

Detecting Fake News

Abinaya. S¹, Sheelavathi. A², Devi. E³, Kalpana. S⁴, Kaviya Tharsini. L⁵

Students, Information Technology, Saranathan College of Engineering, Trichy^{1, 2, 3, 4}

Assistant Professor, Information Technology Saranathan College of Engineering, Trichy⁵

Abstract: Internet is one of the important inventions and a large number of persons are its users. These persons use this for different purposes. There are different social media platforms that are accessible to these users. Any user can make a post or spread the news through these online platforms. FAKE news has proliferated to a big crowd than before in this digital era, the main factor derives from the rise of social media and direct messaging platform. Fake news detection is important research to be done for its detection but it has some challenges too. Some challenges can be due to a smaller number of resources like an available dataset. We propose in this project, a fake news detection using deep learning technique. And implement a novel automatic fake news credibility inference model using deep learning algorithm with Natural Language processing steps which including text mining steps. Based on a set of explicit and latent features extracted from the textual information, deep learning algorithms builds a deep diffusive network model to learn the representations of news articles, creators and subjects simultaneously. There is a Kaggle competition called as the "Fake News Challenge" and social network is employing AI to filter fake news stories out of users' feeds. Combatting the fake news is a classic text classification project with a straight forward proposition. And evaluate the performance of the system in terms of accuracy.

Keywords: Fake, Real, News, Dataset.

I. INTRODUCTION

A different way to detect fake news is through stance detection which will be the focus of our study. Stance Detection is the process of automatically detecting the relationship between two pieces of text. In this study, we explore ways to predict the stance, given a news article and news headline pair. Depending on how similar the news article content and headlines are, the stances between them can be defined as 'agree', 'disagree', 'discuss' or 'unrelated'. We experimented with several traditional machine learning models to set a baseline and then compare results to the state-of-the-art deep networks to classify the stance between article body and headline Fake news can be come in many forms, including: unintentional errors committed by news aggregators, outright false stories, or the stories which are developed to mislead and influence reader's opinion. While fake news may have multiple forms, effect that it can have on people, government and organizations may generally be negative since it differs from the facts. Detecting fake news is hard for many reasons.

First, manual task of identifying fake news is very subjective. Assessing the veracity of a news story is a complex and cumbersome task, even for trained experts. News is not only spread through traditional media outlets anymore but also through various social media channels. Automated solution requires understanding the natural language processing which is difficult and complex. These complexities make it a daunting task to classify text as fake news. In the current fake news corpus, there have been multiple instances where both supervised and unsupervised learning algorithms are used to classify text. However, most of the literature focuses on specific datasets or domains, most prominently the politics domain. Therefore, the algorithm trained works best on a particular type of article's domain and does not achieve optimal results when exposed to articles from other domains.

Since articles from different domains have a unique textual structure, it is difficult to train a generic algorithm that works best on all particular news domains. In this paper, we propose a solution to the fake news detection problem using the machine learning ensemble approach. Our study explores different textual properties that could be used to distinguish fake contents from real.

II. EXISTING SYSTEM

Hidden Markov Models:

Extract features to represent both headlines and bodies of the news articles.

Bilateral Multi-Perspective Matching Models (BiMPM): Provide full attention mechanism between words in body text and headlines.

LSTM Model:

To provide global features and local word embedding features of news dataset.

DISADVANTAGES

- Accuracy is less
- Need large number of datasets to train the data
- Provide high number of false positive rate

III. PROPOSED SYSTEM

"Fake News" is a term used to represent fabricated news or propaganda comprising misinformation communicated through traditional media channels can implement text mining algorithm to extract the key terms based on natural language processing and also include classification algorithm such as deep learning algorithm named as Multi-layer perceptron.

