Social Media Fake News Detection Using Machine Learning

Vandita Singh Gautam
Computer Science and Engineering
Galgotias College of Engineering and Technology
Greater Noida, India
vanditasingh329@gmail.com

Shanu Katiyar
Computer Science and Engineering
Galgotias College of Engineering and Technology, Greater Noida, India
aahan.katiyar@gmail.com

Kavita Sharma
Computer Science and Engineering
Galgotias College of Engineering and Technology, Greater Noida, India
kavitasharma_06@yahoo.co.in

Rudra Pratap Ojha
Computer Science and Engineering
G.L Bajaj Institute of Technology and Management,
Greater Noida, India
rpojha@gmail.com

Abstract— **Nowadays it has become very difficult to tell the legitimacy of news on social media news blogs and online newspaper. Hence there is a need of tools that can check these sources to identify fake news. The automatic detection of false content in news stories is the main topic of this research. We start by introducing a dataset for the false news detection job. To categorize fake news, we applied language processing methods based on logistic regression. Tokenizing, stemming, stopwords removal and data quality checks (i.e., null or missing values), are some of the tasks carried out by the pre-processing algorithms. We have also used SVM algorithm for fake news identification. We also determine which of the two algorithms Logistic Regression or SVM offer greater performance overall.**

***Keywords*—Hoax, ML, Social media, Stemming, Logistic Regression, SVM, Tokenizing, TF-IDF.**

I. INTRODUCTION

News is significant because it keeps the general people informed of events and developments both locally and globally. According to statistics, the majority of adults choose digital news sources over traditional media, such as social media and conventional search engines. News with potential to misguide readers is considered as fake news,categorised into three categories of fake news: Those of a purely false nature, with the express purpose of perplexing the reader; Rumours, which are unconfirmed truths that are generally accepted; and those of a satirical nature, which employ irony and sarcasm to produce parodies and satires.

Social media is characterised as an internet tool created to encourage interactivity between people .It has many advantages and disadvantages. People use social media for access of news because it is free, accessible, and spreads quickly, but some people use it to spread misleading information.

Hence, in order to stop the spread of fake news on any social platform development of a classification model is essential. We will be able to check the authenticity of the news and also determine the reliability of the source of the news. Therefore, in order to build such a system we need to study information reviews performed in the past to foretell whether a file containing information is likely to be fake or not.

We provide an overview of false news detection and talk about possible research trajectories in this post. The following is a summary of the main driving factors behind this survey:

- Although the term "fake news" has been there about on social media for a while, no accord on what it means exists. Clarifications are needed for future researches for fake news detection
- Online sites are the biggest source for the spread of fake news. There are few means that can be used in social media to identify bogus news. Reviewing the current false news identification strategies in various social media contexts will help you gain a basic understanding of the most reliable fake news detection approaches.
- Fake news detection system still needs a lot of work, and many other difficult problems needs a lot of research work. It is important to discuss about the researches done before us to enhance ability to identify and counteract fake news.

We have seen how misleading information can have major implications in recent years, from influencing political campaigns to driving public health problems. Combating fake news is a vital approach for us to encourage ethical journalism and protect the public.

A. MACHINE LEARNING

There has been a great deal of research on machine learning methods for detecting fraud, most of which has focused on online reviews and publicly available posts. A lot of attention is now being paid to the problem of spotting "fake news" in the literature.

Research questions that are answered in this project are:

- Why we have used machine learning for detecting inauthentic news?
- Which ML algorithms is used to detect unauthentic news?
- How are the used classifiers trained for detecting fake news?

1. Why we have used machine learningfor detecting inauthentic news?

People who are well informed about a situation can only asses for news legitimacy, while others will automatically start believing the false information.

Machine learning algorithms are used to detect such unauthentic news. Various academics are competing to develop the best machine learning classifier with higher accuracy as it offers greater performance. The training of the classifier determines its accuracy. A good training method can improve a model's accuracy.

2. Which ML algorithms is used to detect unauthentic news?

Logistic Regression: It is a supervised ML learning algorithm that is used to determine authenticity of statements. It predicts probabilistic value of true or false. Value can be either true that is 0 or false that is 1.

