„Aller guten Dinge sind drei."

Am 13. Februar 2019 fand das dritte Treffen des Roundtable Rhein-Neckar in den Räumen des paritätischen Wohlfahrtsverbands im Zentrum von Heidelberg statt. Die Veranstaltung wurde von Tanja Kenda im Vorfeld sehr gut organisiert und konnte mit 24 Teilnehmern einen neuen Teilnehmerrekord verbuchen.

Roland Mannshardt begrüßte die Teilnehmer und stellte kurz die verschiedenen Angebote und Aktivitäten des Vereins vor.

Kay Sakkiettibutra, Vice President Data Science bei Verivox, und Dr. Arno Becker, Data Scientist bei der HMS Analytical Software GmbH, stellten in ihrem Vortrag „Mit Predictive Modelling zur gezielten Produktempfehlung bei Verivox" vor, wie Produktabschlüsse ausgehend von einem regelbasierten Algorithmus basierend auf Web-Analytics Daten durch den Einsatz von selbstlernenden Algorithmen auf Grundlage einer breiten Datenbasis gesteigert werden konnten. Verivox hat hierzu in 2 Projekten mit HMS Analytical Software GmbH zusammengearbeitet. In dem ersten Projekt wurden neben der Evaluierung eines geeigneten analytischen Ansatzes die Datenvorverarbeitung auf breitere Füße gestellt und ein „next best product" Modell entwickelt, mit dem die Produktaffinität eines Kunden vorhersagt werden kann. In dem zweiten Projekt wurde der Ansatz generalisiert, weitere Daten hinzugefügt und die Vorhersage konnte auf Produkttarife verfeinert werden.

Herr Becker lieferte im weiteren Vortrag die Beschreibung der Aufgabenstellung, Produkt-abschlüsse durch Next- Best-Action Marketing zu steigern. Nach der Erklärung der Begriffe „Next Best Product", „Next Best Filter" und „Next Best User" stellte er die Bestandteile der analytischen Modellierung vor. Die Herausforderungen, die es beim Design des Datenmodells zu meistern gab, rührten in erster Linie aus der Vielzahl an verschiedenen Quellsystemen und der daraus resultierenden Komplexität der ETL Prozesse her.

Das Design und die Implementierung hierfür machten mehr als die Hälfte des Projektaufwands aus. Für die analytischen Modelle wurden insgesamt mehr als 300 Features abgeleitet, die sich in die Bereiche Web-Analytics, Abschlussdaten, Produkt- und Tarifdaten der Abschlüsse, Kundendaten und sozioökonomische Daten einteilen lassen. Diese werden zur Vorhersage von inzwischen 36 Zielgrößen verwendet. Durch die Entwicklung eines flexiblen Metadatenmodells ist es einfach möglich, Features und zu berechnende Modelle auszuwählen, sowie neue Features und Target-Variablen hinzuzufügen, ohne aufwändige Anpassungen am Code durchführen zu müssen. Dies würde bei mehreren Hundert Features und einer steigenden Zahl von Modellen sonst sehr

komplex werden. Die Metadaten bilden die zentrale Abstimmungsstelle zwischen Data Scientists, Domänenexperten und der Datenbank-Administration. Nebenbei liefern sie eine einfach nachvollziehbare Dokumentation. Im Folgenden erläuterte Herr Becker die Vorzüge des gewählten Algorithmus XGBoost (Boosted Trees). Neben der hohen prädiktiven Leistungsfähigkeit ermöglicht die Robustheit des Algorithmus gegenüber korrelierten und irrelevanten Variablen, Ausreißern, sowie fehlenden Werten die Verwendung des gleichen Feature-Vektors für alle NBX-Modelle. Als vorletzter Punkt ging er auf die Operationalisierungs-Aspekte der Lösung ein. Kernstück ist das geschilderte Metadatenmodell, welches sowohl für das monatliche Training der Modelle, als auch für das tägliche Scoring verwendet wird. Abschließend stellte er nochmals die kritischen Erfolgsfaktoren des Projektes vor: saubere und historisch korrekte Daten, die metadatengesteuerte Datenverarbeitung und den robusten Machine Learning Algorithmus.

In der anschließenden Fragerunde und auch schon während des Vortrags entwickelte sich eine rege Diskussion. Dabei ging es um Themen wie Ansätze bei der analytischen Modellierung, einflussreiche Features, Auswahl des Algorithmus, Validierung der Ergebnisse, Datenschutz und den Einsatz der entwickelten Modelle.

Die Gelegenheit zum anschließenden Networking wurde rege genutzt. Prof. Dr. Haneke vom KIT erklärte sich bereit, bei der Organisation und Durchführung des Roundtables zu unterstützen. So wird es in Zukunft auch Veranstaltungen in Karlsruhe geben.

Die Termine für die nächsten Roundtable stehen bereits fest: Sie finden statt am **05.06.2019** und **20.11.2019**.

Roland Mannshardt