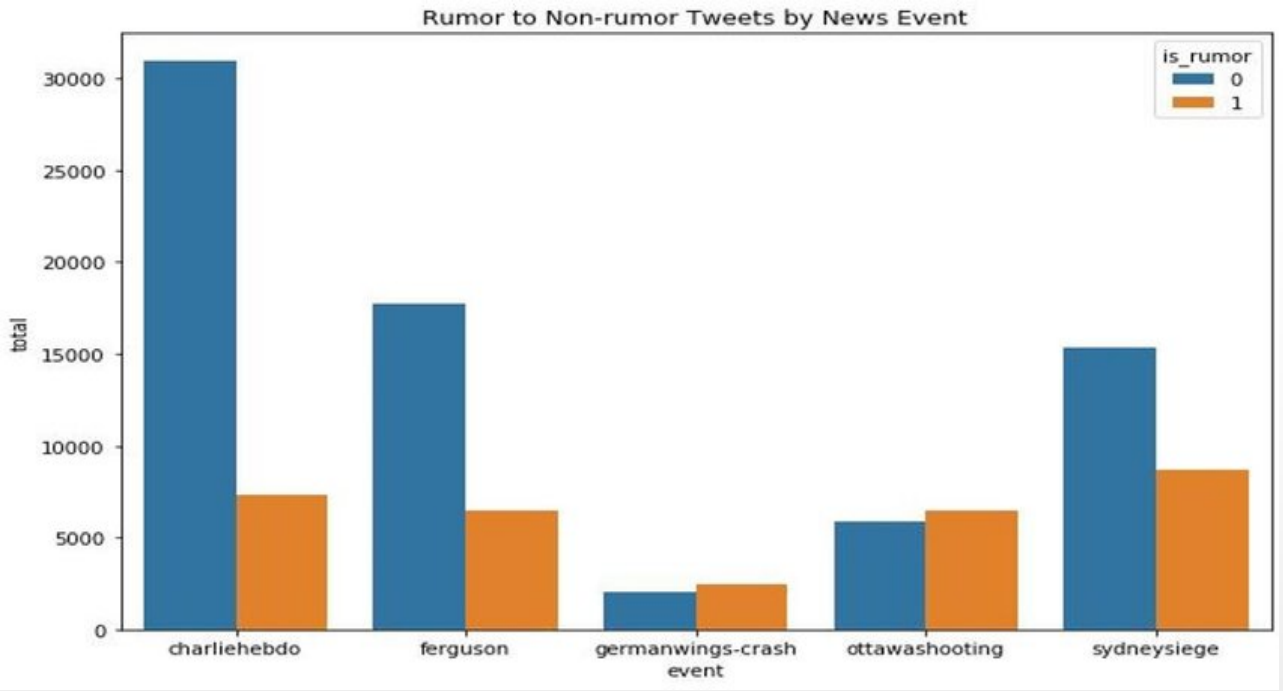
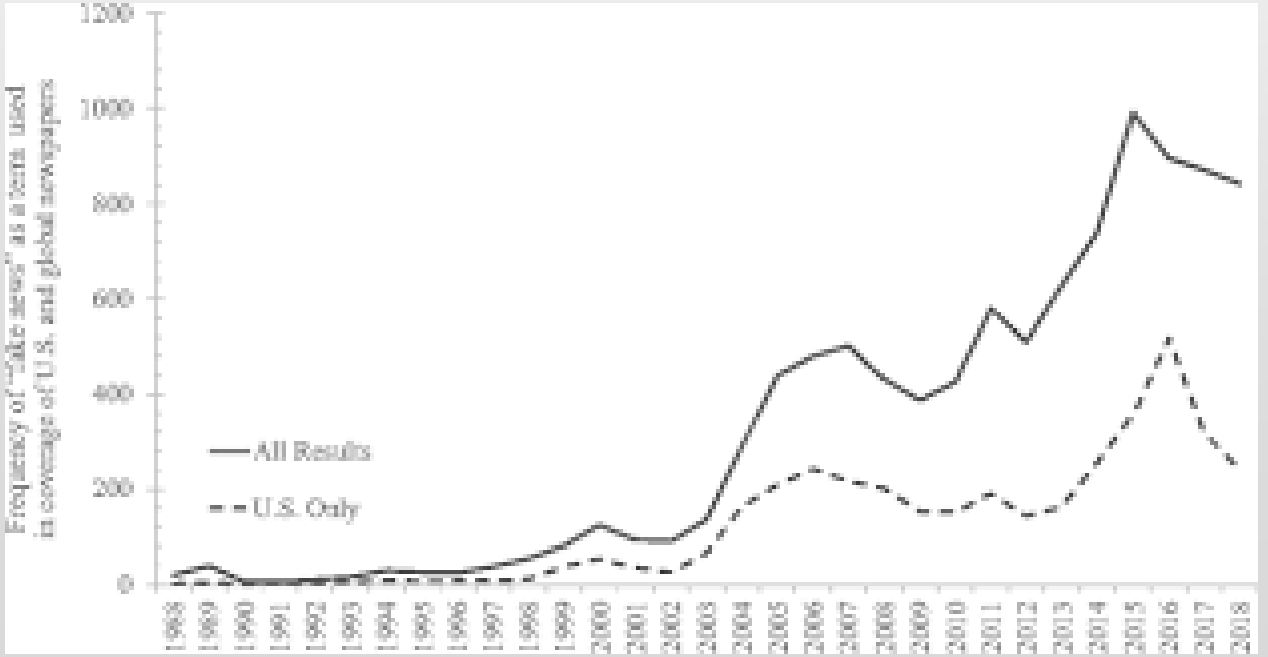
AIDS PROJECT:
FAKE NEWS DETECTION

