

Predictive Analytics and Modelling of Data: Telco Churn Prediction

Description

In the telco industry churn is a common problem and major reason for loss of revenue and growth. Especially in the telco sector, most customers have multiple options from which to choose within a geographic area. Since it is in general 6 times more expensive to attract new customer than to retain, thus preventing customer from churning is a hot topic in the telco industry. Moreover, churn rates even help a company determine how it is measuring up to its competitors. Predicting churn will help save money and/or generate new revenue. Having information about the customer behaviours that precede churn, as well as churn risk ratings for individual customers, can help to directly target these customers. These customers can be contacted in advance and pampered with specific actions to retain them. Hence, a well-constructed model is of vital importance for the survival telco companies.

It is your task to build a predictive model for churn. The company wants you to provide an answer to (some of) the following questions: How can we predict who will churn, what are the reasons for churn, when will customers churn, how much churn will I see next month, and what campaigns help avoid churn?

Methodology

You can use the most promising angle in the exploratory data analysis. Try to come up with relevant graphs and summary statistics to pinpoint that there is a churn problem. Next, you should build several models –as seen in class– predicting churn. You can choose to create new variables, to apply data transformation and/or data reduction. Remember that in the end the goal is to build a churn model that has a high predictive power on a test set. You should evaluate your models with the relevant performance metrics, such as accuracy, lift, AUC, and others. It is also important that you interpret your results and that you perform sensitivity analyses on your results. Remember that often your first predictive model will be far from the best predictive model.

Data

The data set is a real-life data collected from a major wireless telecom operator in South Asia. Features of the dataset include complaint count, network age, user type amongst others. Dataset containing continuous and categorical variables describing various usage behaviour. The data set also has a number of missing values; it is up to the students to effectively handle these. The data set has a balanced number of churn/active classes (labels) (i.e., a 50/50% distribution).

Deliverables

The company is looking for 3 main aspects in your work:

1. EDA (exploratory data analysis)
2. Model that detects and predicts churn accurately (including code & accuracy)
3. Reasons for churn (interpretation of your model and feature importance)

For an extra challenge the company has created test data without labels. They will share a samples submission file to show you what the output of your model prediction has to look like. Please share submission files as *student_id.csv*.

Besides the analyses part, it also important to include an academic angle to your work. Therefore, you should look up relevant literature of churn in the telco industry and try to use some of these academic results in your own work.

The expected deliverables according to the course handbook are a *written report* of maximum 10 pages (without appendices), font size 12 (Arial), and spacing 1.5. In the report you will go over all the steps of the CRISP-DM model and describe how you have done this in your group project. More specifically, the report must contain the following:

- *A management summary*. Here you introduce the problem, summarize your methodology, your results and the managerial insights and recommendations in short. Ideally, this is a one-pager.
- *Problem definition*. Here you dig deeper into the problem of churn in the telco industry. In this section you try to link your case with what has already been done in churn in the telco industry. Specifically try to answer questions like what has been done in churn in the telco industry, what methods did they use, what were there insights and how can we use them in our study? What is the impact of these studies on my work?
- *Data understanding and preparation*. This part contains describing exploratory data analyses and data transformation. Describe the most important insights related to customer churn. Next, dig deeper into the feature engineering, how you handle missing values and solved other data quality issues.
- *Modelling and evaluation*. In this section you discuss the predictive models and evaluation methods that you used, and try to describe them shortly. Next, you discuss the results in terms of model performance and interpretation. This part should answer questions such as: how good can we predict churn, which algorithms are important and what explains churn.
- *Managerial recommendations*. In this section you describe the impact of your results from a managerial perspective. Given your churn models and results, what can the company do to manage churn? This part also contains sensitivity analyses.
- *Appendices*. Here you include figures, tables, explanations and the reference list that could not be added in the 10-page report.

Besides a written report, the students should also make a *presentation*. This should be a business presentation where you describe the business problem, go briefly over your methodology, discuss the results and the impact on firm performance. It is important to come

up with clear recommendations for the company. Remember that this should be a business presentation so don't make it too technical.

The slides for the presentation have to be submitted together with the written report by **4pm 23th of November 2018**.

Grading

The written report accounts for 60% (including 10% peer assessment) of the total mark. Students are also required to grade their peers. This can have an impact on the score on the report and has an impact of 60% of that component. The best way to approach peer review is to consistently record the interactions with your team (i.e., meetings, Skype calls, etc., and indicate what personal input you deliver).

The presentation accounts for 10% of the total mark. These marks will be individual.

References

Some references about churn in the telco industry to get you started.

Óskarsdóttir, M., Bravo, C., Verbeke, W., Sarraute, C., Baesens, B., & Vanthienen, J. (2017). Social network analytics for churn prediction in telco: Model building, evaluation and network architecture. *Expert Systems with Applications*, 85, 204-220.

Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, 218(1), 211-229.

Verbeke, W., Martens, D., & Baesens, B. (2014). Social network analysis for customer churn prediction. *Applied Soft Computing*, 14, 431-446.