AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.2(2026)
www.ajmimc.com

CUSTOMER CHURN PREDICTION USING MACHINE LEARNING

Mrs.K.Helini
Assistant Professor,Dept of AI&DS
VMTW,India

T. Sathvika
Dept of AI&DS
VMTW,India

G.Shiva Joshna
Dept of AI&DS
VMTW,India

D.Vijaya Laxmi
Dept of AI&DS
VMTW,India

A.Sruthi
Dept of AI&DS
VMTW,India

ABSTRACT

Customer churn is a major concern for organizations in industries such as telecom, banking, e commerce, OTT platforms. Churn occurs when customers discontinue a service or shift to competitors, resulting in revenue loss and higher marketing expenses. Traditional methods for handling churn rely on rule-based analysis and manual monitoring, which are often reactive, less accurate, and costly. This project proposes a machine learning-based churn prediction system that analyzes customer behavior, service usage, and demographic information to identify individuals who are likely to leave. By detecting churn risks in advance, the system enables companies to design personalized retention strategies, reduce unnecessary promotional costs, and improve customer satisfaction. The proposed solution not only improves the accuracy of churn prediction but also enables organizations to take proactive measures, reduce and enhance customer satisfaction. By offering insights into the underlying reasons for churn, it strengthens customer loyalty, minimizes revenue loss, and ensures long-term sustainable growth.

Keywords: Customer Retention, Telecom, Banking, Preprocessing,Customer Satisfaction,Revenue loss.

I. INTRODUCTION

Customer churn prediction is an important application of machine learning and data analytics in modern business environments. Organizations today operate in highly competitive markets where retaining customers is just as important as acquiring new ones. Research indicates that acquiring a new customer costs significantly more than retaining an existing one. Therefore, businesses invest heavily in customer retention strategies.Customer churn occurs when customers stop using a company's services or products. This can happen due to several reasons such as poor service quality, high pricing, better competitor offerings, lack of customer support, or dissatisfaction with product features. Identifying customers who are likely to churn enables companies to take preventive actions.

II. PROBLEM STATEMENT

Customer churn represents a major challenge for businesses across various industries. When customers leave a company's service, it leads to financial losses and reduces the overall customer base.Traditional methods of churn analysis rely on manual data analysis and simple statistical techniques. These approaches often fail to capture complex patterns in customer behavior. Additionally, they are time-consuming and inefficient when dealing with large datasets.

Therefore, there is a need for an intelligent system capable of analyzing large volumes of customer data and predicting churn accurately. Machine learning techniques provide a promising solution to this problem by identifying patterns and trends that indicate potential churn behavior.

III. LITERATURE REVIEW

1. "Customer Churn Prediction with Self-Attention Mechanism" (2022)

This 2022 study proposed a deep learning-based churn prediction model using a self- attention mechanism. The project focused on improving feature extraction by capturing relationships between customer attributes. It used neural networks with embedding layers to process high-dimensional telecom data. The results showed that the model outperformed traditional algorithms like logistic regression and decision trees. The study highlighted that attention mechanisms help in identifying important features influencing churn. It also reduced the dependency on manual feature engineering. The model achieved higher accuracy and better generalization. This project proved the effectiveness of deep learning in complex churn prediction problems.

2. "Explainable Customer Churn Prediction" (2023)

This 2023 project focused on combining machine learning with explainable AI techniques. It used XGBoost for prediction and SHAP (SHapley Additive exPlanations) for interpreting results. The main goal was not only to predict churn but also to understand the reasons behind it. The study identified key features such as customer tenure, service usage, and complaints. The results showed high accuracy along with better interpretability. Businesses could use these insights to design targeted retention strategies. This project emphasized the importance of transparency in machine learning models. It marked a shift towards explainable and user-friendly AI systems.

3. "Multimodal Customer Churn Prediction Using Behavioral and Sentiment Data" (2023)

This project introduced a multimodal approach by combining different types of customer data. It integrated behavioral data, transaction history, and sentiment analysis from customer feedback. Machine learning models were trained on this combined dataset to improve prediction performance. The study showed that including sentiment data significantly increased accuracy. It proved that customer emotions and feedback play a vital role in churn behavior. The model outperformed single-data-source approaches. This research highlighted the importance of holistic data integration. It opened new directions for using social media and text data in churn prediction.

