

FAKE NEWS ANALYSIS IN SOCIAL MEDIA USING IBM WATSON

An Industrial/Practical Training Report

Submitted to the Faculty of Engineering of
**JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA,
KAKINADA**

In partial fulfillment of the requirements for the award of the Degree of

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

CH. SAI KUMAR
(19481A0542)

CH. SAJAN
(19481A0534)

B. HIMA KRISHNA
(19481A0516)

B. BALA SAI
(19481A0520)

Under the Enviabale and Esteemed Guidance of
Mrs. P. Reshma, M.Tech
Assistant Professor, Department of CSE



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SESHADRI RAO GUDLAVALLERU ENGINEERING COLLEGE

(An Autonomous Institute with Permanent Affiliation to JNTUK, Kakinada)

SESHADRI RAO KNOWLEDGE VILLAGE

GUDLAVALLERU – 521356

ANDHRA PRADESH

2022-23

SESHADRI RAO
GUDLAVALLERU ENGINEERING COLLEGE
(An Autonomous Institute with Permanent Affiliation to JNTUK, Kakinada)
SESHADRI RAO KNOWLEDGE VILLAGE, GUDLAVALLERU

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING



CERTIFICATE

This is to certify that the project report entitled “**FAKE NEWS ANALYSIS IN SOCIAL MEDIA Using IBM Watson**” is a bonafide record of work carried out by CH.SAI KUMAR (19481A0542), CH.SAJAN(19481A0534), B.HIMA KRISHNA (19481A0516), B. BALA SAI (19481A0520) under the guidance and supervision of **Mrs. P. Reshma**, M.Tech in the partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Computer Science and Engineering of Jawaharlal Nehru Technological University Kakinada, Kakinada** during the academic year 2022-23.

Project Guide
(P. Reshma)

Head of the Department
(Dr. M. Babu Rao)

External Examiner

ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion of any task would be incomplete without the mention of people who made it possible and whose constant guidance and encouragements crown all the efforts with success.

We would like to express our deep sense of gratitude and sincere thanks to **Mrs. P. Reshma, M.Tech**, Assistant Professor, Department of Computer Science and Engineering for her constant guidance, supervision and motivation in completing the project work.

We feel elated to express our floral gratitude and sincere thanks to **Dr. M. Babu Rao**, Head of the Department, Computer Science and Engineering for his encouragements all the way during analysis of the project. His annotations, insinuations and criticisms are the key behind the successful completion of the project work.

We would like to take this opportunity to thank our beloved principal **Dr. G. V. S. N. R. V. Prasad** for providing a great support for us in completing our project and giving us the opportunity for doing project.

Our Special thanks to the faculty of our department and programmers of our computer lab. Finally, we thank our family members, non-teaching staff and our friends, who had directly or indirectly helped and supported us in completing our project in time.

Team members

Ch. Sai Kumar (19481A0542)
Ch. Sajan (19481A0534)
B. Hima Krishna (19481A0516)
B. Bala Sai (19481A0520)



INTERNSHIP REPORT APPROVAL FORM

Date

With immense pleasure, this is to approve that the students of Seshadri Rao Gudlavalleru Engineering College i.e.,

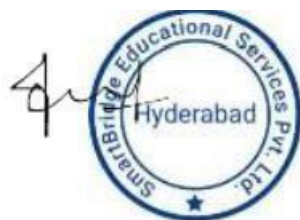
CH.SAI KUMAR (19481A0542), CH.SAJAN(19481A0534), B.HIMA KRISHNA (19481A0516), B.BALA SAI(19481A0520)

successfully completed their Project and Project Report on **“FAKE NEWS ANALYSIS IN SOCIAL MEDIA USING IBM WATSON”** under our guidance.

We are highly impressed with the work that they have done and commend them on their quick grasping skills. They have shown good intent to learn and have put the knowledge gained into application in the form of this project. We appreciate the hard work and commitment shown by them.

We, hereby approve that this document is completely checked and accepted by SmartBridge Technical Team. It's been an absolute pleasure to educate and mentor these students. We hope that this document will also serve as a Letter of Recommendation, to whomsoever applied.

We wish them success in all future endeavors and a great career ahead.\\



Jayaprakash.ch,
Program Manager

ABSTRACT

In our modern era where the internet is being everywhere at the same time, everyone relies on various online resources for news. Along with the increase in the use of social media platforms like Facebook, Twitter, etc. News spread rapidly among millions of users within a very short span of time. The spread of fake news has far-reaching consequences like the creation of biased opinions.

Moreover, spammers use appealing news headlines to generate revenue using Advertisements. We aim to provide the user with the ability to classify the news as fake or real.

Internet and social media have made the access to the news information much easier and comfortable. Often Internet users can pursue the events of their concern in online form, and increased number of the mobile devices makes this process even easier. But with great possibilities come great challenges. Mass media have an enormous influence on the society, and because it often happens, there's someone who wants to require advantage of this fact. Sometimes to realize some goals mass-media may manipulate the knowledge in several ways. This results in producing of the news articles that isn't completely true or maybe completely false. There even exist many websites that produce fake news almost exclusively.

They intentionally publish hoaxes, half-truths, propaganda and disinformation asserting to be real news – often using social media to drive web traffic and magnify their effect. The most goals of faux news websites are to affect the general public opinion on certain matters (mostly political). Samples of such websites could also be found in Ukraine, United States of America, Germany, China and much of other countries. Thus, fake news may be a global issue also as a worldwide challenge. Many scientists believe that fake news issue could also be addressed by means of machine learning and AI.

INDEX

TITLE	PAGENO
CHAPTER 1: INTRODUCTION	
1.1 INTRODUCTION	1
1.2 OBJECTIVES OF THE PROJECT	2
1.3 PROBLEM STATEMENT	2
CHAPTER 2: LITERATURE REVIEW	4
CHAPTER 3: REQUIREMENT ANALYSIS	
3.1 FUNCTIONAL REQUIREMENTS	5
3.2 NON-FUNCTIONAL REQUIREMENTS	5
3.3 SOFTWARE REQUIREMENT SPECIFICATION	5
3.4 HARDWARE SPECIFICATION	5
CHAPTER 4: PROPOSED METHOD	
4.1 METHODOLOGY	6
4.2 DATA PREPARATION	9
4.3 IMPLEMENTATION	17
4.4 ADVANTAGES	25
4.5 DISADVANTAGES	25
CHAPTER 5: RESULTS	
5.1 RESULT	26
CHAPTER 6: CONCLUSION	31
REFERENCES	32

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Fake News contains misleading information that could be checked. This maintains lie about a certain statistic in a country or exaggerated cost of certain services for a country, which may arise unrest for some countries like in Arabic spring. There are organizations, like the House of Commons and the Crosscheck project, trying to deal with issues as confirming authors are accountable.

However, their scope is so limited because they depend on human manual detection, in a globe with millions of articles either removed or being published every minute, this cannot be accountable or feasible manually. A solution could be, by the development of a system to provide a credible automated index scoring, or rating for credibility of different publishers, and news context. This paper proposes a methodology to create a model that will detect if an article is authentic or fake based on its words, phrases, sources and titles, by applying supervised machine learning algorithms on an annotated (labeled) dataset, that are manually classified and guaranteed. Then, feature selection methods are applied to experiment and choose the best fit features to obtain the highest precision, according to confusion matrix results.

We propose to create the model using different classification algorithms. The product model will test the unseen data, the results will be plotted, and accordingly, the product will be a model that detects and classifies fake articles and can be used and integrated with any system for future use.

1.2 PROJECT OBJECTIVES

- Know fundamental concepts of NLP(Natural Language Processing).
- Gain a broad understanding of data set.
- Know how to pre-process/clean the data using different data pre- processing techniques.
- Know how to build a web application using Flask framework.

1.3 PROBLEM STATEMENT

Fake news detection is a subtask of text classification and is often defined as the task of classifying news as real or fake. The term „fake news“ refers to the false or misleading information that appears as real news. It aims to deceive or mislead people. Fake news comes in many forms, such as clickbait (misleading headlines), disinformation (with malicious intention to mislead the public), misinformation (false information regardless of the motive behind), hoax, parody, satire, rumour, deceptive news.

