Fake News on COVID-19 Detector, CovFaq

by

Nur Alia Shafiah binti Azian Sufardi

17000395

Dissertation submitted in partial fulfilment of the requirement for the

Bachelor of Information System (Hons)

September 2021

Universti Teknologi PETRONAS

32610 Seri Iskandar

Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Fake News on COVID-19 Detector, CovFaq

By

Nur Alia Shafiah binti Azian Sufardi

17000395

A project dissertation submitted to the

Universiti Teknologi PETRONAS

In partial fulfilment of the requirement for the

BACHELOR OF INFORMATION SYSTEM (Hons)

Approved by,

_____

Ts. Faizal Ahmad Fadzil

UNIVERSITI TEKNOLOGI PETRONAS

SERI ISKANDAR, PERAK

SEPTEMBER 2021

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

Alia Shafiah

Nur Alia Shafiah binti Azian Sufardi

ABSTRACT

SARS-CoV-2-caused respiratory tract illness (RTI) is the primary mode of transmission for COVID-19, which spreads mostly by droplets in the air. The number of people infected with COVID-19 is rising at a fast pace in Malaysia. It's not only that COVID-19 has been the subject of a lot of disinformation on social media, but there has also been a lot of it (World Health Organization,2020).

Fake News detection on COVID-19 is the focus of this dissertation. Fake news is a major problem, and the goal of this project is to design a system that prevents people from sharing it. During the project's development, the machine learning technique is the best option. Planning, analysis, design, development, testing, development, maintenance, and feedback are all necessary steps in this project's completion.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

COVID-19	Coronavirus Disease 2019
WHO	World Health Organization
FYP 1	Final Year Project 1
FYP 2	Final Year Project 2
CSV	Comma Separated Values
NLP	Natural Language Processing

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Coronavirus illness, or COVID-19, The SARS-CoV-2 influenza virus, which was initially discovered in Wuhan, China, is responsible for a respiratory sickness epidemic in 2019. Infectious illness caused by a recently identified coronavirus, COVID-19. Most persons who are infected with the COVID-19 virus get symptoms ranging from the mild to the severe. If you're above the age of 60, you're more likely to have a major illness, such as heart disease, diabetes, or cancer, than someone younger.

World Health Organization (WHO) has warned of an on-going "info-demic" or an oversupply of information, particularly incorrect information, during an epidemic, due to the transmission of inaccurate information about the virus. Misinformation, fake news, and disinformation are all examples of false information. Spreading misinformation does not necessary mean that the information in question is malicious. Disinformation, on the other hand, is information that has been deliberately fabricated to deceive or mislead others. People who hear this information will be harmed by it since it is based on biassed information.

The main point is fake news which is lies or fabricated information or news that is non-verifiable throughout sources, facts that includes of conspiracy theories, clickbait pages and photos or post that have been manipulated or edited that make sure to look real that can keep popping up all over the social media account like Facebook, Instagram, Twitter and many more.

COVID-19 Fake News in Bahasa Detector is a good system that can be implement it. The aim for this system is to detect whether the news about COVID-19 in Bahasa is fake or not. For the first part of this system is to write a python script to detect the fake news.

1.2 Problem Statement

There are two main concerns face in this research that includes of

- There is no platform to verify all the COVID-19 news is fake or not
- Mostly people just share the COVID-19 news without checking it first before spreading it to the others.

As for now, there still no platform to verify all the COVID-19 news is fake or not. This is because not everyone is aware about the importance of not to share fake news to the others. There only have a campaign in social media such as campaigns #ThinkBeforeSharing at Twitter, YouTube COVID-19 Medical Misinformation Policy a YouTube that can tackles the content that contradicts World Health Organization (WHO).

Next problem is mostly people just share the COVID-19 news without checking it first because they want to be a first person to share the news. Mostly they do this for the sake of fame and to get attention for other people.

1.3 Objectives

In order to complete the COVID-19 Fake New Detector project, below are the project's objectives:

- To develop an input system called CovFaq that able to store various COVID19's confirmed fake news into a database
- To develop a platform for people to check fake news utilizing CovFaq

1.4 Scope of Study

Fake news regarding COVID-19 is being distributed throughout the public, and this research of COVID-19 Fake News Detector is designed to assist authorities stop it. Prior to constructing a system, it's essential to identify the individuals who will be using it and to identify the target audience. At the start, the project's major users were designated as the government and the general public.

Other than that, this project must be finished within amount of time, which is eight months, between May 2021 until December 2021. During the first four months, in Final Year Project 1 (FYP 1), student need to perform the

analysis about the project while during the final four months of the project, that known as Final Year Project 2 (FYP2), the student needs to do the implementation of the system, such as the delivery of the working systems must be complete in this time.

1.5 Significance of Project

In the problem statement section discussed the matters has led to the study of this research topic and will regards to the significant matter can be brought to this project. The importance of this research is to provide a good technology that can help people that use it. Next, this project is very relevant because no more fake news will be share around person in the public. Therefore, it is very crucial to have a system that can ensure all people not share the fake news of COVID-19 that can trigger others.

CHAPTER 2

LITERATURE REVIEW

A literature review is an examination of academic materials, such as articles, theses, and books, that are relevant to the topic at hand. To begin any research endeavor, you'll need to do a review of the existing literature. Useful in undergraduate and post-graduate tasks particularly when writing an academic paper such as a thesis or a dissertation Analysis's fundamental principles are not overlooked, and this demonstrates this to readers.

2.1 Fake News About COVID-19

Zarocostas (2020) asserts that it is more difficult for individuals to get trustworthy and credible information when they want it as a result of the On social media, misinformation concerning COVID-19 has spread extensively, ranging from the marketing of bogus "cure" methods such as gargling with lemon or salt water. Even defining COVID-19 disinformation is a difficult task, since fresh information about the virus's origins and treatments emerges every time. Nonetheless, it is becoming clearer that extensive COVID-19 misinformation exists. For instance, a study performed by Ofcom in the United Kingdom revealed that more than half of the public (46 percent) had been exposed to misleading information about the coronavirus (Ofcom, 2020). In the United States, the Pew Research Center found comparable findings (48 percent) (Mitchell and Oliphant, 2020).

Not only that, in the social media account, social media users' motivations for information sharing, socialization, information seeking, and pass time predicted the

sharing of false information about COVID-19. In contrast, no significant association was found for entertainment motivation. Therefore, through shares, likes, and retweets, social media users can spread news, so they are invariably exposed to an unmanageable type of information, particularly news from independent authors. As a result, social media has become a vehicle for quickly disseminating false information and fake news (Rampersad et al., 2019).

Numerous rumours and misleading news reports concerning the COVID-19 are spreading. It is getting more difficult to separate false news from reliable sources. Misinformation spread through social media has fanned public alarm over the COVID-19 epidemic, forcing governments and authorities to warn individuals to verify the authenticity of news reports before spreading them (Huynh, 2020, Hou et al., 2020). The spread of fake news has accelerated on social media, which many experts feel is adding to the pandemic's risks (Lampos et al., 2020).
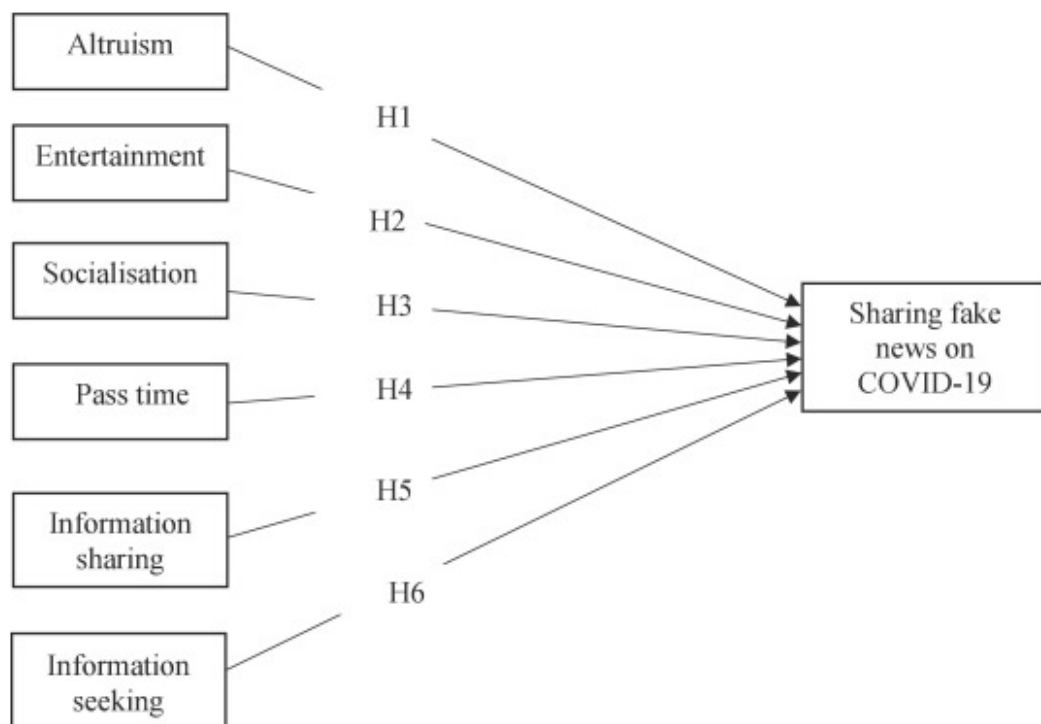
Figure 1: Research Model for Fake News

