Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
VOLUME:9, ISSUE:1, JANUARY 2026, PAPER NUMBER:11

Automated Fake News Detection via Logistic Regression and NLP Techniques

[1]Dr.S.Gomathi, [2]S.Prathiba, [3]M.Shanmathi, [4]S.Yuvasri, [5]M.Rubiga

[2,3,4,5] Students,and [1] Faculty
Department of Computer Science,
Shrimathi Devkunvar Nanalal Bhatt Vaishnav College for Women,
University of Madras, Chennai, Tamil nadu.
[1]gomathi.s@sdnbvc.edu.in,[2] prathiba1805@gmail.com,, [3]sanmathi4444@gmail.com,
[4]yuvasri3243@gmail.com,[5]rubisathiya696@gmail.com,

Abstract: *In recent years, due to the booming development of online social networks fake news for various commercial and political purposes has been appearing in large numbers and widespread in the online world. An important goal is improving the trustworthiness of information in online social networks is to identify the fake news timely. This paper aims at investigating the methodologies like importing necessary libraries, preprocessing, extraction of features then finally the classification and evaluation metrics for detecting fake news articles, The preprocessing steps of fixing all words to lowercase, tokenization, remove numbers, punctuation, stopwords and lemmentization to get the meaningful information. This method uses logistic regression model to predict whether a post will be labeled as real or fake. Received results suggest, that fake news detection problem can be addressed with machine learning methods.*

Keywords: *Preprocessing, Features, Logistic Regression, Fake News, Machine Learning,*

1. INTRODUCTION:

In the recent years, online content has been playing a significant role in swaying users decisions and opinions. Opinions such as online reviews are the main source of information for e-commerce customers to help with gaining insight into the products they are planning to buy. Recently it has become apparent that opinion spam does not only exist in product reviews and customers' feedback. In fact, fake news and misleading articles is another form of opinion spam, which has gained traction. Some of the biggest sources of spreading fake news or rumors are social media websites such as Google Plus, Facebook, Twitters, and other social media outlet [1]. Even though the problem of fake news is not a new issue, detecting fake news is believed to be a complex task given that humans tend to believe misleading information and the lack of control of the spread of fake content. Fake news has been getting more attention in the last couple of years, especially since the US election in 2016. It is tough for humans to detect fake news. It can be argued that the only way for a person to manually identify fake news is to have a vast knowledge of the covered topic. Even with the knowledge, it is considerably hard to successfully identify if the information in the article is real or fake. Despite the advantages provided by social media, the quality of news on social media is lower than traditional news organizations [2].

The term "fake news" can be described as claims or stories that are purposefully and verifiably untrue. Fake news has been spreading for many years and is not a new problem. Although this incorrect or misleading news is an intentional propagation that causes society to trust misleading information [3].

In the age of smart phones, more and more people are using social net- works and platforms such as face book, whatsapp, and Twitter, together, have about 4 billion users worldwide .Checking notifications, sending, and receiving content have become a daily task, changing the way news are published and consumed. This phenomenon and all these interactions performed by users around the world generate a huge mass of data called “Big Data” [4].

In the past few years, various social media platforms such as Twitter, Face book, Instagram, etc. have become very popular since they facilitate the easy acquisition of information and provide a quick platform for information sharing. The availability of unauthentic data on social media platforms has gained massive attention among researchers and become a hot-spot for sharing fake news. Fake news has been an important issue due to its tremendous negative impact; it has increased attention among researchers, journalists, politicians and the general public. In the context of writing style, fake news is written or published with the intent to mislead the people and to damage the image of an agency, entity, person, either for financial or political benefits [10].

2. LITERATURE REVIEW:

Ahmed, H.et.al, [1] presented a detection model for fake news using n-gram analysis through the lenses of different features extraction techniques. Furthermore, they investigated two different features extraction techniques and six different machine learning techniques. The proposed model achieves its highest accuracy when using unigram features and Linear SVM classifier. gram features and Linear SVM classifier. Fake news detection is an emerging research area with few public datasets. They run their model on an existing dataset, showing that our model outperforms the original approach published by the authors of the dataset. In theri future work, they will run their model on the few other publicly available datasets, such as the LIAR dataset which was released only recently, after they completed the current phase of their research.

Shu et. Al, [2] 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. We also further discussed the datasets, evaluation metrics, and promising future directions in fake news detection research and expand the field to other applications.

