

**ADVANCED CUSTOMER CHURN PREDICTION IN BANKING USING USING
CLASSICAL AND QUANTUM LEARNING MODELS****Mr.N.V.Sagar**Assistant Professor, Department of Computer Science and Engineering,
J.B Institute of Engineering and Technology, Moinabad**Anga Ruchitha, Ganne Shivani, Dande Akshara**UG Students, ¹²³⁴Department of Computer Science and Engineering,
J.B Institute of Engineering and Technology, Moinabad**ABSTRACT**

Customer churn prediction has become a crucial task in the banking sector due to increasing competition and the need for customer retention. This study presents a comprehensive approach for predicting customer churn using Machine Learning (ML), Deep Learning (DL), and Quantum Machine Learning (QML) techniques. Traditional statistical approaches often fail to capture complex customer behavior patterns. Therefore, this work utilizes advanced algorithms such as Logistic Regression, Random Forest, Gradient Boosting, Artificial Neural Networks, and Variational Quantum Classifiers.

The proposed system processes customer demographic, transactional, and behavioral data to predict churn probability. Experimental results demonstrate that ensemble-based ML models and deep learning techniques provide high predictive accuracy, while QML offers a future-oriented exploratory approach. The system enables banks to identify high-risk customers and implement targeted retention strategies, thereby improving profitability and customer satisfaction.

INTRODUCTION

The banking and financial services industry has undergone a significant transformation with the rise of digital technologies, increasing competition, and evolving customer expectations. In such a dynamic environment, customer retention has become a crucial factor for ensuring long-term profitability and sustainability. Customer churn refers to the phenomenon where customers discontinue their relationship with a bank by closing accounts or switching to alternative financial institutions. High churn rates not only lead to direct revenue loss but also increase the cost associated with acquiring new customers. Therefore, predicting and preventing customer churn has emerged as a key strategic priority for banks.

With the rapid adoption of online banking systems, mobile applications, and digital payment platforms, banks generate massive volumes of customer data daily. This data includes demographic information, transaction history, credit behavior, account usage patterns, and customer interaction records. Traditional rule-based systems and statistical methods are often insufficient to analyze such large and complex datasets effectively. As a result, advanced data-driven techniques such as Machine Learning (ML) and Deep Learning (DL) are increasingly used to uncover hidden patterns and predict customer behavior with higher accuracy.

Machine Learning techniques, including Logistic Regression, Decision Trees, Random Forests, and Gradient Boosting, are widely applied for churn prediction because of their ability to handle structured data efficiently and provide interpretable results. These models learn from historical customer data and identify relationships between various attributes and churn outcomes. However, classical ML models may struggle with capturing highly complex and non-linear relationships without extensive feature engineering.

Deep Learning models, particularly Artificial Neural Networks (ANNs), overcome this limitation by learning hierarchical feature representations through multiple layers. These models automatically identify complex interactions between variables, leading to improved predictive performance. Techniques such as dropout, regularization, and hyperparameter tuning further enhance model accuracy and generalization.

In addition to ML and DL, Quantum Machine Learning (QML) has emerged as an innovative research area that combines quantum computing principles with machine learning algorithms. Quantum computing leverages phenomena such as superposition and entanglement to process information more efficiently than classical systems for certain tasks. Although still in its early stages, QML provides promising opportunities for solving complex optimization and classification problems.

This project integrates ML, DL, and QML approaches into a unified framework for bank churn prediction. It follows an end-to-end pipeline including data collection, preprocessing, feature engineering, model training, evaluation, and comparison. The study not only focuses on achieving high prediction accuracy but also emphasizes interpretability, scalability, and practical applicability in real-world banking systems. Overall, the project highlights the importance of intelligent predictive analytics in enhancing customer retention strategies and improving decision-making processes in the banking sector.

RELATED WORK

Customer churn prediction has been extensively studied in the banking and financial services domain due to its direct impact on customer retention, profitability, and long-term business sustainability. Early research in this field primarily focused on traditional statistical methods and data mining techniques, which aimed to identify patterns in historical customer data. These approaches provided foundational insights into customer behavior but were limited in handling large-scale datasets and complex feature interactions. With the advancement of computational capabilities, Machine Learning (ML) techniques have become the dominant approach for churn prediction. Logistic Regression has been widely used as a baseline model due to its simplicity, efficiency, and interpretability in binary classification problems. Decision Tree-based models further enhanced predictive performance by capturing non-linear relationships and handling mixed data types effectively. Ensemble methods such as Random Forest and Gradient Boosting have been shown to outperform individual models by combining multiple learners to reduce variance and improve generalization.