2.2 Fake News Detection Using Machine Learning
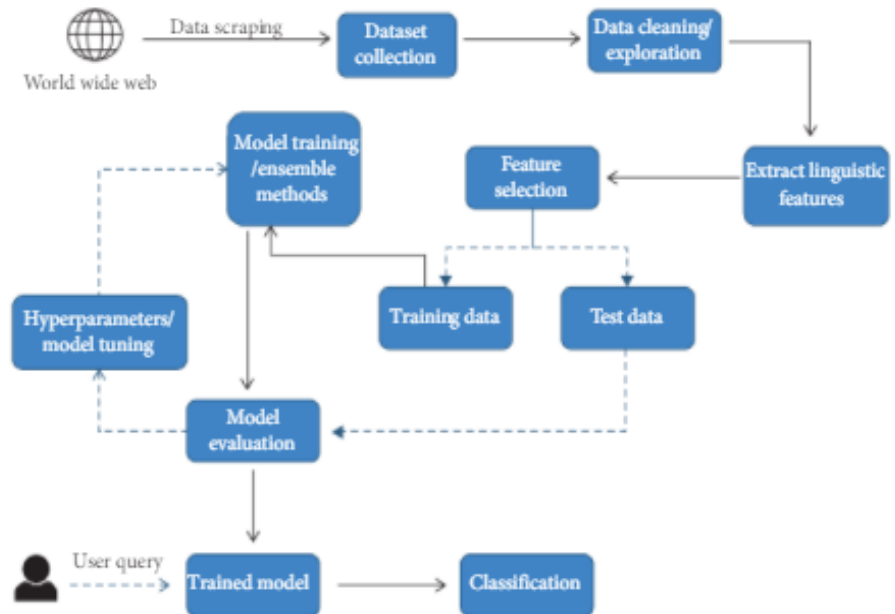
Figure 2: Workflow for training algorithms and classification of news

Machine learning is used to ensemble strategy to classify news stories automatically. It is also to investigate several textual features that can be utilized to distinguish bogus from a real information. Next, is to train a combination of the several machine learning algorithms that using a various ensemble approaches based on the properties.

In order to create a model of a product with supervised machine learning algorithm, that can classify fake news as true or false, by using tools like python, NLP for textual analysis. This process will result in feature extraction and vectorization.
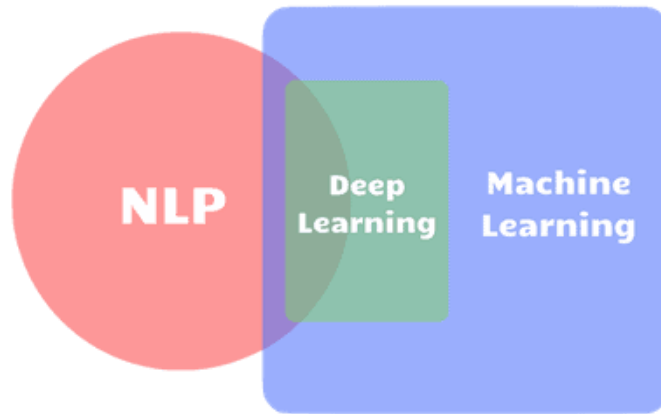
Figure 3: Nature Language Processing.

Natural Language Processing is mostly used to examine one or more specialties of a system or method. An algorithmic system's Natural Language Processing (NLP) rating permits the merging of voice interpretation and speech creation. Additionally, it can be used to identify activities in a variety of languages, making it a good fit for our project.

CHAPTER 3

METHODOLY AND PROJECT WORK

3.1 Research Methodology

Research technique is a practical guide on "how" to do research; it is about how a researcher structures a study in order to ensure that the results are accurate and that the study's aims and objectives are satisfied. This research methodology should provide light on all design selections by proving that the procedures selected are completely compatible with the study aims and objectives. A solid research process produces scientifically sound findings, while a poor research methodology produces none.

3.2 Chosen Methodology

Figure 4: Agile Methodology

Agile technique will be used to organise and plan the whole development stages. This Agile technique consists of four phases: planning, design, programming, and testing. Release phase and feedback phase, during which user testers provide input. Agile technique will be used to construct this project since it is the most efficient and effective way for software development. The following are some of the benefits of adopting an agile methodology:

- The capacity to make necessary changes quickly throughout the development period.
- Capable of minimizing the danger of defects during the last phases of development owing to the agility with which modifications may be made.
- To save more time and resources via effective planning and management.

In the phase one, which is planning phase where all the requirements such as project goals, timelines and budgets, all are defined and be finalized. After that is phase two, the design phase, where all the development begins. All the prototypes that design in the phase two, will be converted into working models during phase 3 which is development phase. After that, the finished product will be testing in the phase 4. Next, the systems will be releasing in the fifth phase and the last phase is feedback phase.

Phase 1: Planning phase = The author summarizes the vision and strategy for designing the system in this phase. Brainstorming ideas and compiling pertinent information on the critical nature of implementing this system. Moreover, at this phase, various boundaries and limitations are established to help prevent the project from expanding or deviating from its primary objective.

Phase 2: Planning phase = In this phase, the software architecture, user interface, development tools on which the software will run, and tools must all be closely researched.

Phase 3: Development phase = This is a critical stage in the development process. The coding process entails creating the source code, as well as finding and correcting errors or glitches.

Phase 4: Testing phase = Testing is done in order to ensure that all functions function properly and to reduce processing time delays. The testing stage helps to reduce the number of errors and bugs.

Phase 5: Deployment and Feedback Phase = In Agile, the development stage is the final step. Maintenance is required when a bug is discovered by a user that was not discovered during the testing phase.

3.3 Machine Learning

Figure 5: Machine Learning.

Machine learning is a data analytics method which is to teach machines to learn from experience in the same way that humans and animals do. This implement analytical methods is to understand knowledge directly from data rather than relying on the model based on a predefined algorithm. Deep learning is a form of machine learning that is highly specialized. For the machine learning, it is consisting of four methods which is supervised learning, unsupervised learning, semi-supervised learning and the last one is reinforcement learning. In this project, the best method for COVID-19 Fake News in Bahasa Detector is supervised learning.

For the classification, it needs data analytics and machine learning. Each of the testing instance's class is determined by combining features and identifying trends common to each of the class in the training data (Amanpreet er al.,2016). All the input data is classified using classification models.

3.4 Flowchart of the System

The creation of the system flowchart is critical because it can describe the activities visually and decisions that the systems run. It also assists in gaining a better understanding of the system.
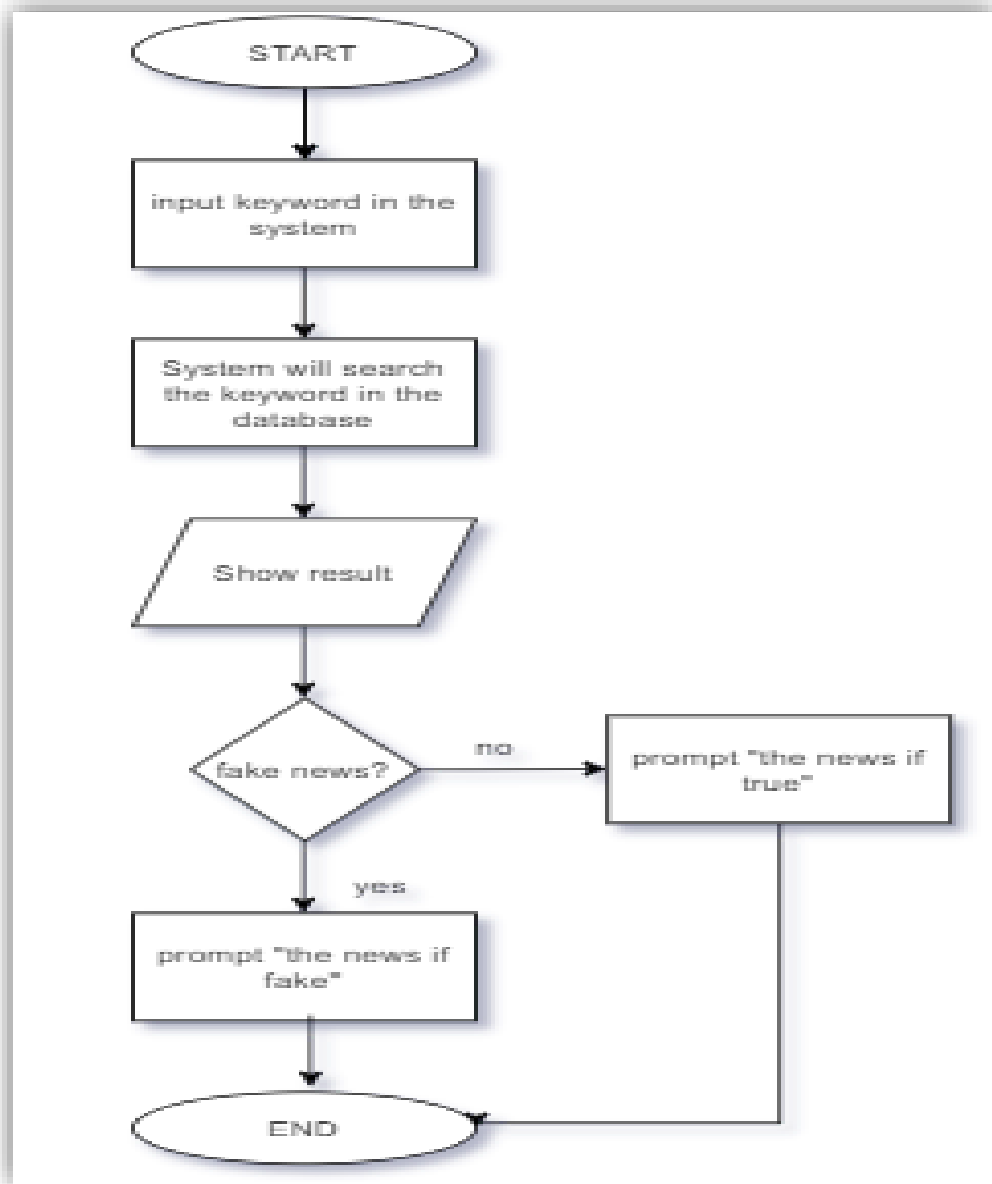
Figure 6: Flow Chart of System.

In the figure above, the author shows the flowchart of the process of Fake News of Covid-19 Detector from the beginning to the end of the system that had been develop.

Based on the figure above, the user needs to put input such as the title of the news or only the part of text. After that, the system will search the keywords in the database and after the systems found the keywords in the database, it will show the result which is the news is fake or the news is real.