SVM: aim is to identify the optimum hyperplane that is the decision boundary that best separates the various classes in the data by the widest margin. SVM finds a hyperplane that is least likely to make classification mistakes on fresh, untested data by maximising the margin.

By utilising a kernel function , SVM can handle datasets that are both linearly separable and non-linearly separable.

3. How are the used classifiers trained for detecting fake news?

To train these classifiers, we used feature extraction methods. In reality. The model is the TF-IDF Model.Training dataset is used for training purpose. Once the unnecessary words have been removed, the words are initially reduced to their simplest form. Hence, trained data should contain only valuable data

Features of fake news: There are two components to fake news: authenticity and intent. When fake news is authentic, it contains untrue information that can be proven to be untrue. The second component, intent, shows that inaccurate matter was written to mislead the readers.

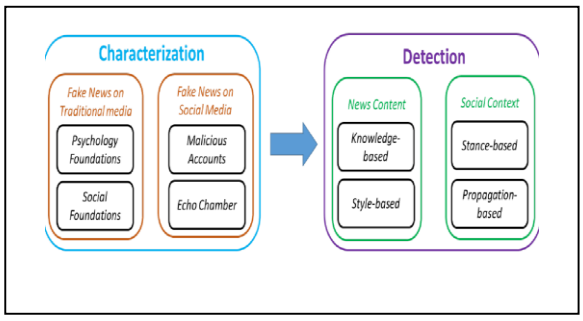

Fig.1. Features of fake news [18]

II. LITERATURE REVIEW

Madalina Erascu et al. [1] They presented machine learning approaches, specifically supervised learning, for detecting bogus news. In their assessment of numerous studies that employed supervised learning techniques to spot bogus news, Erascu et al. For instance, Horne and Adali (2017) classified bogus news using SVMs by taking into account textual characteristics including the article's length, the existence of proper nouns, and the frequency of punctuation. Shu et al. (2017) used decision trees in another study to distinguish between legitimate and fake news based on features like sentiment analysis, readability, and topicality. emphasised the potential of ML methods for spotting false news and talked about the many opportunities and problems in this field. Their review of the literature revealed the state of the art at the time and suggested some potential future study topics.Training data consists of 66% of the data and testing data consists of the rest of the data.

Ms. Chuma Devi et al. [2] Main purpose is to develop a model to detect fake news. The authors talked about certain difficulties in identifying fake news, including the fluidity of internet material, the availability of propaganda, and the requirement for huge datasets. They put out a system that classifies news stories as authentic or fraudulent based on numerous characteristics, including word frequency, n-grams, and part-of-speech tags. they evaluated the performance of their system by calculating accuracy, precision score, recall, and F1 score. They compared their results with several baseline models and found that their system achieved better performance in detecting fake news.Three algorithms are used here. They are Naïve Bayes, Logistic regression, SVM. The accuracy obtained is 78%.

P. Yogendra Prasad et al. [3] They provided a full analysis of the prior studies on fake news detection systems. The authors talked about the many methods that have been utilized in the literature, such as supervised and unsupervised learning, as well as the numerous features that have been applied to false news identification. For the purpose of identifying fake news, they examined the effectiveness of numerous machine learning techniques, like Naive Bayes, Logistic Regression, etc. The authors also covered the methodologies' drawbacks and difficulties, such as the availability of labelled data and the existence of biassed data. Systems performance was tested using a merged dataset (containing a true news and a fake news dataset).

Uma Sharma et al. [4] used 3-fold cross validation for this experiment where around 66% of the data is used for training the model. They provided a full analysis of the prior studies on fake news detection systems. The authors talked about the many methods that have been utilized in the literature, such as supervised and unsupervised learning, as well as the numerous features that have been applied to false news identification. For identifying unauthentic news, they examined the performance of various ML algorithms, including Logistic Regression, SVM etc. The authors also covered the methodologies' drawbacks and difficulties, such as the availability of labelled data and the existence of biassed data.Logistic Regression came out to be best with accuracy of 65%.

