

Statistische Modelle für Umsatzprognosen bei limitierter Datenverfügbarkeit

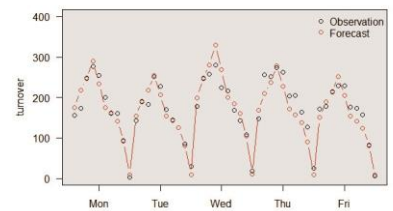
In der Schweiz landet rund ein Drittel aller produzierten Lebensmittel im Abfall. Das Start-Up-Unternehmen Prognolite GmbH vertritt in der Gastronomiebranche eine eigens entwickelte Software, mit der u.a. die voraussichtliche Anzahl nachgefragter Menus mit einem Machine-Learning-Verfahren prognostiziert werden kann. Diese Software hilft den Kunden, Personalkosten, Lebensmittelabfälle und Wartezeiten zu reduzieren. Basis für die Prognosen sind zeitaufgelöste Umsatzzahlen des Gastronomiebetriebes sowie Wetterdaten über die letzten zwei Jahre. Das Ziel dieser Arbeit war die Entwicklung eines statistischen Verfahrens, welches Gastronomiebetrieben auch brauchbare Prognosen erlaubt, wenn deutlich weniger Daten als von zwei Jahren zur Verfügung stehen. Die Wetterdaten werden dabei nicht berücksichtigt.

Anhand der statistischen Datenanalyse wurden verschiedene multiple lineare Regressionsmodelle entwickelt und als R-Programm implementiert. Der Algorithmus entscheidet je nach Datenlage, welches der Regressionsmodelle für die aktuelle Prognose angewendet wird. Um die Prognosegenauigkeit über das ganze Jahr analysieren zu können, wurden rollende Prognosen erstellt, bei welchen immer gleich viele Daten für die Schätzung der Regressionskoeffizienten verwendet wurden. Die Prognosegüte wurde durch einen modifizierten "Absolute Percentage Error" (Modified APE) berechnet. Dieses Gütekriterium wurde von der ZHAW entwickelt und sowohl von den Autoren dieses Berichts als auch vom Industriepartner verwendet. Anhand dieses Gütekriteriums wurde auch die minimal notwendige Beobachtungslänge an empirischen Kundendaten bestimmt, um eine optimale Prognosequalität zu erreichen. Für den untersuchten Betrieb zeigte sich, dass der implementierte Algorithmus mit Daten von nur acht Wochen die besten Vorhersagen liefert. Der Median-APE der Umsatzzahlen für den Tagesumsatz betrug 9% und das 90%-Quantil des APE 21%. Die entsprechenden Zahlen für den Algorithmus des Industriepartners lauten 7% und 21%. Diese Ergebnisse zeigen, dass im Vergleich zu modernen Machine-Learning-Verfahren auch mit klassischen, relativ einfachen statistischen Verfahren ein vergleichbares Resultat erzielt werden kann, wobei der Bedarf an Daten für die statistischen Verfahren deutlich bescheidener ist.

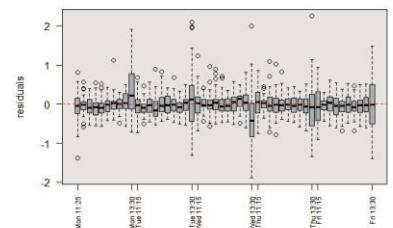


Diplomierende
Loran Halit Avcı
Timo Dominic Harlacher

Dozierende
Christoph Hofer
René Locher



Diese Grafik zeigt viertelstündliche Prognosen und die dazugehörigen Beobachtungen für eine Woche.



Diese Grafik zeigt die Residuen im logarithmierten Raum gegen die Zeit pro Wochentag. Es sind keine systematischen Fehler zu beobachten, wobei die Residuen jeweils um 13:30 Uhr breiter streuen.