Kathmandu University

Department of Computer Science and Engineering

Dhulikhel, Kavre

A Project Report
on
"Loan Approval Prediction System (LAPS)"

[COMP 207]

(For partial fulfillment of 2nd Year/ 2nd Semester in Computer Engineering)

Submitted by

Aavash Adhikari	**(Roll No: 01)**
Nishant Ghimire	**(Roll No: 17)**
Ram Krishna Ghimire	**(Roll No: 20)**
Nishchal Raj Subedi	**(Roll No: 54)**

Submitted to:

Mr. Dhiraj Shrestha

(Department of Computer Science and Engineering)

Submission Date: 2024/01/01

Bona fide Certificate

This project work on

“Loan Approval Prediction System (LAPS)”

is the bona fide work of

“AAVASH ADHIKARI”

“NISHANT GHIMIRE”

“RAM KRISHNA GHIMIRE”

“NISHCHAL RAJ SUBEDI”

who carried out the project work under my supervision.

Project Supervisor

_____

Sudan Jha, PhD

Professor

Department of Computer Science and Engineering

Date: 2024/01/01

Acknowledgment

We are indebted to our project coordinator **Mr. Dhiraj Shrestha,** Department of Computer Science and Engineering of Kathmandu University for transforming our study and research capabilities by allowing us to do the project work as our coursework.

We are heartily thankful to our project mentor, **Mr. Sudan Jha** whose guidance and motivation were crucial for completing this work effectively within the given time. We are indebted to each of the individuals including our family members whose involvement in our lives has brought immense energy and inspiration for us to enhance our abilities, and friends **(Samman Pathak, Darshan Lamichhane, and Sushan Adhikari)**, whose astute observation with an eye for perfection and honest feedback on the program made the project good, enhanced and what it is now.

Abstract

Banks heavily depend on loans for revenue, but identifying reliable applicants remains challenging. Manual assessments often lead to errors in selecting the right candidates. We're building a machine learning-powered website for loan approval prediction to address this problem. This system benefits banks and applicants by automatically identifying qualified candidates.

Our project covers designing a loan approval prediction system. Using real-time data such as education, loan amount, credit history, and income, we have automated the process of checking loan eligibility. Since our goal is loan approval classification, we'll employ supervised machine learning algorithms like logistic regression, decision trees, and random forest. We have handled data preparation, cleaning, and feature enhancement. A key feature is creating a simple user interface (UI) where applicants can easily predict their loan approval chances.

To sum up, our project makes the loan applicant assessment process easier by using Machine learning tools and the web framework of Python i.e. Flask to predict outcomes through automated predictions, reducing errors and increasing efficiency.

Keywords: *Supervised Machine Learning,, Logistic Regression, Bank Loans, Safe Customers, Loan Approval Prediction.*

Table of Content

Abstract i
List of Figures iii
List of Tables iv
List of Abbreviations v
Chapter 1 Introduction 1
1.1 Background 1
1.2 Objectives 2
1.3 Motivation and Significance 3
Chapter 2 Related Works 4
2.1 PretHeure 4
2.2 Sapatisewa 6
Chapter 3 Design and Implementation 8
3.1 System Requirement Specifications 13
3.1.1 Hardware Specifications 13
3.1.2 Software Specifications 13
Chapter 4 Discussion on the Achievements 14
4.1 Challenges 14
4.2 Findings 16
4.3 Features 17
Chapter 5 Conclusion and Recommendation 18
5.1 Limitations 18
5.2 Future Work and Enhancement 19
Chapter 6 Project Planning and Scheduling 21
6.1 Gantt Chart 22
References 23
Appendix 24

List of Figures

Figure 2.1. 1 PretHeure's System View 5
Figure 2.1. 2 PretHeure's System View 6

Figure 2.2. 1 Sapatisewa's System View 6
Figure 2.2. 2 Sapatisewa's System View 7

Figure 3. 1 Flowchart of the project 'Loan Approval Prediction System' 10
Figure 3. 2 Use Case Diagram 11
Figure 3. 3 Prototype Design1 12
Figure 3. 4 Prototype Design2 12

Figure Appendix i Front Page 24
Figure Appendix ii Heat Map 24
Figure Appendix iii Box Plot 25
Figure Appendix iv Scatter Plot 25

List of Tables

Table 1 Gantt Chart For System Development 22

List of Abbreviations

ADA	AdaBoosted Decision Trees
DBMS	Database Management System
EDA	Exploratory Data Analysis
IDE	Integrated Development Environment
ML	Machine Learning
UI	User Interface
VS Code	Visual Studio Code

Chapter 1 Introduction

The system “Loan Approval Prediction System” is a program that predict whether provide loan or not to the applicants or customers. This program works on machine learning algorithm where the customers financial, transactional and many activities are taken into consideration to make an optimal prediction for loan approval. The program is aimed to help financial institutions from bad loan approval and serve the common populace to get the appropriate loan amount.

1.1 Background

In the contemporary world of financial transactions and credit demands, the evaluation and approval of loan applications have become pivotal components of financial institutions operations. However, traditional methods of assessing loans often involve manual processes that result in prolonged processing times, inconsistent decisions, and even potential biases. To address these challenges and work in greater efficiency and accuracy to the loan approval process, our project proposes the development and implementation of an advanced Loan Approval Prediction System.

Our attempt aligns with the growing significance of Machine Learning across diverse industries, ranging from the Smartphone Industry to Healthcare and Banking. In our project, we utilize the power of Machine Learning in collaboration with data analysis techniques to contribute to this transformative trend. With a specific focus on financial institutions like banks, where precise and prompt decision-making is paramount, we aim to revamp the loan approval process. By integrating machine learning algorithms and data analysis, our project aims to harness insights from existing customer data. Through in-depth analysis, we aim to uncover concealed patterns, insights, and relationships that conventional methods might overlook. This data-driven approach centers on predicting the probability of loan approval for potential applicants, streamlining the decision-making process, and ensuring rapid responses for applicants while providing institutions a structured and unbiased basis for making loan approval decisions.

