

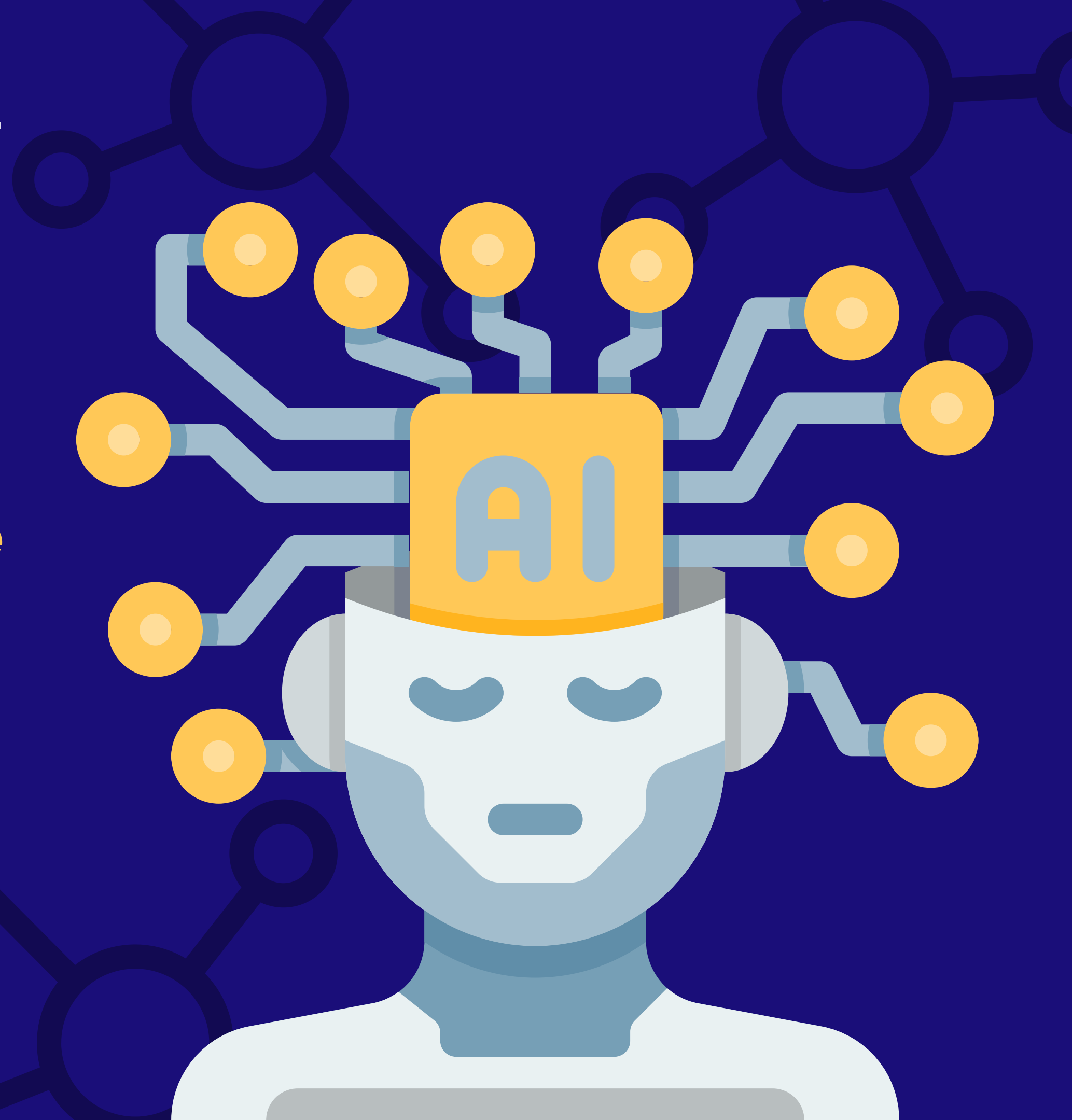


Universidad Industrial de Santander
Inteligencia Artificial I

DeepWine: optimización de la calidad del vino mediante modelos de ML y DL

Proyecto final de curso IA-20251-B1

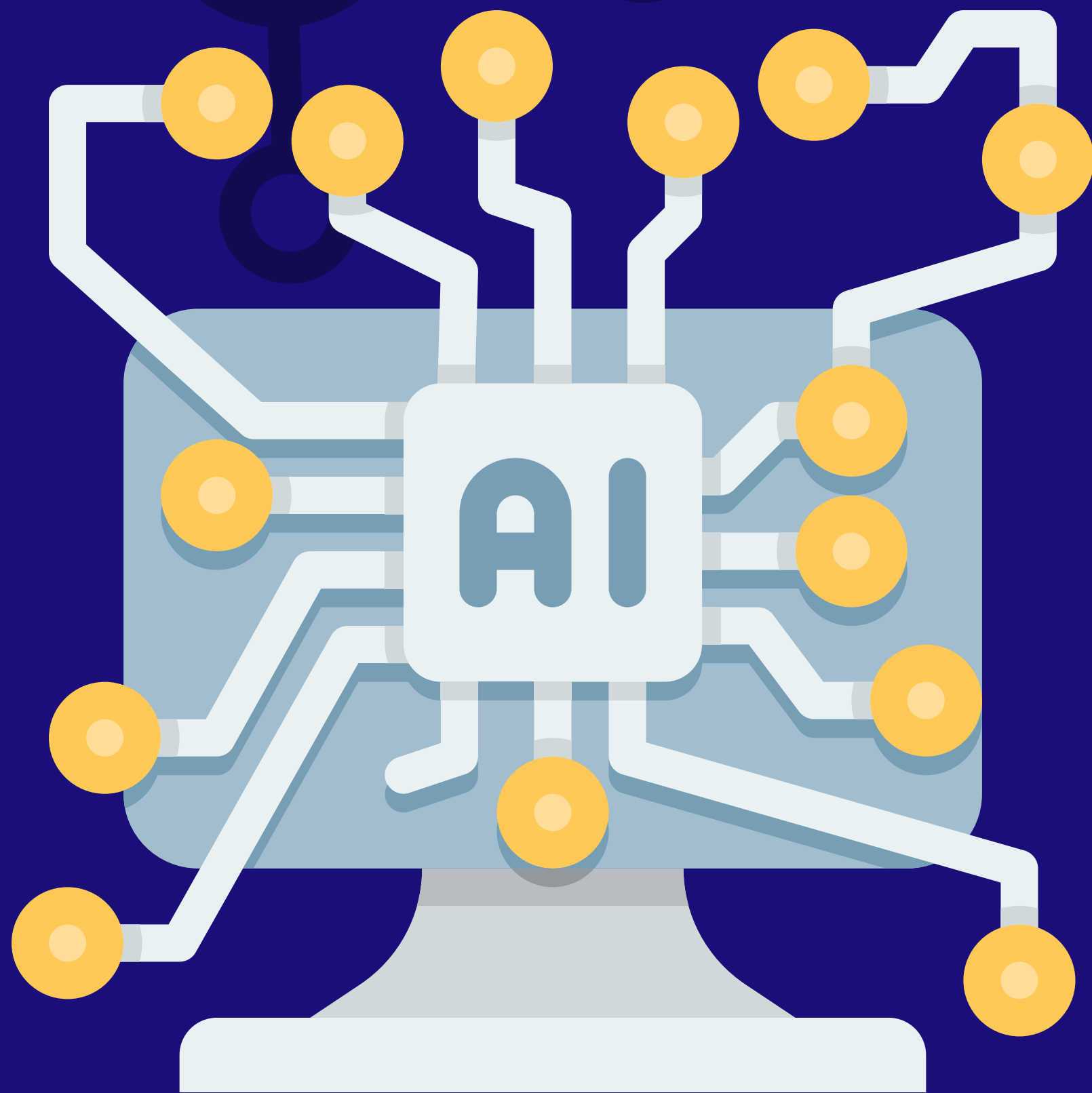
Luis Mario Toscano Palomino - 2220072
Oscar Carreño Serpa - 2220039
Daniel Santiago Convers C. - 2221120





Índice

- 01.** Motivación
- 02.** Objetivo
- 03.** Información del Dataset
- 04.** Procesamiento del Dataset
- 05.** Estadísticas del Dataset
- 06.** Tipos de Modelos Usados





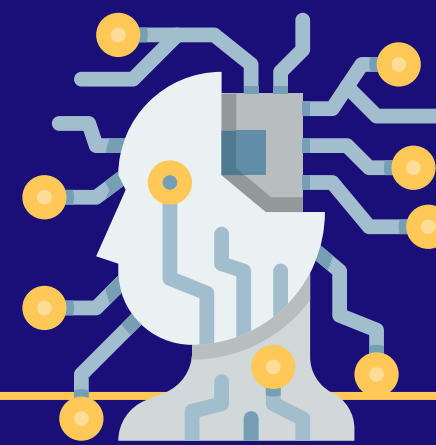
MOTIVACIÓN

Este proyecto busca usar la ciencia de datos para entender mejor cómo se hace el vino tinto. Se plantea usar inteligencia artificial, para ver cómo las diferentes características del vino (cantidad de alcohol, acidez) afectan su composición. La idea principal es entender y analizar el equilibrio entre las “recetas” para mejorar la calidad del vino tinto, usando modelos de aprendizaje, y herramientas computacionales.

