

DESIGN AND DEVELOPMENT OF AN AI TOOL FOR FAKE NEWS DETECTION AND VALIDATION USING NLP

¹ A.NARESH, ² Y.VIVEK KUMAR, ³ V.VARUN KUMAR, ⁴ M.PARUSHURAM, ⁵ T.VARSHITH REDDY

¹Assistant Professor, Department of AIML, Sri Indu College Of Engineering & Technology,Hyderabad.

^{2,3,4,5} U.G. Scholar, Department of CS, Sri Indu College Of Engineering & Technology, Hyderabad.

Abstract: Fake news refers to information that is intentionally false or misleading but presented in the form of legitimate news. The spread of such content is often influenced by human behavior, as studies suggest that individuals are naturally attracted to novel and surprising information, which stimulates higher cognitive engagement. Furthermore, the concept of motivated reasoning contributes to the rapid dissemination of misinformation, as people tend to accept and share content that aligns with their beliefs. This behavior encourages the circulation of deceptive material, often characterized by sensational headlines and clickbait titles.

The proposed study focuses on detecting fake news using machine learning and natural language processing (NLP) techniques, particularly targeting content originating from unreliable sources. The research utilizes the ISOT dataset, which consists of both real and fake news articles gathered from multiple sources. Additionally, web scraping techniques are employed to collect current news data from online platforms, which is then incorporated into the dataset for analysis. The collected data undergoes preprocessing and feature extraction to prepare it for model training. Dimensionality reduction techniques are applied to improve efficiency and performance. The system then employs several classification algorithms, including Rocchio classifier, Bagging classifier, Gradient Boosting classifier, and Passive Aggressive classifier, to distinguish between real and fake news. A comparative analysis of these models is conducted to identify the most effective approach for accurate fake news prediction, thereby enhancing the reliability of information filtering systems.

Keywords: Fake News detection, Machine Learning, Natural Language Processing, Singular Value Decomposition, TF-IDF

I. INTRODUCTION

Fake news is false or misleading information presented as news. The proposed study uses machine learning and natural language processing approaches to identify false news—specifically, false news items that come from unreliable sources.

Fake news and disinformation are ongoing problems that may be found all around us in biased software that amplifies just our viewpoints for a "better" and smoother user experience. Fake news and misinformation are becoming more of a problem as the internet and social media platforms become more main stream. A common goal of fake news is to harm someone or something's reputation or to profit through advertising. The propagation of these ideas may have been influenced by a variety of factors, but they all present humanity with the same underlying problem: a misunderstanding of what is real and what is false. This confusion could result in additional problems, such a medical emergency.

Satire websites or sensationalistic "click-bait" is where news site proprietors most frequently steal from fake. Satire websites frequently publish outlandish news parodies, and visitors to these websites are aware that they should not be taken seriously. Tabloids are what click-bait news articles resemble. Even though the actual tales are relatively mundane, their sensationalist and dramatic headlines entice people to click to learn more. These kinds of headlines draw readers to buzz feed and up worthy.

In this paper, we are focusing on the fake news detection in text media. Machine learning and deep learning techniques for fraud detection has been the subject of extensive study, most of which has concentrated on categorising online reviews and publicly accessible social media posts. Some of the drawbacks of the fake news are shift in public opinion, defamation, false perception and many more.

To overcome the drawbacks of the fake news, a model is created to distinguish the real news from the fake news. The proposed method uses the ISOT dataset. Web scraping is also done on 4 websites and the scraped data is further added into the dataset. The data undergoes data pre-processing, feature extraction, dimensionality reduction and finally the data is sent to the classification models i.e. Rocchio classification, Bagging classifier, Gradient boosting classifier and Passive Aggressive Classifier to train the model which is further used to detect the fake news.

The paper is organised as follows. The section covers the previous research on methods for identifying fake news. The proposed approach is described in the third section. Section four describes the evaluation and analysis of the proposed methodology, and the final section provides the paper's conclusion.

II. LITERATURE SURVEY

There have been several previous works done to detect the fake news. Some of the previous works are given below. A fair amount of research has been done different aspects of this project domain.

