ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

ISSN 2454-9940
www.ijasem.org
INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
Vol 19, Issue 2, 2025

A Smart Method for Detecting Rumors in Social Networks Using Multiple Models

[1]Mr.Ch.Surya Prakash,[2]Gellanki Manikanta,,

[2] Students,Dept.of MCA, Rajamahendri Institute of Engineering & Technology, Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107.

[1] AssociateProfessor,Dept.of MCA, Rajamahendri Institute of Engineering & Technology, Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107.

Abstract—

Society as a whole feels the effects of how social media shapes public opinion. The sheer volume of messages on social media platforms like Twitter and Facebook makes quality control a challenge, despite the fact that these platforms provide channels for sharing news and ideas. Numerous services, including content production and distribution, are offered by these popular platforms. Unfortunately, not everything you read online can be trusted. In an effort to sway public opinion, many individuals propagate inaccurate and misleading information. Various methods for rumor detection, especially for detecting bogus news, are reviewed in this work. It also suggests ways to identify and categorize false news and how to deal with it. The unsubstantiated claims made in rumors and other forms of misinformation may have devastating effects. Despite their popularity, rumors may easily spread on social media due to their unregulated nature. Automatically identifying rumors from tweets and posts is a highly sought-after research field in social media analytics. Term Index—Social media, rumors, false news, reduction

I. INTRODUCTION

Using social media makes it easier for information to be distributed or propagated [1]. Globally, a large number of people utilize social media sites such as Twitter and Facebook. The proliferation of online news sources and social media has greatly simplified the process of keeping up with current events. Many interesting occurrences may be followed on the Internet. The increasing number of people using mobile devices has made this procedure more accessible. Nevertheless, a substantial amount of duty accompanies substantial potential. Society is profoundly affected by the media. Someone is trying to take advantage of it, as is typical. For their own ends, the media may use a variety of information manipulation techniques. This leads to the publication of news articles that are somewhat truthful but not totally accurate. Just a small number of websites are dedicated primarily to disseminating misinformation. They use social media to boost their traffic and impact, and they deliberately create false content that masquerades as news. The main goal of fake news websites is to influence public opinion on certain issues. Therefore, disinformation is an issue and a task on a worldwide scale. A large number of academics hold the view that ML and AI may help combat disinformation. Reason being, more comprehensive datasets are more easily accessible, and technology is becoming more affordable. On some classification tasks, such as image recognition and speech detection, artificial intelligence systems have now started to surpass humans. False news and other misleading facts may manifest in many ways because information influences our views of the cosmos. Intelligence studies show that less informed individuals are hardwired to believe false information spread via social media, which in turn leads them to make poor choices. Some people use false news to make others afraid, spread racist ideas, and incite violence and bullying against others who aren't guilty. Anyone may rapidly send news, putting its dependability at risk, due to the increasing number of clients in web-based life. The substance of news stories alone is not enough to identify fake news since its purpose is to mislead the audience. Almost all of these methods center on the idea of classifying texts as either genuine or false, or as part of a larger effort to detect rumors as a danger [2]. You may find the specifics in the literature. In certain cases, methods based on deep learning and machine learning have shown promise [3]. Research shows that SVMs, which use content-based characteristics including visual and linguistic elements, are more effective than most supervised machine learning (SML) algorithms when it comes to text deception identification. There are several potential uses for detecting false news, particularly in the realm of social media and related applications like online marketing and recommender systems [4]. Using a variety of machine learning techniques, we propose a large number of rumor categories and methods for detecting rumors in this study [5]. Using machine learning techniques and classifiers, this paper's main contributions are as follows: recommending the best machine learning classifier; and classifying suspicious material as either rumor or non-rumor. The rest of the article is structured as follows: Section II reviews the research on several rumor detection systems that use a machine learning methodology. Section IV delves into the validation approach and outcomes of the suggested rumor detection method, which was discussed in Section III. Lastly, Section V serves as the article's conclusion.

