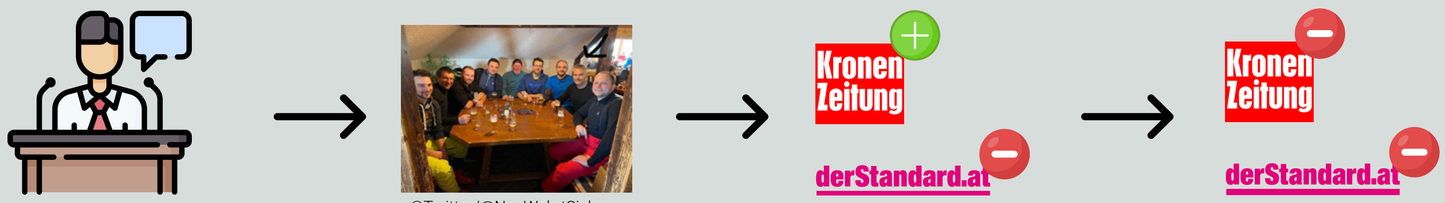
Dynamic Sentiment Analysis for Measuring Media Bias

Thomas Elmar Kolb
Masterstudium Business Informatics

TU Wien Informatics
Institute of Information Systems Engineering
Research Unit: E-Commerce
Supervisor: Mag.rer.nat. Dr.techn. Julia Neidhardt
Contact: thomas.kolb@tuwien.ac.at

What Is This All About?



Relevant questions:
- Media bias towards certain politicians?
- Consistent attitude towards politicians of the same party?
- Sentiment change over time and its relation to external events?

Aim of This Work

Analyzing the sentiments of social news and news media is a highly dynamic research area in which various challenges arise. This master thesis aims to unveil the sentiments towards persons of public interests, who are often presented in polarized contexts, for different media and over time with a focus on Vienna.

Research Questions

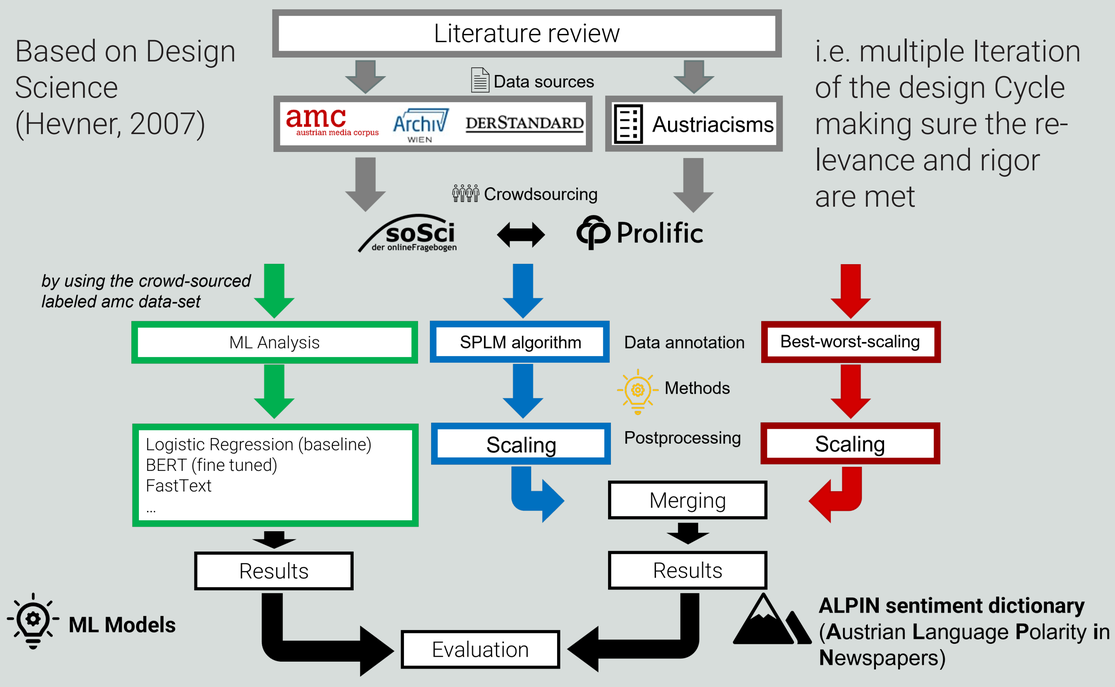
RQ1: To what extent is it possible to predict the polarization of politicians over time in different media?

RQ2: How well do different approaches of machine learning perform in predicting the polarization of politicians in the context of sentiment analysis in the Austrian news media?

Methodology

Design Cycle

Based on Design Science (Hevner, 2007)



i.e. multiple Iteration of the design Cycle making sure the relevance and rigor are met

Alan Hevner. A Three Cycle View of Design Science Research. Scandinavian Journal of Information Systems, 19, January 2007

Rigor Cycle

Literature Review
- State-of-the-Art (Machine Learning (ML) & Lexicon Based Approaches)
- Political polarization
- Media bias
-Austrian German & News Media

Contributions
- New language resource „ALPIN“
- Different ML models developed especially BERT fine tuned models for this domain
- Novel approach to integrate crowd sourcing into Best-Worst Scaling

Relevance Cycle

- Austrian German not well researched
- Analysis of news media not directly comparable with e.g. tweets (Twitter)
- Dynamic setting (i.e., change of polarization of politicians over time) typically not considered

