LOAN APPLICATION ANALYSIS

A Project Report

Submitted in the partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology in

Department of Computer Science and Engineering

By

M.Rushitha Sri (2010030362)

|  |  |
| --- | --- |
| K.Kavya | (2010030550) |

under the supervision of Dr. K. Sreenivasa Rao Assistant Professor,CSE



Department of Computer Science and Engineering

K L University Hyderabad,

Aziz Nagar, Moinabad Road, Hyderabad – 500 075, Telangana, India.

March, 2022

# DECLARATION

The Project Report entitled “LOAN APPLICATION ANALYSIS” is a record of the bonafide work of M. Rushitha sri (2010030362), K. Kavya (2010030550), submitted in partial fulfillment for the award of B. Tech in the Department of Computer Science and Engineering to the K L University, Hyderabad. The results embodied in this report have not been copied from any other Departments/University/Institute.

M. Rushitha Sri (2010030362)

K. Kavya (2010030550)

# CERTIFICATE

This is to certify that the Project Report entitled “LOAN APPLICATION ANALYSIS” is being submitted by M. Rushitha Sri (2010030362), K. Kavya (2010030550) submitted in partial fulfillment for the award of B. Tech in Computer Science Engineering to the K L University, Hyderabad is a record of bonafide work carried out under our guidance and supervision.

The results embodied in this report have not been copied from any other departments/ University/Institute.

**Signature of the Supervisor**

Dr.K.Sreenivasa Rao

Assistant Professor

**Signature of the HOD Signature of the External Examiner**

# ACKNOWLEDGEMENT

It is great pleasure for me to express my gratitude to our honorable President **Sri. Koneru Satyanarayana**, for giving the opportunity and platform with facilities in accomplishing the project-based laboratory report.

I express my sincere gratitude to our Principal **Dr. L. Koteswara Rao** for his administration towards our academic growth.

I express sincere gratitude to our Coordinator **Mr. K. Sreenivasa Rao** for his leadership and constant motivation provided in the successful completion of our academic semester. I record it as my privilege to deeply thank you for providing us with the efficient faculty and facilities to make our ideas into reality.

I express my sincere thanks to our project supervisor **K. Sreenivasa Rao** for his/her novel association of ideas, encouragement, appreciation, and intellectual zeal which motivated us to venture into this project successfully.

Finally, it is pleased to acknowledge the indebtedness to all those who devoted themselves directly or indirectly to make this project report success.

M. Rushitha Sri

K. Kavya

|  |  |
| --- | --- |
| **CHAPTER NO** | **CONTENT** |
| **1** | **INTRODUCTION** |
| **2** | **ABSTRACT** |
| **3** | **EXISTING WORK** |
| **4** | **PROBLEM STATEMENT** |
| **5** | **PROPOSED WORK** |
| **6** | **HARDWARE AND SOFTWARE REQUIREMENTS** |
| **7** | **DATASET** |
| **8** | **METHODOLOGY** |
| **9** | **RESULT** |
| **10** | **CONCLUSION** |

**INTRODUCTION**

Banks are vitally crucial in a market economy. The bank assesses if a borrower is excellent or bad before issuing credit as (defaulter)/(non-defaulter). Predicting whether a debtor will fail or not default in the future is a challenging task for any firm or bank. Loan default prediction is fundamentally a binary classification problem.

In the present scenario, a loan needs to be approved manually by a representative of the bank which means that the bank employee will be responsible for predicting whether the person is eligible for the loan or not by calculating the risk associated with it. As it is done by a human it is a time-consuming process and is susceptible to errors. If the loan is not repaid, then it accounts as a loss to the bank and banks earn most of their profits from the interest paid to them. If the banks lose too much money, then it will result in a banking crisis. This banking crisis affects the economy of the country. So it is very important that the loan should be approved with the least amount of error in risk calculation while taking up the least time possible. So a loan prediction model is required that can predict quickly whether the loan can be passed or not with the least amount of risk possible.

