

Ermittlung von Wiederbeschaffungszeiten anhand eines Machine Learning Prototypen

Sven Fischer, Wirtschaftsinformatik (Masterstudiengang)
Prof. Dr. Herbert Neuendorf, DHBW Mosbach



Beschreibung

Die aktuellen instabilen Lieferketten führen dazu, dass Wiederbeschaffungszeiten sich regelmäßig stark verändern. Debitoren erwarten allerdings, dass über die Wiederbeschaffungszeit Auskunft erteilt werden kann. Bei dem Optima-Konzern (Sondermaschinenbauer in der Verpackungsmaschinenbranche) betrifft dies im besonderen den Bereich Service. Aufgrund der hohen Anzahl an Materialstämmen und regelmäßigen Neuanlagen ist die manuelle Pflege mit sehr hohem Aufwand verbunden. Ein deterministisches Vorgehen ist aufgrund der hohen Komplexität nicht zielführend.

Zielsetzung

Die Hypothese „**Die Ermittlung einer ausreichend genauen Wiederbeschaffungszeit kann durch bereits vorliegende strukturierte Daten vorgenommen werden**“ ist im Bezug auf den Kontext des dualen Partners zu verifizieren. Da von dem Anwendungsfall nicht auf ein ML-Verfahren zu schließen ist, wird zudem die Forschungsfrage: „**Welches der geprüften Verfahren des maschinellen Lernens ist das effektivste zur Vorhersage der Wiederbeschaffungszeit?**“ betrachtet.

Methodisches Vorgehen

Die Forschungsfrage setzt voraus, dass verschiedene ML-Verfahren analysiert werden. Das methodische Vorgehen der Erstellung der Modelle orientiert sich an dem CRISP-DM-Modell. Diese experimentellen Prototypen sind in den Bereich des überwachten Lernens einzuordnen. Die Datenbasis stellen Daten des ERP-Systems dar. 20.000 Datensätze wurden für die Modellerstellung manuell gelabelt. Damit die Ergebnisse der Modelle vergleichbar sind, wird das CRISP-DM-Modell für diese wissenschaftliche Arbeit in abgewandelter Form verwendet.

Prototypische Implementierung

Die Implementierung wurde in der Programmiersprache Python und primär unter Verwendung der Bibliothek scikit-learn umgesetzt. Die Wiederbeschaffungszeiten werden in fünf bzw. sieben Klassen eingeteilt, die vorhergesagt werden sollen. Somit handelt es sich um eine Klassifikationsaufgabe mit einem Label. Im Zuge der Arbeit wurden zwei Iterationen mit den fünf ausgewählten Verfahren (Logistische Regression, SVM, Random Forest, k-nächste-Nachbarn, KNN) durchgeführt. Zusätzlich wurde ein Dummy-Modell erzeugt, welches die ERP-Daten nicht kennt und somit einen Vergleichswert darstellt. Ohne Optimierung der Hyperparameter wurde die erste Iteration durchgeführt. Die zweite Iteration zeichnete sich durch die Verbesserung der Hyperparameter aus. Im folgenden werden die Ergebnisse der Evaluation der zweiten Iteration dargestellt:

Verfahren	Sensitivität	Relevanz	F1-Score	MAE	RSME
Logistische Regression	61,81%	61,81%	61,49%	1,26	1,12
SVM	63,84%	63,77%	63,64%	1,21	1,1
Random Forest	69,74%	69,70%	69,59%	0,99	0,99
k-nächster-Nachbar	62,56%	62,25%	62,25%	1,31	1,14
KNN	62,01%	61,92%	61,92%	1,25	1,12
Dummy	30,71%	9,43%	14,43%	1,63	1,28

