

Article

Not peer-reviewed version

Computational Modelling of Reputational Risk in Ugandan Government Megaproject Narratives: A Sentiment-Informed Approach

[Audrey Karungi Ngambeki](#)*, [Halimu Chongomweru](#), [Florence N. Kivunike](#)

Posted Date: 21 January 2026

doi: 10.20944/preprints202601.1620.v1

Keywords: computational model; sentiment analysis; reputational risk; natural language processing; feature-based classification



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Computational Modelling of Reputational Risk in Ugandan Government Megaproject Narratives: A Sentiment-Informed Approach

Audrey Karungi Ngambeki *, Halimu Chongomweru and Florence N. Kivunike

College of Computing and Information Sciences, Makerere University

* Correspondence: audkarungi@gmail.com; Tel.: +256751564747

Abstract

Reputational risk in textual narratives is a vital aspect of understanding stakeholder perceptions of megaprojects; however, formal computational methods for measuring it remain underdeveloped. This study introduces a computational model for sentiment-based reputational risk, defined as a feature-based supervised classification task. The proposed model combines sentiment polarity, polarity intensity, topic distributions, content length, and structural textual features into a structured mathematical model, enabling systematic evaluation and reproducible predictions. Data were collected from online news and social media, then processed through cleaning, tokenisation, lemmatisation, and sentiment annotation. An ensemble model, merging Random Forests, Gradient Boosting, Logistic Regression, and Support Vector Machines via soft voting, was trained and assessed using accuracy, precision, recall, F1-score, and Cohen's Kappa. Results suggest that reputational risk can be reliably inferred from the interaction between sentiment, topics, and textual structure. Analyses of feature importance highlight polarity intensity, risk scoring, content length, and topic distribution as key predictors. These findings demonstrate the potential of formal computational models to quantify and predict risk within textual data. Future research could expand this model with transformer models and multilingual datasets to improve context-aware insights, explainability, and scalability, thereby laying a foundation for generalised computational approaches to reputational risk modelling.

Keywords: computational model; sentiment analysis; reputational risk; natural language processing; feature-based classification

1. Introduction

Megaprojects are large, collaborative ventures aimed at fostering national development [1]. Delphine et al. [2] describe them as projects that exceed typical initiatives in scale, complexity, duration, stakeholder involvement, and impact [3]. According to the mid-term review of Uganda's Third National Development Plan, government megaprojects face significant implementation challenges: only about 10% are on schedule, 83% of externally funded projects are delayed, and nearly 96% of public building projects incur cost overruns [4]. Examples include the Karuma Hydropower Project, the Standard Gauge Railway, the East African Crude Oil Pipeline (EACOP), and the Kampala-Jinja Expressway. These projects pose substantial social risks, particularly for local communities affected by land conversion, displacement, food insecurity, and rising living costs. Such impacts often lead to increased public dissatisfaction and threaten reputational damage for government institutions, as negative perceptions spread via digital platforms. Studies of Uganda's pre-production oil sector indicate that companies utilise corporate social responsibility (CSR) to gain legitimacy and manage their reputation. Neglecting community concerns can damage trust, weaken the social licence to operate, and increase opposition to major projects like the EACOP, risking financing, partnerships, and project continuity [5].

Natural Language Processing (NLP) is used to analyse large text datasets for sentiment, narrative structures, and reputational risks [6–8]. Techniques such as sentiment analysis, topic modelling, and social network analysis transform unstructured text into features suitable for predictive modelling [9–11]. In Uganda, the use of NLP and machine learning is limited by factors like multilingualism, informal communication styles, and a lack of organised datasets [12,13]. Existing models mainly focus on sentiment polarity, often overlooking the complex interactions among tone, topic, structure, and reputational outcomes [14–16].

The presence of over 40 languages, prevalent informal communication, and a limited number of local language datasets pose significant challenges for standard NLP models in Uganda [17]. Advanced NLP techniques have the potential to monitor public discourse, identify emerging risks, and provide early warnings of reputational damage in large-scale projects [18,19]. Despite this potential, research on NLP applications in Uganda remains limited, with most studies concentrating on megaprojects in developed countries or parts of Asia. Initiatives such as Afrisenti highlight the lack of sentiment analysis tools tailored to African languages, slang, and code-mixing [20]. Therefore, adapting NLP tools to the Ugandan context is crucial for the early detection of risks associated with public engagement in megaprojects.

This research introduces a computational model for sentiment-based reputational risk modelling in government megaproject narratives. The model employs an ensemble of machine learning algorithms, including Random Forest, Gradient Boosting, Logistic Regression, and Support Vector Machines, combined through soft voting to classify reputational risk based on sentiment, topical, and structural textual features. This approach offers an interpretable and generalisable early-warning system for monitoring reputational risk in complex sociotechnical systems. Using Ugandan megaprojects as a case study, the findings demonstrate how sentiment measures derived from natural language processing can be transformed into actionable insights for detecting and reducing public opposition [21,22].

1.1. Problem Statement

Government megaprojects in Uganda face significant reputational risks, mainly due to public opposition often worsened by false narratives spread through social and online media. These narratives reduce public trust, causing project delays, stricter regulatory requirements, and increased costs [3,23,24]. For instance, in the Standard Gauge Railway project, public resistance in 2020 led the Ugandan government to allocate around USD 90 million to compensate Project-Affected Persons, as mandated by China Exim Bank before approving the loan [25].

