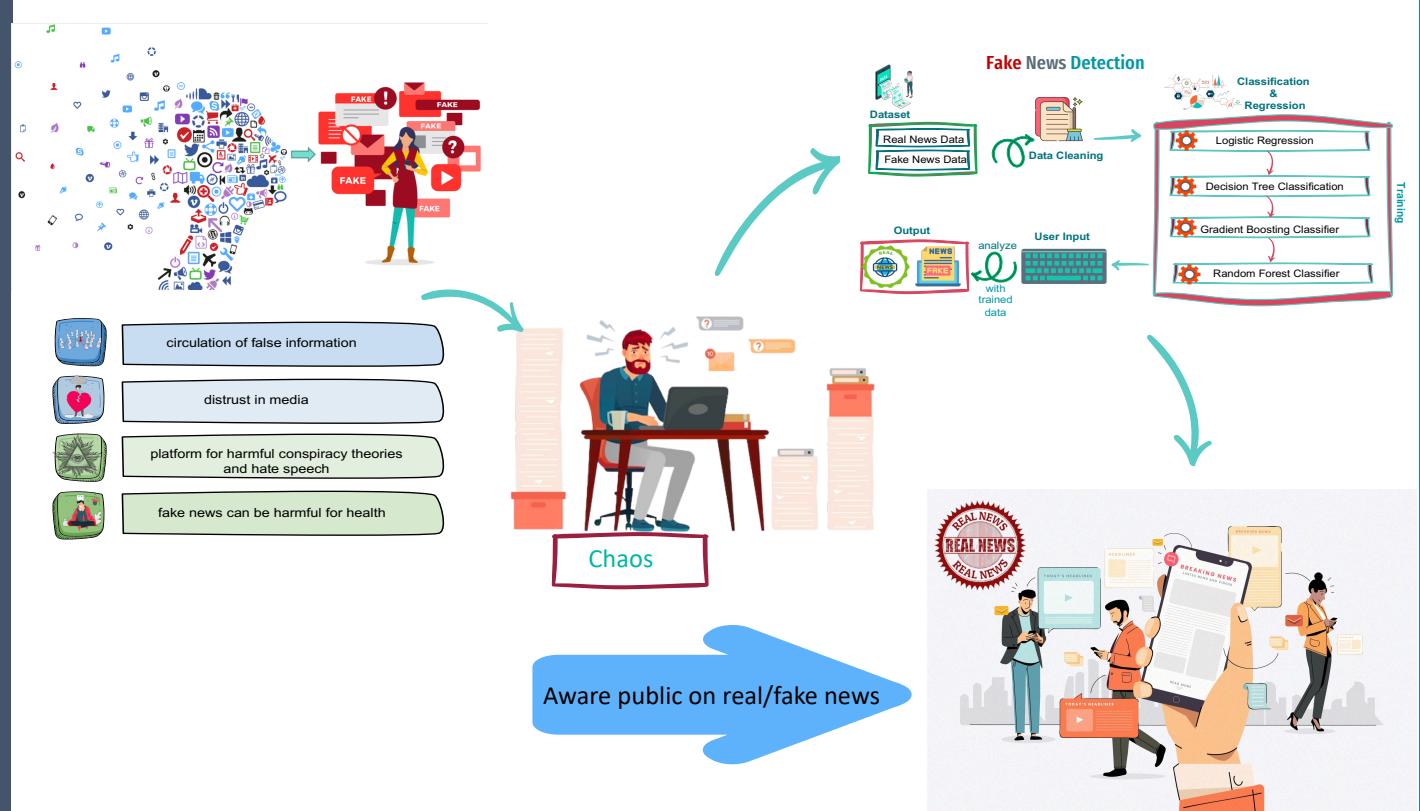


Research on Fake News Detection using classification and regression in Machine Learning

Softwarica college of IT and E-Commerce



Submitted By:
Bijaya Acharya
CU ID: 10269235
SID: 190374

Submitted To:
Manoj Shrestha
Lecturer, Softwarica college of IT and E-Commerce

Deceleration

I hereby state that this paper is entirely my own original creative work, which I have developed independently. All sources, references, and pieces of literature consulted in the course of developing this work have been accurately cited and listed with full references to each one.

Bijaya Acharya

Acknowledgement

First and foremost, I would want to express my appreciation to Mr. Manoj Shrestha, who acted as my thesis' supervisor, for his persistent optimism and advice, which helped me successfully complete this research. In addition, he brought a sense of exploration and enthusiasm to his research and teaching; without him, this dissertation would not have been as interesting and knowledgeable as this.

I want to express my gratitude to Coventry University, Softwarica College of It & Ecommerce, and every individual who helped make my attendance at this institution possible.

I also want to express my gratitude to my family and friends for their support and assistance in helping me accomplish my goal.

Abstract

It is getting harder and harder to tell the difference between real and incorrect information in the age of social media and online news sources. There is proof that spreading fake news has been utilized in the past to sway public opinion and convince people in order to meet individual or group interests. This makes the subject of false news and its accurate identification crucial. The broad consensus is that the person receiving the information should investigate the sources. The veracity of the news, however, could be a challenging issue that calls for more than a single viewpoint based on a news source. In this thesis, Classification & Regression in Machine Learning in Jupyter Notebook are used to build an algorithm for a mockup. For training purposes, we gathered about 20,000 false news stories and 20,000 actual news stories. We used Anaconda environment to start the Jupyter notebook and python environment. After that we successfully built a working system. Even if it isn't always correct, utilizing those facts to train our system has allowed us to create one that can distinguish between false and real news. As the system has more data to assess, the likelihood that it will provide an appropriate categorization rises as the number of words grows. After the multiple trials and errors, we also build it in a way that user can input the news and it categorize the given information. So, this thesis summarizes the purpose for building this system along with the solution that we build to tackle this problem.

Keywords

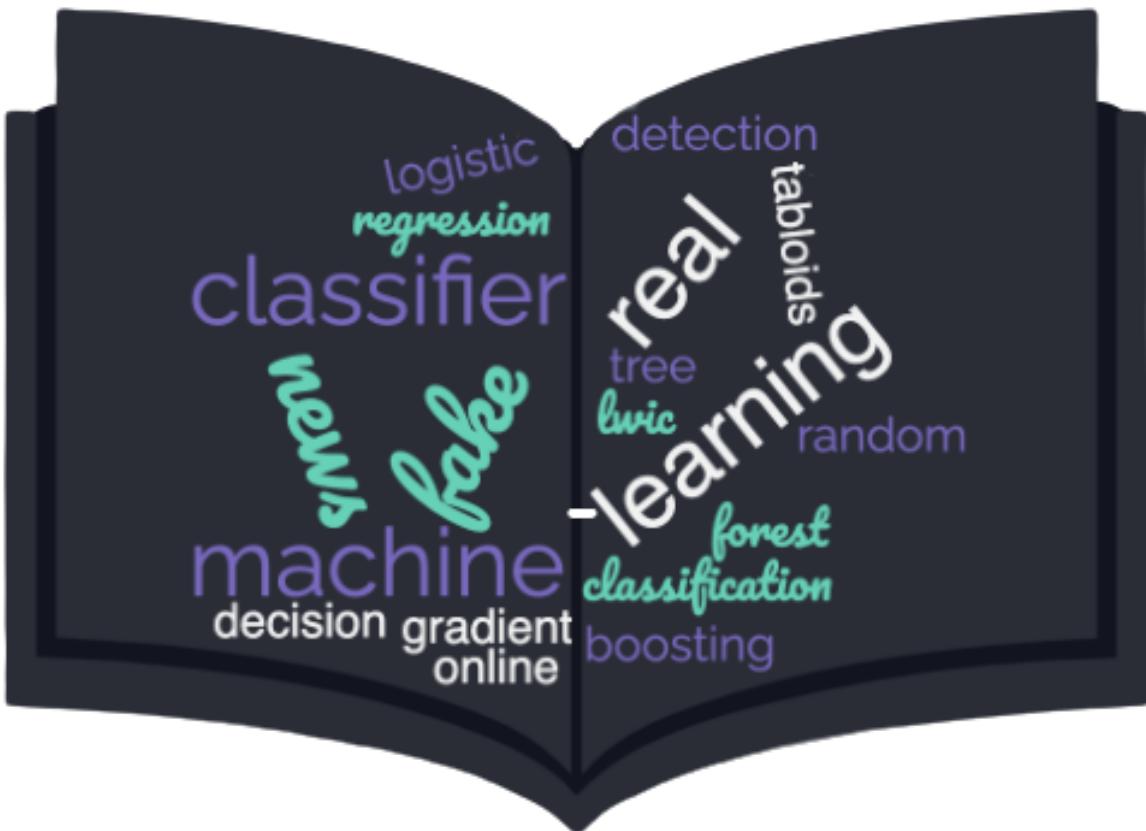


Figure i: Keywords

Table of Contents

<i>Acknowledgement</i>	2
<i>Abstract</i>	3
<i>Keywords</i>	4
<i>Introduction</i>	7
Aim	9
Objectives.....	10
<i>Justification</i>	10
Problem.....	10
Solution	11
<i>Research Question</i>	12
<i>Scope</i>	13
<i>Ethical Consideration</i>	13
<i>Literature Review</i>	15
Research Methodology	15
Desk-based.....	15
<i>Development Methodology</i>	25
Agile	25
Why Agile?.....	26
<i>Tools & Technology</i>	27
<i>Findings</i>	29
<i>Future Work</i>	30
<i>Project and Issue Management</i>	32
Project	32
Issue Management	32
<i>Conclusion</i>	33
<i>References</i>	33
<i>Appendix</i>	37

Table of Figures

Figure i: Keywords.....	4
Figure ii: Fake News	7
Figure iii: us election 2016 fake news status (Rogers, 2020).....	8
Figure iv: Fake news spread reason (FRSA, 2020).....	9
Figure v: Aim	9
Figure vi: Objectives	10
Figure vii: problems of fake news	11
Figure viii: solutions	12
Figure ix: project scope	13
Figure x: Ethical considerations	14
Figure xi: Desk-based Research	15
Figure xii: Automatic Fake news Detection Using LIWC.....	16
Figure xiii: fake news sources	17
Figure xiv: Fake News Detection in Social Media	18
Figure xv: Conceptual Fake News Detection Model	18
Figure xvi: Fake News Detection Using NLP	20
Figure xvii: LTSM detection graph.....	21
Figure xviii: Kauwa-kaate Fake News Detection System: Demo	22
Figure xix: computational fact checking.....	23
Figure xx: fake news detection using LIAR dataset.....	24
Figure xxi: agile	26
Figure xxii: why Agile?.....	27
Figure xxiii: Tools used in project.....	28
Figure xxiv: technologies used in research.....	29
Figure xxv: future work	30
Figure xxvi: Gantt chart	32
Figure xxvii: issues.....	32

Introduction

"I have said that propaganda, misinformation, and disinformation have always been part of political warfare. Social media and other new platforms have given it a new life and reach through which the fake news phenomenon can reach everywhere." - Bilawal Bhutto Zardari

Information is now more accessible than ever thanks to social media and quick and easy global news circulation, as well as the internet's growing accessibility. By 2023, there will be around 5.3 billion internet users, up from the around 3.9 billion in 2018 (Cissco, 2020). Along with the rising global social penetration rate, which peaked at around 71% in East Asia and North America and reached around 49% globally (Dixon, 2021). The social media interactions, such as tweets, shares, comments, likes, and responses, are effective methods for spreading knowledge among peers and the wider public. Because of these developments, news may be delivered to the recipient relatively quickly, and the variety of information sources may be too much for the wide public (Persia, 2017). These elements contribute to the rapid spread of misleading information and purposely created false news, raising the question of what one should believe in. Then that false news results in the having negative impact in the world outside of the digital world. Therefore, I believe that creating an algorithm that can identify potential false news or developing a system that alerts individuals that the media being displayed could be fake news is the simplest way to notify people about the false information available on the internet.

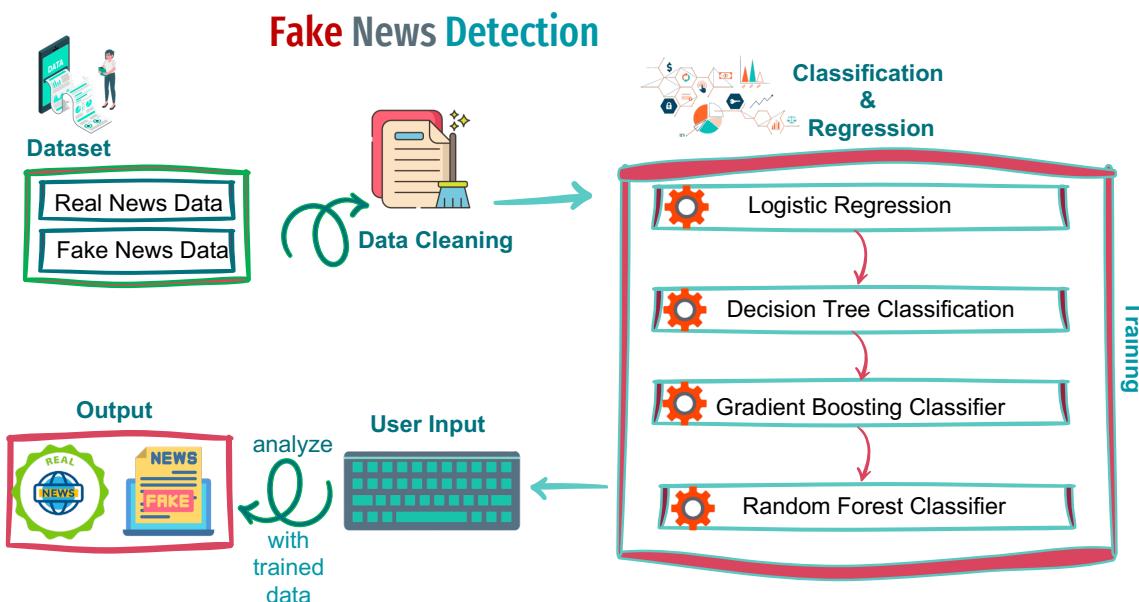


Figure ii: Fake News

False news, according to the Cambridge Dictionary (University, 2022), is "the misleading information or information generated with the goal to deceive."

The 2016 U.S. presidential elections were one instance in recent years where the deliberate dissemination of false information was employed to sway public opinion. The amount of misinformation spreading in relation to the presidential election quickly exceeded the

number of articles produced by reliable news organizations (Silverman, 2016). This study shows that two of the ten most popular engagements were posted by one of the 100 websites operated out of the Republic of North Macedonia in order to generate cash from advertisements (Lawrence Alexander, 2016). In a different investigation (Makse, 2019), researchers discovered that 25% of tweets had a link to a news source that was spreading false or biased information. This finding was used to explain how fake news spreaders influenced Trump and Hillary Clinton supporters.

