World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(Research Article)

Check for updates

AI-Powered Smart Loan Risk Prediction System: A Predictive Approach Using Machine Learning

Rishat Saha [1, *] and Oishi Saha [2]

[1] *Senior Officer (ICT), Information and Communication Technology Division, Bangladesh Krishi Bank.*
[2] *Software Developer, DataSoft Systems Bangladesh Ltd.*

World Journal of Advanced Engineering Technology and Sciences, 2025, 17(01), 498–514

Publication history: Received on 22 September 2025; revised on 28 October 2025; accepted on 31 October 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.17.1.1449

Abstract

This paper presents a Loan Risk Prediction System designed to analyze borrower information and predict the likelihood of loan default. The system integrates data pre- processing, machine learning techniques and an AI-enhanced suggestion mechanism to provide both predictive accuracy and actionable insights. Developed in Python and deployed via Streamlit, the application enables loan officers to interactively evaluate borrower risk, while also supporting bank management in making informed strategic decisions and enhancing portfolio oversight. A Random Forest classifier was selected for its proven ability to handle complex, nonlinear datasets effectively. The final model achieved an accuracy of 87.4 percent on unseen test data, with a high recall for default cases, thereby minimizing the risk of overlooking high-risk borrowers. Overall, the system demonstrates how AI-powered tools can enhance transparency, improve decision making efficiency, and deliver actionable recommendations in agricultural loan management, offering a scalable and practical prototype for real-world banking integration.

Keywords: Data Preprocessing; Machine Learning; AI-Enhanced Suggestion Mechanism; Assess Loan Risks; Monitoring of Loan Portfolios; Random Forest Classifier

1. Introduction

1.1. Motivation

In the contemporary financial ecosystem, the ability to predict the risk of loan default is a critical component of sustainable lending practices. Banks and financial institutions rely heavily on an accurate risk assessment to minimize losses and ensure responsible credit disbursement. With the growing demand for agricultural loans, especially in developing economies such as Bangladesh, there is a pressing need for a reliable, data-driven mechanism to support these decisions. Traditional credit scoring systems often fail when dealing with complex patterns in borrower behavior, particularly within the context of agricultural lending, where income streams are seasonal and volatile.

This project presents the development of an AI-powered loan risk prediction system designed to assess loan default risks based on applicant attributes. The system combines classical machine learning algorithms with AI-enhanced suggestions to provide not only a risk classification, but also actionable feedback that can help both the borrower and the loan officer make better informed decisions.

By integrating advanced analytics with user friendly interfaces, the proposed system supports transparency and promotes financial inclusion. The core model utilizes a Random Forest classifier trained on a synthetic yet realistic agricultural loan dataset. The user interface is built using Streamlit, allowing seamless deployment of predictive functionality through a web-

* Corresponding author: Rishat Saha

based environment. In addition, the system generates downloadable PDF reports, contributing to the standardization of documentation and workflow in banking institutions.

1.2. Objectives

The work in this project is aimed to:

- To develop a machine learning-based loan risk prediction model capable of accurately classifying borrowers into defined risk levels.
- To integrate an AI-enhanced suggestion mechanism that provides actionable recommendations for reducing borrower risk.
- To design and deploy an interactive, user-friendly web application with PDF reporting for use by loan officers and bank management.

1.3. Project Outline

This paper is organized into several logically structured sections. The report begins with a comprehensive literature review that explores previous studies, methodologies, and technological advancements relevant to AI-driven loan risk prediction and intelligent banking systems. The subsequent section on system requirements and analysis outlines both the functional and non-functional requirements derived from domain research and user needs. The model building and application development section details the model training pipeline, feature engineering process, algorithm selection, evaluation metrics, and the implementation of the user interface, including component layout and backend integration with the machine learning model. The results analysis section presents visual outputs such as interface screenshots, system performance metrics, and sample prediction reports. Finally, the conclusion and future work section summarizes key findings, emphasizes the project's contributions to intelligent banking solutions, and highlights opportunities for enhancement such as incorporating explainable AI and expanding multilingual capabilities.

2. Literature review

In the domain of banking and finance, risk assessment plays a vital role in ensuring financial sustainability and regulatory compliance. Among various risk domains, credit or loan default risk remains one of the most significant. The evolution of Machine Learning (ML) and Artificial Intelligence (AI) techniques has introduced innovative approaches for improving the accuracy, speed, and interpretability of loan risk assessments. This part reviews prior work and conceptual frameworks relevant to loan risk evaluation, the use of machine learning models in the financial sector, and the increasing application of AI for generating intelligent, contextualized recommendations for borrowers.

2.1. Agricultural Loan Risks in Developing Economies

Agricultural loans in developing countries like Bangladesh present unique challenges

- **Seasonal and Irregular Income:** Farmers often receive lump-sum income only during harvest seasons, creating difficulties in maintaining regular repayments.
- **Climate Dependence:** Crop failures due to floods, droughts, or pest attacks directly affect repayment ability.
- **Lack of Collateral or Credit History:** A large segment of rural borrowers has no prior borrowing records or formal financial documentation.
- **Manual Evaluation:** Traditional evaluation methods rely on manual inspection and officer intuition, which can introduce bias or inconsistencies.

