iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35271

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Detecting Fake Covid 19 News

Miss Himanshi Rathore[1], Miss Sakshi Sharma[2], Miss Pankhuri Sharma[3], Prof. Ruchi Vyas[4], Dr. Mayank Patel[5]
[1, 2, 3]UG Student , [4]Assistant Professor, [5]Professor & Head, Department of Computer Science & Engineering, Institute of Technical Studies Udaipur, Rajasthan 313001, India

Abstract: ***The fake news Detection program exists to help its users distinguish between useful information and baseless rumours. It helps one to verify it themselves. In the current coronavirus disease (COVID-19) pandemic, misinformation is particularly prevalent, leading to people believing false and potentially harmful statements and posts. The spread of panic and misunderstanding among the public can be reduced if fake news is detected quickly. This covid 19 fake news detection model is specifically built to identify fake news.***

I INTRODUCTION

The Fake News Detection Program will help users to easily distinguish between fake and real news. Automatically detecting fake news poses challenges that defy existing content-based analysis approaches. One of the main reasons is that often the interpretation of the news requires the knowledge of political or social context or 'common sense, which current natural language processing algorithms are still missing. This project aims to detect COVID 19 Fake News across social media platforms to make people aware of the facts about Covid-19, be it home remedies, vaccines, symptoms, or any other form of treatment.

We aim to give users a very simple and easy source to detect whether a given news article is real or not, leading to conscious efforts to discontinue spreading it, thus making the users realize what to believe in.

II EXISTING SYSTEM

In the existing system, the user cannot trace out if the information displayed before him is counterfeit or not. There is no way of checking its authenticity until and unless one searches the inter-webs for it. It may not seem like a cumbersome task at first. But we should keep in mind we are bombarded with thousands of pieces of information every day and doing the above for every piece of information is exhausting. Also, the news may be true in some other perception but we aim to know if it is reliable in the current scenario or not.

III PROBLEM STATEMENT

COVID-19 brought the entire world to a halt. People who were paranoid and scared began sharing home remedies for avoiding this disease. Numerous home remedies for the same can be found on social media platforms such as Twitter and WhatsApp. The world came to a halt as a result of COVID-19. People who were paranoid and terrified started sharing home remedies for avoiding this disease. On social media, you can find a plethora of home remedies for the same.

But the question is, which of these remedies works? Aside from that, the internet was flooded with countless COVID-19 symptoms, raising the question of what the actual symptoms are. Everything about it is ambiguous.

IV TECHNOLOGY STACK

A. Pandas

A fast and efficient DataFrame object with integrated indexing for data manipulation. Reading and writing data between in-memory data structures and various formats, including CSV and text files, Microsoft Excel, SQL databases, and the fast HDF5 format. Data sets can be reshaped and pivoted in a variety of ways. Intelligent label-based slicing, fancy indexing, and large data set subsetting, among other things.

B. Numpy

NumPy is a Python library that provides an n-dimensional array, which is a simple yet powerful data structure. This is the foundation upon which almost all of Python's data science toolkit's power is built.

NumPy extends the computational power of languages such as C and Fortran to Python, a language that is much easier to learn and use. With this power comes simplicity: a NumPy solution is often clear and elegant.

C. *Sklearn*

Scikit-learn is probably the most useful library for machine learning in Python. The sklearn library contains a lot of efficient tools for machine learning and statistical modelling including classification, regression, clustering and dimensionality reduction.

D. *Re*

A RegEx, or Regular Expression, is a sequence of characters that forms a search pattern.
RegEx can be used to check if a string contains the specified search pattern. Python has a built-in package called re, which can be used to work with Regular Expressions.

E. *NLTK*

The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing (NLP) for English written in the Python programming language. NLTK includes graphical demonstrations and sample data. It is accompanied by a book that explains the underlying concepts behind the language processing tasks supported by the toolkit.

