Customer Churn Analysis for the Telecom Industry

Introduction

In today's competitive telecom sector, customer retention is more cost-effective than acquisition. Identifying customers likely to churn allows telecom companies to take proactive measures to retain them. This project uses machine learning to predict customer churn and recommend actionable strategies based on user behavior and model interpretability.

Abstract

This project builds a binary classification model to predict customer churn using telecom user data. It involves data aggregation via SQL and machine learning using Python's Scikit-learn. Techniques like ELI5 and SHAP were employed to interpret the model's output. Based on behavioral and model-derived insights, customers were segmented into At Risk, Loyal, and Dormant groups, allowing targeted retention strategies.

Tools Used

- Python: Scikit-learn (ML model), Pandas, SHAP/ELI5 (interpretability)
- SQL: Data aggregation (call duration, complaints, recharge patterns)
- PowerPoint: Report and visualization
- Jupyter Notebook: Model development and experimentation

Steps Involved in Building the Project

1. Data Collection & Preprocessing

   - SQL queries were used to aggregate customer-level metrics: average call duration, number of complaints, and recharge frequency.

   - The data was cleaned, encoded, and split into training and testing sets.

2. Model Building

- A binary classification model (e.g., Logistic Regression or Random Forest) was trained to predict churn.

- Hyperparameter tuning was performed using GridSearchCV for model optimization.

3. Model Evaluation

- Metrics used: Accuracy, Precision, Recall, and ROC-AUC.

- SHAP and ELI5 were used to explain predictions and identify key churn drivers (e.g., high complaints, low recharge frequency).

4. Customer Segmentation

- Based on churn probability and usage patterns:

  - At Risk: High churn probability, recent complaints.

  - Loyal: Low churn probability, consistent recharges.

  - Dormant: Low activity, moderate churn risk.

5. Reporting & Visualization

- Final results and recommendations were compiled into a PowerPoint presentation for stakeholders.

- Visual aids highlighted churn rates, feature importance, and segment profiles.

Conclusion

The project successfully identified churn-prone customers and segmented them for targeted retention strategies. The use of explainable AI tools ensured transparency and trust in model decisions. Implementing focused retention actions-like reward programs for loyal users or quick complaint resolution for at-risk customers-can significantly improve customer retention and reduce revenue loss.