IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14431

Customer Churn Prediction In Subscription-Based Platforms Using Machine Learning

YUVA DHARSHINI M[1], MALA BHARUMATHI M[2]

M.sc Data Science and Business Analysis, Department of Computer Science, Rathinam College of Arts and Science, Coimbatore, Tamilnadu-641021[1]

Department of Computer Science, Rathinam of College of Arts and Science, Coimbatore, Tamilnadu-641021[2]

Abstract: Customer churn is a major challenge in subscription-based platforms, where customers discontinue their services, leading to revenue loss. This project aims to predict customer churn and understand the key factors influencing customer behavior so that businesses can take actions to improve customer retention. In this project, a subscription-based customer dataset will be used, which includes features such as subscription type, usage patterns, login frequency, and payment details. The data will be preprocessed by handling missing values and encoding categorical variables. Exploratory Data Analysis (EDA) will be performed to identify patterns and trends related to customer churn. Machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, and XGBoost will be implemented to build predictive models. A comparative analysis will be carried out using evaluation metrics such as accuracy, precision, recall, and F1-score to identify the best-performing model. XGBoost is included as an advanced algorithm because it improves prediction accuracy by combining multiple weak models and handling complex data patterns effectively. By performing this analysis, the project will identify high-risk customers and the key factors leading to churn, enabling businesses to design effective retention strategies such as personalized offers and improved customer engagement. As an enhancement, an interactive dashboard will be developed to visualize churn patterns and monitor customer risk levels for better decision-making.

Keywords: Customer Churn Prediction, Machine Learning, XGBoost, Model Comparison, Subscription-Based Platforms

I. INTRODUCTION

In recent years, subscription-based platforms such as Netflix, Spotify, and Amazon Prime have experienced rapid growth due to their convenience and recurring revenue models. These platforms depend heavily on retaining customers for sustained profitability. However, increasing competition and evolving customer preferences have led to a significant rise in customer churn, making retention a major concern for businesses.

Customer churn refers to the situation where users discontinue their subscription over a given period. High churn rates directly affect revenue and increase the cost of acquiring new customers. Therefore, identifying the factors that influence churn and predicting customer behavior in advance has become essential. Customers may churn due to reasons such as low engagement, high subscription costs, poor service experience, or better alternatives offered by competitors.

The dataset used in this project, Customer Subscription Churn and Usage Patterns, provides valuable insights into customer behavior through features such as subscription type, tenure, monthly revenue, usage frequency, engagement level, support interactions, payment failures, and recent activity. These factors play a crucial role in determining churn. For example, customers with low usage, frequent payment issues, or repeated complaints are more likely to leave, whereas long-term and highly engaged users tend to remain loyal.

To address this problem, machine learning techniques are used to analyze customer data and predict churn effectively. Models such as Logistic Regression, Decision Tree, Random Forest, and XGBoost help identify patterns and classify customers based on their likelihood of churn. This project aims to develop a predictive system that enables businesses to take proactive retention measures such as personalized offers, improved support, and targeted engagement strategies, thereby reducing churn and improving overall business performance.

