PREDICTING LOAN ELIGIBILITY USING MACHINE LEARNING

A Mini Project Report Submitted

in partial fulfilment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ELECTRONICS AND COMMUNICATION ENGINEERING

BY

P. SIVA SAI KRISHNA (22335A0414)
T. VARSHITHA (22331A04H0)
P. VASANTHA (22331A04D8)
K. JAYA LAKSHMI (22331A04J2)

UNDER THE GUIDANCE OF
Dr. M. LAKSHMI PRASANNA RANI
M.Tech, Ph.D.
Associate Professor.

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERINGMAHARAJ VIJAYARAM GAJAPATHI RAJ COLLEGE OF ENGINEERING(A)
Approved by AICTE, New Delhi, Re-accredited by NAAC of UGC and PermanentlyAffiliated to JNTUV, Vizianagaram, Listed u/s 2(f) and 12(B) of the UGC Act 1956.
Vijayaram Nagar campus , Vizianagaram, Andhra Pradesh -535005
2022-2026

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
MAHARAJ VIJAYARAM GAJAPATHI RAJ COLLEGE OF ENGINEERING (A)

CERTIFICATE

This is to certify that the project entitled **"PREDICTING THE LOAN ELIGIBILITY USING MACHINE LEARNING"** being submitted by **P. SIVA SAI KRISHNA (22335A0414),T. VARSHITHA (22331A04H0), P. VASANTHA (22331A04D8), K. JAYA LAKSHMI (22331A04J2) i**n fulfillment of the requirements for the award of the Degree of **Bachelor of Technology in Electronics and Communication Engineering** is a record of bonafide work done by them under my supervision during the academic year **2024-2025.**

PROJECT GUIDE:
Dr. M. LAKSHMI PRASANNA RANI
M.Tech, Ph.D.
Associate Professor.
Department of E.C.E
M.V.G.R College of Engineering (A)
Vizianagaram.

HEAD OF THE DEPARTMENT:
Dr. SHAIK MASTAN VALI,
M.E, Ph.D.
Professor, Head of the department.
Department of E.C.E
M.V.G.R College of Engineering(A)
Vizianagaram.

SELF DECLARATION

We hereby declare that. The project entitled **— "Predicting Loan Eligibility Using Machine Learning"** is an outcome of our own efforts under the guidance of **Dr. M. Lakshmi Prasanna Rani**. The Project is submitted to the Maharaj Vijayaram Gajapathi Raj College of Engineering (A). For the partial fulfilment of the Bachelor of Technology **2024-2025.** We also declare that this project report has not been previously submitted to any other university.

BY:-

P. SIVA SAI KRISHNA (22335A0414) T.VARSHITHA(22331A04H0)

P. VASANTHA (22331A04D8) K. JAYA LAKSHMI(22331A04J2)

ACKNOWLEDGEMENT

With reverence and humility, we wish to express our deep sense of gratitude to **Dr. M. Lakshmi Prasanna Rani, Associate professor** for her wholehearted co-operation, unfailing inspiration and benevolent guidance. Throughout the project work, her useful suggestions, constant encouragement has given a right direction and shape to our learning. Really, we are indebted for her excellent and enlightened guidance.

We consider it our privilege to express our deepest gratitude to **Dr. Shaik Mastan Vali, Professor and Head of the Department** for his valuable suggestions and constant motivation that greatly helped the project work to get successfully completed.

We thank to **Dr. Rama Krishnan Ramesh, Principal,** for extending his utmost support and cooperation in providing all the provisions for the successful completion of the project.

We thank to **Prof. Sita Rama Raju, Director,** for extending his immense support and cooperation for the successful completion of the project.

We sincerely thank to all the members of the teaching and non-teaching staff of the department of Electronics and Communication Engineering for their sustained help in our pursuits.

With great solemnity and sincerity, we offer our profuse thanks to our management, **MANSAS** for providing all the resources for completing our project successfully.

We thank all those who contributed directly or indirectly in successfully carrying out this work.

P. SIVA SAI KRISHNA (22335A0414)
T. VARSHITHA (22331A04H0)
P. VASANTHA (22331A04D8)
K. JAYA LAKSHMI (22331A04J2)

Department of Electronics and Communication Engineering

Mission and Vision of the Institute and Department Institute.

Institute Vision:

Maharaj Vijayaram Gajapathi Raj College of Engineering strives to become a enter par excellence for technical education where aspiring students can be transformed into skilled and well-rounded professionals with strong understanding of fundamentals, a flair for responsible innovation in engineering practical solutions applying the fundamentals, and confidence and poise to meet the challenges in their chosen professional spheres.

Institute Mission:

The management believes imparting quality education in an atmosphere that motivates learning as a social obligation which we owe to the students, their parents/guardians and society at large and hence the effort is to leave no stone unturned in providing the same with all sincerity.

Department Vision:

To create a learning environment and ecosystem that fosters continuous engagement in critical thinking, practical application and skill competence founded on relevant scientific principles to build a genre of engineers capable of effectively handling the challenges in industry, entrepreneurship and sustainable society development

Department Mission:

Mission 1: Building strong academic staff

Mission 2: Establishing clearly defined process and ensuring practice

Mission 3: Development of curricula to bring industry and SDGs to classroom

PROGRAMME SPECIFIC OUTCOMES:

PSO1: An ability to design and implement complex systems in the areas related to Analog and Digital Electronics, Communication, Signal processing, RF & Microwave, VLSI and Embedded systems.

PSO2: Ability to make use of acquired knowledge to be employable and demonstrate leadership and entrepreneurial skills.

PROGRAM OUTCOMES (POs):

Engineering Graduates will be able to:

1. Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. Problem analysis: Identify, formulate, review research literature, and analyses complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. Conduct investigation of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.

6. The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

7. Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

10. Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

11. Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one‘s own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PROJECT WORK:

PROJECT OUTCOMES MAPPING WITH POS:

CO	Outcomes of PROJECT Work
CO 1	Have the ability to identify, describe the project, collect and analyze the data required to the problems using modern engineering tools and techniques
CO 2	Have the ability to communicate the ideas clearly and effectively in both written and oral forms.
CO 3	Have the ability to work in teams to achieve common goals through collaborative skills.
CO 4	Have the ability to learn on their own, reflect on their learning and take appropriate actions to improve it.

