Fake News Detection using Machine Learning

Jasmine Shaikh
Department of Electronics and Telecommunication,
K.J. Somaiya College of Engineering,
Mumbai, India
jasmine.shaikh@somaiya.edu

Rupali Patil
Department of Electronics and Telecommunication,
K.J. Somaiya College of Engineering,
Mumbai, India
rupalipatil@somaiya.edu

2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC) | 978-1-7281-8880-5/20/$31.00 ©2020 IEEE | DOI: 10.1109/iSSSC50941.2020.9358890

***Abstract* — This work helps us to detect the accuracy of the fake news using different classification techniques. Fake news is significantly affecting our social life, in fact in every field mainly in politics, education. In this work, we have presented the solution for Fake news problem by implementing fake news detection model by using different classification techniques. Fake News Detection becomes complicated when it comes to resources. Resources like datasets are limited. In this model, we have used classification techniques like Support Vector Machine(SVM), Naïve Bayes, Passive Aggressive Classifier. Output of our model using feature extraction techniques as Term Frequency-Inverted Document Frequency (TF-IDF) and Support Vector Machine (SVM) as classifier, has accuracy of 95.05%.**

Keywords — ***svm, classifier, naive bayes, passive aggressive classifier.***

I. INTRODUCTION

In the recent years, Social Media has been dominant in everyone's life. Fake news spreads mostly through social media. Fake news is threat to the politics, finance, education, democracy, business. Although fake news is not a new problem but today humans believe more in social media which leads to believe in fake news and then spread of the same fake news. It is becoming tough nowadays to distinguish between true and false news which creates problems, misunderstanding. It is difficult to manually identify fake news, its only possible when the person identifying the news has a vast knowledge on the topic of news. Due to recent advancements in computer science, it is easier nowadays to create and spread fake news but it is considerably hard to distinguish the information as true or false. This fake news can affect some products, business if fake news is spread about the products. In politics too, fake news can affect someone's career. “2019 has been a unique year where fact checkers continuously kept moving from one event to the other, and this has been the busiest year for us so far,” said Jency Jacob, managing director at Mumbai-based fact-checking website BOOM, which works with Facebook to check stories and tags specific posts spreading misinformation on the platform. We have compare three different supervised classification technique, Naïve Bayes Classifier, Support Vector Machine (SVM), Passive Aggressive Classifier. We have used a dataset which contains real and fake news and it yields best results.

II. LITERATURE REVIEW

The main motive behind our work is to find best classification algorithm for detecting the fake news and calculating its accuracy. We have studied different classification algorithm and in our model we have used support vector machine, Passive Aggressive Classifier, Naive Bayes. Out of these three, support vector machine gives highest accuracy but time required for svm is high as compared to passive aggressive, naive bayes. A brief review of papers related to fake news detection was done. Due to exponential rise in the use of social media nowadays, it has become a very easy platform to spread fake news as the news reaches the social media user in no time. And it is often found out that people tend to believe in the news. Fake news is the major concern today. As the dataset we used contains text, various NLP models were applied to convert data into required form for data modelling. Veronica Perez-Rosas Bennett Kleinberg, Alexandra Lefevre, Rada Mihalcea [1] Proposed a model to identify fake news automatically in online news. They have developed computational models and resources for fake news detection. They used two datasets, one directly from internet and the second by combination of manually finding data and taking help from internet. Using these datasets, they performed various analysis to identify linguistic properties that are mainly present in fake content. Building a fake news detectors that rely on linguistic features, they have obtained an accuracy upto 78%. Manisha Gahirwal [2] proposed a SVM based model for fake news detection with five predictive feature like humour, negative effect, absurdity, grammar and punctuation. It aimed to determine the authenticity of the content of news article. They first took the URL of the article that is to be detected as fake or not .Then text is extracted from the URL. Text is then passed on to various data preprocessing units. The factors determining whether the news is fake or not are the stance of article and results of google search. Their model achieved an accuracy of 87%. Hadeer Ahmed [3] introduced a machine learning model and n-gram analysis for fake news detection. Best results were yield using Term Frequency-Inverted Document Frequency (TF-IDF) technique, and Linear Support Vector Machine (LSVM) classifier. Accuracy achieved was 92%. MyKhailo Granik [4] followed a simple approach of using naive Bayes classifier for fake news detection. They implemented it a software and tested it against a data set containing Facebook news posts. They accomplished grouping exactness of 74% around. Horne et al. [5] demonstrated it is evident to recognize phony and genuine news stories dependent on the perceptions, that phony news titles have lesser stop-words and things, while having more action words and things. Wang et al. [6] presented LIAR, another dataset that can be utilized for programmed counterfeit news recognition. In spite of the fact that LIAR is a gigantic dataset regarding size however dissimilar to other informational collections, it doesn't contain full