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.

By using Natural Language Processing(NLP) we can predict the news "real" or "fake".

CHAPTER-2

LITERATURE REVIEW

2.1 Existing problem:

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.

Conroy, Rubin, and Chen ^[1] 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 (POS) 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.

2.2 Proposed solution:

By using NLP we can identify the news “fake” or “real”. In this paper a model is build based on the count vectorizer or a tfidf matrix (i.e) word tallies relatives to how often they are used in other artices in your dataset) can help. The actual goal is in developing a model which was the text transformation (count vectorizer vs tfidf 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 countvectorizer or tfidf-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”.

CHAPTER 3

REQUIREMENTS ANALYSIS

3.1 FUNCTIONAL REQUIREMENTS

- News dataset
- Input data : Title, Label.

3.2 NON FUNTIONAL REQUIREMENTS

- High Performance
- Highly Reliability
- High Maintainability
- Less Usability

3.3 SOFTWARE REQUIREMENT SPECIFICATIONS

- Anaconda3
- Jupiter Notebook
- Spyder
- Any Browser
- Languages – Python

3.4 HARDWARE SPECIFICATIONS

- Processor – i5
- RAM – 8GB (minimum)

CHAPTER 4

PROPOSED METHOD

4.1 METHODOLOGY:

This literature review is written for answering some research questions. So the methodology that is used is the systematic literature review. This methodology helps in answering the research questions. The papers were collected from various databases to be discussed in this literature review. **Machines and especially computer systems can simulate human intelligence processes (AI), enabling it to mimic even human behavior.** The software finds application in areas such as Computer Vision, Natural Language Processing, Robotics, Speech Recognition, and much.

Natural Language Processing:

Natural Language Processing (NLP) is a sub-branch of linguistics, computer science, data engineering, and artificial intelligence. NLP relates to the interaction between humans and computers. NLP is a method for processing and analyzing large amounts of natural language data. NLP has many applications such as machine translation, speech recognition, sentiment analysis, automatic question and answer generation, automatic message digest, chatbot, intelligence, text classification.

In the NLP, one crucial step is text extraction, a preprocessing step for using the analysis of text, documents, news, and information before implementing the clustering, classification, or other machine learning tasks . The fundamental preprocessing step for NLP includes word segmentation, tokenization, word stopping, word stemming, term frequency weighting, term frequency, and inverse document frequency weighting. Some advanced NLP techniques may include more complex tasks in the pipeline, such as parts of speech tagging, dependency parsing, named entity recognition, and conference

resolution . Advanced NLP techniques may employ lexical analysis, syntactic analysis, semantic analysis, disclosure integration, and pragmatic analysis.

Naïve Bayesian (NB) is an easy learning probabilistic-based algorithm that uses Bayes' rule in conjunction with the explicit assumption that attributes are conditionally independent of each other. Based on training data, NB estimates the posterior probability $P(y|x)$ of each class, y , of a given object, x . We can use the estimation for classification applications. Because of its computational efficiency and many other desirable properties, NB appears as an acceptable solution in many practical implementation.

Technical Architecture:

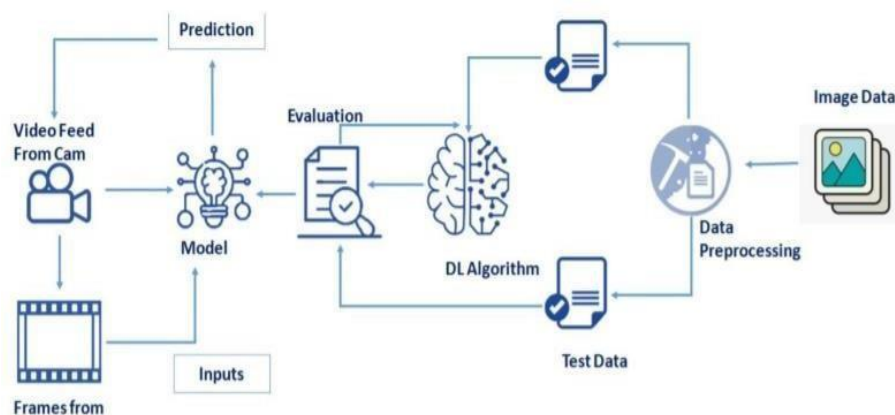


Fig:1

Block diagram:

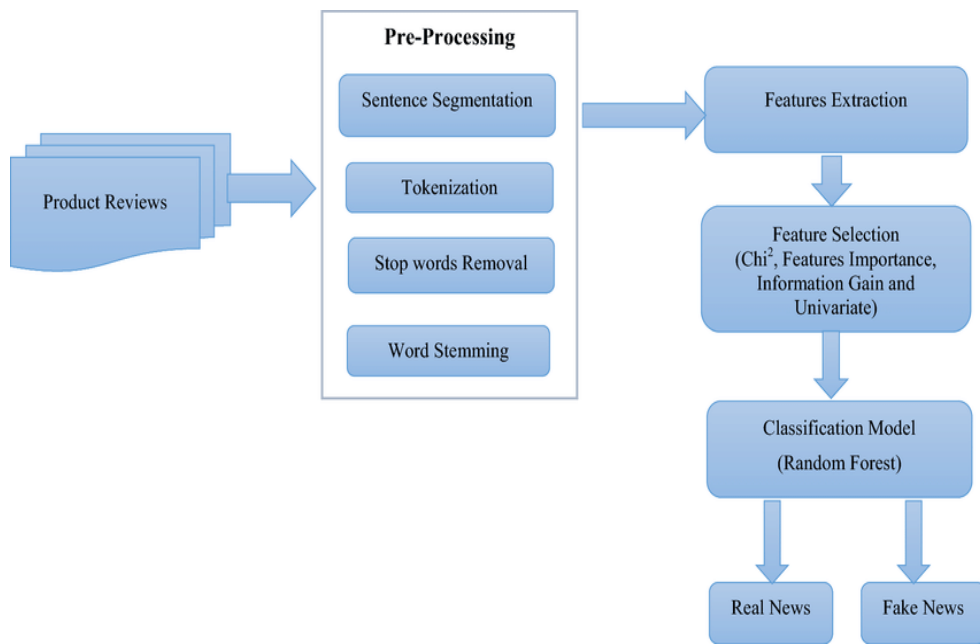
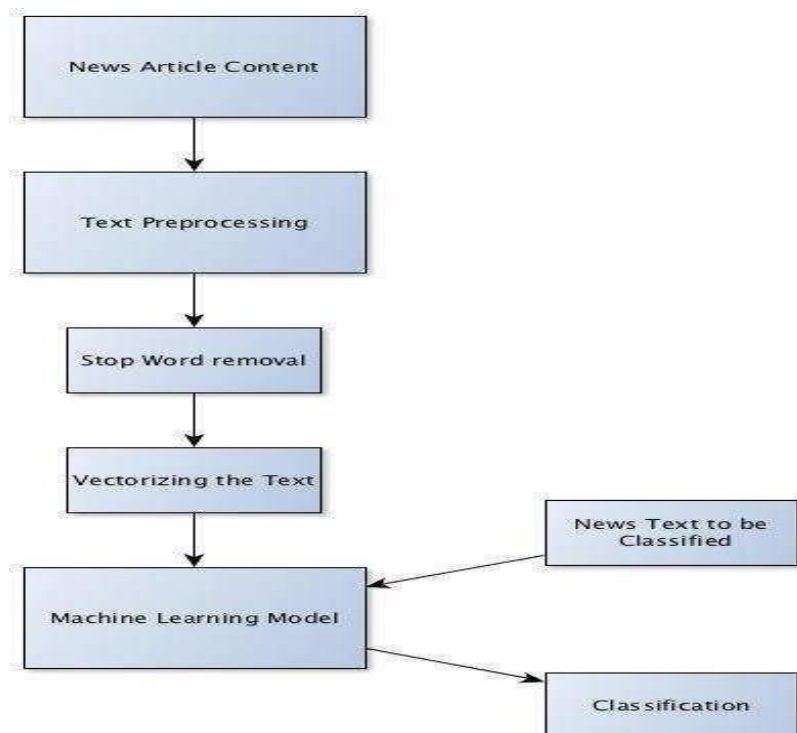


Fig:2

FLOWCHART:



4.2 DATA PREPARATION

4.2.1 Data Preprocessing

Data Pre-processing includes the following main tasks

- Lower case text
- Tokenization
- Stemming
- Applying ImageDataGenerator functionality to the trainset and test set.
- Stop word removal.
- Download the Dataset.
- Importing the required Libraries.
- Loading the dataset.
- Countvectorizer for text classification.
- TF-IDF vectorizer for text classification.
- Inspecting the vectors.

