# BANK CUSTOMER CHURN PREDICTION

[1]Yamini Chouhan, [2] K.S.Murali Sai Gangadhar, [3]Oddala Manoj, [4]Pavan Kumar Reddy
[1]AssistantProfessor, [234]Students
Department of Computer Engineering(Software Engineering)
Siddhartha Institute of Technology & Sciences, Narapally
yaminichouhan_cse@siddhartha.co.in,23tq1a5602@siddhartha.co.in,
23tq1a5613@siddhartha.co.in, 23tq1a5654@siddhartha.co.in

**Abstarct**

Customer retention is a critical challenge for modern banking institutions, as maintaining existing customers is more cost-effective than acquiring new ones and plays a key role in long-term profitability. Customer churn, which refers to customers leaving a bank's services, can lead to significant financial losses if not addressed proactively. Therefore, predicting customer churn has become an essential task for banks to enhance customer satisfaction and implement effective retention strategies.

This project focuses on developing a machine learning-based system to predict customer churn using various demographic and financial attributes of bank customers. The model analyzes features such as credit score, age, tenure, account balance, number of products, credit card ownership, active membership status, and estimated salary to determine whether a customer is likely to leave the bank.

Multiple machine learning algorithms, including Logistic Regression, Random Forest, and Gradient Boosting, were implemented and evaluated to identify the most effective model for churn prediction. Among these, the Random Forest classifier demonstrated superior performance due to its ability to capture complex relationships within the data and reduce overfitting through ensemble learning techniques.

## I. Introduction

The main problem addressed in this project is the development of an automated system capable of predicting whether a bank customer is likely to churn based on historical data. Customer churn refers to the situation where customers stop using a bank's services, which can significantly impact the bank's revenue and growth. In the highly competitive banking sector, understanding customer behavior and preventing churn has become a crucial challenge.

By applying machine learning algorithms to customer data, it becomes possible to analyze patterns in customer behavior such as transaction history, account usage, and demographic details. These patterns help in identifying customers who are at a higher risk of leaving the bank. The proposed automated prediction system aims to improve accuracy in identifying such customers and assist banks in making data-driven decisions.

This system enables banks to take proactive measures such as offering personalized services, improving customer experience, and implementing targeted retention

strategies. As a result, it helps in reducing customer attrition, increasing customer satisfaction, and improving overall profitability.

## I. Literacture Survey

Several research studies have explored the use of machine learning and data mining techniques for predicting customer churn in banking systems, emphasizing its importance in improving customer retention and business profitability. Singh et al. (2024) conducted a comparative analysis of various machine learning algorithms, including Logistic Regression, Random Forest, and Gradient Boosting, and found that ensemble methods such as Random Forest achieved higher accuracy due to their ability to manage complex and nonlinear relationships within customer data. Similarly, Boozary (2025) focused on advanced analytics for churn prediction using models like Random Forest, XGBoost, and Decision Trees, concluding that ensemble learning techniques significantly enhance prediction performance and model stability. Huang et al. (2023) applied data mining approaches using Support Vector Machines and Neural Networks to analyze customer behavior, identifying key factors such as transaction activity, account balance, and service usage patterns as major contributors to customer churn.

In addition to these studies, recent research has also highlighted the importance of feature engineering, data preprocessing, and model optimization in improving prediction accuracy. Techniques such as handling missing values, encoding categorical variables, and balancing imbalanced datasets using methods like SMOTE have shown to enhance model performance. Furthermore, deep learning approaches and hybrid models combining multiple algorithms are gaining attention for their ability to capture complex behavioral patterns in large-scale banking datasets.

Moreover, recent advancements have introduced the use of real-time analytics and big data technologies to enhance churn prediction systems. The integration of predictive models with customer relationship management (CRM) systems allows banks to identify high-risk customers instantly and take timely actions.

## II. System Analysis

System analysis is the process of studying, understanding, and evaluating an existing system or problem in order to design an improved and efficient solution. It involves identifying system requirements, analyzing current workflows, detecting limitations, and proposing better approaches to achieve desired outcomes. The main goal of system analysis is to ensure that the developed system meets user needs, improves performance, and solves real-world problems effectively.

In the context of software and data-driven applications, system analysis plays a crucial role in defining how data is collected, processed, and utilized. It helps in identifying inputs, outputs, processing methods, and system constraints. For machine learning-based systems, system analysis also includes understanding data sources, preprocessing requirements, model selection, and evaluation techniques.

