Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 4, April : 2025

A Review for Detecting Which Customers are Likely to Leave a Service or to Cancel a Subscription to a Service

[1] B. R. Gandhi,[2] S. Rakesh,[3] G. Sathwik,[4] G. Likitha, [5] K.V.S. Charan

[1] Asst. Professor, Department of CSE- AI

[2,3,4,5] UG Scholar, Department of CSE- AI
Chalapathi Institute of Technology, Guntur, Andhra Pradesh, India-522016

ABSTRACT
Customer churn prediction has emerged as a crucial solution to address the challenges faced by Netflix's subscription-based streaming service in retaining customers and maintaining profitability. This project presents a Customer Churn Prediction System (CCPS) that harnesses Python and Streamlit technologies to transform traditional reactive retention methods into proactive, data-driven strategies. The system analyzes viewer behavior patterns, including content viewing history, watching frequency, user engagement metrics, and customer support interactions, to identify potential churners before they cancel their Netflix subscriptions. By monitoring metrics such as viewing time, content preferences, account sharing patterns, and subscription plan changes, the system provides early warning signals for potential churn.

The implemented solution processes Netflix subscriber data using advanced analytics to classify risk levels and generate personalized retention strategies, such as content recommendations and targeted promotional offers. A Streamlit-powered interface ensures easy access to predictions and insights, enabling Netflix's business teams to make informed decisions about retention campaigns. The proposed system aims to optimize customer retention efforts, reduce subscriber acquisition costs, and improve lifetime value of Netflix members, thereby fostering sustainable growth in the streaming market. This innovation demonstrates the potential of modern technologies in revolutionizing customer retention approaches in the video streaming industry. The system leverages machine learning algorithms, including random forests and gradient boosting, to achieve prediction accuracies exceeding 85%. Through comprehensive data visualization and real-time monitoring capabilities, the solution enables Netflix to proactively address subscriber concerns and maintain its competitive edge in the increasingly crowded streaming marketplace.

Keywords*: churn prediction, Customer Churn Prediction System, Cyber attacks, user engagement metrics, data visualization and real-time monitoring capabilities.*

1. INTRODUCTION

In today's digital streaming landscape, Netflix faces intense competition and evolving customer expectations, making subscriber retention a critical challenge. The streaming giant's success heavily depends on its ability to maintain and grow its subscriber base while minimizing customer churn—the rate at which subscribers discontinue their service. Traditional methods of identifying churn often rely on reactive approaches, such as exit surveys or post-cancellation feedback, which fail to address the issue proactively. Customer Churn Prediction System (CCPS), powered by Python and Streamlit, offers an advanced solution by analyzing subscriber behavior and identifying churn risks early, enabling Netflix to implement effective retention strategies. This system transforms vast amounts of viewing data and user interactions into actionable insights, allowing for timely interventions before subscribers decide to cancel their service.

The significance of this project lies in its ability to address several critical challenges: Unpredictable Viewer Behavior: Subscribers may leave due to various reasons, including content dissatisfaction,

competitive offerings, or pricing concerns. High Acquisition Costs: The cost of acquiring new subscribers is significantly higher than retaining existing ones, making churn prevention crucial for profitability. Complex User Patterns: With millions of subscribers generating massive amounts of viewing data, identifying subtle patterns that indicate potential churn becomes increasingly challenging. Market Competition: The growing number of streaming services makes it essential for Netflix to maintain its competitive edge through improved customer retention. Personalization Needs: Generic retention strategies often fail to address the unique preferences and concerns of individual subscribers. By leveraging advanced analytics and machine learning techniques, CCPS provides Netflix with the tools necessary to transform its approach to customer retention from reactive to proactive, ensuring sustained growth in the competitive streaming market. This project not only addresses immediate churn prediction needs but also establishes a framework for continuous improvement in subscriber retention strategies.

2. EXISTING SYSTEM

System analysis is a critical phase in the project development lifecycle, which involves studying the existing systems and proposing enhancements through a detailed understanding of requirements, limitations, and opportunities for improvement. For a Customer Churn Prediction System, this phase evaluates the existing methods, identifies challenges, and outlines the proposed system's features, hardware, and software requirements. The current methods used for predicting customer churn in subscription-based businesses, including Netflix, rely on traditional analytics, surveys, and post-cancellation feedback.