DOI: 10.17148/IJIREEICE.2026.14431

II. LITERATURE REVIEW

1. Imani, M., et al. (2025) conducted a comprehensive review of customer churn prediction techniques, comparing traditional statistical approaches with modern machine learning and deep learning models. The study highlighted that ML and DL models are more effective in analyzing large-scale customer data and capturing complex behavioral patterns. It emphasized that these models significantly improve prediction accuracy compared to traditional methods.
2. Idris, A., et al. (2019) proposed a churn prediction model using big data analytics, focusing on handling large volumes of customer data. The study highlighted the importance of scalable data processing and feature selection in improving model performance. Machine learning algorithms were applied to enhance prediction accuracy and efficiency. The results showed that integrating big data technologies with ML models leads to better churn prediction outcomes
3. Kumar, P., et al. (2025) performed a comparative study of machine learning algorithms such as Decision Tree, Random Forest, and XGBoost. The study found that ensemble models provide higher accuracy and robustness compared to individual models. It emphasized the importance of evaluating models using multiple performance metrics. The research concluded that ensemble techniques are more reliable for real-world churn prediction tasks.
4. Chen, X., et al. (2023) introduced an explainable machine learning model using XGBoost and SHAP values. The study focused on improving model transparency by identifying key factors influencing churn. It highlighted how explainable AI helps businesses understand predictions and make informed decisions. The results demonstrated that combining accuracy with interpretability enhances the usability of churn prediction systems.
5. Ahn, J., et al. (2023) developed a deep learning-based churn prediction system using neural networks. The study showed that deep learning models outperform traditional machine learning approaches in capturing complex customer behavior. It emphasized the ability of neural networks to handle large and complex datasets effectively. The results indicated improved prediction performance in subscription-based environments.
6. Reddy, S., et al. (2025) explored deep learning approaches such as Artificial Neural Networks (ANN) and Long Short-Term Memory (LSTM) models. The study focused on capturing temporal and sequential patterns in customer usage data. It demonstrated that LSTM models are particularly effective for time-series analysis. The findings showed improved prediction accuracy when customer behavior changes over time.
7. Li, H., et al. (2022) proposed attention-based deep learning models to improve churn prediction in subscription-based platforms. The study emphasized analyzing sequential customer activity and assigning importance to key features. Attention mechanisms helped improve model focus and prediction accuracy. The results showed better performance and deeper insights into customer behavior
8. Singh, R., et al. (2024) implemented an ensemble learning framework combining multiple machine learning algorithms. The study highlighted that ensemble techniques improve prediction accuracy and reduce overfitting. It emphasized the importance of leveraging strengths of different models. The results demonstrated better performance compared to individual models.
9. Gupta, A., et al. (2024) compared various machine learning models for churn prediction and found that XGBoost consistently outperforms others in terms of accuracy, precision, and recall. The study highlighted the effectiveness of boosting techniques in handling complex datasets. The results support the use of XGBoost as a reliable model for churn prediction.
10. Sharma, K., et al. (2023) emphasized the importance of feature engineering in churn prediction models. The study identified key variables such as tenure, subscription type, usage frequency, payment behavior, and customer interactions as major indicators of churn. It highlighted that proper feature selection improves model performance significantly. The findings stress the role of preprocessing in accurate prediction
11. Zhang, Y., et al. (2024) focused on customer retention strategies in subscription-based platforms. The study highlighted that combining churn prediction with personalized recommendations improves retention rates. It emphasized the role of predictive analytics in identifying at-risk customers. The results showed improved customer engagement and reduced churn
12. Patel, D., et al. (2024) discussed integrating predictive analytics with business strategies for churn management. The study suggested combining churn prediction models with actionable retention measures such as targeted offers and customer engagement strategies. It emphasized converting predictions into business actions. The findings showed improved retention performance.

III. PROBLEM STATEMENT

In today's competitive digital environment, subscription-based platforms such as Netflix, Spotify, Amazon Prime, and SaaS platforms like Salesforce and Zoom face significant challenges in retaining customers. Customer churn, where users discontinue their subscriptions, leads to revenue loss and increased customer acquisition costs.

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026
DOI: 10.17148/IJIREEICE.2026.14431

Background and Context: Subscription and SaaS platforms depend on recurring revenue, making customer retention essential for business growth. Analyzing customer data can help identify potential churn.
Challenges in Churn Prediction: Churn is influenced by factors such as tenure, usage patterns, subscription type, and payment behavior, which are complex and dynamic.
Limitations of Existing Systems: Existing methods rely on basic models and lack the ability to capture complex patterns or provide real-time insights.
Data Challenges: Datasets often contain missing values, imbalanced data, and noisy features like usage and support interactions.
Business Impact: High churn reduces revenue, customer lifetime value, and overall business performance.
Research Gap and Motivation: There is a need for an effective machine learning system that can accurately predict churn and support proactive retention strategies in subscription and SaaS platforms.

IV. PROPOSED SYSTEM

The main objectives of this research are:

- To develop a machine learning–based system for accurate prediction of customer churn in subscription-based and SaaS platforms.
- To analyze customer behavior using features such as tenure, usage patterns, subscription type, payment history, and support interactions.
- To compare the performance of different machine learning models such as Logistic Regression, Decision Tree, Random Forest, and XGBoost.
- To identify key factors influencing customer churn using feature importance and data analysis techniques.
- To provide actionable insights and decision support for businesses to implement effective customer retention strategies.
- To enhance customer satisfaction and reduce revenue loss by enabling proactive measures such as personalized offers and improved service quality.
- To develop a scalable and efficient system that can support real-time churn prediction and business decision-making.

