

Asignación VI de Inteligencia Artificial – Decisiones y Hallazgos

José Angel Graciano Hernández 100039989

En esta práctica se aplicaron técnicas de **aprendizaje supervisado y no supervisado**, así como **reglas de asociación**, utilizando distintos datasets. El objetivo principal fue entrenar modelos de clasificación para vinos, comparar su desempeño, explorar patrones en datos no etiquetados y extraer relaciones frecuentes entre productos.

Se decidió trabajar con **árboles de decisión y Random Forest** para clasificación, **KNN y regresión logística** para comparación de algoritmos, **K-Means** para clustering, y el algoritmo **Apriori** para reglas de asociación.

1. Decisiones de modelado

Aprendizaje Supervisado

Para la clasificación de vinos se decidió:

- Transformar la variable quality en una **clase binaria** (buena/mala calidad).
- Dividir los datos en **entrenamiento (80%) y prueba (20%)**.
- Normalizar los datos con StandardScaler, ya que algoritmos como KNN son sensibles a la escala de las variables.
- Probar distintos modelos:
 - **Árbol de decisión**: fácil de interpretar, pero con riesgo de sobreajuste.
 - **Random Forest**: combina múltiples árboles para mejorar estabilidad y precisión.
 - **KNN**: captura patrones locales, útil para datasets con estructuras complejas.
 - **Regresión logística**: modelo lineal simple, permite interpretar la relación de cada variable con la clase.

Selección de hiperparámetros

- Para KNN se probaron distintos valores de n_neighbors.
- Random Forest se configuró con 100 árboles y profundidad ajustable.

- En todos los modelos se fijó la **semilla random_state=42** para garantizar reproducibilidad.

2. Hallazgos principales

Resultados supervisados

- Random Forest obtuvo **mayor precisión y f1-score**, confirmando su capacidad de generalización frente a un solo árbol de decisión.
- KNN mostró buen desempeño en la clase mayoritaria, pero menor sensibilidad a la clase minoritaria.
- Regresión logística presentó resultados razonables, aunque con menor f1-score para la clase positiva.

Resultados no supervisados

- K-Means permitió identificar **grupos naturales** en los datos de peces (FISH dataset), mostrando que el algoritmo detecta patrones sin necesidad de etiquetas.
- La interpretación se facilitó mediante reducción a **2 dimensiones con PCA**, generando gráficos de dispersión claros.
- **Reglas de asociación**
- Se encontraron relaciones significativas entre productos, como: Huevos y Cereal.
- Métricas:
 - **Support (0.14)**: el 14% de transacciones incluyen ambos productos.
 - **Confidence (0.41)**: si alguien compra huevos, hay 41% de probabilidad de comprar cereal.
 - **Lift (1.25)**: la compra de huevos aumenta 25% la probabilidad de comprar cereal en comparación con la aleatoriedad.
- Estas reglas son útiles para **marketing, promociones y ubicación de productos** en la tienda.

LINK DE GITHUB DONDE ESTA CARPETA, NOTEBOOK Y TODO

<https://github.com/gracianogeek/Asignacion-VI>