Rumour Detection in Social Networks Using e-LM (Enhanced Language Models)

Vijaya Bhaskar Reddy B., Lahari B., Chinmayee Sruthi B., Naga Kavya T. and Durga R.
Department of CSE-AI & ML, Srinivasa Ramanujan Institute of Technology, Rotarypuram Village, B.K. Samudram Mandal, Anantapur District, Andhra Pradesh, India

Keywords: Rumor Detection, Misinformation Tracking, Social Network Analysis, Information Propagation, Data Analytics, False Information, Digital Credibility, Fact-Checking.

Abstract: In today's fast-paced digital world, social networks serve as powerful platforms for information exchange. However, alongside accurate news, misinformation and rumors spread just as rapidly, often causing confusion, damaging reputations, and influencing public perception. The ability to trace the origin of such false claims is critical for mitigating their impact. Our study introduces an innovative approach to identifying the original source of rumors within social networks using advanced data analytics and network analysis. By examining the flow of information and analyzing dissemination patterns, we aim to track misinformation back to its source, providing valuable insights into how rumors evolve and spread. This research is essential for developing effective tools for early detection and containment of false information. By mapping out misinformation pathways, we enable social media platforms, fact-checkers, and policymakers to take proactive measures in curbing its spread. Strengthening the trustworthiness of online information, our findings contribute to building a more reliable digital space. Ultimately, this approach enhances transparency and accountability in digital communication, ensuring that accurate and credible information prevails over misleading content.

1 INTRODUCTION

The widespread prevalence of social network has enabled the fast dissemination of information allowing individuals from all over the world to engage with news, perspectives and conversations instantaneously. But it also means that misinformation and rumor can spread faster than fact-checking methods or verification systems can keep up. Social media, unlike traditional media, allows instant access to content and many times this information was shared before it has even been confirmed. This allows an environment in which false narratives are gain quickly traction and influencing public perception much faster than any corrective measure can be taken.

Misinformation is a serious problem, endangering public knowledge, trust in reliable news sources, physical safety, organizations and governments. Misinformation can occupy multiple domains: politics, public health, finance, security all of which have influenced public opinion and led to real-life consequences. Tell me more in some cases, misinformation can incite panic, wrap opinion or destroy reputations, making it all the more important to combat this emerging challenge. The tricky part is that it's not just a question of tearing holes in false claims; it is also about understanding how falsehoods arise and propagate through social networks.

In particular, social media platforms are dominated by large numbers of users along with real-time communication, which makes it very vulnerable to the rapid spread of rumors. Engagement-boosting algorithms amplify what is popular over accuracy, giving a foothold to misinformation. Add to the fact that social media is designed to be highly viral, with content often being shared indiscriminately, due to people caring less related to verification. This creates an echo chamber effect in which misinformation is repeated by people within certain communities that makes it more difficult to correct false narratives.

In response to this escalating threat, researchers and practitioners are working to trace misinformation sources and analyze its propagation dynamics. Techniques such as advanced data analytics, artificial intelligence and network analysis are being used to

B., V. B. R., B., L., B., C. S., T., N. K. and R., D.
Rumour Detection in Social Networks Using e-LM (Enhanced Language Models).
DOI: 10.5220/0013898500004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 3*, pages 380-385
ISBN: 978-989-758-777-1

identify providers of false claims, as well as track their evolution. So, by tracing the early spreaders of small rumors – in other words, targeting those who are early in the information diffusion process –fact-checkers and policymakers can step in try to stamp out misinformation before it can go viral. Another for me to propose is enhancing digital literacy and responsible sharing/information consumption of users are essential steps in lessening the impact of false information.

Addressing misinformation plan is critical to maintaining the trustworthiness of online discourse. They must take active steps like building better content moderation systems; and including fact-checking tools to reduce the luminosity of falsehoods on their platforms. This collaborative effort between researchers, policymakers and technology companies can help humankind to find a way out of this digital degradation by prompting transparency and accountability in digital degradation by promoting transparency and accountability in digital spaces; ultimately leading to a well-informed onl8ine environment where facts triumph over falsehood.

