International Journal of Innovations in Engineering and Technology (IJIET)
http://dx.doi.org/10.21172/ijiet.154.04

Comparative Study on Fake News Accuracy Prediction Using Naive Bayes, SVM and ANN

K.Suneetha, Balaga Susmitha, Dasari Poornima
Department of Computer Science and Engineering
Gayatri Vidya Parishad College Of Engineering For Women, Visakhapatnam, Andhra Pradesh, India

Abstract- Fake news is one of the biggest banes in our digitally connected world. Fake news spreads like wildfire and is impacting millions of people every day .With the advancement of digitized society, information is free for everyone and it can be manipulated by anyone. This lead to the enlargement in the spread of fake news and misinformation leading to many problems. In order to filter the information and to verify the trustworthiness of a news, we need a freely available tool. Our dataset is initially preprocessed and then it is tested using Machine Learning algorithms such as Naïve Bayes, SVM and ANN to predict the accuracy and compare them.

Keywords – Fake news, Naïve Bayes, SVM, ANN

I. INTRODUCTION

Fake news is an inaccurate, sometimes sensationalistic report that is created to gain attention, mislead, deceive or damage a reputation. Unlike misinformation, which is inaccurate because a reporter has confused facts, fake news is created with the intent to manipulate someone or something. Fake news can spread quickly. Social media websites in particular have proved to be an easy venue for distributing fake news. When fake news is used to spread propaganda, it can be dangerous. In addition to shaping public opinion and behavior, it can also cause mistrust, encourage dissent and deflect attention from real news. In response to this spread of fake news, here we are going to implement a project on detecting fake news accuracy to protect ourselves from wrong information.

II. PROPOSED ALGORITHM

2.1 Naïve Bayes Algorithm-

Naive Bayes classifier is a probabilistic machine learning model that's used for classification task. It is based on the Bayes theorem.Using Bayes theorem, we can find the probability of A happening, given that B has occurred. Here, B is the evidence and A is the hypothesis. The assumption made here is that the predictors/features are independent. That is presence of one particular feature does not affect the other. Hence it is called naive.

Naive Bayes algorithms are mostly used in sentiment analysis, spam filtering, recommendation systems etc. They are fast and easy to implement but their biggest disadvantage is that the requirement of predictors to be independent. In most of the real life cases, the predictors are dependent, this hinders the performance of the classifier. Naive Bayes algorithm can be defined as a supervised classification algorithm which is based on Bayes theorem with an assumption of independence among features**.**

Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred. Bayes' theorem is stated mathematically as the following equation:

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

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

- Basically, we are trying to find probability of event A, given the event B is true. Event B is also termed as evidence.
- P(A) is the priori of A (the prior probability, i.e. Probability of event before evidence is seen). The evidence is an attribute value of an unknown instance (here, it is event B).
- P(A|B) is a posterior probability of B, i.e. probability of event after evidence is seen.

2.2. Support Vector Machine-

The objective of the support vector machine algorithm is to find a hyper plane in an N-dimensional space (N — the number of features) that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyper planes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.

Hyper planes are decision boundaries that help classify the data points. Data points falling on either side of the hyper plane can be attributed to different classes. Also, the dimension of the hyper plane depends upon the number of features. If the number of input features is 2, then the hyper plane is just a line. If the number of input features is 3, then the hyper plane becomes a two-dimensional plane. It becomes difficult to imagine when the number of features exceeds 3. Support vectors are data points that are closer to the hyper plane and influence the position and orientation of the hyper plane.

2.3 Artificial Neural Networks-

Artificial neural networks (ANNs) are one of the various data mining techniques used to forecast the power output of a wind farm using meteorological information predicted by NWP models. ANNs attempt to copy the behavior of biological neural networks. In analogy to the structure of the brain, ANNs consist of single processing units called neurons. In the network structure, the neurons are arranged in layers. Each of the neurons in the input layer receives one of the variables (e.g., wind speed and direction, humidity, temperature, and atmospheric pressure) on which the variables that we wish to forecast depend. The neurons of the output layer return the values of the variables that we wish to forecast (e.g., the power output of the wind farm at subsequent instants). There can also be a series of intermediate layers, called hidden layers. The manner in which the neurons interconnect is known as the connectivity pattern or architecture of the network.

