Loan Prediction

(PHASE 0)

**Submitted By :**

**Kollisetty Pushpak (AM.EN.U4CSE17308)**

**Tungala Yashwanth (AM.EN.U4CSE17340)**

**Yerramsetti Raghuram (AM.EN.U4CSE17359)**

**ABSTRACT**

Loan prediction is a very common real-life problem that every bank faces at least once in its lifetime. If done correctly, It can save the company by giving loans to only the creditworthy customer.

**Formal description**

A computer program is said to learn from experience E with respect to some class of tasks T and

performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

For example:

• Task (T): Classify a customer whether the loan is approved or not.

• Experience (E): Learns from the past whether the pre -predicted loans have been paid or not.

• Performance (P): Classification accuracy, the number of customers loans predicted correctly out of all customers loans considered as a percentage.

**Assumptions:**

Assumptions are that if successfully implemented, it would reduce people's work and make the process fast with more accurate results.

**1. INTRODUCTION:**

Motivation :

This problem needs to be solved because it would make the process of giving loans automated, which will be very fast and easy. Otherwise, people have to wait for longer intervals of time to get their loans approved or not approved. With this prediction system, a lot of time can be saved too. In some other cases, some people give loans to customers(for some other reasons like by taking bribes, etc..) who can not repay them and this will lead to many problems.

So, in order to prevent all these things, this loan prediction process needs to be automated. This is our motivation in solving the problem.

**Benefits of the solution :**

This is a lot of benefits to the people and employees of the company as mentioned above as it would save a lot of time and work of the people. Also, it would make it supportive to say that technology has advanced a lot.

While talking about personal advantages, it would be useful for us(if implemented successfully) when we take loans. As computers are much faster, it would increase the overall pace of the entire process.

**Solution use:**

With the help of this model, we can identify the persons who are creditworthy. Using this prediction we can save the banks from loan defaulters.

## **2. Dataset finalization**

The dataset for this project is retrieved from Kaggle, the home of Data Science.

1. What is the data about?

The dataset that we are using for this python project will be stored in a CSV file. The dataset contains a corpus of news with features like id, Self\_Employed, LoanAmount.

The dataset for this project is retrieved from Kaggle, the home of Data Science.

[data-set 1](https://www.kaggle.com/altruistdelhite04/loan-prediction-problem-dataset?select=test_Y3wMUE5_7gLdaTN.csv) : https://www.kaggle.com/ninzaami/loan-predication

[data-set 2](https://www.kaggle.com/altruistdelhite04/loan-prediction-problem-dataset?select=train_u6lujuX_CVtuZ9i.csv) : https://www.kaggle.com/ritikmeena/prediction-of-loan-status

[data-set 3](https://www.kaggle.com/zaurbegiev/my-dataset?select=credit_test.csv) : https://www.kaggle.com/itsmesunil/bank-loan-modelling

2. What is the number of features and describe each of the features and explain the importance?

|  |  |
| --- | --- |
| **Variable** | **Description** |
| Loan\_ID | Unique Loan ID |
| Gender | Male/ Female |
| Married | Applicant married (Y/N) |
| Dependents | Number of dependents |
| Education | Applicant Education (Graduate/ Under Graduate) |
| Self\_Employed | Self-employed (Y/N) |
| ApplicantIncome | Applicant income |
| CoapplicantIncome | Coapplicant income |
| LoanAmount | Loan amount in thousands |
| Loan\_Amount\_Term | Term of the loan in months |
| Credit\_History | credit history meets guidelines |
| Property\_Area | Urban/ Semi Urban/ Rural |
| Loan\_Status | Loan approved (Y/N) |