Importing the Libraries

- Import the downloaded data set by using pandas library.
- To visualize the data using Seaborn/Matplotlib.
- Print the shape of the dataset.
- Import the Numpy Library for Multidimensional Arrays.
- Import the NLTK Library for news classification.
- To remove the Stoplist words, Punctuation and Tokenization.
- Import the Count_vectorizer for collection of text documents to a vector of token counts.
- Import the Tfidfvectorizer for Text classification

4.2.2 Model Building

In Natural Language Processing by using Naive Bayes classifier.

In this milestone, we start building our model by:

- Splitting the data into train and test.
- Training and testing the model with Count_vectorizer & predicting the result.
- Training and testing the model TF_IDF vectorizer & predicting the result.
- Improving the model.
- Inspecting the model.
- Saving the model.

Split the data into train and test.

Here, we are reading the dataset(.csv) from the system using pandas and storing it in a variable „df“. It's time to begin building your text classifier! The data has been loaded into a DataFrame called df. The .head() method is particularly informative.

Divide the model into Train and Test data by Dependent and Independent Columns. Here in the dataset we need to separate the dependent and independent variables.

1. The independent variable in the dataset would be considered as 'x'
2. The dependent variable in the dataset would be considered as 'y'

Then we will split the data of independent and dependent variables.

```
df = pd.read_csv('news.csv')
#print the first several row of data
df.head()
```

	Unnamed: 0		title	text	label
0	8476	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fello...	FAKE	
1	10294	Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg LinkedIn Reddit Stumbleu...	FAKE	
2	3608	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	REAL	
3	10142	Bernie supporters on Twitter erupt in anger ag...	— Kaydee King (@KaydeeKing) November 9, 2016 T...	FAKE	
4	875	The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners...	REAL	


```
# Create training set and test set
X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                    test_size=0.33, random_state=53)
```

You can see your data set has three columns: title, text, and label. We make use of Text and label columns to build the classifier.

Training the model with Count_vectorizer.

You can see your data set has three columns: title, text, and label. We make use of Text and label columns to build the classifier.

Machines cannot understand characters and words. So when dealing with text data we need to represent it in numbers to be understood by the machine. Countvectorizer is a method to convert text to numerical data. To show you how it works let's take an example:

```
text = ['Hello my name is james, this is my python notebook']
```

The text is transformed to a sparse matrix as shown below.

	hello	is	james	my	name	notebook	python	this
0	1	2	1	2	1	1	1	1

```

from sklearn.naive_bayes import MultinomialNB

from sklearn.metrics import accuracy_score, confusion_matrix

# Instantiate a Multinomial Naive Bayes classifier: nb_classifier
nb_classifier = MultinomialNB()

# Fit the classifier to the training data
nb_classifier.fit(count_train, y_train)

# Create the predicted tags: pred
pred = nb_classifier.predict(count_test)

# Calculate the accuracy score: score
score = accuracy_score(y_test, pred)
print(score)

# Calculate the confusion matrix: cm
cm = confusion_matrix(y_test, pred, labels=['FAKE', 'REAL'])
print(cm)

```

0.893352462936394
[[865 143]
[80 1003]]

Training model with Tfidf Vectorizer

- Similar to the sparse CountVectorizer created in the previous step, we'll work on creating tf-idf vectors for your documents. You'll set up a TfidfVectorizer and learn some of its features.
- Machine learning algorithms operate on a numeric feature space, expecting input as a two-dimensional array where rows are instances and columns are features.
- In order to perform machine learning on text, we need to transform our documents into vector representations such that we can apply numeric machine learning.
- This process is called *feature extraction* or more simply, *vectorization*, and is an essential first step toward language-aware analysis.

- Not only will many of the same words appear in both, they will not share many words in common with articles about casseroles or quantitative easing. This model, while simple, is extremely effective and forms the starting point for the more complex models we will explore.
- TFIDF works by proportionally increasing the number of times a word appears in the document but is counterbalanced by the number of documents in which it is present.
- Hence, words like „this“, “are” etc., that are commonly present in all the documents are not given a very high rank.

```
from sklearn.feature_extraction.text import TfidfVectorizer

# Initialize a TfidfVectorizer object: tfidf_vectorizer
tfidf_vectorizer = TfidfVectorizer(stop_words='english', max_df=0.7)

# Transform the training data: tfidf_train
tfidf_train = tfidf_vectorizer.fit_transform(X_train)

# transform the test data: tfidf_test
tfidf_test = tfidf_vectorizer.transform(X_test)

# Print the first 10 features
print(tfidf_vectorizer.get_feature_names()[:10])

# Print the first 5 vectors of the tfidf training data
print(tfidf_train.A[:5])
```

```
['00', '000', '0000', '00000031', '000035', '00006', '0001', '0001pt', '000ft', '000km']
[[0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
```

```
0.8565279770444764
```

```
[[ 739  269]  
 [  31 1052]]
```

```
#plot the confusion matrix for tf-idf vectorizer  
import seaborn as sns  
sns.heatmap(cm, annot=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x286b09e21c8>
```

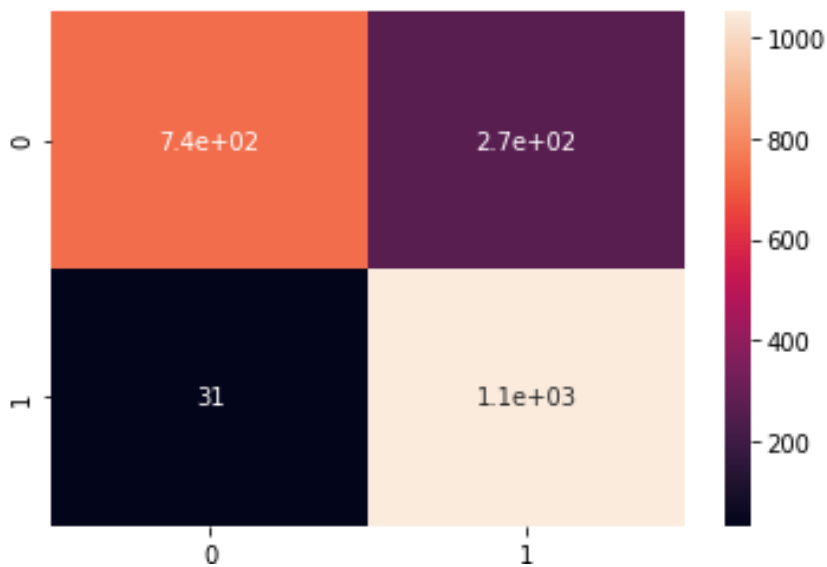


Fig:4

Improving the model

Improving the model. Our job is to test a few different alpha levels using the Tf-Idf vectors to determine if there is a better-performing combination.

```
alphas = np.arange(0, 1, 0.1)

# Define train_and_predict()
def train_and_predict(alpha):
    # Instantiate the classifier: nb_classifier
    nb_classifier = MultinomialNB(alpha=alpha)

    # Fit to the training data
    nb_classifier.fit(tfidf_train, y_train)

    # Predict the Labels: pred
    pred = nb_classifier.predict(tfidf_test)

    # Compute accuracy: score
    score = accuracy_score(y_test, pred)
    return score

# Iterate over the alphas and print the corresponding score
for alpha in alphas:
    print('Alpha: ', alpha)
    print('Score: ', train_and_predict(alpha))
    print()
```

Output of Improving model:

Alpha: 0.0

C:\Users\Shivam\anaconda3\lib\site-packages\sklearn\naive_bayes.py:512: UserWarning: alpha too small will result in numeric errors, setting alpha = 1.0e-10
'setting alpha = %.1e' % _ALPHA_MIN)

Score: 0.8813964610234337

Alpha: 0.1
Score: 0.8976566236250598

Alpha: 0.2
Score: 0.8938307030129125

Alpha: 0.30000000000000004
Score: 0.8900047824007652

Alpha: 0.4
Score: 0.8857006217120995

Alpha: 0.5
Score: 0.8842659014825442

Alpha: 0.6000000000000001
Score: 0.874701099952176

Alpha: 0.7000000000000001
Score: 0.8703969392635102

Alpha: 0.8
Score: 0.8660927785748446

Alpha: 0.9
Score: 0.8589191774270684

Inspecting the model

Now that we have built a "fake news" classifier, we'll investigate what it has learned. We can map the important vector weights back to actual words using some simple inspection techniques.

```
class_labels = nb_classifier.classes_

# Extract the features: feature_names
feature_names = tfidf_vectorizer.get_feature_names()

# Zip the feature names together with the coefficient array
# and sort by weights: feat_with_weights
feat_with_weights = sorted(zip(nb_classifier.coef_[0], feature_names))

# Print the first class label and the top 20 feat_with_weights entries
print(class_labels[0], feat_with_weights[:20])

# Print the second class label and the bottom 20 feat_with_weights entries
print(class_labels[1], feat_with_weights[-20:])
```

FAKE [(-11.316312804238807, '0000'), (-11.316312804238807, '000035'), (-11.316312804238807, '0001'), (-11.316312804238807, '0001pt'), (-11.316312804238807, '000km'), (-11.316312804238807, '0011'), (-11.316312804238807, '006s'), (-11.316312804238807, '007'), (-11.316312804238807, '007s'), (-11.316312804238807, '008s'), (-11.316312804238807, '0099'), (-11.316312804238807, '00am'), (-11.316312804238807, '00p'), (-11.316312804238807, '00pm'), (-11.316312804238807, '014'), (-11.316312804238807, '015'), (-11.316312804238807, '018'), (-11.316312804238807, '01am'), (-11.316312804238807, '020'), (-11.316312804238807, '023')]

REAL [(-7.742481952533027, 'states'), (-7.717550034444668, 'rubio'), (-7.703583809227384, 'voters'), (-7.654774992495461, 'house'), (-7.649398936153309, 'republicans'), (-7.6246184189367, 'bush'), (-7.616556675728881, 'percent'), (-7.545789237823644, 'people'), (-7.516447881078008, 'new'), (-7.448027933291952, 'party'), (-7.411148410203476, 'cruz'), (-7.410910239085596, 'stat'), (-7.35748985914622, 'republican'), (-7.33649923948987, 'campaign'), (-7.2854057032685775, 'president'), (-7.2166878130917755, 'sanders'), (-7.108263114902301, 'obama'), (-6.724771332488041, 'clinton'), (-6.5653954389926845, 'said'), (-6.328486029596207, 'trump')]

Saving the model

After building the model we have to save the model.

Pickle in Python is primarily used in serializing and deserializing a Python object structure. In other words, it's the process of converting a Python object into a byte stream to store it in a file/database, maintain program state across sessions, or transport data over the network. wb indicates write method and rd indicates read method.

This is done by the below code

```
: #saving the model
import pickle
pickle.dump(nb_classifier,open('model.pkl','wb'))
```

4.3 IMPLEMENTATION

TECHNOLOGY DESCRIPTION

ANACONDA SOFTWARE:

- Anaconda Individual Edition contains conda and Anaconda Navigator, as well as Python and hundreds of scientific packages. When you installed Anaconda, you installed all these too.
- Conda works on your command line interface such as AnacondaPrompt on Windows and terminal on macOS and Linux.
- Navigator is a desktop graphical user interface that allows you to launch applications and easily manage conda packages, environments, and channels without using command-line commands.
- You can try both conda and Navigator to see which is right for you to manage your packages and environments. You can even switch between them, and the work you do with one can be viewed in the other.

JUPYTER NOTEBOOK:

- The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. Its uses include data cleaning and transformation, numerical simulation, statistical modelling, data visualization, machine learning, and much more.
- According to the official website of Jupyter, Project Jupyter exists to develop open-source software, open-standards, and services across dozens of programming languages.

INSTALLATION STEPS

Download the Install Anaconda

- Go to your Downloads folder and double-click the installer to launch. To prevent permission errors, do not launch the installer from the Favorites folder. Click **Next**.



Fig:5



- Read the licensing terms and click **I Agree**.

- It is recommended that you install for **Just Me**, which will install Anaconda Distribution to just the current user account. Only select an install for **All Users** if you need to install for all users' accounts on the computer (which requires Windows Administrator privileges).

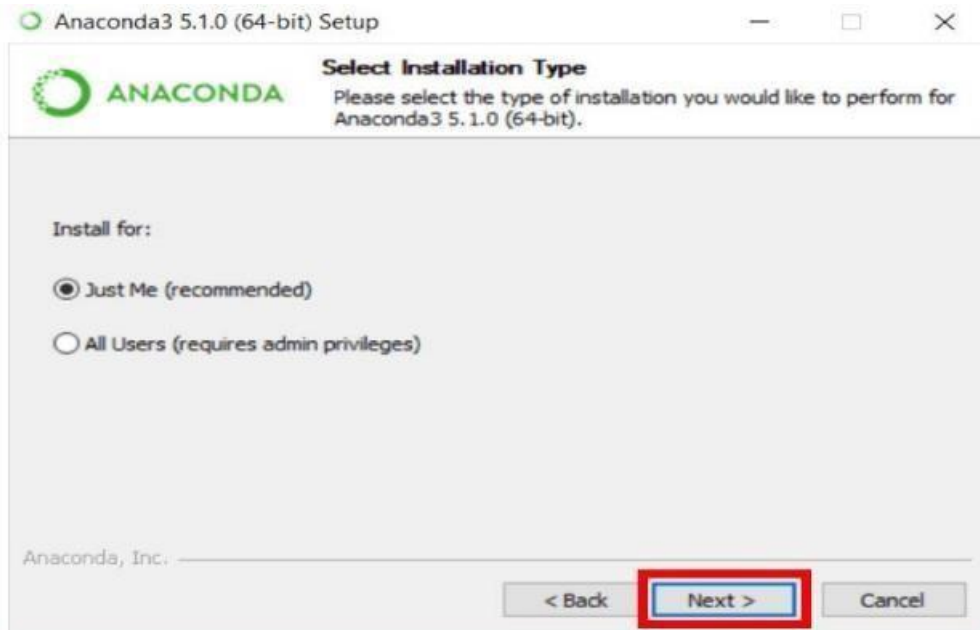


Fig : 6

- Click **Next**.
- Select a destination folder to install Anaconda and click **Next**. Install Anaconda to a directory path that does not contain spaces or uni-code characters. For more information on destination folders, see the FAQ.

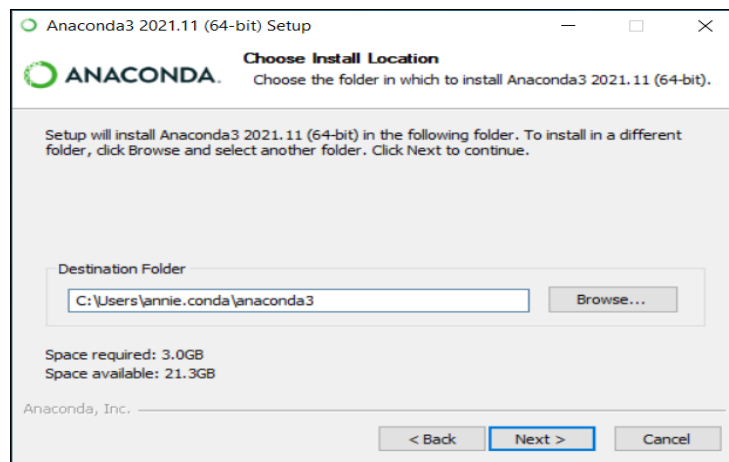


Fig : 7

- Choose whether to add Anaconda to your PATH environment variable or register Anaconda as your default Python. We **don't recommend** adding Anaconda to your PATH environment variable, since this can interfere with other software. Unless you plan on installing and running multiple versions of Anaconda or multiple versions of Python, accept the default and leave this box checked.
- Instead, use Anaconda software by opening Anaconda Navigator or the Anaconda Prompt from the Start Menu.



