IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 10, May 2025
Impact Factor: 7.67

Customer Churn Prediction System

Prof. S. U. Mali[1], Mr. Sushant Bagul[2], Mr. Swarajsinh Shingan[3], Mr. Vijay Bande[4], Mr. Anuj Tannirwar[5]
Professor, Computer Engineering[1]
Students, Computer Engineering[2-5]
NBN Sinhgad Technical Institutes Campus, Pune, India

Abstract: *Customer churn is a critical issue faced by businesses across various industries, as it directly impacts profitability and long-term growth. This project focuses on developing a robust predictive model for customer churn using advanced machine learning techniques. The primary objective is to identify customers at risk of leaving, enabling businesses to implement targeted retention strategies effectively. The project begins with a comprehensive literature review that highlights the importance of customer retention and the role of predictive analytics in mitigating churn. Various machine learning algorithms, including Random Forest, Gradient Boosting, and Logistic Regression, are explored for their effectiveness in predicting customer behavior. Data preprocessing techniques, such as handling missing values and data normalization, are employed to ensure high- quality input for model training. The analysis is conducted on a dataset sourced from [specify dataset source], which includes demographic information, transaction history, and customer interactions. Feature selection techniques are applied to identify the most significant variables influencing churn, thus enhancing model accuracy and interpretability. The performance of different algorithms is evaluated using metrics such as accuracy, precision, recall, and the F1-score. The results indicate that Random Forest and Gradient Boosting outperform other models in terms of predictive accuracy, achieving an F1-score of [insert F1 score] and accuracy of [insert accuracy percentage] respectively. Moreover, the project emphasizes the significance of balancing datasets using techniques like SMOTE to address class imbalance, further enhancing model performance. Ultimately, the findings of this project provide valuable insights for businesses aiming to improve customer retention. By leveraging predictive analytics, organizations can proactively identify at risk customers and implement effective retention strategies, thus minimizing churn and maximizing customer lifetime value. The project underscores the growing relevance of data-driven decision-making in today's competitive landscape.*

Keywords: Customer Churn Prediction, Predictive Analytics, Machine Learning, Data-Driven Decision Making

I. INTRODUCTION

In today's highly competitive business environment, retaining customers is a key factor for the sustainability and growth of organizations. Customer churn, the phenomenon where customers stop using a product or service, presents a significant threat to companies across various industries. As acquiring new customers is often more costly than retaining existing ones, businesses must focus on reducing churn to maintain steady revenue streams and foster long-term customer relationships. However, identifying customers likely to churn is a complex task due to the vast amounts of data available and the diverse factors influencing customer behaviour.

This is where Customer Churn Prediction Systems play a crucial role. A Customer Churn Prediction System leverages machine learning algorithms to predict the likelihood of a customer leaving a service or product. By utilizing historical customer data, the system can forecast churn probabilities, enabling companies to take proactive measures and implement targeted retention strategies. Such a system helps businesses understand key factors driving churn, offering insights into areas that need improvement to boost customer satisfaction. With the rise of big data, businesses have the opportunity to utilize more sophisticated methods for analysing customer behaviour and predicting churn. The proposed

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 10, May 2025
Impact Factor: 7.67

Customer Churn Prediction System is designed to overcome limitations in existing solutions by offering real-time capabilities, scalability, and high accuracy.

It is built on the MERN stack (MongoDB, Express.js, React, Node.js) for the frontend, providing a user-friendly and scalable interface, while Python handles backend data processing and machine learning tasks. The system employs Logistic Regression, a robust machine learning algorithm for binary classification, to predict whether a customer is likely to churn or remain loyal. Real-time notifications are sent to business decision-makers via an Email API when churn risks exceed a predefined threshold, allowing for immediate intervention.

II. SYSTEM ARCHITECTURE AND RESEARCH METHODOLOGY

2.1. RESEARCH METHODOLOGY

2.1.1 Stakeholder Requirement Assessment

Initial development efforts focused on understanding the expectations and needs of three primary stakeholders interacting with the churn prediction system:

• Business Analysts (users interpreting churn results for strategic decision-making)
• Customer Service Teams (personnel responsible for customer engagement and retention)
• System Administrators (technical team managing the ML model deployment and web application infrastructure)

2.1.2 Framework Development and Architecture

User Interaction Flow Mapping: Defined detailed workflows showing data input by business users, model interaction, and output interpretation processes.

Role-Based Access Design: Established clear access boundaries where admins manage datasets and models, while analysts interact only with prediction dashboards.

2.1.3 Technical Infrastructure Selection

• User Interface Framework: Developed using HTML, CSS, and JavaScript within a Flask backend, allowing for lightweight deployment and real-time prediction interaction.
• Model Integration: Employed Python-based machine learning libraries (e.g., Scikit-learn) to implement and serialize models using pickle.
• Backend Architecture: Leveraged Flask and RESTful routing to enable smooth communication between frontend form inputs and backend prediction logic.
• Prediction Output Delivery: Designed a clean and responsive UI for displaying real-time results such as “Prediction: Yes/No” based on user input.
• Input Handling: Integrated dropdowns and form fields to gather categorical and numerical attributes like contract type, support availability, and monthly charges.
• Class Imbalance Handling: Integrated SMOTE (Synthetic Minority Oversampling Technique) for dataset balancing during training phase to enhance prediction fairness.