Sasibhushana Rao Pappu et al. [5] developed a machine learning-based hybrid method for identifying false news. They gathered a dataset of news pieces from several sources, both authentic and fraudulent. By eliminating stop words,

stemming, and changing the text's case, the writers preprocessed the data. Then they retrieved a number of variables, including readability, sentiment, and TF-IDF. These features were utilised by the authors to train ML algorithms. To boost the system's efficiency, they also developed a hybrid strategy that incorporates various machine learning models. The hybrid technique beat the separate machine learning models, according to the results, obtaining an accuracy of 93.7%. After that, they compared their strategy with a number of other false news detecting techniques. Using a variety of criteria, including accuracy, precision, the authors assessed the effectiveness of their method.

Alim Al Ayub Ahmed et al. [6] provided a full analysis of the existing research on fake news detection using ML techniques. They discussed the different approaches, including supervised and unsupervised learning, and the various features used to detect misleading unauthentic information. The authors analyzed performance of several ML algorithms, like Naive Bayes, SVM, etc. used for fake news detection. They also discussed the challenges faced by these methods, including the availability of labeled data and the presence of biased data. offered a thorough analysis of the prior research on machine learning algorithms for detecting unauthentic news. Their research emphasised the need to create efficient strategies to stop the circulation of misleading unauthentic news as well as the many techniques and difficulties in fake news identification.

John Curci et al [7] presented a trilateral strategy to tackle fake news, which consists of technological, legal, and media literacy education. They claimed that to stop the spread of misleading information, a combination of these three strategies is required. They conducted a case study in the United States to examine the effects of technical advancements, regulatory changes, and media literacy education on the propagation of false information. They discovered that while technological and legal solutions were ineffective at stopping the spread of misleading news, media literacy education did. presented a triangular strategy to put a stop on the spread of unauthentic news and carried out a case study to examine the effects of media literacy instruction, technology advancements, and governmental initiatives on this phenomenon in the United States. Algorithms used are Naive Bayes, SVM, LSTM etc., among which LSTM achieves highest F1 score

Afrin Jaman Bonny et al[8]went through a thorough analysis of the literature on ML algorithms for detecting fake news. They examined the numerous strategies put out in the literature, such as supervised and unsupervised learning, as well as the distinct features employed in fake news detection. They also assessed the effectiveness of a number of machine learning algorithms used to identify fake news, such as Naive Bayes, SVM, and Random Forest. They talked about the difficulties these techniques encountered, such as the lack of labelled data, the existence of biased data, and the interpretability of the findings. They talked about how social media affects the dissemination of fake news and the necessity to create efficient detection techniques. They also emphasised the value of media literacy and user education in halting the spread of false information. LR achieves highest performance with an accuracy of 93.9%.

A. LIMITATION

Although ML can be a useful tool for spotting fake news, this method has some drawbacks. To name a few:

1. Limited Data Availability: Access to trustworthy and labelled training data is one of the biggest obstacles to detect unauthentic news. It is challenging to train models that can reliably identify bogus news because there aren't enough large datasets with precise labelling.

2. Adversarial Examples: These are inputs that are specifically created to lead machine learning models to fail. This could imply that malevolent actors purposefully create news stories to avoid detection by machine learning algorithms in the context of detecting fake news.

3. Contextual Understanding: Assessing the reliability of news pieces requires an understanding of the context in which they are published. Machine learning models, however, have trouble comprehending context, particularly when it comes to sarcasm, irony, and other sophisticated language usage.

4. Language Barriers: Machine learning models that have been trained on news stories written in English may not be successful in identifying false information in other languages. This is due to the fact that various languages have various structures, grammars, and vocabularies, which might make it difficult to develop a general strategy.

5. Bias: Depending on the data that machine learning models are trained on, they may occasionally be biassed. The resulting models will be biassed if the training data is biassed, which can produce unreliable findings.

6. Rapidly Changing tactics: Machine learning models may find it difficult to keep up with the rapidly changing tactics used to produce fake news. This means that for models to stay useful, they may need to be updated frequently.

III. PROBLEM IDENTIFICATION

Fake news is an important issue for societies around the world because it can have negative effects on people, organizations, and public policy. Through social media and other internet venues, fake news may spread quickly, resulting in confusion, polarization, and even violence. Additionally, it has the potential to weaken public confidence in journalism and democratic institutions as well as the authority of reliable news sources.

