

## ABSTRACT

---

In recent years machine learning has become increasingly important in the pharmaceutical industry. Also, in the pharmaceutical-commercial sales force, machine learning helps to objectively analyze strategic and tactical questions, such as demand assessment and market potentials. This thesis will investigate for Merck KGaA, Darmstadt, to what extent statistical methods and machine learning can help to make the two oncological drugs Avelumab (trade name: Bavencio®) and Cetuximab (trade name: Erbitux®) available to an even larger number of eligible patients.

With its launch in October 2017, Bavencio® is still quite new on the market, which is why a market potential analysis for customer identification and segmentation is relevant. Erbitux®, on the other hand, is an established product on the market, which is why a forecast to identify order volumes and frequencies of sales figures is required.

Internal sales figures and data on contacts with healthcare professionals by sales representatives as well as the hospital quality reports of the Federal Joint Committee serve as the data basis.

For the market potential analysis of Bavencio®, correlations between measured variables, such as sales figures and patient volume (number and type of procedures performed) per hospital or communication data and sales figures are systematically examined aiming at identifying (potential) hospitals with high numbers of patients. Consequently, it was possible to test approaches for classifying hospitals, by means of which discriminating characteristics could be found between categories and thus potentially driving factors could be identified.

For the sales forecasting of Erbitux®, a time series analysis is performed to identify ordering behavior based on sales figures. This analysis showed that accurate forecasting using ARIMA models is currently not possible. The next step would be to try other approaches, such as Generalized Linear Models. In summary, initial challenges such as sales forecasting, customer identification, and segmentation could be systematically addressed with machine learning. To further substantiate the results, the improvement of data quality and availability is of utmost importance, as well as the testing of further possible algorithms and the verification of the results in practice.

*Keywords*— Analysis of market potential, Quality reports of hospitals, Pharmaceutical-commercial sector, Exploratory data analysis, Regression analysis, Sales forecast

## ZUSAMMENFASSUNG

---

In den vergangenen Jahren hat maschinelles Lernen in der pharmazeutischen Industrie stark an Bedeutung gewonnen. Auch im pharmazeutisch-kommerziellen Außendienst hilft maschinelles Lernen dabei, strategische und taktische Fragen, wie zum Beispiel Bedarfsabschätzung und Marktpotenziale objektiv zu analysieren. In dieser Thesis soll für Merck KGaA, Darmstadt, untersucht werden, inwiefern statistische Methoden und maschinelles Lernen dabei helfen können, die beiden onkologischen Präparate Avelumab (Handelsname: Bavencio®) und Cetuximab (Handelsname: Erbitux®) für eine noch größere Zahl an betroffenen Patient\*innen verfügbar zu machen.

Mit seiner Markteinführung im Oktober 2017 ist Bavencio® noch recht neu auf dem Markt, weshalb eine Marktpotenzialanalyse zur Kundenidentifizierung und -segmentierung von Relevanz ist. Erbitux® hingegen ist ein etabliertes Produkt am Markt, weswegen eine Prognose zur Erkennung von Bestellmengen und -frequenzen von Absatzzahlen benötigt wird.

Als Datengrundlage dienen interne Absatzzahlen und Daten über Kontaktaufnahmen mit medizinischem Fachpersonal durch Außendienstmitarbeiter\*innen, sowie die Krankenhaus-Qualitätsberichte des Gemeinsamen Bundesausschusses.

Für die Marktpotenzialanalyse von Bavencio® werden Korrelationen zwischen Messgrößen, wie z.B. Absatzzahlen und Patientenaufkommen (Zahl und Art durchgeführter Prozeduren) pro Krankenhaus oder auch Kontaktaufnahmen und Absatzzahlen systematisch untersucht mit dem Ziel (potenzielle) Krankenhäuser mit hohen Betroffenenzahlen zu identifizieren. Damit konnten Ansätze zur Einordnung der Krankenhäuser geprüft werden, mittels derer zwischen den Kategorien diskriminierende Merkmale gefunden und somit potenziell treibende Faktoren identifiziert werden konnten.

Für die Absatzprognose von Erbitux® wird eine Zeitreihenanalyse zur Identifikation von Bestellmustern anhand der Absatzzahlen durchgeführt. Diese Analyse ergab, dass eine präzise Vorhersage mithilfe von ARIMA-Modellen aktuell nicht möglich ist. Der nächste Schritt wäre nun weitere Ansätze, wie Generalized Linear Models auszuprobieren.

Zusammenfassend konnten erste Herausforderungen wie Absatzprognose, Kundenidentifizierung und -segmentierung mit maschinellem Lernen systematisch in Angriff genommen werden. Um die Ergebnisse weiter zu konkretisieren ist die Optimierung der Datenqualität und -verfügbarkeit von äußerster Wichtigkeit, ebenso wie die Prüfung weiterer infrage kommender Algorithmen und der Verifizierung der Ergebnisse in der Praxis.

*Keywords*— Marktpotenzialanalyse, Qualitätsberichte der Krankenhäuser, Pharmazeutisch-kommerzieller Sektor, Explorative Datenanalyse, Regressionsanalyse, Absatzprognose