

Exploring the Performance of Algorithms That Are Dedicated To the Detection of Fake News around the Globe Using Machine Learning

G Naga Kumar Kakarla

Associate Professor, Department of Computer Science and Engineering, Malla Reddy College of Engineering for Women, Hyderabad -500100, Telangana, India

Abstract - Due to the COVID-19 pandemic, several health and economic challenges come to play awkwardly. This has introduced misinformation and confusion around the globe. The issues of fake news have attained an increasing eminence in the diffusion of shaping news stories. Many of them stop to depend on the newspapers, magazines, etc and started to rely on social media completely. Social media became the main news source for millions of people due to their easy access, cheap, more attractive and rapid dissemination. The fake content started to spread at a large pace to gain popularity over social media to distract people from the current critical issues, in some occasions spreading more and faster than the true information. People spread fake news on social media for financial and political gain. Fake data in all forms need to be detected as soon as possible to avoid a negative impact on society. This project makes an analysis of the research related to fake news detection; we trained and tested different machine learning algorithms separately to demonstrate the efficiency of the classification on the dataset. This project was implemented in the Jupiter notebook platform and performance was evaluated.

Keywords: SVM, PAC, Fake News, Social Media, Machine Learning, Classifiers.

1. INTRODUCTION

Social media has been in our lives for centuries and has reached even in remote villages. Even though social media has made life in the view of interacting with people, some people spreading and posting fake news has been a major problem for the past few decades. 90% of the population depends on social media for their news reading because of the availability of the internet and the use of smart devices. Facebook and Google are constantly taking measures considering these issues. For example finding out fake news by flagging them as fake, use of hoax sites, fact-checking labels etc. These techniques have not yet gained their purpose, that is why people need to be aware of what to believe and not believe, even though the line between the true and fake is thin, moreover the spreading rate of these fake news is faster which give greater obstacle to predicting their credibility. There arises a need for fake news detection. The motive of this publication is to reach a solution that can be used by people to identify and scrutinize the websites that contain false and misleading information. Natural language processing is a part of artificial intelligence (AI), which comprises techniques that can use text, create models and algorithm which helps in prediction. This work aims to create a model that can use the information or data of the past or present news reports and predict whether the news is fake or not. This project demonstrated the ability of machine learning and AI to be useful for this task [1]. These machine learning techniques are applied with the compilation of Natural language processing's feature extraction method. Analyzing the performance of each technique is done which also helps to see their accuracy. This machine learning algorithm helps to train system to predict the credibility and reliability based on the text, words used and stop word, etc.

Our work is on text-based fake news where we have used three datasets to analyze the best detector. We carry out the process through machine learning algorithms. Three publically available datasets were chosen from the websites. The classification of fake or real news was done using the framework. Datasets were trained using natural language processing techniques for feature extraction to acquire the best accuracy. After the processing classification on the algorithms done and the performance was noted down. The best method is selected based on accuracy, f1-score, recall and precision [4].

2. LITERATURE REVIEW

The article by Monther Aldwairi et. al. [10] comes up with a solution that can be utilized by users to detect and filter out sites containing false and misleading information. Click baits are phrases that are designed to attract the attention of a user who, upon

clicking on the link, is directed to a web page whose content is considerably below their expectations. This leads to irritation and waste of time by the user. The solution includes the use of a tool that can identify and remove fake sites from the results provided to a user by a search engine or a social media news feed. These tools can be directly downloaded and installed by the user in their system.

Farzana Islam et. al.[7] in their paper proposed a model for detecting fake news in the Bengali language. This work is done on fake news classification in the context of Bangladesh and South Asia. They have used data mining algorithms as classifiers. A Bengali newspaper scrapper was developed to create a Bengali news dataset. Text mining is used to create a new corpus dataset. Word cloud is shown as a part of data visualization. Experiments are done with varied features and models. This project is creating an end-to-end pipeline of data collection, ingestion, and web-based demonstration of fake news classification along with visualization. In S D Samantaray's et. al. [8] work the proposed system is divided into two subparts first is text analysis and then performance evaluation. Text analysis is done for the transformation of text into numerical features. The found out numerical features are used for matching the similarity between queried articles and other articles.

3. DATASET

The first step included in this work was to find the dataset that can be used to achieve the goal. News data can be gathered by Expert journalists, Fact-checking websites, Industry detectors, and Crowd-sourced workers. News datasets for our work have been found out from Kaggle. These datasets were used in different research papers for determining the specification of news. Three datasets have been used. Real-world data are incomplete, inconsistent or non-relatable, and are likely to contain many errors. So we had done some error corrections to carry out the classification. After the dataset has been imported data pre-processing is done to get better results through the classification algorithms. Data pre-processing like removing stop words, count vectors, TF-IDF vectorization, word embedding etc is also done on the dataset. The TF-IDF vectorization converts the text to number format which helps to fit the machine learning algorithms easily. In the inputted dataset, no missing values are there and the input dataset will be tokenized. The tokenized dataset will be again processed and unwanted information will be removed from the dataset.