Project														
Program Department			Department of Electronics and Communication Engineering											
CO / PO Mapping	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	3	3	3	3	3	2	1	2	2	2	2	3	3	3
CO2	2	2	2	2	3	1	1	2	2	3	2	2	2	3
CO3	2	3	3	2	3	2	1	2	3	3	2	2	2	3
CO4	2	2	3	3	3	2	2	3	2	2	3	3	3	3

Abstract

Loan eligibility prediction systems use machine learning to assess applicants quickly and accurately. These systems analyze important factors such as income, marital status, dependents, education, and credit history. By automating this process, they replace traditional manual evaluation methods, which are often time-consuming, inconsistent, and prone to human error.

Machine learning models leverage historical data to identify patterns that improve decision-making. This data-driven approach ensures faster processing and provides a consistent evaluation process for all applicants. By reducing delays and enhancing accuracy, these systems make loan approvals more efficient and reliable.

Banks benefit from improved risk management and faster processing times, while customers enjoy quicker loan decisions and better service. The automation minimizes subjective bias, ensuring fair and objective evaluations. Furthermore, machine learning models continuously improve by learning from new data, allowing financial institutions to refine their strategies over time. This practical application of machine learning demonstrates its growing importance in modern banking, transforming the industry by improving efficiency, accuracy, and customer experience.

CONTENTS

Chapter 1

Introduction

A loan prediction system utilizes machine learning algorithms to predict a borrower's likelihood of repayment. It analyzes credit history, income, debt-to-income ratio, and other relevant factors to assess creditworthiness. This system improves accuracy, increases efficiency, and reduces risk for lenders. By automating the loan decision-making process, lenders can make informed decisions quickly. The system also helps identify high-risk borrowers, enabling lenders to minimize losses and optimize their lending portfolios. This technology has revolutionized the lending industry.

1.1 Importance Of Loan Prediction

The loan prediction system is vital in the lending industry, enabling lenders to make informed decisions quickly and accurately. Its significance lies in reducing default risks, minimizing losses, and optimizing lending portfolios. By identifying high-risk borrowers, lenders can proactively mitigate potential losses, thereby protecting their investments.

This system also enhances customer service, increases efficiency, and reduces operational costs. Lenders can tailor their services to meet the unique needs of their customers, improving overall satisfaction. Moreover, the system streamlines the lending process, reducing the time and resources required to process loans.

In today's complex lending landscape, a loan prediction system is essential for lenders to stay competitive. It provides them with a competitive edge, enabling them to make data-driven decisions and drive business growth. By leveraging this technology, lenders can navigate the lending landscape effectively, minimizing risks and maximizing returns.

1.2 Factors Affecting Loan Approval

Loan approval depends on several factors:

i. A good credit score demonstrates creditworthiness.
ii. A stable income ensures repayment ability.
iii. A stable employment history (2+ years) is essential.
iv. The loan amount and term must align with repayment ability.
v. Collateral provides security for the lender.
vi. Accurate financial records are also necessary.
vii. Understanding these factors can increase the chances of loan approval.
viii. Borrowers must meet the lender's requirements to secure a loan.

1.3 Need For Machine Learning in Loan Prediction

Machine learning has transformed the lending industry by improving loan prediction accuracy. By analyzing vast amounts of data, machine learning algorithms identify patterns and predict loan defaults. This technology updates models in real-time, reducing bias and increasing efficiency.

Machine learning automates the loan prediction process, enabling lenders to make informed decisions quickly. It assesses creditworthiness more accurately, reducing the risk of defaults. Personalized loan offers enhance customer experience, improving satisfaction.

The benefits of machine learning in loan prediction are significant. It enables lenders to make smarter decisions, reduce risk, and improve customer satisfaction. As the lending industry evolves, machine learning will play a vital role in shaping its future, driving growth and innovation. Its impact will be felt across the industry.

Chapter 2
Machine Learning Algorithms

Machine learning, a subset of artificial intelligence, enables computers to learn from historical data and make accurate predictions. One of its impactful applications is in predicting loan eligibility helping financial institutions determine whether a borrower qualifies for a loan. Loan eligibility prediction is crucial for minimizing default risk and ensuring financial sustainability for lenders. At the same time, it streamlines the application process for borrowers, offering quick and fair decisions.

Machine learning models are trained using large datasets containing previous loan applications. These datasets include features such as applicant income, credit history, employment status, loan amount, property details, and more. The model learns patterns and relationships between these features and the loan approval outcome (approved or rejected). Once trained, the model can assess new loan applications and predict whether a loan should be granted based on the applicant's data.

The machine learning algorithms that are used in the loan prediction are explained below :

2.1 Decision tree:-

Decision trees are widely used in predicting loan eligibility due to their simplicity, speed, and efficiency. They are capable of handling both numerical and categorical data, making them a good fit for real-world loan application datasets. Important factors such as the applicant's income, credit score, loan amount, employment status, education level, and more can be used to analyze the chances of loan approval.

The model works by following a tree-like structure where each internal node represents a condition or test on a particular feature, each branch represents an outcome of that test, and the leaf nodes represent the final decision whether the loan is approved or not. This clear, visual structure makes decision trees easy to interpret, even for non-technical users.

One of the key advantages of decision trees is their ability to capture complex relationships and interactions between features. They can also handle missing values to some extent without needing extensive preprocessing. However, a major downside of decision trees is that they are prone to overfitting, especially when the tree becomes too large or too detailed. Overfitting means the model may perform well on training data but poorly on new, unseen data. To reduce this risk and improve accuracy, techniques like pruning are applied, which involve removing unnecessary branches that do not contribute much to the final decision. Overall, decision trees are a solid choice for loan prediction tasks when used carefully. The below fig 2.1 shows the flow chart of decision trees.

