Detect and Remove Fake News through Online Social Networks by using Bayes classifier

Anjaneyulu Kunchala[1],Duggineni Srinivasarao[2],D V V Brahmachari[3]Rajesh Vemulakonda[4]

[1] Department of IT, Vigna's Foundation for Science, Technology & Research, Vadlamudi, AP, India.
[2]Department of CSE,Narasaraopet Engineering college,Narasaraopet, AP, India.
[3]Department of CS,Vignan Degree and PG college,Guntur, AP, India.
[4]Department of CSE,Prasad V Potluri Siddhartha Institute of Technology,Vijayawada, AP, India.
kunchalaanjaneyulu6@gmail.com ,dugginenisrinivas.cse@gmail.com, db.achari72@gmail.com

Abstract.

Social media and other platforms are rife with fake news, which is a matter for alarm due to the potential damage it may do to society and the country. Lots of research has previously been done on detection. Research on counterfeit news location and customary AI models are inspected to figure out which is the most incredible to foster an item model that can characterize counterfeit news as obvious or misleading, utilizing apparatuses like Python Scikit-Learn and Natural Language Processing (NLP) for printed examination, to foster an item model with administered AI calculation.Features will be extracted and vectorized as a consequence of this procedure. Python's scikit-learn package, which gives valuable instruments like Count Vectorizer and Tiff Vectorizer, is the most ideal choice for dissecting text information. To augment precision, we will use include choice techniques to test and pick the best-fitting elements from the disarray network results.

1. INTRODUCTION

The following is a brief summary of the main pointsFake News gives information that may or may not be accurate. This causes instability in certain nations, such as the Arab Spring, by perpetuating a false statistic or exaggerating the cost of particular government services. Concerns about author responsibility are being addressed by groups like the House of Commons and the Crosscheck initiative. However, the scope of their use is severely limited since they are based on human detection. A world in which millions of objects are erased or published every minute does not make this practicable or responsible. A solution might be the production of a reliable automated index score, or rating, for different news organisations' dependability and context. This review tells the best way to utilize directed AI calculations on a physically grouped and ensured dataset to develop a model that can assess regardless of whether an article is true in light of its words, expressions, sources, and titles.Features are then selected using feature selection approaches, based on results from the confusion matrix,

in order to maximise accuracy. The model should be built using a variety of classification strategies. Likewise, the thing model will be a model that can recognize and portray counterfeit articles that can be used and facilitated with any structure later on, in view of already obscure information.

2. Related Work

2.1 Fake News and Social Media

Many different types of sites and technologies exist in the world of social media that may be categorised as "social media." [1][2]. As a result of unintended variables like educational shock or inadvertent behaviour, some experts feel that false news arises. This is similar to the situation of the Nepal Earthquake. [3][4]. By the year 2020, false health information would be widely disseminated, endangering the health of everyone on the planet. After COVID-19, the World Health Organization issued a warning that there was a big "infodemic," or burst of both real and fake news, which included much deception, early in February 2020.

2.2 Natural Language Processing

For the most part, NLP is used for research into certain areas of systems and algorithms. A system's Natural Language Processing (NLP) score allows for the combination of voice interpretation and speech generation to be done by the algorithmic system. Additionally, it may be used to designate activities in a wide range of different languages. Language pipelines For example, an Emotion Analyzer and Detection, a Named Entity Recognition, and a Chunking and Semantic Role Labeling were coupled in a new ideal approach.

Sentiment analysis [7] is a method for figuring out how people feel about a certain subject matter. Tracking down a particular expression, separating the feeling, and consolidating it with association examination is the course of opinion investigation. Two languages are used in the sentiment analysis process. Sentiment models database and dictionary of terminology are analytical resources.For constructive and destructive terms, A scale of -5 to 5 is used to try categorization For dialects, for example, Sanskrit [8], Hindi [9], and Arabic, grammatical forms taggers instruments for dialects, for example, European dialects are being explored. It is possible that this will be effective. In order to better organise your vocabulary, categorise the words you use as nouns, adjectives, verbs, etc. European languages have several speech methods that work, however in Asian and Arabic languages, they don't work at all.The tree-bank approach is used in part of the Sanskrit word "speak." SVM [10] is a technique for automatically identifying symbols and components of speech, as well as automatically exposing fundamental phrases in Arabic text [11].

