German Newspapers
Analyzing Sentiment and Bias

Dogà, S., Gomes, G., Morales, R., Vu, A.H.

This project, "German Newspapers: Analyzing Sentiment and Bias," examines how **Bild**, one of Germany's leading newspapers, manages **sentiment** and **bias** in its reporting. Using **spaCy.io**, we analyze content to uncover editorial strategies and trends. The goal is to clarify Bild's stance and its influence on readers' perceptions.

DOES BILD PRODUCE BIASED NEWS?

Main Findings:

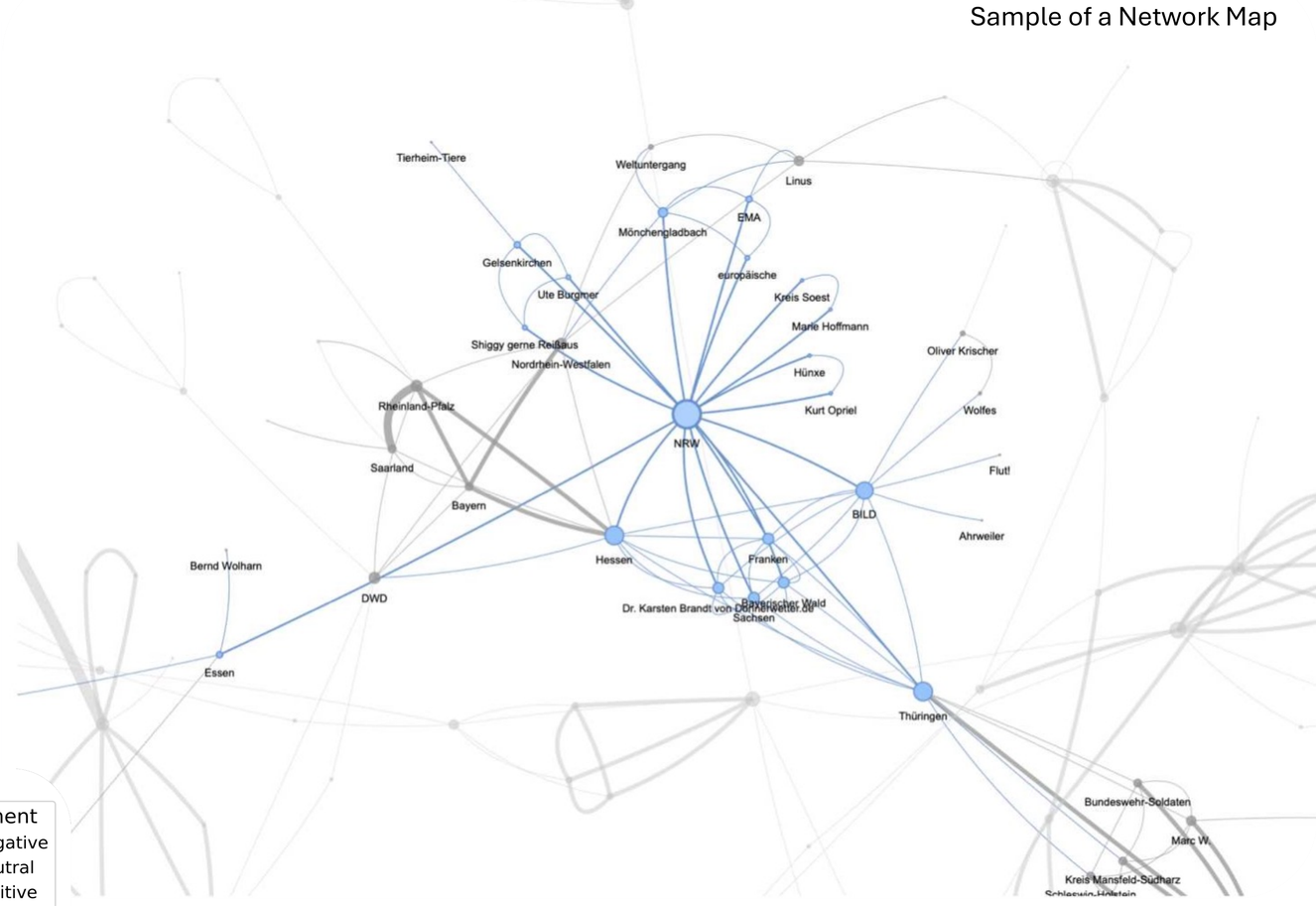
Popular Topics and Prominent News:
The analysis highlights popular topics like politics and key entities such as "Scholz" (PER) and "Bundestag" (ORG), reflecting public and media interest.