The work flow of the proposed system is shown in the Figure 1.Block Diagram given below.

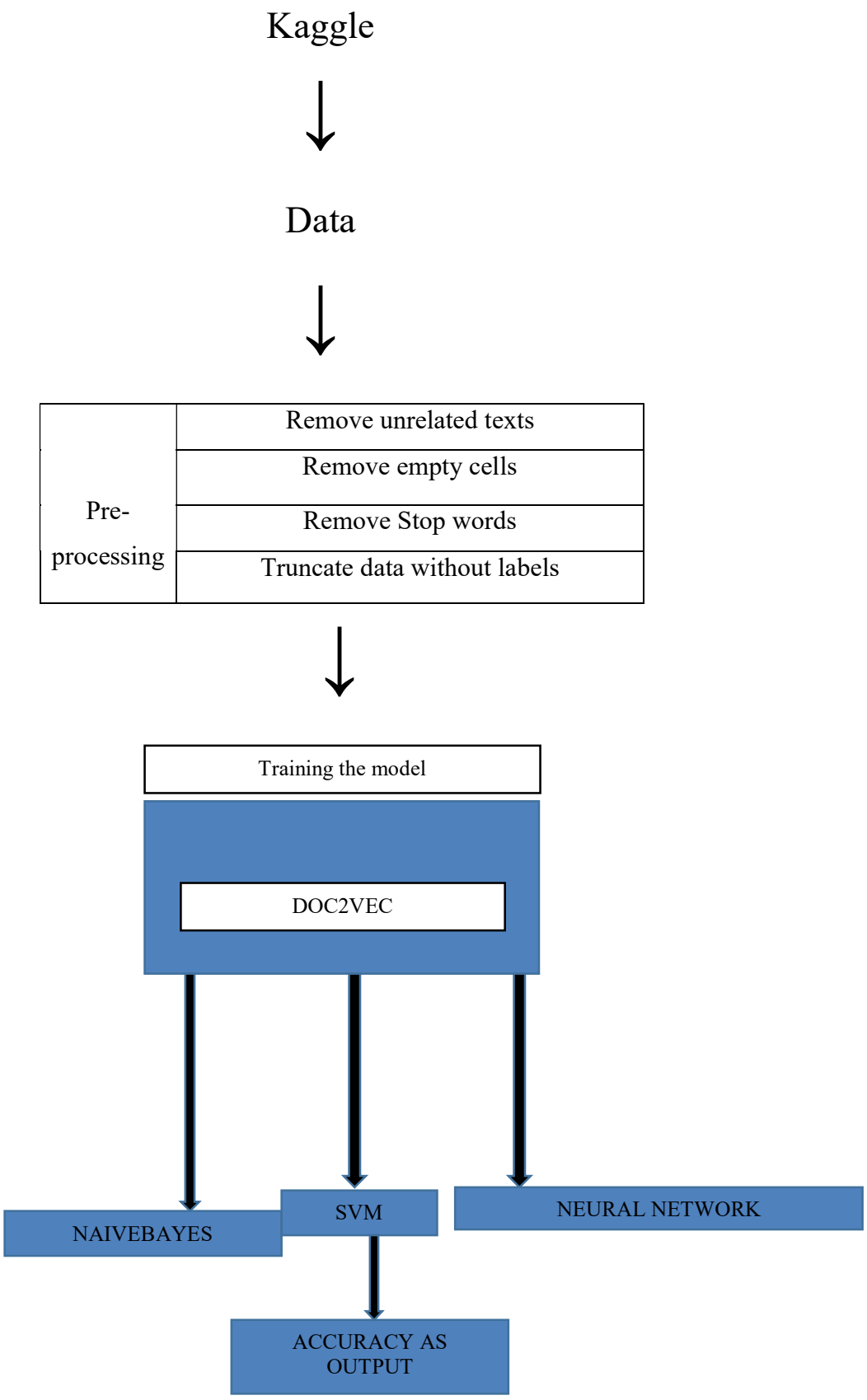


Figure 1 : Block Diagram

http://dx.doi.org/10.21172/ijiet.154.04

III. EXPERIMENT AND RESULT

The dataset is initially preprocessed and then it is trained so as to predict the accuracy. The accuracies of the three algorithms are diagrammatically represented as follows. Naive Bayes algorithm gives us an accuracy of 72% as shown in figure 3 whereas using Support Vector Machine we get an accuracy of around 88% as mentioned in figure 4. The highest accuracy is produced in the case of Artificial Neural Networks which is 92% as shown in figure 5.

