FAKE NEWS DETECTION

A Project Report submitted in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

GITAM UNIVERSITY, VISAKHAPATNAM

Submitted by

GURRALA RAM KAUSHAL	MEKALA VISHNU TEJA
(121710311015)	(121710311029)
MANI SAI RAM REDDY	ARREPU SUMANTH
(121710311024)	(121710311005)

Under the esteemed guidance of

Dr. Angara S V Jayasri

Assistant Professor, C.S.E, GIT

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GITAM INSTITUTE OF TECHNOLOGY

GITAM (Deemed to be University), VISAKHAPATNAM

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GITAM INSTITUTE OF TECHNOLOGY

GITAM (Deemed to be University), VISAKHAPATNAM

GITAM
(DEEMED TO BE UNIVERSITY)
(Estd. u/s 3 of the UGC Act, 1956)
VISAKHAPATNAM ✹ HYDERABAD ✹ BENGALURU

DECLARATION

We, hereby declare that the mini-project report entitled **"FAKE NEWS DETECTION"** is an original work done in the Department of Computer Science and Engineering, GITAM Institute of Technology, GITAM (Deemed to be University) submitted in partial fulfillment of the requirements for the award of the degree of B.Tech in Computer Science and Engineering. The work has not been submitted to any other college or university for the award of any degree or diploma.

Date:

Place : Visakhapatnam

Name and Reg. no's:

GURRALA RAM KAUSHAL (121710311015)

MANI SAI RAM REDDY (121710311024)

MEKALA VISHNU TEJA (121710311029)

ARREPU SUMANTH (121710311005)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GITAM INSTITUTE OF TECHNOLOGY

GITAM (Deemed to be University), VISAKHAPATNAM

GITAM
(DEEMED TO BE UNIVERSITY)
(Estd. u/s 3 of the UGC Act, 1956)
VISAKHAPATNAM ✹ HYDERABAD ✹ BENGALURU

CERTIFICATE

This is to certify that the project entitled **"FAKE NEWS DETECTION"** is a bonfide record of work carried out by

GURRALA RAM KAUSHAL
MANI SAI RAM REDDY
MEKALA VISHNU TEJA
ARREPU SUMANTH

submitted in partial fulfillment of the requirement for the award of the degree of Bachelors of Technology in Computer Science and Engineering.

Project Guide	**Project Coordinator**	**HOD**
Dr.Angara S V Jayasri Assistant Professor	1. Mr. Amarnadh A.M.C, Assistant Professor	Dr. Konala Thammi Reddy HOD, Professor
	2. Dr. G. Srinivas Associate Professor	

ACKNOWLEDGMENT

We would like to thank our project guide **Dr.Angara S V Jayasri**, Assistant Professor, Department of C.S.E, GIT, for his stimulating guidance and profuse assistance. We shall always cherish our association with him for his advice, encouragement, and valuable suggestions throughout this work's progress. We consider it a great privilege to work under his guidance and constant support.

We also express our thanks to the project reviewers **Mr. Amarnadh**, A.M.C**,** Assistant Professor, **Dr. G. Srinivas**, Associate Professor, Department of C.S.E, GIT, for their valuable suggestions and guidance and helped us a lot in completing our project and project report.

We consider it a privilege to express our deepest gratitude to **Dr. Konala Thammi Reddy**, Head of the Department, Professor, C.S.E, GIT, for his valuable suggestions and constant motivation that immensely helped us to complete this project.

Our sincere thanks to **Dr. C. Dharma Raj**, Principal, GIT, for inspiring us to learn new technologies and tools.

We also express our thanks to **Mr. Vijay Shekar C**, Dean of Engineering, GIT, for their constant motivation that immensely helped us to complete this project.

Finally, we deem it a great pleasure to thank one and all that helped us directly and indirectly throughout this project.

We perceive this opportunity as a significant milestone in our career development. I will strive to use gained skills and knowledge in the best possible way, and I will continue to work on their improvement to attain desired career objectives. Hope to continue cooperation with all of you in the future.

GURRALA RAM KAUSHAL (121710311015)
MANI SAI RAM REDDY (121710311024)
MEKALA VISHNU TEJA (121710311029)
ARREPU SUMANTH (121710311005)

TABLE OF CONTENTS

ABSTRACT

In recent years information sharing through internet and social media have grown abundantly and it is hard to find the authenticity of the news. This paper aims to detect the fake news using machine learning and web scraping methodologies. The proposed methodology performs the fake news classification on the labeled dataset based on one of the Machine learning algorithms i.e. logistic regression. Theproposed system also scrapes multiple news websites and checks for keywords mentioned in the input news article and calculate a percentage combining the results from the machine learning approach which uses the logistic regression algorithm and the web scraping approach. Also, a major factor to weigh in terms of fake news is the sentiment or the emotion behind the news article, the news article may be biased concerning a specific political propaganda, hence in favor of a side. But true news must be unbiased; it must be factual, not accusing or declarative. Thus, to inform the user about the bias in the news article, the proposed system performs sentimental analysis involving pre-processing, NLTK lemmatization, porter stemmer and uses naïve Bayesian algorithm and then presents the user with the sentiment behind the news article. The two above mentioned process will present the user with two outputs: the percentage of how accurate the news article combining the results from the logistic regression algorithm and web scraping approach .

INTRODUCTION

The increase in access to internet and the boom in social networking and media such as WhatsApp, Facebook, Instagram etc. made the access to the news information much easier, portable and faster. Often the general public with access to internet can now follow news, their articles of interest and much more, right on their fingertips, anywhere and anytime. News, or content, or media in general have a huge influence on the society, it is both capable to sway opinions, change mindsets as we say in the US elections in 2016 and thus, there arises a high chances that someone wishes to exploit this opportunity in their favor. Sometimes to achieve personal gains, mass-media may manipulate the information in different ways. This leads to producing of the news articles that are not completely true or even completely false. Hence, the core objective of fake news is to reshape or mold the public opinion on certain matters, mostly targeted at satisfying a personal agenda which often include political motives.Furthermore, the feeling or sentiment behind the news article is a major factor to consider in terms of fake news; the news article may be skewed with respect to a specific political agenda, hence in favor of one hand. True news must be impartial; it must be factual, not accusatory or declarative. The proposed system detects the fake news using machine learning algorithm, scraps multiple news websites and keyword checks mentioned in the input news article and calculates a percentage that combines both approach results.