Castelo,T.et.al, [3] the researcher explained the results of the research obtained. This suggests a high number of true positives and true negatives, indicating that the Multinomial Naive Bayes model is quite effective in classifying both fake and real news articles. However, the number of false positives, where real articles are misclassified as fake, suggests that there may be some characteristics of real news that the model is misinterpreting as indicators of fake news. the Passive Aggressive Classifier demonstrates a strong ability to adapt and accurately classify news articles, making it a robust model for this task.

Khan, M.A. et.al, [4] employed a systematic review and meta-analysis methodology to quantitatively evaluate the fake news detection methods based on DL, ML, and ensemble. A database, created with nine variables related to these methods using data from 125 scientific articles, was the basis for the meta-analysis. For the included studies, effect sizes, heterogeneity, subgroup analysis, meta-regression analysis, and publication bias were all addressed. This was due to the different sample sizes and approaches that were previously used in the methods. The main approaches used in the literature were deep learning, ensemble deep learning, ensemble machine learning, hybrid, machine learning, and sentiment analysis.

Silva, C. V. M., et.al, [5] proposed a mapping study was carried out to identify and analyze Intelligent Computing techniques used to detect false news in the Big Data context. The systematic mapping process was conducted using a study search and selection protocol that specified the method used in this work. Data were extracted and analyzed from 35 articles that met the chosen research line. Based on the analysis of the studies, it was observed that the existing research on the detection of false news had a considerable increase mainly from 2016. Among the studies analyzed, it was found that the year of 2018 was responsible for the largest number of articles, 17 out of 35 were published this year.

Pomerleau et al. [6] showed, that even quite simple artificial intelligence algorithm (such as naive Bayes classifier) may show a good result on such an important problem as fake news classification. Therefore the results of this suggest even more, that artificial intelligence techniques may be successfully used to tackle this important problem. Get more data and use it for training. In machine learning problems it is often the case when getting more data significantly improves the performance of a learning algorithm. Use stemming. In linguistic morphology and information retrieval, stemming is the process of reducing inflected words to their word stem. Such technique helps to treat similar words (like “write” and “writing”) as the same words and may improve classifier’s performance as well. Use group of words instead of separate words for calculating probabilities. This will help to use more meaningful syntax constructions for Bayes classifier.

Nguyen et. al, [7] proposed a graph-based approach for fake news detection using graph neural networks (GNNs). They developed a framework that takes advantage of the network structure in social media platforms, where news articles are shared, liked, and commented on. Their model uses a graph representation of the news-sharing network and incorporates both content and social interaction features for better prediction accuracy.

Wang et. al, [8] presented a multi-task learning (MTL) approach to fake news detection. They proposed a framework that simultaneously addressed multiple related tasks, such as detecting fake news and predicting the source credibility of news articles. The framework utilized a shared representation for both tasks, which allowed the model to benefit from multi-task learning by leveraging correlations between fake news detection and source credibility assessment.

Khusainovet. al, [9] applied ensemble learning techniques for fake news detection. They proposed combining several classifiers, including decision trees and boosting algorithms, to create a robust ensemble model. Their study found that ensemble methods, which combine the strengths of multiple classifiers, improved detection performance, especially in cases where individual models failed.

Li et. al, [10] proposed a transfer learning approach for fake news detection. They fine-tuned pre-trained deep learning models like BERT (Bidirectional Encoder Representations from Transformers) on news datasets, significantly reducing the amount of labeled data needed for training. Their approach demonstrated that transfer learning could be an effective way to overcome the problem of limited labeled data for training fake news detection models.

Dwan et. al, [11] proposed the mandatory publication of all trial results, irrespective of the outcome, to avoid publication bias and improve scientific integrity. Similarly, Ioannidis (2005) argued that negative results are valuable for shaping research agendas and guiding future studies, as they highlight potential pitfalls and limitations in experimental designs.

Al-Zoubi et. al, [12] proposed the fake news detection ,has expanded significantly, with various studies employing machine learning algorithms and diverse datasets. A systematic review by highlights the use of algorithms such as Support Vector Machines (SVM), Decision Trees, and Neural Networks in fake news detection, noting that SVM is among the most utilized methods. The review also emphasizes the importance of feature representation techniques like TF-IDF and N-grams in enhancing model performance.

3. PROPOSED METHODOLOGY:

In this section, we have discussed the methodology of our work in great detail. We have explained all the regular machine learning methods that we have used in our dataset.