V. METHODOLOGY AND CALCULATION DETAILS

The proposed system follows a systematic approach to predict customer churn and analyze customer behavior in subscription-based platforms using machine learning techniques.

A. Data Collection

In this project, a customer subscription churn and usage patterns dataset is collected from Kaggle. The dataset contains information such as tenure, subscription type, monthly revenue, usage frequency, engagement level, support interactions, and payment behavior. The target variable is Churn, which indicates whether a customer leaves the service or not. This dataset is used to analyze customer behavior and build predictive models.

B. Data Preprocessing

The dataset is first cleaned to ensure accuracy and consistency. Missing values are handled using techniques such as mean or median imputation, or by removing incomplete records if necessary. Duplicate records are removed to avoid bias. Categorical variables like subscription type and payment method are converted into numerical format using encoding techniques such as Label Encoding or One-Hot Encoding. Numerical features are scaled using normalization or standardization to ensure equal contribution to the model. If the dataset is imbalanced, resampling methods such as oversampling or undersampling are applied to balance the churn and non-churn classes.

C. Exploratory Data Analysis (EDA)

Exploratory Data Analysis is performed to understand the data distribution and identify patterns. Statistical summaries and visualizations are used to analyze features such as tenure, revenue, and usage frequency. Relationships between features and churn are examined using correlation analysis and graphical plots. This step helps identify key factors that influence customer churn and supports better model building.

D. Feature Engineering

New features are created to improve model performance. For example, an engagement score is calculated by combining usage frequency and interaction level to represent customer involvement. Customer activity level is also derived to capture overall behavior. These features provide more meaningful information, allowing models to better understand customer patterns and improve prediction accuracy.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14431

E. Model Selection and Training

The dataset is split into training and testing sets, typically in an 80:20 ratio. Multiple machine learning algorithms are used to build and compare churn prediction models.

- **Logistic Regression:**

Logistic Regression is a statistical model used for binary classification. It calculates the probability of churn using a logistic (sigmoid) function. The model is simple, fast, and interpretable, making it suitable as a baseline model.

- **Decision Tree:**

Decision Tree is a tree-based model that splits the dataset into branches based on feature values. It forms decision rules that help classify customers as churn or non-churn. It can capture non-linear relationships but may overfit if not controlled.

- **Random Forest:**

Random Forest is an ensemble method that combines multiple decision trees. Each tree is trained on a different subset of data, and the final prediction is based on majority voting. This reduces overfitting and improves accuracy compared to a single decision tree.

- **XGBoost (Extreme Gradient Boosting):**

XGBoost is an advanced boosting algorithm that builds models sequentially by correcting the errors of previous models. It uses gradient boosting techniques to optimize performance and handle complex data patterns efficiently. Due to its high accuracy and ability to manage large datasets, it is expected to perform better than other models.

F. Model Evaluation and Comparison

After training, the models are evaluated using the test dataset. The performance of each model is measured using the following metrics:

Accuracy = (TP + TN) / Total Predictions
Precision = TP / (TP + FP)
Recall = TP / (TP + FN)
F1-score = 2 × (Precision × Recall) / (Precision + Recall)

Where TP = True Positive, TN = True Negative, FP = False Positive and FN = False Negative.

A comparison is made between all models based on these metrics. The model with the highest performance is selected as the best model. XGBoost is expected to achieve the highest accuracy due to its boosting mechanism and ability to reduce prediction errors.

G. Churn Prediction

The selected best-performing model is used to predict churn probabilities for customers. Based on these probabilities, customers are classified as high-risk (likely to churn) or low-risk (likely to continue). This helps in identifying customers who require attention.

H. Interpretation and Retention Strategies

Feature importance techniques are applied to understand which factors contribute most to churn. Based on these insights, businesses can design retention strategies such as personalized offers, improved engagement, and better customer support to reduce churn.

