



Fake News Detection And Classification Using Distinct Machine Learning Algorithms

¹ Himanshu Kiran Garud, ² Apurva Avinash Kombde

¹Student, ²Student

^{1,2}Computer Engineering,

^{1,2}K.K. Wagh Institute of Engineering, Education and Research, Nashik, India

Abstract: Social media interaction and online news is becoming popular nowadays. The spread of fake news and hoaxes are becoming a major point of concern as very few people actually check the facts and just continue sharing news articles. At times, this results in serious consequences like social pressure on particular topics and affects the public's point of view on certain problems. Therefore identifying fake news has become a major subject of study in the field of science. This paper aims to discuss various machine learning algorithms for the detection and classification of fake news.

Index Terms - Fake news detection, Social media, Classification, Machine Learning, Deep Learning, Natural Language Processing.

I. INTRODUCTION

Due to advancement in technology and the internet, access to news information is much easier and user friendly. Social media has become the main source of news online with more than 2.4 billion internet users, nearly 64.5 percent receive breaking news from Facebook, Twitter, YouTube, Snapchat, and Instagram instead of traditional media [7]. A major part of such news articles, that people read online particularly on social media platforms, appear to be true but are not. Fake news is any misleading information, story or hoax created particularly to deceive users. People or organizations intentionally spread this news for personal benefits. Fake news on social media raises havoc and causes damage in no time. One example of such a situation is the coronavirus pandemic era. When the entire world was busy battling with COVID-19, many news articles involving home remedies, vaccinations, etc were doing rounds on various social media platforms.

The internet and social media have become an important source of fake news. The users continue to deal with sites containing false information and whose participation tends to affect the reader's ability to seize with real news. The subsequent development of social media platforms like Facebook, Whatsapp, Twitter allows for efficient and fast sharing of information, and thus, users can share false information within a short period. Most of the sites that contain such information also have a sharing option which leads to mass forwarding of misleading news.

Much artificial intelligence and machine learning algorithms are proved to be effective and powerful in identifying and classifying such fake news. This is due to the recent improvement of machine learning algorithms in classification problems, availability of cheaper hardware, and a large number of available datasets.

This paper aims to study machine learning algorithms like the Naive Bayes classifier, Logistic Regression, Support Vector Machine (SVM), Convolutional Neural Network(CNN), Recurrent Neural Network(RNN), etc. These algorithms showed good and accurate results in fake news detection.

II. LITERATURE SURVEY

Muhammed Umer, Zainab Imtiaz, Salem Ullah, Arif Mehmood, Gyu Sang Choi, Byung-won on [1], proposed to use of a hybrid neural network architecture that integrates the proficiency of CNN and LSTM with two separate dimensionality reduction approaches, Chi-Square and Principle Component Analysis (PCA). The dataset utilized is from the Fake News Challenges website consisting of 4 types of thoughts: agree, disagree, unrelated, and discuss. The idea was to reduce the dimensionality of feature vectors, utilizing dimensionality reduction techniques, before passing them to the classifier. Initially, the non-reduced feature set, with and without preprocessing to the neural network, is passed. Then the dimensionality reduction techniques are applied and eventually the results are compared. The performance evaluation metrics involved: accuracy, precision, recall, and F1-score. The results rendered that PCA was more accurate (97.8%) than Chi-Square(91.49%).

Dr. Priyanka Harjule, Akshat Sharma, Sachin Chauhan, Shashank Joshi [2], studied 2 different models for the detection of fake news.

The first one being text-based classification where the model is trained on refined data for 4 models namely, Support Vector Machine(SVM), Naive Bayes Classifier, CNN, and RNN. For this model, a dataset from Kaggle is taken into consideration. The results revealed that RNN gave the best accuracy of 93% and proved to be most efficient as compared to others for the same testing data set.

The second is done by crowd analysis where 2 methods were taken into consideration namely, the Probability method and Parameter Tuning method. Here, the Twitter dataset is being used. The Parameter tuning method resulted in more accuracy of 80% and recognized fake conclusions much better.

Kushal Agarwalla, Shubham Nandan, Varun Anil Nair, D. Deva Hema [3], observed and presented 3 models to detect fake news. These models include Support Vector Machine(SVM), Logistic regression, and Naive Bayes classifier with Lidstone smoothing.

The model which resulted in the highest accuracy is the Naive Bayes classifier (with Lidstone smoothing) with an accuracy of 83%. It was followed by Support Vector Machine(SVM) with an accuracy of 81.65% and the Logistic regression with the lowest accuracy of 65.88%.

The Naive Bayes classifier with Lidstone smoothing determined the fake news articles based on spelling mistakes, jumbled sentences, punctuation errors, etc. The dataset used is from kaggle.com. The size of the dataset is 4008*4 where 4008 is the number of rows and 4 is the number of columns. The dataset consists of 2136 fake news articles and 1872 genuine news articles.

