FAKE NEWS DETECTION USING NLP

by BHASKAR REDDY

Submission date: 07-Nov-2022 09:30AM (UTC-0700)
Submission ID: 1947253331
File name: B198-Review_Paper_FINAL_111.docx (273.96K)
Word count: 1716
Character count: 9554

FAKE NEWS DETECTION USING NLP

BHASKAR REDDY
C TECH
SRM UNIVERSITY
CHENNAI, INDIA

sp4781@srmist.edu.in

GNANESH REDDY
C TECH
SRM UNIVERSITY
CHENNAI, INDIA

mr1409@srmist.edu.in

Dr.R.S.PONMAGAL

ASSOCIATE PROFESSOR,C TECH
SRMIST

CHENNAI,INDIA
ponmagas@srmist.edu.in

Abstract — Using Naive Bayes classification, this work enables us to identify bogus news with more accuracy. The data in this instance broken down into a train data and a test data with latter being further subdivided into groups of data that share similarities. The accuracy the outcome of the test then were verified by matching it to these categories and using the Naive Bayes classifier. It is useful to know whether a new article is accurate or not. It offers the highest level of accuracy and aids in identifying bogus news.

Keywords — Probability, classification of fake news, and machine learning..

1-INTRODUCTION

Accessing news information is made simple by the web and social media.Users find it handy to follow events that interest them online.

The mass media has a significant impact on society, and as is customary, certain people attempt to profit from this influence.
To achieve their objectives, mass media occasionally modifies the content in their own unique ways. Numerous websites disseminate misleading information. They intentionally attempt to disseminate propaganda, hoaxes, and false information while passing it off as real news.
This is due to the fact that artificial intelligence is becoming increasingly popular and that there are numerous technologies available to examine it in part.
The Internet and social media make it simple to access news information. It is practical for the user to
Their main goal is to spread false information so that the public will believe it. Such websites can be found everywhere in the world. Therefore, bogus news has an impact on people's minds. According to a study, several artificial intelligence systems may be able to assist in exposing bogus news.

The data set that will be used to categorise the news is loaded. Following this, the data is broken into training and testing sets, and method is done to evaluate correctness.

I. LITERATURE SURVEY

S.NO	TITLE	Year	Description	PROS	CONS
1	Efficient Fake News by Ahmad, T.,Faisal ,M.S., Rizwan, A., Alkanhel,R., Khan, P.W., and Muthanna Detection Mechanism Using Enhanced Deep Learning Model.	2022	SVM, Random Forest, Logistic Regression, Gaussian Nave Bayes, and Recurrent Neural Network's are some of the Machine Learning techniques used for rumour detection.	For rumour detection, this study employs a fresh collection of content- and social-based traits.	The dataset contains relatively little data. Therefore, as the dataset does not have a lot of tagged data, the accuracy can be lower.
2	Meesad, P. Thai Fake News Detection Based on Information Retrieva	2021	The author of this research suggests using natural language processing to identify .	The aim of false information identification is to give the user the option of categorising the news	We are conducting research to determine ways to improve deep learning's

#	Title	Year			
	l, Natural .			as true or fake.	compreh ension of new concepts .do.
3	A hybrid CNN-RNN-based deep learning techniqu e to fake news detectio n was develope d by Jamal Abdul Nasir, Osama Subhani Khan, and Iraklis Varlami s.	20 21	The TI-CNN Model has been recommende d for deployment.	Evaluating the fake news informatio n in online sources is challengin g. The innocent people can be protected by utilising an algorithm for spotting fake news.	If news includes both sound and video, the machine must appropri ately assess and react.
4	"Fake Media Detectio n Based on Nlp," by Z. Shahbaz and y-C. Byun.	20 21	Developed a robust decision-making architecture using the reinforceme nt learning combined with the block chain frameworks, smart contracts, as well as a special resolution method.	Sharing fake news is one of the more well-known research issues in modern technolog y, and it stems from a linadequat e security ,confidenc e with the veracity in the material that is share on social media.	Existing software tools for microblo gging platform s were primaril y created to address the issue of fake news.
5	Fake News Detectio n Using Machine Learning Approac hes, B N Alwasel 1, H Sirafi1 a nd M Rashid, Z Khanam.	20 21	In order to develop a model of a product with supervised machine learning algorithm, this study analyses the research on fake news detection and investigates standard machine learning models.	In order to develop the best model with supervised machine learning algorithm, the author evaluated the best traditional machine learning models.	Existing software tools for microblo gging platform s were primaril y created to address the issue of fake

3- PROPOSED WORK

The suggested method use the naive Bayes, which has a greater level of accuracy, to identify bogus news. The information is split into two sections here (test and train). The training data is divided into groups using datasets that are similar.After the training data has been processed.The group which has the most in accordance with the group is given the test data.Accuracy of bogus news is now detected using the naive Bayes method, which can be used to stop spam. Each individual word in this passage is assigned a weight, with the most essential word receiving the greatest weight and the least significant word receiving the least. The overall number of words count of distinct words and values given to each word are all calculated using the tfidf vector. By excluding the irrelevant terms from consideration and matching and detecting the accuracy of just relevant words from the dataset, the accuracy of the news may be determined more quickly.