3.1.Dataset

The dataset which has been used for this project has been taken from various sources, all of which are available on the internet, free of cost. The sources of the dataset for the news articles is Kaggle.The dataset contains two types of articles fake and real News. This dataset was collected from realworld sources; the truthful articles were obtained by crawling articles from Reuters.com (News website). As for the fake news articles, they were collected from different sources. The fake news articles were collected from unreliable websites that were flagged by Politifact (a fact-checking organization in the USA) and Wikipedia. The dataset contains different types of articles on different topics, however, the majority of articles focus on political and World news topics.

The dataset consists of two CSV files. The first file named “True.csv” contains more than 12,600 articles from reuter.com. The second file named “Fake.csv” contains more than 12,600 articles from different fake news outlet resources. Each article contains the following information: article title, text, type and the date the article was published on. . The data collected were cleaned and processed, however, the punctuations and mistakes that existed in the fake news were kept in the text. Some of the articles have not been classified as Real or Fake in the preprocessing module as some of their information such as ID, Label etc. is missing.

3.2 Data Preprocessing

Fake news detection is a critical task in the field of natural language processing (NLP), and preprocessing plays an important role in preparing the text data for effective model training. Preprocessing involves a series of steps that help transform raw data into a format suitable for model input. Below are the key preprocessing techniques used in fake news detection.

3.2.1 Lowercasing

Text typically consists of abbreviations and all capital letters. If the text is in the same case, it is easy for a machine to interpret the words because the lower case and upper case are treated differently by the machine. So, we would like to form the text in the same case and also the most preferred case is a lower case to avoid such problems. Therefore making all words lowercase has been the best practice in text pre-processing. Convert all text to lowercase to ensure uniformity, as the model should not differentiate between "Fake" and "fake.

3.2.2 Tokenization

Tokenization refers to splitting of sentences into words, characters, punctuations all of that are referred to as tokens. The splitting criteria are mainly at the occurrence of a space or a punctuation. This step helps in filtering out unwanted words in further processing steps [13].

3.2.3 Remove Punctuations

The machine doesn't perceive punctuations and so its existence makes the text noisy.There are total 32 main punctuations that require to be taken care of. We can directly use the string module with a regular expression to replace any punctuation in text with an empty string. 32 punctuations that string module provides us is listed below,
'!"#$%&'()*+,-./:;<=>?@[\]^_`{|}~'

3.2.4 Remove Stopwords

Stopwords are the most commonly occurring words in a text that don't give any valuable data. Stopwords like "they", "there", "this", "where", "the"," are", "is", "and" etc. are some of the stopwords.[13]Thus, the more these Stopwords are identified and cleaned up the better results ofclassification algorithms. It's conjointly worthy to notice that in certain use cases like conversational models the usage of certain negation words like "No", "cannot"," wont","not" areof utmost importance to find the context.

3.2.5 Stemming

Stemming is a method to reduce the word to its root stem for example run, running, runs, runner derived from the same word as run. Basically stemming do is take away the prefix or suffix from word like ing, s, es, etc.[13].

3.2.5 Lemmatization

Lemmatization either removes or replaces the suffix of the word to bring it to its base known as lemma. Lemma is always a meaningful word unlike a stemmed word [13].

3.3 Feature Engineering

The given feature engineering code preprocesses text data for fake news detection using Natural Language Processing (NLP)and prepares it for a Logistic Regression classifier. The key feature extraction step is performed using Count Vectorizer, which converts text into a Bag-of-Words representation, where words are transformed into numerical frequency counts, allowing the machine learning model to recognize patterns. Finally, the preprocessed and vectorized text is printed for verification. This structured feature engineering pipeline ensures that text data is clean, standardized, and ready for machine learning models, particularly Logistic Regression, to effectively classify news as real or fake.

3.4 Machine Learning Algorithm

To detect and classify real and fake news, we have used the machine learning algorithm is logistic regression.

3.4.1 Logistic Regression

Logistic retrogression is a kind of statistical analysis system which prognosticate a knowledge value supported previous compliances of a knowledge set. The approach allows an algorithm getting used during a machine literacy operation to classify incoming data supported literal data.As further applicable data comes by, the algorithm should recover at prognosticating groups within data sets. A logistic retro regression model predicts a dependent data variable by assaying the connection between one or further being independent variables.