Our project emphasizes the modernization of the loan approval process by using the predictive capabilities of machine learning algorithms and data analysis techniques. Through the predictive insights obtained from historical loan data, we intend to empower financial institutions with an automated, efficient, and objective tool to evaluate the feasibility of granting loans to future applicants. In essence, our project not only accelerates decision-making but also guide in a more precise and equitable approach to loan approvals, aligning with the evolving demands of today's technology-driven era and guiding in a new era of lending practices.

1.2 Objectives

It has a wide range of goals, some of which are mentioned below in order of importance:

1. To improve the accuracy and consistency of loan approval decisions by using machine learning algorithms.
2. To help improve the financial health of the banking sector.
3. To evaluate loan applications based on personal and financial information of the borrower.
4. To increase conversion rates by providing faster response times, easier pre-approvals, and less friction for the applicant.

1.3 Motivation and Significance

The inspiration to create a Loan Approval Prediction System arises from the limitations of traditional loan assessment methods. These conventional approaches often result in slow decision-making and inconsistencies, causing delays for applicants and unequal treatment. Additionally, the difficulty of evaluating risk can complicate matters further. To address these concerns, we're turning to Machine Learning and data analysis for a solution. By automating processes, we aim to enhance efficiency, fairness, and accuracy in loan approvals.

Our Loan Approval Prediction System carries substantial potential. It seeks to benefit both applicants and financial institutions. Swift responses can aid those in need of loans urgently, while streamlining operations for banks. Adopting a data-driven approach holds the promise of reducing biases and promoting inclusivity. Moreover, accurate risk assessment can lead to improved portfolio performance. Importantly, our project demonstrates the commitment of financial institutions to modernization and equitable lending practices in an increasingly technology-oriented landscape.

Chapter 2 Related Works

The idea of a loan approval prediction system is not new in the current world. There are some similar projects or apps based on a loan approval prediction system to find eligible candidates for loan customers.

Existing System

The world is transforming into a digital world and people prefer online transactions rather than physical transactions. And in the existing system, everyone has used random forest and decision trees machine learning algorithms. Although we can get good efficiency using decision trees and random forests to some extent, we can get great results using logistic regression. The lender must manually review each application, supporting the main points provided by the applicant like gender, legal status, education, number of dependents, income, and loan amount credit.

2.1 PretHeure

PretHeure is a Canadian loan approval prediction system in which people fill their genuine details and get approved for loans. There is an age restriction for loan customers and loan customers also have to verify their steady employment and payroll verification.

Disadvantages

1. **Inaccuracy**

   We are not aware of the accuracy of this system. Lack of transparency: This lack of transparency can make it challenging to explain loan decisions to borrowers and ensure fairness.

2. **Lack of transparency**

   This lack of transparency can make it challenging to explain loan decisions to borrowers and ensure fairness.

Features

1. **Faster processing**

   PretHeure, Loan approval prediction systems can automate much of this process, allowing lenders to make decisions more quickly and efficiently. They provide loan within an hour.

2. **Data-driven insights**

   This system uses machine learning algorithms to analyze large datasets of past loan applications and identify patterns that can predict the likelihood of future loan defaults.

3. **Data security**

   Customer's data are fully secured in this system. They assure customer full security of their data.