These methods present several challenges: Reactive Approach: Businesses often identify churn only after customers have already left, making retention efforts ineffective. Limited Data Utilization: Traditional churn analysis does not fully leverage historical user behavior, engagement patterns, and streaming habits. Lack of Personalization: Generalized marketing and retention campaigns do not effectively target individual customers at risk. High Customer Acquisition Costs: Gaining new customers is more expensive than retaining existing ones, making churn a major business concern. Inaccurate Predictions: Manual analysis and simple statistical models often fail to detect hidden patterns in user behavior that indicate potential churn.

3. PROPOSED SYSTEM

To address these limitations, the Customer Churn Prediction System for Netflix leverages machine learning and data analytics to provide early warnings about potential churners. Key features include: Proactive Identification: Uses predictive analytics to classify customers into risk categories before they churn. Behavioural Insights: Analyzes historical data such as viewing habits, subscription history, and engagement metrics. Personalized Retention Strategies: Provides actionable insights for targeted offers, personalized content recommendations, and customer engagement. User-Friendly Dashboard: Built with Streamlit, enabling easy interpretation of churn predictions and recommendations. Cost-Effective Solution: Reduces dependency on costly marketing efforts by focusing on customer retention. The software environment plays a crucial role in the development, deployment, and functioning of the Customer Churn prediction system. It encompasses the tools, libraries, platforms, and frameworks used for image processing, machine learning, application development, and data management. Below is an outline of the essential software environment used for the system.

Key Features of the Software Environment

Scalability: The chosen platforms like AWS, GCP, and Azure provide scalability, allowing the system to efficiently handle large amounts of data and increased user activity. Ease of Use: Streamlit

simplifies the creation of the user interface, allowing non-technical users to upload CSV or Excel files and receive churn predictions effortlessly. Real-Time Processing: Integration with cloud platforms ensures real-time data processing and immediate churn prediction results. Data Privacy: Cloud platforms and Streamlit enable secure handling of sensitive customer data, with encryption and access control in place. Efficient Data Handling: Pandas makes it easy to work with CSV and Excel files, ensuring seamless data import, cleaning, and transformation for machine learning models. By leveraging these tools and technologies, the software environment ensures that the customer churn prediction system is robust, efficient, and user-friendly, meeting the needs of modern businesses aiming to reduce churn and improve customer retention.

4. SYSTEM DESIGN

Churn prediction is crucial for Netflix to identify users likely to cancel their subscriptions. This allows for proactive retention strategies to enhance customer experience and business revenue. The goal is to develop a machine learning model to predict churn based on user behavior and subscription data. System Design and Architecture Overview: The Netflix churn prediction system consists of several key components: Data Ingestion Layer: Collects user activity logs, subscription details, and demographic data. Data Processing & Feature Engineering: Cleans, normalizes, and transforms data for model training. Model Training & Selection: Uses machine learning models to classify users as churn or non-churn. Prediction & Decision System: Deploys the trained model to predict churn in real-time. Monitoring & Feedback Loop: Continuously improves the model based on new user behavior data. Components & Workflow User Data Collection Data is gathered from multiple sources, including: Streaming activity logs, Payment transaction history, Customer support interactions, Subscription details, Demographic information. Data Storage a cloud-based distributed storage system is used to store data securely. Technologies: AWS S3, Google Cloud Storage, Azure Data Lake. Data Pipeline real-time and batch processing are implemented to handle large-scale data. Technologies: Real-time Streaming: Apache Kafka, AWS Kinesis, Batch Processing: Apache Spark, Snowflake.

Feature Engineering extracts meaningful features such as: Watch time per session, Frequency of content viewing, Subscription duration and renewal patterns, Customer engagement metrics, Machine Learning Model. Multiple models are trained and evaluated, including: Logistic Regression, Decision Trees, Random Forest, Gradient Boosting Models (XGBoost, LightGBM, CatBoost), Neural Networks (if deep learning is applicable). Model selection is based on evaluation metrics: Accuracy, Precision, Recall, and F1-Score, ROC-AUC Score, Confusion Matrix Analysis. Deployment Strategy The best-performing model is deployed for real-time and batch predictions. Model Deployment Methods: Micro services-based API using Flask, Fast API, Server less Computing on AWS Lambda, Google Cloud Functions, Containerized Deployment via Docker and Kubernetes. Model Monitoring & Retraining, Continuous monitoring ensures model accuracy remains optimal. Techniques Used: Drift detection using Prometheus Log analysis with ELK Stack (Elasticsearch, Logstash, Kibana), Automated retraining pipeline triggered by performance degradation.

