iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VII **Month of publication:** July 2020

DOI: https://doi.org/10.22214/ijraset.2020.29630

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VII July 2020- Available at www.ijraset.com

Fake News Detection Methods: Machine Learning Approach

Dr. Madhavi Ajay Pradhan[1], Ankita Shinde[2], Rohan Dhiman[3], Shreyas Ghorpade[4], Swapnil Jawale[5]
[1]*Research Guide, Associate Professor, Computer Engineering , AISSMS COE , Pune, India*
[2, 3, 4, 5]*Computer Engineering, AISSMS COE , Pune, India*

Abstract: ***Lots of stuff you scan online particularly in your social media feeds could seem to be true, typically isn't. Fake news is news, stories or hoaxes that are created to intentionally mislead or deceive the readers. Usually, such kind of stories or news are created to either influence peoples views, some wrong belief for some political issue or cause confusion and might be a kind of money earning business for the online fake news publishers. Fake news stories will deceive individuals by looking like trustworthy internet sites or using similar names and web addresses of reputed news organisations. The widespread of fake news has the potential for terribly negative impacts on society and people. Therefore, fake news detection on social media has recently become a very crucial agenda that has attracted a lot of attention. Fake news detection on social media presents distinctive characteristics and challenges that make existing detection algorithms from ancient print media ineffective or not applicable. Machine learning is a method of analyzing data which automates analytical model building. It is one of the most important branches of artificial intelligence which is based on the idea that systems can learn from data, identifies the patterns and make decisions with minimal human intervention. ML can handle things like growing volumes and varieties of available data, have good computational processing that is cheaper and more powerful, and affordable data storage. All of the above things conclude that it's possible to quickly and automatically produce models that can analyze bigger, more complex data and deliver faster with more accurate results. In this project, we have a huge dataset to analyze and train the model accordingly. So ML would be a very efficient and faster method to proceed in this project which will help in better development of the desired system.***
Index Terms: fake news detection, machine learning.

I. INTRODUCTION

Fake news is a term often used to refer to fabricated news. This type of news is found in traditional news, social media platform or various websites of fake news, It has no basis but it is represented true factually. Social media for news consumption has both favorable and unfavorable consequences

On one hand, its easy to access, has low cost, and rapid spread of information lead people to look for and consume news from various social media platforms. On the other hand, it authorizes the widespread of fake news, that is low- quality news with intentionally erroneous information. Firstly, fake news can break the originality balance of the whole news ecosystem. Secondly, fake news intentionally prevails consumers to accept biased or untrue beliefs. Fake news is usually manipulated by the promoter to convey political messages or influence .

Thirdly, fake news changes the way people interpret and respond to real news. For example, few fake news was just created to trigger the publics distrust and make them demented , impeding their abilities to distinguish what is true and what is not . To help diminish the negative effects caused by fake news , both to benefit people and the news ecosystem. Its crucial that we develop various methods to automatically detect fake news on social media. In this 21st century, the influence of fake news became widespread, also the usage of the term.

The opening of the Internet to the public in the 90s was a movement that meant to allow them access to information. Over time, as it is evident the Internet has grown to unimaginable heights with tons of facts coming in continuously which allows the Internet to be a platform as a host for plenty of untruthful, unwanted and information that is misleading that can be made by anyone. Fake news has been growing from being sent via emails to attacking social media platforms.

Besides referring to the stories that are made up designed to trick readers into clicking on links, maximizing traffic and the profit, the term has also referred to ironic news, whose purpose is not exactly to mislead but rather to inform viewers and share humorous commentary about real news and the media.

©IJRASET: All Rights are Reserved

Maps and pictures of Australia's unprecedented bushfires have spread widely on social media.

