Fake News Detection Using Machine learning

A MINI PROJECT REPORT

Submitted by

ATTE SREEJA (21TK1A0510)

In partial fulfillment for the award of the degree of

BACHELORS OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

Under the esteemed guidance of

Mr.K .Yakub Reddy

Asst.Professor, Department of CSE

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

SVS GROUP OF INSTITUTIONS

(Approved by AICTE, New Delhi & Affiliated to JNTU,Hyderabad & Accredited by NAAC)

BHEEMARAM(V),Hanamkonda T.S. India - 506015

JAWAHARALAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

JANUARY 2025

SVS GROUP OF INSTITUTIONS

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that **ATTE SREEJA** HT.No. **21TK1A0510** of the B.Tech has satisfactorily completed the dissertation work entitled **"FAKE NEWS DETECTION USING MACHINE LEARNING"** in the partial fulfillment of the requirements of the B. Tech degree during this academic year **2024-2025.**

INTERNAL	**HOD/CSE**	**PRINCIPAL**
Mr. K. YAKUB REDDY	Dr.V MADHAVI	Dr. R RAMESH
Asst.Professor	Asso.Professor	Dept of Mech&AIML
Internal Guide	**Head of the Department**	**SVSGOI**
Dept of CSE	Dept of CSE	BHEEMARAM

External Examiner

DEPARTMENT OF COMPUTER SCIENCE&ENGINEERING

SVS INSTITUTE OF TECHNOLOGY

BHEEMARAM ,HANAMKONDA- 506015

DECLARATION BY THE CANDIDATE

I **ATTE SREEJA** bearing the H.T No. **21TK1A0510** to hereby certify that the Project Report entitled **"FAKE NEWS DETECTION USING MACHINE LEARNING"**, under the guidance of **Mr. K. Yakub Reddy , Prof.** Department of CSE, SVS Institute of Technology & Science, Hanamkonda, submitted in partial fulfillment of the requirements for the Award of the Degree of Bachelors of Technology in Computer Science & Engineering.

This is a record of bonafide work carried out by me and the results embodied in this Project Report has not been produced / copied from any source. The results embodied in this Project Report have not been submitted to any other University or Institute for the Award of any other Degree or Diploma.

PLACE:
DATE:

ATTE SREEJA

DEPARTMENT OF COMPUTER SCIENCE&ENGINEERING

SVS INSTITUTE OF TECHNOLOGY

BHEEMARAM , HANAMKONDA- 506015 TS

CERTIFICATE FROM THE HEAD OF THE DEPARTMENT

This is to certify that the Project Report entitled **"FAKE NEWS DETECTION USING MACHINE LEARNING"** being submitted by **ATTE SREEJA** bearing the H.T No.**21TK1A0510** in partial fulfillment of the requirements for the Award of the Degree of the Bachelors of Technology in Computer Science & Engineering is a record of bonafide work carried out by him under my guidance.

The results of investigation enclosed in the report have been verified and found satisfactory. The results embodied in this thesis have not been submitted to any other University for the award of any degree or diploma.

Place:
Date:

Dr. V MADHAVI
Asso Professor &HOD

ACKNOWLEDGMENT

The completion of this project work gives me an opportunity to convey my gratitude to all those who helped me to complete the project successfully.

First, I gratefully acknowledge my deep sense of gratitude to Almighty for spiritual Guidance blessings shown to complete the Project. I thank my parents for unconditional support to improve myself throughout my life.

I owe my respectable thanks to Dr.R RAMESH, Principal of SVS Group of Institutions, for providing all necessary facilities and encouraging words for completion of this project.

I gratefully acknowledge Mr.Dr.V MADHAVI, Associate Professor, Head of the Department, Computer Science and Engineering, for his encouragement and advice during the this project work.

My sincere thanks to Mr.K YAKUB REDDY ,Assistant Professor, Guide SVS Group of Institutions, for continuous support for doing this project.I would like to express my thanks to all the faculty members of Department of Computer Science and Engineering,and Non-teaching staff of SVS Group of Institutions who have rendered valuable help in making this project successful.

ATTE SREEJA

LIST OF FIGURES

ABSTRACT

Fake News has become one of the major problems in the existing society. Fake News has high potential to change opinions, facts and can be the most dangerous weapon in influencing society. The proposed project uses NLP techniques for detecting the 'fake news', that is, misleading news stories which come from non-reputable sources. By building a model based on a K-Means clustering algorithm, the fake news can be detected. The data science community has responded by taking actions against the problem. It is impossible to determine whether the news was real or fake accurately. So, the proposed project uses the datasets that are trained using the count vectorizer method for the detection of fake news and its accuracy will be tested using machine learning algorithms. In this research, we concentrate on how to spot fake news in internet news sources. We are dedicated in two ways. In order to determine the percentage of correct news that is phony, we will use multiple datasets of actual and fake news. We provide a thorough description of the selection, justification, and approval process as well as a few exploratory analyses on the observable evidence of etymological differences in false and legitimate news material. In order to create precise false news identifiers, we focus a lot of learning studies. Additionally, we provide close examinations of the automatic and manual evidence of bogus news. Python can be used to spot fake news posted on social media

CHAPTER 1

INTRODUCTION

World is changing rapidly. No doubt we have a number of advantages of this digital world but it also has its disadvantages as well. There are different issues in this digital world. One of them is fake news. Someone can easily spread a fake news. Fake news is spread to harm the reputation of a person or an organization. It can be a propaganda against someone that can be a political party or an organization. There are different online platforms where the person can spread the fake news. This includes the Facebook, Twitter etc. Machine learning is the part of artificial intelligence that helps in making the systems that can learn and perform different actions. A variety of machine learning algorithms are available that include the supervised, unsupervised, reinforcement machine learning algorithms. The algorithms first have to be trained with a data set called train data set. After the training, these algorithms can be used to perform different tasks. Machine learning is using in different sectors to perform different tasks. Most of the time machine learning algorithms are used for prediction purpose or to detect something that is hidden**.**

Fake News Detection is a natural language processing task that involves identifying and classifying news articles or other types of text as real or fake. The goal of fake news detection is to develop algorithms that can automatically identify and flag fake news articles, which can be used to combat misinformation and promote the dissemination of accurate information.Fake news detection refers to the process of identifying and classifying online information as either accurate or false using computational methods, aiming to combat the widespread dissemination of misleading content that can have significant social and political impacts; this is achieved by analyzing textual features like language patterns, source credibility, and contextual cues within news articles to automatically flag potentially fake news items, often utilizing machine learning algorithms to train models on large datasets of labeled news content. The accuracy of the detection achieved by the system is around 70%. This text describes an easy fake news detection method supported one among the synthetic intelligence algorithms – Naive Bayes classifier, Random Forest and Logistic Regression.

Fig: FAKE NEWS

MOTIVATION

We will be training and testing the data, when we use supervised learning it means we are labeling the data. By getting the testing and training data and labels we can perform different machine learning algorithms but before performing the predictions and accuracy, the data is need to be pre processing i.e. the null values which are not readable are required to be removed from the data set and the data is required to be converted into vectors by normalizing and tokening the data so that it could be understood by the machine. Next step is by using this data, getting the visual reports, which we will get by using the Mat Plot Library of Python and Sic-kit Learn. This library helps us in getting the results in the form of histograms, pie charts or bar charts.

OBJECTIVE

Our project's primary goal is to determine the veracity of news in order to determine if it is real or phoney. the development of a machine learning model that would allow us to recognise bogus information. It can be difficult and difficult to identify fake news only based on its content since it is intentionally produced to influence readers to believe false information.

By applying a range of methods and models, machine learning makes it easy to detect bogus news. Additionally, to examine the relationship between two words, we will apply deep learning-based NLP. You may eliminate stop words using this method as well.

The objectives of the Fake News Detection System project using Multinational Naive Bayes are:

- To develop a machine learning model that can accurately detect and classify fake news articles.
- To create a user-friendly web-based platform for users to submit news articles for detection and receive results.

- To improve awareness of the prevalence of fake news and the importance of fact-checking in modern society.

SCOPE OF PROJECT

- Massive amounts of data gave birth to AI systems that are already producing human-like synthetic texts, powering a new scale of disinformation operation. Based on Natural Language Processing (NLP) techniques, several lifelike text- generating systems have proliferated and they are becoming smarter every day.
- Even though TF - IDF classifiers worked well, there are possibilities of exploring other features to improve the model and make it a generic fit.
- While the project focused on text-based news articles and language models, AI algorithms can also analyze other features such as images, videos, date and time, sources, website, and domain for valuable information.

Problem Definition

The problem of "fake news detection" is to develop algorithms that can accurately identify and classify news articles or online content as either "real" or "fake" based on their textual features, aiming to combat the spread of misinformation by automatically flagging potentially false information across various platforms like social media and news websites.

About detecting fake news with Python. This advanced python project of detecting fake news deals with fake and real news. Using sklearn, we build a Tfidfvectorizer on our dataset. Then, we initialize a Passive Aggressive Classifier and fit the model. In the end, the accuracy score and the confusion matrix tell us how well our model fares.

CHAPTER 2

LITERATURE SURVEY

The available literature has described many automatic detection techniques of fake news and deception posts. Since there are multidimensional aspects of fake news detection ranging from using chat bots for spread of misinformation to use of click baits for the rumor spreading . There are many click baits available in social media networks including Facebook which enhance sharing and liking Proceedings of posts which in turn spreads falsified information. Lot of work has been done to detect falsified information.

MEDIA RICH FAKE NEWS DETECTION: A SURVEY

In general, the goal is profiting through click baits. Click baits lure users and entice curiosity with flashy headlines or designs to click links to increase advertisements revenues. This exposition analyzes the prevalence of fake news in light of the advances in communication made possible by the emergence of social networking sites. The purpose of the work is to come up with a solution that can be utilized by users to detect and filter out sites containing false and misleading information. We use simple and carefully selected features of the title and post to accurately identify fake posts. The experimental results show a 99.4% accuracy using logistic classifier.

WEAKLY SUPERVISED LEARNING FOR FAKE NEWS DETECTION ON TWITTER

The problem of automatic detection of fake news in social media, e.g., on Twitter, has recently drawn some attention. Although, from a technical perspective, it can be regarded as a straight-forward, binary classification problem, the major challenge is the collection of large enough training corpora, since manual annotation of tweets as fake or non-fake news is an expensive and tedious endeavor. In this paper, we discuss a weakly supervised approach, which automatically collects a large-scale, but very noisy training dataset comprising hundreds of thousands of tweets. During collection, we automatically label tweets by their source, i.e., trustworthy or untrustworthy source, and train a classifier on this dataset. We then use that classifier for a different classification target, i.e., the classification of fake and nonfake tweets. Although the labels are not accurate according to the new classification target (not all tweets by an untrustworthy source need to be fake news, and vice versa), we show that despite this unclean inaccurate dataset, it is possible to detect fake news with an F1 score of up to 0.9.

FAKE NEWS DETECTION IN SOCIAL MEDIA

Fake news and hoaxes have been there since before the advent of the Internet. The widely accepted definition of Internet fake news is: fictitious articles deliberately fabricated to deceive readers". Social media and news outlets publish fake news to increase readership or as part of psychological warfare. In general, the goal is profiting through clickbaits. Clickbaits lure users and entice curiosity with flashy headlines or designs to click links to increase advertisements revenues. This exposition analyzes the prevalence of fake news in light of the advances in communication made possible by the emergence of social networking sites. The purpose of the work is to come up with a solution that can be utilized by users to detect and filter out sites containing false and misleading information. We use simple and carefully selected features of the title and post to accurately identify fake posts. The experimental results show a 99.4% accuracy using logistic classifier.

