International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

Fake News Detection using Passive Aggressive and Naïve Bayes

Sakshi Puranik[1], Suzan Khan Pathan[2], Ronak Patel[3], Prof. Yogita Shelar[4]

[1][2][3] *Student, Department of Information Technology, Atharva College of Engineering, Mumbai*
[4] *Professor, Dept. of Information Technology, Atharva College of Engineering, Mumbai*

Abstract – *In this era where the propagation of content through social media is at its peak, the population has started relying heavily on the internet for rapid deliverance of news, making it the primary source of news now. But a downside to this prompt news delivery is now being wrongly used by individuals, organizations, and parties, by spreading fake news to serve their propaganda, to benefit from the confusion caused by it, or to just misguide a large scale of the population with a single click. Not only this, malicious sites use lucrative headlines to lure readers into clicking on the news article to generate revenue. Hence, in this paper, we aim to tackle this issue with the help of Machine Learning and Natural Language Processing concepts to differentiate between fake and real news across the internet.*

Key Words: **Machine Learning, Natural Language Processing, Fake News, Passive Aggressive, Naïve Bayes, Classification Algorithms**

1. INTRODUCTION

The chokehold fake news has on the people should not be taken lightly, and measures must be taken to nip this problem in the bud. This project makes use of Artificial Intelligence, Machine Learning algorithms, and Natural Language Processing techniques to make a model that can uncover records that are, with high probability, fake news stories and articles. So far, techniques have been enforced to block out news from sources that are known to facilitate the spread of fake news, but people can also post and publish articles anonymously or under fake names to avoid taking responsibility for it. In such cases the blocking system doesn't stand to do a good job, hence making it necessary to rely on ML and NLP-based solutions, for those responsible citizens that want to confirm the validity of the news they received before circulating it.

2. LITERATURE SURVEY

2.1. Media-Rich Fake News Detection: A Survey

S.B Parikh and P.K Atrey gave a comprehensive approach to the identification of a news story in the modern generation combined with the diverse content types of a news article and its impact on readers. They also gave a comprehensive analysis of already existing approaches to deal with fake news detection and their heavy reliance on text-based analysis. They also introduced various sources from which reliable datasets can be procured to create a dependable model. They concluded their paper by identifying the main areas in which research can be progressed to achieve better results in this endeavor.[1]

2.2. Fake News Detection using One-Class Classification

P. Faustini and T. Covões proposed to detect fake news by using purely text features regardless of the source and platform of the news, and independent of the language used to convey the news on the platform. They experimented with various datasets that contained both texts and social media posts in various languages like German, and Slavic, and compared their analysis. Their method produced good accuracy when compared to benchmarks that previous models had already set. They compared the results obtained through a custom set of features and with other popular techniques when dealing with natural language processing, such as bag-of-words and Word2Vec. [2]

2.3. A *Tool for Fake News Detection*

B. Al Asaad and M. Erascu proposed the use of Supervised Machine Learning Algorithms to detect fake news. They used datasets comprising of fake and real news to train machine learning models using the Scikit-learn library in Python. For feature extraction, they used Bag of words and Term Frequency-Inverse Document Frequency (TF-IDF), and Bi-gram frequency. They further used two different approaches, which are probabilistic classification and linear classification. They performed this on the title and the content to identify fake news correctly.[3]

2.4. *Detecting Misleading Information on COVID-19*

Mohamed K. Elhadad, Kin Fun Li, and Fayez Gebali propose a misleading information detector constructed with information from renowned institutions such as the World Health Organization, UNICEF, and the United Nations as their groundwork as well as epidemiological material which was collected from a range of fact-checking websites. This increased the reliability of the dataset, and consequently of the model, greatly. They also performed a 5-fold cross-validation to check the validity of the collected data and report the evaluation of twelve performance metrics. [4]

