

# LOAN ELIGIBILITY PREDICTION THON PROJECT



## Loan Eligibility Prediction

---



---

Feynnlabs

Authored by: Chandrashekar R

---

# Abstract

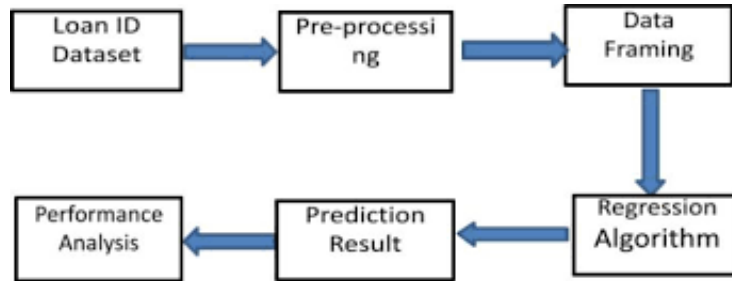
In the banking system, banks have a variety of products to provide, but credit lines are their primary source of revenue. As a result, they will profit from the interest earned on the loans they make. Loans, or whether customers repay or default on their loans, affect a bank's profit or loss. The bank's Non-Performing Assets will be reduced by forecasting loan defaulters. As a result, further investigation into this occurrence is essential. Because precise forecasts are essential for benefit maximization, it's crucial to analyze and compare the various methodologies. The logistic regression model is an important predictive analytics tool for detecting loan defaulters. The two most pressing issues in the banking sector are: 1) How risky is the borrower? 2) Should we lend to the borrower given the risk. Banks make loans to customers in exchange for the guarantee of repayment. Some would default on their debts, unable to repay them for a number of reasons. The bank retains insurance to minimize the possibility of failure in the case of a default. The insured sum can cover the whole loan amount or just a portion of it. At that time, making a decision would take a long time. As a result, the loan prediction machine learning model can be used to assess a customer's loan status and build strategies. This model extracts and introduces the essential features of a borrower that influence the customer's loan status. Finally, it produces the planned performance (loan status). These reports make a bank manager's job simpler and quicker.

## 1. Problem Statement

Loans are the core business of banks. The main profit comes directly from the loan's interest. The loan companies grant a loan after an intensive process of verification and validation. Even after this process, the banks are not assured of loan recovery. They face a huge loss as the borrower delays or completely ignores to pay the loan taken from the bank. This may arise due to many reasons some of them would be misinterpretation or favoritism with the borrower or collectively can be considered as human error. To overcome this problem, we can develop a model which predicts whether the borrower would pay the loan on time using Machine Learning. When seen from the borrower's side even, there are many instances where the borrower is capable of repaying the loan promptly but the loan gets rejected due to some orthodox decisions where the loan providing is decided on some single factor.

## 2. Business Need Assessment

As mentioned above, Loans are the primary income for the banks as their main profit comes directly from the loan's interest. If the Machine Learning model is able to predict the extent of the loan recovery from the individual efficiently, Then the banks would have a better profit as they would assess the data of the borrower and check the prediction of the model, decide whether to provide loan to the borrower or not.



Block Schematic of Loan Prediction

### 3. Target Specification and Characterization

The target audience is the banks that provide the loan. Most of the banks would be willing to provide as much as loans as possible which would increase their profit but recovery is always the biggest fear which makes them a bit more conscious in providing the loan.

Factors considered in this model are-

- Gender (M or F)
- Marital status (Yes or No)
- Dependents (number of persons dependent)
- Education (education of the applicant)
- Self-employed (Yes or No)
- Co-applicant income (income of the-applicant)
- Loan amount (required amount of the borrower)
- Loan amount term (term duration)
- Credit history (ability to repay the loan)
- Property area (urban, semi-urban or rural)

### 4. External Search

The model uses a dataset from a repository on GitHub which consists of 614 rows of data that shows on what basis the loan was provided to the applicant. Based on this data the model would predict whether the applicant (new data) is eligible for the loan or not.