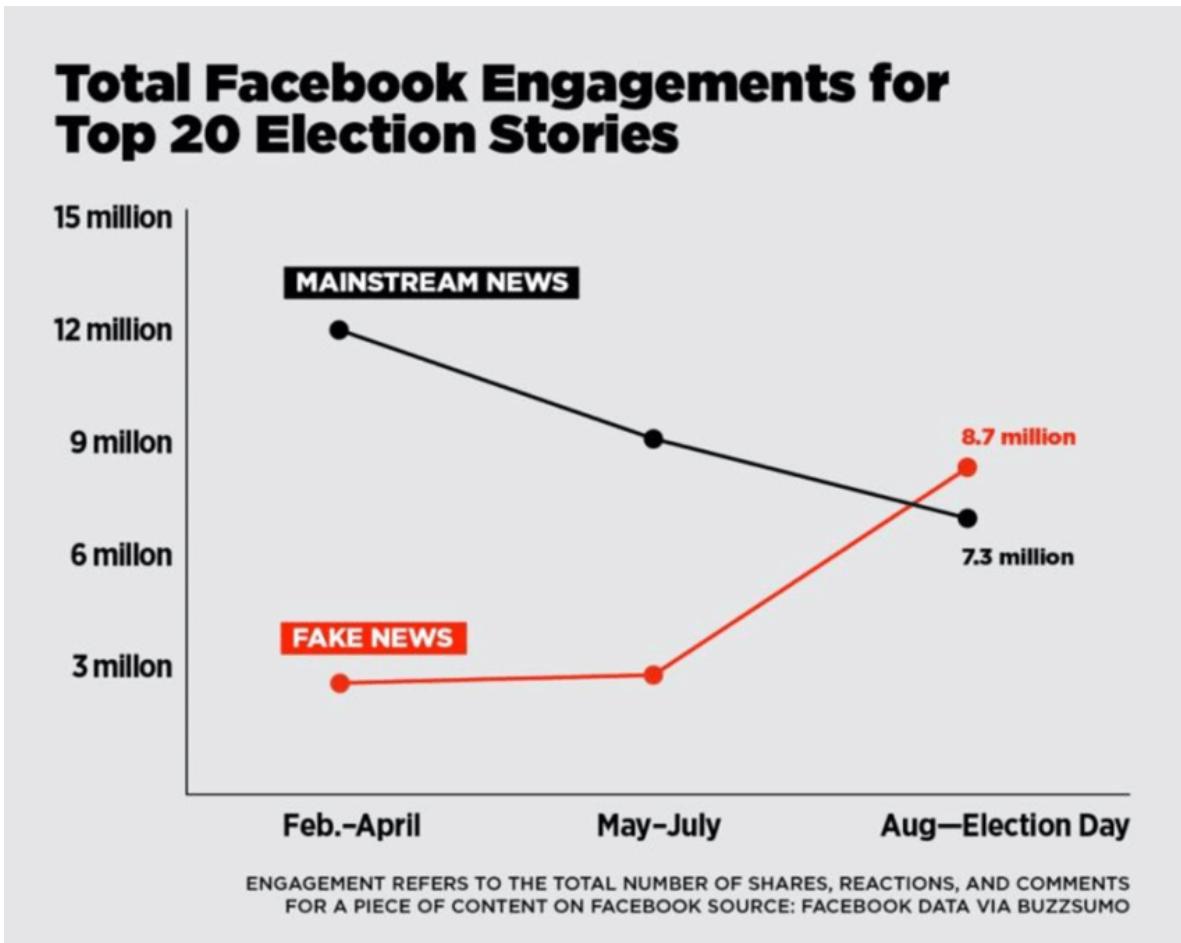


Figure iii: us election 2016 fake news status (Rogers, 2020)

The coronavirus pandemic of 2020 is another significant occurrence that is linked to the rapid spread of false information. Since the Spanish flu pandemic of 1918, this worldwide event the largest viral outbreak has had an impact on people all over the world. The pandemic, according to the Nature article (Fleming, 2020), has caused the spread of a variety of false information, from the life-threatening information about the intravenous use of disinfectant as a form of treatment to the news that Russian President Vladimir Putin released lions to force people to stay at home. Misinformation is disseminated for political and financial purposes. According to the article, more than 23% of Americans think the virus was developed in a Chinese laboratory. This assertion can be regarded as incorrect since there is no supporting data and it can have political motivations. Additionally, it lists the eight signs of false information. Specifically, the dependability of the source, foul language, strong feelings, the quantity of sources, the dependability of social media spreaders, excessive sharing, and potential for one side to benefit.



Figure iv: Fake news spread reason (FRSA, 2020)

For the purpose of detecting fake news, in this project we build computational tools and models in this study. We present two dataset collections that include both authentic and false news in separate files. While the second dataset is gathered straight from the web, the first dataset was compiled using a combination of manual and crowdsourcing annotation techniques. We create false news detectors relying on linguistic features that reach accuracies of up to 76% using these datasets, and we do multiple exploratory studies to find linguistic traits that are largely prevalent in fake news material. We contrast the performance of the created classifiers with an actual human baseline to put our findings into context.

Aim

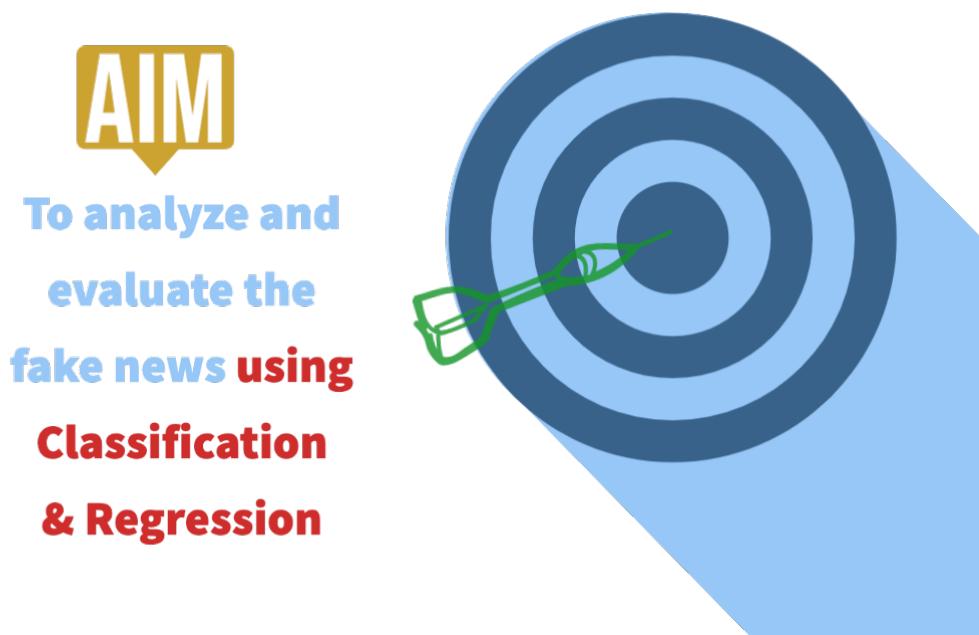


Figure v: Aim

Objectives

Objectives

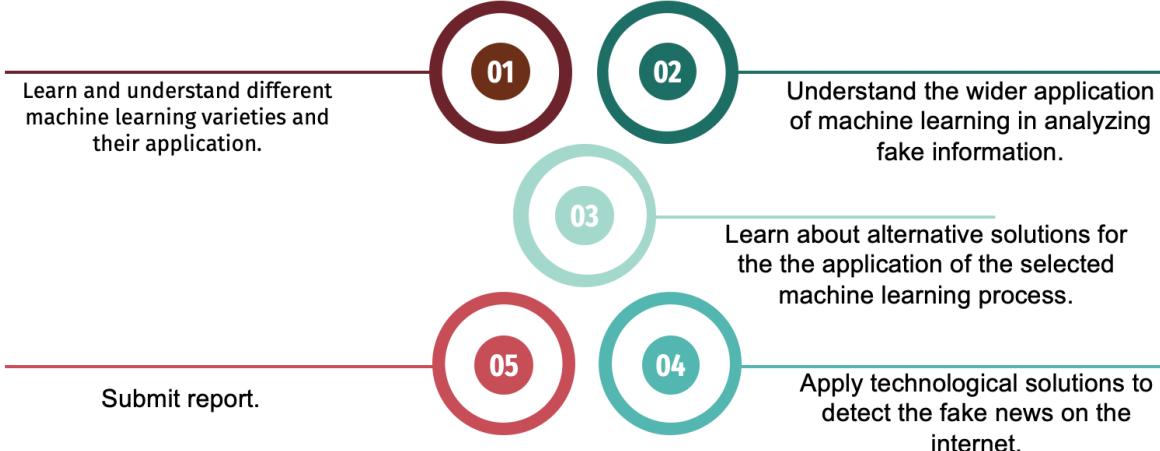


Figure vi: Objectives

Justification

Problem

Fake news is a serious problem in today's society of internet where all age group of people can easily access information. It can deceive and lead people with mis-information. It can even ruin one's life which often happens with celebrity, not only that but it can also flip the country upside down as happens in US election 2016 (Hunt Allcott, 2017). The slight spread of mis information sometimes can create such problem as mentioned above or fill people with false information which may bring them to difficult situations. Since fake news can spread through websites, applications or any other internet sources, users might access to that same information from multiple platform so, it would require a system that can be accessible from multiple platform which is another problem as even in multiple platform it require different system for browsers, software or applications.

Problems

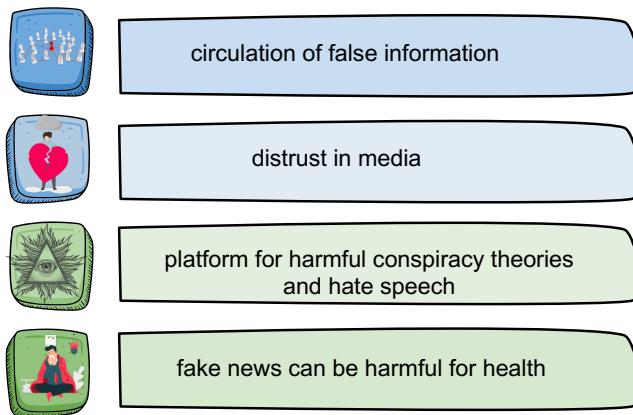


Figure vii: problems of fake news

The false news on the internet also have huge impact on students (Library, 2022). According to studies, a large portion of Americans are unable to distinguish between false and true news. This may lead to misunderstandings and uncertainty concerning significant social and political concerns. Numerous news reports on medical procedures and serious illnesses like cancer or diabetes are false or misleading. If you believe these untrue claims, you could make choices that are bad for your health. According to a Pew Research Center research (Graham, 2017), the meaning of "fake news" varies depending on where one stands politically on the political spectrum. "The Pew study suggests that fake-news panic, rather than motivating people to reject philosophical outlets and the fringe, may actually be speeding up the process of polarization: It is driving consumers to stop using some outlets, to simply consume less information overall, and even to cut out social relationships."

Although people are keenly aware that there is an issue to be solved when talking about false news or disinformation in general, people have not always been clear about the specifics of that problem. Authors have been eager to clarify the key terminology (such as mis- vs. disinformation), but they haven't always been as thorough in deconstructing some of the presumptions made regarding their sources and effects. But not only for building solutions, but also for matching a specific problem with its solution, a better knowledge of the problem(s) is required (Sullivan, 2019).

Solution

There are relatively few answers to the significant problem of false information spreading like wildfire. Humans are blessed with the gift of common sense, which in most cases enables us to assess the news on our own, but it is insufficient to spot every false claim made on the internet. Therefore, we need a system that can detect a false information in the internet. The best solution to it right now is using Machine Learning, collecting the fake news and real news data from the internet and training the system to be able to evaluate the difference between fake and real news.

Solutions

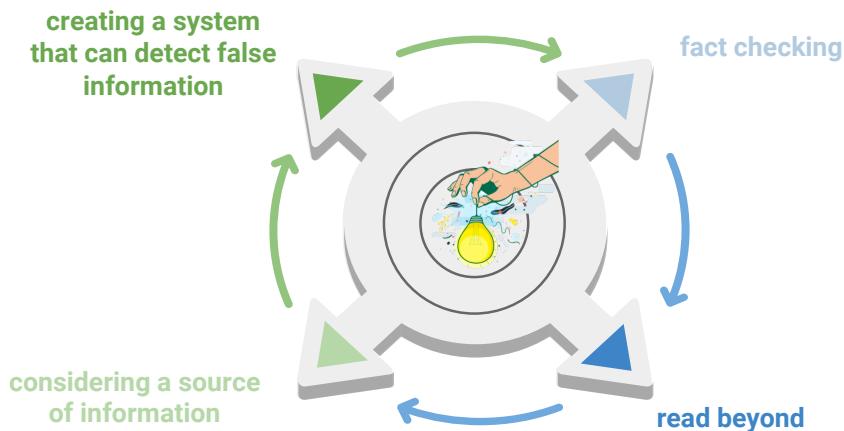


Figure viii: solutions

A large chunk of the dissemination of fake news may be attributed to algorithms since the software that powers social networks can disseminate appealing but misleading articles to new audiences (Wendling, 2017). However, some programmers believe that computer code may also contribute to the answer. According to Claire Wardle of journalism non-profit First (First Draft, 2022), "from an algorithmic standpoint it's conceivable for social networking platforms to detect that website was only launched two weeks ago, thus it's probably likely that this is a less trustworthy site." First Draft is collaborating with Google and Facebook to see if they can add coding to prevent the spread of false information. Wardle insists that modifying algorithms is not the same as censorship.

In brief, the current state of fake news identification significantly drives the need for methods that mine news content in-depth and rely less on how fake news spreads (XINYI ZHOU, 2020). For interpretability considerations, such methods should look into how social and forensic theories might be used to identify bogus news. By creating a theory-driven fake news detection model that focuses on news content, we hope to overcome these issues and be able to identify fake news before it spreads on social media. The model represents news stories using a collection of manual characteristics that, through performing an interdisciplinary research, capture both content structure and style across language levels (i.e., lexicon-level, syntax-level, semantic-level, and discourse-level). Following that, features are used inside a supervised machine learning framework for false news identification.

Research Question

- What kinds of problems may machine learning be used to solve?

- What are the alleged advantages of fake news identification using ML, and how can the ideal solution be put into practice so that you may browse the internet free of false information?
- What ethical issues are raised by the use of fake news detection in machine learning?

Scope

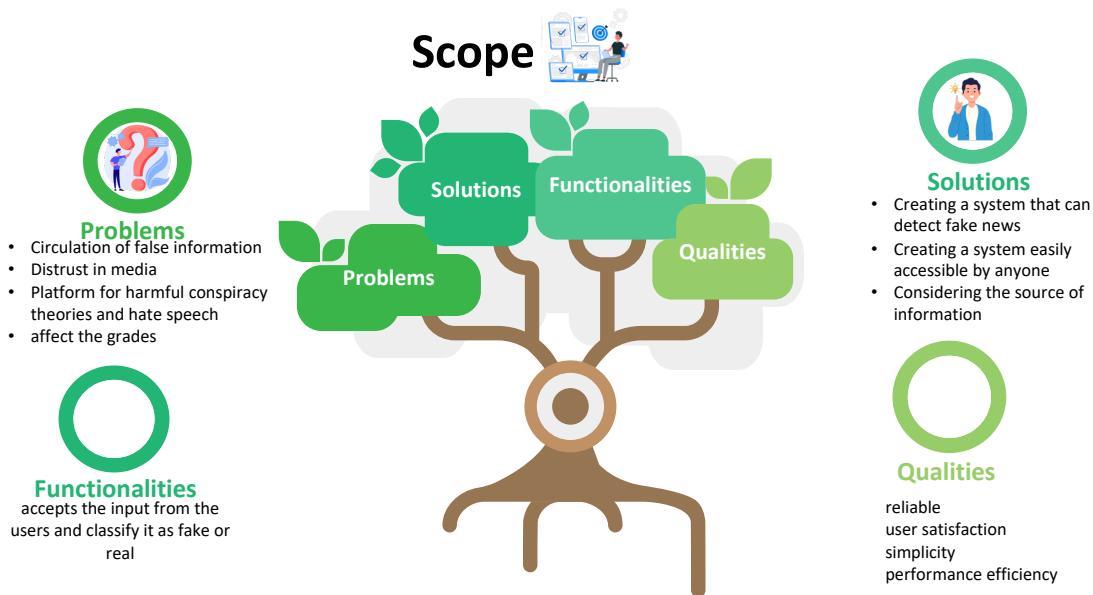


Figure ix: project scope

Ethical Consideration

"A collection of moral standards and ideals that address the issue of what is right and wrong in human affairs. Ethics looks for justifications for doing or refraining from doing something, for endorsing or dis-endorsing behavior, for accepting or rejecting a claim about good or bad laws or virtuous or cruel behavior." - (Australian Government, 2003)

Ethics is a phrase that frequently refers to the study and examination of moral challenges and concepts (University, 2022). Ethics has always been a topic of study for philosophers and religious academics. More lately, academics from many fields have entered the area, developing fresh methods for studying ethics, including behavioral ethics and applied ethics. The term ethics may also be used to describe laws or principles that define what behavior is acceptable for both people and communities. For instance, codes of conduct outline pertinent ethical norms for professionals in a variety of industries, including accountancy, law, journalism, and medicine.