2.3 Data Mining

Both unsupervised and supervised methods of data mining may be used. It is possible to foresee hidden actions using supervised learning, which takes use of previously learned data. Endeavoring to find stowed away information models without the utilization of preparing information, like sets of info names and classes, is called unaided information mining It is a great illustration of solo information mining to utilize total mines and a partnered base.

2.4 Machine Learning (ML)

Characterization It is feasible to work on the precision of programming frameworks without having to reprogramme them utilizing Machine Learning (ML). Information researchers characterize the progressions or traits that the model necessities to assess and use to make expectations. While the preparation is finished, the framework makes new information from the recently scholarly levels [11]. Using six algorithms, fake news may be identified in this study.

2.5 Decision Tree

Using a flowchart-like structure, the decision tree is a powerful tool for solving categorization problems. An attribute is "tested" at each node of the decision tree, and the branching is dependent on the criteria and outcomes of those tests. Eventually, all ascribes are processed and the leaf hub has a class name. The distance between the root and the leaf addresses the order rule.It's wonderful that it can work with both a dependent and a category variable. Using them, you may quickly and clearly see the connections between the most important elements. New factors and elements that are valuable for information investigation and appropriately expect the objective variable are made by them.Predictive models that use supervised learning techniques often use tree-based learning algorithms to achieve high accuracy. They excel at tracing non-linear correlations. They are also known as CART [13][14][15] because they are very good at solving classification or regression issues.

2.6 Random Forest

Random Forests are based on the idea of constructing a large number of decision tree algorithms, each of which produces a unique outcome. The random forest takes up the outcomes that are anticipated by a huge number of decision trees. To guarantee that the decision trees are varied, the random forest picks a subcategory of attributes from each group at random. [16][17] When used to uncorrelated decision trees, Random Forest is most effective. A single decision tree may be used across comparable trees and get the same result.Bootstrapping and feature randomness may be used to create uncorrelated decision trees.

2.7 Support Vector Machine (SVM)

Based on this, the SVM approach uses a representation of each item as a point in a range of dimensions n, with the value of a given attribute equal to the number of supplied coordinates [13]. With n features, SVM technique depicts data item in n-dimensional space, where the coordinates of each feature correspond to the value of that feature.The data is classified using the hyper-plane that was used to distinguish the two groups.

2.8. Naive Bayes

This approach, which is based on Bayes theory and is employed in a variety of machine learning situations [18], operates on the assumption that it is free of predictors. Simply expressed, Naive Bayes posits that no two functions in a category are related. When a fruit is red in colour, has spirals, and has a diameter of less than 3 inches, it is categorised as an

apple. Whether these functions are dependent on one other or on distinct functions, and whether they are dependent on one another or on different capacities, Naive Bayes accepts that they all have their own evidence of the apples. [14]

Step:

1. Analyze the data in Training Dataset T
2. For each class, ascertain the mean and standard deviation of the indicator factors.
3. Repeat
4. In each class, deciding the probability of utilizing the Gauss thickness recipe;
5. Until the probabilities of all the indicator factors have been determined(f1, f2, f3,..., fn),.
6. I computed the probabilities for each category; and
7. Acquire the best chance possible;
8. (Researchgate.net, 2018)

Random Forest (RF) and Nave Bayes (NB) are fundamentally different because of the size of the model. The limited model size of NB models makes them only usable for a certain kind of data since they are ineffective at expressing complex behaviour. Overfitting might occur if the Random Forest model's model sixe is too large. Since of this, NB is an excellent choice for new data input because it can readily be moulded. However, RF may need a forest rebuild after every alteration.

2.9. KNN (k- Nearest Neighbors)

For new locations, KNN categorises them by analysing how often surrounding ks make a sound in connection to them. According to the distance function [15], the class position is exceptionally fundamentally unrelated among the closest neighbors K.

KNN is a kind of supervised learning algorithm with applications such as pattern recognition and intrusion detection. Since no assumptions are made about the distribution of the data, it is nonparametric. It is assumed that the data is distributed Gaussianly in GMM.

3. Methodology

It is explained in this part how to categorise the data. To detect fake articles, a technique has been created based on this methodology . A managed AI calculation is utilized to order the information. The initial phase in this arrangement challenge is to assemble the datasets, after which preprocessing, feature selection, dataset preparing and testing, lastly running the classifiers will happen. [35][36][37][38][39]. Figure 1 depicts the system method that has been recommended. Classifiers like Random forest, SVM, and Nave Bayes are used, as well as voting methods like majority voting and other classifiers, to conduct a large number of tests on a dataset. For maximum accuracy and precision, tests are led on every calculation exclusively and in mix. [40][41][42].