<pre> 1 train = pd.read_csv(r"train_data.csv") 2 train.head() </pre>													
✓ 1.1s													
	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status
0	LP001002	Male	No	0	Graduate	No	5849	0.0	NaN	360.0	1.0	Urban	Y
1	LP001003	Male	Yes	1	Graduate	No	4583	1508.0	128.0	360.0	1.0	Rural	N
2	LP001005	Male	Yes	0	Graduate	Yes	3000	0.0	66.0	360.0	1.0	Urban	Y
3	LP001006	Male	Yes	0	Not Graduate	No	2583	2358.0	120.0	360.0	1.0	Urban	Y
4	LP001008	Male	No	0	Graduate	No	6000	0.0	141.0	360.0	1.0	Urban	Y
<pre> 1 test = pd.read_csv(r'test.csv') 2 test.head() </pre>													
✓ 0.2s													
	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	
0	LP001015	Male	Yes	0	Graduate	No	5720	0	110.0	360.0	1.0	Urban	
1	LP001022	Male	Yes	1	Graduate	No	3076	1500	126.0	360.0	1.0	Urban	
2	LP001031	Male	Yes	2	Graduate	No	5000	1800	208.0	360.0	1.0	Urban	
3	LP001035	Male	Yes	2	Graduate	No	2340	2546	100.0	360.0	NaN	Urban	
4	LP001051	Male	No	0	Not Graduate	No	3276	0	78.0	360.0	1.0	Urban	

## 5. Benchmarking alternate products

There are many independent loan eligibilities predicting firms and banks also maintain their own system to predict whether an applicant is eligible to receive the loan or not. One such independent Organization is “allcloud.in”. It provides various functionalities in which the “Loan Originating system” is such program which gives streamlines processes across various channels and products thus enabling banks to issue loans and manage underwriting services in a few minutes, paperless.

## 6. Applicable Patents

The organization “allcloud.in” is open-source software and it is free to integrate with any other platforms.

## 7. Applicable Constraints

- All necessary libraries in python like pandas, NumPy, sci-kit learn, and matplotlib to execute the code cell is required
- A cloud space to store, update and retrieve the datasets
- An AI/ML expert body to monitor the working of the model and improvise it timely for better performance
- An easy-to-use UI for the client
- Continuous data collection and maintenance.
- Product usage can be found not so reassuring as models have low accuracy, Optimization is required in this case.
- Marketing the product to a Training center or a Software development company.
- The product developed should be responsive such that customers who are not educated will be able to easily use the product.

---

## 8. Business Opportunity

If a company provides a better prediction of loan eligibility than the banks then most of them would prefer using the services from the company as their profits increase as the loan defaulters are kept as minimum as possible. This would increase the company's engagement with the banks and hence a good revenue is generated as the banks trust the predictions of the company.

## 9. Concept Generation

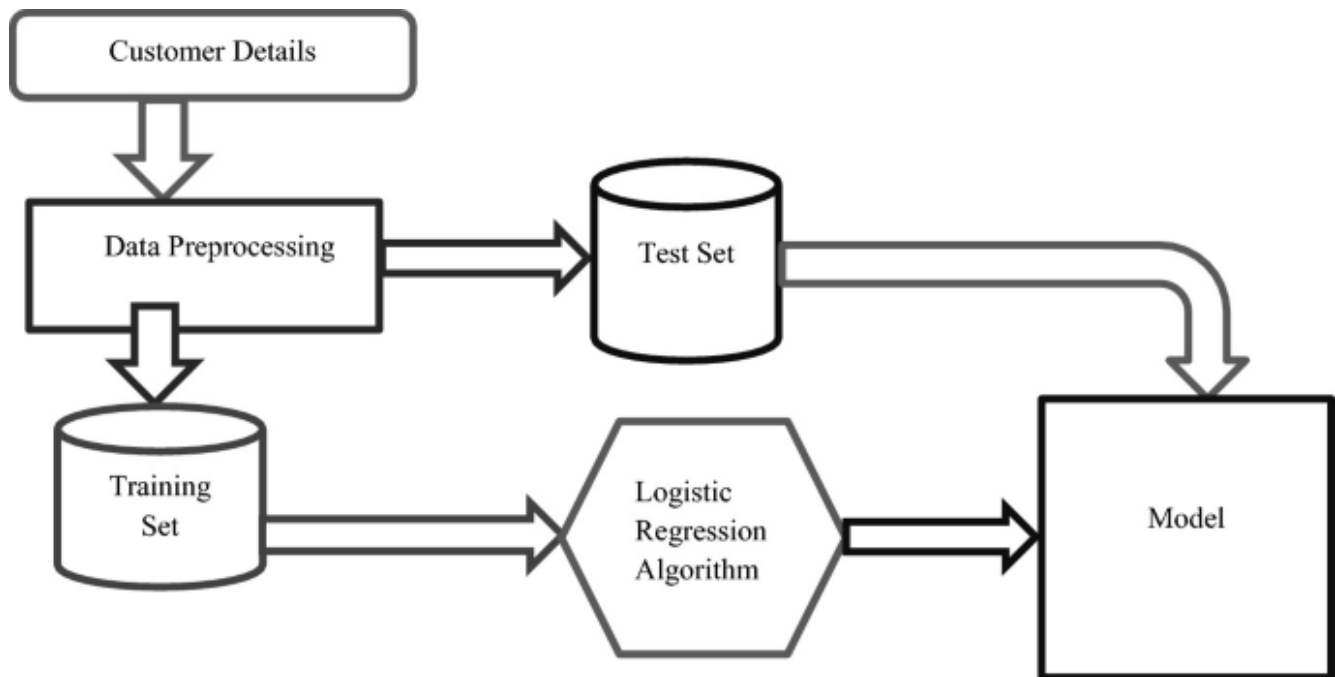
The term "Loan" is both a give and take thing, Banks need to keep providing loans to increase their income of it and the applicant would require a loan for his/her personal reasons. Hence an easy way to help both banks and applicants in providing and getting loans respectively would be by using this machine learning model.