2 RELATED WORKS

In 2013, F. Peter reported on the financial impact of a false tweet about an explosion at the White House, published by The Telegraph under the Finance/Market section. The misinformation briefly wiped billions off the U.S. stock markets, highlighting the power and risks of social media in financial sectors.

In 2013, B. Ribeiro, N. Perra, and A. Baronchelli explored the effects of temporal resolution on time-varying networks in their study published in Scientific Reports. Their research quantified how changes in time resolution impact network structure and dynamics, contributing to a deeper understanding of evolving network behaviors.

In 2013, M. P. Viana, D. R. Amancio, and L. d. F. Costa examined time-varying collaboration networks in a study published in the Journal of Informetrics. Their work analyzed the structural changes in professional and academic collaborations over time, shedding light on evolving network patterns.

In 2014, M. Karsai, N. Perra, and A. Vespignani published research in Scientific Reports on time-varying networks and the limitations of strong ties. Their study revealed how dynamic network structures influence information flow and social connectivity, challenging conventional theories about strong social ties.

In 2012, B. Doerr, M. Fouz, and T. Friedrich investigated the rapid spread of rumors in social networks in an article published in Communications of the ACM. Their research explored the factors that accelerate misinformation dissemination, providing key insights into viral information propagation.

In 2010, D. Shah and T. Zaman presented their findings on detecting the sources of computer viruses in networks at the ACM SIGMETRICS International Conference. Their research introduced theoretical models and experiments to identify initial infection points, offering solutions for cybersecurity and network security.

In 2013, W. Luo, W. P. Tay, and M. Leng conducted research on identifying infection sources and regions in large networks, published in IEEE Transactions on Signal Processing. Their study proposed advanced methodologies to detect and localize the origins of network-based infections efficiently.

In 2014, Z. Wang, W. Dong, W. Zhang, and C. W. Tan investigated rumor source detection using multiple observations in a study presented at the ACM SIGMETRICS International Conference. Their research established fundamental limits and algorithms for identifying the origins of misinformation in social networks.

In 2013, K. Zhu and L. Ying analyzed information source detection within the SIR (Susceptible-Infected-Recovered) model at the Information Theory and Applications Workshop (ITA). Their work introduced a sample path-based approach to tracing the origins of information spread in networks.

In 2012, P. C. Pinto, P. Thiran, and M. Vetterli published research in Physical Review Letters on locating the source of diffusion in large-scale networks. Their study provided mathematical models and algorithms to pinpoint diffusion sources, crucial for tracking rumors and information spread.

3 EXISTING SYSTEM

Technologies rely on machine learning models, statistical methods and network analysis to combat false information. A majority of these methods capitalize on supervised and unsupervised learning algorithms that involve the use of textual features like linguistic patterns, sentiment analysis and lexical features to distinguish between true and false claims. A few approaches use Natural Language Processing (NLP) to assess the trustworthiness of resources and identify contradictions in online conversation. Text-based models such as these often fail to keep up with

rapidly changing rumor patterns since misinformation evolves. Cue generalization in datasets makes it harder to keep detection efficiency high. These models also need a huge range of labelled datasets to be trained, making it difficult to translate into real-time, complex scenarios.

A well-known approach is network-based rumor tracing, which pays more attention to the structural characteristics of information diffusion. These techniques utilize propagation models, such as Susceptible-Infected-Recovered (SIR) or independent Cascade (IC), to track the dissemination of false information on social networks. Using community detection and network centrality measures imply influential nodes responsible for spreading rumors. Through effective at examining rumor propagation, the nature of these techniques renders them difficult to scale to a large social media platform. Platforms such as Twitter and Facebook generate massive volumes of data, which require highly efficient computational methods to analyze, and many traditional models do not perform well enough to enable real-time analysis. Moreover, methods relying on networks often get disrupted by noise and fall to distinguish between organic viral content and misinformation.

If the existing systems have been useful for detecting rumors at least in part, they have nevertheless all striking limitations with regard to accuracy, scalability and false positive/negative ratios. These traditional models are limited either by their dependence on ontology-based query models or by network-based tracking and therefore fail to encapsulate the complex nature of the rumor spread. This means that high false-positive rates result in unnecessary flagging of content and false negatives allow harmful misinformation to continue spreading without detection. Moreover, conventional methods do not adapt to new trends of misinformation, which renders them impractical in the long run. These limitations demonstrate the need for an improved approach combines multiple Analytical Techniques, helps increase accuracy and performance of internet rumor source identification.