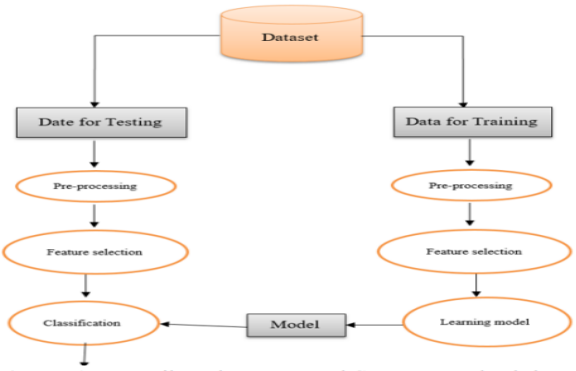


Figure 1. Describes the Proposed System Methodology

Arrangement techniques are used to foster a portrayal model that can be used as a scanner for fake news, and the model is embedded in a Python program that can be used to reveal counterfeit news data [43][44].An optimised Python code was created by doing the required refactorings [25][26].

Despite k-Nearest Neighbors (k-NN), XGBoost, Naive Bayes, and Decision Trees, this model moreover uses Random Forests and Support Vector Machines (SVM).Efforts are made by all of these algorithms to be as exact as feasible. Compare the average of all of them whenever feasible.

Figure 2 shows how the dataset is input into different algorithms to detect false news. In order to arrive at the final result, the accuracy of the obtained data is verified.

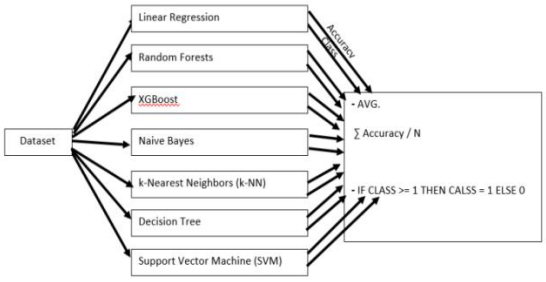


Figure 2. The Classification Algorithms

It's important to recognise political falsehoods throughout the model-building process. NLTK (Natural Language Toolkit) is used to coordinate POS and part assurance ensuing to getting a political news dataset (the Liar dataset is used for the model). Computer based intelligence moves close (Nave Bays and Random Forest) are then used to deliver the classifier model from the enlightening list. Preprocessing of the dataset is done in the structure after the NLTK is applied, as portrayed in Fig 2;an email is sent to the trained section for applying algorithms. The model with a response message is produced when N.B. and Random forest are applied to the system response. Precision is checked for approval after testing on a dummy dataset and validating the results. Thereafter, the model is applied to beforehand obscure information that the client indicates. In this model, a big part of the information is misleading and the other half is genuine, bringing about a 50 percent reset exactness for the model At least 80% of the information from the two sets will be selected at random for inclusion in our model, with the remaining 20% serving just to test it once it has been completed. Prior to applying a classifier to message information, we should initially eliminate the commotion by using Stanford NLP

(Natural language taking care of) for POS (Part of Speech) dealing with and tokenization of words, and a while later encode the resultant data as numbers and floating point values to be recognized as a commitment to AI computations.Python's scikit-learn package was used to tokenize and extract features from text data since it includes helpful apparatuses like the Count Vectorizer and Tiff Vectorizer.There is a matrix of confusion to help visualise the data. Figure 3 shows the situation.

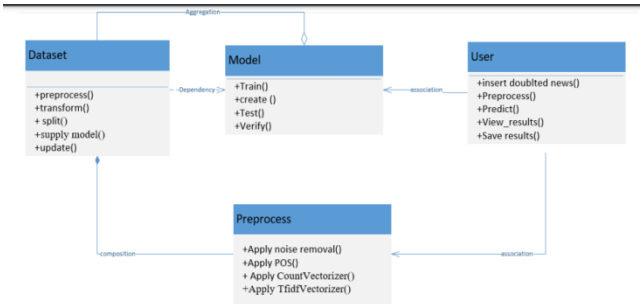


Figure 3. Fake Detector Model

This section goes through the dataset that was selected, The LIAR-PLUS Master that was utilized to clean and concentrate the information,as well as the methodologies that were applied. The evidence sentences in this dataset were manually extracted from the full-text judgment article distributed by columnists on Politifact. True values, as well as parts of speech, were utilised to gather extra four features (nouns, prepositions, and sentences), and each record was named by class mark (0, 1, 2, 3) to be utilized in preparing the model as displayed in Figure 4. The accompanying cycles were taken to evaluate the exactness of the news.