Fig. 1. This False Post Was Shared Widely In A Very Short Span Of Time

II. LITERATURE SURVEY

The available literature has described many automatic detection techniques of fake news and deception posts. Since there are many multidimensional aspects of fake news detection ranging from chatbots to use of click baits for the rumor spreading and spreading wrong information. There are many click baits available in social media networks which include Facebook which enhances sharing, liking and commenting on posts which in turn spreads wrong information. A lot of work has been done to detect and analyze falsified information. The authors in [6] have described linguistic cue approaches(with machine learning) and network analysis approaches. Their aim was to create a hybrid approach that combines linguis- tic cue and machine learning, with network based behavioral data. Furthermore they have described deep syntax analysis and semantic analysis under linguistic cue approach, linked data analysis and social network behaviour under network analysis approach. Even though the approaches they used gives high accuracy classification, they are limited only to textual data and do not explore the unstructured data. The authors in [7] have described and compared 4 models for fake news detection namely nave bayes, SVM, LSTM(Long- Short Term Memory) which is an extension of RNN and Keras based neural network. They used the data from kaggle which contained 33000 articles, as the data collected was noisy they performed pre-processing before applying the model. They compared the models on the basis of accuracy and calculated f1 score and derived a confusion matrix, later represented it in a tabular format. Fig 2. shows the result of the accuracy they observed. Evidently LSTM gives the highest accuracy which is an advantage only when the text is inherently a serialized object. Otherwise it wont give the same accuracy. The authors in [8] make use of one unified word vector for the key sentences of article by extracting them to the question from article and then merging the word vector for each key sentence. They state that it can efficiently perform the sentence matching by executing matching operations between the contextual information obtained from the word vectors of question and key sentences through bidirectional long short term memory. They have achieved an accuracy ranging from 64 to 69%. Their future plan is to develop a more advanced model which applies the model independently to each key sentence. They have achieved an accuracy ranging from 65 to 70%. Their future plan is to develop a more advanced model which applies the model independently to each key sentence. In this paper[9] authors have proposed simple ap- proach for fake news detection using naive Bayes classifier and tested it against a data set of Facebook news posts. They have achieved classification accuracy of approximately 74% on the test set with this model.Naive Bayes typically use bag of words features to identify fake news, an approach commonly used in text classification. Naive Bayes classifiers work by correlating the use of tokens (typically words, or sometimes other constructions, syntactic or not), with fake and genuine news and then using Bayes theorem to calculate a probability that the news is fake or not.A refined issue (or ”disadvantage”) with Naive-Bayes is that if you have got no occurrences of a category label and an explicit attribute worth along (e.g. class=”nice”, shape=” sphere”) then the frequency-based chance estimate are zero. Given Naive-Bayes’ conditional independence assumption, once all the possibilities are increased you’ll get zero and this may have an effect on the posterior chance estimate.

©IJRASET: All Rights are Reserved

Model	Accuracy
Naïve Bayes	71.84%
Support Vector Machine	87.37%
LSTM	94.27%
Keras Based Neural Network	90.62%

Fig. 2. Comparison Of Accuracy

III. TYPES OF DATA ON SOCIAL MEDIA

A. Text

It is analyzed by computational linguistics and focuses on the origin of the text semantically and systematically. Nowadays most of the post or news are spread which contain 70% text therefore the need to concentrate on this type is equal as that of multimedia.

B. Multimedia

Multimedia basically comprises of images, videos, gifs, audios like podcasts on a different platform. Images are a more informative way of any news by showing the happening of an incident falser by making so photoshopped changes. Videos are the series of events showing the event but similar to images can be modified but not unto that extent.

C. Hyperlink

Hyperlinks are the path to a website that consists of news that is falsely spread. Many individuals can click on the links accidentally or by will and go on the websites and read or see the news , may or may not believe it and may even forward it to different groups.

IV. TYPES OF FAKE NEWS

A. Visual Based

Visuals are often being used as persuasive vehicles for mis- as well as dis- information. And pictures or photographs are not the only kind of visuals to consider. Graphs and charts with their uniform and scientific format are increasingly used as means for misinterpreting the information, as they can give a false sense of authority. These trends are concerning. Recent research has been analyzed and demonstrated that people are very poor at identifying manipulated or edited images and while the technology companies have taken into consideration to tackle the problem of the sharing of misinformation and disinformation, their efforts have focused totally on textual data only, which is much easier to discover , interpret and analyze.

	#words	title	text	label
0	8476	You Can Smell Hillarys Fear	Daniel Greenfield, a Shillman Journalism Fello...	0
1	10294	Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg Linkedin Reddit Stumbleu...	0
2	3608	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	1
3	10142	Bernie supporters on Twitter erupt in anger ag...	Kaydee King (@KaydeeKing) November 9, 2016 Th...	0
4	875	The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners...	1

Fig. 3. Structure Of The Dataset

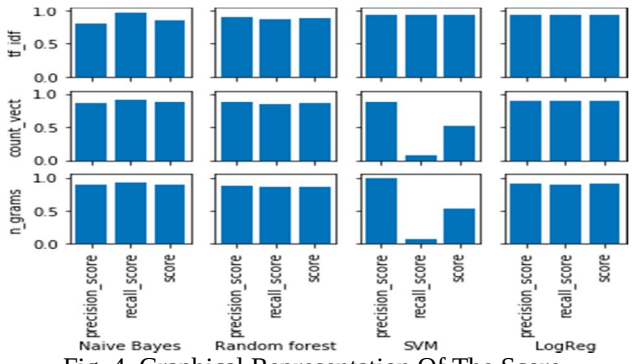


Fig. 4. Graphical Representation Of The Score

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VII July 2020- Available at www.ijraset.com

B. User Based

The user based theories carry out a systematic inquiry of fake news from a user's perspective, considering how users engage with fake news. This type of fake news is targeted at a specific type of audience. The audience can be placed in a particular group based on age, gender, etc.

