



The Journal of *News Science*

Vol. 15, No. 2, Ser.58, summer 2026, P. 35- 39

Journal homepage: <https://www.mjourcom.ir/>

DOI : <http://doi.org/10.22034/lrsi.2026.550018.1443>

Open Access

ORIGINAL ARTICLE

Big Data–Driven Public Opinion Mining: A Conceptual Framework for Evidence-Informed Policymaking

Ghasem Safaeinejad^{1✉} , Seyed Mehdi Sharifi² 

1. Ph.D. in Media Management, Alborz Campus, University of Tehran, Tehran, Iran. Email: safaeinejad@ut.ac.ir
2. Associate Prof., Department of Media Management, College of Management, University of Tehran, Tehran, Iran. Email: Sharifee@ut.ac.ir

Received: September 30, 2025

Accepted: February 23, 2026

EXTENDED ABSTRACT

Introduction:

The rapid expansion of digital ecosystems and social media platforms has fundamentally transformed the production and circulation of public information. Citizens continuously generate vast quantities of textual, behavioral, and networked data through their interactions on social media, messaging applications, news websites, and other digital environments. These data offer unprecedented opportunities for understanding public attitudes, concerns, affective states, and emerging social issues. Concurrently, conventional public opinion polling methods—which rely predominantly on questionnaires, interviews, and field-based sampling—face several inherent limitations, including relatively high operational costs, time-intensive data collection procedures, restricted geographical and temporal coverage, declining response rates, and susceptibility to various response biases.

Against this backdrop, the present study aimed to develop a conceptual framework for national opinion mining grounded in big data analytics, and to examine how data-driven approaches can complement—and, in certain functional domains, progressively substitute for—traditional opinion polling. Particular attention

was devoted to the contextual requirements of applying such a framework within the Iranian setting, including the linguistic idiosyncrasies of Persian-language digital content, constraints in accessing platform data, ethical and privacy considerations, the representativeness of social media data, and the imperative to translate computational outputs into actionable intelligence for policymakers. Accordingly, the study sought to establish an integrated framework that links data acquisition, computational analysis, contextual interpretation, validation, and policy application into a coherent operational model.

Method:

This study was developmental–theoretical in purpose and adopted a qualitative–inferential approach. A systematic qualitative meta-synthesis was employed to integrate findings from prior research and policy documents, and to derive the constituent components of the proposed conceptual framework. The study population comprised peer-reviewed scientific publications and credible policy reports addressing big data, opinion mining, public opinion analysis, social media analytics, and evidence-informed policymaking.

A systematic literature search was conducted across international databases-including Google Scholar, Web of Science, ScienceDirect, and Scopus-as well as Iranian repositories such as Noormags, Magiran, and the Comprehensive Portal of Humanities. The search covered publications from 2010 to 2025. Keywords comprised combinations of "Opinion Mining," "Sentiment Analysis," "Topic Modeling," "Big Data," "Social Media Data," "Social Network Analysis," "Evidence-Informed Policy," "Conceptual Model," and "Framework," along with their Persian equivalents.

The initial search yielded 245 documents. Following the removal of duplicates and screening of titles and abstracts, 98 documents of direct relevance to the research problem were retained. Full-text assessment was subsequently conducted based on criteria of scientific credibility, procedural orientation, and comprehensiveness, culminating in the final selection of 29 documents for in-depth analysis and coding.

Data were analyzed using qualitative content analysis with a combined deductive–inductive strategy. The analytical procedure encompassed open coding, axial coding, and selective coding. Extracted concepts were documented in an evidence matrix to ensure traceability of conceptual inferences. To enhance credibility, portions of the source material were independently coded by two researchers; discrepancies were resolved through discussion and inter-coder consensus. The resulting conceptual model was subsequently reviewed by six experts in data science and media management, and their feedback was incorporated into the final structure.

Findings:

The synthesis of the selected studies yielded a five-layer conceptual framework for big data–driven national opinion mining. The first layer, *data collection and ethical monitoring*, concerns the continuous multi-source acquisition of data from domestic and international social media platforms, messaging applications, news websites, and relevant databases. Diversity of sources is emphasized to mitigate representational bias, while privacy-by-design principles and data anonymization protocols are embedded within the data acquisition process.

The second layer, Persian-language data cleaning and preprocessing, addresses noise reduction and the linguistic complexities inherent in Persian digital content. It encompasses the removal of duplicate and bot-generated content, character normalization, stemming and lemmatization, and the development and ongoing updating of indigenous sentiment lexicons capable of recognizing colloquial expressions, idioms, irony, and culturally specific semantic nuances.

