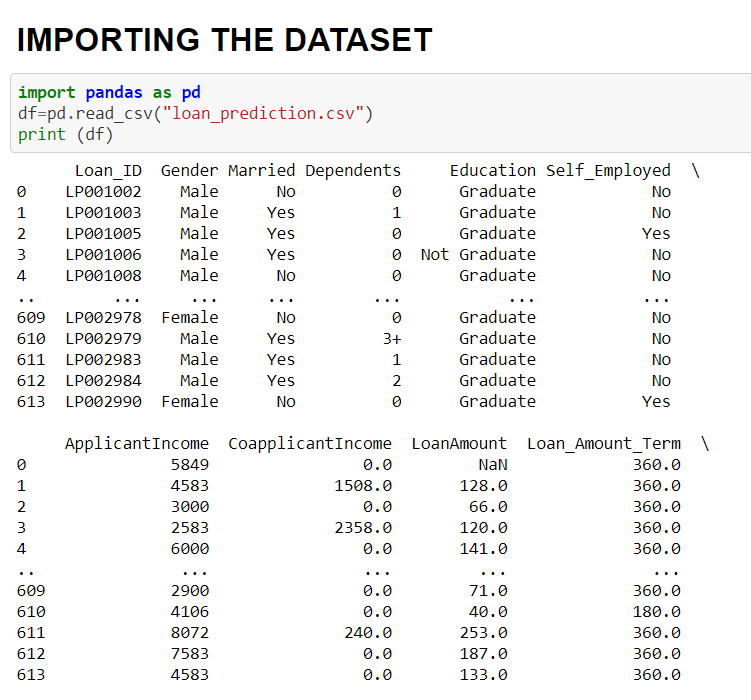
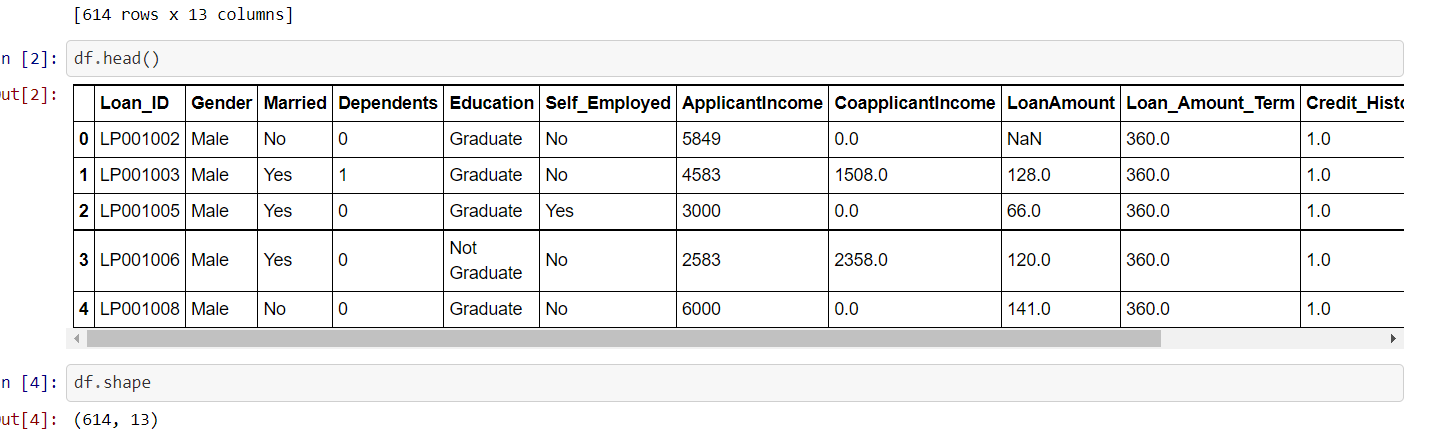
**PROBLEM DESCRIPTION:**

The **goal is to predict the loan status type.**

****

**We have imported the pandas library and csv file is loaded into a variable i,e df . Other various libraries such as ,numpy, Label Encoder, seaborn,train test split,randomforest classifier,svc,dtc etc for various aspects such as encoding,model building etc.**

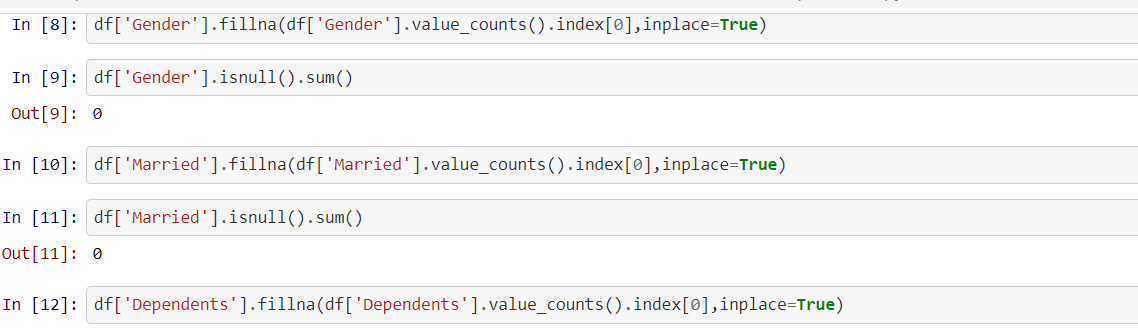
****

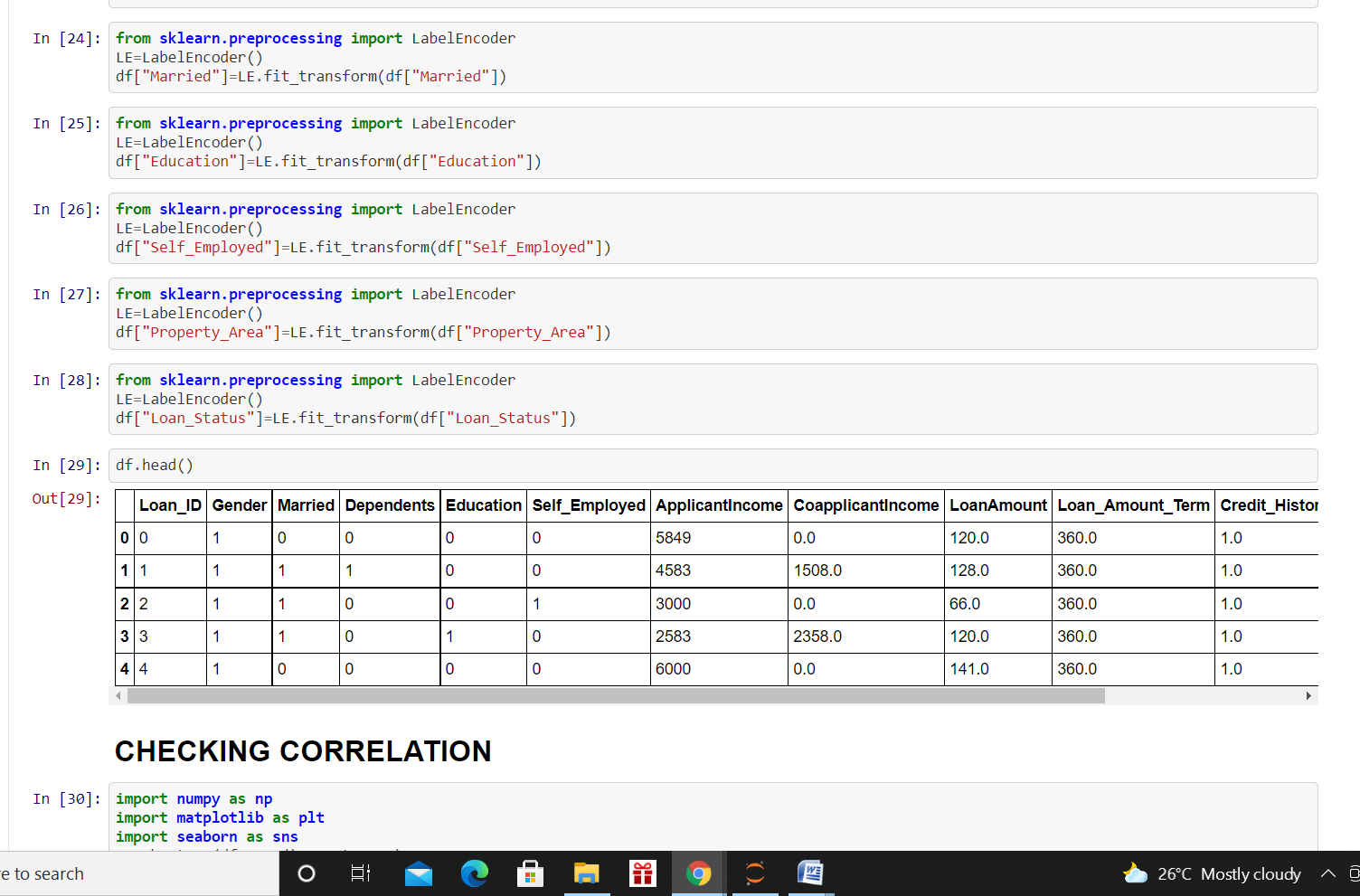
**I am processing the first 5 rows of different columns and checked the shape .We have found that the dataset contains 614 rows and 13 columns.**

**DATA ANALYSIS:**

**At first we have checked the nullvalues and found that gender,married,dependents,self employed,loan amount,loan amount term,credit histry these columns have null values of the data set contains null values.Using the fillna function we have filled the null values.We have seen that the dataset contains categorical columns so using label encoder we have encoded the categorical columns.**

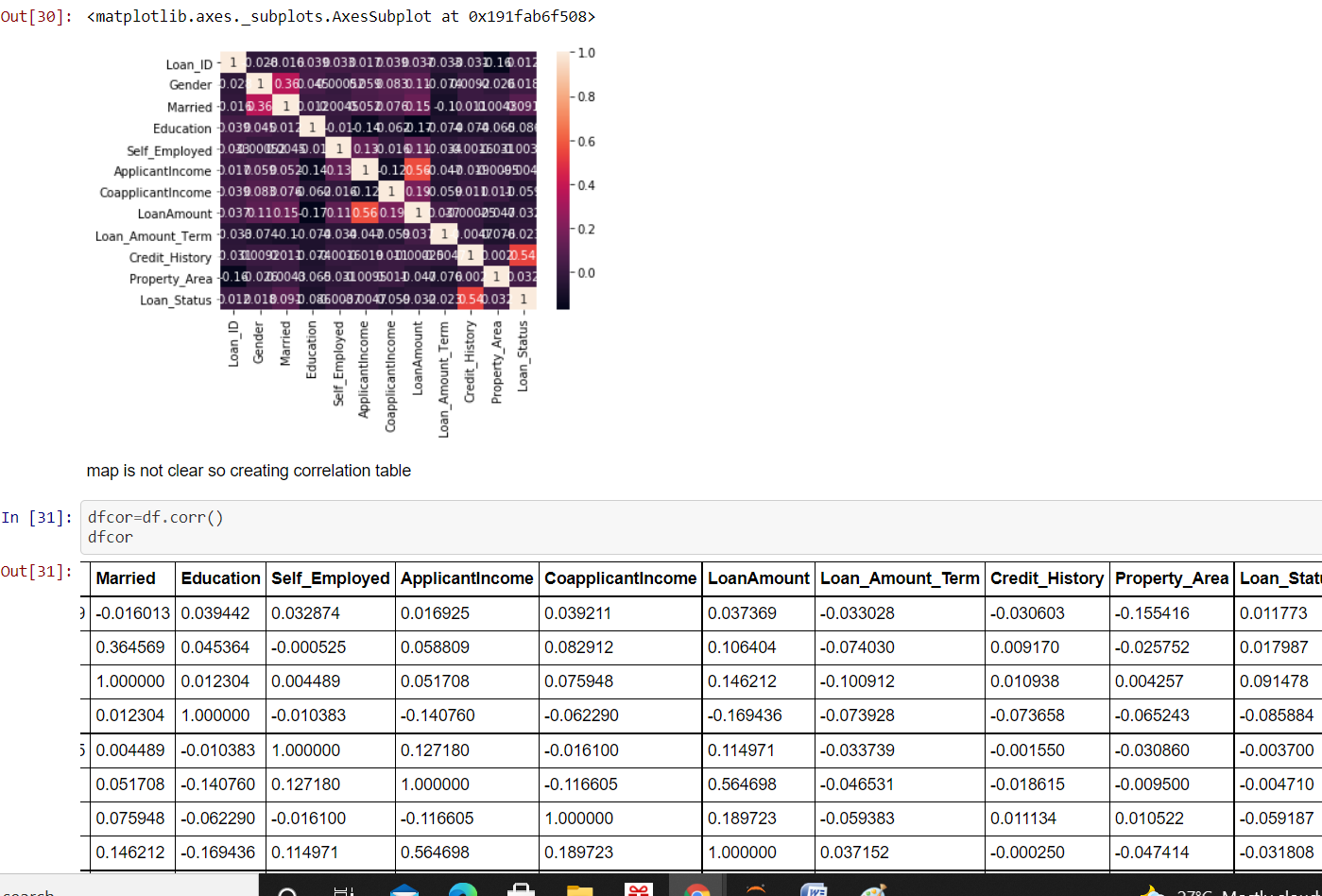
**We checked the first five columns and checked the steps performed.**

****

****

**We have checked the correlation of different variables using heatmap. At first we have imported the matplotlib and seaborn function for visualisualing the data in form of map**

**In this map the black colour indicates low correlation values and violet colour indicates variables which have correlation values from 0.2 to 0.4 pink colour indicates which have correlation values in the range 0.4 to 0.8 orange colour indicates who have some good correlation values but the map is very indistinct so we are using the corr function.From this table we have noticed that the correlation values of the columns with each other is very low and there correlation with the target variable is also very weak.**

****

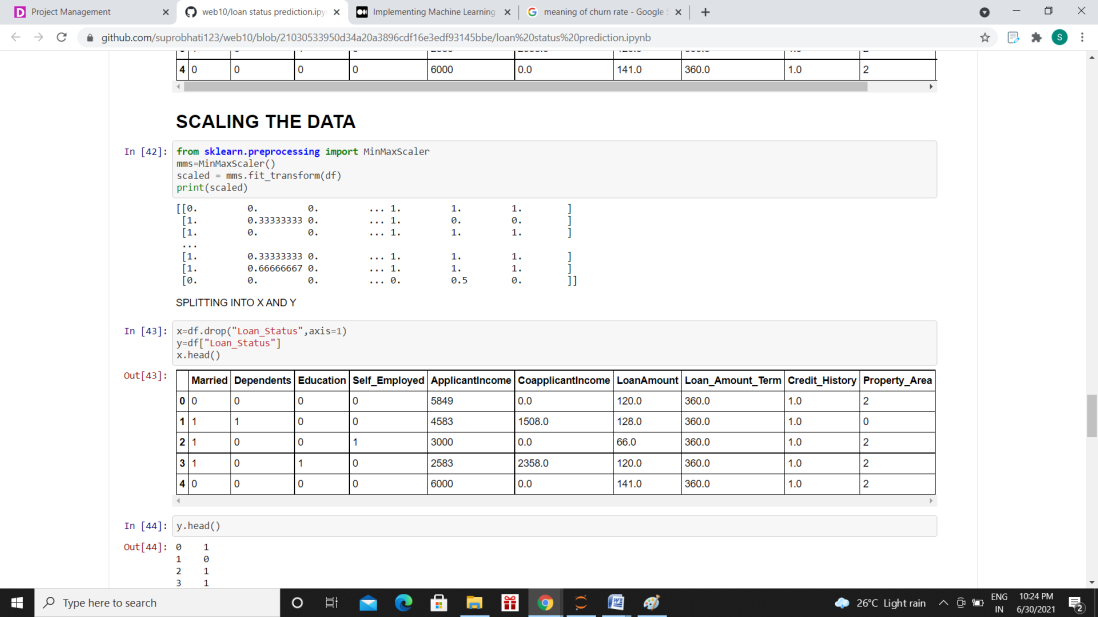
**We have checked the data types again and noticed that dependents column has special ‘+3’ character so using the string.replace function we removed it. We have checked the data types again .all are int and float type of variable .Now the data is ready for further processing.**

**EDA CONCLUDING REMARKS:**

**The datest contains both categorical and numerical data which has 614 rows and 13 columns. Most of the columns have null values that’s why we are using fillna function to fill the null values. After encoding the categorical columns we are using heatmap to see the correlation function . We have seen from the correlation table that most of the columns have week correlation values with the target variable .**

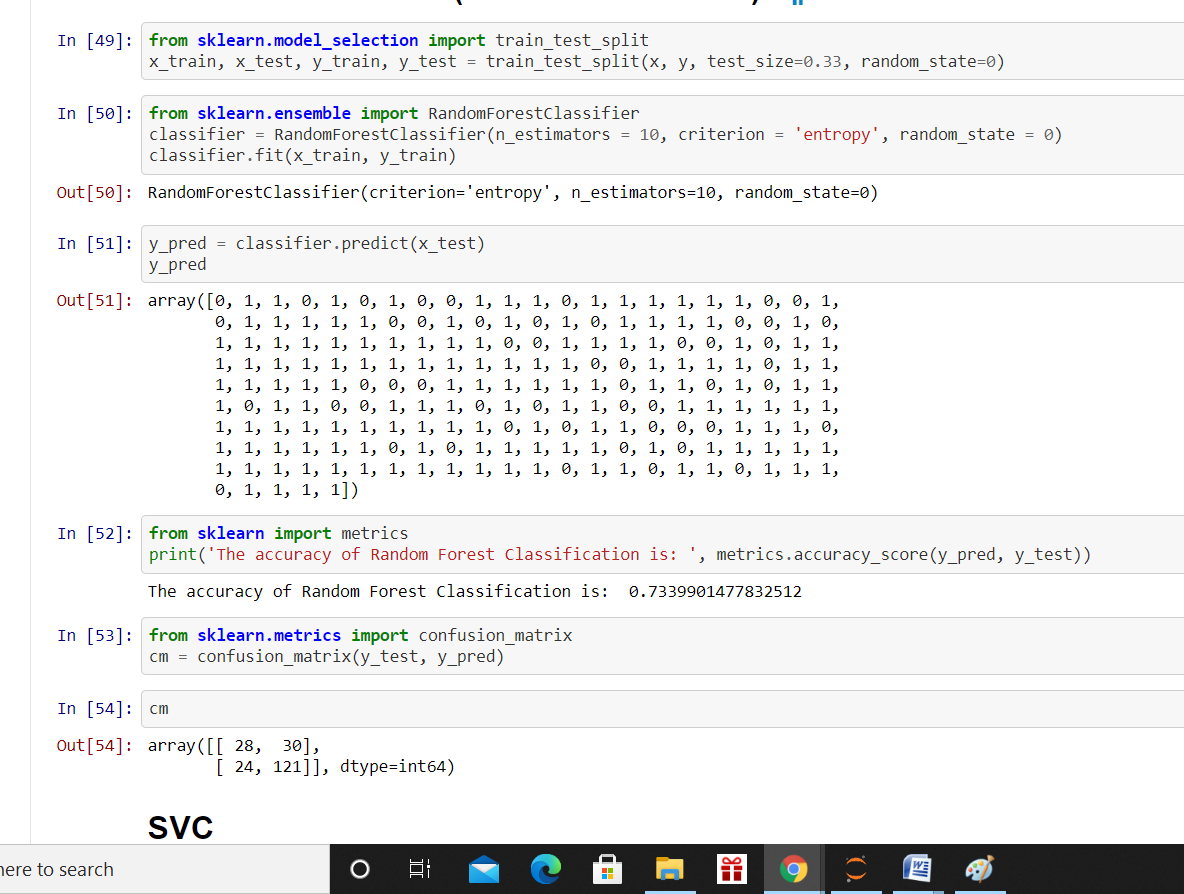
**DATA PREPROCESSING:**

**Before building various models to predict the target variable we are scaling the data using min max scaler .We are importing the min max scaler from sklearn.preprocessor then using fit\_transform function we have scaled the dataset. Then we have splitted the data into x and y column from that with respect to the target variable. After this part we are using various models to predict the target variable I,e loan status.**

****

**MODEL BUILDING:**

**We have splitted the data into train and test data before using various models.**

****

**From sklearn.model\_selection we are importing the train test split function and using it to split the dataset. Then we are importing random forest classifier to predict the target variable .We are using the classifier.predict function to predict the target variable.We are importing the metrics function to check the accuracy state of the model. We are getting 73% accuracy of the model .Then we have imported the confusion\_matrix function and printed the confusion metrics.**

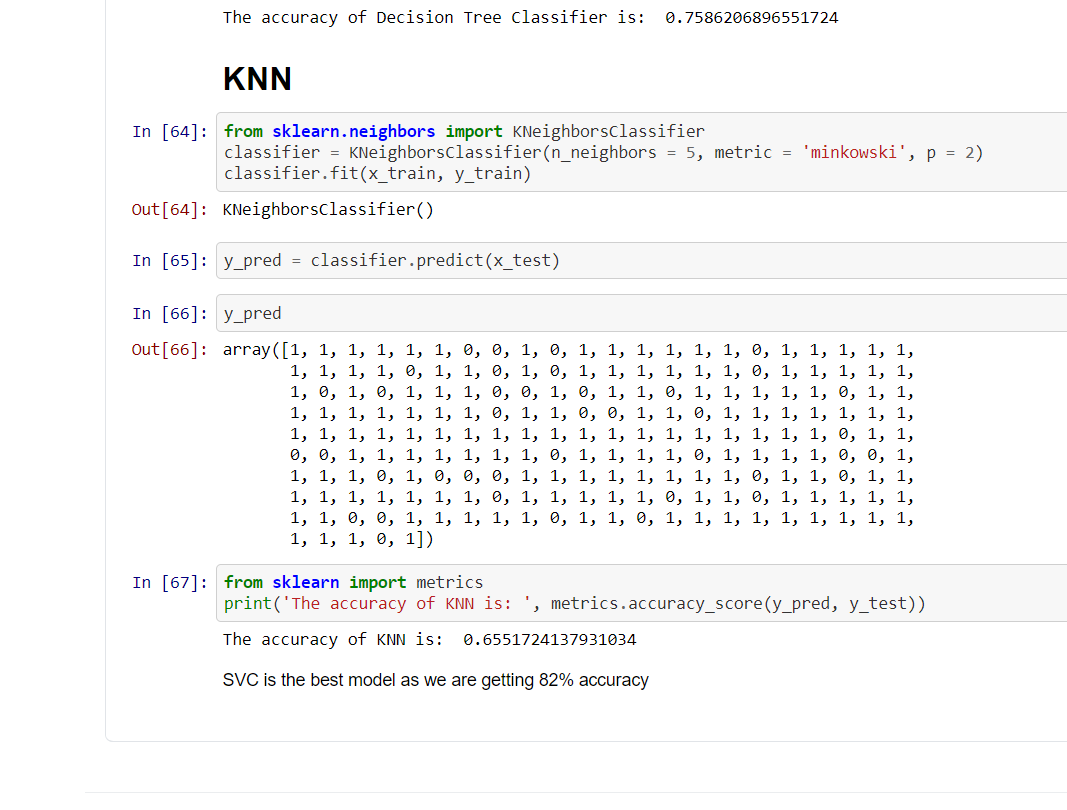
**From sklearn.svm we have imported SVC i,e support vector classifier .Using classifier.fit we have used the model in our train data. We have predicted the y values using classifier.predict function. Then importing the metrics function we are testing the accuracy stae of our model which is 82% .Then we have created the confusion matrics.**

**From sklearn we have imported tree .sklearn.tree we have imported the decisiontreeclassifier .Using classifier.fit function we are applying decision tree algorithm to predict our target variable. We are using classifier.predict function to determine the y values.**

**We are getting the accuracy state of the model as 75%.**

**Again we are using sklearn.neighbours to import the k nearest neighbor algorithm .We are predicting y values using the algorithm and predicting the accuracy state which is 65%.**

**We are using support vecter classifier as the best model to predict the y values which has the highest accuracy state whch is 82% . So we are using this model for futher prediction.**

****

**CONCLUSION: Following are the EDA ,data preprocessing ,and model building part of my project.As we are getting the SVC algorithm as the best model for predicting the target variable we are using this algorithm to predict loan status type.**