



A Project Report

On

Fake News Detection using PassiveAggressiveClassifier

In partial fulfillment for the award of the degree of

MASTER OF COMPUTER APPLICATIONS

Submitted by
PREETHAM HD

MCA
[18MCAL0016]

Under the Guidance of
Dr. Vijayakumar

Department of MCA
School of Computer Science & IT,
Jain Knowledge Campus, Jayanagar 9th Block, Bangalore

May 2021



CERTIFICATE

This is to certify that the project entitled **Fake News Detection using PassiveAggressiveClassifier** is the bonafide record of work done **PREETHAM HD[18MCAL0016]** of **6TH SEMESTER** **MASTER OF COMPUTER APPLICATIONS** during the year **2020-2021**. All the suggestions as required have been incorporated by him/her in the current work. Further this work or any part thereof has not been presented for the award Degree of any University or Organization.

Project Guide

Head of the Department (MCA)

Head School of (CS&IT)

Examiner-1

Examiner-2

Department of MCA
School of Computer Science & IT,
Jain Knowledge Campus, Jayanagar 9th Block, Bangalore

DECLARATION

I affirm that the project work titled **Fake News Detection using PassiveAggressiveClassifier** being submitted in partial fulfillment for the award of **MASTER OF COMPUTER APPLICATIONS** is the original work carried out by me. It has not formed the part of any other project work submitted for award of any degree or diploma, either in this or any other University.

(Signature of the Candidate)

Preetham HD

[USN: 18MCAL0016]

ACKNOWLEDGEMENT

I do acknowledge the support and encouragement of all people who helped me throughout the completion of this project.

I would wish to give thanks **Dr. M N Nachappa, Dean of Commerce** Knowledge Campus, Bangalore, Karnataka for proving the facilities to try to analysis work. His leadership and management skills are continuously a supply of inspiration.

I would wish to give thanks **Dr. Dinesh Nilkhant**, Director – JGI, Knowledge Campus, Bangalore, Karnataka for proving the facilities to try to analysis work. His leadership and management skills are continuously a supply of inspiration.

I conjointly wish to give thanks **Dr. Bhuvana J**, Program co-coordinator of School of Computer Science & IT, Jain deemed to be university, Knowledge campus, Bangalore, Karnataka for his support and cordial cooperation.

Moving further, I would like to give my sincere gratitude to Project Coordinators, **Dr. Lakshmi JVN and Dr. Gangothri R** Department of Master of Computer Application for sharing their experience which helped me in completing my thesis in the best possible way. In addition,

they also helped in critically reviewing and proof reading my work and my project thesis.

I am highly indebted to my guide **Dr. Vijayakumar.** of Department of Master of Computer Application for providing a valuable guidance throughout the course of the project. I am glad to all or any college members, project scholars and staff of Department of Master of computer application Jain deemed to be university for his or her co-operation and encouragement.

I am indebted to my **Parents and Friends;** whose constant support and blessings gave a driving force to go on and to all my friends for their continued moral and material support throughout the course of the project. I am grateful to the Almighty for giving me strength and ability to complete the project successfully.

**PREETHAM HD
(18MCAL0016)**

ABSTRACT

The hazardous development of fake news and its disintegration to vote based system, peddling fake information on social media expanded the interest for fake news detection. Recognize fake news is important to pull in and join scientists in related fields to lead research on fake news, Online news platforms enormously impact our daily life and culture in both positive and negative ways. As online media turns out to be the main source of data, so a lot of fake news is spreading on the web, that widespread with individuals following it with no earlier or complete data of occasion validness.

.

Keywords— Flask, News detection,TFIDFVectorizer

CONTENTS

	Abstract	i
	List of Figures	ii
	Data processing	8
	Work-flow diagram	
Chapter1	Introduction	
	1.1 Introduction	1
	1.2 Existing System	2
	1.3 Proposed System	3
Chapter 2	Literature Review	4
	2.1 Literature Review	5
Chapter3	Hardware Equipment's and Software used	
	3.1 Hardware and Software Requirements	6
Chapter 4	System Design And Methodology	
	4.1 System Design And Methodology.....	7
Chapter 5	Coding	
	5.1 Coding	9-13
Chapter 6	Results	
	6.1 Results	14-16
Chapter 7	Conclusion	
	7.1 Conclusion	17
References		18

CHAPTER 1

1.1 Introduction

Online media for news utilization is a blade that cuts both ways. From one viewpoint, its ease, simple access, and quick scattering of data lead individuals to search out and consume news from online media. Then again, it empowers the widespread of phony news. The broad spread of fake news contrarily affects people and society. Most of our life is spent interacting with people online and share information rather than seeking information from traditional news platforms, so the spread of fake news propagation through online increased massively. In this paper we have proposed a method to detect fake news on basis of the news source, author, and organization, to identify the fake news we have used some machine learning algorithms to label the news as fake or real, to increase readership and attract viewers and collect advertising revenue, some news sources publish fake news and many political parties are spreading fake news and rumors for their party benefits, so fact-checking and fake news detection models are very important to identify the validness of news.