14

3.5 Technology Tools

This are the following tools that will be used in this study:

- Python is the programming language that will be use
- Jupyter is the Integrated Development Environment (IDE) that support Python programming language

Figure 7: Jupyter Notebook

Jupyter Notebook is an open-source web tool that can integrates the live code, visualization, text and many more to the single document. The author uses Jupyter Notebook for this project because it can serve as an Integrated Development Environment (IDE) for testing, writing and debugging code. IDE is usually included in the source code editor and a debugger.

Figure 8: Python

Author used Python as a language because Python is a language that can interpreted and high-level programming language that has constantly semantics. Not only that, but it is also easy to learn, and it can reduce program maintenance costs. It is also free to download the interpreter and can distribute in binary or source form for all the major platforms.

3.6 Project Timelines

3.6.1 Key Milestones

Key Milestone	Completion Week
Proposal Submission	Week 2 of FYP1
Progress Assessment 1	Week 3 of FYP1
Proposal Defence	Week 8 of FYP1
Progress Assessment 2	Week 11 of FYP1
Interim Report	Week 12 of FYP1
Early Development	FYP 2
Progress Assessment 1	Week 3 of FYP 2
VIVA	Week 14 of FYP 2
Progress Assessment 2	Week 12 of FYP 2
Dissertation	Week 12 for softbound and Week 14 for hardbound

Table 1: Key milestone

3.6.2 Gantt Chart

The table below shows the Gantt chart throughout FYP1:

Week / Task	1	2	3	4	5	6	7	8	9	10	11	12
Select project title	■	■										
Refine project proposal			■	■	■							
Do literature search		■	■	■	■	■	■	■	■	■	■	
Background & Introduction					■	■						
Identify objectives &Literature Review					■	■	■					
Collect the data and analysis				■	■	■						
Plan the methodology						■	■					
Present Proposal Défense								■				
Prepare the Interim Report Submission									■	■	■	■
Submission the Interim Report												■

Table 2: Gantt Chart FYP1

The table below shows the Gantt chart throughout FYP2:

Project Development														
Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Activities	Sept				Oct				Nov				Dec	
Developing system software	■	■	■	■	■	■								
Training the Detector	■	■	■	■	■									
Training the Classifier		■	■	■	■	■								
Applying the Model			■	■	■	■								
Prototype Testing							■	■						
Fixed Testing									■					
System Implementation									■	■				
System Maintenance										■	■			
Submission of Dissertation (Soft bound)									■					
Viva												■		
Submission of Dissertation (Hard bound)														■

Table 3: Gantt Chart FYP2

CHAPTER 4

RESULT AND DISCUSSION

4.1 Preliminary Survey

Preliminary surveying comprises obtaining sufficient information to construct a plan/map of the region that can be utilized for planning and design. This kind of survey involves a rapid investigation of a route that has been experimentally selected based on reconnaissance survey information and suggestions.

For this project, the author needs to know how people will response and react to this project. Not only that, but the author also needs to know what it can be benefits on the people. So, the author has prepared a survey for people to answer. The survey is mostly about how people concern about fake news around them. This survey consists of 7 questions and the responded in total of 100 people. All the questions are about Covid-19 fake news and their concern about it. So, below are the result and further explanation for the questions.
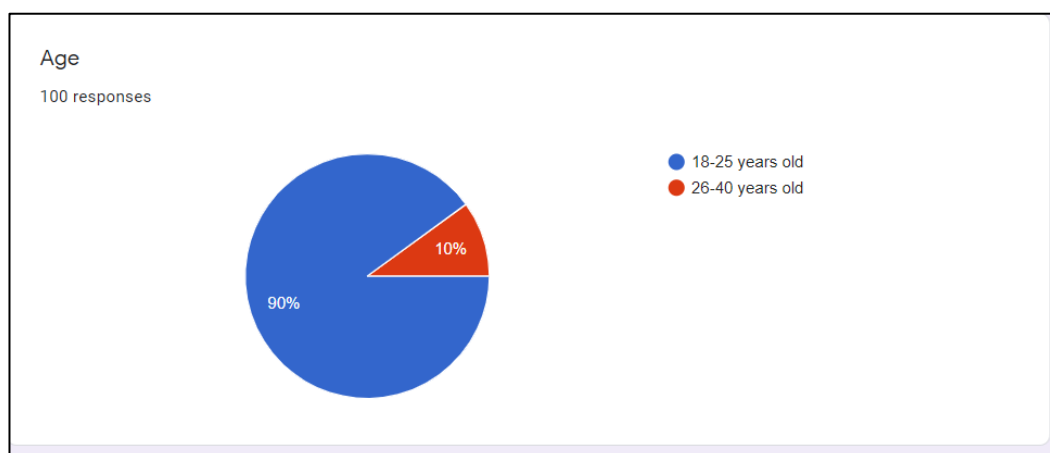
Figure 9: Survey Result Question 1

For the first question, the respondent's age is required. The author divided the age range in this question into two categories: 18 to 25 and 26 to 40 years old. According to the pie chart above, 90% of respondents are between the ages of 18 and 25, accounting for 90 out of 100 respondents, while just 10% are between the ages of 26 and 40, accounting for ten respondents.
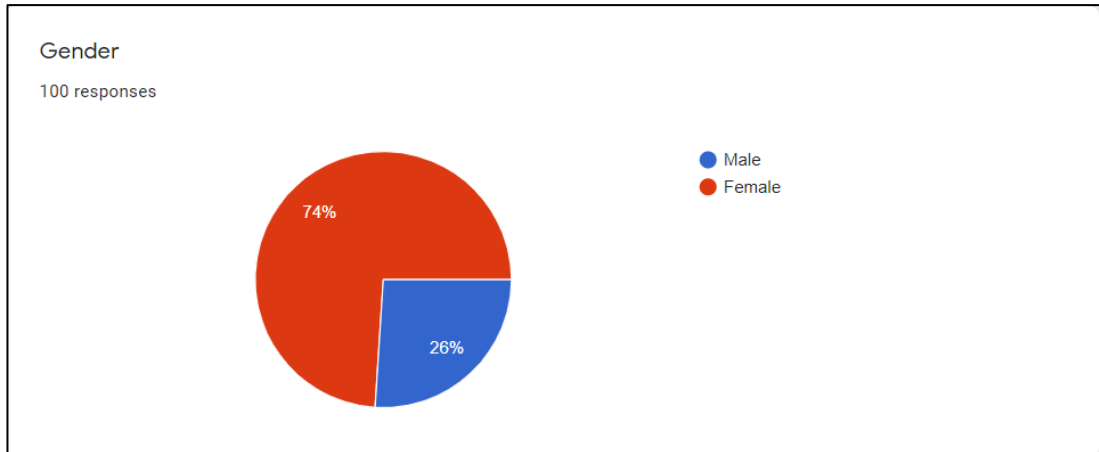
Figure 10: Survey Result Question 2

The second question is about their gender. Mostly the responders are female because around 74% which is consists of 74 respondents while only 26% of the respondents are male which is consists of 26 out of 100 respondents.
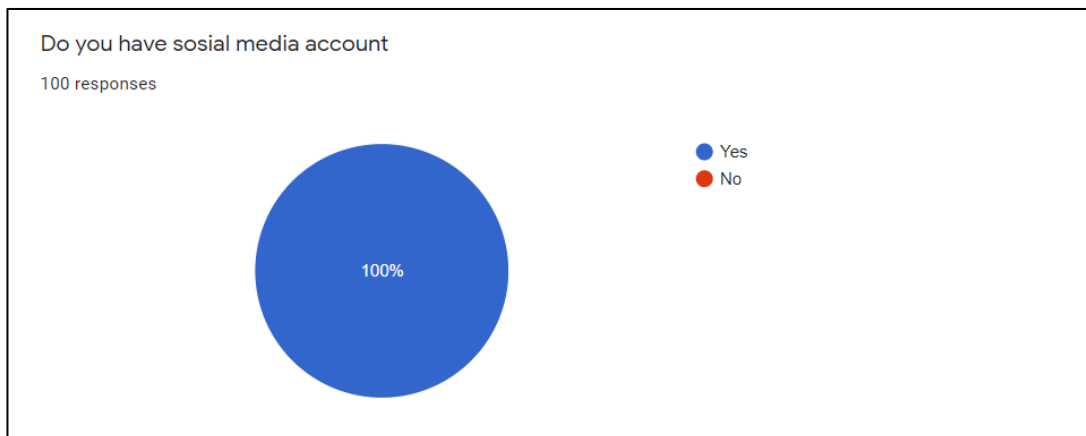
Figure 11: Survey Result Question 3

For the third question, it is about social media account. As the author can conclude is all the respondents have the social media account such as Facebook, Instagram, Twitter and many more.
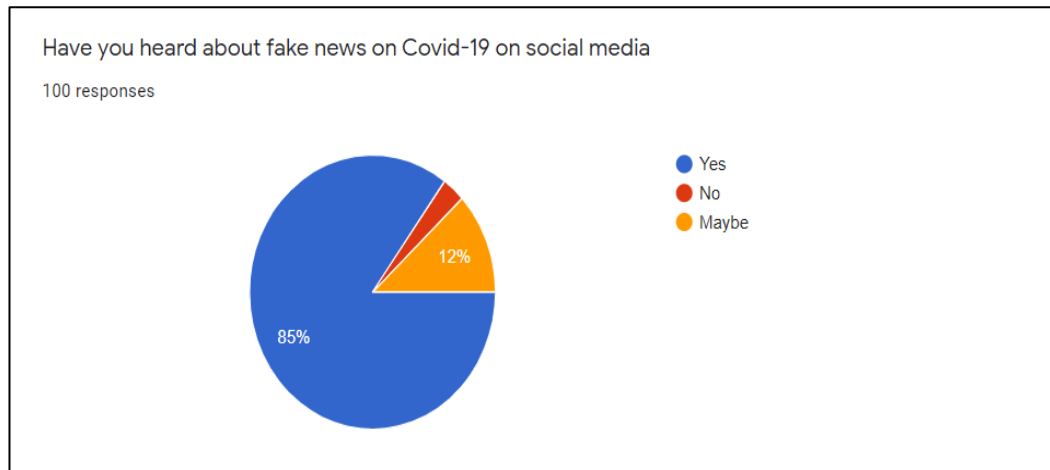
Figure 12: Survey Result Question 4