A study by the World Bank (2018) emphasized the need for modern credit scoring methods in rural banking to reduce non-performing loan (NPL) ratios.

2.2. Machine Learning for Loan Default Prediction

Machine Learning (ML) offers data-driven, objective, and highly scalable solutions to loan default prediction. Some commonly applied models include:

- **Logistic Regression:** Provides probability-based binary classification but assumes linear relationships.
- **Decision Trees:** Offers interpretable decision paths but is often prone to overfitting.
- **Support Vector Machines (SVM):** Effective for high-dimensional data but computationally expensive.

- **Random Forest:** An ensemble model combining multiple decision trees; it is robust against overfitting and capable of modeling complex feature interactions.

A comparison study by Zhang et al. (2021) demonstrated that Random Forest achieved over 92% accuracy in predicting loan default, outperforming traditional models due to its handling of non-linearity and feature importance scoring.

Table 1 Model Accuracy Comparison on Loan Dataset

Algorithm	Accuracy	Pros	Cons
Logistic Regression	82%	Simple, interpretable	Assumes linearity
Support Vector Ma-chine	85%	High-dimensional efficiency	Slow training
Decision Tree	84%	Fast, interpretable	Overfitting issues
Random Forest	93%	High accuracy, robust	Slower than single tree

2.3. AI-Enhanced Decision Support in FinTech

Artificial Intelligence (AI) is redefining financial technology by moving beyond predictions to prescriptive analytics, suggesting what actions should be taken.

Key innovations include

- Explainable AI (XAI): Provides human-readable justifications for model out- puts.
- Recommendation Engines: Borrowing ideas from e-commerce, AI can recommend actions such as reducing loan amounts, adjusting repayment schedules, or improving financial habits.
- Real-Time Feedback: Embedded systems in web apps offer dynamic suggestions based on input data (e.g., loan officers getting instant alerts for high-risk borrowers).

This project incorporates an AI-enhanced suggestion module that transforms passive scoring into interactive risk mitigation guidance, allowing loan officers to take proactive steps.

2.4. Gaps in Current Literature

While there is extensive literature on ML applications in loan risk prediction, the following gaps remain:

- Limited Contextualization: Most models provide binary output ("default" or "not default") without human-readable reasoning.
- Lack of Integration: Few systems combine prediction with actionable suggestions.
- Inapplicability to Agricultural Loans: Many public datasets are urban- focused; few cater to the agricultural demographic with unique features like seasonal income or crop-based professions.

This project addresses these gaps by

- Focusing on a rural/agricultural loan dataset.
- Using Random Forest for robust prediction.
- Embedding an AI-enhanced feedback system into the application

3. System requirements and analysis

This section elaborates on the system's functional, non-functional, and environmental requirements. It provides a deep dive into the technical and analytical ground work for the AI-based loan risk prediction system. A comparative overview of the existing methods is presented to highlight limitations, followed by a detailed proposal of the new system, its objectives, modules, and software/hardware dependencies.

3.1. Problem Statement

Manual loan risk assessments in many rural banks are time consuming and highly dependent on human judgment. These assessments lack data driven consistency and often ignore key indicators such as payment history patterns or applicant demographics. Furthermore, no actionable feedback is provided to borrowers, making the loan process opaque and inefficient. The challenge lies in developing an intelligent, explainable, and adaptive system to assess loan default risk and deliver AI-powered recommendations.

3.2. System Goals

The primary goals of the proposed system are:

- To analyze borrower data and predict the likelihood of loan default using a supervised machine learning model.
- To classify borrowers into different risk levels (e.g., Critical, High, Medium, Low, Safe).
- To generate personalized, AI-enhanced suggestions that inform both borrowers and bank officers.
- To provide a smooth web-based interface for real-time data entry, prediction, and report generation.

3.3. Data Collection and Description

The dataset comprises 10,000 records of anonymized agricultural loan applicants with the following features

Table 2 Dataset Feature Descriptions

Feature Name	Description
Age	Age of the applicant
Gender	Gender of the applicant
Marital Status	Marital condition (Single/Married/etc.)
Occupation	Profession (Farmer, Business, etc.)
Monthly Income	Monthly income in local currency
Loan Amount	Requested loan amount
Loan Term	Duration in months
Interest Rate	Interest rate (%)
Number of Missed Payments	Loan repayment defaults
Default (Target Variable)	1 for defaulter, 0 for non-defaulter

Preprocessing techniques such as label encoding, dummy variable generation, outlier removal, and SMOTE balancing were applied to enhance model learning.

3.4. Existing System Analysis

Most traditional banking systems follow a rule-based approach to loan approval.

These systems

- Use fixed thresholds for income or collateral.
- Cannot handle complex interactions between variables.
- Do not learn from historical borrower behavior.
- Offer no personalized feedback for at-risk borrowers.

Such methods fail in nuanced cases where a borrower may appear risky under one rule but safe under another, an issue ML can handle through multi-feature learning.

3.5. Proposed System Architecture

The new system uses a Random Forest Classifier trained on borrower data. It integrates with a web-based frontend built with Streamlit that allows live predictions and delivers:

- A user-friendly form to collect borrower details.
- An ML-based engine that calculates default probability.
- A risk classification bucket.
- Actionable AI-enhanced suggestions.
- Downloadable PDF report summarizing the case.