1.2 Existing system

There is no fact-checking in social media post, people are mainly dependent on social media for information and the fake news is posted all around social media it's getting hard for people to find out which one is true and which one is fake news, fake news will trigger people or some community and it's hard to maintain peace in the society, the project is concerned with the finding solution to the fake news which has widely spread over the social media.

1.3 Proposed system

We have proposed a model to detect fake news by using filters and machine learning algorithms to predict the news as real or fake based on the author and news source and organization, we built a model to detect it with more accuracy to help people and prevent social media misleading the people, the model label the news as fake or real based on the model trained which helps the readers to distinguish between fake and real information. most of the news sources are published on the web, so many political parties, some organizations, and publishers spread fake news to their benefits so fake news detection and fact-checking models are necessary

CHAPTER 2

LITERATURE REVIEW

2.1 Literature Review

Online media for news utilization is a blade that cuts both ways. From one viewpoint, its ease, simple access, and fast dispersal of data lead individuals to search out and consume news from web-based media. Then again, it empowers the wide-spread of fake news, very-low quality news with lots of bogus data. The broad spread of fake news adversely affects people and society. Therefore, fake news discovery on web-based media has as of late become arising research that is drawing in huge attention, fake news on social media has unique challenges so that existing algorithms won't work on them, fake news indirectly affects the democracy, government, journalism so the comprehensive survey of fake news has done to control the spreading of fake news, the algorithms have been developed to detect the source of the fake news spreading journals and authors and label them as not trustworthy or fake news producers then it can be stopped and it will alert the readers, nobody will do a fact-checking before reading news and social media has become the source of many news information so spread of fake information on social media has grown rapidly. Machine learning algorithms and AI has played a vital role

in classification with some limitations, many supervised models were trained to detect fake news on social media and to do a fact check on publishing tweets on Twitter, post on Instagram and Facebook, Online media for news utilization is getting mainstream these days. The minimal effort, simple access, and quick data scattering of web-based media bring benefits for individuals to seek out news timely. It additionally causes the wide-spread of fake information, people consume news from social media rather than traditional news sources, those news stories with purposefully fake data, are created online for many of purposes, like finance and political gain. Early detection and prevention of fake news are much needed for the welfare of society.

CHAPTER 3

HARDWARE EQUIPMENT'S AND SOFTWARE USED

3.1 Hardware Equipment's and Software used

Hardware requirement :

RAM : 1GB

ROM : 80GB

Processor : Core i3

Software requirement :

Language : Python

IDE : Jupyter Notebook

GUI : Flask with HTML

Module Requirements:

scikit-learn == 0.22.1

pandas

numpy

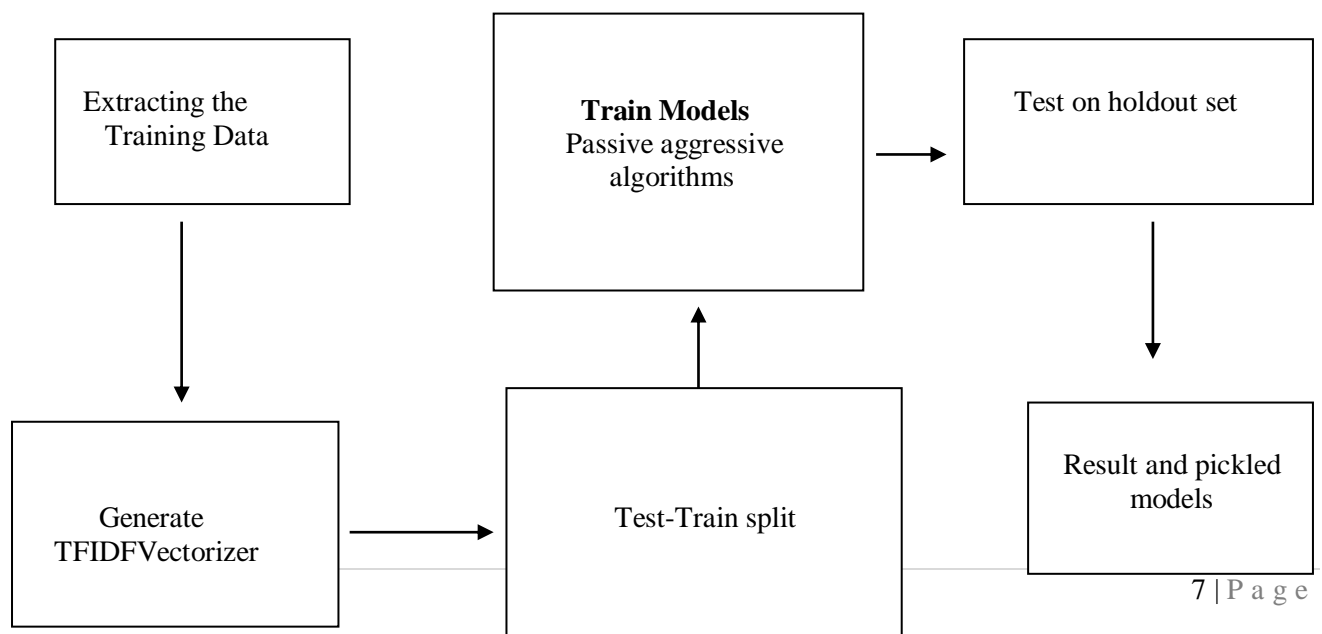
flask

CHAPTER 4

SYSTEM DESIGN AND METHODOLOGY

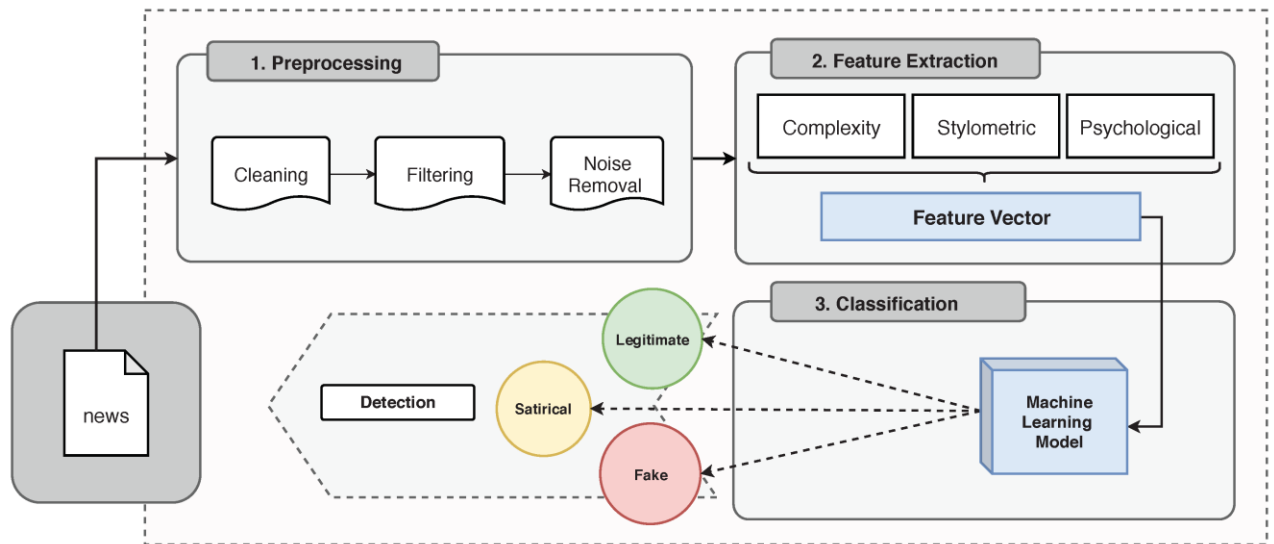
4.1 System Design and Methodology

In the project, the point is to discover the exhibition of the proposed system on fake or real news classifier. The first step was to load the dataset which is collected from kaggle.com which is a huge dataset based on US journal news articles that includes article ID, title, and news labels, 80% of the dataset is used as training data, and the remaining 20% is used as testing data, at first the dataset is loaded and split method is used, then that data will be passed to TfidfVectorizer and then our model is trained using PassiveAggressiveClassifier and it will predict based on the trained model as fake or real, the proposed model was able to attain the 93.76% accuracy, in the second step the Flask has to be installed to create GUI for the proposed model.

