Churn Analysis in Telecom Industry Using Predictive Analytics(USA)

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

By

GUTTA BHARATH

ABSTRACT

Customer churn is a critical challenge in the highly competitive telecom industry, where retaining existing customers is more cost-effective than acquiring new ones. This project focuses on analyzing customer behavior and predicting churn using data-driven and machine learning-based approaches. The primary objective of the study is to identify key factors influencing customer churn and to develop predictive models that can classify customers based on their likelihood of leaving the service.

The study utilizes a telecom customer dataset containing demographic details, service usage patterns, billing information, and contract-related attributes. Data preprocessing and exploratory data analysis (EDA) were performed to uncover meaningful patterns and relationships associated with churn. Multiple machine learning models, including Logistic Regression and Random Forest, were developed and evaluated using performance metrics such as accuracy, precision, recall, and F1-score. Model explainability techniques were applied to understand the contribution of key features influencing churn decisions.

Additionally, customers were segmented into loyal, medium-risk, and at-risk groups based on churn probability. Interactive dashboards were created using Power BI to visualize churn trends, customer segments, and revenue at risk, enabling stakeholders to make informed decisions. The results demonstrate that predictive analytics combined with visualization can effectively identify high-risk customers and support proactive retention strategies. This study provides valuable insights for telecom companies to enhance customer retention, optimize business strategies, and improve overall profitability.

TABLE OF CONTENTS

1. Introduction

1.1 Introduction

The telecommunications industry plays a vital role in modern society by enabling communication, connectivity, and access to digital services. Over the years, rapid technological advancements and market liberalization have led to the emergence of multiple service providers offering similar voice, data, and value-added services. This intense competition has increased customer expectations while reducing brand loyalty, making customer retention a major challenge for telecom companies. In such a dynamic environment, understanding customer behavior and preventing churn have become essential for sustaining profitability and long-term growth.

1.2 Telecom Industry Overview

The telecom industry is characterized by intense competition, rapid innovation, and continuous price pressure. Service providers offer comparable plans, network coverage, and promotional offers, giving customers the flexibility to switch providers with minimal effort. The low switching cost, along with mobile number portability and aggressive marketing strategies, has further increased customer mobility. As a result, telecom companies must constantly focus on improving service quality, pricing strategies, and customer experience to retain their customer base and remain competitive in the market..

1.3 Customer churn concept

. Customer churn refers to the loss of customers when they discontinue their services with a telecom provider and switch to a competitor or stop using the service altogether. Churn can be broadly classified into voluntary and involuntary churn. Voluntary churn occurs when customers actively choose to leave due to dissatisfaction, better offers, or changing needs. Involuntary churn occurs due to reasons such as non-payment, fraud, or regulatory actions. Understanding the type and causes of churn is crucial for designing targeted strategies to reduce customer attrition.

1.4 Importance of Churn analysis

Churn analysis is critical because acquiring new customers is significantly more expensive than retaining existing ones. High churn rates lead to revenue loss, increased marketing costs, and reduced customer lifetime value. By analyzing churn patterns, telecom companies can identify high-risk customers and understand the factors influencing their decisions. Proactive churn analysis enables organizations to implement timely retention strategies, improve customer satisfaction, and maintain a stable revenue stream

1.5 Motivation

The growing availability of customer data and advancements in analytics have created opportunities for data-driven decision-making in the telecom industry. Traditional reactive approaches to churn management are no longer sufficient in a highly competitive market. There is a strong need to leverage data analytics and machine learning techniques to predict churn in advance and derive actionable insights. This study is motivated by the need to transform raw customer data into meaningful information that supports proactive retention strategies, enhances customer experience, and drives business growth.

2.Problem Statement and Objectives

2.1 Problem Statement

The telecom industry faces a persistent challenge of high customer churn due to intense competition, similar service offerings, and low switching costs. Customers frequently change service providers in search of better pricing, improved service quality, or enhanced benefits. This frequent customer attrition leads to significant revenue loss and increased customer acquisition costs for telecom companies. Traditional churn management approaches are mostly reactive, identifying churn only after customers have already left the service, which limits the effectiveness of retention strategies.

The problem addressed in this study is the absence of a proactive, data-driven system capable of predicting customer churn in advance. Telecom companies require an analytical framework that can process large volumes of customer data, identify key churn drivers, and accurately predict customers who are likely to leave. Additionally, presenting these insights in an interpretable and actionable format for business stakeholders remains a critical challenge.