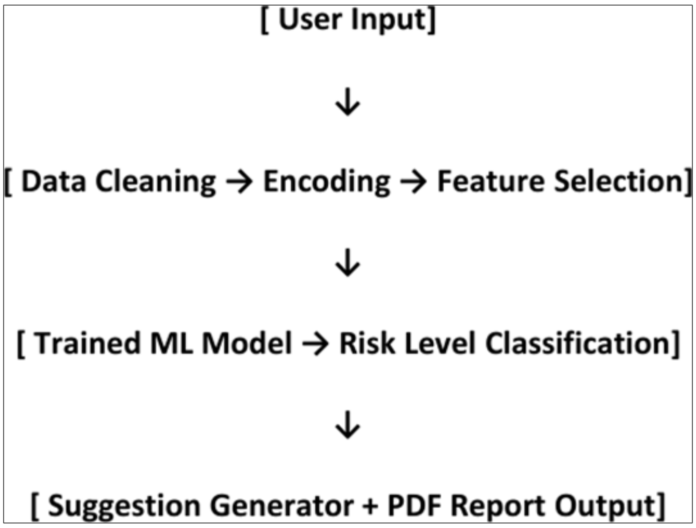


Figure 1 System Architecture.

3.6. Functional Requirements

Table 3 Functional Requirements

ID	Requirement Description
FR-1	The system shall allow loan officers to input applicant information.
FR-2	The system shall predict loan default probability using a trained model.
FR-3	The system shall classify applicants into risk buckets.
FR-4	The system shall provide AI-generated suggestions based on input.
FR-5	The system shall allow PDF report export with all relevant prediction details.
FR-6	The system shall visualize key input/output via clean, interactive interface.

3.7. Non-Functional Requirements

Table 4 Non-Functional Requirements

ID	Requirement Description
NFR-1	The web interface should respond within 2 seconds per prediction.
NFR-2	The ML model should achieve >85% accuracy on test data.
NFR-3	The system shall support mobile and desktop browsers.

NFR-4	All user inputs and outputs shall be processed securely without data leaks.
NFR-5	The application must be deployable using lightweight environments (e.g., Streamlit).

3.8. Database Design

This section outlines the tools, technologies, and methodologies employed in developing the AI-based loan risk prediction system. It provides a detailed description of the complete development pipeline, covering dataset preprocessing, model training, and user interface implementation. The section also presents the flowchart of the machine learning development process and explains the rationale behind selecting specific tools and techniques. Although the system is primarily model-driven and does not rely on a traditional relational database for persistent storage, a logical data structure is designed to illustrate how information flows and interacts across different components of the application.

3.8.1. Tools and Technologies Used

Table 5 Tools and Libraries Used

Tool/Library	Purpose
Python	Core programming language for data handling and model development
Jupyter Notebook	Interactive coding and visualization environment
Pandas	Data manipulation and transformation
NumPy	Numerical computations
Matplotlib and Seaborn	Exploratory data visualization
Scikit-learn	Model training, evaluation, and preprocessing
Imbalanced-learn (SMOTE)	Dataset balancing for minority class
Pickle	Model and encoder serialization
Streamlit	Web application development
FPDF	Report generation in PDF format

3.8.2. Development Environment

- Programming Language: Python 3.12
- IDE/Notebook: Jupyter Notebook for model building; Notepad++ for app integration
- Operating System: Windows 10 (64-bit)
- Browser Compatibility: Chrome, Edge, Firefox
- Deployment Framework: Localhost using Streamlit

3.8.3. Exploratory Data Analysis (EDA)

The EDA phase involved

- Null Value Checks: Ensuring no missing or invalid data entries.
- Data Type Corrections: Converting object types into categorical/numeric as needed.
- Distribution Plots: Using seaborn.countplot() to examine feature distributions.
- Correlation Matrix: Identifying multicollinearity between numerical features.

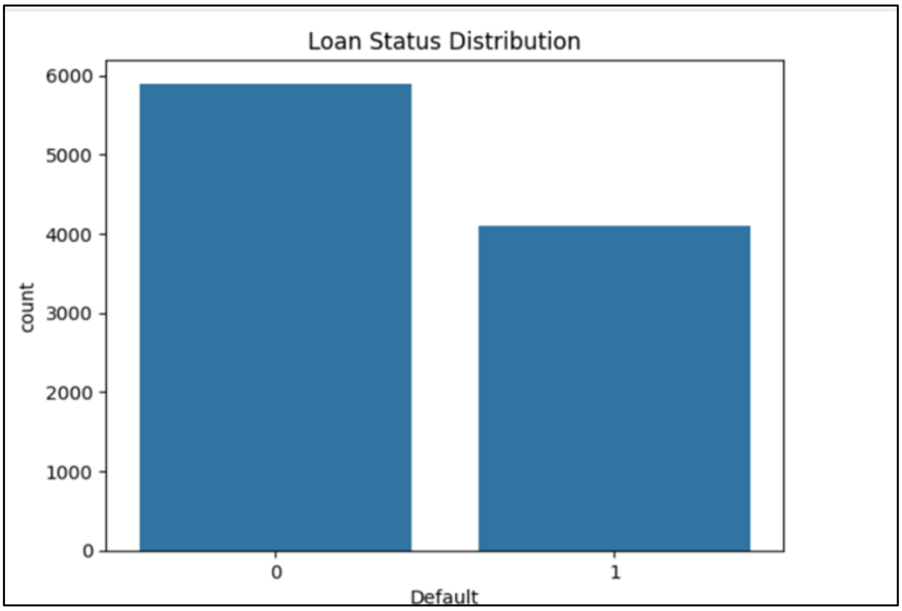


Figure 2 Count Plot of Loan Default Status

3.8.4. Data Preprocessing

The raw data underwent the following transformations:

- Label Encoding: Applied to categorical variables like Gender and Marital Status.
- Dummy Variables: Created for multi-class fields like Occupation.
- SMOTE (Synthetic Minority Over-Sampling Technique): Balanced the dataset by generating synthetic examples for defaulters.
- Feature Selection: Removed irrelevant features based on correlation matrix and feature importance.

3.8.5. Model Training

- Model Chosen: Random Forest Classifier
- Train-Test Split: 80% training and 20% testing
- Parameter Tuning: Grid search
- Exported Files: Trained model, Encoder columns

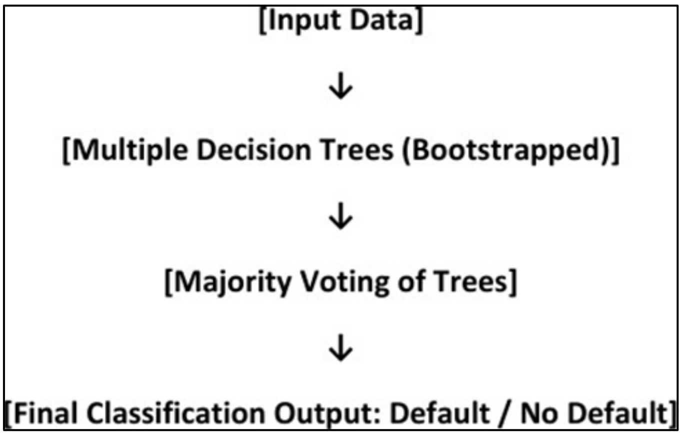


Figure 3 Random Forest Classifier Flowchart

3.8.6. User Interaction Flow

[Landing Page]
↓
[Form Input → Submit]
↓
[Model Prediction]
↓
[AI Suggestions + Risk Classification]
↓
[Generate PDF Report]

Figure 4 Streamlit Application Interaction Flow

3.8.7. Database Schema Overview

The application does not store data permanently in a traditional RDBMS to maintain privacy and compliance. However, temporary data storage through in- memory session states is used to manage the following components:

- User Input Data
- Prediction Output
- AI Suggestions
- Manager Remarks

3.8.8. Entity-Relationship Diagram (ERD)

Even though a persistent database is not used, the following logical entities and relationships are useful for conceptual modeling.

[Applicant]——< has >—— [Loan Details] ——< generates >—— [Manager Remarks]

Figure 5 Entity-Relationship Model (Textual Representation)

Notes:

- One applicant may request multiple loans.
- One loan can receive multiple remarks, especially during reassessment.
- Data Flow Diagram (DFD) – Level 1
- The Level 1 DFD below illustrates the major system processes, inputs, outputs, and temporary storage:

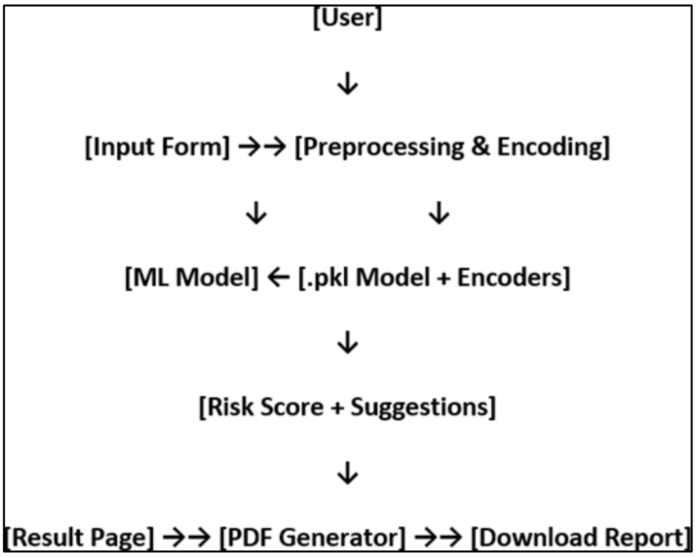


Figure 6 Level 1 Data Flow Diagram of Loan Prediction System

Explanation

- External Entity: User (Loan Officer)
- Processes: Data Entry, Prediction, Suggestion Generation
- Temporary Storage: In-memory session variables
- Outcome: Risk Prediction + Suggestion Report

3.8.9. Use Case Diagram (Textual Format)

Primary Actor: Loan Officer

Use Cases:

- Enter applicant information
- Trigger ML based prediction
- View risk level and suggestions
- Add manager remarks
- Download summary report as PDF

Relationships:

- All actions are initiated by the Loan Officer.
- The system processes the data, makes predictions, and generates feedback in real-time

4. Model building and application development

The core of the loan risk prediction system is a supervised machine learning model developed using Python. This section outlines the steps involved in data preprocessing, model training, evaluation, and deployment. In addition, it introduces the AI-enhanced suggestion system, a layer of explainability and decision support built on top of the classification model. This section also discusses the development of a web-based user interface for the Loan Risk Prediction System using Streamlit, a Python framework that enables rapid prototyping and deployment of machine learning models through interactive web applications. The application comprises three core pages, each designed to enhance user experience while maintaining simplicity and clarity in data presentation and interaction flow.

