DAYANANDA SAGAR UNIVERSITY

KUDLU GATE, BANGALORE – 560068

Bachelor of Technology
in
COMPUTER SCIENCE AND ENGINEERING

Major Project Phase-II Report

(Fake News Detection)

By
Sachin – ENG19CS0271
Shivam Kumar – ENG19CS0298
Shivansh Chaurasia – ENG19CS0299

Under the supervision of

Dr. Niranjan A.

Associate Professor Department of CSE

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING,
SCHOOL OF ENGINEERING
DAYANANDA SAGAR UNIVERSITY,
BANGALORE

(2022-2023)

DAYANANDA SAGAR UNIVERSITY

School of Engineering
Department of Computer Science & Engineering
Kudlu Gate, Bangalore – 560068
Karnataka, India

CERTIFICATE

This is to certify that the Phase-II project work titled **"FAKE NEWS DETECTION"** is carried out by **Sachin (ENG19CS0271), Shivam Kumar (ENG19CS0298), Shivansh Chaurasia (ENG19CS0299),** bonafide students of Bachelor of Technology in Computer Science and Engineering at the School of Engineering, Dayananda Sagar University, Bangalore in partial fulfillment for the award of degree in Bachelor of Technology in Computer Science and Engineering, during the year **2022-2023**.

Dr. Niranjan A.	**Dr. Girisha G S**	**Dr. Udaya Kumar Reddy K R**
Associate Professor Dept. of CS&E, School of Engineering Dayananda Sagar University	Chairman CSE School of Engineering Dayananda Sagar University	Dean School of Engineering Dayananda Sagar University
Date:	Date:	Date:

Name of the Examiner	**Signature of Examiner**
1.	
2.	

DECLARATION

I , **Sachin (ENG19CS0271), Shivam Kumar (ENG19CS0298), Shivansh Chaurasia (ENG19CS0299),** are student's of seventh semester B.Tech in **Computer Science and Engineering**, at School of Engineering, **Dayananda Sagar University**, hereby declare that the Major Project Stage-1 titled **"Fake News Detection"** has been carried out by us and submitted in partial fulfilment for the award of degree in **Bachelor of Technology in Computer Science and Engineering** during the academic year **2022-2023**.

Student **Signature**

Name1: Sachin

USN : ENG19CS0271

Name2: Shivam Kumar

USN : ENG19CS0298

Name3: Shivansh Chaurasia

USN : ENG19CS0299

Place : Bangalore

Date :

ACKNOWLEDGEMENT

It is a great pleasure for us to acknowledge the assistance and support of many individuals who have been responsible for the successful completion of this project work.

First, I take this opportunity to express our sincere gratitude to School of Engineering & Technology, Dayananda Sagar University for providing us with a great opportunity to pursue our Bachelor's degree in this institution.

I would like to thank ***Dr. Udaya Kumar Reddy K R, Dean, School of Engineering & Technology, Dayananda Sagar University*** *for his constant encouragement and expert advice.*

It is a matter of immense pleasure to express our sincere thanks to ***Dr. Girisha G S, Department Chairman****,* ***Computer Science and Engineering****,* ***Dayananda Sagar University,*** *for providing right academic guidance that made our task possible.*

I would like to thank our guide ***Dr. Niranjan A. Associate Professor****,* ***Dept. of Computer Science and Engineering****,* ***Dayananda Sagar University****, for sparing his/her valuable time to extend help in every step of our project work, which paved the way for smooth progress and fruitful culmination of the project.*

I would like to thank our ***Project Coordinator Dr. Meenakshi Malhotra & Dr. Pramod Naik*** *and all the staff members of Computer Science and Engineering for their support.*

I am grateful to our family and friends who provided us with every requirement throughout the course.

I would like to thank one and all who directly or indirectly helped us in the Project work.