V. HIGHLIGHTS OF THE MODEL

A. *Importing all the Required Libraries*

```
In [1]: import nltk
        import pandas as pd
        import numpy as np
        import re
        import string
        from nltk.corpus import stopwords
        from nltk.tokenize import word_tokenize
        from sklearn.metrics import classification_report
        nltk.download('stopwords')
        stop_words = set(stopwords.words('english'))
        nltk.download('punkt')

        [nltk_data] Downloading package stopwords to
        [nltk_data]     C:\Users\HP\AppData\Roaming\nltk_data...
        [nltk_data]   Package stopwords is already up-to-date!
        [nltk_data] Downloading package punkt to
        [nltk_data]     C:\Users\HP\AppData\Roaming\nltk_data...
        [nltk_data]   Package punkt is already up-to-date!

Out[1]: True
```

B. *Loading The Fake News Dataset And Dropping Unnecessary Columns From The Dataset*

```
In [3]: df_text = pd.read_csv('news.csv', encoding='latin-1')
        # Uwe have textual data, but computers work on numbers. So first is required to convert them to numbers using encoding='latin-1'
        #Latin-1 is a superset of ASCII
        df_text.columns = ['id', 'title', 'text', 'label']
        df_text.drop(['id', 'title'], axis=1)
```

Out[3]:

	text	label
0	Daniel Greenfield, a Shillman Journalism Fello...	FAKE
1	Google Pinterest Digg Linkedin Reddit Stumbleu...	FAKE
2	U.S. Secretary of State John F. Kerry said Mon...	REAL
3	â□□ Kaydee King (@KaydeeKing) November 9, 2016...	FAKE
4	It's primary day in New York and front-runners...	REAL
...	...	...
6330	The State Department told the Republican Natio...	REAL
6331	The â□□Pâ□□ in PBS Should Stand for â□□Plutocr...	FAKE
6332	Anti-Trump Protesters Are Tools of the Oligar...	FAKE
6333	ADDIS ABABA, Ethiopia â□□President Obama conve...	REAL
6334	Jeb Bush Is Suddenly Attacking Trump. Here's W...	REAL

6335 rows × 2 columns

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

C. *The data must then be cleaned. It's important to remember that we're teaching a computer to tell the difference between what's real and what's not. Right now, we have text data, but computers work with numbers. So, first, we must convert them to numbers, and then we must ensure that we are only converting texts that are required for comprehension.*

The remaining symbols, which are punctuations, must be removed next. Likely, the news isn't true if there are a lot of punctuations, such as overuse of exclamations. The words must then be tokenized and stemmed to their root. The process of converting each sentence into a list of words or tokens is known as tokenization. Tokens are converted into meaningful numbers. This is also referred to as feature extraction.

```
In [5]: def wordopt(text):
            text = re.sub(r"http\S+|www\S+|https\S+"," ", text, flags=re.MULTILINE)
            text = re.sub(r'\@\w+|\#'," ", text)
            text = text.translate(str.maketrans(" " ," ", string.punctuation))
            return text

In [6]: df_text["text"] = df_text["text"].apply(wordopt)

In [7]: tokens = word_tokenize(str(df_text["text"]))
        # used str so that the following error does not occur
        # TypeError: expected string or bytes-like object
        words = [w for w in tokens if not w in stop_words]
        # Tokenization means to make every sentence into a list of words or tokens.
```

D. *The machine learning pipeline is the next step.*

We need to get the data into X and Y, just like in a typical ML pipeline. To do so, we use X as the TF-IDF vectoriser's output matrix, which needs to be flattened. This encoder converts label texts to numbered targets. Assume we have a list of labels that looks like this: ['real,' 'fake,' 'fake,' 'real’].