A well-defined system analysis ensures that the proposed system is efficient, scalable, accurate, and capable of handling real-time scenarios.

**Existing System**

In the traditional banking system, customer churn prediction is mostly handled using basic statistical methods or manual analysis. Banks rely on historical reports, customer complaints, and simple rule-based approaches to identify customers who may leave their services. These methods often depend on limited data and human intuition rather than advanced analytics. In many cases, banks take action only after customers have already churned, making the process reactive instead of proactive. Additionally, the lack of automation makes it difficult to process large volumes of customer data efficiently.

**Disadvantages of Existing System**

- Lack of automation in analyzing customer data
- Low prediction accuracy due to traditional methods
- Inability to handle large and complex datasets
- Reactive approach instead of proactive churn prevention
- Limited use of customer behavioral data
- Time-consuming and prone to human errors

**Proposed System**

The proposed system is a machine learning-based customer churn prediction system designed to analyze customer data and predict the likelihood of churn. It utilizes various features such as customer demographics, financial details, and account activity to build predictive models. The system applies advanced algorithms like Logistic Regression, Random Forest, and Gradient Boosting to improve prediction accuracy.

The system includes data preprocessing steps such as handling missing values, encoding categorical variables, normalization, and feature selection. After training, the model can automatically classify customers as likely to churn or not. This enables banks to identify high-risk customers in advance and take proactive actions such as personalized offers, improved customer service, and targeted retention strategies. The system can also be integrated with existing banking systems to provide real-time insights.

**Advantages of Proposed System**

- Improved accuracy using machine learning techniques
- Automated processing of large-scale customer data
- Early identification of customers at risk of churn
- Enables proactive decision-making and retention strategies
- Reduces manual effort and human errors
- Capable of analyzing complex and hidden patterns in data
- Scalable and adaptable to large banking systems
- Enhances customer satisfaction and business profitability

