

Customer Churn Prediction and Analysis

Made by :

Atharva Singh (01215611622)

Komal Goel (04015611622)



Introduction

- Customer churn also known as customer attrition or defection, is the rate at which customers end their relationship with a company.
- In the telecom industry, it specifically refers to the percentage of subscribers who cancel their contract or service within a given time period.
- It's a key metric that indicates customer satisfaction and the ability of a company to retain its user base.



Problem Statement and Objectives

Telecom operators face churn rates of 15-25% annually, causing significant financial loss.

Early identification of churn-prone customers is crucial to sustain profitability through targeted retention.

Objectives include data analysis, exploratory visualization, machine learning implementation, evaluation, deployment, and actionable retention strategies.

Technology Stack and Tools Used

The project uses Python for data processing and machine learning, with libraries like Pandas, NumPy, Matplotlib, and Scikit-learn.

XGBoost enhances prediction accuracy. Pickle handles model persistence, Flask provides deployment, and Power BI enables interactive dashboard visualization and decision support.

This stack is chosen for scalability, cost-efficiency, and ease of integration.

Methodology and Workflow

The project follows a structured methodology to ensure reliable and scalable results.

- Data collection includes demographics, billing, and service usage patterns.
- Preprocessing handles missing data and encodes variables.
- Exploratory Data Analysis utilizes Power BI for churn visualization.
- Feature engineering creates new relevant metrics.
- Model training uses Logistic Regression, Random Forest, Gradient Boosting, and XGBoost.
- Evaluation includes confusion matrices and ROC curves.
- Deployment is through Flask API with Pickle serialization.
- Continuous monitoring and retraining maintain accuracy.



Data Preprocessing and Feature Engineering

- Data preprocessing involves handling missing values, encoding categorical features, and normalizing data.
- High-quality inputs are ensured using Pandas and NumPy.
- Feature engineering includes creating variables such as average call duration, complaint counts, and billing consistency.
- These features enhance model prediction capabilities by capturing customer behavior nuances.

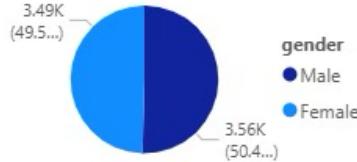
Exploratory Data Analysis Insights

- Power BI dashboards reveal churn distribution patterns across variables like tenure, plan type, and billing history.
- Visualizations identify high-risk customer segments and churn trends.
- EDA insights guide targeted retention strategies and feature selection for modeling.

TELECOM CUSTOMER CHURN ANALYSIS

CUSTOMER PROFILE

Entire- Gender Distribution



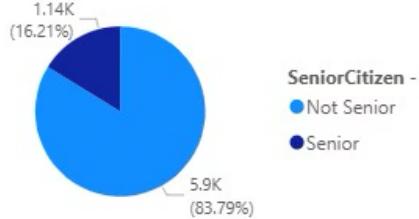
7043

Count of Entire custo...

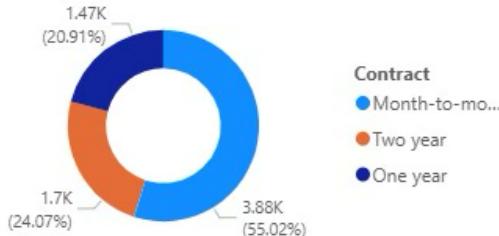
1869

Count of Churners

Count of SeniorCitizen - 1 by SeniorCitizen - 1

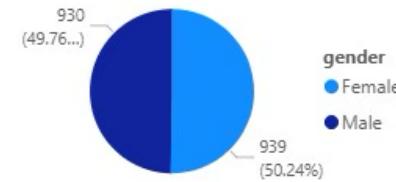


Count of Contract by Contract

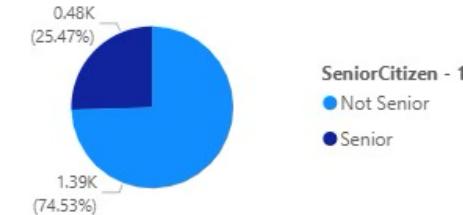


CHURNER PROFILE

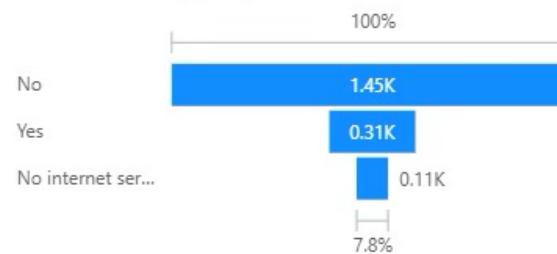
Churner- Gender Distribution



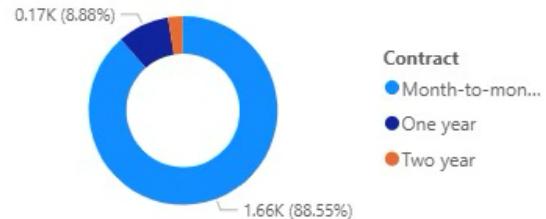
Churner SeniorCitizen - 1 by SeniorCitizen - 1



Count of TechSupport(churn)



Count of Contract by Contract(churn)



Customer Churn Prediction

Senior Citizen (0/1)

Monthly Charges

Total Charges

Tenure (Months)

Gender

Male

Partner

Yes

Dependents

Yes

Phone Service

Yes

Multiple Lines

Yes

Internet Service

DSL

Online Security

Yes

Online Backup

Yes

Device Protection

Yes

Tech Support

Yes

Streaming TV

Yes

Streaming Movies

Yes

Contract

Month-to-month

Paperless Billing

Yes

Payment Method

Electronic check

Predict Churn

Machine Learning Models and Evaluation Metrics

- Models trained include Logistic Regression, Random Forest, Gradient Boosting,
- Performance is measured using accuracy, precision, recall, F1-score, and AUC-ROC.
- Confusion matrices and precision-recall analysis assist in model assessment.
- XGBoost demonstrates superior accuracy for churn prediction in this context.



Conclusions and Future Scope

- Advanced machine learning and visualization techniques offer powerful insights into customer churn.
- The project's predictive model identifies high-risk customers, enabling proactive retention efforts.
- The technology stack supports scalability and integration with CRM systems.
- Future enhancements include real-time predictions via SQL integration and expanded industry applications.
- This solution presents a practical, enterprise-grade tool for reducing telecom churn and maximizing profitability.

Thank you

Do you have any questions ??