

Representation of Political Topics in Search Engines

A Case Study for the 2024 German State Election in Saxony



Mazlum Demirpence

University of Duisburg-Essen, Department of Data Science

mazlum.demirpence@uni-due.de

UNIVERSITÄT
DUISBURG
ESSEN

Offen im Denken

Introduction

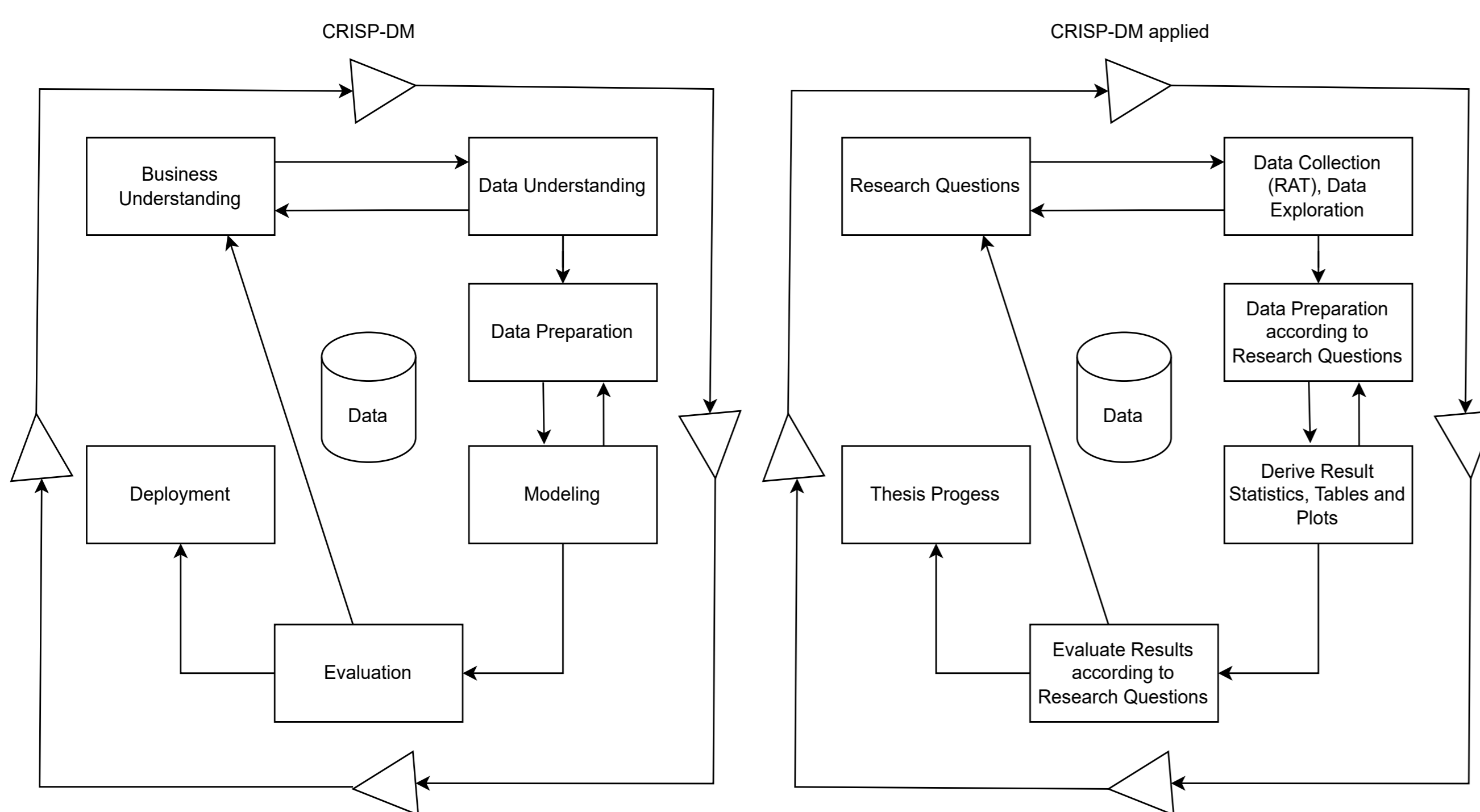
Search engines are gatekeepers of political information (Pradel, 2021), and users generally trust the results of search engines (Pan et al., 2007). The ranking of political search results can significantly influence voter behavior (Epstein & Robertson, 2015). In the 2021 German federal elections, almost all candidates, including lesser-known ones, used SEO to enhance their online visibility (Hinz et al., 2022). The representation of politicians may shift before and after elections (Pradel, 2021), and inequalities in news coverage exist (Unkel & Haim, 2021). Dominance in SERPs is often held by Wikipedia and major news sites (Yagci et al., 2022). This study investigates whether the change of representation of political topics and politicians also apply to smaller elections, such as the German state elections in Saxony two weeks before the election up to eight days after the election

