Shri B V V Sangha's

BASAVESHWAR ENGINEERING COLLEGE(AUTONOMOUS), BAGALKOT- 587102

2021-2022

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

7th Semester

PROJECT PHASE-I REPORT (UIS733P)

DESIGN AND DEVELOPMENT OF CONTEXT AWARE FAKE NEWS DETECTION SYSTEM USING MACHINE LEARNNG LEARNING

Project Guide:

Dr. Lokesh B. Bhajantri

Head of Department:

Dr. Shobha R. Patil

Submitted By

PROJECT ASSOCIATES:

Sl. No	Name of Student	USN
01	Madhuri M Manganahalli	2BA18IS030
02	Madhu Muttal	2BA18IS029
03	Karthik M Manganahalli	2BA18IS025
04	Deepak Kamatad	2BA18IS022

Shri B. V. V. Sangha's

BASAVESHWAR ENGINEERING COLLEGE (AUTONOMOUS),

BAGALKOT-587102

2021-2022

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that Project work entitled "DESIGN AND DEVELOPMENT OF CONTEXT AWARE FAKE NEWS DETECTION SYSTEM USING MACHINE LEARNING" is a bonafide work carried out by "Miss. Madhuri M Manganahalli, Miss. Madhu Muttal, Mr. Karthik M Manganahalli, Mr. Deepak Basavaraj Kamatad" of Department of Information Science and Engineering, Basaveshwar Engineering College (Autonomous), Bagalkot affiliated to VTU Belagavi during the academic year 2021-2022. The Project report has been approved as it satisfies the academic requirements with respect to project work prescribed for 7th semester.

PROJECT GUIDE	Head of Department
Dr. Lokesh B. Bhajantri	Dr. Shobha R. Patil

Project Associates:

Sl. No	Name of Student	USN
01	Madhuri M Manganahalli	2BA18IS030
02	Madhu Muttal	2BA18IS029
03	Karthik M Manganahalli	2BA18IS025
04	Deepak Kamatad	2BA18IS022

Internal Examiner   External Examiner

ACKNOWLEDGEMENT

We believe that our project will be complete only after we thank the people who have contributed to make this project successful.

We are thankful to **Dr. S. S. Injaganeri**, Principal, Basaveshwar Engineering College. Bagalkot, for providing firm support and necessary facilities in the college.

The very outset words are inadequate to express a deep sense of gratitude to **Dr. Shobha R. Patil**, Head of Department of Information Science and Engineering Department for her constant affectionate guidance and her acumen throughout the course of this study.

We wish to express sincere thanks to our Guide **Dr. Lokesh B. Bhajantri** Associate Professor, Department of Information Science and Engineering, Basaveshwar Engineering College, Bagalkot, for guiding us from time to time.

We also thank the teaching and non-teaching staff of Department of Information Science and Engineering for their timely co-operation and much needed assistance.

Finally, we feel to express our sincere thanks to our parents, friends and to all those who were directly or indirectly involved, and who rendered valuable and dedicated advice for successful completion of this PROJECT PHASE – I.

<div align="right">

Your Sincerely

Madhuri M M

Madhu Muttal

Karthik M M

Deepak B K

</div>

Abstract

Most of the smart phone users prefer to read the news via social media over internet. The question is how to authenticate the news and articles which are circulated among social media like WhatsApp groups, Facebook Pages, Twitter and other micro blogs & social networking sites. It is harmful for the society to believe on the rumours and pretend to be a news. The need of the proposed system is to stop the rumours especially in the developing countries like India, and focus on the correct, authenticated news articles. This paper demonstrates a model and the methodology for fake news detection. With the help of Machine learning, it is tried to authenticate the news and later determine whether the news is real or fake using Naive Bayes Algorithm.