After the training and testing splits are completed. To accomplish this we imported the train test split function and used sklearn's preprocessing package.

```
In [8]: from sklearn.feature_extraction.text import TfidfVectorizer
        tf_vector = TfidfVectorizer(sublinear_tf=True)
        tf_vector.fit(df_text["text"])

Out[8]: TfidfVectorizer(sublinear_tf=True)

In [9]: X_text = tf_vector.transform(df_text["text"].ravel())
        y_values = np.array(df_text["label"].ravel())

In [10]: from sklearn import preprocessing
         le = preprocessing.LabelEncoder()
         le.fit(y_values)
         le.transform(y_values)

Out[10]: array([0, 0, 1, ..., 0, 1, 1])

In [11]: from sklearn.model_selection import train_test_split
         X_train, X_test, y_train, y_test = train_test_split(X_text, y_values, test_size=0.15, random_state=120)
```

E. *Using Logistic Regression for Fake news Detection*

```
In [12]: # Using Logistic Regression

         from sklearn.metrics import accuracy_score
         from sklearn.linear_model import LogisticRegression

         model = LogisticRegression(solver="lbfgs")
         model.fit(X_train, y_train)
         y_predict = model.predict(X_test)
         print(accuracy_score(y_test, y_predict))
         print(classification_report(y_test, y_predict))

0.9274447949526814
              precision    recall  f1-score   support

        FAKE       0.90      0.95      0.93       464
        REAL       0.95      0.90      0.93       487

    accuracy                           0.93       951
   macro avg       0.93      0.93      0.93       951
weighted avg       0.93      0.93      0.93       951
```

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

F. Function to Detect Fake News

```
In [13]: def output_label(n):
             if n == "FAKE":
                 return "Fake News"
             elif n == "REAL":
                 return "Real News"

         def manual_testing(news):
             testing_news = {"text":[news]}
             new_def_test = pd.DataFrame(testing_news)
             new_def_test["text"] = new_def_test["text"].apply(wordopt)
             new_x_test = new_def_test["text"]

             new_xv_test = tf_vector.transform(new_x_test)
             pred_LR = model.predict(new_xv_test)


             return print("\n\nLR Prediction: {}".format(output_label(pred_LR[0])))
```