# ABSTRACT

Loans area unit the core business of banks. The most profit comes directly from the loan’s interest. The loan corporations grant a loan with an intensive method of verification and validation. However, they still don’t have the assurance that the customer is in a position to repay the loan with no difficulties. Loan Prediction is extremely useful for the worker of banks and for the customer additionally. The aim of this Project is to supply fast, immediate method for the eligible candidates. It will offer special advantage to the bank. The Loan Prediction System will calculate the load of every option participating in the loan process.

# EXISTING WORK

In today’s world, obtaining loans from financial institutions has become a very common phenomenon. Every day many people apply for loans, for a variety of purposes. But not all applicants are reliable, and everyone cannot be approved. Every year, there are cases where people do not repay the bulk of the loan amount to the bank which results in huge financial losses. The risk associated with making a decision on a loan approval is immense. Hence, the idea of this project is to gather loan data and use machine learning techniques on this data to extract important information and predict if a customer would be able to repay the loan or not. In other words, the goal is to predict if the customer would be a defaulter or not. Considering the magnitude of risk and financial loss involved, it is essential for banks to give loans to credible applicants who are highly likely to pay back the loan amount.

# PROBLEM STATEMENT

By predicting the loan defaulters, the bank can reduce its Non Performing Assets. This makes the study of this phenomenon very important. Previous research in this era has shown that there are so many methods to study the problem of controlling loan default. But as the right predictions are very important for

the maximization of profits, it is essential to study the nature of the different methods and their comparison. We are trying to automate the loan eligibility process (real-time) based on customer details provided while filling out an online application form. These details are Gender, Marital Status, Education, number of Dependents, Income, Loan Amount, Credit History, and others. We have applied Logistics Regression and Random Forest to analyze and predict. Logistics regression and Random Forest gives us the probability of whether a customer should get a loan or not. Depending on the accuracy of these models we will select which one will best fit our data.

# PROPOSED WORK

Considering the financial institution’s situation whether or not they are willing to lend a loan to a client is a risky factor for them.

The following are the two most important banking issues:

1) What is the borrower's risk level?

2) Given the danger, should we credit to the borrower?

The lender's interest rate is determined by the answer to the first question. The interest rate, together with other factors (such as the payback period), assesses the borrower's riskiness; the higher the rate of interest, the riskier the consumer. depending upon the interest rate, we will determine if the applicant is eligible for the loan.

**HARDWARE AND SOFTWARE REQUIREMENTS**

# SOFTWARE REQUIREMENT:

* + - Operating system: Windows
    - Tools: VISUAL STUDIO

# HARDWARE REQUIREMENT:

* RAM: 8.00 GB (7.87 GB usable)
* Processor: Intel(R) Core (TM) i5-10300H CPU @ 2.50GHz 2.50 GHz
* System-type: 64-bit operating system, x64-based processor
* Version: 20H2
* Edition: Windows 10 Home Single Language

# DATASET

**DATASET DESCRIPTION:**

It consists of these details - Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and others.

**TYPE OF DATA:**

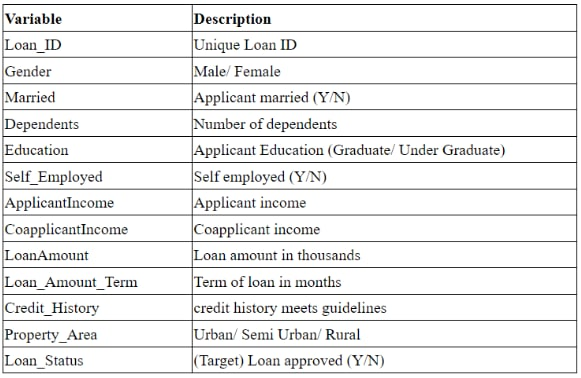
* **Categorical Columns:** Gender (Male/Female), Married (Yes/No), Number of dependents (Possible values:0,1,2,3+), Education (Graduate / Not Graduate), Self-Employed (No/Yes), credit history(Yes/No), Property Area (Rural/Semi-Urban/Urban) and Loan Status (Y/N)(i. e. Target variable)
* **Numerical Columns:** Loan ID, Applicant Income, Co-applicant Income, Loan Amount, and Loan amount term