4-THE USES OF THE NAIVE CLASSIFICATION ALGORITHM

Simple machine learning in artificial intelligence includes naive Bayes classifiers. A well-known technique for determining whether news is accurate and true or fake takes use of execution and naïve bayes NB ideas. This is not the only method as there are several that focus on popular questions. used to train those classifiers. You can use naive Bayes to determine whether the news is authentic or phoney. This particular algorithm is used to categorise texts. In a naive Bayes classifier, The Bayes theorem is applied to assess the quality of the news in respect with the use of symbols and news that may or may not be fraudulent.

II. IMPLEMENTATION

1.IMPORT LIBRARIES/DATASETS

- we need to import all the necessary libraries
- datasets are to be added accordingly

  pandas as pd

  import numpy as np

  declare seaborn as sns

  by googletrans importing Translator

  from sklearn.feature_extraction.text import CountVectorizer

  from sklearn.model_selection import train_test_split

  from sklearn.naive_bayes import MultinomialNB

  from sklearn.metrics import confusion_matrix,accuracy_score

  import nltk

```
[1] import numpy as np
    import pandas as pd
    import seaborn as sns

[2] dataset = pd.read_csv('IFND.csv',encoding='ISO 8859-1'
```

- This are all the necessary libraries and modules for implementing the project.

- The entire project will be consists of several libraries which will be installed before importing them.

Code :

```
#checking empty values

dataset.isna().sum()

dataset.dropna(axis = 0, inplace = True)

dataset.shape

dataset.isna().sum()

dataset.reset_index(inplace = True)
```

```
[5] #checking empty values

    dataset.isna().sum()

    id            0
    Statement     0
    Image         0
    Web           0
    Category      0
    Date      11321
    Label         0
    dtype: int64
```

```
[6] dataset.dropna(axis = 0, inplace = True)
```
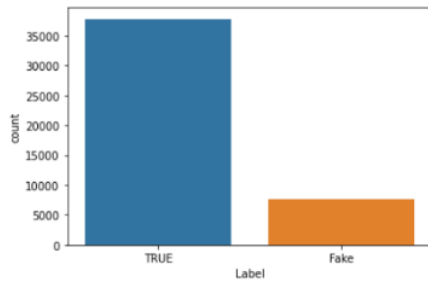
```
[7] dataset.shape

    (45393, 7)
```

```
dataset.isna().sum()

    id           0
    Statement    0
    Image        0
    Web          0
    Category     0
    Date         0
    Label        0
    dtype: int64
```

```
#balanced data or imbalanced

sns.countplot(dataset['Label'])
```

```python
data = dataset['Statement'][0]
data

#regular expression

import re

data = re.sub('[^a-zA-Z]', ' ' , data)

#changing to lower case

data = data.lower()
data

#spliting text

list = data.split()
list

!pip install nltk
import nltk

nltk.download('stopwords')
```



```python
[20] #remove the stopwords from the text if any

    from nltk.corpus import stopwords
    from nltk.stem.porter import PorterStemmer

[21] ps = PorterStemmer()

    review = [ps.stem(word) for word in list if word not in set(stopwords.words('english'))]

review

['prais',
 'india',
 'aarogya',
 'setu',
 'app',
 'say',
 'help',
 'identifi',
 'covid',
 'cluster']

[23] ' '.join(review)

'prais india aarogya setu app say help identifi covid cluster'

[24] corpus = []

    for i in range(0, len(dataset)):
        review = re.sub('[^a-zA-Z]' , ' ', dataset['Statement'][i])
        review = review.lower()
        list = review.split()
        review = [ps.stem(word) for word in list if word not in set(stopwords.words('english'))]
        corpus.append(' '.join(review))

[25] corpus[0]

'prais india aarogya setu app say help identifi covid cluster'
```

```python
#bag of words model

from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer()
x = cv.fit_transform(corpus).toarray()
print(x.shape)
print(x)


y = dataset['Label']
print(y.shape)
x.shape
```



```python
[27] #bag of words model

    from sklearn.feature_extraction.text import CountVectorizer
    cv = CountVectorizer()
    x = cv.fit_transform(corpus).toarray()
    print(x.shape)
    print(x)

    (36631, 15755)
    [[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 0 0 ... 0 0 0]]

[28] x[0]

    array([0, 0, 0, ..., 0, 0, 0])

[29] y = dataset['Label']
    print(y.shape)
    x.shape

    (36631,)
    (36631, 15755)
```
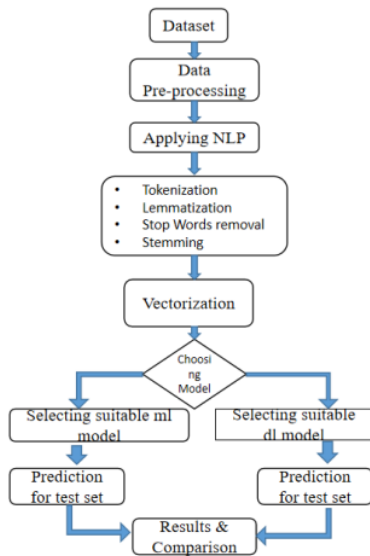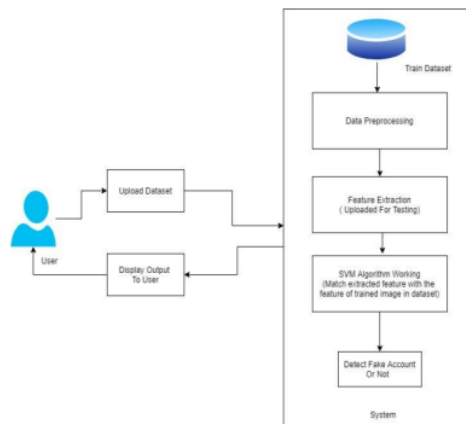
III. SYSTEM ARCHITECTURE

