Customer Churn Analysis Report

Project Title:

Customer Churn Prediction in Telco Dataset Using Logistic Regression and Decision Tree

ADVISED BY

Docator Faheem Chaudhary

BACHELOR OF SCIENCE

IN

COMPUTER SCIENCE

DEPARTMENT OF COMPUTER SCIENCE

Table of Contents

1) Abstract:-

This study focuses on predicting customer churn using the Telco Customer dataset. The objective is to identify key features contributing to customer attrition and use machine learning techniques to forecast churn behavior. Data preprocessing, exploratory data analysis, and modeling were performed using Logistic Regression and Decision Tree classifiers. The models achieved respectable accuracy, and important factors such as contract type, monthly charges, and tenure were found to be strong indicators of churn. This research highlights the power of data analytics in customer retention strategies for telecom companies.

2) Introduction:-

In the telecom industry, customer churn ,the rate at which users stop doing business with a company ,is a major concern. It directly affects profitability and long-term sustainability. The ability to predict churn can help businesses take preemptive actions to retain valuable customers.

This report aims to apply data analytics and machine learning techniques to build a churn prediction model. The main objectives include identifying which customers are likely to churn and discovering which features most strongly influence this behavior. For this purpose, we use the Telco Customer dataset and employ classification algorithms such as Logistic Regression and Decision Tree.

3) Dataset Description:-

The dataset used for this analysis is the Telco Customer Churn dataset from Kaggle. It contains 7,043 customer records and 21 features. The features include customer demographics, services used, payment methods, and churn status.

Below is a summary of key columns:

Feature	Description
gender	Gender of the customer
SeniorCirizen	Whether the customer is a senior citizen (1 = Yes, 0 = No)
tenure	Number of months the customer has been with the company
Contract	Type of contract (e.g , Month-to-month, One year)
MonthlyCharges	The monthly fee charged to the customer
Churn	Whether the customer left the company (Yes/No)

The target variable is **Churn**, which we convert into binary form (1 = Yes, 0 = No) for modeling.

4) Data Preprocessing:-

Before modeling, the dataset underwent several preprocessing steps:

- Removed missing values from the **TotalCharges** column.
- Dropped duplicate rows.
- Converted categorical variables into numerical format using label encoding and one-hot encoding.
- Standardized the **MonthlyCharges** and **TotalCharges** columns for uniformity.
- Converted **Churn** column into 0 and 1 for binary classification.

5) Exploratory Data Analysis (EDA):-

Exploratory analysis helped reveal patterns and relationships in the data:

- Customers with **month-to-month contracts** had the highest churn rate.
- **Tenure** had a negative correlation with churn; newer customers are more likely to leave.
- Customers with **high monthly charges** also showed a higher tendency to churn.

A correlation heatmap and several bar charts were generated to visualize these patterns. For example, churn was much higher among those without online security or tech support services.

6) Modeling and Evaluation:-

Two classification models were implemented:

- **Logistic Regression:** A statistical method for binary classification.

- **Decision Tree Classifier:** A tree-based model that segments the data based on decision rules.

  Both models were trained and evaluated using accuracy scores and confusion matrices.

Model	Accuracy
Logistics Regression	~79.5%
Decision Tress	~81.2%

The Decision Tree provided slightly better performance but showed a risk of overfitting. Feature importance was extracted from the tree model, with contract type, tenure, and monthly charges being the most impactful features.

7) Results & Discussion:-

The modeling process confirmed that churn is influenced primarily by:

- Short tenure with the company
- Month-to-month contracts
- High service charges

Logistic Regression provided good baseline performance and interpretability, while the Decision Tree offered insight into feature interactions.

While performance was acceptable, improvement could be made using ensemble techniques like Random Forest or XGBoost.

8) Business Implications:-

Based on the findings, telecom companies can:

- Offer **loyalty discounts** to short-tenure users
- Encourage **annual contract upgrades**
- Bundle services to reduce churn among high-paying customers

A customer churn dashboard could be created to monitor and act upon these insights in real time.

9) Conclusion:-

The research demonstrates how data analytics can be used to predict customer churn with reasonable accuracy. By identifying critical patterns in customer behavior, businesses can proactively engage and retain customers.

Future improvements may include hyperparameter tuning, testing more advanced models, and using real-time data integration for actionable insights.

10) References:-

1) Kaggel Dataset:-

https://www.kaggle.com/datasets/blastchar/telco-customer-churn

2) Scikit-learn Documentation:-
https://scikit-learn.org/

3) Towards Data Science Article: *Predicting Customer Churn*
https://towardsdatascience.com/

11) **Acknowledgment:-**

I would like to express my sincere gratitude to my advisor, **Dr. Faheem Chaudhary**, for his valuable guidance and continuous support throughout this project. I also thank my peers and **faculty members** in the **Department of Computer Science** for their encouragement and insightful feedback.

12) **Appendix:-**

Appendix A – Model Code Snippets

- Logistic Regression Implementation (Python)
- Decision Tree Classifier Code

Appendix B – Extra Visualizations

- Heatmaps
- Churn by Contract Type Graph

13) **Final Note:-**

This project not only enhanced my understanding of machine learning models but also showed me how data can support real-world business decisions. I look forward to exploring more advanced techniques in future projects to improve predictive accuracy and contribute to customer retention strategies in real-world industries.

14) **Plagiarism:-**

I hereby declare that this project report titled ***“Customer Churn Prediction in Telco Dataset Using Logistic Regression and Decision Tree”*** is the result of my own work and effort, except where acknowledged, and has not been submitted elsewhere for any academic purpose.