Signature of Students

USN : ENG19CS0271

Name: Sachin

USN : ENG19CS0298

Name – Shivam Kumar

USN – ENG19CS0299

Name – Shivansh Chaurasia

TABLE OF CONTENTS

NOMENCLATURE USED

ML	Machine Learning
DL	Deep Learning
NLP	Natural Language Processing
US	United States
DT	Decision Tree
RF	Random Forest
k-NN	k-Nearest Neighbour
CNN	Convolutional neural network
LSTM	Long short-term memory
SVM	Support Vector Machine

LIST OF FIGURES

Abstract

In our modern era where the internet is ubiquitous, 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 to swaying election outcomes for the benefit of certain candidates. Moreover, spammers use appealing news headlines to generate revenue using advertisements via click-baits. We aim to perform binary classification of various news articles available online with the help of concepts pertaining to Artificial Intelligence, Natural Language Processing and Machine Learning. We aim to provide the user with the ability to classify the news as fake or real and also check the authenticity of the website publishing the news.

CHAPTER 1
INTRODUCTION

CHAPTER 1. INTRODUCTION

We consume news through several mediums throughout the day in our daily routine, but sometimes it becomes difficult to decide which one is fake and which one is authentic. Because of this reason we are developing a model which will help to detect the fake news with a good accuracy.

There has been a rapid increase in the spread of fake news in the last decade, most prominently observed in the 2016 US elections. Such proliferation of sharing articles online that do not conform to facts has led to many problems not just limited to politics but covering various other domains such as sports, health, and also science. One such area affected by fake news is the financial markets, where a rumor can have disastrous consequences and may bring the market to a halt.

Our ability to take a decision relies mostly on the type of information we consume; our world view is shaped on the basis of information we digest. There is increasing evidence that consumers have reacted absurdly to news that later proved to be fake. One recent case is the spread of novel corona virus, where fake reports spread over the Internet about the origin, nature, and behavior of the virus. The situation worsened as more people read about the fake contents online. Identifying such news online is a daunting task.

1.1 SCOPE OF THE PROJECT

- To segregate news on the basis of their authenticity
- Can be used in various fields as news is essential everywhere.
- Can be applied to verify the foundations of different matters.
- Can be very helpful in emergency situations like security threats

1.2 SOCIETAL/ENVIRONMENT IMPACTS

- Talking about Societal impact , this would beneficial for all the sections of people
- Since there would be more amount of losses on ground and the network due to the propagation of fake news in society , so this idea would help in reducing the losses incurred in all forms.
- It would be extremely beneficial to the society as resources would not be wasted on taking actions on fake news .
- In general everyone would benefit from such systems.

1.3 PROJECT PURPOSE

Learning from data and engineered knowledge to overcome fake news issue on social media. To achieve the goal a new combination algorithm approach shall be developed which will classify the text as soon as the news will publish online. In developing such a new classification approach as a starting point for the investigation of fake news we first applied available data set for our learning. The first step in fake news detection is classifying the text immediately once the news published online. Classification of text is one of the important research issues in the field of text mining. As we knew that dramatic increase in the content available online gives raise problem to manage this online textual data. So, it is important to classify the news into the specific classes i.e., Fake, Non fake, unclear.

CHAPTER 2

PROBLEM DEFINITION

CHAPTER 2. PROBLEM DEFINITION

There has been a rapid increase in the spread of fake news from the last decade so to overcome this issue we are making a system which helps to identify fake news with good accuracy from a whole bunch of news.

CHAPTER 3.
LITERATURE SURVEY

CHAPTER 3. LITERATURE SURVEY

- In Reference 1 , we see that preliminary analysis of the effectiveness of typical methods of feature reduction and the construction of stream models for an entirely new problem. In the preprocessing step we get an idea about algorithm that can be used to get higher accuracy.

- In Reference 2 ,author is focused on classifying fake news on social media sites with textual content within the natural language processing, a challenging field. In this paper they are divided the methods into three stages, first stage is pre-processing, during second stage they are extracting numerical features, and in last stage they are using ML & DL algorithm classifiers to group the items in the dataset.

- In the Reference 3 , we evaluated in this paper different classification algorithms such as k-nearest neighbor (k-NN), decision tree (DT), random forest (RF), etc. Then stacking method is used to improve individual model performance. we evaluated machine learning models and three deep learning models on two fake news datasets of different size in terms of accuracy, precision, recall, F1-score.