978-1-7281-8880-5/20/$31.00 ©2020 IEEE

articles. It contains 12800 physically marked short articulation picked from politicalFact.com. Rubin et al. [7] proposed a model to recognize humor and parody news stories. They noticed 360 Satirical news stories in spaces science, diversion, civics, business. Their SVM model were based on features like Punctuation, Humour, Grammar, Absurdity and Negative Affect. The most noteworthy exactness for example 90% was achieve utilizing mixes of three highlights Absurdity, Grammar, and Punctuation. Volkova et al. (2017) [8] fabricated a model that predicts 130K news stories as checked or phony. Subtypes of phony news were misleading content, tricks, parody, publicity. Conroy et al. (2015) [9] proposed a mixture approach that consolidates both semantic prompts and AI with network-based conduct information . This crossover approach utilizes both sack of words (BOW) and n gram procedures to speak to information. Chhabra et al. (2011) [10] proposed a URL static component based identification strategy for recognizing noxious sites. IP addresses were considered as outside component centered. The creator has focussed on outer highlights, for example, IP addresses. Vector development VSM is picked as the URL vector model. The dataset utilized in this model comprises of noxious URLs downloaded from the 'Phishtank' , a phishing stage Anushaya Prabha [11] proposed a supervised learning model to detect the fake news articles published during 2015 and 2016 U.S. election cycle. Supervised machine learning algorithms like Passive Aggressive Classifier, *K*-Nearest Neighbours, Naive Bayes, Support Vector Machine were used. These models were trained using contents or the title of the article. They achieve highest accuracy using PAC classifier i.e. 94.63%.

III. METHODOLOGY

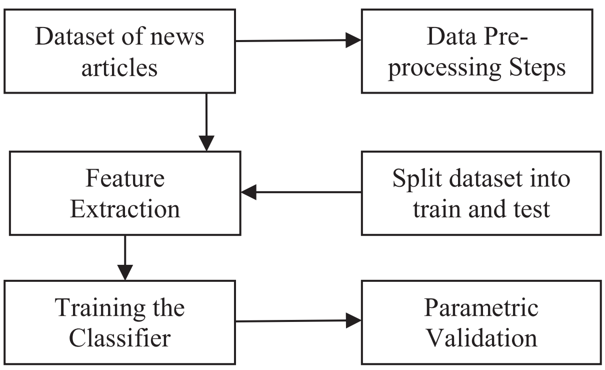

Fig.1. Flowchart for fake news detection

Text data requires preprocessing so that it can be converted into form appropriate for data modelling. There are various techniques widely used to convert text data and the one we used is Natural Language processing techniques (NLTK). The data preprocessing steps we used for news headlines and article are Stop Words Removal, punctuation removal, Stemming. This will decrease the real information size by eliminating the superfluous data that is available in the information. Stops words like as, a, the, an, are, as, at, for are used to construct sentences. They don't have any significance if used as a feature in text classification. Stops Words can be processed and filtered, so removing stop words is the very important step in NLP. We have used Natural Language Toolkit – (NLTK) library to remove stop word. Punctuation like comma just add meaning to sentence and must be filtered from text as they have no more importance. Stemming is a technique to remove suffixes and prefixes from a word. Stemming reduce the word to the base word, into original form. For example, "Knowing"," knew"," knows" will be reduced to "know".

