

DH Projekt

Gruppe: Thilo Brummerloh

19. Januar 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 Baran und Dimitrov 2020 wird herangezogen. Der Liste¹ der Autoren entsprechend, wurden darin tweets aus dem Zeitraum Oktober 2019 bis April 2020 abgelegt.

4 Methoden

4.1 Vorbereitung der Daten

Die im Datensatz enthaltenen Tweets müssen zum sentiment scoring aufbereitet werden. Mit folgenden Methoden werden tweets zu schnell bewertbaren Wortfolgen umgeformt. (Letter Casing Tokenizing Noise removal Stopword removal normalization stemming lemmatization) Zur Aufbereitung wird die Pythonbibliothek spaCy verwendet.

Die Wortlisten der einzelnen tweets können automatisch mit Python zu einer Zahl zusammengefasst werden, die das sentiment des tweets widerspiegelt. Mit Vergleich

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

der Wörter aus den tweets mit einer Liste wie AFINN-en-165, die Valenzbewertungen von Wörtern enthält, kann das sentiment einzelner tweets bestimmt werden.

Die sentiments können über einen Zeitraum in einem Graphen dargestellt werden. Auffälligkeiten müssen gefunden werden und sollen mit einer Bibliothek wie ggplot dargestellt werden.

Literaturverzeichnis

Baran, Erdal und Dimitar Dimitrov (Juni 2020). *TweetsCOV19 - A Semantically Annotated Corpus of Tweets About the COVID-19 Pandemic*. Zenodo. DOI: 10.5281/zenodo.3871753. URL: <https://doi.org/10.5281/zenodo.3871753>.