MACHINE LEARNING AND PATTERN RECOGNITION

**LAB WEEK 6**

Import the following libraries:

*• Pandas*

*• Matplotlib.pyplot*

*• Numpy*

• Import the training and test datasets using *pandas*.

• Plot the training and test data sets having ‘Rating’ on the y-axis and ‘Duration’ on the x-axis. Also, label the data into the two classes ‘Action’ and ‘Drama’ and annotate each data point with the name of the movie:

Model Development and Evaluation:

• Import *KNeighborsClassifier* from *sklearn.neighbors*

• Import *confusion\_matrix, ConfusionMatrixDisplay* from *sklearn.metrics*

• Define a list: *knn\_neighbors* = [1, 3, 5, ……., 499]

• for each *k* in *knn\_neighbors*:

• apply the k-NN algorithm on the train set

• use the test dataset to the find the predictions

• use the test dataset and the predictions to get the confusion matrix for each *k*

• use the Confusion Matrix to calculate the accuracy for each *k*:

Accuracy = (True Negatives + True Positives) / Total no. of data points

• Plot Accuracy vs No. of Neighbors (*k*):

• Find the value of *k* for which the accuracy is the maximum. Print optimal k and corresponding accuracy.

• For that value of *k*, display the confusion matrix and calculate the performance metrics i.e. precision, recall, overall\_precision, overall\_recall and F\_score.

DELIVERABLES:

• Train dataset plot.

• Test dataset plot annotated with movie names.

• Accuracy vs No. of Neighbors plot

• Final confusion matrix and performance metrics(screenshot)