HOT STAR CUSTOMER CHURN ANALYSIS AND PREDICTION

[1] A Veerender, [2] L Alekhya, [3] G Shravani, [4] M Sreekanth, [5] Y Venkatesh
[1]Assistant Professor, [2345]Students
Department of CSE
Siddhartha Institute of Technology & Sciences, Narapally
veerender@siddhartha.org.in, 23tq1a0504@siddhartha.co.in, 24tq5a0508@siddhartha.co.in, 23tq1a0505@siddhartha.co.in, 23tq1a0513@siddhartha.co.in,

Abstract

The HOTSTAR CUSTOMER CHURN ANALYSIS AND PREDICTION system is designed to analyze customer behavior and predict whether a customer is likely to stop using or subscribing to the Hotstar streaming platform. With the rapid growth of online entertainment services, customer retention has become an important challenge for streaming platforms. Customers may discontinue their subscriptions due to factors such as subscription cost, content availability, viewing frequency, service experience, and competition from other platforms. The proposed system uses historical customer data to identify patterns and factors associated with customer churn. The collected data may include customer details, subscription plan, subscription duration, viewing frequency, watch time, payment method, content preferences, and engagement level. The data is first preprocessed by handling missing values, duplicate records, inconsistent data, and categorical attributes. Exploratory data analysis is performed to understand customer behavior and identify important churn patterns. Machine-learning algorithms such as Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine can be used to build the prediction model. The trained model predicts whether a customer is likely to Churn or Stay based on their characteristics and usage behavior. Model performance can be evaluated using accuracy, precision, recall, F1-score, and ROC-AUC. The system can also provide charts, graphs, and dashboards to visualize customer demographics, subscription patterns, engagement levels, and churn rates. The prediction results can help identify high-risk customers and support targeted customer-retention strategies. Overall, the HOTSTAR CUSTOMER CHURN ANALYSIS AND PREDICTION system demonstrates how data analytics and machine learning can be used to understand customer behavior, predict potential churn, and support effective customer retention and data-driven business decisions.

I. Introduction

The rapid growth of online streaming platforms has transformed the entertainment industry by providing users with convenient access to movies, television programs, web series, sports, and other digital content. Hotstar is a popular streaming platform that provides a wide variety of entertainment and live content to users. With increasing competition among streaming services, retaining existing customers has become an important challenge. Customers may discontinue their subscriptions because of subscription cost, lack of preferred content, low engagement, poor service experience, or availability of alternative platforms. Therefore, analyzing customer behavior and identifying users who are likely to leave the platform can help improve customer retention.

The HOTSTAR CUSTOMER CHURN ANALYSIS AND PREDICTION system is developed to analyze customer information and identify patterns related to customer churn. Customer churn refers to a situation in which a customer stops using or subscribing to a service. The system analyzes historical customer data to understand the factors that may influence churn. Information such as subscription plan, subscription duration, viewing frequency, watch time, payment method, content preferences, customer demographics, and engagement level can be considered during the analysis.

The system begins with the collection of historical customer data from suitable datasets or records. The collected data may contain missing values, duplicate entries, inconsistent information, and categorical variables. Data preprocessing techniques are applied to clean and transform the dataset into a suitable format. After preprocessing, exploratory data analysis is performed to understand customer characteristics and identify relationships between customer behavior and churn. Various charts, graphs, and statistical techniques can be used to represent churn patterns clearly.

Machine-learning techniques are then used to predict the likelihood of customer churn. Classification algorithms such as Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine can be trained using historical customer records. The models learn the patterns of customers who have remained active and those who have stopped using the service. After training, the selected model can analyze the information of an existing or new customer and predict whether the customer is likely to churn or remain active.

II. Literature Survey

The increasing popularity of Over-The-Top (OTT) streaming platforms has created a strong need for understanding customer behavior and predicting customer churn. Customer churn refers to the situation in which a subscriber stops using or renewing a streaming service. Researchers have applied machine-learning and data-mining techniques to identify customers who are likely to leave and to understand the factors responsible for churn. A recent review of OTT customer-churn research examined 185 studies published between 2018 and 2024 and highlighted the growing importance of machine learning for customer retention in online streaming and other subscription-based industries.

Mohan and Jahav investigated customer churn prediction on OTT platforms, particularly customers who maintain subscriptions with multiple service providers. Their study examined factors influencing OTT customer churn and explored machine-learning classifiers including Decision Tree, Random Forest, AdaBoost, and Gradient Boosting. The research demonstrates that machine-learning classification can be used to identify customers who are at greater risk of discontinuing their OTT subscriptions.

