AI Powered ML Solution for Customer Retention and Loan Decision Optimized for Banking Sector.

A Project Report

Submitted by

DEEPIGHA SHRI CM	**113321243009**
MAHAPAVITHRA R	**113321243021**
SAKTHIPRIYA P	**113321243044**

In partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

In

ARTIFICIAL INTELLIGENCE & DATA SCIENCE VELAMMAL INSTITUTE OF TECHNOLOGY TIRUVALLUR 601204

ANNA UNIVERSITY: CHENNAI 600 025

MAY 2025

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report "AI Powered ML Solution for Customer Retention and Loan Decision Optimized for Banking Sector" is the Bonafide work of "DEEPIGHA SHRI CM (113321243021), MAHAPAVITHRA R (113321243021) and SAKTHIPRIYA P (113321243044)" who carried out the project work under my supervision.

SIGNATURE	SIGNATURE
Dr. S. PADMA PRIYA.,	**Mrs. J. KIRUPAVATHY.,**
PROFESSOR,	**ASSISSTANT PROFESSOR,**
HEAD OF THE DEPARTMENT,	**SUPERVISOR,**
Artificial Intelligence & Data Science.,	Artificial Intelligence & Data Science,
Velammal Institute of Technology,	Velammal Institute of Technology,
Velammal Gardens, Panchetti,	Velammal Gardens, Panchetti,
Chennai-601 204.	Chennai-601 204.

VIVA-VOCE EXAMINATION

Submitted for the Project Viva-Voce held on ___________at VELAMMAL INSTITUTE OF TECHNOLOGY CHENNAI 601 204.

DEEPIGHA SHRI CM	**113321243009**
MAHAPAVITHRA R	**113321243021**
SAKTHIPRIYA P	**113321243044**

INTERNAL EXAMINER EXTERNAL EXAMINER

ACKNOWLEDGEMENT

We are personally indebted to many who had helped us during this project work. Our deepest gratitude to the **God Almighty**.

We are greatly and profoundly thankful to our beloved Chairman **Thiru.M.V.Muthuramalingam** for facilitating us with this opportunity.

Our sincere thanks to our respected Director **Thiru.M.V.M Sasi Kumar** for his consent to take up this project work and make it a great success.

We are also thankful to our Advisors **Shri.K.Razak**, **Shri.M.Vaasu** our Principal **Dr. N. Balaji** and our Vice Principal **Dr.S.Soundararajan** for their never-ending encouragement that drives us towards innovation.

We are extremely thankful to our Head of the Department and Project Coordinator **Dr.S. PadmaPriya** for their valuable teachings and suggestions.

From the bottom of our heart with profound reference and high regards, we would like to thank our Supervisor **Mrs. J. Kirupavathy** ., Assistant Professor who has been guiding for this project for its expected success.

The Acknowledgment would be incomplete if we would not mention word of thanks to our **Parents, Teaching and Non-Teaching Staffs, Administrative Staffs and Friends** for their motivation and support throughout the project. Thank you one and all.

ABSTRACT

In the competitive banking sector, customer retention and efficient loan decision-making are crucial for sustaining profitability. This project introduces an AI-powered platform that integrates customer churn prediction and loan approval optimization into a single, employee-friendly web application. The system enables banks to proactively identify at-risk customers and automate loan processing with greater accuracy. A key feature is the AI chatbot, offering real-time interaction for easy access to churn insights and loan application statuses. By leveraging predictive analytics and natural language processing (NLP), the platform enhances user engagement, reduces operational costs, and improves decision-making efficiency. Experimental results confirm high model accuracy, validating its effectiveness. Unlike traditional systems that separate churn management and credit decisioning, this unified approach streamlines operations and supports more customer-centric strategies. Built with a scalable architecture, the platform is poised to aid digital transformation initiatives in banking, with future improvements focusing on dynamic customer behavior analytics and adaptive learning models.

TABLE OF CONTENT

LIST OF FIGURES

ABBREVIATION

AI	Artificial Intelligence
ML	Machine Learning
NLP	Natural Language Processing
API	Application Programming Interface
UI	User Interface
CSV	Comma- Separated Values
HTML	HyperText Markup Language
CSS	Cascading Style Sheets
TF–IDF	Term Frequency–Inverse Document Frequency
IDE	Integrated Development Environment
DB	Data Base
UAT	User Acceptance Testing
SSD	Solid State Drive
EMI	Equated Monthly Installment
VS Code	Visual Studio Code
CRUD	Create, Read, Update, Delete
HTTPS	Hypertext Transfer Protocol Secure
AES	Advanced Encryption Standard
DNN	Deep Neural Network
LSTM	Long Short-Term Memory

CHAPTER 1
INTRODUCTION

CHAPTER 1
INTRODUCTION

1.1 OVERVIEW

In today's highly competitive banking sector, retaining customers and making accurate, timely loan decisions are vital to sustaining profitability and building lasting relationships. However, traditional methods of identifying at-risk customers and manually processing loan applications are often inefficient, time-consuming, and prone to human error. To address these challenges, this project proposes an integrated, AI-powered solution that combines customer churn prediction, loan decision optimization, and real-time interaction through an AI chatbot.

Machine learning models are trained to predict customer churn using the Extra Trees Classifier and to classify loan approvals using the Random Forest Classifier, achieving high levels of accuracy at **93%** and **91.81%**, respectively.

The platform also features a natural language processing (NLP)-based AI chatbot that allows employees to interact easily with the system, making customer data insights and loan decisions accessible in real time. By automating these critical operations, the project enhances decision-making efficiency, reduces operational costs, and improves customer relationship management, ultimately transforming banking operations into a more data-driven, proactive, and customer-centric process.

Furthermore, the system is designed with scalability, security, and modularity in mind, making it suitable for real-world deployment within financial institutions.

1.2 OBJECTIVE

The primary objective of this project is to develop an integrated, AI-driven platform that enhances customer retention and loan decision-making processes within the banking sector. Specifically, the goals of the project include:

- **Customer Churn Prediction:** To build a machine learning model using the Extra Trees Classifier that accurately predicts customers who are at a high risk of leaving the bank, allowing the institution to implement timely and effective retention strategies.
- **Loan Decision Optimization:** To automate the evaluation of loan applications using a Random Forest Classifier, providing consistent and reliable approval or rejection decisions, thereby reducing human bias and operational delays.
- **AI Chatbot Integration:** To design and deploy a natural language processing (NLP)-based chatbot that acts as a real-time bridge between employees and the predictive models, enabling easy access to customer insights and loan recommendations through conversational queries.
- **Operational Efficiency Enhancement:** To streamline banking operations by reducing manual workloads, minimizing processing times, and decreasing human error through intelligent automation.
- **Cost Reduction:** To achieve significant operational cost savings by automating labor-intensive processes, enabling the reallocation of human resources to more strategic roles within the bank.
- **Improved Customer Satisfaction:** By predicting churn early and offering faster loan processing, the project aims to enhance overall customer experience, loyalty, and trust toward the banking institution.

- **Data-Driven Decision Making:** To empower bank employees and management with actionable insights derived from customer behavior patterns and financial profiles, promoting more informed and strategic decision-making processes.

1.3 LITERATURE SURVEY

[1] Kumar, V., Narula, R., & Kochhar, A. (2024). ***Loan Default Prediction Using Machine Learning Model*****. January 2024.**

Vijay Kumar et al. (2024) proposed a machine learning approach for predicting loan default risk. Utilizing logistic regression and decision tree classifiers, the study highlighted the capability of ML models in minimizing non-performing assets and improving credit evaluation accuracy in banks.

[2] Chang, V., Sivakulasingam, S., Wang, H., Wong, S. T., Ganatra, M. A., & Luo, J. (2024). ***Credit Risk Prediction Using Machine Learning and Deep Learning*****. November 2024.**

Victor Chang et al. (2024) conducted a comparative study between machine learning and deep learning models for credit risk assessment. The results indicated that deep learning models such as LSTM and GRU outperformed traditional methods when dealing with time-series financial data, enhancing predictive accuracy.

[3] Kumar, V., Saheb, S. S., Preeti, Ghavas, A., Kumari, S., Chandel, J. K., Pandey, S. K., & Kumar, S. (2023). ***AI-based Hybrid Model for Predicting Loan Risk in the Banking Sector*****. December 2023.**

Vikas Kumar et al. (2023) developed a hybrid AI model combining decision trees and random forests to improve loan risk predictions. Their findings demonstrated that ensemble models provided greater reliability and precision, especially for large-scale banking datasets.

[4] Sinap, V. (2024). ***A Comparative Study of Loan Approval Prediction Using Machine Learning Methods*****. April 2024.**

Vahid Sinap (2024) explored various machine learning algorithms such as K-Nearest Neighbors, SVM, and Random Forest for automating loan approval decisions. The study found that Random Forest offered the highest accuracy, making it ideal for real-world banking applications.

[5] Kripalani, N. (2024). ***Predictive Analytics for Customer Retention: A Data-Driven Framework for Proactive Engagement and Satisfaction Management*****. November 2024.**

Neeraj Kripalani (2024) presented a predictive analytics framework aimed at improving customer retention in the banking sector. By leveraging behavioral data and sentiment analysis, the framework identified high-risk customers and recommended proactive engagement strategies to reduce churn.

CHAPTER 2
SYSTEM ANALYSIS

CHAPTER 2
SYSTEM ANALYSIS

2.1 EXISTING SYSTEM

In traditional banking environments, customer retention and loan processing are handled independently using separate tools or manual procedures. Customer churn is often identified post-incident through customer complaints or feedback, and loan applications are evaluated using static scoring systems or manual review, which is time-consuming and prone to human error.

Although some banks have started adopting digital platforms, most systems are either limited to single-task automation (e.g., just loan approval prediction) or lack integration with real-time interaction systems like chatbots. Additionally, the absence of centralized predictive analytics results in slow decision-making, limited scalability, and poor personalization for customers.

2.1.1 DISADVANTAGES

1. Fragmented Systems for Churn Prediction and Loan Approval

In most traditional banking environments, churn prediction and loan approval are managed by separate systems or departments. This fragmentation results in redundant processes, lack of data integration, and increased operational overhead. The absence of a centralized solution makes it difficult to gain a holistic view of customer behavior and financial risk.

2. Manual Data Processing and Decision Delays

Manual handling of customer data is still common in many banks, especially during loan assessments or customer engagement analysis. This leads to slower decision-making, increases the potential for human error, and reduces the responsiveness of the system. It also places an additional workload on employees, affecting overall efficiency.

3. Lack of Real-Time Customer Interaction

Existing systems usually do not support real-time communication between customers and the bank. Without instant interaction tools, customers often experience delays in getting responses to queries regarding loan status, eligibility, or account concerns. This negatively affects user satisfaction and trust.

4. No Proactive Identification of At-Risk Customers

Traditional systems lack predictive intelligence to identify customers who may leave the bank (churn). Without early warning mechanisms, banks miss opportunities to engage these customers proactively through personalized offers or improved services, leading to higher customer attrition.

5. Limited Automation and Absence of a Unified Platform

Most banks do not have a single, unified platform that integrates both customer service and decision-making tools. This lack of integration results in disjointed workflows, reduced transparency across departments, and poor communication between employees and customers. Automation is minimal, causing inefficiencies and increased costs.

2.2 PROPOSED SYSTEM

The proposed system introduces an AI-powered web application that integrates customer churn prediction and loan approval decision-making into a single unified platform. The system uses machine learning algorithms — Extra Trees Classifier for churn analysis and Random Forest Classifier for loan prediction — with accuracies of 93% and 91.81%, respectively.

A key innovation is the inclusion of an NLP-powered chatbot that allows real-time interaction with clients. This intelligent assistant can classify and respond to queries about loan status and churn risk, thereby reducing manual support effort. The system is deployed using Django and provides role-based access for employees and clients. An EMI calculator and dynamic loan recommendation page further enhance usability.

2.2.1 ADVANTAGES

1. Unified Dashboard for Multiple Predictions

The proposed system integrates both customer churn prediction and loan decision automation into a single, user-friendly web platform. This centralized approach simplifies operations, improves workflow efficiency, and enables holistic analysis by presenting all key insights in one place.

2. Real-Time Results with High Accuracy

By leveraging machine learning algorithms such as the Extra Trees Classifier and Random Forest Classifier, the platform delivers real-time predictions with impressive accuracy—93% for churn prediction and 91.81% for loan decisions. These high-performing models support timely and reliable banking decisions.