I. Visualization and Dashboard

The results are visualized using charts such as bar graphs, pie charts, and heatmaps to clearly present churn patterns. An interactive dashboard is developed using Power BI to display key insights such as churn rate, high-risk customers, and model comparison results. This enables effective monitoring and data-driven decision-making.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14431

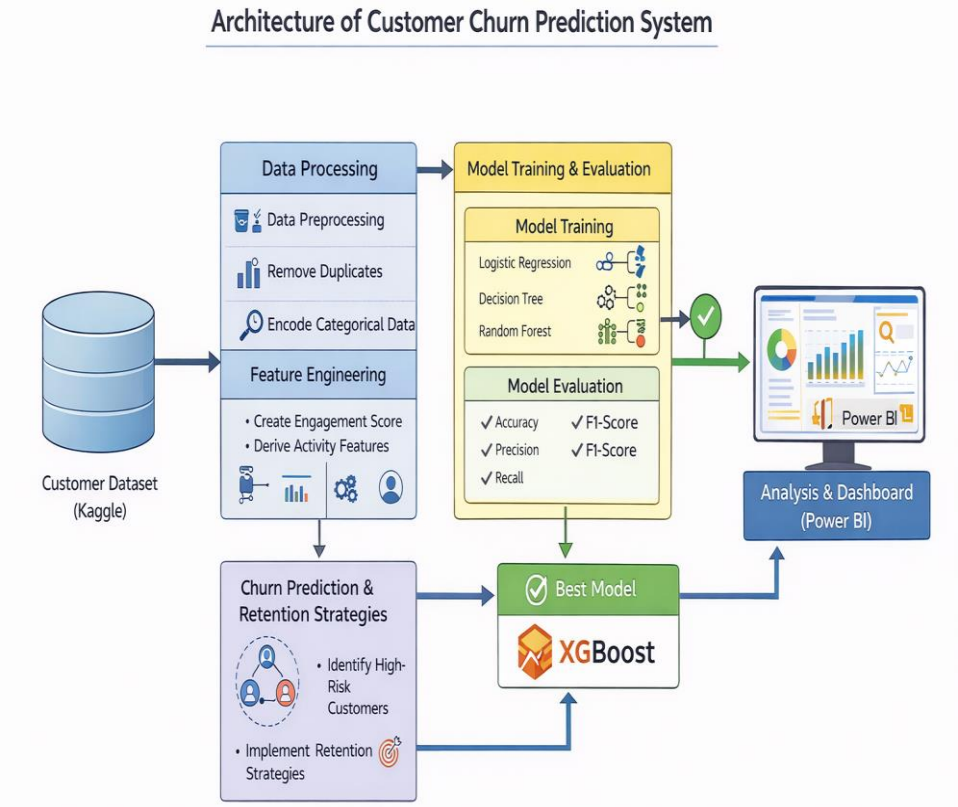
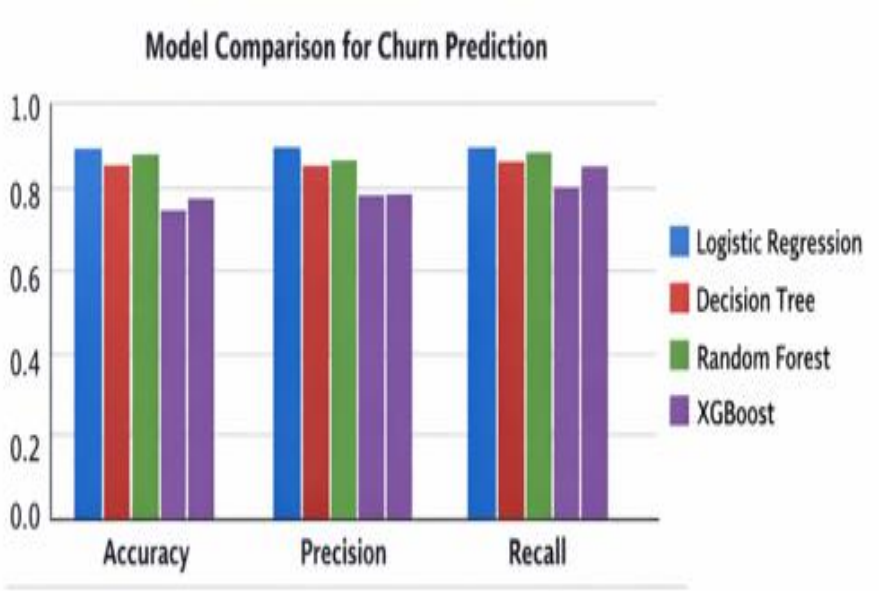


