

ABSTRACT

Today's competitive economy is characterised by companies increasingly relying on a data-driven business model, where data is the basis for decision-making. Customer data is particularly valuable, especially in the insurance industry where the customer is the core of the business. Therefore, it is crucial for every insurance company to know its customers well in order to retain profitable customers in the long term and to plan the limited budget efficiently. In this context, knowing the future value of a customer is crucial for successful customer management.

In this thesis, the efficiency of machine learning methods in predicting customer value in an insurance company was investigated. After using several visualisation and ad hoc analysis methods to identify the variables that could have a significant predictive power, two main types of models were constructed. First, classical regression models based on the conditional mean were used. Secondly, a quantile modelling approach based mainly on the median was also implemented. The regression by mean and the regression by median were compared here because of the asymmetry of the density function of the customer value.

The results showed that in general the simplest model, which assumes a constant customer value, was sufficient at first sight to obtain satisfactory predictions for the entire data set. However, this was only due to the fact that the customer value remained constant for about half of the customers. The detailed analysis of different customer groups proved the effectiveness of the machine learning models when the value of a customer actually changes from one year to the next, especially for customers with a high value. Overall, the quantile regression approach proved to be the best choice among all machine learning methods.

Keywords: Customer Value, Linear Regression, Quantile Regression, Asymmetric Distribution

ZUSAMMENFASSUNG

Die heutige wettbewerbsorientierte Wirtschaft zeichnet sich dadurch aus, dass sich Unternehmen zunehmend auf ein datengesteuertes Geschäftsmodell stützen. Eine solche Ausrichtung ist dadurch gekennzeichnet, dass Daten die Grundlage für die Entscheidungsfindung bilden. Kundendaten sind besonders wertvoll, vor allem in der Versicherungsbranche, in der der Kunde der Kern des Geschäfts ist. Daher ist es für jedes Versicherungsunternehmen entscheidend, seine Kunden gut zu kennen, um profitable Kunden langfristig zu binden und das begrenzte Budget effizient zu planen. In diesem Zusammenhang ist das Wissen um den zukünftigen Wert eines Kunden entscheidend für ein erfolgreiches Kundenmanagement.

In dieser Arbeit wurde die Effizienz von Machine Learning-Methoden bei der Vorhersage des Kundenwerts in einer Versicherungsgesellschaft untersucht. Zunächst wurden mithilfe verschiedener Visualisierungs- und Ad-hoc-Analyseverfahren die Variablen ermittelt, die eine signifikante Vorhersagekraft haben könnten. Anschließend wurden zwei Hauptformen von Modellen konstruiert. Zum einen wurden klassische Regressionsmodelle auf der Basis des bedingten Mittelwerts verwendet. Zum anderen wurde auch ein Quantil-Modellierungsansatz implementiert, der hauptsächlich auf dem Median basierte. Die Regression nach dem Mittelwert und die Regression nach dem Median wurden hier wegen der Asymmetrie der Dichtefunktion des Kundenwerts verglichen.

Die Ergebnisse haben gezeigt, dass im Allgemeinen das einfachste Modell, das von einem konstanten Kundenwert ausgeht, auf den ersten Blick ausreicht, um zufriedenstellende Vorhersagen für den gesamten Datensatz zu erhalten. Dies war jedoch nur darauf zurückzuführen, dass der Kundenwert für etwa die Hälfte der Kunden konstant blieb. Die detaillierte Analyse verschiedener Kundengruppen hat die Wirksamkeit der Machine Learning-Modelle bewiesen, wenn sich der Wert eines Kunden tatsächlich von einem Jahr zum anderen ändert, insbesondere bei Kunden mit einem hohen Wert. Insgesamt erwies sich der Ansatz der Quantilsregression als die beste Wahl unter allen Verfahren.

Schlagwörter: Kundenwert, Lineare Regression, Quantilsregression, Asymmetrische Verteilung