- In the Reference 4 , fake news detection requires a lot of experimentation using machine learning techniques on a wide range of datasets. This paper provides us with various results of experimenting done on datasets and provides a general view to detect fake news

CHAPTER 4
PROJECT DESCRIPTION

CHAPTER 4. PROJECT DESCRIPTION

In our proposed framework, we are expanding on the current literature by introducing ensemble techniques with various linguistic feature sets to classify news articles from multiple domains as true or fake. The ensemble techniques along with Linguistic Inquiry and Word Count (LIWC) feature set used in this research are the novelty of our proposed approach.

There are numerous reputed websites that post legitimate news contents, and a few other websites such as PolitiFact and Snopes which are used for fact checking. In addition, there are open repositories which are maintained by researchers to keep an up-to-date list of currently available datasets and hyperlinks to potential fact checking sites that may help in countering false news spread. However, we selected three datasets for our experiments which contain news from multiple domains (such as politics, entertainment, technology, and sports) and contain a mix of both truthful and fake articles. The datasets are available online and are extracted from the World Wide Web. The first dataset is ISOT Fake News Dataset; the second and third datasets are publicly available at Kaggle.

4.1. PROPOSED DESIGN

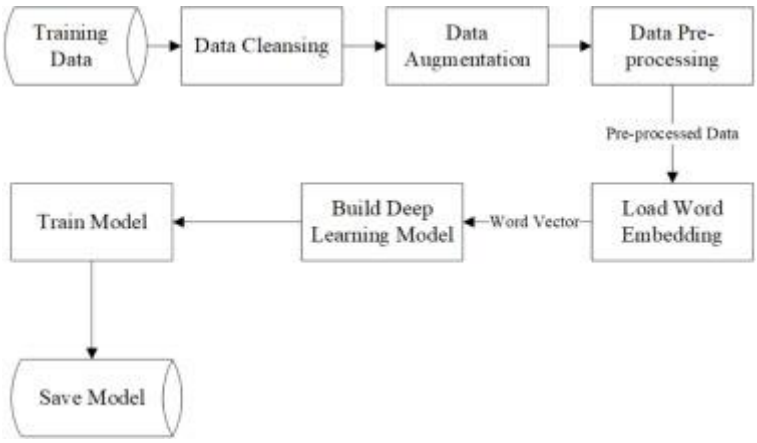


Fig 1 – Proposed block diagram

Initital steps in design is to work on data by training , cleansing etc. After work on data is done DL model is build and then the data is fed to the model and the result is showed.

Because of the multi-dimensional nature of fake news, the recognizing the category of news is not so easy. It is obvious that a practical technique must contain a few perspectives to precisely handle the issue. This is the reason the proposed strategy is a mix of Naïve Bayes classifier, Support Vector Machines, and semantic investigation. The proposed strategy is completely made out of Artificial Intelligence draws near, which is basic to precisely order between the genuine or the fake, rather than utilizing calculations that can't mirror subjective capacities. The three-section strategy is a blend between Machine Learning calculations that subdivide into managed learning procedures, and characteristic language preparing techniques.

4.2. ASSUMPTIONS AND DEPENDENCIES

Assumptions

Fake news detection project as the name suggests predicts fake news with help of machine algorithm. Our assumption of this project is that the system developed will be able to classify between fake news and real news using binary classification .This system will learn from its previous inputs and hence its accuracy would improve from time to time.

Dependencies

Here are some of the dependencies of this project

Jupyter notebook
Python libraries
Research papers
Datasets

CHAPTER 5

REQUIREMENTS

CHAPTER 5. REQUIREMENTS

5.1 Functional Requirements

- Dataset of news to segregate.
- Language support on the particular system
- Detect fake news on the basis of dataset and other factors
- Provide UI for result display

5.2 Non-Functional Requirements

- Availability
- Reliability
- Portability
- Usability
- Security

5.3 Software/System Requirements

Hardware Requirements

- Processor Minimum 1GHz recommended 2GHz or above.
- Hard Drive Minimum 2Gb recommended 10Gb or above.
- Memory (RAM) Minimum 1Gb recommended 4Gb or above.