Figure x: Ethical considerations

Ethical considerations can be addressed at individual and at societal levels. Ethics should be thought of in terms of how people are impacted by other people's actions. When a person learns that his father had Huntington's disease and that there is a 50% possibility that he has inherited the genetic mutation, the results can be emotionally and personally damaging. When determining what information to share and how to release it, the risk of harm to that individual becomes a crucial ethical factor. It will be necessary to weigh the danger against the moral imperatives of respecting the affected person's autonomy and right to know. A collection of rules that direct your study designs and procedures are known as ethical considerations in research. When gathering data from individuals, scientists and researchers must always abide by a set of ethical principles. Understanding real-world occurrences, researching efficient therapies, examining behaviors, and enhancing lives in various ways are frequently the objectives of human research. There are important ethical issues in both what you chose to investigate and how you perform that research. These factors contribute to (Bhandari, 2022):

- Defend the rights of participants in research
- Boost the credibility of your research
- Maintain scientific integrity

Literature Review

Incorrect information has been a major problem on the internet recently, which has prompted researchers and scholars to look at a variety of approaches for identifying false information online. Using machine learning algorithms, several studies have been conducted on various subjects. The following are some of the studies I have conducted for the understanding for the project ideas: Machine learning-based false news identification based on content Automatic Fake News Detection, Authors, and Additional Influence from Social Media: Fact-checking software and detection of fake news based on credibility The Kauwa-Kaate System for Detecting Fake News.

Research Methodology

Desk-based

Desk based research mainly involves gathering information from already-existing sources (Travis, 2016); as a result, it is frequently seen as a low-cost strategy when compared to field research, with the primary costs being executive time, telephone costs, and directories. However, if the researcher lacks the necessary understanding of how the study is carried out, it might also be a complete waste of time and money. Desk research is extremely efficient and may be carried out in the first stages of market research since it is quick and affordable, and the majority of the fundamental data can be quickly retrieved and used as a benchmark in the research process.



Figure xi: Desk-based Research

It makes useful to look at what individuals have previously accomplished in relation to the domain of the product before doing a field visit, creating a prototype, conducting a usability test, or starting any project that you want to be user-centered. Although it's improbable that anybody has conducted the specific research activity you're planning, there have almost definitely been attempts to find answers to questions in the same vein. The simplest and least expensive method to comprehend the domain is to read this research.

Content-based fake news detection using machine learning: NLP, Classification & Regression

Automatic Fake News Detection presented two new data sets (Verónica Pérez-Rosas, 2018). The first one was built by crowdsourcing AMT employees using real articles that were pulled from reputable news sources across six key domains. The second one incorporated both true and false celebrity news collected from online tabloids. The characteristics in this study included classifier in the form of punctuation from the LIWC lexicon, psycholinguistics from the LIWC lexicon, readability, and syntax. They also included vector space representations of n-grams using the Term Frequency-Inverse Document Frequency (TF-IDF) metric. Support Vector Machine (SVM) classifier was used to identify. Depending on the dataset utilized, the outcomes were different. Punctuation and Readability were the topic modeling features that produced the best classification results for the first dataset, whereas the whole LWIC lexicon and vector space representation features produced excellent classification results for the second dataset.

Fake News Detection Using Classification & Regression

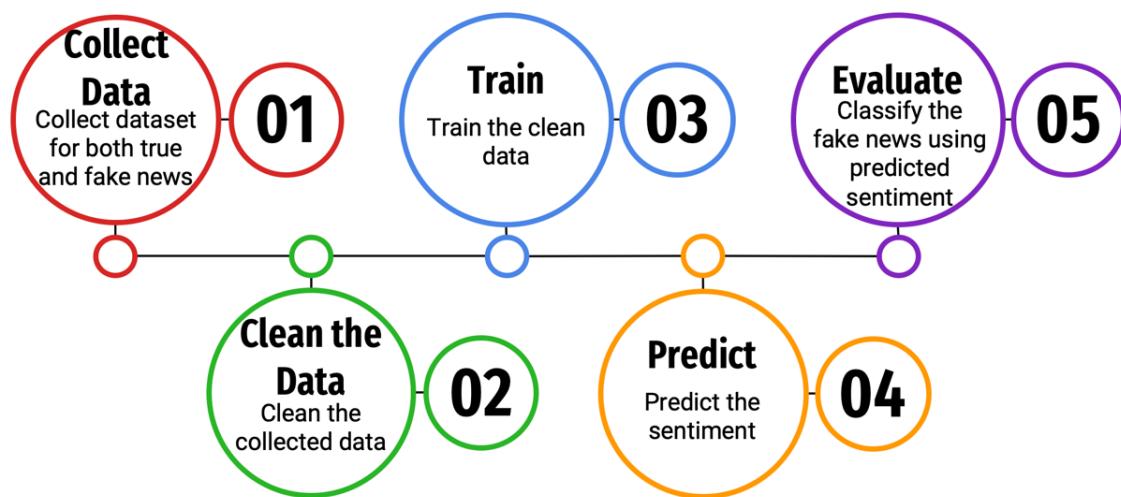


Figure xii: Automatic Fake news Detection Using LIWC

The emotions of the gathered data are also examined using the LIWC lexicon approach. However, the situation with regard to automatic false news identification is different. They gathered their genuine and fake news data for this technique from online tabloids. The ratio of the five emotions in fake news is then analyzed using LIWC, and the same is done for real

news. Now that we have five different emotion ratios, we may compare them with other pieces of content. The system then analyzes the news and classifies it as either being false news or true news by comparing the emotional ratio with the learned data. The best thing about this system is that it keeps training the system with the analyzed news that were not used for training purposes earlier which then rises the accuracy of the system as it gets more data to train on.



Figure xiii: fake news sources

Automatic Fake News Detection is built in python and Js. It uses the machine learning libraries from python like pandas, sklearn, matplotlib, LWIC etc. to develop an algorithm. It was created as a browser extension for the research purpose. For the categorization of news, it makes use of vector space representation features and the LWIC lexicon. To examine the differences in accuracy among various projects, several data and procedures are used. It focuses on a dataset of real and fake celebrity news that was gathered from online tabloids. The same team of individuals that were participating in the research also gathered the data. Even if gathering data independently is the most difficult component of training the algorithm, we don't need to disbelieve the data in order for news categorization accuracy to decline. Identifying fake news Applying Sentiment Analysis, NB, RF classifier trained on LIAR dataset, TF-IDF, sentiments, and cosine similarity scores are used. The model's accuracy, according to them, is increased by the use of emotional score.

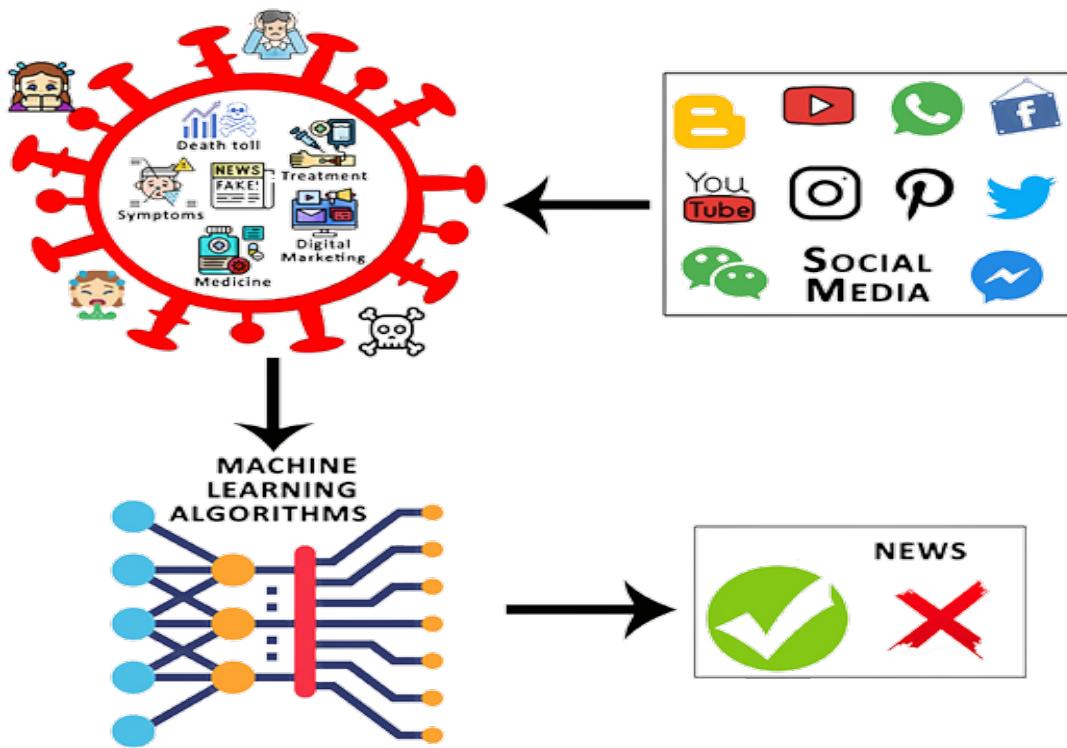


Figure xiv: Fake News Detection in Social Media

Credibility-based the study in Fake News Detection (Niraj Sitaula, 2019) examines how specific source and content factors affect the news's credibility. The greater number of coauthors is a significant signal of legitimacy from the perspective of the source, where anonymous material was typically untrue. The co-authorship graph revealed that false news writers are more likely to collaborate on articles than the opposite is true. The same holds true for the writers' connections to reliable organizations and their past with false information. The stylometry elements listed above, including sentiment analysis, subject knowledge, argumentation, readability, character/word/sentence count, and presence of typos, made up the content-based features. By blending these two reliable sources, they were able to significantly increase the accuracy of the final model with the addition of just three source-based characteristics.

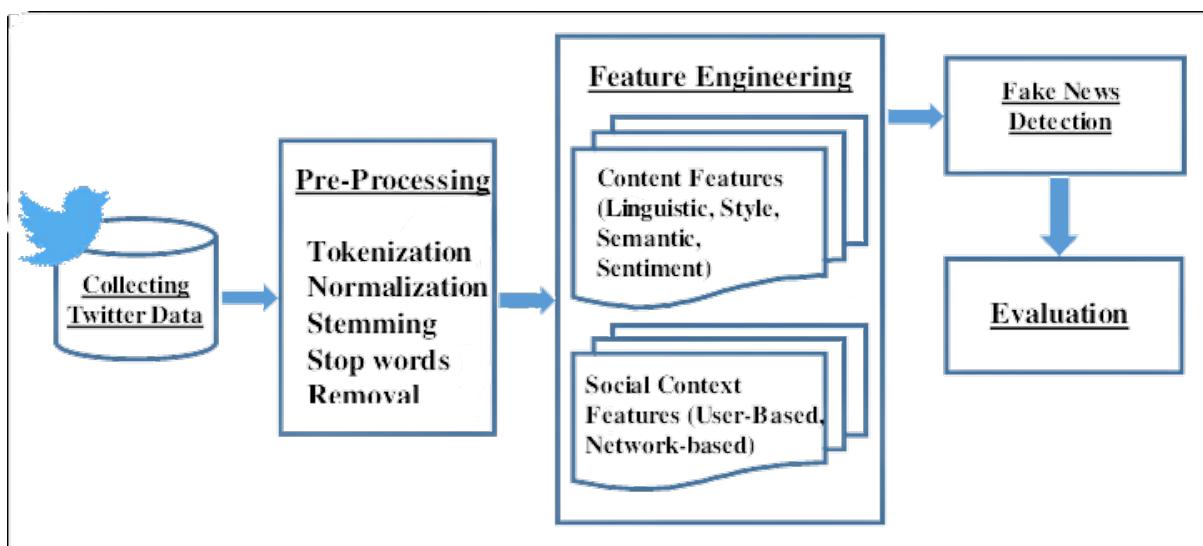


Figure xv: Conceptual Fake News Detection Model

Fake News Detection with Collective User Intelligence is proposed by Neural User Response Generator. The two-level CNN with User Response Generator (TCNN-URG) is used to examine the news' authenticity based on both its content and readers' responses to similar items in the past, and to produce those users' potential responses to the new information. When actual user reactions are unavailable, this strategy is helpful for the early detection of bogus news. In addition to using their own Twitter dataset, they utilize the Weibo dataset. The Conditional Variational Auto encoder is the foundation of the User Response Generator.

The textual data used in CSI: A Hybrid Deep Model for Fake News Detection is combined with user feedback from articles like the one stated above as well as information about the individuals who pushed the news as sources. It consists of three modules: the first one uses an RNN network to process the text and response; the second analyzes user and group information on the reliability of the source; and the third combines these methods. tested on data sets from Weibo and Twitter. For future study, it suggests using reinforcement learning.

The study done focuses on evaluating the websites' overall reputation (Diego Esteves, 2018). It examines the present state of this field's research as well as recent setbacks, such as the price of external APIs and Google PageRank's discontinuation. They exclude user-based social characteristics due to the significant bias they contribute into the final model, which is supported by the ANOVA test, and instead focus on the web credibility model using just content-based features. The resulting model was tested against two databases, the Microsoft Dataset and the Content Credibility Corpus, both of which contained URLs and corresponding Likert 5-star credibility ratings. Readability, PageRank data, General Inquirer (a dictionary similar to LIWC), Vader lexicon (sentiment), Lexical Categories (Lex Rank, LSA), Authority data (address, contact email, etc.), social tags, whether the webpage is open-sourced, and their HTML2seq feature in the form of a bag-of-tags were all content-based features they used (based on BoW). Regression and classification were the two configurations used for the credibility prediction. As a consequence, they put their model to the test on a real-world fact-checking problem and discovered that it was able to distinguish between reputable and unreliable websites based on the assertions made in support of and opposition to each.

The findings showed that the strategy is extremely advantageous since it helps to categorize false news and identifies critical qualities that may be used for accurate fake news identification. Contrary to techniques that employ NLP, content-based false news uses comparable methodologies to categorize the material, making it easier to interpret.

Fake news detection using Machine Learning through NLP (Neuro Language Processing)

The field of computer science known as "natural language processing" (NLP) is more particularly the branch of "artificial intelligence" (AI) concerned with providing computers the capacity to comprehend written and spoken words in a manner similar to that of humans (Education, 2020). NLP integrates statistical, machine learning, and deep learning models with computational linguistics rule-based modeling of human language. With the use of these technologies, computers are now able to interpret human language in the form of text or

audio data and fully "understand" what is being said or written, including the speaker's or writer's intentions and mood.

