Analysis of COVID-19 cases and people emotions using Machine Learning

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY

IN

INFORMATION SYSTEMS

Submitted By:

SANDEEP SURI

(2K18/ISY/11)

Under the supervision of
Prof. Kapil Sharma

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

JULY, 2020

CANDIDATE’S DECLARATION

I, Sandeep Suri, Roll No. 2K18/ISY/11 student of M.Tech Information Systems, hereby declare that the project Dissertation titled “Analysis of COVID-19 cases and people emotions using Machine Learning” which is submitted by me to the Department of Information Technology, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any degree, Diploma Associateship, Fellowship or other similar title or recognition.

Sandeep

Place: Delhi Sandeep Suri

Date:

CERTIFICATE

I hereby certify that the Project Dissertation titled “Analysis of COVID-19 cases and people emotions using Machine Learning” which is submitted by Sandeep Suri, Roll No 2K18/ISY/11 Information Technology, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

15/7/2020

Place: Delhi **Prof. Kapil Sharma**

Date: **SUPERVISOR**

ABSTRACT

In recent days the entire world is facing and fighting towards a severe health problem that is known as COVID-19. Earlier COVID-19 is known as novel coronavirus. As of now, there are a total of 9,711,197 cases that have been reported and found positive in the entire world among more than 200 countries. The number of deaths reported are a huge in number. Four lakhs ninety-one thousand and seven hundred ninety-three (4,91,793). The total number of cases that have been recovered are 5,247,173 in numbers.

Doing analysis on such a vast dataset is itself an exciting and challenging task. In recent days and from the beginning of the year 2020 sentiment and emotions of people have also changed and developed in a various manner due to events happening around in their environment and surroundings. People's sentiment, emotions, and opinions are a beneficial medium for analysing the recent trend in society.

People's views emotions and ideas also convey knowledge and information about how people are reacting to a particular event in the community, states, cities, countries, and the world. One such event is COVID-19, COVID-19 is also known as Coronavirus Disease. These opinions shared by people can vary in many forms, such as videos, images, podcasts, audio, and text. People generally share their emotions on social media platforms those are widely used now a days and are Facebook, Instagram, Twitter, YouTube, Blogs, etc. In this research work, we have focused on text data extracted by twitter using twitter API during the period of COVID-19. We will be finding sentimental analysis of twitter users during the period of COVID-19. With data that is vast and significant, which is extracted from twitter using various hashtags related to coronavirus, COVID-19, China, Italy, Trump, etc. We will be performing an analysis of positive as well as negative sentiment displayed in people, various user tweets around the globe marked with different hashtags that more likely to be delivered via the negative emotion. Notably, we introduce a stage-based method to entertain wherewith the negative sentiment changes simultaneously with unique various development frames of COVID-19, which changed from a society residential community epidemic into the domestic level and a worldwide public health emergency furthermore later, into the global pandemic. At each stage of COVID-19 Coronavirus, sentiment analysis allows us to understand the sentiment from tweets that can be majorly negative in essence with various hashtags. Furthermore, the extraction of keywords renders for the development of ideas in the definition of negative emotion through specific tweets.

ACKNOWLEDGEMENT

I manifest my thanks to my major project guide Prof. Kapil Sharma, Professor, lT Dept., Delhi Technological University, for providing the valuable support and guidance whenever needed, he was just a call away, and he provided valuable guidance at each and every step in executing this major project. I am always thankful to my project guide for his constructive guidance, positive responses, and a good insight without which this project will not be completed.

I additionally extend my appreciation to all the faculty members associated with this unit and all the office staff for providing their important guidance and time whenever it was needed.

Sandeep

SANDEEP SURI

Roll No. 2K18/ISY/11

M.Tech (Information Systems)

E-mail: sandeepsuri001@gmail.com

.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Symbols, Abbreviations and Nomenclature

COVID - Coronavirus

SARS – Severe Acute Respiratory Syndrome

WHO – World Health Organization

P(x) – Probability Of X

SARS-CoV-2 – Severe acute respiratory syndrome coronavirus 2

HCov-19 – HumanCoronavirus2019

EVD – Ebola Virus Disease

AR – Augmented Reality

VR – Virtual Reality

ML – Machine Learning

IG – Instagram

NLP – Natural Language Processing

SA – Sentimental Analysis

Fn- Function

Fig- Figure

μ – Mean

σ – Variance

Σ – Summation

CHAPTER 1 INTRODUCTION

1.1 CORONAVIRUS

According to medicine term, a virus is an agent that is sub-microscopic, which is not visible from the naked eyes of humans and is communicable. It produces and reproduces only inside the living organisms [1]. A virus is infectious in nature and can affect all kinds of living organisms that are present on this earth. From Plant to animal and microorganisms to bacteria and humans, all can get infected by a virus. It is still unclear how virus is being originated in the world. A virus is tiny in size and cannot be seen by humans. A virus is even smaller than a bacterium. The effect of an infection from a virus on humans can be cold, cough, headache, fever, fatigue, illness, etc. Identical cells copy of the original virus is being originated by host cell forcefully when a virus attacks on living organisms and a residing organism get infected [2]. There are many numbers of ways by which viruses can spread and attack the living organisms. One pathway is that disease-bearing organisms like viruses can spread from plants, from animals, and from humans.

A new virus that was earlier known as coronavirus and now disease spread by that virus is known as COVID-19 has originated. The coronavirus strain that causes the COVID-19 disease is SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2) [3]. Earlier it was also known as 2019 novel coronavirus and previously also been referred to as (HCov-19) HumanCoronavirus2019. A positive-sense single-stranded RNA virus is present inside SARS-COV2 that is moreover infectious in humans [4]. The illustration of the SARS-COV-2 virus can be seen in the figure below.

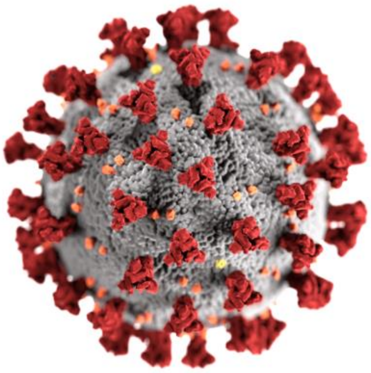
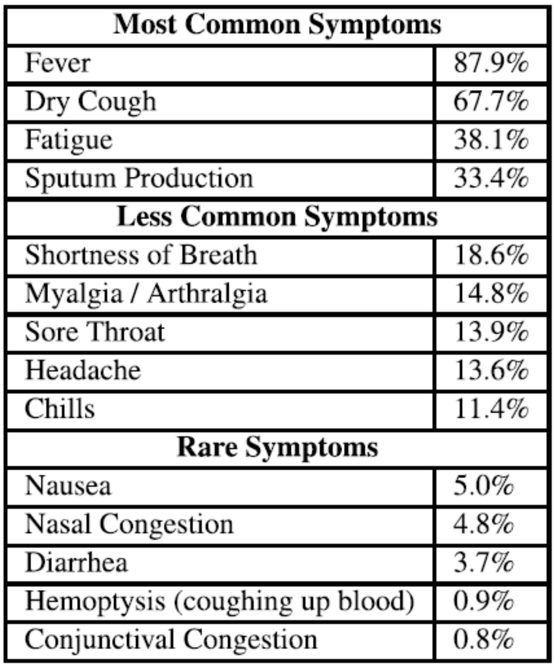


Figure 1.1 SARS-COV-2 Virus

Disease COVID-19 is a compact form or acronym of Coronavirus Disease- 2019. In the earlier stage, this disease COVID-19 is likewise beforehand known as the 2019 novel coronavirus. This disease is named by the WHO (World Health Organization) [5]. WHO (World Health Organization) uses various names for infection and the viruses responsible for the condition. SARS-CoV-2is the virus that is responsible for COVID-19. SARS-CoV-2 stands for severe acute respiratory syndrome coronavirus 2. In the late end of December 2019, as reported by cases in China and media, a novel coronavirus, SARS-CoV-2, developed and produced quickly spreading around the part of Wuhan in China. During the primary explosion in Wuhan, China the virus was usually referred to as Wuhan Virus. Media and research have stated that this novel coronavirus, which is now known as COVID-19, began in Wuhan City in China, and the first case was reported in December 2019. After this, more cases were reported near the radius of the same location in people with the same symptoms where most of the people have a link to the wholesale market in Wuhan City of China. Many of these patients were stall owners, employees of that market area, and residential near-market area and those with some relation with that market either directly or indirectly. As this virus stands for SARS-CoV-2, this severe acute respiratory syndrome coronavirus-2 directly affects the human respiratory system and causes illness in the respiratory organ of patients and humans.

Severe acute respiratory syndrome coronavirus-2 SARS-CoV-2 is a contiguous virus that, moreover, is related to subdivision of single-stranded, positive-sense RNA viruses recognized necessarily as coronaviridae [6]. SARS-CoV-2 strikes the respiratory system of humans and generates diseases such as cold, dry cough, mild fever, mild fatigue, and patient also feel difficulty in breathing that is known as breathlessness. Symptoms shown in human which is affected by COVID-19 are shown in figure below. They are divided in 3 categories a) Most Common Symptoms b) Less Common Symptoms c) Rare Symptoms. The symptoms can be found in the figure given below:



Most Common Symptoms	
Fever	87.9%
Dry Cough	67.7%
Fatigue	38.1%
Sputum Production	33.4%
Less Common Symptoms	
Shortness of Breath	18.6%
Myalgia / Arthralgia	14.8%
Sore Throat	13.9%
Headache	13.6%
Chills	11.4%
Rare Symptoms	
Nausea	5.0%
Nasal Congestion	4.8%
Diarrhea	3.7%
Hemoptysis (coughing up blood)	0.9%
Conjunctival Congestion	0.8%

Figure 1.2 COVID-19 Symptoms

1.1.1 COVID-19

The World Health Organization (WHO) office in China was initially informed of the earlier unknown SARS-CoV-2 ‘Corona’ virus on the 31st December 2019, five months ago. Due to the rapid increment in the number of cases reported by this virus and a massive increase in the mortality rate of each country. The mortality rate is a measure that is defined as a weighted aggregate of the age-specific death measures per 1,00,000 bodies. Hence the WHO decided to announce and declared the disease triggered by the infection – COVID-19 – as a pandemic [7]. Many various diseases have been called an epidemic by WHO. They are described in the table below.

