



Fake News Detection using Machine Learning Algorithm

Jithin Joseph, Sagil P, Amina P.M, Shirin Shahana, Ms. Rani Saritha

Saintgits Institute of Computer Applications, India, jithinjoseph4793@gmail.com , sagilprasannan@gmail.com, aminamuhsina25@gmail.com, shirinshahana3410@gmail.com, rani.saritha@saintgits.org

ABSTRACT

In our modern society where internet is ubiquitous, everyone relies on various online resources for news. Along with the rise in use of social media platforms like Facebook, Twitter etc. news spread rapidly among various users within a really short span of your time. The spread of fake news has far reaching consequences like creation of biased opinions to swaying election outcomes for the benefit of certain candidates. Moreover, spammers use appealing news headlines to get revenue using advertisements via click-baits. During this project, we aim to perform a binary classification of varied news articles available online with the assistance of concepts per Decision tree algorithm and Naive Bayes Classification. Fake data detection is that the most vital problem to be addressed within recent years, there's lot of research happening during this field due to its serious impacts on the readers, researchers, government and personal agencies working together to resolve the problem. This project represents a hybrid approach for fake data detection using the multinomial voting algorithm. The list of algorithms that are used here is Decision Tree and Naïve Bayes algorithms. All these two algorithms use training data because the bag of words model which was created using Count vectorizer. Experimental data has collected from the kaggle data world. Python is used as a language to verify and validate the result.

Key words : Fake news detection, decision tree algorithm, naïve bayes classifier.

1. INTRODUCTION

The detection of fake news on social media poses several new and challenging research problems. Though fake news itself isn't a replacement problem nations or groups are using the print media to execute propaganda or influence operations for centuries—the rise of web-generated news on social media makes pretend news a more powerful force that challenges traditional journalistic norms. There are several characteristics of same problem that make it uniquely challenging for automated detection. The fake news is intentionally written to mislead readers, which makes it nontrivial to detect simply based on news content. The content of pretend news is very diverse in terms of topics, styles and media platforms, and fake news attempts to distort truth with diverse linguistic styles while simultaneously mocking true news. For instance, fake

news may cite true evidence within the correct context to support a non-factual claim. Thus existing hand crafted and data specific textual features are not generally sufficient for fake news detection. Other auxiliary information must even be applied to improve detection, like knowledge domain and user social engagements. Second, exploiting this auxiliary information actually ends up in another critical challenge: the standard of the information itself. Fake news is sometimes associated with newly emerging, time critical events, which cannot be properly verified by existing knowledge bases thanks to the shortage of evidence or claims. Additionally the user's social engagements with fake news produces data that's big, incomplete, unstructured and noisy. Effective methods to differentiate credible users, extract useful post features and exploit network interactions are an open area of research and wish further investigations.

2. LITERATURE REVIEW

Fake Data Analysis and Detection Using Ensembled Hybrid Algorithm

In this paper represents a hybrid approach for fake data detection using the multinomial voting algorithm. This algorithm was tried with numerous phony news dataset which brought about a precision score of 94 percent which is a benchmark in the machine learning field where the other algorithms are at a range of 82 to 88 percent. The rundown of algorithms that have been utilized here is as per the following Naive Bayes, Random Forest, Decision Tree, Support Vector Machine, K Nearest Neighbors. Each one of these algorithms use preparing information as the bag of words model which was made utilizing count vectorizer. Test information has gathered from the kaggle data world. Python is used as a language to verify and validate the results. Tableau is used as a visualization tool. Implementation is carried out using default algorithm values.

Data or information is the most valuable asset in this century. The most important problem to be solved is to evaluate whether the data is relevant or irrelevant. Fake data has a huge impact on lot of people and organizations that may even lead to the end of the organization or panic the people. Machine learning researchers believe that this problem can be solved using the machine learning algorithms and there is lot of on-going research in this field which lead to the new branch called Natural Language Processing. This classification is not that

simple there are lot of challenges to go through in order to succeed. Let’s start with few of them machine learning works with the data if you are having huge and clean data then there was a great chance of creating great classifier. In order to create a real time application, the algorithm should be fed with the most recent data. Data is of different sizes so that should be properly cleaned to get better results.

