

# NATURAL LANGUAGE PROCESSING AUF EINEM JOBPORTAL

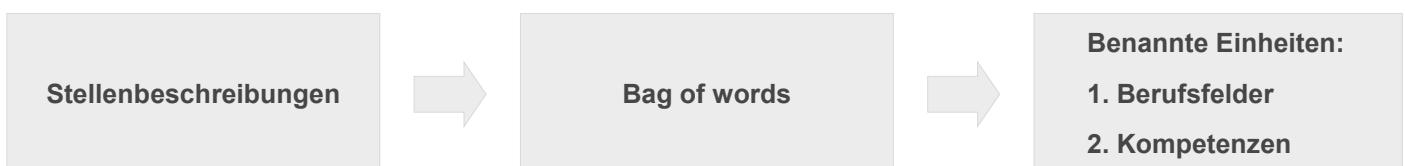
## PROJEKTZIEL

Die Analyse von Arbeitsmarkttrends oder die Suche nach einem geeigneten Kandidaten ist immer eine zeitraubende Aufgabe. So führt die Suche nach einem Thema schnell zu Hunderten oder gar Tausenden von Ergebnissen. Ziel des Projekts war es, ein Modell zu entwickeln, das Informationen über Berufsfelder und Kompetenzen der Stellenanzeigen gemäß den Stellenausschreibungen auf einem Jobportal extrahiert.

Die Implementierung gliederte sich in zwei Schritte: Erfassung der Daten von Stellenausschreibungen von mehr als 100 Unternehmen durch das Webcrawling eines Jobportals und die Entwicklung einer Deep-Learning-Lösung zur Erkennung der benannten Zielgrößen der 30 DAX-Mitglieder.

## ZUR VERFÜGUNG GESTELLTE DATEN

Das Projekt basiert auf einem Datensatz von veröffentlichten Stellenausschreibungen, die durch Webcrawling ermittelt wurden.



## HERAUSFORDERUNGEN

Viele Jobportale untersagen das Webcrawling. Zudem hat das verwendete Portal permanente Modifikationen an Quellcodes vorgenommen, was das Crawling erheblich erschwerte.

Beim Training eines Algorithmus wurde eine ausreichende Anzahl von Wörtern mit breitem Bedeutungsspektrum erkannt. Soweit nur bestimmte Eigenschaften von Interesse waren, wurden alle nicht relevanten herausgefiltert.

## ANGEWANDTE METHODEN

Ein strukturierter Datensatz mit Beschreibungen von Positionen, Firmennamen, Standorten und Erscheinungsdaten wurde durch einen Python-basierten Web-Crawler erstellt. Die Datenbank stand als CSV-Datei zum Download zur Verfügung.

Die Berufsfelder und Kompetenzen wurden mit Hilfe des Daturks-Tools gelabelt und im JSON-Format heruntergeladen.

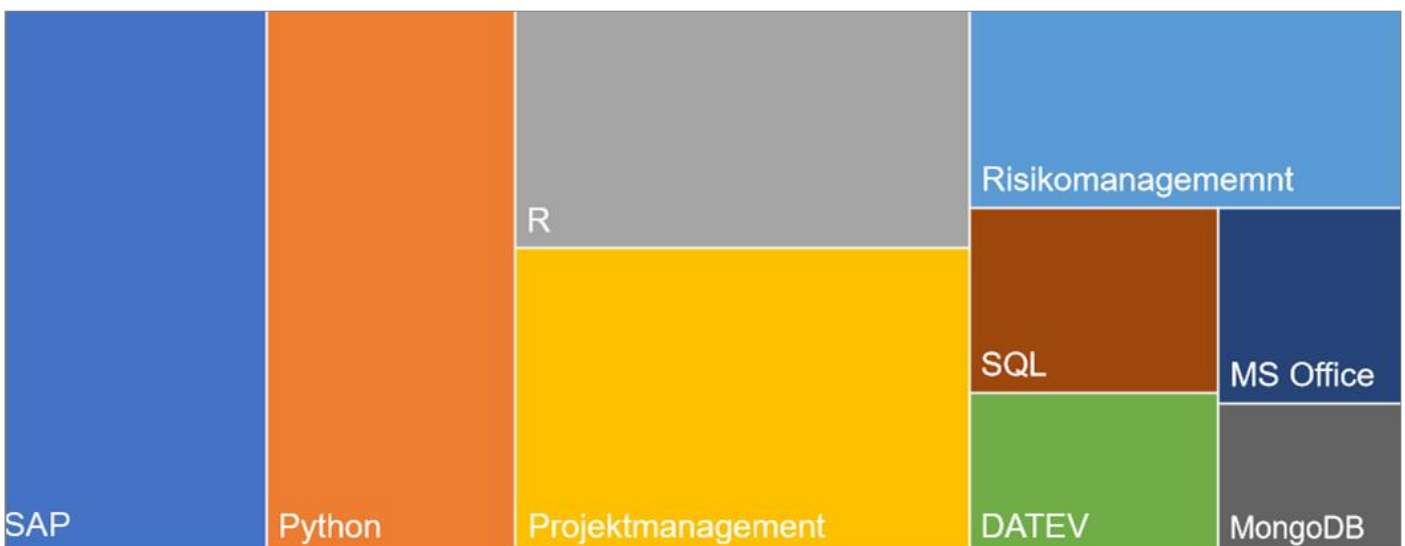


Die gelabelten Entitäten wurden in Trainings- und Testdatensätze unterteilt. Mit Hilfe der spaCy-Bibliothek für NLP wurde der Algorithmus trainiert. Das trainierte Modell wurde gespeichert und kann jeder Zeit für die anschließende Analyse geladen werden. Der Testdatensatz wurde von sklearn-metrics mit hoher Genauigkeit (95,1%) ausgewertet. Schließlich wurden auch die 30 DAX-Mitglieder im Web gecrawlt und die genannten Entitäten entsprechend den Berufsfeldern und Kompetenzen gespeichert. Da sowohl einzelne Wörter als auch Ausdrücke mit mehreren Worten gelabelt waren, erschienen einige „Stoppwörter“ wie "und", "plus" und "auf". Diese Stoppwörter wurden entfernt, um nur die sinnvollen Entitäten zu erhalten.

### PROJEKTERGEBNIS

Die erzielten Ergebnisse werden in einem benutzerfreundlichen Dashboard dargestellt, so dass der Benutzer je nach Bedarf ein Unternehmen, Kompetenzen oder Berufsfelder betrachten kann. Die Filterung nach Unternehmen liefert ein Diagramm, in dem die Größe der Rechtecke der Anzahl der offenen Positionen im Unternehmen entspricht.

### GESUCHTE KOMPETENZEN BEI ADIDAS



### LABORERFAHRUNG

Durch die Auswahl einer bestimmten Qualifikation oder Berufserfahrung wird der Prozentsatz von offenen Positionen im jeweiligen Unternehmen angezeigt. Weitere Analysen, wie z.B. kombinierte Kompetenzen, können ebenfalls durchgeführt werden.

Der erstellte Algorithmus kann erweitert werden, um die große Anzahl von Bewerbungen zu scannen, die ein Unternehmen erhält.

Die Ergebnisse stellen eine einfache Lösung dar, um Ressourcen für Marktforschungsaktivitäten zu sparen.