Disease	Agent	Year	Death	Classification
COVID-19	SARS-CoV-2	2019-Present	3620279	Pandemic
Ebola Virus Disease (EVD)	Zaire Ebolavirus	2014-2016	11,325	Epidemic
SARS	SARS-CoV	2002-2004	774	Outbreak
Spanish Flu	H1N1	1918-1919	~50 Million	Pandemic
Asian Flu	H2N2	1957-1968	~1.1 Million	Pandemic
MERS	MERS-CoV	2012-Present	871*	Outbreak
Asian Lineage Avian Influenza	H7N9	2013-2017	~605	Epidemic
Swine Flu	H1N1(new strain)	2009-2010	~151700 to 575400	Pandemic
Hong Kong Flu	H3N2	1968-1969	~1 Million	Pandemic

Table 1.1 Disease classification by WHO

1.1.2 Cases Timeline

According to the WHO, viral infections, especially those entireties generated through unconventional coronaviruses, recapitulate to arise moreover profess a critical common health issue. Coronaviruses are rounded positive-sense Ribonucelic acid microorganisms covering 600Å - 1400Å within diameter [8]. They have spikes on the upper head in the form of protein extruding from there covering exterior surface, which admits some structure that is crown-like when viewed by electron microscope. The COVID-19 break appeared over information on 30 December 2019 during 27 incidents regarding pneumonia, and unexplored ethology held proclaimed near the WHO's country office in China. The entire timeline of cases can be view in figure.

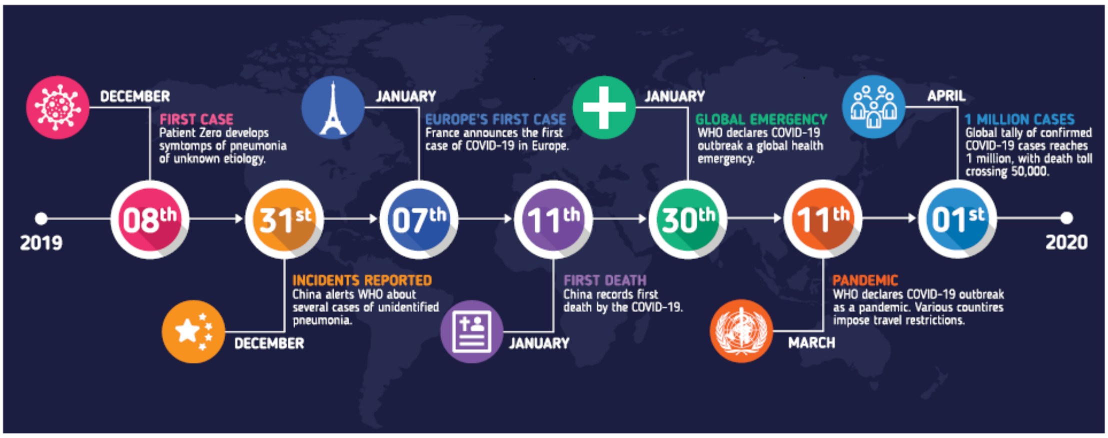

DECEMBER
FIRST CASE
Patient Zero develops symtomps of pneumonia of unknown etiology.
JANUARY
EUROPE'S FIRST CASE
France announces the first case of COVID-19 in Europe.
JANUARY
GLOBAL EMERGENCY
WHO declares COVID-19 outbreak a global health emergency.
APRIL
1 MILLION CASES
Global tally of confirmed COVID-19 cases reaches 1 million, with death toll crossing 50,000.
2019
08th
31st
07th
11th
30th
11th
01st
2020
INCIDENTS REPORTED
China alerts WHO about several cases of unidentified pneumonia.
DECEMBER
FIRST DEATH
China records first death by the COVID-19.
JANUARY
PANDEMIC
WHO declares COVID-19 outbreak as a pandemic. Various countires impose travel restrictions.
MARCH

Figure 1.3 Primary Timeline of Cases

As shown in figure 1.3 the first case was reported in early December 2019. Patient zero develops the symptoms of pneumonia of unknown etiology [9]. Patient zero is the patient who is the first attacker by the virus. Patient zero is also called to the person who has got infected by the virus, and his case is reported for the first time to date. In late December, it was identified, and the incident is reported; hence China alerts WHO about several instances of unidentified pneumonia. China was in light of WHO because cases reported in China was increasing rapidly. Due to the travel history of people in Europe from China, it spread in part of Europe, such as France and Italy. France reported the first coronavirus case in the first week of January [10]. France announces the first case of coronavirus, COVID-19 in Europe. Meanwhile, on January 11, China reported the early death in Wuhan City because of the novel coronavirus. China started hiding the result of death by the novel coronavirus and imposed, lockdown in Wuhan wet market in China [11]. In late January WHO (World Health Organization) declared it as an emergency. COVID-19 was declared as an emergency by WHO in late January.

1.2 MECHANISM OF TRANSMISSION

Although there are numerous investigations toward every region concerning COVID-19's pathophysiological characteristics, its circulation criteria, persists moderately tricky. While those of opening COVID-19 incidents remained connected among some direct revelation concerning people through infected animals, the winged revolution concerning the epidemic has stirred the locus of the investigation into human-to-human conveyance. The list of COVID-19 indications is shown in Table 1.

They are fundamentally transferred among personalities of the range concerning respiratory droplets throughout sneezing, including coughing [12]. Those respiratory droplets possess significant potential to reach a length of about up to 2 meters (6 feet). Accordingly, any character in the tight association, including an infected person, remains in jeopardy of transpiring endangered through the respiratory droplets and enlargement. Although representative bodies possess been within the primary root of SARS-CoV-2 carrying, there is also an opportunity to transmit via asymptomatic personalities. Immediate and secondary contact, including infected superficies, has been identified as different possible causes of COVID-19 transmission [13].

Once a specific virus penetrates inside a healthy body, it reaches within the entrance from nasal upon the snot layers which is existing within the neck moreover, it fastens itself over some body's cellular receptors. Amidst some aides like the spikes already upon its covering, the SARS-CoV-2 breaks each membrane about the cell, and multiple copies of itself are being produced forcefully [14]. These newly generated copies burst out of the cell and infect other cells in the body. Then this COVID-19 virus goes below the bronchial tubes and arrives these lungs, wherever it rigorously reduces every host's air pouches.

The table below shows and gives a detailed view regarding the disease that has been marked and declared as a pandemic by the WHO(World Health Organization) [13]. WHO (World Health Organization) has categorized virus and their disease in majorly three categories, Outbreak, Epidemic, and Pandemic.

The outbreak is the class in which death or mortality caused by the virus in humans is less than 1000. An Epidemic is a disease caused by the virus by which a large community gets affected. A vast number of people inside society or community get affected by illnesses, then that disease is classified as an epidemic. The pandemic is a state that comes after an epidemic state when a virus and infection is spread in the entire nation. Of the whole nation, it has been covered in the majority of the country and their part, moreover when their affect is vast . The number of people getting infected is increasing rapidly than that disease is classified as pandemic disease.

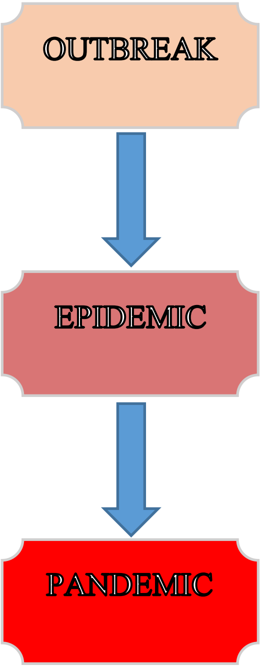


Figure 1.4 Mechanism of Disease

The various disease that has been declared as Outbreak, Epidemic, and Pandemic by WHO are described in the table below.

Disease	**Agent**	**Year**	**Death**	**Classification**
COVID-19	SARS-CoV-2	2019-Present	3620279	Pandemic
Ebola Virus Disease (EVD)	Zaire Ebolavirus	2014-2016	11,325	Epidemic

SARS	SARS-CoV	2002-2004	774	Outbreak
Spanish Flu	H1N1	1918-1919	~50 Million	Pandemic
Asian Flu	H2N2	1957-1968	~1.1 Million	Pandemic
MERS	MERS-CoV	2012-Present	871*	Outbreak
Asian Lineage Avian Influenza	H7N9	2013-2017	~605	Epidemic
Swine Flu	H1N1(new strain)	2009-2010	~151700 to 575400	Pandemic
Hong Kong Flu	H3N2	1968-1969	~1 Million	Pandemic

Table 1.2 Major Disease from Virus

1.3 STAGES OF COVID-19

As stated by the WHO(World Health Organization), the novel coronavirus or the COVID-19 pandemic is considered essentially holding four principal forms concerning transportation. These four ways remain same and uniform everywhere in the entire world for better communication, for better results, for helping each and everyone, for better awareness and for fighting together in these difficult situations and for better conclusion amongst each and every country.

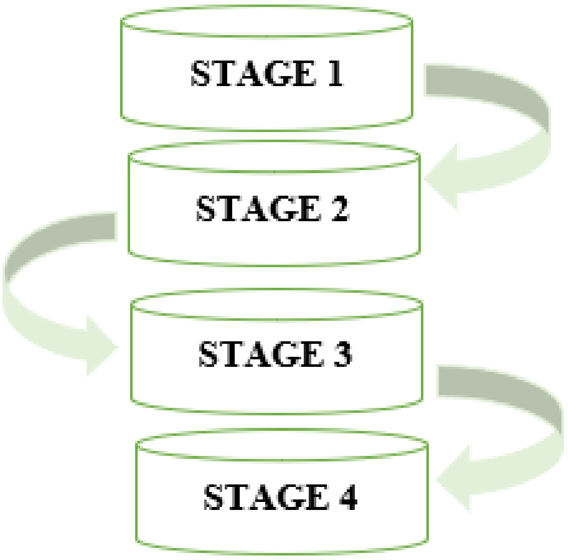


Figure 1.5 Stages of Coronavirus

The WHO (World Health Organization) has classified these stages as stage 1 is called as imported cases only, stage 2 is known as local transmission, whereas stage 3 is furthermore known as cluster of cases and stage 4 the final stage is named as community transmission. The detailed view of each stage is shown in the figure below.