PROBLEM STATEMENT

- The rise of social media and digital platforms has made it easier than ever to create and spread fake news, which can have serious consequences for individuals, organizations, and even society as a whole.
- The problem of detecting fake news has become increasingly urgent, as it is difficult for people to discern what information is true and what is not. Therefore, the development of reliable and effective tools for detecting fake news has become a pressing need.
- The successful development of such a model could have significant implications for the detection and prevention of fake news, as well as for the broader goal of improving the quality and reliability of information in the digital age.

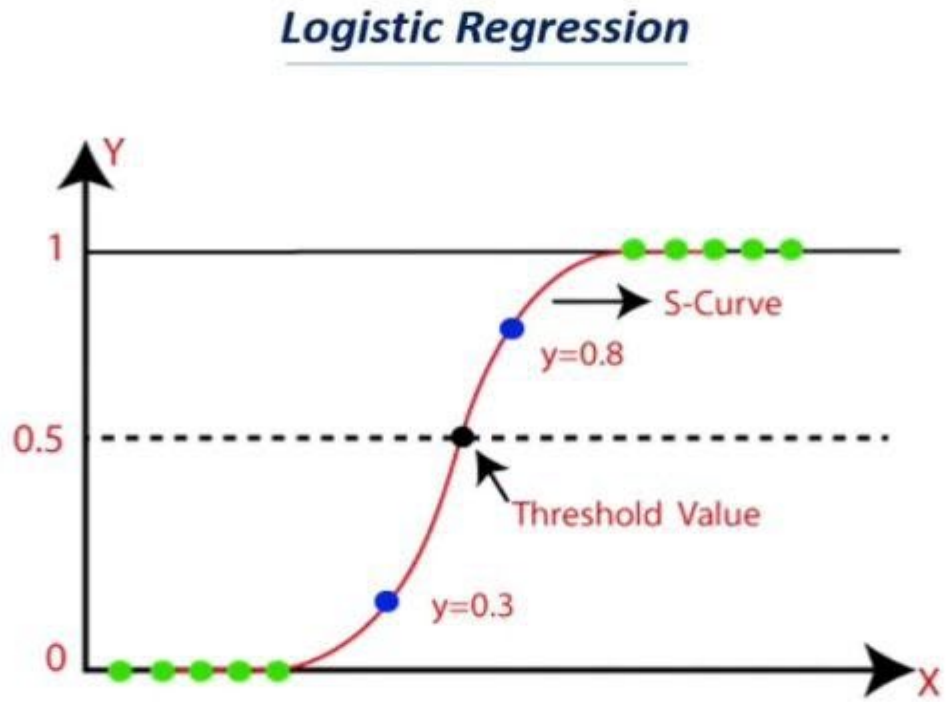

Rumor to Non-rumor Tweets by News Event
is_rumor
0
1
total
30000
25000
20000
15000
10000
5000
0
charliehebdo
ferguson
germanwings-crash
ottawashooting
sydneysiege
event


Frequency of "fake news" as a term used in coverage of U.S. and global newspapers
All Results
U.S. Only
1200
1000
800
600
400
200
0
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018

SOFTWARE AND HARDWARE REQUIREMENTS

- Python for cleaning the data which is given in the form of csv files(use of numpy and pandas library)
- For graphs and piecharts visualisation –matplotlib
- Machine learning model which predicts future data.