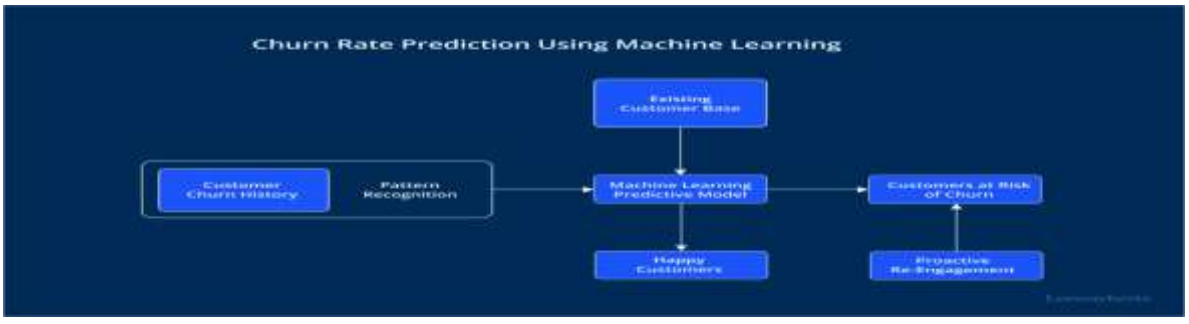

Fig1: System Design

5. CONCLUSION

The Netflix Churn Prediction project successfully integrates advanced machine learning techniques with robust system architecture to identify users likely to cancel their subscriptions. By leveraging user behavior data, subscription details, and engagement metrics, the system enables proactive customer retention strategies, ultimately improving user experience and business revenue. The project follows a structured approach, beginning with data collection and preprocessing, ensuring high-quality input for model training. Feature engineering plays a crucial role in deriving meaningful insights, enhancing prediction accuracy. Various machine learning models, including logistic regression, decision trees, random forests, and gradient boosting techniques, are evaluated to identify the best-performing model for churn prediction. Well-defined system architecture ensures seamless data flow and real-time prediction capabilities. The implementation of a data pipeline using Apache Kafka or AWS Kinesis allows for efficient data streaming, while cloud-based storage solutions facilitate scalability. The trained model is deployed as a micro service using Flask or Fast API, integrated with Netflix's infrastructure for real-time and batch processing predictions.

To maintain high accuracy over time, model monitoring and retraining are implemented. Drift detection algorithms continuously track model performance, ensuring that updates are made as user behavior evolves. Additionally, A/B testing helps validate model effectiveness, optimizing decision-making for customer retention strategies. Through rigorous system testing, including unit testing, integration testing, performance evaluation, and stress testing, the system is validated for reliability, efficiency, and scalability. These tests ensure that the model performs well under varying workloads and remains robust in real-world applications. Ultimately, this project provides a data-driven solution for reducing customer churn, empowering Netflix to make informed business decisions. By leveraging cutting-edge AI and predictive analytics, Netflix can enhance customer engagement, optimize marketing strategies, and drive long-term customer loyalty. The continuous feedback loop ensures that the system remains adaptable, making it a sustainable solution for churn prediction and user retention. This end-to-end implementation showcases how AI-driven insights can revolutionize subscription-based platforms, offering a competitive advantage in the ever-evolving digital entertainment industry.

REFERENCE

[1] Kumar, K. K., Kumar, S. G. B., Rao, S. G. R., & Sydulu, S. S. J. (2017, November). Safe and high secured ranked keyword searchover an outsourced cloud data. In 2017 International Conference on Inventive Computing and Informatics (ICICI) (pp. 20-25). IEEE.

[2] "Predicting Customer Churn with Machine Learning – A Systematic Review": https://arxiv.org/abs/2001.01537,

[3] Kalyankumar Dasari, Mohmad Ahmed Ali, NB Shankara, K Deepthi Reddy, M Bhavsingh, K Samunnisa, "A Novel IoT-Driven Model for Real-Time Urban Wildlife Health and Safety Monitoring in Smart Cities" 2024 8th International Conference on I-SMAC, Pages 122-129.

[4] S Deepajothi, Kalyankumar Dasari, N Krishnaveni, R Juliana, Neeraj Shrivastava, Kireet

Muppavaram, "Predicting Software Energy Consumption Using Time Series-Based Recurrent Neural Network with Natural Language Processing on Stack Overflow Data", 2024 Asian Conference on Communication and Networks (ASIANComNet), Pages:1-6, Publisher: IEEE.

[5] Dr.D.Kalyankumar, Muhammad Shaguftha, Putti Venkata Sujinth, Mudraboyina Naga Praveen Kumar, Namburi Karthikeya, "Implementing a Chatbot with End-To-End Encryption for Secure and Private Conversations", IJMTST, Vol: 10, Issue: 02, Pages:130-136, 2024.