II. LITERATURE STUDY

The field of rumor detection has produced several useful studies. The writers use a variety of machine learning methods in [6], including logistic regression, decision trees, Bayesian networks, lazy learners, support vector machines, and

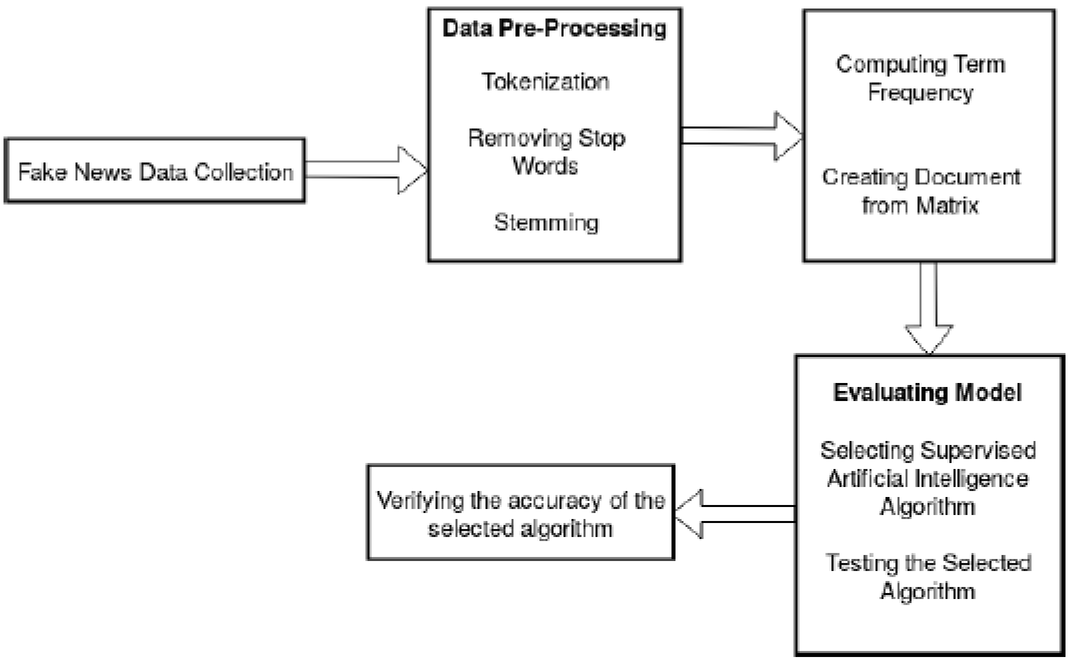


Fig. 1. Rumor Detection and Classification Process

Train rule-based systems to identify fraudulent financial statements. In comparison to the other models, the decision tree achieves the highest classification accuracy, as shown by the results. Integrating qualitative data into the input vector has the potential to raise the accuracy rate. In [7], the authors use a big data platform to identify financial rumors. Architecture that enables the effective identification of financial rumors was also suggested by the authors. To go further into the rumor detection framework, many case studies may be used. Finding rumors on social media using context is the focus of this study. The techniques used for rumor identification include Naive Bayesian and Support Vector Classifier. The suggested model has enhanced inaccuracy, according to experiments. Various classification techniques might be used to ensure precision. The authors of [8] primarily discuss supervised rumor detection and classification using the J48 classifier. This classifier is integrated into the WEKA platform and operates under both the single-step Rumor Detection Classification (SRDC) and two-step RDC (TRDC) schemes. he findings show that SRDC has a worse F-measure compared to TRDC for the MIX dataset. Because of constraints imposed by the model, the pre-processing activity is severely limited. Incorporating the SVM classifier into a supervised learning framework, the authors of [9] present a novel model referred to as Propagation Kernel Tree (PKT). From the Twitter database. Both early and wide rumor detection are improved upon by the proposed method compared to the state-of-the-art baseline. The authors of [10] used three ML algorithms—Random Forest, Naive-