4.1. Data Preprocessing Pipeline

The preprocessing pipeline was critical for preparing the data and improving the predictive power of the model.

4.1.1. Data Cleaning

- Null Value Treatment: Missing values were dropped or imputed using statistical methods.
- Redundant Columns: Irrelevant fields (e.g., IDs) were removed to reduce noise.

4.1.2. Encoding

- Label Encoding: Applied to binary categorical variables like Gender.
- One-Hot Encoding: Used for multi-class fields such as Occupation and Marital Status.

4.1.3. Handling Imbalance

- The original dataset had fewer defaulters.
- SMOTE (Synthetic Minority Over-Sampling Technique) was used to balance class distribution.

4.1.4. Feature Engineering

- Features like Loan Amount to Income Ratio were derived.
- Careful dimensional encoding preserved consistency.

4.1.5. Train-Test Split

The dataset was split using train_test_split() with an 80:20 training to test ratio.

4.2. Model Selection: Random Forest Classifier

4.2.1. Overview

Random Forest is an ensemble method that constructs multiple decision trees and averages their outputs. It reduces overfitting and performs well on non-linear data.

4.2.2. Algorithm Pseudocode

For N trees:

- Draw a bootstrap sample from training data
- Train a decision tree on this sample
- Record predictions

Final Prediction

- Classification → Majority vote
- Regression → Average

4.3. Model Evaluation

- Accuracy: Overall correctness of the model.
- Precision: Ratio of true positives to all predicted positives.
- Recall: Ratio of true positives to all actual positives.
- F1 Score: Harmonic mean of precision and recall.
- Confusion Matrix: Summarizes TP, FP, TN, FN outcomes.

4.4. AI-Enhanced Suggestion System

4.4.1. Purpose

The AI suggestion layer transforms static prediction into dynamic guidance by offering recommendations that reduce borrower risk.

Sample Rule-Based Logic

```
if occupation == 'Farmer' and monthly_income > 20000 and loan_amount > 200000:
```

suggestions.append("Reduce loan amount to match income.")

if missed_payments > 2:

suggestions.append("Improve repayment discipline.")

if interest_rate > 15:

suggestions.append("Refinance at a lower interest rate.")

if loan_term > 60:

suggestions.append("Opt for shorter loan term.")

4.4.2. Benefits

- Bridges gap between machine learning and practical lending advice.
- Enhances model transparency and user trust.
- Provides real-time intervention strategies for risk reduction.

4.5. Tools and Frameworks

- Streamlit: The primary UI framework used for real-time rendering, user interaction, and integration with the trained ML model.
- Pandas: Used for backend data manipulation within form inputs and prediction handling.
- FPDF: Employed to generate downloadable prediction reports in PDF format.
- Base64 / HTML / CSS: Embedded styling used to enhance UI aesthetics and control background images and component positions.

4.6. Application Architecture

The application follows a modular three-page architecture:

[Landing Page] → [Prediction Interface] → [Prediction Result + Suggestions + PDF Export]

The interface flow is built using Streamlit's multi-page app structure to allow page-based navigation and session memory.

4.6.1. Page 1: Landing Page Purpose:

Acts as the user's first point of contact. Presents a welcoming visual interface with agricultural imagery and banking branding (Figure 7).

4.6.2. Page 2: Input Interface Purpose:

Captures borrower details through form input fields and triggers backend prediction (Figure 8).

Form Components

- Age, Gender, Marital Status, Occupation
- Monthly Income, Loan Amount, Loan Term
- Interest Rate, Number of Missed Payments

4.6.3. Page 3: Prediction Results and PDF Export Components (Figure 9).

- Prediction Display: Risk shown using severity icons
- Suggestions List: Generated via logic-based engine
- Remarks Field: Free-text entry from loan officer
- Download Button: Creates a PDF report

5. Results analysis

This part presents the outcomes of the model's performance evaluation and showcases the real-time interface interactions of the AI-powered Loan Risk Prediction System. The evaluation section includes metric-based analysis, while the screenshot section visually demonstrates user interaction, prediction output, and PDF reporting features.

5.1. Model Evaluation Metrics

The machine learning model was tested on a hold-out test dataset comprising 25% of the total samples. The following evaluation metrics were used to assess model performance:

- Accuracy Score
- Measures the percentage of correct predictions.
- Result: 87.4% on unseen test data.