- 1) According to "Fake News Detection Using Machine Learning Ensemble Methods" paper [1] the goal is to develop a system or model that can use historical data to forecast if a news report is fake or not. The dataset used here is ISOT dataset. The model used in this method is Random Forest Classifier. A large number of decision trees are built during the training phase of the random forests or random decision forests ensemble learning approach, which is used for classification, regression, and other tasks. The class that the majority of the trees chose is the output of the random forest for classification problems. Accuracy is one factor to consider when evaluating categorization models. The accuracy of the proposed solution is 90.64.
- 2) The author T. Dinesh. et. al. [2] has introduced a training model that can be used to predict the class or value of the target variable by learning simple decision rules developed from training data. The approach of classifying the fake political news manually requires more knowledge of the domain. In this research, the problem of classifying fake political news articles using machine learning models is discussed. The mean accuracy and standard deviation for the Decision Tree algorithm is 99.6990 and 0.10577. For Naive Bayes algorithm is 95.3870 and 0.00061. The accuracy of innovative fake news detection for political news detection using Decision Tree algorithms has better accuracy in comparison with Naive Bayes algorithms.
- 3) According to the author Kasra Majbouri. et. al. [3], the purpose is running K-Means clustering to see if the algorithm can successfully cluster the news into Real and Fake using just the words in the articles. The proposed method of choosing features and detecting fake news has four main steps. The first step is computing similarity between primary features in the fake news dataset. Then, features are clustered based on their similarities. Next, the final attributes of all clusters are selected to reduce the dataset dimensions. Finally, fake news is detected using the k-means approach. The accuracy of the K-means clustering algorithm in the detection of fake news is approximately 87%.
- 4) The goal of fake news identification, according to the author Monther Aldwahedi et al. [4], is to develop a method that users may use to identify and filter out websites that contain inaccurate and misleading information. A simple and carefully selected feature of the title and post to accurately identify fake posts. The experimental results showed a 99.4% accuracy using the logistic classifier.

III. FAKE NEWS DETECTION USING NATURAL LANGUAGE PROCESSING

Most text and documents contain many terms that are redundant for text classification, such as stop words, misspellings, slangs, and so on. Hence, data pre-processing has to be done before the data is sent to the classification models. After that, the dataset's dimensionality is decreased in order to save time and storage space. When the dimensions are reduced, it becomes easier to visualise. The data is then used to train classification models, which can be used to predict whether or not the presented data is fraudulent.

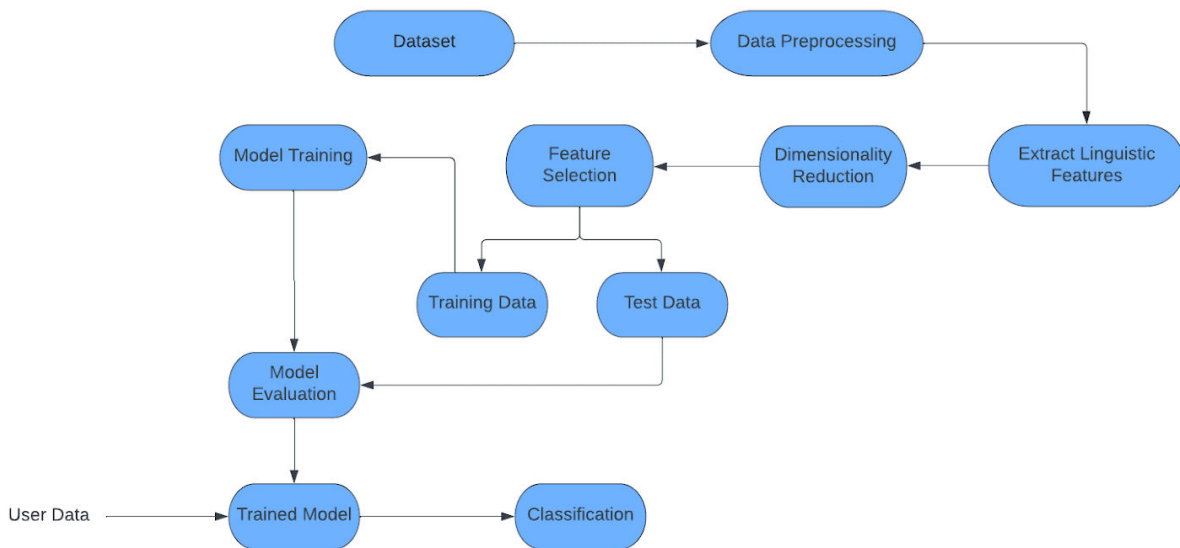


Fig 1. Flow Chart

A. Description of Dataset

The dataset used in this paper is ISOT dataset. In this dataset, there are two types of articles: fake news and real news. The dataset was gathered from real-world sources, and true articles were retrieved via crawling articles from Reuters.com. The fake news articles came from a variety of sources. Politifact and Wikipedia were used to gather the fake news items. Although the majority of the articles in the collection are about politics and foreign events, they cover a wide range of topics. The dataset consists of two CSV files. True.csv is the first file, and it contains almost 12,600 reuter.com stories. Fake.csv, the second file, comprises about 12,600 items obtained from various fake news sites.