Fig:8

- Click **Install**. If you want to watch the packages Anaconda is installing, click Show Details.
- Click **Next**.
- Optional: To install Dataspell for Anaconda, click <https://www.anaconda.com/dataspell>.

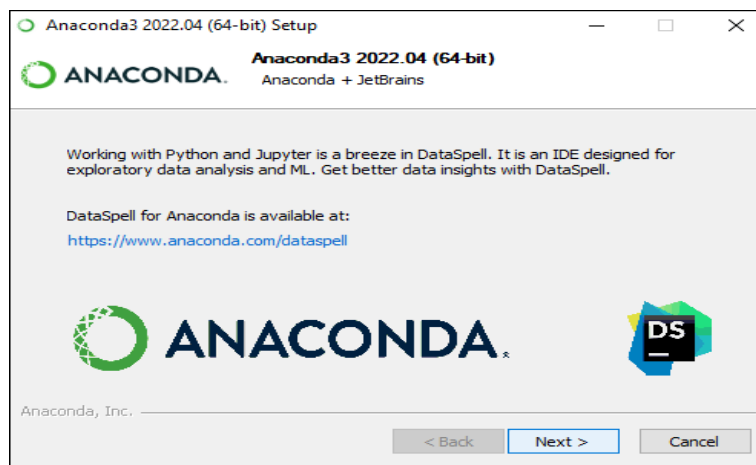


Fig:9

- After a successful installation you will see the “Thanks for installing Anaconda” dialog box:

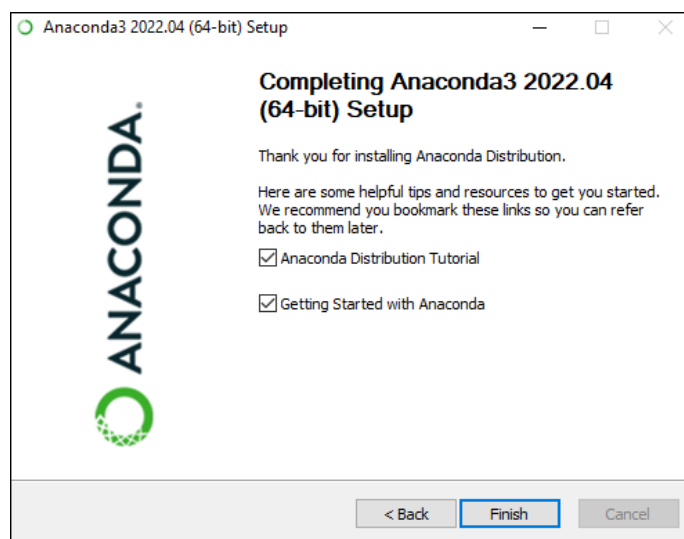


Fig:10

- If you wish to read more about Anaconda.org and how to get started with Anaconda, check the boxes “Anaconda Distribution Tutorial” and “Learn more about Anaconda”. Click the **Finish** button.

PROCEDURE FOR EXECUTION

1. Launch a Notebook To launch a Jupyter notebook, open your terminal and navigate to the directory where you would like to save your notebook. Then type the command `jupyter notebook` and the program will instantiate a local server at `localhost:8888` (or another specified port).

2. Jupyter Interface Now you're in the Jupyter Notebook interface, and you can see all of the files in your current directory. All Jupyter Notebooks are identifiable by the notebook icon next to their name. If you already have a Jupyter Notebook in your current directory that you want to view, find it in your files list and click it to open.

EXECUTION:

Training the Model

At this point, we have training data and a fully configured neural network to train. All that is left is to pass the data to the model for the training process to commence, a process that is completed by iterating on the training data.

Training begins by calling the `fit ()` method.

Save the model

The model is to be saved for future purposes. This saved model also is integrated with an android application or web application in order to predict something.

Predictions:

The last and final step is to make use of our saved model to do predictions.

Then copy and paste the link in the website

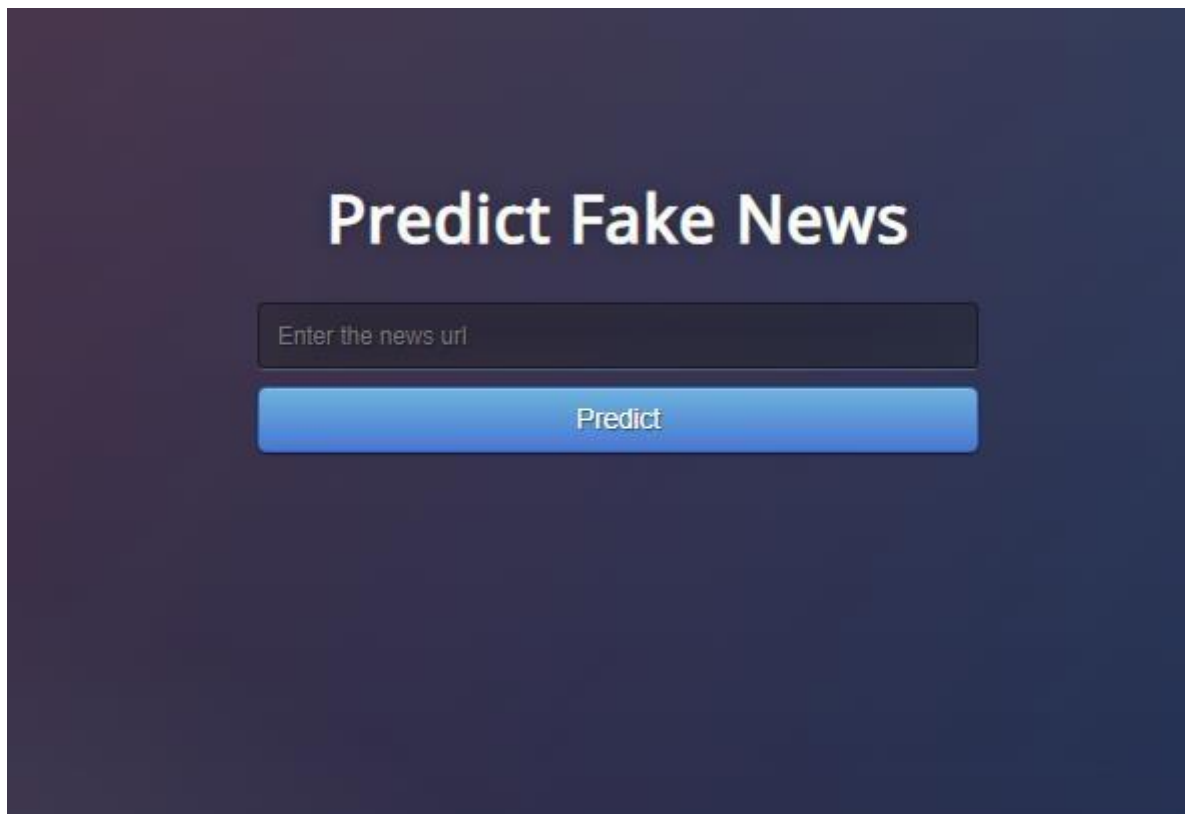


Fig:11

4.4 ADVANTAGES

- Advertisers take the advantage of fake news.
- Influencers also take benefits of fake news.
- Fun and entertainment.
- Political warfare.

4.5 DISADVANTAGES

- Defamation is among the disadvantages of fake news.
- Change in public opinion.
- Fake news may lead to Social unrest.
- Fake news cost lives.
- False Perception.

CHAPTER 5

RESULTS

5.1 RESULT

The Proposed system was procedured by buliding a flask application using web technologies such as Html,css,bootstrap,Flask etc.

