

Projektbericht

Entwicklung einer Lösung maschinellen Lernens zur Imputation fehlender Sendungsdaten in der Logistik

Der Kunde

Fr. Meyer's Sohn – oder kurz FMS – ist ein globales Speditions- und Logistikunternehmen in Familienbesitz, das sowohl See-, Luft-, als auch Landtransporte organisiert und somit für seine Kunden die gesamte Transportkette abdeckt. Weltweit gehört FMS zu den 10 größten Seefrachtspediteuren.



Die Herausforderung

Ein wesentlicher Teil der Arbeit besteht darin, Kundenaufträge im Transportmanagementsystem (TMS) zu erfassen, um Sendungen anzulegen. Fehlende Informationen werden hierbei von Disponenten durch ihr intrinsisches Wissen vervollständigt.

Zielsetzung des Projektes ist eine Lösung, die mittels statistischer Methoden und maschinellen Lernens fehlende Informationen aus Kunden-E-Mails möglichst genau schätzt (*imputation*) und eine Schnittstelle zum automatischen Erfassen von Sendungsdaten in E-Mails ermöglicht.

Besondere Herausforderungen entstehen durch das unregelmäßige Fehlen von Daten und durch heterogene Datentypen, für die einzelne Modelle trainiert werden müssen. Und darin, die Felder der Maske fachlich sinnvoll zu befüllen.

Die Umsetzung

Das Projekt wurde in zwei Phasen umgesetzt.

(1) In der ersten Phase wurden eine Teilmenge der vorhandenen Daten analysiert und bestehende Geschäftsregeln erhoben. Neben Standardaufgaben wie der Datenbereinigung wurde hier ein besonderes Augenmerk auf das *Association Rule Mining* gelegt, um verdeckte Regel zu identifizieren. Wir wählten *MissForest* als geeignetes Modell zum Vervollständigen aus. Es wurde implementiert und anhand einer Testfunktion mit zufällig fehlenden Werten evaluiert. Des Weiteren haben wir eine Pipeline zur Datenvorbereitung konzipiert und mit Hilfe von maßgeschneiderten Transformern - also einzelnen Datenvorbereitungsschritten - umgesetzt.

(2) In der zweiten Phase des Projektes ging es intensiv um die Auswertung und Verbesserung der Ergebnisse. Dazu haben wir ein weiteres Modell sowie eine Benchmark zum Vergleichen implementiert. Die Modelle passten wir auf den vollständigen Datensatz an und überarbeiteten die Implementation der Pipeline, um die Kompatibilität mit bestehenden Transformatern aus *sklearn* zu gewährleisten. Das Deployment der Modelle wurde beispielhaft mit Hilfe des Frameworks *flask* als API aufgezeigt.

Das Ergebnis

Eine unvollständige Sendung kann jetzt als JSON mit einem POST-Request an die API geschickt werden und wird mit allen Pflichtfeldern vervollständigt als JSON zurückgegeben. Für die Imputation stehen zwei Modelle zur Verfügung, welche nun in der Praxis erprobt werden, um Sachbearbeiter zukünftig durch fachlich stimmige Vorschläge zu unterstützen.

Die Projektdetails auf einen Blick

- > Automatische Anlage von Sendungen
- > Imputation mittels KI
- > Unterstützung von Disponenten
- > Projektlaufzeit: ca. 10 Monate

Die eingesetzten Technologien

Vorgehen	CRISP-DM
Schritte	Datenanalyse, Modelling und Evaluierung
Umgebung	JupyterLab (Python 3)
Module	scikit-learn, flask, pandas, numpy

Der Ansprechpartner

Tim Soller
E | tim.soller@hec.de
T | 0421 20 750 370