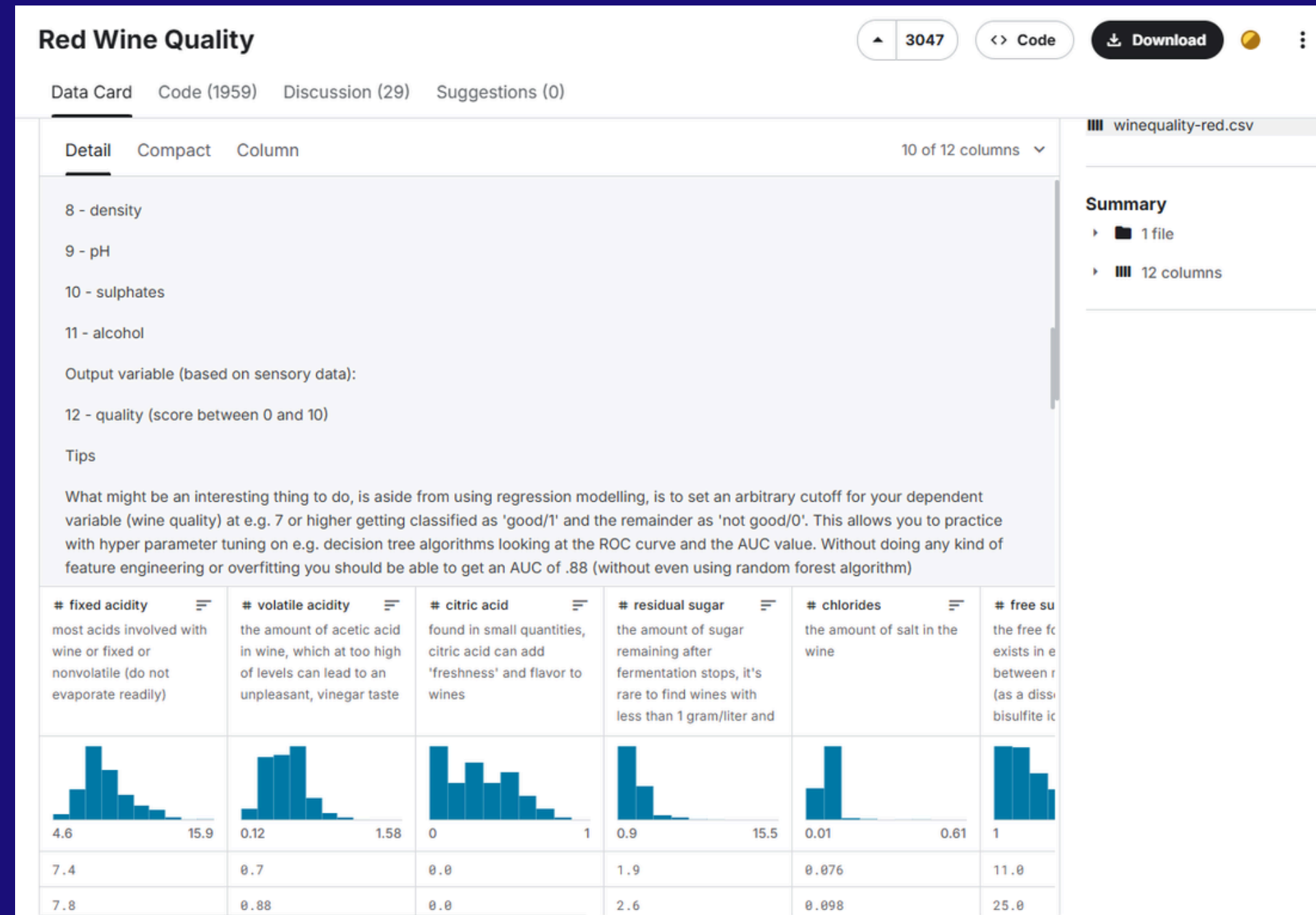


OBJETIVO

Crear un modelo de optimización, basado en modelos de inteligencia artificial, destinado a la identificación de las combinaciones más favorables de atributos fisicoquímicos que conduzcan a la maximización de la calidad del vino tinto.



INFORMACIÓN DEL DATASET





Preprocesamiento

```
▼ Revision de datos nulos en las columnas
```

```
#@title Revision de datos nulos en las columnas  
print("Valores nulos en cada columna:")  
df.isnull().sum()
```

```
↔ Valores nulos en cada columna:
```

	0
fixed acidity	0
volatile acidity	0
citric acid	0
residual sugar	0
chlorides	0
free sulfur dioxide	0
total sulfur dioxide	0
density	0
pH	0
sulphates	0
alcohol	0
quality	0