- build a flask file “Fake_news.ipynb” which is a web framework written in python for server-side scripting. Let’s see the step by step procedure for building the backend application.
- You can also run this on spyder, given as app.py
- App starts running when the “__name__” constructor is called in main.
- Render_template is used to return html files.
- “GET” method is used to take input from the user.
- “POST” method is used to display the output to the user.
- Main.html web pages are given.

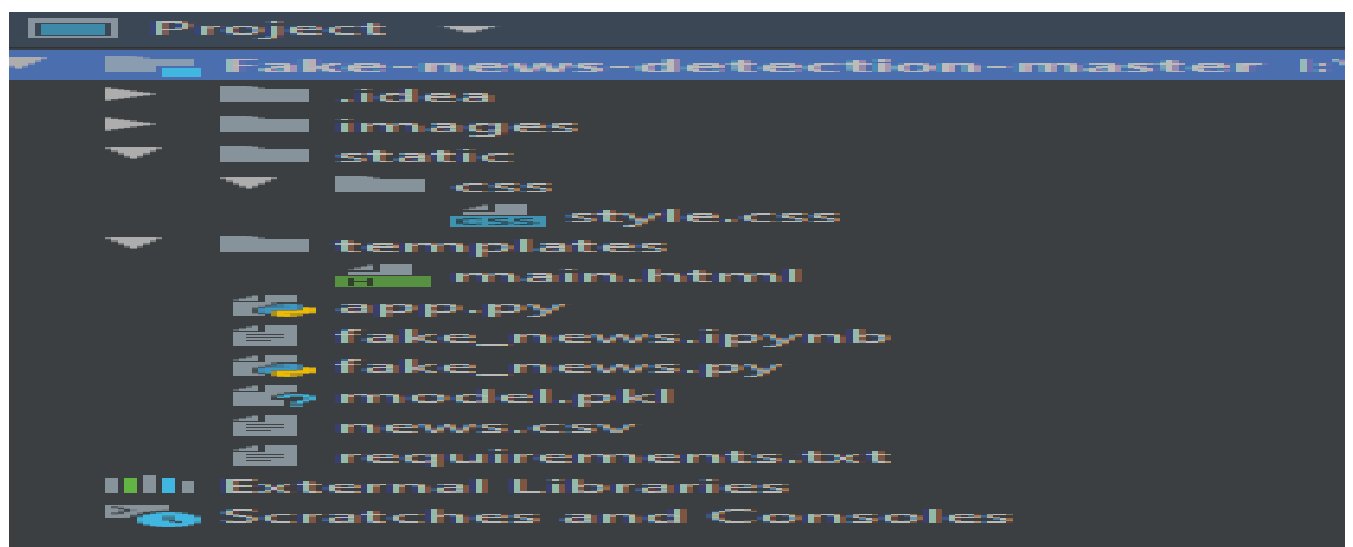


Fig:12

- Import the libraries.

```
#Importing the Libraries

#flask is use for run the web application.
import flask
#request is use for accessing file which was uploaded by the user on our application.
from flask import Flask, request, render_template
from flask_cors import CORS

#Python pickle module is used for serializing
# and de-serializing a Python object structure.
import pickle

#OS module in python provides functions for interacting with the operating system
import os

#Newspaper is used for extracting and parsing newspaper articles.
#For extracting all the useful text from a website.
from newspaper import Article

#URllib is use for the urlopen function and is able to fetch URLs.
#This module helps to define functions and classes to open URLs
import urllib
```

Fig:13

- Load the Flask and assign the model variable.

```
#Loading Flask and assigning the model variable
app = Flask(__name__)
CORS(app)
app=flask.Flask(__name__,template_folder='templates')

with open('model.pkl', 'rb') as handle:
    model = pickle.load(handle)
```

Fig:14

- Finding the accuracy of Algorithm

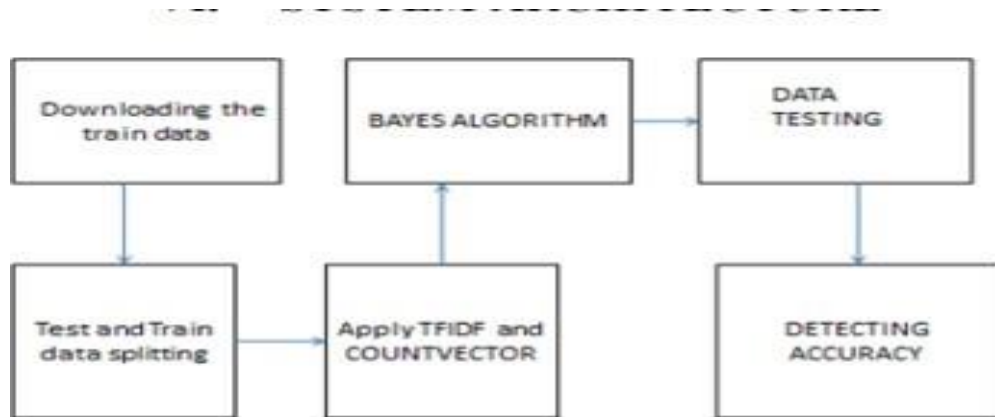


Fig:15

The first step in the detection of fake news is extracting the training data either by downloading it from a file or from online.

Accuracy results of all algorithms.

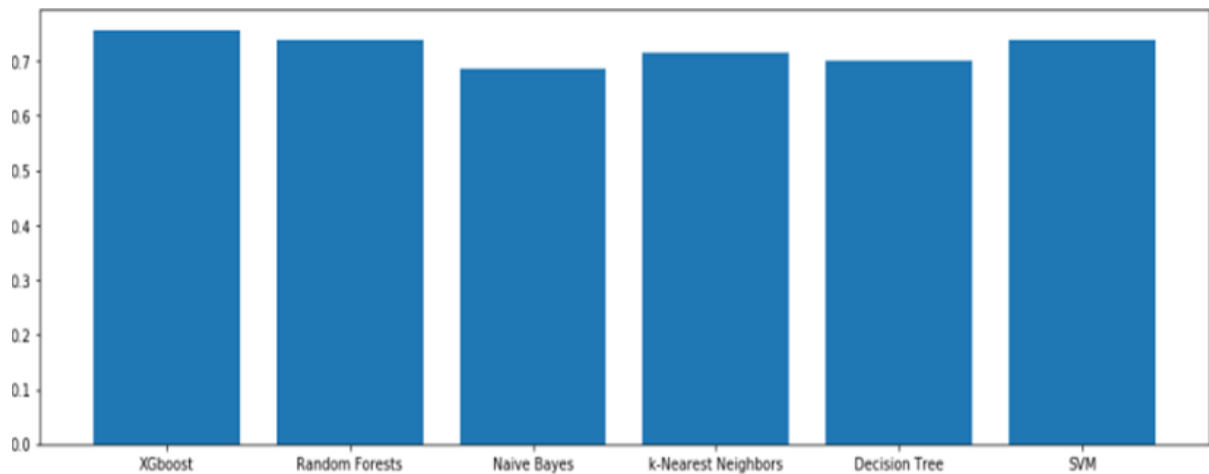


Fig:16

- Confusion Matrix for various Algorithms.