Fig 1: Architecture

VI. COMPARISONS OF DATA

This section presents a comparative analysis of machine learning models along with an analysis of churn behavior across different subscription types. The models, including Logistic Regression, Decision Tree, Random Forest, and XGBoost, are evaluated using performance metrics such as accuracy, precision, and recall to identify the best-performing model. In addition, the relationship between subscription type and churn rate is analyzed to understand how churn varies across different plans. This combined analysis helps in selecting an effective predictive model and identifying high-risk customer segments for better retention strategies



IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026
DOI: 10.17148/IJIREEICE.2026.14431

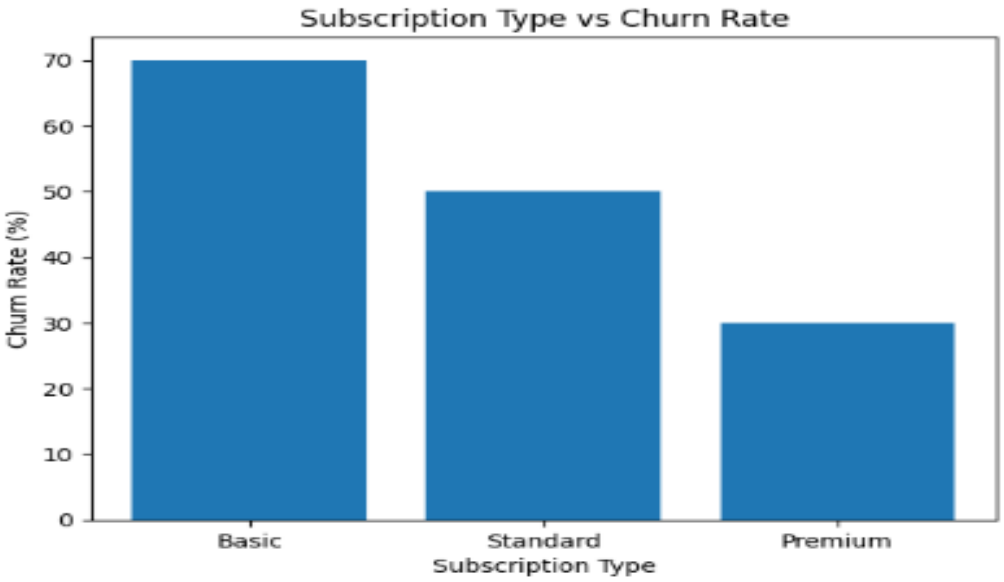


Fig 2: comparison of existing data

VII. RESULT AND ANALYSIS

The implementation of the customer churn prediction system provided meaningful insights into customer behavior in subscription-based and SaaS platforms. By analyzing the dataset Customer Subscription Churn and Usage Patterns, various patterns influencing churn were identified.

A. Churn Distribution:
The analysis shows that a majority of customers remain active, while a smaller percentage of customers churn. However, this smaller group significantly impacts business revenue, making churn prediction essential.

B. Tenure Analysis:
Customers with lower tenure (new users) are more likely to churn compared to long-term customers. This indicates the importance of engaging users early in their subscription period.

C. Usage and Engagement Analysis:
Customers with low usage frequency and low engagement levels show a higher probability of churn. Highly active users tend to remain loyal to the platform.

D. Subscription Type Analysis:
Basic subscription users exhibit a higher churn rate compared to premium users. This suggests that customers receiving more features and value are less likely to leave.

E. Payment Behavior Analysis:
Frequent payment failures and billing issues are strongly associated with churn. Customers facing payment difficulties are more likely to discontinue their subscriptions.

F. Model Performance:
Among the applied models, XGBoost achieved the highest accuracy and best overall performance, making it the most suitable model for churn prediction in this dataset.