Automatic Online Fake News Detection Combining Content and Social Signals

The proliferation and rapid diffusion of fake news on the Internet highlight the need of automatic hoax detection systems. In the context of social networks, machine learning (ML) methods can be used for this purpose. Fake news detection strategies are traditionally either based on content analysis (i.e. analyzing the content of the news) or - more recently - on social context models, such as mapping the news‟ diffusion pattern. In this paper, we first propose a novel ML fake news detection method which, by combining news content and social context features, outperforms existing methods in the literature, increasing their already high accuracy by up to 4.8%. Second, we implement our method within a Facebook Messenger chatbot and validate it with a real-world application, obtaining a fake news detection accuracy of 81.7%.

In recent years, the reliability of information on the Internet has emerged as a crucial issue of modern society. Social network sites (SNSs) have revolutionized the way in which information is spread by allowing users to freely share content. As a consequence, SNSs are also increasingly used as vectors for the diffusion of misinformation and hoaxes. The amount of disseminated information and the rapidity of its diffusion make it practically impossible to assess reliability in a timely manner, highlighting the need for automatic hoax detection systems.

THE SPREAD OF FAKE NEWS BY SOCIAL BOTS

The massive spread of fake news has been identified as a major global risk and has been alleged to influence elections and threaten democracies. Communication, cognitive, social, and

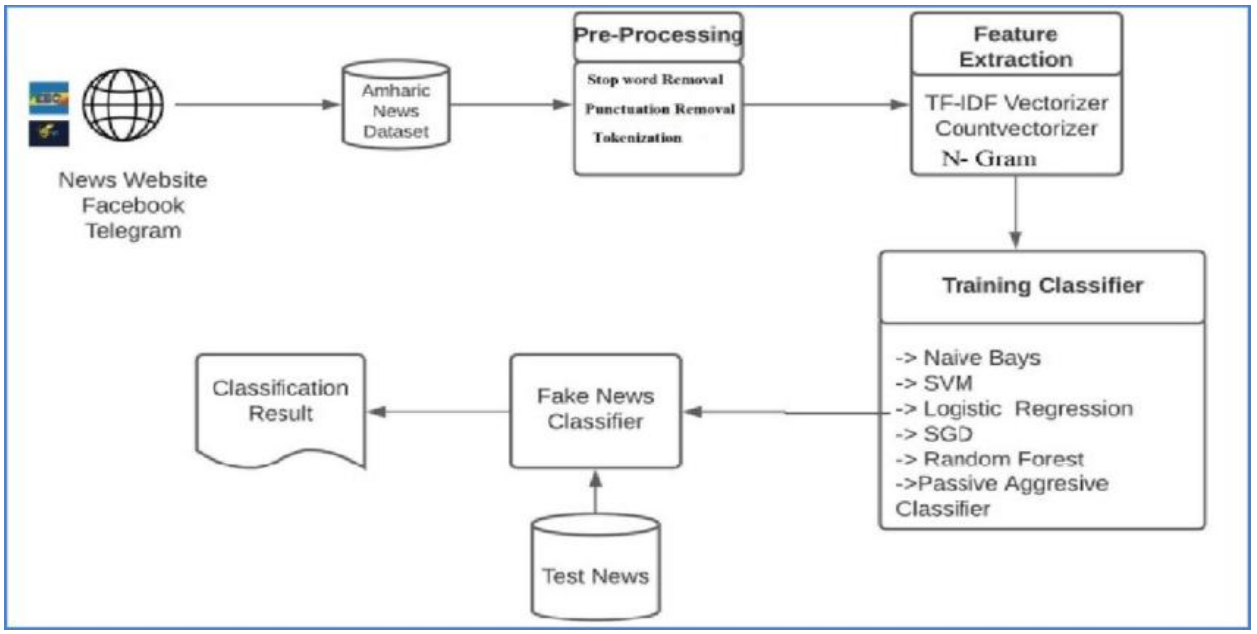
computer scientists are engaged in efforts to study the complex causes for the viral diffusion of digital misinformation and to develop solutions, while search and social media platforms are beginning to deploy countermeasures. However, to date, these efforts have been mainly informed by anecdotal evidence rather than systematic data. Here we analyze 14 million messages spreading 400 thousand claims on Twitter during and following the 2016 U.S. presidential campaign and election. We find evidence that social bots play a key role in the spread of fake news. Accounts that actively spread misinformation are significantly more likely to be bots. Automated accounts are particularly active in the early spreading phases of viral claims, and tend to target influential users. Humans are vulnerable to this manipulation, retweeting bots who post false news. Successful sources of false and biased claims are heavily supported by social bots. These results suggests that curbing social bots may be an effective strategy for mitigating the spread of online misinformation.

MISLEADING ONLINE CONTENT

Big Data Analytics and Deep Learning are two high-focus of data science. Big Data has become important as many organizations both public and private have been collecting massive amounts of domain-specific information, which can contain useful information about problems such as national intelligence, cyber security, fraud detection, marketing, and medical informatics. Companies such as Google and Microsoft are analyzing large volumes of data for business analysis and decisions, impacting existing and future technology. Deep Learning algorithms extract highlevel, complex abstractions as data representations through a hierarchical learning process. Complex abstractions are learnt at a given level based on relatively simpler abstractions formulated in the preceding level in the hierarchy. A key benefit of Deep Learning is the analysis and learning of massive amounts of unsupervised data, making it a valuable tool for Big Data Analytics where raw data is largely unlabeled and un-categorized. In the present study, we explore how Deep Learning can be utilized for addressing some important problems in Big Data Analytics, including extracting complex patterns from massive volumes of data, semantic indexing, data tagging, fast information retrieval, and simplifying discriminative tasks. We also investigate some aspects of Deep Learning research that need further exploration to incorporate specific challenges introduced by Big Data Analytics, including streaming data, high-dimensional data, scalability of models, and distributed computing.

CHAPTER 3
METHODOLOGY

This paper explains the system which is developed in three parts. The first part is static which works on machine learning classifier. We studied and trained the model with 4 different classifiers and chose the best classifier for final execution. The second part is dynamic which takes the keyword/text from user and searches online for the truth probability of the news. The third part provides the authenticity of the URL input by user. In this paper, we have used Python and its Sci-kit libraries. Python has a huge set of libraries and extensions, which can be easily used in Machine Learning. Sci-Kit Learn library is the best source for machine learning algorithms where nearly all types of machine learning algorithms are readily available for Python,



- **Data Collection and Pre-processing** A labeled data set containing fake and real news articles was used to train and test the model. The text data underwent pre-processing steps such as tokenization, removal of stop words, stemming, and vectorization using Term Frequency Inverse Document Frequency (TF - IDF). These steps ensured that the input features were optimized for the classifier.

- **Passive-Aggressive Classifier** The Passive-Aggressive Classifier is a linear model designed for online learning. Its dual behavior enables it to remain passive when predictions are correct and update aggressively in response to errors. The hinge loss function was employed to measure errors, and the model parameters were updated incrementally based on this loss. This approach ensures quick adaptation to new data while maintaining stability.

- **Implementation and Evaluation** The model was implemented using Python and Scikit-learn. The dataset was divided into training and testing sets, and performance metrics such as accuracy, precision, recall, and F1-score were used to evaluate the model. Cross-validation was employed to ensure robustness and generalizability.
- **Results analysis**: Identifying strengths, limitations, and future improvements for each model.

EXISTING SYSTEM

There exists a large body of research on the topic of machine learning methods for deception detection, most of it has been focusing on classifying online reviews and publicly available social media posts. Particularly since late 2016 during the American Presidential election, the question of determining 'fake news' has also been the subject of particular attention within the literature. Convoy, Rubin, and Chen outlines several approaches that seem promising towards the aim of perfectly classify the misleading articles. They note that simple content-related n-grams and shallow parts-of-speech tagging have proven insufficient for the classification task, often failing to account for important context information. Rather, these methods have been shown useful only in tandem with more complex methods of analysis. Deep Syntax analysis using Probabilistic Context Free Grammars have been shown to be particularly valuable in combination with n-gram methods. Feng, Banerjee, and Choi are able to achieve 85%-91% accuracy in deception related classification tasks using online review corpora.

PROPOSED SYSTEM

In this paper a model is build based on the count vectorizer or a fidget matrix (i.e) word tallies relatives to how often they are used in other articles in your dataset) can help . Since this problem is a kind of text classification, Implementing a Naive Bayes classifier will be best as this is standard for text-based processing. The actual goal is in developing a model which was the text transformation (count vectorizer vs fidget vectorizer) and choosing which type of text to use (headlines vs full text). Now the next step is to extract the most optimal features for count vectorizer or fidget-vectorizer, this is done by using a n-number of the most used words, and/or phrases, lower casing or not, mainly removing the stop words which are common words such as “the”, “when”, and “there” and only using those words that appear at least a given number of times in a given text dataset

Problems

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

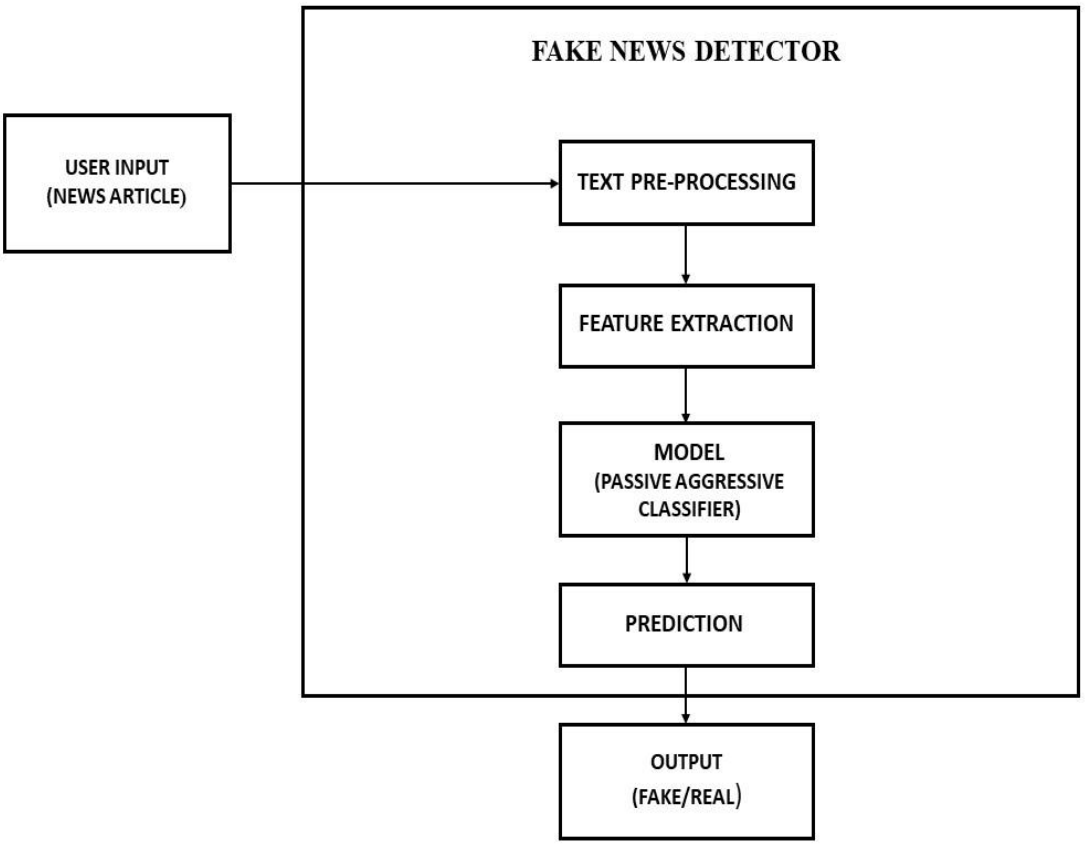
quality news with intentionally false information. The extensive spread of fake news has the potential for extremely negative impacts on individuals and society. Therefore, fake news detection on social media has recently become an emerging research that is attracting tremendous attention. Fake news detection on social media presents unique characteristics and challenges that make existing detection algorithms from traditional news media ineffective or not applicable. First, fake news is intentionally written to mislead readers to believe false information, which makes it difficult and 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. Second, exploiting this auxiliary information is challenging in and of itself as users' social engagements with fake news produce data that is big, incomplete, unstructured, and noisy.

