

## AN ANALYSIS ON HOW TO EVALUATE FAKE NEWS USING DATA SCIENCE APPLICATIONS

<sup>1</sup>Jyothsna Segu, <sup>2</sup>Naga Priyanka Chadalavada, <sup>3</sup>Priyanka Korathaluri, <sup>4</sup>Sujana B  
<sup>1,2,3</sup>Assistant Professor, <sup>4</sup>UG Student, <sup>1,2,3,4</sup>Department of Information Technology, Rishi MS Institute of Engineering and Technology for Women, Kukatpally, Hyderabad.

### ABSTRACT

In this study, we created a new paradigm for identifying false news, an important issue that has lately emerged in social media. We evaluated the effectiveness of several machine learning techniques. Effectively identifying bogus news turns out to be a difficult task. Big data categorization issues are well suited to Apache Spark's machine learning environment, which allows for the simultaneous use of numerous processors. Following tests on huge datasets we received from Kaggle utilizing Naive Bayes, Neural Network, Logistic Regression, and Support Vector Machine revealed that our programmed can claim up to 99% accuracy rates.

### INTRODUCTION

The deliberate misrepresentation of facts under the appearance of respectable journalism, sometimes known as "fake news," is an issue that influences opinion formation, decision-making, and voting habits globally. The majority of false information is first spread through social media channels like Facebook and Twitter before making its way to traditional news outlets like television and radio. Key language traits common to fake news articles that are first disseminated on social media platforms include an overuse of unsupported exaggeration and unattributed cited information. This paper presents and discusses the findings of a fake news identification investigation that shows the effectiveness of a fake news classifier.

### Motivation

In this paper, the research process, technical analysis, technical linguistics work, and classifier performance and results are presented. The paper concludes with a discussion of how the current system will evolve into an influence mining system. The fake news stories that are initially seeded over social media platforms share key linguistic characteristics such as excessive use of unsubstantiated

hyperbole and non-attributed quoted content. The results of a fake news identification study that documents the performance of a fake news classifier are presented and discussed in this paper.

#### Objective:

Fake news has been demonstrated to be problematic in multiple ways. It has been shown to have real influence on public perception and the ability to shape regional and national dialogue. It has harmed businesses and individuals and even resulted in death, when an individual responded to a hoax. It has caused some teenagers to reject the concept of media objectivity and many students can't reliably tell the difference between real and faked articles. It is even thought to have influenced the 2016 United States elections. Fake news can be spread deliberately by humans or indiscriminately by bot armies, with the latter giving a nefarious article significant reach. Not just articles are faked, in many cases fake, mislabeled or deceptive images are also used to maximize impact. Some contend that fake news is a "plague" on society's digital infrastructure. Many are working to combat it. Farajtabar, et al., for example, has proposed a system based on points, while Haigh, Haigh and Kozak have suggested the use of "peer-to-peer counter propaganda

#### Proposed System:

In this paper author is describing concept to detect fake news from social media or document corpus using Natural Language Processing and attribution supervised learning estimator. News documents or articles will be uploaded to application and then by using Natural Language Processing to extract quotes, verbs and name entity recognition (extracting organizations or person names) from documents to compute score, verbs, quotes and name entity also called as attribution. Using supervised learning estimator we will calculate score between sum of verbs, sum of name entity and sum of quotes divided by total sentence length. If score greater than 0 then news will be consider as REAL and if less than 0 then new will be consider as FAKE.

#### Advantages of proposed system:

- It is desirable to use COX data for phylogenetic exploration.

- We use the data of COX experimental values.
- Security

**II. Literature survey:**

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter, 19(1), 22-36.

Nikiforos M.N., Vergis S., Styliou A., Augoustis N., Kermanidis K.L., Maragoudakis M. (2020) Fake News Detection Regarding the Hong Kong Events from Tweets. In: Maglogiannis I., Iliadis L., Pimenidis E. (eds) Artificial Intelligence Applications and Innovations.

AIAI 2020 IFIP WG 12.5 International Workshops. AIAI 2020. IFIP Advances in Information and Communication Technology, vol 585. Springer, Cham. [https://doi.org/10.1007/978-3-030-49190-1\\_16](https://doi.org/10.1007/978-3-030-49190-1_16)

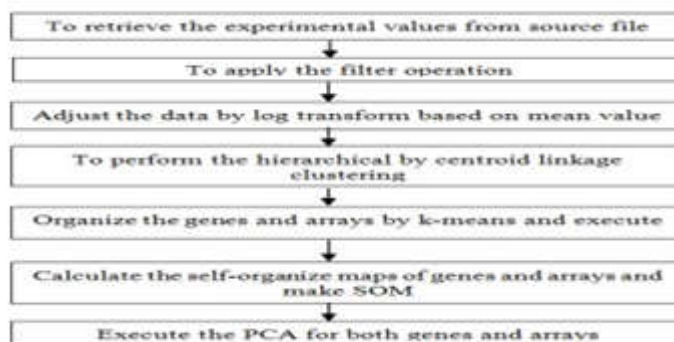
Thota, A., Tilak, P., Ahluwalia, S., & Lohia, N. (2018). Fake news detection: a deep learning approach. SMU Data Science Review, 1(3), 10.

