Customer Churn Analysis
Telecom Industry

Data Analysis & Machine Learning Project by
Makkina Ramya Priya

Project Objective

- • Predict which telecom customers are likely to churn.
- • Analyze usage and service patterns to find key churn indicators.
- • Recommend strategies to retain high-risk customers.

Tools & Technologies Used

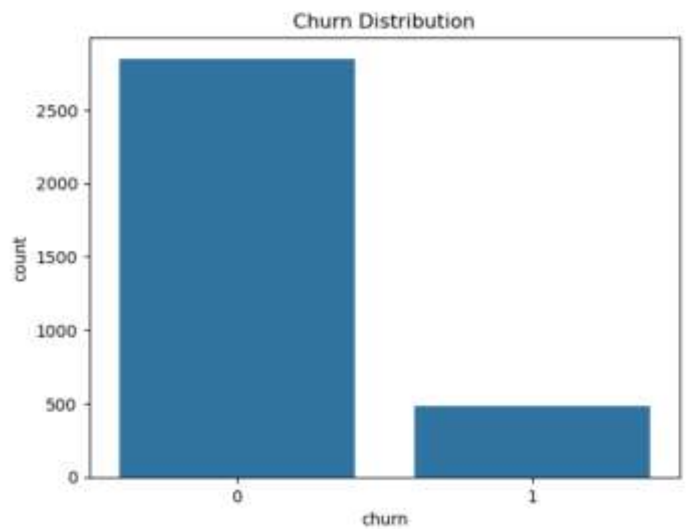
- • Python (Pandas, Seaborn, Scikit-learn (Random Forest))
- • Jupyter Notebook
- • Random Forest Classifier
- • ELI5
- • Data Cleaning & EDA
- • Churn Probability Segmentation
- Model Evaluation (Accuracy, Classification Report)

Dataset Overview

- • 3333 customer records
- • Features include: call durations, charges, plans, customer service calls
- • Target variable: churn (Yes/No)

Churn Distribution: Majority Customers Stay, But Losses Matter:

- This chart shows the number of customers who churned (1) vs. those who did not (0).
- The dataset is **imbalanced**, with more customers not churning.
- Imbalance can bias the model if not handled properly (e.g., using class weighting).



Exploratory Data Analysis (Insights)

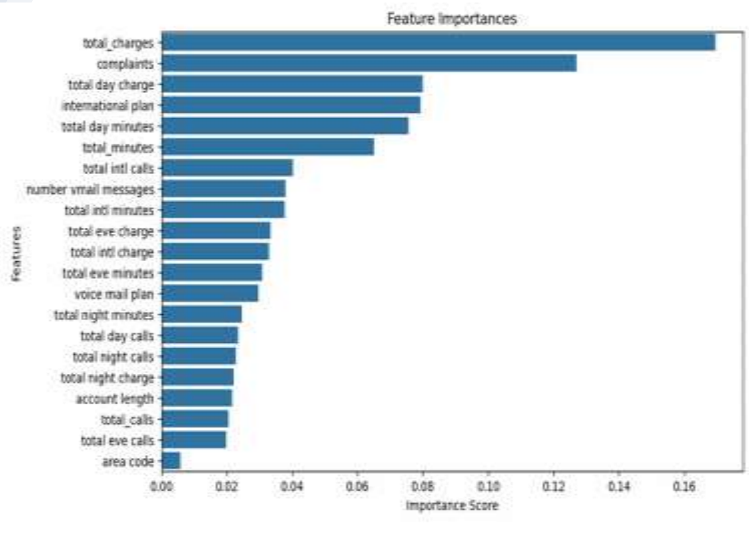
- • Higher churn among customers with:
-   - International plan = Yes
-   - Many customer service calls
-   - High complaints and low total call time
- • Churned users show different call behavior (total_minutes).

Model Summary

- • Algorithm: Random Forest Classifier
- • Train/Test Split: 80/20
- • Accuracy: 97.45%
- • Evaluation: Confusion Matrix & Classification Report.

Top Predictors of Customer Churn

- The chart highlights which customer features most influenced churn prediction based on the Random Forest model:
- 🔹 total_charges, complaints, and international_plan were the most important drivers of churn. 🔹 These insights guide actionable strategies, such as improving complaint handling and reviewing pricing plans.
- The feature importance was calculated using model-based scoring, helping prioritize business areas that directly impact churn.



Model Explainability with ELI5

•ELI5 was used to explain how the Random Forest model makes predictions.

•It highlighted **total charges**, **complaints**, and **international plan** as key drivers of churn.

•This helps non-technical teams understand and trust model outcomes.

Weight	Feature
0.1695 ± 0.1823	total_charges
0.1270 ± 0.0658	complaints
0.0800 ± 0.1125	total day charge
0.0792 ± 0.0643	international plan
0.0756 ± 0.1124	total day minutes
0.0650 ± 0.1016	total_minutes
0.0402 ± 0.0464	total intl calls
0.0379 ± 0.0631	number vmail messages
0.0377 ± 0.0446	total intl minutes
0.0334 ± 0.0406	total eve charge
0.0329 ± 0.0378	total intl charge
0.0309 ± 0.0400	total eve minutes
0.0297 ± 0.0578	voice mail plan
0.0247 ± 0.0230	total night minutes
0.0234 ± 0.0210	total day calls
0.0229 ± 0.0237	total night calls
0.0222 ± 0.0247	total night charge
0.0218 ± 0.0216	account length
0.0205 ± 0.0200	total_calls
0.0197 ± 0.0197	total eve calls
… 1 more …	

Business Recommendations
Use churn segments for targeted offers (At Risk customers)

- • Focus on customers with frequent service calls
- • Offer loyalty benefits to high-usage long-tenure customers
- • Reduce churn by improving support for international plan users
- • Use churn segments for targeted offers (At Risk customers)

Churn Segmentation Overview

- • Churn probabilities calculated using model.predict_proba()
- • Segments:
-   - At Risk → $\text{churn\_prob} > 0.7$
-   - Loyal → $0.3 \le \text{churn\_prob} \le 0.7$
-   - Dormant → $\text{churn\_prob} < 0.3$
- • Enables targeted marketing and retention strategies

Final Recommendations

- Target "At Risk" customers with personalized retention plans
- Improve service for high-complaint users to reduce churn
- Offer flexible recharge plans for low-engagement users
- Re-engage dormant users with free data or discounts
- Reward loyal users to improve brand advocacy

Thank You