MLOps Assignment

1. For each of the following scenarios explain the best matching algorithm among :

a. Regression

b. Classification

c. Clustering

d. Association Rule Mining

e. Unsupervised Feature Extraction

1. If we form segments among customers with different financial backgrounds, poverty levels, spending habits and income levels
2. Suppose the camera has to automatically recognize bus from car
3. Suppose you create a feature in facebook to automatically recognize your friends without labeling
4. If we are to identify the accomplices of a thief in a stealing incident based on the call records
5. If we are trying to predict future prices of stock based on volume of sales and difference between high and low
6. For the given data set boston.csv (for each measure MSE and R Sq) (Submit .ipynb files as attachment and screenshots )
   1. Find Discrete and Continuous variables
   2. Find descriptive analytics of all the variables
   3. Generate a linear regression program for the dataset (MV is the output value)
   4. Normalize the input fields and generate a SGD regressor for the same and find R – Square.
   5. Do feature selection by dropping pairwise high correlation fields and run the linear regression and SGD regression after removing irrelevant fields and dropping high corelated fields
   6. Run PCA on the inputs above and perform both SGD and Linear Regression with varying different no of components
   7. Perform one hot encoding of discrete variables and perform linear regression
   8. Do a optuna based hyperparameter search for KNN regression, Random Forest Regression and Ridge regression for the same data and check the best hyperaparameter values
   9. Develop a small web page in flask based on input data and predicting MV output with different input data given