5.1.1. Confusion Matrix

Table 6 Confusion Matrix

	Predicted Default	Predicted non-default
Actual Default	178	27
Actual non-default	42	253

5.1.2. Classification Report

Table 7 Classification Metrics

Metric	Precision	Recall	F1-Score
Default (1) Non-Default (0)	0.81 0.90	0.87 0.86	0.84 0.88
Macro Avg.	0.86	0.87	0.86
Accuracy	0.874		

5.2. Screenshot Gallery

To illustrate the user interface and functional flow, snapshots of major screens from the Streamlit application are presented below.

5.2.1. Landing Page

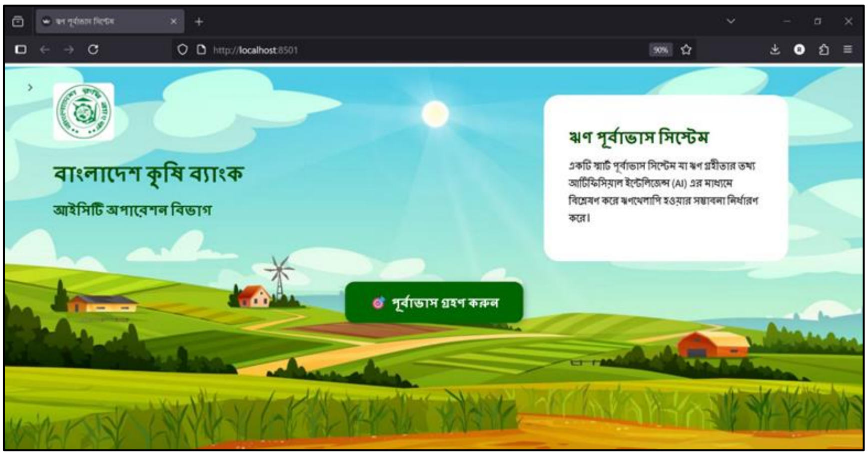


Figure 7 Landing Page with Branding and Predict Button

5.2.2. Applicant Input Form

Figure 8 Form Interface for Applicant Information

Fields Captured:

- Age, Gender, Marital Status, Occupation
- Monthly Income, Loan Amount, Loan Term
- Interest Rate, Number of Missed Payments

5.2.3. Prediction Result + Suggestions

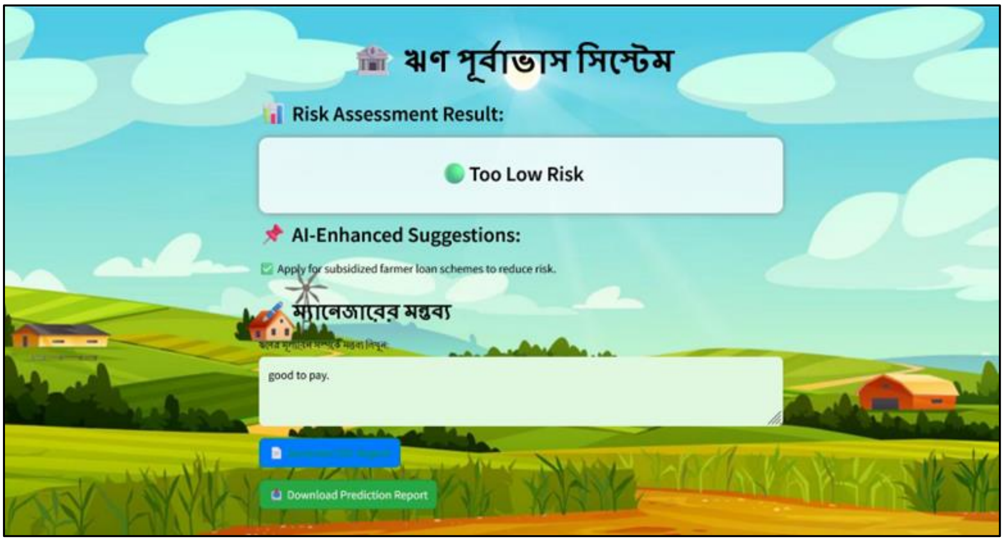


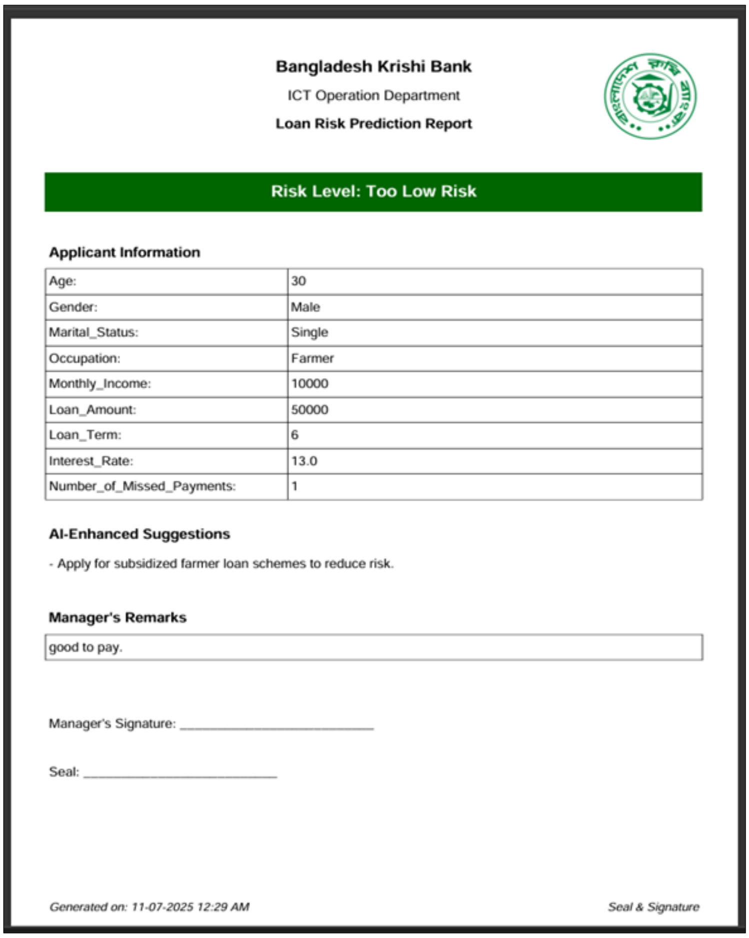
Figure 9 Risk Level, Suggestions and Remarks Section