List of Contents

List of Figures

Chapter 1

Introduction

With the onset of social media, for example Facebook, Twitter and Instagram, the role of print media for example. Newspaper and electronic media for example. Television, Radio News channels in communicating has reduced. The growth of social media platforms has played an important role in this transformation [1]. Anyone can post content over the web. Unfortunately, false news has gained a lot of attention over the web, particularly via social networking sites. Individuals get swayed and don't rethink before sending such misinformation out to other people. The same type of activities is not good for society where some rumours or vague news extinguish all negative feelings between people or a specific group of people. In the rapidly evolving world of technology, preventative measures are required to counteract these types of activities. Broadcast communications are affecting the general population on a massive scale, so a few people attempt to take advantage of it. There are plenty of websites that present false data to the public. They deliberately try to bring out purposeful publicity, deceptions, and falsehood in the name of news. Their role is to manipulate the information that can make people have confidence in it. There are a lot of such sites in numerous places around the world. Therefore, fabricated news affects the brain of an individual. According to scientists, man-made calculations can help in spotting fake news. [2]

We use fake news detection system to stop rumours being spread through various media platforms, whether it is social media or messaging. This is done to prevent spreading fake news that leads to events such as mob lynching. This fake news detection is aimed at detecting fake news and stopping activities like these, thereby protecting society from unwanted acts of violence.[3]

The detection of fake news in a social media is important in today's society. As fresh news content is rapidly being produced as a result of the abundance of available technology. In day to day life over 600 vedios are uploaded on YouTube, 600000 status updates and posts are created,98000 tweets are shared on tweeter,11 million Instant messages are sent,1500 Blogs entries are created and 160 million mail are sent over every 60 seconds on online. All these cannot be true all the time. So, it is important to detect the fake news.
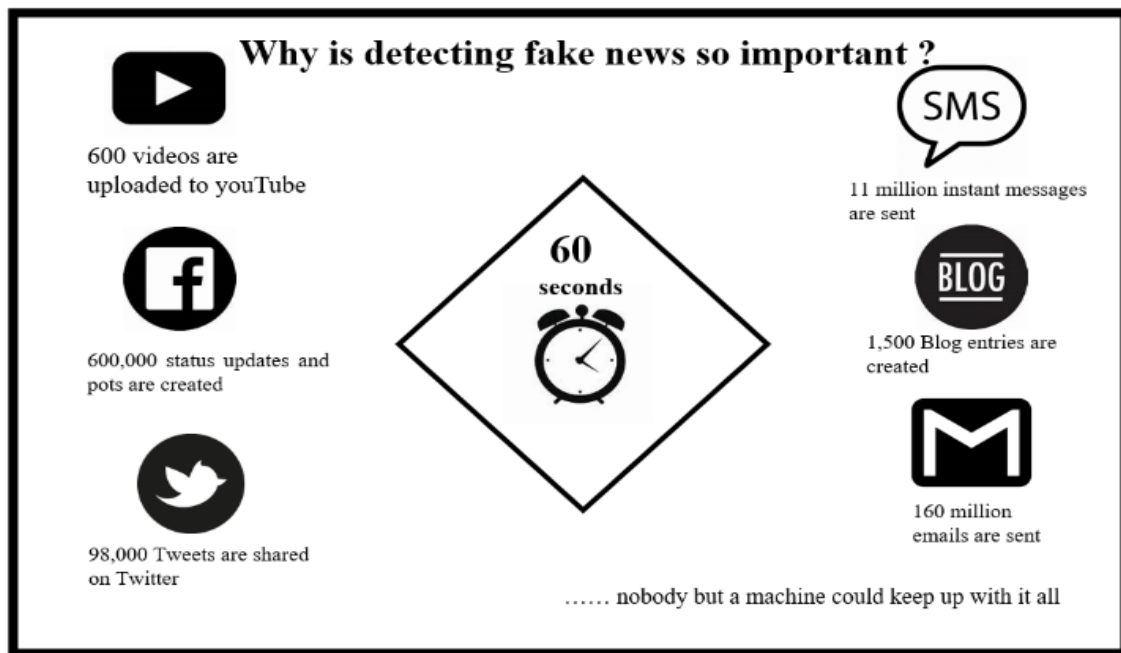
Figure 1.1: Importance of Fake News Detection.

1.1 Objectives of the Project

The objectives of Fake News Detection System are: -

- To create system that can use the data of past news reports and predict the news report being fake or real.
- To find the authenticity of the news in the social media.
- To analyse and design an algorithm for detecting fake news from the social media network.
- To provide accurate data quickly as compared to other services.