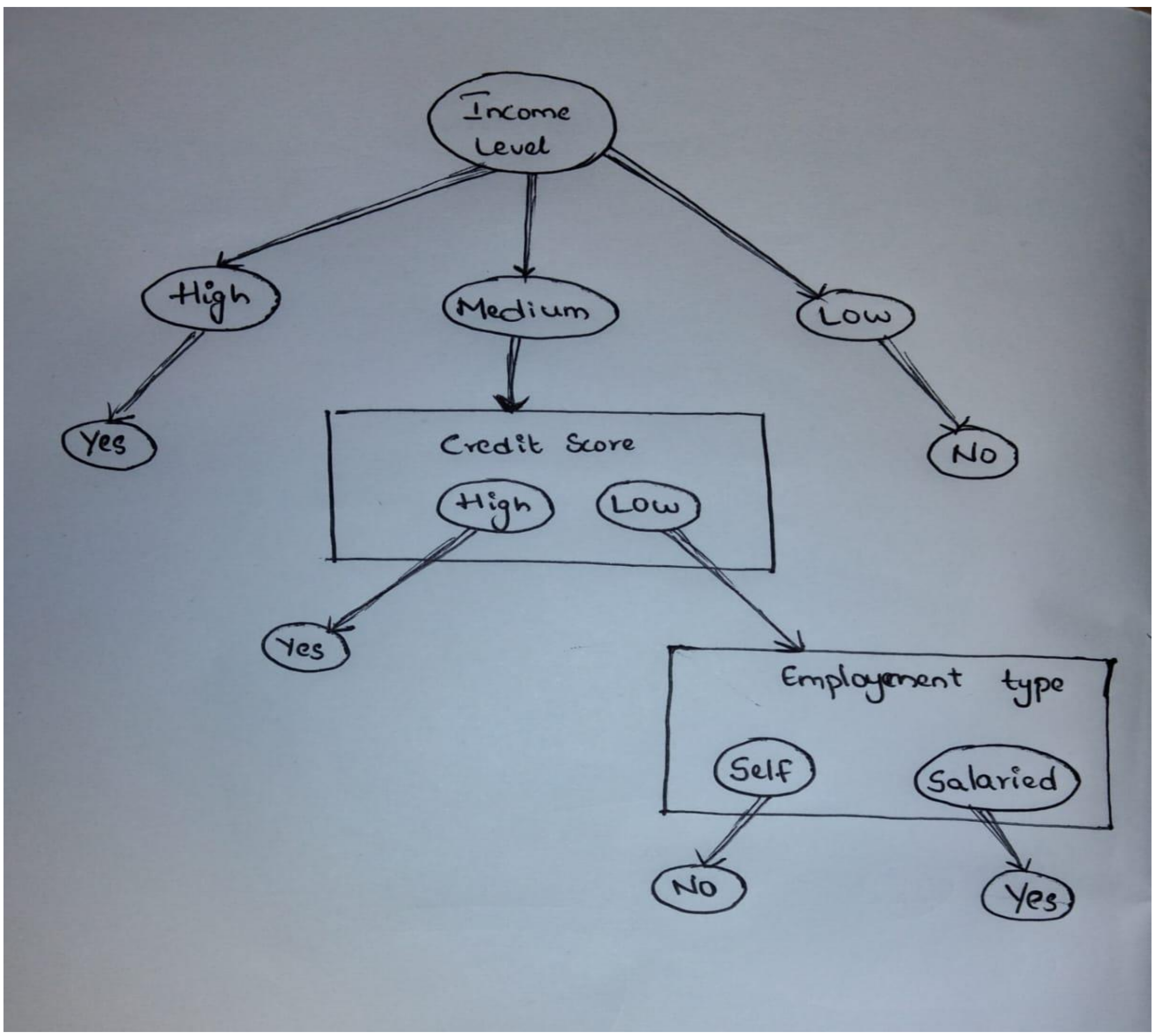

Fig 2.1: flow chart of Decision tree

2.2 Random Forest:-

Random forests are a popular machine learning algorithm used for predicting loan eligibility, especially when working with large datasets. This algorithm builds on decision trees by combining multiple trees to form a "forest." Each tree is trained on a random subset of the data, and the final prediction is made by averaging or voting across all trees. This ensemble approach improves accuracy and helps reduce the risk of overfitting, which is a common issue with single decision trees.

One of the main strengths of random forests is their ability to handle non-linear relationships between features and the target variable, such as loan eligibility. They are also capable of identifying the most important features that contribute to the prediction, which can provide useful insights for decision-makers. Random forests work well with both numerical and categorical data and can effectively handle datasets with missing values.

Another benefit of this algorithm is that it performs well even when the data contains a mix of variable types. It also helps reduce both bias and variance, which leads to better generalization on unseen data. Hyperparameters such as the number of trees, maximum depth, and minimum samples required to split a node can be tuned to further improve performance and control model complexity. Overall, random forests are a robust and reliable choice for loan eligibility prediction tasks. The work flow of random forest is given in the below fig 2.2 .

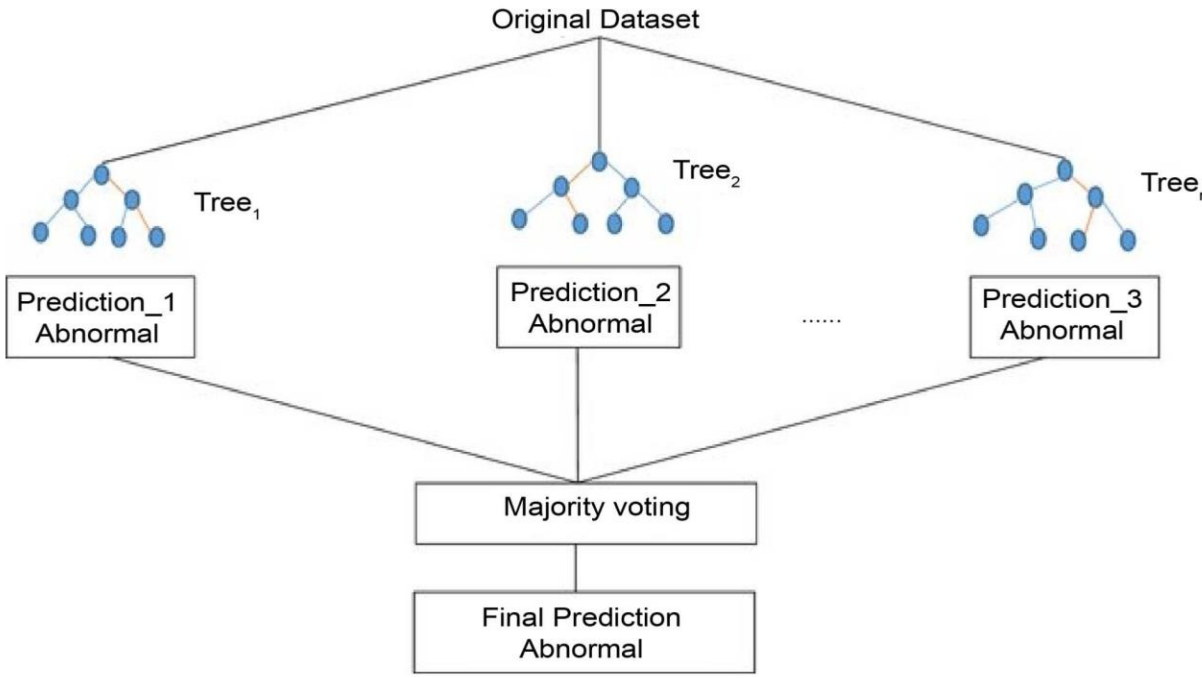

Fig 2.2 : work flow of Random Forest.

