Problem Definition:

Credit/Loan underwriting is a tedious process involving analysing various aspects/features of the borrower manually. The process is lengthy, time-consuming and prone to human error thereby resulting in defaults and losses for the financial institution.

The machine learning model can be used in A**utomating Credit Decisions** thereby reducing loan officers’ data input burden with less time spent on each loan application and lower risk of human error. Machine learning provides a faster, accurate and cost-effective solution to the credit underwriting approvals that can result in minimizing defaults and losses for the financial institutions.

Objective

**Objective** **of the project** is to develop a machine learning model that can automate the loan approval process. **Objective** **of the model** is to predict if the loan application can be approved basis the multiple inputs including Age, Gender, Income, Education and Number of dependents, Self Employed (complete list below) to predict the output Loan\_Status which is approved or not.

X= **Input Variables**

**y** = Loan\_Status

Input Variables

* Loan\_ID
* Married
* Dependents
* Education
* Self\_Employed
* ApplicantIncome
* CoapplicantIncome
* LoanAmount
* Loan\_Amount\_Term
* Credit\_History
* Property\_Area

Machine learning model

This is a classification problem with binary output so Logistic regression, Decision trees and Random Forest classier can be utilized to analyse the model (supervised learning) can be used. Then, we choose which model gives the best result.

**Logistic regression** model evaluates the probability of an occurrence. If the probability is higher than 0.5, the application is classified as approved, or if the probability is less than that, the application is classified as denied. There is a strict cutoff line that divides one classification from the other.

In logistic regression, the target

Decision Tress:

Decision Tree (DT) like the Logistics Regression (LR) both solve the Classification Problem. Some benefits of decision tree are:

1. Categorical data works well with Decision Trees, while continuous data work well with Logistic Regression.
2. Decision Trees handle skewed classes nicely if we let it grow fully.

Random Forest

We also plan to use random forest model as decision trees sometimes are prone to overfitting and single tree may become complex and prone to errors, but many of them can be combined to form a stronger model. One nice byproduct of the random forest algorithm is to rank the features by their importance, which allows us to see which features have the most impact on the decision.