

Phony News Detection & Classification using NLP Approach: A Survey

Sagar Sarvade¹, Shraddha Pawar², Fariha Naz³, Shraddha Tarapurkar⁴

Prof. Vibha Lahane⁵, Prof. Meghana Solanki⁶

¹²³⁴ (Student, Department of Information Technology.)

⁵⁶ (Professor, Department of Information Technology.)

Dr. Dy patil college of Engineering Ambi, Pune, Maharashtra, India.

Abstract – The phony News Detection topic has gained a great deal of interest from researchers around the world. When some event has occurred, many people discuss it on the web through social networking. Very few people know the real fact of the event while most people believe the forwarded news from their credible friends or relatives. There are a lot of news articles, where the news is false or modified according to the individuals mind. Such type of activities are not good for the society where some rumors or vague news evaporates the negative thought among the people or a specific category of people. As fast the technology is moving, at the same pace the preventive measures are required to deal with such activities. So, there is a need for an automated system to analyze the truthfulness of the news.

This paper gives a survey of different techniques used by the researcher to determine whether the news is real or fake.

Keywords: Phony News Detection, Machine Learning techniques, negative effects, Natural Language Processing, online social Media, Classification.

I. INTRODUCTION

The large use of social media has a tremendous impact on our society, culture, business with potentially positive and negative effects. Nowadays, due to the increase in the use of online social networks, fake news for various commercial and political purposes has been emerging in large numbers and widely spread in the online world. People immediately start expressing their concerns or sharing their opinion as soon as they come across a post, without verifying its authenticity. These further results in the spreading of it. This fake news spreads like wildfire and is impacting millions of people every day. There are a lot of news articles, where the news is false or cooked up. It's a prevalent and pressing issue that needs to be acknowledged. Therefore, it is essential to make effort to identify these facts and avoid them from spreading.

News sources like web-based life channels, news sites, and online papers have created demanding to verify reliable news sources due to the enhancement of misleading information. Therefore, detecting fake news has become a crucial problem in attracting tremendous research effort

With so many advances in Natural Language Processing and machine learning, it is possible to detect if a piece of the news article is genuine or fake.

II. LITERATURE REVIEW

A. Machine Learning Approaches

Abdullah-All-Tanvir et al. [1] propose a model for recognizing forged news messages from twitter posts, by figuring out how to anticipate precision appraisals using Twitter datasets. A comparison of five algorithms like Support Vector Machine, Naïve Bayes Method, Logistic Regression, and Recurrent Neural Network models is done to demonstrate the efficiency of the classification performance on the dataset. SVM and Naïve Bayes classifier outperform the other algorithms.

Aswini Thota et al [3] introduced an automated detection system for fake news detection he said is difficult work to complete as it needs representation to recognize nuances in natural language. To address the gaps between related and unrelated news, they introduce neural network architecture to exactly guess the posture that is to be provided headline and article body

Nicollas R et al. [4] propose a computational stylistic analysis based on natural language processing, efficiently applying machine learning algorithms like SVM to detect fake news in texts extracted from social media. The analysis considers news from Twitter, from which approximately 33,000 tweets were collected, assorted between real and proven false. In assessing the quality of detection, 86% accuracy, and 94% precision stand out even employing a dimensional reduction to one-sixth of the number of original features. The database's composition includes both real and fake news, collected from specific Twitter accounts.

ShuoYang et al. [5] presented an unsupervised approach to detect fake news. The Bayesian network model is used to capture the conditional dependencies among the truths of news. an efficient collapsed Gibbs sampling approach is used to infer the truths of news and the users' credibility without any labeled data. The results show that the proposed method significantly outperforms the compared unsupervised methods. The key idea is to extract the users' opinions on the news by exploiting the auxiliary information of the users' engagements with the news tweets on social media and aggregate their opinions in a well-designed unsupervised way to generate our estimation results.

Veronica Perez-Rosas et al. [6] focus on the automatic identification of fake content in online news. For that two novel datasets are constructed here covering seven different news domains and finally accurate fake news detector model is build using a linear SVM classifier and five-fold cross-validation. One dataset is collected via crowdsourcing, and covers six news domains; the second dataset is obtained directly from the web and covers celebrity fake news.