3. METHODOLOGY

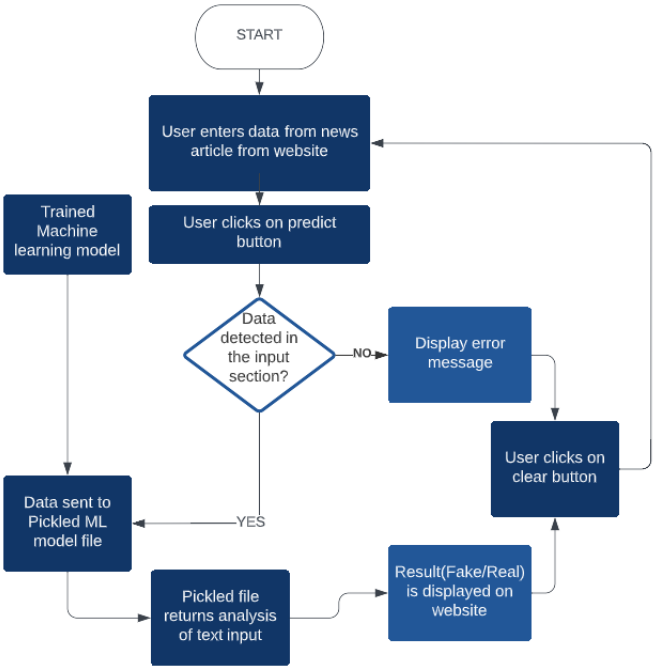

Fig-1: System Design

The data flow between the front-end and back-end moves in the way described in the above diagram. The connection for the same can be done by using many technologies (example: Django) but with a project such as ours, with a definitive task and on a small scale, Flask was used.

3.1. Algorithms

We used the following supervised machine learning algorithms in conjunction with our proposed methodology to evaluate the performance of fake news detection classifiers.

3.1.1 Passive-Aggressive Machine Learning Algorithm

Passive-Aggressive algorithms are usually used for training on a large scale. It is one of the few available 'online-learning algorithms 'in machine learning. In such algorithms, the input data comes in sequential order and the model is updated step-by-step, as opposed to batch learning, where the entire training dataset is used at once. This is used in situations where the amount of data is extremely huge and it is not feasible to process the entire dataset in one go. It is also handy in scenarios where the new data is generated constantly and frequent changes are required. It does not require a learning rate, but it does require a regularization parameter.

Important parameters of this algorithm are:

C: regularization parameter or maximum step size. It is of type float and denotes the penalization the model will make on an incorrect prediction. Its default value is 1.0.

max_iter: it is the maximum number of passes a model makes over the data which is used for training it and only impacts behavior in the fit method. It is of type Int and its default value is 1000.

fit-intercept: it describes whether the intercept should be estimated or not. It is of type Bool and by default is set to True.

tol: stopping criterion. If the value of this parameter is set to None, the model will stop. It is of type Float or None and its default value is 1e-3. [5]

The Passive-Aggressive model has two main functions:

Passive: If the prediction is correct, keep the model as it is and do not make any changes to it, i.e., the data in the example is not enough to cause any changes to the model.

Aggressive: If the prediction is incorrect, make changes to the model, i.e., some changes to the model might improve it and make future predictions more accurate.

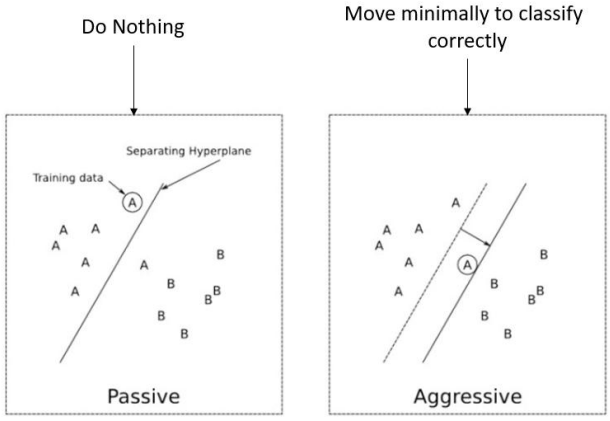

Fig-2: Working of Passive Aggressive

3.1.2 Naïve Bayes Machine Learning Algorithm

Naive Bayes classifier is a class of supervised machine learning algorithms based on **Bayes' Theorem**. All the algorithms belonging to this class share the same main concept, i.e., no two features that are used in the classification process are dependent on each other.

Bayes' Theorem is used to calculate how likely is it that an event will occur given that the probability of another event has already occurred. Mathematically it is represented as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

where A and B are events mentioned above and P(B) is not equal to 0.

Event B is termed as evidence. P(A) is the priori of A, i.e., the probability of an event before the evidence is seen. Here, the evidence, which is the event B in the above formula, is an attribute value of an unknown instance. P(A|B) is a posteriori probability of B, i.e., probability of event after evidence is seen. [6]

The core concept that Naïve Bayes functions on is that each feature that is used to classify an entity is not dependent on any other feature and makes an equal contribution to the outcome of the classifier.