The list of algorithms that has been used here are as follows:

- a) Naive Bayes, Random Forest
- b) Decision Tree
- c) Support Vector Machine
- d) K Nearest Neighbors

Fake News Detection Using Machine Learning Ensemble Methods

In this paper they utilized some computational strategies, for example, characteristic language handling (NLP) can be utilized to identify oddities that different a content article that is beguiling in nature from articles that depend on realities. Different procedures include the investigation of spread of phony news in contrast with genuine news. All the more explicitly, the methodology breaks down how a phony news story engenders diversely on an organization comparative with a genuine article.

The response that an article gets can be separated at a hypothetical level to group the article as genuine or phony. Various examinations have principally engaged on discovery and characterization of phony news via online media platforms like Facebook and Twitter.

At conceptual level, fake news has been grouped into various kinds; the information is at that point extended to sum up machine learning (ML) models for multiple domains. The investigation by Ahmed et al. included extricating etymological features, for example, n-grams from literary articles and training multiple ML models including K-nearest neighbor (KNN), support vector machine (SVM), logistic regression (LR), linear support vector machine (LSVM), decision tree (DT), and stochastic gradient descent (SGD), achieving the highest accuracy (92%) with SVM and logistic regression. Concurring to the examination, as the quantity of n expanded in n-grams determined for a specific article, the by and large exactness diminished. The marvel has been noticed for learning models that are utilized for classification.

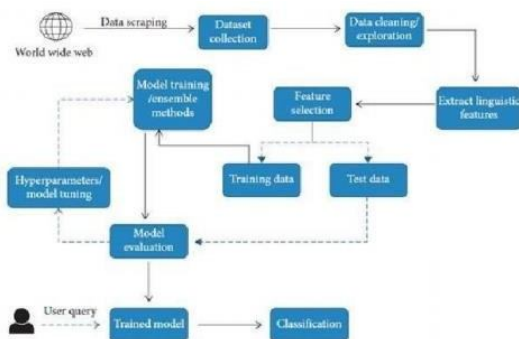


Fig.1 Workflow for training algorithms and classification of news articles

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

Logistic Regression

As we are arranging text based on a wide feature set, with a binary output (true/false or true article/fake article), a logistic regression (LR) model is utilized, since it gives the natural condition to characterize issues into binary or multiple classes. We performed hyper parameters tuning to get the best outcome for all individual datasets, while different parameters are tried prior to securing the greatest exactnesses from LR model. Numerically, the logistic regression hypothesis function can be characterized as follows:

$$h_{\theta}(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

Support Vector Machine

Support vector machine (SVM) is another model for binary classification problem and is accessible in different kernels functions. The target of a SVM model is to appraise a hyperplane (or decision boundary) based on feature set to group data points. The dimension of hyperplane differs as indicated by the number of features. As there could be numerous opportunities for a hyperplane to exist in a N dimensional space, the assignment is to recognize the plane that isolates the data points of two classes with maximum margin. A numerical portrayal of the cost function for the SVM model is characterized as surrendered and appeared in

$$J(\theta) = \frac{1}{2} \sum_{j=1}^n \theta_j^2,$$

such that

$$\theta^T x^{(i)} \geq 1, \quad y^{(i)} = 1,$$

$$\theta^T x^{(i)} \leq -1, \quad y^{(i)} = 0.$$

Multilayer Perceptron

A multilayer perceptron (MLP) is an artificial neural network, with an input layer, at least one hidden layers, and an output layer. MLP can be pretty much as basic as having every one of the three layers; nonetheless, in our analyses we have calibrated the model with different parameters and number of layers to produce an ideal foreseeing model. A fundamental multi-layered perceptron model with one hidden layer can be addressed as a function as demonstrated beneath:

$$f(x) = g(b^{(2)} + W^{(2)}(s(b^{(1)} + W^{(1)}x)))$$

K-Nearest Neighbors (KNN)

KNN is an unsupervised machine learning model where a reliant variable isn't needed to anticipate the outcome on a specific data. We give sufficient training data to the model and let it choose to which specific neighborhood a data point

belongs. KNN model estimates the distance of a new data point to its closest neighbors, and the value of K gauges the majority of its neighbors' votes; if the value of K is 1, then the new data point is doled out to a class which has the nearest distance.