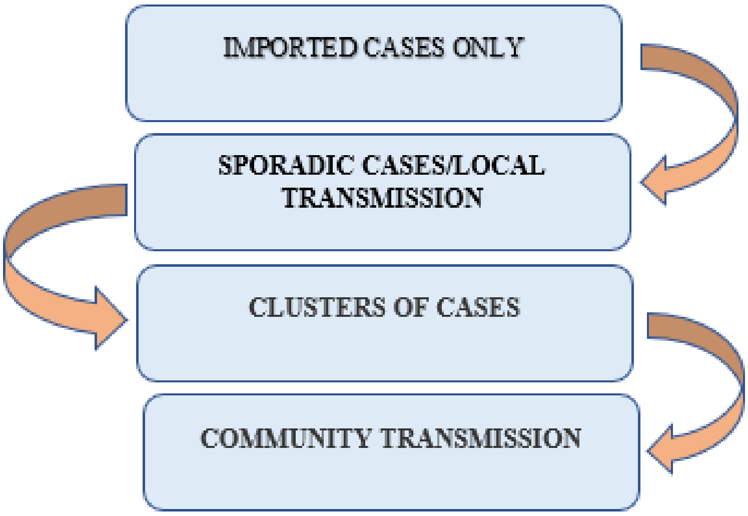


Figure 1.6 Detailed stages of transmission

The primary, first, and beginning step of the COVID-19 outbreak in an appropriate community remains described through the first recorded event concerning the disease. In particular step, the virus residing and disease does not develop in the local community, and the infection spread by COVID-19 virus is ordinarily restrained through these characters beside travel records to a previous region that is being affected by this virus.

The secondary stage regarding the COVID-19 disorder happens if there remain several uncommon circumstances concerning the disease in the nation or state or city or particular country.

It appears meanwhile human who is previously affected beside the disease expanded to characters among all who all appear into contact, ordinarily direct family affiliates, associates, and co-workers.

At this step, this is conceivable to complete communication reproduction furthermore narrow the scope concerning the disease through separating or isolating or quarantining the people or humans or characters who have got infected.

The further stage after the previous stage concerning the COVID-19 explosion in a country remains characterized through the appearance concerning different groups of COVID-19 cases, i.e., during that disease-causing virus begins broadcasting inside a geographic position and affects people who have not a history of a journey neither came in association or communication or contact with someone which have the history of journey or travel history within a country or outside from a country to another country.

By this step, it matures and becomes harder to determine or trace or investigate the root concerning the virus synchromesh, and geographical lockdown or terrestrial lockdown grows and becomes extremely important and necessary to counter and to prevent the explosion from entering, reaching and arriving at stage IV.

The final and fourth step concerning the COVID-19 pandemic in a nation or country remains amalgamated and associated with neighborhood transmission which is also known as community transmission, i.e., greater outbreaks more extensive outbreaks of local synchromesh, local transmission in a particular place, state, city, nation, and country, heading over a remarkably great and extremely high figure or numbers of reported events, detailed incidents and deaths.

At the aforementioned final stage, the outbreak goes out of power, goes out of control, goes out of command, and finding a remedy, searching for a cure or vaccine remains the sole approach to decrease significant impression concerning the disease. Countries like

India, USA, Brazil, and Russia remain currently into this fourth stage concerning the COVID-19 pandemic.

1.3.1 Impact on Society

Due to lack of proper treatment facility that is being provided to the patient who has got infected by novel coronavirus or COVID-19, Due to lack of transparency, Due to lack of social distancing, Due to lack of vaccine, Due to lack of testing kits, Due to lack of concrete plan and proper health facility, the number of cases tested being positive are increasing day by day in a swift number. The COVID-19 will be going to impact almost all the sectors in this world, such as the automotive industry, aviation industry, tourism industry, construction firms, the oil industry, food industry, telecommunication industry, etc. These are the industry which will be going to face a major impact of COVID-19.

The automotive business has witnessed significant separations in merchandise due to poignant lockdown standards implemented in numerous countries worldwide so that to make the effect of virus less and to contain the pandemic. As many countries have strictly imposed social distancing, social distancing is strengthened everywhere, some zones are made as a contaminated zone where none can go out, and no one can come inside those zones. Challenges faced in sales is shown in figure below.

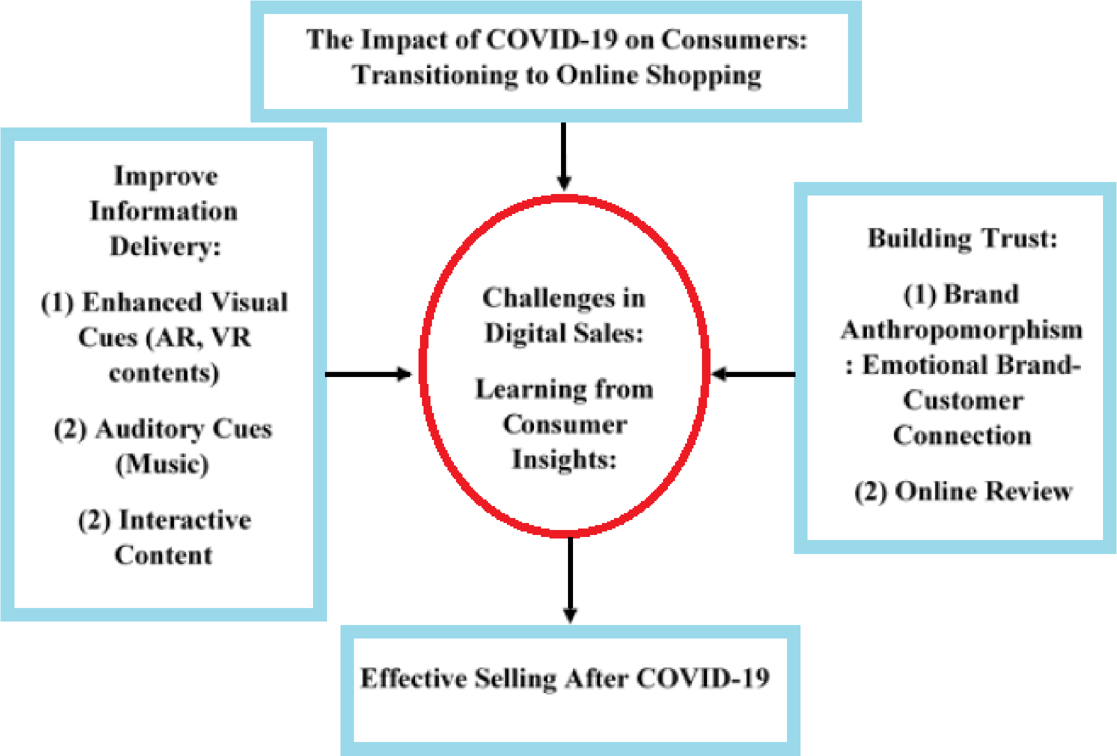
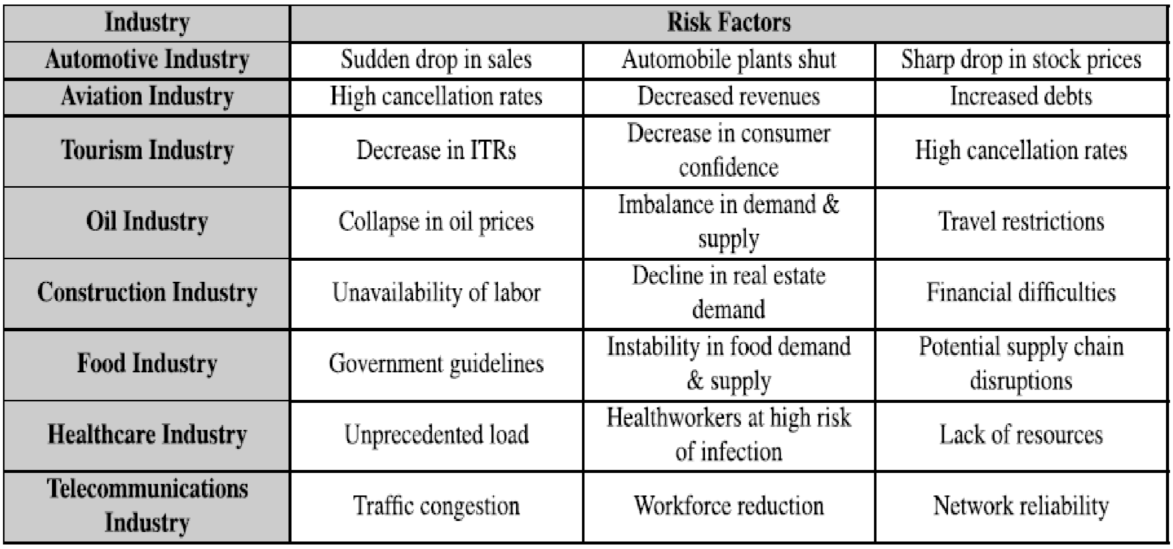

The Impact of COVID-19 on Consumers: Transitioning to Online Shopping
Improve Information Delivery:
(1) Enhanced Visual Cues (AR, VR contents)
(2) Auditory Cues (Music)
(2) Interactive Content
Challenges in Digital Sales:
Learning from Consumer Insights:
Building Trust:
(1) Brand Anthropomorphism : Emotional Brand-Customer Connection
(2) Online Review
Effective Selling After COVID-19

Figure 1.7 Challenges in sales

Lockdown is being imposed where the people are ordered to stay in their homes. Hence the practice of automobiles, including both public & private transportation, possesses declined beyond the globe. Only essential services warriors are allowed to use their vehicles in the lockdown and pandemic. The COVID-19 pandemic possesses a huge impression and a massive impression on the aviation industry. Affected countries, including around all those countries, have been ordered, forced for imposing travel prohibitions on both foreign and domestic commuter flights. The only effective airways combine critical supply routes that are related with medical facility or ambulance or doctors or essential products or support cargo and freight aircraft. Since there is a restriction on traveling by airways, railways and personal vehicles, one of the most affected industries is the tourism industry. It is affected worst. Revenue generated by the tourism industry is almost null in the phase of COVID-19 pandemic. As there is a ban on traveling and only essential services are allowed, only essential vehicles are allowed to commute. Those are very limited in number and the abandonment of foreign and national commuter airways crosswise the world has appeared in extreme deterioration during the usage of aviation fuel and as well as normal fuel that is being used by commuters for traveling from office to home and way back. Hence there is a sharp decline in revenue and profit generated by the oil field industry. These all decline and majors can be seen in the figure below.