Results

Model	Accuracy	Precision	Recall	F1
DummyClassifier v1	0,52	0,54	0,51	0,52
DummyClassifier v2	0,53	0,53	1	0,69
DummyClassifier v3	0,52	0,54	0,57	0,56
PassiveAggresiveClassifier	0,63	0,65	0,66	0,65
RidgeClassifierCV	0,62	0,63	0,76	0,68
LogistivRegressionCV	0,61	0,61	0,72	0,66
SGDClassifier	0,60	0,62	0,63	0,62
SVC	0,62	0,64	0,66	0,65
MLPClassifier	0,61	0,62	0,65	0,64
FastText	-	0,68	0,67	0,67
BERT (distilbert-base-german-cased) Finetuned with the AMC data-set	**0,80**	0,81	**0,80**	**0,81**
BERT (dbmdz/bert-base-german-cased) Finetuned with the AMC data-set	0,78	**0,82**	0,76	0,79
ALPIN (dictionary based approach)	0.70	0.74	0.70	0.72

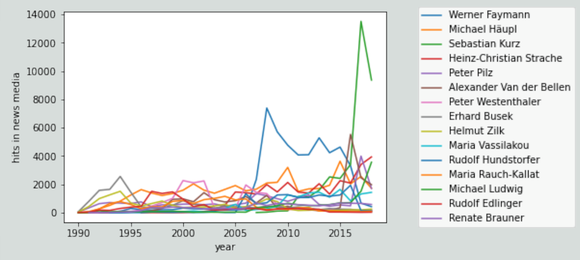
BERT is able to outperform all applied methods. Nevertheless transparancy and explainability is very important.

Therefore the created sentiment dictionary ALPIN is used in the web application.

Discussion

Distribution of Politicians

The distribution follows a power law whereby seven out of the top 15 politicians belong to the “Social Democratic Party of Austria (SPÖ)” (hit = occurence in a certain paragraph)

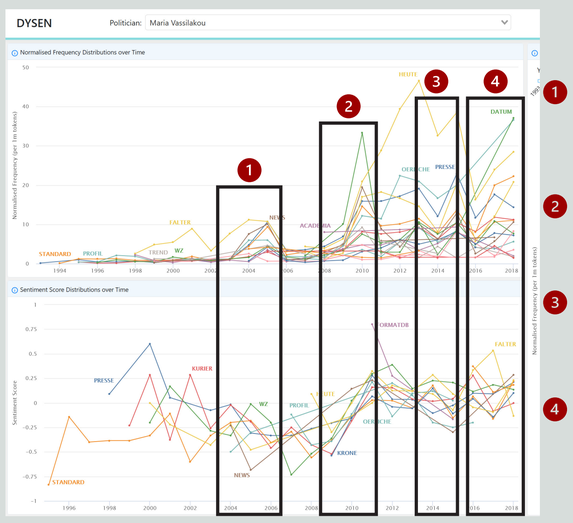


Ethical Questions

By building machine learning based artifacts it is important to raise the following questions:
- Can we trust a system that makes predictions based on a classifier?
- How is e.g. “Bias” defined and which data were used to train the classifier?

This thesis describes the entire process from the beginning to the final cassifier. By developing and applying a sentiment dictionary, this work allows a critical look at what constraints are present and what assumptions were met.

Web Application



Politician: Maria Vassilakou

1 **2004** elected to the federal executive committee of the Green Party
2005 top candidate of the Greens for the municipal elections

2 **2010** top candidate in the state parliament and municipal council election - Elected as vice mayor

3 **2015** state parliament and municipal council election – Controversy surrounding her declaration to resign if Green Party loses vote share

4 **2017** controversial high-rise project at the Heumarkt in Vienna; UNESCO sets the City of Vienna onto the Red List of World Heritage in Danger
2018 Announcement that she will not run in the next state parliament and municipal council election

DYSEN Website based on the models developed by this thesis. Link: https://dysen-tool.acdh-dev.oeaw.ac.at/

Conclusion

ad. RQ1: This question is addressed qualitatively by the web application and quantitatively by the evaluation of the used models and algorithms. The results confirm that tendencies and dynamics can be captured well.

ad. RQ2: State-of-the-art ML models perform well on this task and show better result as traditional dictionary-based approaches. However, with our created language resource good results are achieved and explainability is enhanced.

DYSEN Project

This master thesis is part of the “Dynamische Sentimentanalyse als emotionaler Kompass für die digitale Medienlandschaft” (DYSEN) project founded within the Digital Humanism Call.

funded by:
Stadt Wien
Grant number: MA7-737909/19

Publications

Kolb, Thomas Elmar, Sekanina, Katharina, Kern, Bettina M. J., Neidhardt, Julia, Wissik, Tanja, & Baumann, Andreas. (2022). The ALPIN Sentiment Dictionary: Austrian Language Polarity in Newspapers (1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.5857151

Kern, B. M., Baumann, A., Kolb, T. E., Sekanina, K., Hofmann, K., Wissik, T., & Neidhardt, J. (2021a). A Review and Cluster Analysis of German Polarity Resources for Sentiment Analysis. In 3rd Conference on Language, Data and Knowledge (LDK 2021). Schloss Dagstuhl-Leibniz-Zentrum für Informatik. Shortlisted for Best Paper
https://doi.org/10.4230/OASIcs.LDK.2021.37

Submitted to LREC 2022 (under review):
Kolb, T. E., Kern, B. M., Sekanina, K., Wissik, T., Neidhardt, J., Baumann, A., (2022)
The ALPIN Sentiment Dictionary: Austrian Language Polarity in Newspaper