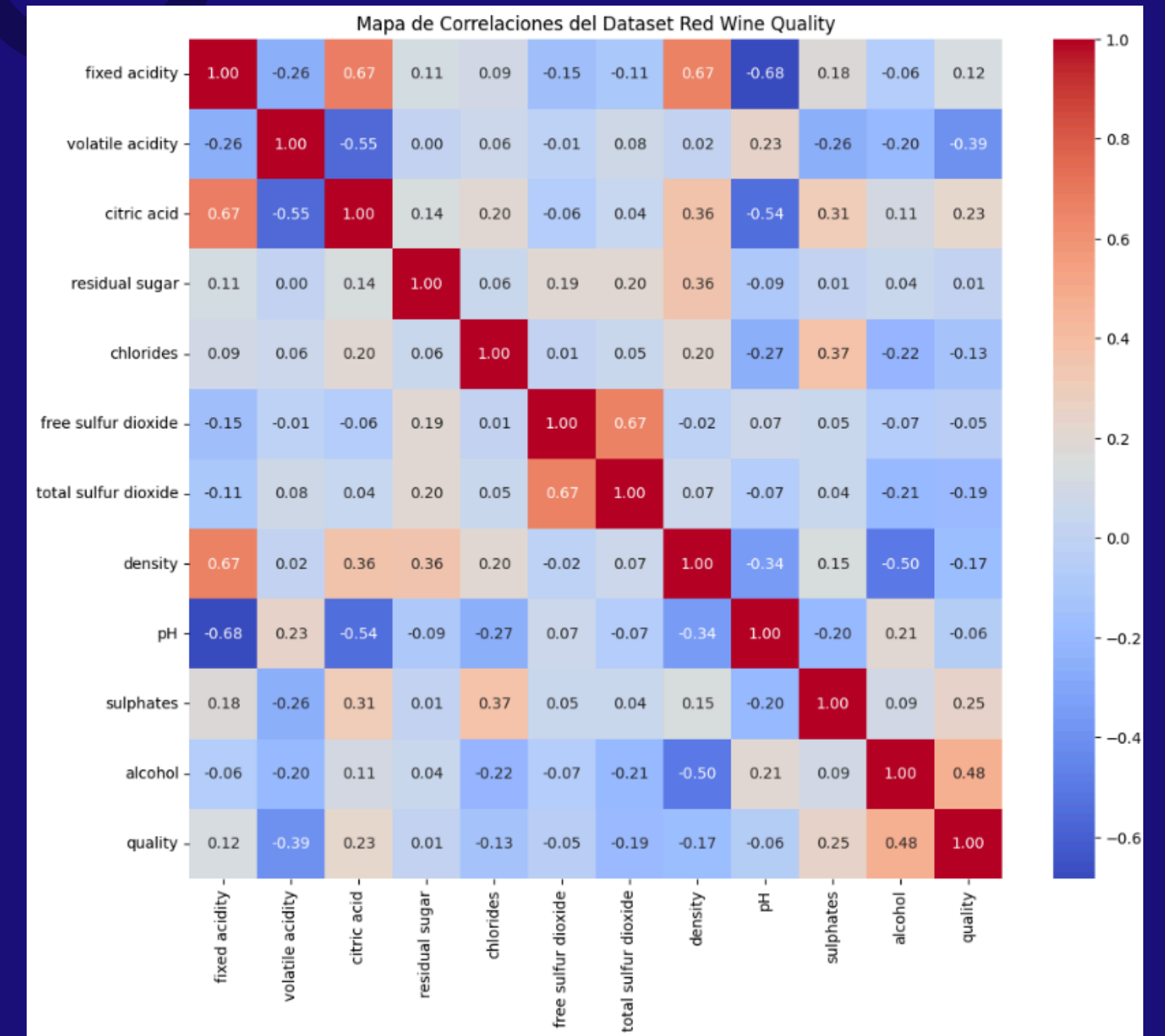
dtype: int64

Se realizó una rectificación de valores nulos en cada columna propuesta en el dataset. Los datos proveídos estaban “limpios” para trabajar.



Estadísticas (y datos)

- Vinos de alta calidad tienden a tener un mayor contenido de alcohol y una menor acidez volátil.
- Correlación positiva moderada entre el contenido de alcohol (alcohol) y la calidad del vino (quality)
- Multicolinealidad entre variables como la acidez fija (fixed acidity) y el ácido cítrico (citric acid), con una correlación de 0.67, y entre el dióxido de azufre total (total sulfur dioxide) y el dióxido de azufre libre (free sulfur dioxide), también con un valor de 0.67

