



## Fake News Detection Using Machine Learning Techniques

<sup>1</sup>S. Leishipem Sophia, <sup>2</sup>Dr. Rajesh Dharmaraj

<sup>1</sup> Student, Department of Master of Computer Applications, Jyoti Nivas College, Bangalore, Karnataka, India

<sup>2</sup>Assistant Professor, Department of Master of Computer Applications, Jyoti Nivas College, Bangalore, Karnataka, India

### ABSTRACT

*Fabricated misinformation is designed to mislead the reader and viewers, it has been a challenging problem and can have negative impacts on the atmosphere. It has become a research challenge for categorizing it as false or true. The following papers reviews various machine learning approaches and techniques to identify the current problem and apply solutions to solve it to an extent and avoid it from happening. A comparison is made to show that the latter is more reliable and effective in terms of determining all kinds of news. It uses various techniques and algorithms.*

**Keywords:** Machine learning techniques, Fake news, NB classifier, PAC, confusion matrix, MNB.

### I. INTRODUCTION

In our modern world there is a vast contribution of the Internet Technology for transmission and internet sharing. The growth of continuous use of media, internet etc. has made it difficult to distinguish between false and true information. An evolution of human history and fake news accumulates a great deal of attention over the internet especially on social media that is spread in order to attract the audience, harm someone's reputation or false advertise and many more. It is a form of news consisting of deliberate disinformation or hoaxes spread via traditional news media or online social media that has caused many consequences. It has become a research challenge for categorizing it as false or true. The following papers reviews various machine learning approaches and techniques to identify the current problem and apply solutions to solve it to an extent and avoid it from happening. Machine learning is one of the recent technologies to recognize and eliminate the Fake news with the help of Algorithms. Methods are employed to detect the credibility of news based on the text content and responses given by users. A comparison is made to show that the latter is more reliable and effective in terms of determining all kinds of news. It uses frequency-based features to train the Algorithms including Support Vector Machine, confusion matrix, Passive Aggressive Classifier, Multinomial Naïve Bayes, Logistic Regression and Stochastic Gradient Classifier. This work can be used as a significant building block for determining the veracity of Fake news. This project will helps us to identify and classify fake news articles. It is necessary to build a model that can differentiate between "REAL" news and "FAKE" news. Our goal is to implement using python and machine learning algorithms. To let the user insert a text to receive an accuracy score. We will analyse and classify whether the news is fake or real.

## II. LITERATURE REVIEW

In paper [1] the methods used are Machine learning classification, Detection and logistic Regression. This paper provide insights on how to heuristically spot fake news. First, regarding the information source, provided statistically significant evidence, secondly, regarding social judgment, about an increase of specific affective cues (love, haha and sad votes) relative to the likes, exhibit a statistically significant decrease on the probability that a posting contains fake news. Furthermore, regarding behavioral cues. And lastly focusing on cognitive cues

In paper [2] the methods used are Rumor Classification, Truth Discovery, Clickbait Detection, Spammer and Bot Detection. In this article, they have explored the fake news problem by reviewing existing literature in two phases: characterization and detection. In the characterization phase, they introduced the basic concepts and principles of fake news in both traditional media and social media. In the detection phase, they reviewed existing fake news detection approaches from a data mining perspective, including feature extraction and model construction. They also further discussed the data sets, evaluation metrics, and promising future directions in fake news detection research and expand the fields to other applications.

In paper [3] the methods used are TriFN Framework, Logistic Regression, Naive Bayes, and Decision Tree. They proposed a novel framework TriFN to model tri-relationship for fake news detection. TriFN can extract effective features from news publisher and user engagements separately, as well as capture the interrelationship simultaneously. Experimental results on real world fake news data sets demonstrate the effectiveness of the proposed framework and importance of tri-relationship for fake news prediction. They mentioned that TriFN can achieve good detection performance in the early stage of news dissemination.

In paper [4] the methods used are TF-IDF Matrix, Passive Aggressive Algorithm and SVM. Their discussion drafts a basic typology of methods available for further refinement and evaluation, and provides a basis for the design of a comprehensive fake news detection tool. Techniques arising from disparate approaches have been utilized together in a hybrid system, whose features were summarized like Linguistic processing, strictly content-based approaches, tools to be designed and contributions in the form of publicly available gold standard data sets should be in linked data format to assist in up-to-date fact checking.

In paper [5] the methods used are Navis Bayes Classifier, Precision matrix. Fake news and Clickbaits interfere with the ability of a user to discern useful information from the Internet services especially when news becomes critical for decision making. They proposed a simple but effective approach to allow users install a simple tool into their personal browser and use it to detect and filter out potential Click baits. The preliminary experimental results conducted to assess the method's ability to attain its intended objective, showed outstanding performance in identify possible sources of fake news.