Thota, Aswini, et al. "Fake news detection: a deep learning approach." SMU Data Science Review 1.3 (2018): 10. [5] Nada, F , Khan, B , Maryam, A , Zuha, N, Ahmed, Z . (2019). FAKE NEWS DETECTION USING LOGISTIC REGRESSION. International Research Journal of Engineering and Technology (IRJET). <https://www.irjet.net/archives/V6/i5/IRJET-V6I5733.pdf> [6] Medium. "Decision Tree Classification".

Access: 4 June 2021. <https://medium.com/swlh/decision-tree-classification-de64fc4d5aac>.

**SYSTEM DESIGN**

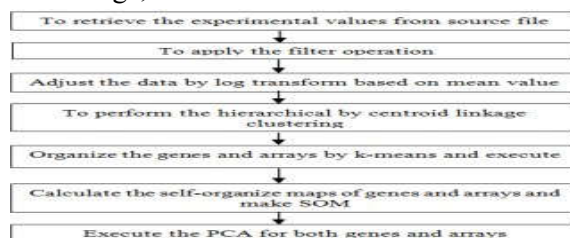
**SYSTEM ARCHITECTURE:**



Algorithms used in this project :-

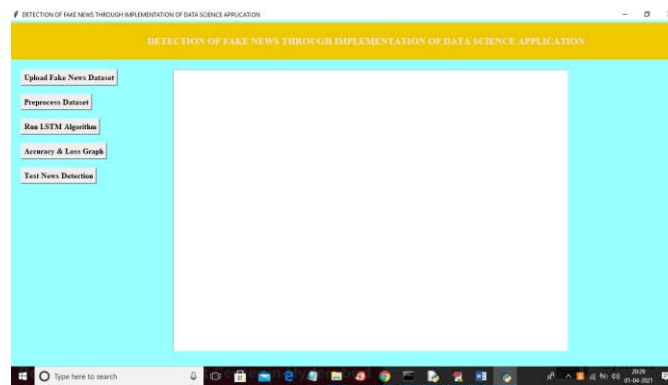
**CLUSTERING ALGORITHM :**

Clustering is an unendorsed culture algorithm that finds the concealed arrangement in the unlabeled data. In this work, we used the filter the values, adjust the data values, then apply the hierarchical method, k-means algorithm, self-organizing maps (SOM), and finally apply the Principal Component Analysis (PCA) for avoid the unwanted values, adjust the data with help of log transform, for clustering genes and arrays with hierarchical clustering by centroid linkage,

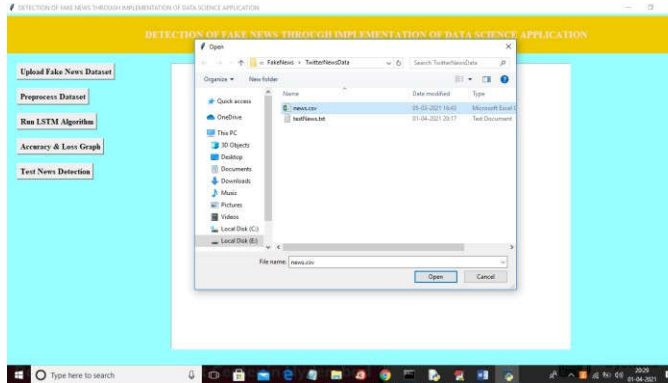


**RESULTS:**

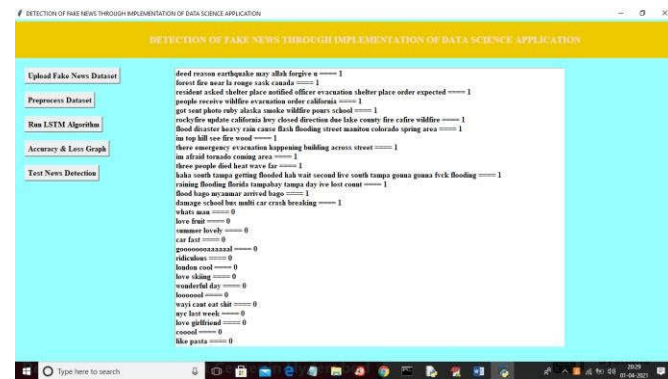
To run project double click on „run.bat“ file to get below screen