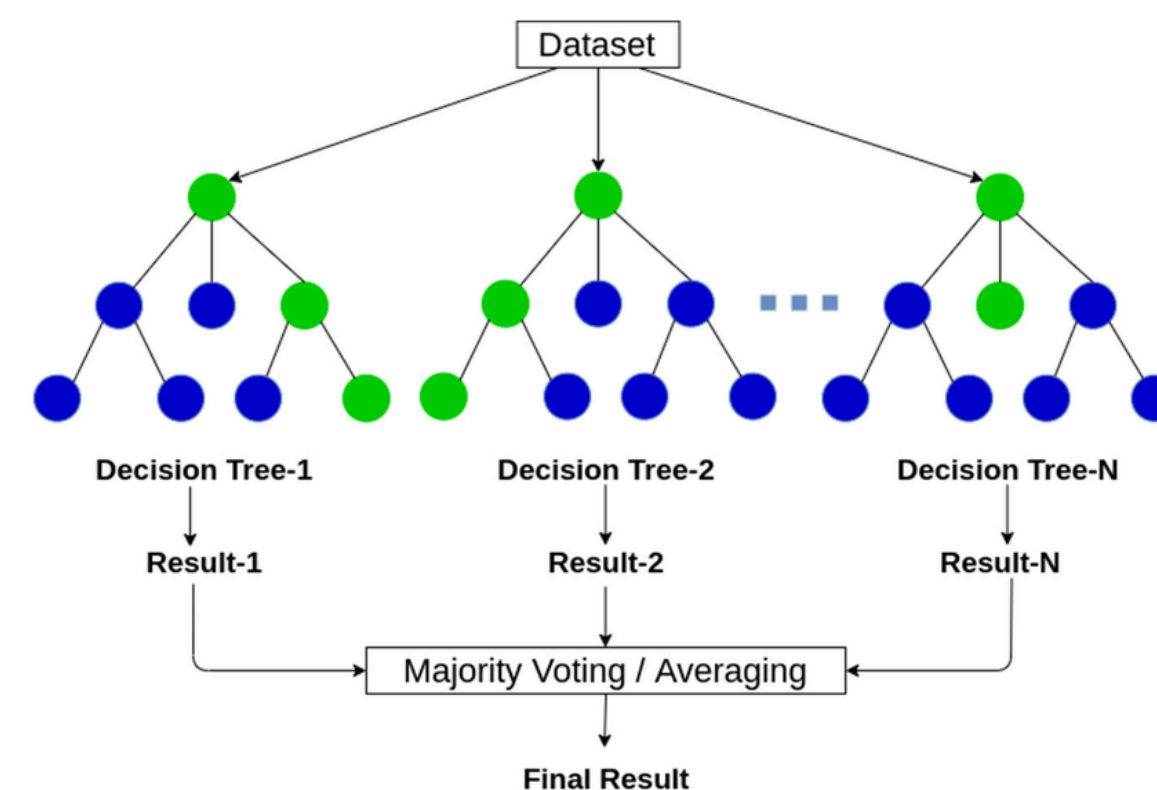




Modelo (ML): Random Forest

Se usó un RF predecir la calidad de un vino en función de sus características fisicoquímicas. El modelo recibe las características de cada vino (como pH, acidez, alcohol, etc.) y devuelve un valor numérico que indica la calidad del vino.