In paper [6] they used explainable machine learning; social network, computing classification system concepts for Social aspects of security and privacy. They proposed a deep hierarchical co-attention network to learn feature representations for fake news detection and explainable sentences/comments discovery. Experiments on real- world data sets demonstrate the effectiveness of the proposed framework

In paper [7] The method used are Support Vector Machine, Passive Aggressive Classifier, Multinomial Naïve Bayes, Logistic Regression and Stochastic Gradient Classifier. In this work, the classification was performed on small number of news items. Adding more data to the dataset will test the consistency of the performance thereby increasing trust of users. To gather more news that appeared fake and using the linguistic based feature. They concluded that by exploring more social media features in our experiments, and combing them we can create an effective and reliable system for detecting Fake news.

### III PROBLEM SPECIFICATION

Fake news is one of the biggest problem and concern in our world where it is all digitalised and digitally connected. Fake news detection is no longer limited it spreads like wildfire. It is impacting millions of people every day. The question is how do we deal with such a sensitive issue? Millions of articles are being given out every day on the internet – how do we tell real from the fake?. By turning to use of machine learning Algorithm, there are more chances to avoid this spread of hoaxes.

### IV. METHODOLOGY

Machine learning process consists of the following steps:

- ⑩ Data collection and preparation
- ⑩ Feature selection
- ⑩ Algorithm choice
- ⑩ Parameters and model selection
- ⑩ Building a model
- ⑩ Testing the model

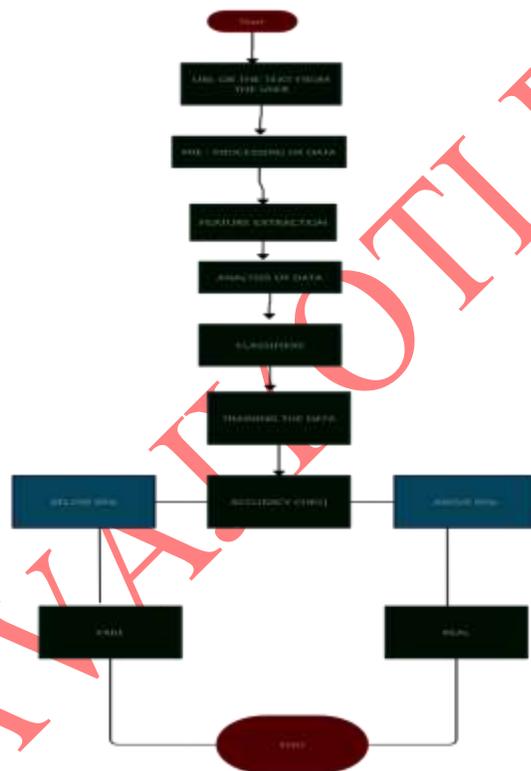


Figure 1: Data Flow Diagram of Fake news Detection

From the research papers gone through, we have learned that they have used various machine learning techniques to predict the fake news. We will implement some the required techniques using datasets, Jupyter notebook, and various python libraries such as pandas, numpy, sklearn etc. and write it in python code. For that we will download few of the datasets from kaggle.com and will try to train and test the data using required machine learning techniques. We will apply feature extraction techniques such as BOW, TFIDF and Count Vectorizer and then select the best classifier by applying classification

techniques like Naïve bayes classifier and PAC. After the best model is selected, we will save in the disk for future predictions and dump the model in the disk using pickle.

## V. WORKFLOW

- **Create a GUI Interface**
  - Where the user can insert a text to receive the result
  - The system will analyse and classify whether the news is fake or real given the input datasets or the contents.
- **Pre-processing the data:** It is a core general skill where there are 4 main important steps.
  - Splitting of the data set in Training and Validation sets
  - Taking care of Missing values
  - Taking care of Categorical Features
  - Normalization of data set
- **Analysing of data**
  - checking excessive punctuation and biased vocabulary
  - read the content and publisher
  - read beyond
  - think before sharing
- **Feature Extraction**
  - Bag of words, TFIDF, Countvectorizer
  - classifiers- NBC, MNB, PAC
  - Confusion matrix
- **Datasets Used**

The column for the dataset “fake\_news.csv” is of size 6335 x 4. This is the dataset used to train our program so that it can later give live predictions. This dataset was downloaded from kaggle.com site where they provide free datasets for the developers. This dataset was first uploaded in the Jupiter notebook folder in order to call the file in the python program when needed. These datasets are then trained and tested using various libraries and machine learning techniques imported.