<URL:-> [**https://www.kaggle.com/datasets/vipin20/loan-application-data**](https://www.kaggle.com/datasets/vipin20/loan-application-data)

**DATASET ROWS AND COLUMNS:**

Training data - We have 614 rows and 13 columns.

Test data - We have 367 rows and 12 columns because the target column is not included in the test data.



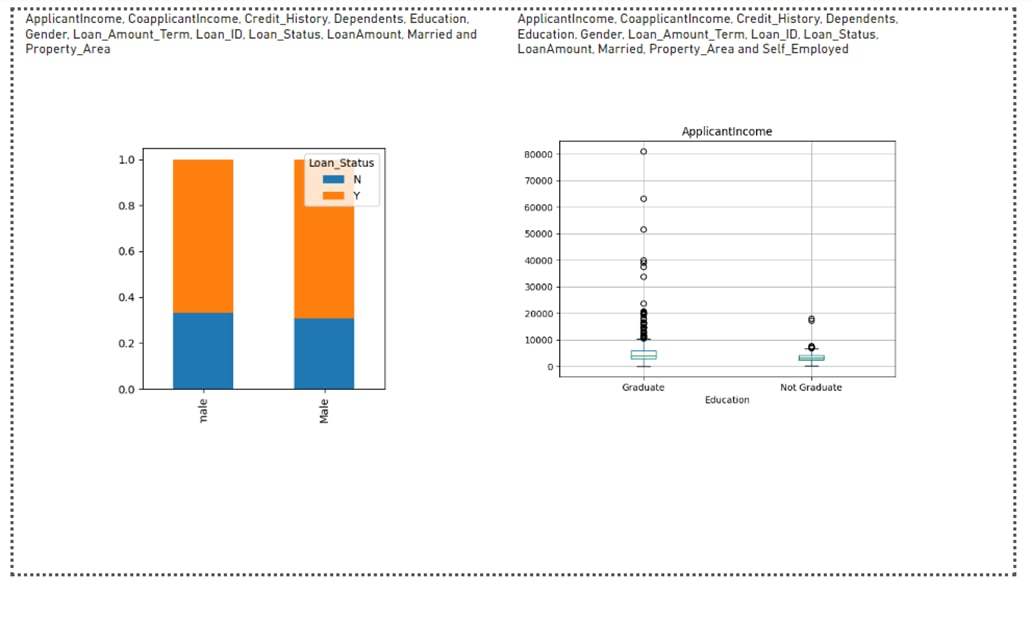
# Fig :- Loan Application Data

# METHODOLOGY

**Logistic regression :-** A very important approach in predictive analytics is used to study the problem of predicting loan defaulters: The Logistic regression model - data is collected from the Kaggle for studying and prediction. Logistic Regression models have been performed and the different measures of performances are computed. The models are compared on the basis of the performance measures such as sensitivity and specificity. The final results have shown that the model produce different results. Model is marginally better because it includes variables (personal attributes of customer like age, purpose, credit history, credit amount, credit duration, etc.) other than checking account information (which shows wealth of a customer) that should be taken into account to calculate the probability of default on loan correctly. Therefore, by using a logistic regression approach, the right customers to be targeted for granting loan can be easily detected by evaluating their likelihood of default on loan. The model concludes that a bank should not only target the rich customers for granting loan but it should assess the other attributes of a customer as well which play a very important part in credit granting decisions and predicting the loan defaulters.