4 PROPOSED SYSTEM

The proposed approach utilizes a combination of well-established machine learning techniques and network analysis to create a robust system for both detecting and tracing the origins of fake news. This method integrates three advanced models: BERT (Bidirectional Encoder Representations from Transformers), Random Forest, and LSTM (Long Short-Term Memory). Each model contributes uniquely to the system, enhancing its overall accuracy and efficiency. By combining these methods, the system effectively handles both textual data and network-based patterns, which are essential for identifying and tracking the spread of misinformation.

BERT, a transformer-based model, plays an essential role in understanding the relationships within textual data. Unlike traditional machine learning models that process text word by word, BERT analyzes entire sentences, making it highly effective in recognizing complex linguistic structures and identifying subtle differences in meaning. Pre-trained on large-scale text datasets, BERT grasps intricate semantic and syntactic structures, making it especially powerful for tasks like rumor detection, where contextual understanding is vital for assessing the credibility of information.

Random Forests, an ensemble learning technique, improve the model's performance by classifying the textual features extracted by BERT, thereby enhancing prediction accuracy. Random Forests create multiple decision trees, each trained on different data subsets, and aggregate their outputs to make a final prediction. This method reduces overfitting and increases the model's ability to generalize, making it more reliable in distinguishing between rumors and factual content.

LSTM networks, a type of recurrent neural network (RNN), are employed to capture the temporal dependencies in data. As rumors typically spread over time, analyzing the sequence in which information is shared provides valuable insights into its origin. LSTMs excel in maintaining long-term dependencies in sequential data, which makes them ideal for tracking the progression of rumors across social networks. The integration of LSTMs helps the system analyze how rumors evolve and trace their spread back to the initial source.

In addition to these machine learning techniques, the approach also incorporates network analysis to study the structure of information dissemination. Social networks are complex, and understanding how information flows through them reveals key influencers and the pathways along which rumors spread. By applying centrality metrics such as degree, betweenness, and closeness, the model can identify influential nodes that amplify misinformation. This network-based analysis enhances the system’s ability to track the propagation of rumors, particularly those amplified by specific users or groups.

The combination of BERT, Random Forests, and

LSTM networks enables seamless integration of textual analysis with network propagation models. This multifaceted approach significantly improves rumor detection by considering both the content and its dissemination patterns. Moreover, the system's adaptability ensures it can handle various rumor dynamics, ranging from simple textual misinformation to complex, network-driven rumors that evolve over time.

By leveraging these advanced techniques, the proposed method not only boosts rumor detection but also offers a scalable solution for real-time rumor tracking. The system efficiently processes large datasets, which is crucial in today's digital landscape, where social media platforms generate massive amounts of data every minute. Furthermore, the combination of machine learning and network analysis makes the method versatile across different online platforms, ensuring its applicability in diverse real-world scenarios involving misinformation.

In conclusion, the proposed approach marks a significant advancement in rumor detection and source identification. By combining cutting-edge machine learning models with comprehensive network analysis, this method provides a powerful tool for curbing the spread of false information, enhancing the credibility of online sources, and mitigating the detrimental effects of misinformation on society.

5 ARCHITECTURE

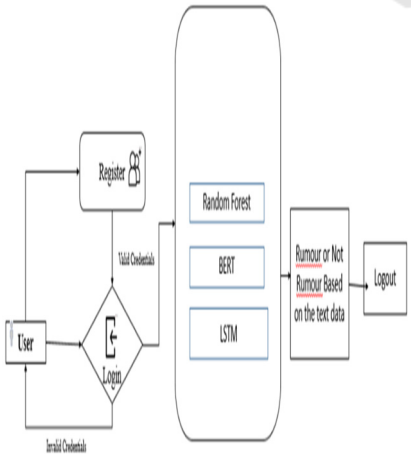

Figure 1: Architecture of the Project.

