

Customer Churn Analysis in the Telecom Industry :

Introduction :

In the highly competitive telecom sector, customer retention is vital to maintaining revenue and market position. This project focuses on analyzing customer behavior to predict churn — the likelihood of customers discontinuing their service. By identifying high-risk customers, telecom companies can take proactive steps to reduce churn and enhance satisfaction.

Abstract :

The goal of this project was to predict customer churn using a real-world telecom dataset of 3,333 records. The data includes customer service usage patterns such as call minutes, charges, complaints, and service plans. A Random Forest Classifier was built to predict churn, followed by churn probability-based segmentation. Exploratory Data Analysis (EDA) and model interpretation using ELI5 helped identify top churn drivers including complaints, total charges, and international plan usage. Based on churn probabilities, customers were segmented into risk categories for targeted retention strategies.

Tools Used :

- Python (Pandas, Matplotlib, Seaborn, Scikit-learn)
- SQL (used offline for aggregations)
- ELI5 (for model explainability)
- Jupyter Notebook
- PowerPoint (for business presentation)

Steps Involved in Building the Project :

1. **Data Preprocessing:**

- Removed irrelevant columns like 'phone number' and 'state'
- Encoded categorical features using LabelEncoder
- Created new features: total_calls, total_minutes, total_charges

2. **Exploratory Data Analysis (EDA):**

- Analyzed churn distribution, complaints, and call behaviors
- Found higher churn rates among international plan users and those with more complaints

3. **Model Building:**

- Used Random Forest Classifier with 80/20 train-test split
- Achieved 97.45% accuracy
- Evaluated with confusion matrix and classification report

4. **Model Explainability:**

- Used ELI5 to understand which features had the greatest influence on churn predictions
- Top features: total_charges, complaints, international_plan

5. **Customer Segmentation:**

- Used predicted churn probabilities to categorize customers:
 - At Risk: $\text{churn_prob} > 0.7$
 - Loyal: $0.3 \leq \text{churn_prob} \leq 0.7$
 - Dormant: $\text{churn_prob} < 0.3$
- Exported results to churn_segments.csv

Conclusion :

The project successfully identified key behavioral and service-related indicators of churn using a machine learning model. With a high accuracy rate and meaningful feature explanations, the model enables business teams to implement precise, data-driven retention strategies. Segmenting customers based on churn probability allows targeted interventions, particularly for high-risk users, improving overall customer lifetime value.