Industry	Risk Factors		
Automotive Industry	Sudden drop in sales	Automobile plants shut	Sharp drop in stock prices
Aviation Industry	High cancellation rates	Decreased revenues	Increased debts
Tourism Industry	Decrease in ITRs	Decrease in consumer confidence	High cancellation rates
Oil Industry	Collapse in oil prices	Imbalance in demand & supply	Travel restrictions
Construction Industry	Unavailability of labor	Decline in real estate demand	Financial difficulties
Food Industry	Government guidelines	Instability in food demand & supply	Potential supply chain disruptions
Healthcare Industry	Unprecedented load	Healthworkers at high risk of infection	Lack of resources
Telecommunications Industry	Traffic congestion	Workforce reduction	Network reliability

Figure 1.8 Industry impacted by COVI-19

In medical terminology, prevention is defined as a measure or steps or procedure that are being followed to reduce or minimize or diminish the chance of getting infected by disease or any virus, COVID-19 is stated declared as a pandemic by WHO (Worl Health Organization), so there is a much need for a proper preventive measure to e followed so as to we can reduce the chance of getting affected by this small but powerful virus. WHO (World Health Organization) with the help of researchers, scientists, scholars, and doctors have issued some guideline for the preventive measure to be followed by each and every citizen and human in this world. They are more elaborated in the upcoming section. precautionary measure to be developed can be seen in the below figure.

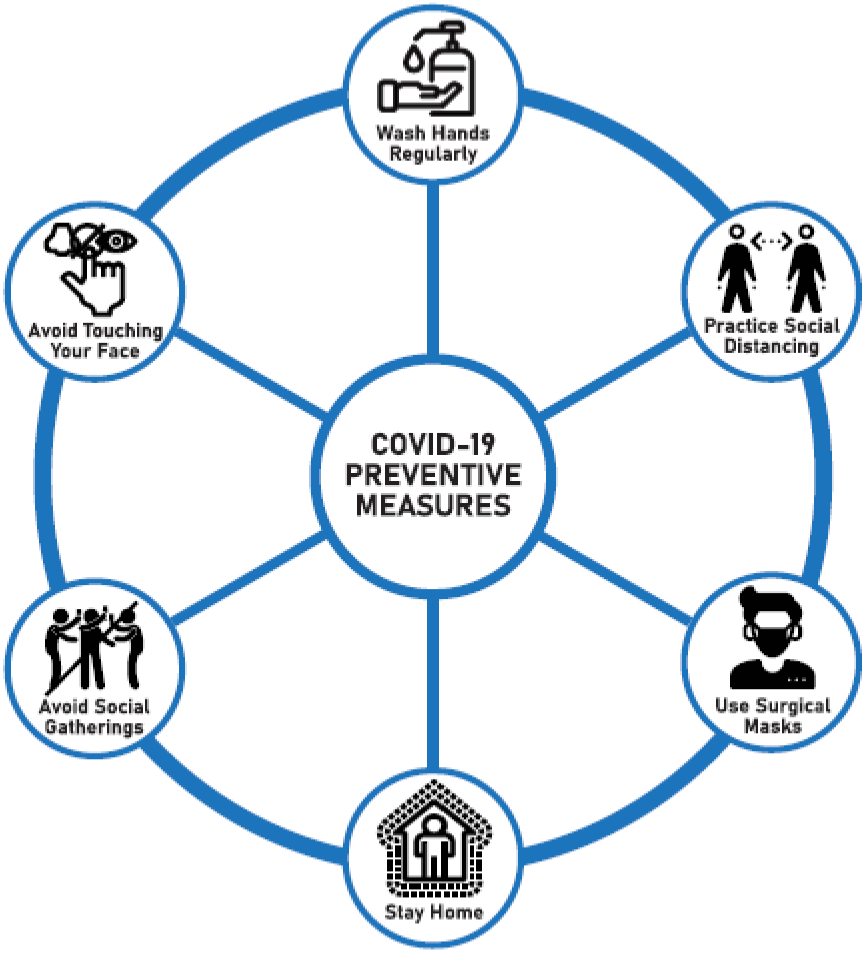


Figure 1.9 COVID-19 Preventive Measure

1.3.2 Unsupervised Learning

In unsupervised learning, we have input data ‘x’ only and no respective output data. The machine is trained without any classified or labelled data. The various types of machine learning techniques that can be used for modelling data and making a machine learn is described in figure below in more detail in a hierarchical manner.

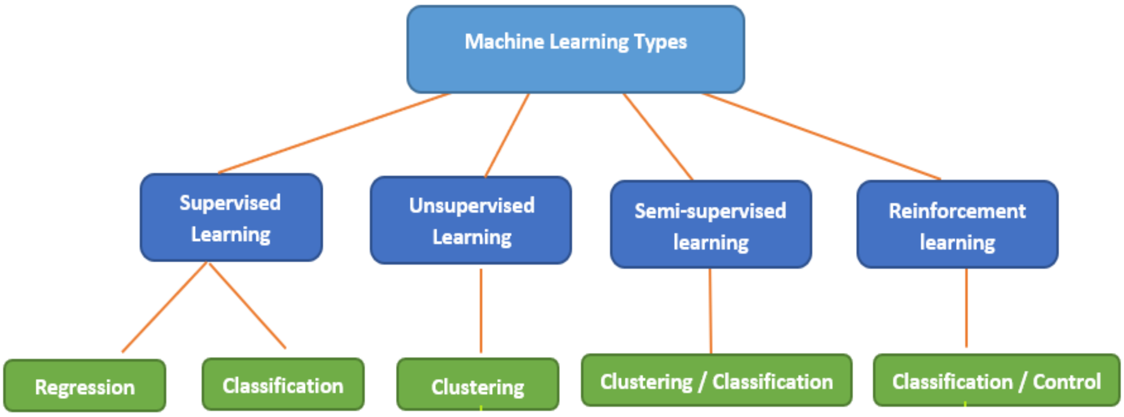


Figure 1.10 Types of machine learning techniques

The machine itself learns to group the information with the help of similarities, patterns or differences in the given data. Unlike supervised learning, there is no supervisor to provide correct labels to the input data and algorithm tries to find hidden structures in the data. Clustering and Association are the examples of unsupervised learning. Some of the unsupervised learning approaches involve K-means clustering, Neural Networks, Principle Component Analysis (PCA), Hierarchal Clustering, etc. In Unsupervised learning, learning happens without a teacher. The training data consisting of {x}, only the inputs are observed and there are no target outputs.

1.3.3 Supervised Learning

In Supervised Learning, learning happens with a teacher. Here, the teacher represents a training dataset consisting of annotated labels to effectively train the algorithm. We know the correct answers; the algorithm iteratively makes predictions on the training data and is corrected by the teacher. Some of the Supervised learning techniques are Decision Trees, Neural Networks and more.

There are several approaches proposed with mentioned techniques. These techniques involve learning from the features extracted from the images and label them to the target. In our experiment, we have considered neural networks to classify between images. Especially, convolutional neural networks (CNNs), due to superior accuracy through recognizing a pattern compared with classical machine learning methods, which rely on hand-crafted features. Also, many methods are formulated using CNNs with high classification accuracy.

1.4 NATURAL LANGUAGE PROCESSING

There has signified an exponential increase in the application concerning textual analytics, natural language processing (NLP), moreover new artificial intelligence systems in investigation furthermore in the expansion of applicability. Notwithstanding, concerning accelerated advancements into and towards natural language processing NLP, arguments encompassing the modifications of specific techniques during explaining central application in manuscript settle. Researchers who are researching at MIT have described whereby smooth single most contemporary NLP mechanisms backside drop compressed and continue "vulnerable to adversarial text." It is consequently essential to assume inherent boundaries regarding text categorization techniques furthermore appropriate machine learning algorithms.

Furthermore, that is necessary to traverse if duplicated exploratory, explanatory, including organization routines comprise corresponding synergies, which instructions subtract us to leverage the "everything is more comprehensive than the entirety of its machines" principle in our hunt for artificial intelligence-driven perspicacity contemporaries from individual interactions. Studies in automated exchanges produce confirmed the effectiveness regarding machine learning toward molding own performance following complicated informational stipulations, highlighting the function regarding specific creation from knowledge in changing individual performance.

The acceleration toward importance on Artificial Intelligence AI systems concerning textual analytics furthermore Natural Language Processing NLP ought to comprehend the colossal improvement in federal reliance upon common media (e.g., Twitter, Facebook, Instagram, blogging, and LinkedIn) toward learning and knowledge, fairly than approaching this established message means. People express their opinions, moods, and activities on social media about diverse social phenomena (e.g., health, natural hazards, cultural dynamics, and social trends) due to personal connectivity, network effects, limited costs, and easy access. Many companies are using social media to promote their product and service to the end-users.

Correspondingly, users yield their encounters furthermore surveys, producing a magnificent repository regarding knowledge saved as subject. Consequently, common communications and welcoming information principles are enhancing essential origins regarding knowledge concerning administering investigation, into the circumstances of expeditious advancement concerning knowledge furthermore information technology. Furthermore, that is necessary to traverse if duplicated exploratory, explanatory, including organization routines comprise corresponding synergies, which instructions subtract us to leverage the "everything is more comprehensive than the entirety of its machines" principle in our hunt for artificial intelligence-driven perspicacity contemporaries from individual interactions. Studies in automated exchanges produce confirmed the effectiveness regarding machine learning toward molding own performance following complicated informational stipulations, highlighting the function regarding specific creation from knowledge in changing individual performance. It is consequently essential to assume inherent boundaries regarding text categorization techniques. The relationship between Natural Language Processing, Artificial intelligence, Machine Learning and Deep Learning is shown in figure below.

