DETECTING FALSE INFORMATION ON COVID-19 USING ENSEMBLE METHODS

[1]K.Pragna, [2]M.Kamal Jostna, [3]A.Alekhya, [4]G.Venkata Kesava, [5]Dr.S.Jayaprada

[1]Student, [2]Student, [3]Student, [4]Student, [5]Associate Professor

[1]Computer Science and Engineering,
[1]Gudlavalleru Engineering College, Gudlavalleru, India

ABSTRACT*:* Now a days covid-19 cases has been increasing very fast and many types of variants are there due to this false information is becoming popular. So people don't know to believe which news. So we are developing a model which will help to predict the news on covid-19 is false or true. The main aim of the paper is to propose a misleading-information detection model that depends on the globe Health Organization, UNICEF, and conjointly the world organization as sources of information, additionally as medical specialty material collected from a spread of fact-checking websites. Getting information from reliable sources ought to assure their validity. The collected ground-truth data is employed to create a detection system that uses machine learning algorithms to spot misleading news. The analysis results indicate the standard and validity of the collected ground-truth information and their effectiveness in constructing models to notice deceptive news on Covid-19.

Keywords: Machine learning, Covid-19, detection model, fact-checking, Accuracy.

I. INTRODUCTION

At the tip of December 2019, the globe Health Organization (WHO) was enlightened of a gaggle of cases of respiratory disease of unknown cause that has been known within the town of the city Wuhan, in Hubei Province, China. Within the initial instance, these were the patients United Nations agency are diagnosed with acute respiratory disease. Most of them have worked for the exchange in the city, and that they had a great deal of the foremost common symptoms-fever, dry cough, tiredness, and, in additional severe cases, wheezing. However, these symptoms failed to have any symptoms of acute respiratory organ inflammation, as has been antecedent thought. At the side of the rise within the range of cases in China at the peak of these of the state and of unknown cause, at the start of Jan 2020. The name of the virus "Severe Acute metastasis Syndrome coronavirus a pair of (SARS-CoV-2" and also the illness itself, 'coronavirus illness (COVID-19)'. COVID-19 could be a world ill health which will need tutelage and strict adherence to non-public and general hygiene and cleanliness all told places. All of the reports revealed by the United Nations agency counsel that the medical specialty studies, since the start of Jan 2020.) As for transportation, it absolutely was straightforward for individuals to move from one place to a different one. This is often in no time, can contribute to the circulation of the blood of COVID-19, creating it a world pandemic. Social networks play a vital and effective role within the unfold of false info in regard to COVID-19, however, it's conjointly within the business of our daily lives, additionally as in an exceeding style of crises and conflicts everywhere the globe. With the emergence of the new virus, its potential isn't totally well-known, additionally because of the state of worry and panic for the overall public, for the distribution and dissemination of false info concerning the virus and its impact, it's still omnipresent.

Misleading info ought to be sent to the breach of a country's economy, reducing people's trust in their government, or promoting a particular product to attain vast profits. This went on with COVID-19. In general, deceptive info concerning the controls, vaccination, and mortality within the delirium of panic-buying of food, sanitizers, masks, and paper product. This has semiconductor diode to some shortages that have no continuous it's provide chain, and also the deepening of the gap between demand and provide, and food insecurity. Additionally, its semiconductor diode to a pointy decline within the world economy, inflicting severe losses to the price of oil and also the fall in world stock markets. Additionally, a number of the individuals haven't any confidence within the government, like in Italia and in Persia because of the unfold of COVID-19, and within the absence of medical protection for all around the world. All of this is often leading the globe in economic decline. It's terribly tough to substantiate the reality, accuracy, and correctness of the data provided, particularly if it's amid a significant illness, which can cause a threat to humanity. The United Nations agency has asked the foremost search engines like Google, Yahoo, and Bing, additionally as a range of social media platforms for the show of official reports that are revealed, and if the highest hits from all the COVID-19-related search term. From this, it follows that need special care and a focus within the choice of sources of knowledge. We've to trust in what's being promoted on social media, however in reliable and unbiased sources of knowledge, like the United Nations agency and also the world analysis organizations, agencies, and non-governmental organizations. Therefore, there's a model got to shield the general public with a tool so as to verify the validity of any of the data in regard to COVID-19. Currently, a day's deceptive info on covid-19 has been enlarged. To scale back that we tend to build a model to notice whether or not the given info is faux or real. During this project, we tend to use the flask framework that acts like a computer program through that we will directly notice the given news is real or fake.