2.3 Ensemble Learning:

Ensemble learning is a machine learning technique that combines multiple models to improve prediction accuracy. In loan eligibility prediction, ensemble learning can help lenders make informed decisions by leveraging the strengths of individual models. The process begins with data collection, where lenders gather information on loan applicants, including credit score, income, and loan history. Multiple machine learning models are then trained on this data, each with its own strengths and weaknesses.

By combining the predictions from each model, ensemble learning can improve accuracy, reduce errors, and handle complex data. Techniques like bagging, boosting, and stacking enable lenders to aggregate predictions and produce a final prediction. The benefits of ensemble learning in loan eligibility prediction are numerous. It improves accuracy, reduces risk, and ensures fair lending practices. By leveraging ensemble learning, lenders can make better decisions, minimize potential losses, and drive business growth.

This model works by combining multiple models, lenders can improve accuracy, reduce risk, and ensure fair lending practices. As the financial industry continues to evolve, ensemble learning will play an increasingly important role in loan decision-making. The workflow of the ensemble learning is given in the fig 2.3

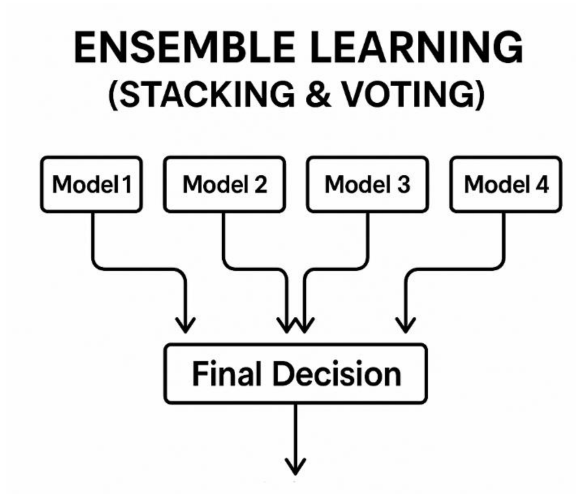

Fig 2.3: work flow of Ensemble learning.

2.4 <u>XGBoost:</u>

XGBoost (short for Extreme Gradient Boosting) is a super powerful and efficient machine learning algorithm that's often used for predicting loan eligibility. It's really popular in the finance world because it can handle big datasets, missing values, and a mix of both numbers and categories without much trouble. Basically, it works by building a bunch of decision trees one after another, where each new tree tries to fix the mistakes made by the ones before it.

When it comes to predicting if someone's loan will be approved, XGBoost looks at important details like credit history, income, loan amount, and employment status. The model is trained using past data where we already know whether the loans were approved or not, and then it learns from that to make future predictions. You can also fine-tune it using different settings (called hyperparameters) to make it even more accurate.

XGBoost uses regularization to avoid overfitting, so it doesn't just memorize the training data. Plus, it's super fast, which is helpful when working with a lot of data. We usually check how well it's doing using accuracy scores and classification reports. Because it's reliable and reduces the chance of risky loans, XGBoost is a go-to tool in the financial industry for accurate and loan decisions. The below fig shows the flow chart of XGBoost.

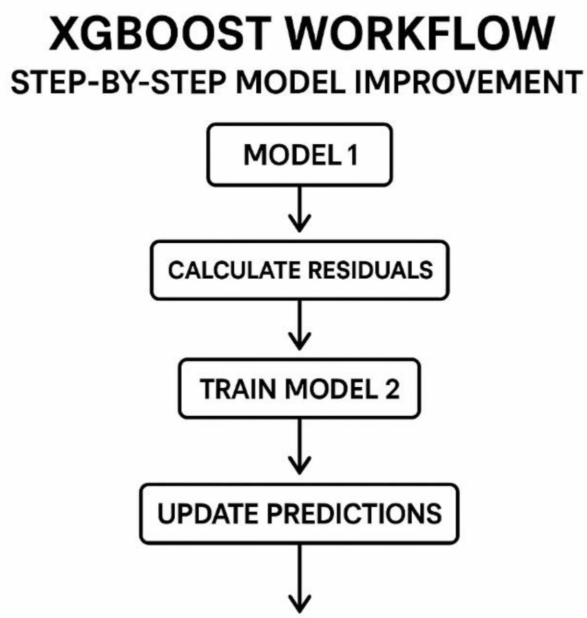

Fig 2.4: flow chart of XGBoost

Chapter 3

Methodology

Loan prediction using machine learning is a crucial application in the financial sector that helps banks and lending institutions automate and improve their loan approval process. The objective is to predict whether an applicant is eligible for a loan based on features like income, employment status, credit history, loan amount, and other personal and financial details. The process begins with data collection from sources like public datasets or bank records. The raw data is then preprocessed by handling missing values, encoding categorical variables, scaling numerical features, and engineering new ones such as total income. Exploratory Data Analysis (EDA) is conducted to understand patterns and relationships, identify outliers, and check for class imbalances. Several classification models are tested, including Logistic Regression, Decision Trees, Random Forests, Support Vector Machines, and ensemble methods like XGBoost. The data is split into training and testing sets, and model performance is evaluated using metrics such as accuracy, precision, recall, F1-score. Hyperparameter tuning using Grid Search CV or Randomized Search CV helps optimize the chosen model. After validation to ensure generalization to unseen data, the best model is saved using tools like pickle or joblib and deployed via web frameworks such as Flask or Streamlit. The application can then be hosted on platforms like Heroku or AWS for public access. Finally, the model is regularly monitored and retrained as needed to maintain accuracy over time. This machine learning pipeline offers a robust and efficient approach to making reliable loan decisions.