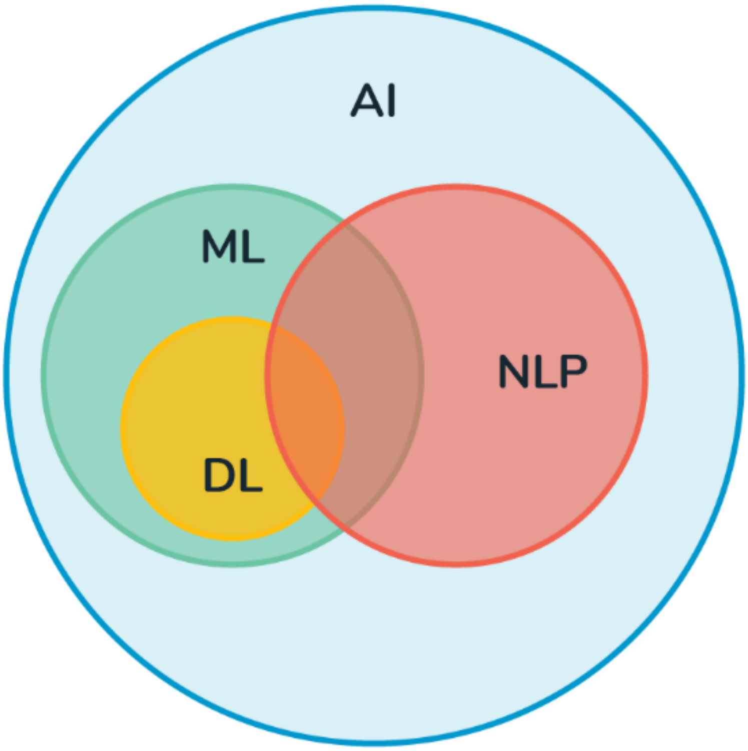


Figure 1.11 Relationship between AI, ML, NLP, DL

The first approach that is being followed by Natural Language Processing (NLP) is as first, and the primary step is to detect the language of the text. It is one of the challenging tasks, and python library NLTK has made this task very easy for us. Once the expression is being detected, the tokenization part is being done. In the tokenization process, each word is being represented as token, and token are being generated using finite automata. The tokenizer is responsible for making and doing tokenization of words in the Natural Language Processing (NLP). The working of tokenizer is shown in the figure below.

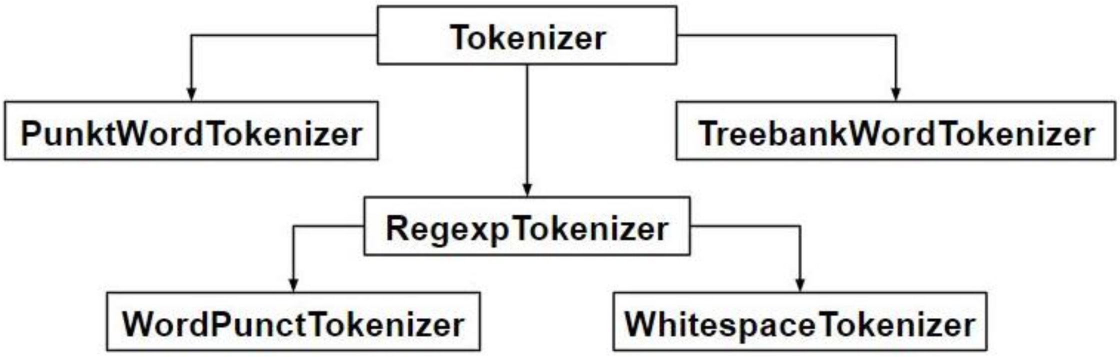


Figure 1.12 Tokenization in NLP

Once the tokenization is done then, NLP does the tagging of each word in the given text. The tag is being associated with each word. Part of Speech is the process done in tagging. Once the tagging is done, NLP uses the removal stop words. Now the model is ready for modelling the data and performing predictions. After this, we can get many forms in the result, such as Sentiment Analysis, Classification, Entity Extraction, Translation, Topic Modelling, etc. Here in this section, we will be focusing on sentimental analysis. For example, if we want to predict next word in the sentence, it is necessary to have the knowledge of the word which comes before it. Here comes the need of connection between the inputs themselves. The detailed steps taken during Natural Language Processing techniques is shown in figure below.

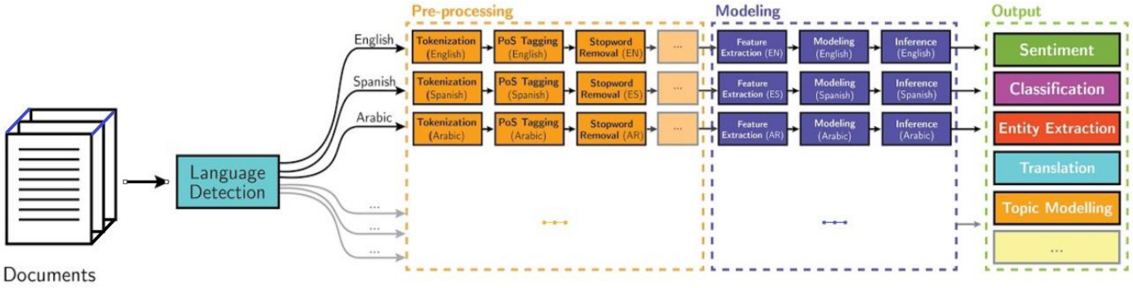


Figure 1.13 NLP Steps

1.4.1 Sentimental Analysis

All the essential aspects and sources about sentimental or opinion or emotions detection are studied. Social media users express their sentiments, views regarding a particular topic, and opinion concerning a topic and emotions towards related issue on social media concerning different cultural phenomena such as environment, review of any product, medical facility, disease, health, disaster, economics towards a company, promotional brand, natural hazards, social trends, etc. People are connected at the personal, emotional, and social levels on social media. The best thing about social media is it is free, available for everyone, freedom of speech, freedom of expressing views towards a topic, freedom of expressing emotions concerning a matter, liberty of giving opinions and suggestions to each other. According to a report published, around 3.8 billion people are using social media. People are actively using these social media websites such as Facebook, Instagram, Snapchat, YouTube, Twitter for expressing their views, opinions regarding a particular issue, emotions towards a topic, and for gaining information.

A recent report revealed by Google has shown that there is an increase in the number of people using Social Media from January 2020 because of lockdown in some nations. There are more active users on the social media website this year as compared to previous years. There is high traffic observed by Google on the internet in early 2020. According to a report revealed by twitter, there are approximately 330 Million accounts on twitter, which itself is a vast number. Due to this actual reason and beauty of twitter regarding tweets and followers, many politicians, leaders, celebrities, etc. are now using twitter to spread their message and to influence people. In this research work, we will be restricting the sentimental analysis of people, their emotions regarding COVID-19, and their views towards COVID-19 on twitter only. Steps involved in this method of doing sentimental analysis on tweets is similar as basic prototype model of doing sentimental analysis. It is as show in figure, Step1 involves Data Collection, collecting of data is being done by using various tools in this research work we have collected data using twitter application tweepy.

We will explain this in a further section. Since here our data is in text form, and those are known as tweets. Tweets of various users are collected during an interval of some days. The scrapping of tweet is done using tweepy. Now step 2 in the basic prototype is text preparation. Preparation of text is being done by removing useless data, useless keywords, unnecessary symbols. This step and phase text is prepared and made ready, and this step is also known as data cleaning. We clean our data in this step and phase. When this step is done our text and cleaned, data is being passed into another stage that is known as Sentiment Detection. In this phase of our prototype, we do the detection of sentiment. Sentiment detection is being done on our cleaned data. Once emotions are detected, then they are classified. Hence sentiment classification is our next phase of a basic prototype model. Sentiments or Emotions are categorized broadly under three categories positive, neutral, and negative. Once the model does classification output is being produced in the last stage.

The sentimental analysis is also known as the extraction of emotions, opinions, or sentiments from data consist of 3 phases. Phase 1 consist of three steps where input is the review data that is being produced by cleaning of data, then sentiment sentence extraction of our tweets are being done. In this phase, the extraction of sentiment on each sentence is being done using various algorithms for this. We have used the nltk library in python. Once this step is completely done, the last stage of phase 1 involves part of speech tagging. Output produced from phase 1 is taken as input in phase 2 of sentimental analysis. Then sentiment phrase identification of this output is made as step one in phase 2. Once the phrase identification of each sentence is made, the score of each word is being computed, and their sentiments are calculated. Based on this score, we generate the feature vector. Feature vector those are calculated in phase 2 are taken as input in phase 3, and sentiment polarity categorization is performed. Once polarity categorization is done, results are interpreted. These steps are shown very clearly in the figure below.