LITERATURE SURVEY

In an assortment of disciplines, including semantics and software and computer science engineering, counterfeit news has become a significant research subject. The authors clarify how the problem is approached from the natural language processing perspective, with the objective of proposing a system to detect unauthentic information in news automatically. The fake news classifier is constructed using logistic regression classifier, wherein the datasets were accumulated by Kaggle News for learning and testing of the system. A comprehensive tutorial-based approach is used for establishing the research and datasets were clearly listed, various detection strategies were coalesced under an intensive framework for counterfeit news detection and state-of-the art patterns and models were employed. FakeNewsTracker, a framework for counterfeit news comprehension and discovery can naturally gather information for news pieces and social setting, which advantages further research of comprehension and anticipating counterfeit news with successful representation procedures. Content-based, source-based, and diffusion-based approaches were presented. The work describes two opposite approaches and suggests an algorithmic solution synthesizing the main concerns. Also, raises awareness of the needs and opportunities of companies currently seeking to help automatically detect fake news through the provision of web services. The authors provide a detailed analysis of the findings of the latest false news. This is characterized by the negative effect of online fake news and state-of -the-art detection methods.Many of these are focused on defining client, content, and background features that suggest misinformation. It has existing repositories that are used to classify fake news. Clickbait, draws in user and their interest with garish features or structures to click connects to expand income from promotions. The work breaks down the commonness of phony news given the advancement made conceivable by the rise of long-range informal communication locales in correspondence. The goal this work is to develop a solution that users can use to identify and remove pages that contain false and misleading information. The main objective of the work is to highlight frameworks , which models distributer news relations and client news connections at the same time for counterfeit news .

PROBLEM IDENTIFICATION

The proposed work suggests a novel and amalgamated method combining some known and well researched methods merging the advantages of AI through the simplistic algorithm, Logistic Regression, which is a simple, easy to understand, quick yet efficient algorithm. Since a major drawback, for ML and AI algorithms, when it comes to processing information such as news articles would be the checking the accuracy of facts or news that have surfaced recently. Since the suggested algorithm for fake news detection only works on a predefined dataset, the training module might not work as efficiently for the same. Thus the proposed system boost the accuracy and bolsters the results through the integration of a web scraping module which is capable of scraping through various news websites, internationally recognized as accurate, for latest news articles and saving them into a text file for matching against the given input by the user. Also, another important aspect that underlies the fake news in the modern times is the use of hate speech viz. text that appeals to the emotions of people, to move their opinion against some issue, usually satisfying a political agenda. Hence to propose a wholesome solution the proposed system counters the propagation of hate speech by curbing it from the source, by felicitating the social media user who perceives the fake news or hate speech article, usually aligning with his political or religious bias, as authentic and forwards it to other citizens. Thus, the proposed system also provides a sentiment analysis module which can recognize the underlying sentiment behind the input text and inform the user for any bias in the sentiment behind the text.

Proposed System

The proposed work is being implemented through the integration of 2 modules, namely the logistic regression module and the web scraping module to detect fake news .

System Methodology

Logistic regression

The logistic regression module performs the simple task of taking the dataset, splitting it into two parts, viz. test and train set. The train dataset which is the Kaggle dataset is used to train the regression model for the user input which is the news to be tested in this case. It is a classification algorithm used for machine learning that predicts the likelihood of a categorical dependent variable, where it will be either fake or authentic henceforth logistic regression will help to describe a relationship between a set of independent variables and categorical dependent variables. The dependent variable in logistic regression is a binary variable that includes data encoded as 1 (the user given news is fake) or 0 (the news is authentic), hence these are the only two classes. The model gives a authenticity value between 0 and 1 later on converted into percentage and hence can be easily categorized as how much the news is authentic or fake.In plain terms it forecasts the possibility of incidence of fakeness in the news set by fitting the data to the logit function which has already been trained by the Kaggle dataset which the proposed system used to train the model. We will continue with the basic linear regression equation with dependent variable included in a relation function to proceed with logistic regression.

Web Scraping

However, the regression module alone is not enough to test against facts which form a major percentage of a news. Hence, to produce more accurate results and to check against newly surfaced news articles, the web scraping module is coupled with this model. The web scraping uses an inbuilt module called 'newspaper' which in turn combines two basic inbuilt modules of python viz. 'Requests' and 'lxml'. The requests module is used to send all kinds of HTTP requests. It is a pretty straightforward module which is imported in python and the necessary news website is requested using the Requests. Get(URL)The usage of requests module has been majorly implemented for its simplicity. The other module as mentioned above as used by the newspaper module is the 'lxml' module. After the page has been requested by the Requests module, the lxml file is used to handle the XML and HTML files. The whole HTML page can be seen in the form of an XML tree, having Elements and Sub-Elements. The text of the article is usually wrapped into some Sub-Element, hence following a tree like hierarchy.

Combining the results from the logistic regression module and the scraping

For more accuracy and better results, the results from both the logistic regression module and the scraping module are combined using pandas Concatenation and complete dataset is used for modelling the Fake News Detector

