Problem Statement

DLFH Finance company deals in home loans. They have offices across all urban, semi urban and rural areas. Customer first apply for home loan after that company validates the customer eligibility for loan.

Company wants to automate the loan eligibility process (real time) based on customer detail provided while filling online application form. These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and others. To automate this process, they have given a problem to identify the customers segments, those are eligible for loan amount so that they can specifically target these customers. Here they have provided a sample data set. This data set is related with a mortgage loan and challenge is to predict approval status of loan .Use KNN algorithm to predict approval status of loan.

Data Files :

1. loan train data : "loantrain.xlsx" ,sheet : "loantraindata"

2. variable description : loantrain.xlsx ,sheet : "variabledescription"

3. loan test data : "loantest.csv"

Follow modelling process to build KNN model.

Questions to be answer:

1. Do you think normalization/statndardization of variables required for the model building.

2. what is the optimal K value.

3. Mr Pradeep used chebyshev distance method to build a KNN classifier.Fortunately he got a accuracy rate more than 95%.Do you think his model is a good model and why?