SyriaTel Churn Prediction - Non-Technical Presentation

Presenter: Joanne Kariuki

Stakeholder: SyriaTel Shareholders and Business Team

Project Overview

I analysed the the customer churn of SyriaTel using data_driven insights with the aim of reducing the customer churn as it leads to substantail revenue losses. I used macine learning to identify patterns in customer behaviour and build predictive models which in turn helped targeted actions to retain high-risk customers

Project Goals

- **Examined** Customer behavior and service usage patterns from the dataset to understand key indicators of churn.
- **Identified** The most influential features contributing to customer churn, such as frequent customer service interactions.
- **Recommended** Strategic actions to proactively retain high-risk customers.

Why This Project Matters

- **Competitive Landscape:** The telecom industry in Syria is competitive, with customers easily switching providers due to minimal switching costs and similar pricing.
- **High Stakes:** Losing a customer costs more than retaining one. Customer churn not only results in immediate revenue loss but also undermines brand loyalty and market share.
- **Opportunity:** By predicting churn before it happens, SyriaTel can proactively intervene, personalize retention strategies, and significantly reduce long-term losses.

Business Understanding

Stakeholder

- **Primary Stakeholder:** Shareholders and Business Tean
  - **Role:** The business team oversees customer retention, gain and service optimization

- **Interest:** find ways to identify customers at risk of leaving and reduce churn-related revenue loss.

Business Questions

1. Can we identify which customers are likely to churn before they do?
2. What customer behaviors and attributes are most associated with churn?
3. How well can we predict churn using historical data?

Data Understanding

Data Sources

2. (**https://www.kaggle.com/datasets/becksddf/churn-in-telecoms-dataset/code**)
   - **File:** `bigml_59c28831336c6604c800002a.csv`

- **Dataset Overview:** 3,333 customer records from SyriaTel, each with multiple service and usage features.
- **Target Variable:** Churn – whether the customer left the service (Yes or No).
- **Key Features Include:**

Service Usage: Day, evening, night, and international call minutes and their charges.

Service Plans: Whether the customer is subscribed to an International Plan or Voice Mail Plan.

Customer Support Interactions: Number of times the customer called customer service.

Data Limitations

- **No Customer Demographics:** The dataset lacks information like age, income, or location, which could offer deeper behavioral insights.
- **Plan Details are Simplified:** Plans are labeled as “Yes/No” without details on pricing, usage limits, or customer preferences.
- **Assumption of Churn Definition:** Churn is binary (Yes/No), but in reality, customer disengagement can be gradual or reversible.

Data Preparation Strategy

Cleaning

- **Handling Missing Values:**
  - No missing values were found in the dataset. All rows were retained.
- **Dropping unnecessary columns**
  - phone number is likely a unique identifier (irrelevant for modeling).
- **Target Variable Encoding:**

  - The target column Churn was originally categorical (Yes, No).

  - We converted it to binary format:Yes → 1 (churned) and No → 0 (retained)

- **Categorical Feature Encoding:**
  - Two features were binary categorical (International Plan and Voice Mail Plan).
  - These were also encoded to Yes → 1, No →
  - No one-hot encoding was needed as there were no multi-class categorical columns
- **Train-Test Split:**
  - Data was split using train_test_split() from sklearn.model_selection: 80% training / 20% testing
  - Stratified sampling was used to maintain the same churn rate distribution across both sets.
- **Feature Scaling:**
  - Logistic Regression is sensitive to the scale of input features.
  - I applied StandardScaler from sklearn.preprocessing to all numerical features: Call minutes, Charges, Number of calls and Account length

Modelling Strategy

1. Baseline Model: Logistic Regression

- Logistic regression is easily interpretable by showing claer coefficients showing how features impact churn
- I Scaled features using StandardScaler, trained with LogisticRegression()and Tuned C parameter for regularization strength.

2. Comparison Model: Decision Tree Classifier

- Decision Tree Classifier will help Capture non-linear patterns and interactions and is also easier to visualize decision rule
- I Used DecisionTreeClassifier() and tuned max_depth and min_samples_split using grid search.

3. Comparison Strategy

- The primary Metric used is Recall in order to identify as many chners as possible
- Accuracy, Precision, F1-Score, and Confusion Matrix are also monitored

4.ROC Curve

- Curve shows how well the model distinguishes between churners and non-churners across different classification thresholds.
- The closer the curve is to the top-left corner, the better the model is at distinguishing churners.

Feature importance and Evaluation

Logistic Regression: Accuracy: ~85% Recall: ~15% F1-Score: Balanced performance **Decision Tree:** Accuracy: ~92% Recall: ~73% Decision Tree is Clearly Superior for Churn Detection

- 73% recall vs logistic's 15% - Identifies more actual churners
- Maintains reasonable 73% precision (only 27% false alarms)

ROC:

- The closer the curve is to the top-left corner, the better the model is at distinguishing churners.This means there's an 83% chance the model ranks a random churner higher than a random non-churner.
- A diagonal line means random guessing (AUC = 0.5).
- Our Logistic Regression AUC = 0.83, indicating strong model performance.

Top Predictive Features:

- International Plan: Customers with this plan are more likely to churn.
- Customer Service Calls: High call volume correlates with dissatisfaction.
- High daytime call usage

Conclusion and Recommendations

Following my analysis:

- The tuned deision tree model works best as it catches 67% of churners(vs 15% with logistic regression)
- The tuned decision tree model is also 88% accurate when predicting churn and makes only 9 false alarms per 566 customers
- The key warning signs for customers likely to churn are:
  - High daytime call usage

- Many customer service calls
- Cutomers that have an international plan

I would recommend the following:

1. Use the tuned decision model to find at risk customers
2. Focus on customers that make lots of daytime calls and offer daytime call discounts
3. Give special support to frequent customer service callers
4. Flag customers with international plans and frequent support calls; assign these accounts to specialized retention teams and create better international plan deals
5. Analyze top customer complaints and address root causes.

Lastly, automate alerts for high-risk profiles.

Next Steps

The expected results would be to find more churners thn before and reduce the chren rate. The next step would be to start with a small test group to prove these results before rolling out to the whole company.

Thank You