Software Requirements

- Windows 10 or Mac OS
- Python 3, Jupiter Notebook
- HTML, CSS, JavaScript, React
- Visual Studio Code

CHAPTER 6
METHODOLOGY

CHAPTER 6. METHODOLOGY

The proposed model makes use of the ability of the CNN to extract local features and of the LSTM to learn long-term dependencies. First, a CNN layer of Conv1D is used for processing the input vectors and extracting the local features that reside at the text-level. The output of the CNN layer (i.e. the feature maps) are the input for the RNN layer of LSTM units/cells that follows. The RNN layer uses the local features extracted by the CNN and learns the long-term dependencies of the local features of news articles that classify them as fake or real. Using this model, a tool is implemented for detecting the fake articles. In this method supervised machine learning is used for classifying the dataset. The first step in this classification problem is dataset collection phase, followed by preprocessing, implementing features selection, then perform the training and testing of dataset and finally running the classifiers. The methodology is based on conducting various experiments on dataset using the algorithms described in the previous section named Random forest, SVM and Naïve Bayes, majority voting and other classifiers. The experiments are conducted individually on each algorithm, and on combination among them for the purpose of best accuracy and precision.

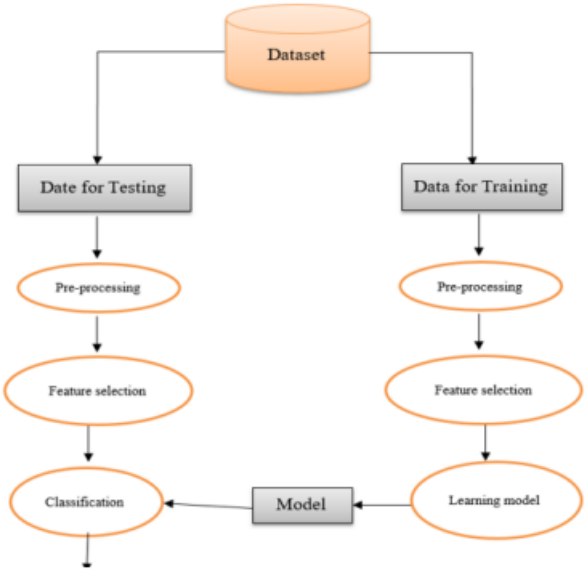


Figure 2- Describes the Proposed System Methodology

The main goal is to apply a set of classification algorithms to obtain a classification model in order to be used as a scanner for a fake news by details of news detection and embed the model in python application to be used as a discovery for the fake news data, Also, appropriate refactoring have been performed on the Python code to produce an optimized code The

classification algorithms applied in this model are k-Nearest Neighbors (k-NN), Linear Regression, XGBoost, Naive Bayes, Decision Tree, Random Forests and Support Vector Machine (SVM). All these algorithms get as accurate as possible. Where reliable from the combination of the average of them and compare them.

CHAPTER 7
EXPERIMENTATION

CHAPTER 7. Experimentation

7.1 Preparing the data –

```
[74]
X_train,X_test,y_train,y_test = train_test_split(data['text'], data.target, test_size=0.2, random_state=42)
```

7.2 Logistic Regression –

```
[82]
from sklearn.linear_model import LogisticRegression

pipe = Pipeline([('vect', CountVectorizer()),
                 ('tfidf', TfidfTransformer()),
                 ('model', LogisticRegression())])

model = pipe.fit(X_train, y_train)

prediction = model.predict(X_test)
print("accuracy: {}%".format(round(accuracy_score(y_test, prediction)*100,2)))
```

```
[89]
cm = metrics.confusion_matrix(y_test, prediction)
plot_confusion_matrix(cm, classes=['Fake', 'Real'])
```