Fake News Detection Using NLP

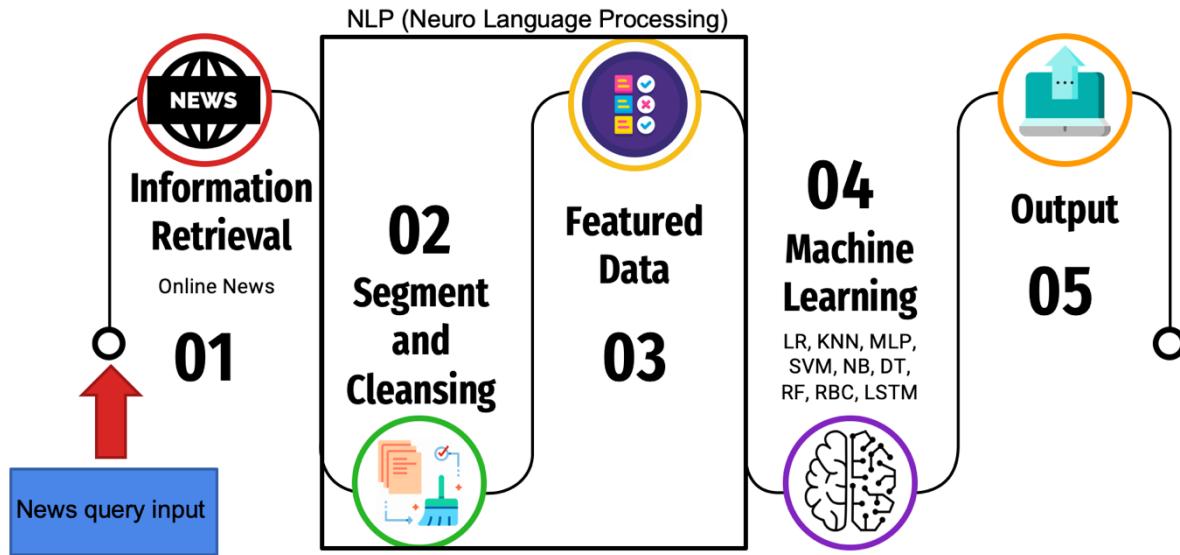


Figure xvi: Fake News Detection Using NLP

In this article (Ray Oshikawa, 2020), they conducted NLP-based research on automated false news identification. They present the general technological issues in false news identification as well as how researchers design various jobs and develop ML solutions to address this issue. They talk about the benefits and downsides of each work as well as any dangers or difficulties that could arise. In further detail, they give a summary of research projects on false news identification and conduct a methodical comparison of their job descriptions, datasets, model building, and results. They also talk about a direction for future study in this area. Social engagement analysis is one of the additional components of this essay.

Researchers in this article (Ray Oshikawa, 2020) presented the first in-depth analysis of Natural Language Processing tools for automatic false news identification. They then reviewed the assumptions and significant concerns for various formulations of the problem before doing a thorough analysis of how fake news detection is linked with current NLP tasks. Then, they provided first-hand knowledge and approachable introductions for new researchers interested in this issue by categorizing and summarizing the datasets, NLP techniques, and outcomes that were available.

The primary objective of this project was to create a machine learning technique to recognize fraudulent or untrustworthy news based on content obtained. To apply machine learning or deep learning algorithms to text data, extra preprocessing is needed. Text data may be transformed using a number of methods that are frequently used to create models-ready formats. In the study (Geetanjali Jain, 2019), stop words were first eliminated from the text data that was accessible. Stop words are more prevalent and retain less relevant information, words (particularly the most widespread terms in a language that do not give much context) can be processed and filtered from the text. Stop words, such as conjunctions like "and," "or," and "but," prepositions like "of," "in," "from," and "to," as well as the articles

"a," "an," and "the," serve more as linking elements in sentences. The removal of stop words as part of data preparation is a crucial initial step in natural language processing since stop words that are of less relevance may consume important processing time. It might be difficult to get the language from the news article's title and body ready for modeling. We must transform unprocessed text into numerical characteristics in order to do text analytics.

True and Fake News

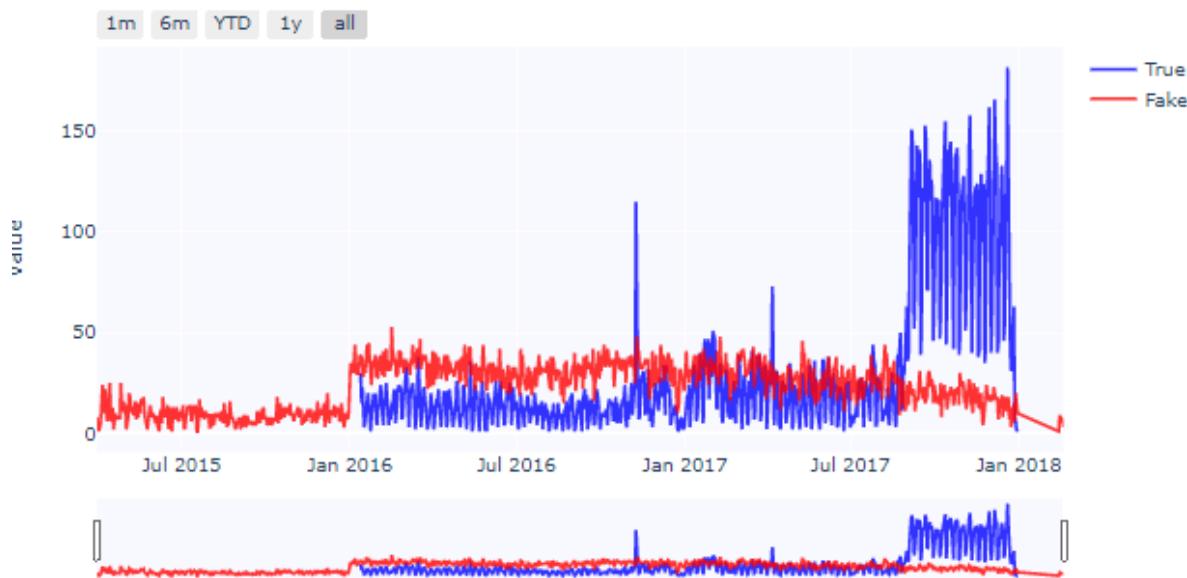


Figure xvii: LTSM detection graph

In order to store the necessary data during the training and processing phases of this project, it makes use of long-short term memory. Hochreiter and Schmidhuber proposed the Long-Short Term Memory (LSTM) unit (Nic Newman, 2015). Because it will selectively memorize the prior input and utilize it, together with the current input, to create predictions, it is effective at categorizing serialized objects. In our issue, the news material (text) is already serialized. The crucial information in a sentence is conveyed by the word order. Therefore, the LSTM model works in the issue they faced during research.

The researchers found the following problems while using NLP in the research. As the input is looped over and over again during RNN (Recurrent Neural Network) training, the weights of the neural network model experience very substantial modifications. This leads in an unstable network as a result of the buildup of error gradients during an update. At their most extreme, weight values have the potential to overflow and produce NaN values. By constantly multiplying gradients via network layers with values greater than 1 or disappearing if the values are less than 1, the explosion is produced by exponential growth. Then the researchers found the mentioned RNN flaw prompted researchers to create and patent a new RNN version known as Long Short-Term Memory. This issue can be resolved by LSTM since it employs gates to regulate the memorization process.

By utilizing LTSM and natural language processing, researchers were able to create a model that can distinguish between true and fake news with great accuracy. The authors advise that fact-checking and linguistic feature analysis be integrated with natural language processing (NLP) to effectively detect false news from accurate news in order to address these weaknesses in fake news (Zhixuan Zhou, 2019). One of the additional advantages of employing NLP is that machine learning text classification, which is used for fake news detection, learns to create classifications based on previous observations rather than depending on human constructed criteria (Monkeylearn, 2022). Machine learning algorithms may understand the many correlations between text fragments and that a specific output (i.e., tags) is anticipated for a specific input by utilizing pre-labeled examples as training data (i.e., text). A "tag" is a pre-established categorization or category that any given text may fit into, making classification easier. Therefore, employing NLP can improve the categorization of fake and true news.

Computational fact-checking using Machine Learning Framework

Fact-checking is used by the Kauwa-Kaate Fake News Detection System: Demo (Abhishek Bagade, 2020) by running queries to fact-checking websites. supports text, picture, and video querying. Although it does not have a completely automatic mechanism for detecting fake news, fact-checkers are supported. When it searches the index of freshly scraped articles from pertinent sources, it employs information retrieval strategy. Using links and tweets to assess the site, they are now attempting to build and employ a bias of the site on a certain domain. They are searching pictures using signature strings from an image matching library. By translating the screenshot to text and querying it as text, they handled it. The smaller scene signatures were utilized to match the videos.

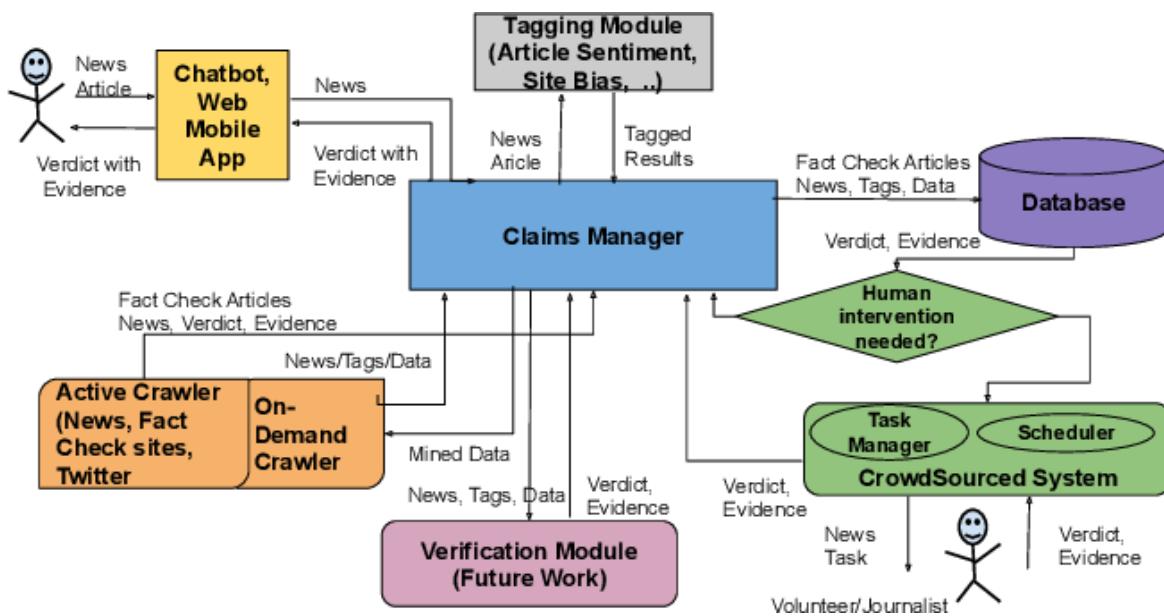


Figure xviii: Kauwa-kaate Fake News Detection System: Demo

Due to the speed at which both true and false information may travel in the current media environment, fact-checking has become more and more crucial (Zhijiang Guo, 2022). In order to automatically anticipate if arguments are real, researchers have been looking into how fact-checking may be automated utilizing methods based on natural language processing,

machine learning, knowledge representation, and databases. The automated fact-checking resulting from natural language processing is surveyed in this work, along with its linkages to related activities and academic fields. To harmonize the many definitions provided and pinpoint shared ideas, researchers offer an overview of the datasets and models that are currently in use in the process. They conclude by highlighting difficulties in further study.

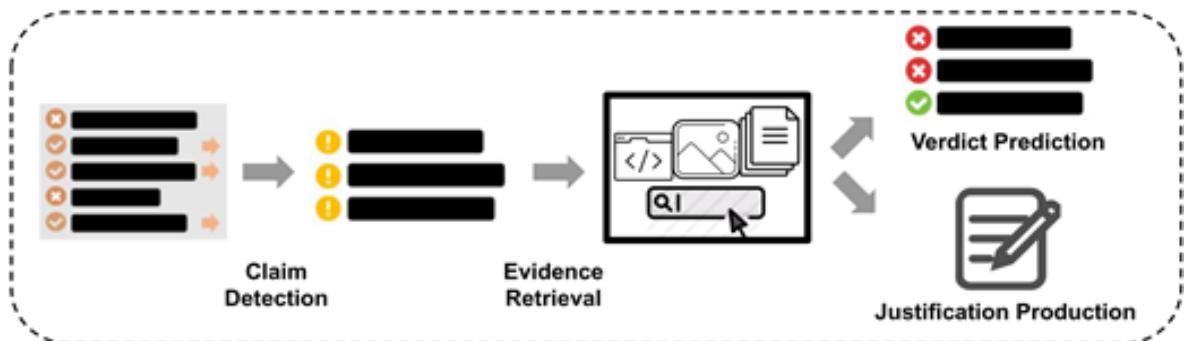


Figure xix: computational fact checking

Numerous articles and software programs have been developed in an effort to automate fact-checking. However, these platforms' inadequacies become clear when used to verify even basic information. They are made to examine particular statements, not a full piece, which is one issue. The second is that they perform badly even when examining a single claim, for a variety of factors, including limits in their capacity to understand plain language. Finally, they typically are unable to provide a strong justification for their choice beyond simply providing web page excerpts, which in our experience does not perform very well. This project consequently holds that, even if it can be helped by computers, fact-checking is now a human activity. The goal of this project is to make it simple for users to determine if a story has been fact-checked or published on a reliable news website.

The crawling, indexing, and utilization of articles from reliable news sources and fact-checking websites for querying purposes after being provided an item to be fact-checked is a crucial component of the architecture. Every three hours, they crawl a variety of well-known fact-checking websites as well as reliable news websites (including the English sites Times of India and The Hindu and the Hindi site NavBharat Times). They scrape the article material using a series of site-specific scrapers because the crawled sites also include superfluous information like adverts and links to other articles in addition to the actual textual content. (They attempted to use generic scrapers but discovered that the quality of their output was not very acceptable.) Their scrapers also pull meta-data from the web pages that are often provided by news websites. Then, they use Solr to index the extracted text from the crawled articles and their meta-data.

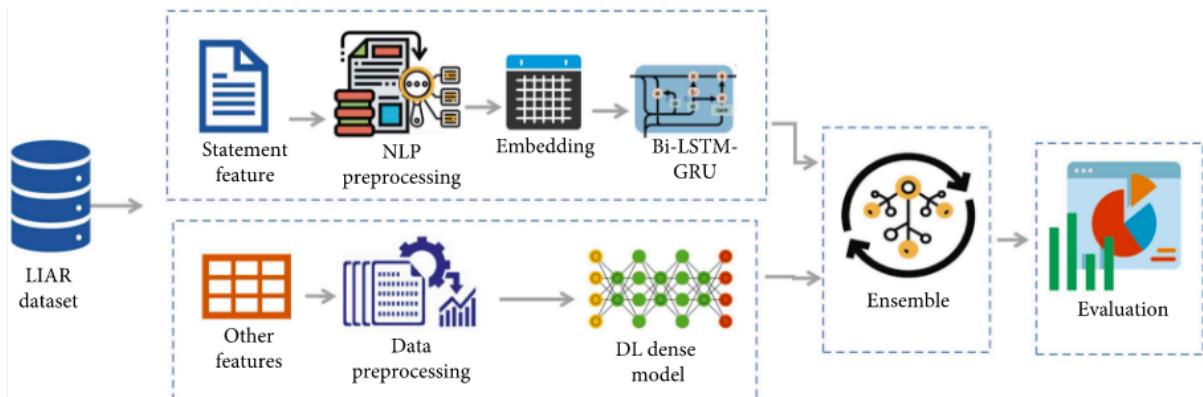


Figure xx: fake news detection using LIAR dataset

The researcher then executed a query on the index using the complete body of an item that has to be fact-checked; by contrast, commercial search engines just utilize the first 32 or so words of the article, which they discovered frequently returns irrelevant content. A photograph from context A may be mistakenly stated to be from context B, but more crucially, fake news is frequently defined by the combination of an image and text. Thus, to determine whether an article is phony, we must search for a text and image combo.

By querying the knowledge graphs built for news stories from knowledge base DBpedia, it is possible to assess the credibility of the news based on the reliability of the content itself. One of the four techniques listed in the survey and used in the aforementioned project was this one. It's thought of as an automatic fact-checking method. The system's uses various pipeline functions in the following ways (Abhishek Bagade, 2020). Basic validation tests are made on the retrieved data by the validation pipeline, and blank articles are removed. Statistics Pipeline gathers data from the crawler run, including the number of articles crawled, and adds it to a database. The crawled data is dumped in JSON files using the WritetoFile Pipeline. The files are arranged according to the site name and the themes they fall under. All of the photos in the crawled articles are downloaded by the ImageInsert Pipeline, which then indexes them and adds information to the Elastic search back-end. The crawled article is indexed in Apache Solr via the IndexSolr Pipeline, which we employ for quick retrieval and search operations. OCR Pipeline employs the Tesseract OCR engine to extract text from pictures crawled in both the English and Hindi languages, while NLU Pipeline uses the IBM Watson NLU API to extract and store different metadata about the article body. Video Pipeline collects signatures from the videos featured in the articles and indexes the signatures in Solr for search and retrieval.