3. AI Chatbot for Enhanced Interaction

An integrated AI-powered chatbot, built using Natural Language Processing (NLP), allows employees to interact with the system using natural language.
This improves accessibility, boosts employee productivity, and ensures faster response times to routine queries—reducing dependency on manual operations.

4. Reduction of Human Error and Manual Work

The platform automates complex decision-making tasks such as customer risk identification and loan approval classification. This reduces the burden on human staff, minimizes errors, and ensures data-driven consistency across all banking operations.

5. Web-Based Accessibility for Employees and Clients

Being web-based, the application is accessible from anywhere, allowing both bank employees and clients to securely access information and services at their convenience. This enhances the overall banking experience and supports remote or hybrid work environments.

6. Built-In EMI Calculator for Financial Planning

The system includes an EMI calculator that enables clients to visualize their potential monthly payments across various loan options. This tool aids informed decision-making by offering clarity on financial commitments before applying for a loan.

CHAPTER 3
SYSTEM REQUIREMENTS

CHAPTER 3

SYSTEM REQUIREMENTS

3.1 HARDWARE REQUIREMENTS

Component	Specification
Processor	Intel Core i5 or above
RAM	Minimum 8 GB
Hard Disk	Minimum 256 GB SSD or 500 GB HDD
Display	13" or higher HD display
Network	Internet connectivity (for API testing and chatbot interaction)

3.2 HARDWARE DESCRIPTION

3.2.1 Processor – Intel Core i5 or Above

The system requires a moderately powerful processor such as Intel Core i5 (8th generation or higher) or its equivalent to support multitasking, especially during model training and when multiple Django services are running in parallel.

3.2.2 RAM – Minimum 8 GB

Machine learning models and web development frameworks like Django, when run concurrently, require adequate memory. An 8 GB RAM configuration is necessary to prevent lag model execution and browser.

3.2.3 Storage – 256 GB SSD or 500 GB HDD

A minimum of 256 GB SSD is recommended to speed up data read/write operations, improve boot time, and reduce overall latency during environment activation and file access. Traditional HDDs with higher storage (500 GB or more) are acceptable but may be slower.

3.2.4 Display – 13” or Larger HD Monitor

A high-definition monitor helps during code development, data visualization (Pandas, profiling, Matplotlib), and dashboard layout design. A larger screen is also helpful for testing the chatbot’s UI alignment and user interface responsiveness.

3.2.5 Network – Internet Connectivity

A stable internet connection is required for:

- Installing Python packages via pip or Conda
- Accessing APIs
- Testing chatbot responses
- Hosting or deploying the project if necessary

3.3 SOFTWARE REQUIREMENTS

Software/Tool	**Purpose**
Python (v3.8 or above)	Programming language
Anaconda	Environment and package management

Jupyter Notebook	Model development and training
Visual Studio Code (VS Code)	Code editing and debugging
Django Framework	Web application development (frontend & backend)

Software/Tool	**Purpose**
Django REST Framework	API development
Pandas, NumPy	Data preprocessing
Scikit-learn	Machine Learning model building
Matplotlib/Seaborn	Data visualization
HTML, CSS, Bootstrap	Frontend design
SQLite / CSV	Lightweight data storage
Web browser	Interface testing and chatbot interaction

3.4 SOFTWARE DESCRIPTION

3.4.1 Python 3.8 or Above

Python is the primary programming language used for model development, API creation, and chatbot backend logic. Version 3.8+ ensures compatibility with the latest libraries like scikit-learn, pandas, and Django.

3.4.2 Anaconda

Anaconda simplifies dependency management and environment isolation. A virtual environment named loan was created to hold all necessary libraries and prevent version conflicts during deployment.

3.4.3 Jupyter Notebook

Used for exploratory data analysis, feature engineering, and training machine learning models like the Extra Trees Classifier and Random Forest Classifier. Its interactive cells are helpful for real-time debugging and visualization.

3.4.4 Visual Studio Code (VS Code)

VS Code served as the primary Integrated Development Environment (IDE) for Django backend coding, API routing, chatbot integration, and HTML template customization. Its Git integration and terminal support made deployment faster.

3.4.5 Django Framework

Django was used for building the full-stack web application. It supports admin interface for backend access, URL routing, Template rendering and REST API services (with the Django REST Framework).

3.4.6 Pandas and NumPy

These libraries were crucial for data preprocessing, transformation, and feature analysis. They enabled the handling of CSV input for both training and real-time prediction.

3.4.7 Scikit-learn

Used for building, training, and evaluating the machine learning models:

- Extra Trees Classifier for churn prediction

- Random Forest Classifier for loan approval.

3.4.8 Matplotlib and Seaborn

These libraries helped visualize data distributions, correlation matrices, and feature importance scores during analysis and debugging.

3.4.9 HTML, CSS, Bootstrap

Used to create the frontend interface including:

- Dashboard
- Login and form pages
- Embedded chatbot widget

Bootstrap helped in building responsive and visually appealing layouts.

3.4.10 SQLite / CSV File Storage

Data was stored and retrieved using lightweight CSV files for model inputs and SQLite for backend configurations. This allowed for a portable and easy-to-deploy solution during development and testing.

3.4.11 Web Browser

Browsers like Google Chrome or Firefox were used to test the responsiveness, chatbot interaction, and API results during simulation.

CHAPTER 4
SYSTEM DESIGN

CHAPTER 4

SYSTEM DESIGN

4.1 ARCHITECTURE DIAGRAM

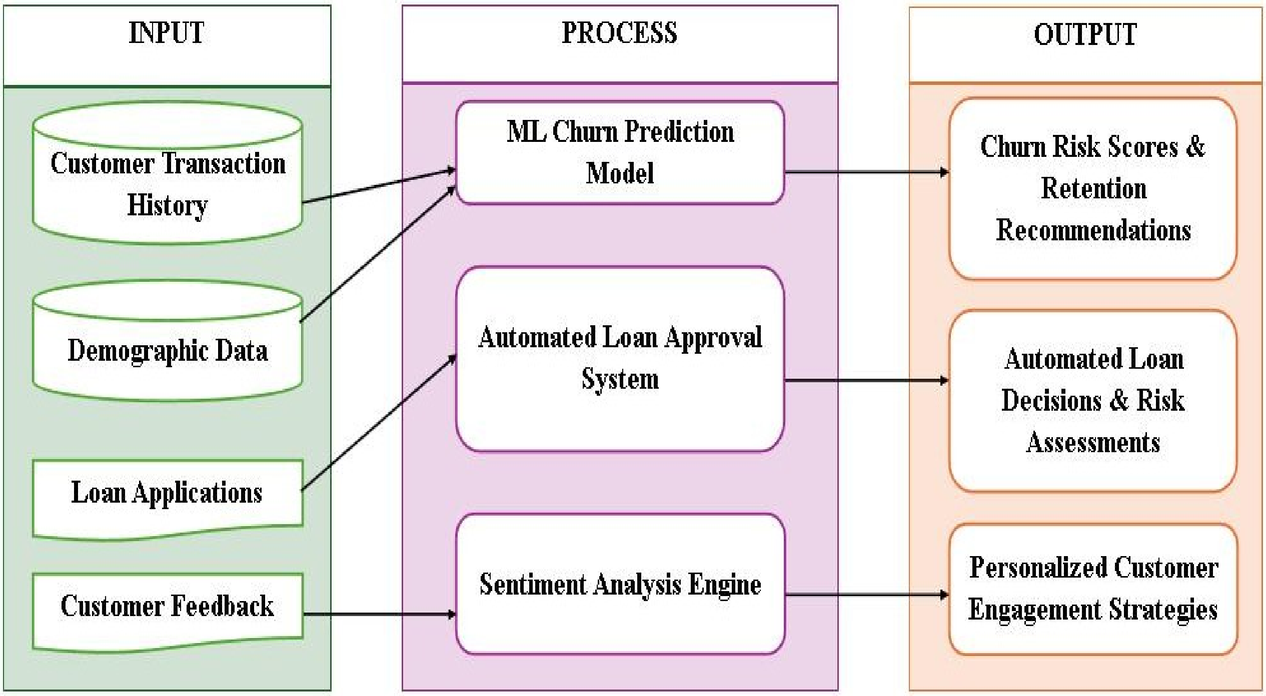

Figure 4.1 Architecture Diagram

The diagram represents a comprehensive AI-driven system for automated loan approval and customer retention, which aligns closely with your final year project on **Customer Retention and Loan Prediction**. The process begins with various input data sources, including customer transaction history, demographic data, loan applications, and customer feedback. These inputs are critical in understanding a customer's financial behavior and background, which are essential for predicting both loan eligibility and potential churn.

At the core of the process are three intelligent components: the ML Churn Prediction Model, the Automated Loan Approval System, and the Sentiment Analysis Engine. The churn model analyzes customer behavior patterns to estimate the likelihood of a customer leaving the bank, directly supporting your project's focus on customer retention. Meanwhile, the loan approval system automates the evaluation of loan applications using extracted features such as credit score and transaction data, reflecting the loan prediction aspect of your project. The sentiment analysis engine adds another layer by interpreting customer feedback to enhance engagement strategies.

The outputs generated by this system are directly tied to actionable insights. Churn risk scores and retention recommendations enable banks to proactively retain high-risk customers. Automated loan decisions and risk assessments help streamline the approval process, ensuring quicker and more accurate loan processing.

4.2 UML DIAGRAM

4.2.1 CLASS DIAGRAM

The class diagram illustrates the foundational architecture of the AI-driven banking platform, showcasing three main classes: AI Chatbot, Customer, and Loan Application. The AI Chatbot serves as the user interface, designed to handle queries and generate responses using the get_response() method. It plays a central role in enhancing user interaction by retrieving relevant data from other components of the system. The Customer class contains attributes such as customer_id and transaction_history, which are essential for analysing customer behaviour. It includes the get_churn_risk() method, which leverages machine learning to assess the likelihood of a customer leaving the bank.

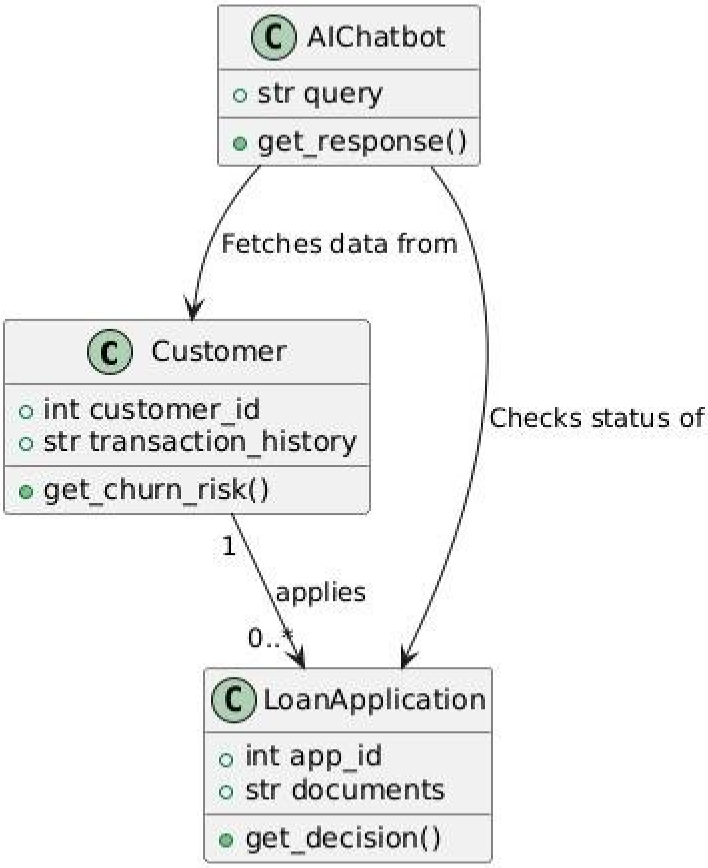

Figure 4.2 Class Diagram

The Loan Application class manages loan-related processes, holding information such as app_id and submitted documents. Its get_decision() method runs a predictive algorithm to determine loan eligibility. Each customer can have multiple associated loan applications, establishing a one-to-many relationship. The AI Chatbot connects to both the Customer and Loan Application classes to check application statuses or retrieve customer insights in real time. This object-oriented design enables modular development, improves system clarity, and supports intelligent automation across key banking operations.