4. "Customer Churn Prediction in Banking Using Machine Learning Techniques" (2024)

This 2024 study focused on applying ensemble learning techniques in the banking sector. Models such as Random Forest, XGBoost, and LightGBM were used and compared. The results showed that ensemble methods provided higher accuracy and robustness. The study also addressed class imbalance using techniques like SMOTE and ADASYN. Feature importance analysis was performed to identify key churn factors. The project demonstrated improved performance in terms of precision, recall, and F1- score. It highlighted that combining multiple models reduces prediction errors. This research proved that ensemble learning is highly effective for churn prediction.

5. "Hybrid Neural Network Model for Customer Churn Prediction with Imbalanced Data Handling" (2024)

This project proposed a hybrid model combining neural networks with data balancing techniques. It focused on solving the issue of imbalanced datasets in churn prediction. ADASYN and SMOTE techniques were used to balance the dataset. The hybrid model improved recall and reduced false negatives. It also achieved better performance compared to standalone machine learning models. The study emphasized the importance of handling data imbalance. It showed that balanced data leads to more reliable predictions. This work contributed to improving model stability and performance in real-world datasets.

6. "Customer Churn Prediction System for Subscription Businesses" (2025)

This 2025 project developed an AI-based system for predicting churn in subscription-based services. It used machine learning algorithms such as decision trees, random forests, and neural networks. The system identified key features like subscription duration, usage patterns, and customer support interactions. Neural networks provided the highest accuracy among all models. The project also focused on real-time prediction and automation. It helped businesses take proactive actions to retain customers. The study emphasized scalability and deployment in real- world systems. This project reflects the growing use of AI in business applications.

7. "Real-Time Customer Churn Prediction Using Big Data Analytics" (2025)

This study introduced a real-time churn prediction framework using big data technologies. It processed streaming customer data to predict churn instantly. The system used scalable machine learning models integrated with cloud platforms. It improved response time and allowed businesses to take immediate action. The study highlighted the importance of real-time analytics in competitive markets. It also focused on automation and continuous learning models. The results showed improved customer retention rates. This project represents the future direction of churn prediction systems.

IV. EXISTING SYSTEM:

In many industries such as telecommunications, banking, insurance, and online service platforms, customer churn has become a significant challenge. Organizations continuously try to understand why customers discontinue their services and switch to competitors. Traditionally, companies relied on conventional business analysis methods to monitor customer behavior and identify potential churn. These traditional approaches form the existing system for churn analysis.

In the existing system, organizations primarily depend on manual data analysis, basic statistical methods, and simple business intelligence tools to understand customer behavior. Customer data such as transaction records, billing information, service usage details, and demographic information are collected and stored in company databases. Analysts then examine this data using spreadsheets or basic reporting tools to identify trends related to customer retention and churn.

One common method used in traditional churn analysis is descriptive analytics, where past customer data is examined to understand patterns and trends. For example, organizations may analyze historical records to determine how many customers discontinued services within a specific time period. However, this approach only provides information about past events and does not accurately predict future churn behavior.

Another approach used in the existing system is rule-based analysis. In this method, predefined business rules are used to identify potential churn customers. For instance, a company may consider a customer as high risk if they have not used a service for a certain number of days or if they have filed multiple complaints within a short period of time. While this method is simple and easy to implement, it often fails to capture complex relationships between different factors that influence customer decisions.

Many organizations also use basic statistical techniques such as regression analysis or correlation analysis to understand the relationship between customer attributes and churn behavior. These techniques can provide some level of insight but are limited in their ability to process large datasets and detect non-linear patterns in customer behavior.

Another limitation of existing churn analysis systems is that they often rely heavily on human expertise and manual interpretation of data. Analysts must examine multiple reports and dashboards to identify possible churn patterns, which can be time-consuming and inefficient. In large organizations with millions of customers, manual analysis becomes extremely difficult and prone to errors.

V. PROPOSED SYSTEM

To overcome the limitations of traditional customer churn analysis methods, this project proposes an intelligent Customer Churn Prediction System using Machine Learning techniques. The proposed system is designed to analyze large volumes of customer data, identify patterns related to customer behavior, and predict whether a customer is likely to discontinue a service. By using advanced machine learning algorithms, the system can provide accurate predictions that help organizations implement effective customer retention strategies.

