

ETIM Klassifikation mit ML-Approach

Vorhersage der ETIM-Klasse basierend auf Daten aus der Geberit Datenbank

Studenten



Etienne Baumgartner



Nathanael Gall

Ausgangslage: Das ETIM Klassifikationsmodell setzt sich auf dem internationalen Markt immer mehr durch. Die ETIM Klassifikation vereinfacht einerseits den Datenaustausch zwischen Händler und Hersteller und andererseits die Klassifizierung der Produkte. Die Geberit AG nutzt diesen Standard für ihren Produktkatalog und die Klassifizierung wird manuell durch einen Sachbearbeiter vorgenommen.

Das Ziel der Arbeit ist die Entwicklung und der Vergleich von zwei Verfahren zur automatischen Bestimmung der ETIM Klasse auf der Basis von maschinellem Lernen.

Vorgehen: Der Erfolg der Arbeit beruht auf der richtigen Verwertung der von Geberit zur Verfügung gestellten Produktdaten. Auf Grund der Analyse wurden zwei Ansätze entwickelt, um die ETIM Klassifizierung zu erlernen.

1. One-Hot-Encoding

Die Daten werden für das Trainieren auf einem Deep Neural Network One-Hot-Encoded. Um die Anzahl Dimensionen des Netzwerks möglichst klein zu halten, wird anhand des Bayes Errors die optimale Spaltenkombination ermittelt. Auf dieser Auswahl wird ausserdem ein Lookup Table erstellt, um eindeutige Datensätze direkt zu klassifizieren.

2. Textembedding

Viele der Spalten weisen einen grossen Anteil an deutschem Text auf. Es wird mit der fastText Library ein Textembedding trainiert, welches die Artikel den jeweiligen ETIM Klassen zuordnet.

Fazit: Nach der Optimierung dieser zwei Ansätze stellt sich heraus, dass der textbasierte fastText Algorithmus genauere Resultate liefert. Mit dem fastText Model wird auf dem vorab abgetrennten finalen Testset eine Top-1-Accuracy von 0.961 und eine Top-3-Accuracy von 0.982 erreicht.

Die höhere Accuracy ist auf die bessere Verwertung der Spaltentexte zurückzuführen, weil das Textembedding die Nähe zweier verschiedener Werte abbilden kann, während beim One-Hot-Encoding alle unterschiedlichen Werte äquidistant sind.

Die Empfehlung ist, die drei wahrscheinlichsten Klassen des fastText Top-3-Algorithmus zu präsentieren und die finale Entscheidung einem Benutzer zu überlassen.

Examinator

Prof. Oliver Augenstein

Experte

Schmid Nico, Jona, St. Gallen

Themengebiet

Verschiedenes

Projektpartner

Geberit AG,
Schachenstrasse 77,
8645 Rapperswil-Jona,
St. Gallen