

Construcción del modelo

Para comparar distintos métodos de clasificación, conectamos tres modelos:

- **Tree** (Árbol de decisión)
- **Logistic Regression** (Regresión logística)
- **Neural Network** (Red neuronal)

Cada modelo fue entrenado con el conjunto de datos de entrenamiento que nos dieron, y su rendimiento se evaluó mediante los widgets **Predictions**, **Confusion Matrix**, **Test & Score** y **ROC Analysis**.

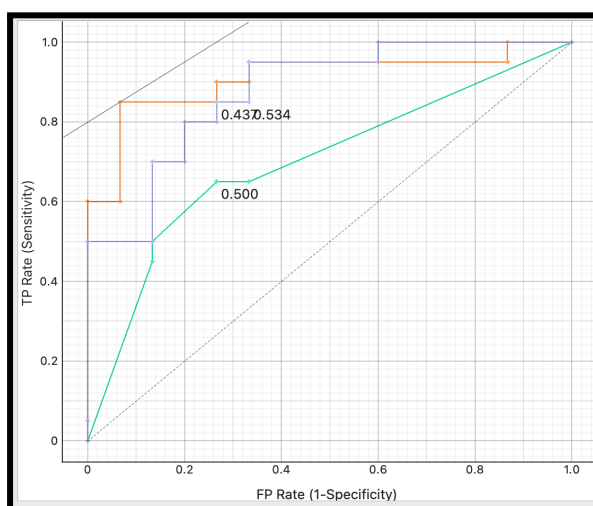
Resultados del modelo

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

Para empezar, el modelo de **regresión logística** obtuvo los mejores resultados generales, con un **AUC de 0.91** y una **precisión (CA) del 85.7%**, junto con valores equilibrados de F1, precisión y recall (0.86). Pudimos ver que estos indicadores muestran que el modelo pudo distinguir de manera confiable entre frutas y verduras, alcanzando un rendimiento robusto y consistente.

Por otro lado, el **árbol de decisión** fue el modelo con menor rendimiento (**AUC = 0.70**; **CA = 0.686**).

Análisis ROC



→ La curva ROC evidencia que la **regresión logística** (línea **naranja**) presenta el mejor desempeño, con la mayor área bajo la curva.

→ La **red neuronal (violeta)** también muestra una buena discriminación, mientras que el **árbol de decisión (verde)** presenta un rendimiento más limitado.

Análisis de las predicciones

Asimismo, el widget **Predictions** permitió visualizar las clasificaciones generadas por cada modelo sobre las mismas observaciones.

Podemos concluir que los tres modelos coincidieron en la mayoría de los casos, aunque se observaron algunas diferencias puntuales. Por ejemplo, alimentos clasificados como fruta por el árbol de decisión y como verdura por la red neuronal.

Matriz de confusión e interpretación

		Actual			
		fruit	vegetable	Σ	
	fruit	13	7	20	
	vegetable	4	11	15	
		Σ	17	18	35

De un total de **35 observaciones**, el modelo logró **24 clasificaciones correctas** (13 frutas y 11 verduras) y **11 clasificaciones incorrectas** (7 frutas confundidas con verduras y 4 verduras confundidas con frutas).

- **Verdaderos positivos (11)**: verduras correctamente clasificadas como *vegetable*.
- **Verdaderos negativos (13)**: frutas correctamente clasificadas como *fruit*.
- **Falsos positivos (7)**: frutas que el modelo clasificó erróneamente como verduras.
- **Falsos negativos (4)**: verduras que fueron clasificadas como frutas.

Conclusión

- Entre los modelos probados, **la regresión logística fue la que mejor funcionó**, con un nivel de acierto del **85,7%**.
- El modelo logró identificar correctamente la mayoría de los casos, aunque hubo algunos errores lógicos: ciertas frutas y verduras tienen valores muy parecidos, lo que puede confundir al sistema..
- El **árbol de decisión** ayudó a entender qué variables fueron más importantes para clasificar, destacando las **calorías** y la **vitamina A** como las que más influyen en la predicción.
- En resumen, los modelos demostraron que es posible clasificar alimentos según sus nutrientes con bastante precisión, y que la regresión logística es la más confiable para este conjunto de datos.