We studied two different features selection methods Term Frequency (TF) and Term Frequency-Inverted Document Frequency (TF-IDF). TF identifies the importance of word based on its occurrence in a document. It describes how common a word is. Thus, document is represented by collection of words. IDF (Inverse Document Frequency) identifies rare word or in other words it describes how rare a word is. The Term Frequency-Inverted Document Frequency (TF-IDF) is a method that finds the importance of word in a document. A word importance increases with its frequency of appearance in the document, however, this is counteracted by the frequency of the word in the corpus. Word with a high TF-IDF score is important for the document. Firstly, we started preprocessing of the data set, by removing unnecessary and insignifiacnt words and characters from the data. The next step is feature Extraction using Term Frequency-Inverse Document Frequency. We investigated three different algorithms, namely, Support Vector Machines (SVM), Naïve Bayes, Passive Aggressive Classifier. We implemented these classifiers using Python Natural Language Toolkit (NLTK). We split the dataset into preparing and testing set. 80% of the dataset is utilized for preparing reason and 20% for testing reason.

A. Naive Bayes

Naïve Bayes is used for calculating conditional probability. It is derived from Bayes theorem. It states that "probability that something will happen, given that something else has already occurred" (Saxena, 2017). In Naïve Bayes, occurrence of one feature is independent of other feature. Naïve Bayes is a type of classifier and it's a supervised learning algorithm. The prediction occurs on the basis of probability. It makes quick predictions for the machine learning models. It works best with text classification. Naïve Bayes Classifier is used for multi-class and binary classifications. But the disadvantage of using it is, classifier fails to learn the relationship between features as it treats all features independent of each other. There are three types of naïve bayes model i.e. Gaussian, Multinomial, Binomial. Gaussian model follows normal distribution. Multinomial model is used mostly for text classification problems and prediction is on the basis of frequency of words. Binomial classifier is similar to multinomial, it is used for classification task. In our model, we have used Multinomial naïve bayes classifier.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \_\_\_\_\_\_\_\_ (1)$$

P(A|B): Probability of event A such that event B has already occurred.

B. Support Vector Machine (SVM)

A support vector machine (SVM), is a managed learning calculation. Hence, the model is constructed after it has already been trained. The main motive of SVM is to categorize new data that comes under. There is a decision boundary or hyperplane that splits dataset into two class. For the considered class, a point is chosen such that it is close to the opponent class. A line is drawn touching the point parallel to hyperplane. Hyperplane is drawn considering maximum margin. SVM are more accurate on smaller dataset. The disadvantage of using SVM on large dataset is training time is high.

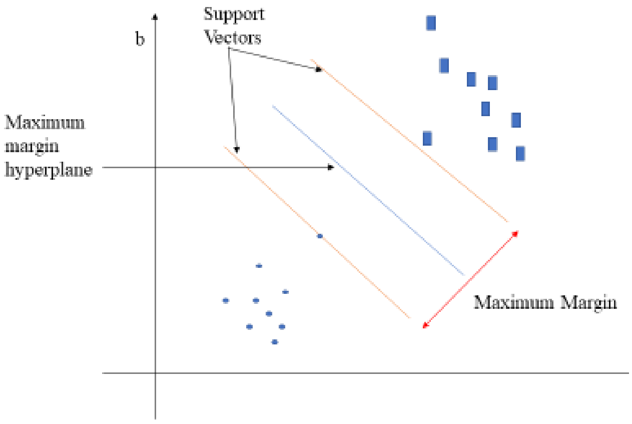

Fig.2. Classification of two different categories using hyperplane.

There are two types of SVM model. Linear SVM is used for linearly separable data. When Single straight line classify two classes of a dataset, such data is called as linearly separable data and the classifier used for this type of data is Linear SVM.