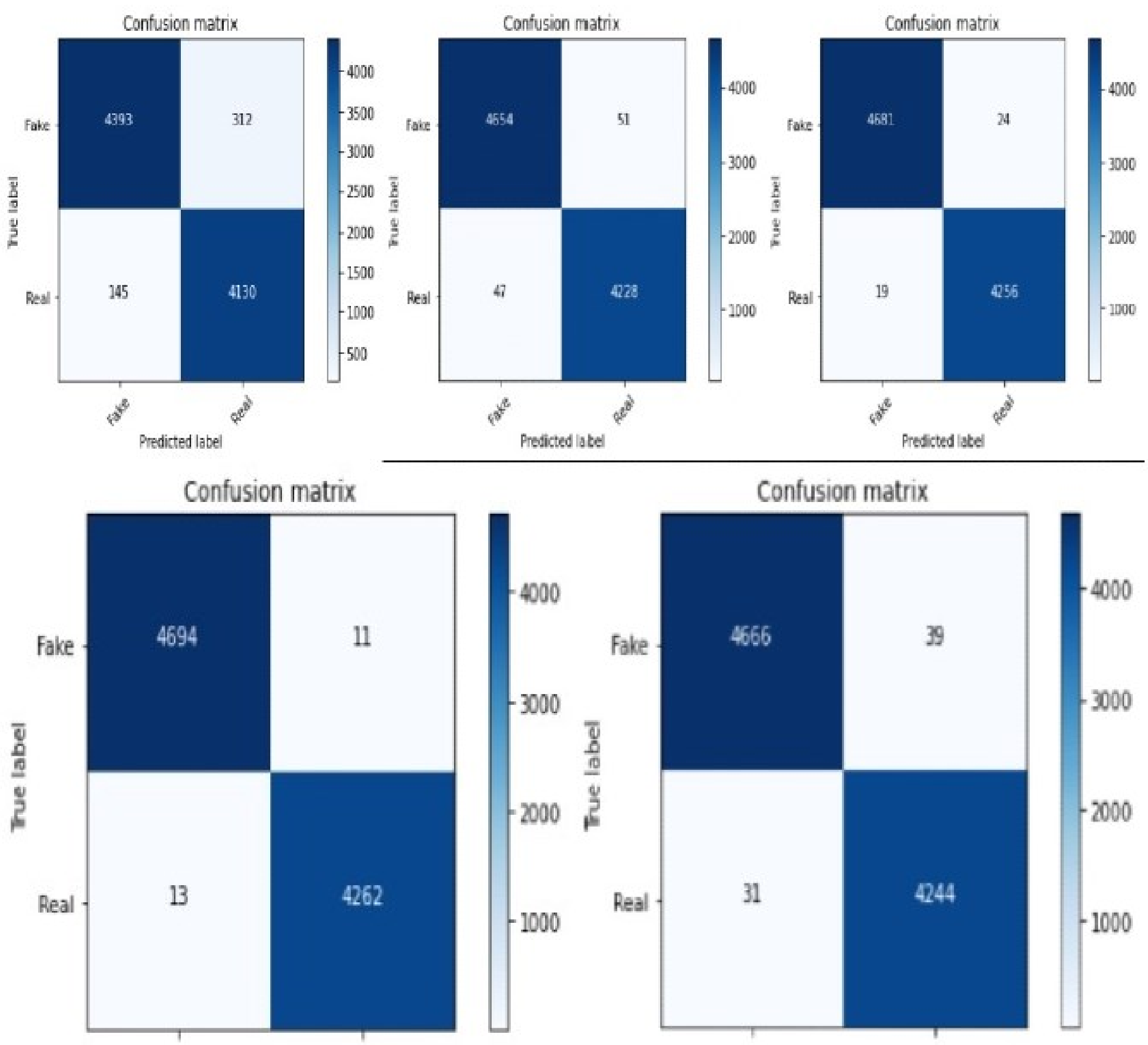
Bayes, and Random Decision Tree—to identify health-related rumors on Twitter. With a recall of 0.944 and an accuracy of 0.946, the Random forest outperforms the other two classifiers in the experiment. We propose a new framework for detecting emerging rumors in social media, the Cross-topic Emerging Rumor Detection (CERT) based sparse representation model [11]. When compared to other methods, experiments show that CERT is superior at identifying spreading rumors. Incorporating crossmodal data may improve the system. A multi-featured automated method for detecting rumors based on hot-topic identification and Rumor Identification. The rumor classification job is subjected to extensive testing comparing logistic regression, random forest, and Naive Bayes. Random forest is the most effective model, according to the experiments, and improving the models' effectiveness may be achieved by exploring a larger set of characteristics and using stronger probabilistic models. Using crowd wisdom, the authors of [12] proposed an effective method for identifying disinformation and bogus news. The findings show that on some datasets, the spread of false information may be significantly reduced. Aggressive conduct, on the other hand, may aid advancement even further. In order to identify rumors on microblogs, the authors of [13] created a content representation technique that made use of a neural network model and a bag of words. Compared to the bag-of-words model's 90% accuracy rate, the neural model's rate is 60% lower, according to the data. The models' performance may be further enhanced by examining other aspects and parameters, however.

