

Detection of Imminent SyriaTel Churn

Predictive Signal Analysis using Machine Learning

Boniface Kimondo Njeri

Moringa School - Data Science, Phase 3

Introduction

- Predicting SyriaTel customer churn using ML techniques
- Churn prediction enables proactive retention strategies
- Analyzes customer behaviors, service use, and support history
- Improves decision-making using historical patterns
- Supports long-term customer satisfaction and revenue growth

Business Background

- Retaining customers is more cost-effective than acquiring new ones
- Churn leads to significant revenue losses
- Churn is often triggered by dissatisfaction or unmet expectations
- Predictive modeling provides data-driven customer insights
- Helps businesses act before customers disengage

Project Objectives

- Develop a churn prediction classifier
- Support proactive customer retention actions
- Analyze key behavioral churn indicators
- Generate insights to improve services and campaigns
- Deploy practical solutions based on model outputs

Tools and Technologies Used

- Python (Pandas, NumPy) for data manipulation
- Matplotlib & Seaborn for visualizations
- Scikit-learn for model building & evaluation
- Jupyter Notebook for analysis & prototyping
- PowerPoint for presenting final results

Dataset Overview

- 3,333 records with 20 features
- Target column: churn (0 = stay, 1 = churn)
- No null values found in dataset
- Features include plans, charges, and call minutes
- Balanced numerical and categorical variables

Data Preprocessing Steps

- Standardized column names to snake_case
- Dropped non-informative columns (e.g., phone_number)
- Converted binary and target values to integers
- Encoded yes/no features for model compatibility
- Verified data types and structure

Key Insights from EDA

- Churn rate is $\sim 14.5\%$ (imbalanced dataset)
- Higher churn associated with international plans
- Frequent customer service calls linked to churn
- Call duration (day) and total charges influence churn
- Important to handle imbalance in evaluation

Churn Distribution

- Majority of users are retained (85.5%)
- Churners form minority ($\sim 14.5\%$)
- Accuracy alone may not reflect model effectiveness
- Class imbalance handled during modeling
- Metrics like F1-score and recall are prioritized

Model Development Steps

- Split dataset into features (X) and target (y)
- Applied train/test split (70%/30%)
- Standardized numerical features using StandardScaler
- Trained a Random Forest Classifier (n=100)
- Tested predictions on unseen data

Why Random Forest?

- Performs well on classification problems
- Handles both categorical and continuous features
- Reduces overfitting through ensemble method
- Works well with imbalanced datasets
- Interpretable feature importance output

Model Evaluation Metrics

- Accuracy: 95.2% (strong overall performance)
- Precision (churn): 94% – few false positives
- Recall (churn): 71% – some churners missed
- F1-score: 0.81 – good balance of precision/recall
- Confusion matrix supports reliable performance

Confusion Matrix Details

- True Positives (TP): 101 churners correctly predicted
- True Negatives (TN): 851 retained customers identified
- False Positives (FP): 6 – over-predicted churners
- False Negatives (FN): 42 – missed actual churners
- Improving recall can reduce missed churners

Model Strengths

- High classification accuracy (95.2%)
- Great at identifying non-churners
- Robust against overfitting via ensemble learning
- Works well without deep hyperparameter tuning
- Provides interpretable results and business insight

Business Insights from Modeling

- High service call frequency is a churn risk indicator
- International plans linked with higher churn
- Long day call durations also increase churn probability
- Feature importance aids in campaign targeting
- Focus on specific behaviors improves retention ROI

Strategic Recommendations

- Improve service quality for international callers
- Proactively engage users with high service complaints
- Incorporate model into CRM to trigger interventions
- Monitor prediction confidence for tiered actions
- Retest model quarterly with new data

Conclusion

- Machine learning enables proactive churn detection
- Random Forest provides accurate and interpretable results
- Model insights guide SyriaTel's retention strategy
- Recall can be improved to detect more at-risk users
- Deploying the model will enhance customer loyalty

Thank You

Questions and Feedback Welcome