**Software Requirements Specification**

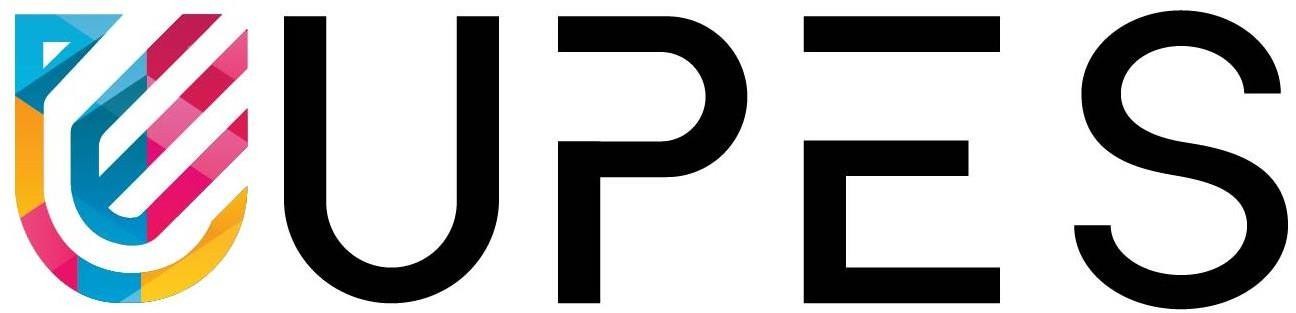
For

# Loan Approval Prediction

6th March, 2025

Prepared by

|  |  |  |
| --- | --- | --- |
| **Specialization** | **SAP ID** | **Name** |
| B.Tech-CSE (DS NON- HONS) | 500105453 | RITIK RAI |
| B.Tech-CSE (DS NON- HONS) | 500107214 | RAGHAV JAIN |
| B.Tech-CSE(DS NON-HONS) | 500105275 | AARON SINGH |



Department of Data Science

School Of Computer Science

UNIVERSITY OF PETROLEUM & ENERGY STUDIES,

DEHRADUN- 248007. Uttarakhand

**Dr. Supreet Singh Virendra Kadyan**

**Project Guide Cluster Head**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Topic** | | **Page No** |
| Table of Content | |  |
| Revision History | |  |
| 1 | Introduction | 4 |
|  | 1.1 Purpose of the Project | 4 |
|  | 1.2 Target Beneficiary | 4 |
|  | 1.3 Project Scope | 5 |
|  | 1.4 References | 5 |
| 2 | Project Description |  |
|  | 2.1 Reference Algorithm | 6 |
|  | 2.2 Data/ Data structure | 6 |
|  | 2.3 SWOT Analysis | 7 |
|  | 2.4 Project Features | 7 |
|  | 2.5 User Classes and Characteristics | 7 |
|  | 2.6 Design and Implementation Constraints | 7 |
|  | 2.7 Assumption and Dependencies | 7 |
| 3 | System Requirements |  |
|  | 3.1 User Interface | 8 |
|  | 3.2 Software Interface | 8 |
|  | 3.3 Database Interface | 8 |
|  | 3.4 Protocols | 8 |
| 4 | Non-functional Requirements |  |
|  | 4.1 Performance requirements | 9 |
|  | 4.2 Security requirements | 9 |
|  | 4.3 Software Quality Attributes | 9 |
| 5 | Other Requirements | 10 |
| Appendix A: Glossary | | 10 |
| Appendix B: Analysis Model | | 10 |
| Appendix C: Issues List | | 10 |

**Revision History**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Change** | **Reason for Changes** | **Mentor Signature** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

1. INTRODUCTION

The Loan Approval Prediction System automates loan eligibility assessment using machine learning, reducing manual effort and ensuring faster, data-driven decisions. It provides a web-based interface where users can enter financial details, and ML models predict loan approval status (Approved, Conditionally Approved, or Rejected).

1.1 Purpose of the Project

The Loan Approval Prediction System aims to simplify and automate the initial screening process for loan applications. Traditional loan approval methods involve manual assessment, leading to delays, inconsistencies, and potential biases in decision-making. This project leverages machine learning (ML) models to provide an efficient, unbiased, and data-driven approach to predicting loan eligibility.

Key Objectives:

1. Enhance Efficiency: The system will significantly reduce processing time by automatically predicting loan approval status based on financial parameters.
2. Improve Accuracy: Using ML algorithms such as Decision Trees and Random Forests ensures more consistent and objective evaluations.
3. User-Friendly Interface: A web-based application will allow users to easily enter their details and receive a decision in real-time.
4. Customizable Thresholds: Financial institutions can set their own approval criteria (e.g., minimum credit score, debt-to-income ratio).
5. Transparency & Explainability: The system will highlight the key factors influencing the decision, ensuring transparency in predictions.

