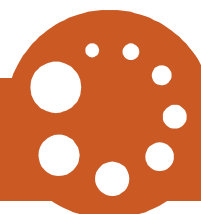




Das Dateninnovationsprojekt «NOGAuto»

Management Summary

EXPERIMENTAL STATISTICS



Neuchâtel, 2024

Herausgeber: Bundesamt für Statistik (BFS)

Auskunft: info.urd@bfs.admin.ch

Redaktion: Lorenz Helbling ; Mathias Constantin ;
Daniele Marx ; Cindia Duc Sfez ; BFS

Themenbereich: 00 Statistische Grundlagen

Originaltext: Französisch

Übersetzung: Sprachdienste BFS

Layoutkonzept: Sektion PUB

Download: www.statistik.ch

Copyright: BFS, Neuchâtel 2024
Wiedergabe unter Angabe der Quelle
für nichtkommerzielle Nutzung gestattet

Das Dateninnovationsprojekt "NOGAuto"

Management Summary

Das Projekt NOGAuto ist Teil der Dateninnovationsstrategie des Bundesamtes für Statistik (BFS) und gehört damit zu den wichtigsten Pilotprojekten zur Verbesserung der Statistikproduktion. Mit NOGAuto soll anhand von Methoden des maschinellen Lernens die Kodierung der wirtschaftlichen Tätigkeit von Betrieben und Unternehmen automatisiert werden. Diese Automatisierung soll somit die Mitarbeitenden des BFS bei der Klassifizierung der Unternehmen nach der Allgemeinen Systematik der Wirtschaftszweige (NOGA) unterstützen. Da viele der für das maschinelle Lernen notwendigen Daten beim BFS bereits vorhanden sind (Erhebungsdaten, Tätigkeitsbeschreibungen im Handelsregister, Erläuterungen zur Klassifikation usw.), konzentrierte sich das Projekt auf deren Nutzung.

In einem ersten Schritt wurden die im Schweizer Betriebs- und Unternehmensregister (BUR) verfügbaren Daten ausgewählt und für die weitere Verwendung nutzbar gemacht. Zu diesen Daten gehören die Beschreibungen der Wirtschaftszweige, die als Hauptquelle für die Zuweisung eines NOGA-Codes zu den Betrieben und Unternehmen dienen. Mithilfe der sogenannten natürlichen Sprachverarbeitung («natural language processing») wurden aus diesen Beschreibungen möglichst viele nützliche Informationen für den Algorithmus herausgefiltert, mit dem sich der NOGA-Code des Unternehmens ermitteln lässt. Anschliessend wurden mit diesen Daten anhand von maschinellem Lernen Modelle gebildet (eines pro Sprache), die an einer Zufallsstichprobe getestet wurden. Dadurch liess sich die Qualität der Prädiktion der Zuteilung beurteilen.

Auf der Grundlage dieser Modelle wurden verschiedene Tools entwickelt, um in die Produktion des BFS überführt zu werden. Ein erstes Tool dient zur Kontrolle von Listen und soll damit die Qualität der in den Registern bereits vorhandenen NOGA-Codes verbessern. Eine weitere Anwendung soll die Mitarbeitenden bei der NOGA-Kodierung von Unternehmen auf Basis der Beschreibungen der jeweiligen wirtschaftlichen Tätigkeiten unterstützen.