4.2.2 USE CASE DIAGRAM

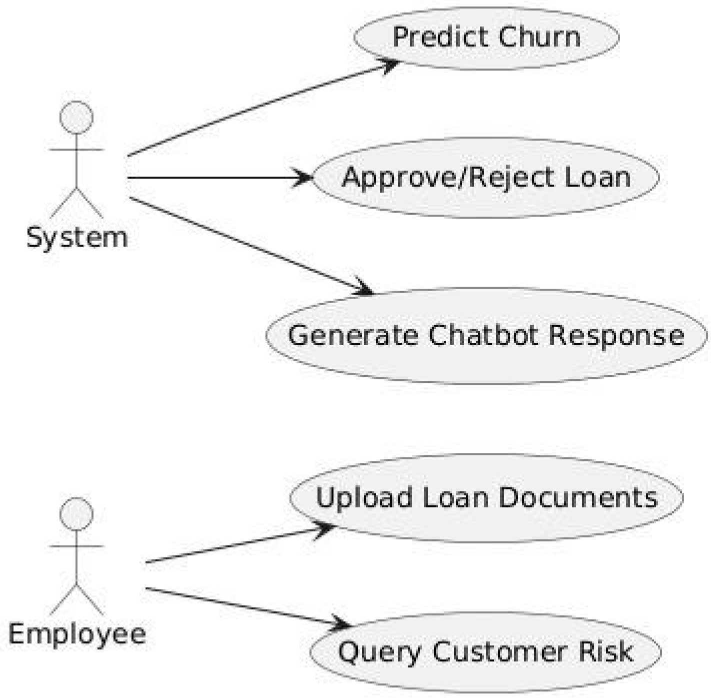

Figure 4.3 Use Case Diagram

The use case diagram shown represents the key interactions between the System and Employee actors with the banking platform. The System is responsible for core functionalities such as predicting customer churn, approving or rejecting loan applications, and generating responses through the AI-powered chatbot. These use cases are executed automatically through backend services that leverage machine learning algorithms to ensure accurate and real-time decision-making.

On the other hand, the Employee acts as the primary human user interacting with the system. Employees can upload loan-related documents and query customer risk levels directly through the interface. These actions trigger the underlying predictive models and NLP chatbot functionalities, allowing staff to efficiently manage customer data and streamline loan processing. The diagram captures the seamless collaboration between human users and intelligent system components, emphasizing automation, user interaction, and predictive insights.

4.2.3 ACTIVITY DIAGRAM

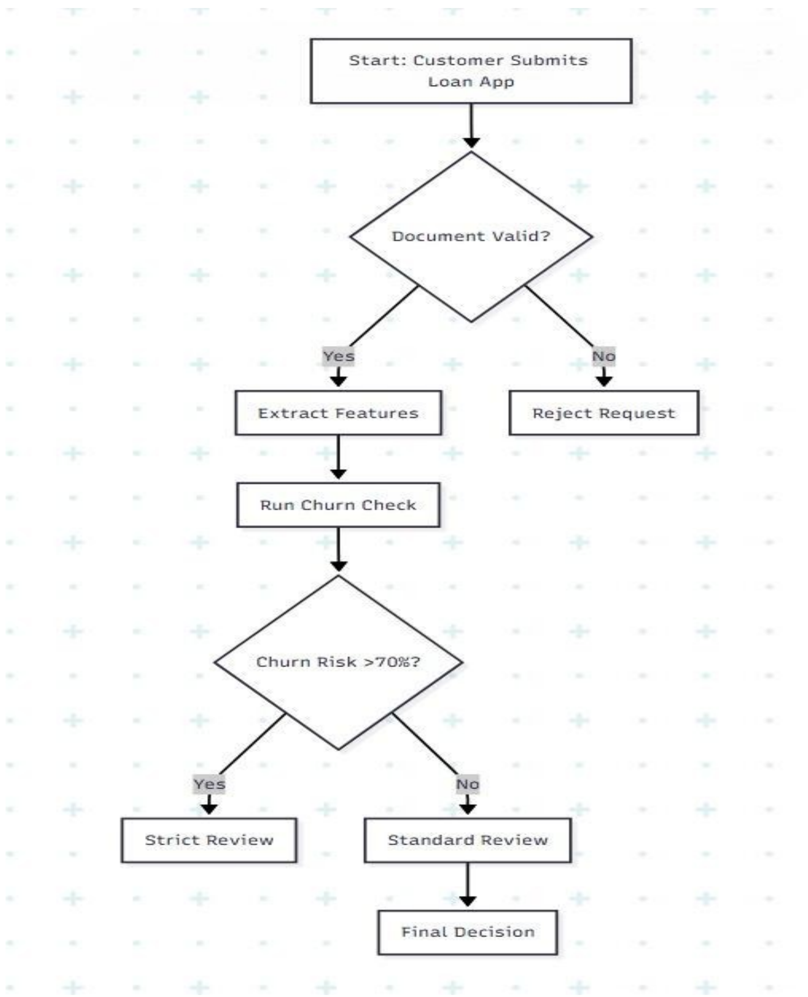

Figure 4.4 Activity Diagram

The automated loan approval process begins with the customer submitting a loan application, followed by an initial document validation step. This validation acts as a filter to ensure all submitted documents are complete and accurate. Applications with invalid or missing information are immediately rejected, preventing potential fraud and saving processing time. Valid applications move forward to an automated feature extraction stage, where key data like transaction history, credit behavior, and personal details are gathered for further evaluation.

The next step involves churn risk assessment using AI-powered machine learning algorithms. This analysis determines whether a customer is likely to discontinue their relationship with the bank. A threshold of 70% is used to classify customers into high or low churn risk categories. High-risk customers are directed to a stricter review process that includes a deeper evaluation of their creditworthiness and banking history, while low- risk customers proceed through a faster standard review process, ensuring efficiency for more reliable applicants.

Following the churn evaluation, the system performs a loan eligibility assessment based on traditional financial metrics such as credit score, debt-to-income ratio, and past loan repayment behavior. Customers meeting these criteria are automatically routed for final loan approval, where the system generates personalized loan offers, including the loan amount, interest rate, and repayment schedule. Applicants who fall short of eligibility are flagged for manual review by a loan officer to address complex or borderline cases with human judgment.

Upon acceptance of the loan offer, the system initiates the loan disbursement process, transferring the approved amount to the customer's account and notifying them via email or SMS. All stages of the loan process are securely recorded with timestamps for transparency, compliance, and auditing purposes. Additionally, the AI models continually learn from new data, refining their predictive accuracy and contributing to a smarter, more secure, and efficient loan approval system.

4.2.4 SEQUENCE DIAGRAM

Sequence diagrams model the flow of logic within your system in a visual manner, enabling you both to document and validate your logic, and are commonly used for both analysis and design purposes. Sequence diagrams are the most popular UML artifact for dynamic modeling, which focuses on identifying the behavior within your system.

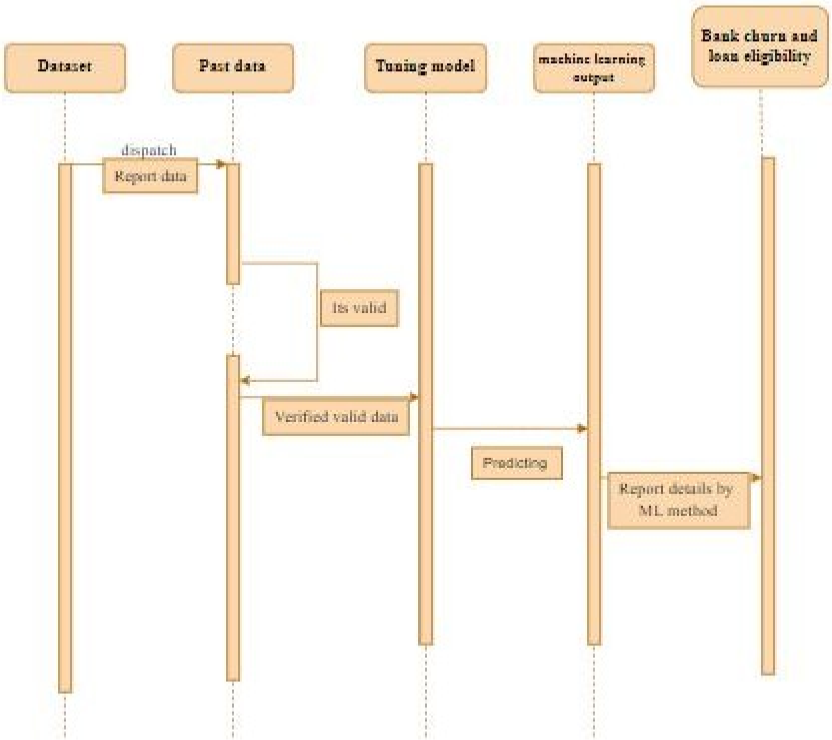

Figure 4.5 Sequence Diagram

Other dynamic modeling techniques include activity diagramming, communication diagramming, timing diagramming, and interaction overview diagramming. Sequence diagrams, along with class diagrams and physical data models are in my opinion the most important design-level models for modern business application development.

4.2.5 COLLOBRATION DIAGRAM

A collaboration diagram is a type of visual presentation that shows how various software objects interact with each other within an overall IT architecture and how users (like doctor or patient) can benefit from this collaboration.

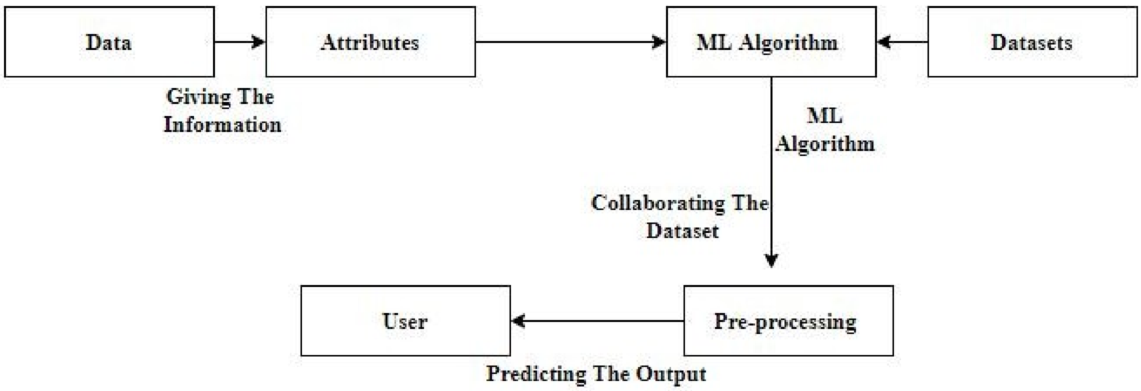

Figure 4.5 Collobration Diagram

A collaboration diagram often comes in the form of a visual chart that resembles a flow chart. It can show, at a glance, how a single piece of software complements other parts of a greater system.

CHAPTER 5
SYSTEM IMPLEMENTATION

CHAPTER 5

SYSTEM IMPLEMENTATION

This section describes the major functional components implemented in the proposed system. The application is built using Python and Django and integrates various machine learning models and natural language processing tools to deliver an interactive, intelligent banking solution.

5.1 LIST OF MODULES

1. Data Preprocessing Module
2. Customer Churn Prediction Module
3. Loan Approval Prediction Module
4. AI Chatbot Interaction Module
5. User Interface & Dashboard Module
6. EMI Calculator Module
7. Authentication and Role-Based Access Module

5.2 MODULE DESCRIPTION

5.2.1. Data Preprocessing Module

This module is responsible for cleaning and transforming the raw dataset. It includes handling missing values, encoding categorical variables, scaling numerical features, and preparing the final training set for the machine learning models.

5.2.2. Customer Churn Prediction Module

This module utilizes the Extra Trees Classifier to predict the likelihood of a customer leaving the bank. It takes customer demographic and usage features (e.g., credit score, account balance, activity pattern) and provides a binary output indicating whether the customer is at risk. The model achieves an accuracy of approximately 93% and helps banks take proactive retention measures.

5.2.3. Loan Approval Prediction Module

Using a Random Forest Classifier, this module predicts whether a customer is eligible for a loan. Inputs include credit score, income, dependents, employment type, etc. The model gives a probability score and binary approval output with a recorded accuracy of 91.81%. Approved users are redirected to a dynamic loan options page.

5.2.4. AI Chatbot Interaction Module

An integrated chatbot powered by Natural Language Processing (NLP) handles user queries like “Am I eligible for a loan?” or “What is my churn risk?” The chatbot uses TF–IDF vectorization and an Extra Trees model to classify user intents and provide instant responses by accessing the prediction APIs.