Advantages

There are several advantages of using machine learning for detecting fake news:

- Algorithms that use machine learning can find links and patterns in data that may not be obvious to people. Machine learning algorithms can precisely identify fake news stories by examining the wording, sources, and social media networks linked to news pieces.
- In order to stop the spread of incorrect information, social media platforms and news organizations can immediately take action thanks to machine learning algorithms' ability to identify fake news stories in real-time.
- Algorithms that use machine learning are able to pick up new information and adapt. Machine learning algorithms may be trained to recognize new trends and identify new sorts of fake news stories as fake news tactics advance.
- The process of identifying false news stories may be automated with machine learning algorithms. Humans will have less work to do, as a result, freeing them up to work on things like fact-checking and investigative journalism.
- The detection of false news stories may be done at a reasonable price using machine learning algorithms. Once taught, the algorithms may be widely used without incurring a lot of expense.

SYSTEM ARCHITECTURE



SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS:

- System - Pentium-IV
- Speed - 2.4GHZ
- Hard disk - 40GB
- Monitor - 15VGA color
- RAM - 512MB

SOFTWARE REQUIREMENTS:

- Operating System - Windows XP
- Coding language - PYTHON

SOFTWARE ENVIRONMENT

PYTHON

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

- **Python is Interpreted** − Python is processed at runtime by the interpreter.

You do not need to compile your program before executing it. This is similar to PERL and PHP.

- **Python is Interactive** − You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
- **Python is Object-Oriented** − Python supports Object-Oriented style or technique of programming that encapsulates code within objects.
- **Python is a Beginner's Language** − Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

History of Python

Python was developed by Guido van Rossum in the late eighties and early nineties at the National Research Institute for Mathematics and Computer Science in the Netherlands. Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, SmallTalk, and Unix shell and other scripting languages. Python is copyrighted. Like Perl, Python source code is now available under the GNU General Public License (GPL). Python is now maintained by a core development team at the institute, although Guido van Rossum still holds a vital role in directing its progress.

PYTHON FEATURES

Python's features include −

- **Easy-to-learn** − Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.
- **Easy-to-read** − Python code is more clearly defined and visible to the eyes.
- **Easy-to-maintain** − Python's source code is fairly easy-to-maintain.
- **A broad standard library** − Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Mac into sh.

- **Interactive Mode** − Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.

- **Portable** − Python can run on a wide variety of hardware platforms and has the same interface on all platforms.
- **Extendable** − You can add low-level modules to the Python interpreter.

These modules enable programmers to add to or customize their tools to be more efficient.

- **Databases** − Python provides interfaces to all major commercial databases.
- **GUI Programming** − Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.

- **Scalable** − Python provides a better structure and support for large programs than shell scripting.

Apart from the above-mentioned features, Python has a big list of good features, few are listed below −

- It supports functional and structured programming methods as well as OOP.

- It can be used as a scripting language or can be compiled to byte-code for building large applications.

- It provides very high-level dynamic data types and supports dynamic type checking.

- It supports automatic garbage collection.
- It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

Python is available on a wide variety of platforms including Linux and Mac OS X.

Let's understand how to set up our Python environment.

Getting Python

The most up-to-date and current source code, binaries, documentation, news, etc., is available on the official website of Python https://www.python.org.

Windows Installation

Here are the steps to install Python on Windows machine.

[1] Open a Web browser and go to https://www.python.org/downloads/.

[2] Follow the link for the Windows installer python-XYZ.msifile where XYZ is the version you need to install.

[3] To use this installer python-XYZ.msi, the Windows system must support Microsoft Installer 2.0. Save the installer file to your local machine and then run it to find out if your machine supports MSI.

[4] Run the downloaded file. This brings up the Python install wizard, which is really easy to use. Just accept the default settings, wait until the install is finished, and you are done.

The Python language has many similarities to Perl, C, and Java. However, there are some definite differences between the languages.

INTERACTIVE MODE PROGRAMMING

Invoking the interpreter without passing a script file as a parameter brings up the following prompt −

```
$ python

Python2.4.3(#1,Nov112010,13:34:43)
[GCC 4.1.220080704(RedHat4.1.2-48)] on linux2

Type"help","copyright","credits"or"license"for more information.
>>>
```

Type the following text at the Python prompt and press the Enter −

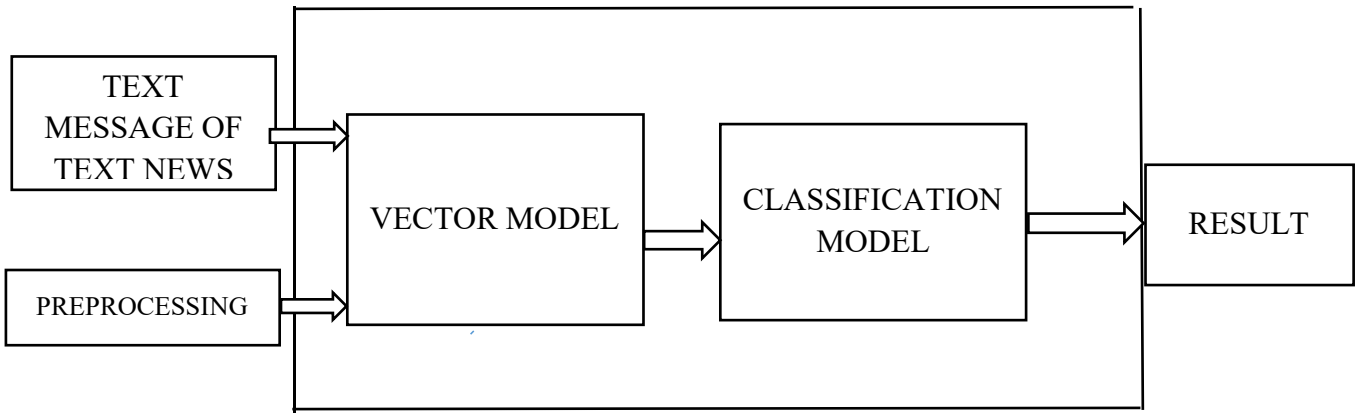
```
>>>print"Hello, Python!"
```

If you are running new version of Python, then you would need to use print statement with parenthesis as in **print ("Hello, Python!");**. However in Python version 2.4.3, this produces the following result −

```
Hello, Python!
```

System Design

The system is developed in two parts. They are Static Implementation and Dynamic Implementation The first component is Static implementation that works on Machine Learning algorithms. Here, what we do is extract the features from the dataset which is already pre-processed The features which are extracted are fed into four different classifiers. The classifiers used are Logistic Regression, Random Forest Classifier, Support Vector Machine, and Passive-Aggressive Classifier. After fitting the model, we compare the accuracies. Model performance is determined with the help of a confusion matrix. The second component is Dynamic implementation which will take the keyword or text present in the news articles from the dataset.



Algorithms

Passive Agreesive Classifier

The Passive-Aggressive Classifier is an algorithm that is used online, and it is perfect for classifying nearly all of the huge data streams. It is very simple to implement, and it is much faster, and it can be explained by an example,Machine Learning-Based Approach for Fake News Detection 517 where learning and gaining from it would be easier and afterward sput it aside. An algorithm of this kind will remain passive for the correctly predicted data of outcomes and will turn more aggressive for the false data of outcomes and will make the update and also will make required adjustments. Unlike other algorithms, it will not converge. These is called Passive-Aggressive algorithms because of the following reasons:

- **PASSIVE:** If it is the correct prediction, then it will keep the model the same and will not make any changes.
- **AGGRESSIVE:** If it is an incorrect prediction, then it will make changes to the model.

INPUT: Aggressiveness parameter C > 0

INITIALIZE: w1 = (0,. . . ,0)

For t = 1,2, . . .

• Receive instance: xt e R n

• Predict: yt = sign (wt, xt)

• Receive correct label: yt e {−1, +1}

• Suffer loss: lt = max {0, 1 − yt (wt,xt)}

• Update

MODULES

A. Data Collection

B. Data Pre Processing

C. Data Cleaning

D. Feature Extraction

E. Training the Classifier

F. Model Evalution

MODULES DESCRIPTION

1. Data Collection

Online news is mostly collected from different types of sources, like press agencies, search engines, and websites of social media. The dataset used in our project is a simple and realistic dataset that contains 6335 news articles simply classified as Fake or Real and later stored in a CSV file. The attributes of the dataset are:

Id: A unique identifier for the article

Title: Article Headline

Text: Article textual content

Class Label: Fake and Real

2. Data Pre-processing

Data that is taken from social media will be mostly not structured and most of which will be an informal type of communication with shortcuts, different slang, and bad grammar. To increase the performance and reliability we have to pre-process the data before using it as a predictive model.

3. Data Cleaning

The data may be in the format of either structured or unstructured. A structured format has patterns that are well defined and the unstructured data do not have a proper structure. Among structured and semi-structured formats,comparison of the better structured and then unstructured way. Text and the data have to be cleaned to highlight the attributes which we want our machine learning system to work on accuracy.

It comprises a few steps:

(1) Punctuation Removal : Punctuations will give the grammatical expression to the sentence which enhances our understandability. The vectorize checks the number of words and cannot be the context that it doesn't add esteem, hence we will eliminate every character which is special. Example: "How are you doing?" Instead "How are you doing".

(2) Tokenization : It divides the context into certain units as sentences split into words like this the unstructured text is given a structure.

Example: "work at the place" is split into wok" "at " "the ".

(3) Removing Stop words : These are the common words which are mostly appeared in any text. They lend much about the info hence we remove them. Example: copper or aluminum is okay for me->copper, aluminum, okay.

(4) Stemming : This process helps us to reduce the length of the word to its stem form. It generally treats words that are related similarly. Suffices, like "er", "ible", "ness", etc. are removed by rule-based approach. After data cleaning is done we perform exploratory data analysis to improve the statistical analysis of the given dataset.

4. Feature Generation

Many features like word count, repetition of distinctive words, repetition of large words, etc can be done using text data. It is done through the creation of the representation of words that catch meanings, relationships, and numerous other types of context which are used within, Computer is made to understand the given text and then do the classification of text. To make machine learning algorithms understand our data vectorizing is done which encodes the text into integers which is numerical form thus to vectors.

- Count Vectorizer tells us about the words that are present in the data which are texts. The result is given as 1 if it is there in the sentence or else 0 is. For each text document bag of words is created along with the document.
- TF-IDF calculates the relative frequency (number of times repeated) of the word that is seen in the document when its frequency is compared within each and every document. TF

addresses the Term Frequency and also computes the frequency of a term appearing within a document. IDF stands for Inverse Document Frequency.

To store each word of relative count in the document matrix TF-IDF is applied to the body.

Number of times t occurs in document ‘d’

TF(t, d) = Total word count of document ‘d’

IDF(t, d) = Number of documents with term t in it Total number of documents

TFIDF(t, d) = T F(t, d) ∗ IDF(t)

To store each word of relative count in the document matrix TF-IDF is applied on the body

5. Training the Model

Using four different machine learning classifiers like Logistic regression, Random forest classifier, Support vector machine, and Passive aggressive classifier, training of the model is done after the features are extracted from the pre-processed datasets. A passive-aggressive classifier is our good performing classifier hence it is selected finally. After this, it is stored in the disk where it is used for fake news classification. It in-takes the article in the dataset from the user as input then predicts the reliability of the news.

6. Model Evaluation

Once fitting the model, we evaluate the performance of each model with the help of a confusion matrix. After comparing the accuracies of the four classifiers, the classifier which is performing best will be taken as a classification model for the detection of news. Most of the approaches will consider this as a problem of classification that estimates if the information is real or fake:

True Positive (TP): when anticipated information that is false is correctly grouped as fake news.

