DSFPT04A Phase 3 Telecommunication Project

—

Laban Kariuki

Content

A breakdown

- Overview
- Business Understanding
- Data Understanding
- Modeling
- Evaluation
- Recommendation
- Next steps

Overview

A model to help predict the likelihood of a subscriber churning from our services based available data.

Business Understanding

Objectives

1. The telecommunication company is looking to identify the likelihood of customers abandoning its services

2. It is more ideal to focus on customers who will be with the company in the foreseeable future than those who will end up leaving

3. By identifying staying customers, resources can thereafter be spent on retention

Data Understanding

Syria Telecoms Company Customer Data

Source and nature of the data

The data has been sourced from Syria Telecommunications Company. It is composed of details relating to

- Phone calls duration
  - Day and night
- Phone calls charges
  - Day and night

The available types in the data

A review of the data reveals the existence of different types of data:

- Characters
- Numerics
- Booleans
- Floats and
- Objects

There was no need to drop any features or records. The data was clean.

Modelling

Options taken

Models used for prediction

- Initial model used the lot transformation
- Later improved by doing a one hot encoding
- Concluded with the bias variance trade off.

Results of the models

- Poor predictor of the likelihood of a customer churning
- Also a poor predictor of the likelihood of our model
- Promising results that could use better or more improved accuracy

Evaluation

Outcome

Results of the models

- Initial model used the lot transformation
- Later improved by doing a one hot encoding
- Finally fit the model with K-Fold cross validation

Improvements made on the models

- Poor predictor of the likelihood of a customer churning
- Also a poor predictor of the likelihood of our model
- Promising results that could use better or more improved accuracy

Recommendations

Key Pickings

The takeout

Our model demonstrated a great deal of underfitting for the most part.

Unfortunately, our K-Fold cross validation did not help reduce the bias.

This is indicating of the majority of the features not being adequate to give a prediction as intended.

Recommendations

Additional records of the same data be used for the analysis

Different data source be utilized.

Reduction in the number of features excluding categorical data

Next Steps

Going forward

Proposal

Data sourcing from similar industry sources with a similar nature of data

Consider utilizing different features in the model.

Consider using fewer features and switching them across models for optimal results

Expectation

- Multiple data sources are available
- Nature of features may vary but a few will be similar
- It might be important to consider extrapolating this model to other types of organizations as well.
- Other models need also be utilized for purposes of establishing the one that best fits our expectation

Conclusion

It is important for us to make use of all the models at our disposal. Although this particular project only utilized two, should time allow, more models can be considered in a bid to establish the one that best fits our data.

Thank You

TELECOM