## 10. Concept Development

Firstly, we collect datasets of all types of applicants, remove all the unwanted data and train the machine learning model with that refined data. Once the model is trained, we can take the values from the user (applicant details are entered by the bank) and give an output of whether the applicant is eligible for the loan or not. We can divide our project into two phases First phase: In this phase, our model is more helpful to banks than the applicants as it only predicts whether the applicant is eligible for getting the loan or not. Second phase: In this phase, the model takes the input, gives the output (whether the applicant is eligible for the loan or not), and also gives the possible ways for the applicant to get the loan (by decreasing the loan amount or increasing the interest, etc).

## 11. Final Product Prototype

Schematic Diagram: The schematic diagram for the above model can be



## Product Details

- **Working**

The machine learning model works using logistic regression which classifies the details entered by the user (applicant details) into whether the applicant can receive the loan or not

- **Data Sources**

We use datasets for training and testing the model from Kaggle

- **Algorithms**

We use the Logistic Regression algorithm as it gives the best accuracy among all other algorithms when tested with the same test data set (83%)

- **Team required to develop**

To develop this project, we need a team that consists of members from the Machine Learning field and banking field where the majority can be members from the Machine Learning field.

---

## Business Model:

For the Business model, we can implement any of the three business models given below

- Direct sales Bm
- Consulting bm
- Freemium bm can be implemented.

In direct sales, we give the customer the product like we give access to the website where he gives inputs and can be sure whether his application for the loan will get approved or not. Consulting is also similar but here he gives input to us and we use the model to give him our opinions on how he improves his situation to get loan approval from the bank. Freemium: we give some access (i.e., some no of inputs accessed out of the total no of inputs our model uses) to our final product based on that he will / not decides to buy our product. This is from a customer point of view; we can sell it to banks which then use it to predict whether a customer pays or not.

## Financial Modelling:

Financial modeling is a technique that builds models that help to forecast a business's financial performance in the future. The forecast is based on assumptions about the future, and the company's historical performance, and requires the preparation of financial statements, the company's expenses, and earnings and may include their supporting schedules. Financial models are tools that are helpful in making business decisions. For: in IPO, whether or not to raise money. This type of financial model involves looking at comparable company analysis in conjunction with an assumption about how much the investors would be willing to pay for the company. The valuation in an IPO model includes "an IPO discount" to make sure the stock trades well in the secondary market. Often used by Investment bankers and corporate development professionals. We are expecting a linear increase in sales for the next 3-4 years based on the situation that our sales might increase exponentially.

## Conclusion

This model is very useful in the banking sector for swift processing of loan eligibility for the applicants. After the Final Submission of test data, my accuracy score was 80%. Feature engineering helped me increase my accuracy. The predictive models based on Logistic Regression, Decision Tree, Random Forest, and Grid Search give the accuracy as 80.12%, 70.84%, 80.12%, and 80.77

GitHub Link

<https://github.com/Chandhucp/FeyNN-Labs/blob/main/Task3/Loan%20Eligibility%20Prediction.ipynb>