

Autor: Joaquin Leon Martínez

Parte 1:

El fichero esta adjunto

Parte 2:

Pregunta 1 - ¿Te costó esfuerzo aprender el presente tema?

Ha sido relativamente sencillo y muy interesante, ya que con estas herramientas se nos abre un gran abanico de posibilidades.

Pregunta 2 - ¿Qué te ha llamado la atención de PyCaret?

Lo rápido y sencillo que puedes realizar estudios en base a diferentes modelos de predicción y la posibilidad de mejorar dichos modelos de una forma tan ágil.

Pregunta 3 - ¿Notas que es más simple haciéndolo así que como se hizo en Fundamentos de Big Data? Si opinas diferente la respuesta no es correcta/incorrecta, siempre y cuando argumentes tu respuesta.

Considero que de esta forma es mucho mas sencillo ya que estas herramientas encapsulan gran parte de la lógica, pero también considero que esta bien haber estudiado antes como obtener estas predicciones de la forma “convencional” para poder comprender como funcionan estas herramientas.

Pregunta 4 - Busca más herramientas AutoML que existan y enumera alguna (simplemente queremos que te familiarices con ello, nada más).

- Google Cloud AutoML
- DataRobot
- Azure Automated Machine Learning
- TPOT

Parte 3:



Curva ROC: La curva de Característica Operativa del Receptor (ROC) es un gráfico que ilustra el rendimiento de un modelo de clasificación binaria a diferentes niveles de umbral de clasificación. Representa la tasa de verdaderos positivos en el eje Y y la tasa de falsos positivos en el eje X. Cada punto en la curva ROC corresponde a un umbral de clasificación diferente. Esto proporciona información sobre el rendimiento global del modelo en términos de sensibilidad y especificidad. Cuanto más cercana esté la curva ROC al rincón superior izquierdo del gráfico, mejor será el rendimiento del modelo. Por lo tanto, un modelo con una curva ROC que coincide con la línea diagonal tiene un rendimiento igual al azar.

Parte 4:

Confusión Matrix: La matriz de confusión, también conocida como tabla de contingencia, es una herramienta utilizada en la evaluación de modelos de clasificación para analizar y visualizar el desempeño del modelo en la tarea de clasificación. En ella se muestra la cantidad de aciertos y errores de clasificación para cada clase. Como norma general, las filas representan las clases reales o etiquetas verdaderas, mientras que las columnas representan las clases predichas por el modelo. Cada celda de la matriz muestra el recuento o la frecuencia de instancias clasificadas en una determinada categoría

Parte 5:

Envía tu predicción a Kaggle y adjunta un pantallazo con tu resultado

Submission and Description		Public Score 
	titanic_pycaret_P9.csv Complete · now · 2	0.622