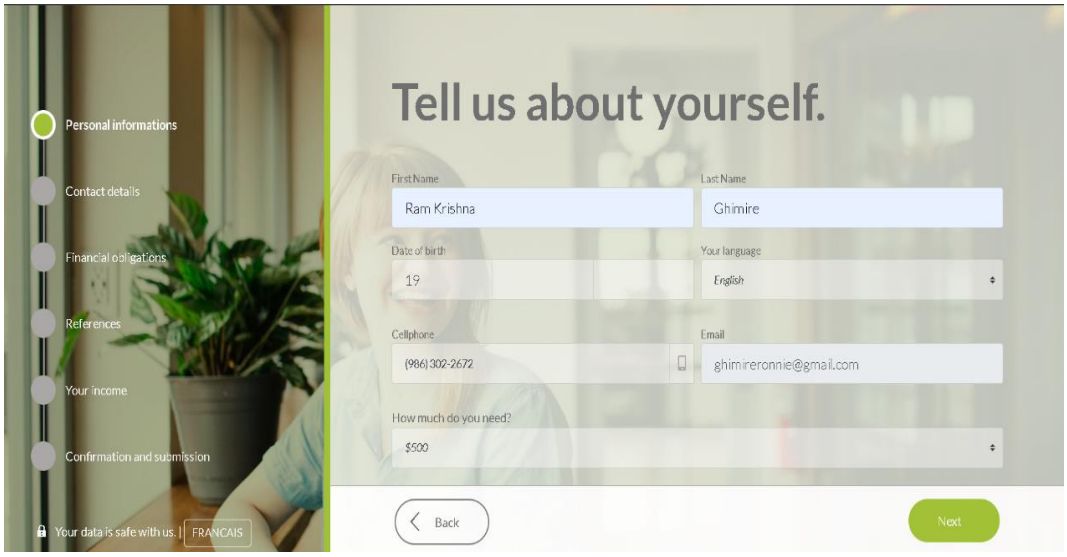

Figure 2.1. 1 PretHeure's System View

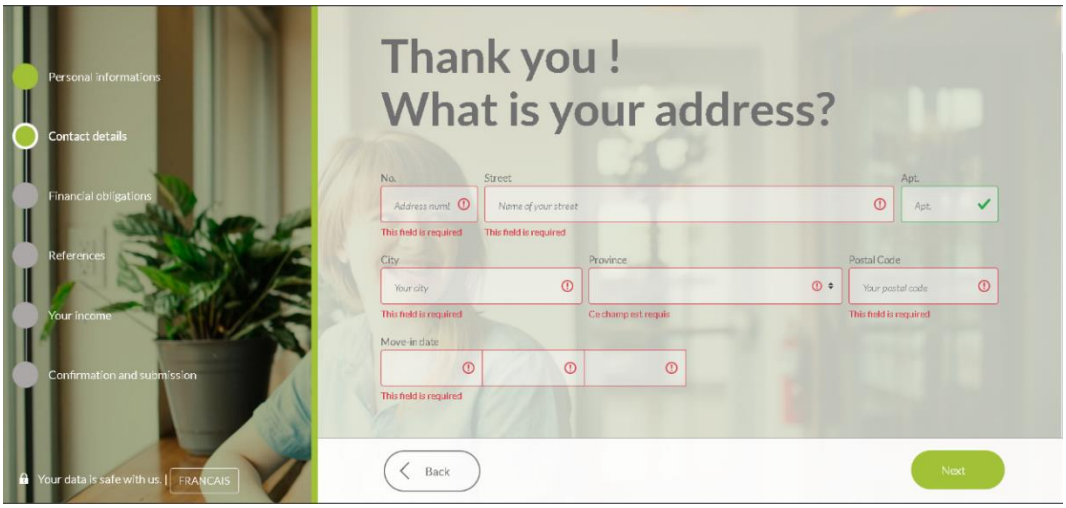

Figure 2.1. 2 PretHeure's System View

2.2 Sapatisewa

Sapatisewa is Nepal's first loan marketplace. It provides loans for business, personnel, auto, education loan, etc. There is placed an eligibility calculator on which customer can provide monthly income, their living expenses, no. of years, and interest and it will give the total amount up to which they are eligible for taking a loan.

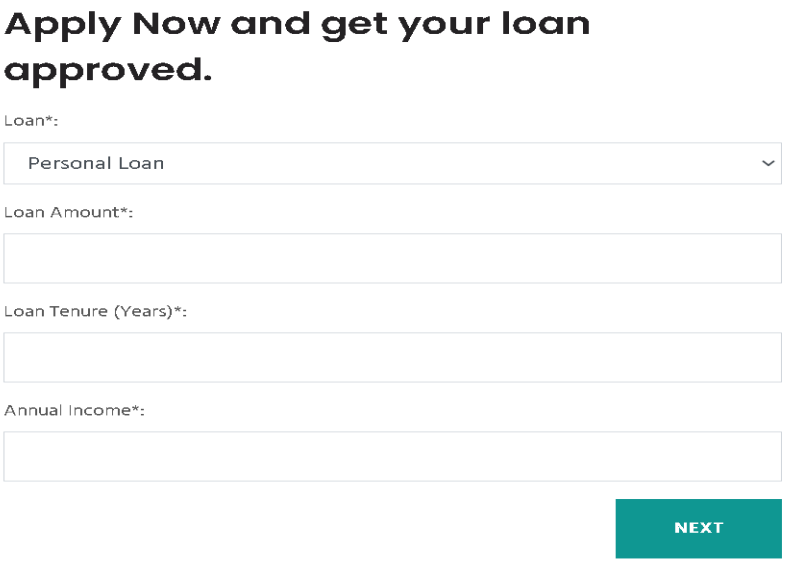

Figure 2.2. 1 Sapatisewa's System View

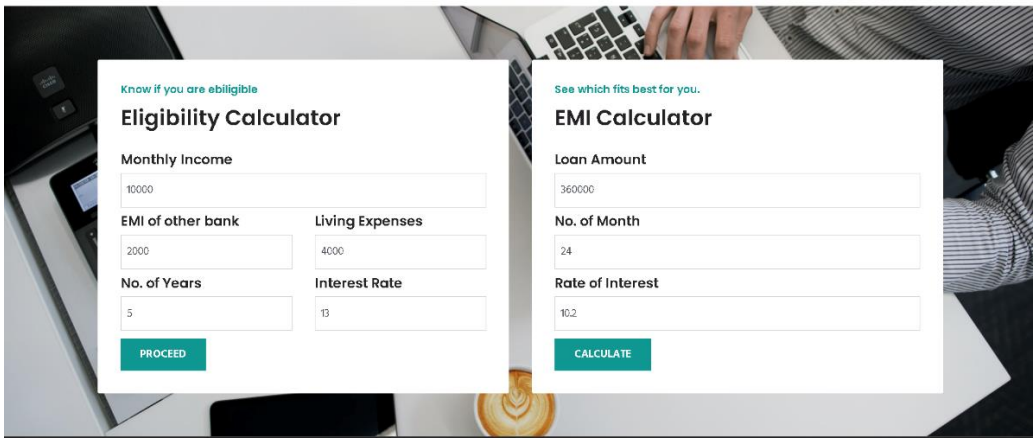

Figure 2.2. 2 Sapatisewa's System View

Chapter 3 Design and Implementation

Firstly, we searched for the relevant title for our project. Then, we skimmed through the topic so as to plan for our project work. After group discussion and analysis of different ideas, we came to the conclusion of selecting the above-discussed project (Loan Approval Prediction System) as our semester project. Since it has wide application in the real world, which we realized after going through its pros and cons, it can affect a myriad of people if it turns out to be a success. After finalizing the project, the proposal regarding all the information about the project is prepared. The programming language needed to convert the idea into a working program (in our case, it is Python and its various libraries like Numpy, Pandas, Scikit learn, MatplotLib, Seaborn) was learned and implemented into our project after the proposal submission.

After having made all the arrangements for turning the concept and idea into substantial existence through our intended program, the coding will commence with the principle of 'divide and conquer', which is mandatory in any group work, by an equal division of portions of the program to all the group members. We will develop a program that makes the best use of the concept of machine learning with the UI design and implementation.

After the completion of the program, the program would be tested by assimilating the chunks of code made by all the members of the group. If the program becomes expansive, it would be made economical and concise. In the unlikely case that the program fails to deliver the required expectations or there is the occurrence of any bugs, the issues was tackled through group discussion and analysis. After the program becomes bug-free and is made to stand out in its name, the final report involving the analysis of the project, program, references, its output, etc. is made and presented to the evaluators.

The algorithm of the code is as:

1. **Data Collection**

   Gather historical loan application data from the financial institution, including approved and rejected loan applications. The dataset should encompass relevant applicant information, credit history, financial attributes, and loan outcomes.

