



Sree Dattha Institute of Engineering and Science  
Department of Computer Science and Engineering

---



# Fake News Detection

Vamshi Ragipani	18E41A05B5
Nikhitha Kollaram	18E41A05B3
Lokesh Ratnala	18E41A05G1
Gayathri Chandragiri	18E41A05C7

Under the Guidances of Ms.B.Gnaneswari  
( Assistante Professor, Department of CSE)



# Contents

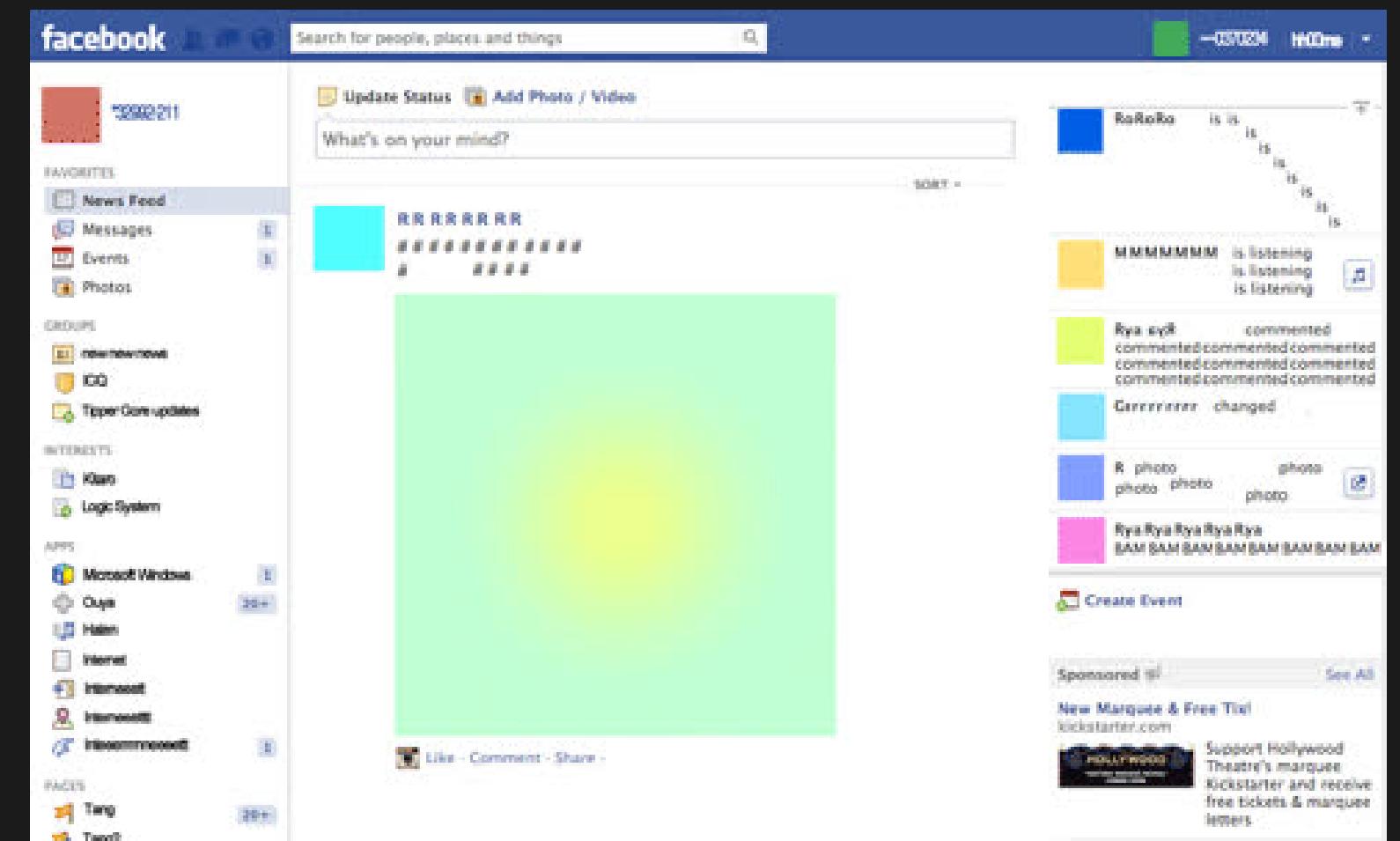
- Abstract
- Introduction
- Problem Statement
- Existing System
- Purpose
- Tools and Technologies
- Data set
- Data Analysis
- Workflow
- Architecture
- Methodology
- Result
- Conclusion

# What is Fake News?

- Fake News is news designed to deliberately spread hoaxes, propaganda, and disinformation

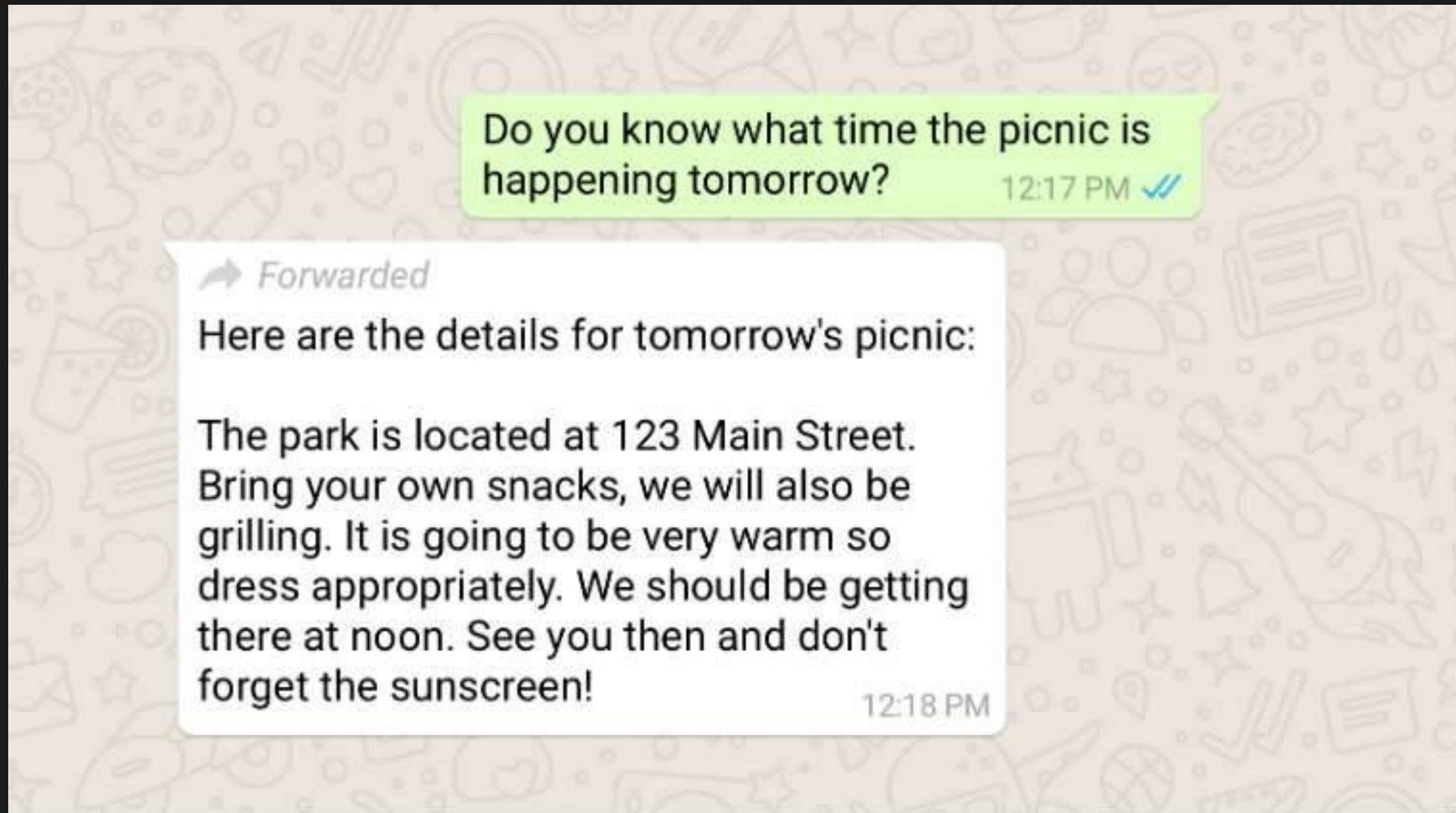


- By clicking on a **clickbait**, users are led to a page that contains false information. Fake news influence people's perceptions.