While previous research has employed machine learning and deep learning techniques to analyse public sentiment regarding megaprojects [26–30], the majority of these studies focus mainly on sentiment polarity, offering limited insight into how online discourse affects reputational risk [31]. At the same time, the rapid growth of social media has generated large volumes of sentiment data, which, if systematically examined, could help in early identification of reputational threats [17–19]. Consequently, this study proposes an integrated natural language processing approach that combines sentiment analysis, topic modelling, and social network analysis to measure reputational risk in Ugandan megaprojects.

1.2. Research Objectives

The primary objective of this study is to develop a natural language processing-based model for sentiment analysis and reputational risk identification in Ugandan government megaprojects. Specifically, the study aims to collect and preprocess sentiment data related to these projects, analyse the processed data to extract meaningful patterns, develop a computational model for reputational risk prediction using natural language processing techniques, and evaluate the model's performance within the context of Ugandan megaprojects.

1.3. Research Questions

This study is guided by the overarching research question: How can a natural language processing model be developed to predict reputational risk in Ugandan megaprojects? To address this question, the study investigates the criteria for collecting and preprocessing sentiment data, the analytical methods suitable for the processed data, the procedures for building a reputational risk prediction model, and the approaches for testing and evaluating the proposed model.

2. Materials and Methods

A corpus of 25,480 publicly available English-language textual narratives on megaprojects in Uganda was compiled from online news outlets (including New Vision, Daily Monitor, The Observer, and Chimp Reports), social media platforms (Facebook and X), and international aggregators such as Google News RSS and the GDELT Project API. Data collection was conducted through automated web-scraping and API-based procedures from 1 May 2000 to 11 September 2025, followed by cleaning to remove duplicates, advertisements, incomplete entries, and non-textual content. The dataset mainly consists of online news articles 96.9%, with social media posts accounting for 3.09%. It covers major projects such as EACOP, Karuma Hydropower Dam, Hoima Airport, and the Standard Gauge Railway. A multi-stage sampling strategy was used, combining purposive selection of governance-, infrastructure-, and community-related content with keyword-based filtering using risk-sensitive terms (e.g., protest, corruption, governance, environment, human rights), and subsequent stratified sampling to ensure balanced sentiment and aspect representation. The final structured corpus is suitable for computational analysis of sentiment and reputational risk in Ugandan megaprojects.

A natural language processing (NLP) framework was used to model reputational risk within megaproject narratives. Text data was standardised through a unified preprocessing and anonymisation pipeline, then transformed into numerical representations using both term-weighted features and semantic sentence embeddings. This method made it possible to analyse data in a high-dimensional vector space. Reputational risk was defined using curated risk indicators, targeted labelling, and feature creation, which enabled the model to distinguish reputational risk from other risk types during training and assessment.

Data collection and processing took place in a reproducible Python environment that included automated web scraping, parallel processing, structured text extraction, and robust error handling. Model development followed established machine-learning lifecycle principles, combining sentence-level embeddings with dimensionality reduction, clustering, and topic modelling to identify key narrative themes, especially those related to reputational risk. Named Entity Recognition helped build actor networks, while organisational co-occurrence and centrality measures analysed stakeholder influence patterns. Sentiment analysis enhanced thematic modelling by assessing polarity, intensity, and trust cues, allowing for a comprehensive evaluation of public perception and reputational risks.

Figure 1 below shows the experimental research design used to develop a natural language processing model for sentiment analysis and reputational risk detection in Ugandan government megaprojects, with reputational risk highlighted as a key part of risk assessment. A deductive method guided the systematic variation of main computational elements, including text preprocessing, feature engineering, and model parameters, to assess their effects on accuracy, precision, recall, and overall model performance.