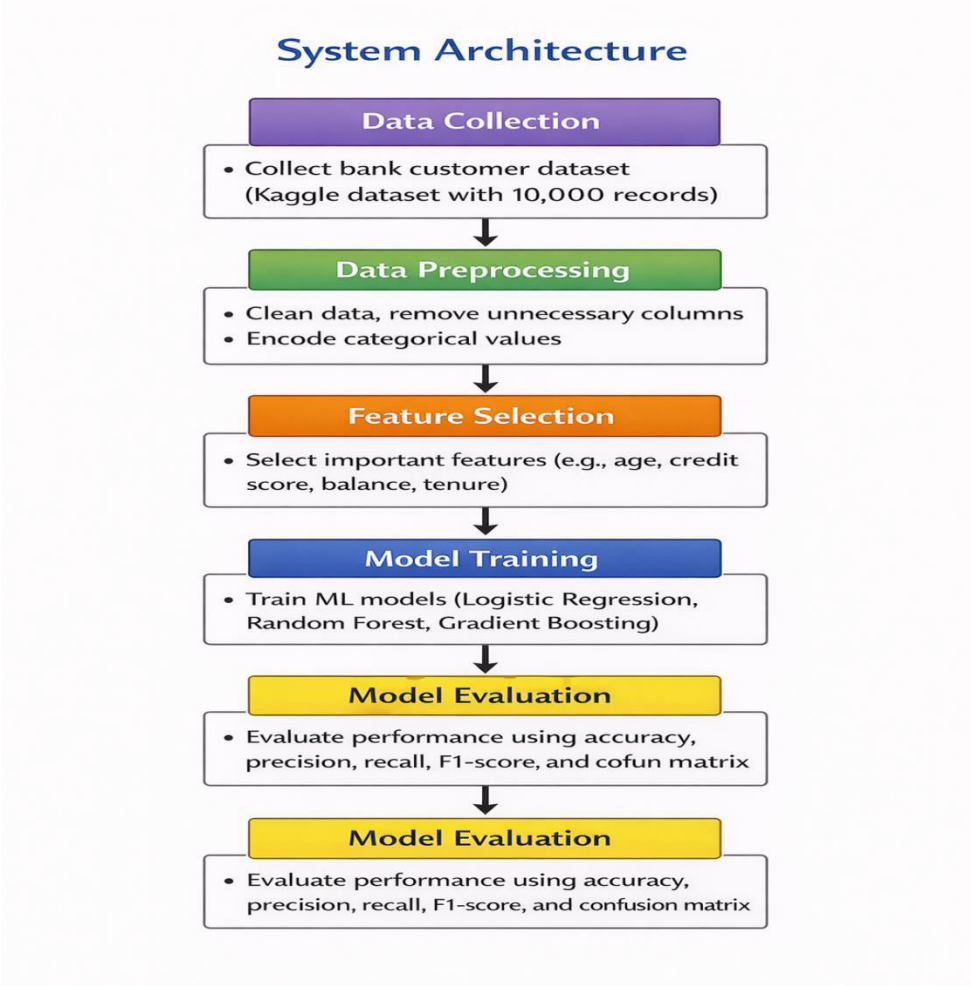
## III. Methodology

The methodology of this project involves a systematic process of collecting, processing, and analyzing customer data to build an accurate churn prediction model. The overall approach is divided into several stages, ensuring efficient data handling and reliable model performance.

Initially, the dataset is collected, which contains customer information such as demographic details, account information, and transaction-related attributes. This data serves as the foundation for the entire prediction system. Once the data is collected, the next step is data preprocessing, where missing values are handled, irrelevant features are removed, and categorical variables are converted into numerical form using encoding techniques. Data normalization and scaling may also be applied to improve model performance.

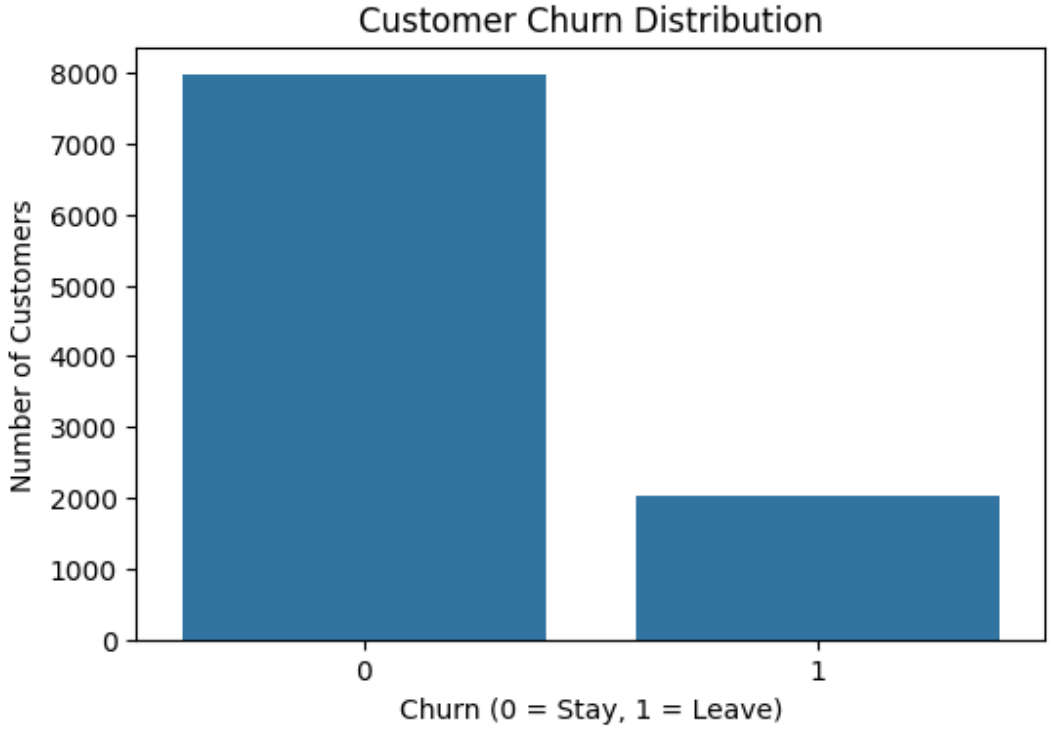
### System Architecture

The system architecture of the Bank Customer Churn Prediction system represents the overall structure and workflow of how data is collected, processed, and used to generate predictions. It consists of multiple layers that work together to provide accurate and efficient churn prediction.

