Prediction of Customer Lifetime Value using machine learning techniques in e-commerce

Nadezhda Zaborskaia (Matr.Nr.3030319)*

Management and Data Science, Leuphana University, Lüneburg

Abstract

Nowadays most successful business operate in the digital realm. E-commerce, food delivery or taxi booking apps are a small part of a big variety of services available online. The Internet services and mobile apps collect a big amount of data every day. The future success of a companies depends on how this data is used by it.

For my master thesis I collaborated with FREE NOW (formerly mytaxi) the taxi app. Since 2009 FREE NOW has been growing up from a German taxi start-up to a leading taxi-hailing app in Europe. FREE NOW has more than 10 million passengers and 100,000 drivers across Europe. FREE NOW provided the data for modelling and technical support to design experiments.

In this research I used different machine learning technics to measure one of the most important indexes in determining the present and future success of the business - Customer Lifetime Value (CLV). CLV index evaluates how valuable a customer will be for a company within a certain time period. The results of CLV can be used in Customer relationship management (CRM) e.g. passenger channel acquisition, target passengers for campaigns and passengers priorities.

The CLV prediction is divided into 2 stages. The first stage is clustering and the second is CLV prediction for segments. A cohort of passengers who registered in a certain time period are divided into k number of clusters. I used different features which have been collected after 7 days of passenger's activity in the app for segmentation. CLV index is calculated for each cluster for 30, 90 and 180 days ahead. I used different approaches to predict the CLV for new passengers. From the first stage each segment of passengers got a CLV index. Thus, when new passengers register in the app, they will be assigned to a cluster with predicted CLV after 7 days.

For the experiment evaluation we used relative error and accuracy score which represents a confident interval with a threshold.

For the baseline we consider all passengers in the cohort to belong to one cluster and as CLV prediction we take the average money spent by the passengers in the cohort in the different time windows.

Submission: March 9th, 2020

*Electronic address: nadezhda.zaborskaia@stud.leuphana.de