

Stimmungsanalyse mit Twitter

Anne Huber, Andreas Franke, Felix Lindner, Burak Özkan, Milomir Soknic

Projektpraktikum Web Science,
Artificial Intelligence Group,
University of Hagen, Germany

17. Dezember 2024



- ▶ Stimmungsanalyse mit Twitter
- ▶ Besonderheiten der Tweets
- ▶ Relevanz zu Web Science

Zielsetzung

- ▶ Vergleich unterschiedlicher ML-Ansätze

1 Daten

2 Methodik

3 Ergebnisse

4 Zusammenfassung

- ▶ Prüfung diverser Datensätze
- ▶ Entscheidung für „Sentiment140“
- ▶ Besonderheiten:
 - ▶ Emojis als Sentiment-Indikatoren
 - ▶ Ausbalancierte Klassen
 - ▶ Bessere Datenqualität
 - ▶ Artikel: „Twitter Sentiment Classification using Distant Supervision“

- ▶ Bereinigung
- ▶ Tokenisierung
- ▶ Transformation
- ▶ Merkmalsextraktion

1 Daten

2 Methodik

3 Ergebnisse

4 Zusammenfassung

- ▶ Logistische Regression
- ▶ *Support Vector Machine*
- ▶ Entscheidungsbäume
- ▶ Naiver Bayes Klassifikator
- ▶ K-nächste Nachbarn

- ▶ „RoBERTa“
- ▶ vortrainiert
- ▶ BERT-basiert
- ▶ bildet die Referenz

1 Daten

2 Methodik

3 Ergebnisse

4 Zusammenfassung

Algorithmus	Genauigkeit
K-nächste Nachbarn	0.68
Entscheidungsbäume	0.78
Logistische Regression	0.79
Naiver Bayes Klassifikator	0.83
<i>Support Vector Machine</i>	0.84
Twitter-RoBERTa-BaseSentiment	0.87

1 Daten

2 Methodik

3 Ergebnisse

4 Zusammenfassung

- ▶ Datensatzauswahl und Bereinigung
- ▶ Training klassischer ML-Verfahren
- ▶ Tests zeigen bessere Ergebnisse mit *Deep Learning* Ansatz
- ▶ Ausblick:
 - ▶ Optimierung der Modelle
 - ▶ *Deep Learning* Ansatz fokussieren

- ▶ Datensatzauswahl und Bereinigung
- ▶ Training klassischer ML-Verfahren
- ▶ Tests zeigen bessere Ergebnisse mit *Deep Learning* Ansatz
- ▶ Ausblick:
 - ▶ Optimierung der Modelle
 - ▶ *Deep Learning* Ansatz fokussieren

Vielen Dank für Ihre Aufmerksamkeit! *Offene Fragen?*