1.2 Scope of the Project

The Scope of the Fake News Detection System are: -

- This project is to control the flow of fake news or information in the social media like Instagram, Facebook and WhatsApp, etc.
- It provides security for individuals and controls criminal activities.

Chapter 2

Literature Survey

Some of the related works are as follows.

The work given in [4] presents a detection of fake news using machine learning. It demonstrates a model and the methodology for fake news detection. With help of machine learning and natural programming language, it is tried to aggregate the news and later determine whether the news is fake or real using Support Vector Machine. This system does not have complete accuracy and it has low efficiency. Analysis of classifiers for fake news detection is presented in [5]. This paper discusses the approach of natural language processing and machine learning in order to solve this problem. Use of bag-of-words, n-grams, count vectorizer has been made, TF-IDF, and trained the data on five classifiers to investigate which of them works well for this specific dataset of labelled news statements. The data is erratic and this means that any type of prediction model can have anomalies and can make mistakes.

Fake news and Message Detection is presented in [6]. This paper aims to perform binary classification of various news articles available online with help of concepts pertaining to Artificial Intelligence, Natural Programming Language and Machine Learning, democracies. It is less accurate than our system.A Survey on Natural Language Processing for Fake News Detection is presented in [7]. This paper describes the challenges involved in fake news detection and also describe related tasks and systematically reviews and compare the task formulations, datasets and NLP solutions that have been developed for this task, and also discuss the potentials and limitations of them. It compares empirical results on classification datasets via various machine learning models in this section. It focuses on three datasets: Liar, Fever and Fake news net. It is because others have limited size than newer datasets, limited numbers of experiments, or have the aspect of rumor detection much more than fake news detection.

Smart System for Fake news Detection Using Machine Learning is presented in [8]. This paper develops a method for fake news detection using machine learning on Twitter by learning to predict accuracy assessments in two credibility-focused Twitter datasets. Apply this method to twitter content sourced from Buzz Feed's fake news dataset and show models trained against crowdsourced workers outperform models based on journalists. use of machine learning techniques to detect Fake news, using Support Vector Machine (SVM). SVM algorithm is less efficient than Naive Bayes algorithm. Fake News Detection on social media is presented in [9].In this paper Fake detection is done by using deep learning and Machine leaning. Here the content filters the originality to give the correct information for the user. They have used SVM algorithm but it is not suitable for large Dataset.

Fake News Detection using Stance Classification is presented in [10]. The main Objective of this paper has been to survey stance classification and fake news detection. And they investigated the process of data gathering and annotation, and how imbalanced data can have significant impact of results obtained in stance classification. Lower shorter training time, High risk of over-fitting than deterministic algorithm.Design Exploration of Fake News: A Transdisciplinary Methodological Approach to Understanding Content Sharing and Trust on social media is presented in [11]. In this paper, Publishers recommended a transdisciplinary transfer in the direction of researching fake news that takes into account algorithmic tactics, psychometric data, and qualitative explorations. The shortcoming of the experiment is that it requires both labelled datasets and communities of experts to help train applications to recognize and categorize contents. So, work is still essential to recognize how human judgement of fake news takes place.

Fake and Spam Messages: Detecting Misinformation during Natural Disasters on social media is presented in [12]. In this paper, Publishers conducted a case study of 2013 Moore Tornado and Hurricane Sandy. Pilot grades showed that the projected tactics classify spam and fake messages with 96.43% accuracy and 0.961 F-measure. One mislaid study is when would be a right prompt to develop fake and spam tweet forecaster while streaming data is looming. Classifying Fake News Articles Using Natural Language Processing to Identify In-Article Attribution as a Supervised Learning Estimator is presented in [13]. In this paper, the research process, methodical analysis, technical semantics work, and classifier performance and results are offered. The paper concludes with a discourse of how the current system will advance into an impact mining system. The attribution-based fake news discovery tool that uses the quote ascription classifier, however, like the attribution classifier, it did not perform well enough for production use. The overall act results for this system are not as robust as desired.

Credibility Assessment of Textual Claims on the Web is presented in [14]. This paper recommends the use of a method lever-ages the joint communication between the language of articles about the claim and the reliability of the basic web sources. Experiments with claims from the popular website snopes.com and from reported cases of Wikipedia hoaxes prove the viability of the projected methods and their superior accuracy over various baselines. Can't examine the role of attribution or speaker information, refined linguistic aspects like denial, and understanding the article's view about the claim. Message Authentication System for Mobile Messaging Applications is presented in [15]. This paper proposes a system enabling users to confirm the authenticity of messages recognised through message sharing applications. The proposed system works as a third-party data authentication service capable of working with a wide-variety of message sharing requests or social network platforms which involve large-scale dataset.

Chapter 3

Proposed Work

3.1 Problem Statement

- Social media for news consumption is a double-edged sword.

- On one hand, its low cost, easy access, and rapid dissemination of information lead people to seek out and consume news from social media.

- On the other hand, it enables the wide spread of "fake news" i.e., low quality news with intentionally false information. The extensive spread of fake news has the potential for extremely negative impacts on individuals and society.

- Therefore, fake news detection on social media has recently become emerging research that is attracting tremendous attention.

- Fake news detection on social media presents unique characteristics and challenges that make existing detection nontrivial to detect based on news content; therefore, we need to include auxiliary information, such as user social engagements on social media, to help make a determination.

3.2 Proposed System Architecture

It comprises the structure and behavior of the system. In this project, the proposed architecture comprises the following steps:

1) To collect the data from the environment,

2) to create the dataset for detecting data using various sources,

3) to pre-processing of dataset,

4) to perform feature extraction,

5) to perform training and testing the dataset using machine learning algorithm,

6) finally, to detect the data, whether fake or real to map with the various sources like websites, newspapers, social media, etc., using proposed machine learning algorithm.
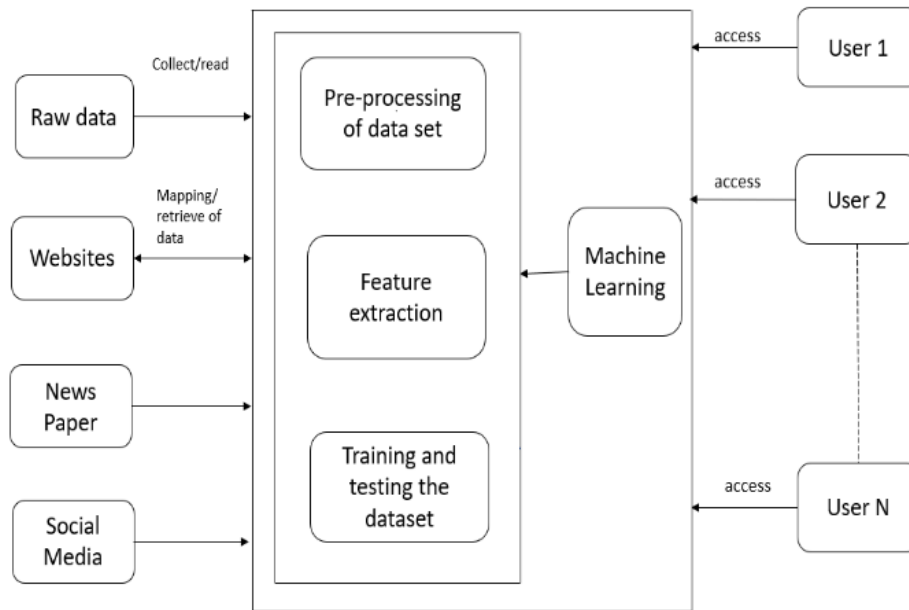
Figure 5.1: System Architecture of Proposed System

The user enters the required news or information in the website, the proposed system is used to check whether the data entered by the user is real or fake using context aware Naïve Bayes algorithm. The administrator extracted the data from different sources for mapping with the raw data to detect the data. The mapped data is stored in a database for training and processing. Those processed data are organized based on their behavior in a dataset. The data can be extracted based on the features and based on the user requirements.

3.3 Methodology

A Naive Bayes classifier is a supervised machine learning algorithm that uses Bayes" theorem. The variables that are used to generate the model are independent of each other. It is proven that this classifier itself provides pretty good results.[16]

$$P(X| C_i) = \prod P (x_k|C_i)$$
$$= P (x_1| C_i) \times P (x_2 |C_i) \times \ldots \times P (x_n| C_i)$$

The classification is conducted by deriving the maximum posterior, which is the maximal P (C_i| X) with the above assumption applying to Bayes theorem. This assumption greatly reduces the computational cost by only counting the class distribution. Naive Bayes is popular algorithm which is used to find the accuracy of the news whether it's real or fake using multinomial Naïve Bayes. There are number of algorithms that focus on common principle, so it is not the only algorithm for training such classifiers. To check if the news is fake or real naive Bayes can be used.

3.4 Use Case Diagram
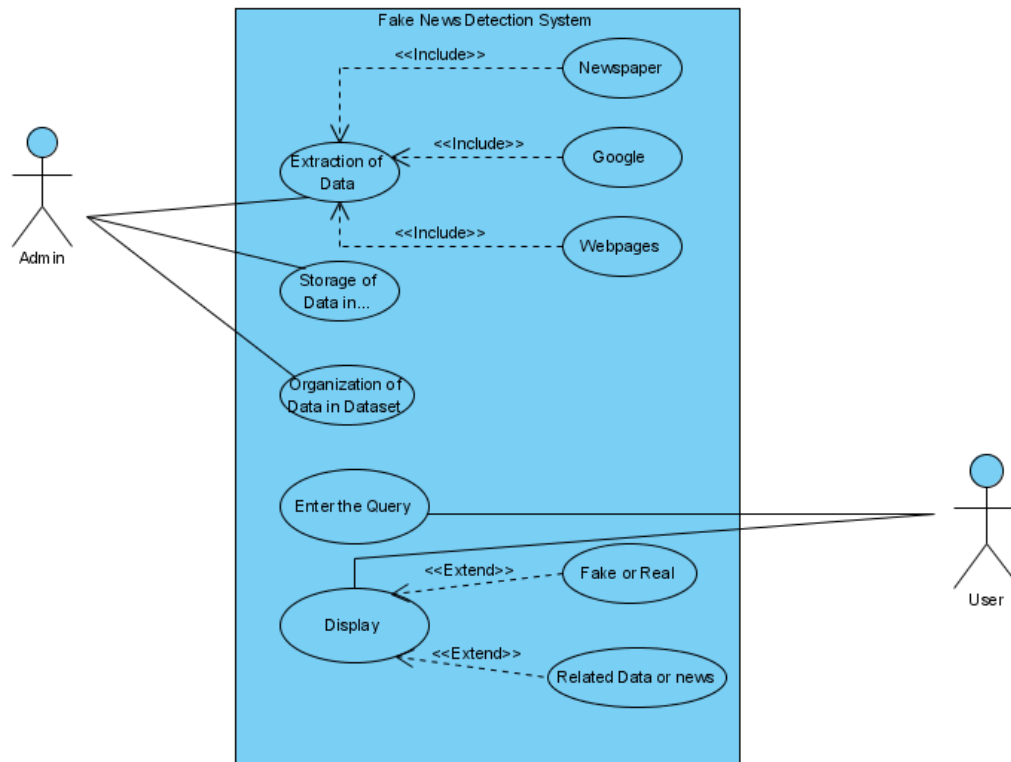
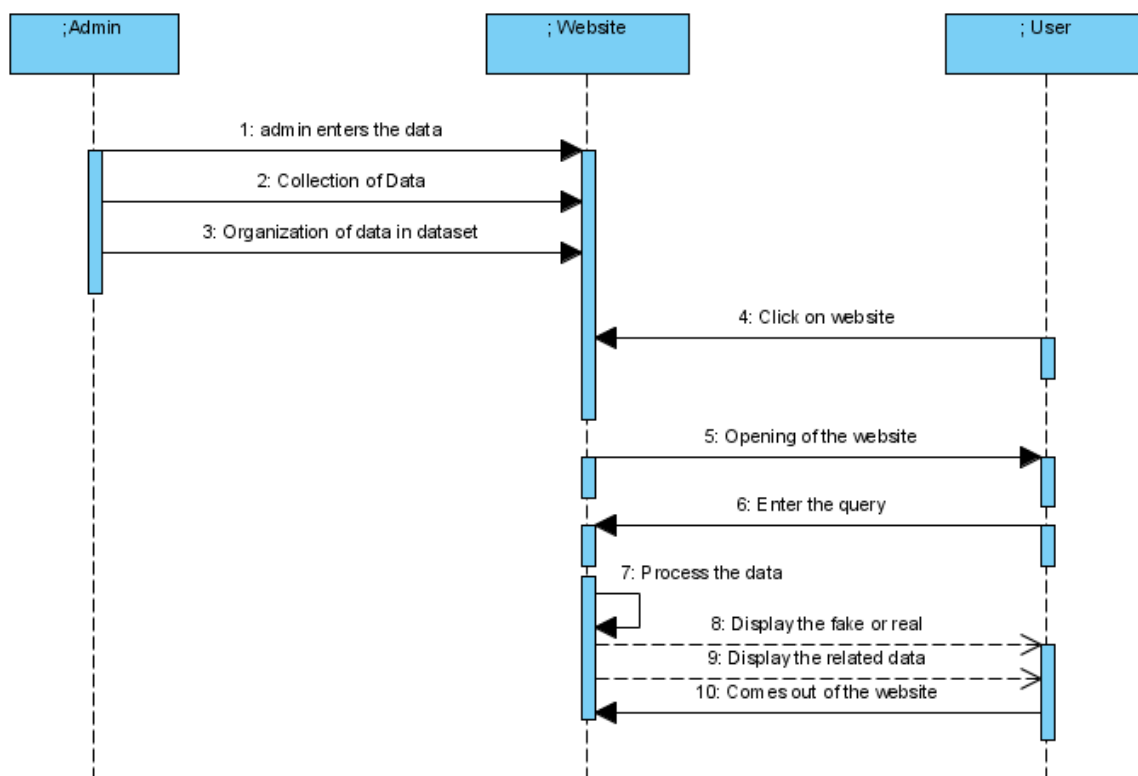**Figure 5.2: Use Case Diagram of Proposed System**

The usecase diagram of proposed system is as shown in the Figure 2. The admin and the user are considered as actors in the usecase model. The following are the usecases of the proposed system – Login, Signup, Data enter, Extraction of data, Create a dataset, Organizationof the dataset, and Display. Inwhich, the admin does the extraction of data and stores the data in the database and extraction of data is done from different sourses like News papers, Google, Web Pages, social media networks, etc., when the user enters the information to be searched in search box the data gets processed using Navie Bayes algorithm and the result is displayed that is Fake or Real, and the related data is displayed.

3.5 Sequence Diagram

The interactions of objects of proposed system is as shown in Figure 3. Here Admin,Website, and user are considered as objects, each object has associates with lifelines. It consists following steps are:

1. **Admin enters the data –** The adminstrator enters the data for identification.
2. **Collection of data:-** The admin collects the data from different sources like Newspaper,google, social media, webpages, etc.

3. **Stores the data in database:-**After collecting the data, the data is stored in the database to create a datasets.

4. **Process of data –** The process of dataset to find the data is fake or real using proposed machine learning algorithm. The proposed system is used to perform training and testing of datasets.

5. **Display fake or real:-**After processing the data, the result will be displayed.

6. **Display the related data:-**The result is displayed along with the related data.

7. **Comes out of the website:-** After completing the above steps, user comes out of the website.

Figure 5.3: Sequence Diagram of Proposed System

3.6 Algorithm:

Step 1: Click on the website

Step 2: Website Opens

Step 3: User enter the Information that is to checked in search engine.

Step 4: Data get processed

Step 5: The result is displayed i.e., real or fake.

Step 6: if Information is fake, then related Correct Information is displayed

Chapter 4

Requirements

4.1 Software Requirements:

- Operating System (Windows, Linux, Mac).

- Back End Tool: MySQL oracle 10g.

- Front End Tool: Python, Natural Programming Language (NPL).