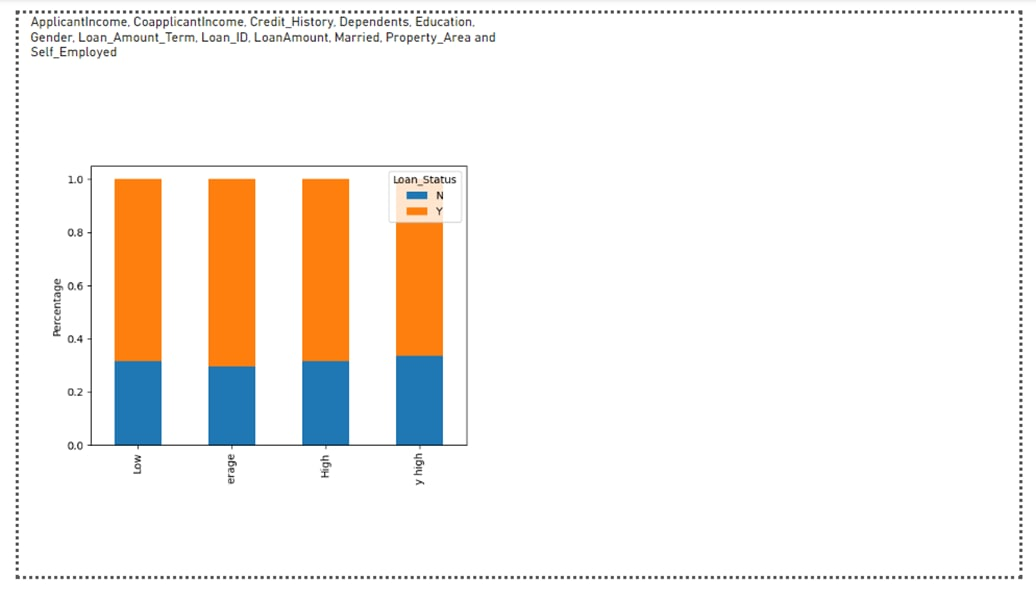
# RESULT

# PYTHON INTEGRATION:



**Figure 1. Study of distribution of data.**

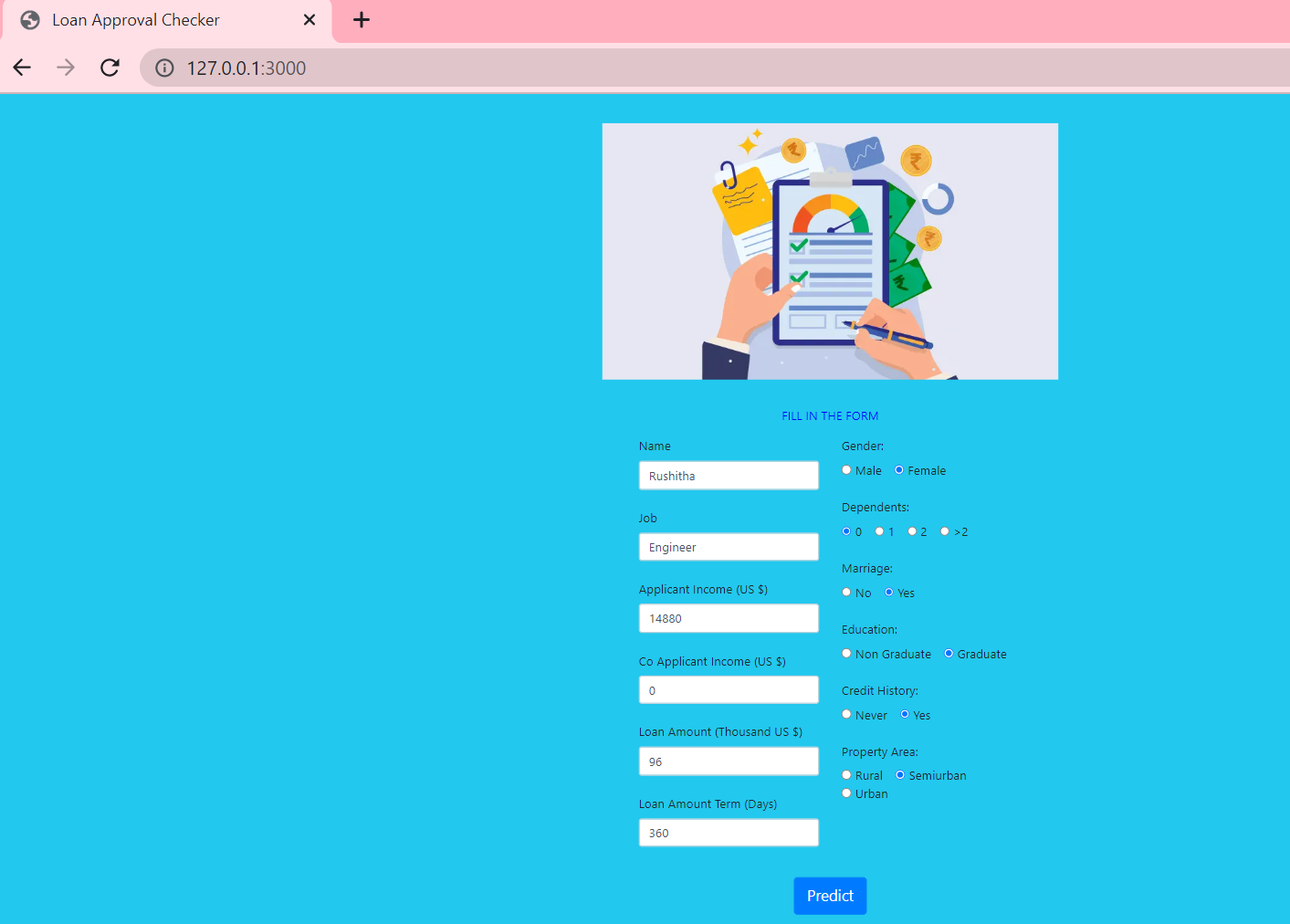
Box plot and histogram are used for study of distribution factors. In the snapshot below one such factor (applicant income) has been used as an example(Fig 1). There are many extreme values due to income gap and difference in education levels.

