Fundamentos de Confusion Matrix

Prof. Jhoan Steven Delgado V.

Intro

Edad	Ingresos	Tiene carro?
24	1’200.000	NO
23	4’500.000	SI
45	1’250.000	SI
32	1’100.000	NO

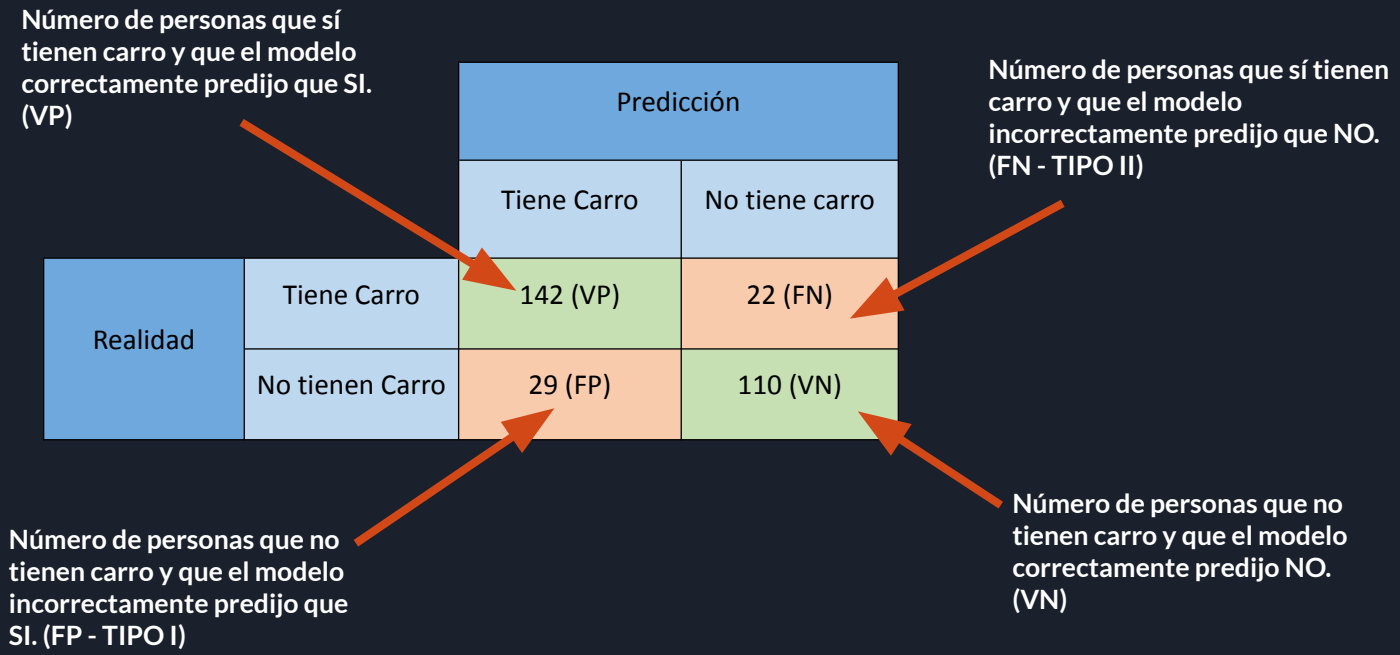
Logistic Regression?

?

KNN?

Confusion Matrix (Matriz de confusión)

*Una matriz de dimensión nxn donde n es la cantidad de clases (o categorías) de la variable objetivo



Confusion Matrix (Matriz de confusión)

KNN

		Predicción	
		Tiene Carro	No tiene carro
Realidad	Tiene Carro	142	29
	No tiene Carro	22	110

Logistic Regression

		Predicción	
		Tiene Carro	No tiene carro
Realidad	Tiene Carro	139	32
	No tiene Carro	20	112

Accuracy

Accuracy= (VP+VN)/(VP+VN+FP+FN) (Fracción de clasificaciones correctas)