Bashar Al Asaad et al. [7] presented a supervised learning-based approach for the detection of fake news. a dataset of fake and real news is used to train a machine learning model using the Scikit-learn library in Python. Features extractions of the text representation model are done using Bag-of-Words, Term Frequency-Inverse Document Frequency (TF-IDF), and Bi-gram frequency. Finally, probabilistic classification and linear classification is used on the title and the content, if the article is fake/real.

B. Deep Learning Techniques

Chaitra K Hiramath et al.[2] proposed system to detect fake news based on classification techniques such as Logistic regression (LR), Naïve Bayes (NB), Support vector machine (SVM), Random Forest (RF), and deep neural network (DNN). Results give that the DNN algorithm is more efficient in terms of accuracy and time kind because rest classifiers require more time and give less accuracy hence DNN is more crucial to detect the fake news.

Terry Traylor et al.[8] presented a model for checking the verification of news extracted from Twitter. The dataset from Twitter threads that mainly includes the information about the Chile earthquake 2010 is fused for the experiment. The fake and real classification is done using the Bayesian Model, Logistic Regression & also Support Vector Machine, Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM). Among which SVM is outperformed that gives 0.94% accuracy for news classification.

Belhakimi et al.[9] propose a merged deep learning model that detects fake articles regarding different characteristics. Word embedding technique and convolutional neural network are used to extract text-based features and compare the different architecture of deep learning while merging two CNNs with different metadata (Text, title, and author). The dataset from kaggle.com is used to perform the experiments. The dataset is divided into 90 % for training and 10 % for testing. The proposed approach is very efficient and allows achieving high performances i.e. highest accuracy score is 96%.

Chandra et al.[10] successfully handled the missing values problem by using data imputation for both categorical and numerical features. For categorical features, missing values are imputed with the most frequent value in the columns. For numerical features, the mean value of the column is used to impute numerical missing values. Besides, TF-IDF vectorization is applied in feature extraction to filter out irrelevant features. Experimental results show that Multi-Layer Perceptron (MLP) classifier with the proposed data preprocessing method outperforms baselines and improves prediction accuracy by more than 15%.

Kushal et al.[11] propose a pipeline that combines processing, feature extraction, and model fusion for a more accurate and automated prediction. Especially the fusion of latent semantic analysis (LSA) and ensemble learning model results using stacking. Experimental analysis of real-world data demonstrates that our pipeline achieves higher accuracy than existing approaches.

III. PROPOSED SYSTEM

Fake news detection is made to stop the rumors that are being spread through the various platforms whether it be social media or messaging platforms. Fake news detection works on the objective of detecting this fake news and stopping activities like this thereby protecting the society from these unwanted acts of violence. It's a prevalent and pressing issue that needs need to be taken care at the earliest. Here an attempt is made to detect whether a piece of news is fake based on data sources.