The subject of false news is intricate and varied, touching on topics like media literacy, political polarization, freedom of speech, and the influence of technology on public dialogue. The abundance of echo chambers and filter bubbles that reinforce people's preexisting views and biases, as well as ease in creating and sharing the content contribute to the problem.

Because of this, identifying false news calls for a collaborative effort from numerous stakeholders, including journalists, researchers, educators, policymakers, and technological firms. In the process of creating tools and strategies to identify fake information and stop it from spreading, data science and machine learning can be quite important.

IV. PROPOSED METHODOLOGY

On Kaggle or many other websites, we can find a large number of datasets for fake news identification. The dataset was obtained through Kaggle. Label column that contains 1 for fake news and 0 for authentic news is present in the datasets. A small number of data are withdrawn from the data set for manual testing, which is used to gauge the system's accuracy. Subject, title, text, and date make up the data set. Since the date are not necessary to determine if the data set is fake or real, we remove that column.

Dataset used consists of the headings as given below

id: unique id for the article
title: summary for the article
author: who wrote it
text: content of the article; could be incomplete
label: shows if the news is true or false:
1: Fake news
0: Real News

Text data contains a number of unsuitable words, special symbols, and other factors that prevent us from using it directly. The ML algorithm has a very difficult time identifying patterns in that text if it is utilised straight without being cleaned, and it occasionally produces an error. Therefore, we must always sanitise text data first.

Next, we must pre-process our data, stemming and removing stop words, and punctuation. After that, we split the data into training and testing groups, with training data making up 80% of the total and testing data making up 20%. Finally, the data must be vectorized using the TF IDF vectorizer before being fed to the machine learning models.

A. IMPLEMENTATION

There are various processes involved in creating a machine learning-based system for detecting fake news. Here is the complete description of the procedures implemented for building our model:

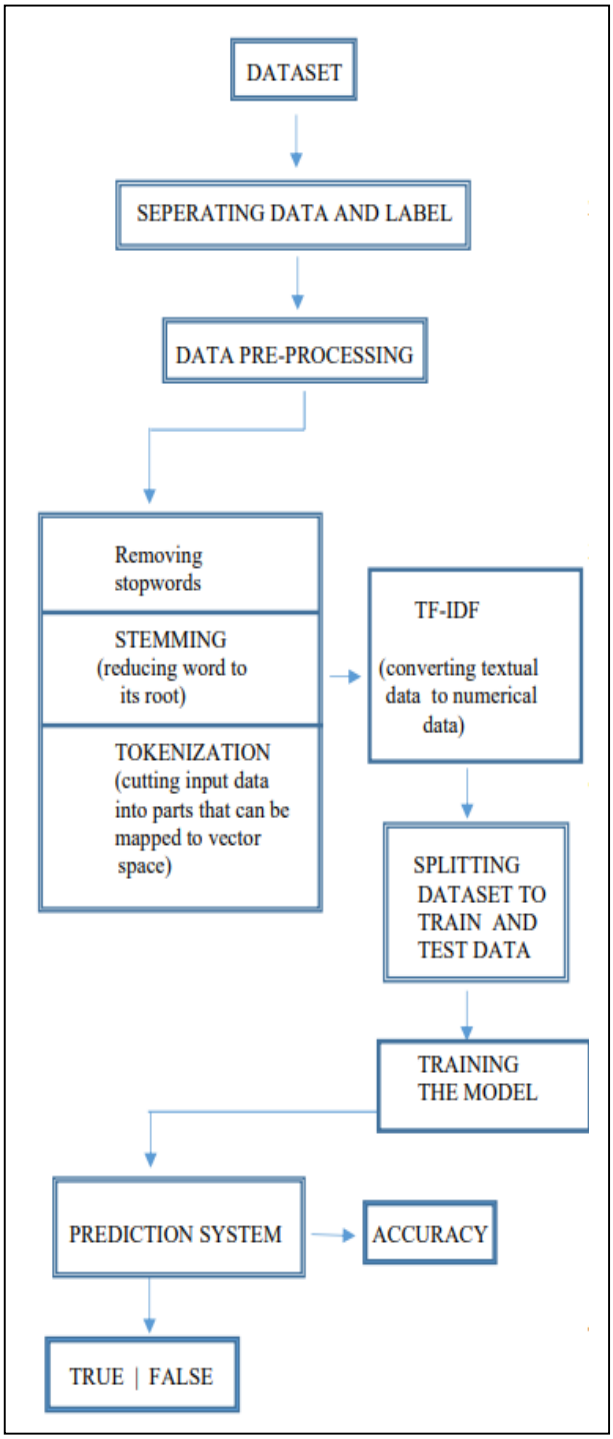

Fig.2. Steps in model training

1. Data Collection: Gather a sizable collection of news stories from various sources, both authentic and fraudulent.

2. Data preprocessing: Remove any extraneous information from the data, including stop words, punctuation, and other irrelevant information. To normalise the text, you may also use alternative methods like stemming and lemmatization.

Stemming: process to reduce any word to its root word.

Example: hung hanged hanging === hang

Steps: lower case, splitting, removing "stopwords", stemming.

3. Feature extraction: It is the process of removing related features from preprocessed data. These characteristics may consist of sentiment analysis, the presence of particular entities (such as persons, locations, and organisations), the frequency of particular words or phrases, or both. We used feature extraction and selection algorithms from the sci-kit learn python libraries in this module.

- TF-IDF: It detects the "relative frequency" of a word in a record in comparison to the frequency of that word across all papers.

  TF stands for Term Frequency. It detects the frequency of appearance of any word in a record which may be greater in a big document than in the short one.

  IDF stands for Inverse Document Frequency, it states that frequently occurring words are not important and hence can be removed. The words "a," "the," "on," and "of," "an," for example, are frequently used hence are of little weightage. IDF is more distinctive the more valuable it is.

4. Model evaluation: assess the trained model's performance on a test dataset. To assess the performance, you may look at the accuracy of the model, precision score and recall and also F1 score.

5. Deployment: As soon as you have a model that works well, utilise it to identify fake news in real time by putting it in a production environment.

The implementation of each phase varies depending on the amount of the dataset, the precise machine learning methods employed and the performance needs of the system.

B. ALGORITHM

1. LOGISTIC REGRESSION

It is a supervised ML algorithm that is used to predict authenticity of any statement. It should give discrete output as 0 or 1 instead gives probabilistic value between 0 or 1. Higher the probability greater chance of to be in class 1.

Logistic function curve is a sigmoid curve. If we consider Y as the threshold value then, all the values above Y are false and all the values below are true. As, we already said through logistic function we are able to get probabilistic value of 1 or 0 for any news.

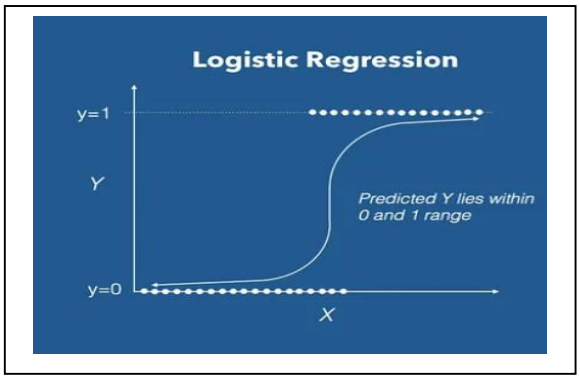

Fig.3. Graphical Representation [16]

2. SVM

The aim of SVM is to identify the optimum decision boundary that separates the various classes in the data by the widest margin. SVM finds a hyperplane that is least likely to make classification mistakes on fresh, untested data by maximizing the margin.

SVM can handle datasets that are both linearly separable and non-linearly separable. SVM can more easily locate a hyperplane that can separate the classes by translating the data into a higher-dimensional feature space.

Main ideas behind SVM are:

- Maximum margin separator: line or hyperplane that maximizes the separation between the separator and the training data to introduce a margin slab.
- Soft margin separator: Draw the optimal separator line when data with several labels are mixed up, taking into consideration instance within the margin slab.
- **Kernel trick**: Allow higher-order polynomials or even non-polynomial functions for more sophisticated models where the data separation border is non-linear.