7.3 Decision Tree Classifier –

```
[91]
from sklearn.tree import DecisionTreeClassifier

pipe = Pipeline([('vect', CountVectorizer()),
                 ('tfidf', TfidfTransformer()),
                 ('model', DecisionTreeClassifier(criterion= 'entropy',
                                                  max_depth = 20,
                                                  splitter='best',
                                                  random_state=42))])
model = pipe.fit(X_train, y_train)

prediction = model.predict(X_test)
print("accuracy: {}%".format(round(accuracy_score(y_test, prediction)*100,2)))
```

```
[92]
cm = metrics.confusion_matrix(y_test, prediction)
plot_confusion_matrix(cm, classes=['Fake', 'Real'])
```

7.4 Random Forest Classifier –

```
[94]
from sklearn.ensemble import RandomForestClassifier

pipe = Pipeline([('vect', CountVectorizer()),
                 ('tfidf', TfidfTransformer()),
                 ('model', RandomForestClassifier(n_estimators=50, criterion="entropy"))])

model = pipe.fit(X_train, y_train)
prediction = model.predict(X_test)
print("accuracy: {}%".format(round(accuracy_score(y_test, prediction)*100,2)))
```

```
[95]
cm = metrics.confusion_matrix(y_test, prediction)
plot_confusion_matrix(cm, classes=['Fake', 'Real'])
```

7.5 Frontend -

```
import './App.css';
import React, { useEffect,useState } from 'react'
import NavBar from './components/NavBar';
import News from './components/News';
import {BrowserRouter as Router, Routes, Route} from "react-router-dom";
import LoadingBar from 'react-top-loading-bar'
import WeatherCard from './components/WeatherCard';
import getFormattedWeatherData from "./services/weatherService";
import StockMarketCard from './components/StockMarketCard';
function App() {
  document.body.style = 'background: #EEECFC;';
  const pageSize = 12;
  const [progress, setProgress] = useState(0);
  const [searchQuery, setSearchQuery] = useState('*');
  const [weather, setWeather] = useState({ details:'Trying to Fetch Latest
Weather Update', icon:'L', temp:273, temp_min:273, temp_max:273, sunrise:0,
sunset:0, speed:0, humidity:0, feels_like:273, timezone:1000, dt:1669793656,
name:'- ', country:'-' });
  useEffect(() => {
    if (navigator.geolocation) {
      navigator.geolocation.getCurrentPosition((position) => {
        let lat = position.coords.latitude;
        let lon = position.coords.longitude;
        // setWeather({details:'Featching Latest Weather Update', icon:'L'});
        fetchWeather(lat,lon);
      }, function errorHandler(err) {
        // eslint-disable-next-line
        if(err.code == 1) {
           alert("Weather update not available : Access to geolocation is
denied!");
           setWeather({details:'Weather update not available : Access to
geolocation is denied!', icon:'E'});
           // eslint-disable-next-line
        } else if( err.code == 2) {
```

```
          alert("Weather update not available : Position is unavailable!");
          setWeather({details:'Weather update not available : Access to
geolocation is denied!', icon:'E'});
        }
     });
    }

    const fetchWeather = async (lat,lon) => {
      await getFormattedWeatherData({lat,lon}).then((data) => {
