IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 6, June 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906297 |

Fake News Detection Using Machine Learning

Puja Sunil Erande[1], Prof. Monika Rokade[2]

[1]PG Student, Sharadchandra Pawar College of Engineering, Junnar, Pune, India

[2]Assistant Professor, Sharadchandra Pawar College of Engineering, Junnar, Pune, India

ABSTRACT:. Fake news has become a significant part of social media and the political realm in recent years. Fake news identification is an important research project, but it comes with its own set of obstacles. Some difficulties may arise as a result of a scarcity of resources, such as a dataset and published literature. In this research, we offer a method for detecting bogus news using machine learning techniques. Three different machine learning classification techniques are compared. Furthermore, we will be using three different models: Logistic Regression, Decision Tree Classifier, and Random Forest Classification.According to the findings of our investigation, we have attained varying levels of accuracy for each method. Our project will be extremely useful in determining whether the given news is true or not.

I.INTRODUCTION

Fake news is raging throughout social media platforms in large quantities. The designation of any article, post, storey, or journal as false or true has become a vital issue in this instance, and it has also piqued the interest of researchers all over the world. According to various research studies conducted to determine the impact of any false or fictional news on us when we return through such fake news facts. Falsified news or information is employed in such a way that an individual's basic brain process is focused on one topic that may or may not be accurate.The pandemic catastrophe that is currently affecting the entire planet is the best illustration of fake news. There are a variety of news items that have been faked and utilised solely to confuse and disrupt people's brains, leading them to believe false news. Is it, however, possible for anyone to tell if it's phoney or real?False information on Indian social media prompted some voters to drink cow wewe or eat dung in order to avoid illness, but in the country, artiodactyl weewee with lime was lauded as a coronavirus-fighting remedy.As a therapy for the probable fatal sickness, the experts looked at completely different stories, such as ingesting garlic, wearing heat socks, and smearing goose fat on one's chest.

II.LITERATURE REVIEW

MahdiehLabaniet. al. [1] proposed a system multivariate filter method for feature selection which is used for various text classification approach. This method focuses on the reduction of redundant features using minimal-redundancy and maximal-relevancy concepts. The proposed method takes into account document frequencies for each term, while estimating their usefulness. It not only selects the features with maximum relevancy but also the redundancy between them is taken into account using a correlation metric. Results obtained using this approach are better than state-of-the-art filter methods.

Asriyanti Indah Pratiwi and Adiwijaya proposed a system [2] Feature Selection and Classification Based on Information Gain for Document Sentiment Analysis. Information Gain Classifier (IGC) is used to extract the various features from movie review dataset. Authors proposed IG-DF-FS based hybrid method called a combination of Information Gain + Document Frequency Feature Selection etc.

Haoyue Liu et. Al. [3] proposed a system of feature selection for imbalanced data. If the dataset is imbalanced, it has a problem of bias-to-majority. This issue is solved in it using Weighted Gini Index (WGI) approach. The WGI approach calculates an impurity reduction score for each feature and features with a high score are considered as important.

Asha S Maneket. al. [4] proposed aspect term extraction for sentiment analysis of movie review dataset. The work is carried out using the Gini index approach for feature selection after NLP processing. It uses SVM classifier to classify test data. This study illustrates a statistical method for weight calculation by Gini Index method for feature selection in sentiment analysis. This framework for sentiment analysis using SVM classifier is compared with other feature selection methods on movie reviews and results have shown that classification by using this efficient method has improved the accuracy.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906297 |

Muhammad ZubairAsgharet. al. [5] proposed a system Aspect-based opinion mining framework using heuristic patterns. The work proposed an integrated framework comprising of an extended set of heuristic patterns generated using POS tags for aspect extraction, a hybrid sentiment classification module with the additional support of intensifiers and negations, and a summary generator. The system obtained classification results with improved precision (0.85) when compared to the alternative methods available. This method is quite generalized and it can classify aspect-based opinions in multiple domains.

Kim Schouten et. al. [6] proposed a system Aspect Category Detection for Sentiment Analysis for supervised as well as unsupervised learning. In this work, the first method presented is an unsupervised method that applies association rule mining on co-occurrence frequency data obtained from a corpus to find aspect categories. The second, supervised, method uses a rather straightforward co-occurrence method where the co-occurrence frequency between annotated aspect categories and both lemmas and dependencies is used to calculate conditional probabilities. If the maximum conditional probability is higher than the associated, trained, threshold, the corresponding aspect category is assigned to that sentence. The accuracy of the system is around 83% for a supervised method.