2. **Data Preprocessing**

   Cleanse and prepare the dataset for analysis. Handle missing values, outliers, and inconsistencies. Encode categorical variables and perform feature scaling to ensure data uniformity.

3. **Feature Selection and Engineering**

   Identify important features that significantly impact loan approval decisions. Utilize domain knowledge and statistical techniques to engineer new features if necessary.

4. **Exploratory Data Analysis (EDA)**

   Perform exploratory analysis to uncover patterns, trends, and relationships within the data. Visualize distributions and correlations to gain insights into factors influencing loan approvals.

5. **Model Selection and Training**

   Carry on with the model-building process after adding additional features (Kumari, *Loan prediction problem from scratch to end* 2022). So, we will start with a logistic regression model and then move over to more complex models like Random Forest. We will build the following models. Logistic Regression Decision Tree Random Forest.

The flow of the program can be explained by the flowchart mentioned below:

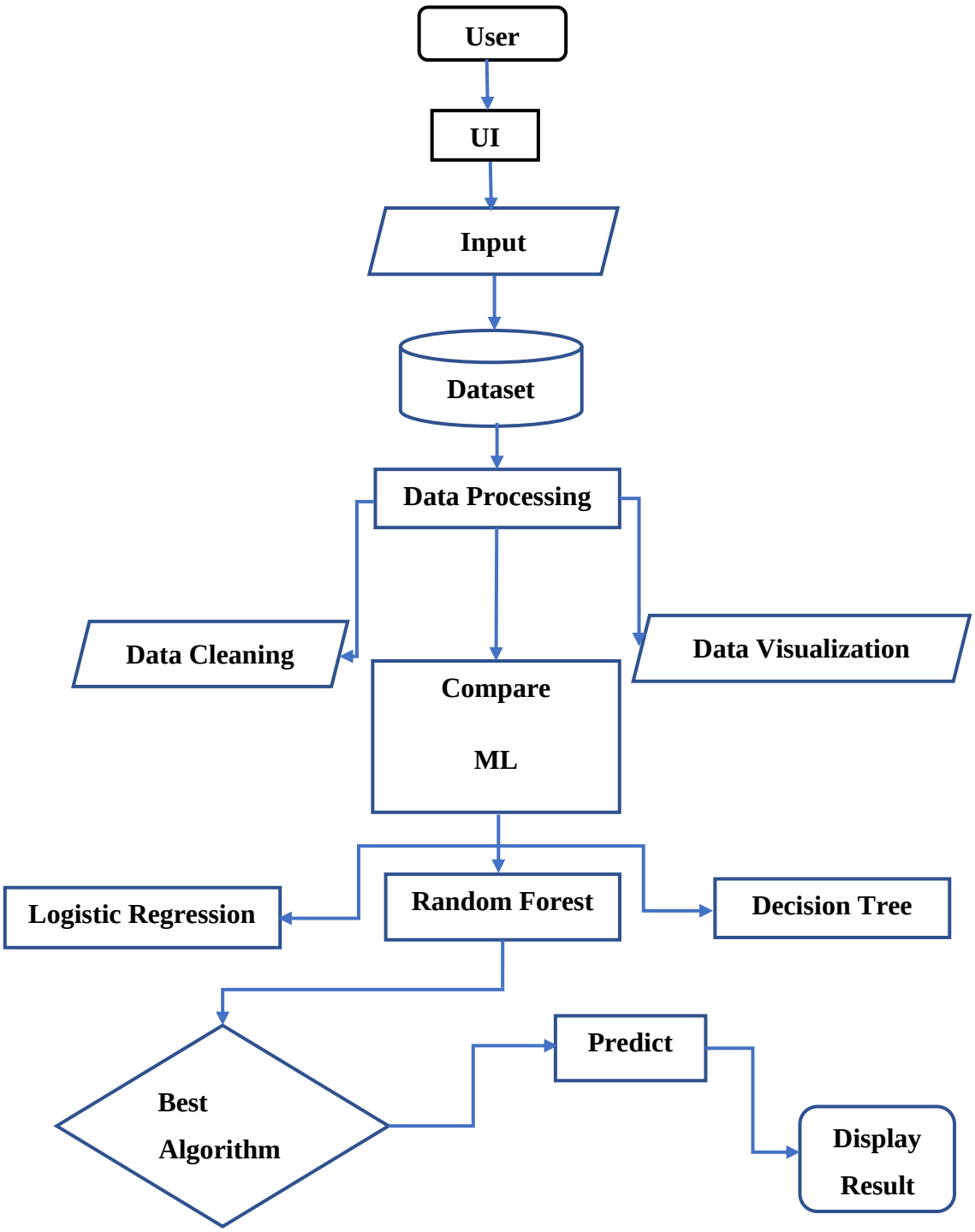

Figure 3. 1 Flowchart of the project 'Loan Approval Prediction System'