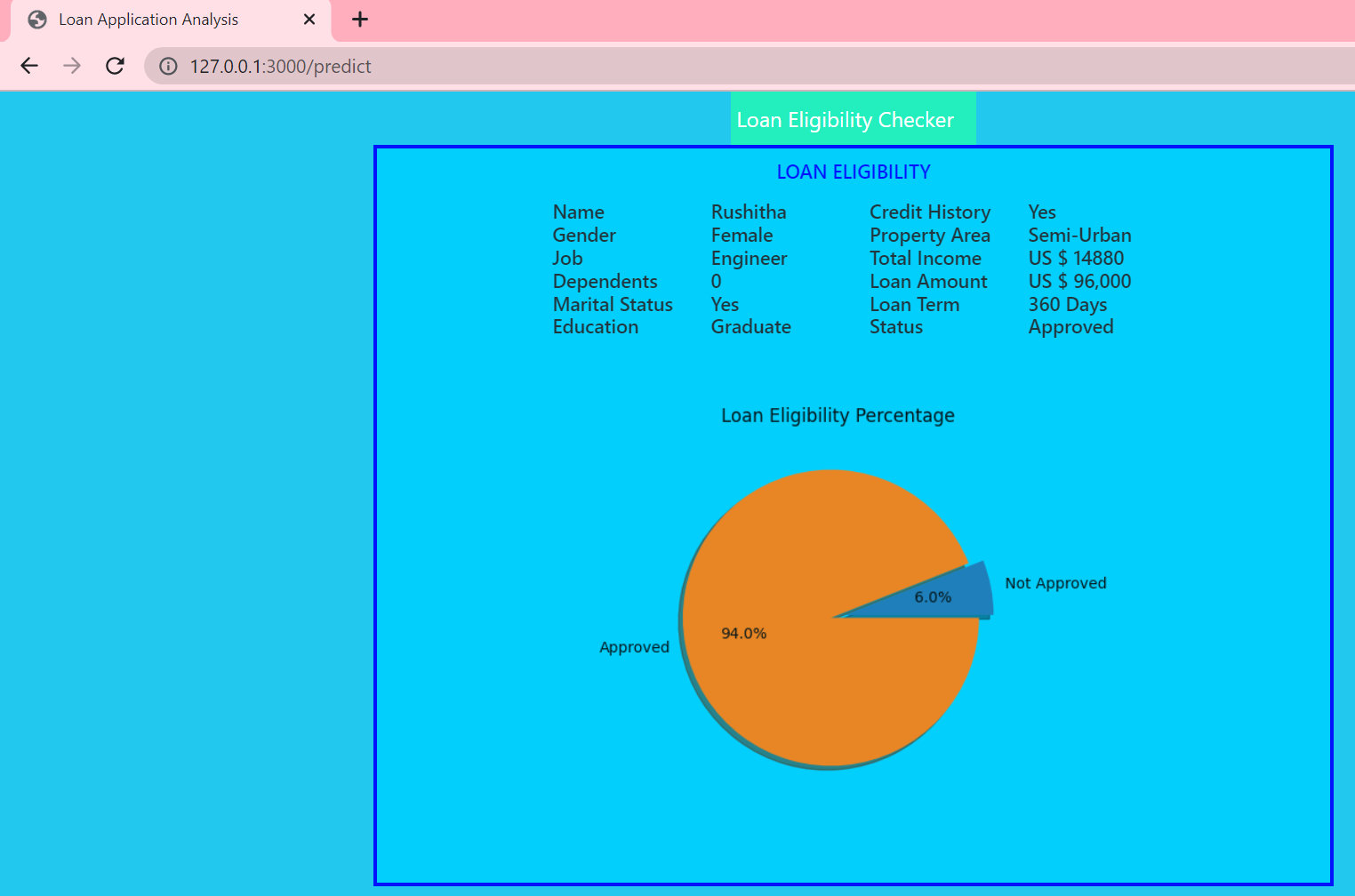


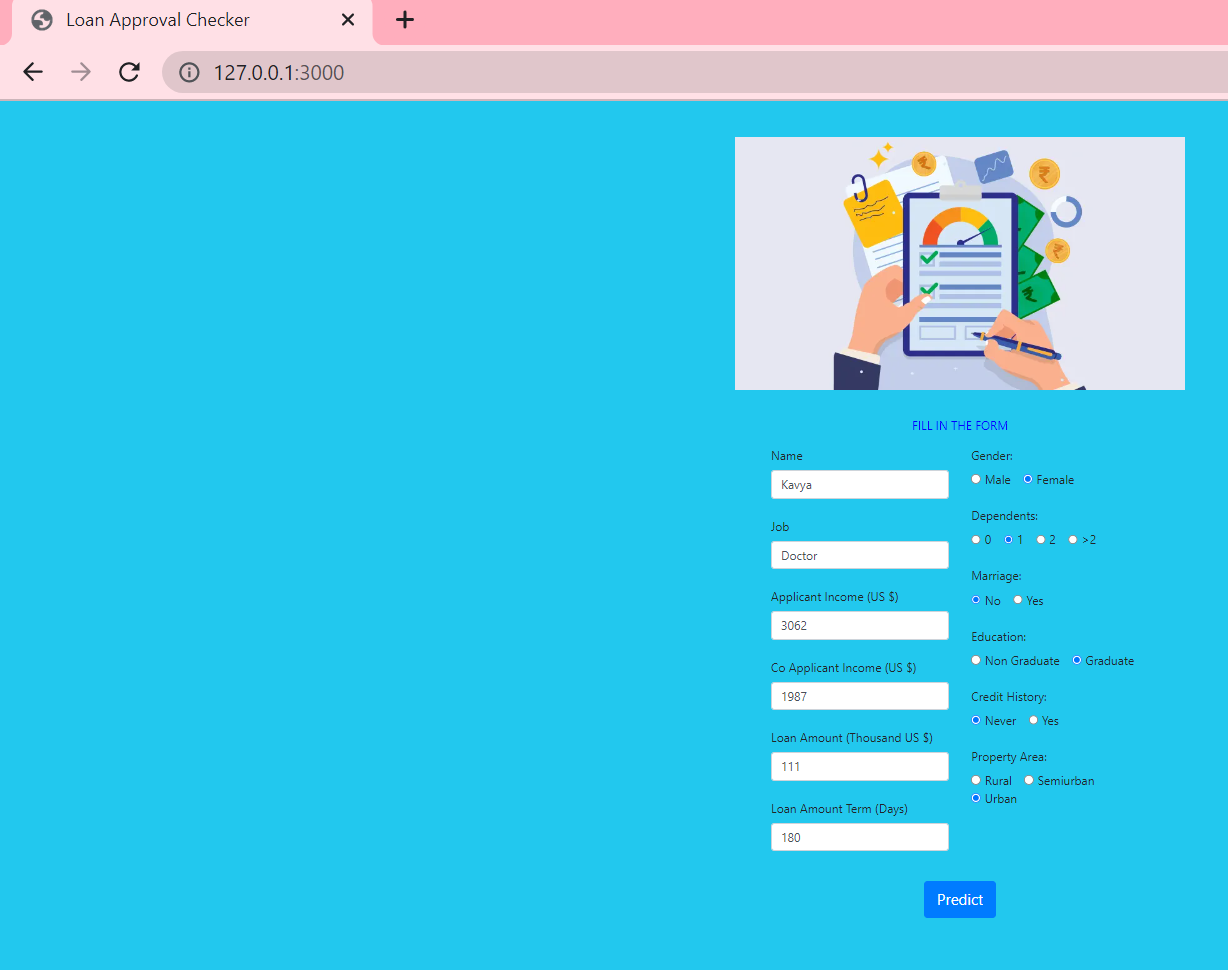
Plotting of graph to infer more results. Plotting relation between factor and result to recognize

patterns. The graph depicts that it’s eight times easier to get a loan if an applicant has a valid credit history.

**OUTPUT SCREENS:**









**CONCLUSION**

The main objective of the Project is to build a Loan Prediction Model that can shorten the loan approval time and decrease the risk associated with it. It is done by predicting if the loan can be given to that person on the basis of various parameters like credit score, income, age, marital status, gender, etc. The prediction model not only helps the applicant but also helps the bank by minimizing the risk. The loan application analysis system can assist the banks in making the best judgement on whether to approve or deny a customer's loan request.

# REFERENCES

# Sheikh, Mohammad Ahmad, Amit Kumar Goel, and Tapas Kumar. "An approach for prediction of loan approval using machine learning algorithm." In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 490-494. IEEE, 2020.

# Maheswari, P., and CH V. Narayana. "Predictions of Loan Defaulter-A Data Science Perspective." In 2020 5th International Conference on Computing, Communication and Security (ICCCS), pp. 1-4. IEEE, 2020.

# Patel, Bhoomi, Harshal Patil, Jovita Hembram, and Shree Jaswal. "Loan default forecasting using data mining." In *2020 International Conference for Emerging Technology (INCET)*, pp. 1-4. IEEE, 2020.

# Chang, Yung-Chia, Kuei-Hu Chang, and Yi-Xin Lin. "Establishment of Business Loan Default Prediction Model by Integrating Survival Analysis with Logistic Regression." Scientia Iranica (2022).

# Alaradi, Mohamed, and Sawsan Hilal. "Tree-Based Methods for Loan Approval." In 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), pp. 1-6. IEEE, 2020.