Assignment – Classification

1. Problem Statement:

By using CKD this data we are going to predict, In Future the customer will be affect by chronic kidney disease or not??.

2. Basic Info:

The dataset is in Excel sheet and it is number so we use Machine Learning Domain. Input and Output is clear so we use Supervised Learning. Output is in Categorical Form so we use Classification.

Machine Learning--→Supervised Learning-→Classification.

This dataset has 399 rows and 25 columns.

3. Preprocessing Method:

In this data pc, pcc, ba, htn, dm, cad, appet, pe, ane, classification Columns are in string type. This name will not be comparable so we use Nominal method to change string into numerical data.

4. Good Model:

I have used all the algorithm of machine learning Like Logistic Regression, SVM , Decision Tree, KNN, Naïve Bayes and Random Forest. I got the best model creation in SVM for this data(Problem Statement).

5.Research Values:

A. Decision Tree: Accuracy =0.99
B. Random Forest: Accuracy=0.99
C. SVM: Accuracy=1.0
D. Logistic Regression=0.99
E. Naïve Bayes=0.98
F. KNN=0.94

6.My Final Model Creation done for this dataset is by using Support Vector Machine(C=10,'Poly').It has good model Without error when compared to other models.