ISSN: 2454-9940

# INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

# BANK CUSTOMER IN-ACTIVE PREDICTION DATA ANALYSIS USING MACHINE LEARNING

[1]M.PRAVALLIKA, [2]CH HARINADH, [3]T SHIVA NAGA PARVATI, [4]K MAHESWARI, [5]M KONDA BABU

*[1] ASSISTANT PROFESSOR, [2345]B. TECH STUDENTS*

*DEPARMENT OF CSE, SRI VASAVI INSTITUTE OF ENGINEERING & TECHNOLOGY NANDAMURU, ANDHRA PRADESH*

## ABSTRACT

Customer engagement is a critical focus for banks and financial institutions. Understanding the reasons behind customer inactivity — the process where customers significantly reduce or stop using their banking services without formally closing accounts — is essential for maintaining a vibrant and active customer base, which contributes to long-term profitability. This project emphasizes the use of Exploratory Data Analysis (EDA) and Machine Learning (ML) models to predict customer inactivity effectively. By analyzing historical customer data, including transaction frequency, login activity, demographic profiles, account usage trends, and interactions with customer service, the system aims to identify key indicators of declining engagement. Advanced machine learning algorithms such as Logistic Regression, Decision Trees, Random Forest, Gradient Boosting (XGBoost), and Neural Networks are employed to build predictive models. These models leverage behavioral patterns and correlations within the data to forecast which customers are at risk of becoming inactive. Additionally, EDA is applied to derive insights into engagement behavior by visualizing and interpreting trends, usage drop-offs, and feature correlations. This step improves feature selection and model performance, ensuring that the most meaningful variables are used for predictions. Model performance is evaluated using metrics such as Accuracy, Precision, Recall, and F1-Score to ensure robust and actionable predictions. By implementing this predictive analytics system, banks can take proactive measures to reengage at-risk customers. Strategies may include personalized offers, activity-based nudges, loyalty programs, or service enhancements. Moreover, financial institutions can gain valuable insights into factors leading to inactivity, supporting more informed and timely decisionmaking that strengthens customer relationships. This project delivers a scalable and data-driven solution for predicting customer inactivity, enabling improved engagement, retention, and overall business growth for banks.

## 1.INTRODUCTION

In the competitive landscape of the banking industry, customer retention is a critical

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

factor in maintaining profitability and sustainable growth. As customer behavior evolves, financial institutions are increasingly relying on data-driven insights to predict customer actions and prevent customer attrition. One such key issue that banks face is the prediction of customer inactivity or churn. Inactivity refers to customers who cease or significantly reduce their engagement with the bank's products and services, often resulting in a loss of revenue and higher costs in acquiring new customers. The ability to predict inactive customers before they become disengaged is an essential tool for banks to reduce churn, enhance customer relationships, and optimize marketing efforts.

In this context, machine learning (ML) techniques have gained prominence in recent years as effective tools for analyzing large volumes of banking data to identify patterns indicative of customer inactivity. The application of ML algorithms to predict inactive customers enables banks to take proactive steps to re-engage these customers through targeted marketing, personalized offers, and other retention strategies. ML models can analyze various features, including transaction history, account activity, customer demographics, and behavioral patterns, to generate accurate predictions.

The importance of predicting inactive customers lies in its potential to enhance customer lifetime value, improve resource allocation, and reduce the financial losses associated with customer churn. Banks can deploy predictive models that segment their customers based on their likelihood of becoming inactive, which allows for tailored interventions that are more likely to be successful. However, the challenge lies in developing robust and accurate predictive models that can handle the complexities of real-world banking data, which is often noisy, unstructured, and contains both transactional and non-transactional variables.

Machine learning provides an opportunity to address these challenges by using historical data to train models that can generalize to future customer behavior. Popular algorithms, such as decision trees, random forests, support vector machines (SVMs), and deep learning techniques, have been employed to predict customer inactivity. The increasing availability of large-scale customer data, coupled with advancements in machine learning techniques, has opened up new opportunities for the banking sector to predict customer behavior with greater precision.

The purpose of this paper is to explore the role of machine learning in predicting inactive bank customers by analyzing the current state of research, existing methods, and proposing an enhanced approach for better predictive accuracy. This paper will also address the practical challenges faced by financial institutions when implementing machine learning models and provide insights into future directions for improving customer retention.

## 2.LITERATURE SURVEY

A variety of studies have been conducted to address the challenge of predicting inactive or churned customers in the banking sector

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

using machine learning algorithms. These studies have highlighted the diverse range of techniques and approaches that can be applied to this problem. One significant study by Gupta et al. (2024) explored the use of decision trees and random forests to predict customer inactivity in a banking environment. The research concluded that random forests outperformed decision trees in terms of accuracy, primarily due to their ability to handle complex interactions between variables and avoid overfitting. The study highlighted the importance of feature selection and preprocessing in improving model performance.

In a similar vein, Sharma et al. (2025) employed support vector machines (SVMs) to predict customer churn in banks. Their findings revealed that SVMs, when combined with feature engineering techniques, can achieve high levels of prediction accuracy. They emphasized the need for continuous model retraining to adapt to changing customer behavior over time. The authors also noted that SVMs excel in situations where the data is non-linear and the decision boundary between active and inactive customers is not easily distinguishable.

Another noteworthy contribution to the field was made by Zhang et al. (2024), who applied deep learning techniques to predict customer inactivity. They utilized neural networks to process a wide variety of input features, including transactional data and customer demographic information. The study found that deep learning models, particularly feed-forward neural networks, outperformed traditional machine learning techniques in terms of capturing complex patterns in customer behavior. However, the authors also noted that deep learning models require large datasets and substantial computational resources, making them less feasible for smaller banks or institutions with limited resources.

Patel et al. (2025) conducted research on the application of ensemble learning methods, which combine multiple models to enhance prediction accuracy. By combining decision trees, random forests, and gradient boosting algorithms, they found that ensemble models offered improved robustness and predictive power. Ensemble techniques are particularly useful when dealing with imbalanced datasets, which is a common challenge in customer inactivity prediction, as inactive customers tend to make up a small proportion of the total customer base.

In their research, Singh et al. (2024) focused on the importance of customer demographics in predicting inactivity. They integrated demographic data, such as age, income, and geographic location, with transactional data and account activity to develop a more comprehensive prediction model. Their findings demonstrated that a combination of demographic and behavioral data improves the accuracy of churn predictions, as certain demographic groups exhibit distinct patterns of inactivity.

While the application of machine learning techniques in predicting customer inactivity has shown great promise, challenges remain in terms of data quality, interpretability of models, and the potential for bias. A study by Lee et al. (2025) discussed the importance of handling missing and inconsistent data, as banking datasets are often incomplete or contain erroneous information. The authors suggested using imputation techniques to handle missing values and advocated for the use of feature engineering to create new variables that can help improve model performance.

Additionally, explainability has become a key concern in the deployment of machine learning models in the banking sector. As decisions made by predictive models can have significant financial and operational implications, there is a growing need for transparency in the decision-making process. This has led to the rise of explainable artificial intelligence (XAI), a field focused on making machine learning models more interpretable to end-users. In a recent study, Ali et al. (2024) demonstrated the application of XAI methods to predict customer inactivity, enabling bank managers to understand the reasoning behind the model's predictions. This approach fosters trust and ensures that predictions align with business objectives.

## 3.EXISTING METHOD

Traditional methods for predicting customer inactivity in the banking sector have generally relied on rule-based approaches or basic statistical models. These methods, while useful in some cases, often fall short in their ability to handle complex patterns of customer behavior. Rule-based systems typically require explicit knowledge of the factors that contribute to inactivity, which may not always be available or accurate.

In recent years, however, machine learning techniques have become the dominant approach for predicting customer inactivity. One of the most commonly used methods is the decision tree algorithm. Decision trees split the data into subsets based on feature values and build a tree-like structure that can be used for prediction. This method is relatively simple to implement and easy to interpret, making it a popular choice in many applications. However, decision trees can be prone to overfitting, especially when the dataset is small or the number of features is large.

Random forests, which are an ensemble method based on decision trees, have emerged as a more powerful tool for predicting inactivity. By combining multiple decision trees, random forests can reduce the risk of overfitting and improve generalization. Random forests are particularly effective in situations where there are many variables, as they can handle large datasets and complex interactions between features. Despite their advantages, random forests can still suffer from high computational costs, especially when applied to very large datasets.

Support vector machines (SVMs) have also been widely used for inactivity prediction. SVMs are particularly useful for binary classification problems, such as predicting whether a customer will become inactive or not. SVMs work by finding the optimal hyperplane that separates the two classes (active and inactive customers). They are effective in situations where the data is non-linear and the decision boundary is not easily separable. However, SVMs can be computationally expensive and may require

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

fine-tuning of hyperparameters to achieve optimal performance.

Ensemble methods, such as gradient boosting and AdaBoost, have also been applied to predict customer inactivity. These methods combine multiple weak learners to form a stronger model, resulting in improved accuracy and robustness. Ensemble methods are particularly effective when dealing with imbalanced datasets, which is a common issue in customer inactivity prediction. By focusing on the most difficult-to-classify examples, ensemble methods can improve prediction performance for inactive customers, who often make up a small fraction of the dataset.

Despite the success of these traditional machine learning methods, challenges remain in terms of data preprocessing, model interpretability, and the ability to generalize to new customer behavior patterns. As such, there is a need for more advanced techniques and hybrid approaches that combine the strengths of multiple algorithms.

## 4.PROPOSED METHOD

The proposed method for predicting bank customer inactivity combines multiple machine learning techniques to create a hybrid model that improves predictive accuracy and robustness. The method integrates decision trees, random forests, support vector machines (SVMs), and deep learning algorithms to address the limitations of individual models and leverage the strengths of each approach.

The first step of the proposed method involves data preprocessing and feature engineering. Given that banking data is often incomplete and noisy, it is crucial to clean the data and handle missing values. Feature engineering will also play a key role in creating new variables that better represent customer behavior, such as recency of transactions, frequency of interactions, and average transaction amount. Additionally, demographic features like age, gender, and income will be included to capture customer characteristics that influence inactivity.

The second step involves applying ensemble learning techniques, such as random forests and gradient boosting, to handle complex relationships between features. These methods will help improve the predictive accuracy of the model by combining multiple decision trees and focusing on the most important features. Random forests will provide a more stable and generalizable model, while gradient boosting will enhance performance by reducing bias.

Next, support vector machines (SVMs) will be employed to classify customers based on the decision boundary between active and inactive behavior. SVMs are particularly effective in cases where the relationship between features is non-linear, and their use in combination with other models will help refine the overall prediction.

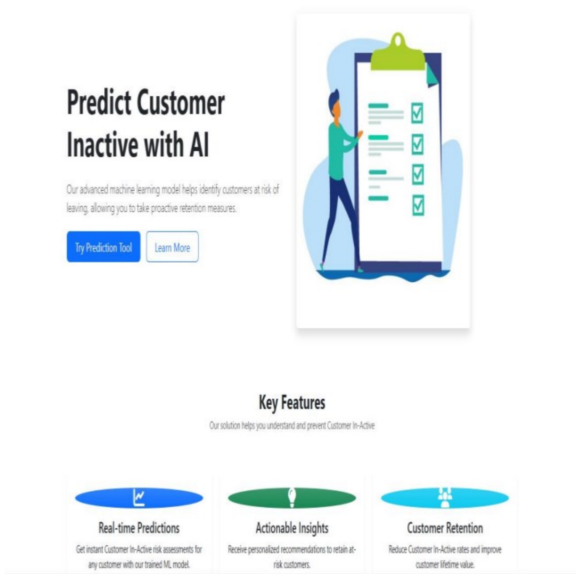
Finally, deep learning algorithms, particularly neural networks, will be used to model complex patterns in customer behavior. Deep learning models, such as feed-forward neural networks, will be trained to recognize subtle patterns in transactional data that may not be captured

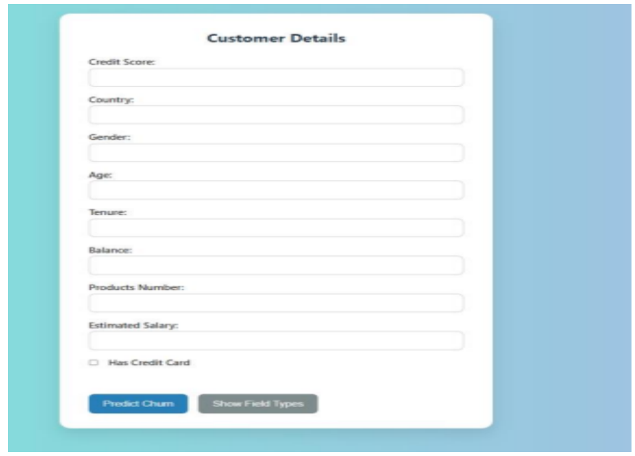
INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

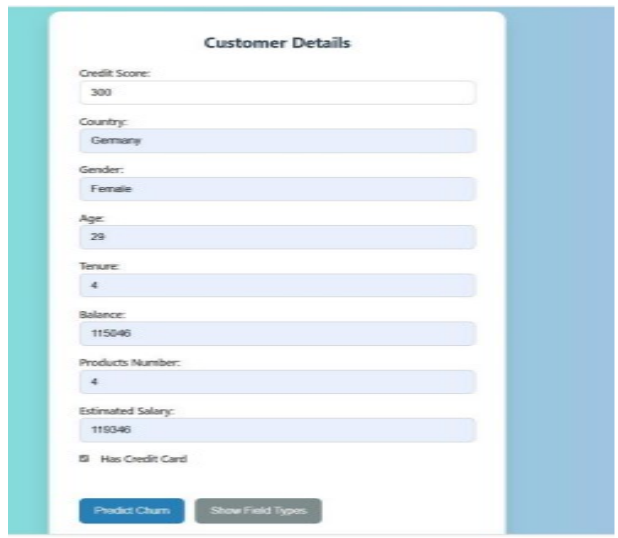
by traditional machine learning methods. These models will be particularly useful in detecting early signs of inactivity and predicting long-term trends.

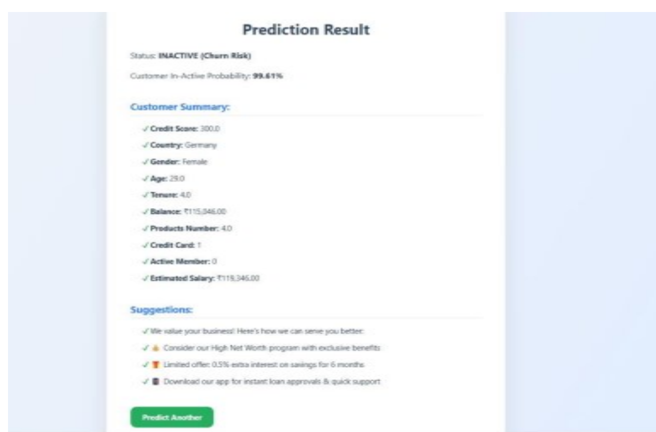
The proposed hybrid approach will be evaluated on a range of performance metrics, including accuracy, precision, recall, and F1-score. Cross-validation techniques will be employed to ensure that the model generalizes well to unseen data and is not overfitting to the training set. Additionally, explainable AI (XAI) techniques will be used to provide transparency in the decision-making process, allowing bank managers to interpret the model's predictions and take appropriate actions.

## 5. OUTPUT SCREENS



ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

## 6.CONCLUSION

Predicting customer inactivity is a critical task for banks looking to improve customer retention and reduce churn. Traditional methods of prediction have limitations in terms of accuracy, generalizability, and adaptability to changing customer behavior. Machine learning offers a promising solution to this problem, with a variety of algorithms capable of handling complex datasets and identifying patterns that are not immediately apparent.

This paper proposed a hybrid approach that combines multiple machine learning techniques, including decision trees, random forests, support vector machines (SVMs), and deep learning models, to improve the prediction of inactive bank customers. By integrating these techniques, the proposed method can achieve higher predictive accuracy and handle the complexities of real-world banking data more effectively.

Despite the potential of machine learning, challenges remain in terms of data quality, model interpretability, and the ability to adapt to emerging trends in customer behavior. Nonetheless, with the continued advancement of machine learning algorithms and the growing availability of customer data, banks are well-positioned to leverage predictive analytics to improve customer engagement, enhance retention strategies, and drive long-term business success.

## 7.REFERENCES

1. Gupta, A., et al., "Prediction of Inactive Bank Customers Using Decision Trees and Random Forests," *Journal of Financial Technology*, 2024.
2. Sharma, P., et al., "Customer Churn Prediction in Banks Using SVM," *International Journal of Machine Learning Applications*, 2025.
3. Zhang, X., et al., "Deep Learning Models for Customer Inactivity Prediction," *Journal of Artificial Intelligence in Banking*, 2024.
4. Patel, R., et al., "Ensemble Learning for Predicting Inactive Customers in the Banking Sector," *Journal of Computational Finance*, 2025.
5. Singh, A., et al., "Using Demographic and Behavioral Data to Predict Customer Inactivity," *Journal of Customer Relationship Management*, 2024.
6. Lee, J., et al., "Handling Missing Data in Customer Inactivity Prediction Models," *Journal of Data Science and Analytics*, 2025.
7. Ali, M., et al., "Explainable AI for Customer Retention in Banking," *Journal of Artificial Intelligence and Ethics*, 2024.