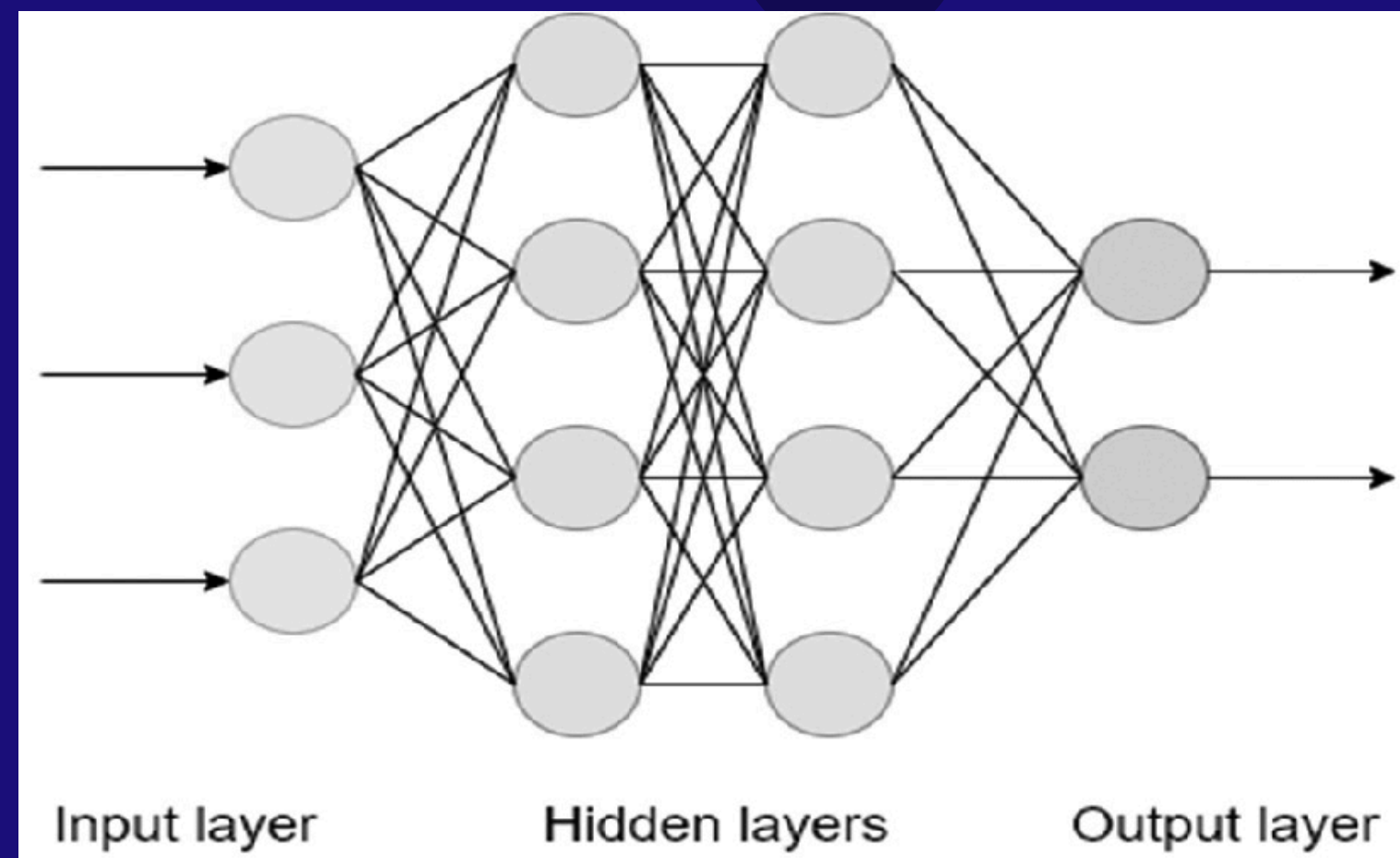
Random Forest





Modelo (DL): DNN

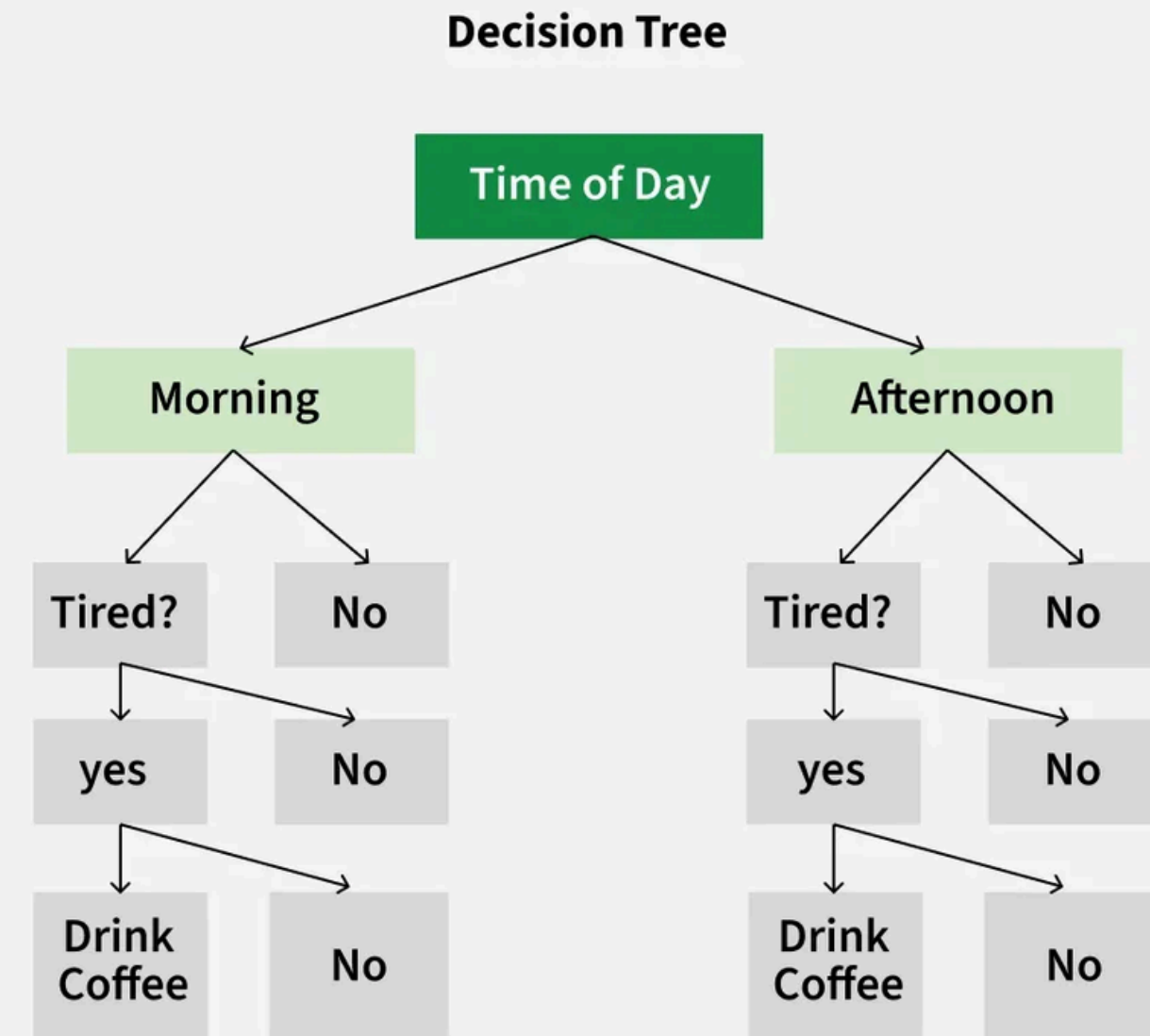
El modelo se utilizó para predecir el contenido de alcohol en el vino en función de sus características fisicoquímicas. El modelo recibe las características del vino (como acidez, pH, azúcares residuales, etc.) y devuelve un valor numérico que indica el contenido de alcohol en porcentaje.