For this question, mostly the respondents already heard about fake news on Covid-19 on the social media which is consists of the 85% and around 12% of the respondents chose maybe as their answers. There are around 3% of the respondents chose "no" as their answers that consists of 3 out of 100 respondents.
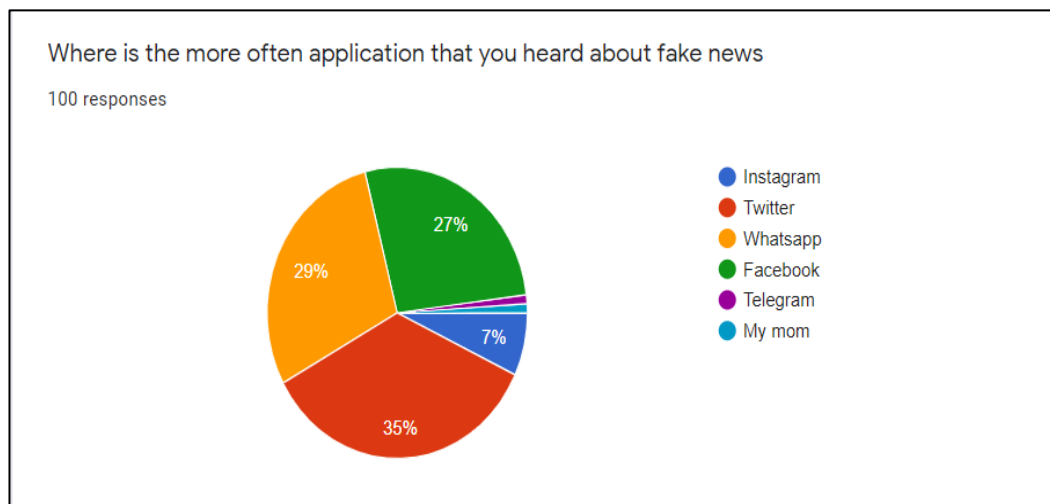
For the question 5, the author asks the respondents where the most often application that the respondents heard about fake news. Mostly the respondents chose "Twitter" as their answers which is around 35% of respondents. This can relate to the question 1 which is the age of the respondents that mostly on younger generation. The least pick answer is around 1% which is from "Telegram" and "others- my mom". The others option such as "Facebook" is around 27%, "WhatsApp" is around 29% and the last one is "Instagram" which is 7%.
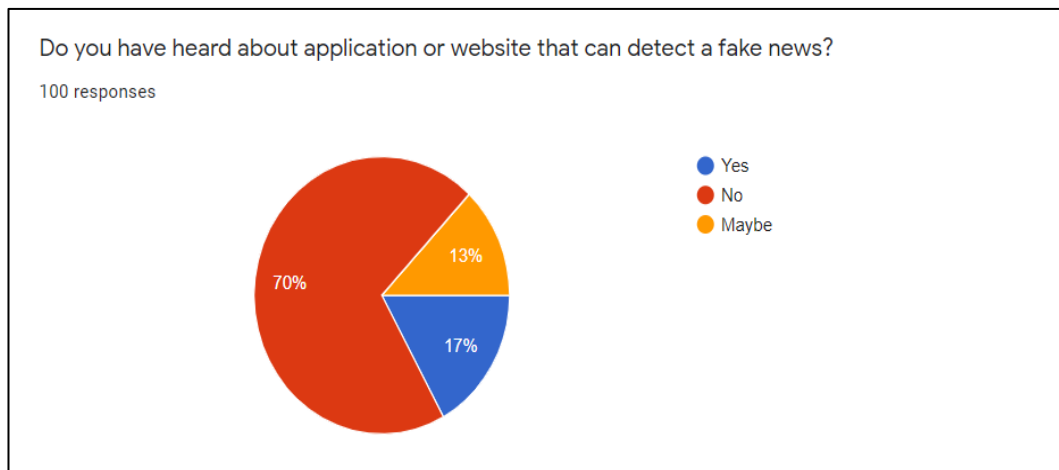
Figure 14: Survey Result Question 6

For the next question, it is about the application or website that can detect a fake news. As a results from this survey, 70 out of 100 respondents chose no that brings to the result mostly they do not hear any kind of the application or website that can detect a fake news. However, around 17 respondents chose yes which is they know the existence of the application and 13% choose "maybe". This question brings the next question which is if the application is will helpful them to detect the fake news.

Do you think it will be helpful if this application can detect fake news?
100 responses

- Yes
- No
- Maybe

90%

10%

Figure 15: Survey Result Question 7

For the last question, mostly the respondents chose "Yes" as their answer which is around 90% that consists 90 out of 100 respondents while only 10 respondents that pick "maybe". No one of the respondents chose "No" as their answer that bring to the mostly all of them agreed that the application will be helpful to them.

4.2 Fake News Covid-19 Detector

In order to run the systems, it is need 2 set of datasets which is "fake" dataset and "real" dataset. In the dataset, it is consisting of title and text of the news.