B. Web Scraping

Large volumes of data can be automatically gathered from websites via web scraping. The majority of this data is unstructured in HTML format and is transformed into structured data in a database or spreadsheet so that it can be used in multiple applications. Here, web scraping is done on 4 websites to get the present news. It is further added into the dataset to detect the present news as fake or not and also to increase the efficiency of detecting the fake news.

C. Text Cleaning and Pre-processing

To prepare the text data for the model building we perform text pre-processing. It is the first step in the Natural Language Processing. Some of the pre-processing steps are:

- 1) *Tokenization*: Tokenization is the process of breaking down a stream of text into tokens, which can be words, phrases, symbols, or any other significant items. This step's major purpose is to extract individual words in a sentence. The tokenization is done on each text in the dataset.
- 2) *Stop Words*: Stop words are the commonly used words and are removed from the text as they do not add any value to the analysis. These phrases have little or no meaning. A list of terms that are regarded as stop words in the English language is included in the NLTK library. All the stop words from the texts are removed.
- 3) *Capitalization*: Sentences can have a combination of capital and lowercase letters. A written document is made up of multiple sentences. One of the method for reducing the issue space is to convert everything to lower case. This aligns all of the words in a document in the same location. Using the python function, all the words are converted to lower case.
- 4) *Stemming*: Stemming is the process of reducing the words to its root form by eliminating extraneous characters. PorterStemmer is one of the stemming model which is used here to convert the words into its root form.
- 5) *Lemmatization*: Text lemmatization is the process of removing a word's superfluous prefix or suffix and extracting the basic word. All the suffixes and prefixes from the words are removed to reduce space.

D. Feature Extraction

TF-IDF stands for Term Frequency-Inverse Document Frequency and it is a measure, used in the fields of information retrieval and machine learning that can quantify the importance or relevance of string representations in a document amongst a collection of documents. The Bag of Words technique, which is useful for text classification or for assisting a machine read words in numbers, is outperformed by the TF-IDF technique when it comes to understanding the meaning of sentences made out of words. Each feature's TF-IDF weights are computed and recorded in a matrix with columns denoting features and rows denoting sentences.

E. Dimensionality Reduction

Dimensionality refers to how many input features, variables, or columns are present in a given dataset, while dimensionality reduction refers to the process of reducing these features. In many circumstances, a dataset has a significant number of input features, which complicates the process of predictive modelling.

In these situations, dimensionality reduction techniques must be used because it is highly challenging to visualise or forecast for the training dataset with a huge number of features. The curse of dimensionality, which is more commonly known, describes how adding more input features frequently makes a predictive modelling task more difficult to model. Since the TF-IDF matrix is a sparse matrix, Singular Value Decomposition is used for dimensionality reduction.

- 1) *Singular Value Decomposition*: Singular Value Decomposition is one of several techniques that can be used to reduce the dimensionality, i.e., the number of columns, of a dataset. A matrix's Singular Value Decomposition is a factorization of that matrix into three other matrices. Finding the ideal set of variables that can most accurately predict the result is the aim of SVD. During data pre-processing prior to text mining operations, SVD is used to find the underlying meaning of terms in various documents.

Mathematics behind SVD,

The SVD of $m \times n$ matrix is given by the formula,

$$A = UWV^T$$

where

U: $m \times n$ matrix of the orthonormal eigenvectors of AA^T

V: transpose of $m \times n$ matrix containing the orthonormal eigenvectors of $A^T A$

W: a $n \times n$ diagonal matrix of the singular values which are the square roots of the eigen values of $A^T A$

The matrix from TF-IDF is given as input to the TruncatedSVD. The columns i.e. features denotes the dimensions whereas the rows in the matrix denotes the points in the space. The dimensions of the matrix are reduced using TruncatedSVD.