V. EXISTING SYSTEM

There are some existing research and projects been done for the classification of fake news . Considering the following dataset and running a few existing algorithm we could find out what accuracy each algorithm gives us .

Applying TF-IDF Vectorizer , Count Vectorizer and N-Gram Tagger on the dataset . (Fig 3.)

TF-IDF stands for Term frequency-inverse document fre- quency. The tf-idf weight is a weight often used in information retrieval and text mining. Variations of the tf-idf weighting scheme are often used by search engines in scoring and ranking a documents relevance given a query. This weight is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the corpus (data-set).

The Count Vectorizer provides a simple way to both tokenize a collection of text documents and build a vocabulary of known words, but also to encode new documents using that vocabulary.

NgramTagger has 3 subclasses UnigramTagger, BigramTag- ger and TrigramTagger. A single token is referred to as a Unigram, for example hello; movie. Bigram is 2 consecutive words in a sentence. Trigram is 3 consecutive words in a sentence.

Now applying the algorithms and finding out the accuracy for all the algorithms to compare which algorithm performs better (Fig 4.). For TF-IDF using the Multinomial Naive Bayes Model gave an accuracy of 86% , the Random Forest Classifier gave an accuracy of 88% , and for Support Vector Machine (SVM) model it gave an accuracy of 93.4%. For Count Vectorizer using the Multinomial Naive Bayes Model gave an accuracy of 88.8% , the Random Forest Classifier gave an accuracy of 85.9% , and for Support Vector Machine (SVM) model it gave an accuracy of 52.4%. For ngrams using the Multinomial Naive Bayes Model gave an accuracy of 89.9% , the Random Forest Classifier gave an accuracy of 86.7% , and for Support Vector Machine (SVM) model it gave an accuracy of 52.4%.

VI. PROPOSED SYSTEM

We have proposed a simple system for the detection of Fake News. For the detection of fake news, we'll be setting up a website over which anyone can search for whether the news they see is fake or real by uploading the news over the website. The website would use one of the techniques that comes out to be the best mentioned above in the paper for the verification process of the news.

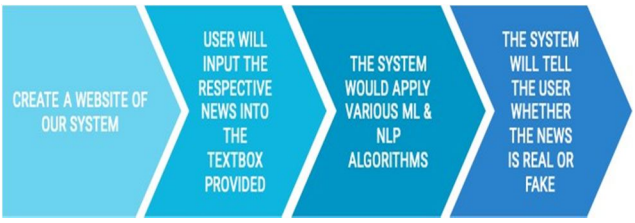


Fig. 5. Proposed System

VII. CONCLUSION

Therefore our research concludes that even though there exists a system for fake news detection, there is a need for an upgrade for the existing systems as the accuracy and efficiency we are aiming to achieve can't be gratified by existing systems. Hence by the Machine learning approach, we aim to achieve the utmost fidelity by using algorithms as mentioned above viz. Naive Bayes, Linguistic, Support Vector Machine, Clustering techniques.