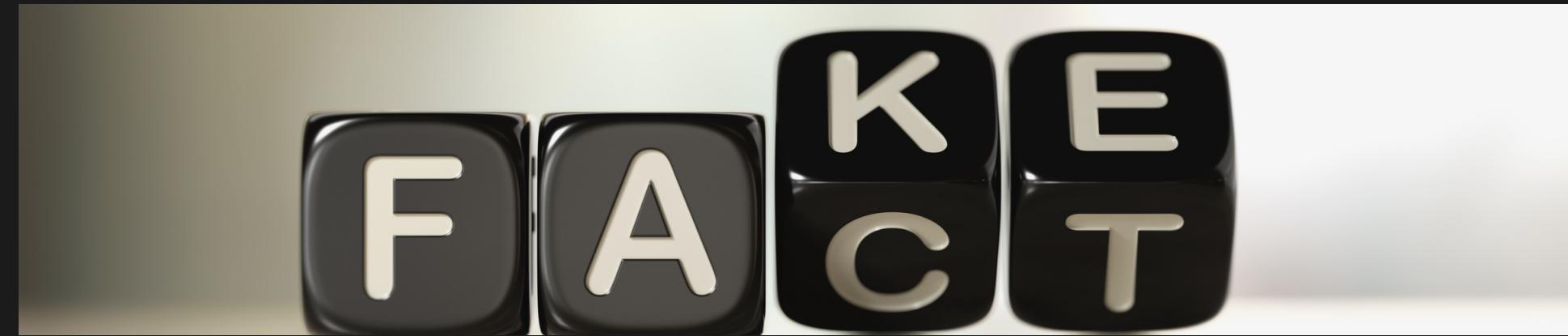
- Fake News stories are usually spread through **Social Media** sites like Facebook, Twitter, Instagram etc..

# How many of you receive forwarded news on WhatsApp?



# Problem Statement

- **Fake news** posted by famous figures on social media spreads faster because people do not fact-check what is shared with them and blindly follow the crowd. Thus, people cannot differentiate between facts and online propaganda.



- which often leads to real-life consequences regardless of being good or bad. People need a transparent medium for fact-checking news to make informed decisions without contributing to a **social problem**

# Why is Detecting Fake News So Important?



98,000

Tweets are shared on Twitter



600

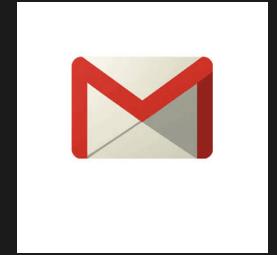
videos are uploaded on  
YouTube



11

Million instant  
messages are sent

Every  
60  
Seconds



160

Million emails are sent



1500

blog entries are created

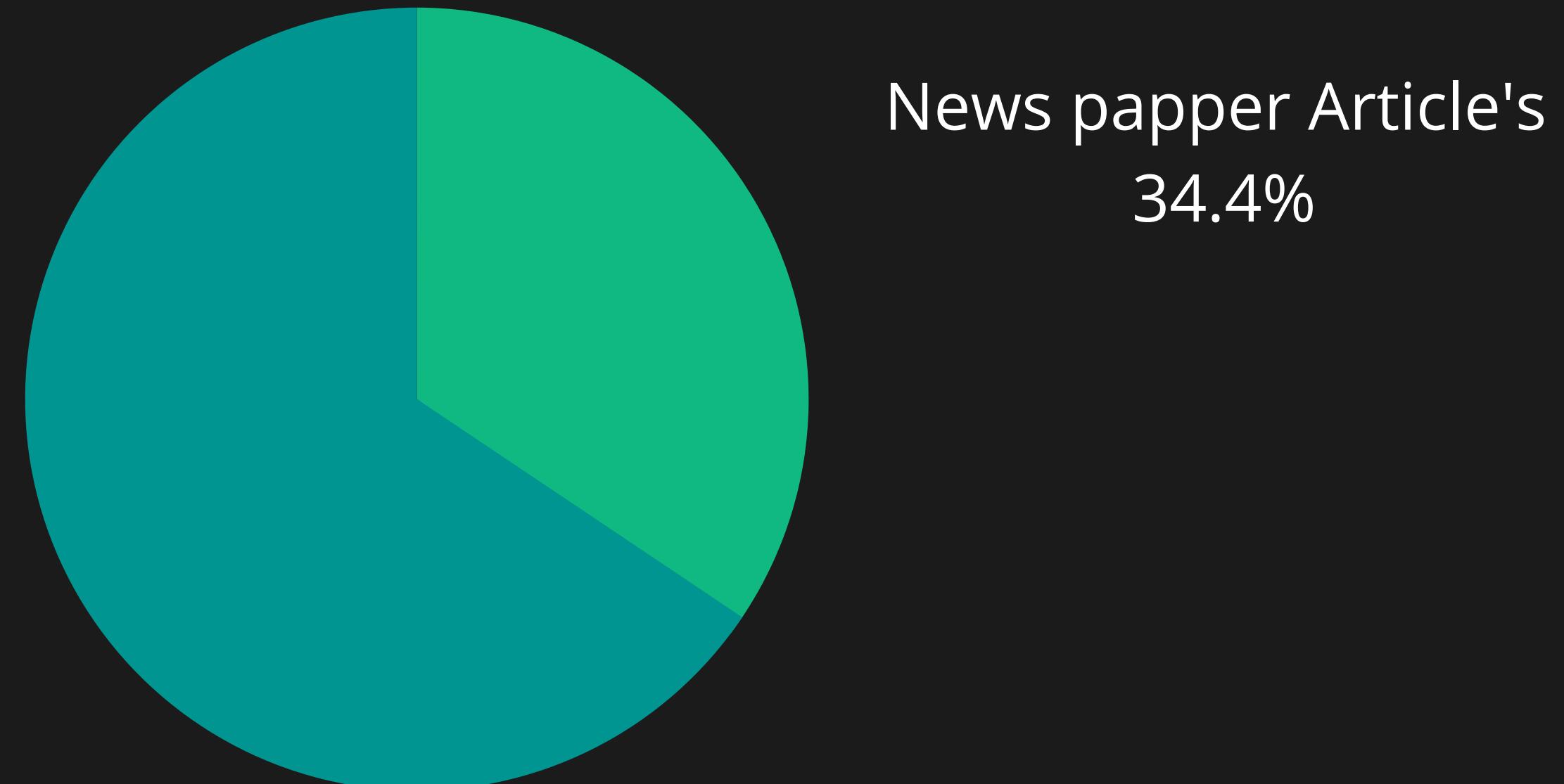


6,00,000

Status updates &  
posts are created

# **Users preferred to use Social Media Apps for getting news rather than Newspaper Articles.**

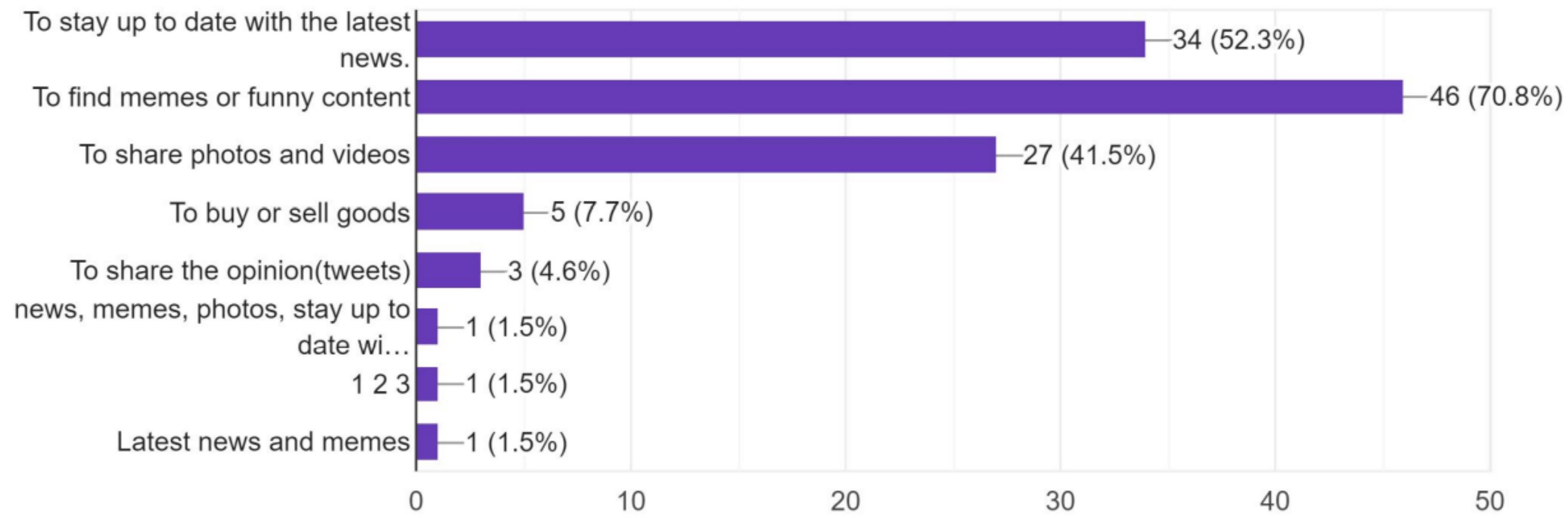
Do you prefer using newspaper articles or social apps to get your preferred news?