F. Classification Techniques

On the basis of training data, the Classification algorithm is a Supervised Learning technique that is used to categorise new observations. The classification algorithms used in this paper is,

- 1) *Rocchio Classification*: A type of Rocchio relevant feedback is Rocchio classification. The centroid of the class of relevant documents is the average of the relevant documents, which corresponds to the most important component of the Rocchio vector in relevance feedback. Rocchio classification, which uses centroids to define the boundaries, is used to compute good class boundaries. Rocchio classification calculates the centroid for each class. When a new text data is given, it calculates the distance from each of the centroid and assigns the data point to the nearest centroid.
- 2) *Bagging*: When the goal is to reduce the variance of a decision tree classifier, bagging is utilised. The goal is to construct different subsets of data from a training sample that was picked at random and replaced. Their decision trees are trained with each group of data. As a result, we have a collection of various models. The average of all the forecasts from various trees is used which is more robust than a single decision tree classifier.
- 3) *Gradient Boosting*: A method for creating a collection of forecasts is called boosting. In order to reduce training errors, boosting is an ensemble learning technique that combines a number of weak learners into a strong learner. A random sample of data is chosen, fitted with a model, and then trained successively in boosting; each model attempts to make up for the shortcomings of the one before it. The weak rules from each classifier are joined during each iteration to create a single, powerful prediction rule. Gradient boosting is a type of machine learning boosting. It relies on the intuition that the best possible next model, when combined with previous models, minimizes the overall prediction error. The key idea is to set the target outcomes for this next model in order to minimize the error. The target outcome for each case in the data depends on how much changing that case's prediction impacts the overall prediction error.
- 4) *Passive Aggressive Classifier*: For large-scale learning, passive-aggressive algorithms are commonly used. It is one of the few 'online-learning algorithms'. In contrast to batch learning, where the full training dataset is used at once, online machine learning algorithms take the input data in a sequential order and update the machine learning model step by step. This is quite helpful when there is a lot of data and training the entire dataset is computationally difficult because of the size of the data. Since, the web scraping is used in this method, it adds the data to the dataset, and the size of the dataset becomes large which makes the Passive Aggressive Classifier model to work efficiently.

IV. RESULT

To assess the effectiveness of the suggested technique on diverse datasets, we ran a number of simulations and experiments using different classifiers. The dataset was divided into training and test set. 80 percent of the dataset is regarded as the training data, and the remaining 20 percent is taken as the test data. The performance of several approaches was compared using the classification's accuracy as the criterion.

Models	Accuracy	Precision	Recall	F1-score
Rocchio Classification	88.09%	0.93	0.84	0.89
Gradient Boosting	86.75%	0.92	0.83	0.87
Bagging Classifier	94.67%	0.95	0.94	0.95
Passive Aggressive Classifier	93.34%	0.95	0.93	0.94

TABLE I. ANALYSIS OF CLASSIFICATION MODELS

Table I and Fig 2 depicts the comparison of the classification report of Rocchio classification, Gradient Boosting Classification, Bagging Classification and Passive Aggressive classification model. The values in the tables specify the accuracy, precision, recall and f1-score of the classification models in the proposed method.

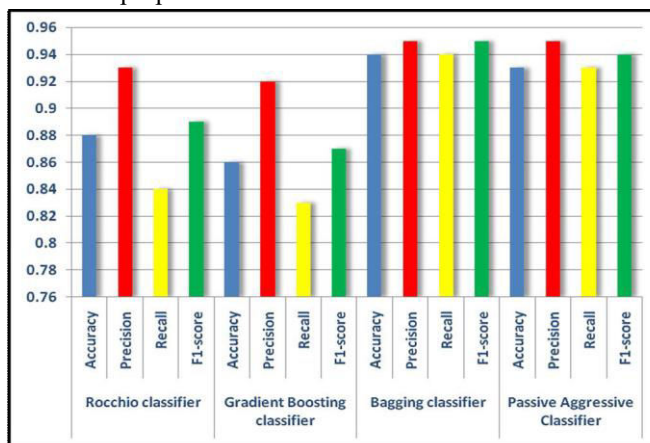


Fig 2. Bar graph for comparison of classification models

As a result of comparison, the BaggingClassifier model has higher accuracy of 94.67% than other classification models like Rocchio Classification model, gradient boosting model and passive aggressive classifier model.