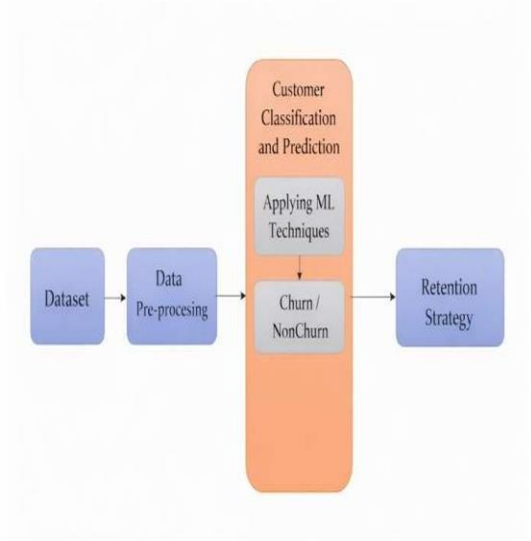
The proposed system utilizes predictive analytics and machine learning algorithms to analyze historical customer data and generate insights about customer behavior. Unlike traditional systems that rely on manual analysis or simple statistical techniques, the proposed system automatically processes large datasets and identifies complex relationships between different customer attributes.

The system works by collecting customer data from organizational databases, preprocessing the data to remove inconsistencies, training machine learning models using historical data, and predicting churn probability for each customer. The predictions generated by the system enable companies to identify customers who are at high risk of leaving the service and take appropriate preventive actions.

One of the key advantages of the proposed system is its ability to handle large-scale datasets efficiently. Modern organizations generate massive amounts of customer data from various sources such as transaction records, service usage logs, billing information, customer feedback, and interaction history. The proposed system can process and analyze this data to identify hidden patterns that indicate potential churn behavior.

VI. SYSTEM ARCHITECTURE

The system architecture of the Customer Churn Prediction project defines the overall structure and flow of data through different components. It includes modules such as data collection, preprocessing, model building, and prediction. This architecture ensures efficient processing and accurate prediction of customer churn.



1. Dataset

The first stage in the system architecture is the Dataset. This stage involves collecting customer-related data from different sources such as company databases, CRM systems, transaction records, and service usage logs.

The dataset contains various attributes that help analyze customer behavior and predict churn. Some common features in the dataset include: Customer ID ,Gender, Age, Tenure (duration of customer relationship),Monthly charges, Total charges, Contract type, Payment method, Internet service type, Customer support interactions, Number of complaints

2. Data Preprocessing

After collecting the dataset, the next step is Data Preprocessing. Raw datasets usually contain missing values, inconsistent data formats, and irrelevant features. Data preprocessing is required to clean and transform the raw data into a structured format suitable for machine learning algorithms.

Several operations are performed during this stage, including:

Data Cleaning

Removing duplicate records, incorrect values, and unnecessary data entries.

Handling Missing Values

Missing data fields are handled by replacing them with appropriate values such as mean, median, or mode.

Data Transformation

Categorical data such as gender or payment method are converted into numerical values using encoding techniques.

Data Normalization

Numerical features such as monthly charges and tenure are scaled to ensure that all features contribute equally to the machine learning model. Feature Selection

Selecting the most important features that influence customer churn while removing irrelevant attributes.

Data preprocessing improves the overall quality of the dataset and ensures better performance of machine learning models.

3. Customer Classification and Prediction

The next stage in the architecture is Customer Classification and Prediction. This stage is the core component of the churn prediction system where machine learning techniques are applied to analyze customer behavior.

The preprocessed data is divided into two parts: Training dataset,Testing dataset

The training dataset is used to train the machine learning model so that it can learn patterns and relationships between customer features and churn behavior.

Several machine learning algorithms can be applied for this purpose, such as: Logistic Regression,Decision Tree,Random Forest,Support Vector Machine,K-Nearest Neighbors

4. Applying Machine Learning Techniques Inside the classification module, machine learning techniques are applied to build predictive models. These models learn from historical customer data and generate predictions for new customers.

The machine learning process typically includes the following steps:

- Training the model using historical data
- Testing the model using unseen data
- Evaluating model performance using metrics such as accuracy, precision, recall, and F1-score
- Optimizing model parameters to improve prediction accuracy

5. Churn / Non-Churn Classification

After applying machine learning algorithms, the system classifies customers into two categories:

Churn Customers

Customers who are likely to discontinue the service or switch to a competitor.