Fake News Detection

$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^k (x_i - y_i)^2},$$

$$\text{Manhattan distance} = \sum_{i=1}^k |x_i - y_i|,$$

$$\text{Minkowski distance} = \left(\sum_{i=1}^k |x_i - y_i|^q \right)^{1/q}$$

In this project, we aim to perform a binary classification of different news articles accessible on the web with the assistance of ideas relating to Artificial Intelligence, Natural Language Processing and Machine Learning. With the growing popularity of mobile technology and social media, information is accessible at one's fingertips. Mobile applications and social media platforms have over-thrown customary print media in the spread of news and information. It is just normal that with the convenience and speed that digital media offers, people express inclination towards utilizing it for their day by day data needs. Not only has it empowered consumers with faster access to di-verse data, it has also provided profit seeking parties with a strong platform to capture a wider audience. With the outburst of information, it is apparently dreary for a layman to distinguish whether the news he consumes is genuine or fake. Fake news is normally distributed with an aim to misdirect or create bias to procure political or monetary benefits. Hence it may tend to have attracting headlines or intriguing content to increase viewership.

The most notable algorithms used by fake news detection systems include machine learning algorithms for example, Support Vector Machines, Random Forests, Decision trees, Stochastic Gradient Descent, Logistic Regression and so on. In this project we have endeavored to carry out two out of these algorithms to train and test our outcomes. We have utilized a blend of both off the shelf datasets as well as expanded it by crawling content on the web. The primary challenge throughout the project has been to fabricate a bunch of uniform clean data and to tune parameters of our algorithms to achieve the greatest precision. They saw that the Random Forests algorithm with a fundamental term frequency-inverse document frequency vector played out the best out of the four algorithms which they attempted.

Prediction Algorithms

They executed two different algorithms from scratch for the prediction model which were: Logistic Regression mod-el and the Naïve Bayes classifier model. The algorithms and the details of implementation have been clarified in the areas

underneath. Along with these we likewise trained and tested our dataset on two other models: Random Forests model and Support Vector Machine model. Given the brief time frame of the project, the last two algorithms were judiciously carried out with the assistance of scikit-learn libraries.

a) Logistic Regression

Logistic Regression is a Machine Learning technique used to estimate relationships among variables using statistical methods. This algorithm is great for binary classification problems as it deals with predicting probabilities of classes, and hence our decision to choose this algorithm as our base-line run. It relies on fitting the probability of true scenarios to the proportion of actual true scenarios observed. Also, this algorithm does not require large sample sizes to start giving fairly good results.

b) Naive Bayes Classifier

This is a straightforward yet incredible classification model that functions admirably. It employments probabilities of the components belonging to each class to form a prediction. The fundamental assumption in the Naive Bayes model is that the probabilities of a attribute having a place with a class is free of other attributes of that class. Henceforth the name 'Naive'.

3. PROPOSED SYSTEM

The aim of this project is to accurately determine the authenticity of the contents of a particular news article. For this purpose, we have devised a procedure which is intended to fetch favorable results. We first take the URL of the article that the user wants to authenticate, after which the text is extracted from the URL. The extracted text is then passed on to the data pre- processing unit. The data preprocessing unit consists of various processes like the Tokenization and Generation of the word cloud. The outputs from these processes play an important role in further analyzing the data. The core deciding factors that we use to determine the output of our project i.e. if a particular news article is fake or not are the stance of the article and comparison of the article with top google search results. The first method is by using stance detection to in order to analyze the stance of the author. Stance is a mental or an emotional position adopted by the author with respect to something. Stance detection is an important part if NLP and has wide applications. The stance of the author can be divided into various categories like Agree, Disagree, Neutral or Unrelated with respect to the title. Giving each of these categories weights can help us in the final conclusion of whether a news article is fake or not. The second method is to use document similarity or tf-idf to know how similar a document is to top search results. This too can give us an insight into the authenticity of a news article. Next, we need to classify the output into various output classes for which we can use classification algorithms or regression models. The output classes can be true, mostly true, false, and mostly false or we can just present it with a number.