II. LITERATURE SURVEY

In [1], Ming and Bhadrachalam, scientists have tried to boost the detection of dishonest messages, as well as the net, to information analysis. We've used deep learning models are supported the direct relationship between neural network (FNN), long and remembering (LSTM), which, together with a range of the vector representations of the words. The models are combined with a time period, data-mining, that is employed to gather info supplements the table of contents/the title of an item (in the admin console, author, etc.). The projected model offers sensible results, in terms of their correctness, accuracy, feedback, precision, and F1-measures, and also the use of LSTM centre, with a mixture of WORD2VEC and presentation.

In [2], F. A., Ozbay b, A. Alatas. We have put together a two-step method for the detection of false information in social networks. The first step is the pre-processing techniques that can be used to collect the data, and to transform the unstructured data sets of structured data, such as the Meeting Frequency Weighting (TF) method of the Document-Term Matrix. The second stage is the implementation of the twenty-three-monitored artificial intelligence, artificial intelligence algorithms, which are based on the information covered in the first phase. Evaluation of their work, the authors used four indicators (precision, accuracy, recall, and F1-measure). The experimental results showed that the use of the J48 algorithm will give the best accuracy of 65.5% for "Buzz Feed's political news in the dataset.

In [3], Agarwal et al. published a paper that describes a deep learning model that can predict the nature of the article, as it is used as an input. Scientists have been discussing and experimenting with the use of the word deposition (wear gloves) to pre-process text for use in the construction of a vector will be the space and set the language of connections. To get started with a pair of false Kaggle messages. A new model is based on the combination of neural networks and recurrent neural network architecture.

Hosni a et al [4] studied the problem of minimizing the adverse impact of the same name that appears in the process from the cutting of the messages that are specific to the distribution of a large amount of noxious information in the shortest amount of time possible. The author has developed a model for the distribution of several rumors, which is called the HISBM model, which captures the process of spreading more rumors on the Internet, on social networking sites.

The spread of fake news false information in the field of health care has had a negative impact on people's lives. Pulido et al. [5] proposed a new application for the analysis of social networks (Reedit, Facebook, and Twitter). The results show that the messages are addressed to false medical information, usually hostile messages on the basis of the evidence of the public's impact, they have to be respectful and transformative, and it is, in the end, in the context of the discussion has taken to social networks to overcome the false medical information. For the generation of the data set, the researchers chose the word "health "as a general matter, and the specific terms" vaccine, "food, "and "Ebola."

In [6], Zhou et al. a proposal for a new model of the theory of the detection of the scams in the context of supervised machine learning. The proposed model can recognize the false reports are in a very early stage. Do not rely on the content of the message, the model is able to detect when there are fake reports that will be published in a news-publishing, even though they do not spread it on social networks. As the experimental results, the authors have shown that the application of the model to two real data sets that their approach is better than that of the state, and the creation of false reports.

III. METHODOLOGY

To detect the false information of covid-19 we used different techniques like Artificial Neural Networks (ANN), Multinomial, Support Vector Machine (SVM), and Passive Aggressive Classifier. And also got different Accuracies for all the techniques.

3.1 ANN Algorithm:

Artificial neural network, that can be tested, and the system is made up of a large number of simple elements called neurons or perceptron's. Any of these can make a simple decision, and it sends the decisions of the other neurons to be organized in layers, with each other. In particular, a neural network that is able to emulate nearly every feature, and find the answer to almost every question is whether there are enough training samples and computing power. An artificial neural network is able to map serval, inputs, multiple outputs, where x and y are the vectors of input and output values, respectively. The function f consists of the relations for the neural network, which is operated by a group of parameters. In the context of the tools, these parameters are often referred to as the weight of the package. The load needs to be adjusted in such a way that, for a given mapping, and this process is easy to understand because of the training course. [7]
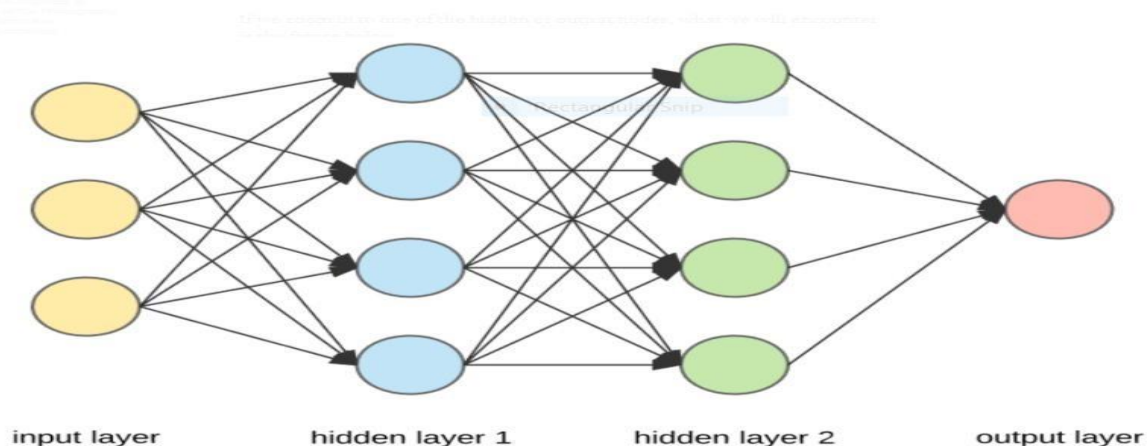

