23/10/2023

Aprendizaje supervisado

* Regresión: valores continuos
* Clasificación: target discreto

Unit 2 – Pandas

Pandas es una librería para tartar con datos y machine learning.

DataFrame es como una table con datos

Para concatenar dos dataframes se utiliza el método concat(). Aconsejable pasar el parametro ‘ignore\_index = True’

Guardar un DataFrame a disco: to\_csv

Cargar un DataFrame desde disco: read\_csv

Funciones importantes

Info() -> información básica general del data-frame. Te indica el tipo del dato, si hay registro con ese dato nulo

Shape() -> Numero de filas, numero de columnas

Head() -> x primeros registros

Tail() -> x últimos registros

Describe() -> de cada columna numérica te da datos estadísticos

corr() -> correlation de los datos. Solo los datos numéricos cuidado

loc() -> Indicar las filas y las columnas que se deberán mostrar

iloc() -> como loc() pero con índices de las filas

columns -> nombre de las columnas

rename() -> utilizar el parámetro ‘inplace=True’ para modificar sin hacer la copia

reindex() -> cambiar el orden de las columnas

insert() -> inserta una nueva columna, se le puede pasar un nombre y un índice

drop()

filtrado -> df.column == ‘test’

df.column.unique() -> saber los valores únicos de una columna

df.isnull() -> numero de nulos

dropna() -> eliminar los nulos

Nota: a las columnas se puede acceder con corchetes o con punto