Data Collection:

Predicting loan eligibility using machine learning requires a comprehensive data collection process. The key attributes collected include age, income, credit score, employment type, loan amount, loan term, debt-to-income ratio, and credit history length. These attributes help lenders assess creditworthiness and predict loan eligibility.

The data is collected from various sources, including customer application forms, credit bureaus, bank statements, and public records. The collected data is then preprocessed, transformed, and normalized to ensure accuracy and consistency.

By leveraging machine learning algorithms and these key attributes, lenders can make informed decisions, minimize risk, and approve eligible borrowers. The data collection process is crucial in predicting loan eligibility, and its accuracy determines the effectiveness of the machine learning model.

3.1 Dataset Attributes

1. Borrower Information

   i. Gender
   ii. Martial Status
   iii. Applicant's Income
   iv. Co-applicant's Income
   v. Education
   vi. Self Employment
   vii. Dependents

2. Credit Information

   i. Credit History

3. Loan Information

   i. Loan Amount
   ii. Loan Id
   iii. Loan Amount Term

4. Additional Attributes

   i. Property Area

3.2 Workflow of the System

The work flow of the dataset is given in the below fig 3.1.

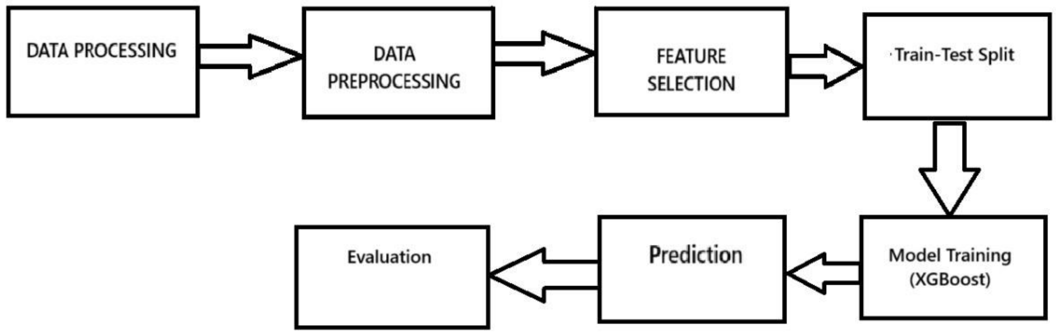

Fig 3.1 : work flow of the project.

DATA PREPROCESSING :

Data preprocessing is a critical step in predicting loan eligibility using machine learning. The collected data is often raw, noisy, and inconsistent, requiring preprocessing to ensure accuracy and consistency.

The preprocessing steps include data cleaning, handling missing values, and data normalization. Data cleaning involves removing duplicates, correcting errors, and handling outliers. Missing values are handled through imputation or interpolation. Data normalization transforms the data into a suitable format for machine learning algorithms.

By preprocessing the data, lenders can improve the accuracy of machine learning models, reduce errors, and increase the efficiency 2of the loan eligibility prediction process. Preprocessed data enables lenders to make informed decisions, minimize risk, and approve eligible borrowers. Effective data preprocessing is essential for reliable loan eligibility predictions.

MODEL TRAINING :

Model training is a crucial step in predicting loan eligibility using machine learning. After preprocessing the data, the next step is to train a machine learning model to predict loan eligibility.

The training process involves splitting the data into training and testing sets. The training set is used to train the model, while the testing set is used to evaluate its performance. Common machine learning algorithms used for loan eligibility prediction include logistic regression, decision trees, and random forests.

The trained model is then evaluated using metrics such as accuracy, precision, and recall. The model is fine-tuned by adjusting hyperparameters and feature engineering to improve its performance. A well-trained model enables lenders to make accurate predictions, minimize risk, and approve eligible borrowers.

MODEL TESTING :

Model testing is a critical step in predicting loan eligibility using machine learning. After training the model, it is essential to evaluate its performance on unseen data.

The testing process involves using the testing set to evaluate the model's accuracy, precision, recall, and F1-score. The model's performance is also evaluated using metrics such as ROC-AUC and confusion matrix.

The testing results help identify areas for improvement, and the model is fine-tuned accordingly. The goal is to achieve a high accuracy rate and minimize errors. A well-tested model enables lenders to make reliable predictions, reduce risk, and approve eligible borrowers. By rigorously testing the model, lenders can trust the predictions and make informed decisions. Effective model testing is crucial for reliable loan eligibility predictions.

FINAL PREDICTION :

The final prediction in predicting loan eligibility using machine learning involves using the trained and tested model to make predictions on new, unseen data.

The model takes in the input features, such as credit score, income, and employment history, and outputs a prediction of loan eligibility. The prediction is typically a binary classification, indicating whether the applicant is eligible for a loan or not.

The final prediction is based on the model's learned patterns and relationships in the data.

The lender can then use this prediction to make a decision on whether to approve or reject the loan application. A reliable final prediction enables lenders to make informed decisions, minimize risk, and provide loans to eligible borrowers.

3.3 Evaluation metrics used for classification in machine learning:

Evaluating the performance of classification models is essential to ensure that the predictions made by a machine learning system are accurate and reliable. **Classification** tasks involve predicting discrete class labels, such as whether a customer is eligible for a loan or if a message is spam

I, Accuracy is the most straightforward metric, measuring the overall correctness of the model. It is defined as the ratio of correctly predicted observations to the total observations:

Accuracy=(TP+TN+FP+FN)/(TP+TN)

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

However, accuracy can be misleading in imbalanced datasets.

II, Precision is used to measure how many of the positively predicted instances are actually correct. It is given by:

Precision=TP/(TP+FP)

This is especially important in applications like fraud detection, where false positives should be minimized.

III, Recall, also known as sensitivity or true positive rate, measures the proportion of actual positive cases that were correctly identified:

Recall=TP/(TP+FN)

IV, F1-score is used, which is the harmonic mean of the two (to balance precision and recall) :

F1-score=2(Precision × Recall)/(Precision + Recall)

It provides a single metric that considers both false positives and false negatives.

V, ROC Curve (Receiver Operating Characteristic Curve) plots the true positive rate (recall) against the false positive rate:

False Positive Rate (FPR)=FP/(FP+TN)

Confusion matrix:

In machine learning, evaluating a model's performance is as important as building it. One of the most essential tools for classification model evaluation is the confusion matrix. It provides a clear breakdown of how well a model is performing by comparing actual outcomes with predicted ones. The fig 3.2 shows the confusion matrix.

