Analysing the European diabetes drug market

Capstone Project Summary

Balint Thaler

MS in Business Analytics

Central European University

Department of Economics and Business

Table of Contents

Introduction

The current document is a public summary of a Capstone project that was conducted in cooperation with a large European multinational pharmaceutical company. The scope of the analysis was the sales of diabetes-treating drugs in 18 selected European countries and aimed firstly at understanding the underlying patterns of the diabetes market with a special focus on one of the sub-categories of diabetes products, called DPP-IV Inhibitors. Secondly, it aimed to understand the behaviour of one of the main – highly successful – competitors of the Client. Thirdly, it aimed to create predictive models that may be used for forecasting or for simulations by the Client to guide marketing decisions. No similar analysis has been conducted at the Client before.

Data sources

Various data sources were utilised to conduct the analysis which included industry-standard IQVIA data on brands, manufacturers and volumes sold, retail price data from the Client's internal databases, public data sources (World Bank, national statistics bureaus), diabetes-related organisations' websites (International Diabetes Federation) and publicly available resources on companies (annual reports, websites, LinkedIn company profiles).

Findings of the analysis

After collecting and joining all the data, a thorough exploratory data analysis was performed. This showed the differences in diabetes therapies across the countries with higher income countries relying more on expensive but more modern therapies, and lower income countries preferring traditional and much cheaper ones. As one of the focus segments – the DPP-IV Inhibitors – have recently had their patents expiry and their markets "genericised", a

closer look was taken at them. This revealed that upon genericisation, the number of companies and brands have greatly increased in many countries, and a shift towards cheaper drugs has started. The genericisation's impact on total volume was inconsistent across the geographies. The previously designated highly successful Competitor has increased their sales massively in this segment by focusing on some of the key countries. The data on retail prices has also confirmed that they have successfully undercut most competitors in these markets which likely also contributed to their success.

Predictive modelling

Various different machine learning models were utilised for the predictions, the best one was chosen based on its performance as measured by the root mean squared error (RMSE). The data was split into training and testing data, the former was used to train the models, the latter was used to evaluate their performance on previously unseen data.

First, a simple linear regression model was used as a benchmark. Then a random forest, a gradient boosting machine, an Xtreme Gradient Booster (XGB) and an AdaBoost model was fitted. Lastly, a more refined linear model was also fitted. The modelling methodology was identical for the general dataset and the DPP-IV dataset as well, however for the latter the AdaBoost and refined linear models were skipped.

In both cases, the XGB model performed with the lowest prediction errors. In the general model, the accuracy was somewhat lower for lower volumes with consistent overpredictions, but then stabilised for medium and high volumes. The best DPP-IV model's performance was more consistent across the board.

Conclusions and recommendations to the Client

Overall, the analysis has concluded that although there were some similarities in how the different countries reacted to the genericisation, more research is required to understand why they were not universally true. Similarly, more data is required to gain insight into the detailed practices (such as marketing campaigns, sales channel used etc.) of the successful competitor to understand the reasons behind the success.

The inclusion of the retail price in the DPP-IV models has contributed to a better accuracy in predictions, however data availability has greatly limited the external validity of the findings (both in time and geographic scope). The Client is advised to conduct further analysis into both the DPP-IV category and the general diabetes therapeutic area with more features around the process of genericisation.

The Client is also advised to utilise the good predictive performance of the general model to forecast sales volumes and to simulate different market entry and exit decisions to support their decision-making process, ultimately leading to higher return on investment of their budget. Caution is advised however in applying the model to countries and therapeutic areas beyond the ones in the data, as no information is available on how similar they may or may not be, and how valid the model's performance might be.

Learning outcomes

The project has given me a chance to engage with a unique industry and support the Client in making decisions that may have a large impact on the organisation's success using (and hence deepening my knowledge of) techniques such as python programming, data visualization and machine learning in a real-life scenario.