Barsotti et al. conducted a comprehensive survey of churn-prediction techniques and reviewed the algorithms, datasets, evaluation metrics, and challenges used in customer-churn research. Their study shows that traditional machine-learning algorithms such as Logistic Regression, Decision Trees, Support Vector Machines, and Random Forests are widely used, while newer research increasingly investigates deep learning and explainable approaches. The researchers also identified challenges including data quality, class imbalance, interpretability, and the need to connect predictive models with effective retention strategies.

Research on customer churn prediction using machine-learning techniques has also demonstrated the importance of analyzing customer usage and behavioral information. Studies have applied classification methods to identify customers likely to discontinue their subscriptions and determine possible reasons for leaving. Such research emphasizes that customer usage patterns, service experience, pricing, and customer satisfaction can provide valuable information for developing churn-prediction models and customer-retention strategies.

Recent systematic research has further shown a movement from traditional statistical models toward ensemble learning and deep-learning techniques. Models such as Random Forest, Gradient Boosting, CNN, RNN, and LSTM have been investigated for capturing complex customer behavior. Researchers have also emphasized that accuracy alone may not be sufficient when evaluating churn models, particularly when the churn dataset is imbalanced. Therefore, measures such as Precision, Recall, F1-score, and ROC-AUC are important when evaluating customer-churn prediction systems.

A 2025 systematic review of customer-churn prediction research found that integrating different data sources can improve predictive capabilities and support customer-retention strategies. The study also highlighted the growing use of deep-learning approaches and identified data privacy and methodological challenges as important areas for further research. These findings are relevant to OTT platforms because streaming services can generate extensive behavioral information through customer subscriptions and content usage.

From the existing literature, it can be observed that considerable research has been conducted on customer-churn prediction in telecommunications and subscription-based services, while OTT-specific churn prediction remains an emerging area. Existing studies commonly focus on predicting churn using customer demographics, subscription information, and usage behavior, but there is an opportunity to combine customer churn analysis, visualization, behavioral analysis, and machine-learning prediction into one integrated system specifically designed around an OTT platform such as Hotstar.

The proposed HOTSTAR CUSTOMER CHURN ANALYSIS AND PREDICTION system addresses this gap by analyzing customer-related data, identifying important churn patterns, visualizing customer behavior, and applying machine-learning algorithms to predict whether a customer is likely to churn or remain active. The system can help identify high-risk customers and provide analytical information that can support customer-retention strategies.

III. System Analysis

The **Hotstar Customer Churn Analysis and Prediction** system is designed to analyze customer behavior and predict whether a customer is likely to stop using the streaming service. OTT platforms generate large amounts of customer data through subscriptions, viewing activity, payment information, and content interaction. The system uses historical customer data to identify patterns related to customer retention and churn. Important attributes such as subscription plan, subscription duration, watch time, viewing frequency, payment method, and customer engagement can be analyzed. The collected data is first cleaned and preprocessed to handle missing values, duplicate records, and inconsistent information. Exploratory data analysis is performed to understand customer characteristics and identify factors associated with churn. Data visualization techniques are used to represent customer trends through charts, graphs, and dashboards. Relevant features are selected and prepared for the machine-learning prediction process. Classification algorithms can then be trained using historical customer records containing churn and non-churn outcomes. The trained model predicts whether a customer is likely to churn or remain active. Model performance can be evaluated using accuracy, precision, recall, F1-score, and ROC-AUC. Overall, the system converts customer data into useful analytical and predictive information that can support customer-retention decisions. Research on OTT churn prediction has similarly explored machine-learning classification methods for identifying factors influencing customer churn.

Existing System

In the existing system, OTT platforms generally collect customer information for subscription management, billing, content delivery, and service administration. Customer records may contain details about subscriptions, payments, viewing activity, and account usage. Traditional systems mainly focus on storing and managing these records rather than predicting future customer behavior. Customer churn may be identified only after a user cancels a subscription or becomes inactive for a certain period. Manual analysis of customer information can become difficult when the number of users and transactions increases. Traditional reporting systems may provide basic information such as active subscriptions, cancelled subscriptions, and revenue. However, they may have limited capabilities for identifying hidden patterns that indicate a customer is likely to leave. Customer behavior can also change over time because of content preferences, pricing, competition, and engagement levels. Without predictive analysis, it can be difficult to identify high-risk customers before they churn. Existing systems may also require manual effort to prepare detailed customer-retention reports. Therefore, traditional approaches may not provide sufficient predictive support for proactive customer-retention strategies. The growing use of machine learning in churn prediction demonstrates the potential for moving from reactive churn identification toward predictive customer analysis.