Use-Case Diagram

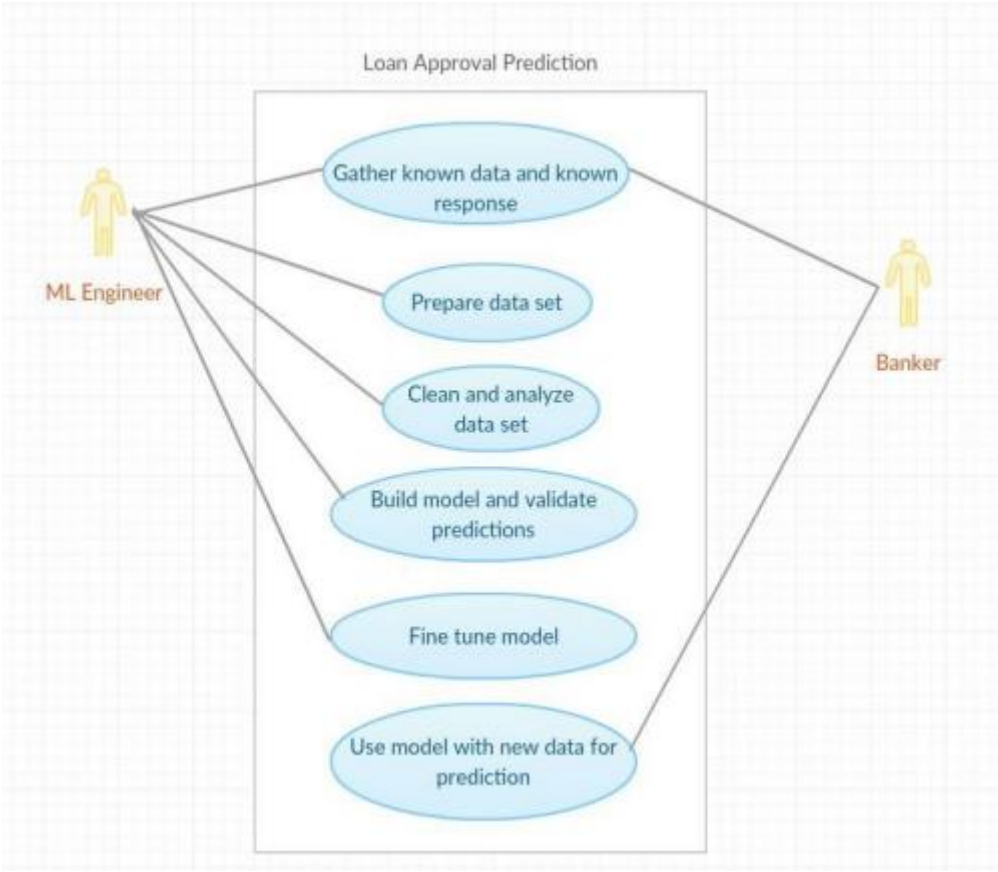

Figure 3. 2 Use Case Diagram

Prototype Design

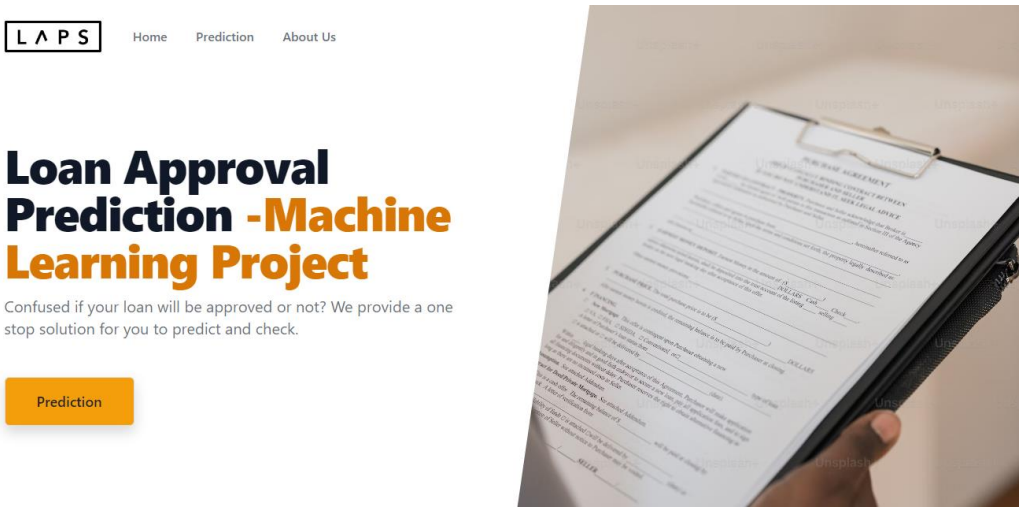

Figure 3. 3 Prototype Design1

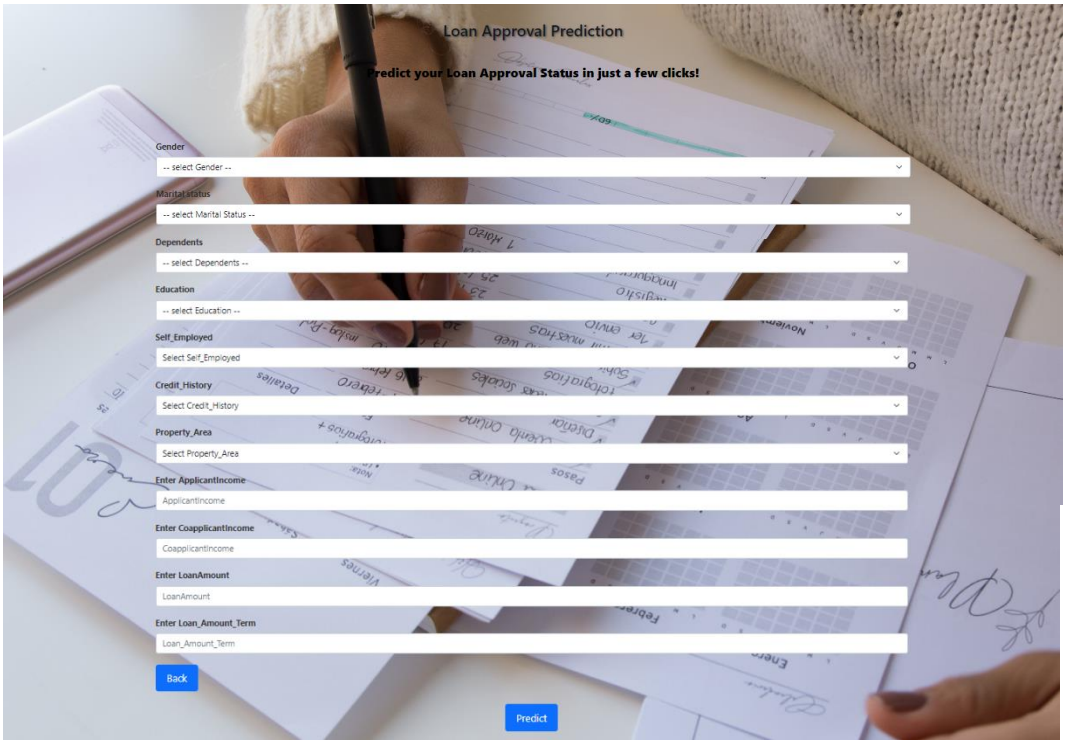

Figure 3. 4 Prototype Design2

3.1 System Requirement Specifications

Since our program is designed to run on a computer, the computer must have its specifications as per the required hardware and software threshold for the smooth execution of the program.