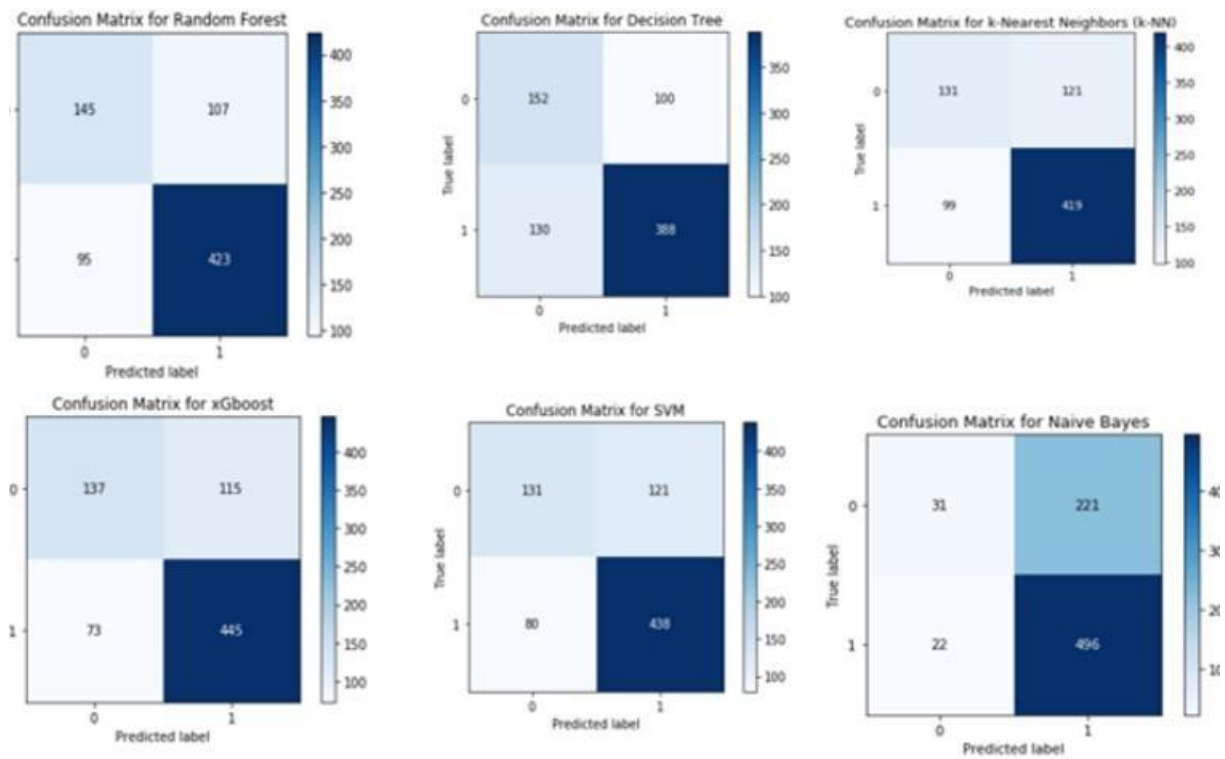


Fig:17

- Let's see our flask application

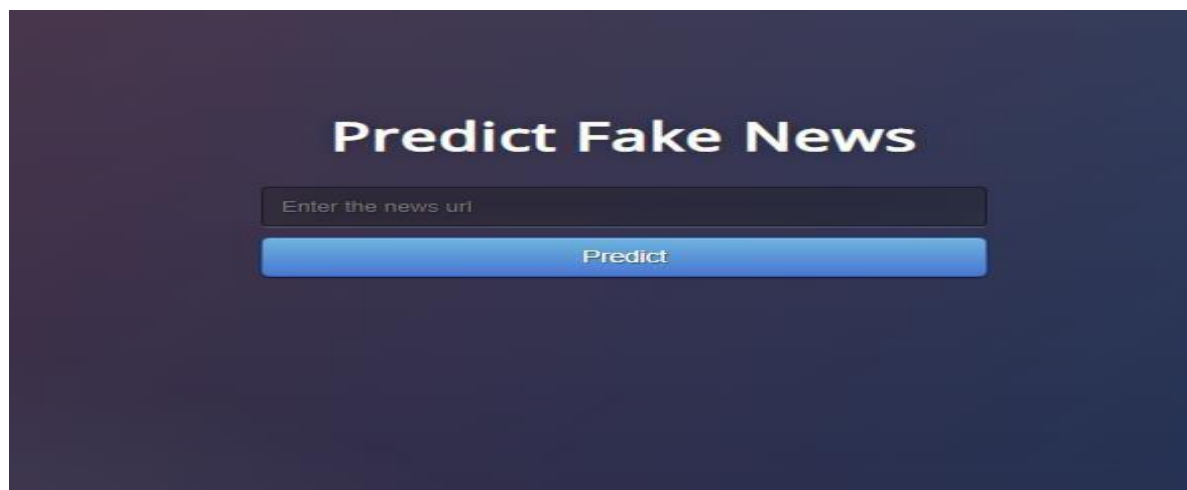


Fig:18

- The flask application predict whether data is “Real” or “Fake”.



Fig:19



Fig:20

CHAPTER 5

CONCLUSION

5.1 CONCLUSION:

Therefore by using naïve Bayes theorem we can conclude that any news from a large or small dataset can be classified as fake or real news by matching it with the previous dataset values in less time which in turn helps the users to believe in a particular news.

To tackle the increasing false information on the internet, the machine learning model created distinguishes an input as real news or fake news. A lot of social media sites like WhatsApp or Facebook are trying to implement such systems into their system to prevent the spread of fake news.

REFERENCE

- 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.
- Fake news websites. (n.d.) Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Fake_news_website. Accessed Feb. 6, 2017
- Rada Mihalcea , Carlo Strapparava, The lie detector: explorations in the automatic recognition of deceptive language, Proceedings of the ACL-IJCNLP
- 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. 380-383
- M. L. Della Vedova, E. Tacchini, S. Moret, G. Ballarin, M. DiPierro and L. de Alfaro, "Automatic Online Fake News Detection Combining Content and Social Signals," 2018 22nd Conference of Open Innovations Association (FRUCT), Jyväskylä, 2018, pp. 272-279.
- 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.
- [13] 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. 436-441
- Dataset- Fake News detection William Yang Wang. "liar, liar pants on fire": A new benchmark dataset for fake news detection.

SESHADRI RAO GUDLAVALLERU ENGINEERING COLLEGE

(An Autonomous Institute with Permanent Affiliation to JNTUK, Kakinada)
Seshadri Rao Knowledge Village, Gudlavalleru

Department of Computer Science and Engineering

Program Outcomes (POs)

Engineering Graduates will be able to:

- 1. Engineering knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
- 2. Problem analysis:** Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
- 3. Design/development of solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
- 4. Conduct investigations of complex problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions., component, or software to meet the desired needs.
- 5. Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
- 6. The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
- 7. Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
- 8. Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

- 9. Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
- 10. Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
- 11. Life-long learning:** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Program Specific Outcomes (PSOs)

PSO1 : Design, develop, test and maintain reliable software systems and intelligent systems. PSO2 : Design and develop web sites, web apps and mobile apps.

PROJECT PROFORMA

Classification of Project	Application	Product	Research	Review
	√			

Note: Tick Appropriate category

Project Outcomes	
Course Outcome (CO1)	Identify and analyze the problem statement using prior technical knowledge in the domain of interest.
Course Outcome (CO2)	Design and develop engineering solutions to complex problems by employing systematic approach.
Course Outcome (CO3)	Examine ethical, environmental, legal and security issues during project implementation.
Course Outcome (CO4)	Prepare and present technical reports by utilizing different visualization tools and evaluation metrics.

Mapping Table

CS1536 : INDUSTRIAL/PRACTICAL TRAINING															
Course outcomes	Program Outcomes and Program Specific Outcome														
	P O 1	P O 2	P O 3	P O 4	P O 5	P O 6	P O 7	P O 8	P O 9	P O 10	P O 11	P O 12		P S O 1	P S O 2
CO1: Acquire technical competence in the specific domain during the training.	3	2	2	2	2			2	2	2	1	2		2	
CO2: Identify the problem statement based on the requirements of the industry	3	3	2	2	1			2	2	2	1	2		2	2
CO3: Adapt project management skills on par with industrial standards.	3	2	2	1	3	1	1	2	2	2	3	2		2	
CO4: Develop a system model to obtain a solution and generate a report.	3	2	3	3	3	2	1	2	2	2	3	2		2	2
INDUSTRIAL/PRACTICAL TRAINING	3	3	3	2	3	1	1	2	2	2	2	2		2	1

Note: Map each project outcomes with POs and PSOs with either 1 or 2 or 3 based on level of mapping as follows:

1-Slightly (Low) mapped 2-Moderately (Medium) mapped 3-Substantially (High) mapped

Date: 30/04/2022

TO WHOMSOEVER IT MAY CONCERN

This is to certify that **Mr./Ms. Sai Kumar Chinnam** pursuing B.Tech, Computer Science and Engineering from Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru has successfully completed his/her Virtual Internship Program from **27/02/2022** to **30/04/2022**.