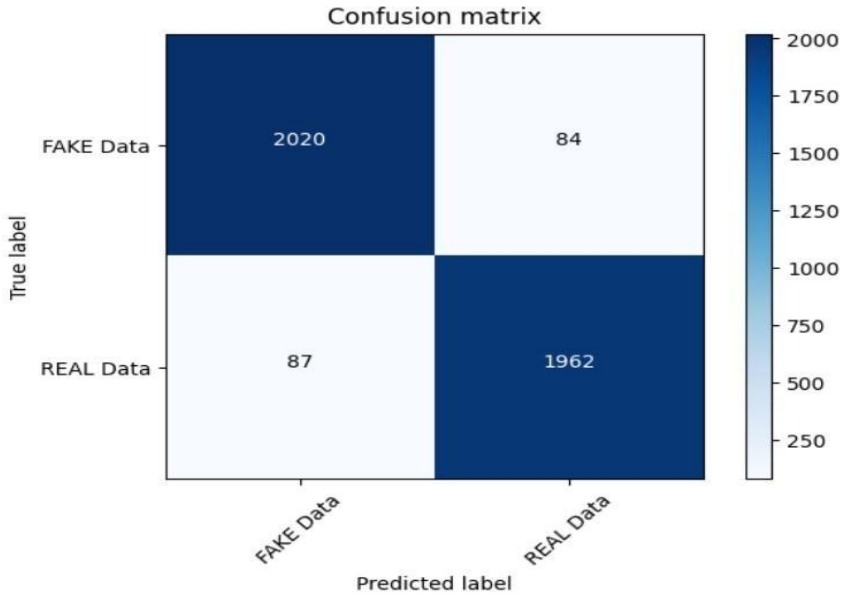
True Negative (TN): when anticipated information that is true is correctly grouped as true news.

False Negative (FN): when anticipated information that is true is not correctly grouped as fake news.

False Positive (FP): when anticipated information that is false is not correctly grouped as true news.

Confusion matrix: A table that outlines the performance of a classifier on a bunch of test information for which the genuine values are known. It is used to visualize algorithm performance. The incorrect and correct numbers of predictions are concluded with value count and further, each class will be broken. A confusion matrix is a summary of prediction results on a classification model. It describes how the model of classification will be confused when it makes predictions. It’s anything but an understanding not just into the

mistakes that are made by the classifier yet more critically the sorts of blunders that are being made.



Formulas for Precision, Recall, F1 score, accuracy:

1. Precision = TP/(TP + FP)
2. Recall = TP /(TP + FN)
3. F1 Score = 2 *((precision*recall) / (precision + recall))
4. Accuracy = (TP + TN) / (TP + TN + FP + FN)

These are the metrics used in machine learning which enable us to evaluate the performance of a classifier model.

Accuracy of classifiers

Model Accuracy

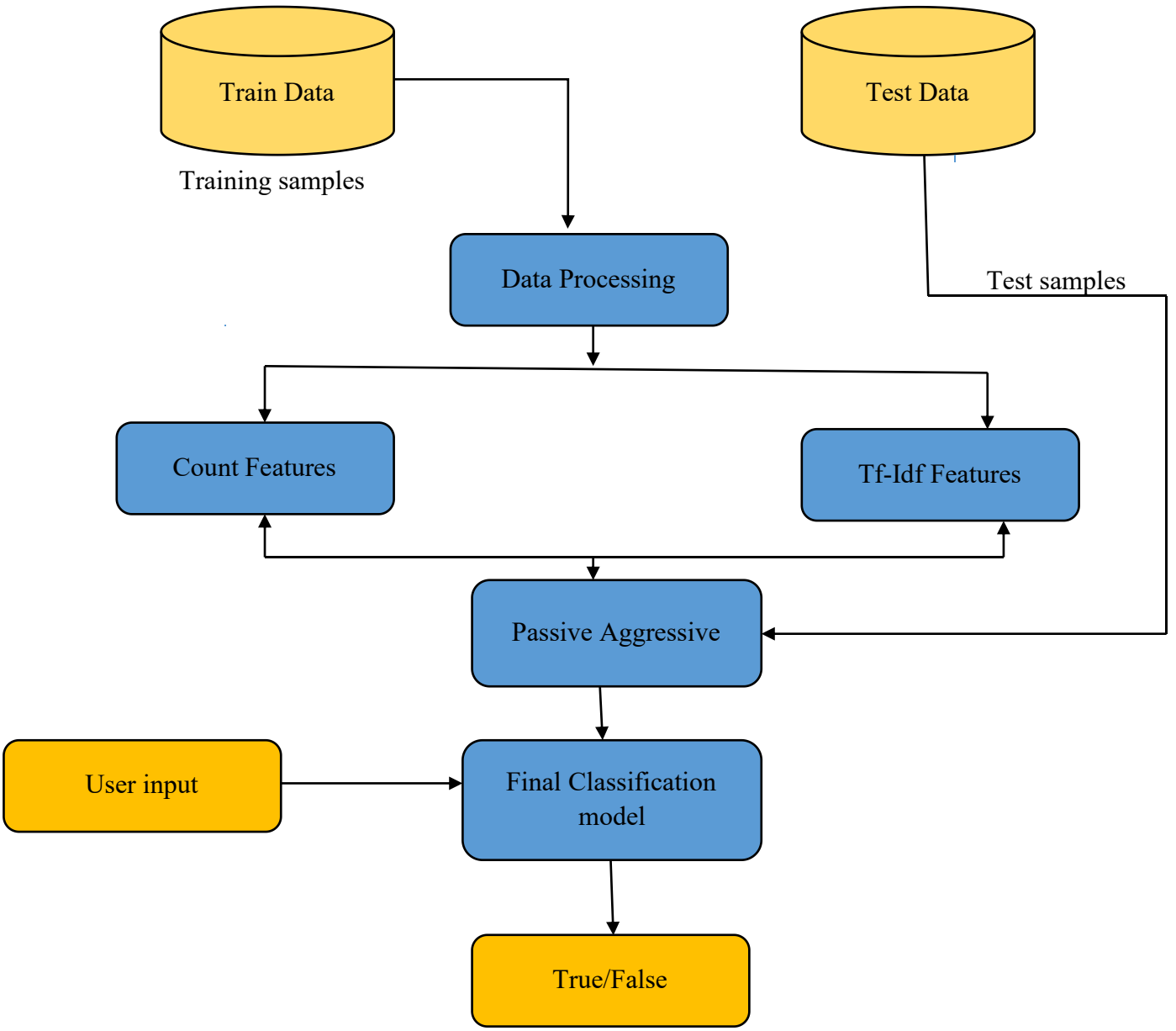
Random Forest Classifier 0.90

Logistic Regression 0.92

Passive Aggressive Classifier 0.94

Support Vector Machine 0.93

WORKFLOW



Implementation Steps

1. Get Dataset from Folder
2. Libraries required for the project.
3. **numpy:** NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.
4. **pandas:** Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

5. **sklearn:** Scikit-learn is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms.
6. **nltk:** The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing for English written in the Python programming language.
7. Remove all non words from news column
8. Remove all stop words using nltk 5. Then create features using TfidfVectorizer
9. Then create X as input variable having feature and Y as output variable having label 0 for fake and 1 for real. 7. Then split data in training and testing.
10. Then pass training data to Passive Aggressive Classifier algorithm. Passive Aggressive Classifier belongs to the category of online learning algorithms in machine learning. It works by responding as passive for correct classifications and responding as aggressive for any miscalculation.
11. **Passive:** If the prediction is correct, keep the model and do not make any changes. i.e., the data in the example is not enough to cause any changes in the model.
12. **Aggressive:** If the prediction is incorrect, make changes to the model. i.e., some change to the model may correct it.
13. Then we will check accuracy of model and deploy it on flask web framework.
14. Our web application will take a news as input and will classify it as Fake or Real according to model.

CHAPTER 4

RESULTS AND DISCUSSION

- Algorithm‟s accuracy depends on the type and size of your dataset. More the data, more chances of getting correct accuracy.
- Machine learning depends on the variations and relations
- Understanding what is predictable is as important as trying to predict it.
- While making algorithm choice , speed should be a consideration factor.

REQUIREMENTANALYSIS

Requirement analysis, also called requirement engineering, is the process of determining user expectations for a new modified product. It encompasses the tasks that determine the need for analysing, documenting, validating and managing software or system requirements. The requirements should be documentable, actionable, measurable, testable and

traceable related to identified business needs or opportunities and define to a level of detail, sufficient for system design.

FUNCTIONAL REQUIREMENTS

It is a technical specification requirement for the software products. It is the first step in the requirement analysis process which lists the requirements of particular software systems including functional, performance and security requirements. The function of the system depends mainly on the quality hardware used to run the software with given functionality.

Usability

It specifies how easy the system must be use. It is easy to ask queries in any format which is short or long, porter stemming algorithm stimulates the desired response for user.

Robustness

It refers to a program that performs well not only under ordinary conditions but also under unusual conditions. It is the ability of the user to cope with errors for irrelevant queries during execution.

Security

The state of providing protected access to resource is security. The system provides good security and unauthorized users cannot access the system there by providing high security.

Reliability

It is the probability of how often the software fails. The measurement is often expressed in MTBF (Mean Time Between Failures). The requirement is needed in order to ensure that the processes work correctly and completely without being aborted. It can handle any load and survive and survive and even capable of working around any failure.

Compatibility

It is supported by version above all web browsers. Using any web servers like localhost makes the system real-time experience.

Flexibility

The flexibility of the project is provided in such a way that is has the ability to run on different environments being executed by different users.

Safety

Safety is a measure taken to prevent trouble. Every query is processed in a secured manner without letting others to know one‟s personal information.

NON- FUNCTIONAL REQUIREMENTS

Portability

It is the usability of the same software in different environments. The project can be run in any operating system.

Performance

These requirements determine the resources required, time interval, throughput and everything that deals with the performance of the system.

Accuracy

The result of the requesting query is very accurate and high speed of retrieving information. The degree of security provided by the system is high and effective.

Maintainability

Project is simple as further updates can be easily done without affecting its stability. Maintainability basically defines that how easy it is to maintain the system. It means that how easy it is to maintain the system, analyse, change and test the application. Maintainability of this project is simple as further updates can be easily done without affecting its stability.

SYSTEM DESIGN AND TESTING PLAN

INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

1.What data should be given as input?

2.How the data should be arranged or coded?

3.The dialog to guide the operating personnel in providing input.

4.Methods for preparing input validations and steps to follow when error occur

OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system‟s relationship to help user decision-making. The output form of an information system should accomplish one or more of the following objectives.