Text-Cleaning

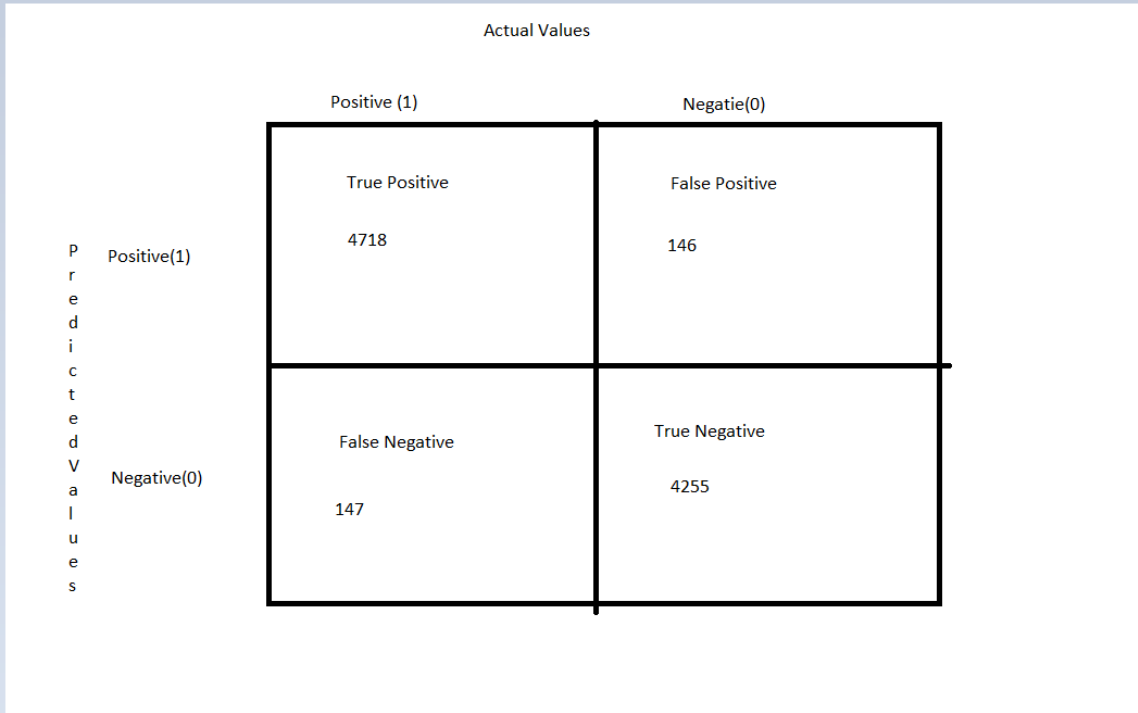
Tools like NLTK (Natural Language Toolkit) are used. It helps to convert raw text into a list of words. We split the text into words, choosing alphanumeric character strings (A-Z, 0-9, a-z and '_'). We remove all punctuations like commas, quotes etc. along with the whitespaces.

TF-IDFVectorizer

Tokenization of the information was completed, and a corpus was made. TF-IDF, term recurrence reverse report recurrence vectorizer from the scikit-learn library is utilized for creating highlights right now. The TF-IDFvectorizer utilizes the corpus created utilizing tokenization. TF-IDFis an estimation plot allotting appraisal or loads dependent on its term recurrence (tf) and the reverse variable recurrence (IDF) for each word in a report. The weight relegated by TF-IDFvectorizer is utilized as a parameter to pass judgment on the pertinence of a word in the document. The words containing higher

weight esteems are regarded progressively significant. The worth expands relatively to how often a term shows upin the content yet is remunerated by the event of the term in the corpus

Results and Discussions



The confusion matrix results are given above which looks highly promising as the results of True Positive (TP) and True Negative (TN) are high**.**True Positives are 4718,true Negatives are 4255,False positives are 146 and false negatives are 147

So the accuracy = (TP+TN)/(TP+TN+FN+FP) * 100

= (4718+4255)/(4718+4255+147+146) * 100

=(8973/9266) * 100

=(0.96837) * 100

= 96.837

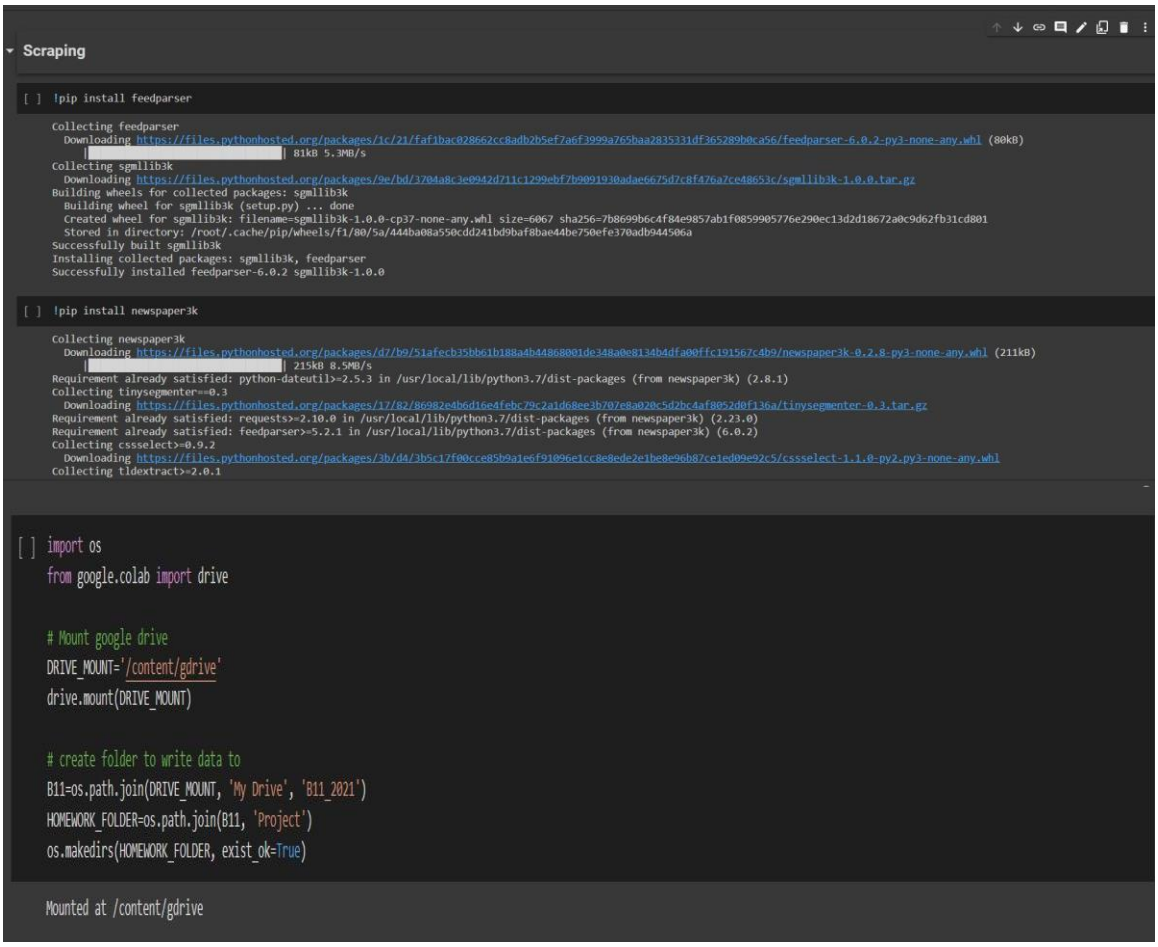
Hardware Requirements:

- Windows 10/8/7/XP

Software Requirements:

- Above 4GB RAM
- Jupyter notebook
- Python
- Colab notebook
- Python 3.6
- Google Drive

IMPLEMENTATION



Scraping

```
[ ] !pip install feedparser

Collecting feedparser
  Downloading https://files.pythonhosted.org/packages/1c/21/faf1bac028662cc8adb2b5ef7a6f3999a765baa2835331df365289b0ca56/feedparser-6.0.2-py3-none-any.whl (80kB)
     |████████████████████████████████| 81kB 5.3MB/s
Collecting sgmllib3k
  Downloading https://files.pythonhosted.org/packages/9e/bd/3704a8c3e0942d711c1299ebf7b9091930adae6675d7c8f476a7ce48653c/sgmllib3k-1.0.0.tar.gz
Building wheels for collected packages: sgmllib3k
  Building wheel for sgmllib3k (setup.py) ... done
  Created wheel for sgmllib3k: filename=sgmllib3k-1.0.0-cp37-none-any.whl size=6067 sha256=7b8699b6c4f84e9857ab1f0859905776e290ec13d2d18672a0c9d62fb31cd801
  Stored in directory: /root/.cache/pip/wheels/f1/80/5a/444ba08a550cdd241bd9baf8bae44be750efe370adb944506a
Successfully built sgmllib3k
Installing collected packages: sgmllib3k, feedparser
Successfully installed feedparser-6.0.2 sgmllib3k-1.0.0

[ ] !pip install newspaper3k

Collecting newspaper3k
  Downloading https://files.pythonhosted.org/packages/d7/b9/51afecb35bb61b188a4b44868001de348a0e8134b4dfa00ffc191567c4b9/newspaper3k-0.2.8-py3-none-any.whl (211kB)
     |████████████████████████████████| 215kB 8.5MB/s
Requirement already satisfied: python-dateutil>=2.5.3 in /usr/local/lib/python3.7/dist-packages (from newspaper3k) (2.8.1)
Collecting tinysegmenter==0.3
  Downloading https://files.pythonhosted.org/packages/17/82/86982e4b6d16e4febc79c2a1d68ee3b707e8a020c5d2bc4af8052d0f136a/tinysegmenter-0.3.tar.gz
Requirement already satisfied: requests>=2.10.0 in /usr/local/lib/python3.7/dist-packages (from newspaper3k) (2.23.0)
Requirement already satisfied: feedparser>=5.2.1 in /usr/local/lib/python3.7/dist-packages (from newspaper3k) (6.0.2)
Collecting cssselect>=0.9.2
  Downloading https://files.pythonhosted.org/packages/3b/d4/3b5c17f00cce85b9a1e6f91096e1cc8e8ede2e1be8e96b87ce1ed09e92c5/cssselect-1.1.0-py2.py3-none-any.whl
Collecting tldextract>=2.0.1
```

```
[ ] import os
    from google.colab import drive

    # Mount google drive
    DRIVE_MOUNT='/content/gdrive'
    drive.mount(DRIVE_MOUNT)

    # create folder to write data to
    B11=os.path.join(DRIVE_MOUNT, 'My Drive', 'B11_2021')
    HOMEWORK_FOLDER=os.path.join(B11, 'Project')
    os.makedirs(HOMEWORK_FOLDER, exist_ok=True)

    Mounted at /content/gdrive
```

```
[ ]  dictionary = {
       "cnn": {
         "link": "http://edition.cnn.com/"
       },
       "bbc": {
         "rss": "http://feeds.bbci.co.uk/news/rss.xml",
         "link": "http://www.bbc.com/"
       },
       "theguardian": {
         "rss": "https://www.theguardian.com/uk/rss",
         "link": "https://www.theguardian.com/international"
       },
       "breitbart": {
         "link": "http://www.breitbart.com/"
       },
       "infowars": {
         "link": "https://www.infowars.com/"
       },
       "foxnews": {
         "link": "http://www.foxnews.com/"
       },
       "nbcnews": {
         "link": "http://www.nbcnews.com/"
       },
       "washingtonpost": {
         "rss": "http://feeds.washingtonpost.com/rss/world",
         "link": "https://www.washingtonpost.com/"
       },
       "theonion": {
         "link": "http://www.theonion.com/"
       }
     }
```

```
[ ]  import feedparser as fp
     import json
     import newspaper
     from newspaper import Article
     from time import mktime
     from datetime import datetime

     # Set the limit for number of articles to download
     LIMIT = 14500

     data = {}
     data['newspapers'] = {}

     # Loads the JSON files with news sites
     with open('NewsPapers.json') as data_file:
         companies = json.load(data_file)

     count = 1

     # Iterate through each news company
     for company, value in companies.items():
         # If a RSS link is provided in the JSON file, this will be the first choice.
         # Reason for this is that, RSS feeds often give more consistent and correct data.
         # If you do not want to scrape from the RSS-feed, just leave the RSS attr empty in the JSON file.
         if 'rss' in value:
             d = fp.parse(value['rss'])
             print("Downloading articles from ", company)
             newsPaper = {
                 "rss": value['rss'],
                 "link": value['link'],
                 "articles": []
             }
             for entry in d.entries:
                 # Check if publish date is provided, if no the article is skipped.
                 # This is done to keep consistency in the data and to keep the script from crashing.
                 if hasattr(entry, 'published'):
                     if count > LIMIT:
                         break
                     article = {}
                     article['link'] = entry.link
                     date = entry.published_parsed
                     article['published'] = datetime.fromtimestamp(mktime(date)).isoformat()
                     try:
                         content = Article(entry.link)
                         content.download()
                         content.parse()
                     except Exception as e:
                         # If the download for some reason fails (ex. 404) the script will continue downloadi
                         # the next article.
                         print(e)
                         print("continuing...")
                         continue
                     article['title'] = content.title
                     article['text'] = content.text
                     newsPaper['articles'].append(article)
                     print(count, "articles downloaded from", company, ", url: ", entry.link)
                     count = count + 1
                             count = count + 1
         else:
             # This is the fallback method if a RSS-feed link is not provided.
             # It uses the python newspaper library to extract articles
             print("Building site for ", company)
             paper = newspaper.build(value['link'], memoize_articles=False)
             newsPaper = {
                 "link": value['link'],
                 "articles": []
             }
```

```
            noneTypeCount = 0
            for content in paper.articles:
                if count > LIMIT:
                    break
                try:
                    content.download()
                    content.parse()
                except Exception as e:
                    print(e)
                    print("continuing...")
                    continue
                # Again, for consistency, if there is no found publish date the article will be skipped.
                # After 10 downloaded articles from the same newspaper without publish date, the company will be skipped.
                if content.publish_date is None:
                    print(count, " Article has date of type None...")
                    noneTypeCount = noneTypeCount + 1
                    if noneTypeCount > 100:
                        print("Too many noneType dates, aborting...")
                        noneTypeCount = 0
                        break
                    count = count + 1
                    continue
                article = {}
                article['title'] = content.title
                article['text'] = content.text
                article['link'] = content.url
                article['published'] = content.publish_date.isoformat()
                newsPaper['articles'].append(article)
                print(count, "articles downloaded from", company, " using newspaper, url: ", content.url)
                count = count + 1
                noneTypeCount = 0
        count = 1
        data['newspapers'][company] = newsPaper

    # Finally it saves the articles as a JSON-file.
    try:
        with open('scraped_articles.json', 'w') as outfile:
            json.dump(data, outfile)
    except Exception as e: print(e)
```

```
Building site for  cnn
1  Article has date of type None...
2  Article has date of type None...
3  Article has date of type None...
```

```
62 Article has date of type None...
63 Article has date of type None...
64 Article has date of type None...
65 Article has date of type None...
66 Article has date of type None...
67 Article has date of type None...
68 Article has date of type None...
69 Article has date of type None...
70 Article has date of type None...
71 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/16/myanmar-bago-killing-hancocks-pkg-intl-hnk-vpx.cnn/video/playlists/around-the-world/
72 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/15/wallet-flyer-milk-sotll-lon-orig-na.cnn/video/playlists/around-the-world/
73 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/16/hong-kong-national-security-education-day-lu-stout-wt-intl-hnk-vpx.cnn/video/playlists/around-the-world/
74 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/16/brazil-coronavirus-tajr-bolsonaro-rio-de-janeiro-darlington-pkg-intl-ldn-vpx.cnn/video/playlists/around-the-world/
75 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/politics/2021/04/15/biden-russia-sanctions-announcement-sot-vpx.cnn/video/playlists/around-the-world/
76 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/15/afghanistan-us-withdrawal-biden-fatima-gailani-intv-anderson-ctw-vpx.cnn/video/playlists/around-the-world/
77 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/15/prince-charles-duchess-cornwall-prince-philip-tributes-ctw-intl-ldn-vpx.cnn/video/playlists/around-the-world/
78 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/14/afghanistan-taliban-life-us-troops-withdrawal-naton-walsh-nkw-pkg-intl-hnk-vpx.cnn/video/playlists/around-the-world/
79 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/sports/2021/04/14/tokyo-olympics-100-days-countdown-essie-pkg-intl-hnk-vpx.cnn/video/playlists/around-the-world/
80 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/13/japan-fukushima-water-release-essie-intl-hnk-vpx.cnn/video/playlists/around-the-world/
81 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/business/2021/04/13/egypt-seizes-ever-given-container-ship-suez-canal-ctw-intl-ldn-vpx.cnn/video/playlists/around-the-world/
82 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/13/china-military-taiwan-biden-xi-jinping-culver-pkg-intl-ldn-vpx.cnn/video/playlists/around-the-world/
83 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/12/zelensky-ukraine-matthew-chance-pkg-intl-ldn-vpx.cnn/video/playlists/around-the-world/
84 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/12/iran-nuclear-incident-natanz-israel-piesteen-gold-intl-ldn-vpx.cnn/video/playlists/around-the-world/
85 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/09/la-soufriere-st-vincent-volcano-eruption-mh-orig.cnn/video/playlists/around-the-world/
86 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/12/prince-philip-yamashiro-special-bond-riviere-pkg-intl-hnk-vpx.cnn/video/playlists/around-the-world/
87 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/15/myanmar-covid-19-coronavirus-watson-pkg-intl-hnk-vpx.cnn/video/playlists/coronavirus-intl/
88 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/16/brazil-coronavirus-tajr-bolsonaro-rio-de-janeiro-darlington-pkg-intl-ldn-vpx.cnn/video/playlists/coronavirus-intl/
89 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/health/2021/04/16/british-mom-pregnant-while-already-pregnant-cohen-pkg-newday-vpx.cnn/video/playlists/coronavirus-intl/
90 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/sports/2021/04/14/tokyo-olympics-100-days-countdown-essie-pkg-intl-hnk-vpx.cnn/video/playlists/coronavirus-intl/
91 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/13/who-coronavirus-reality-check-pandemic-ian-kerknove-berman-intv-newday-intl-ldn-vpx.cnn/video/playlists/coronavirus-intl/
92 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/business/2021/04/12/ewbs-england-reopen-coronavirus-lockdown-stewart-intl-ldn-vpx.cnn/video/playlists/coronavirus-intl/
93 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/12/india-covid-ganges-river-ctw-intl-ldn-vpx.cnn/video/playlists/coronavirus-intl/
94 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/07/lucha-libre-wrestlers-masks-mexico-covid-orig-lk.cnn/video/playlists/coronavirus-intl/
95 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/04/02/brazil-coronavirus-tajr-bolsonaro-darlington-lok-intl-ldn-vpx.cnn/video/playlists/coronavirus-intl/
96 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/03/31/serbia-coronavirus-covid-19-vaccine-hub-pleitgen-pkg-intl-hnk-vpx.cnn/video/playlists/coronavirus-intl/
97 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/03/31/who-findings-concerns-coronavirus-covid-19-intl-hnk-vpx.cnn/video/playlists/coronavirus-intl/
98 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/03/29/cuba-coronavirus-vaccines-speakman-pkg-intl-ldn-vpx.cnn/video/playlists/coronavirus-intl/
99 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/03/26/coronavirus-brazil-record-hospitals-shortages-rivers-ctw-intl-ldn-vpx.cnn/video/playlists/coronavirus-intl/
100 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/03/25/eu-coronavirus-vaccine-export-controls-weber-intv-anderson-cte-intl-ldn-vpx.cnn/video/playlists/coronavirus-intl/
101 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/world/2021/03/23/uk-coronavirus-lockdown-anniversary-black-ctw-pkg-intl-ldn-vpx.cnn/video/playlists/coronavirus-intl/
102 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/politics/2021/04/17/marjorie-taylor-greene-america-first-caucus-sot-vpx-ac360.cnn/video/playlists/this-week-in-politics/
103 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/politics/2021/04/16/biden-white-house-refugee-cap-sot-vpx.cnn/video/playlists/this-week-in-politics/
104 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/politics/2021/04/16/joe-biden-shooting-reaction-press-conference-sot-vpx.cnn/video/playlists/this-week-in-politics/
105 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/politics/2021/04/16/russia-expel-10-us-diplomats-lavrov-biden-sanctions-dougherty-intv-king-ie-vpx.cnn/video/playlists/this-week-in-politics/
106 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/politics/2021/04/16/reality-check-100-days-since-capitol-riot-avlon-newday-vpx.cnn/video/playlists/this-week-in-politics/
107 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/politics/2021/04/16/foreign-policy-russia-sanctions-joe-biden-ian-bremmer-cpt-sot-vpx.cnn/video/playlists/this-week-in-politics/
108 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/politics/2021/04/15/training-reform-police-reality-check-avlon-sot-newday-vpx.cnn/video/playlists/this-week-in-politics/
109 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/politics/2021/04/15/governor-michelle-lujan-grisham-former-staffer-settlement-todd-pkg-vpx.cnn/video/playlists/this-week-in-politics/
110 articles downloaded from cnn  using newspaper, url: http://edition.cnn.com/videos/media/2021/03/25/ted-cruz-reporter-mask-vaccinations-sanjay-gupta-newday-sot-vpx.cnn/video/playlists/this-week-in-politics/
```

```
[ ] import pandas as pd
    for i, site in enumerate(list(d['newspapers'])):
        articles = list(d['newspapers'][site]['articles'])
        if i == 0:
            df = pd.DataFrame.from_dict(articles)
            df["site"] = site
        else:
            new_df = pd.DataFrame.from_dict(articles)
            new_df["site"] = site
            df = pd.concat([df, new_df], ignore_index = True)
```

```
[ ] df.shape
```

```
(1438, 5)
```

```
[ ] df
```