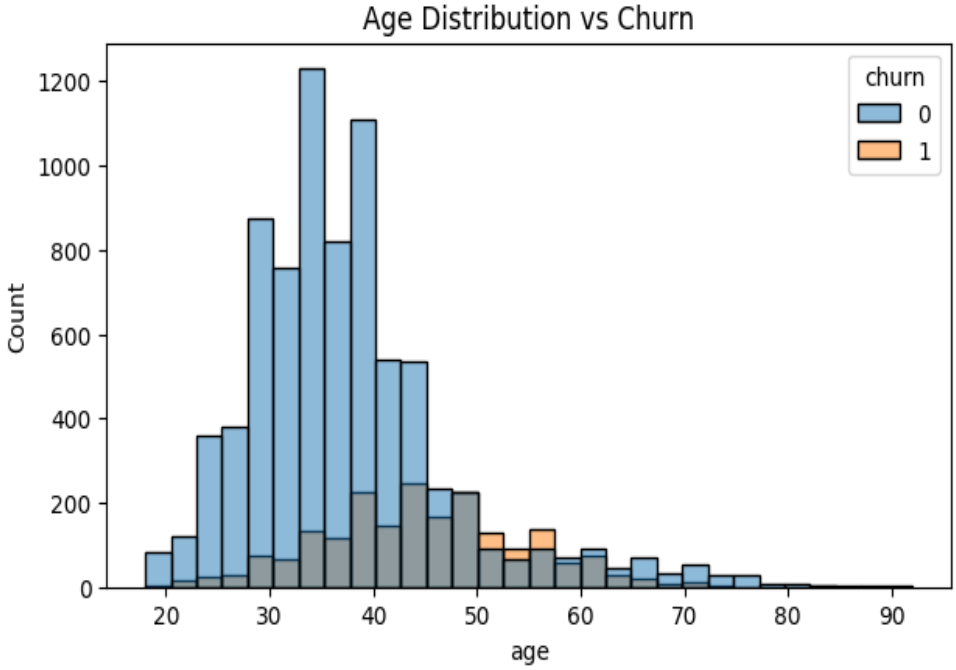


## IV. Result and Output

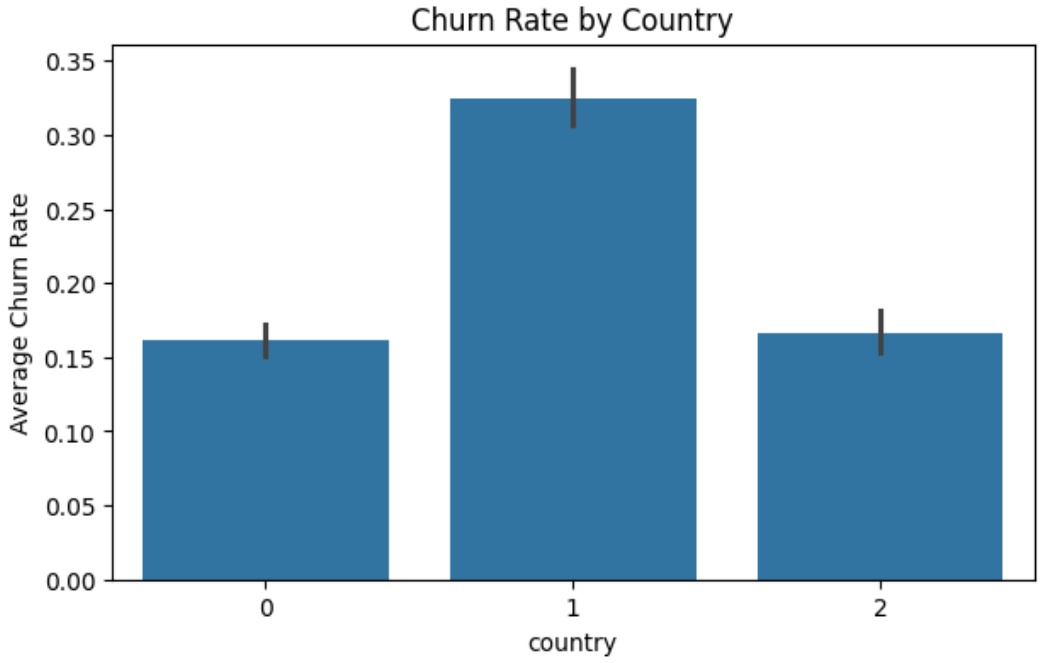
**Customer Churn Distribution :** The first analysis focuses on the distribution of the target variable **Churn**. A count plot was generated to visualize the number of customers who stayed with the bank and those who left.



