# Project Proposal

**COMP 7118**

**Data Mining**

**Fall 2022**

**Ramesh Kumar Surapuredy**

**U00831213**

**Contact email: rsrprddy@memphis.edu**

**AND**

**Hemanth Venumbaka**

**U00870132**

**Contact email: hvnmbaka@memphis.edu**

**What is the problem/application?**

Loan Prediction – To predict the loan amount

**What methods will be tested or implemented?**

In this project we are going to predict the status of a loan by considering the features of a customer like Education, dependencies, application income and co application income .. etc. By using decision tree classifier, random forest classifier and logistic regression we are going to implement the project by comparing the models. And we also use data mining techniques to preprocess the data and for filtering the data.

**What data sets will be used?**

We are using a Loan prediction Dataset from Kaggle. The features of a dataset are `'Loan_ID', 'Gender', 'Married', 'Dependents', 'Education', 'Self_Employed', 'ApplicantIncome', 'CoapplicantIncome', 'LoanAmount','Loan_Amount_Term', 'Credit_History', 'Property_Area', 'Loan_Status'.`

Dataset: Loan_Prediction_dataset

**What are the potential challenges for implementation?**

Feature engineering and choosing the right algorithm are the potential challenge for this implementation.

**What are the expected deliverables?**

The expected result is to determine whether the customer is eligible for the loan or not. As we observed, linear regression is performing well when compared to the other models.

**Conclusions:**

This project is mainly focused on predicting the status of a loan. From this project we came across many data mining and feature engineering techniques.