

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit der Verbesserung eines Empfehlungssystems für Verspätungsgründe im Schienenverkehr durch die automatisierte Generierung räumlicher und zeitlicher Features sowie den Einsatz verschiedener Data-Matching-Methoden. Besonderer Fokus liegt dabei auf der Balance zwischen Modellgüte und Interpretierbarkeit, da beide Faktoren essenziell für das Vertrauen und die Akzeptanz der Nutzenden sind.

Zur Generierung der Features wurde die `tsfresh`-Bibliothek eingesetzt, die eine automatisierte, datengetriebene Extraktion relevanter Merkmale ermöglicht. Für das Data Matching wurden sowohl maschinelle Lernmethoden als auch regelbasierte Ansätze evaluiert. Historische Verspätungsdaten dienten als Grundlage, um die Verknüpfungsqualität, die Erklärbarkeit der Modelle sowie die Effektivität der verwendeten Ansätze systematisch zu bewerten.

Die Ergebnisse zeigen, dass die gewählte automatisierte Feature-Generierung zu signifikanten Verbesserungen der Modellgüte führt, ohne die Interpretierbarkeit der Features zu beeinträchtigen. Im Bereich des Data Matching konnte der Machine-Learning-Ansatz eine deutliche Steigerung der Modellgüte um bis zu 39 % erreichen, während regelbasierte Methoden vor allem durch ihre Einfachheit und Transparenz überzeugten.

Abschließend wird festgestellt, dass die Kombination aus interpretierbaren Features und leistungsstarken Matching-Methoden eine entscheidende Grundlage für die Verbesserung der Qualität von Empfehlungssystemen im Schienenverkehr darstellt.