4. SYSTEM OVERVIEW

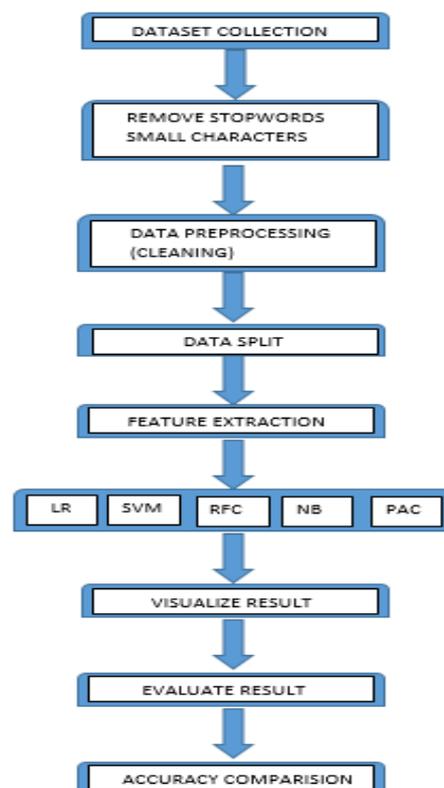


Fig.1. Overview of work

5. EXPERIMENTAL RESULT

As there is an increase in social media users, the spread of fake news also increased the trice. People are getting distracted and confused most of the time because of fake news. Experiments are going on to resolve this problem. As a contribution, we dedicate this work towards fake news detection. In this work, we have used three the dataset data dataset which was collected from a public source. Along with the well-known machine we had learning we evaluated the performance of five algorithms.

	LR	SVM	RFC	PAC	NB
DATASET-1	98.72	99.46	98.87	99.54	94.11
DATASET-2	89.92	92.78	89.67	92.65	83.09
DATASET-3	97.21	98.8	96.71	99.0	94.32

The comparative study was carried out by looking at the accuracy value. Depending on the accuracy the best model was selected. As the initial stage, the dataset was divide into training and t t testing datasets. The training data set is used to fit the model and the prediction is done on the test data set. The performance measure of five different machine learning methods is shown in the following tables.

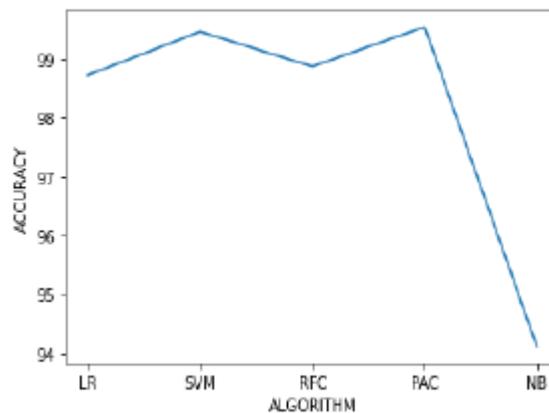


Fig.3. Accuracy comparison on DATASET-1

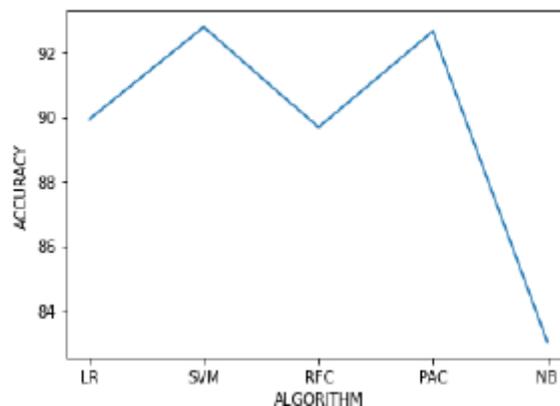


Fig.4. Accuracy comparison on DATASET-2

6. CONCLUSIONS

In this paper, we have examined and explored the performance of five algorithms that are dedicated to the detection of fake news. From our critical analysis, we have reached an interpretation that the Naïve Bayes classifier has shown the least performance when comparing to the other four models. This paper has also revealed the accuracy of each model in identifying fake news. As

mentioned earlier the use of social media has been spread vastly, this research can be used as a skeleton for other investigators to interpret which models are precisely and accurately completing its mission in identifying fake news. However it is to be highlighted that we have some ways or mechanisms for the detection of fake news, or a way to aware people to know that everything is they read is not true, so we need critical thinking and evaluation. In that way, we can help people to make choices so that they won't be tricked or fooled into thinking what others want to guide or exploit into our thoughts.

REFERENCES

- [1] V. Agarwal, "Analysis of Classifiers for Fake News Detection," ICRTAC, pp. 377-383, 2019.
- [2] F. Islam, "Bengali Fake News Detection," IEEE, pp. 281-287, 01 October 2020.
- [3] S. D. Samantaray, "Fake News Detection using Text Similarity Approach," IJSR, vol. 08, no. 01, pp. 1126-1132, January 2019.
- [4] Vogel, "Detecting Fake News Spreaders on Twitter from a Multilingual Perspective," IEEE, pp. 599-606, 22 December 2020.
- [5] M. Aldwairi, "Detecting Fake News in Social Media Networks," EUSPN, pp. 215-222, 2018.
- [6] K. ArunKumar, "A Study of Fake News Detection using Machine Learning Algorithms," IJTES, vol. 11, no. 01, pp. 1-7, 2020.
- [7] P. B. Petkar, "Fake News Detection: A Survey of Techniques," IJITEE, vol. 09, no. 09, pp. 383-386, July 2020.
- [8] U. Sharma, "Fake News Detection Using Machine Learning Algorithms," IJCRT, vol. 08, no. 06, pp.1394-1402, 6 June 2020.
- [9] S. Aphiwongsophon and P. Chongstitvatana, "Detecting Fake News with Machine Learning Method".
- [10] K. Agarwalla, "Fake News Detection using Machine Learning and Natural Language Processing," IJRTE, vol. 07, no. 06, pp. 844-847, March 2019.