Several studies have emphasized the importance of handling class imbalance in churn datasets, where churned customers typically represent a minority class. Research by Dal Pozzolo et al. demonstrated that standard classifiers tend to be biased toward the majority class if imbalance is not properly addressed. Techniques such as resampling, cost-sensitive learning, and ensemble approaches were proposed to improve recall and reduce false negatives, which is critical in identifying potential churners. In addition to classical ML models, other algorithms such as Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN) have been explored for churn prediction. While these models can perform well under certain conditions, their effectiveness depends heavily on parameter tuning, feature scaling, and dataset characteristics. Overall, machine learning approaches have demonstrated strong performance when combined with effective preprocessing and feature engineering techniques. The emergence of Deep Learning (DL) has further advanced churn prediction capabilities by enabling models to learn complex and non-linear relationships in high-dimensional data. Artificial Neural Networks (ANNs) have been widely applied in financial prediction tasks due to their ability to automatically learn feature representations. Studies have shown that deep learning models achieve higher accuracy and recall compared to traditional ML models, particularly when large datasets are available. However, these models often suffer from reduced interpretability, making it difficult for financial institutions to understand the reasoning behind predictions.

PROBLEM STATEMENT

Customer churn is a major concern in the banking sector, as it directly impacts revenue, customer lifetime value, and overall business growth. Retaining existing customers is more cost-effective than acquiring new ones, yet banks face increasing difficulty in maintaining customer loyalty due to intense competition and the availability of better alternatives. With the rise of digital banking services, customers can easily switch between financial institutions, making churn prediction a critical task for banks.

Banks generate large volumes of customer data, including demographic details, transaction history, credit behavior, and account activity. However, traditional statistical and rule-based approaches are often analyzing such complex and high-dimensional data. These methods struggle to identify hidden patterns and non-linear relationships that influence customer decisions. As a result, banks are unable to proactively identify customers at risk of leaving and implement effective retention strategies.

Therefore, there is a need for an advanced, data-driven solution that can accurately predict customer churn. This project aims to develop a robust churn prediction system using Machine Learning, Deep Learning, and Quantum Machine Learning techniques to enhance predictive accuracy, support decision-making, and improve customer retention strategies.

PROPOSED SYSTEM

The proposed system aims to develop an intelligent, scalable, and data-driven solution for predicting customer churn in the banking sector using a combination of Machine Learning (ML), Deep Learning (DL), and Quantum

Machine Learning (QML) techniques. The system is designed to analyze large volumes of structured banking data and identify customers who are likely to discontinue their relationship with the bank. By leveraging advanced predictive analytics, the system enables banks to take proactive measures to improve customer retention and enhance overall business performance. The proposed system follows a modular and layered architecture, ensuring flexibility, scalability, and ease of maintenance. It consists of multiple interconnected components, including data collection, preprocessing, feature engineering, model development, evaluation, and prediction generation. Each component performs a specific function, contributing to the overall efficiency of the system.

In the initial stage, customer data is collected from reliable sources and includes attributes such as age, gender, credit score, account balance, tenure, transaction activity, and product usage. This data is then passed to the preprocessing module, where it is cleaned and transformed into a suitable format for analysis. Missing values are handled, categorical variables are encoded, and numerical features are normalized to ensure consistency and improve model performance. The feature engineering module plays a critical role in enhancing predictive capability by selecting relevant attributes and generating new features based on domain knowledge. This step reduces dimensionality and improves model efficiency.

The core of the proposed system lies in the model layer, where multiple predictive models are developed and trained. Classical ML models such as Logistic Regression and Random Forest provide baseline performance, while DL models such as Artificial Neural Networks capture complex nonlinear relationships. Additionally, QML models are explored using hybrid classical-quantum approaches to evaluate their feasibility in real-world applications. The evaluation module assesses model performance using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. Based on these metrics, the best-performing model is selected for deployment. The final output of the system is a prediction indicating whether a customer is likely to churn.

Overall, the proposed system provides a comprehensive and efficient solution for churn prediction, integrating multiple advanced techniques to improve accuracy, interpretability, and scalability.

SYSTEM ARCHITECTURE

The system design of the Bank Churn Prediction project provides a structured framework for implementing the predictive system. It defines how different components interact with each other to transform raw data into meaningful predictions. The design follows a layered and modular architecture, ensuring flexibility, scalability, and efficient data flow. The architecture is divided into several layers, including the input/data layer, preprocessing layer, feature engineering layer, model layer, evaluation layer, and output layer. Each layer performs a specific function and contributes to the overall workflow of the system.

The input layer is responsible for collecting raw customer data, which includes demographic, financial, and behavioral attributes. This data is then processed in the preprocessing layer, where cleaning, encoding, and normalization are performed. Ensuring data quality at this stage is essential for achieving accurate predictions.

The feature engineering layer focuses on selecting relevant features and transforming them to improve model performance. This includes removing redundant features, creating derived variables, and applying normalization techniques. The model layer is the core of the system, where Machine Learning, Deep Learning, and Quantum Machine Learning models are implemented. Each model is trained independently and evaluated based on performance metrics. This layered approach allows easy comparison and integration of different models.

The evaluation layer measures model performance using various metrics and identifies the best-performing model. Finally, the output layer generates predictions indicating whether a customer is likely to churn.

The design also considers important factors such as scalability, efficiency, interpretability, and data security. By adopting a modular approach, the system allows easy updates and integration of new technologies, making it suitable for real-world banking application

OBJECTIVES

The primary objective of this project is to design and develop an intelligent and reliable system for predicting customer churn in the banking sector using advanced computational techniques such as Machine Learning, Deep Learning, and Quantum Machine Learning. Customer churn prediction is essential for financial institutions because retaining existing customers is significantly more cost-effective than acquiring new ones. By identifying customers who are likely to leave, banks can implement targeted retention strategies and improve customer satisfaction.

One of the key objectives is to understand the concept of customer churn and analyze the factors that influence customer behavior. These factors include demographic attributes such as age and income, financial characteristics

such as account balance and credit score, and behavioral aspects such as transaction frequency and product usage. By studying these factors, the system can identify patterns associated with customer attrition.

Another important objective is to collect and preprocess customer data to ensure its quality and suitability for model training. Data preprocessing involves handling missing values, removing duplicates, encoding categorical variables, and normalizing numerical features. Proper preprocessing is essential for improving model performance and ensuring accurate predictions.

The project also aims to implement and evaluate various Machine Learning algorithms, including Logistic Regression, Random Forest, and Gradient Boosting. These models serve as baseline approaches and provide insights into the effectiveness of classical techniques for churn prediction. In addition, Deep Learning models such as Artificial Neural Networks are developed to capture complex and non-linear relationships in the data.

A unique aspect of this project is the exploration of Quantum Machine Learning techniques. Hybrid classical-quantum models are implemented to demonstrate the potential of quantum computing in predictive analytics. Although these models are experimental, they provide valuable insights into future technological advancements. Another objective is to compare the performance of ML, DL, and QML models using standard evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. This comparison helps in identifying the most effective approach for churn prediction.

METHODOLOGY

The methodology adopted in this project follows a structured and systematic approach to develop an effective bank churn prediction system. The process begins with data collection, where a relevant dataset containing customer information is obtained. This dataset includes demographic details, account information, transaction behavior, and a churn label indicating whether a customer has left the bank.

The next step is data preprocessing, which is crucial for improving data quality and preparing it for analysis. During this stage, missing values are handled, duplicates are removed, and inconsistencies are corrected. Categorical variables such as gender and geography are encoded into numerical form, while numerical features are normalized or standardized. Additionally, class imbalance is addressed using resampling techniques to ensure that the model does not become biased toward the majority class.

After preprocessing, exploratory data analysis (EDA) is performed to understand data distribution, correlations, and patterns related to churn. Visualization techniques such as histograms, scatter plots, and correlation matrices are used to gain insights into the dataset. Feature engineering is then carried out to select relevant attributes and create new features that enhance model performance.

The dataset is subsequently divided into training and testing sets to evaluate model generalization. In the Machine Learning phase, algorithms such as Logistic Regression, Random Forest, and Gradient Boosting are implemented and optimized using hyperparameter tuning techniques. These models serve as baseline approaches for churn prediction.

ALGORITHM

- Step 1: Upload or collect customer dataset through the system.
- Step 2: Validate the input data (check for missing or incorrect values).
- Step 3: Perform data preprocessing (handle missing values, remove duplicates).
- Step 4: Encode categorical variables (e.g., gender, geography).
- Step 5: Normalize/scale numerical features for better performance.
- Step 6: Split the dataset into training and testing sets (e.g., 80:20).
- Step 7: Train the model using Machine Learning algorithms (Random Forest, Logistic Regression) or Deep Learning (ANN).
- Step 8: (Optional) Apply Quantum Machine Learning model for advanced prediction.
- Step 9: Extract important features and learn patterns from training data.
- Step 10: Apply classification to predict churn (Yes/No).
- Step 11: Compute prediction probability (confidence score).
- Step 12: Evaluate model using metrics (accuracy, precision, recall, F1-score).
- Step 13: Display prediction result and confidence score to the user.
- Step 14: (Optional) Store prediction results for future analysis.

