Titel: Einführung in den Vorhersage-Wettbewerb

Autor: Lukas Küppers, Hendrik Kaiser und Vincent Schmalor

Datum: 19.12.2019

**Einleitung**

*Literaturrecherche? (Nachfragen am Freitag)*

Dieses Dokument stellt sowohl die Durchführung, als auch die Protokollierung des Vorhersagewettbewerbs dar. Die Aufgabenstellung besteht darin, ein Modell aus einem Grunddatensatz zu entwickeln und dieses Modell auf einen Anwendungsdatensatz anzuwenden. Dabei wird aus mehreren Einflussvariablen eine abhängige Variable geschätzt. Der konkrete Anwendungsfall ist ein Auszug aus einem Face-to-Face Interview über das Konsum- und Schenkverhalten zu Weihnachten. Das Grunddatenset beinhaltet 400 Beobachtungen mit 17 erklärenden Variablen und die abhängige Variable des geschätzten Budgets, die es im Anwendungsdatenset zu modellieren gilt.

In dem folgenden Kapitel wird diese Aufgabe bearbeitet, indem zunächst ein Überblick über die Daten in einer explorativen Datenanalyse gegeben wird. Anhand verschiedener Analysen und der Untersuchung der semantischen Zusammenhänge wird dann eine Vorgehensweise erarbeitet und beschrieben.

Das Ergebnis ist eine csv-Datei, die das Anwendungsdatenset mit der modellierten, abhängigen Variable beinhaltet. So wird eine Prognose über diese Variable für jede Beobachtung gegeben, für die das tatsächliche Budget nicht vorliegt. Eine Zusammenfassung der Ergebnisse findet sich auch auf dem Plakat der genannten Autoren.

**Explorative Datenanalyse**

*Rohdatenanalyse und -strukturierung*

*Bereinigung der Datenmenge*

*Erweiterung um abgeleitete Metriken*

*Visualisierungen der Datenmenge*

*Lineare Modelle*

*Übersicht*

*Detailanalyse*

**Methodik**

**Modellbildung und -anwendung**

**Ergebnis**

**Fazit**