Challenge Theme: Industry

Predicting Customer Churn for Telecommunication

Problem statement

Customer Churn is when a subscriber of a product or service cancels their subscription or decides to switch to a competitor. This is not ideal for any company as it impacts their revenue, growth, and reputation.

The goal of this challenge is to develop a model that can identify customers who are likely to churn, and when. This can help the company take preventative measures before losing customers. The data is a mix of features as well as time series data, so take that into account when developing your model. Also, note that the dataset is much more likely to contain customers that will not churn rather than customers that will. How can you deal with this class imbalance?

Data description

Data for this challenge would be very similar to the one IBM has published on telco customer churn data here: https://community.ibm.com/community/user/businessanalytics/blogs/steven-macko/2019/07/11/telco-customer-churn-1113

Like in the IBM dataset, the data can be roughly broken into four different schemes.

Table	Attributes
Demographics	- CustomerID - Gender - Age - Married # Dependants
Location	- CustomerID - Country - State - City - Postal Code - Latitude - Longitude - Population (based on zip code)
Services	- CustomerID - Quarter (Q1,…,Q4) - Referred a friend? - Tenure in months (how long has customer been with company) - Did customer accept marketing offer? (None, Offer A, …) - Phone Service (Yes/No) - Long Distance phone cost - Internet Service - Average Monthly GB download # times contacted tech support - Unlimited Data (Yes/No)

	• Contract type (Month-to-Month, One year, etc.) • Monthly Cost for all services
Status	• CustomerID • Quarter • Customer Satisfaction score (1 to 5) • Churn Label (Yes=Left this quarter, No=Otherwise) • Churn Probability • Churn Reason (None, Competitor, Unsatisfied, Price, Other)

Suggested outputs

Some potential routes for expressing findings that you may wish to consider:

- How to best deal with a mix of time series and feature data
- Feature engineering
- Visualisations of when/where customers frequently churn
- Explainability: What factors are most likely to influence customer churn?