Research Questions

RQs always consider Google, Bing and DuckDuckGo two weeks before the election to eight days after the election compare the results of the election period.

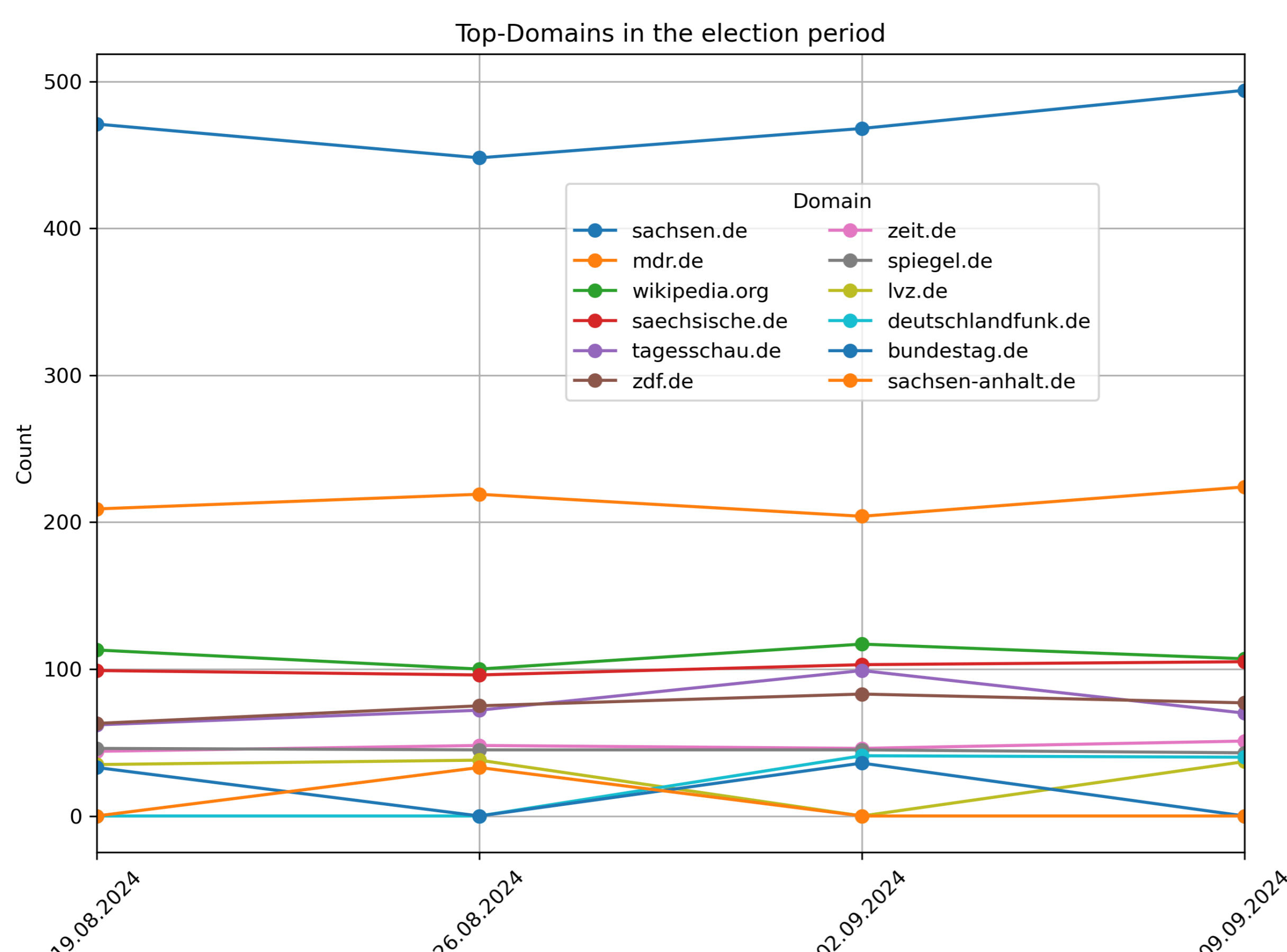
1. "Which domains and topics dominated the search results?"
2. "Which parties and politicians were most frequently represented, and how significantly did the use of SEO influence the ranking of this content?"
3. "What SEO factors were particularly influential in the positioning of political content in the search results?"
4. "Is there any evidence of a political bias regarding the sentiment towards parties or politicians in the search results?"