III. THE PROPOSED SYSTEM

In order to distinguish between real and fraudulent news pieces across several domains, we provide an ensemble method that makes use of separate sets of linguistic features. The LIWC feature set was recommended by the framework as a way to assemble the approaches. You may find fact-checking news articles published by a number of respectable websites. In addition, researchers keep track of all the datasets that are presently accessible in open repositories. These repositories also provide links to fact-checking websites that might be useful in the battle against fake news. However, in order to conduct the experiments, we combined three enormous datasets that included a variety of subjects (e.g., politics, entertainment, technology, and sports) and genuine and false tales. Kaggle was used to get the datasets. Studies tracking the development of methods to identify and validate rumors have gained traction with the proliferation of social media. There are often several phases involved in conducting a fact-checking method for a particular claim, but generally speaking, they include: _ gathering potentially relevant documents to support the claim _ determining the document's position in relation to the claim _ assigning a trustworthiness grade to the documents _ drawing a conclusion based on the preceding procedures. A rumor classification system consists of four parts: detecting rumors, monitoring them, classifying their attitude, and classifying their truth. The development of a rumor categorization system is heavily influenced by temporal characteristics, such as the emergence of new rumors amid breaking news. The most recent rumors that circulate as a consequence of breaking news tend to be completely novel. Since a rumor categorization system's training data could not match what it would encounter in the real world, it is essential that rumors be found automatically and that the system can handle new, unseen rumors. Under these conditions, real-time evaluation of a stream of communications is essential for early rumor identification and resolution. There have been whispers about this for quite some time. It may take a long time for the veracity of some rumors to be confirmed. People are curious about these rumors even though it's hard to tell which ones are true. On top of that, the system may utilize the classifier it generated for older data to classify newer data, which is useful since the vocabulary is less likely to change in continuing rumor exchanges. Posts do not need real-time processing because, unlike freshly emerging murmurs, long-standing rumors are often analyzed retroactively. The first stage of the four-step process for rumor detection that is presented here is to collect data from all of the social media platforms that are being considered. For the purpose of extracting useful characteristics, this data must be consistently structured. The pre-processing steps include cleaning, reduction, transformation, and consolidation. Specific to the network, content-based, and pragmatic features are culled. Using a range of machine learning techniques, including Naive Bayesian, Support Vector Machines, Decision Tree, Random Forest, and Logistic Regression, each piece of data is classified as a rumor or not. Part A: Pre-Processing Dates Collecting data for false news detection classification is a complex endeavor in and of itself. If information with fake labels is difficult to spot during categorization, then the data and its natural label are as predicted. There are a lot of records that are obviously incorrect. We take these processes in order to fine-tune the records: _ Remove any items with no value. _ Eliminate any duplicate records and retain just one, marking it as 1 (positive). Put all numbers into standard records and set all unknown records to zero. Following data collection, we pre-process by following this sequence: Data consolidation is the process of merging all of the collected data into one cohesive database. Additionally, the procedure includes transforming all obtained data into a uniform format. _ Data-Cleaning: A great deal of irrelevant information (noise) exists in the data and must be eliminated. Data cleaning involves removing any unnecessary

