

## ABSTRACT

---

Accurate forecasts of demand play an important role in many businesses and industries. Especially in the retail sector, these forecasts build the basis for planning various supply-chain activities such as stock management or the allocation of scarce resources and personnel. In this context, the underlying data are often time series. What makes time series data special is that successive observations are usually not independent. In order to make predictions of the future course of time series, established time series models therefore attempt to model the inherent structures and patterns based on past observations. Potentials that external data sources may offer in the form of leading indicators are often neglected.

Therefore, this work addresses the research question whether and under which conditions the integration of external data contributes to improving the accuracy of forecasts. Based on a similarity measure, leading indicators from an external online open data source are determined for the time series of a German retail company. Within an experiment, these leading indicators are incorporated individually as external regressors into a linear time series model. The comparison of the forecasting performance between the univariate and bivariate models is intended to provide information on which factors are responsible for the successful inclusion of leading indicators.

The results of the experiment reveal that a certain time series pattern has a statistically significant influence on the outcome of the inclusion. When both the time series of the retail company and the time series of their corresponding leading indicators exhibit this pattern, the integration of the external regressors can be particularly beneficial for the forecast accuracy. Moreover, the similarity of the time series pairs plays a subordinate role as the results additionally show.

*Keywords*— Forecasting; Leading Indicator; Similarity Matching; Google Trends; ARIMAX model.

## ZUSAMMENFASSUNG

---

Möglichst genaue Nachfrageprognosen spielen in vielen Unternehmen und Branchen eine wichtige Rolle. Insbesondere im Einzelhandel bilden robuste Prognosen von Abverkäufen die Grundlage für die Planung verschiedener Aktivitäten innerhalb einer Lieferkette, wie z. B. die Lagerverwaltung oder die Personal- und Ressourcenzuteilung. Häufig handelt es sich bei den zugrundeliegenden Daten um Zeitreihen. Das Besondere an Zeitreihendaten ist, dass aufeinanderfolgende Beobachtungen in der Regel nicht unabhängig voneinander sind. Etablierte Zeitreihenmodelle versuchen daher die der Zeitreihen inhärenten Strukturen und Muster auf Basis historischer Werte zu modellieren, um Vorhersagen für den zukünftigen Verlauf zu treffen. Mögliche Potenziale, die externe Datenquellen in Form von Frühindikatoren bieten können, werden dabei oftmals vernachlässigt.

Die vorliegende Arbeit beschäftigt sich daher mit der Forschungsfrage, ob und unter welchen Bedingungen die Integration von externen Daten zur Verbesserung der Prognosegenauigkeit beitragen kann. Hierfür werden auf Basis eines Ähnlichkeitsmaßes Frühindikatoren aus einer Online-Datenquelle für die Zeitreihen eines deutschen Einzelhandelsunternehmens bestimmt. Im Rahmen eines Experiments werden diese Frühindikatoren anschließend einzeln als externe Regressoren in ein lineares Zeitreihenmodell aufgenommen. Der Vergleich der Vorhersagegenauigkeit zwischen den univariaten und bivariaten Modellen soll Aufschluss darüber geben, welche Faktoren für die Einbeziehung von Frühindikatoren erfolgskritisch sind.

Die Ergebnisse des Experiments zeigen, dass ein bestimmtes Zeitreihenmuster einen statistisch signifikanten Einfluss auf den Ausgang der Einbeziehung hat. Wenn sowohl die Zeitreihen des Einzelhandelsunternehmens als auch die Zeitreihen ihrer entsprechenden Frühindikatoren dieses Muster aufweisen, kann die Integration der externen Regressoren besonders vorteilhaft für die Prognosegenauigkeit sein. Wie die Ergebnisse ebenfalls veranschaulichen, spielt die Ähnlichkeit der Zeitreihenpaare dabei eine untergeordnete Rolle.

*Schlagworte*— Forecasting; Leading Indicator; Similarity Matching; Google Trends; ARIMAX model.