4. METHODOLOGY

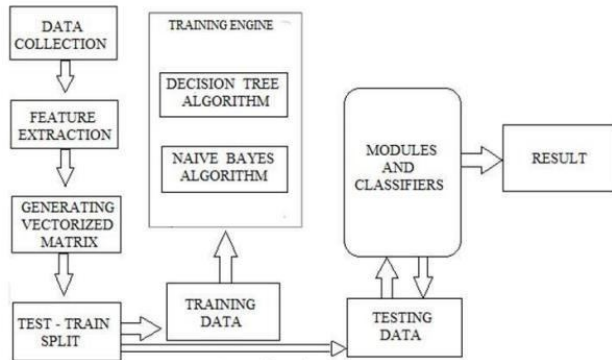


Fig.2 Workflow of Fake News Detection

Step 1: Feature Extraction

News content features portray the meta information identified with a piece of news. A list of representative news content attributes are recorded beneath:

- a) Source: Author or publisher of the news article.
- b) Headline: Short title text that plans to grab the eye of per users and portrays the primary theme of the article.
- c) Body Text: Main text that elaborates the details of the news article; there is typically a significant case that is specifically highlighted and that shapes the point of the publisher.
- d) Image/Video: Part of the body content of a news article that provides visual cues to outline the story. In light of these raw content attributes, various types of feature representations can be worked to extract discriminative characteristics of fake news. Normally, the news content we are taking a gander voluntarily generally be phonetic - based and visual-based.

Term Frequency - Inverse Document Frequency

The tf-idf is a statistical measure that reflects the significance of a specific word regarding a document in a corpus. It is regularly utilized in information retrieval and text mining as one of the components for scoring documents and performing searches. It is a weighted measure of how often a word occurs in a document comparative with how frequently it happens across all documents in the corpus. Term frequency is the number of times a term occurs in a document. Inverse document frequency is the inverse function of the number of documents in which it occurs.

$$tf-idf(t, d) = tf(t, d) * \log(N/(df + 1))$$

Hence a term like “the” that is common across a assortment will have lesser tf-idf values, as its weight is diminished by the idf component. Thus the weight computed by tf-idf addresses the significance of a term inside a document. The tokenized

data was utilized to generate a sparse matrix of tf-idf features for portrayal. This represented our feature vector and was utilized in subsequent prediction algorithms.

Step 2: Model Construction

Since fake news endeavors to spread bogus claims in news content, the most straightforward means of detecting it is to check the honesty of major claims in a news article to decide the news veracity. Knowledge-based approaches intend to utilize external sources to fact-check proposed claims in news content. The goal of fact-checking is to assign a truth value to a claim in a specific context. Fact-checking has attracted increasing attention, and numerous efforts have been made to build up a feasible automated fact-checking system. Existing fact-checking approaches can be categorized as expert-oriented, crowdsourcing-oriented, and computational-oriented.

5. DATASET

Information has been gathered kaggle information world. It has four segments and they are file, title, text and mark of the different news stories of different diaries. Among the four content is utilized as an autonomous variable and mark is utilized as a reliant variable. There is no adjustment of the precision scores even with the utilization of titles in light of the fact that the title words will get reshaped in the content section. The train and test information split were in the proportion of 80 – 20 utilizing irregular capacity where the train set is utilized for the preparation reason and test set is utilized for the testing reason. The dataset for this venture was worked with a blend of both genuine and phony news. The majority of the information was physically lithered and removed, though some were utilized off the rack. The whole dataset added up to 4050 news stories. To proficiently gather such tremendous information, we made a multi-strung web crawler. We ran the crawler utilizing around 100 strings all at once and download the crude HTML body substance of the slithered pages. The wellsprings of genuine news incorporate Yahoo News, AOL, Reuters, Bloomberg and The Guardian among many. Hotspots for counterfeit news incorporate the Onion. UsaNewsFlash, Truth-Out, the Controversial Files, etc. To remove significant substance from the slithered pages we utilized two procedures. First was to lessen clamor by eliminating.