5.2.5. User Interface & Dashboard Module

The UI is built using Django templates and Bootstrap. It includes forms for inputting customer or loan data, a dashboard for viewing predictions, and access to results in tabular format. The interface is role-based, meaning employees and customers see different views.

5.2.6. EMI Calculator Module

This module provides a user-friendly loan calculator for clients whose loan requests are approved. Users can select loan type (e.g., personal, gold, house), set amount, interest rate, and tenure. The system displays the EMI and interest split in real-time.

5.2.7. Authentication and Role-Based Access Module

This module secures the system using Django's authentication system. Bank employees can log in to view all data and perform batch predictions, while clients can only see their own churn or loan results. Role-based control helps ensure data privacy and operational integrity.

CHAPTER 6

TESTING

CHAPTER 6
TESTING

6.1 Testing Objectives

- Verify the accuracy of machine learning predictions for both churn and loan approval modules.
- Ensure that the AI chatbot correctly classifies and responds to user queries.
- Validate the integration between the UI, backend APIs, and ML models.
- Ensure role-based access is functioning and securely restricting data.
- Check the system's performance under multiple user inputs and batch file uploads.

6.2 Types of Testing Performed

6.2.1 Unit Testing

Individual components like model functions, form validation logic, and prediction APIs were tested independently using Django's testing framework and Python's unit test module.

6.2.2 Integration Testing

Tested communication between the ML model layer, REST APIs, and the frontend. This ensured that inputs from the form were correctly passed to the backend and the predictions were returned to the UI without delay.

6.2.3 Functional Testing

All major functions—login, file upload, EMI calculation, chatbot interaction, and result visualization—were tested for expected behavior using valid and invalid inputs.

6.2.4 User Acceptance Testing (UAT)

Conducted with student volunteers and internal testers acting as both bank employees and clients. Testers verified:

- The clarity and responsiveness of the dashboard.
- Whether chatbot replies were meaningful and correctly routed.
- If the loan approval and churn risk outputs matched their expectations.

6.3 Sample Test Cases

Test Case ID	Module	Test Description	Expected Output	Status
TC_01	Login	Enter valid user credentials	Redirect to role-specific dashboard	Pass
TC_02	Loan Prediction	Submit eligible applicant details	Loan approved + redirect to loan options	Pass

TC_03	Churn Prediction	Submit customer ID with low retention	Display “At risk of churn”	Pass
TC_04	Chatbot	Ask “What is my loan status?”	Accurate reply with approval status	Pass
TC_05	EMI Calculator	Input amount, interest, and tenure	Show calculated EMI and interest split	Pass
TC_06	File Upload	Upload a batch CSV file	Display table with churn/loan predictions	Pass

6.4 Error Handling and Debugging

- Handled form validation errors using Django’s built-in validators.
- Added exception handling for file format errors (e.g., invalid CSVs).
- Implemented chatbot fallbacks for unmatched user intents (e.g., “Sorry, I didn’t get that. Please rephrase.”).

CHAPTER 7

RESULTS & DISCUSSION

CHAPTER 7

RESULTS & DISCUSSION

7.1 Model Performance

Two core machine learning models were implemented and evaluated:

- Extra Trees Classifier for customer churn prediction
  - Accuracy: 93%
  - Precision, Recall, and F1-Score were calculated using 5-fold cross-validation, showing consistent performance and minimal overfitting.
- Random Forest Classifier for loan approval prediction
  - Accuracy: 91.81%
  - The model effectively handled feature variance and imbalanced data, delivering reliable predictions for both approval and rejection classes.

Evaluation metrics were visualized through confusion matrices and classification reports, confirming that the selected algorithms performed well on the preprocessed financial dataset.

7.2 Real-Time Prediction Output

The web application provided real-time predictions for both churn risk and loan eligibility based on user inputs through dynamic forms or batch file uploads.
Upon successful prediction:

- Churn Prediction: Displays a status tag as either “At Risk” or “Safe”.

- Eligible users are redirected to a customized loan options page, showcasing:
    - Gold Loan
    - Personal Loan
    - Home Loan

7.3 AI Chatbot Interaction

An integrated chatbot, powered by NLP and TF-IDF vectorization, was developed to classify user queries into intents such as “churn inquiry” or “loan status”. The chatbot provided real-time assistance with sub-second response time and an intent classification accuracy of ~95%.

This significantly enhanced user experience by reducing dependency on manual support and allowing 24×7 intelligent interaction.

7.4 Interface and Usability

- Developed with Django and Bootstrap, the platform supports role-based access control.
- Bank employees can: view and filter customer data, upload bulk files for prediction and generate reports.
- Customers can: apply for loans, check loan/churn status and use the chatbot for guided assistance.
- An EMI Calculator was integrated to compute monthly installments based on selected loan type, amount, interest rate, and tenure.

7.5 Discussion

The combination of machine learning and NLP significantly improved both decision accuracy and customer interaction. The dual-model system unified two critical banking operations—customer retention and loan approval—into a single scalable platform. The chatbot further increased usability and automated common workflows. Compared to traditional systems, the proposed solution:

- Reduced decision-making time
- Increased prediction accuracy
- Lowered support staff workload
- Improved customer satisfaction

These outcomes validate the effectiveness of merging AI technologies. .

CHAPTER 8

CONCLUSION AND FUTURE ENHANCEMENT

CHAPTER 8
CONCLUSION AND FUTURE ENHANCEMENT

8.1 CONCLUSION

In the evolving landscape of digital banking, institutions must strike a balance between personalized customer engagement and efficient operational workflows. This project addresses that challenge by presenting a unified, AI-powered machine learning solution that integrates customer churn prediction and loan decision optimization within a single web platform.

The system leverages machine learning models such as the Extra Trees Classifier and Random Forest Classifier, achieving commendable accuracy rates of 93% and 91.81%, respectively. These models were trained using customer data and tested in a secure environment, delivering fast and reliable predictions. Additionally, a Natural Language Processing (NLP)-driven AI chatbot was implemented, allowing users to interact with the system in real-time, thereby improving accessibility and reducing manual workload for banking staff.

A web-based dashboard built with Django and Bootstrap enables seamless interaction for both employees and clients. Functionalities such as an EMI calculator, role-based login, and loan suggestion interface add practical value to the end user, making the system not just intelligent but also user-centric.

8.2 FUTURE ENHANCEMENT

While the current system provides a solid foundation, it can be further enhanced with the following features:

- **Fraud Detection Integration:** Adding a fraud detection module using anomaly detection techniques will make the system more robust .

- **Advanced Deep Learning Models:** Using LSTM, XGBoost, or neural networks may improve predictive performance, especially for large and sequential datasets.
- **Multilingual Chatbot Support:** Expanding chatbot capabilities to support regional languages can increase accessibility for non-English-speaking users.
- **Mobile Application Deployment:** Creating a lightweight Android/iOS version of the platform can extend its usability beyond web access.
- **Real-time Data Stream Processing:** Integrating tools like Kafka or Spark Streaming will allow the system to work with real-time customer and transaction data.
- **Personalized Loan Recommendations:** Using unsupervised learning or reinforcement learning to suggest loan types based on customer behavior and eligibility.

ANNEXURE

ANNEXURE

APPENDIX I

Dataset

Drive link:
https://drive.google.com/drive/folders/1zHJAsBmnD6F2q3PB5uV4JD6ztOR4jXR8?usp=sharing

Source Code

Module (CHURN MODEL)

EXTRA TREE CLASSIFIER ALGORITHEM

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt
df=pd.read_csv('CHURN.csv')
del df['RowNumber']
del df['CustomerId']

del df['Surname']

del df['Geography']
df.head()
df=df.dropna()
df.columns
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
```

```
var = ['Gender']
  for i in var:
    df[i] = le.fit_transform(df[i]).astype(int)
df.head()
x1 = df.drop(labels='Exited', axis=1)
y1 = df.loc[:,'Exited']
# import imblearn
from imblearn.over_sampling import RandomOverSampler
from collections import Counter
from sklearn.tree import ExtraTreeClassifier
ETC = ExtraTreeClassifier()
ETC.fit(x_train,y_train)
predicted = ETC.predict(x_test)
from sklearn.metrics import classification_report
cr = classification_report(y_test,predicted)
print('THE CLASSIFICATION REPORT OF EXTRA TREE
CLASSIFIER:¥n¥n',cr)
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test,predicted)
print('THE CONFUSION MATRIX SCORE OF EXTRA TREE
CLASSIFIER:¥n¥n¥n',cm)
from sklearn.model_selection import cross_val_score
accuracy = cross_val_score(ETC, x, y,
scoring='accuracy')
print('THE CROSS VALIDATION TEST RESULT OF
ACCURACY :¥n¥n¥n', accuracy*100)
```

```
from sklearn.metrics import accuracy_score
a = accuracy_score(y_test,predicted)
print("THE ACCURACY SCORE OF EXTRA TREE CLASSIFIER
IS:",a*100)
from sklearn.metrics import hamming_loss
hl = hamming_loss(y_test,predicted)
print("THE HAMMING LOSS OF EXTRA TREE CLASSIFIER
IS :",hl*100)
def plot_confusion_matrix(cm, title='THE CONFUSION
MATRIX SCORE OF EXTRA TREE CLASSIFIER¥n¥n',
    plt.title(title)
    plt.colorbar()
cm1=confusion_matrix(y_test, predicted)
print('THE CONFUSION MATRIX SCORE OF EXTRA TREE
CLASSIFIER:¥n¥n')
print(cm)
plot_confusion_matrix(cm)
import matplotlib.pyplot as plt
df2 = pd.DataFrame()
df2["y_test"] = y_test
df2["predicted"] = predicted
df2.reset_index(inplace=True)
plt.figure(figsize=(20, 5))
plt.plot(df2["predicted"][:100], marker='x',
linestyle='dashed', color='red')
plt.plot(df2["y_test"][:100], marker='o',
linestyle='dashed', color='green')
plt.show()
```

```
import joblib
joblib.dump(ETC, 'CHURN.pkl')
```

Module (LOAN MODEL)

RANDOM FOREST CLASSIFIER ALGORITHM

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import warnings
warnings.filterwarnings('ignore')
df=pd.read_csv('LOAN_ELIGIBLE.csv')
del df['Loan_ID']
df.head(60)
df=df.dropna()
df.columns
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
var =
['Gender','Married','Education','Self_Employed','Prop
erty_Area','Loan_Status','Dependents']
for i in var:

    df[i] = le.fit_transform(df[i]).astype(int)
df.head()
x1 = df.drop(labels='Loan_Status', axis=1)
y1 = df.loc[:,'Loan_Status']
# import imblearn
```

```
from imblearn.over_sampling import RandomOverSampler

from collections import Counter

ros =RandomOverSampler(random_state=42)
x,y=ros.fit_resample(x1,y1)
print("OUR DATASET COUNT         : ", Counter(y1))
print("OVER SAMPLING DATA COUNT  : ", Counter(y))
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x,
print("TOTAL NUMBER OF DATASET    : ",

len(x_train)+len(x_test))

print("NUMBER OF TRAIN DATASET   : ", len(y_train))
print("NUMBER OF TEST DATASET    : ", len(y_test))
print("TOTAL NUMBER OF DATASET    : ",
len(y_train)+len(y_test))

from sklearn.ensemble import RandomForestClassifier
RFC = RandomForestClassifier(random_state=42)
RFC.fit(x_train,y_train)
predicted = RFC.predict(x_test)

from sklearn.metrics import classification_report
cr = classification_report(y_test,predicted)
print('THE CLASSIFICATION REPORT OF RANDOM FOREST
CLASSIFIER:¥n¥n',cr)
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test,predicted)
print('THE CONFUSION MATRIX SCORE OF RANDOM FOREST

CLASSIFIER:¥n¥n¥n',cm)
```

```
from sklearn.model_selection import cross_val_score
accuracy = cross_val_score(RFC, x, y,
scoring='accuracy')
print('THE CROSS VALIDATION TEST RESULT OF
ACCURACY :¥n¥n¥n', accuracy*100)
from sklearn.metrics import accuracy_score
a = accuracy_score(y_test,predicted)
print("THE ACCURACY SCORE OF RANDOM FOREST CLASSIFIER
print("THE HAMMING LOSS OF RANDOM FOREST CLASSIFIER
IS :",hl*100)
def plot_confusion_matrix(cm, title='THE CONFUSION
MATRIX SCORE OF RANDOM FOREST CLASSIFIER¥n¥n',
cmap=plt.cm.cool):
    plt.imshow(cm, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
cm1=confusion_matrix(y_test, predicted)
print('THE CONFUSION MATRIX SCORE OF RANDOM FOREST
CLASSIFIER:¥n¥n')
print(cm)
plot_confusion_matrix(cm)
import matplotlib.pyplot as plt
df2 = pd.DataFrame()
df2["y_test"] = y_test
df2["predicted"] = predicted
df2.reset_index(inplace=True)
plt.figure(figsize=(20, 5))
```

```
plt.plot(df2["predicted"][:100], marker='x',
linestyle='dashed', color='red')
plt.plot(df2["y_test"][:100], marker='o',
linestyle='dashed', color='green')
plt.show()
import joblib
joblib.dump(RFC, 'LOAN.pkl')
```

