

# Entwicklung eines Data-Mining-Modells zur Bestimmung der Abhängigkeit zwischen Verspätungen und Spritverbrauch von Bussen

Lukas Ehl

Technische Hochschule  
Mittelhessen

Fachbereich MNI  
Wiesenstr. 14  
35390 Gießen  
E-Mail:

[lukas.ehl@mni.thm.de](mailto:lukas.ehl@mni.thm.de)

Prof. Dr. Harald Ritz

Technische Hochschule  
Mittelhessen

Fachbereich MNI  
Wiesenstr. 14  
35390 Gießen  
E-Mail:

[harald.ritz@mni.thm.de](mailto:harald.ritz@mni.thm.de)

Ingo Nobbers

Technische Hochschule  
Mittelhessen

Fachbereich MNI  
Wiesenstr. 14  
35390 Gießen  
E-Mail:

[ingo.nobbers@mni.thm.de](mailto:ingo.nobbers@mni.thm.de)

## Kategorie

Bachelorarbeit

## Schlüsselwörter

Data Mining, Python, Telematiksysteme, Logistik, Pandas, Digitalisierung, Big Data, Business Intelligence, CRISP-DM

## Zusammenfassung

Die Menge der gesammelten Daten hat sich in den letzten Jahren vervielfacht. Unternehmen sammeln immer mehr Daten zur Informationsgewinnung, um daraus einen Mehrwert ziehen zu können.

Data Mining ermöglicht nicht nur das Sammeln von großen Datenmengen, sondern auch die Analyse.

In der vorliegenden Bachelorarbeit besteht die Aufgabenstellung darin, die Telematikdaten per GPS-Koordinaten und durch die Fahrt festgelegte Zeitspannen mit den Daten eines Echtzeitdatensystems zusammenzuführen.

Dazu werden die theoretischen Grundlagen erarbeitet, um anschließend ein Data Mining-Modell mit Hilfe von CRISP-DM zu entwickeln. Die verschiedenen Voraussetzungen der Datensätze in Hinblick auf Datenqualität und -quantität erfordert von Beginn an eine Herangehensweise, welche auf ein ressourcenschonendes Modell abzielt.

Die große Menge an Telematikdaten muss für diese Zielsetzung eingeschränkt werden, ohne dabei an Informationsgehalt zu verlieren und anschließend durch die Nutzung des selbst entwickelnden Modells mit den Echtzeitdaten zusammengeführt. Die Entwicklung und Inbetriebnahme des Systems geschieht mit Hilfe von Python.

Nach der Entwicklung eines funktionierenden Modells werden in dieser Bachelorarbeit die Fahrtstatistiken analysiert. Mit Hilfe einer Regressionsanalyse und der Berechnung des Korrelationskoeffizienten nach Pearson wird die Forschungsfrage beantwortet, ob eine

Abhängigkeit zwischen der Verspätung und dem Spritverbrauch von Bussen besteht.

Die Ergebnisse der Entwicklung dieses Modells können bei Folgeanalysen eingebracht werden und somit zu einer effizienteren Steuerung des Unternehmens beitragen.