G. Function to Take Input From the User

```
In [*]: news = str(input())
        manual_testing(news)
```

H. Testing the Model

```
In [17]: news = str(input())
         manual_testing(news)
```

Project Veritas just released yet another video of the CNN producer calling voters stupid as sh*t and even mocking on of thei r top news anchors Chris Cuomo!The undercover reporter stroked Carr s ego saying he had just as much gravitas as Cuomo and h e responded saying, Just can t stand the guy when he talks, we re like shut up. The Veritas reporter asked for clarification a s to who he was talking about and he reaffirmed that he was referring to Cuomo.Carr backed up his past comments on other videos such as calling the American electorate stupid as sh-t, and admitting to pushing stories on Russia strictly for ratings. I me an granted, anything I ve said to you I would defend on the record. I don t think I ve said anything, wrong, said Carr.OUR PRE VIOUS VIDEOS AND REPORTS FROM PROJECT VERITAS ON CNN:Project Veritas newest video from the American Pravda: CNN series exposes Jimmy Carr, the Associate Producer for CNN s New Day attacking President Donald Trump and admitting that CNN has a left-leaning bias. When asked by an undercover journalist if CNN is impartial, Carr plainly responded, In theory. We reported on the previo us videos from Project Veritas: James O Keefe s Project Veritas reporters went undercover at CNN to investigate the #VeryFakeNe ws network to determine the motivation behind CNN s Trump-Russia collusion obsession. Since the inauguration, CNN has mentioned Russian story nearly 16,000 times.Project Veritas reporters can be seen in the video below having a conversation with CNN s s upervising producer John Bonifield. The reporter talked to Bonifield first about the constant barrage of Trump-Russia stories o n CNN. Bonifield admitted that although CNN has no evidence or proof of Trump involvement with Russia, they continue to make it their top story on CNN simply for ratings and to make money. Bonifield has worked as a journalist and as a producer for nea rly 15 years at CNN, making him a pretty reliable source when it comes to how things work at CNN.Bonifield actually confessed t o the undercover Project Veritas journalist that President Trump is correct when he says the media is on a witch hunt to take him down. Bonifield told the reporter: I think the President is right to say, like, look you are witch hunting me You have n o real smoking gun, you have no proof. Bonifield told the undercover reporter for Project Veritas that CNN is biased and is pla ying to their audience by attacking President Trump, and actually admitted that Trump is good for business right now. The Pro ject Veritas reporter asked Bonifield, But honestly, you think the whole Russia shit is bullshit? to which Bonifield replied: Could be bullshit. I mean, it s mostly bullshit right now. Like, we don t have any big, giant proof. The way these leaks happe n, they d leak it. It d leak. If it was something really good, it would leak. Watch the unbelievable video here:Did the three r esignations that took place at CNN today have anything to do with the impending release of O Keefe s explosive video proving th at CNN manufactures Trump-Russia collusion stories for ratings, or for money?The story about the three CNN employees resigning for publishing fake news broke yesterday. Is ti a coincidence that today, O Keefe released this explosive bombshell video?Here s the story that broke yesterday: CNN employees have resigned for their role in pushing fake news about the Trump-Russia story that claimed a member of the Trump transition team was under investigation.Reporter Thomas Frank, editor Eric Lichtblau and exe cutive editor Lex Haris all resigned from their positions following a company-wide investigation into the single-sourced story that was quickly debunked, The Washington Post first reported. CNN quickly followed up with a story of their own on the resigna tions. Lichtblau reportedly oversaw the false story, while Haris headed up CNN s investigative unit. In the aftermath of the re traction of a story published on CNN.com, CNN has accepted the resignations of the employees involved in the story s publicatio n, CNN s coverage quotes a company spokesman as saying.Philly.com On Thursday evening, CNN investigative reporter Thomas Fra nk published a potentially explosive report involving an investigation of a Russian investment fund with potential ties to seve ral associates of President Donald Trump.But by Friday night, the story was removed from CNN s website and all links were scrub bed from the network s social media accounts. That story did not meet CNN s editorial standards and has been retracted, CNN sa id in an editors note posted in place of the story. Links to the story have been disabled.

LR Prediction: Fake News

The main goal of this project is to address the problems associated with false news detection while also developing a brand new innovative smart system that can convince the institution. In this project, a smart web app will be developed that can identify the source of each piece of news and then record the information into a database system. With this, efforts are being made to automate the task of detecting fake news. The most common of these actions is the blacklisting of untrustworthy sources and authors. Although these tools are useful, to produce a progressive complete end-to-end solution, we must represent more difficult cases in which reliable sources and authors are responsible for the dissemination of fake news. The goal of this project is to create a model that can be used to recognise language patterns that can be used to classify fake and real news of covid-19 using ML (machine learning) techniques. Fake news is now such a major issue that it is affecting our society as well as our facts and opinions. AI and machine learning techniques can be used to solve the problem at hand