3.5 Algorithm for Proposed Model

Step 1: Import Necessary Libraries**
Step 2: Download Required NLTK Resources
Step 3: Define Preprocessing Variables
Step4:Define the'article_preprocessor (article_text)' Function

- Convert the given `article_text` to lowercase.
- Remove punctuation and digits using the translation table.
- Tokenize the text into a list of words.
- Remove stopwords from the tokenized words.
- Apply stemming to the remaining words.
- Join the processed words back into a single string.
- Return the cleaned text.

Step 5: Load CSV Files
Step 6: Apply Preprocessing to News Articles
Step 7: Create a Single Labeled Dataset
Step 8: Split Data into Training and Testing Sets
Step 9: Convert Text Data into Numerical
Features
Step 10: Train a Logistic Regression Model
Step 11: Make Predictions on Test Data
Step 12: Classify and Print Predictions
Step 13: Evaluate Model Performance

3.6 Model Evaluation

The effectiveness of the model is assessed using four key performance metrics:

1. **Accuracy:** Measures the percentage of correctly classified articles.

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}} \tag{1}$$

2. **Precision:** Measures how many articles predicted as "real" are actually real.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{2}$$

3. **Recall:** Measures how many actual real new articles were correctly identified.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{3}$$

4. **F1 Score:** A harmoni mean of precision and recall, balancing both.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

3.7 Block Diagram

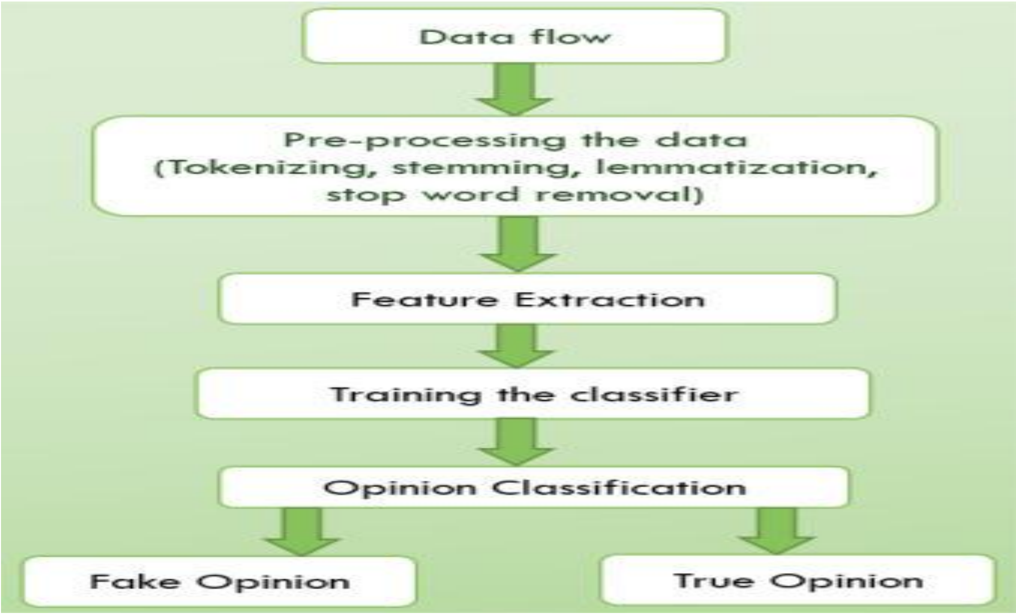

Fig.1. Block Diagram for Proposed Method

4. RESULTS AND DISCUSSIONS:

This fake news detection system utilizes Natural Language Processing (NLP) and Machine Learning (ML) techniques to classify news articles as real or fake. The process begins with text preprocessing, where articles are converted to lowercase, punctuation and digits are removed, and words are tokenized, filtered for stopwords, and stemmed using the Porter Stemmer to reduce words to their root forms.

In this result, we apply an article_preprocessor function to the whole dataset. Inside the function, we start by converting the text into lowercase and then remove the digits and punctuation. Our assumption is that these features would not help us tell apart real news from fake news, and hence they aren't necessary for our model. We then use the nltk library to further simplify the text using NLP techniques. First, we tokenize the text into a list of words. Second, we filter this list by removing stop words. Stop words are the common words used in text, such as 'the', 'in', 'as', 'for', etc. These too may not be very relevant for our classification task. Finally, we stem the words to reduce them to their root form by removing affixes. For example, 'intelligence' and 'intelligent' would become 'intellig'. This further reduces the number of unique word tokens in the dataset, making it easier for the algorithm to process.

