Determination of political views

using NLP on social media postings

Lennart Hallenberger, Florian Siepe

Department Mathematics and Computer Science Philipps University of Marburg

July 17, 2022



Outline



- 1. Goal (revisited)
- 2. Overview of the dataset
- 3. Benchmarks
- 4. Summary & Outlook

Goal (revisited)

Goal (revisited)



Main objective

Classify social media postings to match them with a political party.

Approach

- Use posts from known politicians with their political party as label
- Data cleaning and pre-processing
- Use LDA to extract topics from postings (Topic Modeling)
- Train classifier, predict and evaluate

Overview of the dataset

Overview of the dataset



Democrat vs Republican Tweets

- 84502 Tweets
- 433 Accounts
- 51% Republican
- 49% Democrat
- 200 last tweets from known members of Democratic and Republican party
- Collected from May 2018

ML-Pipeline

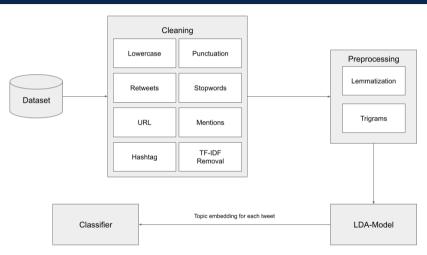


Figure: ML-Pipeline

Benchmarks

Perplexity

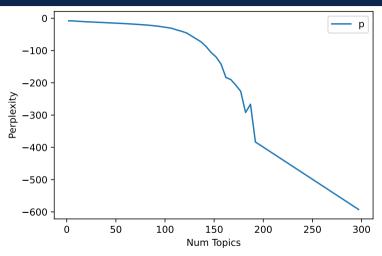
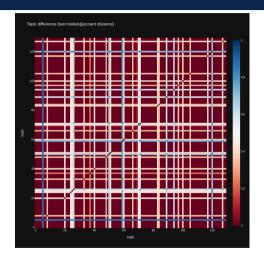


Figure: Perpexlity over # topics

Distinction matrix



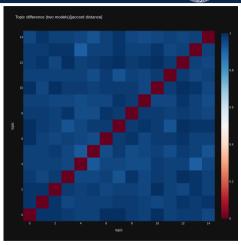


Figure: 130 Topics Figure: 15 Topics

Precision & Recall



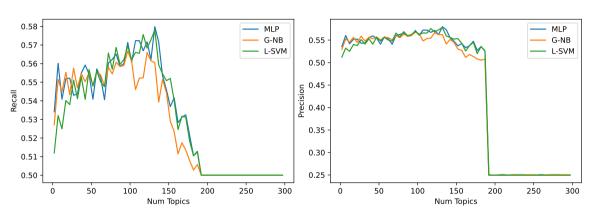


Figure: Recall over # topics

Figure: Precision over # topics

Confusion & F₁-Score



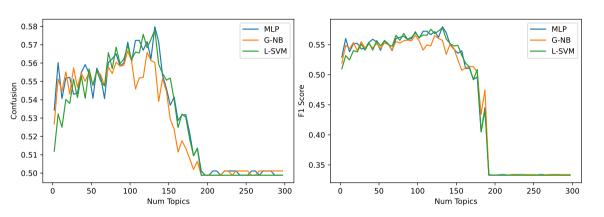


Figure: Confusion over # topics

Figure: F₁-Score over # topics

LVM vs MLP vs G-NB (1)



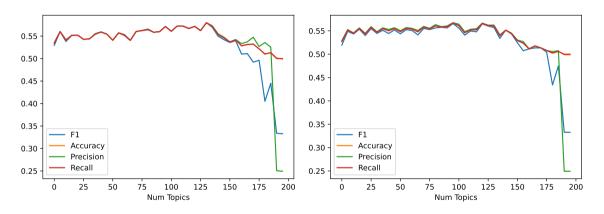
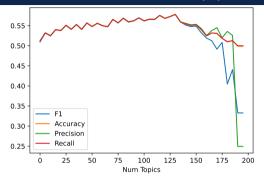


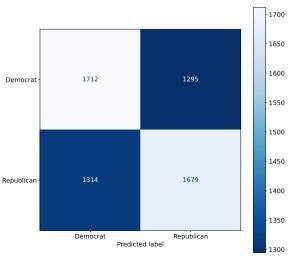
Figure: MLP performance over # topics

Figure: G-NB performance over # topics

LVM vs MLP vs G-NB (2)







Democrats vs Republicans





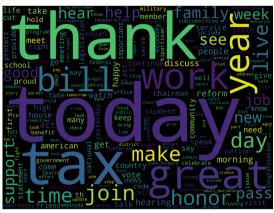
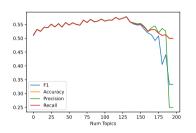


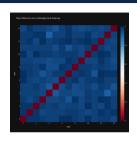
Figure: Wordcloud Democrat

Figure: Wordcloud Republican

Summary & Outlook

Summary & Outlook





Outlook

- Refine pre-processing (modify TF-IDF or manually remove overlapping words)
- Adjust parameters of the LDA model and LVM classifier

Result

• First results do not look promising for reliable classification of posts.