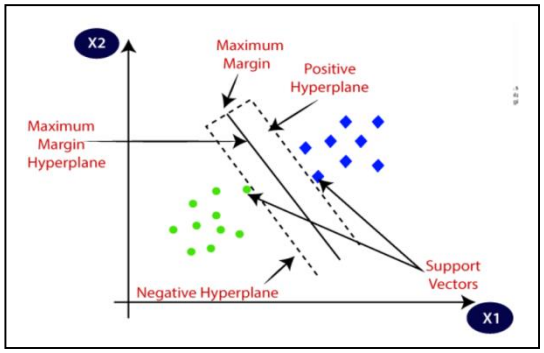

Fig.4. Graphical Representation [17]

C. TOOLS AND LIBRARIES

Logistic Regression-based false news detection system can be implemented utilizing a variety of tools and packages. Here are a few that are frequently used:

1. Python: Python offers a broad collection of libraries for working with text data, machine learning methods, and web development. It is a popular programming language for machine learning and data science activities.

2. Flask: It is a python package that is used to create web application. This framework is simple but at the same time extensible. Getting started with it is simple because there isn't a steep learning curve.

3. TF-IDF: It is a tool that computes the TF-IDF count for each word in a record and represents the resulting scores as a vector. This vector acts as the input for machine learning algorithms. It is a common technique used to evaluate the importance of words in a document or corpus. It measures term frequency, and how much it discriminates against more common words (inverse document frequency). TfidfVectorizor and TfidfTransformer class from `sklearn.feature_extraction.text` module can be used.

4. NLTK: A natural language processing toolkit that includes instruments for text processing tasks including tokenization and stemming.

5. Matplotlib: A tool for creating graphs, histograms, and other data visualisations.

6. Scikit-learn: This machine learning library offers resources for developing categorization models, such as logistic regression

Fig.5. Scikit learn Logo [14]

7. NumPy: A library for mathematical operations used in computation and numerical analysis.

Fig.6. NumPy Logo [15]

8. Common terms called ‘stopwords’ are frequently purged from text data during natural language processing in order to reduce noise and boost analysis performance. The following words—"the," "a," "an," "in," "of," "and," "to," "for," "that," and "is"—are typically regarded as having little to no meaning on their own. By removing these terms from the text data, it may be possible to concentrate on the analysis of the text's most significant and relevant words. It is crucial to remember that the list of ‘stopwords’ can change based on the application and domain.

9. sklearn.metrics` is a module in the scikit-learn library that provides a wide range of metrics for evaluating machine learning models. It includes functions for classification, regression, clustering, and pairwise metrics. Some of the commonly used metrics in this module are:

   - Classification metrics
   - Regression metrics

   These help to evaluate the performance of any ML model on various datasets. By comparing the metrics of different models, we can determine which model performs the best on the given dataset and choose that as our final model.

   Overall, from data pre-processing and feature extraction to model creation, evaluation, and deployment, the aforementioned modules and libraries can be used to implement a false news detection system

V. RESULT ANALYSIS

A logistic regression and SVM algorithm result analysis can shed light on the effectiveness of the system and possible areas for development. For instance, if the system has high recall but low accuracy, it might be missing some instances of bogus news that could be detected by more advanced characteristics or a more complex model. The system may be recognizing too many false positives if it has a high recall but low precision, which might be decreased by implementing more stringent thresholds or additional validation stages.

Overall, a fake news detection system's result analysis utilizing logistic regression and SVM can help to inform the creation of more precise and trustworthy models and contribute to ongoing efforts to stop the spread of misleading information.

On testing the trained dataset using SVM and Logistic Regression algorithm we get the following results

Logistic Regression:

- Accuracy – 97.67
- F1 Score – 97.7
- Precision Score – 96.38
- Recall Score – 99.07

SVM (Support Vector Mechanism):

- Accuracy – 98.9
- F1 Score – 98.9
- Precision Score – 98.64
- Recall Score – 99.17

Hence, we could see from above observation that SVM algorithm offers more accuracy than Logistic Regression
Accuracy (SVM) >>> Accuracy (Logistic Regression)

Hence, SVM has better performance than Logistic Regression algorithm.

