SyriaTel
Predicting Customer Churn

George Nyang’aya

Project Overview

- SyriaTel is facing customer retention challenges in a highly competitive telecom market.
- Use machine learning techniques to build a model that uses customer data to predict whether the customer churned or not.
- Create a model that in future can be used to highlight customers at risk of churning so it can’t be actively prevented.

Problem Statement

- When customers stop using SyriaTel's services, it has a major impact on revenue loss.
- Retaining customers is more cost-effective than acquiring new ones.
- Currently SyriaTel have no definitive way to predict the customers most likely to churn.
- This project aims to build a data-driven solution, to identify customers likely to churn, so counteractive measures can be put in place.

Business Understanding

- The telecom industry is highly competitive with similar pricing across providers.
- Cost of marketing to non-customers as well as setup fees makes obtaining new customers more expensive than customer retention.
- SyriaTel's success depends on maximising market share and loyalty.

Data Understanding

- The dataset included the customers:
  - Phone number (unique identifier but not relevant to modelling)
  - Usage patterns
  - Packages subscribed to
  - Use of services
  - Churn status
- Dataset had class imbalance regarding churn, significantly more non-churners than churned customers.

Modelling

- We tested multiple models to find the most accurate and appropriate one.
- The focus was on **classification**: predicting a "yes" or "no" to whether a customer will churn.
  - Basic Logistic Regression - a simple, linear model to establish a baseline.
  - Balanced Logistic Regression - to address the imbalance of churners vs non-churners.
  - Decision Tree - a more flexible model that can identify non-linear/complex relationships with churn.
  - Tuned Decision Trees - optimised aspects of the model for better performance.

Evaluation

- Since the number of churners was very small compared to non-churners, measuring accuracy wouldn’t be as helpful.
- Key metrics:
  - **Recall**: out of all customers that churned, how many did the model catch – important as we don’t want to miss anybody that churned
  - **Precision**: out of all the customers the model predicted churned, how many were accurately predicted, to avoid false alarms
  - **AUC Score**: how well the model separates churners from non-churners.
- We also used accuracy when comparing training and testing model performance to ensure the models would be general enough to apply against different datasets.

Final Model – Tuned Decision Tree

- Churn **recall** reached **0.83**, meaning we successfully identified 83 out of every 100 churners.
- High **precision** of **0.79** meant we didn't over-predict churn (few false alarms).
- The **AUC score** was also the highest on the model, proving the model reliably separates churners and non-churners.
- Most importantly, the model performs just as well on testing data as it does on training data. This means the model should perform well on the fresh data obtained for the real world.

Recommendations

- Continue trying other types of regression models, e.g. CART regression, to see if the model can further improved.
- Try adjusting using new features to tune the model.
- Use other techniques to address the imbalance of churners and non-churners and see the impact on the model.

Next Steps

- Use the model to highlight customers most likely to churn, using real time data.
- Focus retention plans on these high-risk customers.
- Monitor model performance over time and retrain as customer behaviour evolves.
- Use the model to see the effectiveness of different retention schemes/campaigns.