Non-Churn Customers

Customers who are likely to continue using thE service.

The classification result helps businesses identify high-risk customers who may leave the company in the near future.

The prediction results can also be represented in the form of probability scores indicating the likelihood of churn.

6. Retention Strategy

The final stage of the system architecture is the Retention Strategy. Once the system identifies customers who are likely to churn, businesses can implement strategies to retain those customers.

Some common customer retention strategies include:

- Offering personalized discounts
- Providing better service packages
- Improving customer support
- Offering loyalty rewards
- Sending promotional offers
- Addressing customer complaints quickly

VII. SYSTEM REQUIREMENTS

A. Hardware Requirements

Processor (CPU)

Recommended: Intel Core i5 / Inte Core i7 / AMD Ryzen 5 or above

Random Access Memory (RAM)

MinimumRequirement:8GBRAM Recommended: 16 GB RAM

Storage (Hard Disk / SSD)

MinimumRequirement:256GB
Recommended: 512 GB SSD or higher

Input Devices

Keyboard,Mouse

Output Devices

Monitor / Display Screen,Printer

B. Software Requirements

Programming Language

Python

Python is the primary programming language used for developing the customer churn prediction model.

Development Environment

Jupyter Notebook / Anaconda / Visual Studio Code

Machine Learning Libraries

Several Python libraries are used to implement machine learning algorithms and perform data analysis.

- NumPy,Pandas,Scikit-learn, TensorFlow/Keras

Data Visualization Tools

Data visualization tools help in understanding patterns and trends in customer data.

Matplotlib, Seaborn

VIII. METHODOLOGY

The methodology of the Customer Churn Prediction project follows a systematic process starting with data collection from reliable sources such as telecom or banking datasets, which include customer demographics, service usage, and churn labels. The collected data is then preprocessed by handling missing values, removing duplicates, encoding categorical variables, and normalizing the data to ensure consistency. Exploratory Data Analysis (EDA) is performed to understand patterns, trends, and relationships between features using visualizations and correlation analysis. Feature engineering and selection are applied to identify the most relevant attributes that influence churn while reducing unnecessary data. The processed dataset is then used to build machine learning models such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and sometimes Neural Networks, where the data is split into training and testing sets for effective learning.

IX. WORKFLOW

1. Data Collection

The workflow begins with collecting relevant customer data from reliable sources such as telecom companies, banking systems, or public datasets. The dataset typically includes customer demographics, account details, service usage, and churn status

2. Data Preprocessing

After collecting the data, preprocessing is performed to clean and prepare it for analysis. Missing values are handled using appropriate techniques such as mean or mode imputation. Categorical variables are converted into numerical format using encoding methods

3. Exploratory Data Analysis (EDA)

EDA is carried out to understand the structure and patterns within the dataset. Various visualization techniques such as bar charts, histograms, and heatmaps are used. This step helps in identifying trends, correlations, and relationships between variables.

4. Feature Engineering and Selection

In this step, important features are selected and new features may be created to enhance model performance. Irrelevant or redundant features are removed to reduce complexity. Techniques like correlation analysis and feature importance ranking are used.

5. Model Building

Machine learning models are developed to predict customer churn. Algorithms such as Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM) are commonly used. The dataset is divided into training and testing sets.

6. Model Evaluation

The trained models are evaluated using performance metrics such as accuracy, precision, recall, and F1-score. A confusion matrix is used to analyze the predictions. ROC-AUC curve helps in understanding classification performance.

7. Model Optimization

Model optimization is performed to improve accuracy and efficiency. Hyperparameter tuning techniques such as Grid Search and Cross-Validation are used. Overfitting and underfitting issues are addressed. Model parameters are adjusted to achieve better results

8. Deployment

After selecting the best model, it is deployed into a real-time system or application. The model can be integrated into web or business applications. It predicts whether a customer is likely to churn. Based on predictions, companies can take preventive measures

9. Monitoring and Maintenance

The final step involves monitoring the model after deployment to ensure consistent performance. New data is collected and used to update the model periodically. Performance metrics are checked regularly. If accuracy decreases, the model is retrained.