Disadvantage of Existing System

- Customer churn may be identified only after the customer has already left.
- Limited ability to predict customers who are likely to churn.
- Manual analysis of large customer datasets can be time-consuming.
- Difficulty in identifying hidden customer behavior patterns.
- Limited use of historical customer data for prediction.
- Traditional reports may provide only basic customer statistics.
- Difficult to identify high-risk customers at an early stage.
- Greater dependence on manual analysis and predefined rules.
- Limited visualization of customer churn trends.

- Difficult to analyze multiple customer attributes simultaneously.
- Customer-retention strategies may not be sufficiently personalized.
- Large datasets can be difficult to analyze efficiently using traditional methods.

Proposed System

The proposed **Hotstar Customer Churn Analysis and Prediction** system uses data analytics and machine-learning techniques to identify customer behavior patterns and predict potential churn. The system collects customer-related information such as subscription plan, subscription duration, viewing frequency, watch time, payment method, content preferences, and engagement level. Historical customer records are used to train the prediction model and learn patterns associated with customers who churn and customers who remain active. The collected dataset is first preprocessed by handling missing values, duplicate records, inconsistent data, and categorical variables. Exploratory data analysis is then performed to understand customer characteristics and identify important factors affecting churn. Visualization techniques are used to display churn rates, subscription patterns, customer engagement, and other important information. Relevant features are selected and transformed into a suitable format for machine-learning algorithms. Classification models such as Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine can be applied to the prepared dataset. The models are evaluated using suitable performance metrics, and an effective model is selected for prediction. The selected model can predict whether an existing or new customer is likely to churn or remain active. The system can also provide a dashboard containing analytical charts, customer statistics, and prediction results. Overall, the proposed system helps identify high-risk customers earlier and provides information that can support proactive customer-retention strategies. Machine-learning research in OTT environments has specifically examined churn prediction and the factors influencing subscriber behavior.

Advantages of Proposed System

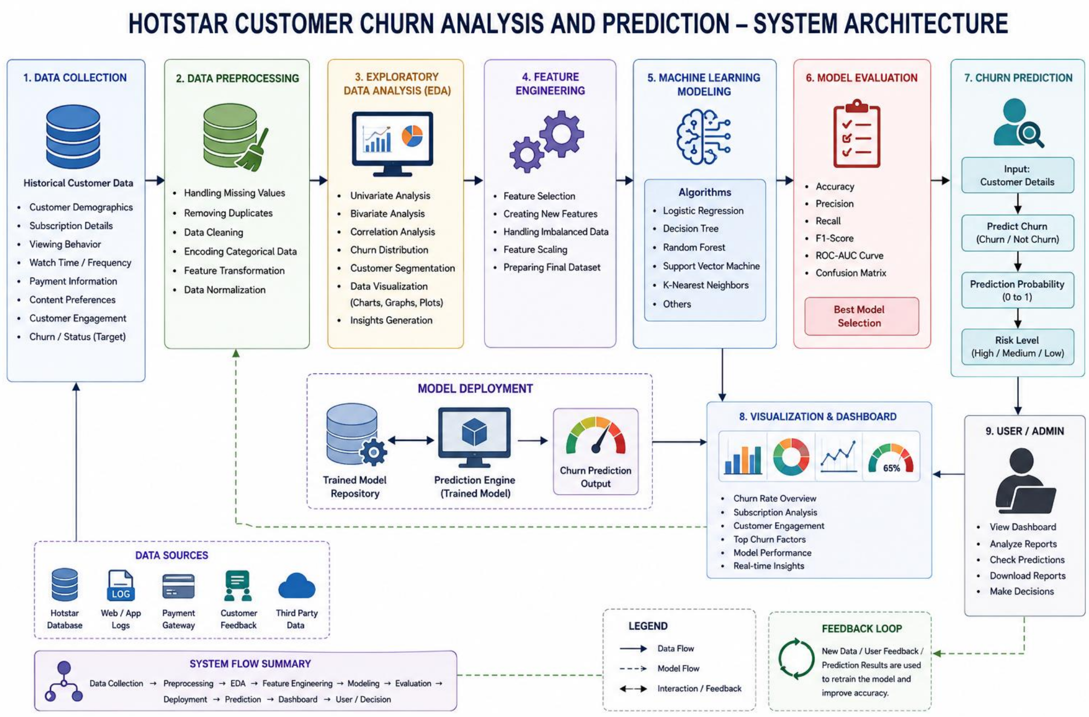
- Predicts customers who are likely to churn.
- Helps identify high-risk customers at an early stage.
- Analyzes large amounts of customer data efficiently.
- Reduces manual effort in customer analysis.
- Identifies important factors associated with customer churn.
- Provides graphical visualization of churn patterns.
- Helps understand customer subscription and viewing behavior.
- Supports data-driven customer-retention strategies.
- Allows comparison of different machine-learning algorithms.
- Provides measurable model performance using evaluation metrics.
- Helps businesses focus retention efforts on high-risk customers.
- Can be extended with advanced machine-learning and deep-learning techniques.
- Improves the ability to make proactive rather than reactive retention decisions.