G. Dashboard Visualization:
The insights obtained from the analysis are effectively presented using an interactive dashboard created in Power BI. The dashboard includes key visualizations such as churn distribution, activity level analysis, subscription type comparison, tenure trends, and payment behavior.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14431

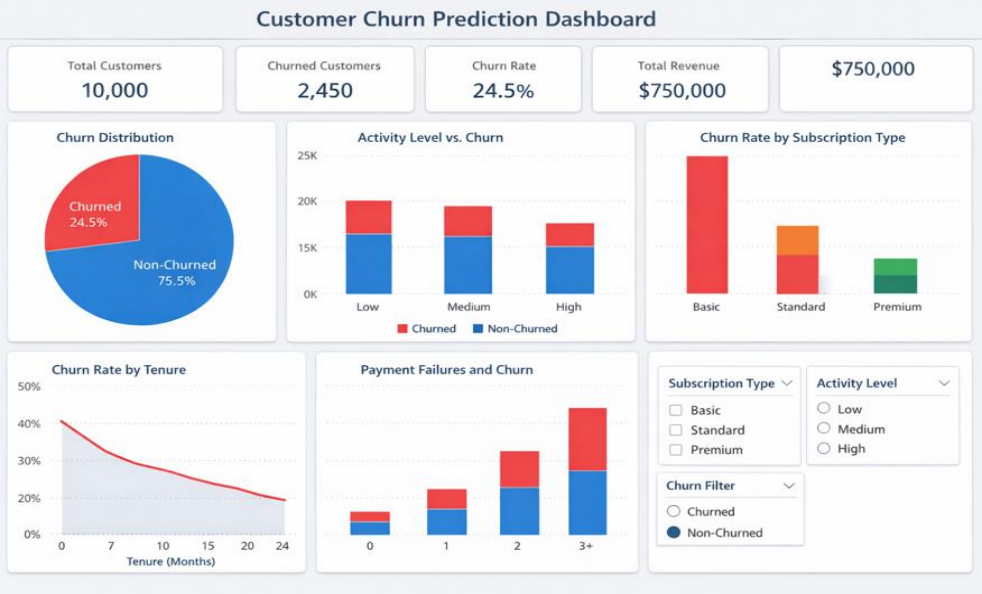


Fig 3: Dashboard

VIII. FUTURE ENHANCEMENTS

The proposed customer churn prediction system can be further improved in several aspects to enhance its performance, scalability, and practical applications.

A. Advanced Model Enhancements

The system can be improved by incorporating advanced deep learning models such as Long Short-Term Memory (LSTM) networks to capture time-based customer behavior. Hybrid models combining machine learning and deep learning techniques can also be explored to achieve higher prediction accuracy and better performance.

B. Real-Time Data Integration

Currently, the system works on static datasets. In future, it can be enhanced by integrating real-time data processing, enabling continuous churn prediction. This allows businesses to identify at-risk customers instantly and take proactive actions.

C. Scalability

The system can be scaled to handle large-scale customer datasets efficiently. It can also be extended to support multiple subscription-based platforms such as SaaS applications and OTT services. Integration with real-time streaming data can further improve its capability.

D. Business Applications

This system can be applied across various industries including OTT platforms, SaaS companies, telecom services, and e-commerce platforms with subscription models. It helps organizations improve customer retention, engagement, and overall business performance.

E. Integration with Business Systems

The churn prediction model can be integrated with Customer Relationship Management (CRM) systems and marketing automation tools. It can also be connected with business intelligence tools like Power BI for effective visualization and monitoring of churn trends.

F. Decision Support System

The project can be extended into a complete decision support system by providing churn risk scores and recommending personalized retention strategies. This will help businesses make data-driven decisions and improve customer satisfaction.

G. Future Research Scope

Further research can focus on explainable AI techniques to improve model transparency and interpretability. Additionally, customer segmentation and customer lifetime value prediction can be included to gain deeper insights into customer behavior.