Modelo (ML): Decision tree

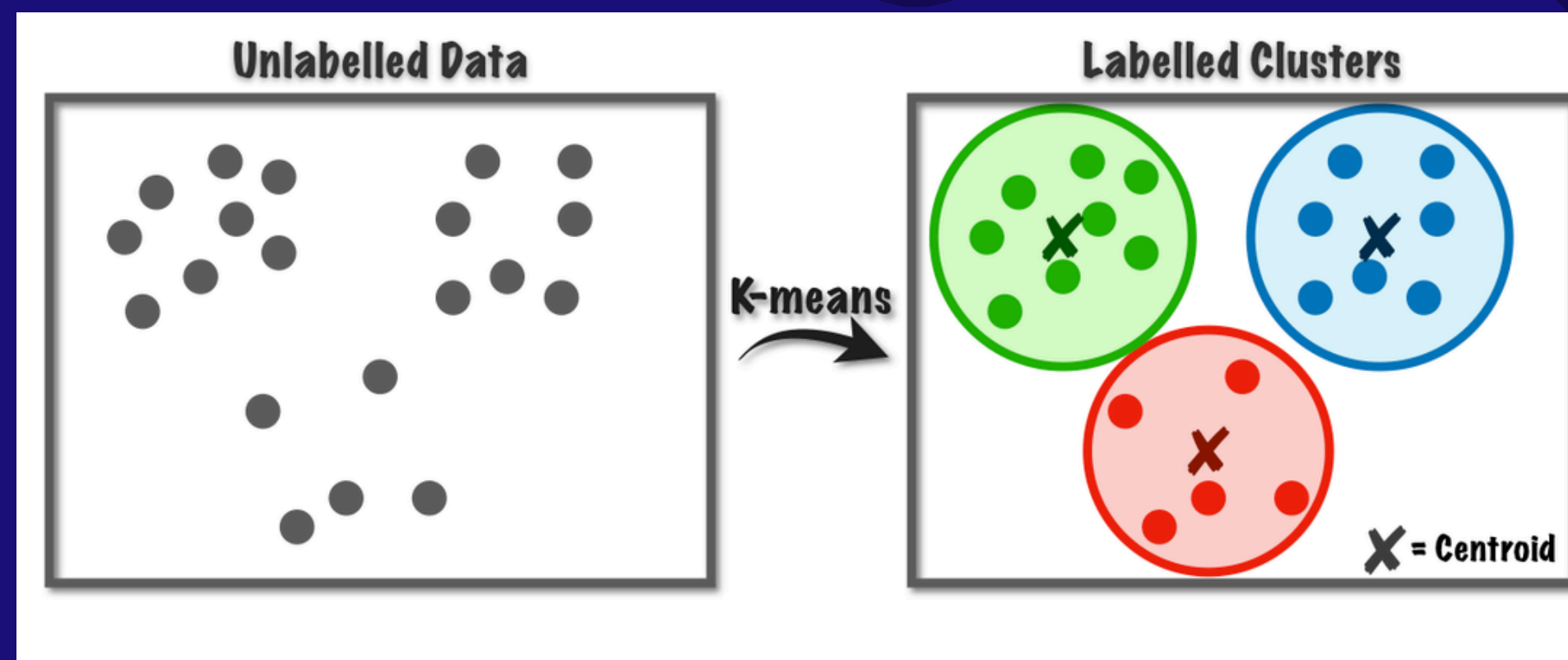
Este modelo busca predecir la acidez fija de un vino en función de su densidad. Este enfoque permite visualizar cómo diferentes profundidades de árboles impactan la predicción, ofreciendo una comparación entre modelos más simples y modelos más complejos.