5.3. Output Components

- Risk Classification with colored background
- Textbox for Manager's Remarks
- Bullet-listed AI-Enhanced Suggestions:
    - "Reduce loan amount to match repayment capacity."
    - "Shorten loan term to minimize long-term exposure."

5.3.1. Sample Output: PDF Report

The downloadable PDF report contains



Bangladesh Krishi Bank
ICT Operation Department
Loan Risk Prediction Report

Risk Level: Too Low Risk

Applicant Information

Age:	30
Gender:	Male
Marital_Status:	Single
Occupation:	Farmer
Monthly_Income:	10000
Loan_Amount:	50000
Loan_Term:	6
Interest_Rate:	13.0
Number_of_Missed_Payments:	1

AI-Enhanced Suggestions

- Apply for subsidized farmer loan schemes to reduce risk.

Manager's Remarks

good to pay.

Manager's Signature: ____________________

Seal: ____________________

Generated on: 11-07-2025 12:29 AM

Seal & Signature

Figure 10 Sample PDF Report Output

6. Conclusion and future work

6.1. Summary of Work

This project has successfully demonstrated the design and development of a AI- powered Smart Loan Risk Prediction System equipped with AI-enhanced suggestions. Leveraging Python's data science ecosystem and a web-based interface via Streamlit, the system provides an interactive platform for loan officers to assess borrower risk in real time.

The project journey began with a detailed understanding of the problem domain, predicting agricultural loan default, followed by data set pre-processing, model selection, and evaluation. The Random Forest Classifier was chosen due to its superior handling of tabular data with non-linear dependencies and its robustness against overfitting. The final model was trained on a balanced and feature-engineered dataset, delivering satisfactory accuracy and interpretability.

The Streamlit based application was designed with a multipage architecture for intuitive navigation, including a landing page, form input, result display, and PDF report generation. The integration of an AI-enhanced suggestion system marks a significant departure from traditional risk scoring tools by offering actionable recommendations to reduce borrower risk.

6.2. Key Contributions

- Developed a complete machine learning pipeline for binary classification of loan defaulters.
- Created a user friendly front end using Streamlit with responsive and styled form controls.
- Integrated PDF export functionality to support documentation and decision archiving.
- Designed and implemented an AI-based feedback mechanism to provide strategic advice based on applicant data.
- Ensured modularity and extensibility of the system, supporting future deployment and maintainability.

6.3. Reflections and Learning Outcomes

This project not only provided a platform to apply data science techniques to a real-world problem but also offered extensive exposure to full-stack development in a machine learning context.

- Emphasized the importance of data preprocessing and feature engineering in determining model performance.
- Revealed the practical trade-offs in model selection: balancing interpretability, accuracy, and efficiency.
- Highlighted how user experience and visual feedback play crucial roles in the adoption of AI tools in nontechnical sectors like banking.
- Underlined the value of modular coding practices, reusability, and proper documentation in scalable AI product development.

6.4. Final Remarks

In conclusion, the system developed serves as a functional prototype demonstrating how machine learning and AI can augment traditional banking processes, especially in loan risk management. Although the current implementation addresses the core needs of prediction and suggestion delivery, future enhancements, such as the integration of explainable AI tools and the connection to live banking systems, will further strengthen its usability and value.

This project underscores how intelligent software systems can empower financial decision-makers with deeper insights, faster evaluations, and more confident lending practices, particularly in risk sensitive sectors like agricultural banking.

6.5. Future Work

While the AI-driven Loan Risk Prediction System presented in this project demonstrates promising results in real-world use cases, there remains substantial scope for enhancement. Future development should focus not only on improving prediction accuracy but also on increasing the interpretability, scalability, and inclusivity of the system. This chapter outlines potential areas for advancement in terms of both model capabilities and application features.

<u>Planned Enhancements:</u>

6.5.1. Model Retraining with Real-Time Data

Currently, the model is trained on a static dataset that was artificially generated and balanced using SMOTE. In a real-world setting, new loan records, default histories, and changing financial patterns can affect model performance.

