

PROJECT – REPORT (INT 247)

Roll No: **RK19ZNA28**

Section: **K19ZN**

MACHINE LEARNING FOUNDATION

TOPIC: FAKE NEWS DETECTION



SUBMITTED BY: **TANISH VERMA - 11907877**

BTECH. CSE SEC: K19ZN

SUBMITTED TO: **DR. SAGAR PANDE**

- 23754

INDEX

1. CERTIFICATE
2. ACKNOWLEDGEMENT
3. ABSTRACT
4. INTRODUCTION
5. LITERATURE REVIEW / RELATED WORKS
6. METHODOLOGY
7. RESULT
8. CONCLUSION
9. SCREENSHOTS
10. REFERENCES

ACKNOWLEDGMENT

I am over whelmed in all humbleness and gratefulness to acknowledge my depth to all those who have helped me to put these ideas, well above the level of simplicity and into something concrete.

I would like to express my special thanks of gratitude to my teacher **Dr. Sagar Pande** as well as to **Lovely Professional University** who gave me the golden opportunity to do this wonderful project on the topic **(Fake News Detection)**, which also helped me in doing a lot of Research and I came to know about so many new things.

I am really thankful to them.

Any attempt at any level can 't be satisfactorily completed without the support and guidance of my Teacher. I would like to thank my teacher who helped me a lot in gathering different information, collecting data and guiding me from time to time in making this project, despite of their busy schedules, they gave me different ideas in making this project unique.

Tanish Verma

School of Computer Science and Engineering, Lovely Professional University, Jalandhar, India.

Email: tanish.verma762@gmail.com

Abstract: The fake news on social media and various other media is widely spreading and is a matter of serious concern due to its ability to cause a lot of social and national damage with destructive impacts. A lot of research is already focused on detecting it. This paper makes an analysis of the research related to fake news detection and explores the traditional machine learning models to choose the best, in order to create a model of a product with supervised machine learning algorithm, that can classify fake news as true or false, by using tools like python, Machine Learning for textual analysis. This process will result in feature extraction and vectorization; we propose using Python scikit-learn library to perform tokenization and feature extraction of text data, because this library contains useful tools like Count Vectorizer and Tfidf Vectorizer. Then, we will perform feature selection methods, to experiment and choose the best fit features to obtain the highest precision, according to confusion matrix results.

1. INTRODUCTION

Fake News contains misleading information that could be checked. This maintains lie about a certain statistic in a country or exaggerated cost of certain services for a country, which may arise unrest for some countries like in Arabic spring. There are organizations, like the House of Commons and the Crosscheck project, trying to deal with issues as confirming authors are accountable. However, their scope is so limited because they depend on human manual detection, in a globe with millions of articles either removed or being published every minute, this cannot be accountable or feasible manually. A solution could be, by the development of a system to provide a credible automated index scoring, or rating for credibility of different publishers, and news context.

This paper proposes a methodology to create a model that will detect if an article is authentic or fake based on its words, phrases, sources and titles, by applying supervised machine learning algorithms on an annotated (labelled) dataset, that are manually classified and guaranteed. Then, feature selection methods are applied to experiment and choose the best fit features to obtain the highest precision, according to confusion matrix results. We propose to create the model using different classification algorithms. The product model will test the unseen data, the results will be plotted, and accordingly, the product will be a model that detects and classifies fake articles and can be used and integrated with any system for future use.

2. RELATED WORK

2.1 Social Media and Fake News

Social media includes websites and programs that are devoted to forums, social websites, microblogging, social bookmarking and wikis [1][2]. On the other side, some researchers consider the fake news as a result of accidental issues such as educational shock or unwitting actions like what happened in Nepal Earthquake case [3][4]. In 2020, there was widespread fake news concerning health that had exposed global health at risk. The WHO released a warning during early February 2020 that the COVID-19 outbreak has caused massive ‘infodemic’, or a spurt of real and fake news—which included lots of misinformation.