1. Preprocessed Liar-dataset (12.8K)

2. The sentences in various circumstances were sourced from POLITIFACT.COM and carefully labelledPython is used to transform the data from TSV to CSV.

3. In order to remove the noise, we will use the NLP NLTK and SAFAR v2 libraries. The suffix is eliminated by stemming the words, which removes ids, dots, commas, and quotations from the noise. The dataset will next be converted into tokens and statistical values using POS (Part of Speech).

4. Select lexical characteristics for feature extraction, the number of words in an article or speech, as well as the average word length, article length, and the number of speech sections (adjective).

5. Using the Tfidf Vectorizer function in Python Sklearn, extract unigram and bigram features. To produce TF-IDF n-gram features, use this feature extraction package.

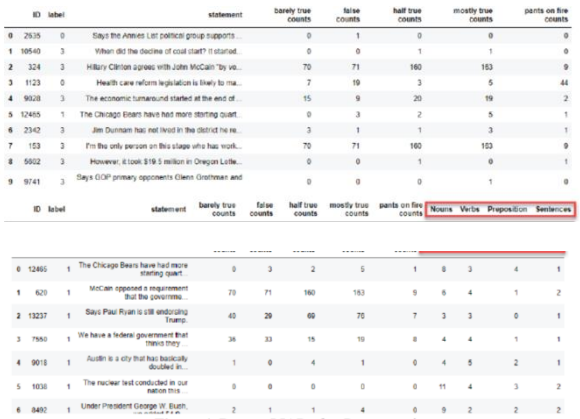
6. Using python sklearn, divide the dataset into 70 percent train and 30 percent test.

7. After applying all of the algorithms, create an ipynb file for the classification model.

8. Create a confusion matrix by testing model precision on the test section of the dataset.

9. Compare false and genuine news for accuracy, precision, recall, and f1-score.

10. Create an interface for users to test previously unseen news.

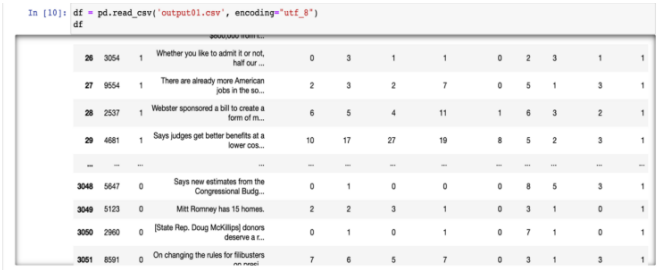


	ID	label	statement	barely true counts	false counts	half true counts	mostly true counts	pants on fire counts
0	2635	0	Says the Annies List political group supports ...	0	1	0	0	0
1	10540	3	When did the decline of coal start? It started...	0	0	1	1	0
2	324	3	Hillary Clinton agrees with John McCain "by vo...	70	71	160	163	9
3	1123	0	Health care reform legislation is likely to ma...	7	19	3	5	44
4	9028	3	The economic turnaround started at the end of ...	15	9	20	19	2
5	12465	1	The Chicago Bears have had more starting quart...	0	3	2	5	1
6	2342	3	Jim Dunnam has not lived in the district he re...	3	1	1	3	1
7	153	3	I'm the only person on this stage who has work...	70	71	160	163	9
8	5602	3	However, it took $19.5 million in Oregon Lotte...	0	0	1	0	1
9	9741	3	Says GOP primary opponents Glenn Grothman and ...	0	0	0	1	0

	ID	label	statement	barely true counts	false counts	half true counts	mostly true counts	pants on fire counts	Nouns	Verbs	Preposition	Sentences
0	12465	1	The Chicago Bears have had more starting quart...	0	3	2	5	1	8	3	4	1
1	620	1	McCain opposed a requirement that the governme...	70	71	160	163	9	6	4	1	2
2	13237	1	Says Paul Ryan is still endorsing Trump.	40	29	69	70	7	3	3	0	1
3	7550	1	We have a federal government that thinks they ...	36	33	15	19	8	4	4	1	1
4	9018	1	Austin is a city that has basically doubled in...	1	0	4	1	0	4	5	2	1
5	1038	1	The nuclear test conducted in our nation this ...	0	0	0	0	0	11	4	3	2
6	8492	1	Under President George W. Bush, ...	2	1	1	4	0	9	2	2	2

Figure 4. Dataset LIAR after Preprocessing

There are two parts to this document: the introduction and the details. The algorithm learns how to discern between real and fake news in the first segment of the data, which makes up around 75% of the total. False news is denoted by the number 0, whereas true news is represented by the number 1. When all the data has been analysed, 25 percent of it will be used for a test to see whether the news is real or phoney, and then the algorithm's percentage will be calculated based on the percentage of accurate and wrong replies.Figures 5 and 6 are examples.



```
In [10]: df = pd.read_csv('output01.csv', encoding="utf_8")
         df
```

26	3054	1	Whether you like to admit it or not, half our ...	0	3	1	1	0	2	3	1	1
27	9554	1	There are already more American jobs in the so...	2	3	2	7	0	5	1	3	1
28	2537	1	Webster sponsored a bill to create a form of m...	6	5	4	11	1	6	3	2	1
29	4681	1	Says judges get better benefits at a lower cos...	10	17	27	19	8	5	2	3	1
...	...	...	...	...	...	...	...	...	...	...	...	...
3048	5647	0	Says new estimates from the Congressional Budg...	0	1	0	0	0	8	5	3	1
3049	5123	0	Mitt Romney has 15 homes.	2	2	3	1	0	3	1	0	1
3050	2960	0	[State Rep. Doug McKillips] donors deserve a r...	0	1	0	1	0	7	1	0	1
3051	8591	0	On changing the rules for filibusters ...	7	6	5	7	0	3	1	3	1

Figure 5. Dataset LIAR after Preprocessing

```
In [13]: from sklearn.model_selection import train_test_split
         X = df[['barely true counts','false counts', 'half true counts','mostly true counts','pants on fire counts','Nouns','Ve
         y = df['label']
         X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=1)

In [75]: # X_train

In [76]: # X_test

In [13]: # y_train

In [14]: # y_test
```

Figure 6. Dataset LIAR after Preprocessing

4 .Results The justification for this adventure is to cover political news data from the Liar dataset, which is a New Benchmark Dataset for False News Detection and is assigned either fake or trustworthy news.We looked at the "Liar" dataset. The six ways of data analysis are shown in the confusion matrix.The following are the six detection algorithms:

• XGboost.

• Random Forests.

• Naive Bayes.

• K-Nearest Neighbors (KNN).

• Decision Tree.

• SVM

The chaos lattice is normally made by Python code using the psychological learning module when the computation code is run on the Anaconda stage.

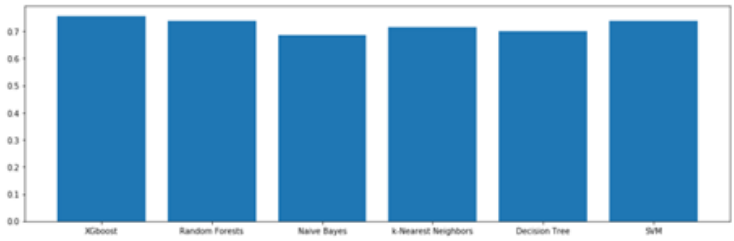


Figure 8. Accuracy Results of all the Algorithms

Figure 7 portrays the Confusion Matrix for the calculations in general:

The accuracy of these algorithms is demonstrated in Figure 8. As indicated, XGBOOST has the greatest accuracy (more than 75%), followed by SVM and Random forest, both of which have a 73 percent accuracy.

5. Conclusion

Two layers of assessment, characterization and disclosure, are used in this research to detect false news. In the first step, social media addresses the fundamental ideas and standards of phony news. Existing strategies for distinguishing counterfeit news utilizing different regulated learning calculations are investigated during the discovery phase When it comes to false news identification in [20], the paper's algorithm is based on text analysis and utilizes models in light of discourse credits and forecast models that aren't viable with different models. Credulous Bayes classifier from [21] was utilized to identify counterfeit news from an assortment of sources with a 74% achievement rate. [22] Incorporating many different types of machine learning techniques, but the accuracy of the probability thresholds they use is only around 85-71%, For unreliable news sources, [23] used the Naive Bayes algorithm to detect fraudulent reports on multiple social media networks. [24] It was 74.5 percent correct data that they got from Kaggle. [27] Detected Twitter spam senders using Naive Bayes algorithms, with accuracy ranging from 70% to 71.2 percent. [28] They used a variety of methods, with a 76 percent accuracy rate. [29] They use three popular methodologies in their studies: Nave Bayes, Neural Networks, and Support Vector Machines (SVM). When it comes to identifying false communications, Nave Bayes has a 96.08 percent accuracy rate. It had a 99.9% success rate for the neural network and machine vector (SVM). [30] By combining KNN and random forests in a false message detection model, they were able to boost their final findings by up to 8%. [31] For the 2012 Dutch races, they tried eight administered AI classifiers on the Twitter dataset. Choice trees are remembered to perform best for this informational index, which has a F score of 88%. [32] N-gram analysis is used to present a counterfeit detection model with the most elevated exactness being used, which consolidates a

unigram and straight SVM exercise manual. At 92 percent precision, this is the best accuracy possible. To summarize our discoveries in the examination rundown and framework investigation, we saw that 70-76 percent of exploration distributions employed the nave bays technique, while the other 30 to 40 percent used on qualitative analysis based on word frequency repetitions and sentiment analysis. [40] [41] [42]. POS textual analysis, a quantitative technique that depends on the addition of numerical statistical values as features, is an addition to these methodologies that we suggest.Increase these attributes and use random forests to improve our results even more, we feel In our proposed dataset, we want to include a number of metrics, such as the number of sentences, the quantity of characters, the normal sentence length, things and relational words, descriptors, etc.

6. References

[1]. Economic and Social Research Council. Using Social Mmedia. Available at: https://esrc.ukri.org/research/impact-toolkit/social-media/using-social-media

[2]. Gil, P. Available at: https://www.lifewire.com/what-exactly-is-twitter-2483331. 2019, April 22.

[3]. E. C. Tandoc Jr et al. “Defining fake news a typology of scholarly definitions”. Digital Journalism , 1–17. 2017.

[4]. J. Radianti et al. “An Overview of Public Concerns During the Recovery Period after a Major Earthquake: Nepal Twitter Analysis.” HICSS '16 Proceedings of the 2016 49th Hawaii International Conference on System Sciences (HICSS) (pp. 136-145). Washington, DC, USA : IEEE. 2016.

[5]. Alkhodair S A, Ding S H.H, Fung B C M, Liu J 2020 “Detecting breaking news rumors of emerging topics in social media” Inf. Process. Manag. 2020, 57, 102018.

[6]. Jeonghee Yi et al. “Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. ”In Data Mining, 2003. ICDM 2003. Third IEEE International Conference (pp. 427-434). http://citeseerx.ist.psu.edu. 200).2003

[7]. Tapaswi et al. “Treebank based deep grammar acquisition and Part-Of-Speech Tagging for Sanskrit m sentences.” Software Engineering (CONSEG), on Software Engineering (CONSEG), (pp. 1-4). IEEE. 2012

[8]. Ranjan et al. “Part of speech tagging and local word grouping techniques for natural language parsing in Hindi”. In Proceedings of the 1st International Conference on Natural Language Processing (ICON 2003). Semanticscholar. 2003

[9]. MonaDiab et al. Automatic Tagging of Arabic Text: From Raw Text to Base Phrase Chunks. Proceedings of HLT-NAACL 2004: Short Papers (pp. 149–152). Boston, Massachusetts, USA: Association for Computational Linguistics. 2004

[10]. Rouse, M. https://searchenterpriseai.techtarget.com/definition/machine-learning-ML May 2018

[11]. Sumeet Dua, Xian Du. “Data Mining and Machine Learning in Cybersecurity”. New York: Auerbach Publications.19 April 2016.

[12]. RAY, S. https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/ 2017, September

[13]. Huang, T.-Q. (n.d.) https://www.researchgate.net/figure/Pseudo-code-of-information-gain-basedrecursive-feature-elimination-procedure-with-SVM_fig2_228366941 2018

[14]. Researchgate.net. Available at: Available at: https://www.researchgate.net/figure/Pseudocode-ofnaive-bayes-algorithm_fig2_325937073. 2018.

[15]. Researchgate.net. Available at: https://www.researchgate.net/figure/Pseudocode-for-KNNclassification_fig7_260397165, 2014.