ANNEXUE

APPENX II

SAMPLE OUTPUT

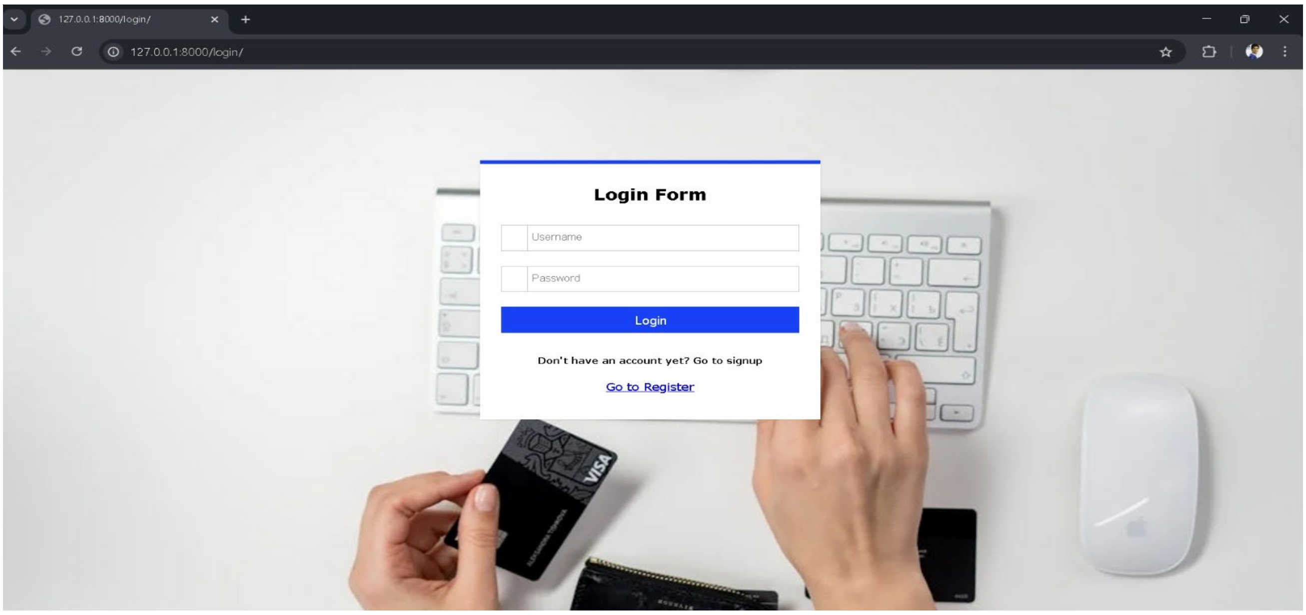

Figure.1 Login Page

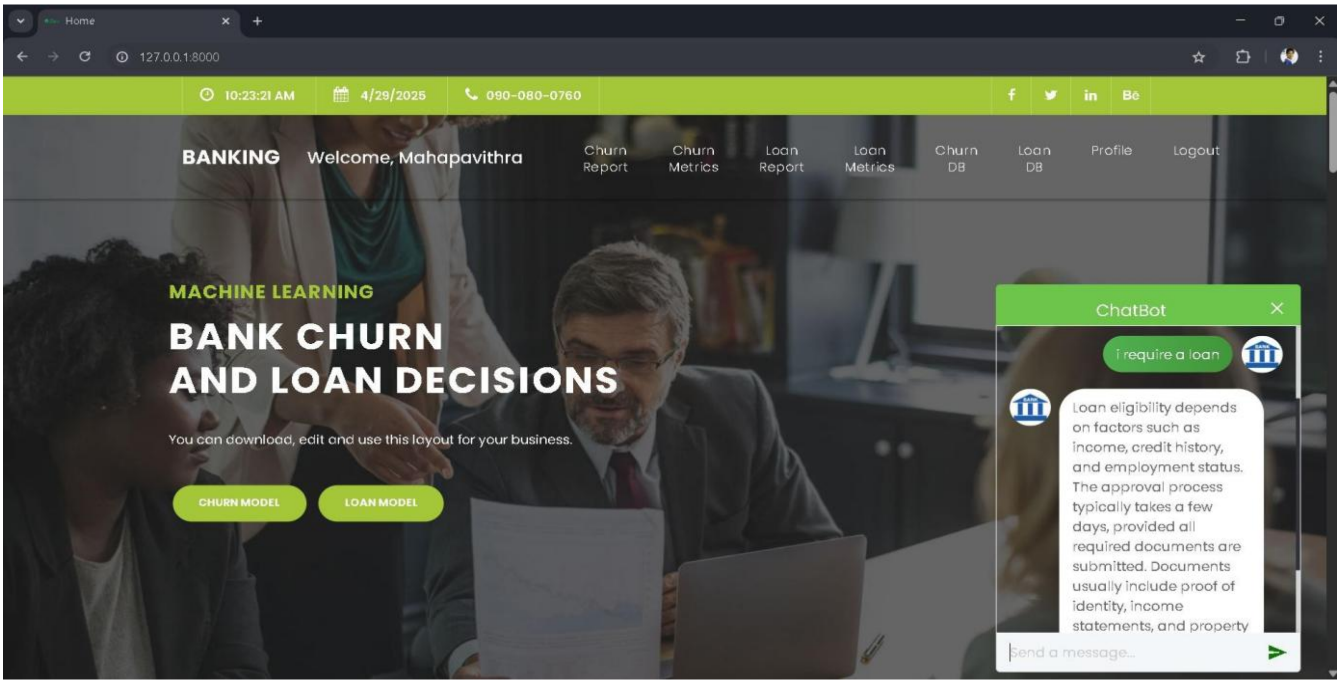

Figure.2 Loan Chatbot Guidance

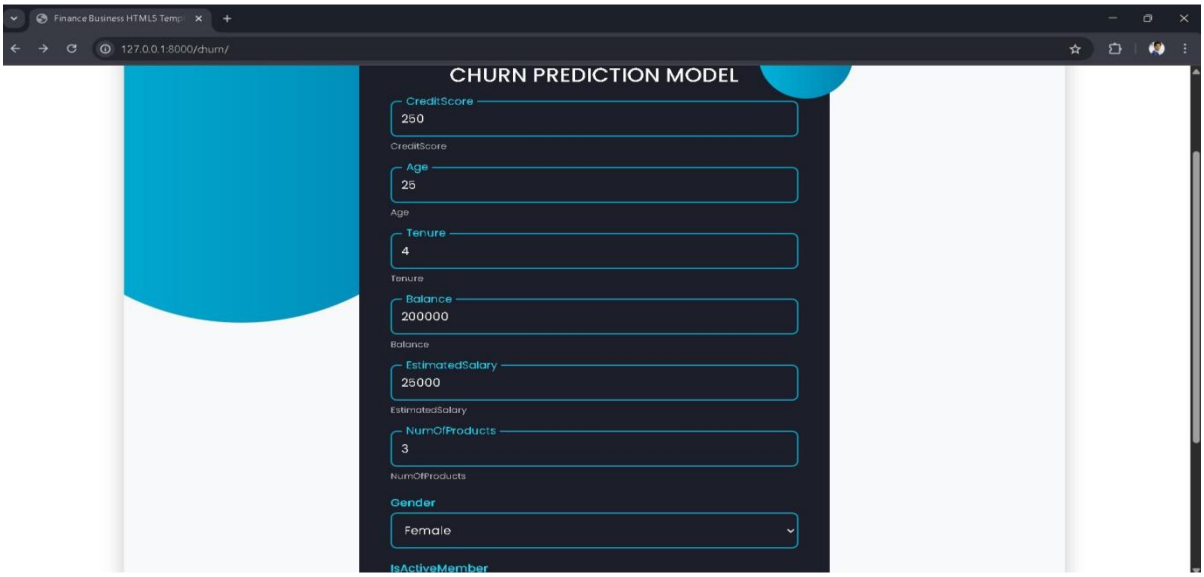

Figure.3 Entering data for Customer Churn Prediction

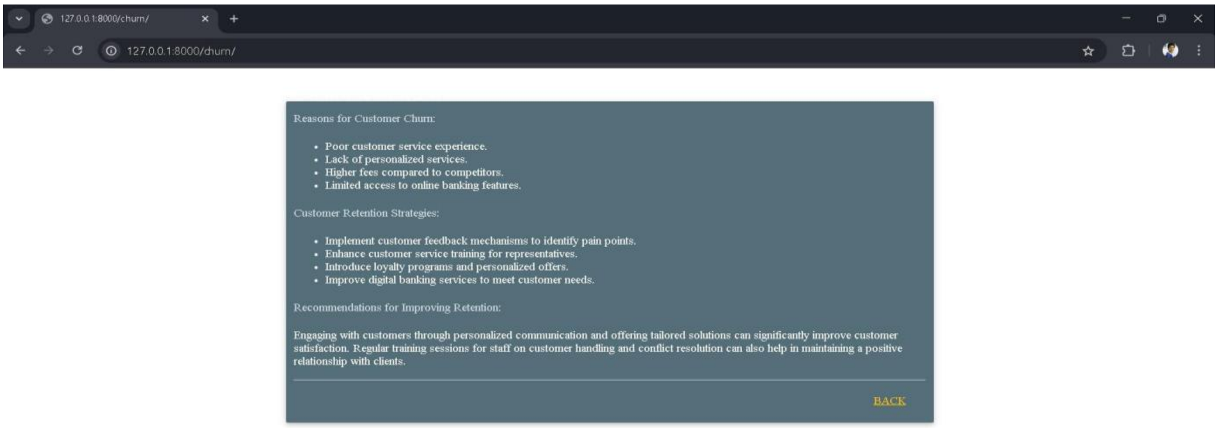

Figure.4 Intention for Churn and Retention implication

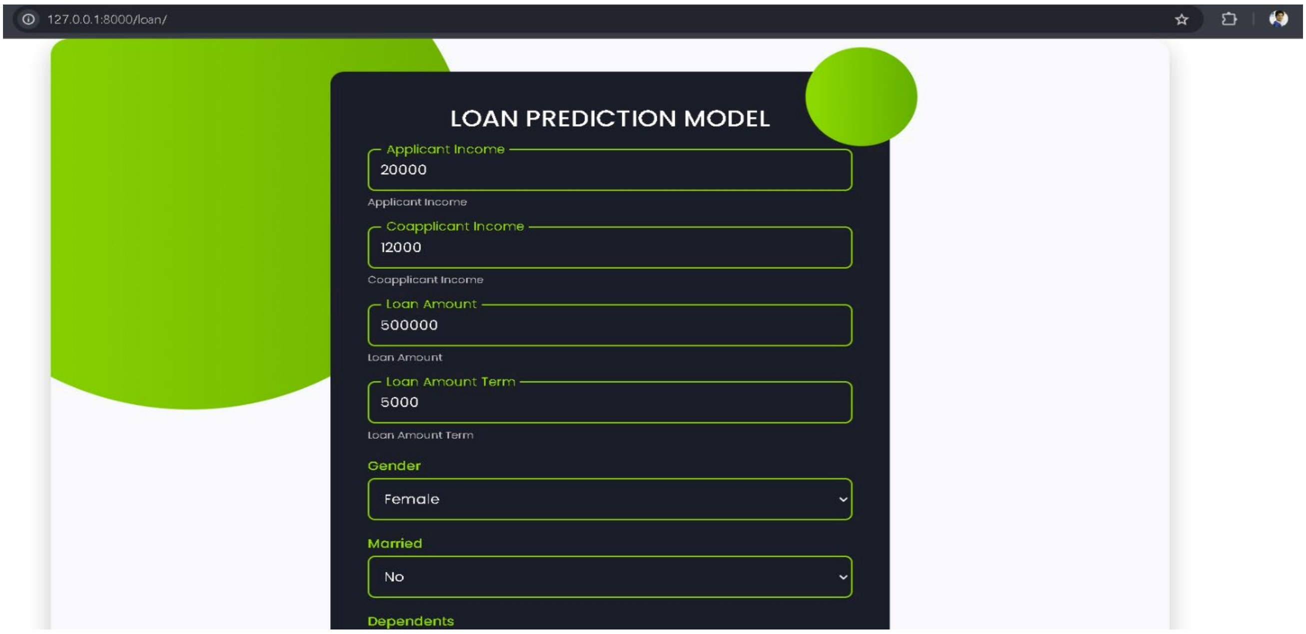

Figure.5 Loan prediction Model Input Interface

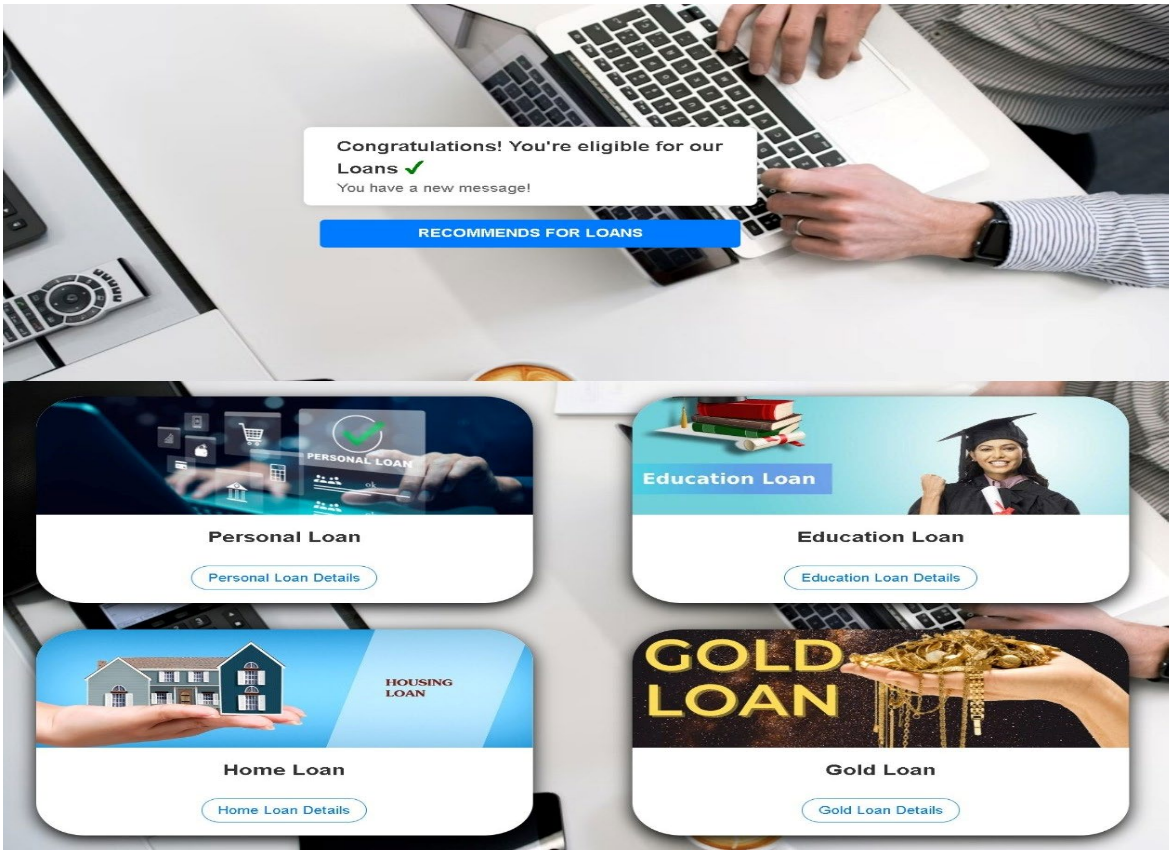

Figure.6 Loan Eligibility Notification

APPENDIX III
JOURNAL PUBLICATION

AI Powered ML Solution for Customer Retention and Loan Decision Optimized for Banking Sector.

Mrs. J. KIRUPAVATHY, M.E., (Assistant Professor), Department of Artificial Intelligence and Data Science, Velammal Institute of Technology, Tiruvallur, India
kirupavathy@velammalitech.edu.in

Ms. DEEPIGHA SHRI CM
Department of Artificial Intelligence and Data Science, Velammal Institute of Technology, Tiruvallur, India
deepighashri@gmail.com

Ms. MAHAPAVITHRA R
Department of Artificial Intelligence and Data Science, Velammal Institute of Technology, Tiruvallur, India
rmahapavithra@gmail.com

Ms. SAKTHI PRIYA P
Department of Artificial Intelligence and Data Science, Velammal Institute of Technology, Tiruvallur, India,
sakthisarange@gmail.com

Abstract*—In the dynamic environment of the banking sector, customer retention and efficient loan decision-making are critical to maintain profitability and cutthroat advantages. This project provides an AI-powered ML solution dwelling both provocations through an integrated platform. The strategy combines customer churn prediction and loan decision optimization into a merged web application, that can be activated easily and managed by the employees in the bank. Using the platform, proactively banks can find the danger prone customers and automate loan acceptance processes with increased accuracy. The key features of the project would be the AI-chatbot that provides the space for real-time reciprocity, allowing employees to handily access churn and loan applications. By employing predictive analytics and NLP, the results enhance user engagement, running cost is reduced and upgrades decision-making productivity. Experimental results demonstrate high model accuracy in churn prediction and loan classification, validating the system's potential to significantly optimize banking operations. This innovative integration of two major banking functions into a single platform offers a novel contribution to the application of AI in the financial sector. Moreover, the proposed system bridges the operational gap typically seen in conventional banking solutions, where churn management and credit decisioning are handled through isolated modules. By unifying these services, the platform not only streamlines workflow but also offers predictive insights that enable banks to adopt more customer-centric strategies. With scalable architecture and real-time responsiveness, the solution is well-positioned to support digital transformation initiative in the banking industry. Future enhancements could involve the integration of dynamic customer behavior analytics and adaptive learning models, further strengthening the system's ability to respond to emerging financial trends.

Keywords—Customer Churn, Loan Decision, AI Chatbot, Banking Analytics, Machine Learning

I. Introduction

The banking industry today is facing unrivalled competition and evolving customer reckoning. Retaining existing customers and efficiently processing loan applications have become vital for maintaining profitability and ensuring long-term growth. However, traditional manual methods for loan decision-making and customer retention strategies are often time-consuming, prone to errors, and lack the ability to adapt to rapidly changing customer behavior and market dynamics. As highlighted in recent studies, manual processes not only delay decisions but also increase operational costs and introduce a higher risk of inaccuracies.[4]

"There is an urgent need for a unified, intelligent platform that can help banks simultaneously predict customer churn risk and optimize loan approval decisions, enabling proactive engagement and smarter credit management". Research has shown that predictive analytics, when applied correctly, can significantly reduce non-performing loans and improve customer loyalty.[2]

Most existing systems either focus only on loan prediction or only on customer retention, requiring multiple platforms and resulting in operational inefficiencies. Additionally, these systems typically offer limited real-time interaction with clients and minimal flexibility for banking employees to manage both areas effectively.

"To address this gap, we propose an Ai-powered machine learning solution that seamlessly integrates customer churn prediction and loan decision optimization into a single, easy-to-use web application".

The platform is designed to be activated and managed by bank employees, allowing them to access predictive insights for both customer retention and loan approvals through a merged dashboard. Recent advancements have demonstrated the power of integrating data-driven insights into decision systems to not only automate but also personalize financial services.[3]

The solution provides a streamlined approach for assessing loan eligibility and identifying at danger prone customers, enabling faster decision-making and proactive customer retention strategies. In the proposed system, machine learning algorithms are employed to ensure accurate predictions. We utilize the *"Extra Trees Classifier, a powerful ensemble learning method known for handling large datasets and complex decision boundaries. Through extensive training* *and testing, the Extra Trees Classifier achieves an impressive accuracy rate of ~93%"* in predicting both customer churn risk and loan approval decisions, ensuring reliable and consistent performance.

