

USE OF PYTHON AND MACHINE LEARNING FOR FAKE NEWS DETECTION

A. Gayathri^{*1}, B. Uma Devi^{*2}, Mr. J. Arun Kumar^{*3}

^{*1} BTech, Department of ECE, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India

^{*2} BTech, Department of ECE, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India

^{*3} Assistant Professor, Department of ECE, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India

ABSTRACT

Everyone now relies on various online news sources because the internet is so widely used. News quickly spread among millions of users in a very short period of time along with the increase in the use of social media platforms like Facebook, Twitter, etc. Additionally, spammers use alluring news headlines to lure readers into clicking on their ads. You occasionally need to confirm the accuracy of information.

A new area of research that is receiving a lot of attention is fake news detection. Due to the limited resources, including datasets, processing, and analysis methods, it does, however, face some difficulties. In this work, we suggest a machine learning-based system for detecting fake news.

Keywords: sklearn, Vectorizer, PassiveAggressive Classifier

I. INTRODUCTION

As more and more of our time is spent talking online through social networking sites, more and more people prefer to look for and consume news on social media sites instead of traditional news outlets. These social media platforms' fundamental characteristics are implied in the causes of this change in consumer behavior. People and society may be significantly harmed by the mass dissemination of erroneous information. Second, bogus news will upset the credibility balance in the news ecosystem. For instance, it is clear that during the 2016 U.S. presidential election, more popular false news than popular mainstream news was disseminated on Facebook.

A new area of research that is receiving a lot of attention is fake news detection. Due to the limited resources, including datasets, processing, and analysis methods, it does, however, face some difficulties. In this work, we suggest a machine learning-based system for detecting fake news.

II. LITERATURE REVIEW MATERIALS AND METHODOLOGY

RELATED WORK

Framework to Identify and Secure the Issues of Fake News and Rumors in Social Networking by Anav Bedi, Nitin Pandey, and Sunil Kumar Khatri According to our analysis, misleading news also continues to spread as publishers do in this cycle as consumers and publishers both increase. Because of the surge in smartphone usage and the popularity of these platforms, people now prefer social media sites like Facebook and Twitter to more conventional forms of media for their news.

By Mohammed Hazim Alkawaz and Sayeed Ahsan Khan, "Use of Fake News and Social Media by Main Stream News Channels of India." Social media fake news is a concern. One of the top 10 threats to society is the dissemination of false information and fake news online.

III. MODELING AND ANALYSIS

1. EXISTING MODEL

Linguistic cue methods

Network analysis methods

Selected methods explored further:

- Naïve Bayes Classifier
- Support Vector Machines(SVM)
- Semantic Analysis

2. PROPOSED METHOD

This study falls under the category of employing image processing to identify fake news. The proposed solution to the fake news issue is the implementation of a tool that can identify and remove fraudulent websites from search engine or social media news feed results.

How to use Python to spot false news

- Make necessary imports.
- Read the data into a data frame.
- Get the data from the data set.
- Split the dataset into training and testing sets.
- Initialize a Vectorizer with stop words from the English language and a maximum document frequency of 0.7

IV. PROCESS AND CONCLUSION

Step 1:

In order to complete this project, include a number of modules, including numpy, pandas, and mostly the sklearn module, commonly known as scikit-learn. In the Python environment, Scikit-learn, an open source data analysis toolkit, is considered to be the pinnacle of machine learning (ML).

```
✓ [6] import numpy as np
import pandas as pd
import itertools
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
```

Step2:

Adding the data collection and determining its size and form

```
#Read the data
df=pd.read_csv('/news.zip')
#Get shape and head
df.shape
df.head()
```

Unnamed: 0		title	text	label
0	8476	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fello...	FAKE
1	10294	Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg LinkedIn Reddit Stumbleu...	FAKE
2	3608	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	REAL
3	10142	Bernie supporters on Twitter erupt in anger ag...	— Kaydee King (@KaydeeKing) November 9, 2016 T...	FAKE
4	875	The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners...	REAL

Step 3:

Getting the data labels whether they are Fake or Real. Data labelling in machine learning is the process of classifying unlabeled data (such as photos, text files, videos, etc.) and adding one or more insightful labels to give the data context so that a machine learning model may learn from it.

```
[ ] #Data - Get the labels
labels=df.label
labels.head()

0    FAKE
1    FAKE
2    REAL
3    FAKE
4    REAL
Name: label, dtype: object
```

Step 4:

Next Processes is dividing the Dataset.

```
[ ] #Data - Split the dataset
    x_train,x_test,y_train,y_test=train_test_split(df['text'], labels, test_size=0.2, random_state=7)
```

Step 5:

Next step is to initialize the Vectorizer as TfidfVectorizer

```
[ ] #Data - Initialize a TfidfVectorizer
    tfidf_vectorizer=TfidfVectorizer(stop_words='english', max_df=0.7)
```

Step 6:

Comparing between the trained data and test data and we obtain a accuracy of 93.3%.

```
#Data - Predict on the test set and calculate accuracy
y_pred=pac.predict(tfidf_test)
score=accuracy_score(y_test,y_pred)
print(f'Accuracy: {round(score*100,2)}%')
```

Accuracy: 93.13%

Step 7:

In the final step we create a matrix defining the percentage of the accuracy of the dataset given to the model

```
] #Data - Build confusion matrix
confusion_matrix(y_test,y_pred, labels=['FAKE','REAL'])

array([[593, 45],
       [ 42, 587]])
```

CONCLUSION

Use of Python for False News Detection We trained a model using a political dataset, a Vectorizer, and a PassiveAggressiveClassifier. We ultimately achieved a magnitude accuracy of 93.13%.

Social media is becoming more and more popular, and more people are using it to obtain their news than they are using traditional news sources. However, fake news has been disseminated via social media and has had a detrimental effect on both individual users and society as a whole. In this study, we looked at the issue of fake news by doing a characterisation and detection analysis of the literature. During the characterisation phase, we presented the core concepts and tenets of false news in conventional and social media.

V. REFERENCES

- [1] <https://www.hindawi.com/journals/complexity/2020/8885861/#data-availability>
- [2] <https://medium.com/swlh/fake-news-detection-using-machine-learning-69ff9050351f>
- [3] <https://data-flair.training/blogs/advanced-python-project-detecting-fake-news/>
- [4] <https://www.youtube.com/watch?v=tdFMIO5lfgA>
- [5] <http://ieeexplore.ieee.org/abstract/document/7100738/>
- [6] Wikipedia article about Naïve Bayes. Available: https://en.wikipedia.org/wiki/Naive_Bayes_classifier
- [7] A proposed way of implementation. Available: <https://www.datacamp.com/community/tutorials/scikit-learn-fakenews>
- [8] Study about Bayes theorem. Availabe:
<http://dataaspirant.com/2017/02/06/naive-bayes-classifiermachine-learning/>