[16]. Rampersad G, Althiyabi T 2020 “Fake news: Acceptance by demographics and culture on social media” J. Inf. Technol. Politics 2020, 17, 1–11.

[17]. NaphapornSirikulviriya; SukreeSinthupinyo. “Integration of Rules from a Random Forest.” International Conference on Information and Electronics Engineering (p. 194 : 198). Singapore: semanticscholar.org. 2011.

[18]. Jasmin Kevric et el. “An ef

[19]. ShivamB.Parikh and PradeepK.Atrey. “Media-RichFake News Detection: A Survey.” IEEE Conference on Multimedia Information. Miami, FL: IEEE. 2018.

[20]. MykhailoGranik and VolodymyrMesyura. “Fake news detection using naive Bayes classifier.” First Ukraine Conference on Electrical and Computer Engineering (UKRCON). Ukraine : IEEE. 2017.

[21]. Gilda, S. “Evaluating machine learning algorithms for fake news detection.” 15th Student Conference on Research and Development (SCOReD) (pp. 110-115). IEEE. 2017.

[22]. Akshay Jain and AmeyKasbe. “Fake News Detection.” 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). Bhopal, India: IEEE. 2018.

[23]. Yumeng Qin et al. “Predicting Future Rumours.” Chinese Journal of Electronics (Volume: 27 , Issue: 3 , 5 2018, 514 - 520.

[24]. ArushiGupta and RishabhKaushal. “Improving spam detection in Online Social Networks.” International Conference on Cognitive Computing and Information Processing (CCIP). semanticscholar.org.2015.

[25]. Khanam, Z., Ahsan, M.N.”Evaluating the effectiveness of test driven development: advantages and pitfalls.”International. J. Appl. Eng. Res. 12, 7705–7716, 2017

[26]. Khanam, Z. “Analyzing refactoring trends and practices in the software industry.” Int. J. Adv. Res. Comput. Sci. 10, 0976–5697, 2018.

[27]. Veronica Perez-Rosas et al. Available at: https://www.researchgate.net/publication/319255985_Automatic_Detection_of_Fake_News August, 2017.

[28]. SupanyaAphiwongsophon et al. “ Detecting Fake News with Machine Learning Method.” 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). Chiang Rai, Thailand, Thailand: IEEE . 2018.

[29]. Prabhjot Kaur et al. “ Hybrid Text Classification Method for Fake News Detection.” International Journal of Engineering and Advanced Technology (IJEAT) , 2388-2392. 2019

[30]. Looijenga, M. S. “The Detection of Fake Messages using Machine Learning.” 29 Twente Student Conference on IT, Jun. 6th, 2018, Enschede, The Netherlands. Netherlands: essay.utwente.nl. 2018.

[31]. I. Traore et al. “Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques.” International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments (pp. 127–138). Springer International Publishing . 2017.

[32]. Khanam Z., Alkhaldi S. “An Intelligent Recommendation Engine for Selecting the University for Graduate Courses in KSA: SARS Student Admission Recommender System.” In: Smys S., Bestak R., Rocha Á. (eds) Inventive Computation Technologies. ICICIT 2019. Lecture Notes in Networks and Systems, vol 98. Springer, Cham. 2019.

[33]. Khanam Z. and Ahsan M.N. “ Implementation of the pHash algorithm for face recognition in secured remote online examination system.” International Journal of Advances in Scientific Research and Engineering (ijasre) Volume 4, Issue 11 November. 2018.

[34]. Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. ACM Transactions on Intelligent Systems and Technology (TIST), 10(3), 1-42.

[35]. Sharma, Karishma, et al. "Combating fake news: A survey on identification and mitigation techniques." ACM Transactions on Intelligent Systems and Technology (TIST) 10.3 (2019): 1-42.

[36]. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter, 19(1), 22-36.

[37]. Shu, Kai, et al. "Fake news detection on social media: A data mining perspective." ACM SIGKDD explorations newsletter 19.1 (2017): 22-36.

[38]. Khanam Z. and Agarwal S. Map-reduce implementations: Survey and Performance comparison, International Journal of Computer Science & Information Technology (IJCSIT) Vol 7, No 4, August 2015.

[39]. Zhang, Jiawei, Bowen Dong, and S. Yu Philip. "Fakedetector: Effective fake news detection with deep diffusive neural network." 2020 IEEE 36th International Conference on Data Engineering (ICDE). IEEE, 2020.