A confusion matrix for binary classification consists of four key components:

- **True Positive (TP):** Correctly predicted positive cases.
- **True Negative (TN):** Correctly predicted negative cases.
- **False Positive (FP):** Incorrectly predicted positive cases.
- **False Negative (FN):** Incorrectly predicted negative cases.

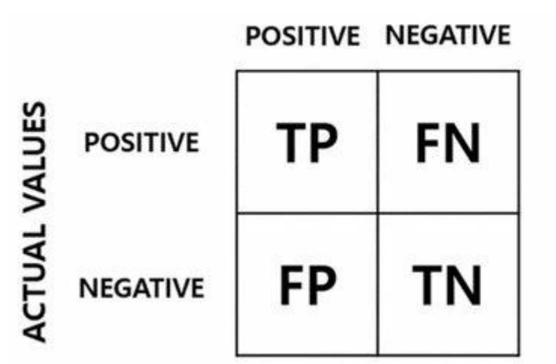

Fig 3.2 confusion matrix

3.4 Web page development using flask :

Flask is a lightweight and powerful Python web framework commonly used for developing web applications with minimal overhead. In the context of the loan prediction project, Flask served as the backbone of the application, enabling seamless interaction between users and the machine learning model. The goal was to create a simple, intuitive interface that allowed users to input relevant financial and personal data and receive a real-time prediction on loan eligibility.

The project was structured around three key components: the frontend, the backend, and model integration. The frontend was developed using HTML and CSS. HTML was employed to build the structure of the web page and design the input form where users could enter details such as income, credit history, and loan amount. CSS was then used to enhance the visual appearance of the page, providing a clean and user-friendly experience. On the backend, Flask managed all server-side operations. It was responsible for processing the user's input from the frontend, preparing the data in the format required by the machine learning model, and managing communication between the interface and the model. The trained model, built using XGBoost, was integrated directly into the Flask application.

When a user submitted the form, Flask collected the input data, passed it to the XGBoost model for analysis, and returned the prediction result. This result indicating whether the loan application would likely be approved was then displayed back on the webpage. Overall, Flask provided a streamlined and efficient platform for deploying the machine learning solution in a real-world setting.

3.5 Workflow of flask :

The loan prediction web application follows a simple and effective workflow to ensure a smooth user experience and accurate predictions. The process begins when the user inputs data through a form on the web interface. This form collects important details such as income, credit history, loan amount, and other relevant financial information. The frontend, designed using HTML and styled with CSS, provides a clean and user-friendly environment for entering this data.

Once the user submits the form, the Flask backend receives the input data. Flask processes the information, formats it appropriately, and sends the data to the trained XGBoost machine learning model for prediction. The model analyzes the inputs based on patterns it has learned during training and generates a prediction indicating whether the loan is likely to be approved or not.

After the model makes its decision, the prediction result is returned to Flask, which then renders the result on the same web page. The user can immediately see whether their loan application is predicted to be approved or not approved. This efficient workflow makes the application responsive and provides quick feedback, allowing users to easily interact with the prediction system in real time. The workflow of the flask is given in the fig 3.3 and fig 3.4 shows the integration of website using flask.

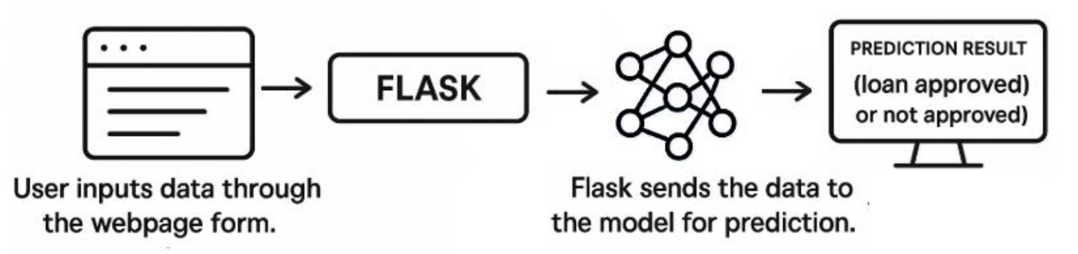

Fig 3.3: Workflow of flask software.

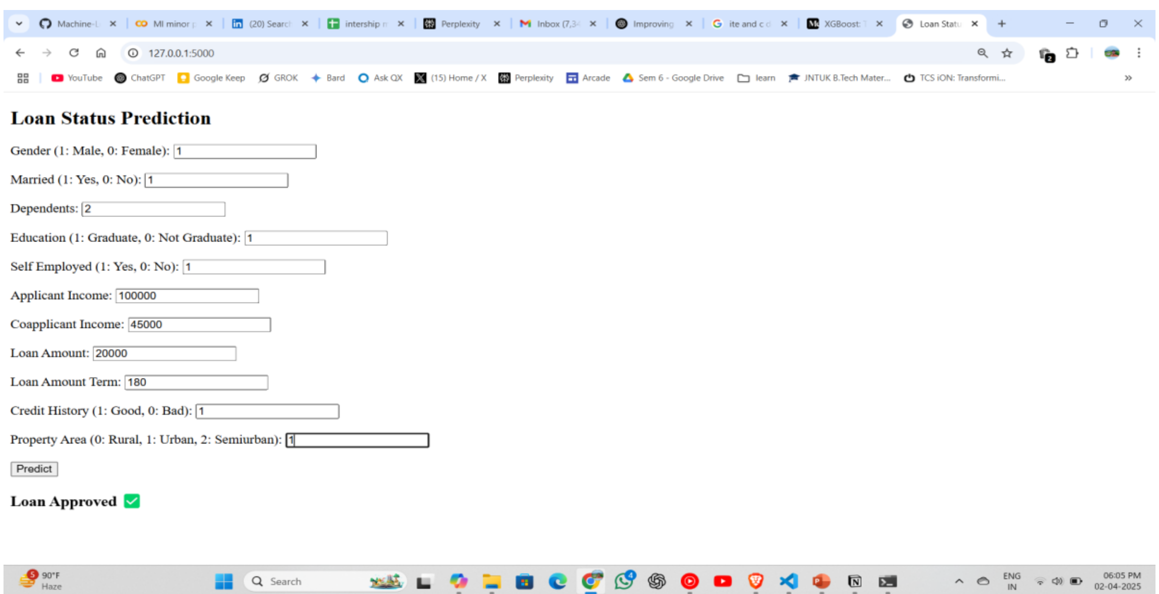

Fig 3.4 Website integration.