VIII. ACKNOWLEDGMENT

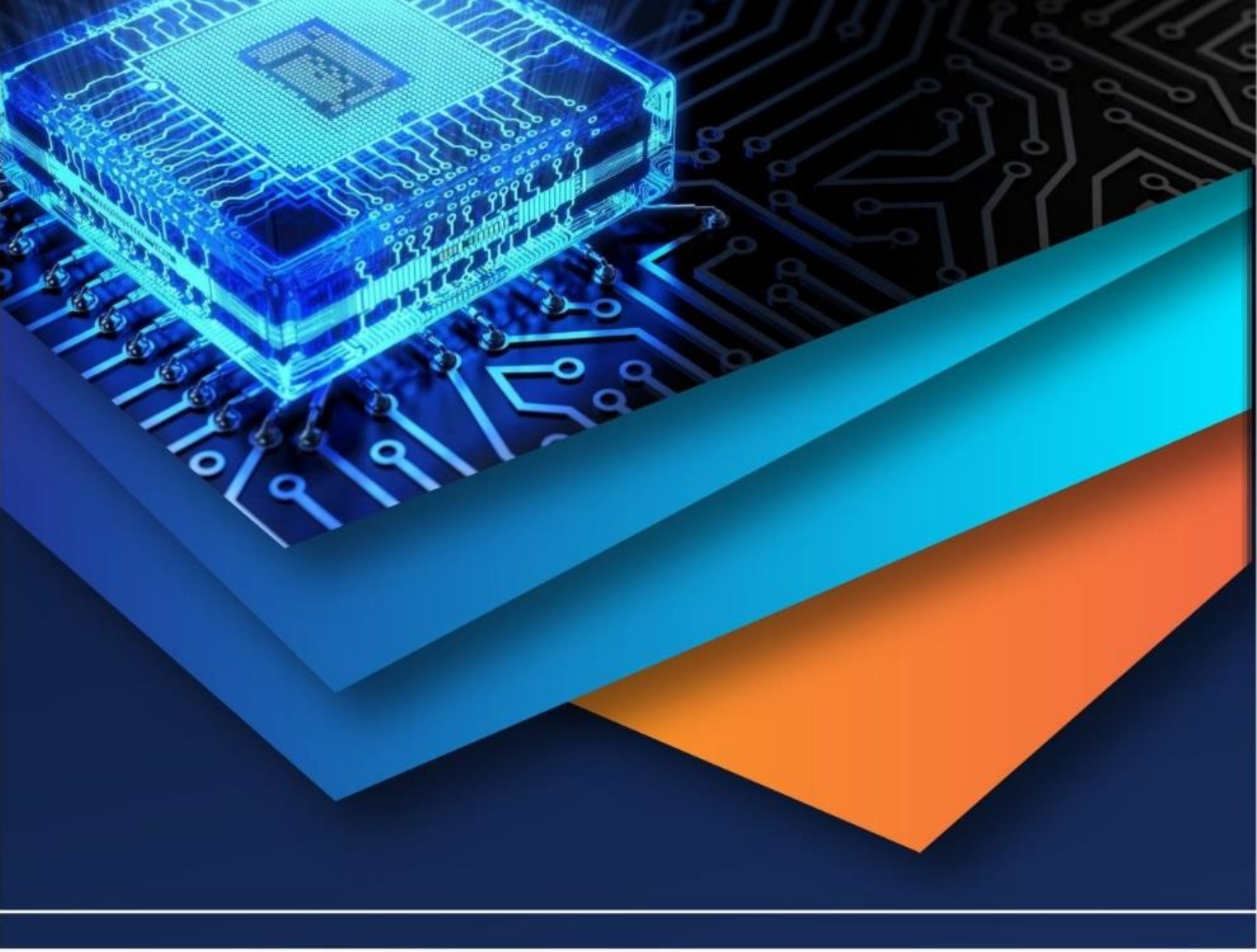
We the authors would like to extend our sincere gratitude and thanks to our guide Prof. M. A. Pradhan, for her invaluable guidance and for giving us useful inputs and encouragement, which inspired us to work harder. Due to her forethought, appreciation of work involved and continuous imparting of useful tips, this research has been successfully completed.

©IJRASET: All Rights are Reserved

REFERENCES

[1] Syed Ishfaq Manzoor , Jimmy Singla and Nikita, Fake News Detection Using Machine Learning approaches: A systematic Review, 2019 3rd In- ternational Conference on Trends in Electronics and Informatics (ICOEI).
[2] Kyeong-Hwan Kim and Chang-Sung Jeong, Fake News Detection System using Article Abstraction,2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)
[3] Zaitul Iradah Mahid , Selvakumar Manickam and Shankar Karup- payah,Fake News on Social Media: Brief Review on Detection Tech- niques,2018 Fourth International Conference on Advances in Computing, Communication Automation (ICACCA).
[4] Mykhailo Granik and Volodymyr Mesyura,Fake News Detection Using Naive Bayes Classifier,2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON).
[5] Shlok Gilda, Evaluating Machine Learning Algorithms for Fake News Detection, 2017 IEEE 15th Student Conference on Research and Devel- opment (SCOReD).
[6] N. J. Conroy, V. L. Rubin, and Y. Chen,Automatic deception detection: Methods for finding fake news,Proceedings of the Association for Infor- mation Science and Technology, vol. 52, no. 1, pp. 14, 2015.
[7] Tijare, Poonam.,A Study on Fake News Detection Using Nave Bayes, SVM, Neural Networks and LSTM.,2019.
[8] Namwon Kim, Deokjin Seo, Chang-Sung Jeong, FAMOUS: Fake News Detection Model based on Unified Key Sentence Information.,2018 IEEE 9th International Conference on Software Engineering and Service Sci- ence (ICSESS).
[9] Mykhailo Granik, Volodymyr Mesyura, Fake News Detection Using Naive Bayes Classifier,2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON) .

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)