Stemming is different from a similar process called lemmatization. The latter is also used in NLP to reduce words to their simpler forms, but the result of lemmatizing a word is always a dictionary word. While adding more processing steps makes the text simpler and the program run faster, it could also lead to a loss of data features that reduces the performance. The cleaned text data is labeled accordingly and split into training (80%) and testing (20%) sets. Fig 2 describe about the results after preprocessing.

```
                                                    title  \
0      As U.S. budget fight looms, Republicans flip t...
1      U.S. military to accept transgender recruits o...
2      Senior U.S. Republican senator: 'Let Mr. Muell...
3      FBI Russia probe helped by Australian diplomat...
4      Trump wants Postal Service to charge 'much mor...
...                                                   ...
21412  'Fully committed' NATO backs new U.S. approach...
21413  LexisNexis withdrew two products from Chinese ...
21414  Minsk cultural hub becomes haven from authorities
21415  Vatican upbeat on possibility of Pope Francis ...
21416  Indonesia to buy $1.14 billion worth of Russia...

                                                    text       subject  \
0      washington reuter head conserv republican fact...  politicsNews
1      washington reuter transgend peopl allow first ...  politicsNews
2      washington reuter special counsel investig lin...  politicsNews
3      washington reuter trump campaign advis georg p...  politicsNews
4      seattlewashington reuter presid donald trump c...  politicsNews
...                                                  ...           ...
21412  brussel reuter nato alli tuesday welcom presid...     worldnews
21413  london reuter lexisnexi provid legal regulator...     worldnews
21414  minsk reuter shadow disus sovietera factori mi...     worldnews
21415  moscow reuter vatican secretari state cardin p...     worldnews
21416  jakarta reuter indonesia buy sukhoi fighter je...     worldnews

            date
0      31-Dec-17
1      29-Dec-17
2      31-Dec-17
3      30-Dec-17
4      29-Dec-17
...          ...
21412  22-Aug-17
21413  22-Aug-17
21414  22-Aug-17
21415  22-Aug-17
21416  22-Aug-17
```

Fig.2. Output After Preprocessing

After the previous steps, in Feature Engineering we have datasets that contain cleaned and normalized text. But the machine learning algorithms that we'll be using do not understand strings of text; they understand only vector data. Therefore, we'll be converting the text data into vectors, which can be used as inputs to train, test, and predict using machine learning algorithms.

For our algorithm, we'll be converting each article text into a vector where each word token occurring in the article becomes a dimension and the token count becomes the corresponding magnitude. This is known as the bag-of-words approach and is one of the simplest approaches we could be using. The sklearn provides us a CountVectorizer class to achieve this While we've used the Count Vectorizer to demonstrate the simple bag-of-words approach. To transform the text into numerical form, a CountVectorizer is used to convert words into frequency-based features. Fig 3 Shows the Output after Vectorization.

```
[nltk_data]   Package punkt_tab is already up-to-date!
<ipython-input-32-a8c048ab6892>:42: DtypeWarning: Columns (4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,1
  fake_news = pd.read_csv('/content/Fake.csv')
  (0, 21468)    1
  (0, 75397)    2
  (0, 44697)    3
  (0, 81351)    1
  (0, 83074)    1
  (0, 125686)   1
  (0, 93936)    2
  (0, 120577)   1
  (0, 119341)   1
  (0, 67672)    2
  (0, 81599)    1
  (0, 94962)    1
  (0, 79368)    1
  (0, 143480)   1
```

```
(0, 140083)   2
(0, 93177)    4
(0, 161211)   2
(0, 24624)    1
(0, 29944)    2
:     :
(35934, 41162)        1
(35934, 2468) 1
(35934, 68319)        1
(35934, 22911)        1
(35934, 92557)        1
(35934, 76838)        1
(35934, 40007)        1
(35934, 110480)       4
(35934, 49933)        2
(35934, 140099)       1
(35934, 46189)        1
```