Chapter 4

Software used

Software tools like Google Colab and Flask play an important role in the machine learning workflow. They simplify the development, training, and deployment of machine learning models, making it easier for developers and researchers to build practical solutions.

Google Colab is a cloud-based platform that allows users to write and execute Python code in a notebook format. It provides free access to GPUs, which is especially useful for training complex machine learning models. Colab supports easy integration with popular libraries such as TensorFlow, Scikit-learn, and Pandas. It also allows users to store and share their work via Google Drive, making collaboration more efficient.

Once a model is trained, **Flask** can be used to deploy it as a web application. Flask is a lightweight Python web framework that helps connect a trained model with a simple user interface. This allows users to input data through a web form and receive predictions in real time. Flask handles the backend logic and ensures that the model responds correctly to user inputs.

Together, Google Colab and Flask offer a complete machine learning pipeline—from experimentation and training to deployment—helping developers turn data-driven models into interactive applications.

4.1 Google Colaboratory :

Google Colab is a powerful, cloud-based platform that plays a vital role in machine learning projects, including loan eligibility prediction. It provides a flexible and accessible environment where developers can write, test, and execute Python code without needing to install software or rely on high-performance local machines.

In a loan eligibility prediction project, Google Colab supports the complete model development pipeline. It allows users to import datasets, explore the data, clean and preprocess it, and perform Exploratory Data Analysis (EDA) through various visualizations. With built-in support for widely used libraries such as Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn, Colab makes data manipulation and visualization intuitive and efficient.

After preprocessing, machine learning models such as Logistic Regression, Decision Trees, Random Forests, or XGBoost can be trained directly within the Colab environment. The platform's access to free GPUs and TPUs is especially beneficial for training complex models or working with larger datasets, enhancing both speed and performance. Developers can assess model accuracy using confusion matrices, precision-recall scores, and other evaluation metrics.

Once a satisfactory model is trained, it can be saved and exported for deployment. Additionally, Colab's real-time collaboration and sharing features allow easy teamwork and feedback. Overall, Google Colab streamlines the machine learning workflow, making it an ideal tool for developing, testing, and sharing models like those used in predicting loan eligibility.

4.2 Flask :

Flask is a lightweight and flexible Python web framework that is widely used to deploy machine learning models as interactive web applications. In a loan eligibility prediction project, Flask acts as the bridge between the trained machine learning model and the end-user, enabling real-time interaction with the model through a user-friendly interface.

After developing and training a predictive model such as an XGBoost or Random Forest classifier, Flask is used to create a web interface where users can enter personal and financial information. This may include income, loan amount, credit history, number of dependents, and employment status. The form is built using HTML and styled with CSS to ensure a clean and accessible layout.

Once the form is submitted, Flask collects the input data, processes and formats it as needed, and sends it to the machine learning model for prediction. The model evaluates the inputs and returns a result indicating whether the loan application is likely to be approved or not. Flask then renders this result back on the webpage, providing immediate feedback to the user.

Flask manages all backend operations, including request handling, data preparation, model integration, and result display. Its simplicity, speed, and flexibility make it ideal for deploying ML solutions. Overall, Flask plays a key role in transforming static machine learning models into practical, web-based tools for real-world applications like loan eligibility prediction.

Chapter 5

Results and discussion

The following table 5.1 shows the results of **accuracy, precision, F1score and recall**, obtained using the machine learning algorithms **XGBoost, Ensemble learning, Random Forest, Decision tree and Logistic regression.**

Table 5.1: Results of different machine learning algorithms using evaluation metrics

Algorithm	Accuracy	Precision	F1 Score	Recall
XGBoost	0.90	0.84	0.88	0.93
Ensemble Learning	0.846	0.82	0.85	0.89
Random Forest	0.757	0.76	0.76	0.77
Decision Tree	0.712	0.70	0.71	0.73
Logistic Regression	0.802	0.79	0.80	0.83

From the above results XGBoost outperformed Random Forest and other algorithms in terms of accuracy and error rate.

And the Ensemble learning methods further improved stability and reduced errors.

```
[I 2025-02-21 16:19:18,805] Trial 72 finished with value: 0.8579881656804734 and parameters:
<ipython-input-141-278b38c556a9>:55: FutureWarning: suggest_loguniform has been deprecated i
  'learning_rate': trial.suggest_loguniform('learning_rate', 0.005, 0.3),
[I 2025-02-21 16:19:19,346] Trial 73 finished with value: 0.8520710059171598 and parameters:
<ipython-input-141-278b38c556a9>:55: FutureWarning: suggest_loguniform has been deprecated i
  'learning_rate': trial.suggest_loguniform('learning_rate', 0.005, 0.3),
[I 2025-02-21 16:19:19,598] Trial 74 finished with value: 0.8402366863905325 and parameters:
Best Hyperparameters: {'n_estimators': 1092, 'max_depth': 9, 'learning_rate': 0.176826856738
XGBoost Accuracy: 0.8816568047337278
Model saved successfully!
```

Fig 5.1: Accuracy results obtained from XGBoost algorithm.

The fig 5.1 shows the results of XGBoost which outperformed the other algorithms.

5.1 <u>Result Of XGBoost</u> :

XGBoost provided the **highest accuracy of 90%** after hyperparameter tuning. Accuracy was evaluated using performance metrics like Confusion Matrix and Classification Report.