ADVANTAGES

- Analyze all types of features
- Improve the accuracy rate
- Time complexity can be reduced

3.1 PROPOSED SYSTEM ARCHITECTURE

System architecture refers to the overall design and structure of a system, which includes hardware, software, network, and other components. It is a conceptual model that describes the relationships and interactions between different components of a system and how they work together to achieve a common goal. System architecture can be classified into different categories based on the type of system being developed. In this architecture, we can train the datasets that are collected from Web sources named as KAGGLE and extract the keywords using Natural language processing. Then build the model using MLP algorithm. In testing phase, user can input the text and classify the text whether it is fake or real

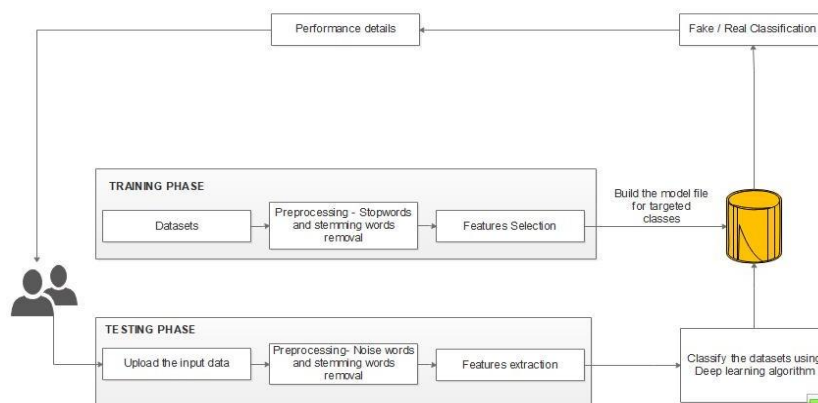


Fig.1: System Architecture

IV. IMPLEMENTATION

For this process, we can upload the datasets from users and upload the news group datasets. The given input document is processed for removing redundancies, inconsistencies, separate words, stemming, and documents are prepared for next step, the stages implement stop words removal, stemming words analysis to structure the document. Can calculate the term frequency and inverse document frequency. In information retrieval, tf-idf or TFIDF, short for term frequency–inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus, extract the relevant features from the uploaded documents. User can input the news datasets or twitter datasets. In this module, implement convolutional neural network algorithm to classify the extract keywords.

MLP algorithm is used to classify the datasets whether it is fake or real. Based on classification, fake news data are predicted. The proposed system provides improved accuracy rate in fake news detection. Accuracy parameter is calculated in terms of true positive and false positive rate.