Accuracy KNN = (142+110)/(142+110+22+29) = 0.83 -> 83 % de todas las predicciones fueron correctas

KNN

		Predicción	
		Tiene Carro	No tiene carro
Realidad	Tiene Carro	142 (VP)	29 (FN)
	No tiene Carro	22 (FP)	110 (VN)

Logistic Regression

		Predicción	
		Tiene Carro	No tiene carro
Realidad	Tiene Carro	139 (VP)	32 (FN)
	No tiene Carro	20 (FP)	112 (VN)

Recall (Sensitivity), Specificity

- Recall (Sensitivity): Porcentaje de verdaderos **positivos** que fueron correctamente predichos por nuestro modelo

  Recall (Sensitivity) = VP / (VP+FN) -> TRP (True Positive Rate)

- Specificity: Porcentaje de verdaderos **negativos** que fueron correctamente predichos por nuestro modelo

  Specificity = VN / (VN+FP) -> TNR (True Negative Rate)

Recall (Sensitivity) KNN= 142/(142+29)= 0.83 -> 83% de las personas con carro, fueron correctamente predichas por nuestro modelo KNN.

Specificity LR = 112/ (112+20) = 0.85 -> 85% de las personas sin carro, fueron correctamente predichas por nuestro modelo LR.

KNN

		Predicción	
		Tiene Carro	No tiene carro
Realidad	Tiene Carro	142 (VP)	29 (FN)
	No tiene Carro	22 (FP)	110 (VN)

Logistic Regression

		Predicción	
		Tiene Carro	No tiene carro
Realidad	Tiene Carro	139 (VP)	32 (FN)
	No tiene Carro	20 (FP)	112 (VN)

Precision

- Precision: Porcentaje de los positivos **predichos** de nuestro modelo que fueron correctamente predichos (correctos)

  Precision = VP / (VP+FP)

Precision KNN=142/(142+22) = 0.87 -> De las 164 personas que predecimos que tenían carro, solo el 87% realmente tienen carro.

Precision nos da una noción de la calidad de los positivos predichos

KNN

		Predicción	
		Tiene Carro	No tiene carro
Realidad	Tiene Carro	142 (VP)	29 (FN)
	No tiene Carro	22 (FP)	110 (VN)

Logistic Regression

		Predicción	
		Tiene Carro	No tiene carro
Realidad	Tiene Carro	139 (VP)	32 (FN)
	No tiene Carro	20 (FP)	112 (VN)

F1 Score

- Favorece a los clasificadores que tienen alto recall y precision.

  F1 Score = 2*{Precision*Recall)/(Precision+Recall)}

KNN

		Predicción: Tiene Carro	Predicción: No tiene carro
Realidad	Tiene Carro	142 (VP)	29 (FN)
	No tiene Carro	22 (FP)	110 (VN)

Logistic Regression

		Predicción: Tiene Carro	Predicción: No tiene carro
Realidad	Tiene Carro	139 (VP)	32 (FN)
	No tiene Carro	20 (FP)	112 (VN)

Resumen

- Accuracy = (VP+VN)/(VP+VN+FP+FN) (Fracción de clasificaciones correctas)
- Mala clasificación = (FP+FN)/(VP+VN+FP+FN): (Fracción de clasificaciones incorrectas)
- Precisión= VP / (VP+FP): (Fracción de verdaderos positivos a positivos predichos)
- Recall (sensibilidad o TPR)= VP / (VP+FN)(Fracción de verdaderos positivos sobre todos los positivos)
- Especificidad (o TNR) = VN / (VN+FP): (Fracción de verdaderos negativos sobre todos los negativos)
- Tasa de falsos positivos (o FPR) = FP / (VN+FP)=1 – TNR

Referencias

- Notas de clase, Prof. Javier Díaz, Fundamentos de analitica.
- The Statquest Ilustrated guide to Machine Learning, Josh Starmer.