It's possible to locate a story on a website they do not trust if the researcher can't find it on a reliable news source or fact-checking website. However, deciding whether to trust a website might be a little difficult. They thought a better strategy is to determine the author's bias using a range of data, such as links to the author's website and tweets that link to the website or post as well as the website's content. In order to assist users in determining the validity of a certain article, then they presently investigated a number of possibilities for determining the site's bias on a particular issue. The front-end and REST API were used to develop the web

app system which accepts the queries, images or videos along with the text. They provide REST API access to their system, allowing users to individually query the picture and text systems. The API combines those data after that. They have a basic Web client that responds to user requests as a proof of concept.

The primary advantage of computational fact checking is that the majority of the frameworks offer the previously trained modal, allowing us to transmit our data directly to the evaluation process, halving the amount of code and simplifying the process. Even while fact checking requires more time, the results are more accurate. It makes use of libraries that the creators have previously trained, and it instructs the system based on the knowledge that other users have used the library for computational fact-checking. As a result, it has a greater possibility of producing results with high accuracy when there is a vast diversity of information being provided each time library is used by other users.

Development Methodology

Agile

Agile is a project management methodology that emphasizes incremental and iterative project completion phases (Coursera, 2022). Short-term development cycles are used to carry out a project's incremental components. Instead of top-down administration and adhering to a predetermined plan, the strategy places a higher priority on speedy delivery, flexibility, and cooperation. Agile procedures provide continuous feedback, giving stakeholders the chance to communicate regularly and team members the chance to adapt to problems as they emerge. Although the Agile method was initially developed for software development, it is now extensively employed in carrying out several different kinds of projects and in managing companies. Since Agile first appeared two decades ago, it has gained enormous popularity, with many businesses striving to "follow Agile" or starting "Agile transformations" to help them produce better products more quickly.



Figure xxi: agile

Why Agile?

We have adopted agile in this project rather than following other conventional approaches as it was found most suitable for our project. Technical burden is reduced to a minimum with the aid of agile software development. Any flaws, feature modifications, or other upkeep activities are put to the product backlog. Agile is also simple to modify. The team does not try to build every feature at once on an Agile project. Instead, the team divides up the features into smaller subsets for each sprint, which aids in the creation of a high-quality product. Following agile methods for the development processes has become a better option thanks to the involvement of several firms and the development of new development techniques based on agile principles.

Why Agile?

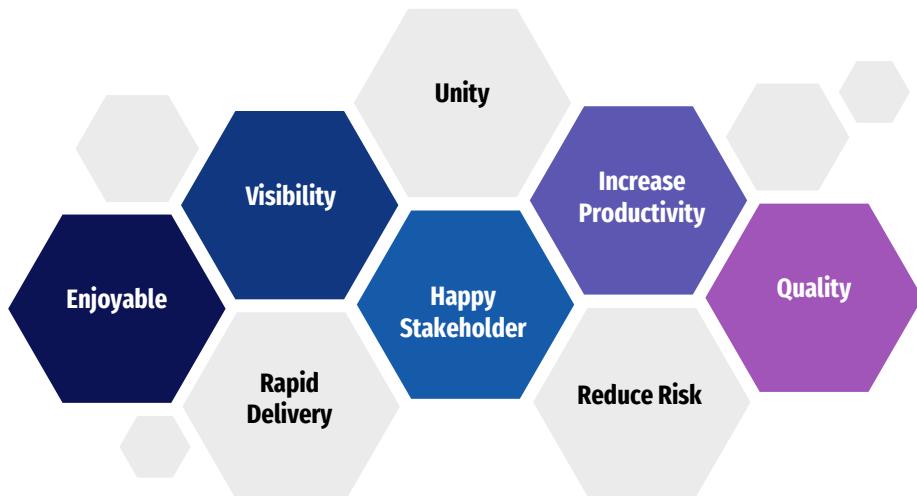


Figure xxii: why Agile?

Agile methodology facilitates the quick delivery of functional software. It enables rapid response to change, protects companies and projects from a rollback, and so aids in project stabilization. Agile adheres to the "fail fast, fail safe" philosophy. The sooner you make a mistake, the sooner you can learn from it and move on. Agile emphasizes quick sprints or iterations, which speed up the production of new products or features. Agile does not treat testing as an afterthought. Agile combines testing concurrently with development, which helps to improve project quality, in contrast to traditional project management, which only does test once the development phase is complete. Agile teams are small because they have high levels of cooperation and self-organization, are cross-functional, and are transparent. The responsibility of each team member boosts morale.

Tools & Technology

In order to complete the project, a variety of tools and technologies were used. We have used easiest tools and technologies to carry out the research so we could conduct the research in the given timeframe with the least problems possible. A list of these tools is given below:

Tools Used in Project

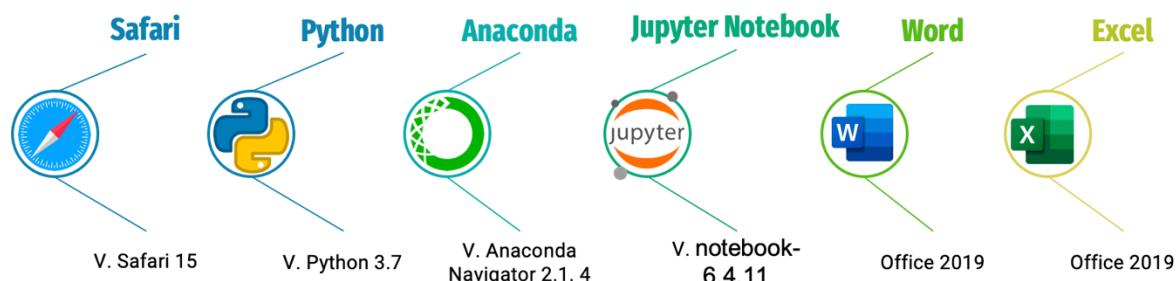


Figure xxiii: Tools used in project

Safari: Developed by Apple, Safari is a graphical web browser. It is built mostly on open-source code, particularly Web Kit. Because we were conducting our research on an Apple device, we utilized the Safari browser to run the Jupyter notebook and continue the research process further.

Python: Python is a strong, flexible, and all-purpose programming language. It makes for a fantastic programming language because it is clear and simple to read. Python has the ability to perform any task, from data science to machine learning to web development. We employed Python to build an algorithm in our project that can recognize fake news and learn from it.

Anaconda: Designed to make package management and deployment easier, Anaconda is a distribution of the Python and R computer languages for scientific computing. It has been applied in our research to build the environment needed for our research endeavor.

Jupyter Notebook: The objective of Project Jupyter is to provide open-source software, open standards, and interactive computing services for a variety of programming languages. Fernando Pérez and Brian Granger separated it from IPython in 2014. Python code for our project is written in a Jupyter notebook.

Word: Developed by Microsoft, Microsoft Word is a word processing program. It is one of the most widely used pieces of software for creating documents. Word was used to create this dissertation for our assignment.

Excel: Microsoft produced Microsoft Excel, a spreadsheet, for Windows, macOS, Android, and iOS. It has calculating or computing capabilities, graphing tools, pivot tables, and the Visual Basic for Applications macro programming language. This has been utilized in our project to store and utilize the datasets that were gathered.



Figure xxiv: technologies used in research

Machine Learning: Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. In anticipation of new output values, machine learning algorithms use past data as input. In order to accurately classify the news in our project, we employ machine learning to train the algorithm using both actual and fraudulent news.

Classification: Classification is the process of grouping a set of data into categories. It may be applied to both structured and unstructured data. Predicting the class of the provided data points is the first step in the procedure. The terms target, label, and classes are frequently used to describe the classes.

Regression: Regression is a method for determining how independent traits or variables relate to a dependent feature or result. It is a technique for machine learning predictive modeling, where an algorithm is used to forecast continuous outcomes.

Findings

What kinds of problems may machine learning be used to solve?

Even if machine learning can't be used for everything, it can still help with a lot of issues. It can be used to address extremely challenging issues like detecting credit card fraud, recognizing and identifying faces, and even enabling self-driving automobiles! Machine learning may assist IT teams in gaining access to the value concealed in vast quantities of operational data, speeding up the identification and diagnosis of problems for today's IT Big Data difficulties (KALUZA, 2022). Businesses may now automatically analyze massive amounts of data in previously unachievable ways to strengthen support for essential business services, optimize IT operations, and avoid outages.

What are the alleged advantages of fake news identification using ML, and how can the ideal solution be put into practice so that you may browse the internet free of false information?

People find it difficult to determine if any given piece of information they find online is true or false given today's circumstances. Therefore, fake news detection systems themselves are advantageous for people because they may serve as barriers that prevent them from being exposed to misleading information. The system may be integrated into blogs, news sites, or social media platforms where numerous writers can post to warn them against publishing misleading material. It can also be utilized in the form of browser extensions to allow users to focus on certain information as they see fit.

What ethical issues are raised by the use of fake news detection in machine learning?

Ethics is a phrase that frequently refers to the study and examination of moral challenges and concepts (University, 2022). Ethics has always been a topic of study for philosophers and religious academics. More lately, academics from many fields have entered the area, developing fresh methods for studying ethics, including behavioral ethics and applied ethics. We cannot claim that the algorithm we have developed is 100% effective, therefore labeling accurate material as false might have an adverse effect on the author. Given that it may end up being the cause of the issue, the ethics in question here may provide a bigger issue for both sides.

Future Work

This project was begun primarily with the goal of developing a quick system that could identify fake news based on user input and it was created within limited time period. As a result, there are several features and functions in this project that can be improve and added. Relating to this, future studies will concentrate on a thorough analysis of the procedures for adding new development approaches into practice. The following are some of the recognized enhancements that may be applied to the false news detection:

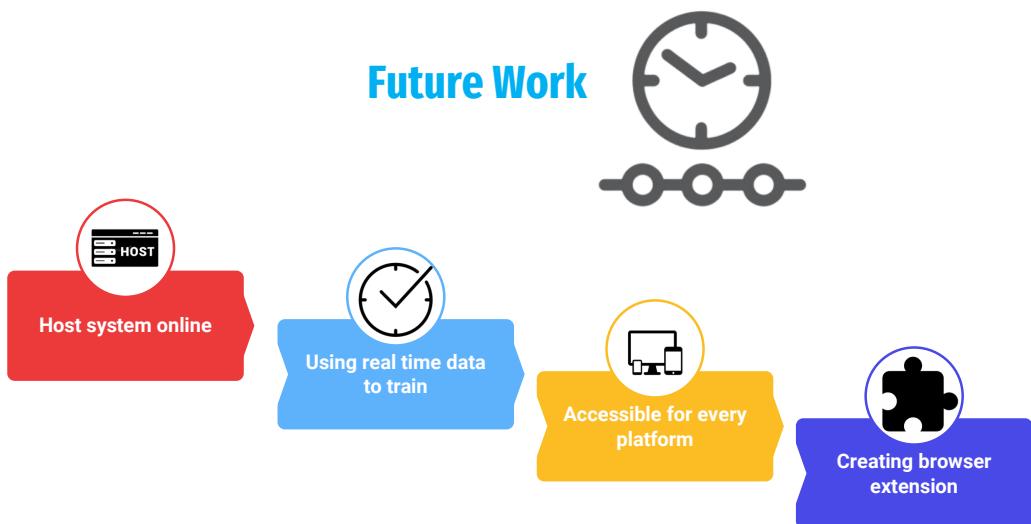


Figure xxv: future work

Can be developed as extension in browsers:

Users may install it on their browsers and receive notifications when they encounter incorrect information on the pages they are browsing if it is made available as an extension. In this way it can be used in real time detection of fake news.

Using real time data to train:

We can make our system to train on real time with the data users are using to evaluate our system. It will be the best ideas as our system will be using different types of data and information of the current trend. It will increase the accuracy of the system detection which will be the plus point for us as well as the users.

Using NLP to get better accuracy in classification:

Along with the classification and regression NLP (Natural Language Processing) can be used to get better accuracy in the evaluation of the fake news. NLP can be used here to understand how people are using their language to post the fake news online.

Accessible in proper user interface instead of jupyter notebook:

Currently, even to get user input we are using jupyter notebook. So, instead we can use web page or use libraries like streamlit or flask to create easy interface. It is very important part as people unaware of these tools and technologies should easily be able to use the system.

Hosting the System Online:

System can be hosted online so the people from all over the world can use the system which would also make it easier to get feedbacks from the people around the world.

Accessible for every platform:

Today with the advancement of the technology almost every people are holding computing devices. People uses laptops, mobile phones or computers as per their situations and requirements. We can made available for every platform, so people using every platform can easily use the system for their benefits.

Project and Issue Management

Project Gantt Chart

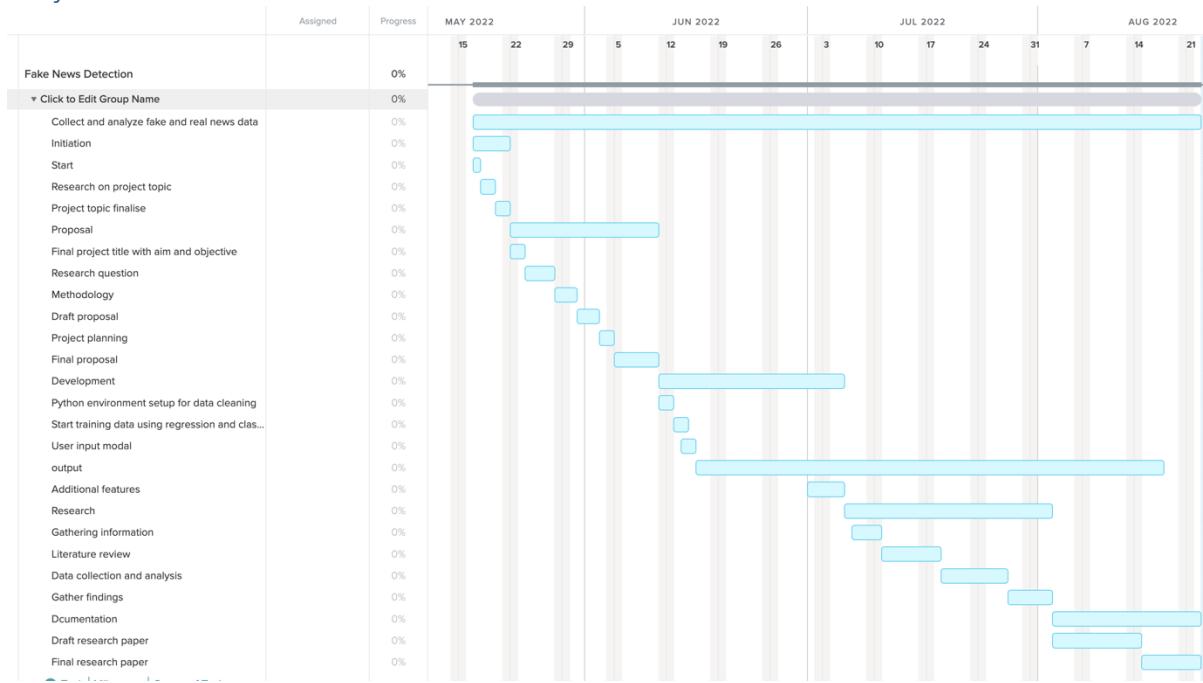


Figure xxvi: Gantt chart

Issue Management

While working on this project, we ran into a few issues. The problems will inevitably arise when we are conducting the research study. Even though we had several problems, we were nevertheless able to finish our assignment.