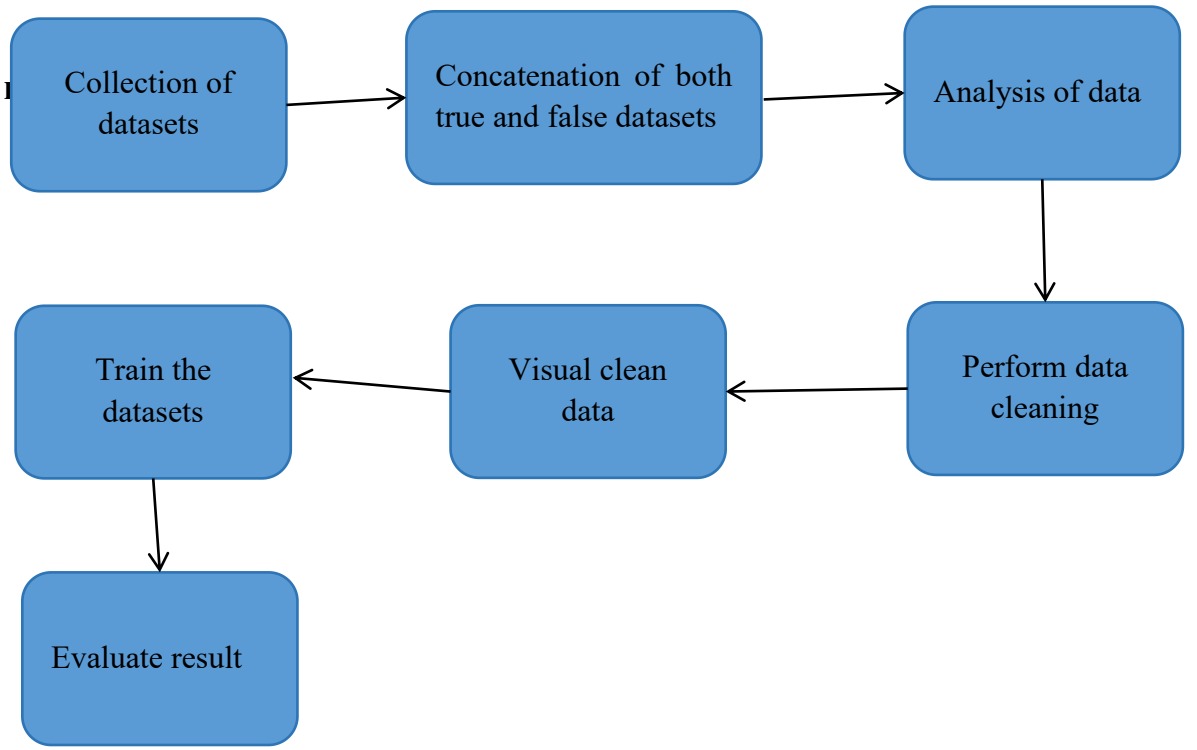
- Convey information about past activities, current status or projections of the
- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

SOCIAL FEASIBILITY

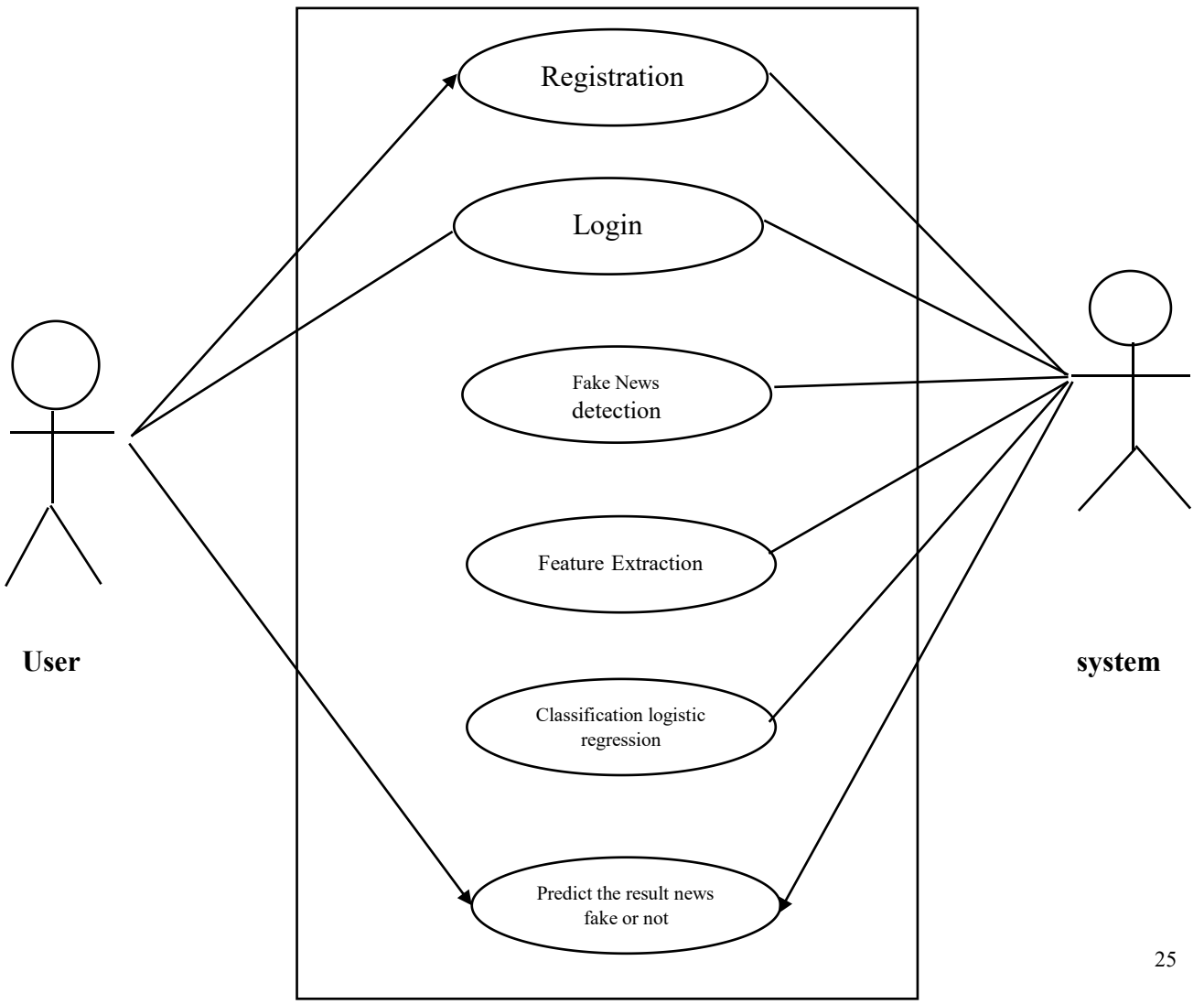
The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

DATAFLOW DIAGRAM

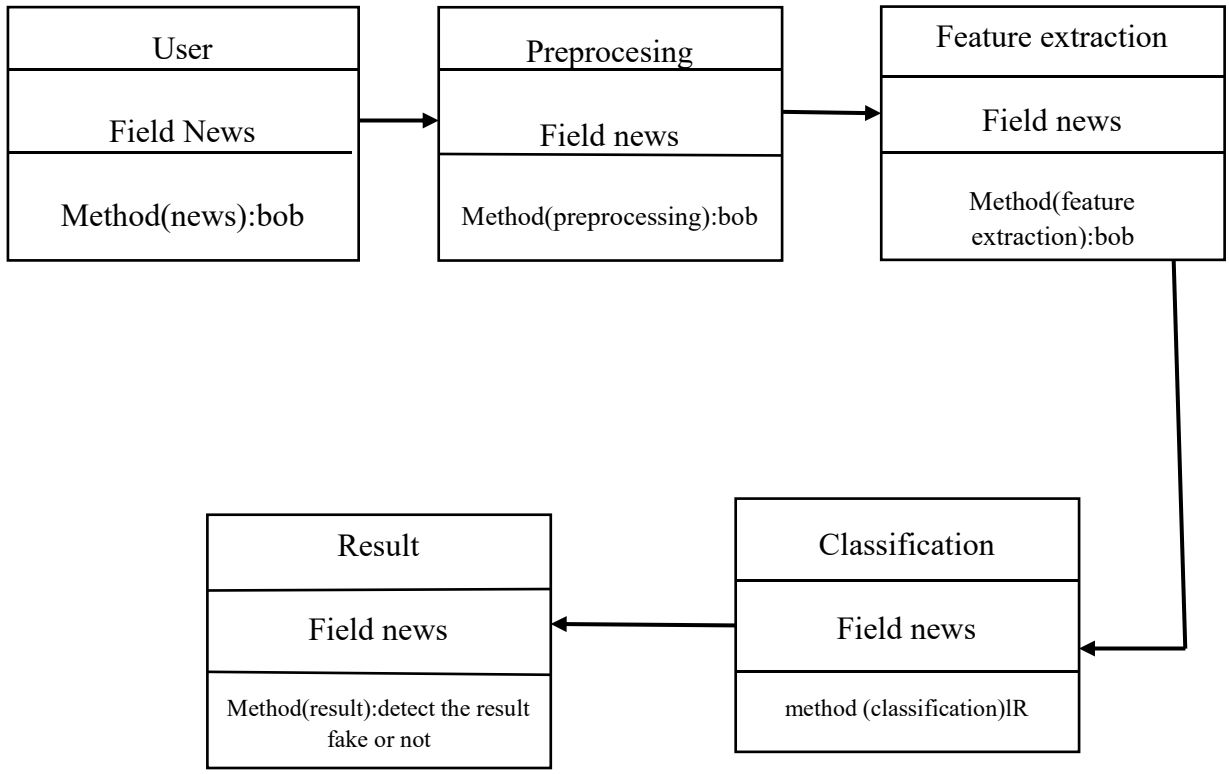
- The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
- The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.



Activity diagram



Class Diagram



TEST PROCEDURE

SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner.

There are various types of test. Each test type addresses a specific testing requirement.

UNIT TESTING

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific

business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

INTEGRATION TESTING

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

FUNCTIONAL TESTING

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input: identified classes of valid input must be accepted.

Invalid Input: identified classes of invalid input must be rejected.

Function: identified functions must be exercised.

Output: identified classes of application outputs must be exercised.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

WHITE BOX TESTING

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other

kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works

ACCEPTANCE TESTING

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Source code

Login.html:

```
       <!doctype html>
<html lang="en">
  <head>
    <!-- Required meta tags -->
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1">

    <!-- Favicons -->
    <link href="static/icons8-facebook-ios-16-filled-32.png" rel="icon">
    <link    href="static/icons8-facebook-ios-16-filled-96.png"    rel="icons8-facebook-ios-
16filled-96.png">
    <!-- Bootstrap CSS -->
    <link href="https://cdn.jsdelivr.net/npm/bootstrap@5.1.0/dist/css/bootstrap.min.css"
rel="stylesheet" integrity="sha384-
KyZXEAg3QhqLMpG8r+8fhAXLRk2vvoC2f3B09zVXn8CA5QIVfZOJ3BCsw2P0p/We" crossorigin="anonymous">
    <link href="https://unpkg.com/tailwindcss@^2/dist/tailwind.min.css" rel="stylesheet">
```

```
    <title>Fake News Detection</title>
  </head>
  <body class="bg-gray-200">
    <header class="text-gray-600 body-font -my-8">
        <div style="position: relative; top: 15px; right: 50px;" class="container mx-auto
flex flex-wrap p-5 flex-col md:flex-row items-center">
          <a class="flex title-font font-medium items-center text-gray-900 mb-4 md:mb-0">
            <svg xmlns="http://www.w3.org/2000/svg" fill="none" stroke="currentColor"
stroke-linecap="round" stroke-linejoin="round" stroke-width="2" class="w-10 h-10 text-white
p-2 bg-green-500 rounded-full" viewBox="0 0 24 24">
              <path d="M12 2L2 7l10 5 10-5-10-5zM2 17l10 5 10-5M2 12l10 5 10-5"></path>
            </svg>
            <span class="ml-3 text-xl">Fake News Detection</span>
          </a>
           <nav style="position: relative; bottom: 13px; left:100px;" class="md:ml-auto flex
flex-wrap items-center text-base justify-center">
            <a href="/" class="mr-5 hover:text-gray-900">Home</a>
            <a href="/predict" class="mr-5 hover:text-gray-900">Prediction</a>
            <a href="#" class="mr-5 hover:text-gray-900">About</a>
            <a href="#" class="mr-5 hover:text-gray-900">Contact us</a>
          </nav>
        </div>
      </header><hr>
      <section  style="position:  relative;  right:  80px;  top:  30px;"  class="text-gray-600
body-font -my-8">
        <div class="container mx-auto flex px-5 py-24 md:flex-row flex-col items-center">
          <div class="lg:flex-grow md:w-1/2 lg:pr-24 md:pr-16 flex flex-col md:items-start
md:text-left mb-16 md:mb-0 items-center text-center">
            <h1 style="position: relative; left: 80px;" class="title-font sm:text-4xl
text3xl mb-4 font-medium text-gray-900" >Fake News Detection
            <br class="hidden lg:inline-block">
              using Machine Learning</h1>

            <p class="mb-8 leading-relaxed">Realiable Source for Verifying News Authencity,
              Here you can Validate the accuracy of the news to uncover the truth.
            </p>
            <div class="flex justify-center">
              <a href="/predict"><button style="position: relative; left: 205px; background:
#10b962;" class="inline-flex text-white bg-green-500 border-0 py-2 px-6 focus:outline-none
hover:bg-green-600 rounded text-lg">Get Started</button></a> </div>
          </div>
          <div style="position:relative; left:90px;" class="lg:max-w-lg lg:w-full md:w-1/2
w-5/6">
            <img class="object-cover object-center rounded" alt="hero"
src="../static/hero_img.svg">
          </div>
        </div>

    <!-- Optional JavaScript; choose one of the two!
```

```
        Option 1: Bootstrap Bundle with Popper -->
    <script
src="https://cdn.jsdelivr.net/npm/bootstrap@5.1.0/dist/js/bootstrap.bundle.min.js"
integrity="sha384-U1DAWAznBHeqEIlVSCgzq+c9gqGAJn5c/t99JyeKa9xxaYpSvHU5awsuZVVFIhvj"
crossorigin="anonymous"></script>

    <!-- Option 2: Separate Popper and Bootstrap JS --> <!-
    -
    <script src="https://cdn.jsdelivr.net/npm/@popperjs/core@2.9.3/dist/umd/popper.min.js"
integrity="sha384-eMNCOe7tC1doHpGoWe/6oMVemdAVTMs2xqW4mwXrXsW0L84Iytr2wi5v2QjrP/xp"
crossorigin="anonymous"></script>
    <script src="https://cdn.jsdelivr.net/npm/bootstrap@5.1.0/dist/js/bootstrap.min.js"
integrity="sha384-cn7l7gDp0eyniUwwAZgrzD06kc/tftFf19TOAs2zVinnD/C7E91j9yyk5//jjpt/"
crossorigin="anonymous"></script>
    -->
  </body>
</html>
```