X. LIMITATIONS AND DISCUSSIONS

Limitations

The Customer Churn Prediction project has several limitations that affect its overall performance. One major issue is the presence of imbalanced datasets, where churn cases are fewer than non-churn cases, leading to biased predictions. Data quality problems such as missing values, noise, and incomplete information can reduce model accuracy. The model may not effectively capture changing customer behavior over time. Complex models like Random Forest and Neural Networks lack interpretability, making it difficult to understand predictions. Additionally, the model may face scalability issues when applied to large real-time systems. Limited features and dependency on historical data also restrict prediction capability.

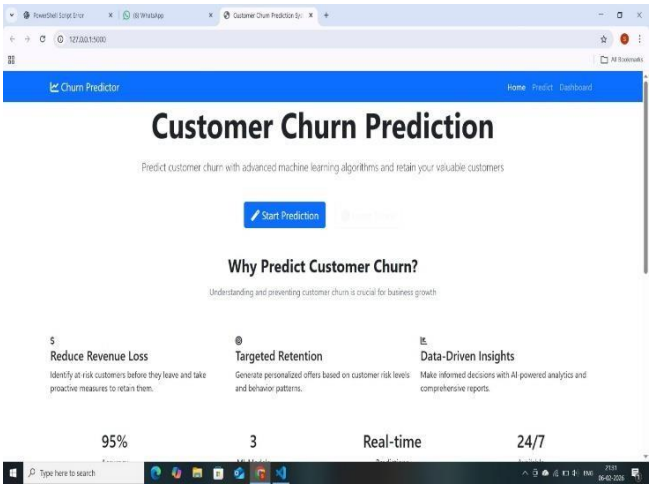
Discussion

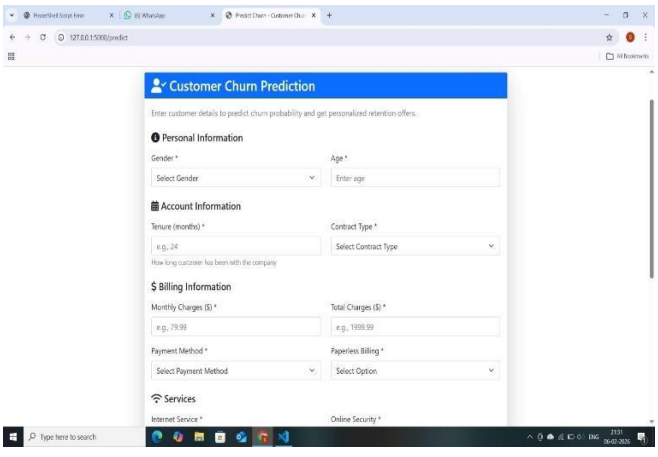
Despite these limitations, the project successfully demonstrates the effectiveness of machine learning in predicting customer churn. Proper data preprocessing and feature engineering improve model performance significantly. Ensemble models like Random Forest provide better accuracy and reliability compared to basic models. The system helps identify key factors influencing churn, enabling businesses to take proactive retention strategies. It supports decision-making by providing actionable insights. With further improvements like handling imbalanced data and incorporating real-time analytics, the model can be made more efficient. Overall, the project proves to be useful for enhancing customer retention and business growth.