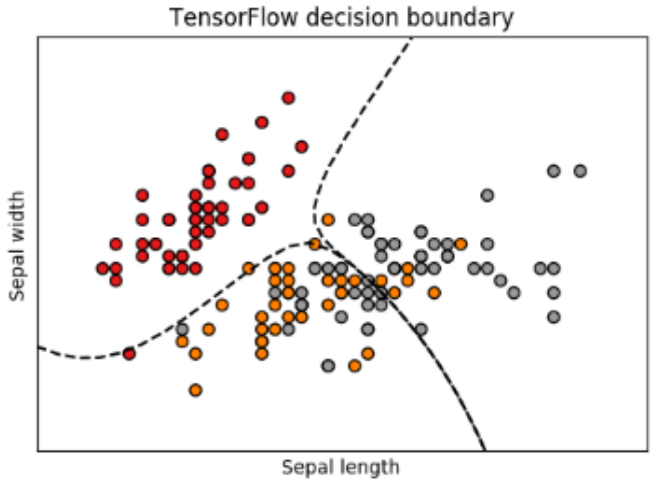

Fig-3: Naïve Bayes Classifier

3.1. Algorithms Training Process

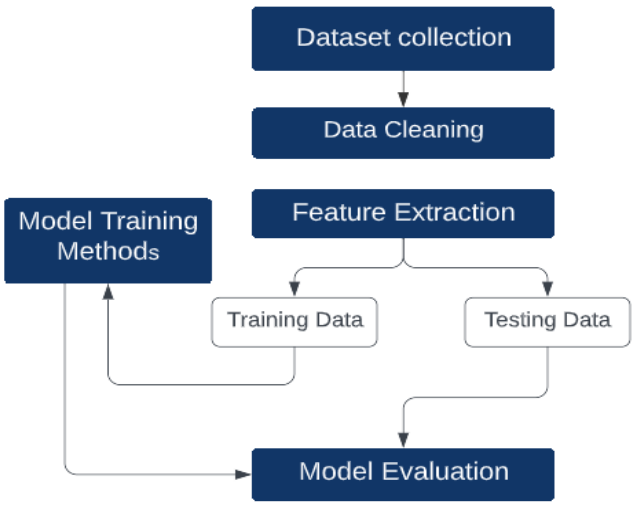

Fig -4: Model Training Process

3.1.1. Dataset Collection

We selected datasets from Kaggle [7] which contained 6335 rows of labeled fake and real news that covered a wide range of topics from multiple domains such as politics, technology, sports, entertainment, etc. Furthermore, we made sure that the dataset was balanced, i.e., it contained equal instances of both fake and real news so the model isn't biased.

3.1.2. Data Cleaning

Before the dataset could be used to train the models, we needed to clean it. The article's unwanted variables such as author and serial number were filtered out. Articles with no body text or body text containing less than 20 words were filtered out as well. Lastly, the dataset was checked for any Null values and duplicates and that every entry of an article had a label with it, without any spelling errors.

3.1.3. Feature Extraction

After the relevant attributes were selected in the data cleaning and exploration phase, we moved on to extracting linguistic features from the data. This involves converting the textual characteristics of data into numerical values to serve as input to the models to be trained. This was done using Bag of words (BoW) and Term-Frequency-Inverse Document Frequency (TF-IDF).

BoW is a score that is used to represent how frequently a word occurs in a document. It involves 2 things: a vocabulary of known words and a measure of the presence of known words. The model is only concerned about the presence of known words in the document, and not where in the document it is present.

In TF-IDF, the main concept is that if a word occurs frequently in the document, but also across many other documents in the dataset, the word might just be a common word that occurs frequently (for example – 'the'), not because it is relevant or important. The TF (term-frequency) is the count of how many instances of a word are in a document. The IDF (Inverse Document Frequency) calculates how frequently the word is used across documents. If a word is common and is appearing many times across documents, the IDF score for that word would be closer to zero. Hence, these scores have the effect of highlighting words that are distinct (contain useful information) in a given document.

3.1.4. Dividing Data for training and testing

The dataset was split into a ratio of 67/33 where the larger chunk was used for training the models and the remaining to check the accuracy of the table. The articles were shuffled to ensure that a fair allocation of fake and real articles was present in the testing and training instances. This was done using random state.