IJETRM

International Journal of Engineering Technology Research & Management (IJETRM)

Journal Article

<https://ijetrm.com/issue/>

EXPERIMENTAL SETUP

The experimental setup was designed to ensure fair and reliable evaluation of all models. The dataset consists of customer attributes such as age, gender, balance, credit score, and transaction activity.

Data preprocessing steps included cleaning, handling missing values, encoding categorical variables, feature scaling, and addressing class imbalance. The dataset was split into training (80%) and testing (20%) sets to evaluate model generalization. Machine Learning models were implemented using Scikit-learn, Deep Learning models using TensorFlow/Keras, and Quantum models using Qiskit. Hyperparameter tuning and cross-validation techniques were applied to improve performance and reduce overfitting. The setup ensures reproducibility and consistency in model evaluation.

PERFORMANCE METRICS

The performance of the models is evaluated using the following metrics:

- Accuracy: Measures the overall correctness of predictions.
- Precision: Indicates how many predicted churners are actually churners.
- Recall: Measures the ability to identify actual churners.
- F1-Score: Harmonic mean of precision and recall.
- ROC-AUC: Evaluates the model's ability to distinguish between churn and non-churn cases.

These metrics provide a comprehensive evaluation of model performance and help in selecting the best model for deployment.

RESULTS AND ANALYSIS

The results of the Bank Churn Prediction system demonstrate the effectiveness of Machine Learning (ML), Deep Learning (DL), and Quantum Machine Learning (QML) approaches in identifying customers likely to churn. The models were trained and tested on structured banking data, and their performance was evaluated using standard classification metrics. Among the models, ensemble methods such as Random Forest and Gradient Boosting achieved high accuracy and stability due to their ability to handle non-linear relationships and reduce overfitting. Deep Learning models, particularly Artificial Neural Networks, showed improved performance in capturing complex patterns and feature interactions, especially when trained on large datasets. However, they required higher computational resources.

Quantum Machine Learning models, implemented using hybrid classical-quantum approaches, demonstrated feasibility but achieved comparatively lower performance due to current hardware limitations.

Overall, the analysis indicates that ML models provide a balance between accuracy and interpretability, while DL models offer higher predictive power. The results validate the effectiveness of the proposed system in predicting customer churn and supporting decision-making in banking applications.

FUTURE ENHANCEMENT

Although the current system provides effective churn prediction, several improvements can be made in the future:

- Integration of real-time data processing for dynamic predictions
- Use of advanced deep learning models such as RNNs and Transformers
- Improvement in Quantum Machine Learning models with advanced hardware
- Implementation of Explainable AI (XAI) for better interpretability
- Deployment of the system as a web or mobile application
- Incorporation of customer feedback and behavioral analytics

These enhancements can further improve accuracy, usability, and scalability of the system.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my project guide and faculty members for their valuable guidance, support, and encouragement throughout the development of this project. Their insights and suggestions have greatly contributed to the successful completion of this work.

I also thank my institution for providing the necessary resources and environment to carry out this project effectively. Special thanks to my friends and team members for their cooperation and support during the project development.

CONCLUSION

The Bank Churn Prediction project successfully demonstrates the application of Machine Learning, Deep Learning, and Quantum Machine Learning techniques in predicting customer churn. The study highlights the importance of data-driven decision-making in the banking sector and provides a comprehensive framework for analyzing customer behavior. The results show that Machine Learning models are effective for structured data and provide interpretable results, making them suitable for practical deployment. Deep Learning models offer improved accuracy by capturing complex relationships but require higher computational resources. Quantum Machine Learning, although still in its early stages, presents promising opportunities for future research and innovation. The project emphasizes the importance of data preprocessing, feature engineering, and model evaluation in achieving accurate predictions. It also demonstrates the value of integrating multiple approaches within a unified framework for comparative analysis.

In conclusion, the proposed system provides a robust and scalable solution for customer churn prediction, enabling banks to improve customer retention strategies and enhance business performance. Future work can focus on improving quantum models, incorporating real-time data, and enhancing interpretability using explainable AI techniques.

REFERENCES

- 1) Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
- 2) Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, 29(5), 1189–1232.
- 3) Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3), 273–297.
- 4) Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). Wiley.
- 5) Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- 6) Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- 7) Chollet, F. (2018). *Deep Learning with Python*. Manning Publications.
- 8) Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.