The gathered information was handled utilizing different content pre-processing measures, as clarified later and put away in CSV documents. The genuine and phony information were then blended and rearranged to get a CSV record containing a united randomized dataset.

Text Pre-processing

Since a large portion of the information was crept and separated physically, we needed to initially go through the information to get association and arranging of text. The information was made uniform and equivalent by changing over it into a uniform UTF-8 encoding. There were a few situations where we experienced unusual images and letters contradictory with the character set which must be taken out. We saw that the information from news stories were frequently coordinated into sections. In this way, we performed managing to dispose of additional areas and void lines in text.

Stage 1: Feature Extraction

News content highlights depict the meta data identified with a piece of information. A rundown of delegate news content credits are recorded beneath:

- a) Source: Author or distributor of the news story.
- b) Headline: Short title text that means to grab the eye of per users and depicts the fundamental subject of the article.
- c) Body Text: Main content that explains the subtleties of the report; there is generally a significant case that is explicitly featured and that shapes the point of the distributor.
- d) Image/Video: Part of the body substance of a news story that gives viewable prompts to outline the story. In view of these crude substance ascribes, various types of highlight portrayals can be worked to extricate discriminative qualities of phony news. Ordinarily, the news content we are seeing will generally be phonetic based and visual-based.

Term Frequency - Inverse Document Frequency

The tf-idf is a factual measure that mirrors the significance of a specific word regarding an archive in a corpus. It is frequently utilized in data recovery and text mining as one of the parts for scoring records and performing look. It is a weighted proportion of how frequently a word happens in a report comparative with how regularly it happens across all archives in the corpus.

Term recurrence is the occasions a term happens in a record. Reverse archive recurrence is the converse capacity of the quantity of reports wherein it happens.

$$\text{tf-idf}(t, d) = \text{tf}(t, d) * \log(N/(\text{df} + 1))$$

Subsequently a term like "the" that is normal across an assortment will have lesser tf-idf esteems, as its weight is reduced by the idf part. Consequently the weight registered by tf-idf addresses the significance of a term inside a record.

The tokenized information was utilized to produce a scanty lattice of tf-idf highlights for portrayal. This addressed our element vector and was utilized in ensuing forecast calculations.

Step 2: Model Construction

Since counterfeit word endeavors to get out bogus cases in news content, the most clear methods for identifying it is to check the honesty of significant cases in a news story to choose the news veracity. Information based methodologies plan to utilize outside sources to certainty check proposed claims in news content. The objective of actuality checking is to allot a reality worth to a case in a specific setting. Actuality checking has pulled in expanding consideration, and numerous endeavors have been made to build up an achievable computerized certainty checking framework. Existing certainty checking approaches can be arranged as master situated, publicly supporting focused, and computational-situated.

Exploratory Design

Datasets : Online news can be gathered from various sources, for example, news organization landing pages, web search tools, and web-based media sites. Nonetheless, physically deciding the veracity of information is a difficult undertaking, generally requiring annotators with area aptitude who performs cautious examination of cases and extra proof, setting, and reports from definitive sources. For the most part, news information with explanations can be assembled in the accompanying manners: Expert writers, Fact-checking sites, Industry finders, and Crowd-sourced labourers.

Assessment Metrics : Assess the presentation of calculations for counterfeit news location issue, different assessment measurements have been utilized. In this subsection, we survey the most broadly utilized measurements for counterfeit news discovery. Most existing methodologies consider the phony news issue as an order issue that predicts if a news story is phony.