V. CONCLUSIONS

The manual classification of false political news requires for a deeper understanding of the field. The problem of predicting and categorizing data in the fake news detection issue needs to be confirmed using training data. Reducing the amount of these features could increase the accuracy of the fake news detection algorithm because the majority of fake news datasets have many attributes, many of which are redundant and useless. As a result, this research suggests a technique for dimensionality reduction-based fake news detection. The dimension-reduced dataset is constructed using the final set of features. After specifying the final set of features, the next step involves utilizing classification models like Rocchio Classification, Bagging, Gradient Boosting Classifier, and Passive Aggressive Classifier to forecast the fake data.

We assessed the performance of the suggested method on the dataset after it had been implemented. With the classification methods, we achieved the highest accuracy with the 94.67 percent accuracy of the TF-IDF feature extraction and the bagging classifier technique.

REFERENCES

- [1] Ahmad, Ifikhar, Muhammad Yousaf, Suhail Yousaf, and Muhammad Ovais Ahmad. "Fake news detection using machine learning ensemble methods." *Complexity* 2020 (2020).
- [2] Dinesh, T., and T. Rajendran. "Higher classification of fake political news using decision tree algorithm over naive Bayes algorithm." *REVISTA GEINTEC-GESTAO INOVACAO E TECNOLOGIAS* 11, no. 2 (2021): 1084-1096.
- [3] Yazdi, Kasra Majbouri, Adel Majbouri Yazdi, Saeid Khodayi, Jingyu Hou, Wanlei Zhou, and Saeed Saedy. "Improving fake news detection using k-means and support vector machine approaches." *International Journal of Electronics and Communication Engineering* 14, no. 2 (2020): 38-42.
- [4] Aldwairi, Monther, and Ali Alwahedi. "Detecting fake news in social media networks." *Procedia Computer Science* 141 (2018): 215-222.
- [5] Zhang, Jiawei, Bowen Dong, and S. Yu Philip. "Fakedetector: Effective fake news detection with deep diffusive neural network." In *2020 IEEE 36th international conference on data engineering (ICDE)*, pp. 1826-1829. IEEE, 2020.
- [6] Parikh, Shivam B., and Pradeep K. Atrey. "Media-rich fake news detection: A survey." In *2018 IEEE conference on multimedia information processing and retrieval (MIPR)*, pp. 436-441. IEEE, 2018.
- [7] Shu, Kai, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. "Fake news detection on social media: A data mining perspective." *ACM SIGKDD explorations newsletter* 19, no. 1 (2017): 22-36.

- [8] Zhang, Xichen, and Ali A. Ghorbani. "An overview of online fake news: Characterization, detection, and discussion." *Information Processing & Management* 57, no. 2 (2020): 102025.
- [9] Rubin, Victoria L., Niall J. Conroy, and Yimin Chen. "Towards news verification: Deception detection methods for news discourse." In *Hawaii International Conference on System Sciences*, pp. 5-8. 2015.
- [10] Chen, Yimin, Niall J. Conroy, and Victoria L. Rubin. "Misleading online content: recognizing clickbait as" false news"." In *Proceedings of the 2015 ACM on workshop on multimodal deception detection*, pp. 15-19. 2015.
- [11] Zhang, Xichen, and Ali A. Ghorbani. "An overview of online fake news: Characterization, detection, and discussion." *Information Processing & Management* 57, no. 2 (2020): 102025.
- [12] Zhou, Xinyi, and Reza Zafarani. "A survey of fake news: Fundamental theories, detection methods, and opportunities." *ACM Computing Surveys (CSUR)* 53, no. 5 (2020): 1-40.
- [13] Reis, Julio CS, André Correia, Fabrício Murai, Adriano Veloso, and Fabrício Benevenuto. "Supervised learning for fake news detection." *IEEE Intelligent Systems* 34, no. 2 (2019): 76-81.
- [14] Singhal, Shivangi, Rajiv Ratn Shah, Tanmoy Chakraborty, Ponnurangam Kumaraguru, and Shin'ichi Satoh. "Spotfake: A multi-modal framework for fake news detection." In *2019 IEEE fifth international conference on multimedia big data (BigMM)*, pp. 39-47. IEEE, 2019.
- [15] Oshikawa, Ray, Jing Qian, and William Yang Wang. "A survey on natural language processing for fake news detection." *arXiv preprint arXiv:1811.00770* (2018).