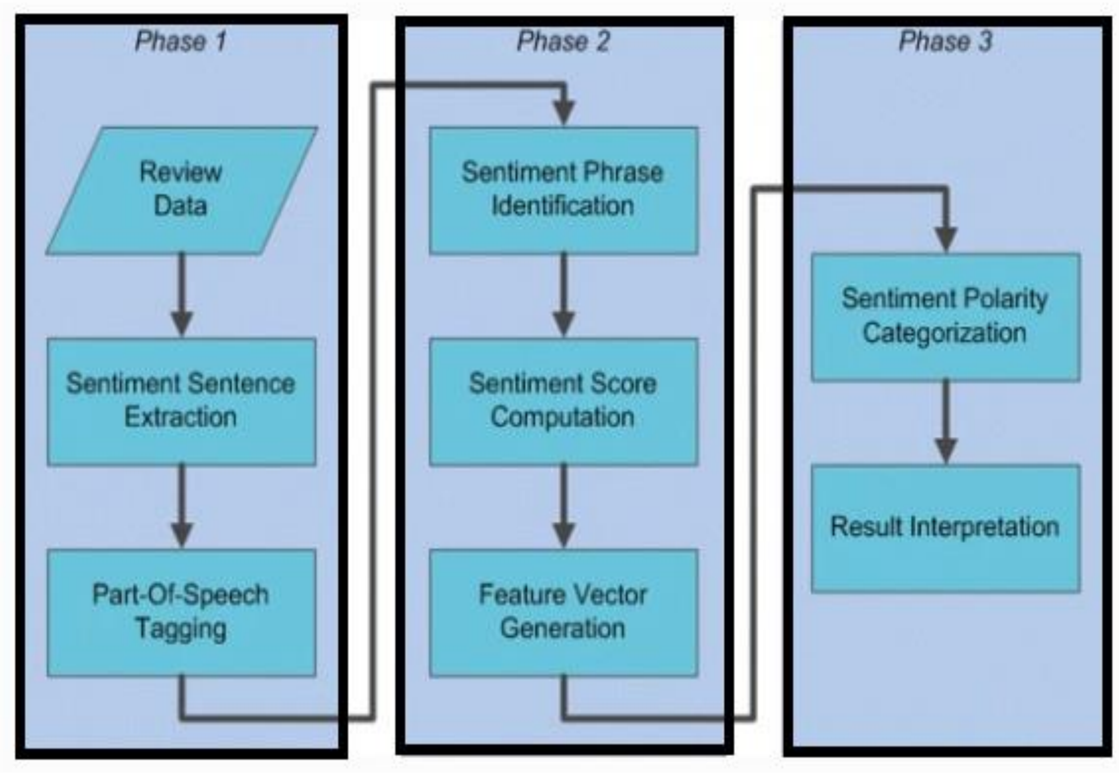
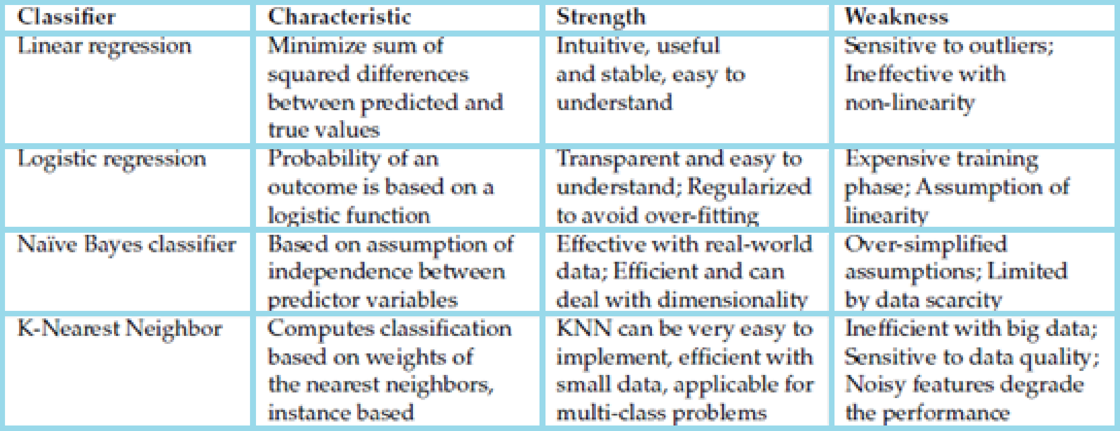

Phase 1
Review Data
Sentiment Sentence Extraction
Part-Of-Speech Tagging
Phase 2
Sentiment Phrase Identification
Sentiment Score Computation
Feature Vector Generation
Phase 3
Sentiment Polarity Categorization
Result Interpretation

Figure 1.14 Phases in sentimental analysis

1.4.2 Classification Methods

Existing study has practiced different textual distribution arrangements to assess sentiment done on social media. Those classifiers remain classified within various classes based upon their relationships. Those classifiers continue classified inside various classes based superimposed their connections. The subdivision that reflects addresses features regarding four indispensable classifiers we evaluated and examined, including linear regression and K-nearest neighbour, furthermore we are comparing Naïve Bayes and logistic regression because of their concepts, strengths and weaknesses. Both these models consist of additional parameters to control memory updating process. They both are able to capture long and short-term dependencies. The below figure gives an overview regarding various classifiers that we will be using and has been extensively used by multiple researchers, scholars or scientists in the field of NLP to find the sentiment. These are the basic and one of the essential model classifiers for doing sentimental analysis.



Classifier	Characteristic	Strength	Weakness
Linear regression	Minimize sum of squared differences between predicted and true values	Intuitive, useful and stable, easy to understand	Sensitive to outliers; Ineffective with non-linearity
Logistic regression	Probability of an outcome is based on a logistic function	Transparent and easy to understand; Regularized to avoid over-fitting	Expensive training phase; Assumption of linearity
Naïve Bayes classifier	Based on assumption of independence between predictor variables	Effective with real-world data; Efficient and can deal with dimensionality	Over-simplified assumptions; Limited by data scarcity
K-Nearest Neighbor	Computes classification based on weights of the nearest neighbors, instance based	KNN can be very easy to implement, efficient with small data, applicable for multi-class problems	Inefficient with big data; Sensitive to data quality; Noisy features degrade the performance

Figure 1.15 Classifier for Machine Learning

CHAPTER 2 RELATED WORK

In this section, we reviewed what work has been done before in the field of price prediction. We studied the type of techniques has been employed for the same and how much they are efficient.

A literature review is a reflection and analysis of previously performed work by other researchers and scholars, which is related or similar to this matter. All the essential aspects and sources about sentimental or opinion or emotions detection are studied. Social media users express their sentiments, views regarding a particular topic, and opinion concerning a topic and emotions towards related issue on social media concerning different cultural phenomena such as environment, review of any product, medical facility, disease, health, disaster, economics towards a company, promotional brand, natural hazards, social trends, etc. People are connected at the personal, emotional, and social levels on social media. The best thing about social media is it is free, available for everyone, freedom of speech, freedom of expressing views towards a topic, freedom of expressing emotions concerning a matter, liberty of giving opinions and suggestions to each other. According to a report published, around 3.8 billion people are using social media. People are actively using these social media websites such as Facebook, Instagram, Snapchat, YouTube, Twitter for expressing their views, opinions regarding a particular issue, emotions towards a topic, and for gaining information. A recent report revealed by Google has shown that there using Social Media since January 2020 because of lockdown in some nations. There are more active users on the social media website this year as compared to previous years. There is high traffic observed by Google on the internet in early 2020. According to a report revealed by twitter, there are approximately 330 Million accounts on twitter, which itself is a vast number. Due to this actual reason and beauty of twitter regarding tweets and followers, many politicians, leaders, celebrities, etc. are now using twitter to spread their message and to influence people. In this research work, we will be restricting the sentimental analysis of people, their emotions regarding COVID-19, and their views towards COVID-19 on twitter only. Steps involved in this method of doing sentimental analysis on tweets is similar as basic prototype model of doing sentimental analysis. It is as show in figure, Step1 involves Data Collection, collecting of data is being done by using various tools in this research work we have collected data using twitter application tweepy. We will explain this in a further section. Since here our data is in text form, and those are known as tweets. Tweets of various users are collected during an interval of some days. The scrapping of tweet is done using tweepy. Now step 2 in the basic prototype is text preparation. Preparation of text is being done by removing useless data, useless keywords, unnecessary symbols. This step and phase text is prepared and made ready, and this step is also known as data cleaning. We clean our data in this step and phase. When this step is done our text and cleaned, data is being passed into another stage that is known as Sentiment Detection. In this phase of our prototype, we do the detection of sentiment. Sentiment detection is being done on our cleaned data. Once emotions are detected, then they are classified. Hence sentiment classification is our next phase of a basic prototype model. Sentiments or Emotions are categorized broadly under three categories positive, neutral, and negative. Once the model does classification output is being produced in the last stage.

