

Big Data Praktikum: Attribute extraction from eCommerce product descriptions

Gruppe: Gregor Pfänder, Thilo Brummerloh

23. April 2021

1 Thema

Auf twitter werden täglich unübersichtliche Mengen an Texten zu unterschiedlichen Themen geschrieben. Um einen Überblick zu erhalten kann eine Sentiment Analyse durchgeführt werden, die diese Datenmengen auf darstellbare Zahlen zusammenfasst.

2 Fragestellung

Die durchzuführende Sentiment Analyse soll das sentiment der Tweets aus dem betrachteten Datensatz herausfinden. Das sentiment soll über den betrachteten Zeitraum anschaulich dargestellt werden und Auffälligkeiten sollen näher betrachtet werden.

3 Daten

Als Datenbasis sollen tweets, die über einen Zeitraum mehrerer Monate gesammelt wurden und in Verbindung mit dem Thema des COVID-19 stehen, verwendet werden. Die Datenbasis von **erdal'baran** wird herangezogen. Der Liste¹ der Autoren entsprechend, wurden darin tweets aus dem Zeitraum Oktober 2019 bis April 2020 abgelegt.

4 Methoden

4.1 Trainieren des Modells

¹<https://data.gesis.org/tweetscov19/keywords.txt>