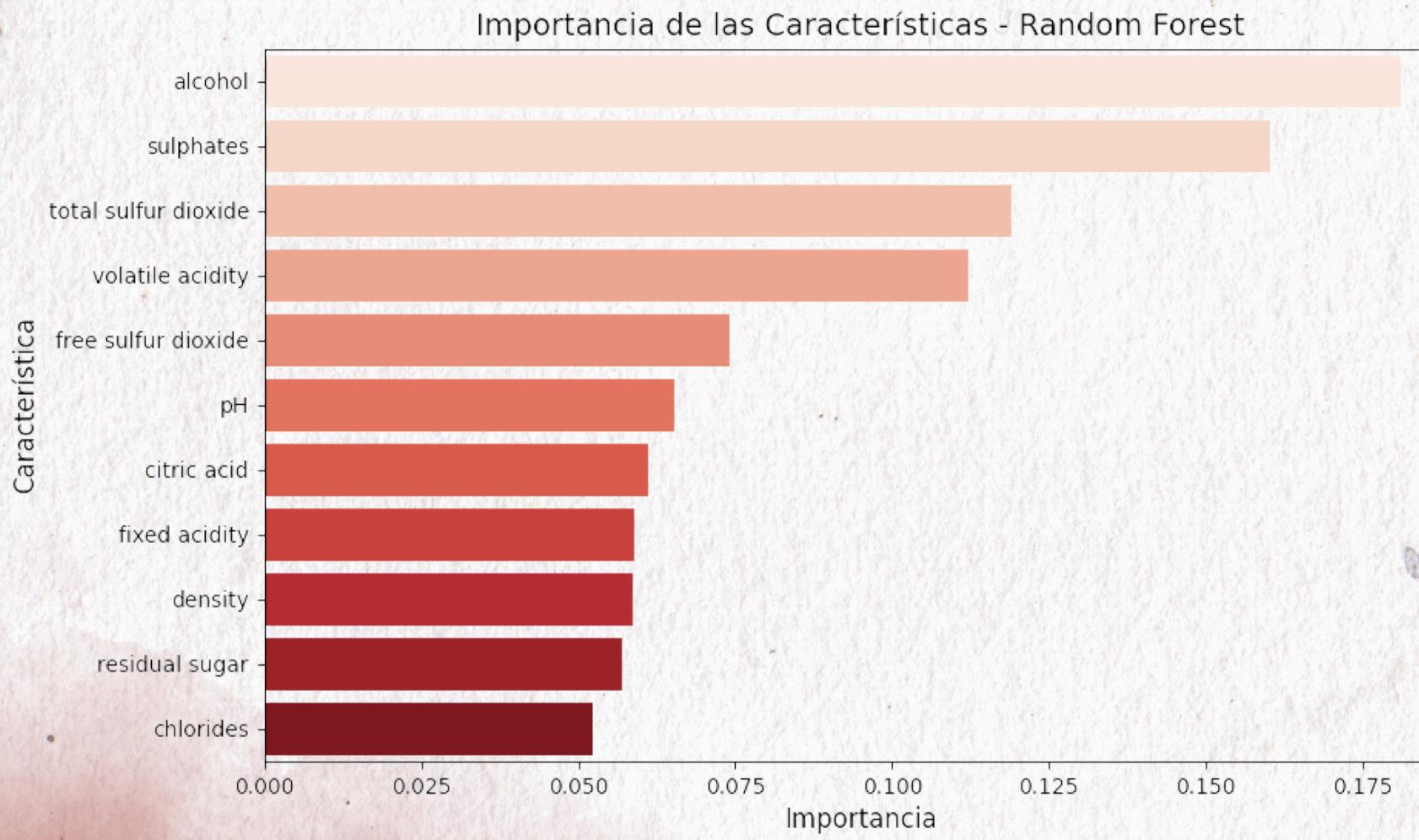
SILHOUETTE SCORE: 0.16 (ALL FEATURES)

SILHOUETTE SCORE: 0.24 (KEY FEATURES)

Valores bajos y cercanos a cero, los clusters no están bien definidos

ALGUNA MÉTRICA MÁS

¿QUÉ ELEGIR?



Model	Accuracy	Precision	Recall	F1-Score	Silhouette Score	Best Score
Random Forest	0.885593	0.69	0.69	0.69	NaN	NaN
Gradient Boosting	0.826271	0.55	0.69	0.61	NaN	NaN
SVM	0.500000	0.24	0.73	0.36	NaN	NaN
KNN	0.597458	0.33	0.77	0.47	NaN	NaN
Logistic Regression	0.588983	0.39	0.88	0.54	NaN	NaN
KMeans	NaN	NaN	NaN	NaN	0.155636	NaN
GridSearchCV	NaN	NaN	NaN	NaN	NaN	0.954733

Model	Accuracy	Precision	Recall	F1-Score	Silhouette Score	Best Score
Random Forest	0.830508	0.63	0.73	0.68	NaN	NaN
Gradient Boosting	0.754237	0.52	0.65	0.58	NaN	NaN
SVM	0.262712	0.33	0.88	0.48	NaN	NaN
KNN	0.665254	0.39	0.77	0.52	NaN	NaN
Logistic Regression	0.516949	0.38	0.85	0.52	NaN	NaN
KMeans	NaN	NaN	NaN	NaN	0.244666	NaN
GridSearchCV	NaN	NaN	NaN	NaN	NaN	0.91893



CONCLUSIONES

Este proyecto sienta las bases para un sistema robusto de clasificación y agrupación de vinos, con amplias posibilidades de mejora (optimización de parámetros, ingeniería de características, ensamblado de modelos) y escalabilidad (modelos Random Forest y Gradient Boosting) para adaptarse a necesidades futuras.

GRACIAS