IV. Methodology

The methodology of the **Hotstar Customer Churn Analysis and Prediction** system consists of several stages, beginning with data collection and ending with churn prediction. First, a historical customer dataset containing subscription and behavioral information is collected. The dataset is examined to understand its attributes, data types, missing values, and churn target variable. Data preprocessing is then performed to handle missing values, duplicate records, inconsistent information, and categorical variables. Categorical data is converted into numerical form using suitable encoding techniques. Exploratory data analysis is performed to understand customer demographics, subscription patterns, engagement levels, and churn distribution. Data visualization is used to identify relationships between customer attributes and churn behavior. Relevant features are selected and transformed to improve the effectiveness of the prediction models. The processed dataset is divided into training and testing datasets for model development and evaluation. Machine-learning algorithms such as Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine can be trained using the historical data. The trained models are evaluated using accuracy, precision, recall, F1-score, and ROC-AUC to determine their performance. Finally, the best-performing model is selected and used to predict whether a customer is likely to churn or remain active. The prediction results are then presented through a user-friendly dashboard along with customer analysis and visualization. Current churn-prediction research emphasizes the importance of model evaluation, class imbalance, interpretability, and behavioral features.

System Architecture

The Hotstar Customer Churn Analysis and Prediction system architecture represents the complete flow of customer data from collection to final churn prediction. The first stage is the Data Collection Layer, which receives historical customer information such as subscription details, viewing behavior, payment information, and engagement data. The collected information is transferred to the Data Preprocessing Layer, where missing values, duplicate records, inconsistent data, and categorical attributes are handled. The cleaned data is then passed to the Exploratory Data Analysis Layer, where customer behavior and churn patterns are analyzed using statistical techniques and visualizations. The Feature Engineering Layer selects and transforms important customer attributes required for prediction. The processed data is then provided to the Machine Learning Layer, where different classification models are trained using historical churn data. The trained models are evaluated using suitable performance metrics to identify an effective prediction model. The selected model is connected to the Churn Prediction Layer, which receives customer information and generates the predicted churn status. The Analytics and Visualization Layer presents customer statistics, churn rates, subscription patterns, and prediction results using charts, graphs, and tables. The User Interface Layer allows administrators or authorized users to enter customer details and view the generated prediction and analytical information. The system can also maintain prediction results for further analysis and model improvement. Thus, the overall architecture follows the flow Customer Data → Data Preprocessing → Exploratory Analysis → Feature Engineering → Machine Learning → Model Evaluation → Churn Prediction → Visualization → User, providing an organized framework for customer churn analysis and prediction. This architecture is consistent with the broader research direction of combining customer data, feature engineering, machine learning, and predictive analytics for churn management.



V. Result and Output

VI. Conclusion

The HOTSTAR CUSTOMER CHURN ANALYSIS AND PREDICTION system provides an effective approach for analyzing customer behavior and predicting customers who are likely to discontinue their subscription or stop using the streaming platform. The system uses historical customer data to identify important patterns related to subscription behavior, viewing activity, customer engagement, payment methods, and other relevant factors. Data preprocessing improves the quality of the dataset by handling missing values, duplicate records, and categorical information. Exploratory data analysis and visualization help in understanding customer characteristics and identifying major factors associated with churn. Machine-learning algorithms can be applied to classify customers into churn and non-churn categories based on their historical behavior. The performance of different models can be evaluated using accuracy, precision, recall, F1-score, and ROC-AUC to select a suitable prediction model. The system also provides dashboards and graphical reports that make customer and churn information easier to understand. By identifying high-risk customers at an early stage, the system can help organizations develop appropriate customer-retention strategies such as personalized content, engagement campaigns, and suitable subscription offers. This can reduce customer loss and improve long-term customer relationships. The system can be further enhanced by incorporating real-time customer activity, advanced deep-learning algorithms, recommendation systems, and explainable AI techniques. Overall, the project demonstrates how data analytics and machine learning can be used to analyze customer churn, predict potential customer loss, and support data-driven retention strategies for OTT streaming platforms.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, “Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks,” in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, “Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology,” in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, “An advanced movie recommender using collaborative filtering and sentiment analysis,” International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, “Automatic crop recommendation system using LightGBM and decision tree machine learning models,” Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, “Smart agriculture through IoT and machine learning for analyzing carbon footprints,” in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, “Soil image classification using transfer learning approach: MobileNetV2 with CNN,” SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.