The third layer, analysis and modeling, transforms processed data into structured analytical intelligence. Core techniques include sentiment analysis, analysis of sentiment dynamics, topic modeling, clustering, social network analysis, and anomaly detection. These methods enable the identification of dominant and emerging topics, emotional trajectories, influential actors, sudden shifts in public attention, and potentially coordinated or non-organic communication patterns.

The fourth layer, contextual interpretation and conceptual validation, underscores the role of interdisciplinary expertise in interpreting computational outputs. Social, cultural, generational, and

geographical contexts must be taken into account to identify possible biases and to prevent purely algorithmic interpretations of public opinion. Expert validation serves as an additional mechanism for assessing the conceptual coherence, applicability, and trustworthiness of the model.

The fifth layer, feedback and policymaking, connects analytical findings to decision-making processes. Outputs can be presented through interactive management dashboards, concise policy briefs, and data visualization tools. The framework also incorporates a feedback loop through which the consequences of policy interventions can inform subsequent data collection procedures and analytical models, thereby enabling a continuously adaptive learning system.

Conclusion:

The proposed conceptual framework demonstrates that big data-driven opinion mining can provide a scalable, multidimensional, and near-real-time complement to conventional public opinion polling. By integrating computational techniques with contextual interpretation, expert validation, ethical safeguards, and policy-oriented feedback mechanisms, the framework transcends purely technical approaches to social media analytics and establishes a more comprehensive model for data-informed public decision-making.

Nevertheless, big data-driven opinion mining should not be regarded as an unconditional replacement for conventional surveys. Digital data may underrepresent population segments with limited online participation and may contain demographic, geographical, platform-related, and behavioral biases. Therefore, particularly during early implementation phases, combining big data analytics with traditional survey results can improve validation and facilitate the estimation and correction of potential biases.

Implementation of the proposed model in Iran requires reliable data infrastructures, advanced computational resources, trained interdisciplinary teams, enhanced Persian natural language processing tools, transparent ethical and legal protocols, and institutional mechanisms for translating analytical findings into policy-relevant evidence. Subject to these prerequisites, the proposed framework can provide a foundation for developing indigenous national opinion-mining systems and strengthening Iran's capacity for evidence-informed and data-driven policymaking.

Data Availability Statement.

Data available on request from the authors.

Acknowledgements

The authors would like to thank anonymous reviewers.

Ethical considerations

Not applicable.

Funding

Not applicable.

Conflict of interest

The authors declare no conflict of interest.

References:

- Abdalla, H. B. (2022). A brief survey on big data: Technologies, terminologies and data intensive applications. *Journal of Big Data*, 9, Article 107, 1–36. <https://doi.org/10.1186/s40537-022-00659-3>
- Akbas, E. (2017). Opinion mining on non-English short text. In M. Kryszkiewicz, A. Appice, D. Ślęzak, H. Rybinski, A. Skowron, & Z. Raś (Eds.), *Foundations of Intelligent Systems: ISMIS 2017* (Lecture Notes in Computer Science, Vol. 10352, pp. 1–10). Springer, Cham. https://doi.org/10.1007/978-3-319-60438-1_41
- Aladeemy, A. A., Alzahrani, A., Algarni, M. H., Khalaff, O. I., Wong, W.-K., & Aqburi, S. (2024). Advancements and challenges in Arabic sentiment analysis: A decade of methodologies, applications, and resource development. *Heliyon*, 10(21), e39786. <https://doi.org/10.1016/j.heliyon.2024.e39786>
- Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using topic modeling methods for short-text data: A comparative analysis. *Frontiers in Artificial Intelligence*, 3, Article 42. <https://doi.org/10.3389/frai.2020.00042>
- Atkin, C., & Gaudino, J. (1984). The impact of polling on the mass media. *The ANNALS of the American Academy of Political and Social Science*, 472(1), 119–128. <https://doi.org/10.1177/0002716284472001011>
- Bovet, A., Morone, F., & Makse, H. (2018). Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump. *Scientific Reports*, 8, Article 6789, 1–16. <https://doi.org/10.1038/s41598-018-26951-y>