Additionally, to enhance user interaction, the platform integrates an *"AI-powered chatbot"* that leverages *"Natural Language Processing"* techniques. While many traditional banking applications lack real-time feedback systems [1], the chatbot allows clients to easily query their loan application status, understand their churn probability, and receive personalized support in real-time. By combining predictive analytics with conversational AI, the system not only reduced the manual workload for bank employees but also significantly improves the overall customer experience.

Thus, the proposed solution offers a novel, efficient and intelligent approach for modern banking institutions to manage critical operations from a single, interactive platform, ultimately leading to better operational efficiency, higher customer satisfaction and competitive advantage.

II. RELATED WORKS

Vijaya Kumar et al. [1] investigated the application of multiple machine learning classifiers- including neural networks, discriminant analysis, naïve Bayes, k-nearest neighbors, logistic regression and ensemble decision trees-to predict individual loan defaults. Their study, conducted on real bank credit data, demonstrated that these models could achieve accuracy rates between 76% and 82%, highlighting the promise of automated loan risk assessment while underscoring the need for improved feature engineering and model robustness.

Victor Chang et al. [2] extended this line of work to credit card customers, comparing a suite of gradient-boosted and deep learning algorithms such as XGBoost, LightGBM, AdaBoost and feed-forward neural networks. Their experiments on a large credit-card dataset showed that XGBoost outperformed other methods, reaching an accuracy of 99.4%, and emphasized the importance of handling class imbalance and hyperparameter tuning for high-stakes financial predictions.

Vikas Kumar et al. [3] proposed an AI-based hybrid framework for mortgage loan risk prediction by combining logistic regression, decision trees and gradient boosting in an ensemble architecture. This approach not only improved prediction accuracy over single-model baselines but also enhanced stability across different market conditions. However, the solution remained focused on backend risk scoring without any front-end, client-facing interaction mechanisms.

Vahid Sinap [4] conducted a relative study of loan approval prediction using machine learning pipelines that integrated feature selection techniques such as recursive feature elimination and k-best ranking with classifiers counting Support Vector Machine, Random Forests and Decision Trees. The research demonstrated that a carefully tuned random forest model, validated via cross-validation, could achieve up to 97.71% delicacy, emphasizing the crucial role of data preprocessing and model evaluation strategies.

Neeraj Kripalani [5] addressed customer retention by developing a predictive analytics framework that analyzes clickstream, usage-pattern and sentiment data to identify users at risk of churn up to six weeks before attrition. His work showed a 15-25% reduction in churn rates through timely, personalized intervention strategies. Despite its effectiveness in retention, the framework operated independently of credit decision processes.

Li et al. [6] proposed the use of machine learning techniques for enhancing customer retention within financial services. Their study emphasized the importance of predictive modeling and customer segmentation in proactively identifying danger prone customers. By applying machine learning classifiers to historical customer behavior data, they successfully reduced churn rates and helped banks tailor personalized retention strategies.

Kumar and Verma [7] conducted a comparative study of AI-based credit scoring and loan decisioning models. They evaluated various machine learning and deep learning algorithms, including Random Forest, XGBoost and Deep Neural Networks (DNN), to predict loan approval outcomes and manage credit risks more effectively. Their findings highlighted that ensemble models such as Random Forest and boosting algorithms achieved superior performance compared to traditional linear models, thus offering greater reliability in credit decision-making.

Zhang and Patel [8] explored the role of conversational AI systems in banking by integrating Natural Language Processing (NLP) techniques. Their research demonstrated that AI-powered chatbots significantly improve customer satisfaction by offering real-time assistance, reducing response times and handling client queries efficiently. They showed that the integration of NLP not only enhanced operational efficiency but also contributed to stronger customer loyalty in digital banking environments.

Gupta and Mehta [9] presented an AI-driven predictive modeling approach specifically targeting customer churn in the banking sector. Their framework employed ensemble learning models such as Random Forest and XGBoost to achieve high churn prediction accuracy. They demonstrated that predictive analytics could significantly enhance customer retention efforts by enabling proactive intervention strategies.

Collectively, these studies illustrate significant advances in discrete banking functions-loan default forecasting [1], credit card risk modeling [2], mortgage ensemble methods [3], optimized approval pipelines [4] and proactive churn management [5] yet none integrate both customer retention and loan decisioning within a unified, real-time interface. Our work fills this gap by combining an Extra Tree Classifier (~93% accuracy) with an AI-driven NLP chatbot in a single web application, enabling seamless, end-to-end management of both critical banking operations. The advancements presented in these studies collectively demonstrate the value of machine learning and AI in modernizing financial services.