        console.log(
          `App.js - Successfully fetched weather for ${data.name},
${data.country}.`
        );
        setWeather(data);
      });
    };

  }, []);

  return (
    <div className="App">
      <h1>hello</h1>
      <Router>
      <NavBar setSearchQuery={setSearchQuery}/>
      <LoadingBar
        height={3}
        color='#f11946'
        progress={progress}
      />
        <Routes>
          <Route  path="/" element={<><div className="container"><div
className="Fake"><a href="https://shivansh3270-fake-news-detection-
streamlitapp-3l6o3h.streamlit.app/">Fake News Detector</a></div><div
className="card-group" style={{ margin: '35px 0px -60px', marginTop: '90px',
padding: '10px' }}><WeatherCard
weather={weather}></WeatherCard><StockMarketCard/></div></div><News
setProgress={setProgress} key="general" pageSize={pageSize} country="IN"
language="en" query={searchQuery} category="news" /></>}/>
          <Route  path="/business" element={<News setProgress={setProgress}
key="business" pageSize={pageSize} country="IN" language="en"
query={searchQuery}  category="business"/>} />
          <Route  path="/economics" element={<News setProgress={setProgress}
key="economics" pageSize={pageSize} country="IN" language="en"
query={searchQuery}  category="economics"/>} />
          <Route  path="/politics" element={<News setProgress={setProgress}
key="politics" pageSize={pageSize} country="IN" language="en"
query={searchQuery}  category="politics"/>} />
```

```
          <Route  path="/entertainment" element={<News
setProgress={setProgress} key="entertainment" pageSize={pageSize} country="IN"
language="en" query={searchQuery}  category="entertainment"/>} />
          <Route  path="/science" element={<News setProgress={setProgress}
key="science" pageSize={pageSize} country="IN" language="en"
query={searchQuery}  category="science"/>} />
          <Route  path="/technology" element={<News setProgress={setProgress}
key="technology" pageSize={pageSize} country="IN" language="en"
query={searchQuery}  category="tech"/>} />
          <Route  path="/sports" element={<News setProgress={setProgress}
key="sports" pageSize={pageSize} country="IN" language="en"
query={searchQuery}  category="sport"/>} />
          <Route  path="/travel" element={<News setProgress={setProgress}
key="travel" pageSize={pageSize} country="IN" language="en"
query={searchQuery}  category="travel"/>} />
          <Route  path="/music" element={<News setProgress={setProgress}
key="music" pageSize={pageSize} country="IN" language="en"
query={searchQuery}  category="music"/>} />
          <Route  path="/food" element={<News setProgress={setProgress}
key="food" pageSize={pageSize} country="IN" language="en"
query={searchQuery}  category="food"/>} />
          <Route  path="/gaming" element={<News setProgress={setProgress}
key="gaming" pageSize={pageSize} country="IN" language="en"
query={searchQuery}  category="gaming"/>} />
          <Route  path="/search" element={<News setProgress={setProgress}
key="search" pageSize={pageSize} country="IN" language="en"
query={searchQuery} category="Search Results"/>} />
        </Routes>
      </Router>
    </div>
  );
}
```