2.2 Natural Language Processing

The main reason for utilizing Natural Language Processing is to consider one or more specializations of system or an algorithm. The Natural Language Processing (NLP) rating of an algorithmic system enables the combination of speech understanding and speech generation. In addition, it could be utilized to detect actions with various languages.[6] suggested a new ideal system for extraction actions from languages of English, Italian and Dutch speeches through utilizing various pipelines of various languages such as Emotion Analyzer and Detection, Named Entity Recognition (NER), Parts of Speech (POS) Taggers, Chunking, and Semantic Role Labeling made NLP good Subject of the search [5][6].

The Sentiment analysis [7] extracts emotions on a particular subject. Sentiment analysis is composed of extracting a specific term for a subject, extracting the sentiment, and pairing with connection analysis. The Sentiment analysis uses dual languages Resources for analysis: Glossary of meaning and Sentiment models database. for constructive and Destructive words and attempts to give classifications on a level of -5 to 5. Parts of speech taggers tools for languages such as European languages are being explored to produce parts of language taggers tools in different languages such as Sanskrit [8], Hindi [9] and Arabic. Can be efficient Mark and categorize words as names, adjectives, verbs, and so on. Most part of speech techniques can be performed effectively in European languages, but not in Asian or Arabic languages. Part of the Sanskrit word "speak" specifically uses the tree-bank method. The Arabic utilizes Vector Machine (SVM) [10] uses a method to automatically identify symbols and parts of speech and automatically expose basic sentences in Arabic text [11].

2.3 Data Mining

Data mining techniques are categorized into two main methods, which is; supervised and unsupervised. The supervised method utilizes the training information in order to foresee the hidden activities. Unsupervised Data Mining is a try to recognize hidden data models provided without providing training data for example, pairs of input labels and categories. A model example for unsupervised data mining is aggregate mines and a syndicate base [12].

2.4 Machine Learning (ML)

Classification Machine Learning (ML) is a class of algorithms that help software systems achieve more accurate results without having to reprogram them directly. Data scientists characterize changes or characteristics that the model needs to analyse and utilize to develop predictions. When the training is completed, the algorithm splits the learned levels into new data [11]. There are six algorithms that are adopted in this paper for classifying the fake news.

2.5 Decision Tree

The decision tree is an important tool that works based on flow chart like structure that is mainly used for classification problems. Each internal node of the decision tree specifies a condition or a “test” on an attribute and the branching is done on the basis of the test conditions and result. Finally, the leaf node bears a class label that is obtained after computing all attributes. The distance from the root to leaf represents the classification rule. The amazing thing is that it can work with category and dependent variable.

They are good in identifying the most important variables and they also depict the relation between the variables quite aptly. They are significant in creating new variables and features which is useful for data exploration and predicts the target variable quite efficiently.

Tree based learning algorithms are widely with predictive models using supervised learning methods to establish high accuracy. They are good in mapping non-linear relationships. They solve the classification or regression problems quite well and are also referred to as CART [13][14][15].

2.6 Random Forest Random

Forest are built on the concept of building many decision tree algorithms, after which the decision trees get a separate result. The results, which are predicted by large number of decision tree, are taken up by the random forest. To ensure a variation of the decision trees, the random forest randomly selects a subcategory of properties from each group [16][17] The applicability of Random forest is best when used on uncorrelated decision trees. If applied on similar trees, the overall result will be more or less similar to a single decision tree. Uncorrelated decision trees can be obtained by bootstrapping and feature randomness. Random Forest Pseudo-code To make n classifiers: For $i = 1$ to n do Sample the training data T randomly with replacement for T_i output Build a T_i -containing root node, N_i Call Build

2.7 Support Vector Machine (SVM)

The SVM algorithm is based on the layout of each data item in the form of a point in a range of dimensions n (the number of available properties), and the value of a given property is the number of specified coordinates [13]. Given a set of n features, SVM algorithm uses n dimensional space to plot the data item with the coordinates representing the value of each feature. The hyper-plane obtained to separate the two classes is used for classifying the data.

2.8. Naive Bayes

This algorithm works on Bayes theory under the assumption that it is free from predictors and is used in multiple machine learning problems [18]. Simply put, Naive Bayes assumes that one function in the category has nothing to do with another. For example, the fruit will be classified as an apple when its color is red, it swirls, and the diameter is close to 3 inches. Regardless of whether these functions depend on each other or on different functions, and even if these functions depend on each other or on other functions, Naive Bayes assumes that all these functions share a separate proof of the apples [14]

2.9. KNN (k- Nearest Neighbours)

KNN classifies new positions based on most of the sounds from the neighbouring k with respect to them. The position assigned in the class is highly mutually exclusive between the nearest neighbours K , as measured by the role of the distance [15]. KNN Pseudo-code Classify (X, Y, x) // X : training data, Y : class labels of X , x : unidentified sample
 For $i = 1$ to m do Calculate distance $d(X_i, x)$ end for
 Calculate set I containing indices for k smallest distances $d(X_i, x)$.
 return majority label for $\{Y_i \text{ where } i \in I\}$ [16]
 KNN falls in the category of supervised learning and its main applications are intrusion detection, pattern recognition. It is nonparametric, so no specific distribution is assigned to the data or any assumption is made about them. For example GMM, assumes a Gaussian distribution of the given data.

2.10 . Combining Classifiers

Achieving the best possible taxonomic performance is the primary goal when planning paradigm detecting systems. For that reason, different classification planners for the models of detecting actions are able to be progressed. Although if one model may perform the highest execution, the style sets correctly categorized by variant classifiers is not important to be overlap. Variant categorization planners can give additional information for the models. With this additional information, the execution of individual models can be improved [19]. 2.11 Related Work on Fake News Detection [20] pointed out various sources of media and made the suitable studies whether the submitted article is reliable or fake. The paper utilizes models based on speech characteristics and predictive models that do not fit with the other current models.

3. METHODOLOGY

This section presents the methodology used for the classification. Using this model, a tool is implemented for detecting the fake articles. In this method supervised machine learning is used for classifying the dataset. The first step in this classification problem is dataset collection phase, followed by pre-processing, implementing features selection, then perform the training and testing of dataset and finally running the classifiers [35][36][37][38][39]. Figure [1] describes the proposed system methodology. The methodology is based on conducting various experiments on dataset using the algorithms described in the previous section named Random forest, SVM and Naïve Bayes, majority voting and other classifiers. The experiments are conducted individually on each algorithm, and on combination among them for the purpose of best accuracy and precision [40][41][42].

Steps: -

- PREPARE DATASET
- DATA PRE-PROCESSING
- TRAIN MACHINE (LOGISTIC REGRESSION MODEL)
- TEST MACHINGE
- TRAINED LOGISTIC REGRESSION MODEL
- MAKING PREDICTIVE SYSTEM

4. RESULTS

The Scope of this project is to cover political news data. A dataset is prepared and then Machine is trained and used for textual fake and true news detection.

And Carried Out Results for the different News.

Our Dataset Includes:

1. ID - UNIQUE ID FOR NEWS ARTICLE
2. TITLE - TITLE OF THE NEWS ARTICLE
3. AUTHOR - NAME OF THE AUTHOR OF NEWS ARTICLE
4. TEXT - TEXT OF THE ARTICLE
5. LABEL - MARKS NEWS IS REAL OR FAKE

1: FAKE NEWS 0: REAL NEWS

Different Algorithms Used for Detection are:

- Decision Tree
- SVM
- KNN
- NAÏVE BAYES

Libraries Use are:

```
import numpy as np
import pandas as pd
import re
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```


5. CONCLUSION

The research in this paper focuses on detecting the fake news by reviewing it in two stages: characterization and disclosure. In the first stage, the basic concepts and principles of fake news are highlighted in social media. During the discovery stage, the current methods are reviewed for detection of fake news using different supervised learning algorithms. As for [20] the displayed fake news detection approaches that is based on text analysis in the paper utilizes models based on speech characteristics and predictive models that do not fit with the other current models. From [21] they utilized Naive Bayes classifier to detect fake news from different sources, with results of accuracy of 74%. [22]

Used combined ML algorithms, but they depend on unreliable probability threshold with 85-91% accuracy. [23] uses the Naive Bayes to detect fake news from different social media websites, but the results were not accurate for the untruthful sources. [24] They got their data from Kaggle with average accuracy of 74.5%. [27] Used Naive Bayes algorithms to detect Twitter spam senders, with accuracy rated from 70% to 71.2%. [28] They tried different approaches with accuracy of 76%. [29]

Three common methods are utilized through their researches: Naïve Bayes, Neural Network and Support Vector Machine (SVM). Naïve Bayes has an accuracy of 96.08% for detecting fake messages. The neural network and the machine vector (SVM) reached an accuracy of 99.9 0%. [30] They used the combination of KNN and random forests that gave the final results improved by up to 8% using a mixed false message detection model. [31] They worked on 2012 Dutch elections fake news on Twitter, they examine the execution of 8 supervised machine learning classifiers in the Twitter dataset. And they assume that the decision tree algorithm works best for the data set used with a F score of 88%. [32] Presented a counterfeit detection model using N-gram analysis achieved the highest accuracy in use contains a unigram and a linear SVM workbook. The highest accuracy is 92%.


In the aforementioned research summary and system analysis, we concluded that most of the research papers used naïve bays algorithm, and the prediction precision was between 70-76%, they mostly use qualitative analysis depending on sentiment analysis, titles, word frequency repetition [40][41][42]. In our approach we propose to add to these methodologies, another aspect, which is POS textual analysis, it is a quantitative approach, it depends on adding numeric statistical values as features, we thought that increasing these features and using random forest will give further improvements to precession results. The features we propose to add in our dataset are total words (tokens), Total unique words (types), Type/Token Ratio (TTR), Number of sentences, Average sentence length (ASL), Number of characters, Average word length (AWL), nouns, prepositions, adjectives etc.

6. SCREENSHOTS

Jupyter FAKE NEWS DETECTION PROJECT Last Checkpoint: an hour ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

FAKE NEWS DETECTION



ABOUT THE DATASET

1. ID - UNIQUE ID FOR NEWS ARTICLE
2. TITLE - TITLE OF THE NEWS ARTICLE
3. AUTHOR - NAME OF THE AUTHOR OF NEWS ARTICLE
4. TEXT - TEXT OF THE ARTICLE
5. LABEL - MARKS NEWS IS REAL OR FAKE

1: FAKE NEWS 0: REAL NEWS

Jupyter FAKE NEWS DETECTION PROJECT Last Checkpoint: an hour ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

IMPORTING REQUIRED LIBRARIES

```
In [2]: import numpy as np
import pandas as pd
import re
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

```
In [3]: import nltk
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

Out[3]: True

```
In [4]: # printing the stopwords in English
print(stopwords.words('english'))
```

['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'a bout', 'against', 'between', 'into', 'through', 'during', 'before', 'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under', 'again', 'further', 'then', 'onc e', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'too', 'ver y', 's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "should've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', 'aren't', 'couldn', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn', "mustn't", 'needn', "needn't", 'shan', "shan't", 'shouldn', "shouldn't", 'was n', "wasn't", 'weren', "weren't", 'won', "won't", 'wouldn', "wouldn't"]

DATA PREPROCESSING

INTERSTING TEST AND TRAIN DATA SET

```
In [5]: df_test = pd.read_csv("test.csv")
df_train = pd.read_csv("train.csv")
```

```
In [6]: # loading the dataset to a pandas DataFrame
news_dataset = pd.read_csv('C:/Users/HP/Downloads/Fake News Detection INT 247 Tanish Verma 11907877/fake-news-prediction/train.csv')
```

```
In [7]: news_dataset.shape
```

```
Out[7]: (20800, 5)
```


```
In [8]: # print the first 5 rows of the dataframe
df_train.head(5)
```

```
Out[8]:
```

	id	title	author	text	label
0	0	House Dem Aide: We Didn't Even See Comey's Let...	Darrell Lucus	House Dem Aide: We Didn't Even See Comey's Let...	1
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Daniel J. Flynn	Ever get the feeling your life circles the rou...	0
2	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, ...	1
3	3	15 Civilians Killed In Single US Airstrike Hav...	Jessica Purkiss	Videos 15 Civilians Killed In Single US Aistr...	1
4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print 'nAn Iranian woman has been sentenced to...	1

```
In [9]: # counting the number of missing values in the dataset
news_dataset.isnull().sum()
```

```
Out[9]: id      0
title    558
author   1957
text     39
label     0
```

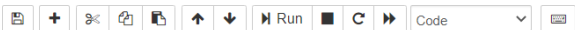
 jupyter FAKE NEWS DETECTION PROJECT Last Checkpoint: an hour ago (autosaved)

 Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted

Python 3



Training the Model: Logistic Regression

```
In [26]: model = LogisticRegression()
```

```
In [27]: model.fit(X_train, Y_train)
```

```
Out[27]: LogisticRegression()
```

Evaluation

accuracy score

```
In [28]: # accuracy score on the training data
X_train_prediction = model.predict(X_train)
training_data_accuracy = accuracy_score(X_train_prediction, Y_train)
```



```
In [29]: print('Accuracy score of the training data : ', training_data_accuracy)
```

```
Accuracy score of the training data : 0.9865985576923076
```

```
In [30]: # accuracy score on the test data
X_test_prediction = model.predict(X_test)
test_data_accuracy = accuracy_score(X_test_prediction, Y_test)
```

```
In [31]: print('Accuracy score of the test data : ', test_data_accuracy)
```

```
Accuracy score of the test data : 0.9790865384615385
```

 **jupyter** FAKE NEWS DETECTION PROJECT Last Checkpoint: an hour ago (autosaved)  Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

Making a Predictive System

```
In [32]: X_new = X_test[3]

prediction = model.predict(X_new)
print(prediction)

if (prediction[0]!=0):
    print('The news is Real')
else:
    print('The news is Fake')

[0]
The news is Real
```

```
In [33]: print(Y_test[3])

0
```

```
In [38]: X_new = X_test[2]

prediction = model.predict(X_new)
print(prediction)

if (prediction[0]!=0):
    print('The news is Real')
else:
    print('The news is Fake')

[1]
The news is Fake
```

```
In [39]: print(Y_test[2])

1
```

7. 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. ASCI-2020 IOP Conf. Series: Materials Science and Engineering1099 (2021) 012040 IOP Publishing doi:10.1088/1757-899X/1099/1/012040 12
- [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>. 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). Semantic scholar. 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]. NaphapornSirikulviriyaa; SukreeSinthupinyo. “Integration of Rules from a Random Forest.” International Conference on Information and Electronics Engineering (p. 194 : 198). Singapore: semantic scholar.org. 2011.

- [18]. Jasmin Kevric et al. "An effective combining classifier approach using tree algorithms for network intrusion detection." *Neural Computing and Applications* , 1051–1058. 2017.
- [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]. Supanya Aphiwongsophon 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. [40]. K Ludwig, M Creation 2020 "Dissemination and uptake of fake-quotes in lay political discourse on Facebook and Twitter" *J. Pragmat*, 157, 101–118.
- [41]. Can Machines Learn to Detect Fake News? A Survey Focused on Social Media. Available at: <https://scholarspace.manoa.hawaii.edu/handle/10125/59713>
- [42]. Cardoso Durier da Silva, F., Vieira, R., & Garcia, A. C. (2019, January). Can machines learn to detect fake news? a survey focused on social media. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*.
- [43]. Bovet, Alexandre, and Hernán A. Makse. "Influence of fake news in Twitter during the 2016 US presidential election." *Nature communications* 10.1 (2019): 1-14. The science of fake news.
- [44]. <https://science.sciencemag.org/content/359/6380/1094.summary> Science 09 Mar 2018:Vol. 359, Issue 6380, pp. 1094-1096 DOI: 10.1126/science.aao2998.