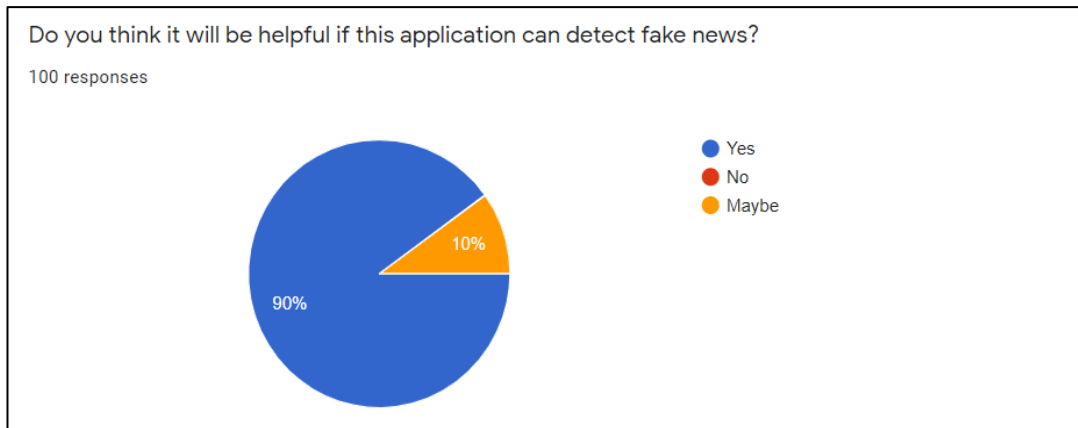|	A	B
1	title	text
2	FACEBOOK DELETES MICHIGAN ANTI-LOCKDOWN GROUP WITH 380,000 MEMBERS	Facebook has shuttered a popular group for Michiganders who oppose their governor's extreme lockdown measures, fueling debate about free speech during the corona
3	Purdue University says students face 'close to zero lethal threat' from COVID-19 as the school plans to re	Purdue University President Mitch Daniels, the former governor of Indiana, said in an April 21 letter he intends to open the West Lafayette, Indiana university for in-perso
4	THE HIGH COST OF LOCKING DOWN AMERICA: "WE'VE SEEN A YEAR'S WORTH OF SUICIDE ATTEMPTS	Locking down much of the country may have helped to "flatten the curve" a bit, but it has also had some very serious consequences that public health officials did not ant
5	NY TIMES WORRIES HALF OF AMERICANS WILL REFUSE TO TAKE CORONAVIRUS VACCINE	The New York Times has published an article by popular author Kevin Roose in which he expresses his concern that half of Americans would refuse to take a coronavirus v
6	CDC adds 6 new possible symptoms of Covid-19	When the coronavirus outbreak began, the Centers for Disease Control and Prevention listed cough, shortness of breath and fever as possible symptoms. Now, the organi
7	Purdue University says students face 'close to zero lethal threat' from COVID-19 as the school plans to re	General view of the campus of Purdue University in October 2018 in West Lafayette, Indiana. Michael Hickey/Getty Images Purdue University President Mitch Daniels sa
8	TRUMP: PELOSI IS A 'SICK WOMAN' WITH A LOT OF 'MENTAL PROBLEMS'	President Trump slammed Nancy Pelosi Tuesday, calling her a "sick woman" with a lot of "mental problems" after Pelosi said Trump should fear for his life taking hydroxy
9	TWO SEPARATE REPORTS CONFIRM 'MOST INTEL AGENCIES' IN US BELIEVE CORONAVIRUS CAME FRO	Two separate reports cited a senior intelligence official as confirming that most of the 17 US intelligence agencies believe that the coronavirus originated in the Wuhan
10	UNIVERSITY GIVING CORONAVIRUS AID MONEY TO ILLEGALS	After the recent Department of Education guidelines clarified that federal funds through the Coronavirus Aid, Relief, and Economic Security (CARES) Act should not be dis
11	NFL 'PREPARED TO MAKE ADJUSTMENTS' IF CALIFORNIA LOCKDOWN EXTENDS INTO SEASON	The NFL has three teams in California, and they're keeping open the possibility of moving them if California's lockdown continues into the regular season. The question of v
12	'Fewer Covid-19 cases in hotter regions'	'Fewer Covid-19 cases in hotter regions' U Tejonmayam	TNN	Apr 26, 2020, 06:25 IST Researchers at IIT-M found that 90% of the total confirmed cases were recorded
13	280 govt labs conducting Covid-19 tests, reporting to ICMR	The Indian Council of Medical Research on Sunday said that a total of 280 operational government laboratories in the country are conducting tests for Covid-19 and repo
14	Colorado governor says he's worried about possible second spike in Covid-19 cases as state set to reopen	Colorado Gov. Jared Polis on Sunday said he is worried about the potential for a second spike in coronavirus cases as the state begins loosening restrictions this week. The
15	We are genetically stronger to take on Covid-19, says scientist	We are genetically stronger to take on Covid-19, says scientist Sukshma Ramakrishnan	TNN	Apr 27, 2020, 12:48 IST The genetic make-up and food habits of Southeast
16	NYT STORY THAT CLAIMS US OFFICIALS 'PRESSURED' SPIES TO LINK LAB TO CORONAVIRUS IS BASED O	A New York Times story published Thursday claimed that Trump officials have unduly pressured intelligence agencies to link the Wuhan bio-lab to the coronavirus outbrea
17	Two more COVID-19 positive cases in Uttarakhand, total tally 50,	Chennai: DEHRADUN: Two more Covid-19 positive cases surfaced on Monday including a nurse from All India Institute of Medical Sciences, Rishikesh.
18	Covid-19: Backyard cinema lights up lockdown gloom for Berliners	Covid-19: Backyard cinema lights up lockdown gloom for Berliners Posted on by admin
19	Assam Covid-19 recovery rate 77%, but first patient still positive	Assam doctors cured and discharged eight more Covid-19 patients from hospitals on Sunday, taking the state's recovery rate over 75% against 21.09% for the country, bu
20	CDC RUSHES OUT PANDEMIC GUIDANCE FOR RELIGIOUS WORSHIP AFTER SURPRISE TRUMP ANNOUN	After weeks of butting heads with the Trump administration over pandemic guidance, the US Centers for Disease Control (CDC) issued guidance on Friday for the reopeni
21	Work on Telugu films with Rs 600 Cr investment stalled due to COVID-19 lockdown	Hyderabad, April 27 (): About 50 Telugu movies invarious stages of making with an investment of approximatelyRs 600 crore have been stalled due to the ongoing lockdo
22	BRONX ER DOCTOR EXPLAINS WHY REOPENING THE ECONOMY NOW MIGHT ACTUALLY SAVE LIVES	Last night, after we reported on a pair of ER doctors arguing that the US states should lift their lockdowns immediately (during the course of their argument, which was r
23	COVID-19 cases in Maha cross 8,000-mark; death toll now 342	The number of COVID-19 cases in Maharashtra crossed the 8,000-mark on Sunday after 440 more people were found infected, a health official said.
24	WOMAN SENTENCED TO WEEK IN JAIL FOR REFUSING TO APOLOGIZE FOR REOPENING SALON	A Texas woman was sentenced to seven days in jail for refusing to apologize for reopening her hair salon in defiance of local lockdown orders.
25	STUDY: LOCKDOWN "WILL DESTROY AT LEAST SEVEN TIMES MORE YEARS OF HUMAN LIFE" THAN IT S	A study has found that the "stay at home" lockdown order in the United States will "destroy at least seven times more years of human life" than it saves and that this nun
26	Vietnam's COVID-19 test kit recognized by WHO - Xinhua	Vietnam's COVID-19 test kit recognized by WHO Source.
27	Covid-19: Impact on the banking sector	Blanket policies announced by the government, intended as relief measures, may do more harm than good. Shutterstock Published at : April 26, 2020 Updated at : April 2
28	SLIPPERY SLOPE TO DESPOTISM: PAVED WITH LOCKDOWNS, RAIDS, AND FORCED VACCINATIONS	"You have no right not to be vaccinated, you have no right not to wear a mask, you have no right to open up your business… And if you refuse to be vaccinated, the state
29	OPERATION HAIRCUT: HAIRSTYLISTS TO PROTEST GOV. WHITMER STRIPPING BARBER'S LICENSE FOR	A conservative group is organizing a civil disobedience demonstration in Michigan after Democrat Governor Gretchen Whitmer stripped a barber of his license for trying
30	TUCKER CARLSON: "THE VIRUS ISN'T NEARLY AS DEADLY AS WE THOUGHT IT WAS"	Fox News host Tucker Carlson appeared to join those demanding a move to re-open society as soon as possible when he said during his show last night "the virus isn't nea
31	MEDIA, LEFTISTS FLIP OUT AFTER TRUMP SAYS HE'S TAKING HYDROXYCHLOROQUINE	The media spent all night in a frothing fit because President Trump told reporters that he is taking the anti-malaria drug hydroxchloroquine, as a precaution against cor
32	WHO CLAIMS U.S. AND BRAZIL IGNORED WARNINGS ABOUT CORONAVIRUS	Despite amplifying Chinese propaganda that there was no "human to human" transmission of coronavirus on January 14, a WHO spokesperson today claimed that the U.
33	Drug-gene testing could give experts insight into COVID-19 treatment	Print Thousands of clinical trials have launched in an effort to better understand the SARS-CoV-2 virus and in search of drugs to successfully treat the symptoms of infectio
34	JOE BIDEN FORGETS WORD FOR 'CORONAVIRUS', LOSES TRAIN OF THOUGHT	Democrat presidential frontrunner Joe Biden again stumbled through a television interview that aired Saturday.
35	TRUMP TO SUSPEND ALL IMMIGRATION INTO US OVER CORONAVIRUS	US President Donald Trump said he will sign an executive order to temporarily suspend immigration into the United States, saying he needs to protect American jobs as c
36	VIDEO: CROWD CHEERS AFTER POLICE REFUSE TO ENFORCE NEW JERSEY LOCKDOWN ORDER AT GYM	A New Jersey gym reopened in defiance of Governor Chris Murphy's shutdown order, and patrons were waiting for the state to respond with a heavy hand.
37	Trumpian Reciprocity in the Era of Covid-19 Views	In the trumpeyed view of the world, it's OK for the United States to send people infected with Covid-19 into other countries, but it's not OK for people from foreign count

Figure 16: "Fake" Dataset

title	text
Other Viewpoints: COVID-19 is worse than the flu	We can now officially put to rest all comparisons of COVID-19 to influenza. Coronavirus is killing more Americans than the flu and almost everything else. According to data con
Bermuda's COVID-19 cases surpass 100	The Ministry of Health in Bermuda has confirmed a newspaper report that 10 residents from the Westmeath Residential and Nursing Home had tested positive for COVID-19, ta
Delhi: Eight nurses test positive for Covid-19 at Kalawati Sara	Delhi: Eight nurses test positive for Covid-19 at Kalawati Saran Hospital. Another nurse at the Kalawati Saran Hospital in central Delhi has tested positive for Covid-19, taking to
Mississippi man recovering at home after 21 days on a ventila	A Pearl River County man is recovering at home after spending three weeks on a ventilator due to COVID-19. Doctors didn't believe there was much chance for survival, but the
Eight nurses test positive for Covid-19 at Kalawati Saran Hosp	Another nurse at the Kalawati Saran Hospital in central Delhi has tested positive for Covid-19, taking to eight the total number of nurses there affected by the highly infectious d
Singapore races to build beds for COVID-19 patients as cases	Singapore is rapidly building bed space for coronavirus patients in cavernous exhibition halls and other temporary facilities as it faces a surge in cases, mainly among its large cor
Coronavirus symptoms: CDC adds six symptoms for COVID-15	Chills, repeated shaking with chills, muscle pain, headache, sore throat and a loss of taste or smell. Those are the six new symptoms the Centers for Disease Control and Prevent
Africa's COVID-19 confirmed cases surpass 30,000, death toll	Africa's COVID-19 confirmed cases surpass 30,000, death toll hits 1,374: Africa CDC Source: Xinhua	2020-04-26 23:08:28	Editor: huaxia Video Player Close
In Detroit, grief runs deep as city grapples with COVID-19 - Sa	Close Image 1 of 3 Jamon Jordan poses with a photo of his mother, Jacquelynne Jordan, in Detroit, Friday, April 24, 2020. Jamon Jordan could not mourn his mother in the tradit
Coronavirus: Minks test positive for COVID-19 at two Dutch f:	Two Dutch mink farms have reported cases of COVID-19 among their animals, the country's Ministry of Agriculture confirmed on Sunday.
Odisha: 3 new Covid-19 cases in Sundargarh, state tally 103	3 new Covid-19 cases in Sundargarh, state tally 103 Minati Singha	TNN	Apr 26, 2020, 22:18 IST Representative image. BHUBANESWAR: With three new positive cases detecte
Gov. Kemp, Walmart, eTrueNorth announce mobile COVID-15	On Sunday, Gov. Brian Kemp announced the launch of a mobile COVID-19 testing initiative to set up drive-thru sites in communities without access to testing.
COVID-19: County releases three-step plan to protect public h	Norfolk County has drafted a three-step plan for how to protect public health during COVID-19.
Coronavirus alert: 12 Fresh New Covid-19 Cases in Karnataka	
Covid 19 coronavirus: Concerns for senior students and their	A high school principal has raised concerns about what prolonging senior students from returning to school will do to their learning and mental wellbeing. It comes after Prime N
People donate RVs to front-line workers who fear of bringing	As COVID-19 brings unprecedented challenges to the world, the West Michigan community responded with unprecedented kindness.
Health Ministry confirms two more cases of Covid-19	Two 50-year-old men from Birgunj tested positive for the virus on Sunday
English River FN makes 'urgent call for support' from province	As the COVID-19 outbreak continues to spread in the Far North region of Saskatchewan, the English River First Nation "is making an urgent call for support to help control and e
Three new COVID-19 cases in Karnataka in last 24 hours, lowe	A 45-year-old woman, who was categorised as a SARI patient, succumbed to COVID-19 on Sunday in Bengaluru.
Military doctors receive COVID-19 aid from China	The Ministry of Defence on Saturday received 16 tonnes of medical supplies from its Chinese counterpart. The aid comes following China's previously held technical training and
Armenian community donates over 5 million meals to familie:	The goal was to donate 1.5 million meals to help families struggling during the COVID-19 pandemic. But thanks to the Armenian American community over 5 million meals have
China donates COVID-19 medical supplies to Rwanda	China donates COVID-19 medical supplies to Rwanda By linkedin
Amazon withdraws from MWC for corona virus reasons	Amazon withdraws from MWC for corona virus reasons – TechCrunch Amazon withdraws from MWC for corona virus reasons – TechCrunch Google + Amazon is the latest com
AntiCancer Inc. to Target Its Oral Recombinant Methioninase	Currently there are no therapies or vaccines for patients infected with 2019-nCoV coronavirus around the world. As the number of 2019-nCoV patients approaches 10,000 and r
Corona virus DETECTED in MUMBAI,INDIA	The Maharashtra government is mulling to set up a coronavirus testing facility at Kasturba Hospital here to reduce the burden on Pune-based National Institute of Virology (NIV
CBCP discourages holding of hands during 'Our Father' prayer	"In praying we invite ourselves with all our brothers and sisters suffering with the disease brought by this virus, bring up to God our longing for them to be restored to full health
The coronavirus has nothing to do with Corona beer. But, son	As new coronavirus outbreak continues to spread, some people seem to think that the deadly outbreak is somehow related to Corona beer. It's not.