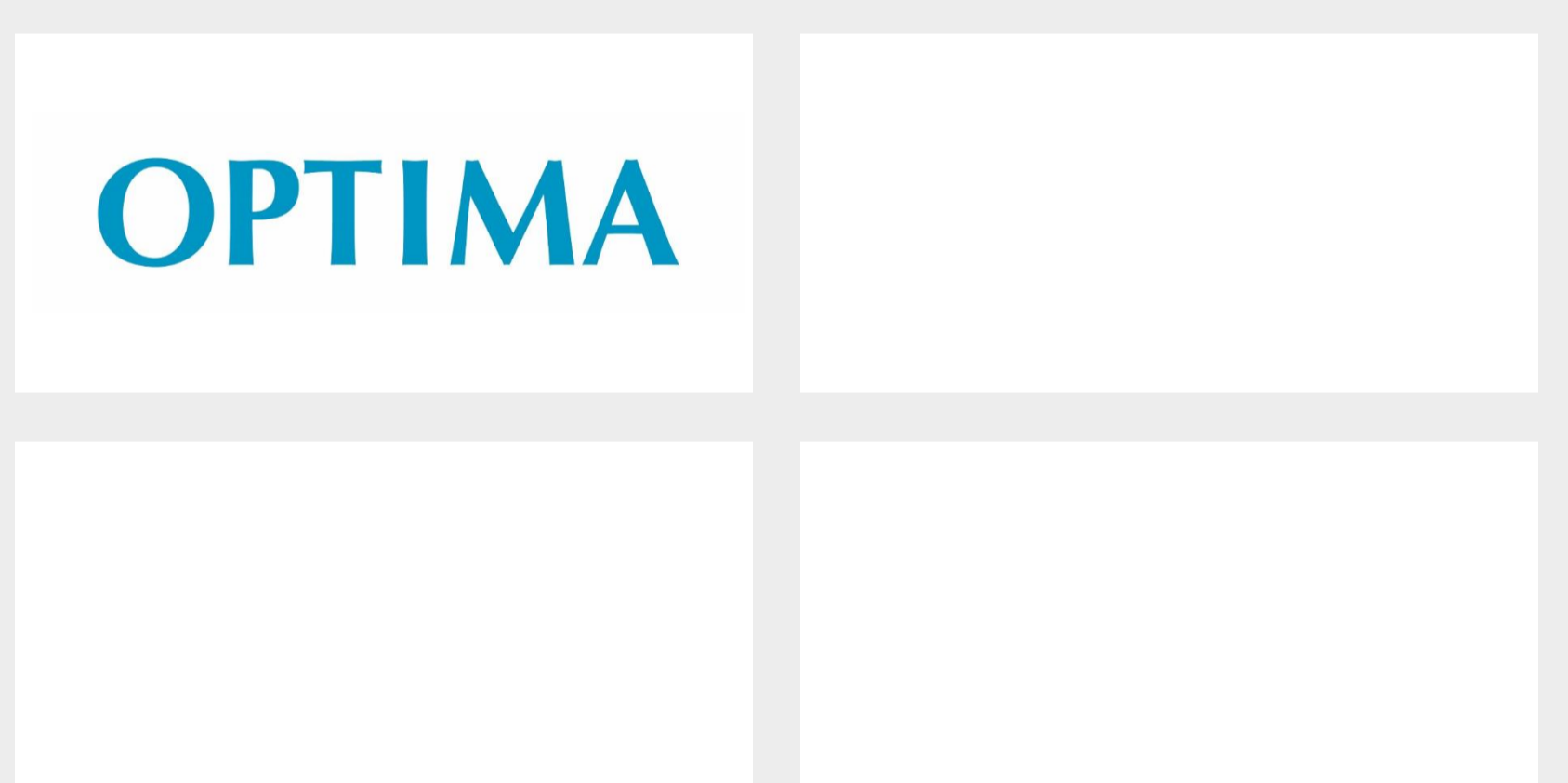
Ergebnisse

- » Aus dem Dataset konnten Rückschlüsse auf die Wiederbeschaffungszeit gezogen werden
- » Eine Definition von „ausreichend genau“, in Bezug auf Auswahl und Wert einer Kennzahl für die Bewertung des Modells, konnte noch nicht exakt ermittelt werden
- » Bei dem aktuellen Grad der Optimierung ist das vorteilhafteste Modell das Random Forest Modell

Ausblick

- Das Thema wird im Unternehmenskontext als experimenteller Prototyp weiterverfolgt. Im Fokus steht hierbei:
- » Implementierung eines Prozesses zur kontinuierlichen Ermittlung gelabelter Daten
 - » Einbindung des ML-Modells in bestehende Prozesse des ERP-Systems
 - » Verbesserung der Modelle
 - » Betrachtung und Automatisierung des Modell-Lebenszyklus

Kooperative Partner



Quellen

- Im folgenden die prägendste Literatur bezüglich der Methodik, Grundlagen- und Praxisliteratur für diese wissenschaftliche Arbeit:
- » Wirth R, Hipp J (2000) CRISP-DM: Towards a standard process model for data mining. Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining.
 - » Russell SJ, Norvig P (2012) Künstliche Intelligenz: Ein moderner Ansatz. Pearson, München.
 - » Géron A (2020) Praxiseinstieg Machine Learning mit Scikit-Learn, Keras und TensorFlow. Konzepte, Tools und Techniken für intelligente Systeme. O'Reilly, Heidelberg.

Kontakt

Duale Hochschule Baden-Württemberg

Autor: Sven.Fischer@optima-packaging.com
Betreuer: Herbert.Neuendorf@mosbach.dhbw.de