2.2 Objectives of the Study

The objective of this study is to develop a predictive analytics framework for analyzing customer churn in the telecom industry. The study aims to examine customer demographics, service usage patterns, and billing information to identify factors influencing churn behavior. It focuses on building and evaluating machine learning models that can accurately predict customer churn and classify customers based on their risk levels. Furthermore, the study seeks to visualize churn trends and insights using interactive dashboards, enabling stakeholders to make informed decisions. Ultimately, the objective is to support telecom companies in designing effective customer retention strategies and improving overall business performance.

3.Dataset Description

The dataset used in this study contains customer-level information collected from a kaggle. It includes a combination of demographic details, service subscription data, usage patterns, billing information, and churn status. The dataset is structured and suitable for performing exploratory data analysis and predictive modeling. Each record represents an individual customer, allowing for detailed analysis of behavior patterns associated with customer churn.

3.1 Data Source

The data used for this analysis is a telecom customer churn dataset commonly used for analytical and machine learning studies. The dataset represents historical customer information and includes both churned and non-churned customers. It provides sufficient attributes to analyze customer behavior and identify key churn drivers.

3.2 Dataset Structure

The dataset consists of multiple rows representing customers and several columns representing different attributes. These attributes include numerical and categorical variables. The target variable, **Churn**, indicates whether a customer has discontinued the service. The dataset structure enables effective segmentation, predictive modeling, and visualization.

3.3 Description of Attributes

The dataset used in this study contains detailed customer-level information related to demographics, service subscriptions, account details, and billing patterns. Each record represents an individual customer and is uniquely identified using the customerID attribute, which serves only as an identifier and is not used for predictive modeling. Demographic attributes in the dataset include gender, senior citizen status, partner status, and dependent information, which help in understanding customer profiles and household characteristics.

The dataset also includes several service-related attributes that describe the type of services subscribed by customers. These attributes capture whether the customer uses phone services, multiple phone lines, and the type of internet service subscribed, such as DSL, fiber optic, or no internet service. Additional service features include online security, online backup, device protection, technical support, streaming TV, and streaming movie services. These variables provide insights into customer engagement levels and service utilization patterns.

Account-related attributes in the dataset include tenure, which represents the number of months a customer has remained with the service provider, and contract type, which indicates whether the customer is on a month-to-month, one-year, or two-year contract. Billing preferences such as paperless billing and payment method are also included, reflecting customer convenience and payment behavior.

Billing-related attributes consist of monthly charges and total charges, which represent the customer's recurring expenses and overall spending with the telecom provider. The target attribute of the dataset is churn, which indicates whether a customer has discontinued the service. This variable serves as the primary outcome for predictive modeling and analysis, enabling the identification of customers who are likely to leave the service provider.

3.4 Data Quality and Assumptions

The dataset is assumed to be accurate and representative of real-world telecom customer behavior. Basic data quality checks such as handling missing values and data type corrections are performed before analysis. The study assumes that historical customer behavior patterns are indicative of future churn tendencies, which forms the basis for predictive modeling.

4. Methodology

The methodology adopted in this study follows a structured and systematic approach to analyze customer churn in the telecom industry. It involves data collection, preprocessing, exploratory analysis, model development, evaluation, and visualization. This step-by-step approach ensures accurate prediction of customer churn and meaningful interpretation of results for business decision-making.

4.1 Data Collection

The dataset used in this study consists of historical telecom customer data containing demographic information, service usage details, billing attributes, and churn status. Each record represents an individual customer, enabling detailed behavioral analysis. The dataset serves as the foundation for predictive modeling and churn analysis.

4.2 Data Preprocessing

Data preprocessing is a critical step to ensure data quality and reliability. This includes handling missing values, correcting data types, and removing inconsistencies. Categorical variables are encoded into numerical formats to make them suitable for machine learning algorithms. Numerical features are scaled where necessary to maintain uniformity and improve model performance.

4.3 Exploratory Data Analysis (EDA)

Exploratory data analysis is performed to understand the underlying patterns and relationships within the dataset. Various statistical summaries and visualizations are used to analyze churn distribution, customer tenure, service usage, and billing behavior. EDA helps in identifying key churn drivers and guides feature selection for model development.