V. EXPERIMENT RESULTS

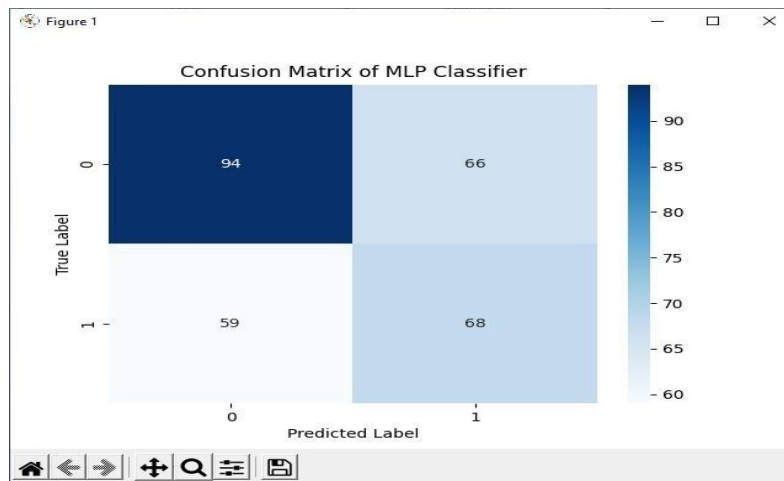


Fig.2

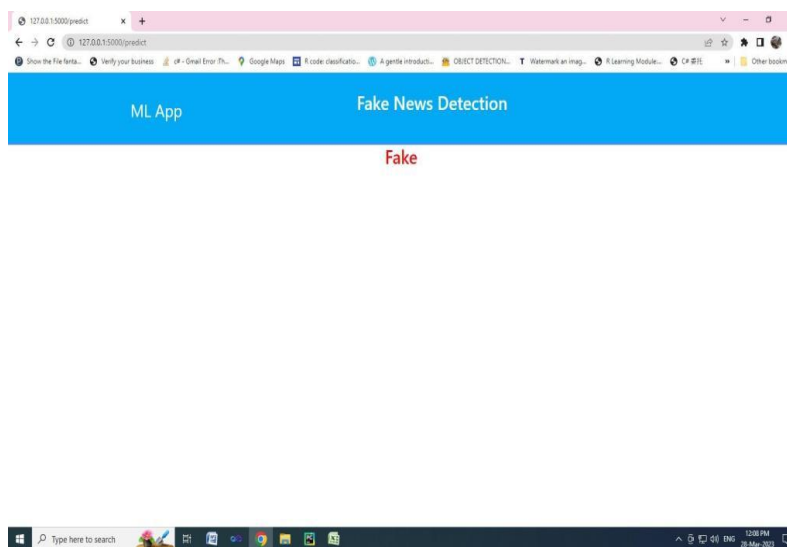


Fig.3

VI. CONCLUSION

In this project, we have studied the fake news article, creator, and subject detection problem. Based on the news augmented heterogeneous social network, a set of explicit and latent features can be extracted from the textual information of news articles, creators, and subjects respectively. Furthermore, based on the connections among news articles, creators and news subjects, a deep diffusive network model has been proposed for incorporate the network structure information into model learning. The accuracy metric presumably would be altogether improved by methods for utilizing progressively complex model. It is worth noting, that even with the given dataset, only part of the information was used.

The current project did not include domain knowledge related features, such as entity-relationships. The proposed system proves that multi-layer perceptron neural network algorithm provides improved accuracy rate We formulated the fake news detection on social media as an inference problem in deep learning model that can be solved using multi-layer neural network algorithm. We can conclude that, the proposed system to provide improved accuracy rate in fake news detection. Experiments on well-known benchmark datasets show that the proposed model consistently improves over the state of the art in fake news detection in both the late and early detection settings.

REFERENCES

- [1] Mihaylov, Todor, Georgi Georgiev, and Preslav Nakov. "Finding opinion manipulation trolls in news community forums." Proceedings of the nineteenth conference on computational natural language learning. 2015.
- [2] Mihaylov, Todor, and Preslav Nakov. "Hunting for troll comments in news community forums." arXiv preprint arXiv:1911.08113 (2019).
- [3] Bourgonje, Peter, Julian Moreno Schneider, and Georg Rehm. "From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles." Proceedings of the 2017 EMNLP workshop: natural language processing meets journalism. 2017.
- [4] Chopra, Sahil, Saachi Jain, and John Merriman Sholar. "Towards automatic identification of fake news: Headline-article stance detection with LSTM attention models." Proc. Stanford CS224d Deep Learn. NLP Final Project. 2017.
- [5] Konstantinovskiy, Lev, et al. "Toward automated factchecking: Developing an annotation schema and benchmark for consistent automated claim detection." Digital threats: research and practice 2.2 (2021): 1-16.
- [6] Ul Haq, Muhammad Aziz, et al. "Indoor localization using improved multinomial naïve bayes technique." International afro- European conference for industrial advancement. Springer, Cham, 2016.
- [7] Lu, Jing, Peilin Zhao, and Steven Hoi. "Online passive aggressive active learning and its applications." Asian Conference on Machine Learning. PMLR, 2015.
- [8] Nayak, Janmenjoy, Bighnaraj Naik, and H. S. Behera. "A comprehensive survey on support vector machine in data mining tasks: applications & challenges." International Journal of Database Theory and Application 8.1 (2015): 169-186.
- [9] Kataria, Aman, and M. D. Singh. "A review of data classification using k-nearest neighbour algorithm." International Journal of Emerging Technology and Advanced Engineering 3.6 (2013): 354-360.
- [10] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," ACM Comput. Surv., vol. 53, no. 5, pp. 1–40, Sep. 2020.