This project is beneficial for loan applicants, banks, and financial analysts, reducing manual work, increasing accuracy, and providing a seamless user experience.

1.2 Target Beneficiary

The Loan Approval Prediction System is designed to benefit various stakeholders involved in the loan approval process. The key beneficiaries include:

1. Loan Applicants (Individuals & Businesses)

* Self-Assessment: Individuals can check their loan eligibility before applying, reducing unnecessary rejections.
* Transparency: Users gain insights into factors affecting their approval chances, such as credit score, income, and debt-to-income ratio.
* Time-Saving: Eliminates the need for multiple visits to financial institutions by providing instant feedback on loan eligibility.

2. Financial Institutions (Banks, NBFCs, Credit Unions, Fintech Companies)

* Automated Pre-Screening: Reduces manual efforts in evaluating loan applications by automating the initial screening process.
* Consistency & Accuracy: Ensures standardized decision-making by eliminating human biases in loan approvals.
* Customizable Criteria: Lenders can set dynamic approval thresholds based on their policies, such as minimum credit score or maximum debt-to-income ratio.

3. Data Analysts & Researchers

* Financial Risk Analysis: Provides insights into risk assessment and trends in loan approvals.
* Machine Learning Model Optimization: Researchers can analyze how different ML algorithms perform in predicting loan approvals.
* Feature Importance Analysis: Helps in understanding the most influential factors affecting loan approval decisions.

4. Government & Regulatory Bodies

* Financial Inclusion Monitoring: Helps in assessing whether loan approvals are fair and unbiased across different demographics.
* Regulatory Compliance: Ensures adherence to lending regulations by implementing transparent and explainable AI models.

By benefiting these stakeholders, the system enhances efficiency, transparency, and fairness in loan approval processes, ultimately improving the overall lending ecosystem.

1.3 Project Scope

* Develop a web-based system for predicting loan approval.
* Provide an intuitive interface where users enter financial details.
* Implement machine learning models such as Decision Tree and Random Forest.
* Allow users to configure dynamic thresholds for key loan parameters.
* Display feature importance to help users understand approval factors.

1.4 References

* KaggleDataset:<https://www.kaggle.com/code/chonlitasai/loan-approval-xgboost-95-accuracy/input>
* Scikit-learn Documentation: <https://scikit-learn.org/>

## 2. PROJECT DESCRIPTION

2.1 Reference Algorithm

The Loan Approval Prediction System utilizes machine learning algorithms to evaluate loan applications based on financial attributes. The primary algorithms used are:

1. Decision Tree Classifier

* A tree-based model that splits data into branches based on conditions.
* Provides an interpretable and easy-to-understand decision-making process.
* Useful for categorizing loan applications into Approved, Conditionally Approved, or Rejected.

2. Random Forest Classifier

* An ensemble learning method that combines multiple decision trees.
* Reduces overfitting and improves prediction accuracy.
* Suitable for handling complex financial datasets.

3. Logistic Regression (for Benchmarking)

* A simple statistical model used for performance comparison.
* Helps determine if more complex models (like Decision Tree/Random Forest) improve accuracy.

2.2 Data/Data Structure

* Input Features:
  + Annual Income (in USD)
  + Credit Score
  + Loan Amount
  + Employment Status (Full-time, Part-time, Self-employed, Unemployed)
  + Loan Duration (in years)
  + Debt-to-Income Ratio
  + Past Defaults
  + Age
  + Education Level
* **Output Categories:**
  + Approved
  + Approved with Conditions (e.g., smaller loan, higher interest rate)
  + Rejected
* **Data Format:** CSV/JSON

2.3 SWOT Analysis

Strengths:

* Automates decision-making, reducing manual effort.
* Provides quick and accurate predictions.
* Scalable and adaptable to different financial institutions.

Weaknesses:

* Accuracy depends on data quality and feature selection.
* Potential biases in training data affecting fairness.

Opportunities:

* Can be extended to other financial products (e.g., credit cards, mortgages).
* Integration with real-time banking APIs for better assessments.

Threats:

* Risk of data breaches due to financial data sensitivity.
* Regulatory constraints around financial decision automation.