CHAPTER 8

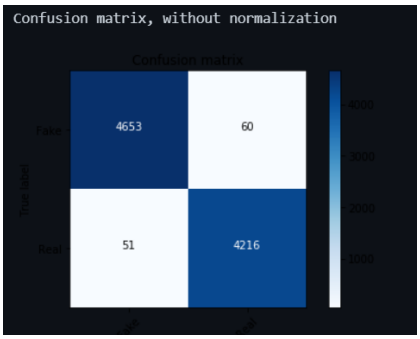
RESULT AND TEST CASES

CHAPTER 8 RESULT AND TEST CASES

8.1 Results

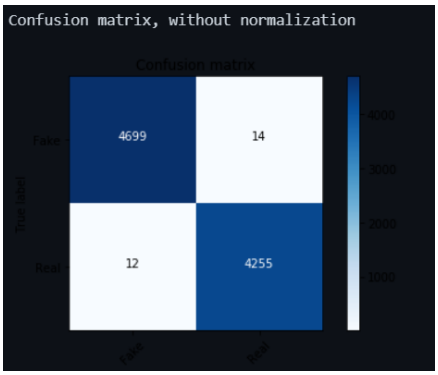
Logistic Regression –

accuracy: 98.76%



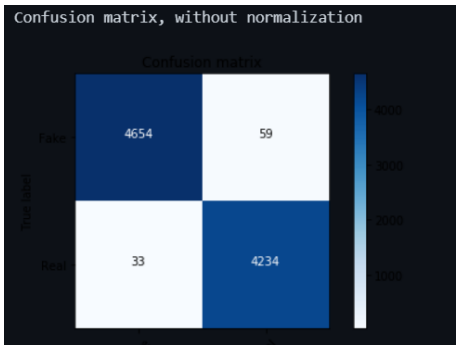
Decision Tree Classifier –

accuracy: 99.71%

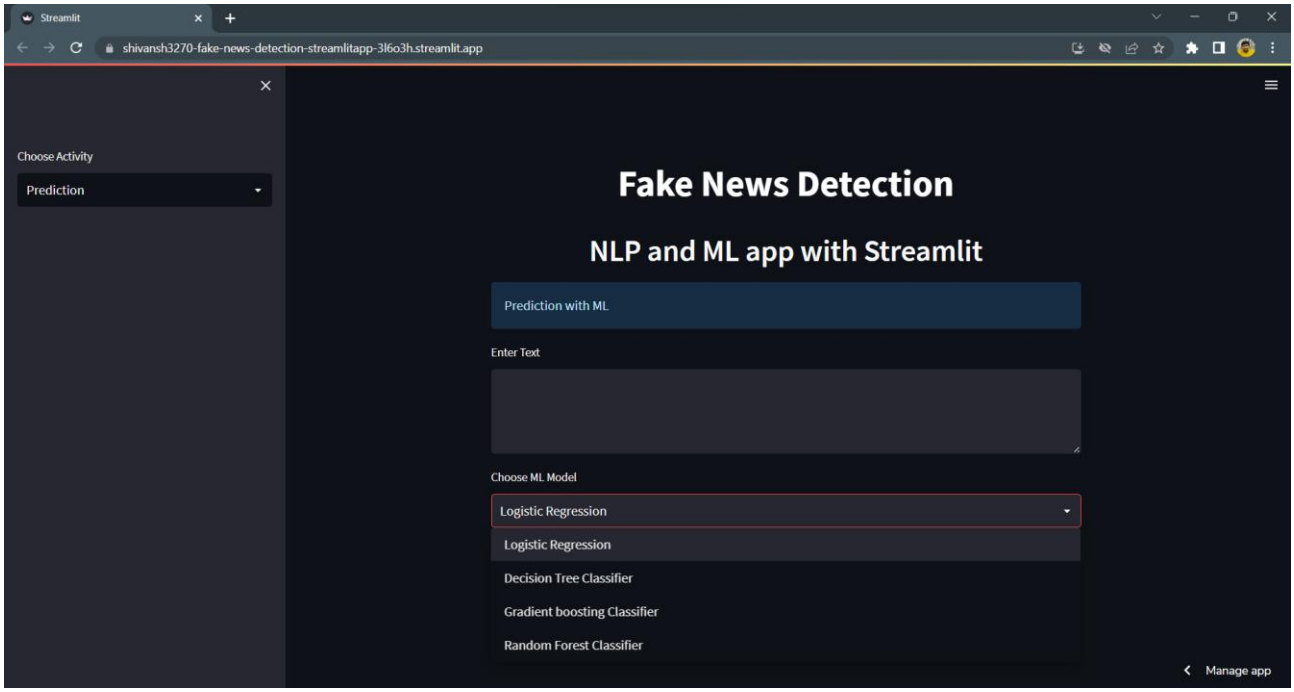


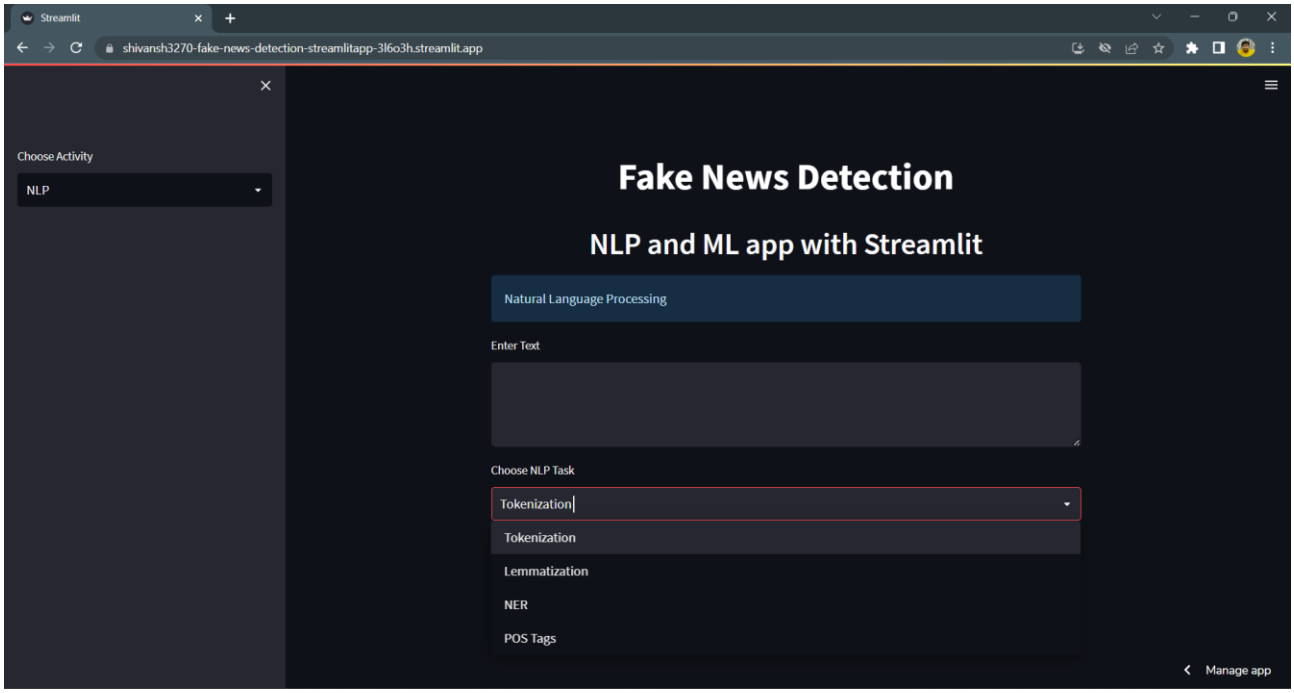
Random Forest Classifier –

accuracy: 98.98%



Web Application


Streamlit
shivansh3270-fake-news-detection-streamlitapp-3l6o3h.streamlit.app
Choose Activity
Prediction
Fake News Detection
NLP and ML app with Streamlit
Prediction with ML
Enter Text
Choose ML Model
Logistic Regression
Logistic Regression
Decision Tree Classifier
Gradient boosting Classifier
Random Forest Classifier
Manage app


Streamlit
shivansh3270-fake-news-detection-streamlitapp-3l6o3h.streamlit.app
Choose Activity
NLP
Fake News Detection
NLP and ML app with Streamlit
Natural Language Processing
Enter Text
Choose NLP Task
Tokenization
Tokenization
Lemmatization
NER
POS Tags
Manage app

8.2 Discussion of results

As we can see in the above screenshots , the user should enter any news to check its validity in text format . After the news is entered , its is checked with the dataset to check whether the news is fake or true . After the computation is done in backend , appropriate results are displayed .

CHAPTER 9
CONCLUSION

CHAPTER 9. CONCLUSION AND FUTURE WORK

