

# ***Bank Loan status prediction using machine learning***

(by, Ankit Kumar(2018018), Arbaaz Khan(2018023) and Mayank Gupta(2018157))

## **MOTIVATION.**

The financial area has advanced a ton with numerous individuals applying for loans. Approval for loan is a vital cycle for banking associations, as advance reimbursement is a significant contributing variable in the bank's budget or financial statement.

1. Loan approval for clients who will not be able to pay it back may affect the economy of our country that's why predicting it right is an important thing to banks.
2. Leaving this choice for bank representatives can be a very ineffective and time-consuming idea.

## **Dataset Used and Preprocessing.**

For the prediction of the bank loan status, we are going to use a dataset that is available on Kaggle as "Bank Loan Status Dataset". To fit the dataset according to our need we will be mapping string values to the numerical values. Also, we need to handle the null/ NA values that are present in the dataset.

## **Learning Techniques.**

Currently, as we have planned we are going to use the given machine learning algorithms to train our model. As a reference we are going to stick with Logistic Regression further we will be testing our model for Random Forest, Decision Tree and Neural Network (Sequential).

We are planning to use the following techniques as there are very few categorical variables in the dataset and Random Forest works well with these types of datasets. Also, the size of the initial dataset is quite large which makes it preferable to use Neural Network techniques.

## **Evaluation Metrics and Error analysis.**

For the given dataset we will be using the binary classifiers to predict whether the person with the provided details will be able to repay his/her bank loan completely or not. To find the accuracy of the model we will be using different machine learning models to find the accuracy and analyse the results from the same. For the analyses of the accuracy and errors from different models we will be finding them: 1.**Accuracy**, 2.**Precision**, 3.**Recall**, 4.**F1-Score**, 5.**AUC-Score**.

## **Deliverables.**

Working code(ipynb), dataset and report

By Ankit: Preprocessing, Base model working, Neural network implementation, Data analysis.

By Arbaaz: Preprocessing, Random forest model implementation, Data analysis.

By Mayank: Preprocessing, Decision Tree implementation, analysis of different methods, Data analysis.