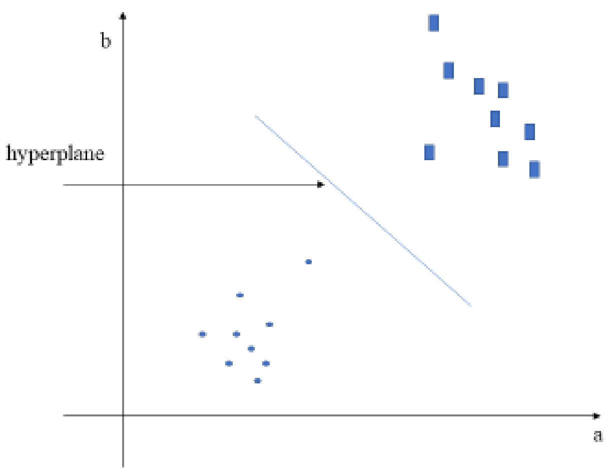

Fig.3. Single Straight line classifies two classes of linear dataset

In figure 3, as it is 2-D space, a single straight line can easily classify two classes. However, there can be multiple lines possible that can separate data. SVM chooses best hyperplane (straight line) considering maximum margin between support vectors for nonlinear data, a single straight line cannot separate data into two class. We have to consider 3-D space for it. For linear SVM, 2-D space was used.

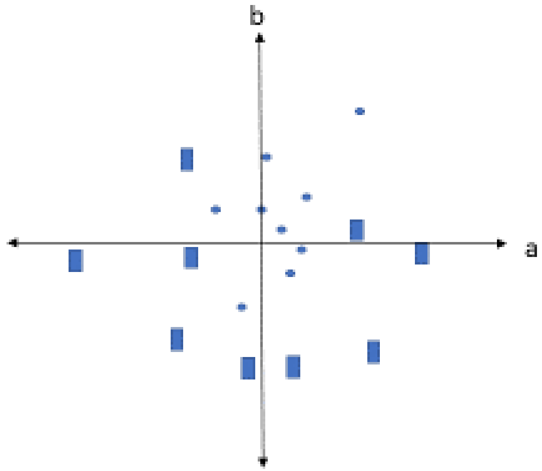

Fig.4. Nonlinear data

C. Passive Aggressive Classifier

Passive Aggressive algorithms are online learning algorithms and used for both regression as well as classification. It is easy to use and work fast as compared to SVM but does not provide high accuracy like SVM.It is mainly used to classify massive data. The algorithm remains passive for a correct classification outcome, and it turns aggressive if it is an incorrect classification, updating and adjusting.

IV. IMPLEMENTATION AND RESULTS

The dataset we took is a news dataset. Shape of dataset (6335,4). We have isolated the label column from the data frame. Preprocessing steps are applied and then we split the dataset into train and test data set. Feature are extracted using tfidf vectorizer. Using passive Aggressive Classifier on test set give accuracy of 92.9%.

Table.1. Confusion matrix using Passive Aggressive Classifier

n=1267	Predicted Yes	Predicted No
Actual Yes(638)	TP(587)	FN(51)
Actual No(629)	FP(40)	TN(589)

We have created a multinomial naive bayes classifier and predicting on tfidf train set. Accuracy of naive Bayes classifier 84.056%.

Table.2. Confusion matrix using Naïve Bayes Classifier

n=1267	Predicted Yes	Predicted No
Actual Yes(638)	TP(450)	FN(188)
Actual No(629)	FP(14)	TN(615)

From sklearn, we imported svm model. Linear kernel is used in creating SVM classifier and response was predicted for test data set. Accuracy of SVM classifier is 95.05%.

Table.3. Confusion matrix using SVM

n=1267	Predicted Yes	Predicted No
Actual Yes(638)	TP(598)	FN(40)
Actual No(629)	FP(48)	TN(581)

From the confusion matrix, we can calculate precision, F-measure, recall. Precision describes the number of class positive that actually belong to the class positive. Recall quantifies the number of positive class predictions that are made out of all positive examples present in dataset. F-Measure shows a balance between precision and recall.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F-Measure} = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \quad (4)$$

where,

TP = True Positives
FP = False Positives
TN = True Negatives
FN = False Negatives

Table.4. Performance parameters of our model

Classifier	Accuracy	Recall	Precision	F measure
Naïve Bayes Classifier	84.056%	70.53%	96.98%	81.666%
Passive Aggressive Classifier	92.9%	92%	93.62%	92.80%
Support Vector Machine	95.05%	93.73%	92.56%	93.141%