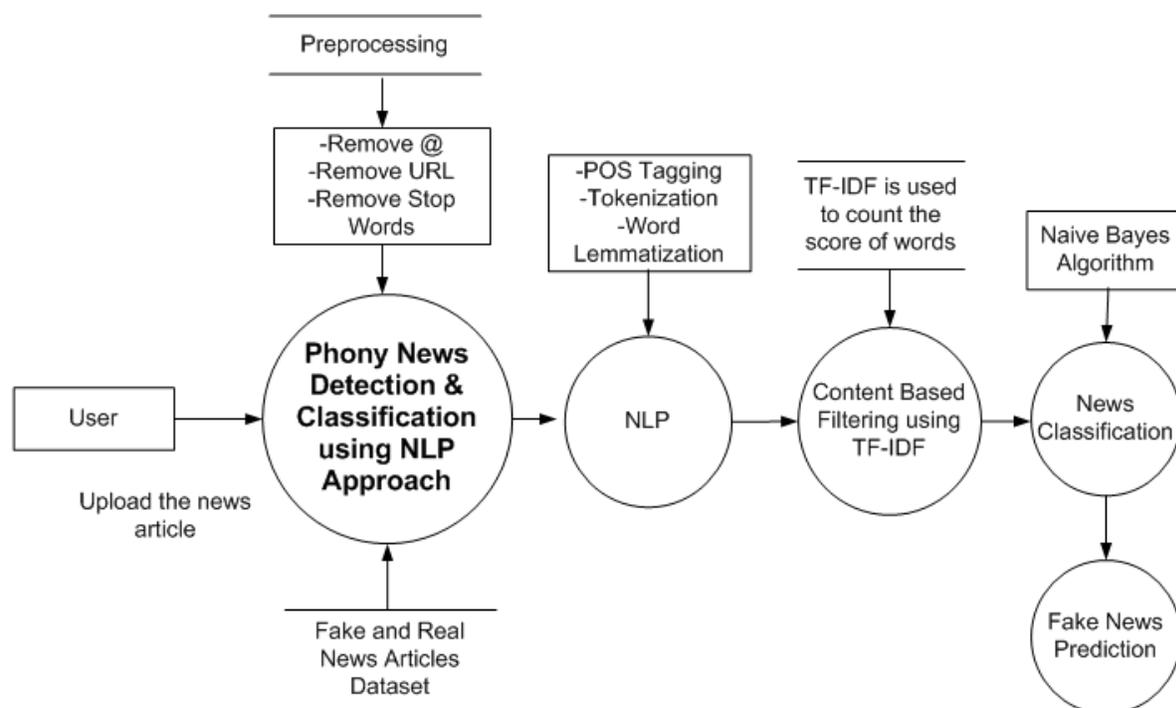


Figure 1: System Architecture

The System extracts the information from the news article. Then Natural language processing (NLP) technologies are used for parsing, tokenizing, stemming and filtering the content of the data. By using TF-IDF is an approach that utilizes the counts of words appearing in the documents to figure out the similarity between documents. News classification as fake news is done by using a Naive Bayes classification algorithm.

VI. CONCLUSION

The main intension of spreading fake news is to trigger people's distrust and make them confused; impeding their abilities to differentiate what is true from what is not. To help mitigate the negative effects caused by fake news and benefit both the public and the news ecosystem, It's crucial to develop a system that automatically detects fake news on social media.

Most existing methods of fake news detection require an extensive amount of time and labor to build a reliably annotated dataset.

Here we are trying to build a classifier that can detect whether a piece of news is fake based on data sources using the Naive Bayes classification algorithm.

REFERENCES

- [1]Abdullah-All-Tanvir ; Ehesas Mia Mahir ; Saima Akhter ; Mohammad Rezwanaul Huq , "Detecting Fake News using Machine Learning and Deep Learning Algorithms", 2019 7th International Conference on Smart Computing & Communications (ICSCC).
- [2] Chaitra K Hiramath, Prof. G.C Deshpande, "Fake News Detection Using Deep Learning Techniques", 2019 1st International Conference on Advances in Information Technology.
- [3]Aswini Thota¹, Priyanka Tilak¹, Simeratjeet Ahluwalia¹, Nibhrat Lohia¹ ,” Fake News Detection: A Deep Learning Approach ” SMU Data Science Review, Vol. 1 , No. 3, Art. 10.
- [4]Nicollas R. de Oliveira, Dianne S. V. Medeiros and Diogo M. F. Mattos, Member, IEEE , "A Sensitive Stylistic Approach to Identify Fake News on Social Networking " ,IEEE SIGNAL PROCESSING LETTERS,2020.
- [5]ShuoYang, KaiShu, SuhangWang, RenjieGu, FanWu, HuanLiu, "Unsupervised Fake News Detection on Social Media: A Generative Approach ", TheThirty-ThirdAAAIConferenceonArtificialIntelligence(AAAI-19).
- [6]Verónica Pérez-Rosas¹, Bennett Kleinberg², Alexandra Lefevre¹Rada Mihalcea, "Automatic Detection of Fake News", 23 September 2017.
- [7]Bashar Al Asaad, Madalina Erascu, "A Tool for Fake News Detection", Conference: 2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC).
- [8]Terry Traylor ; Jeremy Straub ; Gurmeet ; Nicholas Snell , "Classifying Fake News Articles Using Natural Language Processing to Identify In-Article Attribution as a Supervised Learning Estimator", 2019 IEEE 13th International Conference on Semantic Computing (ICSC).
- [9]Belhakimi Mohamed Amine; Ahlem Drif ; Silvia Giordano , "Merging deep learning model for fake news detection", 2019 International Conference on Advanced Electrical Engineering (ICAEE).
- [10]Neelam Majgaonkar, Ruhina Hodekar, Priyanka Bandagale, "Fake News Detection Enhancement with Data Imputation "IEEE 2018.
- [11]Abhishek S. Parab, Amol Joglekar, "Implementation of Home Security System using GSM module and Microcontroller", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015, 2950-2953.