International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 3314–3317

# News and Information Credibility Analysis

Manish Ambre 1 , Pranit Borkar 2 , Neha Ban 3 , and Mayur Gore 4

*1Student, PVPIT, Computer Engineering Department*

*2Student, PVPIT, Computer Engineering Department*

*3Student, PVPIT, Computer Engineering Department*

*4Student, PVPIT, Computer Engineering Department*

***Abstract***

*In today's social networking world, people use their social media accounts not just for connecting with each other but as a source of news and information. News and information is available onvarious sources on the Web for free and easily, which creates problems regarding the credibility of a news item or an article. News is shared over social media very quickly making the content sharedviral and difficult to stop from spreading if it is fake or dishonest in any way.In this paper, we discuss a system that uses various Natural Language Processing techniques andMachine Learning Classification algorithms to predict and classify Fake news articles or headlinesusing the Sci-Kit libraries. We also aim to use a custom made news data-set that will be created by scraping news data from several sources on the internet.*

***Keywords: Fake News, Machine Learning, Natural Language Processing, Web Scrapping***

## 1. Introduction

Fake News has become an important problem considering the increasing use of the internet messaging and social media platforms like WhatsApp and Twitter to read, write and share news articles all over the world quickly. It has become very difficult to assess the credibility of a news article for common people who often fall prey to fake news and its significant consequences such as panic, distress, etc. in the society.

## 2. Literature Review

In the paper[1], 'Fake News Detection Using A Deep Neural Network', author Rohit Kumar Kaliyar discusses using Deep Neural networks to detect fake news. In this project, the author explored different Machine learning models such as Naive Bayes, K nearest neighbors, Decision tree, Random forest and several Deep Learning networks such as Shallow Convolutional Neural Networks (CNN), Very Deep Convolutional Neural Network (VDCNN), Long Short-Term Memory Network (LSTM), Gated Recurrent Unit Network (GRU), etc.

In the paper[2], 'Fake News Pattern Recognition using Linguistic Analysis', authors Amitabha Dey, Rafsan Zani Rafi, Shahriar Hasan Parash, Sauvik Kundu Arko, Amitabha Chakrabarty discuss using Linguistic Analysis to detect fake news. The authors proposed a general framework that can identify an author's bias. They used a corpus of 200 tweets on 'Hilary Clinton', while performing veracity assessment, 'text normalization' on tweets, and explored techniques for feature extraction to classify news into categories. They performed a comprehensive linguistic analysis on the corpus of tweets, extracting a bag-



ISSN: 2233-7857 IJFGCN

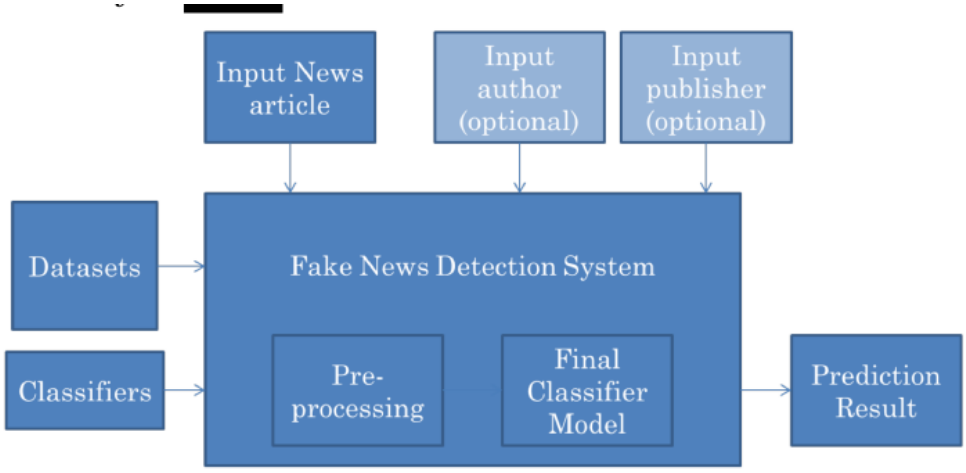
of-words to find a pattern, to apply the k-nearest neighbor algorithm for classifying polarized or biased news from credible news.

In the paper [3], 'Fake news detection using naive Bayes classifier', authors Mykhailo Granik, VolodymyrMesyura discuss using the Naive-Bayes Classifier to detect fake news.This paper shows a simple approach for fake news detection using the naive Bayes classifier. It was implemented as a software system and tested against a corpus of Facebook news posts. They achieved classification accuracy of approximately 74% on the test set.