Figure 2 : Front Page

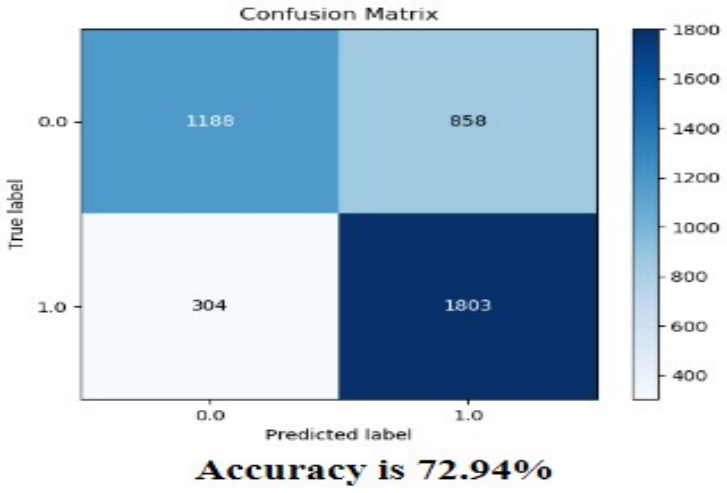


Figure 3 : Naïve Bayes Result

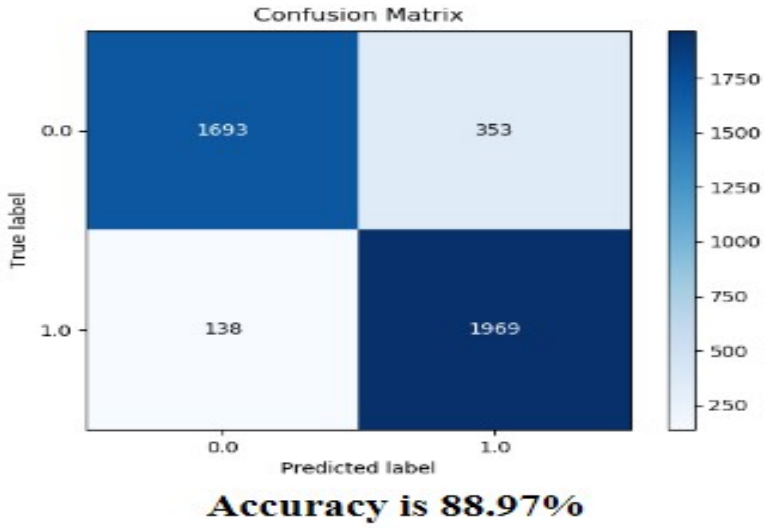


Figure 4 : SVM Result

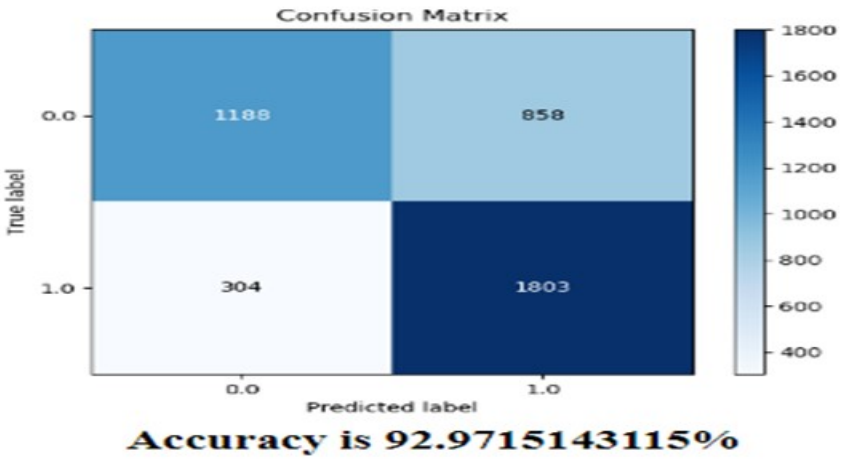
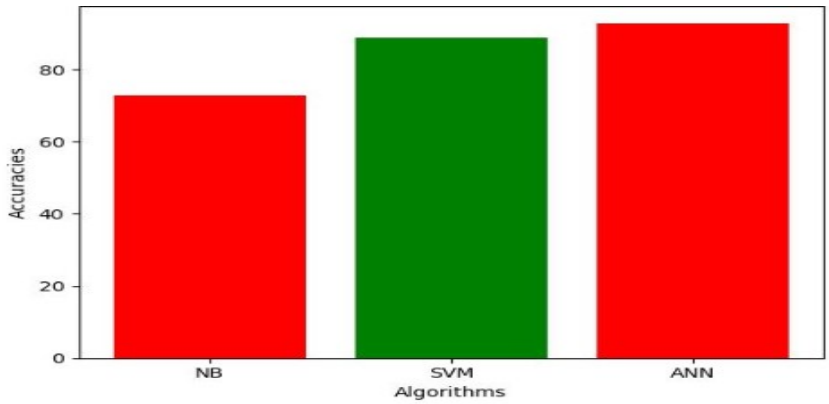


Figure 5 : ANN Result



Comparative Study On Fake News Accuracy Prediction

Figure 6 : Graph Result

Table -1 Experiment Result

Model	Accuracy
Naïve Bayes Algorithm	**72.94%**
Support Vector Machine	**88.97%**
Artificial Neural Networks	**92.9715%**

Table 1 show the accuracy of our proposed method of three models, where artificial neural networks has a better performance than others.

IV.CONCLUSION

Combination of machine learning and artificial neural networks are useful for fake news accuracy prediction as it looks like that fake news may be the most challenging area of research in the coming years. Here, we build a model for fake news accuracy prediction using machine learning algorithms. For text processing, we applied Naive Bayes Algorithm, SVM and ANN. Artificial Neural Networks gives the best accuracy out of all.

REFERENCES

