

Memoria Machine Learning Meteo

ML Meteo

El proyecto consiste en determinar con la máxima precisión posible si va a llover o no el próximo día en función de todos nuestros parámetros. Los datos son de diferentes estaciones meteorológicas de Australia.

Evolución del proyecto

La primera parte consiste en realizar un EDA de nuestro dataset para prepararlo y que nuestros modelos lo puedan procesar. Para ello revisamos todos nuestros parámetros y el target y realizamos los cambios que sean necesarios. Por ejemplo, tratado de missings, outliers, cambio de variables a numérico, etc.

Una vez que tenemos el dataset limpio, la siguiente parte consiste en realizar un preprocesado de los datos y el aplicado de la separación de train y de test.

Por último, la elección de que modelos vamos a utilizar. En este caso, nuestro problema es de clasificación por lo que utilizamos modelos que nos sirvan para este cometido. Una vez que los modelos han entrenado se utilizan en la parte de test para determinar con que precisión han trabajado.

Librerías utilizadas en el proyecto

Para la lectura de datos y la limpieza de estos se ha utilizado Numpy y Pandas. Para la representación gráfica de los mismos se ha utilizado seaborn y matplotlib.

Para la parte de machine learning se ha utilizado la librería de sklearn. Dentro de esta se han utilizado diferentes herramientas y modelos. Entre ellos destacar regresión logística, árboles de decisión, random forest y xgboost.

Sources empleadas para la búsqueda de datos

Los datos utilizados en el proyecto han sido sacados de la plataforma kaggle. En concreto el dataset llamado Rain in Australia.

Link: <https://www.kaggle.com/jsphyg/weather-dataset-rattle-package>