- Cen, Y., & Wang, Y. (2016). Social public opinion analysis and decision making support with big data. *Data Analysis and Knowledge Discovery*, 32(7), 3–11. <https://doi.org/10.11925/infotech.1003-3513.2016.07.02>
- Cortis, K., & Davis, B. (2021). Over a decade of social opinion mining: A systematic review. *Artificial Intelligence Review*, 54(7), 4873–4965. <https://doi.org/10.1007/s10462-021-10030-2>
- Farrokhi, M. (2004). Public opinion polling in Iran. Majlis Research Center (Iranian Parliament Research Center). (In Persian)
- Farzindar, A., & Louis, A. (2016). Natural language processing for social media [Book review]. *Computational Linguistics*, 42(4), 833–836. https://doi.org/10.1162/COLI_r_00270
- Featherstone, J. D., & Barnett, G. A. (2020). Validating sentiment analysis on opinion mining using self-reported attitude scores. In 2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 1–4). Paris, France. <https://doi.org/10.1109/SNAMS52053.2020.9336540>
- Heidari, E. (2020). A study of big data in social networks. *New Research Quarterly in Humanities, New Series*(26), 109–130. (In Persian)
- Lasswell, H. D. (1948). The structure and function of communication in society. In L. Bryson (Ed.), *The communication of ideas* (pp. 37–51). Harper & Row.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.
- Miller, P. V. (2017). Is there a future for surveys? *Public Opinion Quarterly*, 81(S1), 205–212. <https://doi.org/10.1093/poq/nfx008>
- Patil, A. S. (2016). Opinion mining techniques for non-English languages: An overview. *International Journal of Computational Linguistics (IJCL)*, 7(2), 1–16.
- Rahmanian, P., Mousavi, S. M. R., & Sadraldini, M. H. (2025). Multimodal sentiment analysis in Persian using transformer-based models and sentiment recognition capsules. *Theoretical and Applied Research in Machine Intelligence*, 3(1), 51–67. <https://doi.org/10.22034/abmir.2025.23347.1137> (In Persian)
- Rajabi, Z., Velvi, M. R., & Hourali, M. (2022). A review of sentiment analysis methods in Persian texts. *Signal and Data Processing*, 10(52), 107–132. (In Persian)
- Sandelowski, M., Barroso, J., & Voils, C. I. (2007). Using qualitative metasummary to synthesize qualitative and quantitative descriptive findings. *Research in Nursing & Health*, 30(1), 99–111. <https://doi.org/10.1002/nur.20176>
- Schaefer, T., & Brooker, R. (2005). *Public opinion in the 21st century: Let the people speak?* Houghton Mifflin Harcourt Publishing Company.
- Shang, S., Shi, M., Shang, W., & Hong, Z. (2015). Research on public opinion based on big data. In 2015 IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS) (pp. 559–562). <https://doi.org/10.1109/ICIS.2015.7166655>
- Shi, Z., & Agrawal, R. (2025). A comprehensive survey of contemporary Arabic sentiment analysis: Methods, challenges, and future directions. In *Findings of the Association for Computational Linguistics: NAACL 2025* (pp. 3760–3772). Albuquerque, NM: Association for Computational Linguistics.
- Stahl, B. C. (2025). The ethics of data and its governance: A discourse theoretical approach. *Information*, 16(6), 497. <https://doi.org/10.3390/info16060497>
- Temporão, M. (2019). *Measuring public opinion using big data: Applications in computational social sciences** [Doctoral dissertation, Laval University]. Canada.
- Wang, Y., Guo, J., Yuan, C., & Li, B. (2022). Sentiment analysis of Twitter data. *Applied Sciences*, 12(22), 11775. <https://doi.org/10.3390/app122211775>
- Zahed, A., & Sakhi, M. R. (2016, December 19). A review of big data techniques [Conference presentation]. First National Conference on Electrical and Computer Engineering, Distributed Systems and Smart Networks, Islamic Azad University, Kashan Branch, Iran.
- Zhao, Z., Liu, W., & Wang, K. (2023). Research on sentiment analysis method of opinion mining based on multi-model fusion transfer learning. *Journal of Big Data*, 10, Article 155. <https://doi.org/10.1186/s40537-023-00837-x>

Cite this article: Safaei Nejad, Gh. & Sharifi, S.M. (2026), Big Data–Driven Public Opinion Mining: A Conceptual Framework for Evidence-Informed Policymaking, *News Science*, 15 (2), 35-39. <http://doi.org/10.22034/lrsi.2026.550018.1443>



© The Author(s) retain the copyright.

DOI: <http://doi.org/10.22034/lrsi.2026.550018.1443>

Publisher: Press Institute for Contemporary International Journalism (PICIJ)