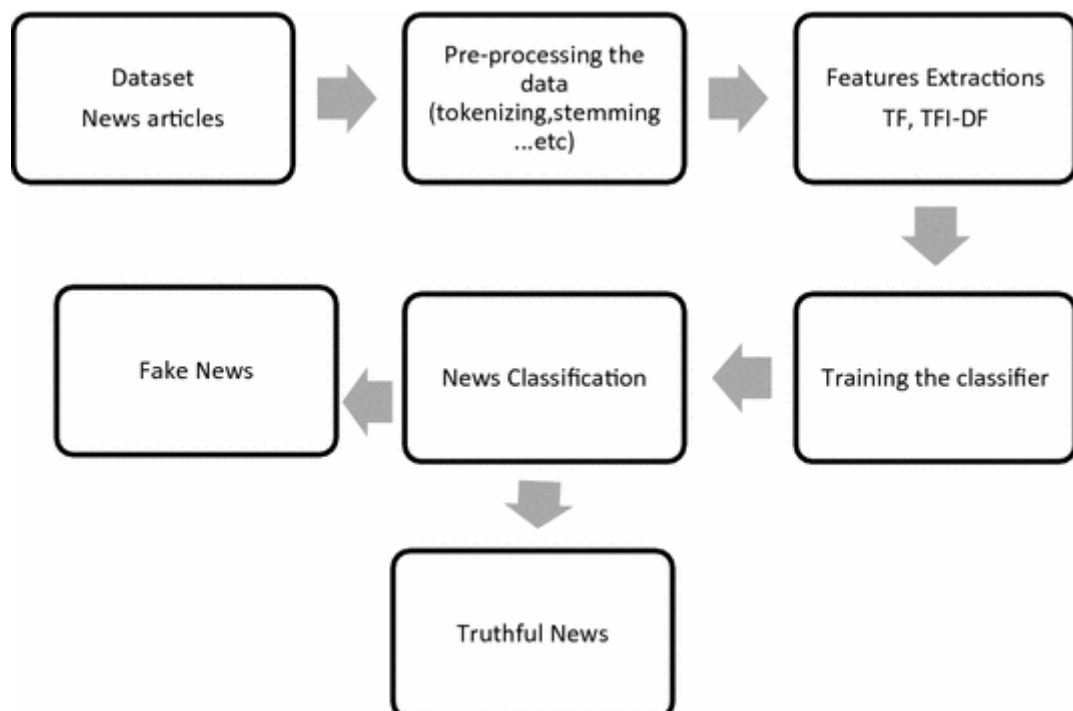


DIAGRAMS

Data-processing :



Workflow:



CHAPTER 5

Coding

5.1 Coding

TfidfVectorizer Explanation

Convert a collection of raw documents to a matrix of TF-IDF features

TF-IDF where TF means term frequency, and IDF means Inverse Document frequency.

```
In [1]: from sklearn.feature_extraction.text import TfidfVectorizer  
text = ['Hello Preetham Bhat here, I love machine learning', 'Welcome to the Machine learning hub' ]
```

```
In [ ]: vect = TfidfVectorizer()
```

```
In [ ]: vect.fit(text)
```

```
In [ ]: ## TF will count the frequency of word in each document. and IDF  
print(vect.idf_)
```

```
In [ ]: print(vect.vocabulary_)
```

A words which is present in all the data, it will have low IDF value. With this unique words will be highlighted using the Max IDF values.

```
In [ ]: example = text[0]  
example
```

```
In [ ]: example = vect.transform([example])  
print(example.toarray())
```

Here, 0 is present in the which indexed word, which is not available in given sentence.

PassiveAggressiveClassifier

Passive: if correct classification, keep the model; Aggressive: if incorrect classification, update to adjust to this misclassified example.

Passive-Aggressive algorithms are generally used for large-scale learning. It is one of the few 'online-learning algorithms'. In online machine learning algorithms, the input data comes in sequential order and the machine learning model is updated step-by-step, as opposed to batch learning, where the entire training dataset is used at once. This is very useful in situations where there is a huge amount of data and it is computationally infeasible to train the entire dataset because of the sheer size of the data. We can simply say that an online-learning algorithm will get a training example, update the classifier, and then throw away the example.


```

In [ ]: import os
        #os.chdir("C:/Fake News Detection")
        os.chdir("F:\Fake_News_Detection-master")

In [ ]: import pandas as pd

In [ ]: dataframe = pd.read_csv('news.csv')
        dataframe.head()

In [ ]: x = dataframe['text']
        y = dataframe['label']

In [ ]: x

In [ ]: y

In [ ]: from sklearn.model_selection import train_test_split
        from sklearn.feature_extraction.text import TfidfVectorizer
        from sklearn.linear_model import PassiveAggressiveClassifier
        from sklearn.metrics import accuracy_score, confusion_matrix

In [ ]: x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=0)
        y_train

In [ ]: y_train

In [ ]: tfvect = TfidfVectorizer(stop_words='english',max_df=0.7)
        tfid_x_train = tfvect.fit_transform(x_train)
        tfid_x_test = tfvect.transform(x_test)

        • max_df = 0.50 means "ignore terms that appear in more than 50% of the documents".
        • max_df = 25 means "ignore terms that appear in more than 25 documents".

In [ ]: classifier = PassiveAggressiveClassifier(max_iter=50)
        classifier.fit(tfid_x_train,y_train)

In [ ]: y_pred = classifier.predict(tfid_x_test)
        score = accuracy_score(y_test,y_pred)
        print(f'Accuracy: {round(score*100,2)}%')

In [ ]: cf = confusion_matrix(y_test,y_pred, labels=['FAKE','REAL'])
        print(cf)

In [ ]: def fake_news_det(news):
        input_data = [news]
        vectorized_input_data = tfvect.transform(input_data)