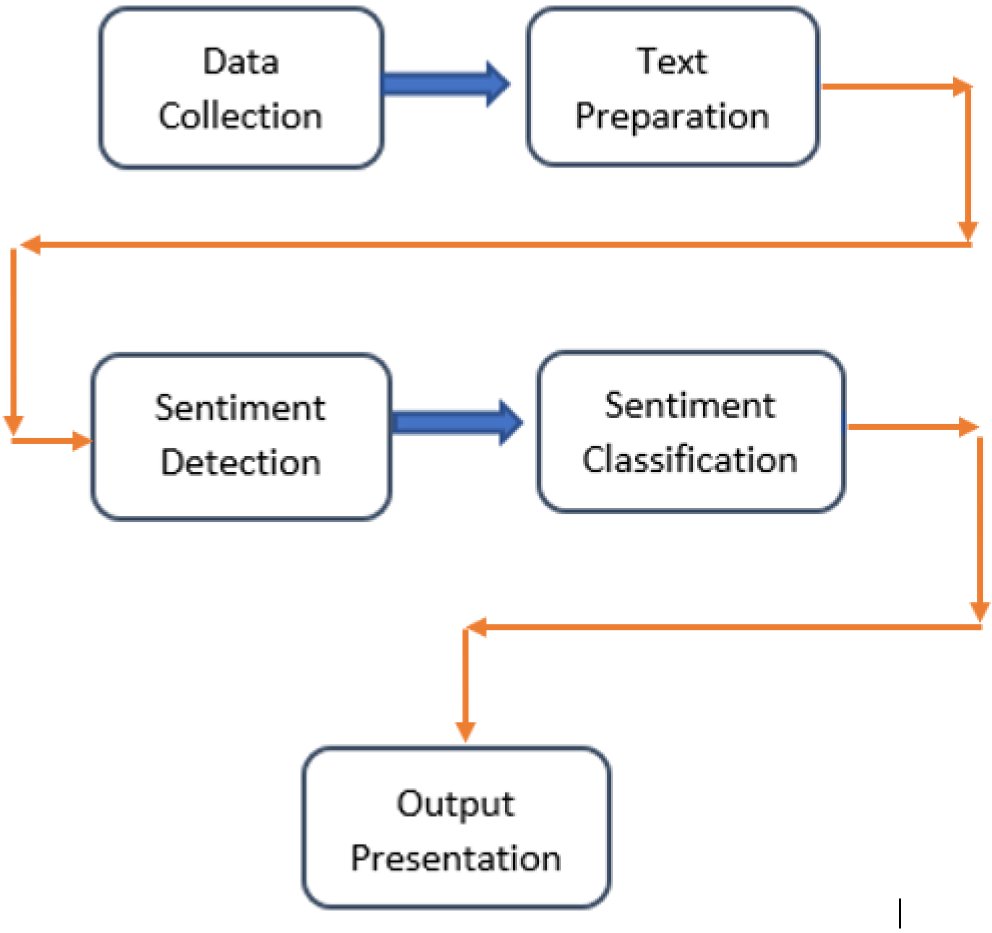


Figure 2.1 Process followed

ARIMA can predict time-series data even with less term time period. It is a parametric method for predicting time series individually. Prices are accurately predicted by using threshold for estimation. Some of the disadvantages of this model are: it cannot work parallel for more than one time series, it is not guaranteed that values estimated are closer to actual ones and lastly, the model fails when accuracy, RMSE etc. parameters are considered. Latent Source Model was developed for binary classification. There is a method called Bayesian Regression for predicting variation in bitcoin price. In combination with Bayesian Regression, LSM determines the patterns in the system. Using trading strategy for making decision either to buy or sell bitcoin, mid-price is predicted. A threshold value is set for this. If average price is less than the threshold, then bitcoin will be sold and if it is equal to greater than bitcoin is purchased. Value for threshold is updated from time to time according to trading strategy. MLP uses binary classification with two classes as 0 or 1 where 1 indicates increment in bitcoin price for the upcoming day and 0 indicates there will drop in price for the same. One more model i.e. NARX is a dynamic recurrent network having feedback connections with many network layers. This model can make predictions even on dynamic data of time series. The attraction of the model is that this model can accept continuous as well as discreet values for making predictions. It is implemented from linear ARX model. Advantages of this model is that it is faster, gives good accuracy, understand system behaviour better and can generalise effectively as compared to other neural networks. Comparing all these models it was found that NARX model gave highest accuracy in predicting prices.

Having a lot of methods and models for price prediction, there is a need to design methods which can improve performance of prediction. In this study, the authors proposed techniques for closing stock price prediction using deep beliefs networks (DBNs) [21].

In our study we have used python library and plotly, Plotly is also known as front end for machine learning model and data analysis. In plotly we can view and plot data in various forms as it enhances the way data is coming and being viewed by various users. Best thing in plotly is we can also zoom in or zoom out the plot which is not possible in general python library. Plotly is used for enhanced user interface in machine learning modeling and data analysis.

To best of our knowledge, this work gives better results which are closer to actual values with a fair computational time.

CHAPTER 3 - EXPERIMENTAL APPROACH

3.1 DATASET

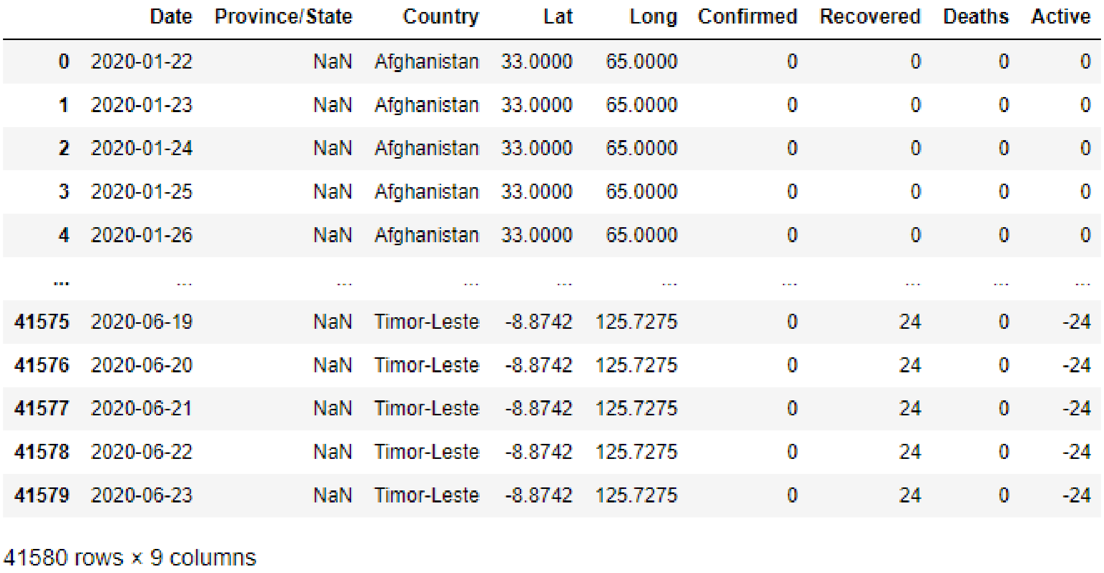
COVID-19 is the trending and recent topic in today's era. Finding data for COVID-19 itself is a challenging task. Even after finding data set of COVID-19 getting accurate data for analyzing trends and analyzing COVID-19 cases concerning various dimensions itself is a big challenge. To overcome all of this difficulty, we have come up with a very basic solution. We have used the official data set that is being approved by World Health Organization(WHO). This data set is being provided by The Center for Systems Science and Engineering (CSSE) the Johns Hopkins Univeristy [22]. This dataset is very accurate and being daily updated in the dataset provided by CSSE at JHU.

This dataset consists of various field. One good thing about this dataset is that it is in CSV format. It is in the form of comma separated file; the CSV format makes it more compatible, readable and easier for a programmer and analyst to analyse this dataset. The dataset consists of various fields and features, such as date, province or state, country, lat, long, confirmed, recovered, deaths, active, etc etc they may be seen in the figure below. The data set include a time series of tracking of humans by date, including confirmed cases the number of people being recovered, mortality rate and etc.

The dataset consists of 41590 rows and nine columns, nine columns that are being used in dataset are as follows.

- Date
- Province/State
- Country
- Lat

- Long
- Confirmed
- Recovered
- Deaths
- Active



	Date	Province/State	Country	Lat	Long	Confirmed	Recovered	Deaths	Active
0	2020-01-22	NaN	Afghanistan	33.0000	65.0000	0	0	0	0
1	2020-01-23	NaN	Afghanistan	33.0000	65.0000	0	0	0	0
2	2020-01-24	NaN	Afghanistan	33.0000	65.0000	0	0	0	0
3	2020-01-25	NaN	Afghanistan	33.0000	65.0000	0	0	0	0
4	2020-01-26	NaN	Afghanistan	33.0000	65.0000	0	0	0	0
...	...	...	...	...	...	...	...	...	...
41575	2020-06-19	NaN	Timor-Leste	-8.8742	125.7275	0	24	0	-24
41576	2020-06-20	NaN	Timor-Leste	-8.8742	125.7275	0	24	0	-24
41577	2020-06-21	NaN	Timor-Leste	-8.8742	125.7275	0	24	0	-24
41578	2020-06-22	NaN	Timor-Leste	-8.8742	125.7275	0	24	0	-24
41579	2020-06-23	NaN	Timor-Leste	-8.8742	125.7275	0	24	0	-24

41580 rows × 9 columns

Figure 3.1 Data in tabular form

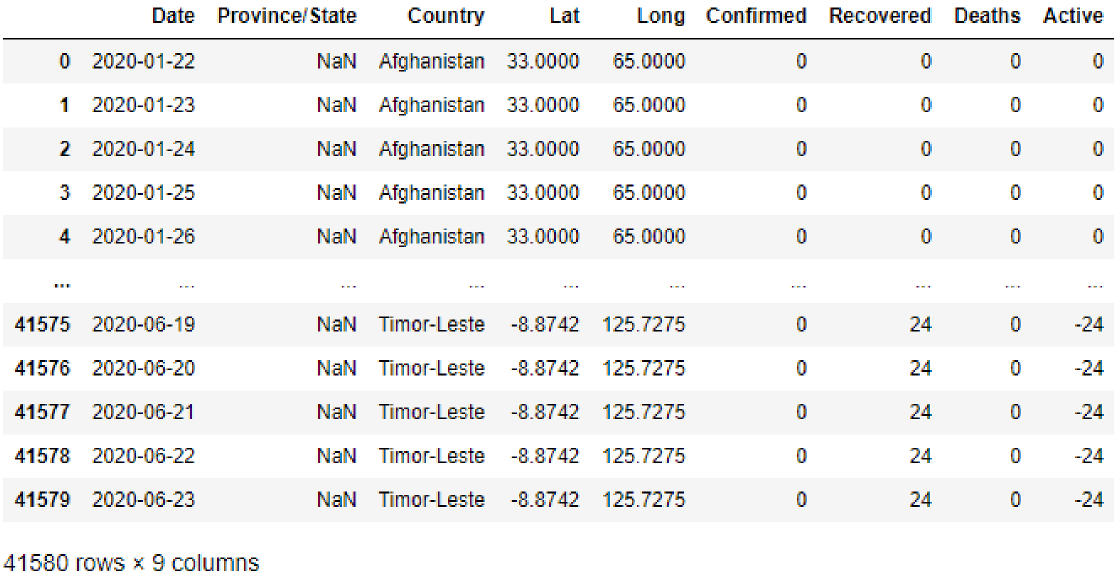
Date tells the date of cases, the date on which case has been reported positive. Province/State is the state or location where the case has been reported, Similarly Country column is used to mark the country on which the case has been reported. Confirmed cases tells the number of cases that has been confirmed so far and are positive. Recovered column signify the number of recovered cases till date.

3.2. DATA PRE-PROCESSING

Most of the time, the real-world data or the raw data is not complete, not consistent, or is full of so many errors. To transform the data into an understandable form, pre-processing of data or information or knowledge is done. Data pre-processing is the method and a procedure or a way for resolving the unwanted issues before applying algorithms. Some of the data pre-processing steps involves cleaning, integration, transformation, reduction, etc.

The dataset consists of 41590 rows and nine columns, Nine columns that are being used in dataset are as follows.