4.2 Hardware Requirements:

- Processor Core: Intel i5 or i3.

- Processor Speed: 1.0 GHz or above

- Hard Disk: 20 GB or above.

- RAM: 1 GB or above

Chapter 5

Conclusion and Future Work

In this paper, we presented the new architecture of fake news detection system based on Naïve bayes. Proposed System illustrates about how fake news are detected. The system collects the information from various sources of social media,news paper and from different website.When user enters the query or the information which they what to know whether it is real or fake,the entered data is processed in a system by using the Navie Bayas algorithm where it compares the enterd data with data in the system stored.Based on the accuracy it displays fake or real and display the related data.

If we get an opportunity to enhance this project, we should like to detect the fake news from videos, images and audios.

References

[1] Kelly Stahl "Fake News detection in social media" , published by california State University Stanislaus at csustan.edu in 2018.

[2] Harsh Khatter "A smart System for Fake News Detection Using Machine learning" in 2019 from www.researchgate.net/publication/339022255.

[3]Anjali Jain, Amey Kasbe "Detection of fake news Using Machine Learning" in 2018 from https://ieeexplore.ieee.org/document/8977659/;jsessionid=Jg0cEABOEDbZ8XqfqJg8TeZ3cWh8FY_sCQ-nokpuchURWbpnFEeb!142814072.

[4] Vasu Agrwal, H. Praveen, Sultan, Shreejan Malhotra "Analysis of classifiers for fake news detection" in 2019 from https://www.researchgate.net/publication/339548191_ Analysis_of_ Classifiers_for_Fake_News_Detection.

[5] Lokesh Parab "Fake News and Message Detection" in 2021 from https://www.researchgate.net/publication/355586673_Fake_News_and_Message_Detection_project _report.

[6] Ray Oshikawa, Jing Qian, William Yang Wang " A survey on Natural Language Processing fir Fake News Detection"in 2018 from https://scholar.google.co.in/scholar?q=ray+oshikawa+a+survey+on+natural+language&hl=en&as_sdt=0&as_vis=1&oi=scholartl.

[7]Prof B.J Deokate "Smart System for Fake News Detection Using Machine Learning" in 2017 from https://www.researchgate.net/publication/339022255_A_smart_System_for_Fake_News_Detection _Using_Machine_Learning

[8]Xinyi Zhou, Reza Zafarani "A survey of Fake News" in https://in.docworkspace.com/d/s IDDi39KVAZ2n5I8G.

[9] Steni Moi TS, Sreeja P S "Fake News Detection on social Media" in https://www.researchgate.net /institution/Hindustan-University.

[10] Anders E. Lillie, Emil R. Middelboe "Fake News Detection using Stance Classification" inhttps://www.researchgate.net/publication/342762209_A_Survey_of_Fake_News_Fundamental_Theories_Detection_Methods_and_Opportunities

[11] Jaigris Hodson, Brian Traynor, "Design Exploration of Fake News: A Transdisciplinary Methodological Approach to Understanding Content Sharing and Trust on Social Media" 2018 IEEE International Professional Communication Conference.

[12] Terry Traylor, Jeremy Straub, Gurmeet and Nicholas Snell, "Classifying Fake News Articles Using Natural Language Processing to Identify In-Article Attribution as a Supervised Learning Estimator," in IEEE (ICSC),2019.

[13]Terry Traylor, Jeremy Straub, Gurmeet and Nicholas Snell, "Classifying Fake News Articles Using Natural Language Processing to Identify In-Article Attribution as a Supervised Learning Estimator," in IEEE (ICSC),2019.

[14]Terry Traylor, Jeremy Straub, Gurmeet and Nicholas Snell, "Classifying Fake News Articles Using Natural Language Processing to Identify In-Article Attribution as a Supervised Learning Estimator," in IEEE (ICSC),2019.

[15] Kashyap Popat, Subhabrata Mukherjee, Jannik Strotgen and Gerhard Weikum, "Credibility Assessment of Textual Claims on the Web" in ACM,2016.

[16] Harsh Khatter "A smart System for Fake News DetectionUsing Machine learning" in 2019 from www.researchgate.net/publication/339022255.