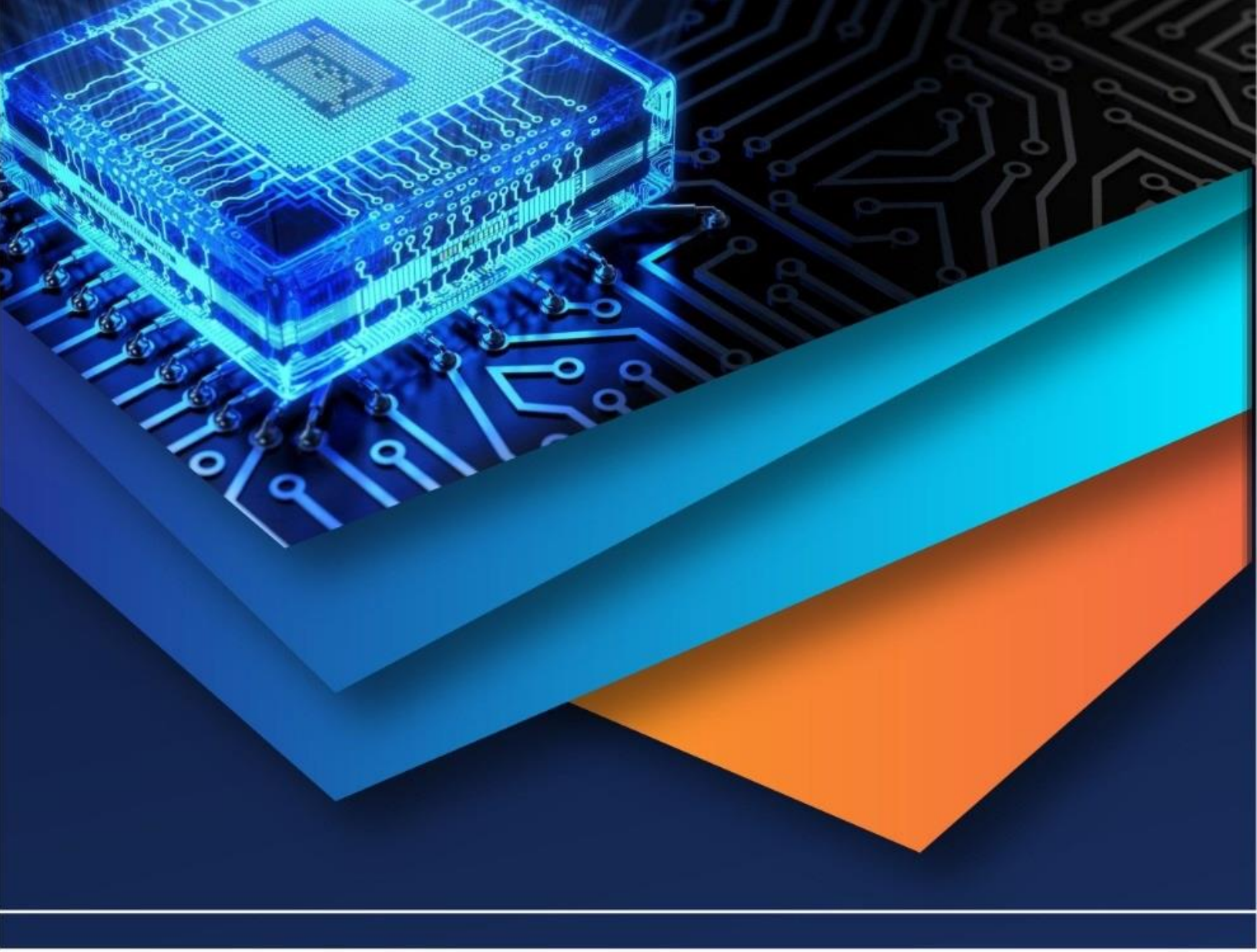
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

VI ACKNOWLEDGMENT

This project demanded a huge amount of research work and dedication which would not have been possible without the support of many individuals and organization. Therefore, we would like to extend our sincere gratitude to all of them. We are also grateful to Prof. (Dr Vikas Mishra), Director, GITS for supporting and motivating us to do quality work. Finally, we are thankful to Dr Mayank Patel (HOD CSE) for helping us in implementing the project.

REFERENCES

[1] https://www.who.int/health-topics/coronavirus#tab=tab_1
[2] https://www.who.int/emergencies/diseases/novel-coronavirus-2019
[3] https://en.wikipedia.org/wiki/Coronavirus_disease_2019
[4] https://en.wikipedia.org/wiki/COVID-19_pandemic
[5] https://www.jhsph.edu/covid-19/projects-and-initiatives/
[6] https://data-flair.training/blogs/advanced-python-project-detecting-fake-news/



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY