

Deutsche Fake News Datensätze:

(1) Kaggle - Fake News Dataset German

<https://www.kaggle.com/datasets/astoeckl/fake-news-dataset-german?resource=download>

Darstellung: Fake oder kein Fake

Statistiken zum Datensatz:

Gesamtanzahl der Zeilen: 63868

Anzahl der 'Fake'-Zeilen: 4627

Anzahl der 'Kein Fake'-Zeilen: 59241

Anteil 'Fake' in Prozent: 7% -> nicht gut

-> enthält zu wenig Fake Artikel

(2) Zenodo - German Fake News Dataset "GermanFakeNC"

<https://zenodo.org/records/3375714>

Darstellung: Wie viele Fakes erhalten die Artikel. Angaben nur als Referenzen (Link + Zeilenangabe)

Statistiken zum Datensatz: 489

Anzahl ‚Fake‘ in Prozent: 100%

-> Datensatz enthält nicht die ‚Fake‘-Texte, was das Auswerten sehr schwierig macht

(3) FANG-COVID: A New Large-Scale Benchmark Dataset for Fake News Detection in German

https://github.com/FANG-COVID/FANG-COVID_DATASET

Darstellung: enthält Titel, header, Daten und fake/kein Fake label

Statistiken zum kombinierten Datensatz:

Gesamte Anzahl Zeilen: 41242

Anzahl 'fake'-Zeilen: 13186

Anzahl 'real'-Zeilen: 28056

Anteil 'fake' in Prozent: 31.97%

-> eignet sich sehr gut!

(4) DeFaktS: A German Dataset for Fine-Grained Disinformation Detection through Social Media Framing

-> nicht öffentlich