

Klassifizierung von Unternehmensbeschreibungen

Daniel Gauermann (71411) & Yannick Maurer (68036)

Beschreibung

Die Plattform Yahoo! Finance verfügt über eine Vielzahl von Informationen zu an der Börse gelisteter Unternehmen. Beispielsweise bietet die Plattform für jedes Unternehmen eine eigene Seite mit Unternehmensdaten und einer Beschreibung an. Ziel dieser Projektarbeit soll es sein, mithilfe der Unternehmensbeschreibungen eine Klassifizierung des jeweiligen Sektors des Unternehmens vorzunehmen. Dazu sollen verschiedene Modelle antrainiert und schließlich das Modell mit der höchsten Genauigkeit bestimmt werden.

Datenbeschaffung (Data Retrieval)

Zunächst werden Unternehmensbeschreibungen und Informationen über Web Scraping heruntergeladen und gespeichert. Bei den Unternehmensbeschreibungen handelt es sich um ungefähr 8.000 Beschreibungen mit ihrem tatsächlichen Sektor.

Datenvorverarbeitung

Die heruntergeladenen Beschreibungen werden anschließend vorverarbeitet (bspw. Tokenisierung, Stemming/Lemmatisierung). Mitunter werden die Textdaten auch in eine einheitliche Form gebracht, da diese für die spätere Klassifizierung entscheidend ist. Nicht benötigte Informationen wie bspw. Stoppwörter sollen aus den Texten gefiltert werden.

Klassifizierung

Nach der Datenbeschaffung sollen verschiedene Klassifizierungsmethoden trainiert werden und die Methode mit der höchsten Genauigkeit bestimmt werden.

Weitere mögliche Themen

Es ist möglich weitere Information aus den Unternehmensbeschreibungen zu extrahieren, sollten die Klassifizierungsmethoden zu kurz ausfallen. Beispielsweise könnten relevante Wörter und die damit verbundenen Technologien visuell hervorgehoben werden. Darüber hinaus könnten weitere Visualisierungen wie bspw. Wordclouds erstellt werden.