LOAN STATUS PREDICTION

The objective of the project is to build a Machine Learning Model to predict the loan to be approved or to be rejected for an applicant based on his/her Income, Education, Credit history and many other factors.

* Analysis of the train data was made initially, from which the feature engineering ideas can be identified.
* Some of the features were having Null values and in this project those values have been imputed and made not null.
* Encoding has been done to the categorical features to proceed further with the project and the unwanted feature that is not required for the prediction have been dropped from the input train data set.
* All the above necessary changes were made in the Feature Engineering part to make the data set balanced to proceed further with the project.
* In the feature observation part the correlations between the features have been identified and in which Credit history holds the primary significance.
* The top five features that have significant role with the prediction of the Loan Status were used in the model building part.
* Here we have used both Logistic Regression and Decision Tree Classifier for the building of the model. In which Logistic Regression had accuracy values of 80% and 85% for train and test. For Decision tree it had values of 82.39% and 83.11% for train and test.
* We have used the decision tree model to predict the test data results since they have nearly identical accuracy values indicating a decent performance by the model.
* With which the project was completed.

Note:

The model can be further improved. More careful analysis in the feature engineering part can aid in providing more balanced train set data that can be used in the model building.

Other classification algorithms like Random Forest, XGBoost can even provide better results which can again improve the model.

By incorporating the above mentioned possibilities better models can be definitely developed which again improves the Scope of the Loan status prediction.