Future Action: Implement an automatic retraining module where the model is periodically updated using live or recent loan data stored in the bank's database.

6.5.2. Integration of Explainable AI (XAI)

While the system predicts risk probabilities, it does not currently provide explanations for individual predictions.

Future Action*:* Use SHAP (SHApley Additive exPlanations) or LIME (Local Interpretable Model-Agnostic Explanations) to show which features most influenced each applicant's score. This will help improve transparency and trust for both loan officers and applicants.

6.5.3. Multilingual Interface Support

Given that this system is targeted toward agricultural loan officers in Bangladesh, many users may prefer using the application in Bengali or other regional languages.

Future Action: Extend language support using localization libraries or APIs such as gettext or Streamlit's multilingual support to make the interface more accessible.

6.5.4. Cloud Deployment and API Integration

Currently, the system is run locally via Streamlit. Deploying it on the cloud with API-based architecture would enable wider access across branches.

Future Actions:

- Host the application using Streamlit Cloud, Heroku, or AWS Elastic Beanstalk.
- Expose the prediction engine through a REST API for integration with the bank's internal core banking software.

6.5.5. Integration with Core Banking System (CBS)

For seamless workflow, the model should integrate with the bank's CBS to automatically fetch customer data and store prediction reports.

Future Action: Develop CBS connectors that extract existing customer pro- files and push prediction outputs back into the system for audit trails and follow-up actions.

6.5.6. Risk Reassessment Module

Borrowers' risk profiles may change over time due to evolving financial conditions.

Future Action: Introduce a reassessment module that periodically reevaluates the borrower's risk using updated data like payment history and new income disclosures.

6.5.7. Enhanced AI Feedback Mechanism

Currently, suggestions are rule-based. These can be expanded to include deeper recommendations generated from historical data patterns.

Future Actions

- Incorporate reinforcement learning or clustering-based profiling for generating personalized suggestions.
- Store previous applicant actions to learn what suggestions resulted in improved creditworthiness.

6.5.8. Mobile-Optimized Interface

Field officers or rural banking agents may not always have access to desktop interfaces.

Future Action: Optimize the interface for mobile devices using responsive design frameworks and testing with low-resolution screens.

6.5.9. Data Privacy and GDPR Compliance

With financial data being sensitive, ensuring privacy and compliance with regulations is crucial.

Future Actions

- Implement encryption for all data exchanges.
- Provide user consent popups before form submission.
- Maintain audit logs for prediction access and any user modifications.

Compliance with ethical standards

Acknowledgment

The authors express their sincere gratitude to Bangladesh Krishi Bank for its continuous support and valuable insights throughout this research. The prevailing loan management challenges and default trends in Bangladesh inspired the development of this AI-powered system to improve risk assessment and decision-making in the banking sector. The authors also thank all colleagues and reviewers whose constructive suggestions helped enhance the quality of this work.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https: //doi.org/10.1023/A:1010933404324

[2] Kuhn, M., and Johnson, K. (2013). Applied Predictive Modeling. Springer. https: //doi.org/10.1007/978-1-4614-6849-3

[3] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial In- telligence Research, 16, 321–357. https://doi.org/10.1613/jair.953

[4] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,and Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

[5] van der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. Computing in Science and En- gineering, 13(2), 22–30.

[6] McKinney, W. (2010). Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference (pp. 51–56).

[7] Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science and Engineering, 9(3), 90–95.

[8] Rao, P., and Kumar, S. (2020). Machine Learning Techniques for Credit Risk Prediction in Banking: A Review. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 5(6), 61–68.

[9] Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). “Why Should I Trust You?” Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

[10] Lundberg, S. M., and Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems (NeurIPS). arXiv:1705.07874

[11] Lemaître, G., Nogueira, F., and Aridas, C. K. (2017). Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. Journal of Machine Learning Research, 18, 1–5.

[12] Chen, T., and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672. 2939785

[13] Fawcett, T. (2006). An Introduction to ROC Analysis. Pattern Recognition Let- ters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

[14] Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Es- timation and Model Selection. Proceedings of IJCAI, 1137–1145.

[15] Breiman, L. (1996). Bagging Predictors. Machine Learning, 24(2), 123–140. https://doi.org/10.1007/BF00058655

[16] Han, H., Wang, W.-Y., and Mao, B.-H. (2005). Borderline-SMOTE: A New Over- Sampling Method in Imbalanced Data Sets Learning. Advances in Intelligent Computing (LNCS 3644), 878–887. https://doi.org/10.1007/11538059_91

[17] He, H., Bai, Y., Garcia, E. A., and Li, S. (2008). ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning. IEEE International Joint Con- ference on Neural Networks (IJCNN), 1322–1328. https://doi.org/10.1109/ IJCNN.2008.4633969

[18] Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T. (2007). Bias in Ran- dom Forest Variable Importance Measures: Illustrations, Sources and a Solution. BMC Bioinformatics, 8, 25. https://doi.org/10.1186/1471-2105-8-25

[19] Thomas, L. C., Edelman, D. B., and Crook, J. N. (2002). Credit Scoring and Its Applications. SIAM.

[20] Basel Committee on Banking Supervision. (2011). Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems. Bank for Interna- tional Settlements.