4.4 Model Development

Multiple machine learning models are developed to predict customer churn, including Logistic Regression and ensemble-based models such as Random Forest. The dataset is split into training and testing sets to evaluate model performance. Hyperparameter tuning is applied to optimize model accuracy.

5. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a crucial step in understanding the structure, patterns, and relationships present in the telecom customer churn dataset. The primary objective of EDA in this study is to gain insights into customer behavior, identify trends associated with churn, and detect important variables that influence customer attrition. Various statistical summaries and visualizations were used to analyze both categorical and numerical attributes in the dataset.

The initial analysis focused on the distribution of the target variable, churn. The dataset shows that a significant portion of customers have not churned, while a smaller but important segment has discontinued the service.Understanding churn distribution helps in framing effective prediction and retention strategies.

Customer tenure was analyzed to examine its relationship with churn behavior. The analysis revealed that customers with shorter tenure are more likely to churn compared to long-term customers. This indicates that new customers are at higher risk of leaving the service, possibly due to unmet expectations or competitive offers. Tenure emerged as one of the most influential factors in churn prediction.

Billing-related attributes such as monthly charges and total charges were also explored. Customers with higher monthly charges showed a higher tendency to churn, suggesting price sensitivity among users. In contrast, customers with higher total charges generally exhibited lower churn rates, indicating long-term engagement with the service provider. These findings emphasize the role of pricing strategies in customer retention.

Service and contract-related attributes were analyzed to understand their impact on churn. Customers on month-to-month contracts showed significantly higher churn rates compared to those on one-year or two-year contracts. Additionally, customers who did not subscribe to value-added services such as online security, technical support, or device protection were more likely to churn. This suggests that bundled services and long-term contracts contribute to customer loyalty.

Overall, exploratory data analysis provided valuable insights into customer behavior and highlighted key churn drivers such as tenure, contract type, service usage, and billing patterns. These insights guided feature selection and model development in subsequent stages of the study, ensuring that predictive models were built on meaningful and impactful variables.

6. Model Development and Evaluation

Model development in this study focuses on building reliable machine learning models to predict customer churn based on demographic, service usage, contract, and billing attributes. The cleaned and preprocessed dataset was first divided into training and testing subsets to ensure unbiased model evaluation. The target variable, churn, was treated as a binary classification problem, where the goal was to predict whether a customer would churn or remain with the service provider.

Multiple classification models were developed to compare performance and identify the most effective approach. Logistic Regression was implemented as a baseline model due to its simplicity and interpretability. It helped in understanding the linear relationship between customer attributes and churn probability. Decision Tree models were also explored to capture non-linear relationships and interactions among features. Additionally, ensemble techniques such as Random Forest were applied to improve prediction accuracy by combining multiple decision trees and reducing overfitting.

Model performance was evaluated using standard classification metrics including accuracy, precision, recall, and F1-score. Since customer churn datasets often exhibit class imbalance, special emphasis was placed on recall and F1-score to ensure that churned customers were correctly identified. Cross-validation techniques were used to validate model stability and generalization capability. Among the models developed, ensemble-based approaches demonstrated better performance in capturing complex patterns within the data.

To enhance trust and interpretability, model explainability techniques were applied to understand the influence of individual features on churn prediction. Key attributes such as tenure, contract type, monthly charges, and availability of support services were identified as significant contributors to churn. These insights not only validated the results obtained during exploratory analysis but also helped translate model outputs into meaningful business insights.

Overall, the model development and evaluation process demonstrated that machine learning techniques can effectively predict customer churn and provide valuable insights for proactive retention strategies. The selected model served as the foundation for customer segmentation and dashboard-based visualization in the subsequent stages of the study.