6. ALGORITHMS USED

We carried out two unique calculations without any preparation for the expectation model which were Decision Tree Algorithm and the Naïve Bayes classifier model. The calculations and the subtleties of execution have been clarified in the areas beneath. Notwithstanding these we additionally prepared and tried our dataset on two different models: Random Forests model and Support Vector Machine model. Given the brief timeframe casing of the undertaking, the last two calculations were wisely executed with the assistance of scikit-learn libraries.

a) Decision Tree Algorithm

A choice tree model is a flowchart looks like construction in which each inward hub addresses a test on a trait where each branch addresses the result of the test, and each leaf hub addresses a class mark. The ways from root hub to leaf hub will make the grouping rules. In dynamic, a choice tree and a firmly related stream outline is utilized as the visual and

scientific choice help instrument, in which the normal benefits of contending choices are determined by utilizing the stream. Choice Tree, deals with the pack of words highlights where the information of various articles gathered is changed over into encoded design by utilizing different vectorization strategies dependent on the necessity some of them are tally vectorizer (CV), term recurrence and backwards report recurrence vectorizer (TFIDF).

b) Naive bayes

Credulous Bayes classifier is a basic likelihood put together classifier based with respect to the Bayes hypothesis with incredible (innocent) autonomy presumption between the information highlights, where class marks picked from some limited set. It's anything but a one single calculation to prepare such classifiers, however an assortment of numerous calculations dependent on one normal standard: each innocent Bayes classifier accepts that the worth of a specific element is autonomous with the worth of some other element, given the class variable.

Guileless Bayes is the most selected measurable method for the models like email separating, spam sifting, etc. Credulous Bayes chips away at the sack of words highlights where the information of various articles gathered is changed over into encoded design by utilizing different vectorization methods dependent on the necessity some of them are check vectorizer (CV), term recurrence and converse record recurrence vectorizer (TFIDF).

The Bag of words will be passed to the Naïve Bayes model as a preparation information and dependent on the information the model will learn.

At that point when any article is passed to order vectorizer will make scanty grid and afterward model will foresee dependent on the word conveyance in the inadequate lattice.

$$Pr(F|W) = \frac{Pr(W|F) \cdot Pr(F)}{Pr(W|F) \cdot Pr(F) + Pr(W|T) \cdot Pr(T)}$$

where:

Pr(F|W) – restrictive likelihood, counterfeit information when the word present in the article; Pr(W|F) – contingent likelihood of discovering the word W in counterfeit information articles; Pr(F) – by and large likelihood that the given information is phony information;

Pr(W|T) – contingent likelihood of discovering the word W in genuine information articles;

Pr(T) – by and large likelihood that given information is genuine information. This equation depends on the Bayes' hypothesis.

Assessment Metrics

To assess the exhibition of calculations for counterfeit news discovery issue, different assessment measurements have been utilized. In this subsection, we survey the most broadly utilized measurements for counterfeit news recognition. Most

existing methodologies consider the phony news issue as an order issue that predicts if a news story is phony.

We utilized the accompanying three measurements for the assessment of our outcomes. The utilization of more than one grid assisted us with assessing the exhibition of the models from alternate points of view.

Characterization Accuracy

This portrays the quantity of exact forecasts made out of the absolute number of expectations made. Characterization precision is determined by isolating the absolute number of right outcome by the complete number of test information records and increasing by 100 to get the rate.

Disarray Matrix

This is an incredible visual approach to portray the expectations as four classes:

- a) True Positive (TP): when anticipated phony news pieces are really commented on as phony news.
- b) True Negative (TN): when anticipated genuine news pieces are really commented on as evident news.
- c) False Negative (FN): when anticipated genuine news pieces are really clarified as phony news.
- d) False Positive (FP): when anticipated phony news pieces are really clarified as obvious news.

By detailing this as a grouping issue, we can characterize:

Accuracy and Recall

Accuracy which is otherwise called the positive prescient worth is the proportion of significant cases to the recovered occurrences.

$$Accuracy = \frac{\text{No. of True Positives}}{\text{No. of True Positives} + \text{No. of False Positives}}$$

Review which is otherwise called affectability is the extent of pertinent cases recovered among the complete number of important occasions.

$$Review = \frac{\text{No. of True Positives}}{\text{No. of True Positives} + \text{No. of False Negatives}}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

These metrics are ordinarily utilized in the AI people group and empower us to assess the exhibition of a classifier from alternate points of view. In particular, precision estimates the similitude between anticipated phony news and genuine phony news.