9.1 Conclusion

Due to increasing use of internet, it is now easy to spread fake news. A huge number of persons are regularly connected with internet and social media platforms. There is no any restriction while posting any news on these platforms. So some of the people takes the advantage of these platforms and start spreading fake news against the individuals or organizations. This can destroy the repute of an individual or can affect a business. Through fake news, the opinions of the people can also be changed for a political party. There is a need for a way to detect these fake news. Machine learning classifiers are using for different purposes and these can also be used for detecting the fake news. The classifiers are first trained with a data set called training data set. After that, these classifiers can automatically detect fake news

CHAPTER 10
REFERENCES

CHAPTER 10. REFERENCES

- Paweł Ksieniewicz, Paweł Zyblewski, Michał Choras, Rafał Kozik, Agata Giełczyk " Fake News Detection from Data Streams".

- Awf Abdulrahman, Muhammet Baykara, "Fake News Detection Using Machine Learning and Deep Learning Algorithm".

- TAO JIANG, JIAN PING LI, AMIN UL HAQ, ABDUS SABOOR, AMJAD ALI "A Novel Stacking Approach for Accurate Detection of Fake News".

- Jamal Abdul Nasir, Osama Subhani Khan, Iraklis Varlamis, "Fake news detection: A hybrid CNN-RNN based deep learning approach".