Product recommendation tactics

Public Project Summary

Attila Gulyás

Abstract

This capstone project for the CEU MSc in Business Analytics program aims to provide actionable product recommendation tactics to the Client. In the project, business-driven data analytics approach was applied.

Background

The Client is a small Hungarian company which operates two online toy stores. The company was growing dynamically for years; however, the revenue growth intensity is moderating after a sharp increase in 2017. The Client has a lot of data, stored in a SQL database, however, the data is limitedly used for analytical purposes. After discussing the possible development directions with the Client, it turned out that an advanced product recommendation system, by using up a significant amount of data in the database, can be a breaking point for the company. However, the Client faced the problem that there was no recommendation concept at all to boost revenue.

Analysis tasks

In the capstone project, a business-driven data analytics approach was applied to address the identified problems. In the project, I worked along the CRISP-DM project methodology, which is an international standard for data analytics projects.

At the beginning of the project, goals and requirements were clarified and business questions were drawn up. From a product recommendation point of view, four questions were found relevant to be answered with data analytics: for whom, on which channel, when, and what to recommend.

Based on the business questions, the necessary data was defined, and data collection took place. For the analysis, transactional data extracted from different tables of a SQL database were used. SQL data was queried with R, then all the tables were downloaded in csv. Then, drawing the data model up was essential in data understanding.

After understanding the data and its structure, it was prepared for the analysis. To do that in the right way, data was explored by implementing descriptive analysis. Data preparation

included several tasks, for example, handling data quality issues, transforming variables, organizing analysis data in tidy format, etc. It was an iterative process during the project.

Next, different analytical technics and machine learning models were applied to find the answers for the business questions. Answering the question "for whom", I did an RFM customer segmentation, dividing customers into 8 actionable segments. I pointed out that this approach is useful, considering that different customer segments require different marketing actions with different focus.

Answering the question "on which channel", I pointed out that product recommendation can take place in different ways on different marketing channels. The project scope was the email marketing channel suggested to be used for personalized e-mail recommendations based on previous transactions. Customers in the contact list (who allowed promotional emails) are the scope of email recommendation. It was also highlighted that there is huge potential in social media retargeting, and it should be exploited in the near future.

Answering the question "when", I elaborated a clustering approach, which is suitable for identifying patterns in the ordering dates of each customers. Each cluster is a potential buying period when the customer may intend to purchase based on previous orders. One of the advantages of the algorithm is that it identifies dense areas of data points as clusters. Also, it is able to identify separated points as noise and not to assign them to any clusters.

One of the key results was the development of product recommendation models for answering the question "what". I used five classic models (random, popularity based, association rule, item based collaborative filtering, user based collaborative filtering), and combining them, I built my own hybrid product recommendation model. All the models used transactional data; recommendations take place based of previous purchases. Each model recommends different item lists by users.

The performance of the different models was evaluated in order to make them comparable and to select the best one. Based on the tests, the hybrid model is the best. I found that it performs 2-4 times better than an intuitive, manual, rule of thumb-based approach (e.g. popularity method). I found that the hybrid model works very well, and it gives relevant recommendations even when there is no hit on the test set.

Considering that this project was a conceptual one, a framework of recommendation tactics was suggested in the end as deployment depicting how to combine the technics along the four business questions.

Key outcomes and benefits to the client

In the project an email product recommendation framework was elaborated, supporting what, when and for whom to recommend. Along the four business questions that the project aimed to answer, the followings are the main outputs and deliverables:

For whom?

- Actionable segment descriptions;
- Action tips by segments;
- Customer new customer id assignment table (csv);
- Customer segment assignment table (csv).

When?

 Order time clusters by customers identifying periods of potential interests to purchase (csv).

On which channel?

• Conceptual framework on how to use marketing channels to recommend.

What?

- 2-4 times better recommendation, than with an intuitive, manual approach;
- Relevant recommendation even when there is no hit;
- Recommendation list by customers (csv).

The whole analysis consists of 18,000 line R program code (~740 A4 pages of clean code). All the codes are ready to be run on a regular (e.g. monthly) basis. At the end of the project, follow-up tasks were suggested to the client to be done in association with product recommendation (A/B testing, going into production, extension to other channels).

Learning experience

Personally, I feel that this project was challenging, and I could gain a lot of hands-on experience in the field of product recommendation. One of the most important experiences is that it is worth taking time for data understanding and data preparation because the results of these activities have a huge impact on model performance in the later stage of the work. Another useful experience is that it is worth trying different modelling approaches or combining them with each other. For example, in this project, a hybrid recommendation model was found to be the best one. Also, it is worth trying out a simple approach too, because it can be used as good benchmark to compare performances of models.