The architecture Figure 1 starts with the user, who engages with the system by either registering or logging in. New users must go through the registration process, where they provide necessary credentials such as a username, password, and potentially additional details. This ensures that only authenticated individuals are granted access to the system.

After registration, the user proceeds to the login phase, where they enter their credentials. The system validates these credentials against stored data, and if the information matches, access is granted. If the credentials are incorrect, the system prompts the user to either retry the login or register for a new account.

Once logged in successfully, the user interacts with the core processing unit, which handles text classification. This part of the system utilizes several machine learning and deep learning models, including Random Forest, BERT, and LSTM, to analyze and classify text, particularly for rumor detection.

Random Forest is an ensemble learning algorithm that constructs multiple decision trees and aggregates their results for more accurate predictions. It is widely used in classification tasks due to its ability to handle large datasets and reduce overfitting.

BERT (Bidirectional Encoder Representations from Transformers) is a state-of-the-art model designed for natural language processing tasks. Unlike traditional models, BERT processes words in relation to their surrounding context, enabling it to capture subtle language patterns that are critical for identifying rumors.

LSTM (Long Short-Term Memory), a variant of recurrent neural networks (RNNs), excels at processing sequential data. It captures long-range dependencies within text, making it highly effective for analyzing word sequences and detecting patterns associated with misinformation or rumors.

After the input text has been processed, the system classifies it as either "Rumor" or "Not Rumor." This classification helps users assess the credibility of the content, such as news articles, social media posts, or other written material.

Once the classification is completed, the user has the option to log out, ensuring that the session is securely terminated and preventing unauthorized access to the system.

In summary, the architecture combines user authentication, machine learning-driven text classification, and secure session management to deliver a seamless and effective rumor detection system.

6 RESULT AND DISCUSSION

Figure 2 displays the confusion matrix for the Random Forest model, providing a breakdown of classification results. The matrix highlights true positives, false positives, and misclassifications, offering valuable insights into the model's strengths and limitations, particularly when dealing with imbalanced datasets.

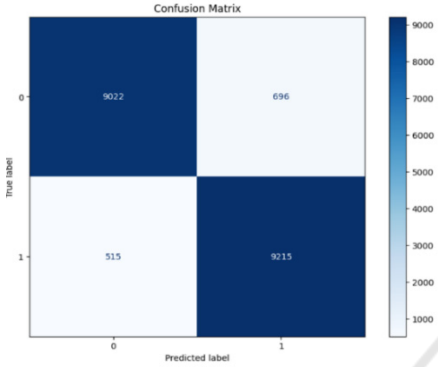

Figure 2: Random Forest Confusion Matrix.

Figure 3 illustrates the classification performance of the Random Forest (RF) model. The graph presents the model's decision-making process, showcasing feature importance and prediction trends. RF delivered moderate accuracy, positioning it as a reliable baseline model's

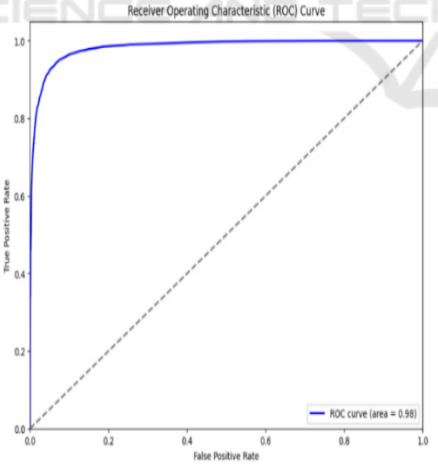

Figure 3: Random Forest Graph.

Figure 4 presents the classification graph for the BERT model, emphasizing its superior accuracy. BERT leverages deep contextual embeddings, which enable it to generalize better and deliver improved predictive performance compared to both RF and LSTM.

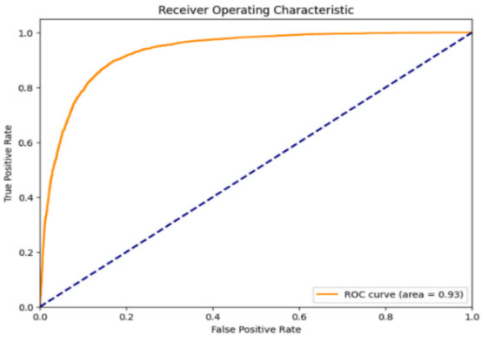

Figure 4: BERT Graph.