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

information, inconsistencies, and noise from the dataset. Data transformation is the process of enhancing data by collecting it, standardizing it, and adding particular properties. _ Reducing Data: Fewer variables and instances are considered. Equalization of the skewed data occurs at this level as well. _ The data is suitable for feature extraction as it has been appropriately shaped following pre-processing. Subsequently, SVM, LR, Naive Bayes, and RF are used to train and fine-tune the models. To integrate all of these models, Voting Classifiers are used. They combine all of these classifiers into an ensemble classifier, which then utilizes the soft voting approach to get the final prediction based on the label and class probabilities. Scalar Multiplier For tokenization, we use spaces and punctuation marks as delimiters after removing the English stopwords from all datasets. After tokenizing the headlines, a sparse matrix is produced with each headline row and the token column. Tokens now reflect not only their morphological but also their contextual use, thanks to the reappearance of multiple n-grams. B. The Ensemble Approach The ensemble method is a multi-model approach that improves accuracy for ensemble learners by reducing the error rate and enhancing performance. Compared to the previously utilized model, the ensemble model is quite comparable. It is standard practice to collect many expert views before settling on a course of action. Which aid in minimizing uncertainty leading up to a final decision or a negative result. A decision boundary that is compatible with the data is generated by a classification algorithm. These algorithms take into account various factors to operate on certain datasets. Everything from the tarining dataset to the machine learning model and the parameters used determine the outcomes. Overfitting or underfitting of the training model occurs depending on the data used and might lead to biased results when applied to new data. Corss validation may help alleviate the overfitting problem. Data models



INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Fig. 2. Prediction Labels using Ensemble model

choose based on past performance by training several datasets with regard to certain factors. That way, classification may be done with the help of a suitable model that has decision and limit parameters already set. So, to fix classification errors and get the best results, ensemble learning methods might be utilized. The number of votes collected by the models trained using various techniques determines the final categorization. One use of classification based on training algorithm results is voting classifiers. 1) Binary Tree: The provided model was fine-tuned using a variety of parameters in order to discover the best one for making accurate predictions. When solving a regression or classification issue, techniques based on decision trees function better. Various algorithms that work with data presented as a tree fall under this category. The tree dataset is processed and classified effectively using the cost estimation approach. Subtracting the provided dtata from the sum of each class's squared probability, the Gini index functions as the cost function and identifies classification difficulties. 2) Bagging Ensemble Classifier: Bagging classifiers were one of the first ensemble approaches to prevent overfitting. In order to solve the classification issue, the bagging classifier uses votes from the chosen class. Selecting a data tree model is how it operates on datasets. This type is also known as bootstrao aggregation. One kind of bagging classifier is the random forest model. As an early ensemble technique, bagging classifiers or bootstrap aggregating helped reduce variance (overfitting) in training sets. The random forest model is among the most often used variations of bagging classifiers. In a classification issue, the bagging model chooses a class based on primary votes assessed by several trees. However, in order to decrease overall variance, the data for each tree is obtained using random sampling with replacement from the full dataset. However, in order to solve regression problems, the bagging model averages several estimates. 3) Boosting Ensemble Classifier: This technique guides learners to correctly classify data points by using an incremental strategy. In the beginning, all data points are classified using the same weighted coefficients. Next time around, the weighted coefficients are lowered for properly labeled data points and raised for wrongly labeled ones. The proper identification of misclassified data points in earlier rounds improves overall accuracy, and each round's subsequent tree learns to lessen the errors of the prior round. 4) The Voting Ensemble Classifier: Compared to bagging and boosting algorithms, the voting ensemble is easier to construct. Using a random selection and replacement process, bagging algorithms create several datasets by dividing the overall dataset into smaller groups. A generic model that can accurately classify the issue is trained via boosting by training numerous models sequentially, with each model learning from the previous by raising weights for misclassified data. A voting ensemble is a group of separate models that work together to make a majority voting prediction based on their categorization findings. Figure 2 shows the predicted occurrence of bogus news using the ensemble technique.