Methods

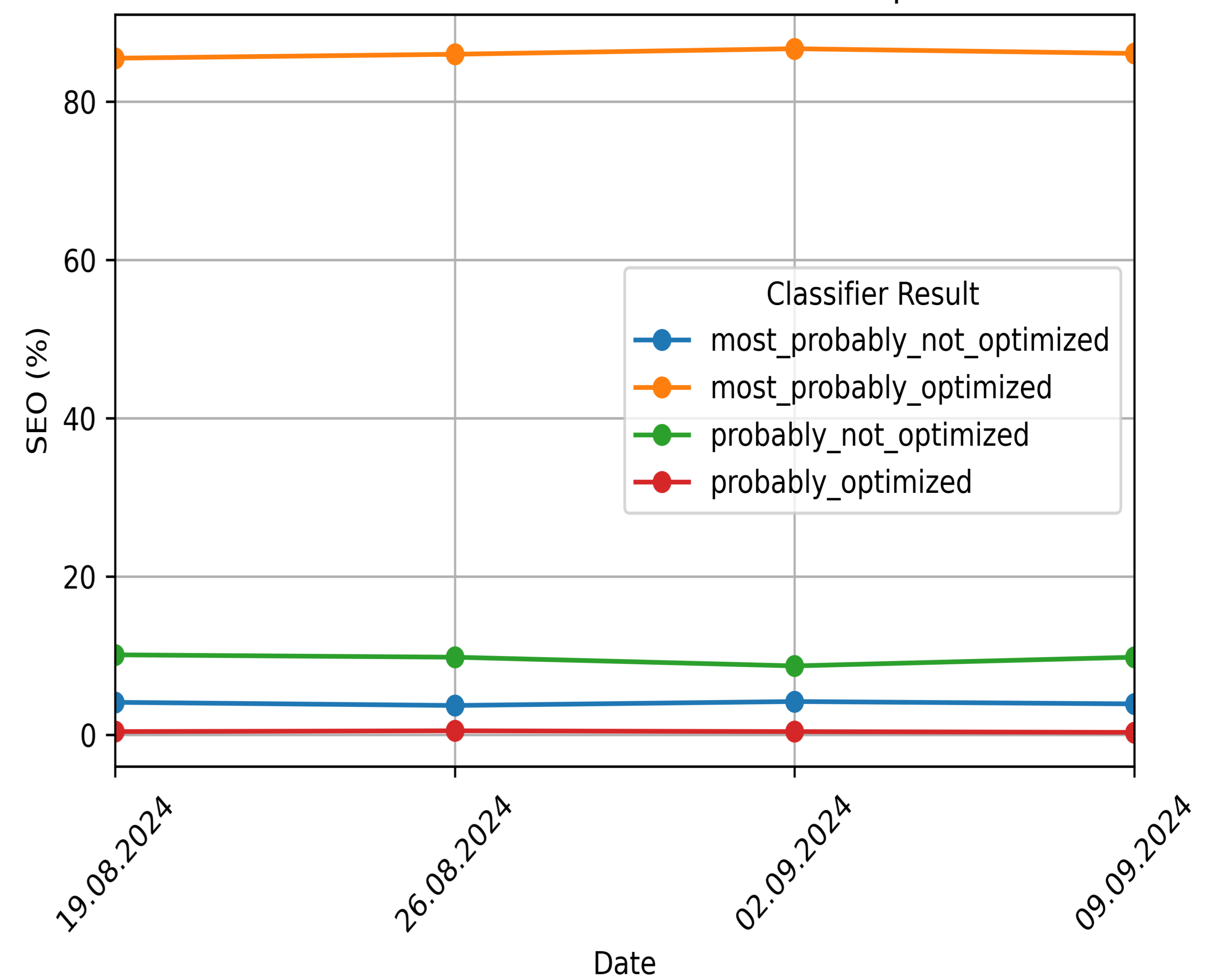


- CRISP-DM for structured work with data, still de facto standard for data science projects (Martínez-Plumed et al., 2021).
- Generate political queries from Google Trends and Wahl-O-Mat.
- Collect SERP data from Google, Bing, and DuckDuckGo.
- RQ1: Count domain occurrences and calculate the Jaccard Similarity Index for domain comparisons across search engines.
- RQ1: Use n-grams to filter topics in titles/descriptions and compare them to topics from Wahl-O-Mat and Google Trends.
- RQ2: Count occurrences of politicians and parties.
- RQ3: Classify domains by SEO usage percentage.
- RQ3: Analyze correlation and perform multiple regression to identify significant SEO factors influencing SERPs.
- RQ4: Conduct sentiment analysis on domain titles/descriptions and use a two-tailed t-test to assess significance.