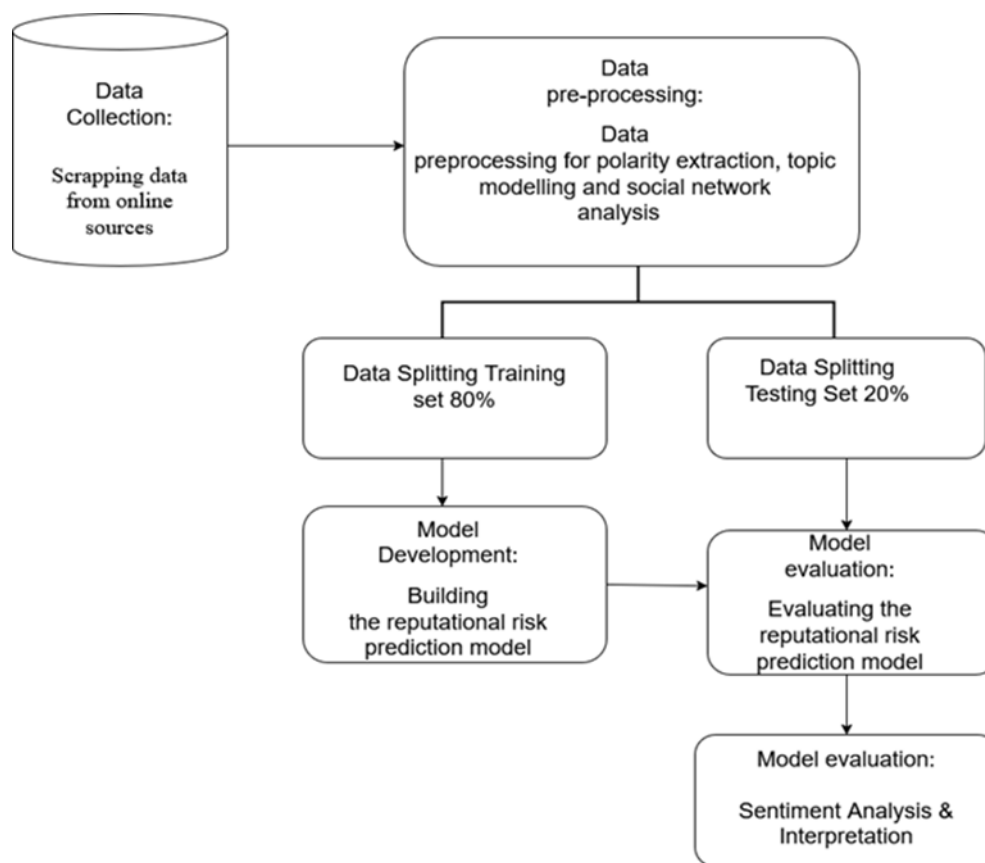


Figure 1. Depicting experimental research design.

2.1. Train-Test Validation Split

The dataset, consisting of 25,480 annotated text records, was split into training (20,384; 80%) and testing (5,096; 20%) subsets. Stratified sampling guaranteed proportional representation of high 19.3%, medium 68%, and low 12.7% risk categories. This method promoted balanced model learning and reliable evaluation.

2.2. Model Development

An ensemble-based natural language processing (NLP) model was created to classify online media narratives about Ugandan megaprojects into categories of low, medium, and high reputational risk. While traditional frameworks focus on technical and financial risks, public sentiment and stakeholder perceptions have become increasingly important [32]. The ensemble combines Random Forest, Gradient Boosting, Logistic Regression, and Support Vector Machine classifiers using a soft-voting mechanism, utilising their complementary strengths to improve prediction accuracy and maintain interpretability [33,34].

Hyperparameters for each base learner were optimised through systematic tuning: Random Forest ($n_estimators=50$, $max_depth=10$), Gradient Boosting ($n_estimators=50$, $learning_rate=0.01$), Logistic Regression ($C=100$, $penalty="l1"$), and Support Vector Machine ($C=1.0$, rbf kernel). The soft-voting approach averaged predicted probabilities, which improved performance on the imbalanced dataset (medium-risk 68%) and produced stable cross-validation results (mean CV= 0.9283 ± 0.0066).

2.3. Reputational Risk Labels

Reputational risk scores were calculated by combining sentiment polarity (using VADER), aspect-category weighting, keyword indicators, and polarity thresholds, resulting in a normalised score between 0 and 1. Scores of 0.7 or higher were classified as high risk, 0.4 to 0.69 as medium risk,

and below 0.4 as low risk. This method captured negative sentiment, sensitive thematic content, and explicit reputational cues.

3. Results

The ensemble model achieved an accuracy of 92.37%, a weighted F1 score of 0.9217, and Cohen's Kappa above 0.80, indicating strong agreement with true labels. Class-specific performance was highest for medium risk (precision 0.92, recall 0.98), while high-risk and low-risk precision were 0.92 and 0.99, respectively. Confusion matrix analysis showed that errors mainly occurred between adjacent categories, with no misclassifications from high to low risk.

The confusion matrix presented in Figure 2 summarises the model's performance across three reputational risk categories. The strong diagonal dominance indicates high classification accuracy. High-risk posts were predominantly identified correctly 845, with remaining errors limited to misclassification as medium risk 138 and none as low risk. Similarly, 478 low-risk instances were correctly classified, with misclassifications occurring only in the medium-risk category 171. The medium-risk category achieved the highest accuracy, with 3,384 correct predictions and minimal confusion with high 77 or low risk 3. Overall, errors were confined to adjacent risk levels, and no confusion occurred between high and low risk, which demonstrates clear class separation. The greater overlap involving medium-risk content is expected due to shared linguistic features; however, the high correct classification rate demonstrates robust detection of nuanced risk signals.

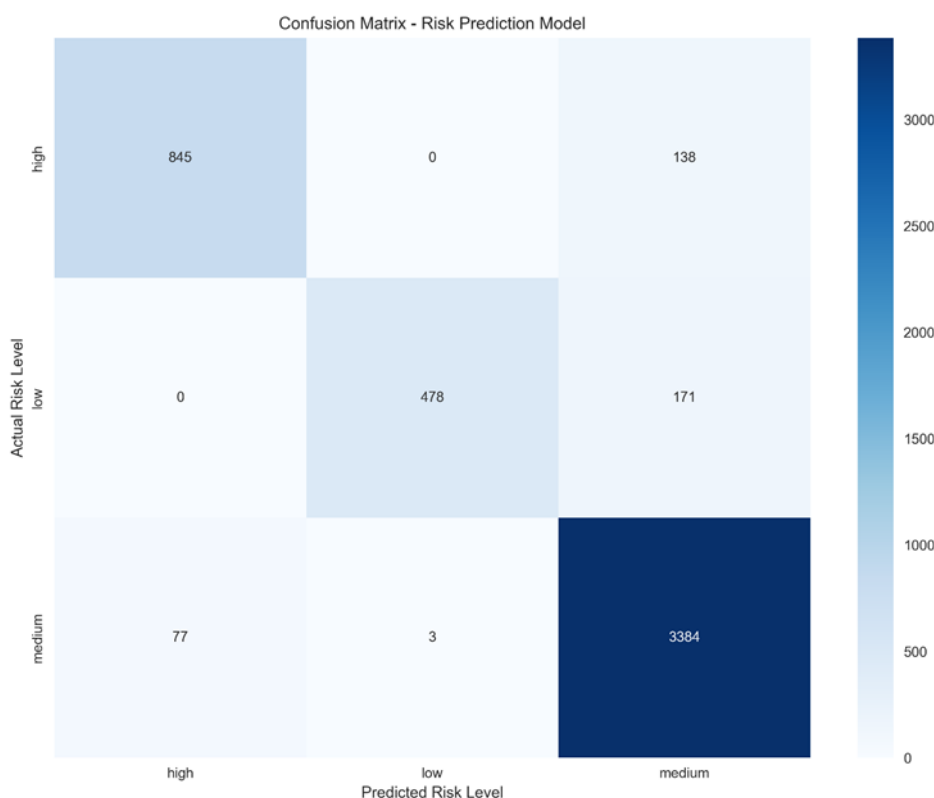


Figure 2. Confusion matrix.

Figure 3 feature importance analysis indicates that risk_score is the most influential predictor, followed by content_length. Several topic-model features, sentiment and polarity measures, and network-based attributes also play a role to varying degrees. These findings collectively imply that reputational risk results from the interaction of composite heuristics, thematic contexts, semantic richness, and emotional cues.

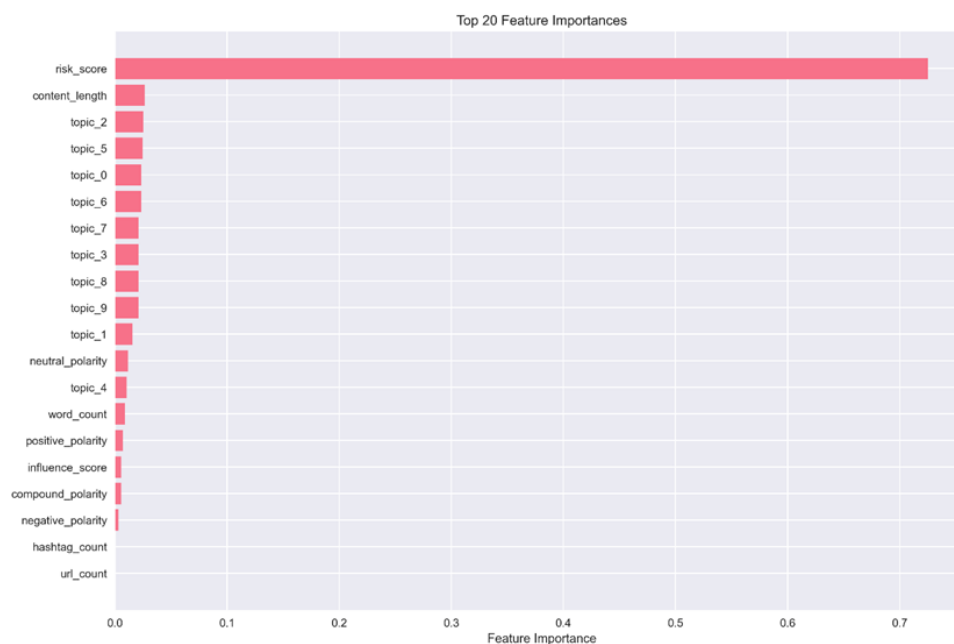


Figure 3. Feature importance analysis.

The heatmap in Figure 4 shows that medium-risk predictions are most prevalent and are associated with neutral and positive posts. High-risk predictions occur across all sentiment polarities, whereas low-risk predictions are primarily associated with positive sentiment. These patterns indicate that although sentiment contributes to reputational risk, the model incorporates a broader spectrum of nuanced linguistic signals.

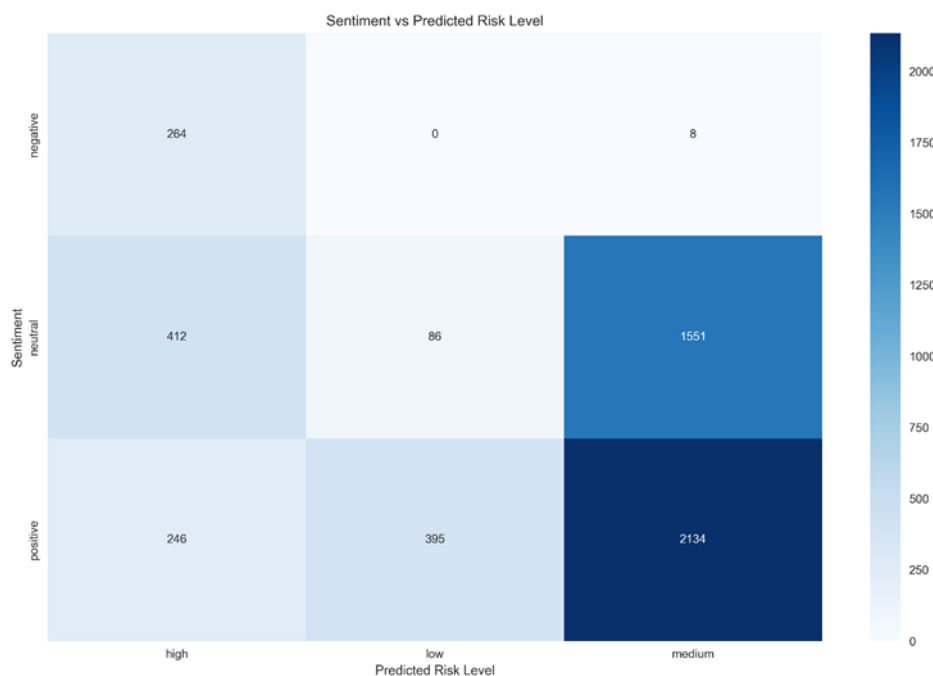


Figure 4. Sentiment Vs Predicted Risk Level Heat Map.

Figure 5 shows that the predicted risk level distribution indicates that medium-risk posts are mainly linked to governance-related content 1,913, with energy 575 and infrastructure 533 also being notable. High-risk predictions are mainly associated with human rights 321, governance 202, and environmental issues 168. Conversely, low-risk posts are mostly classified as “other” 462. These

findings demonstrate the model's ability to recognise governance, energy, infrastructure, and human rights as key factors influencing reputational risk.



Figure 5. Aspect vs Predicted Risk Level Heatmap.

Figure 6 shows that the predicted risk-level distribution includes 3,693 medium-risk, 922 high-risk, and 481 low-risk posts. This spread indicates a dominance of moderately critical discourse. High-risk posts tend to have strongly negative narratives, while low-risk posts are more supportive or neutral. These results match the dataset's features and support the model's contextual accuracy.

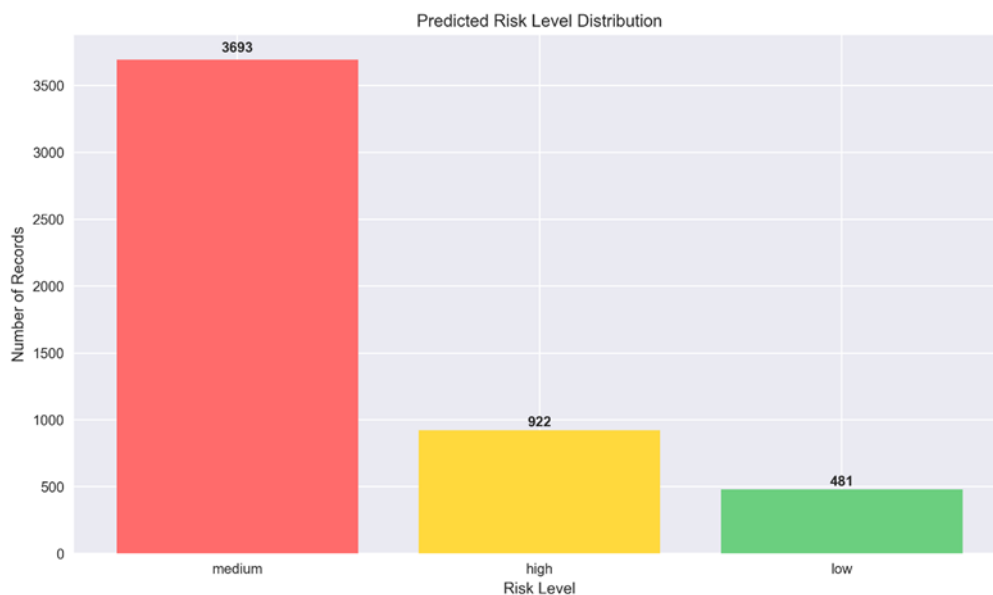


Figure 6. Predicted risk level distribution.

4. Discussion

The ensemble classifier achieved an accuracy of 92.37%, a weighted F1-score of 0.9217, and Cohen's Kappa over 0.80, demonstrating consistent performance across five-fold cross-validation (mean CV = 0.9283 ± 0.0066). The model performed best on medium-risk posts (precision 0.92, recall 0.98) and maintained high precision for both high-risk 0.92 and low-risk 0.99 categories. Feature importance analysis identified risk_score, content length, topic distribution, and polarity intensity as

the main predictors. These findings confirm that reputational risk results from the interaction of tone, context, and thematic content, highlighting the model's usefulness for early-warning monitoring of megaproject-related discourse.

4.1. Comparison with Related Studies

Recent research demonstrates that online sentiment closely mirrors social and institutional events, enabling digital public opinion to both reflect and anticipate broader societal dynamics [35,36], as a result, sentiment analysis has been widely applied to address reputational and governance challenges in infrastructure projects. These applications include reputational risk detection in UK infrastructure [37], analysis of public responses to hydropower development in China [38], and assessment of major transport initiatives such as the Padma Setu in Bangladesh [26], the Thailand-China and Laos-China railways [30], and Kenya's Standard Gauge Railway [39]. This study advances the literature by focusing on reputational risk, public trust, and institutional credibility, rather than solely on financial metrics. The use of a hybrid ensemble model that emphasises interpretability and computational efficiency is vital for real-time public-sector applications in Uganda. Further studies have examined sustainability risks in megaprojects [40] and established sentiment analysis as an effective early-warning mechanism for reputational risk [21].

This study advances the literature by prioritising reputational risk alongside institutional credibility and public trust, rather than focusing exclusively on financial performance. While deep learning models such as BERT [41] often achieve higher accuracy on large English datasets, the proposed hybrid ensemble model is designed to improve interpretability and computational efficiency. These attributes are essential for real-time public-sector applications in Uganda. The model's transparent structure enables stakeholders to identify the linguistic cues that inform risk classification, thereby supporting accountable and evidence-based decision-making in the implementation of megaprojects.

4.2. Implications of Findings

4.2.1. Theoretical Contribution

The integration of natural language processing and aspect-based sentiment analysis provides a strategic tool for monitoring reputational risk in Ugandan megaprojects. This approach also delivers a publicly annotated dataset to support future research on policy communication, public perception, and trust modelling.

4.2.2. Practical Contribution

Ongoing monitoring of citizen discourse enables early identification of reputational risks related to governance, human rights, or environmental issues. This supports timely, evidence-based interventions by ministries and agencies. The interpretability of the model facilitates the identification of linguistic cues and thematic drivers of reputational tension, thereby promoting transparency, accountability, and participatory governance.

4.3. Limitations

1. Prediction Failures: Despite strong overall performance, the model showed several limitations. Misclassifications most often occurred between the medium- and high-risk categories due to overlapping vocabulary and common themes in discussions of governance, infrastructure, and public protest. Some low-risk posts were labelled as medium risk because neutral sentiment in supportive content produced weak polarity signals, leading the model to rely more on context. These inconsistencies reveal the inherent challenge in distinguishing subtle differences in reputational implications.

2. **Data Challenges:** Several dataset-specific challenges caused prediction weaknesses. The frequent use of Ugandan English, local slang, and hybrid dialects limited the effectiveness of sentiment analysis tools trained on standard Western English. Short-form posts, especially those from X, offered minimal linguistic context, reducing the reliability of both polarity measurements and topic modelling. Additionally, LDA topic modelling was unstable with very short or lexically sparse texts, further weakening the semantic features supplied to the model.
3. **Linguistic Complexity:** Complex linguistic behaviors, including sarcasm, humor, euphemisms, and coded political language, posed significant challenges. Sarcastic statements were often assigned positive polarity scores despite conveying criticism, leading to mismatches between sentiment and meaning. Without domain-specific lexicons or context-aware transformer models capable of interpreting these subtleties, these linguistic nuances remained difficult to model computationally.
4. **Model-Level Weaknesses:** Several model-specific limitations also impacted performance. The ensemble design reduced interpretability because aggregated probability outputs made it difficult to trace predictions back to individual classifiers. Additionally, the model analyzed each text independently without considering temporal context, which prevented it from capturing how reputational risk evolves over time. The risk scoring function's reliance on manually defined aspect weights and keyword lists also limited scalability, as performance depended on the completeness and relevance of these engineered features.

4.4. Key Findings and Contributions

The study demonstrates that sentiment-aware ensemble models function as effective early-warning tools for reputational risk in public-sector megaproject communication. These models accurately classify discourse into low, medium, and high-risk categories by integrating sentiment polarity, topical context, and engagement indicators. The interpretable ensemble architecture facilitates ethical and transparent AI use, providing a data-driven foundation for proactive reputation management. This approach enhances transparency, accountability, and stakeholder trust within Uganda's governance and infrastructure sectors.

This research offers a replicable methodology for sentiment-based reputational risk detection, positioning sentiment analytics within governance and development communication contexts. It introduces a validated sentiment-risk taxonomy, an annotated Ugandan dataset, and a customised feature-extraction pipeline to support government agencies, project managers, policy analysts, and AI developers in localised monitoring. The integration of interpretable machine-learning components ensures that decision-making processes remain transparent and accountable.

5. Conclusions

Advancing the model with transformer-based architectures such as BERT, RoBERTa, and XLNet, along with attention visualisation, could improve contextual semantic understanding while maintaining interpretability. Incorporating local languages, including Luganda, Runyankore, and Swahili, as well as continuous learning systems, would support adaptive, inclusive, and scalable monitoring of public sentiment. This progression moves toward a real-time, explainable AI-driven reputational surveillance system suitable for Ugandan megaprojects and similar developing contexts.