IV. EXPERIMENTATION AND RESULT

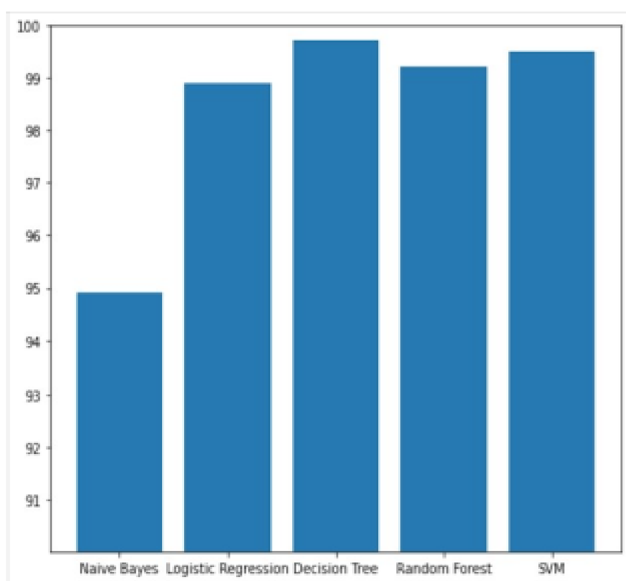


Fig. 3. Comparison of Ensemble model

In Fig. 3, we can see a statistical comparison of various algorithms' accuracy levels. By thinking about the finished dataset, we can see the summary. Through the use of decision trees, a maximum accuracy of 99.73% was attained. In second position, with an accuracy of 99.52%, is the support vector machine (SVM). With a random forest accuracy of 99.22%, it ranks third. With a 98.91% success rate, logic regression outperforms naïve bayes, which comes in at 94.91%. On the final dataset, all of the algorithms have almost identical performance, with decision tree displaying the best accuracy and naive bayes the worst.

V. CONCLUSION

Negative social effects, such as animosity and panic, may be disseminated via rumors. So, it is necessary to debunk rumors. This paper provides a concise overview of the psychology research on rumors, current techniques for identifying rumors, and the evaluation matrix for assessing the efficacy of these approaches. Studies aimed at identifying rumors have increased in number in tandem with the use of social media. We need a complete system that can promptly identify new rumors in the making as existing methods can't handle stream data effectively or automatically spot new rumors on social media.

REFERENCES

[1] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, “The role of social networks in information diffusion,” in Proceedings of the 21st international conference on World Wide Web, 2012, pp. 519–528.
[2] T. Takahashi and N. Igata, “Rumor detection on twitter,” in The 6th International Conference on Soft Computing and Intelligent Systems, and The 13th International Symposium on Advanced Intelligence Systems. IEEE, 2012, pp. 452–457.
[3] A. Kumar and S. R. Sangwan, “Rumor detection using machine learning techniques on social media,” in International conference on innovative computing and communications. Springer, 2019, pp. 213–221.
[4] N. M. Santhosh, J. Cheriyan, and M. Sindhu, “An intelligent exploratory approach for product recommendation using collaborative filtering,” in 2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS), 2021, pp.
232–237.
[5] G. Liang, W. He, C. Xu, L. Chen, and J. Zeng, “Rumor identification in microblogging systems based on users behavior,” IEEE Transactions on Computational Social Systems, vol. 2, no. 3, pp. 99–108, 2015.
[6] D. Wang, “Analysis and detection of low quality information in social networks,” in 2014 IEEE 30th International Conference on Data Engineering Workshops, 2014, pp. 350–354.
[7] X. Liu, A. Nourbakhsh, Q. Li, R. Fang, and S. Shah, “Realtime rumor debunking on twitter,” in Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, ser. CIKM ’15. New York, NY, USA: Association for Computing Machinery, 2015, p. 18671870. [Online]. Available: https://doi.org/10.1145/2806416.2806651
[8] G. Cai, H. Wu, and R. Lv, “Rumors detection in chinese via crowd responses,” in 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), 2014, pp. 912–917.
[9] M. Lukasik, T. Cohn, and K. Bontcheva, “Estimating collective judgement of rumours in social media,” CoRR, vol. abs/1506.00468, 2015. [Online]. Available: http://arxiv.org/abs/1506.00468
[10] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, “Prominent features of rumor propagation in online social media,” in 2013 IEEE 13th International Conference on Data Mining, 2013, pp. 1103–1108.
[11] F. Yang, Y. Liu, X. Yu, and M. Yang, “Automatic detection of rumor on sinaweibo,” in Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, ser. MDS ’12. New York, NY, USA: Association for Computing Machinery, 2012. [Online]. Available: https://doi.org/10.1145/2350190.2350203
[12] Z. Yang, C. Wang, F. Zhang, Y. Zhang, and H. Zhang, “Emerging rumor identification for social media with hot topic detection,” in 2015 12th Web Information System and Application Conference (WISA), 2015, pp. 53–58.
[13] S. Wang and T. Terano, “Detecting rumor patterns in streaming social media,” in 2015 IEEE International Conference on Big Data (Big Data), 2015, pp. 2709–2715.