

Predictions - Orange

Show probabilities for **Classes known to the model**

	Neural Network	Logistic Regression	Tree	
1	0.88 : 0.12 → fruit	0.99 : 0.01 → fruit	1.00 : 0.00 → fruit	
2	0.00 : 1.00 → vegetable	0.02 : 0.98 → vegetable	1.00 : 0.00 → fruit	
3	0.99 : 0.01 → fruit	0.99 : 0.01 → fruit	1.00 : 0.00 → fruit	
4	1.00 : 0.00 → fruit	1.00 : 0.00 → fruit	1.00 : 0.00 → fruit	
5	0.13 : 0.87 → vegetable	0.23 : 0.77 → vegetable	0.00 : 1.00 → vegetable	
6	0.92 : 0.08 → fruit	0.96 : 0.04 → fruit	1.00 : 0.00 → fruit	

	name	vitamin A %	vitamin C %	calcium %	iron %	magnesium %	calories (per 100g)	potassium (mg)	protein (g)	fiber (g)
?	1	154	3	1	4	61	213	1.1	3.0	
?	15	300	2	11	3	20	202	2.2	2.1	
?	0	43	2	3	5	53	151	1.1	7.0	
?	2	16	1	3	7	160	485	2.0	7.0	
?	0	8	1	1	5	43	325	1.6	2.8	
?	26	16	1	1	2	50	173	1.0	1.6	

Estamos analizando cuando entrenamos con una dataset al modelo y le agregamos información externa para el testeo.

Usamos 3 modelos: Neural Network, Logistic Regression and Tree.

- En principio podemos ver que en el 80% de los casos, las tres variables devuelven la misma predicción.
- En el caso de Tree, las probabilidades binarias (Sí o No), esto puede darse a un overfitting del modelo porque no generaliza, sino que aprendió perfectamente las variables y no puede distinguir cuando ingresan datos nuevos.
- En los modelos de Neural Network y Logistic Regression, notamos que puede generalizar más porque salen fuera del 0 o 1.
- Al tener poca cantidad de datos de prueba, deberemos ver cómo se comporta Neural Network con mayor volumen. Aun así Logistic Regression nos da mejores predicciones porque tiende más a los extremos sin llegar a ser binarias (como en el caso de Tree), por lo cual podremos confiar más en ella por su precisión y generalización.