

Experimento Propio

Patricio Guledjuan

Azul Noguera

Rocio Gonzalez

2023-08-22

Introducción

El objetivo de nuestro experimento consistió en explorar y comprender el impacto de las variables en el rendimiento del modelo. A través del análisis de las variables más influyentes, buscamos obtener una visión más profunda de cómo se toman decisiones en el contexto de cada conjunto de datos. Además, evaluamos si la selección de características relevantes podría potencialmente mejorar el desempeño del modelo.

Nuestro enfoque implicó identificar las variables más significativas en función del comportamiento del modelo original y, posteriormente, elegir las variables más influyentes para construir un nuevo modelo. Esta estrategia nos permitió comparar el rendimiento del nuevo modelo con el del original y analizar cualquier diferencia resultante.

En una primera etapa, evaluamos la importancia de los atributos en el modelo original utilizando rpart. En este contexto, una alta puntuación de importancia sugiere que la variable ejerce una mayor influencia en las decisiones tomadas por el árbol, y por ende, en las predicciones resultantes.

##	Churn_Importance
## complains	236.2103361
## seconds_of_use	204.9099906
## frequency_of_use	171.5954940
## customer_value	149.2826469
## distinct_called_numbers	132.2503734
## status	112.0612624
## frequency_of_sms	97.1680251
## call_failure	80.4796559
## subscription_length	62.5382885
## age_group	49.6025298
## age	41.0622159
## charge_amount	33.6214038
## tariff_plan	0.3453161

##	Heart_Importance
## ST_Slope	90.452387
## Oldpeak	57.272199
## MaxHR	46.758755
## ChestPainType	43.218971
## Age	36.473852
## ExerciseAngina	36.327523
## Cholesterol	30.218954
## RestingBP	24.524261
## Sex	8.631311
## FastingBS	5.295776
## RestingECG	4.607943

##		Wine_Importance
##	alcohol	213.69668
##	density	177.25551
##	free.sulfur.dioxide	138.41699
##	residual.sugar	136.51164
##	volatile.acidity	134.56902
##	total.sulfur.dioxide	129.91177
##	chlorides	114.29425
##	fixed.acidity	90.97684
##	citric.acid	80.10864
##	pH	72.57032
##	sulphates	54.38498

A partir de estos valores de importancia, establecimos una estrategia para determinar qué variables serían incluidas en el nuevo conjunto de datos. Para ello, calculamos el valor promedio de importancia de todas las variables presentes en el conjunto de datos original. Posteriormente, seleccionamos aquellas variables cuya importancia superara este promedio. Como resultado, cada uno de los conjuntos de datos originales se redujo a un conjunto de seis variables predictoras de mayor relevancia.

##	Dataset	Importancia_Promedio
## 1	ChurnRate	105.47135
## 2	HeartDisease	34.88927
## 3	WineQuality	122.06333