2.4 Project Features

1. User Input Form: A web-based interface to collect user financial details.
2. Prediction Engine: Uses an ML model to analyze inputs and generate results.
3. Dynamic Thresholds: Admins can set limits for credit scores, loan amounts, etc.
4. Feature Importance Display: Shows how each input affects the decision.
5. Visualization: Graphical and textual representation of the decision tree.

2.5 User Classes and Characteristics

* Loan Applicants: Users with minimal technical knowledge who input their details.
* Admin Users: Financial institutions managing threshold settings.

2.6 Design and Implementation Constraints

* Data Quality: The system depends on high-quality, diverse training data.
* Regulatory Compliance: The system must adhere to financial regulations and privacy laws.
* Performance: The model should generate predictions in under 2 seconds.

2.7 Assumption and Dependencies

* Users will provide accurate financial information.
* The model will be updated periodically to improve accuracy.
* The web application will be hosted on a cloud platform.

## 3. SYSTEM REQUIREMENTS

3.1 User Interface

* Web-based Interface: Users will interact with the system through a web application.
* Form-Based Input: Users will enter financial details like income, credit score, loan amount, etc.
* Result Display: The system will show loan approval results, including reasons for approval/rejection.
* Interactive Visuals: Feature importance and decision tree visualization will be displayed graphically.
* Customizable Settings: Admins can modify loan approval thresholds dynamically.
* Responsive Design: The system will be accessible via desktops, tablets, and mobile devices.

3.2 Software Interface

* Frontend: Developed using HTML, CSS, JavaScript (React.js or Flask for lightweight UI).
* Backend: Python-based (Flask/Django) to handle data processing and model predictions.
* Machine Learning Library: Scikit-learn for Decision Tree, Random Forest, and Logistic Regression models.
* Visualization Tools: Matplotlib, Seaborn for data visualization.
* Database: PostgreSQL or MySQL for storing user data and application logs.

3.3 Database Interface

* User Data Table: Stores applicant details and loan application history.
* Model Training Data: Stores historical loan approval data for retraining the model.
* Admin Configuration Table: Stores dynamic threshold settings (minimum credit score, max loan amount, etc.).
* Logging and Analytics Table: Stores system logs, user queries, and prediction outcomes.

3.4 Protocols

* HTTP/HTTPS: Used for web-based communication between frontend and backend.
* REST API: Used to send user input data to the backend and retrieve prediction results.
* Database Queries: SQL-based queries to fetch and update loan prediction data.
* Authentication Protocols: OAuth2 or JWT for secure user authentication.

## 4. NON-FUNCTIONAL REQUIREMENTS

### 4.1 Performance Requirements

### Response Time: The system should process and return predictions within 2 seconds for optimal user experience.

### Scalability: The system should handle at least 1,000 concurrent users without significant performance degradation.

### Throughput: The model should be able to process at least 100 loan applications per minute.

### Model Efficiency: The machine learning model should be optimized for inference speed while maintaining accuracy.

4.2 Security Requirements

* Data Encryption: All user inputs and stored data should be encrypted using AES-256 encryption.
* Authentication & Authorization: Users must log in using secure authentication mechanisms (JWT, OAuth).
* Data Privacy: The system should comply with GDPR & other financial data protection regulations.
* Logging & Monitoring: Track unauthorized access attempts and system errors for security audits.

4.3 Software Quality Attributes

* Usability: The web application should have an intuitive design, easy navigation, and accessibility compliance (WCAG 2.1).
* Maintainability: The codebase should follow modular design principles for easy updates and debugging.
* Reliability: The system should have 99.9% uptime, with proper failover mechanisms in place.
* Portability: The system should be deployable on cloud-based servers (AWS, GCP) or local hosting environments.

5. Other Requirements

5.1 Legal and Compliance Requirements

* Ensure compliance with GDPR and CCPA for data privacy.
* Maintain audit logs for transparency in decision-making.

5.2 Scalability and Maintenance

* Support increasing user traffic.
* Allow future model enhancements.

5.3 Disaster Recovery and Backup

* Regular database backups to prevent data loss.
* Recovery mechanisms for unexpected failures.

Appendix A: Glossary

* ML (Machine Learning): Predictive analytics method.
* API (Application Programming Interface): Enables system communication.
* Feature Importance: Determines which variables impact loan approval most.

Appendix B: Analysis Model

* Analyzes historical loan data to determine approval trends.
* Uses feature importance analysis to rank influencing factors.
* Implements cross-validation for reliable model performance.

Appendix C: Issues List

* Data Bias: Ensuring fair training data for unbiased predictions.
* User Experience: Improving UI for better accessibility.
* Performance Optimization: Reducing model execution time.
* Regulatory Compliance: Adapting to financial regulations.