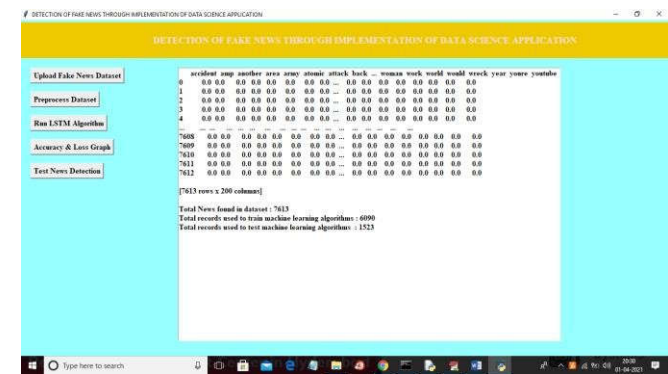
In above screen click on „Upload Fake News Dataset“ button to upload dataset



In above screen selecting and uploading „news.csv“ file and then click on „Open“ button to load dataset and to get below screen

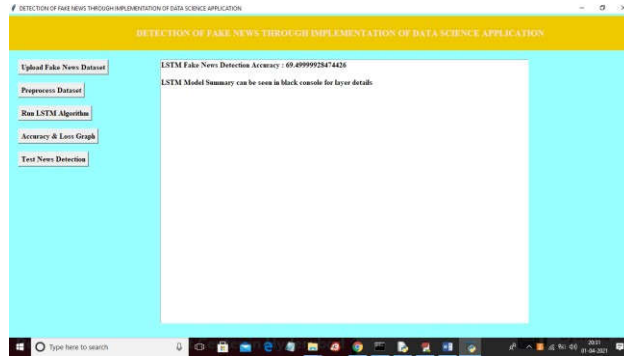


In above screen dataset loaded and then in text area we can see all news text with the class label as 0 or 1 and now click on „Preprocess Dataset & Apply NGram“ button to convert above string data to numeric vector and to get below screen

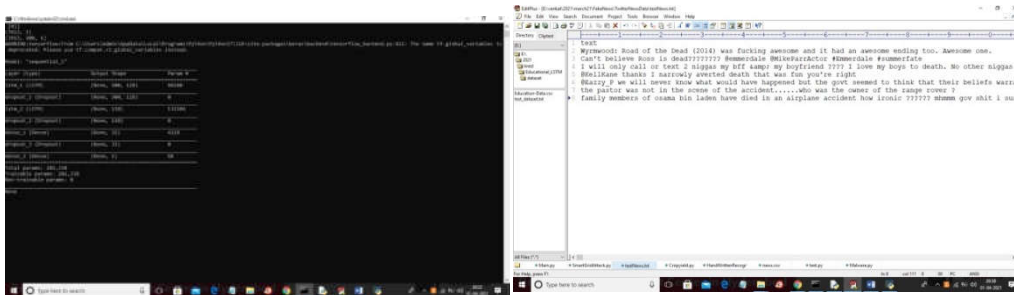


In above screen all news words put in column header and if that word appear in any row then that rows column will be change with word count and if not appear then 0 will be put in column. In above screen showing some records from total 7612 news records and in bottom lines we can see dataset contains total 7613 records and then application using 80% (6090 news records) for training and then using 20% (1523 news records) for testing and now dataset is ready with numeric record and now click on

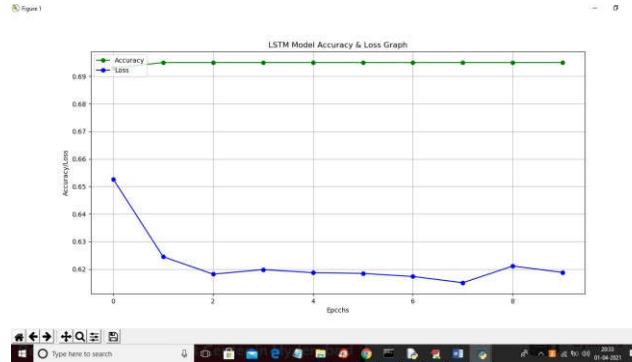
„Run LSTM Algorithm“ button to train above dataset with LSTM and then build LSTM model and then calculate accuracy and error rate