	title	text	link	published	site
0	Eyewitnesses recount bloody crackdown in Bago,...	At least 82 anti-coup protestors were killed b...	http://edition.cnn.com/videos/world/2021/04/16...	2021-04-16T00:00:00	cnn
1	This Welsh river turned white due to a milk spill	Photos and videos of the River Dulais in Wales...	http://edition.cnn.com/videos/world/2021/04/16...	2021-04-16T00:00:00	cnn
2	Hong Kong police showcase 'Chinese-style goose...	Hong Kong marked the first National Security E...	http://edition.cnn.com/videos/world/2021/04/16...	2021-04-16T00:00:00	cnn
3	In Brazil, coronavirus killed 3 people every m...	Experts warn Brazil could soon suffer an "unim...	http://edition.cnn.com/videos/world/2021/04/16...	2021-04-16T00:00:00	cnn
4	New sanctions imposed on Russia in response to...	The Biden administration targeted Russia with ...	http://edition.cnn.com/videos/politics/2021/04...	2021-04-15T00:00:00	cnn
...	...	...	...	...	...
1433	12 Steps to Starting a Small Business	Getty Images\n\nBeing your own boss can be imm...	https://www.nbcnews.com/veteran-services/next-...	2019-05-02T18:39:00+00:00	nbcnews
1434	Military families say this is their top concern	Members of the military face hurdles every day...	https://www.nbcnews.com/veteran-services/next-...	2019-05-28T17:17:00+00:00	nbcnews
1435	CNBC AND ACORNS ANNOUNCE STRATEGIC PARTNERSHIP	CNBC TO PROVIDE EDITORIAL AND PRODUCTION EXPER...	https://www.cnbc.com/2019/01/28/cnbc-and-acorn...	2019-01-28T00:00:00	nbcnews
1436	'Captain Tom': Funeral held for U.K. war veter...	LONDON — "I told you I was old," will be the e...	https://www.nbcnews.com/news/world/captain-tom...	2021-02-27T16:19:00+00:00	nbcnews
1437	Why veterans struggle to share their stories w...	Lt. Col. Paul Huszar had been on a mission for...	https://www.today.com/series/veterans/veterans...	2020-11-10T21:42:00+00:00	nbcnews

1438 rows × 5 columns

```
[ ] !cp scraped_articles.json "/content/gdrive/My Drive/B11_2021/Project/"
```

Exploratory Analysis + Pre-Processing

```
[ ] import json
    import pandas as pd
    import numpy as np
    import seaborn as sns
    import matplotlib.pyplot as plt
    import nltk
    from nltk.corpus import stopwords
    from nltk.tokenize import word_tokenize
    import spacy
    import wordcloud
    from wordcloud import WordCloud
    import statistics
    from collections import Counter
    import datetime
    import textblob
    from textblob import TextBlob
    import sklearn
```

```
[ ] import os
    from google.colab import drive

    # Mount google drive
    DRIVE_MOUNT='/content/gdrive'
    drive.mount(DRIVE_MOUNT)

    # create folder to write data to
    B11=os.path.join(DRIVE_MOUNT, 'My Drive', 'B11_2021')
    HOMEWORK_FOLDER=os.path.join(B11, 'Project')
    os.makedirs(HOMEWORK_FOLDER, exist_ok=True)
```

```
Mounted at /content/gdrive
```

```
[ ] fake = pd.read_csv('/content/gdrive/MyDrive/B11_2021/Project/Fake.csv')
    true = pd.read_csv('/content/gdrive/My Drive/B11_2021/Project/True.csv')
    with open('/content/gdrive/My Drive/B11_2021/Project/scraped_articles.json') as json_data:
        scraped = json.load(json_data)
```

```
[ ] fake.head()
```