Fig 3.1.1: ANN Architecture

3.2 Multinomial Naive Bayes:

Naive Bayes is used in the processing of natural language. The naive Bayes classifier is a family of algorithms, which make use of Bayesian theorems. It is naive to assume that all features are independent of each other. Bayes ' theorem to compute the probability $P(c \mid x)$, where c is the class of all the possible outcomes, and x is the instance to be classified. $P(c \mid x) = P(x \mid c) P(c)$ $P(x)$ is, According to our data, the level of which is equal to either 0 or 1, where 0 means false reports " and " 1 " means that it's the real news. The news of x, we have that: P (it's the real news is $\mid$x) and P (false news $\mid$times). If P is true, new $\mid$x) > P (false news for a $\mid$x), then the system assumes that this is real news. Otherwise, the message will be predicted to be false. [8]

3.3 Support Vector Machine:

Support Vector Machine or SVM could also be a linear model for classification and regression problems. SVM model takes the data within the training set and maps it to data points in space so that there is a transparent gap between points belonging to different categories. This gap is made as wide as possible to reinforce the performance of the model. Whenever fresh information is given to the model, it maps the aim to the identical space and predicts the category supported by the side of the gap on which they fall.
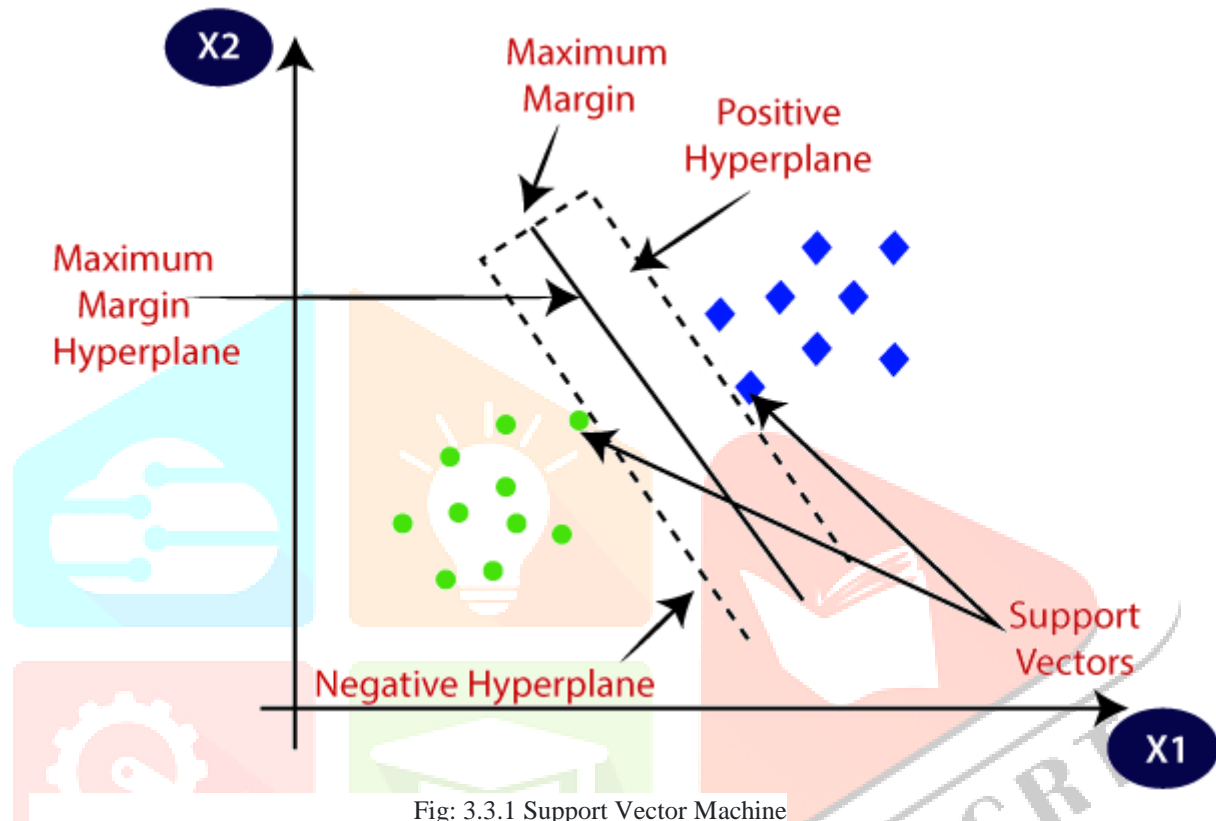