# A Major portion of the users used social media platforms to keep track of the latest news after searching content for entertainment

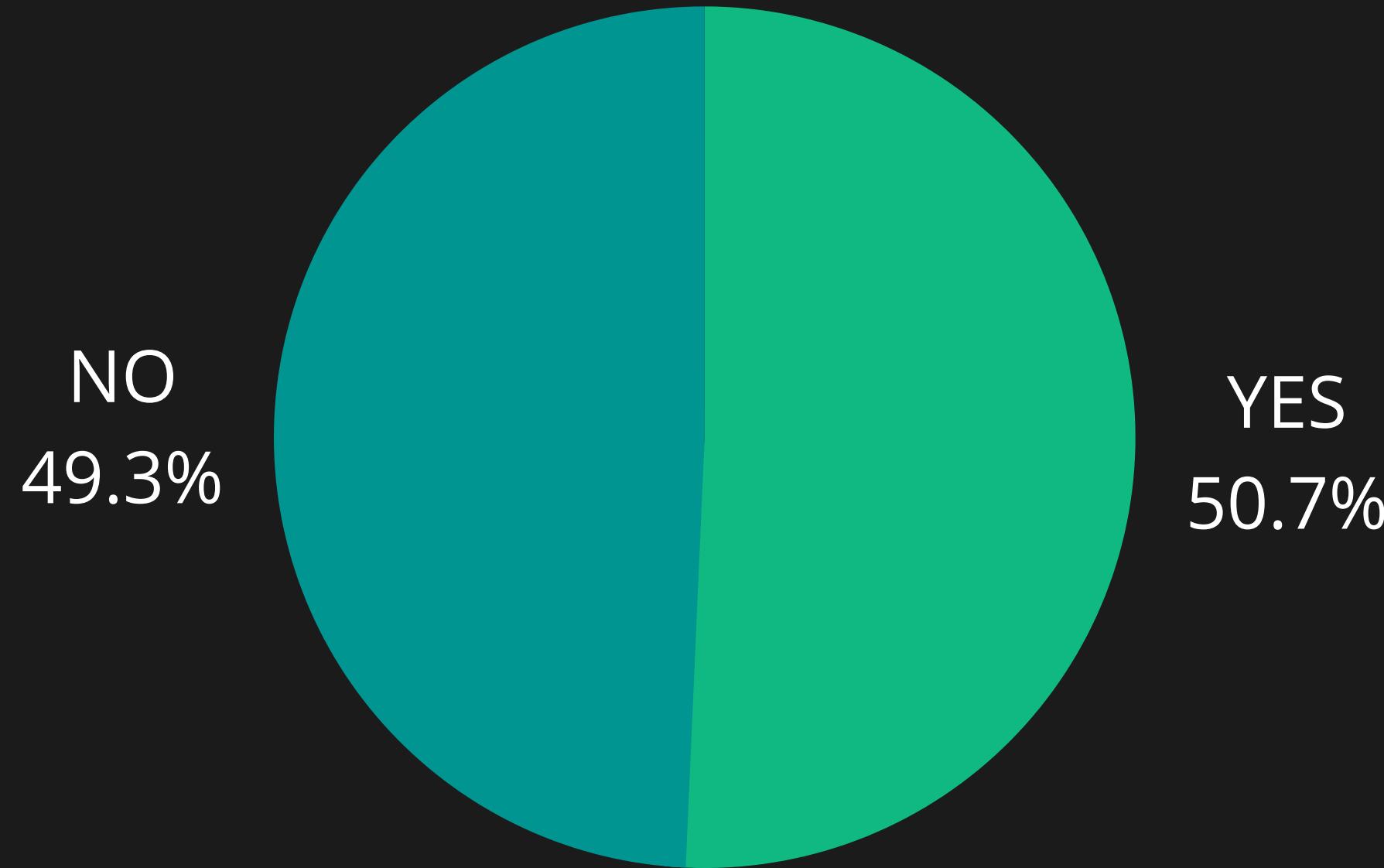
For what purposes do you use social media platforms?(Mark all that apply)

65 responses



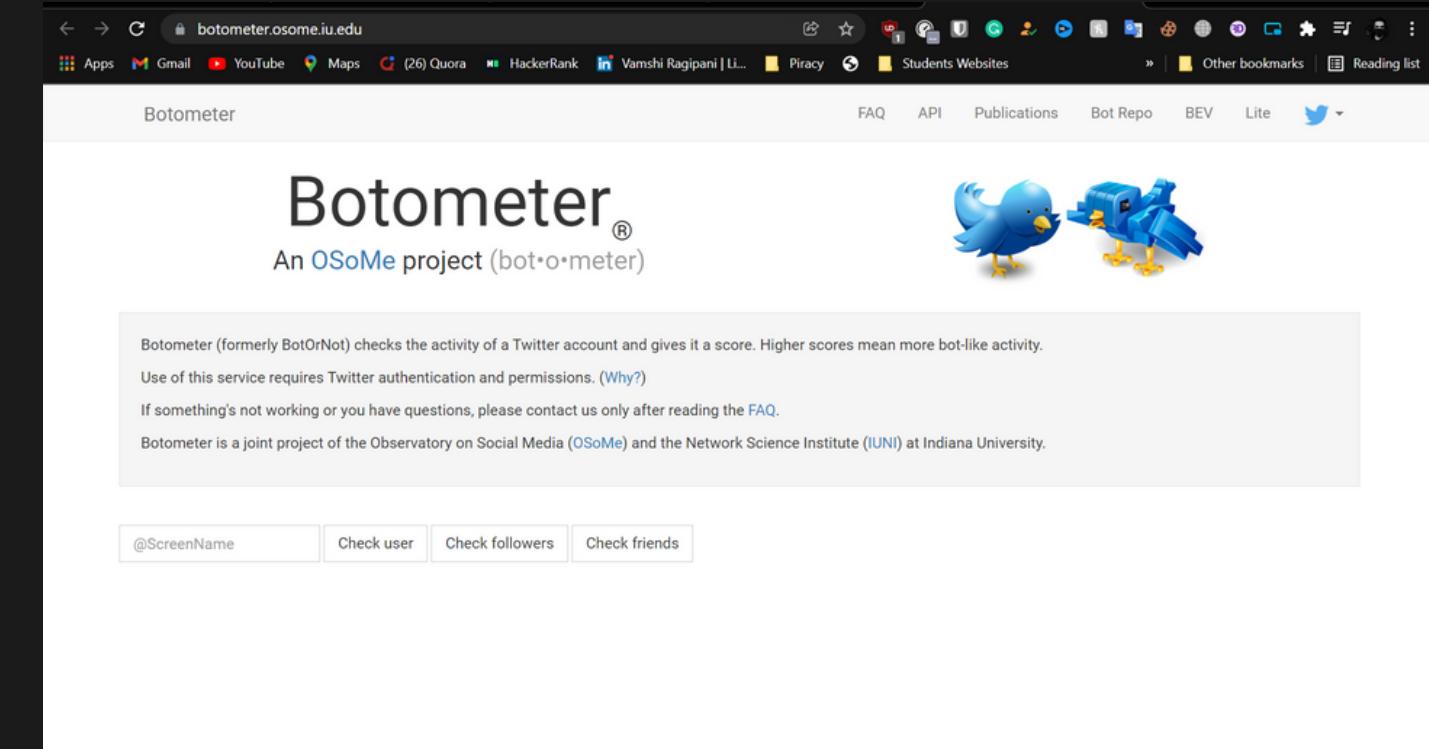
# More than 50% of the users fall for Fake news on social media platforms

Have you ever been the victim of fake news?

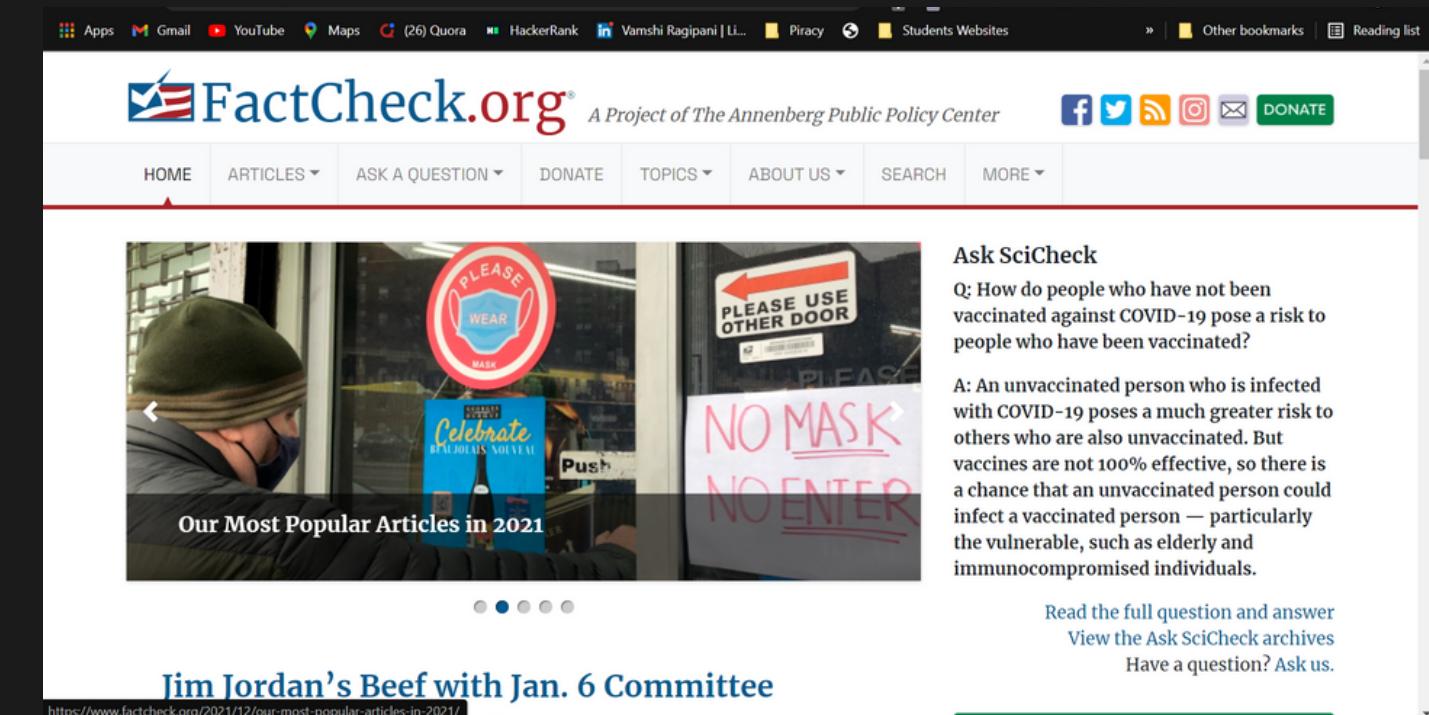


# Existing System

**Botometer:** Indiana University created Botometer (formerly “BotOrNot”) as a response to the prevalence of fake bots on Twitter. The site rates account on a scale of one to five — one being real and five being fake — based on its history, tweets, and mentions.