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017

```
[ ] true.head()
```

	title	text	subject	date
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017

```
[ ] print(list(scraped['newspapers']))
```

```
['cnn', 'bbc', 'theguardian', 'breitbart', 'infowars', 'foxnews', 'nbcnews', 'washingtonpost', 'theonion']
```

```
for i, site in enumerate((list(scraped['newspapers']))):
    articles = list(scraped['newspapers'][site]['articles'])
    if i == 0:
        df = pd.DataFrame.from_dict(articles)
        df["site"] = site
    else:
        new_df = pd.DataFrame.from_dict(articles)
        new_df["site"] = site
        df = pd.concat([df, new_df], ignore_index = True)
```

```
[ ] df
```

	title	text	link	published	site
0	Eyewitnesses recount bloody crackdown in Bago,...	At least 82 anti-coup protestors were killed b...	http://edition.cnn.com/videos/world/2021/04/16...	2021-04-16T00:00:00	cnn
1	This Welsh river turned white due to a milk spill	Photos and videos of the River Dulais in Wales...	http://edition.cnn.com/videos/world/2021/04/16...	2021-04-16T00:00:00	cnn
2	Hong Kong police showcase 'Chinese-style goose...	Hong Kong marked the first National Security E...	http://edition.cnn.com/videos/world/2021/04/16...	2021-04-16T00:00:00	cnn

```
[ ] print(list(scraped['newspapers']))
```

```
['cnn', 'bbc', 'theguardian', 'breitbart', 'infowars', 'foxnews', 'nbcnews', 'washingtonpost', 'theonion']
```

```
for i, site in enumerate((list(scraped['newspapers']))):
    articles = list(scraped['newspapers'][site]['articles'])
    if i == 0:
        df = pd.DataFrame.from_dict(articles)
        df["site"] = site
    else:
        new_df = pd.DataFrame.from_dict(articles)
        new_df["site"] = site
        df = pd.concat([df, new_df], ignore_index = True)
```

```
[ ] df
```

	title	text	link	published	site
0	Eyewitnesses recount bloody crackdown in Bago,...	At least 82 anti-coup protestors were killed b...	http://edition.cnn.com/videos/world/2021/04/16...	2021-04-16T00:00:00	cnn
1	This Welsh river turned white due to a milk spill	Photos and videos of the River Dulais in Wales...	http://edition.cnn.com/videos/world/2021/04/16...	2021-04-16T00:00:00	cnn
2	Hong Kong police showcase 'Chinese-style goose...	Hong Kong marked the first National Security E...	http://edition.cnn.com/videos/world/2021/04/16...	2021-04-16T00:00:00	cnn
3	In Brazil, coronavirus killed 3 people every m...	Experts warn Brazil could soon suffer an "unim...	http://edition.cnn.com/videos/world/2021/04/16...	2021-04-16T00:00:00	cnn
4	New sanctions imposed on Russia in response to...	The Biden administration targeted Russia with ...	http://edition.cnn.com/videos/politics/2021/04...	2021-04-15T00:00:00	cnn
...	...	...	...	...	...
1433	12 Steps to Starting a Small Business	Getty Images\n\nBeing your own boss can be imm...	https://www.nbcnews.com/veteran-services/next-...	2019-05-02T18:39:00+00:00	nbcnews
1434	Military families say this is their top concern	Members of the military face hurdles every day...	https://www.nbcnews.com/veteran-services/next-...	2019-05-28T17:17:00+00:00	nbcnews
1435	CNBC AND ACORNS ANNOUNCE STRATEGIC PARTNERSHIP	CNBC TO PROVIDE EDITORIAL AND PRODUCTION EXPER...	https://www.cnbc.com/2019/01/28/cnbc-and-acorn...	2019-01-28T00:00:00	nbcnews
1436	'Captain Tom': Funeral held for U.K. war veter...	LONDON — "I told you I was old," will be the e...	https://www.nbcnews.com/news/world/captain-tom...	2021-02-27T16:19:00+00:00	nbcnews
1437	Why veterans struggle to share their stories w...	Lt. Col. Paul Huszar had been on a mission for...	https://www.today.com/series/veterans/veterans...	2020-11-10T21:42:00+00:00	nbcnews

1438 rows x 5 columns

```
[ ] scraped = df
    scraped["label"] = scraped["site"].apply(lambda x: "fake" if(x == "breitbart" or x == "infowars" or x == "theonion") else "true")
    scraped.drop(labels=["link", "site"], axis = 1, inplace=True)
```

```
[ ] mask = scraped['label'] == 'fake'
    scraped[mask]
```

```
[ ] #Create column with labels
    fake["label"] = "fake"
    true["label"] = "true"
```

```
[ ] true.head()
```

	title	text	subject	date	label
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017	true
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017	true
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017	true
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017	true
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017	true

```
[ ] fake.head()
```

	title	text	subject	date	label
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017	fake
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	fake
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	fake
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	fake
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	fake

```
[ ] #Check shapes of dataframes. I scraped 1844 articles to add to this dataset.
    print (fake.shape)
    print (true.shape)
    print (scraped.shape)
```

```
(23481, 5)
(21417, 5)
(1438, 4)
```

```
[ ] scraped["published"] = scraped["published"].apply(lambda x: x[0:10])
    scraped["published"] = scraped["published"].apply(pd.to_datetime)
```

```
[ ]
    print (fake["subject"].unique())
    print (true["subject"].unique())
```

```
['News' 'politics' 'Government News' 'left-news' 'US_News' 'Middle-east']
['politicsNews' 'worldnews']
```

```
[ ]
```

```
[ ] fake.drop("subject", axis=1, inplace=True)
    true.drop("subject", axis=1, inplace=True)
```

```
[ ] #Combine scraped with current datasets
    scraped.rename(columns={"published": "date"}, inplace=True)
    scraped_f = scraped[scraped["label"] == "fake"]
    fake = pd.concat([fake,scraped_f], axis=0, ignore_index=True)

    scraped_t = scraped[scraped["label"] == "true"]
    true = pd.concat([true,scraped_t], axis=0, ignore_index=True)
```

```
[ ] #Remove articles with only pictures / no text
    true = true[true["text"] != ""]
    fake = fake[fake["text"] != ""]
```

```
[ ] #The datasets are pretty balanced, which is good!
    print(fake.shape)
    print(true.shape)
```

```
fake.drop("title", axis=1, inplace=True)
true.drop("title", axis=1, inplace=True)
```

```
[ ] fake.head()
```

	text	label
0	Donald Trump just couldn t wish all Americans ...	fake
1	House Intelligence Committee Chairman Devin Nu...	fake
2	On Friday, it was revealed that former Milwauk...	fake
3	On Christmas day, Donald Trump announced that ...	fake
4	Pope Francis used his annual Christmas Day mes...	fake

```
[ ] true.head()
```

