

Análisis de Modelos de Predicción

En este documento se pasmarán los resultados obtenidos del análisis de los modelos de predicción de un dataset que tiene datos que ayudan a clasificar a los aspirantes en admitido y no admitido.

Antes de esto se debe tener en claro las variables que usaremos para medir la fiabilidad de los modelos.

Actual Class	Predicted class	
	Class = Yes	Class = No
	Class = Yes	Class = No
	Class = Yes	Class = No
	True Positive	False Negative
	False Positive	True Negative

True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN)

Accuracy: es la medida de rendimiento más intuitiva y es simplemente una relación entre la observación predicha correctamente y el total de observaciones.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$$

Precision: es la relación entre las observaciones positivas pronosticadas correctamente y el total de observaciones positivas pronosticadas.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall: es la proporción de observaciones positivas predichas correctamente a todas las observaciones en la clase Yes.

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1 score: es el promedio ponderado de precisión y recuperación. Por tanto, esta puntuación tiene en cuenta tanto los falsos positivos como los falsos negativos, no es tan fácil de entender como el accuracy, pero F1 suele ser más útil que el accuracy, especialmente si tiene una distribución de clases desigual.

$$\text{F1 score} = \frac{2(\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})}$$

Reportes de Modelos

Reporte del modelo KNN				
	precision	recall	f1-score	support
0	0.59	0.68	0.63	28
1	0.53	0.43	0.48	23
accuracy			0.57	51
macro avg	0.56	0.56	0.55	51
weighted avg	0.56	0.57	0.56	51

Reporte del modelo Decision tree				
	precision	recall	f1-score	support
0	0.55	0.64	0.59	28
1	0.44	0.35	0.39	23
accuracy			0.51	51
macro avg	0.49	0.50	0.49	51
weighted avg	0.50	0.51	0.50	51

Reporte del modelo Logistic Regression				
	precision	recall	f1-score	support
0	0.57	0.46	0.51	28
1	0.46	0.57	0.51	23
accuracy			0.51	51
macro avg	0.51	0.51	0.51	51
weighted avg	0.52	0.51	0.51	51

Reporte del modelo Support Vector Machines				
	precision	recall	f1-score	support
0	0.62	0.54	0.58	28
1	0.52	0.61	0.56	23
accuracy			0.57	51
macro avg	0.57	0.57	0.57	51
weighted avg	0.58	0.57	0.57	51

Reporte del modelo Artificial Neural Network				
	precision	recall	f1-score	support
0	0.56	0.79	0.66	28
1	0.50	0.26	0.34	23
accuracy			0.55	51
macro avg	0.53	0.52	0.50	51
weighted avg	0.54	0.55	0.52	51

Llegamos a la conclusión después de analizar nuestros modelos que ninguno de los analizados puede ser útil al momento de tomarlo como base para realizar la aplicación para darle al futuro estudiante la certeza de que será o no aceptado en la universidad a la que se postuló, esto no porque los algoritmos estén incorrectos sino debido a que el dataset con el que fueron entrenados los modelos no es suficientemente bueno para entrenar modelos predictores, por lo que hemos tomado la decisión de buscar un nuevo dataset, con el mismo tema pero diferentes features, para así buscar mejores resultados, porque no lo que tenemos ahora sería casi igual tirar una moneda al aire y que ella haga la predicción.