

AIS BEWERBERPROZESS

Lösungs- präsentation

NOAH WOLFAHRT

ERGEBNISÜBERSICHT

Vollständigkeitsprüfung:

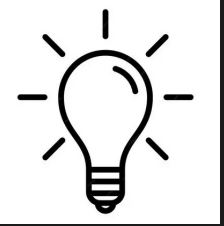
Der Algorithmus erkennt insgesamt 287 fehlende Anlagen in der Kundendatei.

EP-Zuordnung:

Der Algorithmus weist mit einer geschätzten Genauigkeit von 92.5 % die richtige Anlagennummer zu.

(Link zur technischen Lösung)

TEIL 1 - VOLLSTÄNDIGKEITSPRÜFUNG



IDEE - CLUSTERING MIT KMEANS

Annahme:

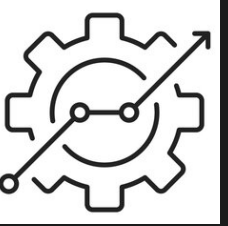
- **verschiedene Typen von Gebäuden** mit verschiedenen Anlagen im Datensatz
(Bürogebäude, Wohnhäuser, Lagerhallen, ...)

Cluster-spezifischer Ansatz:

- verschiedene Cluster von KI ermitteln lassen
und jedem Gebäude ein Cluster zuweisen
- pro Cluster **die wichtigsten Anlagen** ermitteln
- neue Gebäude zunächst zuordnen und dann auf
clustertypische Anlagen überprüfen



TEIL 1 - VOLLSTÄNDIGKEITSPRÜFUNG



ERGEBNIS UND OPTIMIERUNGSPOTENZIAL

Es werden durchschnittlich 47 Anlagen pro Cluster als relevant bewertet. In der Kundendatei werden durchschnittlich 41 Anlagen pro Gebäude als fehlend markiert.

Der Algorithmus könnte noch folgendermaßen optimiert werden:

- **Anpassung des Relevanzniveaus**

Was sind wichtige Anlagen?

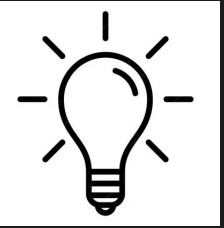
(hier: Anlagen, die in 90 % aller Gebäude desselben Clusters vorkommen.)

- **Anpassung der Clusteranzahl**

Mit wie vielen verschiedenen Gebäudetypen soll kalkuliert werden?

(hier: 4)

TEIL 2 - ZUORDNUNG



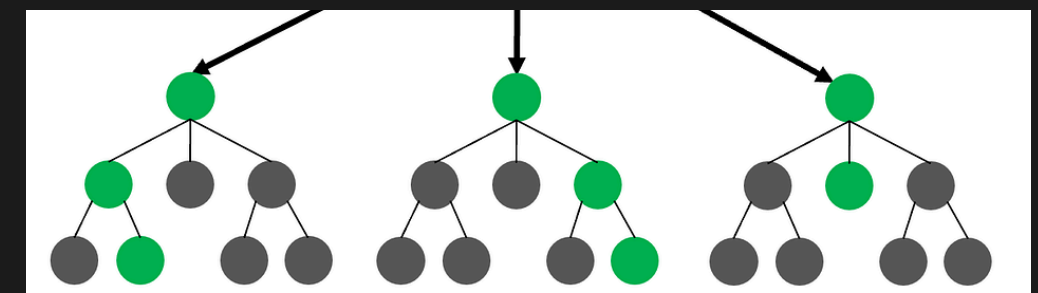
IDEE - TFIDF/LOGREG UND DECISION TREES

Zentrale Erkenntnis:

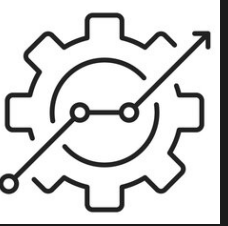
- simpler Algorithmus (TFIDF + LogReg) erkennt bereits relativ zuverlässig die Normalgruppe (NG) einer Anlage (98.2 %), nicht jedoch die Normalposition (NP) (ca. 70 %)

Mehrstufiger Ansatz:

- Trainiere für jede NG ein eigenes Decision-Tree-Modell.
- Lasse zunächst die NG vorhersagen. Je nach Resultat wird die NP anschließend mit einem auf die spezifische NG angepassten Decision-Tree-Modell vorhergesagt.



TEIL 2 - ZUORDNUNG



ERGEBNIS UND OPTIMIERUNGSPOTENZIAL

Im internen Test Set werden jetzt 92.5 % aller Anlagen richtig zugeordnet.

Der Algorithmus könnte noch folgendermaßen optimiert werden:

- **Überprüfung der Relevanz bestimmter Felder / entsprechende Gewichtung**
(z. B. Wie relevant ist das Feld Kostengruppen-Bezeichnung?)
- **Auswahl des Textkonverters**
(verschiedene Algorithmen denkbar)
- **Feinkalibrierung der einzelnen Decision Trees**