Figure 17: "Real" Dataset

25

From the dataset above, the author imports the dataset into the system. Not only that, but the author also imports the library that needed into the system.

Importing required library

Here I am going to importing some of the required library, if extra library is required to install It will be install later on.

```python
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
```

Figure 18: Import the required Library

Inserting fake and real dataset

```python
In [2]: df_fake = pd.read_csv('C:/Users/win10/Documents/FYP/Fake_news/Fake.csv',encoding='latin1')
        df_true = pd.read_csv('C:/Users/win10/Documents/FYP/Fake_news/True.csv',encoding='latin1')
```

Figure 19: Inserting the Dataset

```python
In [3]: df_fake.head(10)
```

Out[3]:

	title	text
0	FACEBOOK DELETES MICHIGAN ANTI-LOCKDOWN GROUP ...	Facebook has shuttered a popular group for Mic...
1	Purdue University says students face 'close to...	Purdue University President Mitch Daniels, the...
2	THE HIGH COST OF LOCKING DOWN AMERICA: □WE□VE ...	Locking down much of the country may have help...
3	NY TIMES WORRIES HALF OF AMERICANS WILL REFUSE...	The New York Times has published an article by...
4	CDC adds 6 new possible symptoms of Covid-19	When the coronavirus outbreak began, the Cente...
5	Purdue University says students face 'close to...	General view of the campus of Purdue Universit...
6	TRUMP: PELOSI IS A □SICK WOMAN□ WITH A LOT OF ...	President Trump slammed Nancy Pelosi Tuesday, ...
7	TWO SEPARATE REPORTS CONFIRM □MOST INTEL AGENC...	Two separate reports cited a senior intelligen...
8	UNIVERSITY GIVING CORONAVIRUS AID MONEY TO ILL...	After the recent Department of Education guide...
9	NFL □PREPARED TO MAKE ADJUSTMENTS□ IF CALIFORN...	The NFL has three teams in California, and the...

Figure 20: df_fake.Head

```
In [4]: df_true.head(10)
```

Out[4]:

	title	text
0	Other Viewpoints: COVID-19 is worse than the flu	We can now officially put to rest all comparis...
1	Bermuda's COVID-19 cases surpass 100	The Ministry of Health in Bermuda has confirme...
2	Delhi: Eight nurses test positive for Covid-19...	Delhi: Eight nurses test positive for Covid-19...
3	Mississippi man recovering at home after 21 da...	A Pearl River County man is recovering at home...
4	Eight nurses test positive for Covid-19 at Kal...	Another nurse at the Kalawati Saran Hospital i...
5	Singapore races to build beds for COVID-19 pat...	Singapore is rapidly building bed space for co...
6	Coronavirus symptoms: CDC adds six symptoms fo...	Chills, repeated shaking with chills, muscle p...
7	Africa's COVID-19 confirmed cases surpass 30,0...	Africa's COVID-19 confirmed cases surpass 30,0...
8	In Detroit, grief runs deep as city grapples w...	Close Image 1 of 3 Jamon Jordan poses with a p...
9	Coronavirus: Minks test positive for COVID-19 ...	Two Dutch mink farms have reported cases of CO...

Figure 21: df_true.Head

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
```

Figure 22: Class of dataset

Based on figure 22, it is show that the author inserting a column that called "class" for the fake and real news dataset. For the fake dataset, it is class df_fake (0) while for the real dataset, it is class of df_true (1).

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

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

Out[6]: ((1058, 3), (2061, 3))

```
In [7]: df_fake_manual_testing = df_fake.tail(10)
        for i in range(1057,1047,-1):
            df_fake.drop([i], axis = 0, inplace = True)
        df_true_manual_testing = df_true.tail(10)
        for i in range(2060,2050,-1):
            df_true.drop([i], axis = 0, inplace = True)
```

Figure 23: Remove 10 Rows for Manual Testing

27

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

```
In [8]: df_manual_testing = pd.concat([df_fake_manual_testing,df_true_manual_testing], axis = 0)
        df_manual_testing.to_csv('C:/Users/win10/Documents/FYP/Fake_news/manual_testing.csv')
```

Figure 24: Merging the Manual Testing

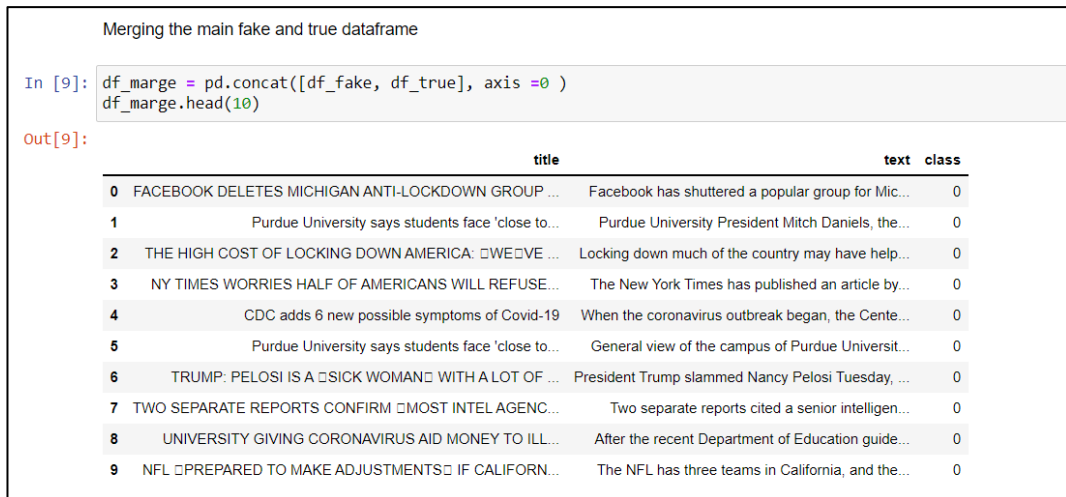| A	B	C	D
	title	text	class
1048	JUST IN!!! Corona Virus Spotted In Owe	JUST IN!!! Corona Virus	0
1049	Corona Virus Uncertainty Clouds Mobile	Corona Virus Uncertainty	0
1050	Corona Virus Â– The African connection	Like every issue of	0
1051	Was the Wuhan corona-virus made in a	It is too early to tell. We do	0
1052	President backs GoÂ's Balik-Probinsya p	President Rodrigo Duterte is	0
1053	Twinbrook Â'pausingÂ' COVID-19 plan a	Apr 24, 2020 at 4:31 PM	0
1054	Experts suggest ventilation, not isolatio	The Coronavirus epidemic for	0
1055	Corona virus and pentagon Quarantine	Here is the technique we use	0
1056	Characterizing Patients Hospitalized Wit	Characterizing Patients	0
1057	Corona Virus WARNING	Just something to be prepared	0
2051	Corona the beer is being linked with cor	Home Offbeat Corona the	1
2052	San Diego Declares Local Public Health	San Diego County officials say	1
2053	China travel ban 'vital' in preventing 201	Imposing a travel ban to and	1
2054	Madhya Pradesh: 'Suspected' corona vir	A 30-year-old patient who	1
2055	Cases show incubation period of 2019-r	Fresh evidence is suggesting tha	1
2056	CDC Outlines US Process of Evaluating F	Recognizing individuals at risk	1
2057	Corona virus disrupting business and tra	Corona virus disrupting	1
2058	2019-nCoV: Health dept. on full alert	Quarantine advised for 80	1
2059	Screening machine for corona virus laur	South Sudan has begun	1
2060	Coronavirus (2019-nCoV)	Coronavirus (2019-nCoV) is a	1

Figure 25: Manual Testing Dataset

Merging the main fake and true dataframe

```
In [9]: df_marge = pd.concat([df_fake, df_true], axis =0 )
        df_marge.head(10)
```

Out[9]:

	title	text	class
0	FACEBOOK DELETES MICHIGAN ANTI-LOCKDOWN GROUP ...	Facebook has shuttered a popular group for Mic...	0
1	Purdue University says students face 'close to...	Purdue University President Mitch Daniels, the...	0
2	THE HIGH COST OF LOCKING DOWN AMERICA: □WE□VE ...	Locking down much of the country may have help...	0
3	NY TIMES WORRIES HALF OF AMERICANS WILL REFUSE...	The New York Times has published an article by...	0
4	CDC adds 6 new possible symptoms of Covid-19	When the coronavirus outbreak began, the Cente...	0
5	Purdue University says students face 'close to...	General view of the campus of Purdue Universit...	0
6	TRUMP: PELOSI IS A □SICK WOMAN□ WITH A LOT OF ...	President Trump slammed Nancy Pelosi Tuesday, ...	0
7	TWO SEPARATE REPORTS CONFIRM □MOST INTEL AGENC...	Two separate reports cited a senior intelligen...	0
8	UNIVERSITY GIVING CORONAVIRUS AID MONEY TO ILL...	After the recent Department of Education guide...	0
9	NFL □PREPARED TO MAKE ADJUSTMENTS□ IF CALIFORN...	The NFL has three teams in California, and the...	0

Figure 26: Merging the main dataset

Randomly shuffling the dataframe