Author Contributions: Conceptualization, A.K.N.; methodology, A.K.N. ; graphical discussion and software, A.K.N. ; validation, A.K.N, H.C. and F.N.K.; formal analysis, A.K.N; investigation, A.K.N; data curation, A.K.N.; writing—original draft preparation, A.K.N; review and editing, H.C. and F.N.K.; supervision, H.C. and F.N.K; funding A.K.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yao, R.; Gillen, A. Public Opinion Evaluation on Social Media Platforms: A Case Study of High Speed 2 (HS2) Rail Infrastructure Project. *UCL Open Environ* 2023, 5, doi:10.14324/111.444/ucloe.000063.
2. Delphine; Witte, P.; Spit, T. Megaprojects – An Anatomy of Perception. *disP - The Planning Review* 2019, 55, doi:10.1080/02513625.2019.1630189.
3. Nyarirangwe, M.; Babatunde, O.K. Megaproject Complexity Attributes and Competences: Lessons from IT and Construction Projects. *International Journal of Information Systems and Project Management* 2019, 7, doi:10.12821/ijispm070404.
4. National Planning Authority (NPA) Mid-Term Review of the Third National Development Plan (NDPIII) 2020/21–2024/25 2023.
5. Smith, L.; Tallontire, A.; Van Alstine, J. Anticipatory CSR: Legitimacy Politics in Uganda’s Pre-Production Oil Sector. *Resources Policy* 2025, 110, doi:10.1016/j.resourpol.2025.105756.
6. Joseph, T. Natural Language Processing (NLP) for Sentiment Analysis in Social Media. *International Journal of Computing and Engineering* 2024, 6, doi:10.47941/ijce.2135.
7. Hou, Y.; Huang, J. Natural Language Processing for Social Science Research: A Comprehensive Review. *Chin J Sociol* 2025, 11, doi:10.1177/2057150X241306780.
8. Jelodar, H.; Wang, Y.; Orji, R.; Huang, S. Deep Sentiment Classification and Topic Discovery on Novel Coronavirus or COVID-19 Online Discussions: NLP Using LSTM Recurrent Neural Network Approach. *IEEE J Biomed Health Inform* 2020, 24, doi:10.1109/JBHI.2020.3001216.
9. Alhazmi, H.N. Text Mining in Online Social Networks: A Systematic Review. *IJCSNS International Journal of Computer Science and Network Security* 2022, 22.
10. Laureate, C.D.P.; Buntine, W.; Linger, H. A Systematic Review of the Use of Topic Models for Short Text Social Media Analysis. *Artif Intell Rev* 2023, 56, doi:10.1007/s10462-023-10471-x.
11. Meštrović, A.; Beliga, S.; Pitoski, D. LLMs for Social Network Analysis: Mapping Relationships from Unstructured Survey Response. *Applied Sciences* 2026, 16, doi:10.3390/app16010163.
12. Hu, S.; Oppong, A.; Mogo, E.; Collins, C.; Occhini, G.; Barford, A.; Korhonen, A. Natural Language Processing Technologies for Public Health in Africa: Scoping Review. *J Med Internet Res* 2025, 27.
13. Owen, B. Natural Language Processing (NLP) for Low- Resource Languages. 2023.
14. Islam, M.S.; Kabir, M.N.; Ghani, N.A.; Zamli, K.Z.; Zulkifli, N.S.A.; Rahman, M.M.; Moni, M.A. “Challenges and Future in Deep Learning for Sentiment Analysis: A Comprehensive Review and a Proposed Novel Hybrid Approach.” *Artif Intell Rev* 2024, 57, doi:10.1007/s10462-023-10651-9.
15. Tan, K.L.; Lee, C.P.; Lim, K.M. A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. *Applied Sciences (Switzerland)* 2023, 13.
16. Zhang, W.; Deng, Y.; Liu, B.; Pan, S.J.; Bing, L. Sentiment Analysis in the Era of Large Language Models: A Reality Check. In Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2024 - Findings; 2024.
17. Bo Pang; Lillian Lee Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval* 2008.
18. El Barachi, M.; AlKhatib, M.; Mathew, S.; Oroumchian, F. A Novel Sentiment Analysis Framework for Monitoring the Evolving Public Opinion in Real-Time: Case Study on Climate Change. *J Clean Prod* 2021, 312, doi:10.1016/j.jclepro.2021.127820.
19. Kavitha, C.R.; Rajarajan, S.J.; Jothilakshmi, R.; Alaskar, K.; Ishrat, M.; Devi, V.C. Study of Natural Language Processing for Sentiment Analysis. In Proceedings of the Proceedings - 2023 3rd International Conference on Pervasive Computing and Social Networking, ICPCSN 2023; 2023.