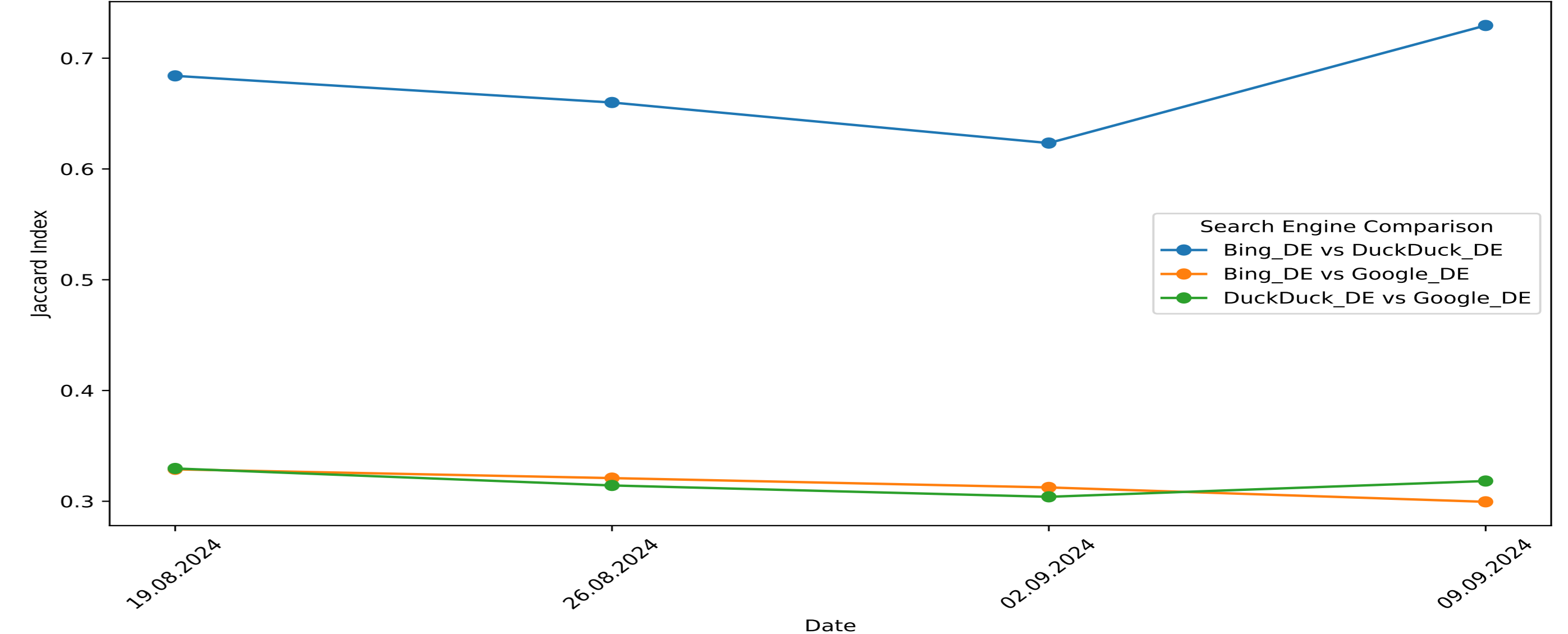
Results



Overall SEO Distribution in the election period



Jaccard Index Across Search Engines for Each Date



Discussion and expected Results

- RQ1: Even in minor elections like Saxony, few domains dominate SERPs.
- RQ2: Conservative political topics are expected to dominate the SERPs.
- RQ3: The current governing coalition is anticipated to dominate the SERPs over other parties.
- RQ4: SEO usage is expected to remain consistently high, reflecting its strong influence on SERPs.
- RQ5: A strong correlation is expected between certain SEO factors, domain rankings, and SEO categorization.
- RQ6: Google sentiment is expected to differ from Bing and DuckDuckGo, potentially reflecting biases toward specific parties or politicians during the election.

References

Epstein, R., & Robertson, R. E. (2015). *The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections*. <https://doi.org/10.1073/pnas.1419828112>

Hinz, K., Sünkler, S., & Lewandowski, D. (2022). SEO im Wahlkampf: Welche Kandidierende durch Suchmaschinenoptimierung ihre Sichtbarkeit zu erhöhen versuchen. In K.-R. Korte, M. Schiffrers, A. Von Schuckmann, & S. Plümer (Eds.), *Die Bundestagswahl 2021* (pp. 1–28). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-35758-0_19-1

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., Ramírez-Quintana, M. J., & Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories [Conference Name: IEEE Transactions on Knowledge and Data Engineering]. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), 3048–3061. <https://doi.org/10.1109/TKDE.2019.2962680>

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication*, 12(3), 801–823. <https://doi.org/10.1111/j.1083-6101.2007.00351.x>

Pradel, F. (2021). Biased Representation of Politicians in Google and Wikipedia Search? The Joint Effect of Party Identity, Gender Identity and Elections. *Political Communication*, 38(4), 447–478. <https://doi.org/10.1080/10584609.2020.1793846>

Unkel, J., & Haim, M. (2021). Googling Politics: Parties, Sources, and Issue Ownerships on Google in the 2017 German Federal Election Campaign. *Social Science Computer Review*, 39(5), 844–861. <https://doi.org/10.1177/0894439319881634>

Yagci, N., Sünkler, S., Häußler, H., & Lewandowski, D. (2022). A Comparison of Source Distribution and Result Overlap in Web Search Engines. *Proceedings of the Association for Information Science and Technology*, 59(1), 346–357. <https://doi.org/10.1002/pra2.758>