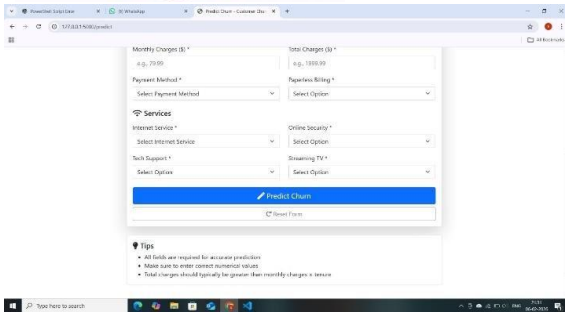
XI. RESULTS AND OUTPUTS

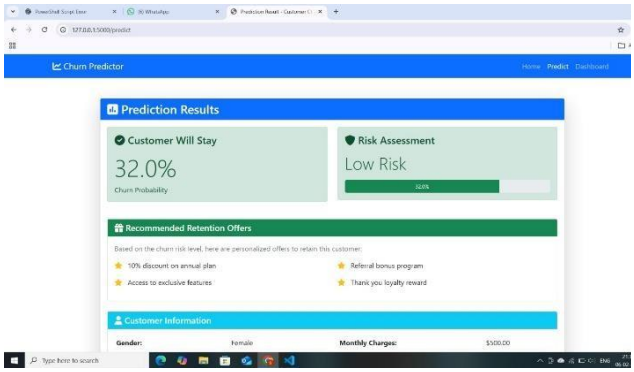
The results of the Customer Churn Prediction project show that machine learning models can effectively identify customers likely to churn. Among the models used, Random Forest achieved higher accuracy and better balance between precision and recall. The output of the system is a prediction indicating whether a customer will churn or not, along with probability scores. These results help businesses identify high-risk customers and take preventive actions. Overall, the system provides reliable and useful insights for improving customer retention.

Web Application Outputs









XII. CONCLUSION AND FUTURE WORK

Conclusion

The Customer Churn Prediction project successfully demonstrates the use of machine learning techniques to identify customers who are likely to leave a service. By applying data preprocessing, feature engineering, and models like Random Forest and Logistic Regression, the system achieves good prediction accuracy. The project highlights important factors influencing customer churn, helping businesses take proactive measures to retain customers. It improves decision-making and supports customer relationship management. Overall, the system proves to be effective, reliable, and useful for reducing churn and increasing business profitability.

Future Work

The project can be further enhanced by using advanced techniques such as deep learning and real-time data analysis. Handling imbalanced datasets more effectively can improve prediction performance. Integration with big data technologies can help process large-scale customer data. The model can also be improved by incorporating customer sentiment analysis from social media. Developing explainable AI models will help in better understanding predictions. Deployment as a web or mobile application can increase usability. Overall, future improvements can make the system more accurate, scalable, and suitable for real-world applications.

ACKNOWLEDGEMENT

We would like to express our sincere thanks to our project guide and faculty members for their guidance and support throughout this project. We are also grateful to our institution for providing the necessary resources. We thank our family and friends for their encouragement and support in completing this project.

We also acknowledge the contribution of all team members for their cooperation and teamwork. Their efforts and coordination helped us successfully complete this project on time.

XIII. REFERENCES

1. T. Verbraken, W. Verbeke, and B. Baesens, “A Novel Profit Maximizing Metric for Measuring Classification Performance of Customer Churn Prediction Models,” *IEEE Transactions on Knowledge and Data Engineering*, 2013.
2. J. Huang, C. Ling, “Using AUC and Accuracy in Evaluating Learning Algorithms,” *IEEE Transactions on Knowledge and Data Engineering*, 2005.
3. I. H. Witten, E. Frank, M. A. Hall, “Data Mining: Practical Machine Learning Tools and Techniques,” Morgan Kaufmann, 2016.
4. T. Hastie, R. Tibshirani, J. Friedman, “The Elements of Statistical Learning,” Springer, 2009.
5. F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, 2011.
6. Kaggle, “Telco Customer Churn Dataset,” Available: https://www.kaggle.com

7. IBM, "Customer Churn Prediction using Machine Learning," IBM Developer Resources.
8. Deloitte, "Customer Analytics and Churn Prediction Report," Deloitte Insights, 2022.
9. S. Moro, P. Rita, B. Vala, "Predicting Customer Churn in Banking Using Data Mining," *Decision Support Systems*, 2015.
10. A. Idris, A. Khan, Y. S. Lee, "Intelligent Churn Prediction in Telecom: Employing mRMR Feature Selection," *Expert Systems with Applications*, 2012.
11. W. Verbeke, D. Martens, C. Mues, B. Baesens, "Building Comprehensible Customer Churn Prediction Models," *Expert Systems with Applications*, 2012.
12. M. Ahmed, S. Mahmood, "Customer Churn Prediction Using Machine Learning Techniques," *International Journal of Advanced Computer Science*, 2023.
13. R. Kumar, P. Singh, "Customer Churn Prediction Using Random Forest and XGBoost," *International Journal of Data Science*, 2024.
14. S. Sharma, A. Gupta, "Explainable AI for Customer Churn Prediction Using SHAP," *IEEE Access*, 2023.
15. P. Reddy, K. Reddy, "Customer Churn Prediction Using Deep Learning Techniques," *International Conference on AI and Data Science*, 2025.