7. CONCLUSION

The issues of phony news and disinformation assume a significant part on these days life. This is on the grounds that the high level of innovation and specialized techniques we've empowered data spreading among individuals with none check. This is a motivation behind why analysts began looking for answers for prevent counterfeit news and disinformation from spreading without any problem. Nonetheless, it is notable that controlling the progression of data online is incomprehensible. The information we used in our work is gathered from the planet Wide Web and contains news stories from different areas to cover up a large portion of the news rather than explicitly grouping political news. The essential point of the examination is to spot designs in text that separate phony articles from genuine news. In this task, we played out an endeavour to check the news stories believability relying upon their attributes. At this point, we carried out a calculation joining a few order strategies with text models. It performed well, and in this manner the precision results were moderately fulfilling.

8. FUTURE SCOPE OF THE PROJECT

As future work, we intend to all the more likely investigation the blend between the element extraction techniques and the classifiers as we will actually want to pick the content portrayal model that performs best with the classifier. Additionally, to accomplish a higher exactness, we should carry out a more modern calculation which may utilize information mining advancements with large information, in light of the fact that making a major dataset including more kinds of news stories with more class factors (names) will help raising the precision score.

REFERENCES

- [1] I. Ahmad, Muhammad Yousaf, SuhailYousaf and M.O. Ahmad, "Fake News Detection Using Machine Learning Ensemble Methods", Complexity in Deep Neural Networks, Hindawi.
- [2] P. B. P. Reddy, M. P. K. Reddy, G. V. M. Reddy and K. M. Mehata, "Fake Data Analysis and Detection Using Ensembled Hybrid Algorithm," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 890-897
- [3] Fake News Detection Using Naive Bayes Classifier, MykhailoGranik, VolodymyrMesyura, Computer Science Department, Vinnitsia National Technical University Vinnitsia, Ukraine.
- [4] Fake News Detection, Akshay Jain, AmeyKasbe, Department of Electronics and Communication Engineering, Maulana Azad National Institute of Technology, Bhopal, India.

- [5] Understanding User Profiles on Social Media for Fake News Detection, Kai Shu, Suhang Wang, Huan Liu, Arizona State University, Tempe, AZ 85281
- [6] A Feature Based Approach for Sentiment Analysis using SVM and Coreference Resolution, Hari Krishna M, Rahamathulla K, Ali Akbar, Department of Computer Science and Engineering, Govt. Engineering College Thrissur, India.
- [7] A Naïve Bayes Approach for working on Gurmukhi Word Sense Disambiguation, Himdweep Walia¹, Ajay Rana², Vineet Kansal³, (1)GNIOT, Gr. Noida, Uttar Pradesh, India, (2)Amity University Uttar Pradesh, Noida, India. (3)CSED, IET, Lucknow, Uttar Pradesh, India.
- [8] A novel Naive Bayes model: Packaged Hidden Naive Bayes, YaguangJi, Songnian Yu, Yafeng Zhang, School of Computer Engineering & Science, Shanghai University Yanchang Rd. 149, 200072 Shanghai, China.
- [9] Accuracy of Classifier Combining Based on Majority Voting, Peng Hong, Lin Chengde, LuoLinkai, Zhou Qifeng, Department of Automation, Xiamen University Xiamen, P.R.China.
- [10] Analysis of ordering based ensemble pruning techniques for Voting based Extreme Learning Machine, Sukirty Jain, Sanyam Shukla, Bhagatsingh Raghuvanshi, Computer Science and Engineering, MANIT, Bhopal, India.
- [11] Detecting Depression Using K-Nearest Neighbour's (KNN) Classification Technique.
- [12] Evaluating Machine Learning Algorithms for Fake News Detection, Shlok Gilda, Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India
- [13] Fake News Detection Enhancement with Data Imputation, Chandra MouliMadhavKotteti, Na Li, LijunQian, CREDIT Center, Prairie View A&M University, Texas A&M University System, Prairie View, TX 77446, USA.
- [14] Identifying tweets with Fake News, Saranya Krishnan, Min Chen, Division of Computing and software systems, School of STEM, University of Washington Bothell, USA.