20. Muhammad, S.H.; Abdulmumin, I.; Ayele, A.A.; Ousidhoum, N.; Adelani, D.I.; Yimam, S.M.; Ahmad, I.S.; Beloucif, M.; Mohammad, S.M.; Ruder, S.; et al. AfriSenti: A Twitter Sentiment Analysis Benchmark for African Languages. In Proceedings of the EMNLP 2023 - 2023 Conference on Empirical Methods in Natural Language Processing, Proceedings; 2023.
21. Schröder, D.; Tieben, M. Sentiment Analysis for Reputational Risk Management. In The Digital Journey of Banking and Insurance, Volume II: Digitalization and Machine Learning; 2021.
22. Kamble, S.T.; Mirajka, R.P.; Patel, P.K.A.; Patil, P.R. A Review of Sentiment Analysis Using Machine Learning Model, Bharati Vidyapeeth's College of Engineering, Kolhapur (MH), India, 2023.
23. Eja, K.M.; Ramegowda, M. Government Project Failure in Developing Countries: A Review with Particular Reference to Nigeria. *Global Journal of Social Sciences* 2020, 19, doi:10.4314/gjss.v19i1.4.
24. Kiyita, M.; Kyambadde, M. Uganda's Parish Development Model: Factors to Prioritise to Guarantee Success in Its Implementation, 2023.
25. Ogwang, T.; Vanclay, F. Resource-Financed Infrastructure: Thoughts on Four Chinese-Financed Projects in Uganda. *Sustainability (Switzerland)* 2021, 13, doi:10.3390/su13063259.
26. Mahmud, T. Al; Sultana, S.; Chowdhury, T.I. Analysis of the Opinion of the People of Bangladesh on the Padma Setu Megaproject. *Int J Comput Appl* 2023, 185, doi:10.5120/ijca2023922896.
27. Kochuieva, Z.; Borysova, N.; Melnyk, K.; Huliieva, D. Usage of Sentiment Analysis to Tracking Public Opinion. In Proceedings of the CEUR Workshop Proceedings; 2021; Vol. 2870.
28. Yin, W.; Fan, H. Perception Changes and the Attribution of the Impact of Lancang-Mekong Hydropower Dams in the Media of Riparian Countries from 1971 to 2020. *J Hydrol Reg Stud* 2023, 45, doi:10.1016/j.ejrh.2022.101302.
29. Kundu, O.; James, A.D.; Rigby, J. Public Opinion on Megaprojects over Time: Findings from Four Megaprojects in the UK. *Public Management Review* 2023, 25, doi:10.1080/14719037.2021.2003107.
30. Nokkaew, M.; Nongpong, K.; Yeophantong, T.; Ploykitikoon, P.; Arjharn, W.; Siritaratiwat, A.; Narkglom, S.; Wongsinlatam, W.; Remsungnen, T.; Namvong, A.; et al. Analyzing Online Public Opinion on Thailand-China High-Speed Train and Laos-China Railway Mega-Projects Using Advanced Machine Learning for Sentiment Analysis. *Soc Netw Anal Min* 2024, 14, doi:10.1007/s13278-023-01168-8.
31. Xu, Q.A.; Chang, V.; Jayne, C. A Systematic Review of Social Media-Based Sentiment Analysis: Emerging Trends and Challenges. *Decision Analytics Journal* 2022, 3, doi:10.1016/j.dajour.2022.100073.
32. Wang, J.; Luo, L.; Sa, R.; Zhou, W.; Yu, Z. A Quantitative Analysis of Decision-Making Risk Factors for Mega Infrastructure Projects in China. *Sustainability (Switzerland)* 2023, 15, doi:10.3390/su152115301.
33. Chithuloori, P.; Kim, J.-M. Soft Voting Ensemble Classifier for Liquefaction Prediction Based on SPT Data. *Artif Intell Rev* 2025, 58, 228, doi:10.1007/s10462-025-11230-w.
34. Deng, H.; Zhou, Y.; Wang, L.; Zhang, C. Ensemble Learning for the Early Prediction of Neonatal Jaundice with Genetic Features. *BMC Med Inform Decis Mak* 2021, 21, doi:10.1186/s12911-021-01701-9.
35. Göçen, A.; Mahat, ; Ibrahim, M.; Ul, A.; Khan, I. Public Attitudes toward Higher Education Using Sentiment Analysis and Topic Modeling. *Discover Artificial Intelligence* 2024, 4, 83, doi:10.1007/s44163-024-00195-4.
36. Abebe, B.T.; Weiss, M.; Modess, C.; Roustom, T.; Tadken, T.; Wegner, D.; Schwantes, U.; Neumeister, C.; Schulz, H.; Scheuch, E.; et al. Mindfulness Virtual Community. *Trials* 2019, 17.
37. ZHAO, Z.; XUE, X.; GAO, X.; LIU, G.; WU, H. Coupling and Evolution Mechanism of Infrastructure Mega-Projects Complex Ecosystem: Case Study on Hong Kong-Zhuhai-Macao Bridge. *Frontiers of Engineering Management* 2018, doi:10.15302/j-fem-2018080.
38. Jiang, H.; Qiang, M.; Lin, P. Assessment of Online Public Opinions on Large Infrastructure Projects: A Case Study of the Three Gorges Project in China. *Environ Impact Assess Rev* 2016, 61, doi:10.1016/j.eiar.2016.06.004.
39. Sauer, J.; Overos, H.; Newlands, J.; Morrison, M.; Stewart, K.; Wood, T. Analyzing Multi-Lingual Discussions of the Standard Gauge Railway Project in Kenya Using Natural Language Processing and Social Media Analysis. *J Asian Afr Stud* 2025, 60, doi:10.1177/00219096231173390.
40. Li, Y.; Xiang, P.; You, K.; Guo, J. Dynamic Network Analysis of the Risks of Mega Infrastructure Projects from a Sustainable Development Perspective. 2022, doi:10.3390/buildings12040434.

41. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference; 2019; Vol. 1.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.