## 3. Proposed System

To develop a Machine Learning Program to predict and classify fake news from real news. We aim to use a custom created corpus of labeled real and fake new articles scraped from several Digital news sources to build a Classifier that can predict the credibility of an input news article. The classification model will focus on identifying fake news. Once a source is labeled as a producer of fake news, we can predict with higher confidence that any future articles from that source will have a better chance of being fake news. The proposed application of the project is for use in applying visibility weights in social media so that they can better filter news floating on the platforms, as well as guide an end-user in identifying fake news quickly and easily

### 3.1 System Architecture




In the proposed system, we have following components what work together to detect a fake news using Machine Learning:
1) Input: The Fake news detection system requires an input news article to be analyzed. User can input a news article in text format and a thorough news article is expected. User can also input additional details about the article, namely the author name and the publisher. These additional details will help in assessing the input article more accurately and will also help in identifying fake news sources.The data set used by our system also has additional details about articles which include author name, publisher name, etc.

2) Fake News Detection System: This component encompasses the Machine Learning and Natural Language Processing modules that process the input article to successfully generate a prediction of whether the input article is fake or not along with the probability



ISSN: 2233-7857 IJFGCN

figures.

i)Pre-processing: This module contains the Natural Language Processing and feature extraction processes. The input article is in Natural language and needs to be processed into necessary format, I .extracted keywords and features. Several sub processes such as Tokenization, Removing stop words, lemmatization, etc. are performed over the input in this component.

ii) Final Classifier Model: This component refers to the final selected classifier model that will be applied on the input article to predict its credibility. This ”final” classifier model is chosen on the basis of the performance of the various classifiers used in our system on the training set and test set. This final classifier has the best f-score of all the classifiers when trained and tested by the data-set used in the
creation of this project.

3) Data-sets: This component represents the data-set that will be used to train and test the Machine Learning system. The data-set will be divided into a training set (to train the system) and a test set(to check and evaluate its performance). The data-sets used are in .csv format, i.e. Comma Separated Values format.

4) Classifiers: This component represents the various classifiers used for this project, i.e. Naive-Bayes classifier, Support Vector Machine classifier, Random Forest classifier and Logistic Regression classifier.
These classifiers will each process the training data-set and test data-set and their f-score performance and confusion matrix will decide which of them will be used as the final classifier for custom user input.
5) Prediction Result: The result of the system will be either ”Fake” or ”Real” and a probabilistic estimate of a user input article of being fake or real will be provided. This will guide the user to understand and classify fake news.

## 4. Results

We compared our models using their Confusion Matrices to calculate various performance metrics such as Precision, Recall and the F1 scores. Following table shows our results.

| Model | Precision | Recall | $F^1 Score$ |
|---|---|---|---|
| Naive Bayes | 0.68 | 0.88 | 0.75 |
| Logistic Regression | 0.73 | 0.83 | 0.77 |
| SVM | 0.86 | 0.91 | 0.88 |
| Random Forest | 0.85 | 0.94 | 0.91 |

We observed that SVM and Random Forest Models gave us the best results. The RF model achieves the highest F1 score in comparison to all the other models, followed by the SVM model.

## 5. Conclusion

A complete, production-quality classier will use many different features in addition to the vectors corresponding to the words in the news text. For fake news detection, we can



ISSN: 2233-7857 IJFGCN

add use features such as the source of the news, including any associated URLs, the topic or genre of news, place of origin, publishing medium (social media, blogs ,print, etc.), publication year, as well as linguistic features such as the use of proper nouns, capitalization, etc. Besides we can also incorporate other more sophisticated AI approaches such as Neural Networks, LSTM, etc.

## Acknowledgments

We are thankful and would like express our gratitude those who supported, helped us to prepare this paper. Special thanks to our guide who helped us to complete this paper hassle free. Your guidance was precious and your deep knowledge in this subject helped us in understanding this topic multidimensional. Also we thank many well-wishers who helped us to complete this paper.

## Reference

[1] .Sohan Mone, Devyani Choudhary, Ayush Singhania. Fake news Identification. Stanford University,2017.

[2] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake News Detection on SocialMedia: A Data Mining Perspective. 2017.

[3] Conroy, N. J., Rubin, V. L., Chen, Y. Automatic deception detection: Methods for finding fake news.2015, Proceedings of the Association for Information Science and Technology, 52(1), 1–4.

[4] Kaggle Fake News NLP Stuff. https://www.kaggle.com/rksriram312/fake-news-nlp-stuff/notebook



ISSN: 2233-7857 IJFGCN