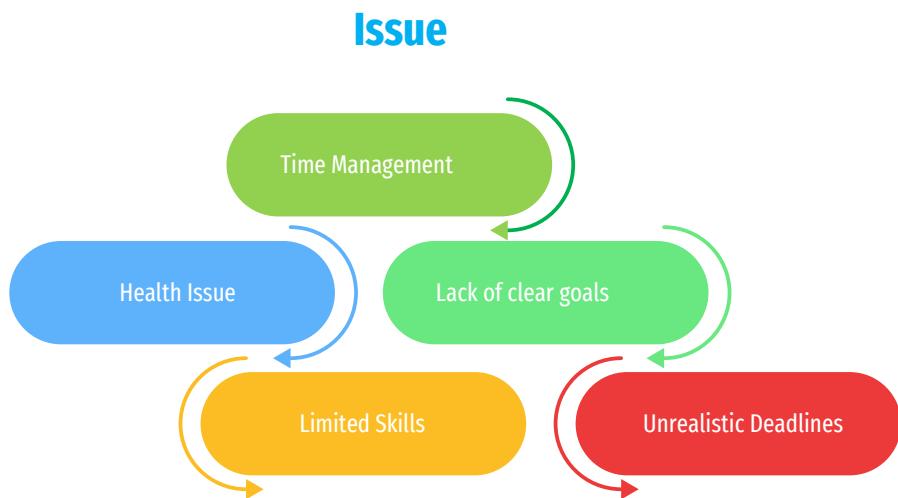


Figure xxvii: issues

Time management was a bigger problem because we couldn't follow the set plan and several other problems made the time go by more slowly. The project was delayed by a week mostly due to health issues. Tensorflow was a challenge for us as well. A free and open-source software library for artificial intelligence and machine learning is called TensorFlow. Although it may be used to many different tasks, deep neural network training and inference are given special attention. Since the software was run on a Mac, the most recent version of TensorFlow was successfully installed, however there was a problem while training the data.

Conclusion

In this project, we'll look into how machine learning's classification and regression techniques may be utilized to spot fake news. To conduct research on multiple libraries while sticking to the notions of the research questions, several machine learning techniques will be applied. We will be performing research while adhering to the project strategy and taking into account any external circumstances that may have an impact on the project. Finally, to complete the research we will be following the procedure, plan, and concepts stated above, we will be undertaking research on developing a false news detection system utilizing classification and regression in machine learning.

References

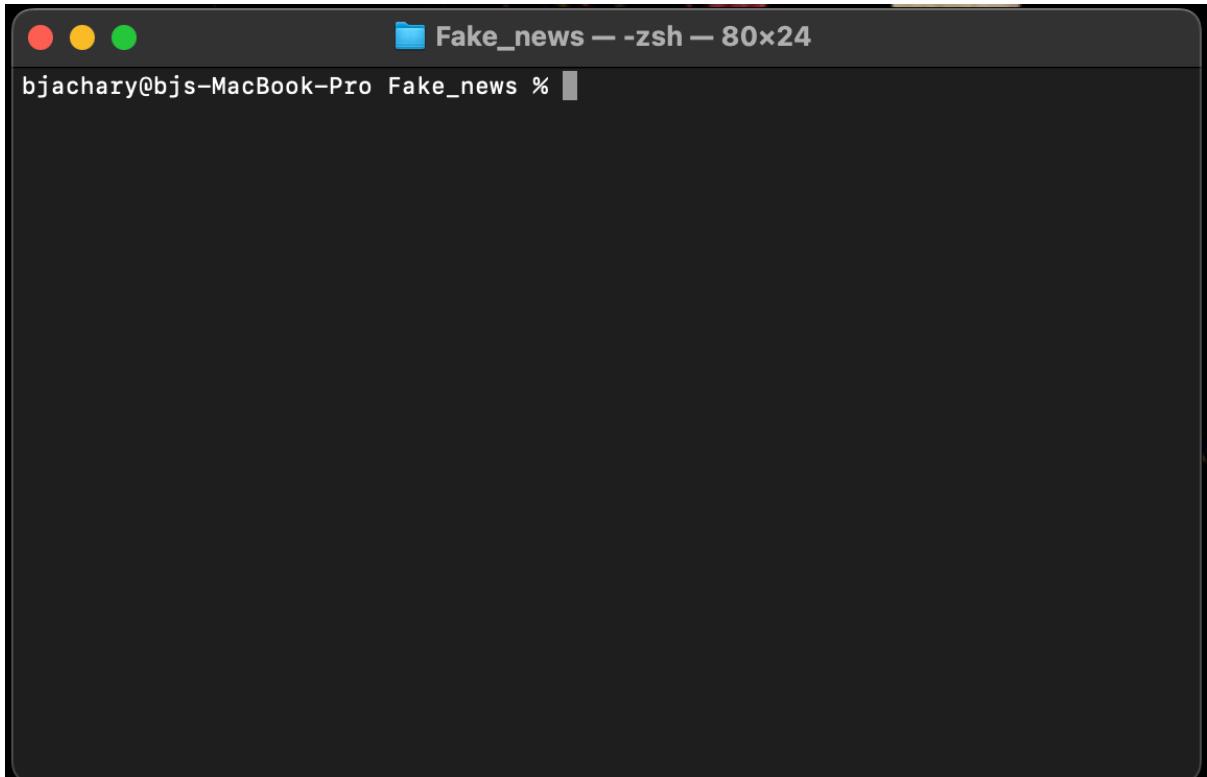
- Abhishek Bagade, A. P. S. S. M. A. S. C. K. C. S. S. O., 2020. *The Kauwa-Kaate Fake News Detection System: Demo*. [Online]
Available at: <https://researchr.org/publication/BagadePSACC020/bibliographies>
- Australian Government, A. C., 2003. *Protection of human genetic information*. [Online]
Available at: <https://www.alrc.gov.au/inquiry/protection-of-human-genetic-information/>
- Bhandari, P., 2022. *Ethical Considerations in Research*. [Online]
Available at: <https://www.scribbr.co.uk/research-methods/ethical-considerations/>
- Brownlee, J., 2021. *Neural Network Models for Combined Classification and Regression*. [Online]

- Available at: <https://machinelearningmastery.com/neural-network-models-for-combined-classification-and-regression/>
- Brownlee, J., 2022. *What is Machine Learning? Machine Learning For Beginners*. [Online] Available at: <https://www.edureka.co/blog/what-is-machine-learning/>
- Cadwalladr, C., 2017. *The great British Brexit robbery: how our democracy was hijacked*. [Online] Available at: <https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robery-hijacked-democracy>
- Cisco, 2020. *Annual Internet Report*. [Online] Available at: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-/index.html>
- Diego Esteves, A. J. R. P. C. J. L., 2018. *Belittling the Source: Trustworthiness Indicators to Obfuscate Fake News on the Web*. [Online] Available at: <https://arxiv.org/abs/1809.00494>
- Dixon, S., 2021. *Social media - Statistics & Facts*. [Online] Available at: <https://www.statista.com/topics/1164/social-networks/>
- Education, I. C., 2020. *Machine Learning*. [Online] Available at: <https://www.ibm.com/cloud/learn/machine-learning>
- Education, I. C., 2020. *Natural Language Processing (NLP)*. [Online] Available at: <https://www.ibm.com/cloud/learn/natural-language-processing>
- First Draft, C. W., 2022. *First Draft*. [Online] Available at: <https://firstdraftnews.org/>
- Fleming, N., 2020. *Coronavirus misinformation, and how scientists can help to fight it*. [Online] Available at: <https://www.nature.com/articles/d41586-020-01834-3>
- FRSA, A. K., 2020. *Internet Matters*. [Online] Available at: <https://www.internetmatters.org/hub/news-blogs/tackling-fake-news-during-the-covid-19-pandemic-and-beyond/> [Accessed 03 07 2022].
- Gallo, A., 2015. *A Refresher on Regression Analysis*. [Online] Available at: <https://hbr.org/2015/11/a-refresher-on-regression-analysis>
- Geetanjali Jain, A. M., 2019. *Natural language Processing Based Fake News Detection using Text Content Analysis with LSTM*. [Online] Available at: <https://ijarcce.com/wp-content/uploads/2019/12/IJARCCE.2019.81119.pdf>
- Graham, D. A., 2017. *Some Real News About Fake News*. [Online] Available at: <https://www.theatlantic.com/ideas/archive/2019/06/fake-news-republicans-democrats/591211/>
- Hannah Rashkin, E. C. J. Y. J. S. V. Y. C., 2017. *Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking*. [Online] Available at: <https://aclanthology.org/D17-1317/>
- Hunt Allcott, M. G., 2017. *Stanford*. [Online] Available at: <https://web.stanford.edu/~gentzkow/research/fakenews.pdf>
- KALUZA, D. B., 2022. *What Kind Of Problems Can Machine Learning Solve?*. [Online] Available at: <https://www.evolven.com/blog/what-kind-of-problems-can-machine-learning-solve.html> [Accessed 17 08 2022].

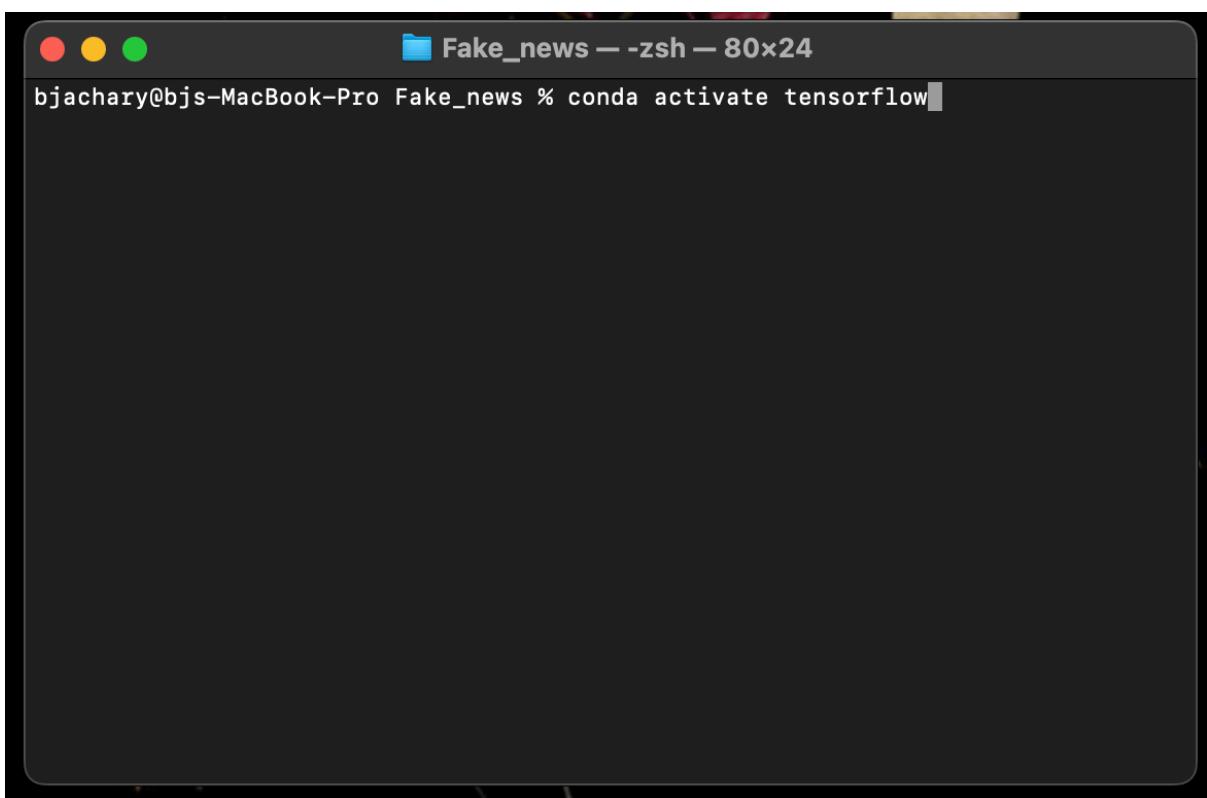
- Lawrence Alexander, C. S., 2016. *How Teens In The Balkans Are Duping Trump Supporters With Fake News*. [Online]
Available at: <https://www.buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo>
- Library, A., 2022. *Fake News and Alternative Facts: Finding Accurate News: Why is Fake News Harmful?*. [Online]
Available at: <https://researchguides.austincc.edu/c.php?g=612891&p=4258046>
- Makse, A. B. & H. A., 2019. *Influence of fake news in Twitter during the 2016 US presidential election*. [Online]
Available at: <https://www.nature.com/articles/s41467-018-07761-2>
- Monkeylearn, 2022. *Text Classification: What it is And Why it Matters*. [Online]
Available at: <https://monkeylearn.com/text-classification/>
[Accessed 15 08 2022].
- Ndapandula Nakashole, T. M. M., 2014. *Language-Aware Truth Assessment of Fact Candidates*. [Online]
Available at: https://www.researchgate.net/publication/270878560_Language-Aware_Truth_Assessment_of_Fact_Candidates
- Nic Newman, D. L. R. K. N., 2015. *Reuters Institute Digital News Report 2015*. [Online]
Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2619576
- Niraj Sitaula, C. K. M. J. G. X. Z. R. Z., 2019. *Credibility-based Fake News Detection*. [Online]
Available at: <https://arxiv.org/abs/1911.00643>
- Persia, M. G. a. F., 2017. *A Survey of Multimedia Recommender Systems: Challenges and Opportunities*. [Online]
Available at: <https://doi.org/10.1142/S1793351X17500039>
- Ray Oshikawa, J. Q. W. Y. W., 2020. *A Survey on Natural Language Processing for Fake News Detection*. [Online]
Available at: <https://aclanthology.org/2020.lrec-1.747/>
- Ray Oshikawa, J. Q. W. Y. W., 2020. *A Survey on Natural Language Processing for Fake News Detection*. [Online]
Available at: <https://arxiv.org/pdf/1811.00770.pdf>
- Rogers, R., 2020. *Research Gate*. [Online]
Available at: https://www.researchgate.net/figure/Fake-news-outperforms-mainstream-news-in-the-months-prior-to-the-2016-US-presidential_fig1_345007859
[Accessed 03 07 2022].
- Rowan Zellers, A. H. H. R. Y. B. A. F. F. R. Y. C., 2019. *Defending Against Neural Fake News*. [Online]
Available at: <https://arxiv.org/abs/1905.12616>
- Silverman, C., 2016. *This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook*. [Online]
Available at: <https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook>
- Sullivan, 2019. <https://dash.harvard.edu/handle/1/42517743>. [Online]
Available at: <https://dash.harvard.edu/handle/1/42517743>
- Tavernise, S., 2016. *Fake News Partisan Republican Democrat*. [Online]
Available at: <https://www.nytimes.com/2016/12/06/us/fake-news-partisan-republican-democrat.html>

- Travis, D., 2016. *Desk research: the what, why and how*. [Online]
Available at: <https://www.userfocus.co.uk/articles/desk-research-the-what-why-and-how.html>
- University, C., 2022. *Fake News*. [Online]
Available at: <https://dictionary.cambridge.org/dictionary/english/fake-news>
- University, T., 2022. *Ethics*. [Online]
Available at: <https://ethicsunwrapped.utexas.edu/glossary/ethics>
- Verónica Pérez-Rosas, B. K. A. L. R. M., 2018. *Automatic Detection of Fake News*. [Online]
Available at: <https://aclanthology.org/C18-1287/>
- Victoria, U. o., 2022. *Fake News*. [Online]
Available at: <https://libguides.uvic.ca/fakenews/consequences>
- Wendling, M., 2017. *Solutions that can stop fake news spreading*. [Online]
Available at: <https://www.bbc.com/news/blogs-trending-38769996>
- XINYI ZHOU, A. J. V. V. P. a. R. Z., 2020. *Fake News Early Detection: An Interdisciplinary Study*. [Online]
Available at: <https://arxiv.org/pdf/1904.11679.pdf>
- Zhijiang Guo, M. S. A. V., 2022. *A Survey on Automated Fact-Checking*. [Online]
Available at: https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00454/109469/A-Survey-on-Automated-Fact-Checking
[Accessed 16 08 2022].
- Zhixuan Zhou, H. G. M. M. B. J. H., 2019. *Fake News Detection via NLP is Vulnerable to Adversarial Attacks*. [Online]
Available at: <https://arxiv.org/abs/1901.09657>
[Accessed 12 08 2022].