Fig: 3.3.1 Support Vector Machine

3.4 Passive Aggressive Classifier:

Passive aggressive classifier may be a web algorithm that learns from massive streams of data. The concept is to urge an example, update the classifier, and throw away the instance. It fast and easy to implement but doesn't provide global guarantees like SVM. Now we'll apply these models to our data. But, text can't be given directly as input to the classifier. Instead, we are getting to convert the text to numbers. Machine learning uses an easy model called bag-of-words to deal with textual data. The thought is to look out all the unique words within the document and build a vector of a size capable of the number of unique words. Each word is assigned an index within the vector. The index like the word is full of the frequency of that word within the document. The foremost drawback with this approach is that it ignores all the knowledge related to the order of the words, and only takes into account the frequency of the words. [10]

IV. IMPLEMENTATION

To implement the model firstly we should collect the dataset and do data pre-processing that means in which we should remove missing values otherwise fill them with mean or median. Later we should divide the dataset into train and test set. And apply algorithms on it and predict the output.

4.1 Getting the Data

In the dataset we have 10200 records and it contains 2 rows one is headlines and the other one is outcomes which contains 2 values 0 means fake news and 1 means real news. We have collected dataset from GitHub which contains news from different websites. [9]

We need a dataset which is used to build our machine learning model .However, in this paper we are using dataset which is stored in .csv format. The dataset consists of headlines and outcomes. The class variable is outcome whether the given news is fake or real.

4.2. Data Implementation

When a data is incomplete or inconsistent or noisy then data preprocessing is needed. There are so many routes to deal with unprocessed data:

1. Data Gathering: Firstly, We have collected the data from the GitHub which has news from different websites.
2. Data Pre-processing: later in Data Preprocessing technique We Will Identify and Handle Missing Values and Null Values.
3. Implementing algorithms: We will implemented algorithms like. Artificial Neural Networks, Support Vector Machine etc.
4. Training and testing the model: For training a model we initially split the model into three sections which are training data, validation data and testing data.
5. Prediction: Finally, we will predict whether the news is real or fake.

V. RESULTS AND DISCUSSIONS

Dataset consists of two sets trainset and test set. In this model we took 80% of data for training and 20% of data for testing.
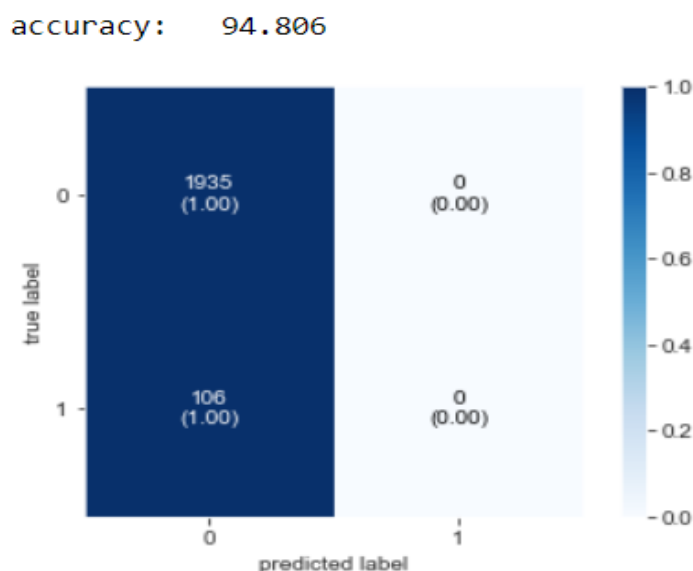