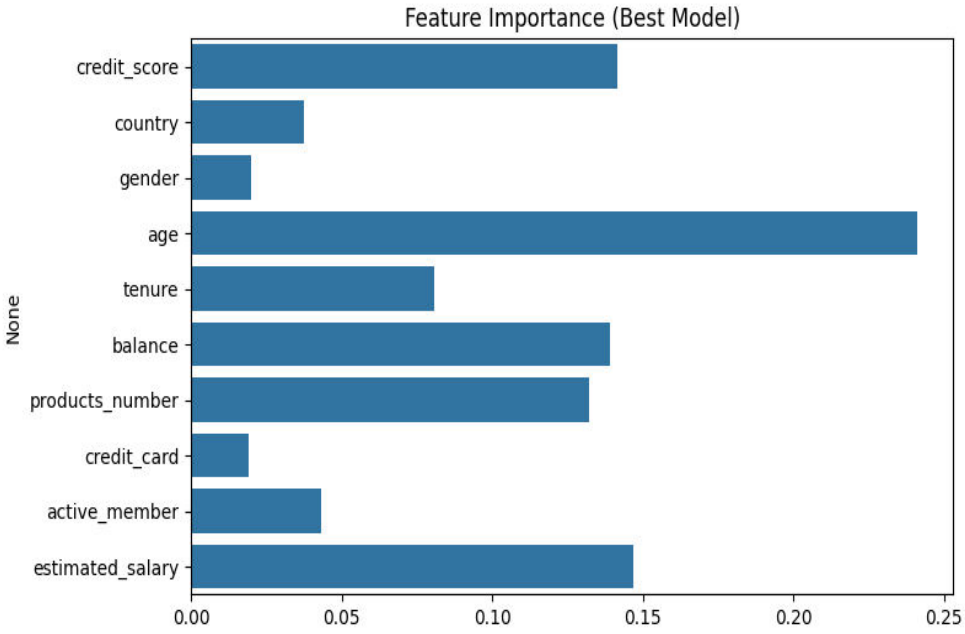

**Age Distribution with Respect to Churn :** A histogram was created to analyze the distribution of customer age and its relationship with churn behavior.



**Customer churn rate by country :** A bar chart was generated to analyze how churn behavior varies across different countries. This visualization helps identify geographical patterns in customer retention.




**Feature Important Analysis :** Feature importance was calculated using the Random Forest classifier to identify which attributes contribute most to predicting customer churn.



## V. Conclusion :

In conclusion, this project successfully developed a machine learning-based system for predicting bank customer churn using both demographic and financial data. The system followed a well-structured pipeline consisting of data preprocessing, exploratory data analysis, model training, and performance evaluation, ensuring accurate and reliable predictions.

Multiple machine learning algorithms, including Logistic Regression, Random Forest, and Gradient Boosting, were implemented and compared. Among these, the Random Forest classifier achieved the best performance, demonstrating higher accuracy and robustness in predicting customer churn. Visualization and analysis of the dataset helped in identifying key factors influencing churn, such as customer age, account balance, credit score, number of bank products, and active membership status.

The results clearly indicate that machine learning techniques are highly effective in analyzing customer behavior and identifying churn patterns. By leveraging such predictive systems, banks can detect high-risk customers in advance and take proactive measures, such as personalized services and targeted retention strategies, to reduce customer attrition.

## References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring
COVID-19 through intensive investigation with supervised machine learning algorithm. In
Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.
[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November).
Machine Learning Techniques in Cloud Based Intrusion Detection. In The International
Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer
Nature Switzerland.
[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on
colorectal cancer with SMOTE on machine learning Algorithm
[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve
1Professor, Department of computer Science & engineering, Anurag University, TS, India.
2Student, Department of computer Science & engineering, Anurag University, TS, India.
[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time

Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT,
Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6,
doi: 10.1109/ICICAT68430.2025.11414670.
[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis
of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial
Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications
in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer,
2025, doi: 10.1007/978-3-031-88304-0_79.
[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting
fake documents from land records using blockchain technology," in Blockchain for Smart
Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.
**[8]** Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," *International Research
Journal of Modernization in Engineering Technology and Science*, vol. 7, no. 7, July 2025,
doi: 10.56726/IRJMETS81618.
**[9] Ravi Kumar Banoth, Ramana Murthy B V**, "Automatic crop recommendation system
using LightGBM and decision tree machine learning models," *Journal of Machine and
Computing*, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
**[10] Ravi Kumar Banoth**, **Dr. B.V. Ramana Murthy**, "Smart agriculture through IoT and
machine learning for analyzing carbon footprints," in *Proc. Int. Conf. Computer Science and
Communication Engineering (ICCSCE)*, Apr. 2025.**[11]** Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer
learning approach: MobileNetV2 with CNN," *SN Computer Science*, vol. 5, art. no. 199,
2024, doi: 10.1007/s42979-023-02500-x.