3.1.1 Hardware Specifications

As our system is built on a computer having processor Intel core i5 RAM 8 GB so we recommend a hardware with these specifications or better than these.

3.1.2 Software Specifications

This system is designed in computer system with 64-bit operating system, x64-based processor. So, we recommend these specifications or better than our computer system for running this website possessing loan approval prediction system.

Front End Tools

HTML, CSS.

Integrated Development Environment: VS Code, Jupyter Notebook.

Back End Tools

Jupyter Notebook, Python and its libraries like NumPy, SciKit-Learn, Pandas, MatplotLib, Seaborn, MS Excel along with its web framework i.e. Flask.

Chapter 4 Discussion on the Achievements

After we have decided our project, we have set our weekly goal and divide task and continuously work for that and achieve according to our plan. In the process we were also getting confused somewhere but with the help of resources we had collected at initial stage we had solved such errors. For completion of projects we had gone through lots of ups and down. We have studied documentations of every topics we needed. We have also watched tutorials where some of them were useful but some of them were not.

Our loan approval prediction project seamlessly integrates a user-friendly interface with advanced machine learning models, including Logistic Regression, Decision Tree and Random Forest. This achievement breaks down barriers, making financial decision-making accessible. The system not only provides real-time loan approval predictions but also contributes to financial literacy by considering crucial factors like education, income, and credit history. In essence, our work signifies a move toward a transparent and technologically empowered future, enabling individuals to make informed financial choices with confidence.

All of these features are the achievement of our team members who have given their best to incorporate all the features into this website.

4.1 Challenges

1. **Availability of Quality data**

   The era of digital wallets started around 2009 in Nepal since then there has been a significant spike in the digital sector of the country(*The trend of digital wallet in Nepal* 2023). The Covid-19 pandemic and nationwide lockdown propelled the trend forcing users to adopt digital mode of receiving and making payments. So, since we only have limited datasets, we cannot ensure accurate predictions.

2. **Data privacy and security**

   Applicants have to enter their personal details like name, date of birth, contact number, address, financial details like income details, bank account information, employment history, loan history, credit history etc. So, managing and protecting the sensitive data of users is a crucial and big challenge for us.

3. **Scalability**

   We have to face challenges regarding server capacity and performance of the model as the volume of customers increases.

4. **Credits scoring challenges**

   Many individuals in Nepal may not have a well-established credit history. This can make it difficult to assess their creditworthiness based on traditional credit scoring methods, which rely heavily on past borrowing and repayment behavior.

5. **Customer education**

   As there is a lack of tech-related knowledge among us we have to provide education and awareness about how the online loan approval system works.

6. **Varieties of loan**

   There are many different types of loan like education loans, business loans, personal loans, agricultural loans, vehicle loans, housing loans and for every type of loan approval, flexible loan approval prediction system build-up is challenging.

7. **Cannot work with live server**

   We have only work on local server. We have not tested our system in live server. So, we do not know anything about the performance of our system in live server.

4.2 Findings

The findings from our loan approval prediction project unveil insightful patterns and trends within the dataset, shedding light on the factors influencing loan outcomes. Here are some key findings:

1. The bar charts for categorical variables show that there are more male applicants than female, most applicants are married, a majority have graduated, a smaller portion is self-employed, and applicants are fairly evenly distributed across different property areas.
2. The boxplots for ApplicantIncome and CoapplicantIncome reveal a significant number of outliers, indicating that there are applicants with incomes much higher than the average.
3. The scatter plot for ApplicantIncome versus LoanAmount does not show a clear linear relationship, suggesting that higher applicant income does not necessarily correlate with higher loan amounts.
4. The correlation heatmap shows that there is a moderate positive correlation between ApplicantIncome and LoanAmount, and a strong positive correlation between Credit_History and Loan_Status, indicating that applicants with a good credit history are more likely to have their loans approved.

These findings provide a comprehensive overview of the dataset's characteristics and have guided for further analysis or predictive modeling.

4.3 Features

1. **User-Friendly Loan Prediction Form**

   We designed an intuitive and user-friendly form where users can input their information, such as education, income, credit history, etc., for the loan prediction.

2. **Predictive Model Integration**

   Our website seamlessly integrates the predictive loan approval model into the backend to provide users with instant predictions based on the information they provide.

3. **Clear Prediction Outcome**

   Our website communicates the loan prediction outcome clearly to users, indicating whether their loan is likely to be approved or not.

4. **Visual Feedback**

   The website incorporates visual elements such as color-coded results or icons to enhance the user experience and make the prediction outcomes visually appealing and easy to interpret.

Chapter 5 Conclusion and Recommendation

Through the process of learning and implementing, we, the LAPS team, were able to successfully complete the intended project right on time.

In conclusion, our loan prediction project is a big move in using tech for smarter financial decisions. With a sleek website and a predictive tool, users can easily foresee their loan approval chances. We kept things simple, clear, and stylish for a better user experience. The model's findings give us key insights into what factors really matter for loan outcomes, benefiting both users and financial institutions. It's a user-friendly solution making financial decisions more straightforward.

5.1 Limitations

As this is the first machine learning related project that we have undertaken, it cannot be argued to be a flawless one. It might contain flaws. It has not been tested with a survey among mass of people, based on our own observation and friends'feedback, the shortcomings of the project are as follows:

1. **Limited Data Quantity**

   The project grapples with the challenge of having a small dataset, which can hinder the model's ability to discern intricate patterns and relationships within the data.

2. **Data Imbalance**

   Imbalances in the dataset, such as a disproportionate number of approved or rejected loans, pose challenges in training the model to make accurate predictions across varied scenarios.