Fig 5.1: Accuracy and confusion matrix of Multinomial NB

In the above figure 4.1 it represents the accuracy and confusion matrix of multinomial navie bayes which got 94.806% accuracy and in the x-axis it has predicted label and in y-axis it has true label.
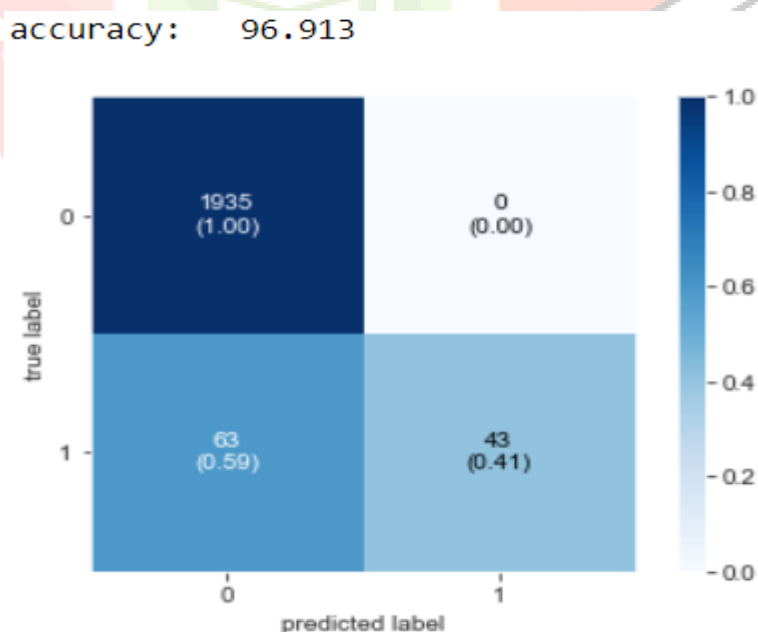

Fig 5.2: Accuracy and confusion matrix of SVM

In the above Figure 4.2 it represents the accuracy and confusion matrix of support vector machine (SVM) which got 96.913% accuracy and in the x-axis it has predicted label and in y-axis it has true label.
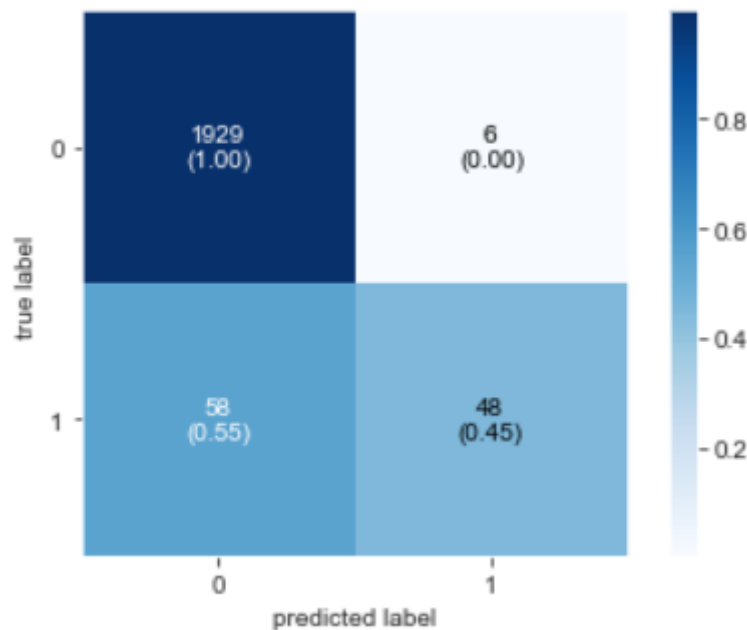
accuracy: 96.864