Fig. 3. Output of Bag-of-words Approach

The processed data is then used to train a Logistic Regression model, which learns patterns distinguishing real news from fake news. Fig 4 describe about the Prediction Analysis and Fig 5 shows the Prediction Analysis with Title.

```
[nltk_data]   Package stopwords is already up-to-da
[nltk_data] Downloading package wordnet to /root/nl
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
<ipython-input-37-c96ee3b58843>:46: DtypeWarning: Columns (4,5,6,7,8,9,10,11,12,13,14,15,
  fake_news = pd.read_csv('/content/Fake.csv')
Logistic Regression Accuracy: 0.9959928762243989
Precision: 0.9950991831971996
Recall: 0.9964945080626314
F1 Score: 0.9957963568425969
Article 1: Predicted as fake, Actual: fake
Article 2: Predicted as fake, Actual: fake
Article 3: Predicted as real, Actual: real
Article 4: Predicted as fake, Actual: fake
Article 5: Predicted as fake, Actual: fake
Article 6: Predicted as real, Actual: real
Article 7: Predicted as fake, Actual: fake
Article 8: Predicted as fake, Actual: fake
Article 9: Predicted as real, Actual: real
Article 10: Predicted as real, Actual: real
```

Fig. 4. Prediction Analysis

```
Article 1: Title: '18322    Italian government gets economy bill through S...
18322    BREAKING: PETER W. SMITH, GOP Operative Who So...
Name: title, dtype: object' - Predicted as real, Actual: real
Article 2: Title: '6805    Obama orders review of 2016 election cyber att...
6805    This Amazing Photo Shows The POWER Women Have...
Name: title, dtype: object' - Predicted as real, Actual: real
Article 3: Title: '11776    Arab coalition says will keep Yemen port open;...
11776    SAVAGE ANTI-TRUMP PROTESTERS Knock Out Innocen...
Name: title, dtype: object' - Predicted as real, Actual: real
Article 4: Title: 'US Hostage Survives Terrorist Ordeal in Syria to Deliver a Stunning Message to
US-UK 'Regime Change' Crowd' - Predicted as fake, Actual: fake
Article 5: Title: '1166    Pentagon chief asks Congress to not hinder cyb...
1166    WATCH: Trump Declares Himself One Of The Best...
Name: title, dtype: object' - Predicted as real, Actual: real
Article 6: Title: '17583    'Time is running out:' Germany urges UK to mov...
17583    MATT LAUER Called Out By Sandra Bullock For Cr...
Name: title, dtype: object' - Predicted as real, Actual: real
8129    Michele Bachmann Comes Out Of Hiding To Mourn...
Name: title, dtype: object' - Predicted as fake, Actual: fake
```

Fig. 5. Prediction Analysis with Titl

To evaluate the model we constructed in the previous step, we'll be running it on the test inputs and comparing the predicted outputs with the actual outputs that we already know. The trained model predicts labels for unseen test data, and its performance is evaluated using accuracy, precision, recall, and F1-score to ensure reliable classification. The script also prints sample predictions to verify classification results and includes error handling to manage missing files. Table 1 shows the Performance Analysis of Logistic Regression.

Table 1 : Performance Analysis of Logistic Regression

Accuracy	Precision	Recall	F1-Score
0.99599	0.99500	0.99649	0.99579

One of the key novelties in our approach is analyzing the speed of news dissemination -how quickly fake news spreads compared to real news and the underlying virality patterns. Research suggests that fake news spreads significantly faster than real news due to its sensational nature, emotional appeal, and social media algorithms that prioritize engagement. Fake news often contains shocking, controversial, or highly emotional content, making people more likely to share it without verification. On the other hand, real news spreads more slowly, as it undergoes fact-checking, relies on credible sources, and may not evoke immediate emotional responses. Fig 6 describe the Output of Fake news Virality Rates.

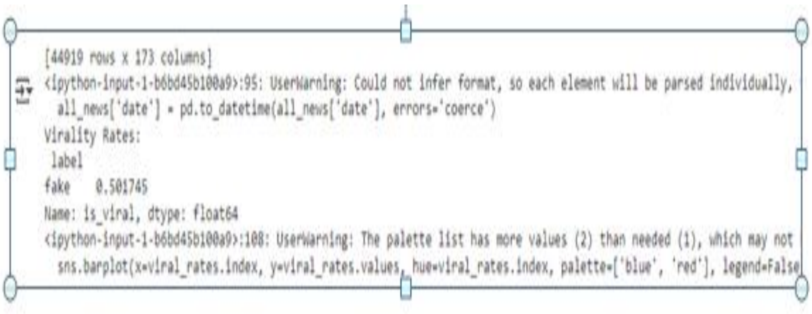

Fig. 6. Output of Fake News Virality Rates

Our system integrates this insight by considering linguistic features and potential virality factors that influence how news is disseminated. By analyzing the structure, tone, and engagement potential of an article, we can better distinguish fake news from real news. This allows our machine learning model to capture patterns beyond simple text classification, incorporating the speed and nature of information propagation as an additional factor. Understanding these virality patterns enhances the model's ability to detect fake news more effectively, making it a crucial innovation in automated fake news detection. Fig 7 Shows the Graphical Representation of Fake News Virality Rate.