Appendix



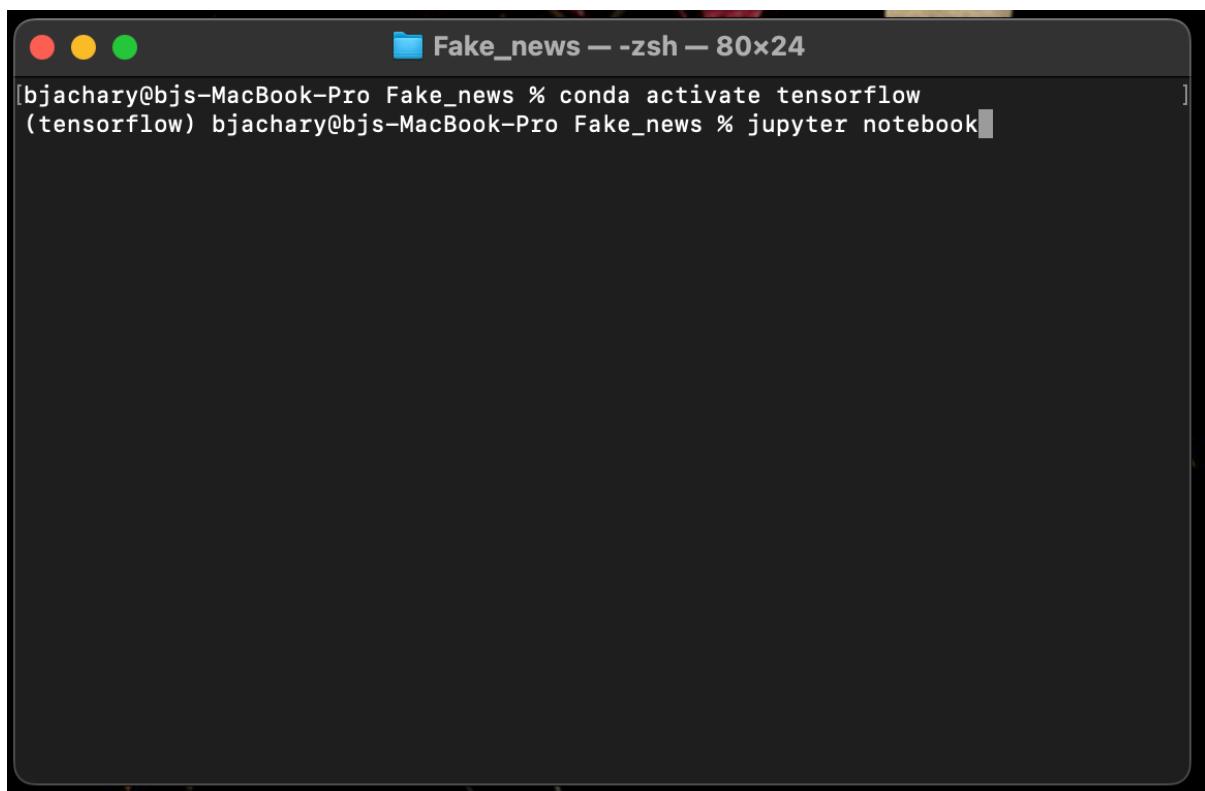
A screenshot of a macOS terminal window titled "Fake_news — -zsh — 80x24". The window has three colored window control buttons (red, yellow, green) at the top left. The title bar shows the folder icon and the text "Fake_news — -zsh — 80x24". The main pane is dark gray and contains the command prompt "bjachary@bjjs-MacBook-Pro Fake_news %". There is no output or text displayed below the prompt.



A screenshot of a macOS terminal window titled "Fake_news — -zsh — 80x24". The window has three colored window control buttons (red, yellow, green) at the top left. The title bar shows the folder icon and the text "Fake_news — -zsh — 80x24". The main pane is dark gray and contains the command prompt "bjachary@bjjs-MacBook-Pro Fake_news % conda activate tensorflow". The text "conda activate tensorflow" is partially visible in the bottom right corner of the terminal window.



```
[bjachary@bjjs-MacBook-Pro Fake_news % conda activate tensorflow  
(tensorflow) bjachary@bjjs-MacBook-Pro Fake_news % ]
```



```
[bjachary@bjjs-MacBook-Pro Fake_news % conda activate tensorflow  
(tensorflow) bjachary@bjjs-MacBook-Pro Fake_news % jupyter notebook ]
```

thesis

Page 1 of 40 8277 words English (United States)

True

A1	B1	C1	D1	E1	F1	G1	H1	I1	J1	K1	L1	M1	N1	O1	P1	Q1	R1	S1	T1	U1	V1	W1	X1	Y1	Z1	AA1	AB1	AC1	AD1	AE1	AF1	AG1																																
1 U.S. to India: WASHINGTONpoliticsNews December 13, 2017	2 U.S. military WASHINGTONpoliticsNews December 13, 2017	3 Trump says he'll meet with Kim Jong Un WASHINGTONpoliticsNews December 13, 2017	4 Senate passes \$1.3 trillion spending bill WASHINGTONpoliticsNews December 13, 2017	5 FBI director pi WASHINGTONpoliticsNews December 13, 2017	6 Trump wants SEATTLE/WI politicsNews December 13, 2017	7 Ukraine crisis: WASHINGTONpoliticsNews December 13, 2017	8 Trump says pi WASHINGTONpoliticsNews December 13, 2017	9 Father of Tru The following politicsNews December 13, 2017	10 Trump signs \$1.3 trillion spending bill WASHINGTONpoliticsNews December 13, 2017	11 Alabama off pi WASHINGTONpoliticsNews December 13, 2017	12 Amazon to open second headquarters in New York City WASHINGTONpoliticsNews December 13, 2017	13 New York pi NEW YORKpoliticsNews December 13, 2017	14 Father of Tru The following politicsNews December 13, 2017	15 Trump signs \$1.3 trillion spending bill WASHINGTONpoliticsNews December 13, 2017	16 Man says he's in Dec. 25 politicsNews December 13, 2017	17 Trump's office (Reuters) - A politicsNews December 13, 2017	18 U.S. to India: WASHINGTONpoliticsNews December 13, 2017	19 Trump on pi The following politicsNews December 13, 2017	20 U.S. to India: WASHINGTONpoliticsNews December 13, 2017	21 Treasury says (Reuters) - A politicsNews December 13, 2017	22 Federal judge WASHINGTONpoliticsNews December 13, 2017	23 Father of Tru The following politicsNews December 13, 2017	24 Trump travel (Reuters) - A politicsNews December 13, 2017	25 Second term: WASHINGTONpoliticsNews December 13, 2017	26 Failed vote: UMA (Reuter) politicsNews December 13, 2017	27 Trump signs WASHINGTONpoliticsNews December 13, 2017	28 Congress vs WASHINGTONpoliticsNews December 13, 2017	29 Trump signs \$1.3 trillion spending bill WASHINGTONpoliticsNews December 13, 2017	30 In victory for NEW YORKpi politicsNews December 13, 2017	31 Senate passes \$1.3 trillion spending bill WASHINGTONpoliticsNews December 13, 2017	32 Trump on pi The following politicsNews December 13, 2017	33 Mexico to pi MEXICO/GT politicsNews December 13, 2017	34 House speaker WASHINGTONpoliticsNews December 13, 2017	35 Alabama off (Reuters) - D politicsNews December 13, 2017	36 House panel WASHINGTONpoliticsNews December 13, 2017	37 Callista Greg VATICAN/G1 politicsNews December 13, 2017	38 U.S. to India: WASHINGTONpoliticsNews December 13, 2017	39 House speaker WASHINGTONpoliticsNews December 13, 2017	40 Failed vote: UMA (Reuter) politicsNews December 13, 2017	41 Congress vs WASHINGTONpoliticsNews December 13, 2017	42 Trump on pi The following politicsNews December 13, 2017	43 House speaker WASHINGTONpoliticsNews December 13, 2017	44 House speaker WASHINGTONpoliticsNews December 13, 2017	45 U.S. House WASHINGTONpoliticsNews December 13, 2017	46 House speaker WASHINGTONpoliticsNews December 13, 2017	47 President U.S. SAN FRANCISCO politicsNews December 13, 2017	48 Senate vs WASHINGTONpoliticsNews December 13, 2017	49 U.S. to India: WASHINGTONpoliticsNews December 13, 2017	50 Short term: WASHINGTONpoliticsNews December 13, 2017	51 Spy pi WASHINGTONpoliticsNews December 13, 2017	52 Trump signs \$1.3 trillion spending bill WASHINGTONpoliticsNews December 13, 2017	53 House gives WASHINGTONpoliticsNews December 13, 2017	54 U.S. to India: WASHINGTONpoliticsNews December 13, 2017	55 U.S. travel WASHINGTONpoliticsNews December 13, 2017	56 Trump comes WASHINGTONpoliticsNews December 13, 2017	57 U.S. to India: WASHINGTONpoliticsNews December 13, 2017	58 Democrat F WASHINGTONpoliticsNews December 13, 2017	59 U.S. travel WASHINGTONpoliticsNews December 13, 2017	60 Trump signs \$1.3 trillion spending bill WASHINGTONpoliticsNews December 13, 2017	61 U.S. to India: (Reuters) - D politicsNews December 13, 2017	62 House speaker WASHINGTONpoliticsNews December 13, 2017	63 White House WASHINGTONpoliticsNews December 13, 2017	64 Democrat F WASHINGTONpoliticsNews December 13, 2017	65 After being (elected) president: WASHINGTONpoliticsNews December 13, 2017

False

Excel screenshot showing a CSV file named "Fake" with 93 rows of data. The columns are labeled A through AA. The data includes news items from various sources like Fox News, NBC News, and CNN, dated from December 1, 2017, to December 29, 2017.

Index	Title	Text	Subject	Date
1	Donald Trump Donald Trump News	Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.		December 31, 2017
2	Drunk Bragg House Intelli News			December 31, 2017
3	Sheriff Davi On Friday, It News			December 30, 2017
4	Trump's So Sad About Christmas News			December 29, 2017
5	Papa John's Papa John's News			December 29, 2017
6	WATCH Fox Republicans News			December 29, 2017
7	Heads Up! The number... News			December 29, 2017
8	Fresh Off Ti Donald Trump News			December 29, 2017
9	Trump Said In The wake... News			December 29, 2017
10	Former FBI Director speaks News			December 29, 2017
11	WATCH! Bre Justify by News			December 29, 2017
12	Papa John's A centerpiece... News			December 29, 2017
13	WATCH! Fox Republicans News			December 29, 2017
14	Bad Blood! Republicans News			December 29, 2017
15	WATCH! Lincoln meets... News			December 29, 2017
16	Heires To Abigail Disney News			December 29, 2017
17	Tone Deaf T Donald Trump News			December 29, 2017
18	The Intercept A new administration... News			December 29, 2017
19	Mario Lopez Speaks About News			December 29, 2017
20	SNI Hilarious Right now, It News			December 29, 2017
21	Republican! Senate Major News			December 29, 2017
22	In A Heartbe It almost see News			December 29, 2017
23	TV's First Lady... News			December 29, 2017
24	Meghan Mc As A Democrat News			December 29, 2017
25	CNN CALLS! Alabama is a News			December 29, 2017
26	White House! A bunch of... News			December 29, 2017
27	Desperate! Donald Trump News			December 29, 2017
28	Accused Chi Ronald Reag News			December 29, 2017
29	WATCH! Fox Judge Jean News			December 29, 2017
30	Liberal Grou Donald Trump News			December 29, 2017
31	Donald Trump's First... News			December 29, 2017
32	BREAKING! In America, A News			December 29, 2017
33	Watch It! Ti New questo News			December 29, 2017
34	Trump Com! On Wednesday News			December 29, 2017
35	Trump Only President... News			December 29, 2017
36	Watch This! Donald Trump News			December 29, 2017
37	Sitting GOP! Arizona Rep News			December 29, 2017
38	Republican! By now, the News			December 29, 2017
39	Donald Trump! All he News			December 29, 2017
40	Jahr! McCain All he News			December 29, 2017
41	This Sarah H Donald J. Tru News			December 29, 2017
42	FBI Agents I Donald Trump News			December 29, 2017
43	Another Of By now, even News			December 29, 2017
44	Leave Em Alone! Donald News			December 29, 2017
45	Trump Jumps Donald Trump News			December 29, 2017
46	Watch This! Donald Trump News			December 29, 2017
47	White House! White D... News			December 29, 2017
48	REAGAN! Donald Trump News			December 29, 2017
49	Trump Gets Donald Trump News			December 29, 2017
50	This Hilario! In Michigan, News			November 30, 2017
51	Joe Scarborough People have News			November 30, 2017
52	Sources Con Jared Kushner News			November 30, 2017
53	Trump! Good! Donald Trump News			November 30, 2017

Jupyter Notebook screenshot titled "Fake news Detection". The notebook interface shows a header with "localhost" and "Fake_news_detection - Jupyter Notebook". The main code cell contains the following:

```

Fake news Detection



```

Importing required library
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import re
import string

```



Inserting fake and real dataset



```

In [1]: df_fake = pd.read_csv("./data/Fake.csv")
df_true = pd.read_csv("./data/True.csv")

```


```

Jupyter Fake_news_detection Last Checkpoint: 16 minutes ago (autosaved)

In [3]: df_fake.head(5)

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year's...	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

In [4]: df_true.head(5)

	title	text	subject	date
0	As U.S. budget fight looms, Republicans flip...	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. Mueller...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017
3	FBI Russia probe helped by Australian diplomati...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donald...	politicsNews	December 29, 2017

Inserting a column called "class" for fake and real news dataset to categories fake and true news.