	text	label
0	The head of a conservative Republican faction...	true
1	Transgender people will be allowed for the fi...	true
2	The special counsel investigation of links be...	true
3	Trump campaign adviser George Papadopoulos to...	true
4	President Donald Trump called on the U.S. Pos...	true

```
[ ] final = pd.concat([true,fake], axis=0, ignore_index=True)
```

```
[ ] final.shape
```

```
(46330, 2)
```

```
[ ] 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
```

```
[ ] x = final['text']
    y = final['label']
```

```
[ ] x
```

```
0         The head of a conservative Republican faction...
1         Transgender people will be allowed for the fi...
2         The special counsel investigation of links be...
3         Trump campaign adviser George Papadopoulos to...
4         President Donald Trump called on the U.S. Pos...
                              ...
46325    Have an important tip? Let us know. Email us h...
46326    Keep up to date with our latest:\n\nHave an im...
46327    Have an important tip? Let us know. Email us h...
46328    Chaz Neal, a prominent BLM activist in Minneso...
46329    Have an important tip? Let us know. Email us h...
Name: text, Length: 46330, dtype: object
```

```
[ ] y
```

```
0        true
1        true
2        true
3        true
4        true
         ...
46325    fake
46326    fake
46327    fake
46328    fake
46329    fake
Name: label, Length: 46330, dtype: object
```

```
[ ] x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=0)
    y_train
```

```
[ ] tfvect = TfidfVectorizer(stop_words='english',max_df=0.7)
    tfid_x_train = tfvect.fit_transform(x_train)
    tfid_x_test = tfvect.transform(x_test)
```

```
[ ] classifier = PassiveAggressiveClassifier(max_iter=50)
    classifier.fit(tfid_x_train,y_train)

    PassiveAggressiveClassifier(C=1.0, average=False, class_weight=None,
                                early_stopping=False, fit_intercept=True,
                                loss='hinge', max_iter=50, n_iter_no_change=5,
                                n_jobs=None, random_state=None, shuffle=True,
                                tol=0.001, validation_fraction=0.1, verbose=0,
                                warm_start=False)
```

```
[ ] y_pred = classifier.predict(tfid_x_test)
    score = accuracy_score(y_test,y_pred)
    print(f'Accuracy: {round(score*100,2)}%')

    Accuracy: 98.29%
```

```
[ ] cf = confusion_matrix(y_test,y_pred, labels=['fake','true'])
    print(cf)

    [[4791   73]
     [  85 4317]]
```

```
[ ] def fake_news_det(news):
        input_data = [news]
        vectorized_input_data = tfvect.transform(input_data)
        prediction = classifier.predict(vectorized_input_data)
        print(prediction)
```

```
[ ] fake_news_det('U.S. Secretary of State John F. Kerry said Monday that he will stop in Paris later this week, amid criticism that no top American officials attended Sunday's unity march against terrorism.')

    ['true']
```

```
[ ] fake_news_det("""Go to Article
    President Barack Obama has been campaigning hard for the woman who is supposedly going to extend his legacy four more years. The only problem with stumping for Hillary Clinton, however, is she's not exactly a candidate easy to get too enthused about.  """)

    ['fake']
```

RESULTS

Binomial Logistic Regression

```
[ ] from sklearn.linear_model import LogisticRegression

    # 2. instantiate a logistic regression model
    lr = LogisticRegression()

    lr.fit(tfid_x_train,y_train)
```

```
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                   intercept_scaling=1, l1_ratio=None, max_iter=100,
                   multi_class='auto', n_jobs=None, penalty='l2',
                   random_state=None, solver='lbfgs', tol=0.0001, verbose=0,
                   warm_start=False)
```

```
[ ] y_pred_lr = lr.predict(tfid_x_test)
    score = accuracy_score(y_test,y_pred_lr)
    print(f'Accuracy: {round(score*100,2)}%')
```

```
Accuracy: 96.84%
```

```
[ ] cf = confusion_matrix(y_test,y_pred_lr, labels=['fake','true'])
    print(cf)
```

```
[[4718  146]
 [ 147 4255]]
```

Support Vector Machines

```
[ ] from sklearn.svm import SVC

    #Linear kernel fits decently well, decided not to use another kernel because of parsimony
    #and because linear has the lowest risk of overfitting
    svc = SVC(kernel='linear', random_state=1)
    svc.fit(tfid_x_train,y_train)
```

```
SVC(C=1.0, break_ties=False, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma='scale', kernel='linear',
    max_iter=-1, probability=False, random_state=1, shrinking=True, tol=0.001,
    verbose=False)
```

```
[ ] y_pred_svm = svc.predict(tfid_x_test)
    score = accuracy_score(y_test,y_pred_svm)
    print(f'Accuracy: {round(score*100,2)}%')
```

```
Accuracy: 97.8%
```

CONCLUSION

With social media increasingly prevalent, more and more people are receiving news from social media rather than traditional news media. Online networking has since been used to disseminate misleading news, which has had significant adverse effects on individual consumers and broader community.In this paper, we discussed the problem of fake news by combining two separate approaches for greater accuracy in identifying false news. Based on the findings discussed above, the results of this analysis indicate that a Fake News Classifier can detect false news with 96% accuracy.

SCOPE FOR FUTURE DEVELOPMENT

There is scope for the future development of this project. Computer technology keeps finding new methods and technologies on a day to day basis. It is dynamic and not static. The skills which are prominent today will become obsolete in a few days. To keep pace with the technical developments, the system may be additionally improved. So, it is not concluded.

Further extensions to this system can be made required with minor modifications. It can be developed for wearable devices which will be much more easier. Yet it will improve with further augmentations. Augmentations can be done effectually. We can even apprise the same with further changes and can be integrated with minimal alteration. Thus the project is flexible and can be improved at any time with more advanced features.

REFERENCES

[1]Torabi Asr, Fatemeh, and Maite Taboada. "Big Data and quality data for fake news and misinformation detection."Big Data & Society6, no. 1 (2019): 2053951719843310. [2]Granik, Mykhailo, and Volodymyr Mesyura. "Fake news detection using naive Bayes classifier." In2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), pp. 900-903. IEEE, 2017.

[3]Zhou, Xinyi, Reza Zafarani, Kai Shu, and Huan Liu. "Fakenews: Fundamental theories, detection strategies and challenges." InProceedings of the Twelfth ACM International Conference on Web Search and Data Mining, pp. 836-837. ACM, 2019.

[4]Shu, Kai, Deepak Mahudeswaran, and Huan Liu. "FakeNewsTracker: a tool for fake news collection, detection, and visualization."Computational and Mathematical Organization Theory25, no. 1 (2019): 60-71.

[5]Figueira, Á., & Oliveira, L. (2017). The current state of fake news: challenges and opportunities.Procedia Computer Science,121, 817–825. doi: 10.1016/j.procs.2017.11.106