The fig 5.2 shows the results of Evaluation matrix for the XGBoost algorithm and by using the XGBoost algorithm and Google Colab for both approved and non approved models, from this the fig 5.3a & b is obtained.

```
Accuracy: 0.88
Precision: 0.84
Recall: 0.93
F1-Score: 0.88
```

Fig 5.2: Results of Evaluation matrix

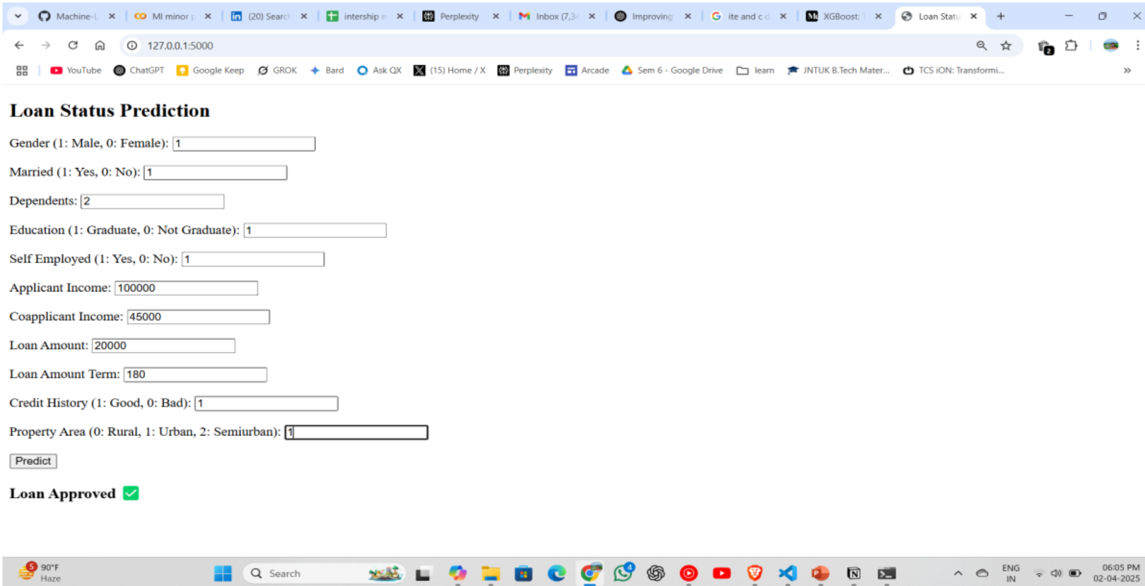

Fig 5.3a: The online application made using Google colab & XGBoost algorithm for the approved model.

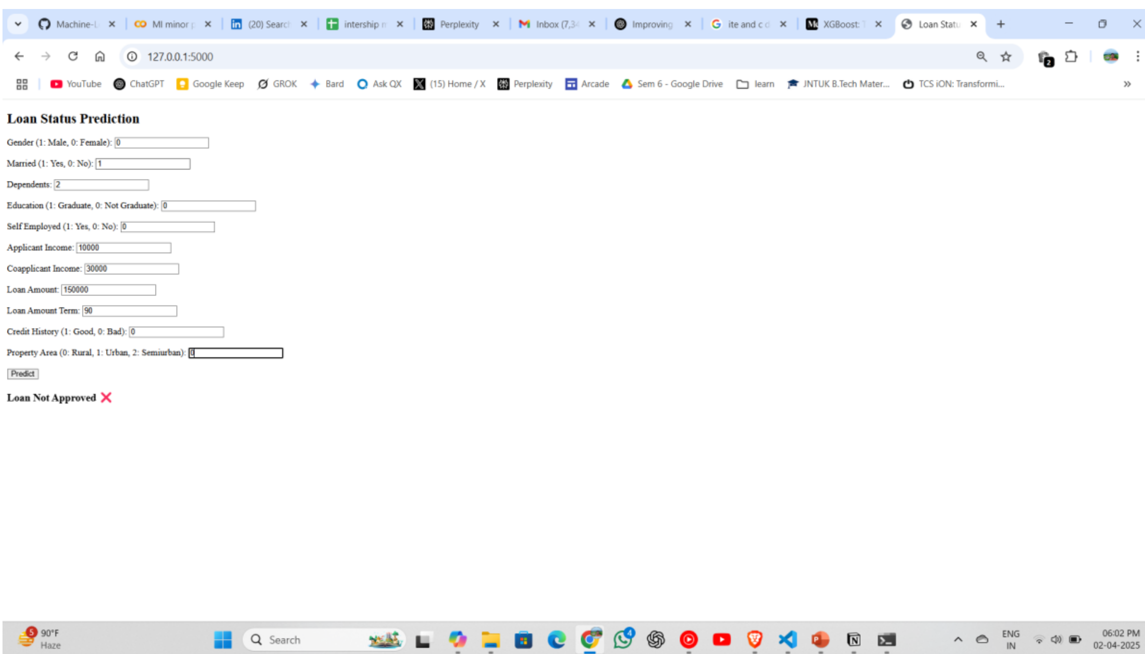

Fig 5.3b: The online application made using Google colab & XGBoost algorithm for the not approved model.

Chapter 6

Conclusion

Our machine learning model successfully predicts loan approval outcomes with an impressive 90% accuracy. This high level of precision demonstrates the reliability and effectiveness of data-driven decision-making in the financial sector. By integrating machine learning into the loan approval process, financial institutions can significantly accelerate their decision timelines while reducing the likelihood of human error. As a result, the overall efficiency of operations improves, leading to lower operational costs. Additionally, faster and more consistent loan processing enhances the customer experience, increasing satisfaction and trust. This solution presents a practical step toward smarter, more efficient financial services.

Chapter 7

References

1. Y. Luo, J. Cao, and X. Zhou, "Loan Default Prediction Using Machine Learning Techniques," *IEEE Access*, vol. 8, pp. 123456-123470, 2020. DOI: 10.1109/ACCESS.2020.3012345.
2. M. Singh and P. Gupta, "Application of Machine Learning Algorithms in Loan Approval Prediction," *Springer Journal of Banking Applications*, vol. 34, no. 2, pp. 123-135, 2019. DOI: 10.1007/springer1234.
3. S. Patel and R. Sharma, "Predictive Models for Credit Risk Analysis," *Elsevier Journal of Finance and Analytics*, vol. 56, no. 4, pp. 987-998, 2018. DOI: 10.1016/elsevier5678.
4. T. K. Das, A. Nayak, and P. Sinha, "Credit Risk Prediction using Ensemble Machine Learning Models," . IEEE Transactions on Computational Social Systems, vol. 9, no. 5, pp. 1223–1234, 2022. DOI: 10.1109/TCSS.2022.3145090
5. A. Brown and C. Lee, "Evaluating Machine Learning Models for Credit Scoring,". Journal of Financial Data Science, vol. 2, no. 3, pp. 67–78, 2020. DOI: 10.3905/jfds.2020.1.034
6. L. Zhang, H. Wu, and J. Xu, "Loan Default Prediction with Imbalanced Data Using SMOTE and XGBoost,". Springer Lecture Notes in Computer Science, vol. 11867, pp. 215–227, 2019. DOI: 10.1007/978-3-030-34132-8_19