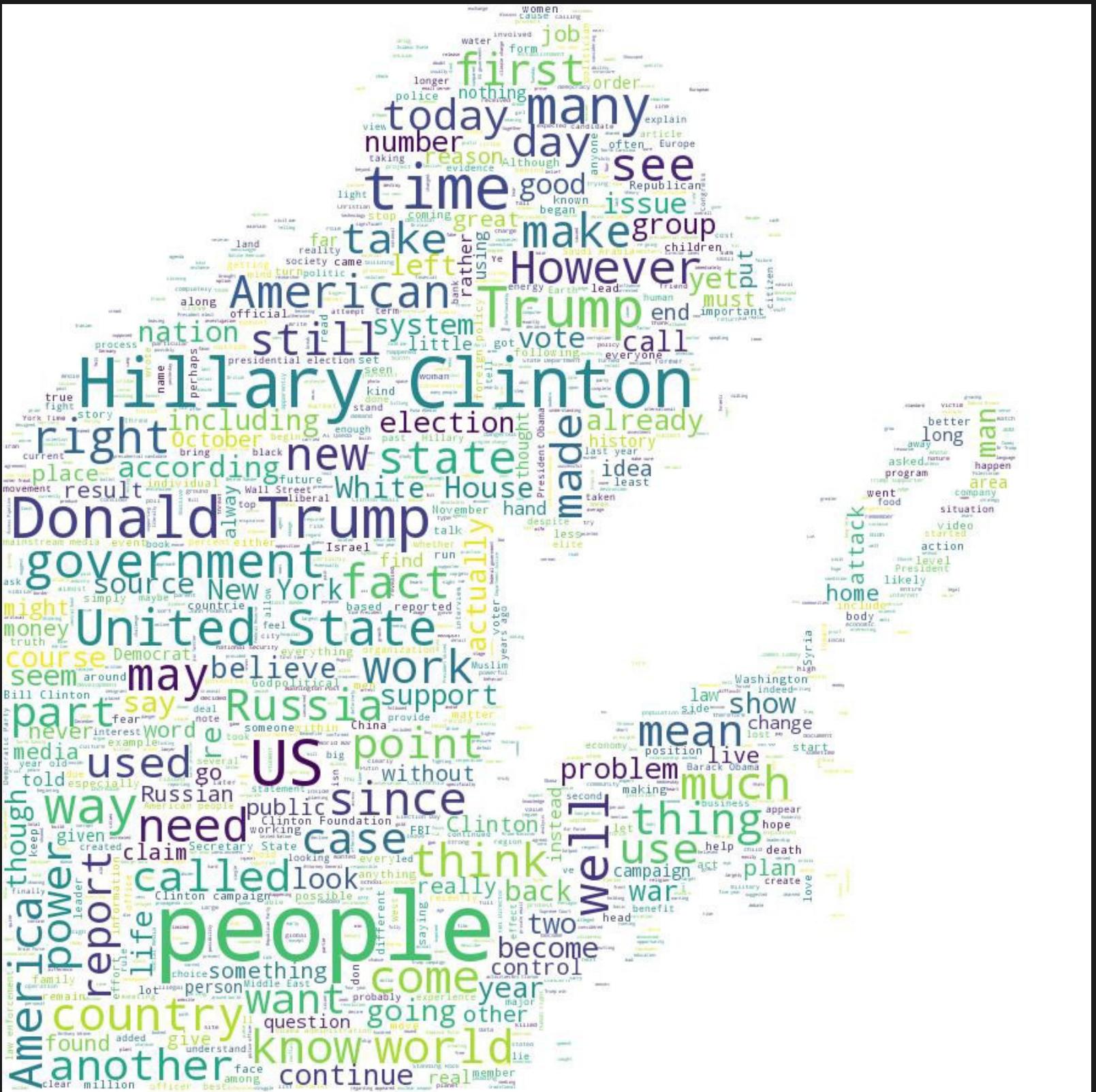
**FactCheck.org:** Users can ask FactCheck.org questions about the validity of political claims and information, and the team behind the site will investigate and explain them truthfully.



# HOW TO RECOGNIZE A **FAKE** **NEWS** **STORY**

- 1 READ PAST THE HEADLINE
- 2 CHECK WHAT NEWS OUTLET PUBLISHED IT
- 3 CHECK THE PUBLISH DATE AND TIME
- 4 WHO IS THE AUTHOR?
- 5 LOOK AT WHAT LINKS AND SOURCES ARE USED
- 6 LOOK OUT FOR QUESTIONABLE QUOTES AND PHOTOS
- 7 BEWARE CONFIRMATION BIAS
- 8 SEARCH IF OTHER NEWS OUTLETS ARE REPORTING IT
- 9 THINK BEFORE YOU SHARE

# Detecting Fake News with Natural Language Processing (NLP)



- As Human beings when we read a sentence or a paragraph we can interpret the words with the whole document and understand the context.
  - Given today's volume of news, it is possible to teach a computer how to read and understand the difference between real news and fake news using NLP. The building blocks are Data Set and Machine Learning Algorithm

# PURPOSE

- This project aims to develop a method for detecting and classifying fake news stories using Natural Language Processing.
- Our goal is to develop a model that classifies given news articles as either fake or true.
- We gathered our data, processed the text, and translated our articles into supervise model features.



# TOOLS AND TECHNOLOGIES

- TOOLS:



WINDOWS (OS)



Jupyter Notebook



Microsoft Office excel.



Visual Studio Code



Spyder

- LANGUAGES:



Python



HTML

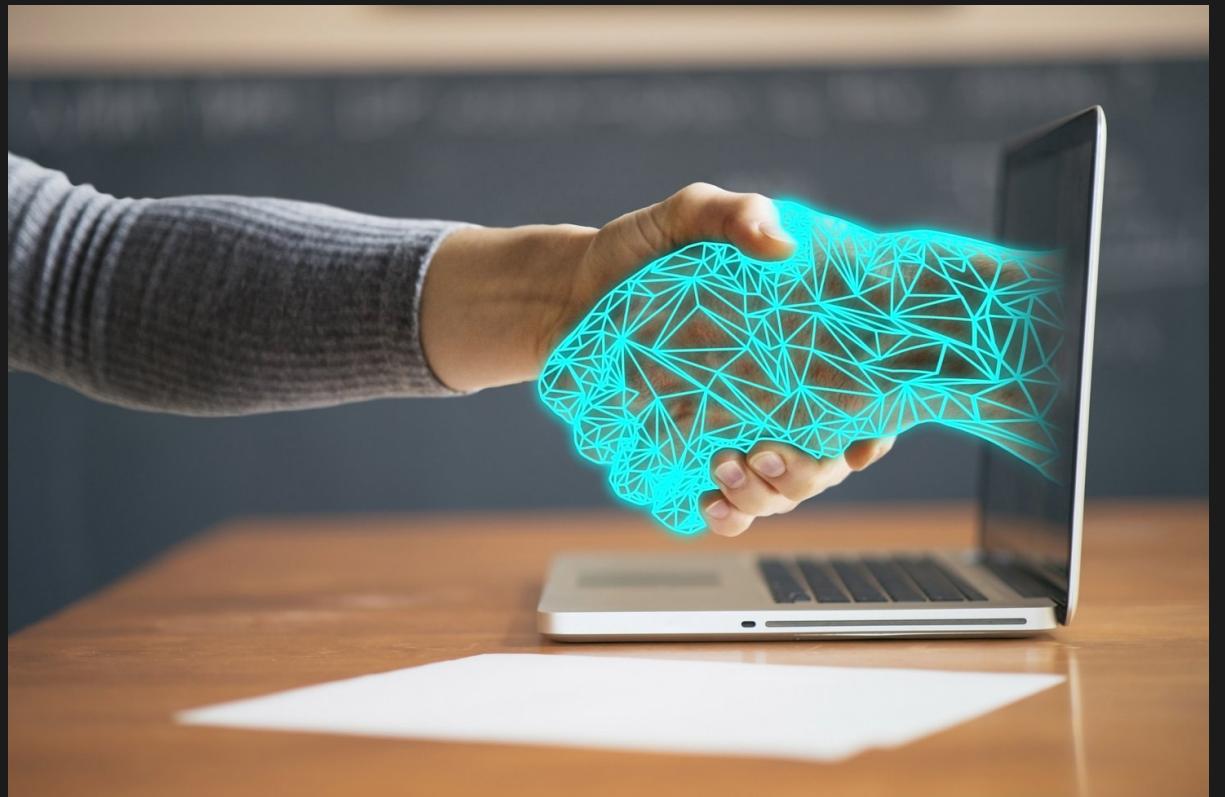


CSS

# TECHNOLOGY

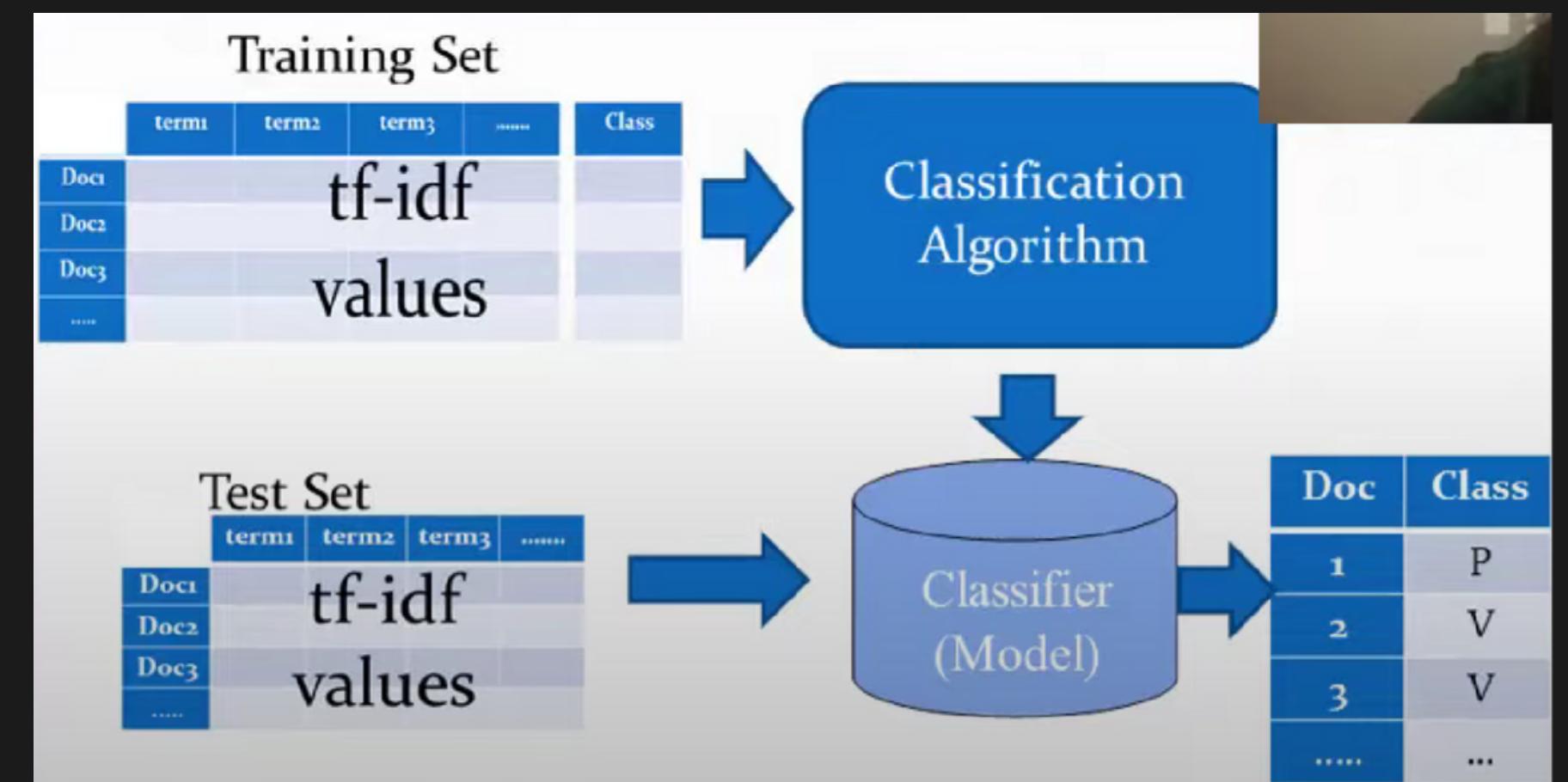
After Vectorization, we split the data into test and train data.

- WE fit ML models to the data
- Natural language processing (NLP)
- Logistic Regression,
- Naive-Bayes,
- Decision Tree,
- Passive-Aggressive Classifier



# ARCHITECTURE

- Extract the features from the pre-processed data using TF-IDF Vectorization
- Split the dataset into training data and testing data
- Using classification algorithm train the model with feature extracted training data
- Test and validate the model with the test data
- We used a naive basis algorithm to model and validate the data
- Compare the accuracy of the model.



# DATA SET

G2	A	B	C	D	E
S	title	text		label	
1	8476 You Can Smell Hillaryâ€™s Daniel			FAKE	
2	10294 Watch The Exact Moment Google Pinterest			FAKE	
3	3608 Kerry to go to Paris in U.S. Secretary of ges			REAL	
4	10142 Bernie supporters on Twitter Kaydee King			FAKE	
5	875 The Battle of New York: It's primary day			REAL	
6	6903 Tehran, USA				
7	7341 Girl Horrified At What She Share This				
8	95 â€˜Britainâ€™s Schindler A Czech stockbroker who saved more than 650.				
9	4869 Fact check: Trump and Clinton				
10	2909 Iran reportedly makes new Iranian				
11	1357 With all three Clintons in CEDAR RAPIDS,				
12	988 Donald Trumpâ€™s Shocking Donald				
13	7041 Strong Solar Storm, Tech Click Here To				
14	7623 10 Ways America Is Prepared for October 31,				
15	1571 Trump takes on Cruz, but Killing Obama administration rules, dismantling				
16	4739 How women lead differently As more women				
17	7737 Shocking! Michele Obama Shocking!				
18	8716 Hillary Clinton in HUGE Tr				
19	3304 What's in that Iran bill the Washington				

- The dataset we are using the relevant news during the US presidential election to analyze, and also the label published by signal media at Kaggle competition as ground truth (supervised learning).
- **30% of the data is used for Training the ML Model**
- **70% of the data is used to test the model**

# DATA ANALYSIS

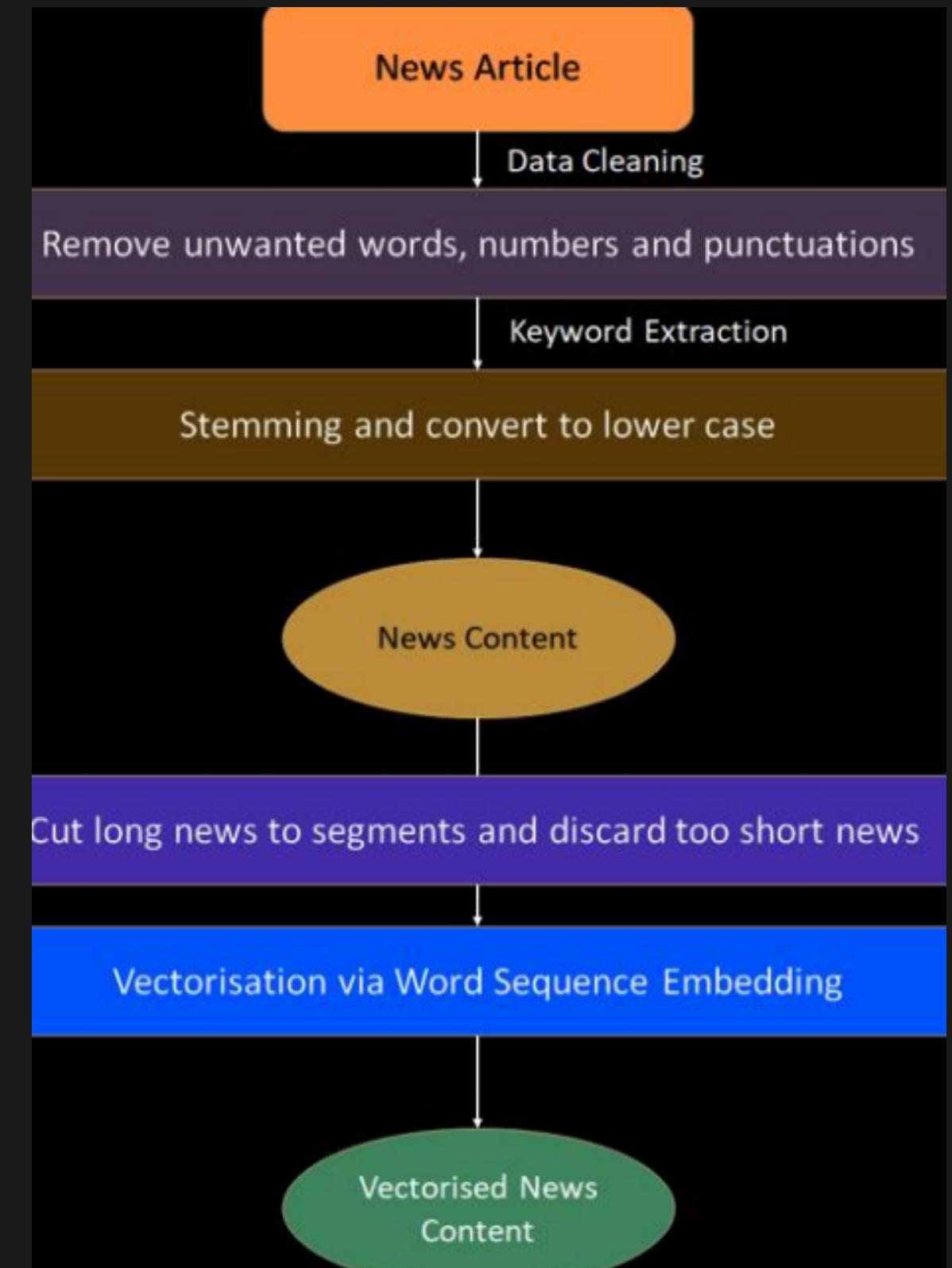
- we have used the CSV dataset. The dataset contains 7796 rows and 4 columns
- This dataset has four columns,
- **Title**: this represents the title of the news.
- **Author**: this represents the name of the author who has written the news.
- **Text**: this column has the news itself.
- **Label**: this is a binary column representing if the news is fake (1) or real (0)