III. PROPOSED METHODOLOGY

The planned system is organized into five main modules: Data Collection & preprocessing, feature engineering & Representation, Machine learning Model, Chatbot Integration and Deployment. Each module is described below.

A. Data Collection & Preprocessing

For Data Sources, we aggregate customer account data, transaction histories and demographic information from the bank's databases. Loan application records- containing features such as income, credit history and requested loan amount-re loaded s a tabular dataset.

For Churn Labeling, customers are labeled as churned if they close all active accounts or become inactive for more than six months. The Churn Rate metric is computed over each quarter:

$$Churn\ Rate = \left(\frac{Customers\ Lost}{Customer\ at\ period}\right) \times 100$$

Missing values in numeric fields are imputed using median values; categorical fields use mode imputation. Outliers beyond three standard deviations are capped.

Table 1. Collected and Processed Dataset

1	customer_	credit_sco	country	gender	age	tenure	balance	products_	credit_car	active_me	estimated	churn
2	15634602	619	France	Female	42	2	0	1	1	1	101348.9	1
3	15647311	608	Spain	Female	41	1	83807.86	1	0	1	112542.6	0
4	15619304	502	France	Female	42	8	159660.8	3	1	0	113931.6	1
5	15701354	699	France	Female	39	1	0	2	0	0	93826.63	0
6	15737888	850	Spain	Female	43	2	125510.8	1	1	1	79084.1	0
7	15574012	645	Spain	Male	44	8	113755.8	2	1	0	149756.7	1
8	15592531	822	France	Male	50	7	0	2	1	1	10062.8	0
9	15656148	376	Germany	Female	29	4	115046.7	4	1	0	119346.9	1
10	15792365	501	France	Male	44	4	142051.1	2	0	1	74940.5	0
11	15592389	684	France	Male	27	2	134603.9	1	1	1	71725.73	0
12	15767821	528	France	Male	31	6	102016.7	2	0	0	80181.12	0
13	15737173	497	Spain	Male	24	3	0	2	1	0	76390.01	0
14	15632264	476	France	Female	34	10	0	2	1	0	26260.98	0
15	15691483	549	France	Female	25	5	0	2	0	0	190857.8	0
16	15600882	635	Spain	Female	35	7	0	2	1	1	65951.65	0
17	15643966	616	Germany	Male	45	3	143129.4	2	0	1	64327.26	0
18	15737452	653	Germany	Male	58	1	132602.9	1	1	0	5097.67	1
19	15788218	549	Spain	Female	24	9	0	2	1	1	14406.41	0

B. Feature Engineering & Representation

We normalize continuous variables (e.g., income, tenure) via min-max scaling. Categorical variables (e.g., employment type) are one-hot encoded.

Client queries and support transcripts are vectorized using TF-IDF:

$$tfidf_{t,d} = tf_{t,d} \times \log\left(\frac{N}{df_t}\right),$$

C. Machine Learning Models

We employ two tree-based ensemble classifiers for our core prediction tasks:

For Churn prediction, an Extra Tree Classifier is trained on the preprocessed customer feature set. Hyperparameters-number of estimators T, maximum tree depth and minimum samples per leaf-are optimized via grid search with 5-fold cross-validation. The resulting model attains an average accuracy of approximately 93% on the hold-out churn dataset.

To approve the loan, loan prediction uses Random Forest Classifier is used for credit decisioning, with hyperparameters tuned in an identical manner. This model achieves an accuracy of 91.81% on its hold-out test set.

Algorithm Formulation:

$$P(y = k \mid x) = \frac{1}{T}\sum_{t=1}^{T} 1(h_t(x) = k),$$

and the final class label is chosen as

$$\hat{y} = \arg\max_k P(y = k \mid x)$$

Where is $tf_{t,d}$ the occurrence t in Document d, df_t is the Number of Documents suppress t and N is the Total Number of Documents.

D. Chatbot Integration

User messages are classified into intents (e.g., “loan status”, “churn risk inquiry”) using the TF-IDF-trained Extra Trees model. For “loan status” queries, the chatbot queries the Random Forest service; for churn inquiries, it returns the extra Trees prediction. The chatbot runs as a Django app endpoint, ensuring sub-second response times.

E. Deployment & Security

All components are built using Python programming language with the Django framework and managed via Anaconda Environments. Visual Studio Code (VS Code) is used as the primary Integrated Development Environment (IDE) for coding, testing and project execution.

The development environment was structured to ensure efficiency and flexibility, utilizing Python 3.x within an Anaconda-managed setup for seamless environment activation, package management and dependency isolation. Visual Studio Code (VS Code) served as the primary integrated development environment (IDE) for writing, editing, debugging and executing both the Django server and machine learning components. Additionally, Jupyter Notebooks, accessed via Anaconda, were employed during the early stages of model experimentation and prototyping to facilitate rapid iteration and testing.

The backend services are powered by the Django REST Framework, which hosts three RESTful API endpoints:

1. /api/churn/ for predicting customer churn using the ETC (Extra Tree Classifier).

2. /api/loan/ for loan approval predictions utilizing the Random Forest Classifier.

3. /api/chat/ for enabling real-time chatbot interactions.

On the frontend, Django templates combined with Bootstrap are employed to design the web dashboard, presenting churn scores, loan approval probabilities and integrating the AI chatbot for live support. To ensure security, all web service operate over HTTPS with TLS encryption and sensitive data such as prediction outputs and customer information is protected at rest using AES-256 encryption standards. Access control mechanisms based on user roles are implemented, granting detailed customer data access solely to authorized banking personnel, while clients securely view only their own information.

To further enhance the platform’s performance and reliability, asynchronous processing techniques have been integrated wherever feasible, particularly for background tasks such as model inference, data preprocessing and user session management. This ensures that real-time services like the chatbot interaction and prediction APIs maintain low latency, even during high-traffic periods. Moreover, a modular development approach has been adopted, enabling independent updates or upgrades of individual system components without disrupting the overall functionality.

Continuous integration and deployment (CI/CD) practices have been incorporated into the workflow, allowing for systematic testing and streamlined deployment of new features or security patches. Emphasis has also been placed on scalability, with the backend architecture designed to accommodate future expansions, such as the integration of additional predictive models, enhance analytics dashboards or support for larger user bases, thereby ensuring the platform's adaptability to evolving business needs.

F. Interface

The suggested system donates a unified, web-based interface created with Django framework, offering uninterrupted interaction for both bank employees and clients.

The interface begins with a *"secure authentication system"*, where users log in with authorized credentials. Role-based access control ensures that bank employees have access to all customer predictions and system management features, while clients are restricted to viewing their own churn probabilities and loan application statuses.

Fig 1. Working of the system

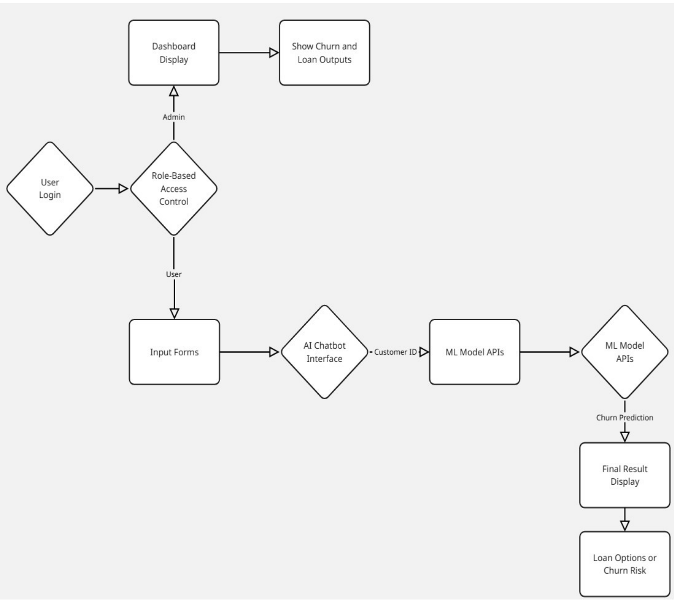

Fig 1. shows the System architecture ease secure user login with role-based access, manage users to input forms and an AI chatbot, while admins access a dashboard. User inputs are processed via ML model APIs to predict churn risk or advise loan options, with results presented intersubjectively.

The *"web dashboard"* serves as the central hub for users. It is designed using Django templates integrated with Bootstrap for responsive layout and ease of navigation. The dashboard displays key outputs from the machine learning models, including the customer's churn risk and loan approval probabilities. These intuitive elements such as color-coded gauges, charts and status panels enabling users to quickly interpret critical information.

The interface provides *"input forms" for* real-time prediction tasks. Employees can search for customer details by entering customer IDs or uploading batch files (CSV format) for bilk churn and loan assessments. Clients can directly apply for loans by filling out simplified application forms and immediately receive their eligibility results based on the Random Forest model prediction.

An important part of the interface is the integration of an *"AI-powered chatbot"*, Accessible through a floating chat widget on all pages, the chatbot uses Natural Language Processing (NLP) to classify user queries into predefined intents such as *"loan status"* or *"churn inquiry"*. Based on the classified intent, the chatbot fetches real-time predictions from the respective APIs and responds to the user with minimal delay, thus enhancing the interactive experience.

IV. Results and Discussions

The development and deployment of the proposed AI-powered banking platform was carried out using Python programming language, with Anaconda serving as the environment management tool. A custom environment named *"loan"* was created within Anaconda to organize all necessary libraries and dependencies required for project execution. For the development of the code and backend setup of API, *"Visual Studio code (VS Code)"* was made used of, providing a clean and powerful IDE.

Meanwhile, *"Jupyter Notebook"* also managed through Anaconda, and it was used for model training, experimentation and hyperparameter tuning of the machine learning algorithms.

In the project, the *"Extra Tree classifier"* was employed for *"customer churn prediction",* while the "Random Forest Classifier" was utilized for *"loan decision optimization"*. The Extra Tree Classifier, an ensemble learning method that aggregates results from multiple randomized decision trees, provides greater generalization and reduces variance compared to standard decision trees.

It was chosen because of its ability to handle large, complex datasets and prevent overfitting through random feature splits. In churn prediction, the Extra Trees model achieved an accuracy of approximately ~93%, enabling the system to proactively identify customers likely to leave the bank.

Fig 2. Entering data for Customer Churn Prediction

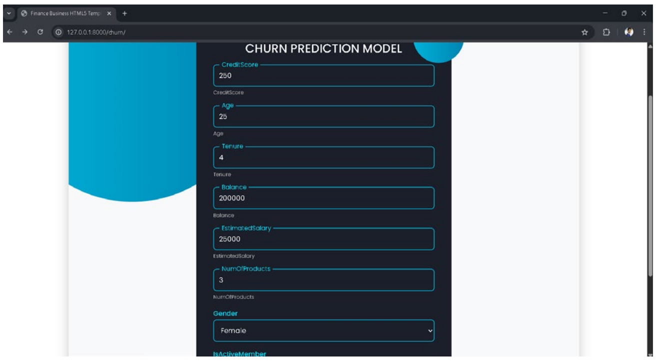

Fig 2. gives the input form of a Churn Prediction Model. Users furnish customer details like credit score, age, balance and gender to predict the likelihood of customer churn utilize a machine learning model.

This predictive capability allows the bank to take targeted retention actions such as personalized offers, improved service communication or loyalty rewards-all based on data-driven insights. The model was seamlessly using RESTful

APIs, allowing real-time access to predictions via the user interface. This integration ensures that bank employees can leverage Ai-driven insights during routine operations without needing technical expertise.

Fig 3. Intention for Churn and Retention implication

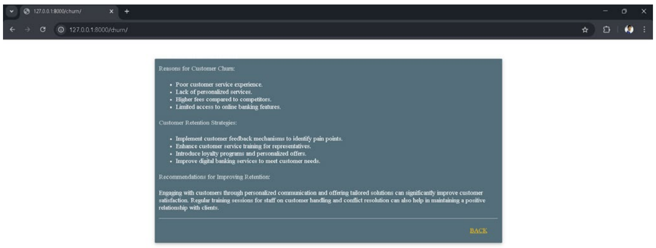

Fig 3. highlights the main motive for the customers to leave a bank, like poor service and increased fees. It also provides plan of actions like loyalty programs and digital service advancement to help conserver the customers

Fig 4. Positive Customer Opinions on Bank Service

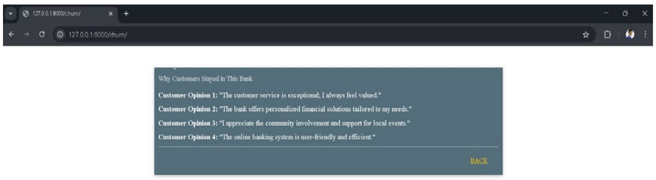

Fig 4. manifest real response from complacent customers. Their comments reflect valuing for personalized support, strong community involvement and a user-friendly online banking system.