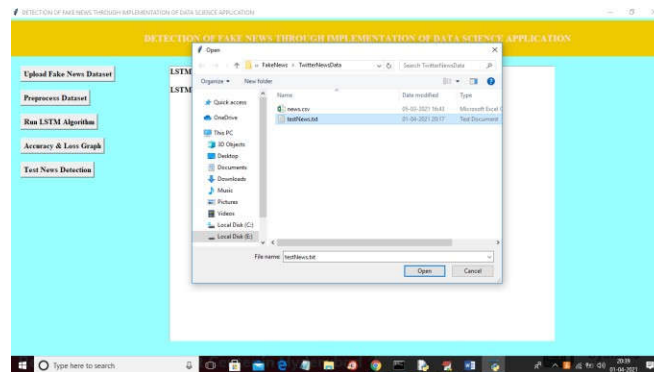
In above screen LSTM model is generated and we got its prediction accuracy as 69.49% and we can see below console to see LSTM layer details



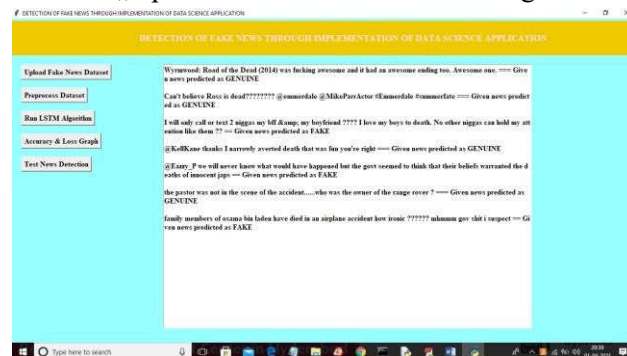
In above screen different LSTM layers are created to filter input data to get efficient features for prediction. Now click on „Accuracy & Loss Graph“ button to get LSTM graph



In above graph x-axis represents epoch/iterations and y-axis represents accuracy and loss value and green line represents accuracy and blue line represents loss value and at each increasing epoch loss values get decrease and accuracy reached to 70%. Now click on „Test News Detection“ button to upload some test news sentences and then application predict whether that news is genuine or fake. In below test news dataset we can see only TEXT data no class label and LSTM will predict class label for that test news. In above screen in test news we have only one column which contains only news „TEXT“ and after applying above test news we will get prediction result



In above screen selecting and uploading „testNews.txt“ file and then click on „Open“ button to load data and to get below prediction result



In above screen before dashed symbols we have news text and after dashed symbol application predict news as „FAKE or GENUINE“. After building model when we gave any news text then LSTM will check whether more words belongs to genuine or fake category and whatever category get more matching percentage then application will predict that class label.

**CONCLUSION:**

This paper presented the results of a study that produced a limited fake news detection system. The work presented herein is novel in this topic domain in that it demonstrates the results of a full-spectrum research project that started with qualitative observations and resulted in a working quantitative model. The work presented in this paper is also promising, because it demonstrates a relatively effective level of machine learning classification for large fake news documents with only one extraction feature. Finally, additional

research and work to identify and build additional fake news classification grammars is ongoing and should yield a more refined classification scheme for both fake news and direct quotes. Future Work The work presented in this paper is also promising, because it demonstrates a relatively effective level of machine learning classification for large fake news documents with only one extraction feature. Finally, additional research and work to identify and build additional fake news classification grammars is ongoing and should yield a more refined classification scheme for both fake news and direct quotes.

## REFERENCES:

1. M. Balmas, "When Fake News Becomes Real: Combined Exposure to Multiple News Sources and Political Attitudes of Inefficacy, Alienation, and Cynicism," *Communic. Res.*, vol. 41, no. 3, pp. 430–454, 2014.
2. C. Silverman and J. Singer-Vine, "Most Americans Who See Fake News Believe It, New Survey Says," *BuzzFeed News*, 06-Dec-2016.
3. P. R. Brewer, D. G. Young, and M. Morreale, "The Impact of Real News about „Fake News“: Intertextual Processes and Political Satire," *Int. J. Public Opin. Res.*, vol. 25, no. 3, 2013.
4. D. Berkowitz and D. A. Schwartz, "Miley, CNN and The Onion," *Journal. Pract.*, vol. 10, no. 1, pp. 1–17, Jan. 2016.
5. C. Kang, "Fake News Onslaught Targets Pizzeria as Nest of Child-Trafficking," *New York Times*, 21-Nov-2016.
6. C. Kang and A. Goldman, "In Washington Pizzeria Attack, Fake News Brought Real Guns," *New York Times*, 05-Dec-2016.