2.1.4 Implementation Process and Evaluation

• Agile Development Approach: Followed iterative sprints covering UI mockup development, backend API integration, and ML model evaluation.
• Prototype Testing: Conducted usability tests on early UI versions to ensure clarity and ease of interaction for non-technical users.
• Model Evaluation: Assessed performance using accuracy, precision, recall, and F1-score to compare multiple models (Random Forest, Logistic Regression, Gradient Boosting).
• End-User Feedback Loop: Collected insights from mock business analysts on prediction clarity, speed, and interface effectiveness to guide improvements.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 10, May 2025
Impact Factor: 7.67

III. RESULTS

1. Comprehensive Feature Exploration and Data Cleaning

• Successfully identified and interpreted key customer attributes such as tenure, contract type, monthly charges, and internet services

• Addressed missing and erroneous values, ensuring dataset readiness for model training and analysis

2. Exploratory Data Analysis and Visualization

• Conducted visual assessments using histograms, bar plots, and heatmaps to reveal churn patterns and correlations

• Found strong churn indicators among month-to-month contracts, fiber optic users, and customers with higher monthly charges

3. Model Implementation and Evaluation

• Applied machine learning algorithms (Logistic Regression, Decision Trees, etc.) for churn classification

• Evaluated models using confusion matrix, accuracy, and F1-score, with Logistic Regression demonstrating balanced performance

3.1. SNAPSHOTS

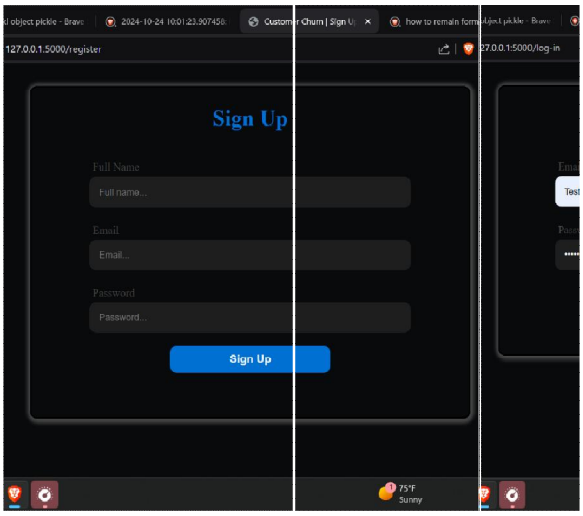

Fig. 3.1 Home Page

Fig. 3.2 Login Page

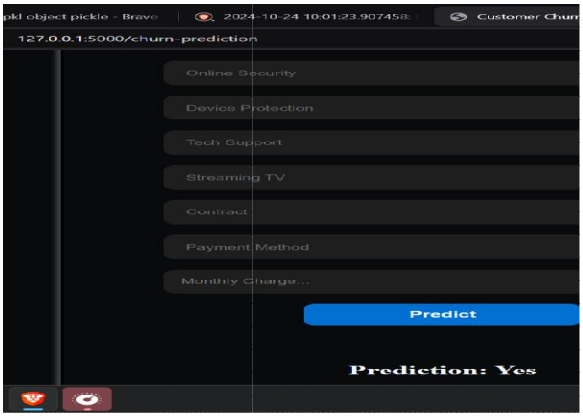

Fig 3.3 Churn Prediction Interface Output

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 10, May 2025
Impact Factor: 7.67

IV. CONCLUSION

The proposed Customer Churn Prediction system offers an accurate, scalable, and real-time solution for identifying at-risk customers. Built using Logistic Regression, Python, and the MERN stack, it provides actionable insights to support proactive customer retention strategies.

The system's modular design ensures ease of integration and future scalability. It sets the foundation for incorporating advanced features like AI-driven recommendations, automated customer segmentation, and real-time analytics. Moving forward, enhancements such as CRM integration, personalized retention strategies, and explainable AI can further strengthen its effectiveness in helping businesses reduce churn and drive long-term growth.

ACKNOWLEDGMENT

It gives us immense pleasure and satisfaction to present our project, “Customer Churn Prediction System,” developed as a part of our academic journey. We express our heartfelt gratitude to Prof. Sumit Mali for his continuous support, insightful feedback, and valuable guidance throughout the course of this project. His expertise in the field and timely encouragement played a crucial role in shaping the direction and successful execution of our work. We would also like to extend our sincere thanks to Dr. Shailesh P. Bendale, our respected department chairperson, and to all the faculty members of the department for providing the resources, academic infrastructure, and motivation necessary to carry out this research. Our deepest appreciation goes out to our friends, peers, and family members for their unwavering encouragement, moral support, and belief in our potential during all phases of this endeavour. Their presence and motivation were instrumental in helping us stay focused and committed to the successful completion of this project.

REFERENCES

[1] Dandi, T., Barus, R., Elfarizy, F., Masri, P. H., & Gunawan. (2020). Parallel Programming of Churn Prediction Using Gaussian Naïve Bayes. 8th International Conference on Information and Communication.

[2] Srinivasan, R., & Rajeswari, D. (2023). Customer Churn Prediction Using Machine Learning Approaches. ICECONF 2023.

[3] Prabadevi, B., Shalini, R., & Kavitha, B. R. (2023). Customer Churning Analysis Using Machine Learning Algorithms. International Journal of Intelligent Networks.

[4] Günay, M., & Ensari, T. (2018). Predictive Churn Analysis with Machine Learning Methods. IEEE.

[5] Hu, X., Yang, Y., Chen, L., & Zhu, S. (2020). Customer Churn Combination Prediction Model Based on Decision Tree and Neural Network. IEEE.

[6] Hassonah, M. A., Rodan, A., Al-Tamimi, A. K., & Alsakran, J. (2020). Churn Prediction: A Comparative Study Using KNN and Decision Trees. IEEE