```

```

In [ ]: def fake_news_det(news):
        input_data = [news]
        vectorized_input_data = tfvect.transform(input_data)
        prediction = classifier.predict(vectorized_input_data)
        print(prediction)

In [ ]: fake_news_det('U.S. Secretary of State John F. Kerry said Monday that he will stop in Paris later this week, amid criticism that n
        ◀ ▶

In [ ]: fake_news_det("""Go to Article
        President Barack Obama has been campaigning hard for the woman who is supposedly going to extend his legacy four more years. The o
        ◀ ▶

In [ ]: import pickle
        pickle.dump(classifier,open('model.pkl', 'wb'))

In [ ]: # Load the model from disk
        loaded_model = pickle.load(open('model.pkl', 'rb'))

In [ ]: def fake_news_det1(news):
        input_data = [news]
        vectorized_input_data = tfvect.transform(input_data)
        prediction = loaded_model.predict(vectorized_input_data)
        print(prediction)

In [ ]: fake_news_det1("""Go to Article
        President Barack Obama has been campaigning hard for the woman who is supposedly going to extend his legacy four more years. The o
        ◀ ▶

In [ ]: fake_news_det1("""U.S. Secretary of State John F. Kerry said Monday that he will stop in Paris later this week, amid criticism tha
        ◀ ▶

In [ ]: fake_news_det1('U.S. Secretary of State John F. Kerry said Monday that he will stop in Paris later this week, amid criticism that
        ◀ ▶

- - -

```

Flask model

```
1 from flask import Flask, render_template, request
2 from sklearn.feature_extraction.text import TfidfVectorizer
3 from sklearn.linear_model import PassiveAggressiveClassifier
4 import pickle
5 import pandas as pd
6 from sklearn.model_selection import train_test_split
7
8 app = Flask(__name__)
9 tfvect = TfidfVectorizer(stop_words='english', max_df=0.7)
10 loaded_model = pickle.load(open('model.pkl', 'rb'))
11 dataframe = pd.read_csv('news.csv')
12 x = dataframe['text']
13 y = dataframe['label']
14 x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=0)
15
16 def fake_news_det(news):
17     tfidf_x_train = tfvect.fit_transform(x_train)
18     tfidf_x_test = tfvect.transform(x_test)
19     input_data = [news]
20     vectorized_input_data = tfvect.transform(input_data)
21     prediction = loaded_model.predict(vectorized_input_data)
22     return prediction
23
24 @app.route('/')
25 def home():
26     return render_template('index.html')
27
28 @app.route('/predict', methods=['POST'])
29 def predict():
30     if request.method == 'POST':
31         message = request.form['message']
32         pred = fake_news_det(message)
33         print(pred)
34         return render_template('index.html', prediction=pred)
35     else:
36         return render_template('index.html', prediction="Something went wrong")
37
38 if __name__ == '__main__':
39     app.run(debug=True)
```

CSV File

	A	B	C	D	E
1		title	text	label	
2	8476	You Can S	Daniel	FAKE	
3	10294	Watch The	Google	FAKE	
4	3608	Kerry to g	U.S.	REAL	
5	10142	Bernie sup	â€”	FAKE	
6	875	The Battle	It's	REAL	
7	6903	Tehran, U		FAKE	
8	7341	Girl Horri	Share	FAKE	
9	95	â€”Britair	A Czech st	REAL	
10	4869	Fact check	Hillary	REAL	
11	2909	Iran repor	Iranian	REAL	
12	1357	With all th	CEDAR	REAL	
13	988	Donald Tr	Donald	REAL	
14	7041	Strong So	Click	FAKE	
15	7623	10 Ways /	October	FAKE	
16	1571	Trump tak	Killing Ob	REAL	
17	4739	How wor	As more	REAL	
18	7737	Shocking!	Shocking	FAKE	
19	8716	Hillary Cli	O	FAKE	
20	3304	What's in	Washing	REAL	
21	3078	The 1 cha	While	REAL	
22	2517	The slippe	With	REAL	
23	10348	Episode #	Novemb	FAKE	
24	778	Hillary Cli	Hillary	REAL	
25	3300	New Sena	Mitch	REAL	
26	6155	â€”Inferno	Mises.or	FAKE	
27	636	Anti-Trum	Washing	REAL	
28	755	Sanders T	Meanwhi	REAL	
29	626	Donald Tr	After a	REAL	
30	691	Pure chao	If you	REAL	
31	5743	Syrian Wa	Syrian	FAKE	
32	1787	GOP insid	On this da	REAL	