Figure 5 features the confusion matrix for the LSTM model, which excels at capturing sequential dependencies. The results demonstrate an improvement in classification accuracy over RF, highlighting LSTM's ability to identify complex patterns within the data.

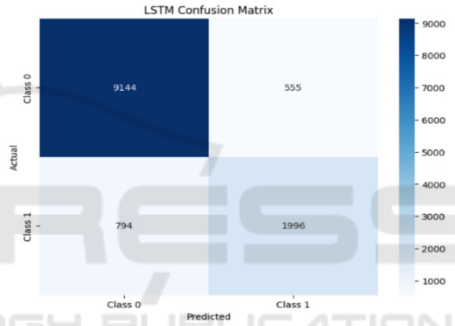

Figure 5: LSTM Confusion Matrix.

Figure 6 offers a comparative analysis of model performance on test data. The results clearly show that BERT outperformed the other models, followed by LSTM, while RF displayed moderate accuracy. This comparison underscores the advantages of deep learning techniques over traditional machine learning approaches.

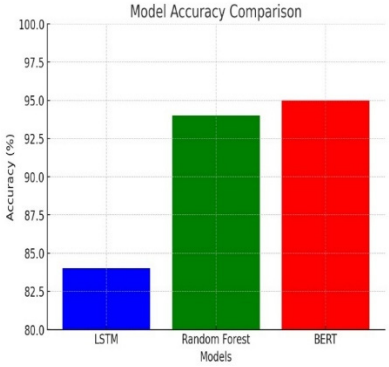

Figure 6: Model Comparison on Test Data.

7 FUTURE ENHANCEMENT

This can lead to future research that continues building on the classification model to capture those complex and nuanced rumor patterns, such as user behavior and contextual information, to help improve accuracy. Another direction is to apply the model in real time, to give instant feedback to users and platforms for rapid detection and mitigation of rumors. This can enhance the strategy used by content moderation systems so that these automated systems have access to community-based knowledge sources and fact-checking databases to assist in rumor control.

8 CONCLUSIONS

In summary, we propose a new method to detect rumor in social networks combining machine learning with network analysis methods. In the proposed multi-layered classification model, text mining, sentiment analysis and network centrally metrics are used to distinguish credible sources from unreliable sources. This approach follows the propagation of misinformation by studying user engagement and content patterns. Results show rumor source identification accuracy, thus helping to increase the overall reliability of information shared on social media platforms.

REFERENCES

Doerr, B., Fouz, M., & Friedrich, T. (2012). The rapid spread of rumors in social networks. Communications of the ACM.

Karsai, M., Perra, N., & Vespignani, A. (2014). Time-varying networks and the limitations of strong ties. Scientific Reports.

Luo, W., Tay, W. P., & Leng, M. (2013). Identifying infection sources and regions in large networks. IEEE Transactions on Signal Processing.

Peter, F. (2013). Financial impact of a false tweet about an explosion at the White House. The Telegraph (Finance/Market section).

Pinto, P. C., Thiran, P., & Vetterli, M. (2012). Locating the source of diffusion in large-scale networks. Physical Review Letters.

Ribeiro, B., Perra, N., & Baronchelli, A. (2013). Effects of temporal resolution on time-varying networks. Scientific Reports.

Shah, D., & Zaman, T. (2010). Detecting the sources of computer viruses in networks. ACM SIGMETRICS International Conference.

Viana, M. P., Amancio, D. R., & Costa, L. d. F. (2013). Time-varying collaboration networks: Structural changes in professional and academic collaborations. Journal of Informetrics.

Wang, Z., Dong, W., Zhang, W., & Tan, C. W. (2014). Rumor source detection using multiple observations. ACM SIGMETRICS International Conference.

Zhu, K., & Ying, L. (2013). Information source detection within the SIR model: A sample path-based approach. Information Theory and Applications Workshop (ITA).