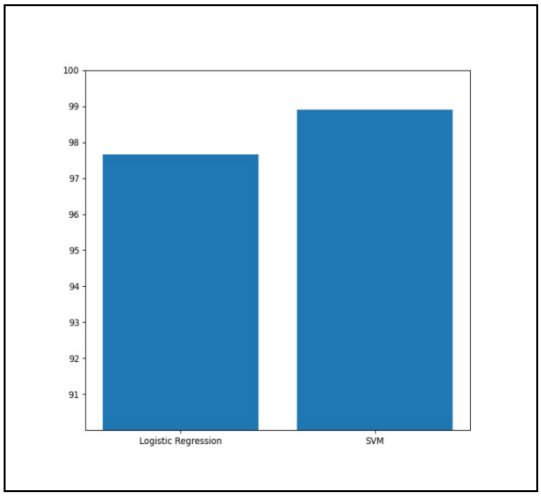

Fig.9 Comparison Bar Graph

CONFUSION MATRIX

It is a graphical representation used to assess the effectiveness of a binary class system, such as a system for spotting false news. It displays four outcomes as total number of system-generated categorization.

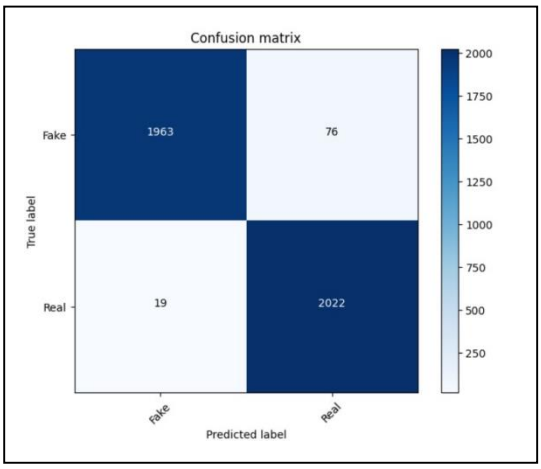

Fig.7 CONFUSION MATRIX
(Logistic Regression)

Here we observe that: true positive=1968, true negative=2022, false positive=19, false negative=76

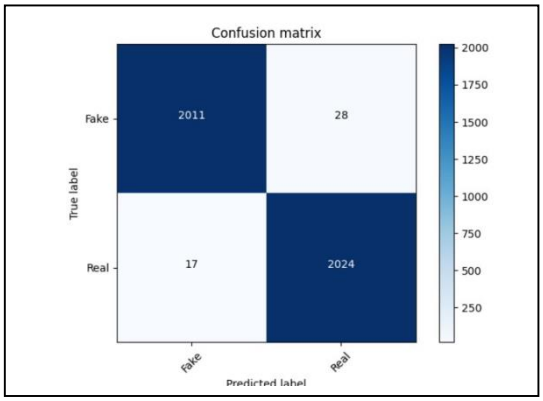

Fig.8 CONFUSION MATRIX
(SVM)

Here we observe that: true positive=2011, true negative=2024, false positive=17, false negative=28

VI. FUTURE SCOPE

Boost accuracy: The current detector version is based on the SVM Classifier, a straightforward machine learning technique. Although it has demonstrated a respectable level of accuracy, more sophisticated models might perform better. More sophisticated models, including neural networks or ensemble approaches that aggregate numerous models for greater accuracy, could be used by developers.

Add more domains: The Fake News Detector was developed using news articles from a particular dataset. It might be enhanced to operate with various text formats, such social network postings or customer evaluations. The project's scope may be significantly increased, but this would necessitate retraining the model on a new dataset.

Create a browser extension: A browser plugin that enables users to rapidly assess the legitimacy of news stories they come across while exploring the internet might be made.

VII. CONCLUSION

SVM algorithm has better performance than logistic regression, we built a model on SVM. SVM based fake news detection system is a useful tool for spotting and preventing the spread of misleading information. Hence we built a model based on SVM algorithm for predicting fake news.

Combining approaches for data collection, preprocessing, feature engineering, model training, evaluation, and deployment can be used to construct such a system. In the process, the right tools and libraries are chosen and used, such as Python, Scikit-Learn, Pandas, NLTK, Flask, and Django, to complete the given task. With the help of this implementation, the system may be trained to spot patterns in text data and determine whether a news piece is likely to be fake or legitimate.