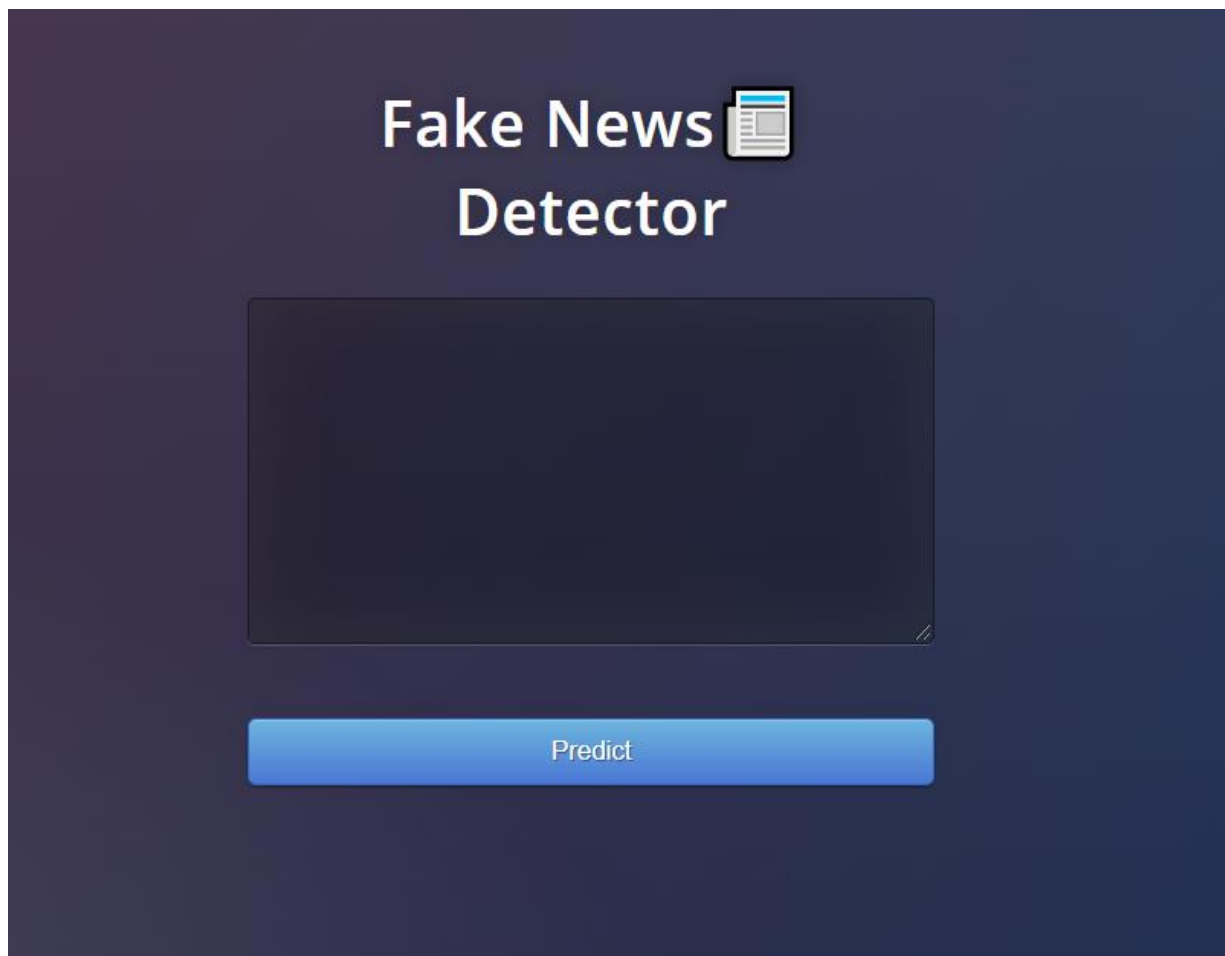
CHAPTER 6

RESULTS

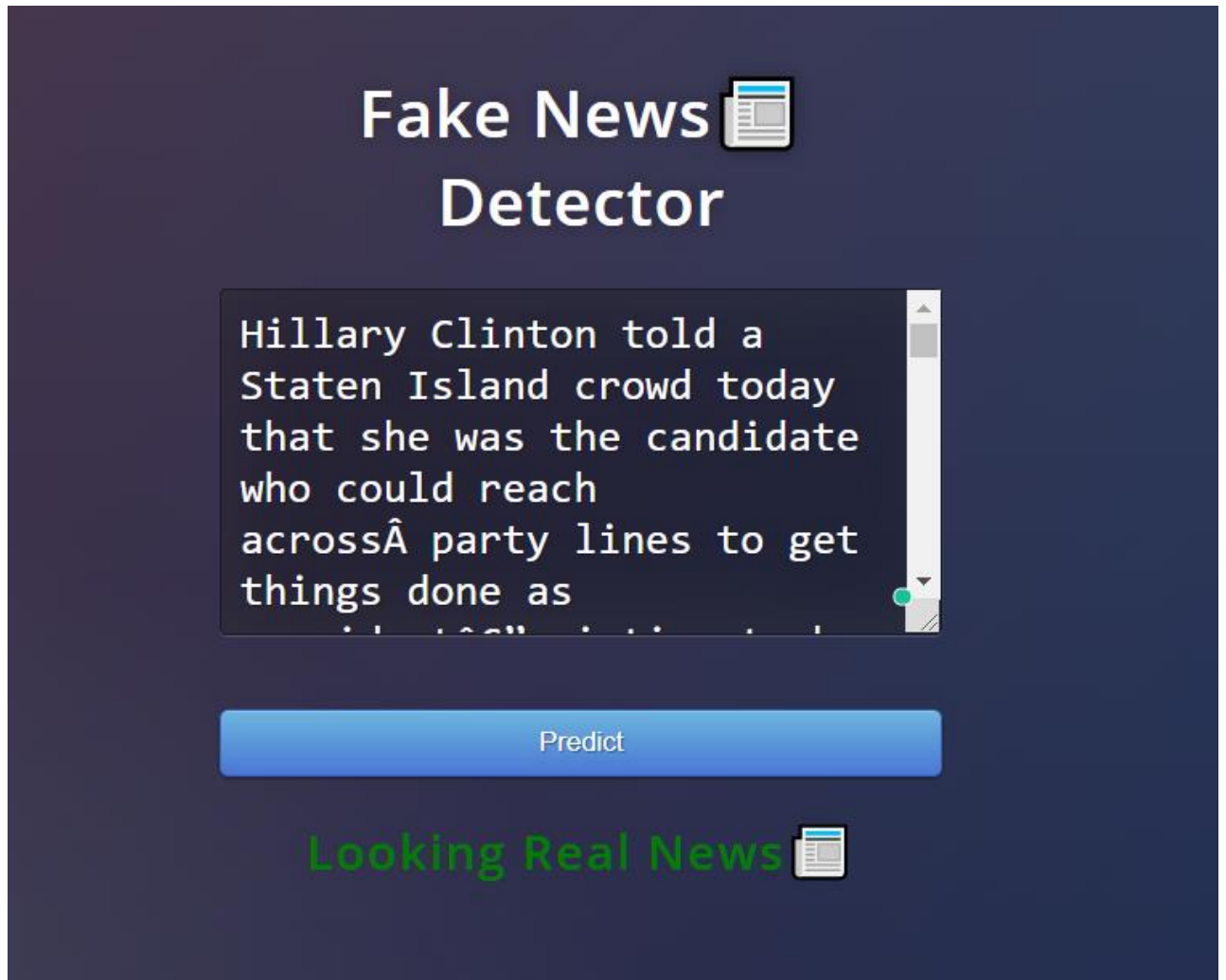
Accuracy :

```
In [20]: y_pred = classifier.predict(tfid_x_test)
score = accuracy_score(y_test,y_pred)
print(f'Accuracy: {round(score*100,2)}%')
```