```
In [13]: df = df.sample(frac = 1)
```

```
In [14]: df.head()
```

Out[14]:

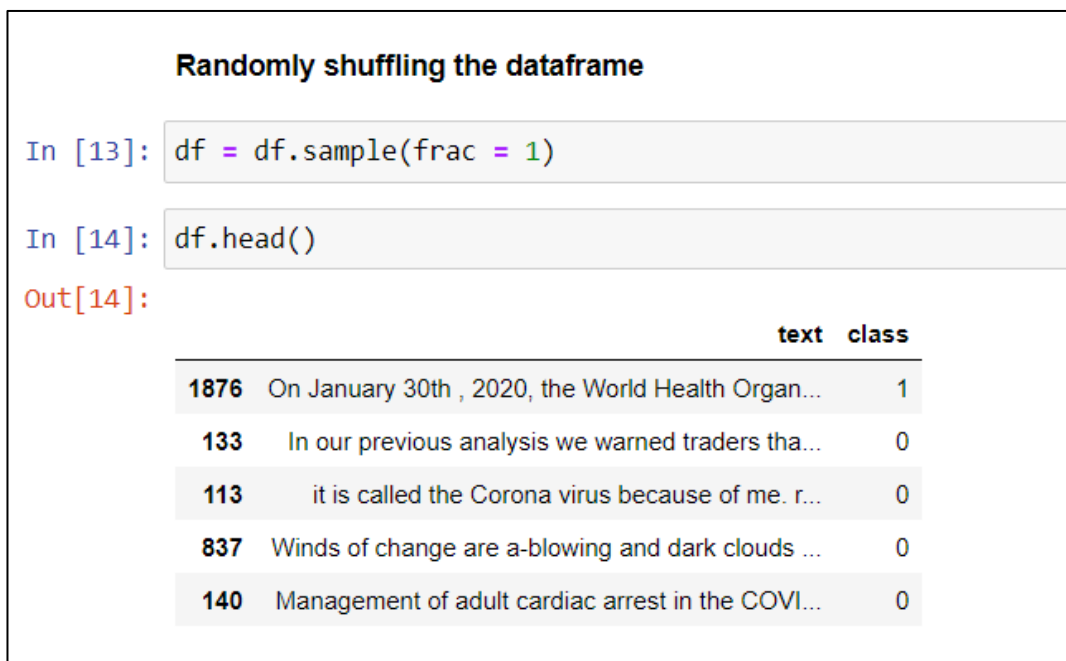|	text	class
1876	On January 30th , 2020, the World Health Organ...	1
133	In our previous analysis we warned traders tha...	0
113	it is called the Corona virus because of me. r...	0
837	Winds of change are a-blowing and dark clouds ...	0
140	Management of adult cardiac arrest in the COVI...	0

Figure 27: Randomly Shuffling the Data frame

29

Model Testing With Manual Entry

News

```
In [44]: def output_lable(n):
             if n == 0:
                 return "Fake News"
             elif n == 1:
                 return "True News"


         def manual_testing(news):
             testing_news = {"text":[news]}
             new_def_test = pd.DataFrame(testing_news)
             new_def_test["text"] = new_def_test["text"].apply(wordopt)
             new_x_test = new_def_test["text"]
             new_xv_test = 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("\n\nLR Prediction: {} \nDT Prediction: {} \nGBC Prediction: {} \nRFC Prediction: {}".format(output_lable(pred_L
                                                                                      output_lable(pred_D
                                                                                      output_lable(pred_G
                                                                                      output_lable(pred_R
```

Figure 28: Model Testing with Manual Entry

```
In [45]: news = str(input())
         manual_testing(news)

         FACEBOOK DELETES MICHIGAN ANTI-LOCKDOWN GROUP WITH 380,000 MEMBERS


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

Figure 29: Manual Testing using "Title"

```
In [46]: news = str(input())
         manual_testing(news)

         Because we have no clear picture as to when theproductions will be resumed, when the movies will be taken tomasses.


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

Figure 30: Manual Testing using some of "Text"

30

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

4.1 Conclusion

In conclusion, the primary goal of this project is to create an input system called CovFaq that is capable of storing multiple COVID19-confirmed false news in a database. According to the research that I conducted during FYP1, the system has not yet been deployed in Malaysia, since I found no evidence of the COVID-19 Fake News Detector's existence. As a result, there is a chance for me to design a system to assist authorities during the COVID-19 epidemic. COVID-19 Fake News Detector might assist authorities in ensuring that everyone adheres to the norms established for the dissemination of false news.

4.2 Recommendations

For the recommendations the author suggests to:

- Fake News Detector into language which are English and Bahasa. This is because the one that the author implement is in the English language so the author recommend doing in Bahasa since not all the users can understand English.

- The author also recommend the system to autosave the database daily.

REFERENCES

Coronavirus disease (COVID-19): Masks. (n.d.). Retrieved March 30, 2021, from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question- and-answers-hub/q-a-detail/coronavirus-disease-covid-19-masks

WHO, "Coronavirus disease (COVID-19) advice for the public", 2020, Available: https://www.who.int/emergencies/diseases/ novel-coronavirus-2019/advice-for-public

SCRUM model for agile methodology. IEEE Xplore. (n.d.). https://ieeexplore.ieee.org/abstract/document/8229928.

van der Linden, S., Roozenbeek, J., & Compton, J. (1AD, January 1). *Inoculating Against Fake News About COVID-19*. Frontiers. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.566790/full.

Apuke, O. D., & Omar, B. (2020, July 30). *Fake news and COVID-19: modelling the predictors of fake news sharing among social media users*. Telematics and Informatics. https://www.sciencedirect.com/science/article/pii/S0736585320301349.

World Health Organization. (n.d.). *COVID-19 Mythbusters*. World Health Organization. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters.

Zarocostas, J. (2020). How to fight an infodemic. *Lancet* 395:676. doi: 10.1016/S0140-6736(20)30461-X

What is fake news and how to stop the spread of misinformation? Settlement.Org | Information Newcomers Can Trust. (n.d.). https://settlement.org/ontario/daily-life/communication/ethnic-and-cultural-media/what-is-fake-news-and-how-to-stop-spreading-misinformation/.

What Is Machine Learning: Definition, Types, Applications and Examples. Potentia Analytics. (2019, December 19). https://www.potentiaco.com/what-is-machine-learning-definition-types-applications-and-examples/.

What are popular platforms doing to stop the spread of fake news online? Internet Matters. (2020, November 18). https://www.internetmatters.org/hub/news-blogs/stopping-the-spread-of-fake-news-on-popular-online-platforms/.

Ofcom (2020). *Half of UK Adults Exposed to False Claims about Coronavirus.* Available online at: https://www.ofcom.org.uk/about-ofcom/latest/features-and-news/half-of-uk-adults-exposed-to-false-claims-about-coronavirus

Mitchell, A., and Oliphant, J. B. (2020). *Americans Immersed in COVID-19 News; Most Think Media Are Doing Fairly Well Covering It. Pew Research Center.* Available online at: https://www.journalism.org/2020/03/18/americans-immersed-in-covid-19-news-most-think-media-are-doing-fairly-well-covering-it/

G. Rampersad, T. Althiyabi, G. Warner-Søderholm, A. Bertsch, E. Sawe, D.D. Lee, T. Wolfe, J. Meyer, J. Engel, U.N. Fatilua, N. Rusmania, Y. Jiang, D.I. Bolnick, M. Kirkpatrick, A.P. Gyaisey, A. Afful-dadzie, R. Boateng, O. Boyd-Barrett, M. Perugini, R.-M. Hage Birds of a Feather: Homophily in Social Networks Computers in Human Behavior, 9 (1) (2019), pp. 1-9, 10.1126/science.aao2998

Lampos, V., Moura, S., Yom-Tov, E., Cox, I. J., McKendry, R., & Edelstein, M. (2020). Tracking COVID-19 using online search. 93, 4–9. http://arxiv.org/abs/2003.08086.

Hou, Z., Du, F., Jiang, H., Zhou, X., Lin, L., Assessment, T., & Commission, N. H. (2020). Assessment of public attention, risk perception, emotional and behavioural responses to the COVID-19 outbreak social media surveillance in China. Doi: 10.1101/2020.03.14.20035956.

T.L.D. Huynh **the COVID-19 risk perception: A survey on socioeconomics and media attention** Economics Bulletin, 40 (1) (2020), pp. 758-764

Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020, October). October 2020Complexity 2020(5):1-11 DOI:10.1155/2020/8885861 Authors: Iftikhar Ahmad Muhammad Yousaf Institute of Management Sciences, Pakistan, Peshawar Suhail Yousaf Vrije Universiteit Amsterdam Muhammad Ovais Ahmad University of Oulu. https://www.researchgate.net/publication/346262009_FakeMuhammad Yousaf_News_Detection_Using_Machine_Learning_Ensemble_Methods.

Khanam1, Z., Alwasel1, B. N., Sirafi1, H., & Rashid2, M. (2021, March 1). *IOPscience*. IOP Conference Series: Materials Science and Engineering. https://iopscience.iop.org/article/10.1088/1757-899X/1099/1/012040.

APPENDICES

- Coding

localhost:8888/notebooks/Documents/FYP/Fake_news/Fake_news_detection.ipynb#

jupyter Fake_news_detection Last Checkpoint: 11 hours ago (autosaved)

Logout

File   Edit   View   Insert   Cell   Kernel   Widgets   Help

Trusted | Python 3 (ipykernel)

Code

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

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

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

```
Out[6]: ((1058, 3), (2061, 3))
```

```python
In [7]: df_fake_manual_testing = df_fake.tail(10)
        for i in range(1057,1047,-1):
            df_fake.drop([i], axis = 0, inplace = True)
        df_true_manual_testing = df_true.tail(10)
        for i in range(2060,2050,-1):
            df_true.drop([i], axis = 0, inplace = True)
```

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

```python
In [8]: df_manual_testing = pd.concat([df_fake_manual_testing,df_true_manual_testing], axis = 0)
        df_manual_testing.to_csv('C:/Users/win10/Documents/FYP/Fake_news/manual_testing.csv')
```

Merging the main fake and true dataframe