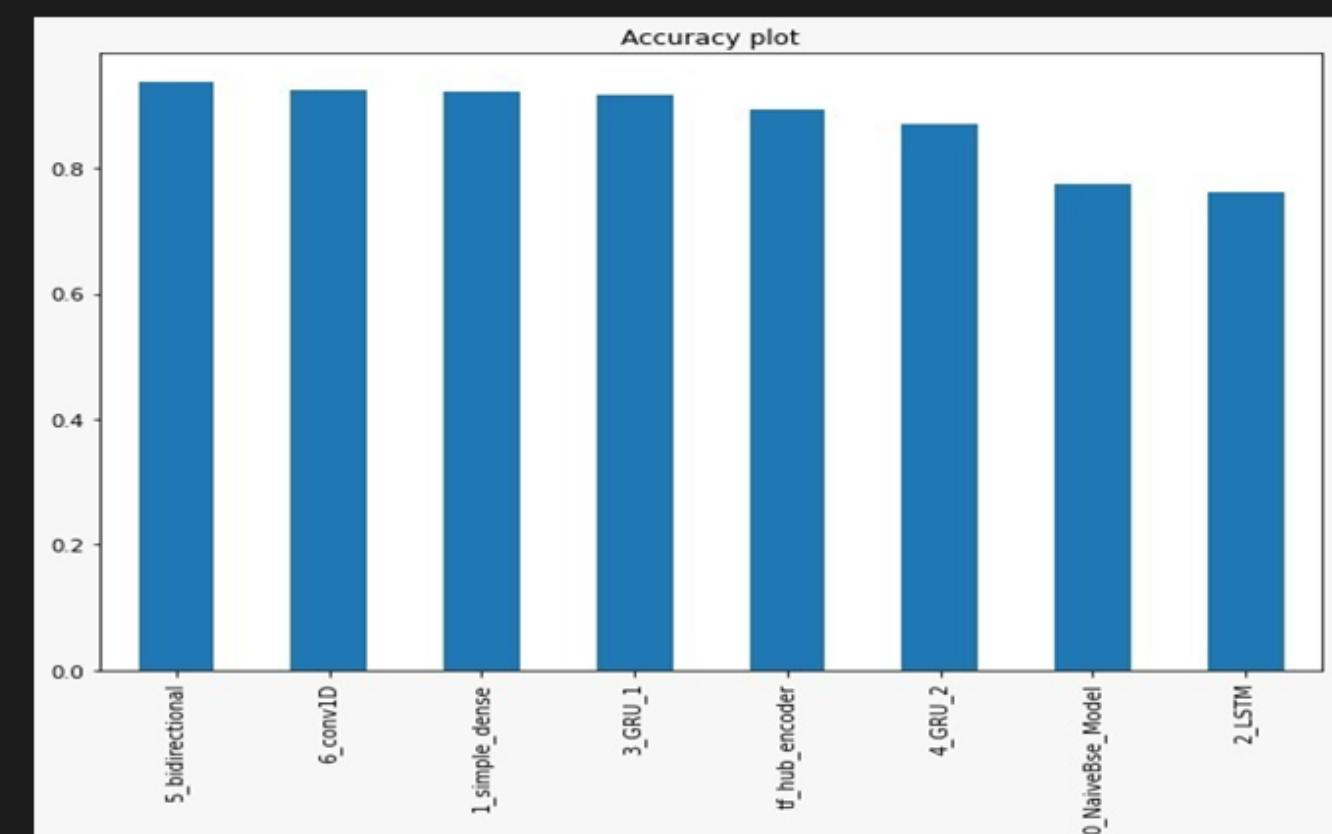
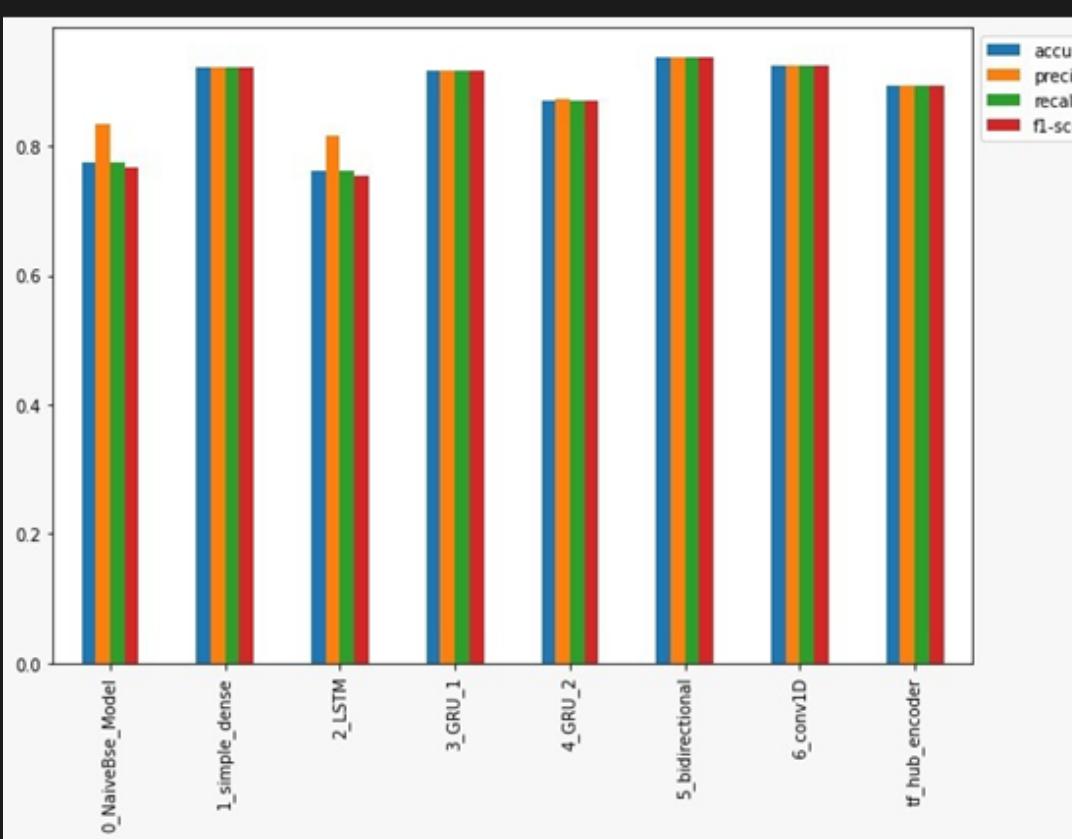
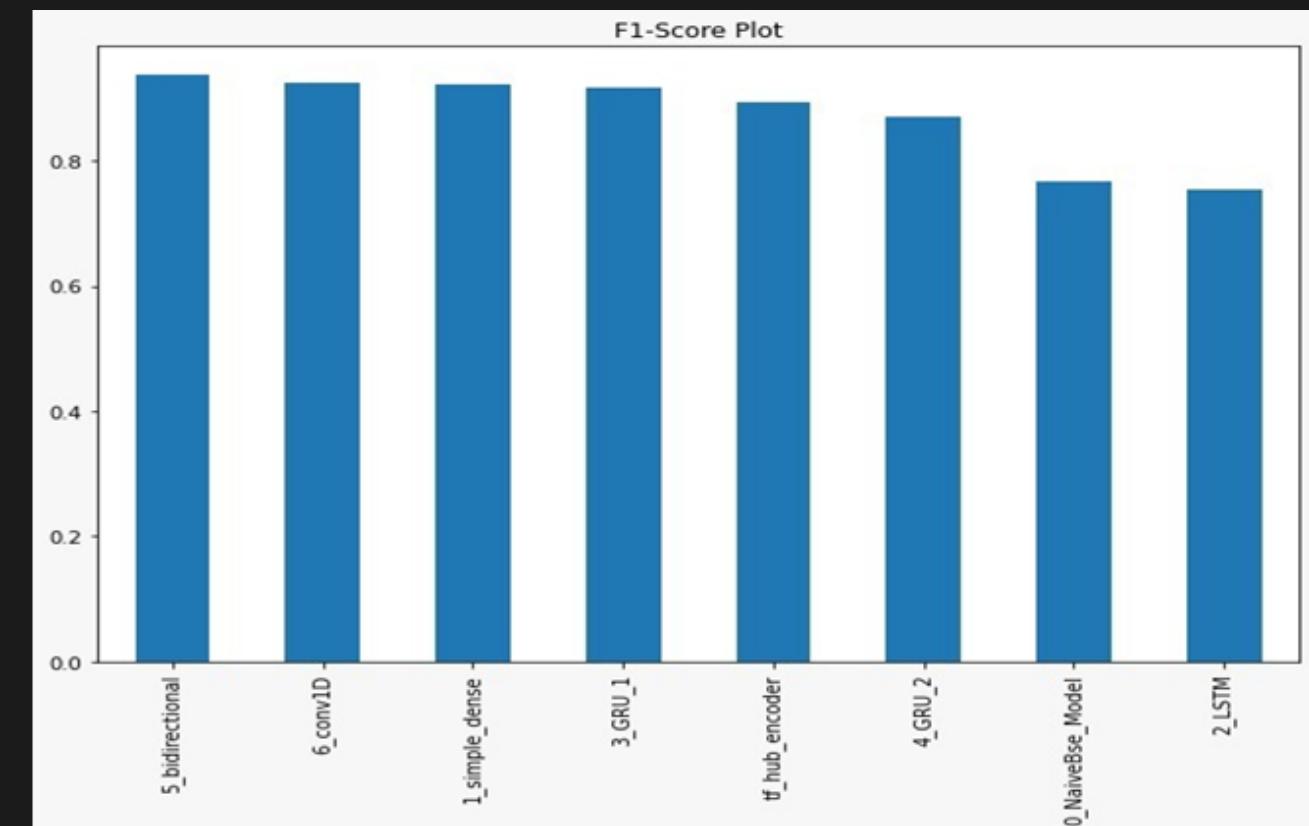
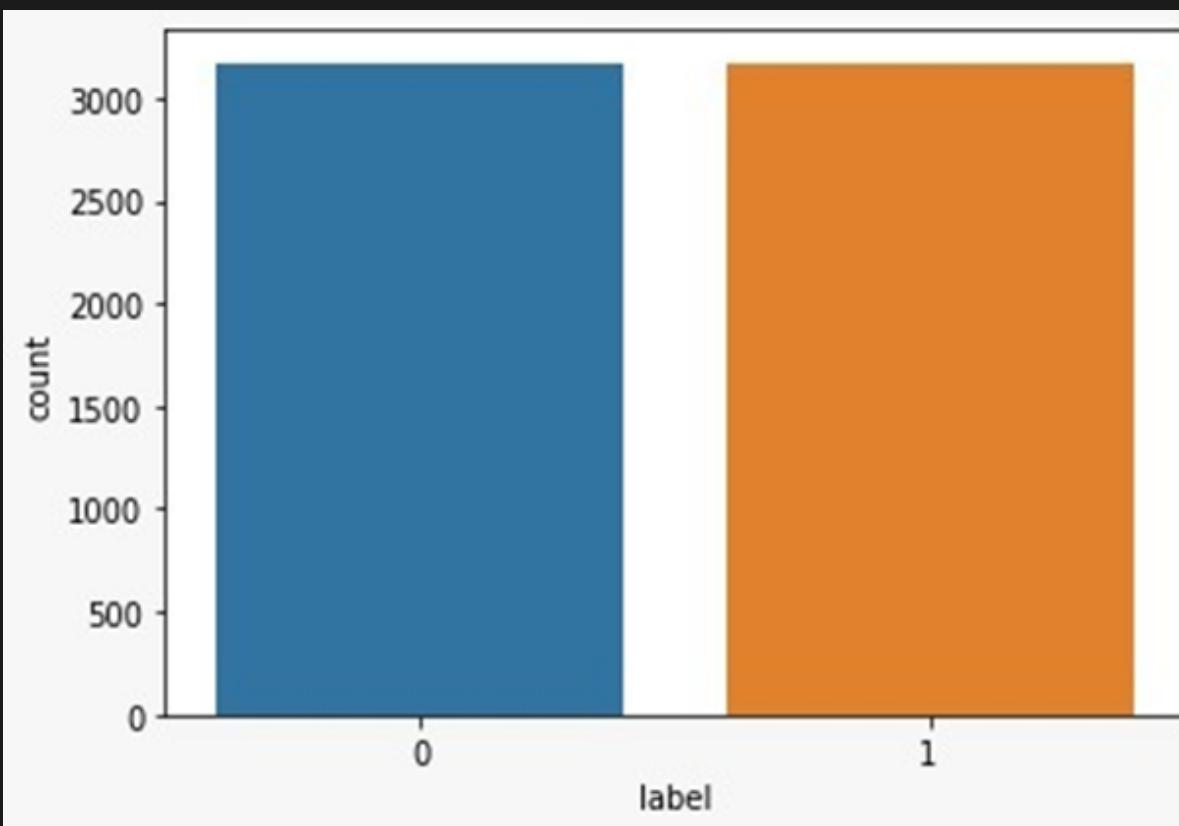
# WORKFLOW