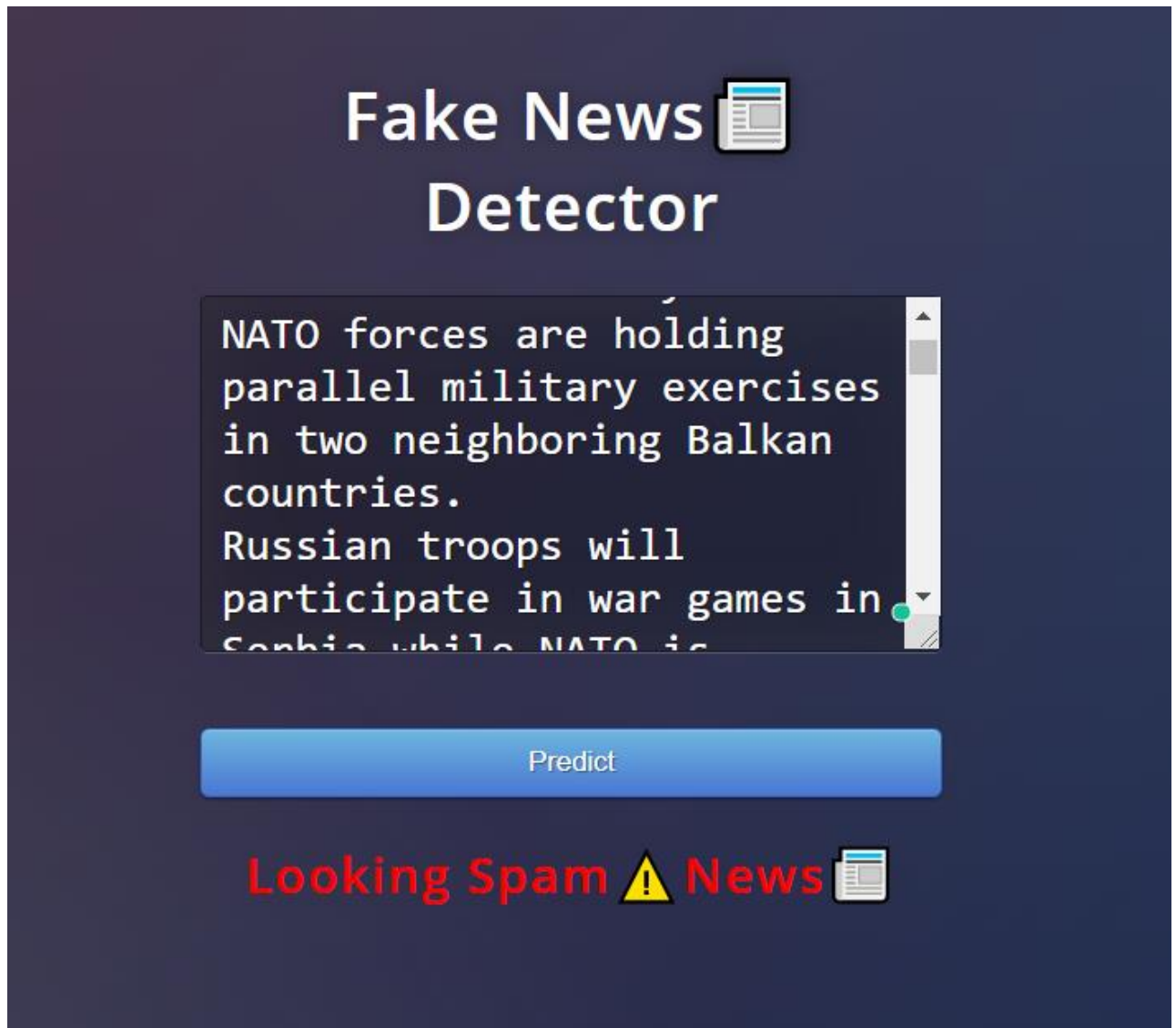
Accuracy: 93.69%



Real News



Fake News



CHAPTER 7

CONCLUSION

7.1 Conclusion

Social media popularity is increasing day by day with the huge number of users and more people consume news from social media rather than traditional news media, and a huge rate of fake news is spread on social media which causes negative impacts on the users. Political parties, some journals, and news channels spread fake news for their gain. Fake information spreads faster than real news so the impact will be very high on the people following it without knowing the authenticity of the news. So the model was proposed with the help of machine learning algorithms to detect and predict fake news. Fake news spreading is one of the biggest threats to democracy and misleads the people so fact-checking and fake news detecting model is necessary.

REFERENCE

- [1] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. *SIGKDD Explor. Newsl.* 19, 1 (June 2017), 22–36. DOI:<https://doi.org/10.1145/3137597.3137600>
- [2] Xinyi Zhou, Reza Zafarani, Kai Shu, and Huan Liu. 2019. Fake News: Fundamental Theories, Detection Strategies and Challenges. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (WSDM '19)*. Association for Computing Machinery, New York, NY, USA, 836–837. DOI:<https://doi.org/10.1145/3289600.3291382>
- [3] J. C. S. Reis, A. Correia, F. Murai, A. Veloso and F. Benevenuto, "Supervised Learning for Fake News Detection," in *IEEE Intelligent Systems*, vol. 34, no. 2, pp. 76-81, March-April 2019, doi: 10.1109/MIS.2019.2899143.
- [4] A. Uppal, V. Sachdeva and S. Sharma, "Fake news detection using discourse segment structure analysis," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2020, pp. 751-756, doi: 10.1109/Confluence47617.2020.9058106.
- [5] S. I. Manzoor, J. Singla and Nikita, "Fake News Detection Using Machine Learning approaches: A systematic Review," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 230-234, doi: 10.1109/ICOEI.2019.8862770.
- [6] Meital Balmas. When fake news becomes real: Combined exposure to multiple news sources and political attitudes of inequality, alienation, and cynicism. *Communication Research*, 41(3):430-454, 2014
- [7] Michele Banko, Michael J Cafarella, Stephen Soderland, Matthew Broadhead, and Oren Etzioni. Open information extraction from the web. In *IJCAI'07*.