## VI. RESULTS

Using the available data sets downloaded from the Kaggle, we were able to train and test the data sets using various Python libraries and Machine learning algorithms in Python code. The available datasets were trained and tested using the techniques so that it will be able to give live predictions in the user interface in the future. We studied various machine learning algorithms and methods and we were able to apply few techniques. We used Bag of words, Tf-idf Vectorizer and Count vectorizer for feature extraction and two classifiers namely Naïve bayes and Passive aggressive classifier to train and test the data and confusion matrix to check the performance of the algorithm by calculating the precision and

recall. The accuracy for the Passive aggressive classifier came out to be the highest with 93.69% accuracy. The best model was saved in the disk for future predictions.

## VII. CONCLUSIONS AND FUTURE SCOPE

Due to the inherent relationship among publisher, news and social engagements during news dissemination process on social media, we propose a novel framework. We worked on different classification model. The model achieved accuracy to predict whether the news is fake or real. While going through the research papers we have come across many of the existing techniques used in machine learning, few have tried to implement and have concluded with the applied technique. We have learned about various other python libraries that is imported for the code to run by exploring more social media features in the experiments, and combining them we can create an effective and reliable system for detecting Fake news. It's better we do our own newsgathering first. Read beyond, check the author and other news outlets. There are several interesting future directions.

First, it's worth to explore effective features and models for early fake news detection, as fake news usually evolves very fast on social media. At last, how to identify low quality or even malicious users spreading fake news is important for fake news intervention and mitigation. Since the best model i.e. Passive aggressive classifier is saved in the disk, therefore for future work, the python code written in Jupyter notebook will be combined together with the HTML and CSS code written in Sublime Text 3 in a VS Code editor through flask. After the best model is saved, using pickle, the model will be dumped into the disk. The interface will be then able to predict whether the news is fake or real.

## REFERENCES

- [1] Automatic Detection of Fake News on Social Media Platforms by Christian Janze Goethe, University Frankfurt Theodor and Marten Risius, University of Mannheim (Pacific Asia Conference on Information Systems(PACIS 2017 Proceedings ))
- [2] Fake News Detection on Social Media:A Data Mining Perspective by Kai Shuy, Amy Slivaz, Suhang Wangy, Jiliang Tang, and Huan Liu Computer Science & Engineering, Arizona State University, Tempe, AZ, USA 2016
- [3] Beyond News Contents: The Role of Social Context for Fake News Detection by Kai Shu Arizona State University, Suhang Wang Penn State University and Huan Liu Arizona State University WSDM '19, Melbourne, VIC, Australia, February 11–15, 2019.
- [4] Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee, and Huan Liu. 2019. dEFEND:Explainable Fake News Detection. In The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '19), August 4–8, 2019, Anchorage, AK, USA. ACM, New York, NY, USA
- [5] <https://data--flair-training.cdn.ampproject.org/v/s/data-flair.training/blogs/advanced-python-project-detecting-fake-news/amp>
- [6] <https://www.pantechsolutions.net/fake-news-detection-using-machine-learning>
- [7] Detecting fake news in Social Media Networks by Monther Aldwairi, Ali Alwahedi, Social Media as information source, 2014

- [8] <https://www.geeksforgeeks.org/passive-aggressive-classifiers/>
- [9] <https://www.geeksforgeeks.org/confusion-matrix-machine-learning/?ref=lbp>
- [10] Fake News Detection Using Machine Learning, Author: Simon Lorent Supervisor: Ashwin Itoo, University Of Liège Faculty Of Applied Science Belgium Accademic Year 2018-2019
- [11] A smart System for Fake News Detection Using Machine Learning, Anjali Jain Avinash Shakya, Harsh Khattar and Amit Kumar Gupta, Conference: 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT) September 2019
- [12] Fake News Detection Using Machine Learning approaches: A systematic Review, Syed Ishfaq Manzoor, Jimmy Singla Lovely Professional University, Nikita, Conference: 2019 3<sup>rd</sup> International Conference on Trends in Electronics and Informatics (ICOEI)
- [13] Machine Learning for Detection of Fake News by Nicole O'Brien, Department of Electrical Engineering and Computer Science May, 17, 2018
- [14] FAKE NEWS DETECTION USING ML, Srishti Agrawal, Vaishali Arora, Ruchika Arora, Pronika Chawla, Madhumita Kathuria, International Research Journal of Engineering and Technology (IRJET) Volume: 07 Issue: 05 | May 2020
- [15] Fake News Detection using Machine Learning and Natural Language Processing, Kushal Agarwalla, Shubham Nandan, Varun Anil Nair, D. Deva Hema, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-6, March 2019
- [16] <https://www.geeksforgeeks.org/detecting-frauds-with-ml-and-ai/>
- [17] [https://en.wikipedia.org/wiki/Detecting\\_fake\\_news\\_online](https://en.wikipedia.org/wiki/Detecting_fake_news_online)