In [5]: df_fake["class"] = 0
df_true["class"] = 1

Removing last 10 rows from both the dataset, for manual testing

In [6]: df_fake.shape, df_true.shape

Out[6]: ((23481, 5), (21417, 5))

In [7]: df_fake_manual_testing = df_fake.tail(10)
for i in range(23481, 23470, -1):
 df_fake.drop([i], axis = 0, inplace = True)
df_true_manual_testing = df_true.tail(10)
for i in range(21416, 21406, -1):
 df_true.drop([i], axis = 0, inplace = True)

In [8]: df_fake.shape, df_true.shape

Out[8]: ((23471, 5), (21407, 5))

Merging the manual testing dataframes in single dataset and save it in a csv file

Jupyter Fake_news_detection Last Checkpoint: 16 minutes ago (autosaved)

In [8]: df_fake.shape, df_true.shape

Out[8]: ((23471, 5), (21407, 5))

Merging the manual testing dataframes in single dataset and save it in a csv file

In [9]: df_fake_manual_testing["class"] = 0
df_true_manual_testing["class"] = 1

```
/var/folders/j/h_9zspfh004nnw2cszy4mc8000gn/T/ipykernel_55719/860779283.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
df_true_manual_testing["class"] = 1

```
/var/folders/j/h_9zspfh004nnw2cszy4mc8000gn/T/ipykernel_55719/860779283.py:2: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
df_true_manual_testing["class"] = 1

In [10]: df_fake_manual_testing.head(10)

	title	text	subject	date	class
23471	Seven Iranians freed in the prisoner swap have...	21st Century Wire says this week, the historic...	Middle-east	January 20, 2016	0
23472	#Hashtag Hell & The Fake Left	By Dady Chery and Gilbert MercierAll writers...	Middle-east	January 19, 2016	0
23473	Astrotroupe: Journalist Reveals Brainwashing ...	Vic Bishop Waking TimesOur reality is careful...	Middle-east	January 19, 2016	0
23474	The New American Century: An Era of Fraud	Paul Craig RobertsThe last years of the 20th...	Middle-east	January 19, 2016	0
23475	Hillary Clinton: 'Israel First' (and no peace ...	Robert Faurita CounterpunchAlthough the United...	Middle-east	January 18, 2016	0
23476	McPalin: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016	0
23477	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It's a familiar theme. ...	Middle-east	January 16, 2016	0
23478	Sunnistan: US and Alled 'Safe Zone' Plan to T...	Patrick Henningsen21st Century WireRemember ...	Middle-east	January 15, 2016	0
23479	How to Blow \$700 Million: Al Jazeera America F...	21st Century Wire says Al Jazeera America will...	Middle-east	January 14, 2016	0
23480	10 U.S. Navy Sailors Held by Iranian Military ...	21st Century Wire says As 21WIRE predicted in ...	Middle-east	January 12, 2016	0

In [11]: df_true_manual_testing.head(10)

	title	text	subject	date	class
21407	Mata Pires, owner of embattled Brazil builder ...	SAO PAULO (Reuters) - Cesar Mata Pires, the ow...	worldnews	August 22, 2017	1
21408	U.S., North Korea clash at U.N. forum over nuc...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017	1

In [11]: df_true_manual_testing.head(10)

	title	text	subject	date	class
21407	Mata Pires, owner of embattled Brazil builder ...	SAO PAULO (Reuters) - Cesar Mata Pires, the ow...	worldnews	August 22, 2017	1
21408	U.S., North Korea clash at U.N. arms forum on ...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017	1
21409	U.S., North Korea clash at U.N. arms forum on ...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017	1
21410	Headless torso could belong to submarine journ...	COPENHAGEN (Reuters) - Danish police said on T...	worldnews	August 22, 2017	1
21411	North Korea shipments to Syria chemical arms a...	UNITED NATIONS (Reuters) - Two North Korean sh...	worldnews	August 21, 2017	1
21412	Fully committed NATO backs new U.S. approach.	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	1
21413	LexisNexis withdraws two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of L...	worldnews	August 22, 2017	1
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of dissolved Sov...	worldnews	August 22, 2017	1
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017	1
21416	Indonesia to buy \$1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukho...	worldnews	August 22, 2017	1

In [12]: df_manual_testing = pd.concat([df_fake_manual_testing, df_true_manual_testing], axis = 0)
df_manual_testing.to_csv("./testing/manual_testing.csv")

Merging the main fake and true dataframe

In [13]: df_marge = pd.concat([df_fake, df_true], axis = 0)
df_marge.head(10)

	title	text	subject	date	class
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn't wish all Americans ...	News	December 31, 2017	0
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	0
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauke...	News	December 30, 2017	0
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	0
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	0
5	Racist Alabama Cops Brutalize Black Boy While...	The number of cases of cops brutalizing and k...	News	December 23, 2017	0
6	Fresh Off The Golf Course, Trump Lashes Out A...	Donald Trump spent a good portion of his day a...	News	December 23, 2017	0
7	Trump Said Some INSANELY Racist Stuff Inside ...	In the wake of yet another court decision that...	News	December 23, 2017	0
8	Former CIA Director Slams Trump Over UN Bully...	Many people have raised the alarm regarding th...	News	December 22, 2017	0
9	WATCH: Brand-New Pro-Trump Ad Features So Muc...	Just when you might have thought we'd get a br...	News	December 21, 2017	0

In [14]: df_marge.columns

Out[14]: Index(['title', 'text', 'subject', 'date', 'class'], dtype='object')

In [14]: df_marge.columns

Out[14]: Index(['title', 'text', 'subject', 'date', 'class'], dtype='object')

"title", "subject" and "date" columns is not required for detecting the fake news, so I am going to drop the columns.

In [15]: df = df_marge.drop(["title", "subject", "date"], axis = 1)

In [16]: df.isnull().sum()

Out[16]: text 0
class 0
dtype: int64

Randomly shuffling the dataframe

In [17]: df = df.sample(frac = 1)

In [18]: df.head()

Out[18]:

	text	class
5859	Ever since the horrific act of senseless viole...	0
4906	ISTANBUL (Reuters) - U.S. Secretary of State R...	1
11259	OUR PREVIOUS REPORT ON JUDGE NAPOLITAN S CLAIM...	0
15103	He d like to build a giant wall, deport illegit...	0
12075	BETHLEHEM, West Bank (Reuters) - A small group...	1

In [19]: df.reset_index(inplace = True)
df.drop(["index"], axis = 1, inplace = True)

In [20]: df.columns

Out[20]: Index(['text', 'class'], dtype='object')

In [21]: df.head()

Out[21]:

	text	class
0	Ever since the horrific act of senseless viole...	0
1	ISTANBUL (Reuters) - U.S. Secretary of State R...	1
2	OUR PREVIOUS REPORT ON JUDGE NAPOLITAN S CLAIM...	0
3	He d like to build a giant wall, deport illegit...	0
4	RETHYMNO, Greek Islands (Reuters) - A small group...	1

Jupyter Fake_news_detection Last Checkpoint: 16 minutes ago (autosaved)

```
In [21]: df.head()
Out[21]:
   text      class
0  Ever since the horrific act of senseless vio...      0
1  ISTANBUL (Reuters) - U.S. Secretary of State R...      1
2  OUR PREVIOUS REPORT ON JUDGE NAPOLITAN S CLAIM...      0
3  He d like to build a giant wall, deport illega...      0
4  BETHLEHEM, West Bank (Reuters) - A small group...      1
```

Creating a function to convert the text in lowercase, remove the extra space, special chr, url and links.

```
In [22]: def wordopt(text):
    text = text.lower()
    text = re.sub(r'[\r\n\t\f\v]', ' ', text)
    text = re.sub(r'[^\w\s]', '', text)
    text = re.sub(r'(http|https)://[\w\W]+', '', text)
    text = re.sub(r'<[^>+>+', '', text)
    text = re.sub(r'[\%]{2} re.escape(string.punctuation), '' ,text)
    text = re.sub(r'\n', ' ', text)
    text = re.sub(r'\w\w\w', ' ', text)
    return text
```

```
In [23]: df["text"] = df["text"].apply(wordopt)
```

Defining dependent and independent variable as x and y

```
In [24]: x = df["text"]
y = df["class"]
```

Splitting the dataset into training set and testing set.

```
In [25]: x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25)
```

Convert text to vectors

```
In [26]: from sklearn.feature_extraction.text import TfidfVectorizer
```

```
In [27]: vectorization = TfidfVectorizer()
xv_train = vectorization.fit_transform(x_train)
xv_test = vectorization.transform(x_test)
```

Jupyter Fake_news_detection Last Checkpoint: 17 minutes ago (autosaved)

```
In [28]: from sklearn.linear_model import LogisticRegression
```

```
In [29]: LR = LogisticRegression()
LR.fit(xv_train,y_train)
```

```
Out[29]:
+ LogisticRegression
LogisticRegression()
```

```
In [30]: pred_lr=LR.predict(xv_test)
```

```
In [31]: LR.score(xv_test, y_test)
```

```
Out[31]: 0.9850267379679144
```

```
In [32]: print(classification_report(y_test, pred_lr))
          precision    recall  f1-score   support
          0       0.99     0.98     0.99    5959
          1       0.98     0.99     0.98    5261
   accuracy                           0.99    11220
  macro avg       0.98     0.99     0.98    11220
weighted avg       0.99     0.99     0.99    11220
```

2. Decision Tree Classification

```
In [33]: from sklearn.tree import DecisionTreeClassifier
```

```
In [34]: DT = DecisionTreeClassifier()
DT.fit(xv_train, y_train)
```

```
Out[34]:
+ DecisionTreeClassifier
DecisionTreeClassifier()
```

```
In [35]: pred_dt = DT.predict(xv_test)
```

```
In [36]: DT.score(xv_test, y_test)
```

```
Out[36]: 0.9959001782531194
```

```
In [37]: print(classification_report(y_test, pred_dt))
```

Jupyter Fake_news_detection Last Checkpoint: 17 minutes ago (autosaved)

```
In [37]: print(classification_report(y_test, pred_dt))
      precision    recall  f1-score   support
          0       1.00     1.00     1.00      5959
          1       1.00     0.99     1.00      5261
   accuracy                           1.00      11220
  macro avg       1.00     1.00     1.00      11220
weighted avg       1.00     1.00     1.00      11220
```

3. Gradient Boosting Classifier

```
In [38]: from sklearn.ensemble import GradientBoostingClassifier
In [39]: GBC = GradientBoostingClassifier(random_state=0)
GBC.fit(xv_train, y_train)
Out[39]: GradientBoostingClassifier(random_state=0)

In [40]: pred_gbc = GBC.predict(xv_test)
In [41]: GBC.score(xv_test, y_test)
Out[41]: 0.9950089126559715

In [42]: print(classification_report(y_test, pred_gbc))
      precision    recall  f1-score   support
          0       1.00     0.99     1.00      5959
          1       0.99     1.00     0.99      5261
   accuracy                           1.00      11220
  macro avg       0.99     1.00     0.99      11220
weighted avg       1.00     1.00     1.00      11220
```

4. Random Forest Classifier

```
In [43]: from sklearn.ensemble import RandomForestClassifier
In [44]: RFC = RandomForestClassifier(random_state=0)
RFC.fit(xv_train, y_train)
Out[44]: RandomForestClassifier(random_state=0)
```

Jupyter Fake_news_detection Last Checkpoint: 17 minutes ago (autosaved)

```
In [44]: RFC = RandomForestClassifier(random_state=0)
RFC.fit(xv_train, y_train)
Out[44]: RandomForestClassifier(random_state=0)

In [45]: pred_rfc = RFC.predict(xv_test)
In [46]: RFC.score(xv_test, y_test)
Out[46]: 0.9863636363636363

In [47]: print(classification_report(y_test, pred_rfc))
      precision    recall  f1-score   support
          0       0.99     0.99     0.99      5959
          1       0.99     0.98     0.99      5261
   accuracy                           0.99      11220
  macro avg       0.99     0.99     0.99      11220
weighted avg       0.99     0.99     0.99      11220
```

Model Testing With Manual Entry

```
News
In [48]: def output_label(n):
    if n == 0:
        return "Fake News"
    elif n == 1:
        return "Not A Fake 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 = vectorization.transform(new_def_test["text"])
    new_x_test = vectorization.transform(new_x_test)
    pred_LR = LR.predict(new_x_test)
    pred_DT = DT.predict(new_x_test)
    pred_GBC = GBC.predict(new_x_test)
    pred_RFC = RFC.predict(new_x_test)

    return print("\nLR Prediction: {} \nDT Prediction: {} \nGBC Prediction: {} \nRFC Prediction: {}".format(output
          output
          output
          output))
```

The screenshot shows a Jupyter Notebook interface with the title "Fake_news_detection". At the top, there's a toolbar with various icons like file operations, run, and help. Below the toolbar, a status bar indicates "localhost" and "Python 3 (ipykernel)". The main area has a header "Model Testing With Manual Entry" and a section titled "News".

Model Testing Results:

	accuracy	0.99	0.98	0.99	5261
macro avg	0.99	0.99	0.99	11220	
weighted avg	0.99	0.99	0.99	11220	

Code for Model Testing:

```
In [48]: def output_label(n):
    if n == 0:
        return "Fake News"
    elif n == 1:
        return "Not A Fake News"

def manual_testing(news):
    testing_news = pd.DataFrame([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 = vectorization.transform(new_x_test)
    pred_LR = LR.predict(new_xv_test)
    pred_DT = DT.predict(new_xv_test)
    pred_GBC = GBC.predict(new_xv_test)
    pred_RFC = RFC.predict(new_xv_test)

    return print("\nLR Prediction: {} \nDT Prediction: {} \nGBC Prediction: {} \nRFC Prediction: {}".format(output_label(pred_LR), output_label(pred_DT), output_label(pred_GBC), output_label(pred_RFC)))
```

Input and Output Cells:

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

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

```
In []:
```

The article reported that Herschel had made these discoveries using new “hydro-oxygen magnifiers” and went on to describe in believable scientific detail, how the discovery was made. Bizarre life forms, inhabitants of the moon, were described, painting a fantastical picture.

“*Of animals, he classified nine species of mammalia, and five of ovipara. Among the former is a small kind of rein-deer, the elk, the moose, the horned bear, and the biped beaver. The last resembles the beaver of the earth in every other respect than in its destitution of a tail, and its invariable habit of walking upon only two feet. It carries its young in its arms like a human being, and moves with an easy gliding motion.*

— GREAT ASTRONOMICAL DISCOVERIES Lately Made, The Sun, Thursday, August 27, 1835

99

```
In [49]: news = str(input())
manual_testing(news)
```

Of animals, he classified nine species of mammalia, and five of ovipara. Among the former is a small kind of reindeer, the elk, the moose, the horned bear, and the biped beaver. The last resembles the beaver of the earth in every other respect than in its destitution of a tail, and its invariable habit of walking upon only two feet. It carries its young in its arms like a human being, and moves with an easy gliding motion.

```
LR Prediction: Fake News
DT Prediction: Fake News
GBC Prediction: Fake News
RFC Prediction: Fake News
```

Donald Trump

Trump sues US government over FBI search of Mar-a-Lago

Ex-president seeks to prevent bureau from reading seized documents until court official weighs in

Martin Pengelly in New York

✉ @MartinPengelly
Mon 22 Aug 2022 22.45 BST



📷 The FBI searched Mar-a-Lago on 8 August. Photograph: Marco Bello/Reuters

Donald Trump on Monday filed suit against the US government over the FBI search of his Mar-a-Lago home, seeking to temporarily stop the bureau reading seized materials until a special court official can be appointed to review documents concerned.

As the Guardian [reported on Saturday](#), citing Trump's lead attorney, Jim Trusty, and two sources familiar with the matter, "the suit argues that the court should appoint a special master - usually a retired lawyer or judge - because the **FBI** potentially seized privileged materials in its search and the Department of Justice (DoJ) should not itself decide what it can use in its investigation".

The suit, [filed in US district court](#) for the southern district of Florida, also "requires the government to provide a more detailed receipt for property; and ... requires the government to return any item seized that was not within the scope of the search warrant".

```
In [51]: news = str(input())
manual_testing(news)
```

Donald Trump on Monday filed suit against the US government over the FBI search of his Mar-a-Lago home, seeking to temporarily stop the bureau reading seized materials until a special court official can be appointed to review documents concerned. As the Guardian reported on Saturday, citing Trump's lead attorney, Jim Trusty, and two sources familiar with the matter, "the suit argues that the court should appoint a special master - usually a retired lawyer or judge - because the **FBI** potentially seized privileged materials in its search and the Department of Justice (DoJ) should not itself decide what it can use in its investigation". The suit, filed in US district court for the southern district of Florida, also "requires the government to provide a more detailed receipt for property; and ... requires the government to return any item seized that was not within the scope of the search warrant". The Mar-a-Lago search, on 8 August, was mounted to look for official records and material from Trump's presidency that the National Archives and DoJ believe was improperly taken from the White House when Trump left office.

LR Prediction: Fake News
DT Prediction: Fake News
GBC Prediction: Fake News
RFC Prediction: Not A Fake News

Source Code link: <https://github.com/achbj/fake-news-detection>

Project Video link: Thesis-Link (https://livecoventryac-my.sharepoint.com/:f/g/personal/acharyab2_unif_c Coventry_ac_uk/EhfKj12vwVFMve3D3yrYfsB8hax3iFr4aIbE1G0GJtTFQ?e=p02lcE)