[1] Conroy, Niall & Rubin, Victoria & Chen, Yimin. (2015). Automatic Deception Detection: Methods for Finding Fake News. . USA
[2] Ball, L. & Elworthy, J. J Market Anal (2014) 2: 187. https://doi.org/10.1057/jma.2014.15
[3] Lu TC. Yu T., Chen SH. (2018) Information Manipulation and Web Credibility. In: Bucciarelli E., Chen SH., Corchado J. (eds) Decision Economics: In the Tradition of Herbert A. Simon's Heritage. DCAI 2017. Advances in Intelligent Systems and Computing, vol 618. Springer, Cham
[4] Rubin, Victoria & Conroy, Niall & Chen, Yimin & Cornwell, Sarah. (2016). Fake News or Truth?Using Satirical Cues to Detect Potentially Misleading News. . 10.18653/v1/W16-0802.
[5] www.kaggle.com
[6] https://deeplearning4j.org/keras-supported-features
[7] https://en.wikipedia.org/wiki/Naive_Bayes_classifier
[8] https://ieeexplore.ieee.org/document/7489220
[9] https://www.kaggle.com/rksriram312/fake-news-nlp-stuff/notebook

[10] https://ieeexplore.ieee.org/document/8546944
[11] https://ieeexplore.ieee.org/document/8862770
[12] https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-4a444fca47