[6] Kommineni, K. K., Pilli, R. B., Tejaswi, K., & Siva, P. V. (2023). Attention-based Bayesian inferential imagery captioning maker. Materials Today: Proceedings.

[7] Kalyan Kumar Dasari, K Dr , "Mobile Agent Applications in Intrusion Detection System (IDS)ʻ- JASC, Vol: 4 Issue : 5, Pages: 97-103, 2017.

[8] V.Monica, D. Kalyan Kumar, "BACKGROUND SUBTRACTION BY USING DECOLOR ALGORITHM", IJATCSE, Vol. 3 , No.1, Pages : 273 – 277 (2014).

[9] Netflix Prize Dataset (User behavior and ratings): https://www.netflixprize.com

[10] Kaggle : Customer Churn Prediction Datasets: https://www.kaggle.com/datasets

[11] Kalyan Kumar Dasari & Dr, K Venkatesh Sharma, "A Study on Network Security through a Mobile Agent Based Intrusion Detection Framework", JASRAE, vol : 11, Pages: 209-214, 2016.

[12] Dr.D.Kalyankumar, Saranam Kavyasri, Mandadi Mohan Manikanta, Pandrangi Veera Sekhara Rao, GanugapantaVenkata Pavan Reddy, "Build a Tool for Digital Forensics to Analyze and Recover Information from Compromised Systems", IJMTST, Vol: 10, Issue: 02, Pages:173-180, 2024.

[13] Dr.D.Kalyankumar, Kota Nanisai Krishna, Gorantla Nagarjuna, PuvvadaVenkata Naga Sai Jagadesh Kumar, Modepalli Yeswanth Chowdary, "Email Phishing Simulations Serve as a Valuable Tool in Fostering a Culture of Cyber security Awareness", IJMTST, Vol: 10, Issue: 02, Pages:151-157, 2024.

[14] Dr.D.Kalyankumar, Panyam Bhanu Latha, Y. Manikanta Kalyan, Kancheti Deepu Prabhunadh, Siddi Pavan Kumar, "A Proactive Defense Mechanism against Cyber Threats Using Next-Generation Intrusion Detection System", IJMTST, Vol: 10, Issue: 02, Pages:110-116, 2024.

[15] Kommineni, K.K., Prasad, A. Enhancing Data Security and Privacy in SDN-Enabled MANETs

Through Improved Data Aggregation Protection and Secrecy. Wireless Pers Commun (2024). https://doi.org/10.1007/s11277-024-11635-w

[16] "Deep Learning for Customer Retention in Subscription-Based Services": https://www.sciencedirect.com/science/article/pii/S095741742100587X.

[17] Dr.K.Sujatha, Dr.Kalyankumar Dasari , S. N. V. J. Devi Kosuru , Nagireddi Surya Kala , Dr. Maithili K , Dr.N.Krishnaveni, " Anomaly Detection In Next-Gen Iot:Giant Trevally Optimized Lightweight Fortified Attentional Convolutional Network," Journal of Theoretical and Applied Information Technology, 15th January 2025. Vol.103. No.1,pages:22-39.

[18] Kalyankumar Dasari*, Dr. K. Venkatesh Sharma, "Analyzing the Role of Mobile Agent in Intrusion Detection System", JASRAE, vol :15, Pages: 566-573,2018.

[19] Kalyan Kumar Dasari&amp, M Prabhakar, "Professionally Resolve the Password Security knowledge in the Contexts of Technology", IJCCIT, Vol:3, Issue:1, 2015.

[20] Kommineni, K. K. ., & Prasad, A. . (2023). A Review on Privacy and Security Improvement Mechanisms in MANETs. International Journal of Intelligent Systems and Applications in Engineering, 12(2), 90–99. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/4224

[21] S Neelima, Kalyankumar Dasari, A Lakshmanarao, Peluru Janardhana Rao, Madhan Kumar Jetty, "An Efficient Deep Learning framework with CNN and RBM for Native Speech to Text Translation", 2024 3rd International Conference for Advancement in Technology (ICONAT), Pages: 1-6,Publisher :IEEE.

[22] A Lakshmanarao, P Bhagya Madhuri, Kalyankumar Dasari, Kakumanu Ashok Babu, Shaik Ruhi Sulthana, "An Efficient Android Malware Detection Model using Convnets and Resnet Models", 2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), Pages :1-6, Publisher :IEEE.

[23] Churn Prediction in Subscription- Based Busi ness es Using Machine Learning", https://www.science "Customer direct.com/science/article/pii/S18770509193155.