

Mat-2.4177 Seminar on Case Studies in Operations Research 2015

Estimation of consumer repurchase behavior

Midterm Report

Client: Microsoft Mobile Oy

17/04/2015

Authors:

Niko Lietzén (Project Manager)

Alessandro Mancuso

Antti Melén

Pekka Alli

1. Project status

The project goal of creating a forecasting model for a customer's next replacement purchase has not changed. At this stage, we have created a preliminary model. The current model is a Pareto/NBD (Negative Binomial Distribution) model. However, we are also still considering alternative models. The Pareto/NBD model is implemented in the R package Buy 'Til You Die (BTYD). We have some difficulties with the computationally slow estimation times of our current model. To get a suitable fit for the model, we require sufficiently many observed customer transactions. However, the computation time grows exponentially as we increase the number of observations. We are currently trying to determine a sufficient amount of observations such that the computation time is acceptable and ensure the quality of the parameters is sufficient for forecasting. We have determined feasible ways to measure the quality of our model.

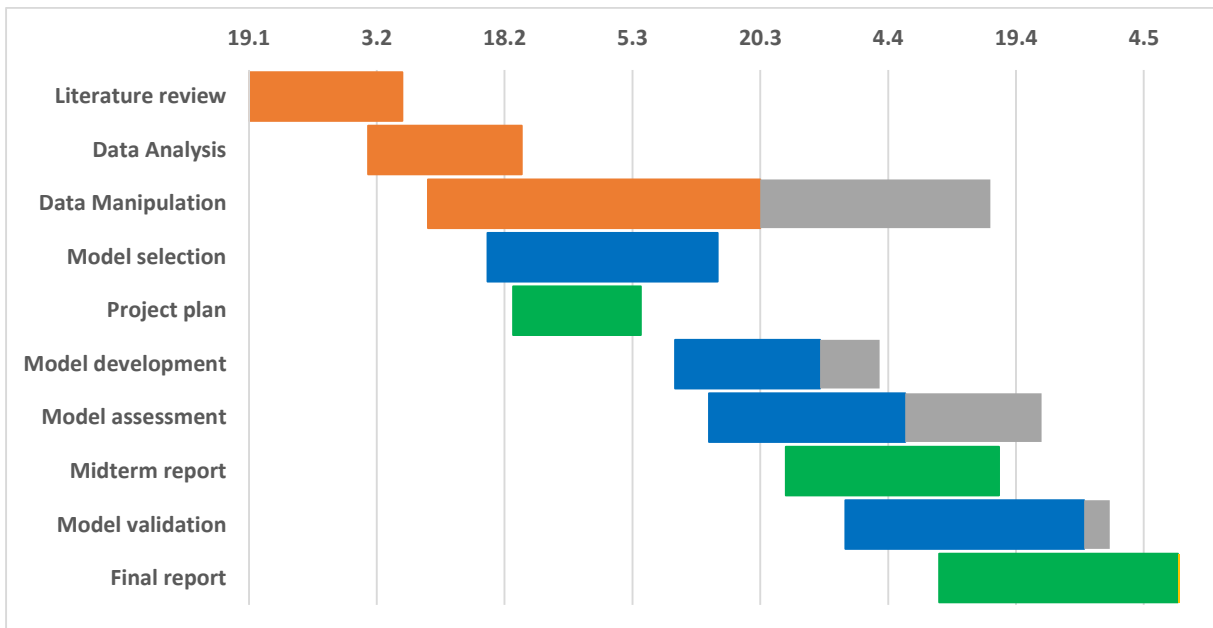
Furthermore, our model would benefit from a data set with more observations regarding one customer. Currently, we have a data set with only one purchase related to one customer. If we decide to only consider customers with more than one purchases, the model might not fit the real life situation. We will ask for the client's opinion and continue the modeling process according to their wishes.

2. Task Status and Updated Schedule

The tasks that we have completed are Literature review, Data Analysis, Model Selection and Project Plan. We completed all of them with no problems and they have not caused any delays to the project.

The task that is behind schedule is Data Manipulation. Some problems have arisen during the task and as of April 8th we still have some incorrect entries that prevent us from proceeding to following tasks such as Model Verification. The overall status of the project is behind the original schedule but the delay is not so critical that it should prevent us from completing the project on time.

Because the tasks have been delayed, we have already started working on other tasks that are not dependent on those. The Midterm report was finished early and we have already started working on the Final report. The updated schedule is presented below. The grey bars present extensions we have reserved for the tasks.



3. Updated Risks

From the four risks identified (data quality, wrong model selection, overly complex model, Alessandro’s departure) in the project plan, two of them realized. The first risk “Poor data quality” did not actually result from poor data but from our teams’ slow data manipulation process. However, the outcome was similar and it delayed our project. The other risk that realized was “Alessandro’s departure”. Alessandro left the country and will not be able participate in our meetings or the seminars.

4. Initial results

We have completed the data analysis of our data set. Time series analysis indicates that there are no visible seasonal patterns in our data. Thus, we have decided to leave out the seasonal component from our model. Furthermore, we have trouble with the computation time and adding extra components to the model would presumably make the estimation part even slower.

We have complete implementations ready for our current model. However, we are currently discussing ways to make our model more user friendly to the client. We still need to adjust our model. The figure below demonstrates the difficulties we are facing. The estimation below was calculated in a reasonable time. However, it is clear that the model does not fit the actual data very well.

The key results that our model gives are the expected number of transactions in a specified time interval, the probability that an individual with observed behavior is still an active customer at a specified time and the expected number of transactions in the future period for an individual in a specified time interval. We are currently working on the presentation of these results. We want to present the results to that the benefit for the client is maximized.

Tracking Weekly Transactions