Prediction.html:

```
<!doctype html>
<html lang="en">
  <head>
    <!-- Required meta tags -->
    <meta charset="utf-8">
```

```
    <meta name="viewport" content="width=device-width, initial-scale=1">

    <!-- Favicons -->
    <link href="static/icons8-facebook-ios-16-filled-32.png" rel="icon">
    <link    href="static/icons8-facebook-ios-16-filled-96.png"    rel="icons8-facebook-ios-
16filled-96.png">
    <!-- Bootstrap CSS -->
    <link href="https://cdn.jsdelivr.net/npm/bootstrap@5.1.0/dist/css/bootstrap.min.css"
rel="stylesheet" integrity="sha384-
KyZXEAg3QhqLMpG8r+8fhAXLRk2vvoC2f3B09zVXn8CA5QIVfZOJ3BCsw2P0p/We" crossorigin="anonymous">
    <link href="https://unpkg.com/tailwindcss@^2/dist/tailwind.min.css" rel="stylesheet">
```

```
    <title>Fake News Detection</title>
  </head>
  <body class="bg-gray-200">
    <header class="text-gray-600 body-font -my-8">
        <div style="position: relative; top: 15px; right: 50px;" class="container mx-auto
flex flex-wrap p-5 flex-col md:flex-row items-center">
          <a class="flex title-font font-medium items-center text-gray-900 mb-4 md:mb-0">
            <svg xmlns="http://www.w3.org/2000/svg" fill="none" stroke="currentColor"
stroke-linecap="round" stroke-linejoin="round" stroke-width="2" class="w-10 h-10 text-
white p-2 bg-green-500 rounded-full" viewBox="0 0 24 24">
              <path d="M12 2L2 7l10 5 10-5-10-5zM2 17l10 5 10-5M2 12l10 5 10-5"></path>
            </svg>
            <span class="ml-3 text-xl">Fake News Detection</span>
          </a>
          <nav style="position: relative; bottom: 13px; left:100px;" class="md:ml-auto
          flex
flex-wrap items-center text-base justify-center">
            <a href="/" class="mr-5 hover:text-gray-900">Home</a>
            <a href="/predict" class="mr-5 hover:text-gray-900">Prediction</a>
            <a href="#" class="mr-5 hover:text-gray-900">About </a>
            <a href="#" class="mr-5 hover:text-gray-900">Contact us</a>
          </nav>
        </div>
      </header><hr>
      <section class="text-gray-600 body-font -my-8">
        <div class="container px-5 py-24 mx-auto">
          <div class="flex flex-col text-center w-full mb-20">
            <h2 style="font-size:small;" class="text-xs text-green-500 tracking-widest
font-medium title-font mb-1">Machine learning project</h2>
            <h1 class="sm:text-3xl text-2xl font-medium title-font mb-4 text-gray-
            900">Fake
News Detection</h1>
            <p class="lg:w-2/3 mx-auto leading-relaxed text-base">"Fake news detection
using machine learning" is a project that involves training a machine learning model to
distinguish between real and fake news articles. The model is trained on a dataset of news
articles using supervised learning algorithms and NLP techniques. The goal of the project
is to help identify and flag fake news articles to prevent their spread and reduce the
impact of misinformation in society.</p>
          </div>
          <br><p style="font-weight:500; text-align: center; position: relative; bottom:
75px;" class="text-primary"><b>{{prediction_text}}</b></p>
```

```
<div style="position: relative; bottom: 40px;" class="center">
            <form action="/predict" method="POST">
                <div class="mb-3">
                  <label for="exampleInputEmail1" class="form-label">Enter news
headlines</label>
                  <input type="text" class="form-control" id="news" name="news" aria-
describedby="emailHelp">
                  <div id="emailHelp" class="form-text">Authenticity of News can be
verified here.</div>
                </div>
                <button style="height: 45px; width: 110px; background:
#10b962;" type="submit" class="inline-flex text-white bg-green-500 border-0 py-2 px-6
focus:outline-none hover:bg-green-600 rounded text-lg">Predict</button> </form>
          </div>
        </div>

      </section>

    <!-- Optional JavaScript; choose one of the two! -->

    <!-- Option 1: Bootstrap Bundle with Popper -->
    <script
src="https://cdn.jsdelivr.net/npm/bootstrap@5.1.0/dist/js/bootstrap.bundle.min.js"
integrity="sha384-U1DAWAznBHeqEIlVSCgzq+c9gqGAJn5c/t99JyeKa9xxaYpSvHU5awsuZVVFIhvj"
crossorigin="anonymous"></script>

    <!-- Option 2: Separate Popper and Bootstrap JS -->
    <!--
    <script src="https://cdn.jsdelivr.net/npm/@popperjs/core@2.9.3/dist/umd/popper.min.js"
integrity="sha384-eMNCOe7tC1doHpGoWe/6oMVemdAVTMs2xqW4mwXrXsW0L84Iytr2wi5v2QjrP/xp"
crossorigin="anonymous"></script>
    <script src="https://cdn.jsdelivr.net/npm/bootstrap@5.1.0/dist/js/bootstrap.min.js"
integrity="sha384-cn7l7gDp0eyniUwwAZgrzD06kc/tftFf19TOAs2zVinnD/C7E91j9yyk5//jjpt/"
crossorigin="anonymous"></script>
    -->
  </body>
</html>
```

App.py

```
from flask import Flask, render_template, request
import pandas as pd import sklearn import
itertools import numpy as np import seaborn as sb
import re import nltk import pickle
```

```
from sklearn.feature_extraction.text import TfidfVectorizer from matplotlib
import pyplot as plt from sklearn.linear_model import
PassiveAggressiveClassifier from nltk.stem import WordNetLemmatizer from
nltk.corpus import stopwords app =
Flask(__name__,template_folder='./templates',static_folder='./static')
```

```
loaded_model = pickle.load(open("model.pkl", 'rb'))
vector = pickle.load(open("vector.pkl", 'rb'))
```

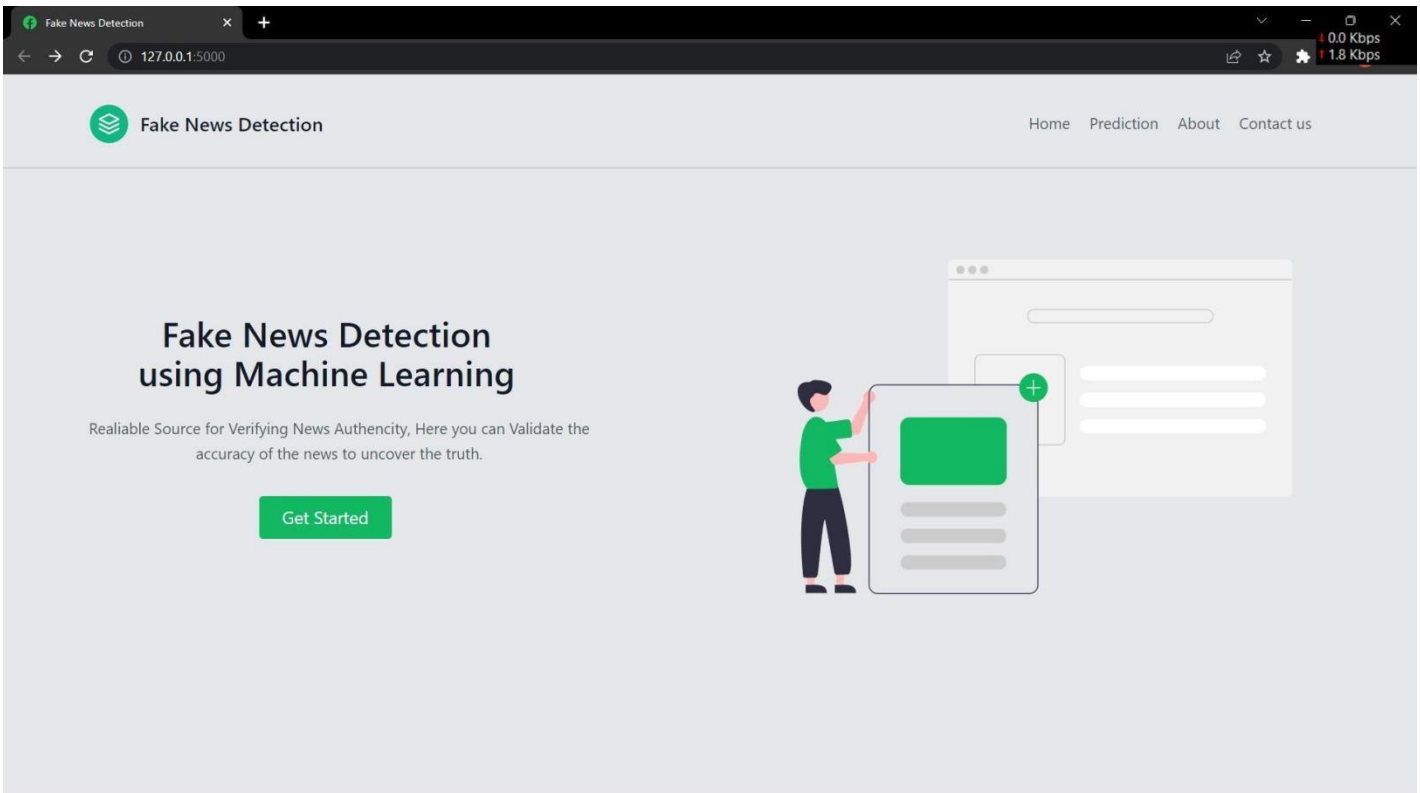
```
lemmatizer = WordNetLemmatizer() stpwrds =
set(stopwords.words('english')) corpus = []
def fake_news_det(news):
    review = news review = re.sub(r'[^a-zA-
    Z\s]', '', review) review = review.lower()
    review = nltk.word_tokenize(review) corpus
    = [] for y in review :
        if y not in stpwrds :
            corpus.append(lemmatizer.lemmatize(y))
    input_data = [' '.join(corpus)] vectorized_input_data =
    vector.transform(input_data) prediction =
    loaded_model.predict(vectorized_input_data) return
    prediction
```