3.1.5. Model Evaluation

The learning algorithms are trained on the dataset using different hyperparameters to increase accuracy for a given dataset, with a good balance between variance and bias.

Then they are evaluated to measure their performance based on various parameters, most of which were based on the confusion matrix. A confusion matrix is a tabular representation of a classification model performance on the test set, which consists of four parameters: true positive, false positive, true negative, and false negative.

		Predicted class: P	Predicted class: N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

Fig -5: Confusion Matrix

Accuracy: It is the metric that is most used among all the parameters to determine the efficiency of an ML model. It represents the percentage of correctly predicted observations, be it true or false. The formula for calculating the accuracy of a model is as follows :

$$Accuracy = (TP+TN)/(TP+TN+FP+FN)$$

Although accuracy is a good parameter to evaluate the working of a model, it has certain limitations and cannot be used alone to provide a complete analysis. For example, while training a classification model like ours, if a true article is predicted as false, or a false article is predicted as true by our model, it can have detrimental effects on its overall performance. Hence, it is advised to use more parameters along with accuracy to ensure that the quality of the model trained is good. That is why, in this project, we have used 3 more parameters that can be derived through the confusion matrix which are Recall, Precision, and F1-Score.

Recall: It is used to calculate the total number of positive classifications out of all factually true records. In our project, it represents the number of articles identified as true out of the total number of true articles in the dataset.

$$Recall = TP/(TP+FN)$$

Precision: In contrast to Recall, the Precision parameter is used to depict the ratio of true positive to all the events that were identified as true by the model. In our project, precision shows the number of articles that were true in the dataset out of all the articles that were identified to be true by our model:

$$Precision = TP/(TP+FP)$$

F1-Score: The F1-score depicts the trade-off between precision and recall by calculating the harmonic mean between each of the two. It takes into account both the false positives and false negatives (i.e, all the records wrongly identified by the model). F1-score can be mathematically represented by:

$$F1\text{-}Score = 2*(Precision*Recall)/ (Precision + Recall)$$ [8]

3. RESULTS AND DISCUSSION

The two models in consideration, namely Naïve Bayes and Passive-Aggressive were compared and their accuracies recorded to identify which model fared better, to convert it into a pickled file and make the connection to our website.

The result for the same is given in table-1.

Table-1: Classification Result of Models

Model	*Accuracy*
Naïve Bayes	0.89
Passive Aggressive	0.93

Fig-6: Fake News detection parameters comparison

Fig-7: Real News detection parameters comparison

4. CONCLUSIONS

The primary aim of the research on fake news detection is to identify patterns in text that can help differentiate between legitimate news articles and fake news articles circulating on the internet. We extracted different textual features from the articles using bag of words and TF-IDF and used the feature set as an input to the models. The learning models were trained on a balanced dataset and then parameter-tuned to obtain optimal accuracy. Some models have achieved higher accuracy than others in the model evaluation phase.

Although a lot of research has been conducted in this field due to the harmful effects of fake news on people's lives, there is much room for improvement. News no longer propagates in only text media, but also in photos and videos that increase the complexity of the problem.

Hence, including samples of fake and real news in image and video format in datasets and training models in them could be the next step in this direction. Also, instead of creating a website for people to use to verify their news, converting it into a chrome extension will allow easier access to them, consequently encouraging a larger number of people to check the nature of their news before making any impactful decisions based on it.

REFERENCES

[1] S. B. Parikh and P. K. Atrey, "Media-Rich Fake News Detection: A Survey," IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2018.

[2] P. Faustini and T. Covões, "Fake News Detection Using One-Class Classification," 8th Brazilian Conference on Intelligent Systems (BRACIS), 2019

[3] B. Al Asaad and M. Erascu, "A Tool for Fake News Detection," 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 2018

[4] M. K. Elhadad, K. F. Li and F. Gebali, "Detecting Misleading Information on COVID-19," in *IEEE Access*, 2020

[5] Crammer, Koby & Dekel, Ofer & Keshet, Joseph & Shalev-Shwartz, Shai & Singer, Yoram. (2006). Online Passive-Aggressive Algorithms. Journal of Machine Learning Research. 7. 551-585.

[6] Chen, H., Hu, S., Hua, R. *et al.* Improved naive Bayes classification algorithm for traffic risk management. *EURASIP J. Adv. Signal Process.* **2021,** 30 (2021)

[7] Dataset from Kaggle: https://www.kaggle.com/datasets/hassanamin/textdb3

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