Abdullah All Tanvir, Ehsas Mia Mahir, Saima Akhter, Mohammad Rezwanal Huq [4], investigated 5 models for forged news messages detection. The work was majorly done on Twitter datasets especially involving the information about the Chile earthquake 2010. One important observation done in psychological research said that word lengths in tweets can be an important feature as fake news involves a lot of subjects, words, and fictional declarations. The classification models executed using the datasets were the Bayesian model, Logistic Regression, and SVM. Also, 2 deep learning methods RNN and LSTM were implemented. To process the datasets three ideas - Count Vectors, TF-IDF (Word level, N-gram level, Character level), Word embedding are considered.

After implementing inter and intra comparison on the above mentioned 5 classification models, the accuracy percentages were 89.06% for Naive Bayes, 89.34% for SVM, 69.47% for logistic regression, 74% for RNN, and 78% for LSTM. Later, the precision, recall, and F1 score were calculated for machine learning algorithms. The precision result is the same for all three of them. SVM outperformed considering recall and F1 score of both Naive Bayes and SVM is high. Therefore, from the investigation SVM proved to perform best for characterization techniques among all the other 5 models used.

Shenhao Zhang, Yihui Wang, Chengxiang Tan [5], put forward a model to identify the fake news on text classification. This model used TF-IDF (Term Frequency-Inverse Documentation Frequency), Word2vec, and Word2vec weighted by TF-IDF. The dataset is taken from some websites and data provided by a research institute. There are 5,581 news items out of which 2,878 are the real news and 2,703 fake news, retaining categories and texts.

The TF-IDF and Word2vec model showed a good performance in classifying real and fake news but the model Word2vec weighted by TF-IDF resulted in higher accuracy than the two models. The effectiveness of the algorithms is measured by precision rate, recall rate, and F1 score. The precision rate of fake news by word2vec is 4.31% higher than the Tf-IDF algorithm and an average F1% score of word2vec is 4.25% higher than TF-IDF.

The precision rate of real news, the Word2vec weighted by TFIDF is 3.11% higher than Word2vec, the average value of P, R, and F1 based on Word2vec weighted by TF-IDF is higher than TF-IDF and Word2vec. This showed the effectiveness of Word2vec weighted by TF-IDF algorithm in identifying the fake news.

Mykhailo Granik, Volodymyr Mesyura [6], suggested a model that detects the fake news using a Naive Bayes classifier. The Naive Bayes classifier was implemented as a software system. The dataset used for learning and testing the naive Bayes classifier contains information about various Facebook posts and each post represents a news article. This is a simple artificial intelligence model that gives an accuracy of 74%.

III. CONCLUSION

The research depicted that even quite easy machine learning algorithms may show a good result on such a crucial issue as the spread of fake news worldwide. The various machine learning and deep learning algorithms prove to be efficient and fast for the identification and classification of fraudulent news articles from authentic ones. Thus, the results of this research show machine learning techniques can be successfully applied to tackle this cardinal problem.

IV. ACKNOWLEDGMENT

With a deep sense of gratitude, I wish to express my sincere thanks to our parents for their immense help in planning and motivating the works. With a sincere feeling of appreciation, I wish to communicate my earnest gratitude to our folks for their tremendous assistance in arranging and executing the works.

REFERENCES

- [1] Muhammed Umer, Zainab Imtiaz, Salem Ullah, Arif Mehmood, Gyu Sang Choi, Byung-won on, “Fake News Stance Detection Using Deep Learning Architecture(CNN-LSTM)”, IEEE Access, September 9, 2020.
- [2] Dr. Priyanka Harjule, Akshat Sharma, Sachin Chauhan, Shashank Joshi, “Reliability Of News”, 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE-2020), February 07-08, 2020.
- [3] Kushal Agarwalla, Shubham Nandan, Varun Anil Nair, D. Deva Hema, “Fake News Detection using Machine Learning and Natural Language Processing”, International Journal of Recent Technology and Engineering (IJRTE), Volume-7, Issue-6, March 2019.
- [4] Abdullah All Tanvir, Ehsas Mia Mahir, Saima Akhter, Mohammad Rezwana Huq, “Detecting Fake News using Machine Learning and Deep Learning Algorithms”, 2019 7th International Conference on Smart Computing & Communications (ICSCC).
- [5] Shenhao Zhang, Yihui Wang, Chengxiang Tan, “Research on Text Classification for Identifying Fake News ”, 2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC).
- [6] Mykhailo Granik, Volodymyr Mesyura, “Fake News Detection Using Naive Bayes Classifier”, 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON).
- [7] How Social Media Has Changed How We Consume News [Online]. Available: <http://www.Forbes.com/>(visited on 07/01/2021)