3. **Model Generalization**

   The modest accuracy of 77.2 percent indicates potential difficulties in the model's generalization capabilities, particularly when faced with unseen data or diverse applicant profiles.

4. **Reduced Predictive Power**

   The restricted dataset size may limit the predictive power of the model, affecting its ability to provide reliable insights into loan approval outcomes.

5.2 Future Work and Enhancement

There's room for improvement in this project, and some areas that could be improved in the future include:

1. **Expanded Dataset**

   Acquire a more extensive and diverse dataset to improve the model's learning capacity and enhance its predictive accuracy. A larger dataset could capture a wider range of loan approval scenarios and improve generalization.

2. **Dynamic Model Updates**

   Implement a mechanism for dynamic model updates to adapt to changing economic conditions and user behavior over time. Regularly updating the model ensures its relevance and effectiveness in different financial landscapes.

3. **Feature Enrichment**

   Explore additional features that could contribute to a more nuanced understanding of loan approval dynamics. Incorporating a richer set of variables might provide a more comprehensive view of applicant profiles.

4. **Explanatory Tools**

   Develop tools or features that provide users with explanations for the model's predictions. Enhancing interpretability can build user trust and facilitate a better understanding of the factors influencing loan outcomes.

5. **Integration with Financial Institutions**

   Explore collaborations with financial institutions to integrate the predictive model into their services. This can expand the tool's impact and provide users with additional resources and options for financial planning.

Chapter 6 Project Planning and Scheduling

At first, a round table discussion was done to work in our proposed project which encapsulates our knowledge of solving a real-world contemporary problem. After that, an analysis of the program was done along with the project plan. The programming needed for transforming the thought into a working program was learned in about three weeks. After having all the tools and knowledge ready for project preparation, the program development was begun, and "core programming" was commenced for about six weeks. A week after that was spent on testing and debugging the program and the following final week was spent on the preparation of the final report and presentations which has concluded the project in a total of twelve weeks.

The Gantt chart in the below table represents the time allocation of the project.

6.1 Gantt Chart

Table 1 Gantt Chart For System Development

Task Summary	WEEKS												Duration
	1	2	3	4	5	6	7	8	9	10	11	12	(Weeks)
Problem Identification													1
Requirements Gathering													1
Prototype Design													3
Learning and Core Programming													6
System Testing													2
Documentation													12

References

Banoula, M. (2023, August 21). *Machine learning tutorial: A [step-by-step] guide for*
https://www.simplilearn.com/tutorials/machine-learning-tutorial

Compare, choose and apply now !!! - sapatisewa. (n.d.-c).
https://sapatisewa.com/apply-now

FreeCodeCamp.org. freecodecamp.org. (n.d.).
https://www.freecodecamp.org

Get the loan you need within the Hour: Prêtheure. Get the loan you need within the hour | PrêtHeure. (n.d.).
https://www.pretheure.com

Your machine learning and Data Science Community. Kaggle. (n.d.).
https://www.kaggle.com

Kumari, K. (2022, June 9). *Loan prediction problem from scratch to end*. Analytics Vidhya.
https://www.analyticsvidhya.com/blog/2022/05/loan-prediction-problem-from-scratch-to-end/

The trend of digital wallet in Nepal. IME Remit. (2023, November 5).
https://imeremit.com.np/blogs/the-trend-of-digital-wallet-in-nepal/#:~:text=The%20era%20of%20digital%20wallets,of%20receiving%20and%20making%20payments.