DATA FLOW DIAGRAM

Architecture Diagram

IV. RESULTS DISCUSSION

A. Data Pre-processing

This has all the information that has to be thoroughly examined and preprocessed. Prior to preprocessing, such as tokenizing and stemming, we first the train, test, and verification file systems must be investigated.. Here, the data is carefully examined to see if any values are missing.

B. Feature Extraction

We used scikit and python feature extraction and selection techniques on this dataset. A technique known as tf -idf is used to pick features. The features were also extracted using word to vector, and pipelining was utilised to make the code easier.

C. Classification

The data is divided into two categories in this instance Test and training data The train dataset can then be separated into groups of similar-entity categories. Following the matching of the test data, each group is assigned to a category, and which it belongs. Next, the Naive Bayes classifier is used, and each word's probability is determined separately. The Laplace smoothing is used in this case If indeed the phrase with chance will be evaluated is lacking again from train data dataset. Finally,data authenticity was determined.

D. Prediction

Method we ultimately chose and found to be the top performer was saved on disc under the name file model.After If we shut this repo, the guess file will transfer this model to the user's machine and use it to correctly identify fake news.A user-provided news story serves as the input, and a final categorization result, along with the likelihood that it is accurate, is displayed to the user.

V. CONCLUSION

This means that every article out of a big or small dataset may be categorized as false or true by rapidly compared it to values from an earlier dataset, allowing consumers to believe in a specific item of news.This conclusion can be drawn using the naive Bayes theorem.

AUTHORS AND AFFILIATIONS

1.Dr.R.S.PONMAGAL,ASSISTANT PROFESSOR, C TECH,SRMIST

2.BHASKAR REDDY, STUDENT, C TECH,SRMIST

3.GNANESH REDDY, STUDENT, C TECH,SRMIST.

REFERENCES

[1]. Meesad, P. Thai Fake News Detection Based on Information Retrieval, Natural Language Processing and Machine Learning. SN COMPUT. SCI. 2, 425 (2021).

[2]. Uma Sharma, Sidarth Saran, Shankar M. Patil, 2021, Fake News Detection using Machine Learning Algorithms, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NTASU – 2020.

[3]. Sakeena M Sirajudeen, Nur Fataihah A Azmi, Adamul Abubakar, Online fake news detection algorithm, journal of theorical and Applied information Technology, 2017.

[4] 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.

[5]. Jamal Abdul Nasir, Osama Subhani Khan, Iraklis Varlamis, Fake news detection: A hybrid CNN-RNN based deep learning approach, International Journal of Information Management Data Insights, Volume 1, Issue 1, 2021.

[6]. Z. Shahbazi and Y. -C. Byun, "Fake Media Detection Based on Natural Language Processing and Blockchain Approaches," in IEEE Access, vol. 9, pp. 128442-128453, 2021.

[7]. K. Shu, S. Wang and H. Liu, "Understanding User Profiles on Social Media for Fake News Detection," 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2018, pp. 430-435.

[8]. Ahmad, T.; Faisal, M.S.; Rizwan, A.; Alkanhel, R.; Khan, P.W.; Muthanna, A. Efficient Fake News Detection Mechanism Using Enhanced Deep Learning Model. Appl. Sci. 2022.

FAKE NEWS DETECTION USING NLP

8 Umair Hassan. "Sentiment analysis using machine learning classification models", Institute of Electrical and Electronics Engineers (IEEE), 2022
Publication

<1 %

9 Akshay Kulkarni, Adarsha Shivananda. "Natural Language Processing Recipes", Springer Science and Business Media LLC, 2021
Publication

<1 %