Telecom Churn Analysis

Justin Lee

Overview

Analyst: Justin Lee

Stakeholder: SyriaTel

Background: A leading mobile telecommunications company in Syria since 2000.

Motive: The company aims to minimize revenue losses caused by customers who discontinue their services prematurely. This analysis seeks to identify and understand patterns that may predict customer churn.

Data Understanding

Dataset features:

- 20 predictor variables
- 3,333 records
- Churn
  - 483 churn
  - 2850 non-churn
  - 14% ratio

```
account length               int64
area code                    int64
phone number                object
international plan           object
voice mail plan             object
number vmail messages        int64
total day minutes          float64
total day calls              int64
total day charge           float64
total eve minutes          float64
total eve calls              int64
total eve charge           float64
total night minutes        float64
total night calls            int64
total night charge         float64
total intl minutes         float64
total intl calls             int64
total intl charge          float64
customer service calls       int64
churn                         bool
```
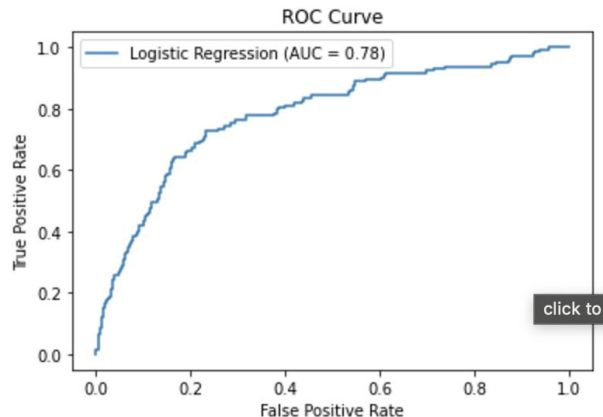
Baseline Model

- Accuracy: 86%
- Non-churn
  - Precision: 87%
  - Recall: 99%
  - F1: 0.92
- Churn
  - Precision: 58%
  - Recall: 8%
  - F1: 0.14
- ROC-AUC: 0.78

```
Classification Report:
              precision    recall  f1-score   support

           0       0.87      0.99      0.92       857
           1       0.58      0.08      0.14       143

    accuracy                           0.86      1000
   macro avg       0.72      0.53      0.53      1000
weighted avg       0.82      0.86      0.81      1000

Confusion Matrix:
 [[849    8]
 [132   11]]
ROC-AUC Score: 0.7826863917879088
```

ROC Curve

Logistic Regression (AUC = 0.78)

True Positive Rate

False Positive Rate

click to

Addressing Churn Class Imbalance

- Accuracy: 70%
- Non-churn
  - Precision: 94%
  - Recall: 69%
  - F1: 0.8
- Churn
  - Precision: 29%
  - Recall: 76%
  - F1: 0.42
- ROC-AUC: 0.778

```
Classification Report (SMOTE):
              precision    recall  f1-score   support

           0       0.94      0.69      0.80       857
           1       0.29      0.76      0.42       143

    accuracy                           0.70      1000
   macro avg       0.62      0.72      0.61      1000
weighted avg       0.85      0.70      0.75      1000

Confusion Matrix (SMOTE):
 [[595 262]
 [ 35 108]]
ROC-AUC Score (SMOTE): 0.7784351004887762
```
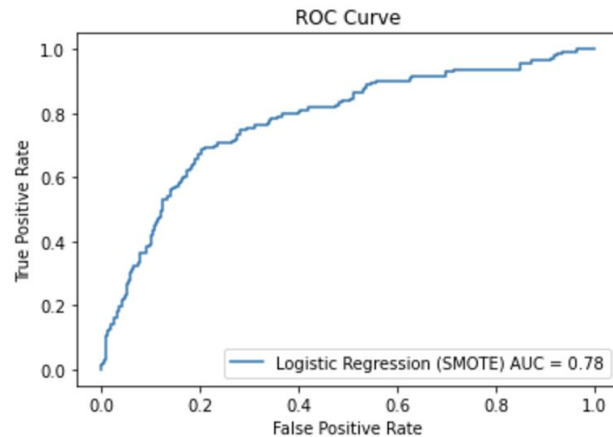
Evaluation

Baseline	Iterated Model
- High accuracy (86%) but low recall (8%) 29% higher precision score for chun than iterated model - ROC-AUC remained consistent at 0.78 - Class imbalance present	- Improved recall for churn (from 8% to 76%) - ROC-AUC remained consistent at 0.78 - SMOTE addressed class imbalance

Recommendations & Next Steps

1

Test on more advanced models like Random Forest or Gradient Boosting.

2

Adjust decision threshold to achieve a better precision-recall balance for churn cases.

3

Improve data collection to better capture factors driving churn.

4

Deploy model in a live environment to predict churn and measure effectiveness of intervention.

Contact Information

Justin Lee

Email: jlee8150@gmail.com

LinkedIn:

https://www.linkedin.com/in/justinlee2023

Thank you