```python
In [9]: df_marge = pd.concat([df_fake, df_true], axis =0 )
        df_marge.head(10)
```

Out[9]:

	title	text	class
0	FACEBOOK DELETES MICHIGAN ANTI-LOCKDOWN GROUP	Facebook has shuttered a popular group for Mic...	0
1	Purdue University says students face 'close to...	Purdue University President Mitch Daniels, the...	0
2	THE HIGH COST OF LOCKING DOWN AMERICA: ⊏WE⊏VE	Locking down much of the country may have help...	0

NY TIMES WORRIES HALF OF AMERICANS WILL REFUSE   The New York Times has published an article by...

24°C Cloudy   ENG 4:05 AM

---

localhost:8888/notebooks/Documents/FYP/Fake_news/Fake_news_detection.ipynb#

jupyter Fake_news_detection Last Checkpoint: 11 hours ago (autosaved)

Logout

File   Edit   View   Insert   Cell   Kernel   Widgets   Help

Trusted | Python 3 (ipykernel)

Code

```python
In [9]: df_marge = pd.concat([df_fake, df_true], axis =0 )
        df_marge.head(10)
```

Out[9]:

	title	text	class
0	FACEBOOK DELETES MICHIGAN ANTI-LOCKDOWN GROUP ...	Facebook has shuttered a popular group for Mic...	0
1	Purdue University says students face 'close to...	Purdue University President Mitch Daniels, the...	0
2	THE HIGH COST OF LOCKING DOWN AMERICA: ⊏WE⊏VE ...	Locking down much of the country may have help...	0
3	NY TIMES WORRIES HALF OF AMERICANS WILL REFUSE...	The New York Times has published an article by...	0
4	CDC adds 6 new possible symptoms of Covid-19	When the coronavirus outbreak began, the Cente...	0
5	Purdue University says students face 'close to...	General view of the campus of Purdue Universit...	0
6	TRUMP: PELOSII IS A ⊏SICK WOMAN⊏ WITH A LOT OF ...	President Trump slammed Nancy Pelosi Tuesday, ...	0
7	TWO SEPARATE REPORTS CONFIRM ⊏MOST INTEL AGENC...	Two separate reports cited a senior intelligen...	0
8	UNIVERSITY GIVING CORONAVIRUS AID MONEY TO ILL...	After the recent Department of Education guide...	0
9	NFL ⊏PREPARED TO MAKE ADJUSTMENTS⊏ IF CALIFORN...	The NFL has three teams in California, and the...	0

```python
In [10]: df_marge.columns
```

```
Out[10]: Index(['title', 'text', 'class'], dtype='object')
```

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

```python
In [11]: df = df_marge.drop(["title"], axis = 1)
```

```python
In [12]: df.isnull().sum()
```

```
Out[12]: text     0
         class    0
         dtype: int64
```

24°C Cloudy   ENG 4:06 AM

36

Randomly shuffling the dataframe

```
In [13]: df = df.sample(frac = 1)
```

```
In [14]: df.head()
```

Out[14]:

	text	class
1876	On January 30th , 2020, the World Health Organ...	1
133	In our previous analysis we warned traders tha...	0
113	it is called the Corona virus because of me. r...	0
837	Winds of change are a-blowing and dark clouds ...	0
140	Management of adult cardiac arrest in the COVI...	0

```
In [15]: df.reset_index(inplace = True)
         df.drop(["index"], axis = 1, inplace = True)
```

```
In [16]: df.columns
```

Out[16]: Index(['text', 'class'], dtype='object')

```
In [17]: df.head()
```

Out[17]:

	text	class
0	On January 30th , 2020, the World Health Organ...	1
1	In our previous analysis we warned traders tha...	0
2	it is called the Corona virus because of me. r...	0
3	Winds of change are a-blowing and dark clouds ...	0
4	Management of adult cardiac arrest in the COVI...	0

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

```
In [18]: def wordopt(text):
             text = text.lower()
             text = re.sub('\[.-?\]', '', text)
             text = re.sub("\\W"," ",text)
             text = re.sub('https?://\S+|www\.\S+', '', text)
             text = re.sub('<.*?>+', '', text)
             text = re.sub('[%s]' % re.escape(string.punctuation), '', text)
             text = re.sub('\n', '', text)
             text = re.sub('\w*\d\w*', '', text)
             return text
```

```
In [19]: df["text"] = df["text"].apply(wordopt)
```

Defining dependent and independent variable as x and y

```
In [20]: x = df["text"]
         y = df["class"]
```

Splitting the dataset into training set and testing set.

```
In [21]: x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25)
```

Convert text to vectors

```
In [22]: from sklearn.feature_extraction.text import TfidfVectorizer
```

```
In [23]: vectorization = TfidfVectorizer()
         xv_train = vectorization.fit_transform(x_train)
         xv_test = vectorization.transform(x_test)
```

1. Logistic Regression

```
In [24]: from sklearn.linear_model import LogisticRegression
```

```
In [25]: LR = LogisticRegression()
         LR.fit(xv_train,y_train)
```

```
Out[25]: LogisticRegression()
```

```
In [26]: pred_lr=LR.predict(xv_test)
```

```
In [27]: LR.score(xv_test, y_test)
```

```
Out[27]: 0.8141935483870968
```

```
In [28]: print(classification_report(y_test, pred_lr))

                       precision    recall  f1-score   support

                   0       0.82      0.54      0.65       249
                   1       0.81      0.94      0.87       526

            accuracy                           0.81       775
           macro avg       0.82      0.74      0.76       775
        weighted avg       0.81      0.81      0.80       775
```

2. Decision Tree Classification

```
In [29]: from sklearn.tree import DecisionTreeClassifier
```

```
In [30]: DT = DecisionTreeClassifier()
         DT.fit(xv_train, y_train)
```

---

2. Decision Tree Classification

```
In [29]: from sklearn.tree import DecisionTreeClassifier
```

```
In [30]: DT = DecisionTreeClassifier()
         DT.fit(xv_train, y_train)
```

```
Out[30]: DecisionTreeClassifier()
```

```
In [31]: pred_dt = DT.predict(xv_test)
```

```
In [32]: DT.score(xv_test, y_test)
```

```
Out[32]: 0.712258064516129
```

```
In [33]: print(classification_report(y_test, pred_dt))

                       precision    recall  f1-score   support

                   0       0.55      0.59      0.57       249
                   1       0.80      0.77      0.78       526

            accuracy                           0.71       775
           macro avg       0.67      0.68      0.68       775
        weighted avg       0.72      0.71      0.71       775
```

3. Gradient Boosting Classifier

```
In [34]: from sklearn.ensemble import GradientBoostingClassifier
```

```
In [35]: GBC = GradientBoostingClassifier(random_state=0)
         GBC.fit(xv_train, y_train)
```

3. Gradient Boosting Classifier

```
In [34]: from sklearn.ensemble import GradientBoostingClassifier
```

```
In [35]: GBC = GradientBoostingClassifier(random_state=0)
         GBC.fit(xv_train, y_train)
```

```
Out[35]: GradientBoostingClassifier(random_state=0)
```

```
In [36]: pred_gbc = GBC.predict(xv_test)
```

```
In [37]: GBC.score(xv_test, y_test)
```

```
Out[37]: 0.792258064516129
```

```
In [38]: print(classification_report(y_test, pred_gbc))

                       precision    recall  f1-score   support

                    0       0.75      0.53      0.62       249
                    1       0.81      0.91      0.86       526

             accuracy                           0.79       775
            macro avg       0.78      0.72      0.74       775
         weighted avg       0.79      0.79      0.78       775
```

4. Random Forest Classifier

```
In [39]: from sklearn.ensemble import RandomForestClassifier
```

```
In [40]: RFC = RandomForestClassifier(random_state=0)
         RFC.fit(xv_train, y_train)
```

```
Out[40]: RandomForestClassifier(random_state=0)
```

```
In [41]: pred_rfc = RFC.predict(xv_test)
```

```
In [42]: RFC.score(xv_test, y_test)
```

```
Out[42]: 0.7935483870967742
```

```
In [43]: print(classification_report(y_test, pred_rfc))

                       precision    recall  f1-score   support

                    0       0.84      0.44      0.58       249
                    1       0.78      0.96      0.86       526

             accuracy                           0.79       775
            macro avg       0.81      0.70      0.72       775
         weighted avg       0.80      0.79      0.77       775
```

Model Testing With Manual Entry

News

```
In [44]: def output_lable(n):
```

News

```
In [44]: def output_lable(n):
             if n == 0:
                 return "Fake News"
             elif n == 1:
                 return "True News"

         def manual_testing(news):
             testing_news = {"text":[news]}
             new_def_test = pd.DataFrame(testing_news)
             new_def_test["text"] = new_def_test["text"].apply(wordopt)
             new_x_test = new_def_test["text"]
             new_xv_test = 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("\n\nLR Prediction: {} \nDT Prediction: {} \nGBC Prediction: {} \nRFC Prediction: {}".format(output_lable(pred_L
                                                                                                  output_lable(pred_D
                                                                                                  output_lable(pred_G
                                                                                                  output_lable(pred_F
```

```
In [46]: news = str(input())
         manual_testing(news)

         Because we have no clear picture as to when theproductions will be resumed, when the movies will be taken tomasses.


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

```
In [ ]:
```

- Survey Question

Email address *

Your answer

Gender *

◉ Male

○ Female

Age *

○ 18-25 years old

◉ 26-40 years old

○ Other:

Fake News on Social Media

Do you have sosial media account *

○ Yes

○ No

Have you heard about fake news on Covid-19 on social media *

○ Yes

○ No

○ Maybe

Where is the more often application that you heard about fake news *

○ Instagram

○ Twitter

○ Whatsapp

○ Facebook

○ Other: _____

42

Do you have heard about application or website that can detect a fake news? *

○ Yes

○ No

○ Maybe

Do you think it will be helpful if this application can detect fake news? *

○ Yes

○ No

○ Maybe

Thank You For Participating.

Back   Submit   Clear form