The steps in this process are as follows:

- Dataset Loading
- Data Pre-Processing (Remove StopWords, Stemming, Drop duplicate, and Remove meaningless char from the text.)
- Features Selection
- Applying classification and model construction
- Classifying the new data..

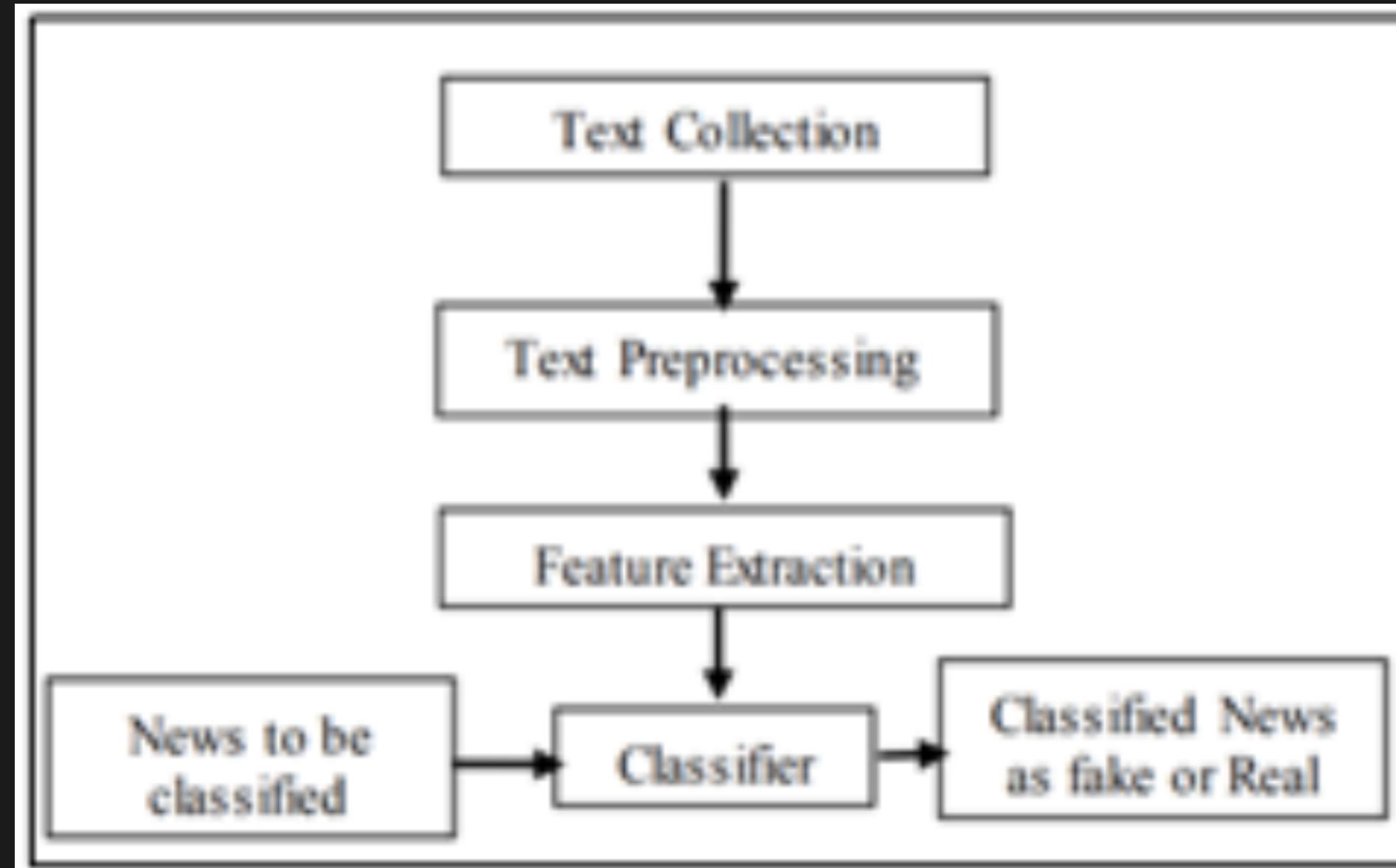


# Graphs



# METHODOLOGY

The Fake news model detection is built using steps like Text Collection, Text Preprocessing, Feature Extraction, and then finally classification using different classifiers.