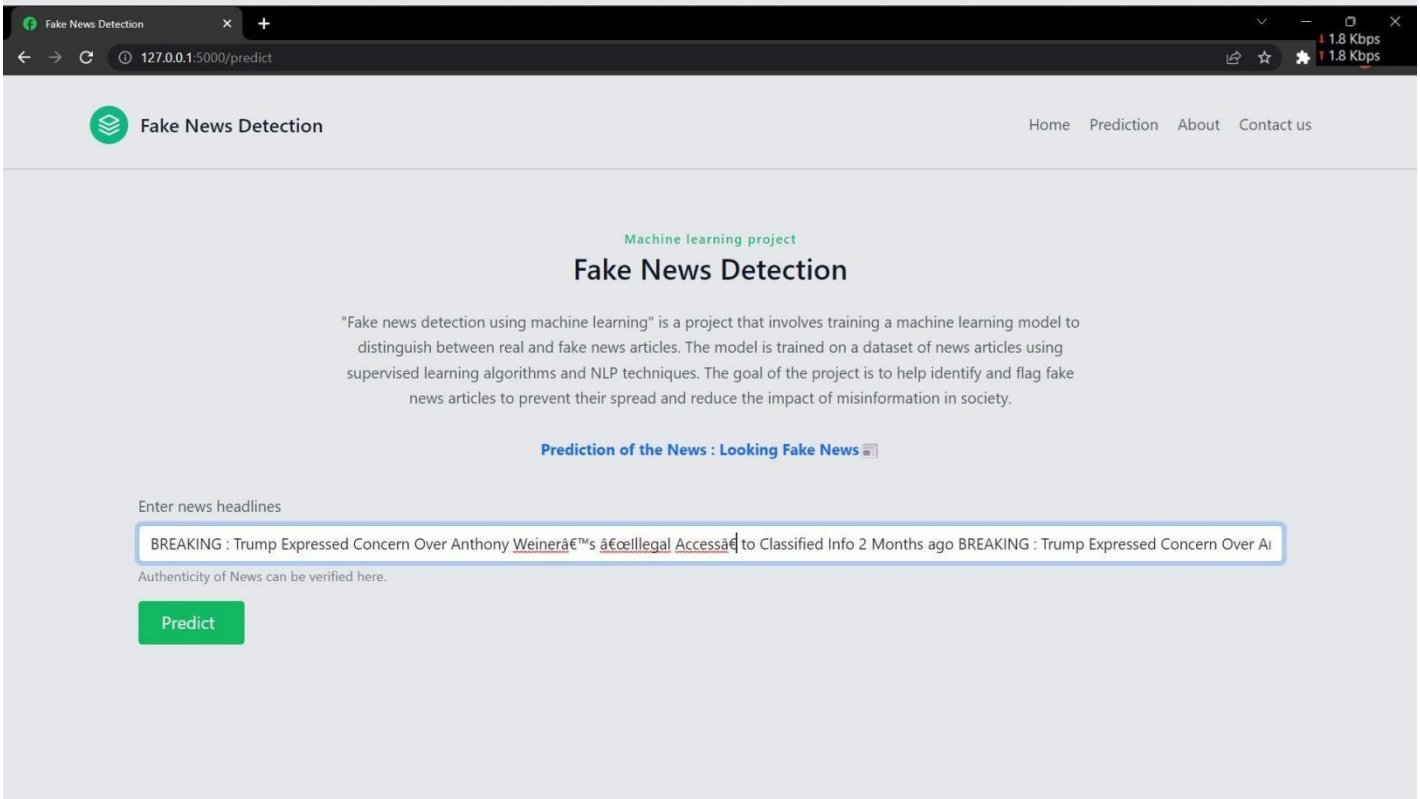
```
@app.route('/')
def home():
    return render_template('index.html')
@app.route('/predict', methods=['GET','POST'])
def predict():
    if request.method == 'POST':
        message =
        request.form['news'] pred =
        fake_news_det(message)
        print(pred) def predi(pred):
            if pred[0] == 1:
              res="Prediction of the News : Looking Fake News "
            else:
              res="Prediction of the News : Looking Real News "
            return res
        result=predi(pred) return render_template("prediction.html",
        prediction_text="{}".format(result))
    else:
        return render_template('prediction.html', prediction="Something went wrong")
if __name__ == '__main__':
    app.run(debug=True)
```

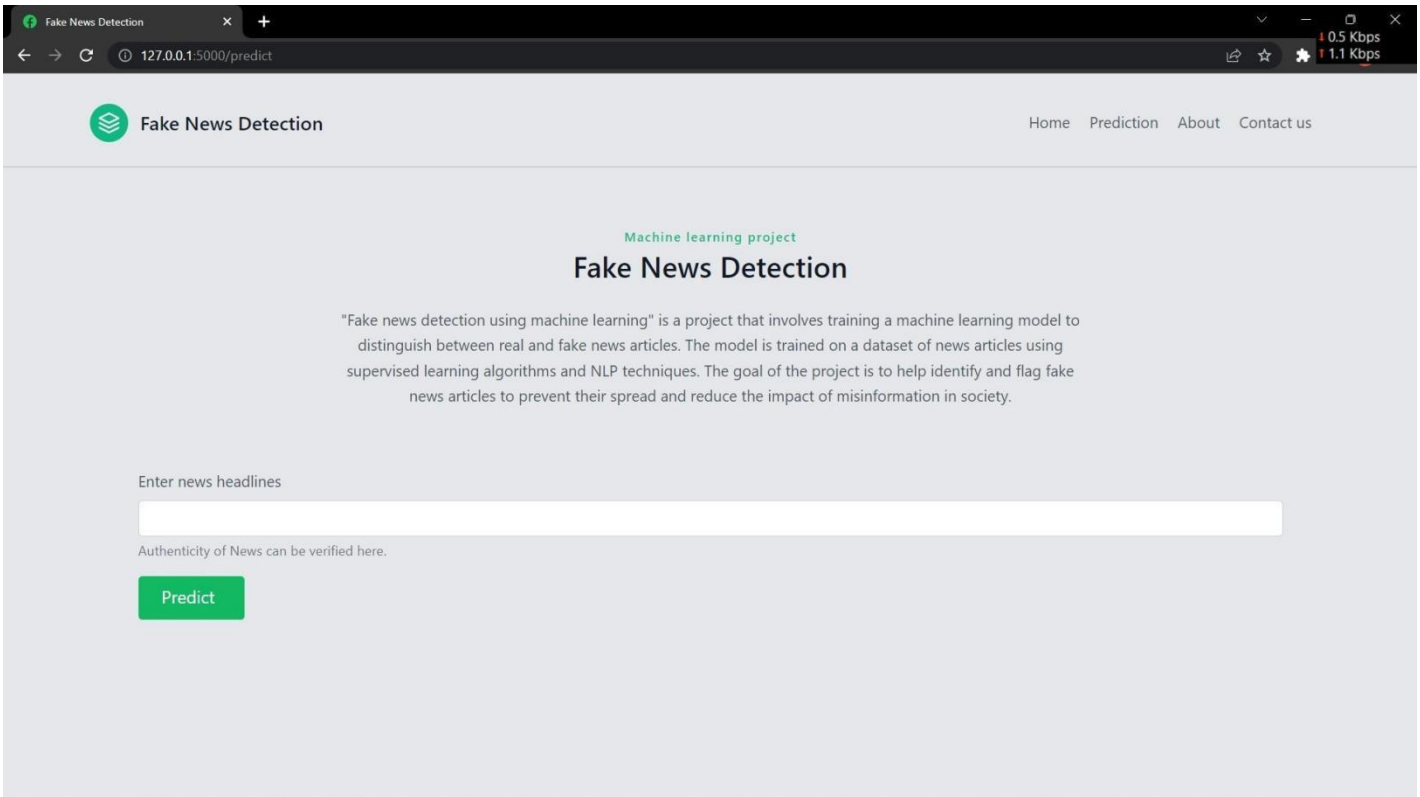
DATASET:

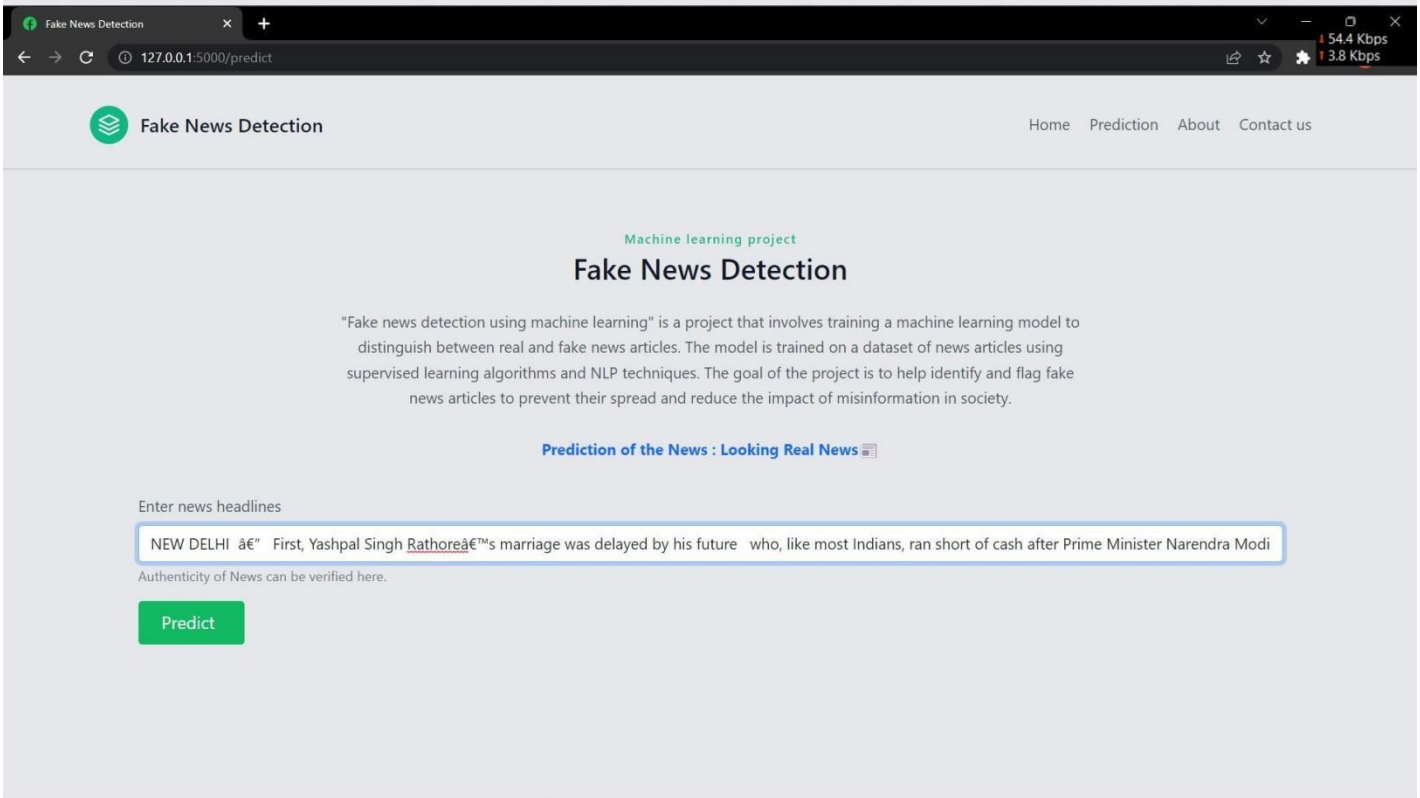
PALO ALTO, Calif. - After years of scorning the political process, Silicon Valley has leapt into the fray. The prospect of a President Donald J. Trump is pushing the tech community to move beyond its traditional role as donors and to embrace a new existence as agitators and activists. A distinguished venture capital firm emblazoned on its corporate home page an earthy epithet. One prominent tech chieftain says the consequences of Mr. Trump's election would "range between disastrous and terrible. " Another compares him to a dictator. And nearly 150 tech leaders signed an open letter decrying Mr. Trump and his campaign of "anger" and "bigotry. " Not quite all the action is . Peter Thiel, a founder of PayPal and Palantir who was the first outside investor in Facebook, spoke at the Republican convention in July. The New York Times reported on Saturday that Mr. Thiel is giving $1. 25 million to support Mr. Trump's candidacy even as other supporters flee. (He also recently gave $1 million to a "super PAC" that supports Senator Rob Portman, the Republican freshman running for in Ohio.) Getting involved in politics used to be seen as clashing with Silicon Valley's value system: You transform the world by making problems obsolete, not solving them through Washington. Nor did entrepreneurs want to alienate whatever segment of customers did not agree with them politically. Such reticence is no longer in style here. "We're a bunch of nerds not used to having a lot of limelight," said Dave McClure, an investor who runs a tech incubator called 500 Startups. "But to quote 'With great power comes great responsibility. '" Mr. McClure grew worried after the Republican and Democratic conventions as Mr. Trump began to catch up to Hillary Clinton in the polls. He wanted Silicon Valley to do more, and so late last month he announced Nerdz4Hillary, an informal effort. An initial group of donors pledged $50, 000 the goal was to ask the "nerdz" for small donations to match that sum. They have not come through yet. "We're kind of optimistic we'll get the other $50, 000 in a few weeks," Mr. McClure said. That relatively slow pace reflects Silicon Valley's shifting position: Even as it becomes increasingly free with its opinions, it has been less free with its checkbook. The most recent data, from late August, shows Mrs. Clinton taking in $7. 7 million from the tech community, according to Crowdpac, a that tracks donations. By that point in 2012, Crowdpac says, President Obama had raised $21 million from entrepreneurs and venture capitalists. Reid Hoffman, the billionaire of the business networking site LinkedIn, offers a snapshot of Silicon Valley's evolving approach to politics. Mr. Hoffman was a top Obama donor, giving $1 million to the Priorities USA political action committee, something several of his peers did as well. Last month, Mr. Hoffman garnered worldwide publicity for saying he would donate up to $5 million to veterans' groups if Mr. Trump released his taxes, a remote possibility that never came to pass. He has castigated Mr. Trump in interviews, saying he was speaking for those who were afraid. Mr. Hoffman's outright donations, however, have been smaller this election cycle. In May, he gave $400, 000 to the Hillary Victory Fund. Asked if there was more recent giving that had not shown up in federal election records, Mr. Hoffman cryptically responded in an email, "Looking at some PACs, etc. " He declined several opportunities to elaborate. Even as Priorities USA has raised $133 million this election cycle, far exceeding its total in 2012, its tech contributions have dwindled. The only familiar tech name this time around is John Doerr of the venture capital firm Kleiner Perkins Caufield Byers, who gave $500, 000. The AOL Steve Case said his September endorsement of Mrs. Clinton, via an in The Washington Post, was the first time he ever publicly declared for a candidate. "I always focused on policy and avoided politics," he said. "But if Trump were elected president, I would be disappointed in myself for not acting. " When he wrote the he was uncertain about donating money to Mrs. Clinton, saying only that it was "probable. " A spokeswoman said Sunday that Mr. Case gave $25, 000 to the Hillary Victory Fund. Mason Harrison, Crowdpac's head of communications, offered a possible reason for Mrs. Clinton's support. "Donors give to support candidates they love, not to defeat candidates they fear," he said. A few billionaires are acting instead of talking. Dustin Moskovitz, a founder of Facebook, said he was giving $20 million to various Democratic election efforts — the first time he and his wife, Cari Tuna, have endorsed a candidate. He declined to be interviewed. Part of the problem for Mrs. Clinton is that, however preferable she may be to Mr. Trump in the tech community, she pales in comparison to President Obama.

OUTPUT:


Fake News Detection
127.0.0.1:5000
0.0 Kbps
1.8 Kbps
Fake News Detection
Home
Prediction
About
Contact us
Fake News Detection using Machine Learning
Realiable Source for Verifying News Authencity, Here you can Validate the accuracy of the news to uncover the truth.
Get Started


Fake News Detection
127.0.0.1:5000/predict
1.8 Kbps
1.8 Kbps
Fake News Detection
Home
Prediction
About
Contact us
Machine learning project
Fake News Detection
"Fake news detection using machine learning" is a project that involves training a machine learning model to distinguish between real and fake news articles. The model is trained on a dataset of news articles using supervised learning algorithms and NLP techniques. The goal of the project is to help identify and flag fake news articles to prevent their spread and reduce the impact of misinformation in society.
Prediction of the News : Looking Fake News
Enter news headlines
BREAKING : Trump Expressed Concern Over Anthony Weinerâ€™s â€œIllegal Accessâ€ to Classified Info 2 Months ago BREAKING : Trump Expressed Concern Over Ai
Authenticity of News can be verified here.
Predict


Fake News Detection
127.0.0.1:5000/predict
0.5 Kbps
1.1 Kbps
Fake News Detection
Home
Prediction
About
Contact us
Machine learning project
Fake News Detection
"Fake news detection using machine learning" is a project that involves training a machine learning model to distinguish between real and fake news articles. The model is trained on a dataset of news articles using supervised learning algorithms and NLP techniques. The goal of the project is to help identify and flag fake news articles to prevent their spread and reduce the impact of misinformation in society.
Enter news headlines
Authenticity of News can be verified here.
Predict


Fake News Detection
127.0.0.1:5000/predict
54.4 Kbps
3.8 Kbps
Fake News Detection
Home
Prediction
About
Contact us
Machine learning project
Fake News Detection
"Fake news detection using machine learning" is a project that involves training a machine learning model to distinguish between real and fake news articles. The model is trained on a dataset of news articles using supervised learning algorithms and NLP techniques. The goal of the project is to help identify and flag fake news articles to prevent their spread and reduce the impact of misinformation in society.
Prediction of the News : Looking Real News
Enter news headlines
NEW DELHI â€" First, Yashpal Singh Rathoreâ€™s marriage was delayed by his future who, like most Indians, ran short of cash after Prime Minister Narendra Modi
Authenticity of News can be verified here.
Predict

CLASSIFICATION REPORT

```
[35]: print(classification_report(Y_test, Y_pred))
```

	precision	recall	f1-score	support
0	0.96	0.96	0.96	2104
1	0.96	0.96	0.96	2049
accuracy			0.96	4153
macro avg	0.96	0.96	0.96	4153
weighted avg	0.96	0.96	0.96	4153

CHAPTER 5

CONCLUSION AND FUTURE WORK

Many people consume news from social media instead of traditional news media. However, social media has also been used to spread fake news, which has negative impacts on individual people and society. In this paper, an innovative model for fake news detection using machine learning algorithms has been presented. This model takes news events as an input and based on twitter reviews and classification algorithms it predicts the percentage of news being fake or real.

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

REFERENCES

[1]. Parikh, S. B., & Atrey, P. K. (2018, April). Media-Rich Fake News Detection: A Survey. In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (pp. 436-441). IEEE.

[2]. Conroy, N. J., Rubin, V. L., & Chen, Y. (2015, November). Automatic deception detection: Methods for finding fake news. In Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community (p. 82). American Society for Information Science.

[3]. Helmstetter, S., & Paulheim, H. (2018, August). Weakly supervised learning for fake news detection on Twitter. In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 274-277). IEEE.

[4]. Stahl, K. (2018). Fake News Detection in Social Media.

[5]. Della Vedova, M. L., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M., & de Alfaro, L. (2018, May). Automatic Online Fake News Detection Combining Content and Social Signals. In 2018 22nd Conference of Open Innovations Association (FRUCT) (pp. 272-279). IEEE.

[6] Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., & de Alfaro, L. (2017). Some like it hoax: Automated fake news detection in social networks. arXiv preprint arXiv:1704.07506.

[7]. Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., & Menczer, F. (2017). The spread of fake news by social bots. arXiv preprint arXiv:1707.07592, 96-104.

[8]. Chen, Y., Conroy, N. J., & Rubin, V. L. (2015, November). Misleading online content: Recognizing clickbait as false news. In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection (pp. 15-19). ACM.

[9]. Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. Journal of Big Data, 2(1), 1.

[10]. Haiden, L., & Althuis, J. (2018). The Definitional Challenges of Fake News.

[11]. [1] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, Fake News Detection on Social Media: A Data Mining Perspective‖ arXiv:1708.01967v3 [cs.SI], 3 Sep 2017

[12]. Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, ―Fake News Detection on Social Media: A Data Mining Perspective‖ arXiv:1708.01967v3 [cs.SI], 3 Sep 2017

[13]. M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev, 2017, pp. 900-903.

[14]. Fake news websites. (n.d.) Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Fake_news_website. Accessed Feb. 6, 2017

[15]. Cade Metz. (2016, Dec. 16). The bittersweet sweepstakes to build an AI that destroys fake news.

[16]. Conroy, N., Rubin, V. and Chen, Y. (2015). Automatic deception detection: Methods for finding fake news‖ at Proceedings of the Association for Information Science and Technology, 52(1), pp.1-4.

[17]. Markines, B., Cattuto, C., & Menczer, F. (2009, April). ―Social spam detection‖. In Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web (pp. 41-48)

[18]. Rada Mihalcea , Carlo Strapparava, The lie detector: explorations in the automatic recognition of deceptive language, Proceedings of the ACL-IJCNLP

[19]. Kushal Agarwalla, Shubham Nandan, Varun Anil Nair, D. Deva Hema, ―Fake News Detection using Machine Learning and Natural Language Processing,‖ International Journal of Recent Technology andEngineering (IJRTE)
ISSN: 2277-3878, Volume-7, issue-6, March 2019
[20]. H. Gupta, M. S. Jamal, S. Madisetty and M. S. Desarkar, "A framework for real-time spam detection in Twitter," 2018 10th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru, 2018, pp. 380383

[21]. M. L. Della Vedova, E. Tacchini, S. Moret, G. Ballarin, M. DiPierro and L. de Alfaro, "Automatic Online Fake News Detection Combining Content and [22]. Social Signals," 2018 22nd Conference of Open Innovations Association (FRUCT), Jyvaskyla, 2018, pp. 272-279.

[23]. C. Buntain and J. Golbeck, "Automatically Identifying Fake News in Popular Twitter Threads," 2017 IEEE International Conference on Smart Cloud (SmartCloud), New York, NY, 2017, pp. 208-215.

[24]. S. B. Parikh and P. K. Atrey, "Media-Rich Fake News Detection: A Survey," 2018 IEEE Conference on Multimedia Information Processing and Retrieval
(MIPR), Miami, FL, 2018, pp. 436-44

[25]. Scikit-Learn- Machine Learning In Python

[26]. Dataset- Fake News detection William Yang Wang. " liar, liar pants on _re": A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648, 2017.

[27]. Shankar M. Patil, Dr. Praveen Kumar, ―Data mining model for effective data analysis of higher education students using MapReduce‖ IJERMT, April 2017 (Volume-6, Issue-4).

[28]. Aayush Ranjan, ― Fake News Detection Using Machine Learning‖, Department Of Computer Science & Engineering Delhi Technological University, July 2018.
[29]. Patil S.M., Malik A.K. (2019) Correlation Based Real-Time Data Analysis of Graduate Students Behaviour. In: Santosh K., Hegadi R. (eds) Recent Trends in Image Processing and Pattern Recognition. RTIP2R 2018. Communications in Computer and Information Science, vol 1037. Springer, Singapore.

[30]. Badreesh Shetty, “Natural Language Processing (NLP) for machine learning‖ at towardsdatascience, Medium.