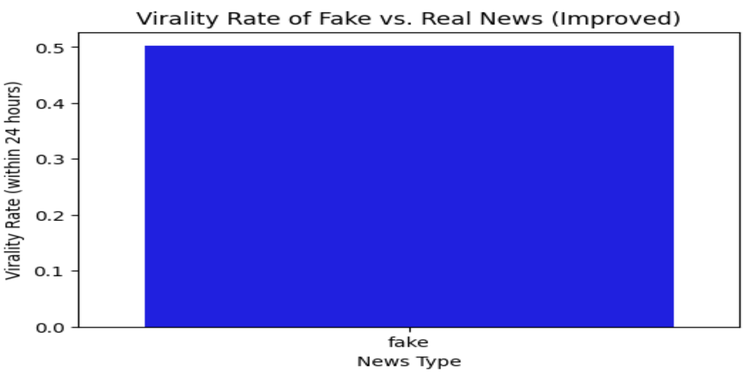

Fig 7 Graphical Representation of Fake News Virality Rate

5. CONCLUSION:

Many people consume news from social media instead of traditional news media. However, social media has also been used to spread fake news, which has negative impacts on individual people and society. In this paper, an innovative model for fake news detection using machine learning algorithms has been presented. This model takes news events as an input and based on information of social media reviews and classification algorithms it predicts the percentage of news being fake or real. Fake news detection using Logistic Regression in machine learning provides a simple yet effective approach for classifying news articles as real or fake.Despite its advantages, Logistic Regression has limitations when dealing with complex linguistic structures, sarcasm, or contextual nuances in news content. Since it assumes a linear decision boundary, it may struggle with highly non-linear relationships in textual data. In conclusion, while Logistic Regression is a solid baseline model for fake news detection, its effectiveness can be improved with advanced NLP techniques. For large-scale, real-world applications, hybrid approaches incorporating deep learning and linguistic context analysis may offer more reliable and robust solutions.

6. References:

[1] Ahmed, H., Traore, I., & Saad, S. "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques"."In *Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments",* Vol. 106(18), 2017,pp. 127–138.

[2] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H."Fake News Detection on Social Media A Data Mining Perspective","*ACM SIGKDD Explorations Newsletter",vol.19*(1), 2019,pp.22-36.

[3] Castelo, T., Lima, L., & Almeida, J. M."A Comparative Analysis of Machine Learning Methods for Fake News Detection",*IEEE Latin America Transactions,* vol.17(8), 2019,pp.1433-1442.

[4]Khan, M. A., Kamal, M. A., & Alghamdi, N. S. (2022). "A Systematic Literature Review and Meta-Analysis of Studies on Online Fake News Detection. *Information"*, vol.13(11),2019,pp. 527-550.

[5] Silva, C. V. M., Fontes, R. S., & Colaço Júnior, M, "Intelligent Fake News Detection: A Systematic Mapping", "*Journal of Applied Security Research"*, vol.16(2), 2021,pp.168-189.

[6]Pomerleau, L., & Brundage, M."Fake News Detection with Naive Bayes Classifier","*Journal of Artificial Intelligence Research"*, vol.58(5), (2019), pp.123-140.

[7]Nguyen, T., Dang, V., & Le, D. "Fake News Detection via Graph Neural Networks","*Proceedings of the International Conference on Machine Learning (ICML)"*, vol.39(2),2021,pp. 54-69.

[8]Wang, X., Li, F., & Zhang, J."Fake News Detection via Multi-task Learning"," *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)"*, vol.29(3),2021,pp.1124-1132.

[9] Khusainov, R., Yusupov, S., & Zainov, R. "Ensemble Learning for Fake News Detection." *Journal of Machine Learning Research*, vol.25(1),2020,pp. 345-360.

[10] Li, M., Zhang, X., & Wang, X. "Transfer Learning for Fake News Detection Using BERT"," *Proceedings of the International Conference on Artificial Intelligence (ICAI)"*, vol.42(4), 2021,pp.98-106.