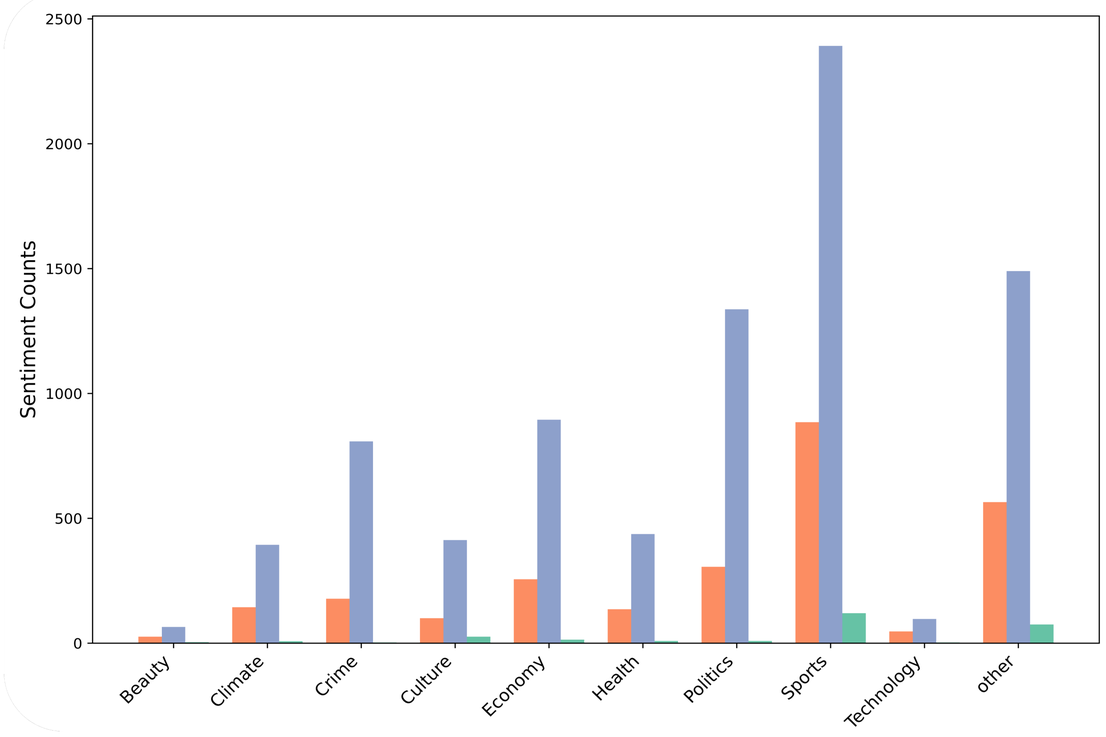
Dominance of Neutral Sentiment:
A substantial 74.1% of articles adopt a neutral tone, indicating a focus on balanced reporting.

Engaging Readers:
The neutral tone may serve as a strategy to encourage reader interaction and allow personal interpretation.

Recreating Events:
Tracking topic relevance and connections over time helps analyze their broader impact across contexts.

Sample of a Network Map





PROJECT TIMELINE

1. Data Collection
Gathered 47,784 articles; retained 11,242 valid ones.

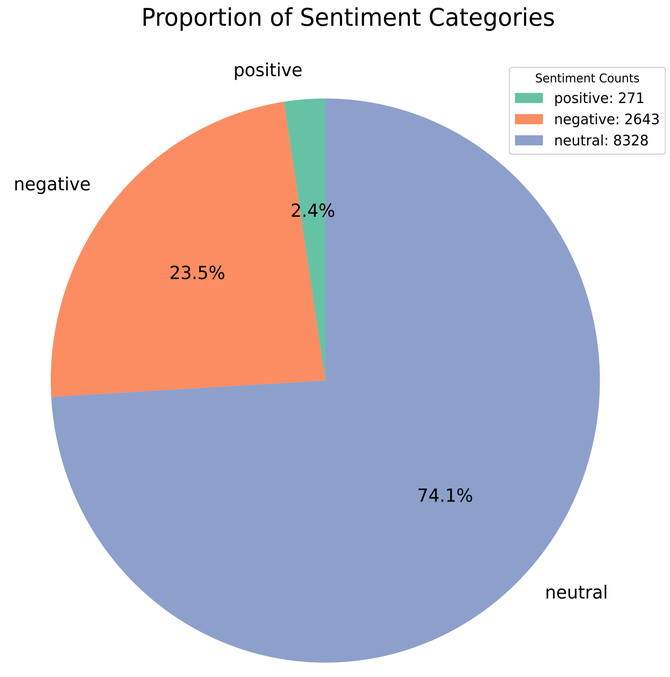
2. Data Cleaning
Removed duplicates, invalid entries, and processed texts.

3. Sentiment Analysis
Used BERT-based model to categorize articles (positive, neutral, negative).

4. Entity Extraction
Identified entities: people, locations, organizations.

5. Visualization
Created bar charts, pie charts, and stacked bar plots.

6. Conclusion
Summarized findings and drafted recommendations.



Methods Used:

- **Data Collection Method:** Utilized custom scraper functions to collect news article URLs from German newspapers, storing them in an SQLite database.
- **Data Extraction Process:** Extracted detailed information from collected URLs, including titles, authors, publication dates, main text, and source domains, using the NewsPlease library and multithreading.
- **Data Storage and Format:** Stored extracted data in a CSV file for further analysis, ensuring a clean and structured dataset.
- **Entity Extraction Method:** Utilized spacy's German model to extract entities from newspaper articles.
- **Entity Extraction Purpose:** Used extracted entities to create network visualizing relationships between places, people, and more.
- **Data Visualization Tools:** Employed matplotlib, networkx, and pyvis for generating pie charts, bar charts, stacked bar plots, and graphs.

Conclusion:

Data analysis of news articles from the first half of 2024 showed most of the neutral sentiment, indicating a **balanced** and **reliable** information delivery. However, a potential bias towards neutral or factual reporting was observed due to the low percentage of articles with positive sentiment.

Learn More:

HAW HAMBURG