

LAB Assignments (Part - II)

Subject: Machine Learning

All the datasets are available in the following link:

<https://drive.google.com/drive/folders/15i6Zy9o3VwQZD823ZdJdmMuM6e4IcUUM?usp=sharing>

6. Using KNN algorithm, predict which category a customer belongs to on the basis of the data provided by a telecommunications firm. Find the accuracy of the KNN algorithm in predicting the category of a customer.

Dataset: teleCust.csv

7. Using Decision Tree algorithm, predict which drug among drug A, drug B, Drug C, drug X and drug Y should be given to a patient. Find the accuracy of the decision tree in predicting the correct drug for the patient.

Dataset: drug.csv

8. Using Naive Bayes algorithms, predict if a person is diabetic or not, based on the features provided. Find accuracy and F1-Scores of both algorithms.

Dataset: pima-indians-diabetes.data.csv

9. Using SVM algorithm, predict if a patient has a benign tumor or malignant tumor (cancer) based on the features provided. Use the following kernel for the SVM algorithm:

a) Linear b) Polynomial c) RBF d) Sigmoid

Find the following metrics for each of the SVM algorithms:

1) Accuracy 2) Recall 3) Precision 4) F1-Score
5) Jaccard Score 6) Error rates 7) Confusion Matrix

Compare all four SVM models using an ROC curve.

Dataset: samples_cancer.csv

10. Applying SVM, Naive Bayes, Decision tree and KNN predict diabetes based on features set. Compare the four classification algorithms with performance metrics such as accuracy, recall, precision, F1- score. Also design the heat map confusion matrix for above algorithms and construct ROC curve for comparison.

Dataset: pima-indians-diabetes.data.csv