

Data Mavericks

Análisis de Modelos Predictivos - Conjunto de Datos de Frutas

Model	AUC	CA	F1	Prec	Recall	MCC
Tree	0.700	0.686	0.687	0.699	0.686	0.380
Logistic Regression	0.910	0.857	0.858	0.859	0.857	0.712
Neural Network	0.877	0.771	0.769	0.771	0.771	0.529

Al comparar el rendimiento de los tres modelos evaluados (Árbol de Decisión, Regresión Logística y Red Neuronal), se observa que la **Regresión Logística** es la que muestra el mejor desempeño general.

- Presenta la mayor precisión (CA = 0.857), el mejor F1-score (0.858) y la mayor precisión (Prec = 0.859), lo que indica que no solo acierta con frecuencia, sino que además tiene un muy buen equilibrio entre verdaderos positivos y negativos.
- Su **Recall** (0.857) también es el más alto, lo cual significa que detecta correctamente la mayoría de los casos positivos reales.
- El valor de **MCC (0.712)**, que resume la calidad de la predicción en un solo número incluso si las clases están desbalanceadas, refuerza la robustez del modelo.

Aunque la **Red Neuronal** también tiene buen rendimiento, la Regresión Logística supera a los demás modelos en todas las métricas relevantes, por lo que es la **opción más recomendable** para este problema.

Al observar los resultados de los distintos modelos aplicados al conjunto de datos, la Regresión Logística se destaca con un desempeño general muy alto,

alcanzando un accuracy del 91%. No obstante, entendemos que en este tipo de problemas no alcanza con solo mirar el accuracy, ya que tanto los falsos positivos como los falsos negativos pueden tener impacto importante. Por este motivo, se utilizó el F1-score como métrica principal, ya que combina precisión y recall en una sola medida. Este enfoque permite evaluar mejor el balance entre los casos correctamente identificados y los errores cometidos. La Regresión Logística volvió a sobresalir, obteniendo también el F1-score más alto con un valor de 85,8%. Esto sugiere que el modelo no solo acierta en general, sino que también mantiene un buen equilibrio entre detectar correctamente los casos positivos sin excederse en errores tipo I. En función de ambos indicadores, se recomienda continuar con la Regresión Logística como modelo principal, aunque sería conveniente explorar mejoras adicionales o probar ensambles si se busca optimizar aún más el rendimiento