For loan decisioning, the *"Random Forest Classifier"* was selected due to its robustness and ability to maintain accuracy even in the presence of missing or noisy data. The Random Forest operates by constructing a multitude of decision trees and outputting the class that is the mode of the classes of the individual trees. It successfully delivered an accuracy of 91.81% when predicting loan approval outcomes, thereby minimizing credit risk for the bank.

This high level of accuracy ensures that loan applications are assessed using consistent, data-driven methods, reducing the chances of human bias or oversight. The model considers a wide range of applicant features-such as credit history, income level, employment status and loan purpose-to make informed and objective predictions.

Fig 5. Loan prediction Model Input Interface

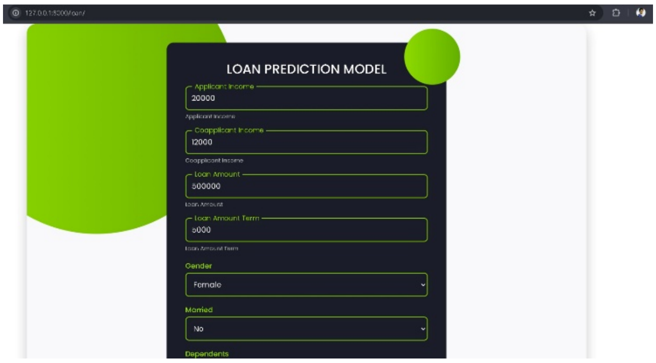

Fig 5. interface allows users to input applicant details like income, loan amount, term, gender, marital status, etc. to predict the approvals.

Fig 6. Loan Eligibility Notification

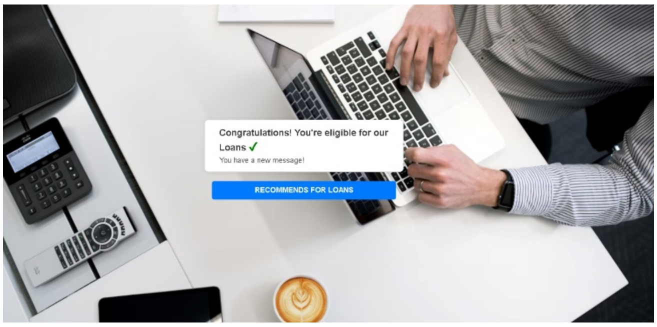

Fig 6. show s that the system successfully recognized eligible users for loan subscriptions based on the input data and Predictive Model Analysis. With this a testimony message is put on view, alerting the user of their eligibility.

Fig 7. Loan option display

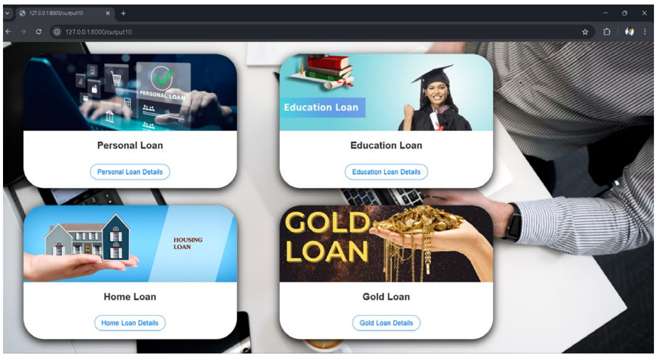

Fig 7. shows the eligible loan varieties-Personal, Educational, Home and Gold after prediction. User can view and choose the required loan through the collaborative interface.

In addition to improving decision accuracy, the model also significantly reduces the time required for processing each loan application, contributing to faster service delivery and improved customer satisfaction. By automating the initial screening phase, bank employees can prioritize more complex or borderline cases, improving resource allocation. The integration of this model into the platform's backend through REST APIs allows real-time predictions to be generated instantly upon user input, ensuring a seamless and interactive user experience.

Fig 8. Loan rejection alert

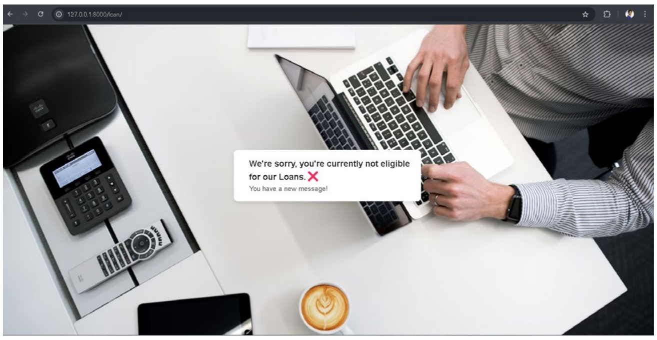

Fig 8. displays the user ineligibility for the loan through the evaluation model and shows the rejection promptly.

Upon successful execution, the system outputs highly intuitive results. If a customer is eligible for a loan based on the model's prediction, the platform automatically navigates them to a page offering detailed loan options such as *"Gold loan, house loan, education loan and personal loan"*. Each option is dynamically customized based on the customer's profile, providing a seamless and customer-centric banking experience.

A crucial enhancement to the platform is the integration of an *"AI-powered chatbot, which plays a pivotal role in improving real-time interaction."*. The chatbot leverages *"Natural Language Processing (NLP)"* techniques to interpret user queries and classify intents accurately.

NLP allows the chatbot to understand natural human language instead of fixed keywords, enabling it to respond intelligently to various inquires like checking loan status or churn risk. The chatbot not only reduces the manual workload for bank employees but also significantly improves customer engagement by offering instant, context-aware responses.

The capability that makes it a valuable tool for providing personalized banking experiences, as it can remember user preference and offer tailored responses or suggestions for loan products based on customer history.

The integration of this AI chatbot results in faster response times, enhanced user satisfaction and more efficient use of bank resources. It can operate 24/7, allowing customers to access banking services at any time, thus improving overall customer retention. As the system evolves, the chatbot could be further enhanced with advanced features, such as voice recognition and multi-language support, making it an even more powerful tool for the bank and its customers.

Fig 9. Loan Chatbot Guidance

Fig 9. show the effective functionally of the Chatbot interface for loan eligibility task. It contributes spontaneous response with the queries specified such as income and credit history.

V. Conclusion and Future Enhancement

Foreword a unified, AI-powered web application meant to rationalize two critical banking functions: customer churn prediction and loan decision automation. By integrating machine learning models- *"Extra Trees Classifier"* for churn prediction and *"Random Forest Classifier"* for loan approval- within a secure, user-friendly Django-based platform, the system enables banking institutions to proactively manage customer retention and credit decisions with high accuracy. The inclusion of an *"AI-powered chatbot"* leveraging *"Natural Language Processing (NLP)"* enhances client interaction and automates responses to common queries in real-time.

This intelligent system not only automates decision-making but also ensures data-driven consistency, reducing human error and improving operational efficiency. It allows banking staff to focus on more strategic tasks while routine predictions and client communications are handled autonomously. Through intuitive dashboards, users can track churn probabilities, loan eligibility, and system performance in a centralized manner.

Experimental results demonstrate that the platform delivers significant accuracy (93% for churn and 91.81% for loan predictions) while maintain scalability and performance. Furthermore, the use of tools like *"Anaconda, Jupyter Notebook and Visual Studio Code"* facilitated modular development, environment isolation and effective testing. The web dashboard and EMI calculators simplify decision-making for both employees and customers, making the platform highly accessible.

"Integration of Fraud Detection Module", Includes a fraud threat classifier would enable a complete credit risk management system.

"Deep Learning Model", restore current model with LSTM or DNNs could brush up predictive strength, mostly for sequential customer behavior data.

"Multi-Language Chatbot Support", supplements the chatbot to succor regional languages for broader convenience.

"Mobile App Version", develops a receptive mobile app version to let banking operations to progress.

"Real-Time Data Pipeline", Incorporates tools like Apache Kafka or Spark Streaming to enable real-time prediction from live transaction data.

These improvements aim to improve the platform from a functional prototype to a scalable, enterprise-grade AI banking assistant capable of transforming customer service and operational intelligence in financial institutions.

References

[1] Vijay Kumar, Rachna Narula and Akanksha Kochhar, "Loan Default Prediction Using Machine Learning Model," Jan 2024.

[2] Victor Chang, Sharuga Sivakulasingam, Hai Wang, Siu Tung Wong, Meghana Ashok Ganatra and Jiabin Luo, "Credit Risk Prediction Using Machine Learning and Deep Learning ," Nov 2024.

[3] Vikas Kumar, Shaiku Shahida Saheb, Preeti, Atif Ghavas, Sunil Kumari, Jai Kishan Chandel, Saroj Kumar Pandey and Santosh Kumar, "AI-based Hybrid Model for Predicting Loan Risk in the Banking Sector," Dec 2023.

[4] Vahid Sinap, "A Comparative Study of Loan Approval Prediciton using Machine Learning Methods," April 2024.

[5] Neeraj Kripalani, "Predictive Analytics for Customer Retention: A Data-Driven Framework for Proactive Engagement and Satisfaction Management," Nov 2024.

[6] Li, Y. and Chen, S., "Machine Learning Techniques for Enhancing Customer Retention in Financial Services," Journal of Financial Data Science,2023.

[7] Kumar, A. and Verma, R., "AI-based Credit Scoring and Loan Decisioning Models: A Comparative study," International Journal of Data Science and Analytics, 2024.

[8] Zhang, L. and Patel, K., "Conversational AI Systems in Banking: Improving Customer Engagement through NLP," IEEE Access, 2023.

[9] Md. Sohel Rana, Anchala Chouksey, Saddam Hossain, Md Sumsuzoha, Proshanta Kumar Bhowmik, Miraz Hossain, MD Fazlay Rabby, Nisha Gurung and MD Abdul Fahim Zeeshan, "AI-Driven Predicitive Modeling for Banking Customer Churn: Insights for the US Financial Sector," Journal of Ecohumanism, Vol. 4, No. 1, 2025, pp. 3478-3497

Gmail

Deepigha Shri <deepighashri@gmail.com>

Paper Selected on GREENCON 2025 - Reg

1 message

Vanitha R <vanithar@care.ac.in> 16 May 2025 at 09:54
To: HOD ECE <hod.ece@care.ac.in>
Bcc: deepighashri@gmail.com

Greetings!

We are pleased to confirm that your Paper has been selected for the **GREENCON 2025: Sustainable Technologies in Computing, Electronics & Electrical Engineering scheduled for May 30 & 31, 2025**

Kindly do the payment on the below link within 20.05.2025

https://forms.gle/ifVEHCHZopJ3tCzQ6

The detailed schedule for the conference will be communicated to you at a later date.

Thank you for your participation and we look forward to your valuable contribution to the conference.

Thanks and Regards,
Vanitha.R ,M.E.,(Ph.D)
Assistant Professor
Department of Electronics and Communication Engineering,
CARE College of Engineering, Trichy - 9.

CARE | **College of Engineering**
care.ac.in | 0431-2690505

Transforming Lives Through Education & Research.

REFERENCE

REFERENCE

[1] Vijay Kumar, Rachna Narula and Akanksha Kochhar, “Loan Default Prediction Using Machine Learning Model,” Jan 2024.

[2] Victor Chang, Sharuga Sivakulasingam, Hai Wang, Siu Tung Wong, Meghana Ashok Ganatra and Jiabin Luo, “Credit Risk Prediction Using Machine Learning and Deep Learning ,” Nov 2024.

[3] Vikas Kumar, Shaiku Shahida Saheb, Preeti, Atif Ghavas, Sunil Kumari, Jai Kishan Chandel, Saroj Kumar Pandey and Santosh Kumar, “AI-based Hybrid Model for Predicting Loan Risk in the Banking Sector,” Dec 2023.

[4] Vahid Sinap, “A Comparative Study of Loan Approval Prediciton using Machine Learning Methods,” April 2024.

[5] Neeraj Kripalani, “Predictive Analytics for Customer Retention: A Data-Driven Framework for Proactive Engagement and Satisfaction Management,” Nov 2024.