# RESULT

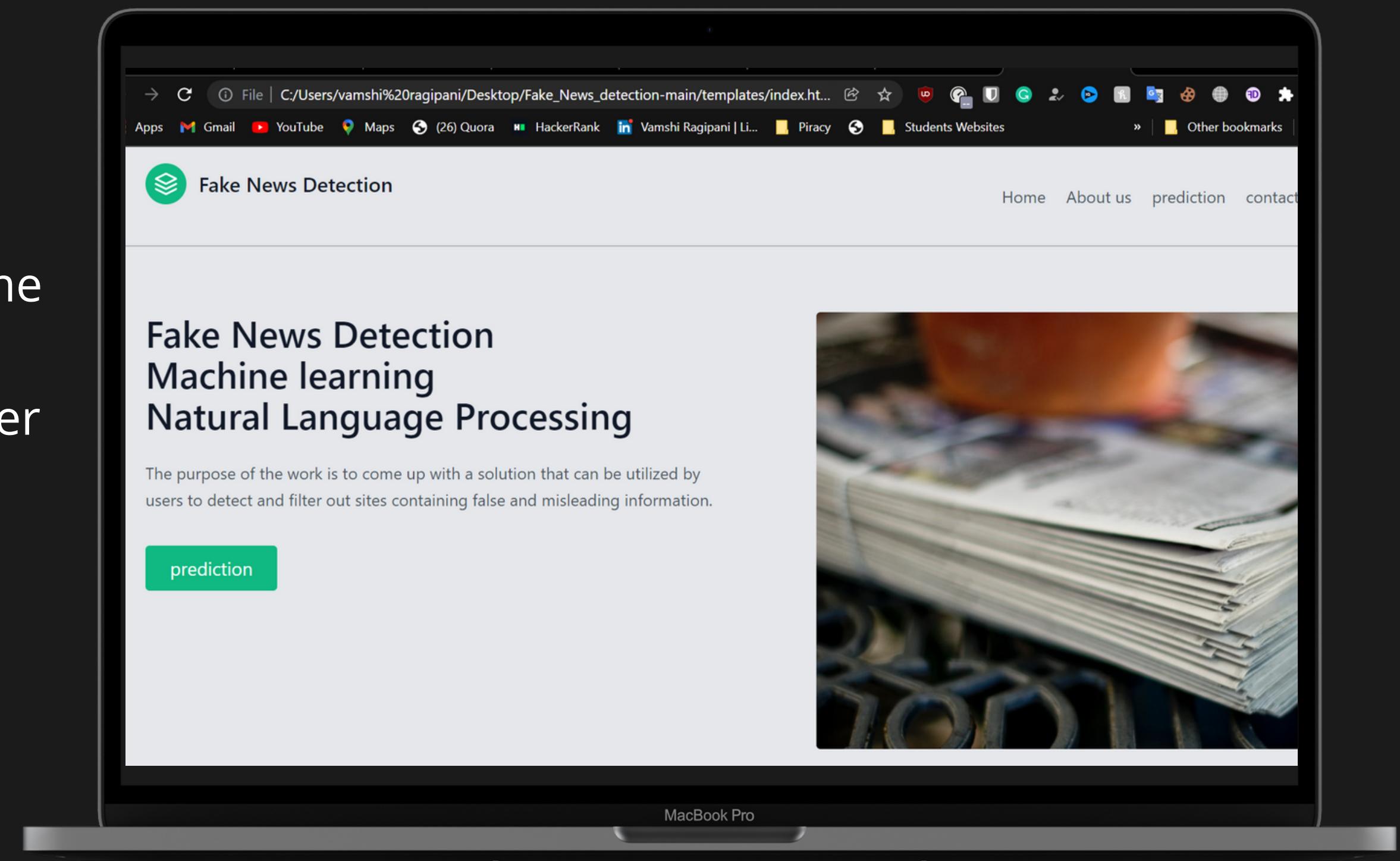
- We use a confusion matrix to calculate the accuracy value.
- PA classifier is more accurate than when data increases Logistic Regression, Naive-Bayes, Decision Tree
- When the training data size is approx. 80 rows, the PA classifier has higher accuracy of 83%, and Logistic Regression 91.73%, Naive-Bayes 82.32 %, Decision Tree 80.49%
- But when the training data size is 7900 rows, the PA classifier accuracy is 96.88%

```
Name: label, Length: 63, dtype: object
Accuracy of PA classifier is: 83.04 %f
Confusion Matrix for PA classifier is:
[[2467 664]
 [ 400 2741]]
Accuracy: of knn 75.51 %f
KNN
```

```
Name: label, Length: 63, dtype: object
Accuracy of PA classifier is: 83.04 %f
Confusion Matrix for PA classifier is:
[[2467 664]
 [ 400 2741]]
Accuracy: of knn 75.51 %f
KNN
```

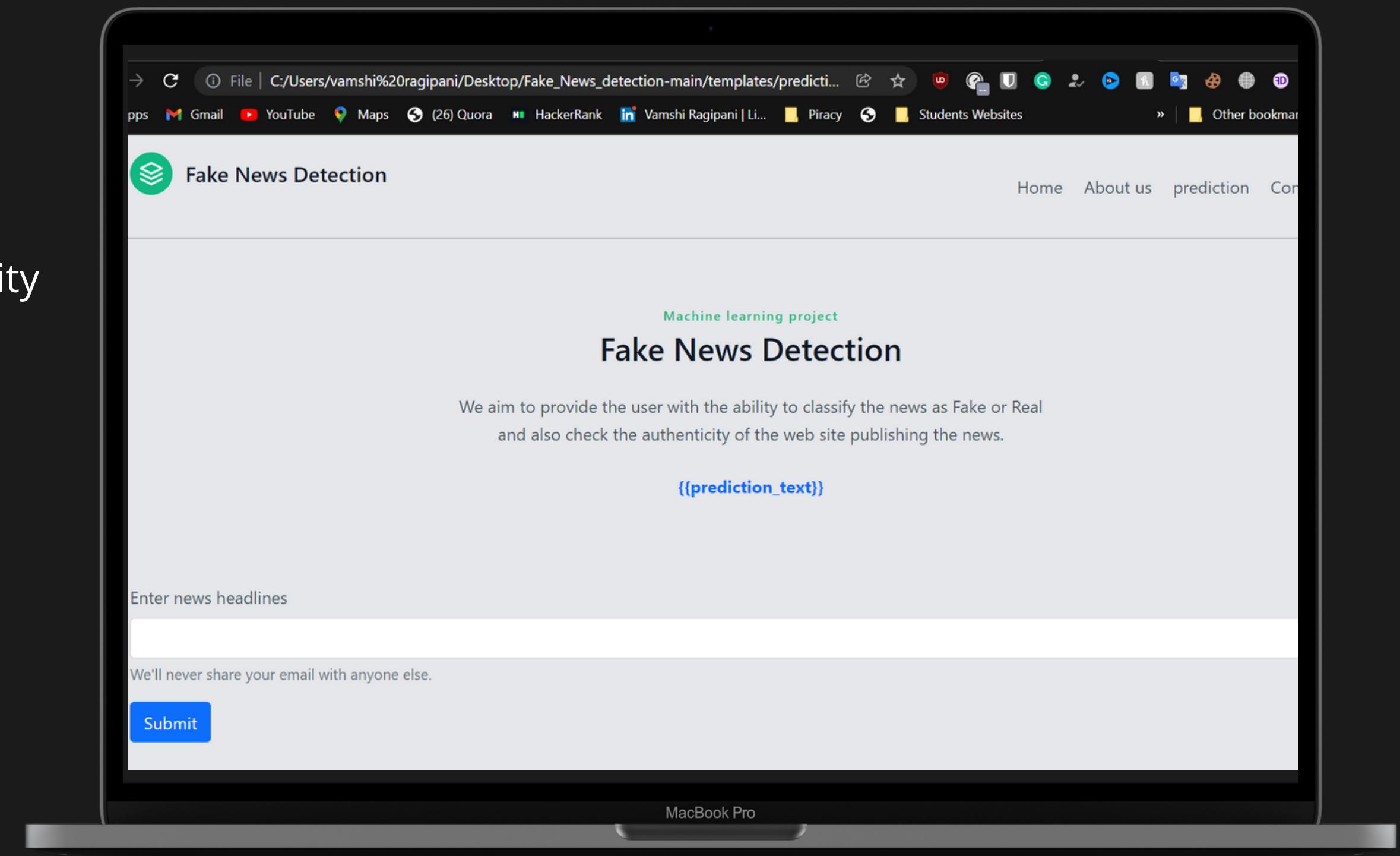
# Front End

The purpose of the work is to come up with a solution that can be utilized by users to detect and filter out sites containing false and misleading information.



# Front End

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.



# CONCLUSION

- The passive-aggressive classifier performed the best here and gave an accuracy of 93.12%.
- We can print a confusion matrix to gain insight into the number of false and true negatives and positives
- Collecting the data once isn't going to cut it given how quickly information spreads in today's connected world and the number of articles being churned out.



# Thank You

---