***Project Description Document:***

***a.General Information on dataset:***

Dataset used: all\_perth\_310121.csv

Number of classes :13

Number of samples : 33656

Number of samples used in training : 28256.0

Testing: 5652.0

b. ***Implementation details:***

* *At feature extraction phase, how many features were extracted, their names, the dimension of resulted features.*

*1 feature ,average\_room,(* 33656,1)

* *Is cross-validation is used in any of implemented models? If yes, specify the number of fold and ratio of training/validation.*

*Yes, with linear regressor , number of fold=7 ,ratio of training/validation is near 1*

* *Hyperparameters used in your model, as initial learning rate, optimizer, regularization, batch size, no. of epochs, etc…*

*As KNN:*

*K=20*

*P=1*

*As KNN grid search:* params : {'n\_neighbors': 10, 'p': 1, 'weights': 'distance'}

best : KNeighborsRegressor(n\_neighbors=10, p=1, weights='distance')

***c.Results details:***

*For each model you should show all these results for your model on testing data (loss curve, accuracy, confusion matrix, ROC curve)*

*Linear regression->* Mde: 69958.36489201686

MSE: 30988708864.10156

RMSE: 176036.10102504987

R2 Score: 0.7561091306885459

reg score ,test : 0.7561091306885459

reg score ,train : 0.7688092825185864

*knn ->*training accuaracy: {82.3550946090696}

testing accuaracy: {78.54884103709976}

**LOGISTIC REGRESSION MODEL :**

IMAGE DATASET : <https://www.kaggle.com/datasets/yaswanthgali/english-fontnumber-recognition?select=Font>

DATASET INFORMATIONS:

DATASET NAME: english-fontnumber-recognition WE WORKED ON DIGITS CLASSES ONLY

Number of classes :10

Number of samples : 10,160

Number of samples used in training : 8128.0

Testing: 2032.0

b. ***Implementation details:***

***Feature Extraction Phase: hog feature extraction***

***Cross validation: yes***

***Number of Folds:*** 6

***Cv\_result: [0.99822904 0.99527745 0.99527466 0.99704666 0.99881867 0.99881867]***

***Hyperparameters:* random\_stateUse a new random number generator seeded by the given integer. Using an int will produce the same results across different calls**

**max\_iter :**

**For estimators involving iterative optimization, this determines the maximum number of iterations to be performed in fit.**

***Confusion matrix;***



***Precision and recall and f1\_score:***



***Roc curve and auc:***



***accuracy\_score:***



**K-means MODEL :**

IMAGE DATASET : <https://www.kaggle.com/datasets/yaswanthgali/english-fontnumber-recognition?select=Font>

DATASET INFORMATIONS:

DATASET NAME: english-fontnumber-recognition WE WORKED ON DIGITS CLASSES ONLY

Number of classes :10

Number of samples : 10,160

Number of samples used in training : 10,160

Testing: 2032.0

b. ***Implementation details:***

***Feature Extraction Phase: there is no feature extraction used***

***Hyperparameters:* random\_stateUse a new random number generator seeded by the given integer. Using an int will produce the same results across different calls**

**number of clusters :the choice of the number of clusters (K) is a critical decision in K-means clustering, and it often requires some exploration and analysis. There isn't a one-size-fits-all answer, and the optimal number of clusters depends on the characteristics of your data and the goals of your analysis.**

**kmeans clustering:**



**elbow graph:**



**Silhouette Score Plot::**