Appendix

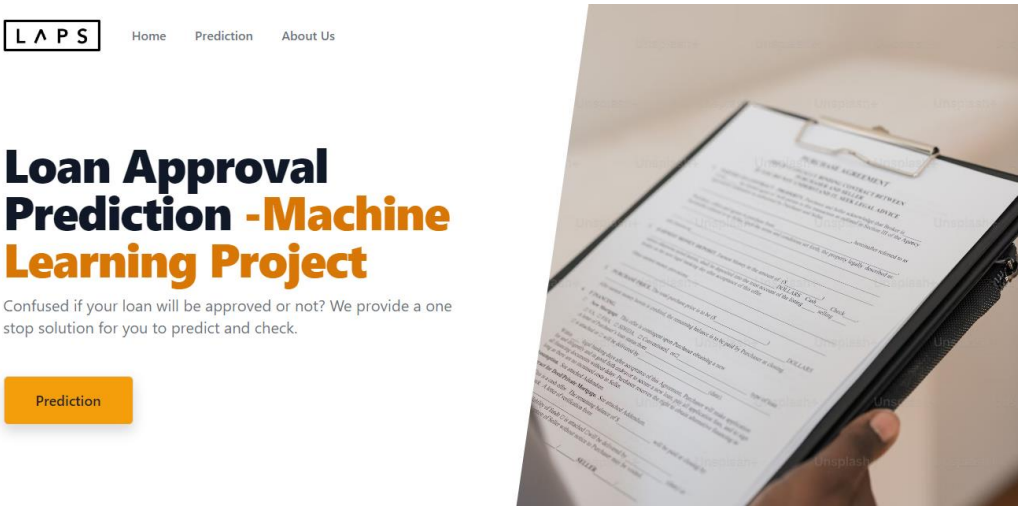

Figure Appendix i Front Page

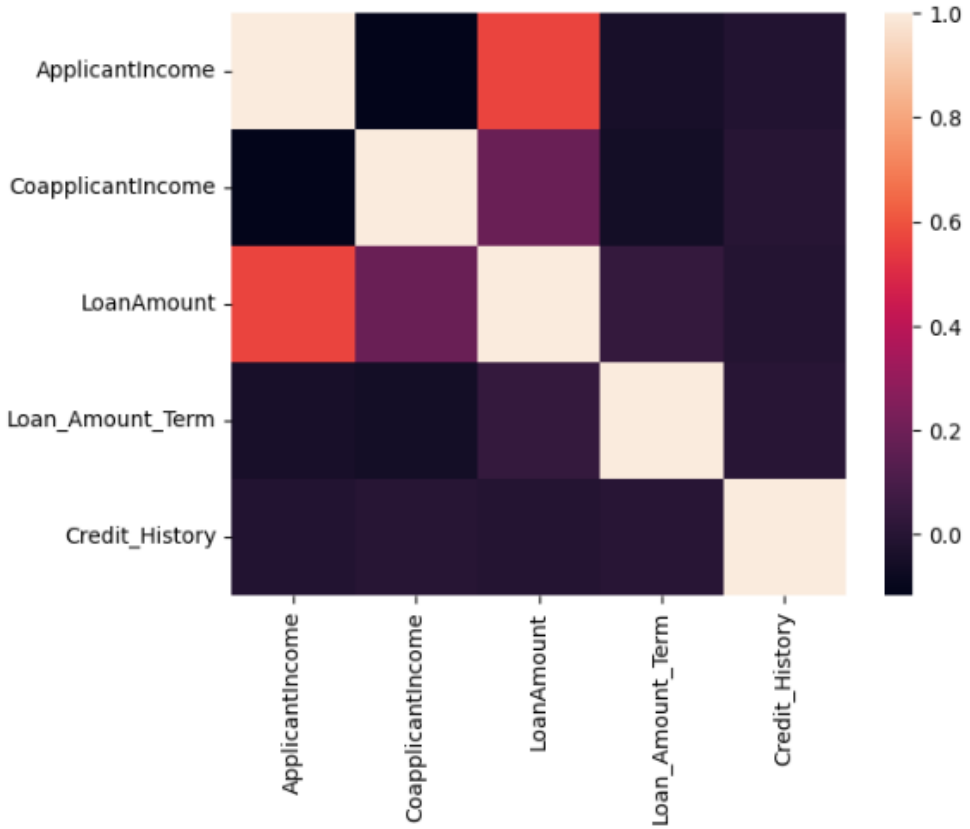

Figure Appendix ii Heat Map

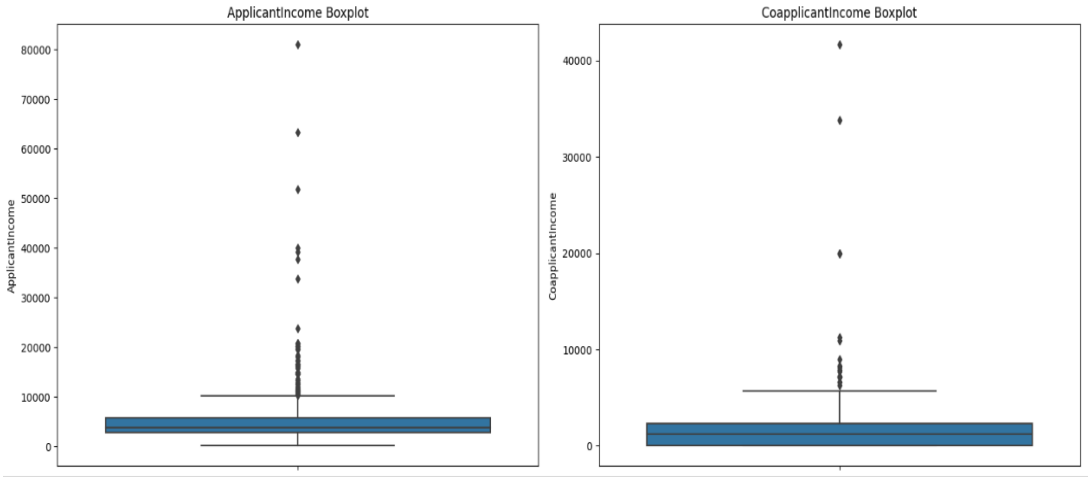

Figure Appendix iii Box Plot

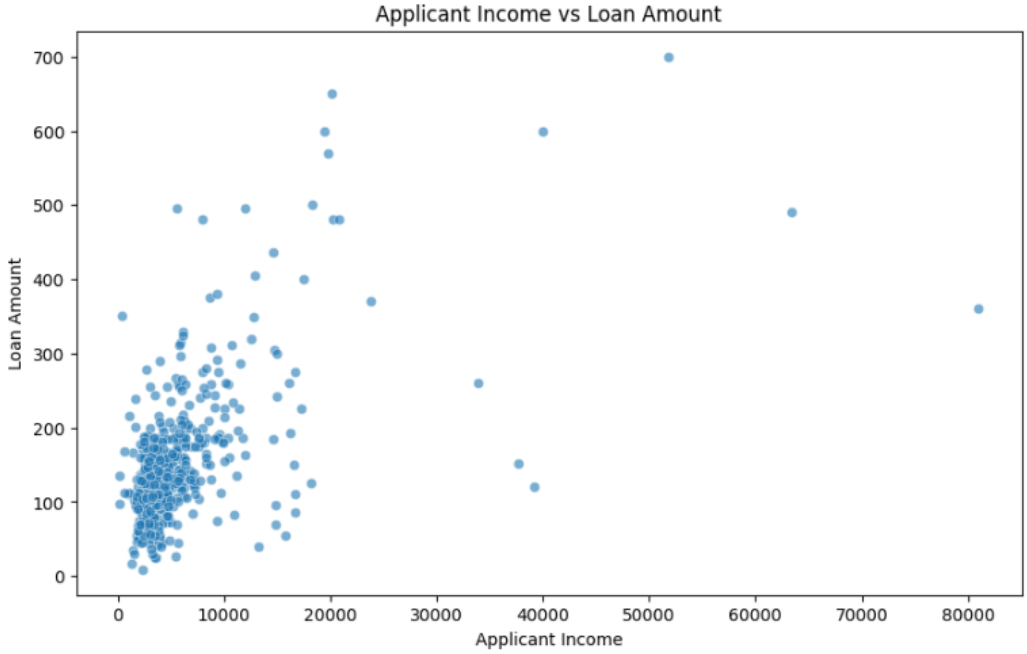

Figure Appendix iv Scatter Plot