TALLINN UNIVERSITY OF TECHNOLOGY
Faculty of Information Technology

Triinu Lepp 166763IVSM

NEW FLEEP USER CHURN PREDICTION USING MACHINE LEARNING

Master's thesis

Supervisor: Ants Torim

PhD

Tallinn 2019

Summary

Customer churn is an important topic and should be monitored by each company who is interested in keeping its customers. Fleep users churn rate is highest within the first days after users signs up and Fleep wants to lower it. We defined that churn occurs when user has been inactive for 30 days. Since we were analysing only the new users, we took all signups within a year and marked whether user had churned within the first 30 days or not. Then, we used the user first day's activities data and trained different models.

The thesis followed CRISP-DM methodology. First, current situation of the business was analysed, business objectives and data mining goals were set and project plan was produced. Next, in the data understanding phase the features used for churn prediction were mapped. Furthermore, in the data preparation phase four new datasets were created: one with user first two days activities, one with user first seven days activities and two more that were subsets of the previous two with an exception of only using the variables that were the most related to the new user churn.

The feature selection for creating the subsets was done using statistical analysis. There we also got a more detailed insight which variables are the most associated with the churn. Knowing this helps company already during the user onboarding process. Next, in the modelling phase models were trained using logistic regression, naïve Bayes classifier, support vector machine, k-nearest neighbours, decision trees, random forest, gradient tree boosting and neural networks. When using the dataset with information about users first seven days activities and gradient tree boosting, we managed to train a model with accuracy 0.84 and F1 score 0.83. Therefore finally, in the evaluation phase was suggested to use this dataset and gradient tree boosting machine learning technique for new user churn prediction.