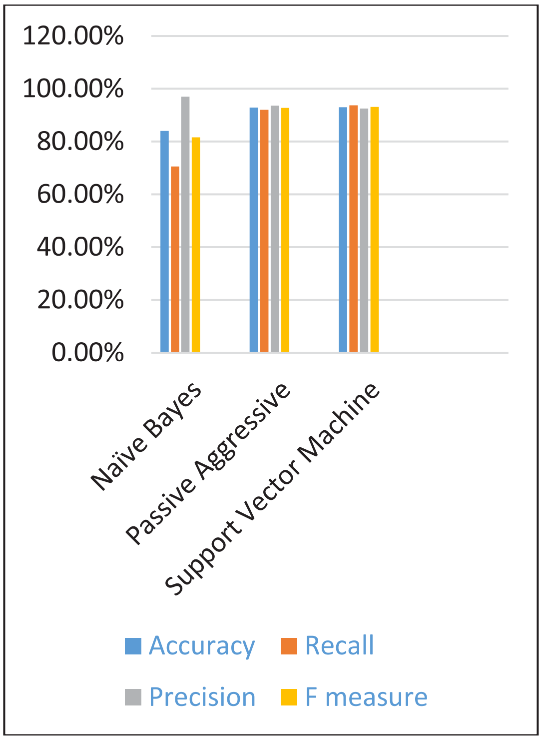

Fig.5. Accuracy, recall, precision, F measure of different algorithms used.

.

Table.5. Comparison with other existing approaches

Classifier	Accuracy
Naïve Bayes Classifier	84.056%
Passive Aggressive Classifier	92.9%
Support Vector Machine (proposed approach)	95.05%
Manisha Gahirwal [2] (SVM)	87%
Anushaya Prabha [11] (PAC)	94.63%
MyKhailo Granik [4] (naïve bayes classifier)	74%

From table 5, it is observed that SVM gives the highest accuracy 95.05%. Manisha Gahirwal [2] SVM model gives accuracy of 87%. They have used predictive features like negative absurd, grammar, etc.

V. CONCLUSION

Most of the fake news were shared on Twitter. Social Media is the major source of Fake news. In this work, we have introduced a discovery model for counterfeit news utilizing highlights extraction procedures and three diverse AI methods. The proposed model accomplishes its most noteworthy precision with SVM classifier. The most noteworthy precision score is 95.05%.

REFERENCES

[1] Pérez-Rosas, Verónica & Kleinberg, Bennett & Lefevre, Alexandra & Mihalcea, Rada. (2017). Automatic Detection of Fake News.

[2] Gahirwal Manisha et. al; International Journal of Advance Research, Ideas and Innovations in Technology ISSN: 2454-132X Impact factor: 4.295

[3] Traore, Issa & Saad, Sherif. (2017). Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. 127-138. 10.1007/978-3-319-69155-8_9.

[4] Mykhailo Granik, Volodymyr Mesyura, "Fake News Detection Using Naive Bayes Classifier", 2017 IEEEFirst Ukraine Conference on Electrical and Computer Engineering (UKRCON).

[5] Horne, B.D., Adali, S.: This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In: the 2nd International Workshop on News and Public Opinion at ICWSM (2017)

[6] Wang, W.Y.: Liar, Liar Pants on fire: a new Benchmark dataset for fake news detection. arXiv preprint (2017). arXiv:1705.00648

[7] Rubin., Victoria, L., et al.: Fake news or truth? Using satirical cues to detect potentially misleading news. In: Proceedings of NAACL-HLT (2016)

[8] S. Volkova, K. Shaffer, J.Y. Jang, N. Hodas, in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), vol. 2 (2017), vol. 2, pp. 647–653

[9] S. Volkova, K. Shaffer, J.Y. Jang, N. Hodas, in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), vol. 2 (2017), vol. 2, pp. 647–653

[10] N.J. Conroy, V.L. Rubin, Y. Chen, Proceedings of the Association for Information Science and Technology 52(1), 1 (2015)

[11] S. Chhabra, A. Aggarwal, F. Benevenuto, P. Kumaraguru, in Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (ACM, 2011), pp. 92–101

[12] Prabha, T. & Aisuwariya, T. & Kiran, M. & Vasudevan, Shriram. (2020). An Innovative and Implementable Approach for Online Fake News Detection Through Machine Learning. Journal of Computational and Theoretical Nanoscience. 17. 130-135. 10.1166/jctn.2020.8639.