Laith Mohammad Abualigahet. al. [7] proposed a feature selection method which is a hybrid of Genetic operators (GA) and particle swarm optimization algorithmfor text clustering. The hybrid approach improved the accuracy of text clustering. The GA is used to solve the unsupervised feature selection problem, called Feature Selection based Genetic Algorithm for Text Classification (FSGATC). This method is used to create a new subset of informative features in order to obtain more accurate clusters on different review text datasets. This method also overcomes the other comparative methods in improving text clustering results based on different common benchmark datasets used in the domain of text mining.

BasantAgarwalet. al. [8] proposed a system Concept-Level Sentiment Analysis with Dependency-Based Semantic Parsing. This system illustrates a fundamental issue of the sentiment analysis task and uses concepts as features. It presents a concept extraction algorithm based on a novel concept parser scheme to extract semantic features that exploit semantic relationships between words in natural language text. The system also extracts the actual concept using ConceptNet ontology like RDF framework. Concepts extracted from the text are sent as queries to ConceptNet to extract their semantics. It selects important concepts and eliminates redundant concepts using the Minimum Redundancy and Maximum Relevance feature selection technique. All selected concepts are then used to build a machine learning model that classifies a given document as positive or negative.

Data Hiding In Audio-Video Using Anti Forensics Technique For Authentication has proposed by Sunil S.Khatal and Yogesh kumar Sharma [9]. This is a secure data hiding approach for hide the text data into video as well as image. Once sender hide data into specific objects while receivers does same operation for authentication. The major benefit of this system can eliminate zero day attacks in untrusted environments.

Sunil S.Khatal and Yogesh Kumar Sharma [10] proposed a system to analyzing the role of Heart Disease Prediction System using IoT and Machine Learning. This is the analytical based system to detection and prediction of heart disease from IoT dataset. This system can able to detect the disease and predict accordingly.

Monika Rokade and Yogesh Patil [11] proposed a system deep learning classification using nomaly detection from network dataset. The Recurrent Neural Network (RNN) has classification algorithm has used for detection and classifying the abnormal activities. The major benefit of system it can works on structured as well as unstructured imbalance dataset.

The MLIDS A Machine Learning Approach for Intrusion Detection for Real Time Network Dataset has proposed by Monika Rokade and Dr. Yogesh Patil in [12]. The numerous soft computing and machine learning classification algorithms have been used for detection the malicious activity from network dataset. The system depicts around 95% accuracy ok KDDCUP and NSLKDD dataset.

Monika D. Rokade and Yogesh Kumar Sharma [13] proposed a system to identification of Malicious Activity for Network Packet using Deep Learning. 6 standard dataset has sued for detection of malicious attacks with minimum three machine learning algorithms.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906297 |

Sunil S. Khatal and Yogesh kumar Sharma [14] proposed a system Health Care Patient Monitoring using IoT and Machine Learning for detection of heart and chronic diseases of human body. The IoT environment has used for collection of real data while machine learning technique has used for classification those data, as it normal or abnormal.

III.METHODOLOGY OF PROPOSED SURVEY

Outlined the many types of fake news in their latest paper, which is given below.
1. Visual-based: These false news pieces make extensive use of graphics as content, which might include manipulated images, doctored video, or a combination of the two.
2. User-based: This type of false news is created by phoney accounts and targeted at certain audiences, such as certain age groups, genders, cultures, and political affiliations**.**
3. Knowledge-based: These types of posts provide scientific (so-called) justification for some unsolved problems, leading viewers to believe they are genuine. Natural therapies for elevated sugar levels in the physical body, for example.
4. Style-based: Posts are written by photojournalists from the UN agency, some of whom are phoney and replicas of licenced journalists.
5. Stance-based: It is the representation of true statements in such a way that their meaning and purpose are altered.
The goal of this research is to create a model for detecting fake news using three machine learning methods. Because the focus of this project is on model building in machine learning using a jupyter notebook, it is not constantly developing new usual package systems. Machine learning often necessitates a large and high-quality dataset, as well as a significant amount of time for model training and testing. In other words, if the model provides predictable results, such as the prediction of fake and actual news, the model is considered to be fairly accurate.

Management of Data
This part collects a body of knowledge (dataset), which could be a collection of report articles, stories, news, or blog postings. Once the dataset has been collected, nltk is used to identify a collection of written or spoken material stored on a computer and used to discover how language is utilised: the data is investigated to obtain a better understanding of its structure, which means stopwords are deleted.

Data Exploration
The charting of graphs according to the fake and true news anticipated by the machine learning algorithm is the major focus of the information exploration section. Word clouds are created, which are simply a visual image technique for conveying text information in which the size of each word represents its frequency or relevance.
A word cloud is used to highlight important information points. Tokenization is performed with this approach.

Model Training
The machine learning model can then be trained after the data has been adequately analysed and controlled. During the Model Training phase, many methodologies are considered, and a learning problem that is a prediction task is determined. Whatever possibilities are available within the training data set are then investigated. The model is then trained using an appropriate algorithm. We employed three algorithms in our case: Logistic Regression, Decision Tree Classifier, and Random Forest Classifier. The dataset is then matched to the algorithm's rule for training purposes, and the testing is completed.

Model Assessment
The output of the model developed is measured in numerous ways while evaluating it. The model's correctness is graded using performance indicators such as F1 score, precision, recall, and accuracy rate, which are based on the confusion matrix report. Various alterations are frequently made to the model till satisfaction is obtained in terms of the model's creation yielding clever precision of output.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906297 |

IV.RESULTS AND DISCUSSION

We built this project around three machine learning methods, each of which has its unique accuracy percentage when applied to the dataset. The following are the accuracy levels for each algorithm:

Classifier	Accuracy
Logistic Regression:	98.8%
Decision Tree Classifier	99.6%
Random Forest Classifier	98.9%

V.CONCLUSION AND FUTURE WORK

Fake news has a harmful influence on society every time it is spread. When it comes to distinguishing between fake and authentic news, there is still a lot of confusion in society. Fake news is a false alarm for everyone since it always misleads the readers, leaving them confused and unable to act appropriately. With their own eyes, they see their daily lives. So, when will our project be able to confidently forecast whether the provided news is fake or not?People will be able to check whether the news they have in front of their eyes is real or not if they consider our project's idea, and they will become more conscious of the spread of fake news. This system was completed in the final year, but it will undoubtedly benefit from more enhancements in the near future, such as the use of a flask.

.

REFERENCES

[1] Labani M, Moradi P, Ahmadizar F, Jalili M. A novel multivariate filter method for feature selection in text classification problems.Engineering Applications of Artificial Intelligence. 2018 Apr 1;70:25-37

[2] Pratiwi AI. On the feature selection and classification based on information gain for document sentiment analysis. Applied Computational Intelligence and Soft Computing. 2018;2018.

[3] Liu H, Zhou M, Lu XS, Yao C. Weighted Gini index feature selection method for imbalanced data. In2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC) 2018 Mar 27 (pp. 1-6). IEEE.

[4] Manek AS, Shenoy PD, Mohan MC, Venugopal KR. Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. World wide web. 2017 Mar 1;20(2):135-54.

[5] Asghar MZ, Khan A, Zahra SR, Ahmad S, Kundi FM. Aspect-based opinion mining framework using heuristic patterns. Cluster Computing. 2017:1-9.

[6] Schouten K, Van Der Weijde O, Frasincar F, Dekker R. Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data. IEEE transactions on cybernetics. 2017 Apr 14;48(4):1263-75.

[7] Abualigah LM, KhaderAT, Al-Betar MA. Unsupervised feature selection technique based on genetic algorithm for improving the text clustering. In2016 7th international conference on computer science and information technology (CSIT) 2016 Jul 13 (pp. 1-6).IEEE

[8] Agarwal B, Poria S, Mittal N, Gelbukh A, Hussain A. Concept-level sentiment analysis with dependency-based semantic parsing: a novel approach. Cognitive Computation. 2015 Aug 1;7(4):487-99.

[9]Sunil S.Khatal ,Dr.Yogesh kumar Sharma, "Data Hiding In Audio-Video Using Anti Forensics Technique ForAuthentication ", IJSRDV4I50349, Volume : 4, Issue : 5

[10]Sunil S.Khatal Dr. Yogesh Kumar Sharma. (2020). Analyzing the role of Heart Disease Prediction System using IoT and Machine Learning. *International Journal of Advanced Science and Technology*, *29*(9s), 2340 - 2346.

[11] Monika D.Rokade ,Dr.Yogesh kumar Sharma,"Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic."IOSR Journal of Engineering (IOSR JEN),ISSN (e): 2250-3021, ISSN (p): 2278-8719

[12] Monika D.Rokade ,Dr.Yogesh kumar Sharma"MLIDS: A Machine Learning Approach for Intrusion Detection for Real Time Network Dataset", 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), IEEE

[13]Monika D.Rokade, Dr. Yogesh Kumar Sharma. (2020). Identification of Malicious Activity for Network Packet using Deep Learning. *International Journal of Advanced Science and Technology*, *29*(9s), 2324 - 2331.

[14] Sunil S.Khatal ,Dr.Yogesh kumar Sharma, "Health Care Patient Monitoring using IoT and Machine Learning.", IOSR Journal of Engineering (IOSR JEN), ISSN (e): 2250-3021, ISSN (p): 2278-8719

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details