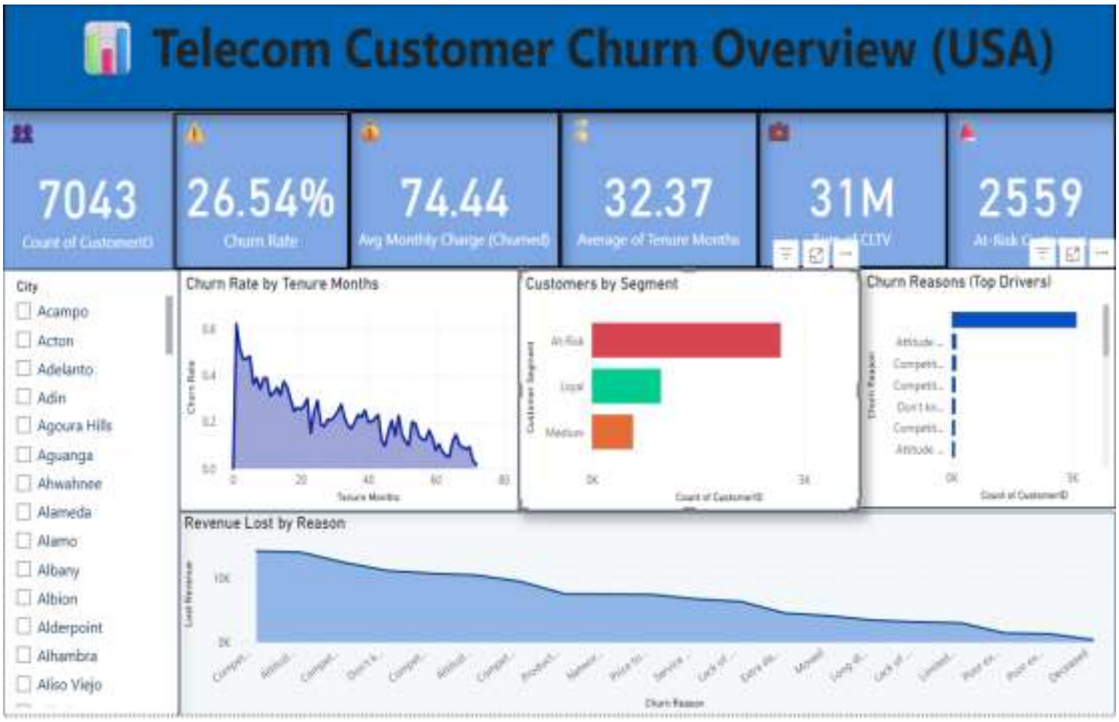
7. Visualization and Dashboard Design

Visualization plays a crucial role in transforming analytical results into meaningful business insights. In this study, Power BI was used to design interactive dashboards that present customer churn patterns, model insights, and customer segmentation in a clear and intuitive manner. The primary objective of the dashboard is to support decision-makers in understanding churn behavior and identifying high-risk customers for targeted retention strategies.

The dashboard includes key performance indicators such as overall churn rate, total number of customers, churned versus retained customers, and revenue at risk. These KPIs provide a quick overview of the current churn situation and help stakeholders assess the severity of customer attrition. Interactive elements such as slicers were incorporated to filter data based on contract type, internet service, tenure range, and payment method, allowing users to explore churn patterns across different customer segments.

Several visualizations were created to highlight important insights derived from the analysis. Bar and column charts were used to compare churn rates across contract types and service subscriptions, revealing higher churn among customers with month-to-month contracts and limited value-added services. Line and distribution charts illustrated the relationship between tenure, monthly charges, and churn behavior. Tables with conditional formatting were included to clearly identify high-risk customers based on churn scores.

Overall, the Power BI dashboard provides a comprehensive and interactive view of customer churn, enabling stakeholders to quickly identify trends, analyze contributing factors, and make data-driven decisions. The integration of predictive insights with visualization enhances the practical value of the study and supports effective customer retention and business planning.



The dashboard presents a comprehensive overview of customer churn behavior in the U.S. telecom market. It shows a total customer base of **7,043**, with an overall **churn rate of 26.54%**, indicating that more than one-quarter of customers have discontinued the service. Churned customers have a relatively high **average monthly charge of 74.44**, suggesting price sensitivity as a contributing factor. The **average customer tenure is 32.37 months**, highlighting that churn is more prevalent among customers with shorter service durations. The estimated **Customer Lifetime Value (CLTV) is around 31 million**, while **2,559 customers are identified as at-risk**, requiring immediate retention focus.

Trend analysis reveals that **churn rate decreases as tenure increases**, confirming that long-term customers are more loyal. Customer segmentation shows that the **At-Risk segment is the largest**, followed by Loyal customers, with a smaller Medium-risk group. Analysis of churn drivers indicates that **attitude-related issues and competitive factors** are the most significant reasons for customer churn. Additionally, revenue loss analysis highlights that competitive pricing and customer dissatisfaction contribute the most to financial loss.

