Class_15.04.2023_ML_Workflow

April 17, 2023

ML Project Workflow

- Getting the data
    - (Amazon – USA) – 50 states – each state constitutes of 2 corporate office and 2 fullfil
    - 200 sites in total. all employees sending request to Amazon HQ – seattle
    - some site – through email / online forms etc etc .. data source here will be primarily
    - some site – handwritten paper
    - Tabular format data is directly going into your database
    - Unstructured data – converted first by the data engineers in place – tabular data will

- ETL (extract , transform , load)

    - Data clusters
    - (a-z) – dept 1 – (a-c) , dept 2 – (d-f)
    - x cluster – (a-c) , y cluster (d-f)

- Import the specific machine learning model – Sklearn/scikit learn
- model (training data , x training data , y training data)
- testing the model (x test data , y test data)
- the score of the model – (75% – 85%)
- evaluate the model ,
    - Regression problems – RMSE , MSE ,MAE (cost functions)
    - Classification problems – cross validation , confusion matrix
- Storing the model in a pickle file
- The pickle file will be sent for deployment
- Published in web , either open for public , or open to specific clients