LOGISTIC REGRESSION FUNCTION:



X – input features

Y – Prediction Probability

w – weights

b - biases

$$Y = \frac{1}{1 + e^{-Z}}$$

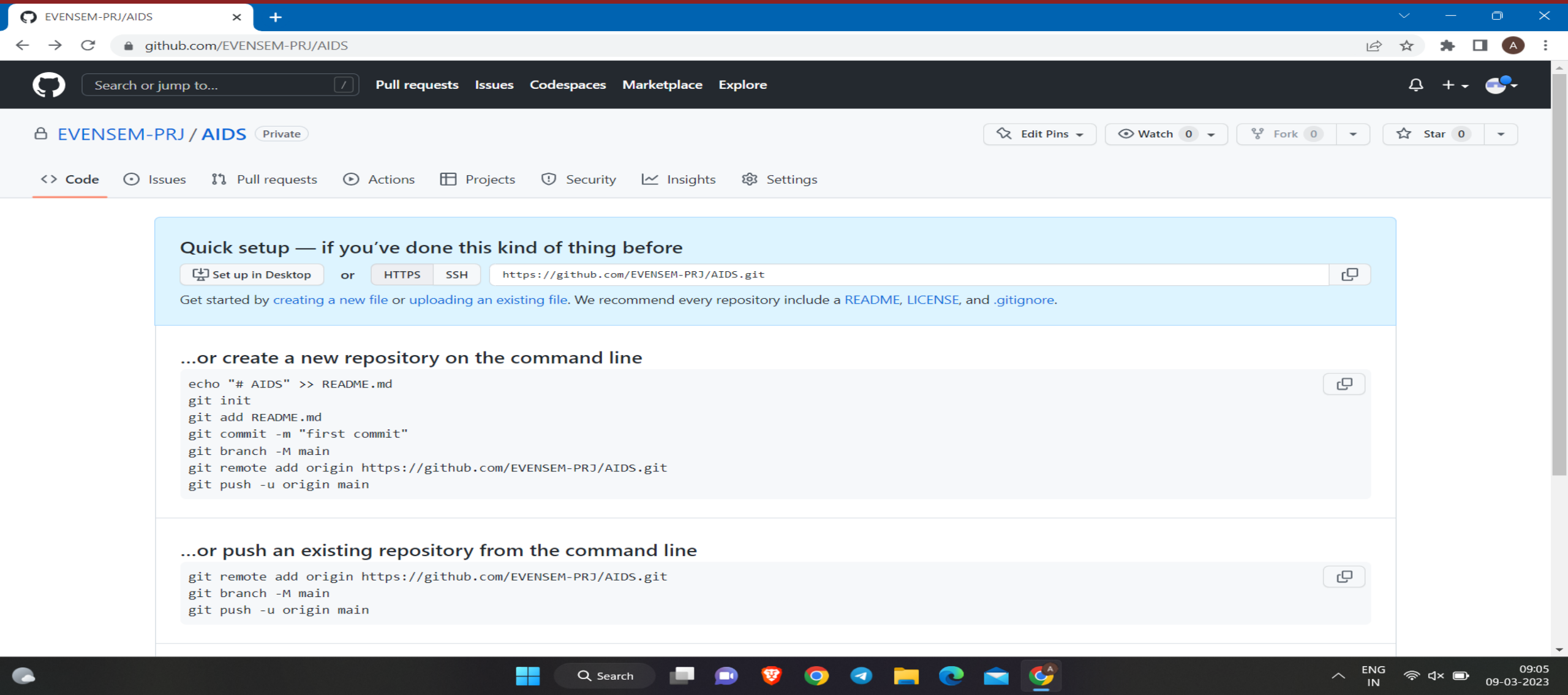
$$Z = w.X + b$$

Sigmoid Function

FUTURE WORK

- After the data is given and analysed visualisation will be done based on the analysis of the data
- A machine learning model will be created using suitable algorithm which predicts the accurate data.
- This can be done by dividing the data into 2 equal parts for effective training of the model. Based on the accuracy the model will be further modified .

GITHUB


EVENSEM-PRJ/AIDS
github.com/EVENSEM-PRJ/AIDS
Search or jump to...
Pull requests
Issues
Codespaces
Marketplace
Explore
EVENSEM-PRJ / AIDS
Private
Edit Pins
Watch 0
Fork 0
Star 0
Code
Issues
Pull requests
Actions
Projects
Security
Insights
Settings
Quick setup — if you've done this kind of thing before
Set up in Desktop
or
HTTPS
SSH
https://github.com/EVENSEM-PRJ/AIDS.git
Get started by creating a new file or uploading an existing file. We recommend every repository include a README, LICENSE, and .gitignore.
...or create a new repository on the command line
echo "# AIDS" >> README.md
git init
git add README.md
git commit -m "first commit"
git branch -M main
git remote add origin https://github.com/EVENSEM-PRJ/AIDS.git
git push -u origin main
...or push an existing repository from the command line
git remote add origin https://github.com/EVENSEM-PRJ/AIDS.git
git branch -M main
git push -u origin main
Search
ENG
IN
09:05
09-03-2023

CODE:

```
import numpy as np
import pandas as pd
import re
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

```
import nltk
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
True
```

```
# printing the stopwords in English
print(stopwords.words('english'))
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she'
```

Data Pre-processing

```
# loading the dataset to a pandas DataFrame
news_dataset = pd.read_csv('/content/train.csv')
```

```
news_dataset.shape
```

```
(20800, 5)
```

```
# print the first 5 rows of the dataframe
news_dataset.head()
```