Overall, the dashboard effectively combines churn metrics, customer segmentation, and revenue impact to support data-driven decision-making and proactive customer retention strategies.

This dashboard provides a detailed view of customer behavior and churn risk across different contract types, payment methods, and service categories. The **average churn probability is 58.70%**, indicating a significant portion of customers are at risk of leaving. Customers are classified into three risk segments, with **2,419 high-risk customers**, **2,945 medium-risk customers**, and **1,679 low-risk customers**, highlighting the need for focused retention strategies, especially for high-risk segments.

Analysis of churn by payment method shows that customers using **electronic check** experience the highest churn rate, while customers using **automatic payment methods** such as bank transfer and credit card have comparatively lower churn rates. This suggests that payment convenience plays an important role in customer retention. The churn score trend across tenure indicates that churn risk is highest during the early months of customer tenure and gradually decreases as tenure increases, reinforcing the importance of early engagement.

The relationship between monthly charges and churn score reveals that customers with **higher monthly charges tend to exhibit higher churn risk**, indicating price sensitivity among customers. Contract-based analysis further shows that **month-to-month contract customers dominate the at-risk segment**, whereas customers on one-year and two-year contracts are more likely to remain loyal. The customer table highlights individual high-risk customers, enabling targeted intervention.

Overall, this dashboard effectively combines behavioral patterns, risk segmentation, and financial indicators to support proactive churn management and data-driven decision-making.

8 . Results and Discussion

The results of this study demonstrate the effectiveness of data analytics and machine learning techniques in predicting customer churn and identifying high-risk customer segments in the telecom industry. Through exploratory analysis, model development, and dashboard visualization, several important insights were obtained that directly support business decision-making.

The predictive models developed were able to identify key churn drivers with reasonable accuracy and reliability. Attributes such as customer tenure, contract type, monthly charges, payment method, and availability of support services emerged as the most influential factors affecting churn. Customers with shorter tenure, higher monthly charges, and month-to-month contracts showed a significantly higher likelihood of churn. These findings are consistent with the trends observed during exploratory data analysis and validate the robustness of the modeling approach.

Customer segmentation based on churn risk further enhanced the interpretability of results. The analysis revealed that a large proportion of customers fall into the high-risk and medium-risk categories, indicating substantial revenue at risk. High-risk customers were predominantly associated with month-to-month contracts, electronic check payment methods, and limited service add-ons. In contrast, loyal customers typically had longer tenure, long-term contracts, and automatic payment methods, highlighting the importance of customer commitment and convenience.

The Power BI dashboards played a crucial role in presenting these results in an interactive and intuitive format. Stakeholders can easily monitor churn trends, filter customers by contract or payment type, and identify individual at-risk customers for targeted retention efforts. The integration of churn probability scores and visual analytics bridges the gap between technical model outputs and business understanding.

Overall, the results confirm that a proactive, data-driven approach to churn management can significantly improve a telecom company’s ability to retain customers. By identifying churn risk early and understanding its key drivers, organizations can design personalized retention strategies, optimize pricing and contract structures, and enhance customer satisfaction.

9. Conclusion and Future Scope

This study successfully demonstrated the application of data analytics and machine learning techniques to analyze and predict customer churn in the telecom industry. By leveraging customer demographic data, service usage patterns, contract details, and billing information, the study identified key factors influencing churn behavior. Exploratory data analysis revealed that customer tenure, contract type, monthly charges, payment method, and value-added services play a significant role in customer retention. The predictive models developed were effective in identifying customers with a high likelihood of churn, enabling proactive intervention.

The integration of Power BI dashboards enhanced the practical value of the study by presenting insights in an interactive and business-friendly manner. Customer segmentation into high-risk, medium-risk, and loyal groups provided clarity on where retention efforts should be focused. The results highlight the importance of data-driven decision-making in reducing churn, optimizing revenue, and improving customer satisfaction. Overall, the study confirms that predictive analytics can serve as a powerful tool for telecom companies to manage customer churn effectively.

In terms of future scope, the study can be extended by incorporating real-time customer data to enable live churn prediction and automated alerts. Additional external factors such as customer feedback, network quality metrics, and competitor pricing can further improve prediction accuracy. Advanced techniques such as deep learning and survival analysis may also be explored to enhance model performance. Integrating the churn prediction system with customer relationship management (CRM) platforms can help telecom companies implement personalized and timely retention strategies, leading to long-term business growth.