Modelo (NSL): K-Means

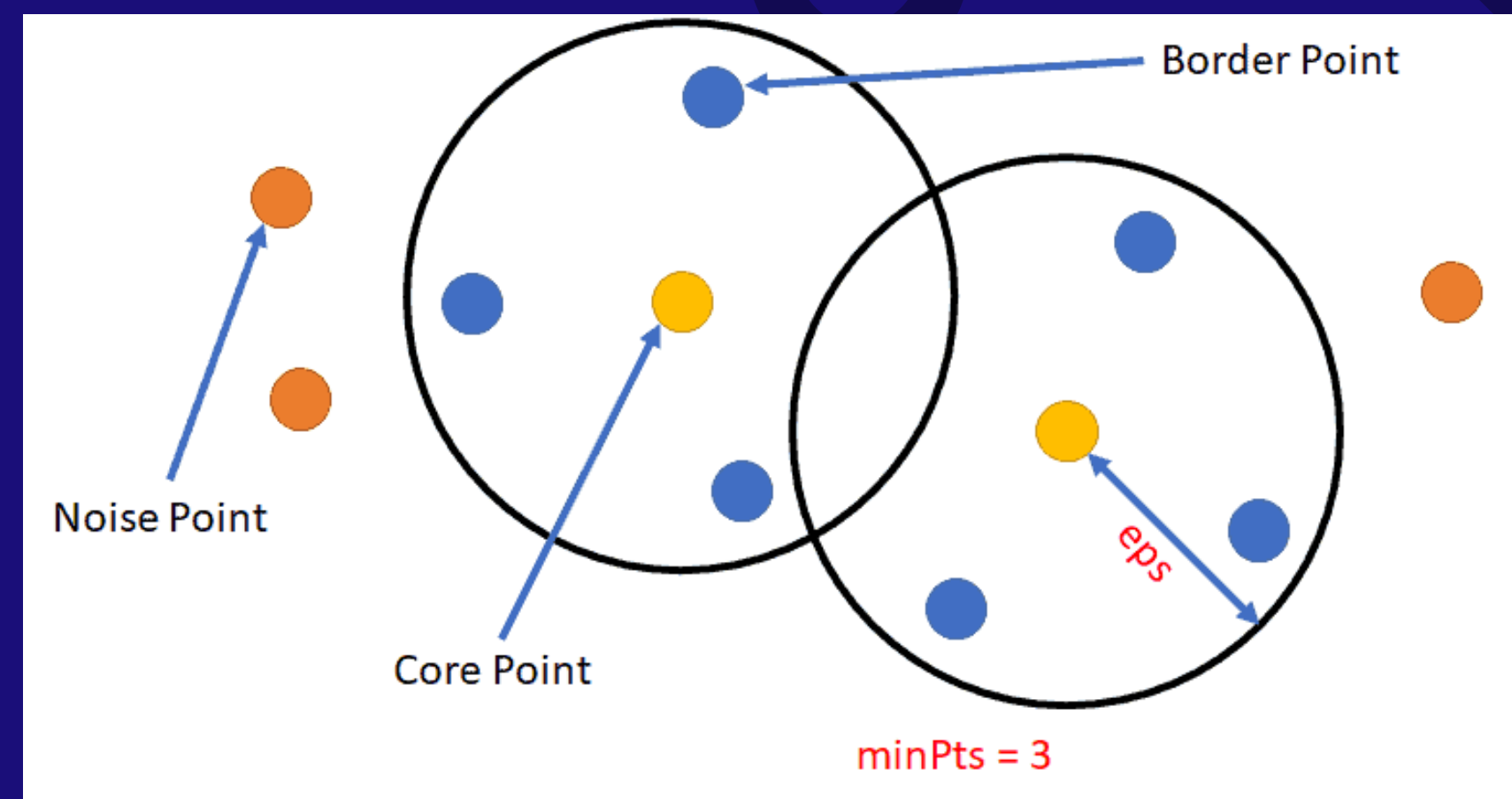
El modelo aplica un enfoque no supervisado utilizando K-Means clustering para agrupar los vinos en función de sus características fisicoquímicas. A diferencia de los modelos supervisados, este modelo no utiliza la variable "quality" para entrenarse.





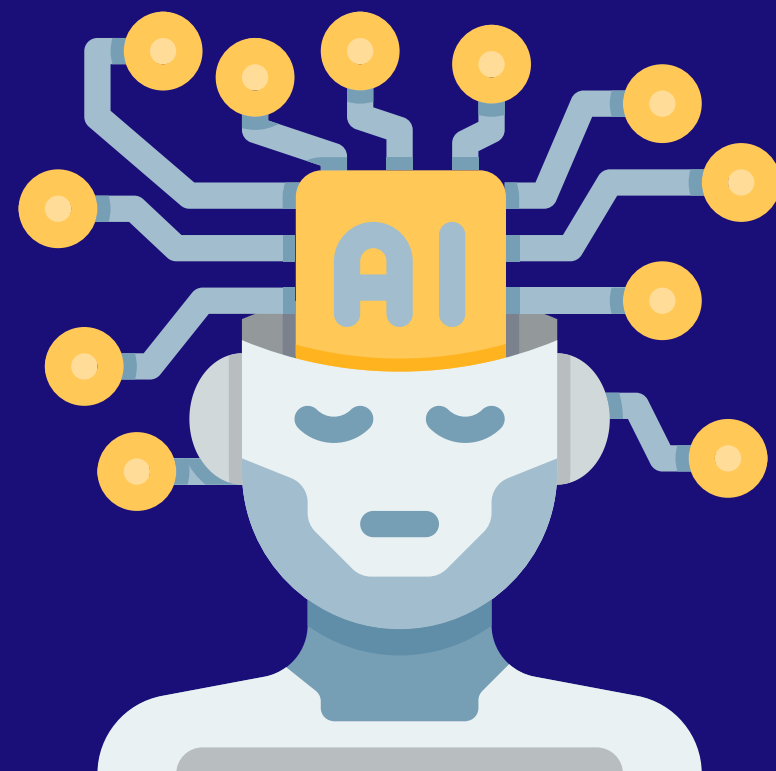
Modelo (NSL): DBSCAN

El modelo aplica un enfoque no supervisado utilizando el algoritmo DBSCAN para agrupar vinos en función de características fisicoquímicas. No usa "quality" durante el entrenamiento, sino que identifica agrupaciones basadas en la densidad de los datos, permitiendo detectar tanto grupos densamente conectados como puntos atípicos (ruido).





Universidad Industrial de Santander
Inteligencia Artificial I



Gracias