VIII. REFERENCES

[1] Pappu, Sasibhushana & Sangeetha, N. & Maniteja, K. & Dilip, Y. & Abhishek, A.. (2022). Analysis of Classifiers for Fake News Detection: A Machine Learning Perspective. International Journal for Research in Applied Science and Engineering Technology. 10. 1573-1580. 10.22214/ijraset.2022.44925.

[2] Hasan Alrikabi, Ali & Ateaa, Heba. (2022). Fake News Detection Based on the Machine Learning Model. Design Engineering (Toronto).

[3] Sharma, Uma & Saran, Sidarth & Patil, Shankar. (2020). Fake News Detection Using Machine Learning Algorithms. 2320-2882.

[4] Ahmed, Alim Al Ayub & Aljarbouh, Ayman & Donepudi, Praveen & Choi, Myung. (2021). Detecting Fake News using Machine Learning: A Systematic Literature Review. Psychology (Savannah, Ga.). 58. 1932-1939. 10.17762/pae.v58i1.1046.

[5] Iftikhar Ahmad, Muhammad Yousaf, Suhail Yousaf, Muhammad Ovais Ahmad, "Fake News Detection Using Machine Learning Ensemble Methods", Complexity, vol. 2020, Article ID 8885861, 11 pages, 2020.

[6] Devi, Ms & Priyanka, R & Surendra, P & Priyanka, B & Nikhila, Ch & Mohammad, Sameer. (2019). FAKE NEWS DETECTION USING MACHINE LEARNING. SSRN Electronic Journal. 6. 104-109.

[7] Bonny, Afrin Jaman & Bhowmik, Puja & Mahmud, Md. Shihab & Sattar, Abdus. (2022). Detecting Fake News in Benchmark English News Dataset Using Machine Learning Classifiers. 1-8. 10.1109/ICCCNT54827.2022.9984461.

[8] Sharma, K., & Gupta, B. B. (2019). Towards privacy risk analysis in android applications using machine learning approaches. International Journal of E-Services and Mobile Applications (IJESMA), 11(2), 1-21.

[9] Sharma, K., & Gupta, B. B. (2018). Taxonomy of distributed denial of service (DDoS) attacks and defense mechanisms in present era of smartphone devices. *International Journal of E-Services and Mobile Applications (IJESMA)*, *10*(2), 58-74..

[10] Amer, Eslam & Kwak, Kyung & El-Sappagh, Shaker. (2022). Context-Based Fake News Detection Model Relying on Deep Learning Models. Electronics. 11. 1255. 10.3390/electronics11081255.

[11] Thompson, Robyn & Joseph, Seena & Adeliyi, Timothy. (2022). A Systematic Literature Review and Meta-Analysis of Studies on Online Fake News Detection. Information. 13. 527. 10.3390/info13110527.

[12] Shahid, Wajiha & Jamshidi, Bahman & Hakak, Saqib & Isah, Haruna & Khan, Wazir & Khan, Khurram & Choo, Kim-Kwang. (2022). Detecting and Mitigating the Dissemination of Fake News: Challenges and Future Research Opportunities. IEEE Transactions on Computational Social Systems. PP. 1-14. 10.1109/TCSS.2022.3177359.

[13] Shrivastava, G., Kumar, P., Ojha, R. P., Srivastava, P. K., Mohan, S., & Srivastava, G. (2020). Defensive modeling of fake news through online social networks. IEEE Transactions on Computational Social Systems, 7(5), 1159-1167.

[14] Scikit learn Logo. adapted from https://vectorseek.com/vector_logo/scikit-learn-logo-vector/

[15] NumPy Logo. adapted from https://www.google.com/search?q=numpy&tbm=isch&chips=q:numpy,g_1:icon:rXh3T8znNZg%3D&hl=enGB&sa=X&ved=2ahUKEwiZsu6ZuvXAhU9jtgFHaneDHEQ4lYoB3oECAEQNw&biw=1519&bih=722#imgrc=zoEApvJJ4gzfYM

[16] Graphical Representation. Adapted from https://towardsdatascience.com/introduction-to-logistic-regression-66248243c148

[17] Graphical Representation. Adapted from https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

[18] Features of fake news. Adapted from https://www.kdnuggets.com/2017/10/guide-fake-news-detection-social-media.html