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026
DOI: 10.17148/IJIREEICE.2026.14431

IX. CONCLUSION

Customer churn prediction plays a vital role in improving customer retention and ensuring business sustainability in subscription-based and SaaS platforms. This project successfully demonstrates the use of machine learning techniques to analyze customer behavior and predict churn based on features such as tenure, usage patterns, subscription type, engagement level, and payment history. Different machine learning models, including Logistic Regression, Decision Tree, Random Forest, and XGBoost, were applied and compared to identify the most effective approach. Among these, XGBoost provided the best performance due to its ability to handle complex patterns and improve prediction accuracy.
The analysis shows that customers with low engagement, low usage frequency, frequent payment failures, and shorter tenure are more likely to churn. By identifying such high-risk customers in advance, businesses can take proactive measures such as personalized offers, improved customer support, and targeted engagement strategies.
Overall, this system provides a data-driven approach to churn prediction, helping organizations reduce customer attrition, improve customer satisfaction, and enhance overall business performance.

REFERENCES

[1]. Raghukumar AM, Narayanan G (2020) Comparison of machine learning algorithms for detection of medicinal plants. In: 2020 fourth international conference on computing methodologies and communication (ICCMC). IEEE, pp 56–60
[2]. Priyanka R, Aravinth J (2021) Comparative analysis of different machine learning classifiers for prediction of diabetic retinopathy. In: 2021 International conference on recent trends on electronics, information, communication & technology (RTEICT). IEEE, pp 233–239
[3]. Akaramuthalvi JB, Palaniswamy S (2021) Comparison of conventional and automated machine learning approaches for breast cancer prediction. In: 2021 third international conference on inventive research in computing applications (ICIRCA). IEEE, pp 1533–1537
[4]. Dalvi PK, Khandge SK, Deomore A, Bankar A, Kanade VA (2016) Analysis of customer churn prediction in telecom industry using decision trees and logistic regression. In: 2016 symposium on colossal data analysis and networking (CDAN). IEEE, pp 1–4
[5]. Shrikhande PA (2018) Performance enhancement of customer churn prediction in telecom sector using decision tree techniques
[6]. Vafeiadis T, Diamantaras KI, Sarigiannidis G, Chatzisavvas KC (2015) A comparison of machine learning techniques for customer churn prediction. Simul Model Practice Theory 55:1–9
[7]. Rahman M, Kumar V (2020) Machine learning based customer churn prediction in banking. In: 2020 4th international conference on electronics, communication and aerospace technology (ICECA). IEEE, pp 1196–1201
[8]. Razak NIA, Wahid MH (2021) Telecommunication customers churn prediction using machine learning. In: 2021 IEEE 15th Malaysia international conference on communication (MICC). IEEE, pp 81–85
[9]. Balan A, Ramanathan T (2022) Comparative analysis of ma- chine learning algorithms to predict solar irradiance. In: 2022 international conference on disruptive technologies for multi- disciplinary research and applications (CENTCON), vol 2. IEEE, pp 167–172
[10]. Devi OR, Pothini SK, Kumari MP, Charan UNS (2023) Customer churn prediction using machine learning: subscription renewal on OTT platforms. In: 2023 2nd international conference on applied artificial intelligence and computing (ICAAIC). IEEE, pp 1025–1029
[11]. Ranjan NM, Bharambe Y, Deshmukh P, Choudhary D (2023) Churn prediction in telecommunication industry
[12]. Blank C, Hermansson T (2018) A machine learning approach to churn prediction in a subscription-based service
[13]. Pebrianti D, Istinabiyah DD, Bayuaji L (2022) Hybrid method for churn prediction model in the case of telecommunication companies. In: 2022 9th international conference on electrical engineering, computer science and informatics (EECSI). IEEE, pp 161–166
[14]. Ulkhaq MM, Wibowo AT, Tribosnia MR, Putawara R, Firdauz AB (2021) Predicting customer churn: a comparison of eight machine learning techniques: a case study in an Indonesian telecommunication company. In: 2021 international conference on data analytics for business and industry (ICDABI). IEEE, pp 42–46
[15]. Nair AJ, Rasheed R, Maheeshma KM, Aiswarya LS, Kavitha KR (2019) An ensemble-based feature selection and classification of gene expression using support vector machine, K-nearest neighbor, decision tree. In: 2019 international conference on communication and electronics systems (ICCES). IEEE, pp 1618–1623
[16]. Barham S, Aweisi N (2023) A review on machine learning-based customer churn prediction in the telecom industry. In: 2023 9th international conference on control, decision and information technologies (CoDIT). IEEE, pp 2659–2