Fig 5.3: Accuracy and confusion matrix of PAC

In the above figure 4.3 it represents the accuracy and confusion matrix of Passive Aggressive Classifier (PAC) which got 96.864% accuracy and in the x-axis it has predicted label and in y-axis it has true label.
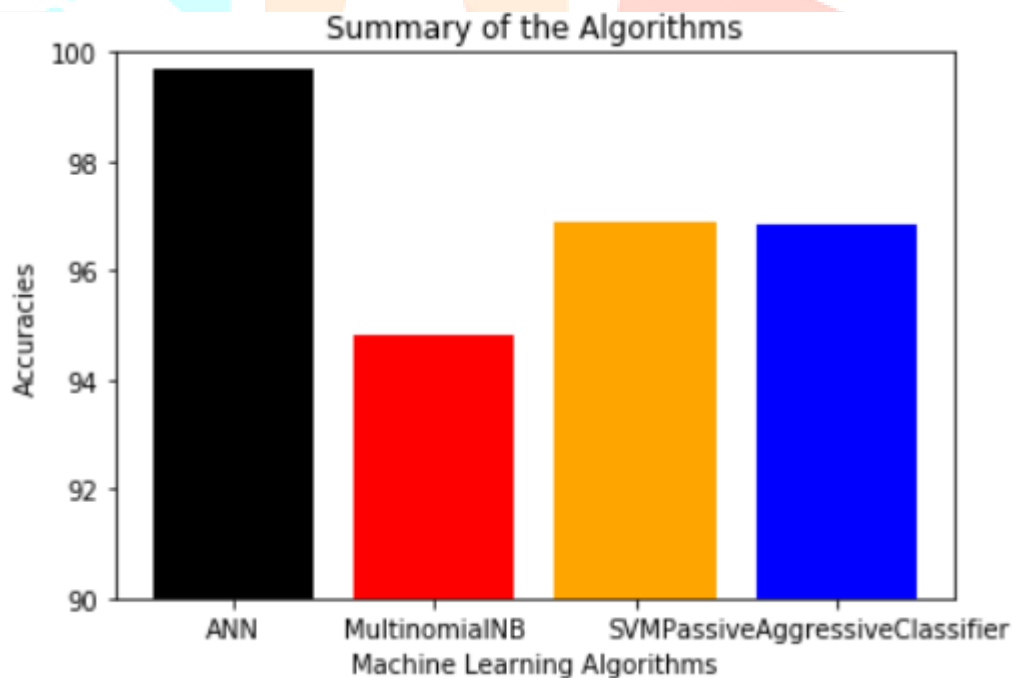

Fig 5.4: Accuracy of algorithms

In the Figure 4.4 it represents the graph of Accuracies of all the algorithms in x-axis we have different algorithms and in y-axis It has different Accuracy values.

DISCUSSIONS

Here we used different techniques to train and test the dataset. In all of these which algorithm gets highest accuracy we will load that model into flask which is a python web frame work which acts as a user interface in which we will predict our output. In Ann we got 99% Accuracy, 94.8% in multinomial, 96.9% in Support Vector Machine and finally 96.8% in Passive Aggressive Classifier.

VI.CONCLUSION

In this system, we have developed a method by which we can predict whether the news on covid-19 is real or fake. Due to the use of our system, we can easily predict the news is real or fake. The reason for developing such system is because now-a-days many news are coming about covid-19 we don't know whether they are true or false so by using this we can detect easily. In different algorithms we got different Accuracies and by using flask framework we will predict the output in a webpage.

REFERENCES

[1] Deepak D., Chitturi Bhadrachalam Deep neural approach to fake-news identification International conference on computational intelligence and data science (ICCIDS 2019), Procedia computer science, Vol. 167 (2020).

[2] Ozbay F.A., Alatas B.Fake news detection within online social media using supervised artificial intelligence algorithms Physical A, 540 (2020).

[3] Agarwal A., Mittal M., Pathak A., et al. "Fake news detection using a blend of neural networks": An application of deep learning SN Computer Science, 1 (2020).

[4] Hosni Adil Imad Eddine, Li Kan Minimizing the influence of rumors during breaking news events in online social networks Knowledge-Based System, 0950-7051, 193 (2020).

[5] Pulido CM. Ruiz-Eugenio L., Redondo-Sama G., Villarejo-Carballido B. A new application of social impact in social media for overcoming fake news in health J Environ Res Public Health, 17 (7) (2020).

[6] Zhou X., Jain A., Phoha V.V., Zafarani R. Fake news early detection: A theory-driven model Digit Threats: 1 (2) (2020).

[7] https://www.geeksforgeeks.org/introduction-to-ann-set-4-network-architectures.

[8] https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB

[9] https://datasetsearch.research.google.com/search?query=Fake%20Content%20Detection&docid=L2cvMTFrZHF6dDF2ZA%3D%3D

[10] https://www.geeksforgeeks.org/passive-aggressive-classifiers.