During this period he/she had learned the concepts of **Machine Learning & Deep Learning** and successfully completed a project "**Fake News Analysis In Social Media Using IBM Watson**".

Refer the enclosed **Certificate of Merit** for his/her performance during the tenure of Virtual Internship Program.

We wish him/her all the best for his/her future endeavours.

For **SmartBridge Educational Services Pvt. Ltd.,**



Jayaprakash.ch,
Program Manager

Certificate of Merit

This is to certify that **Mr./Ms. Sai Kumar Chinnam** has completed his/her Virtual Internship Program from **27/02/2022** to **30/04/2022** with **Machine Learning & Deep Learning** as the specialization and secured a SKILL INDEX **9** of 10.

Career Readiness Factor (CRF)

Evaluation Metrics: (on a scale of 1 to 4)

1- Rarely/poorly displays characteristic 2- Occasionally displays characteristic

3- Frequently displays characteristic 4- Always displays characteristic

NA – Not Applicable

Motivation/Enthusiasm	4
Leadership Qualities	4
Flexibility towards work	4
Professionalism/Work Ethics	3
Self-Confidence	3
Ability to work independently	3
Oral/written communication	3
Problem solving skills	3
Over All Score	27

Date: 30/04/2022



Jayaprakash. ch

Program Manager

TO WHOMSOEVER IT MAY CONCERN

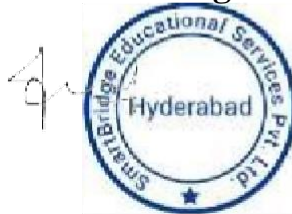
This is to certify that **Mr./Ms. Sajan Chadalavada** pursuing B.Tech, Computer Science and Engineering from Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru has successfully completed his/her Virtual Internship Program from **27/02/2022** to **30/04/2022**.

During this period he/she had learned the concepts of **Machine Learning & Deep Learning** and successfully completed a project "**Fake News Analysis In Social Media Using IBM Watson**".

Refer the enclosed **Certificate of Merit** for his/her performance during the tenure of Virtual Internship Program.

We wish him/her all the best for his/her future endeavours.

For **SmartBridge Educational Services Pvt. Ltd.**,



Jayaprakash.ch,
Program Manager

SmartBridge Educational Services Pvt. Ltd.

Plot No 132, Above DCB Bank , 2nd Floor, Bapuji Nagar, Habsiguda, Nacharam Main Road, Hyd - 500076.



info@thesmartbridge.com



[/thesmartbridgeiot](https://www.facebook.com/thesmartbridgeiot)



[/smartbridgeiot](https://www.instagram.com/smartbridgeiot)



[/thesmartbridge](https://twitter.com/thesmartbridge)

Certificate of Merit

This is to certify that **Mr./Ms. Sajan Chadalavada** has completed his/her Virtual Internship Program from **27/02/2022** to **30/04/2022** with **Machine Learning & Deep Learning** as the specialization and secured a SKILL INDEX **9** of 10.

Career Readiness Factor (CRF)

Evaluation Metrics: (on a scale of 1 to 4)

1- Rarely/poorly displays characteristic 2- Occasionally displays characteristic

3- Frequently displays characteristic 4- Always displays characteristic

NA – Not Applicable

Motivation/Enthusiasm	3
Leadership Qualities	4
Flexibility towards work	4
Professionalism/Work Ethics	3
Self-Confidence	4
Ability to work independently	3
Oral/written communication	4
Problem solving skills	3
Over All Score	28

Date: 30/04/2022



Jayaprakash. ch

Program Manager

TO WHOMSOEVER IT MAY CONCERN

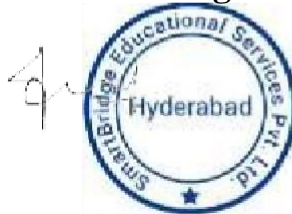
This is to certify that **Mr./Ms. Bandreddi Hima Krishna** pursuing B.Tech, Computer Science and Engineering from Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru has successfully completed his/her Virtual Internship Program from **27/02/2022** to **30/04/2022**.

During this period he/she had learned the concepts of **Machine Learning & Deep Learning** and successfully completed a project “**Fake News Analysis In Social Media Using IBM Watson**”.

Refer the enclosed **Certificate of Merit** for his/her performance during the tenure of Virtual Internship Program.

We wish him/her all the best for his/her future endeavours.

For **SmartBridge Educational Services Pvt. Ltd.**,



Jayaprakash.ch,
Program Manager

SmartBridge Educational Services Pvt. Ltd.

Plot No 132, Above DCB Bank , 2nd Floor, Bapuji Nagar, Habsiguda, Nacharam Main Road, Hyd - 500076.



info@thesmartbridge.com



[/thesmartbridgeiot](https://www.facebook.com/thesmartbridgeiot)



[/smartbridgeiot](https://www.instagram.com/smartbridgeiot)



[/thesmartbridge](https://twitter.com/thesmartbridge)

Certificate of Merit

This is to certify that **Mr./Ms. Bandreddi Hima Krishna** has completed his/her Virtual Internship Program from **27/02/2022** to **30/04/2022** with **Machine Learning & Deep Learning** as the specialization and secured a SKILL INDEX **9** of 10.

Career Readiness Factor (CRF)

Evaluation Metrics: (on a scale of 1 to 4)

1- Rarely/poorly displays characteristic 2- Occasionally displays characteristic

3- Frequently displays characteristic 4- Always displays characteristic

NA – Not Applicable

Motivation/Enthusiasm	3
Leadership Qualities	3
Flexibility towards work	3
Professionalism/Work Ethics	4
Self-Confidence	3
Ability to work independently	4
Oral/written communication	3
Problem solving skills	3
Over All Score	26

Date: 30/04/2022



Jayaprakash. ch

Program Manager

TO WHOMSOEVER IT MAY CONCERN

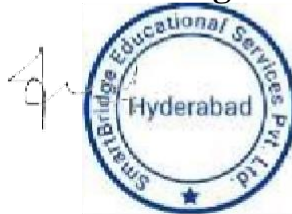
This is to certify that **Mr./Ms. Bala Sai Battula** pursuing B.Tech, Computer Science and Engineering from Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru has successfully completed his/her Virtual Internship Program from **27/02/2022** to **30/04/2022**.

During this period he/she had learned the concepts of **Machine Learning & Deep Learning** and successfully completed a project **“Fake News Analysis In Social Media Using IBM Watson”**.

Refer the enclosed **Certificate of Merit** for his/her performance during the tenure of Virtual Internship Program.

We wish him/her all the best for his/her future endeavours.

For **SmartBridge Educational Services Pvt. Ltd.**,



Jayaprakash.ch,
Program Manager

SmartBridge Educational Services Pvt. Ltd.

Plot No 132, Above DCB Bank , 2nd Floor, Bapuji Nagar, Habsiguda, Nacharam Main Road, Hyd - 500076.

 info@thesmartbridge.com  [/thesmartbridgeiot](https://www.facebook.com/thesmartbridgeiot)  [/smartbridgeiot](https://www.instagram.com/smartbridgeiot)  [/thesmartbridge](https://twitter.com/thesmartbridge)

Certificate of Merit

This is to certify that **Mr./Ms. Bala Sai Battula** has completed his/her Virtual Internship Program from **27/02/2022** to **30/04/2022** with **Machine Learning & Deep Learning** as the specialization and secured a SKILL INDEX **9** of 10.

Career Readiness Factor (CRF)

Evaluation Metrics: (on a scale of 1 to 4)

1- Rarely/poorly displays characteristic 2- Occasionally displays characteristic

3- Frequently displays characteristic 4- Always displays characteristic

NA – Not Applicable

Motivation/Enthusiasm	4
Leadership Qualities	4
Flexibility towards work	3
Professionalism/Work Ethics	3
Self-Confidence	3
Ability to work independently	3
Oral/written communication	4
Problem solving skills	3
Over All Score	27

Date: 30/04/2022



Jayaprakash. ch

Program Manager