- Date
- Province/State
- Country
- Lat
- Long
- Confirmed
- Recovered
- Deaths
- Active

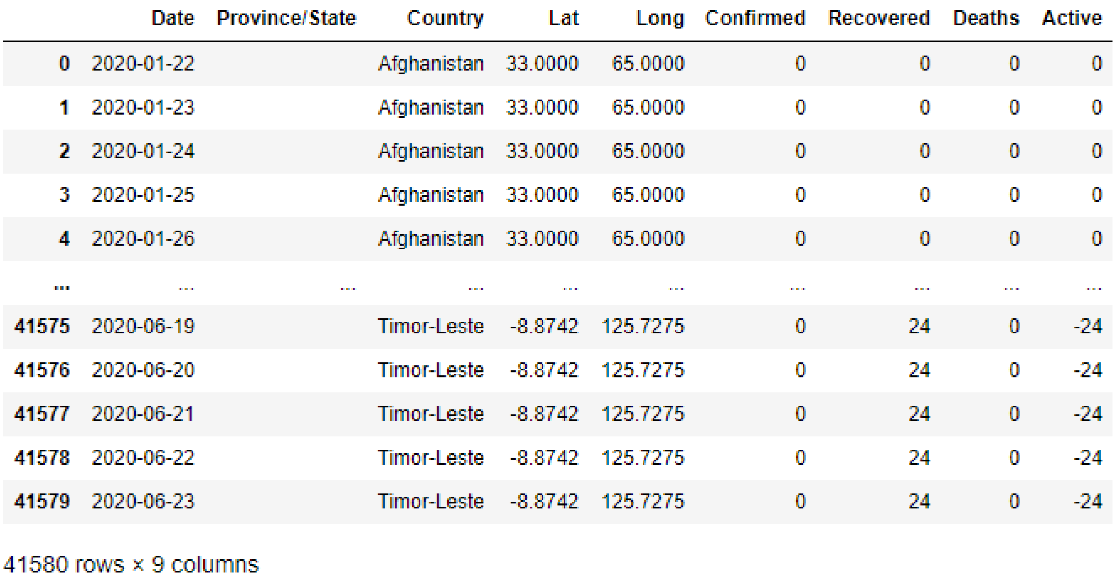


	Date	Province/State	Country	Lat	Long	Confirmed	Recovered	Deaths	Active
0	2020-01-22	NaN	Afghanistan	33.0000	65.0000	0	0	0	0
1	2020-01-23	NaN	Afghanistan	33.0000	65.0000	0	0	0	0
2	2020-01-24	NaN	Afghanistan	33.0000	65.0000	0	0	0	0
3	2020-01-25	NaN	Afghanistan	33.0000	65.0000	0	0	0	0
4	2020-01-26	NaN	Afghanistan	33.0000	65.0000	0	0	0	0
...	...	...	...	...	...	...	...	...	...
41575	2020-06-19	NaN	Timor-Leste	-8.8742	125.7275	0	24	0	-24
41576	2020-06-20	NaN	Timor-Leste	-8.8742	125.7275	0	24	0	-24
41577	2020-06-21	NaN	Timor-Leste	-8.8742	125.7275	0	24	0	-24
41578	2020-06-22	NaN	Timor-Leste	-8.8742	125.7275	0	24	0	-24
41579	2020-06-23	NaN	Timor-Leste	-8.8742	125.7275	0	24	0	-24

41580 rows × 9 columns

Figure 3. 1 Dataset after processing

As seen in the figure of dataset State is null and our data is not cleaned, Data contains so much of noise and things which is of no use for us. Such things in future can led to less accuracy of analysis of result. Result can be less accurate because of noise present in the dataset. Hence to overcome this and for good result we are cleaning our data and making dataset more précised. First of all, we will be making State entry as "blank" as they are "NAN" in the figure above. After making all those entries blank the view of our dataset is being shown in figure below.



	Date	Province/State	Country	Lat	Long	Confirmed	Recovered	Deaths	Active
0	2020-01-22		Afghanistan	33.0000	65.0000	0	0	0	0
1	2020-01-23		Afghanistan	33.0000	65.0000	0	0	0	0
2	2020-01-24		Afghanistan	33.0000	65.0000	0	0	0	0
3	2020-01-25		Afghanistan	33.0000	65.0000	0	0	0	0
4	2020-01-26		Afghanistan	33.0000	65.0000	0	0	0	0
...	...	...	...	...	...	...	...	...	...
41575	2020-06-19		Timor-Leste	-8.8742	125.7275	0	24	0	-24
41576	2020-06-20		Timor-Leste	-8.8742	125.7275	0	24	0	-24
41577	2020-06-21		Timor-Leste	-8.8742	125.7275	0	24	0	-24
41578	2020-06-22		Timor-Leste	-8.8742	125.7275	0	24	0	-24
41579	2020-06-23		Timor-Leste	-8.8742	125.7275	0	24	0	-24

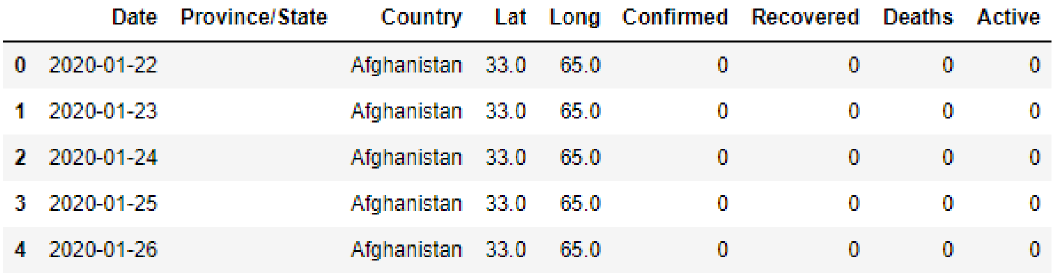
41580 rows × 9 columns

Figure 3. 2 Plotting average values from 'low' and 'high' values

Deploying country wise data now Country day wise, country wise and day wise. Now we will be managing data country wise so that we can get information regarding country wise data.

df.head() will return this output.

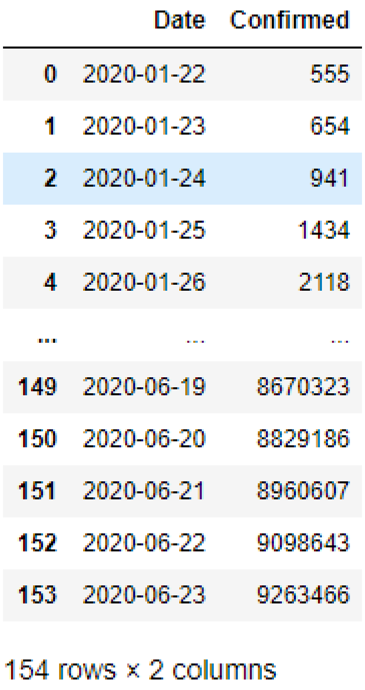
Since there is no date in country wise data, so we have removed our parser from country wise data



	Date	Province/State	Country	Lat	Long	Confirmed	Recovered	Deaths	Active
0	2020-01-22		Afghanistan	33.0	65.0	0	0	0	0
1	2020-01-23		Afghanistan	33.0	65.0	0	0	0	0
2	2020-01-24		Afghanistan	33.0	65.0	0	0	0	0
3	2020-01-25		Afghanistan	33.0	65.0	0	0	0	0
4	2020-01-26		Afghanistan	33.0	65.0	0	0	0	0

Figure 3.4 Head of data

Now we will be analyzing confirmed cases that too date wise. The output return by the query is shown in figure below.



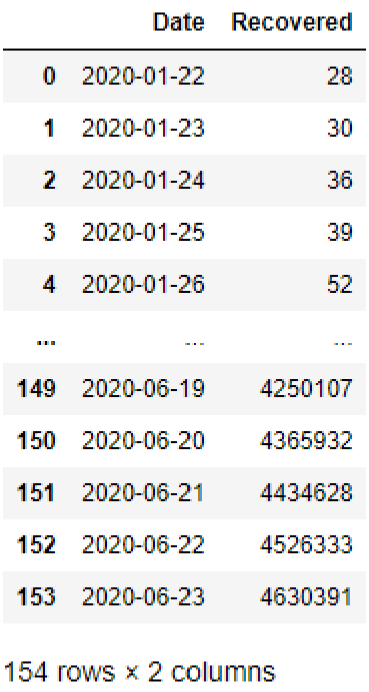
	Date	Confirmed
0	2020-01-22	555
1	2020-01-23	654
2	2020-01-24	941
3	2020-01-25	1434
4	2020-01-26	2118
...	...	...
149	2020-06-19	8670323
150	2020-06-20	8829186
151	2020-06-21	8960607
152	2020-06-22	9098643
153	2020-06-23	9263466

154 rows × 2 columns

Figure 3.5 Confirmed case till date

This data reveals that and output of query tells that as shown in figure above this data tell that there is around 555 cases on 22nd January 2020 this is surprise and shocking to know by our output data that number of cases of covid-19 positive has been increase exponentially in a very short time power data output reveals are there are around 9263466 bases on 23rd June 2020 Reported worldwide in more than 200 countries.

Now we will be looking at the cases of covid-19 who have been recovered so far, output data tell that on 22nd January 2020 there are around 28 cases who have been recovered by covid-19 disease if you look at data more accurately we will come to know that there are 4630391 On 23rd June 2020 itself is a good number.

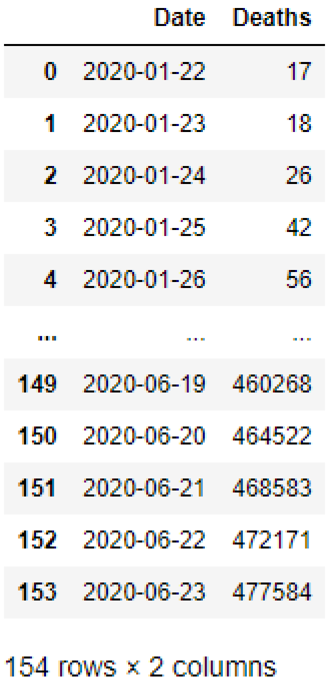


	Date	Recovered
0	2020-01-22	28
1	2020-01-23	30
2	2020-01-24	36
3	2020-01-25	39
4	2020-01-26	52
...	...	...
149	2020-06-19	4250107
150	2020-06-20	4365932
151	2020-06-21	4434628
152	2020-06-22	4526333
153	2020-06-23	4630391

154 rows × 2 columns

Figure 3.6 People recovered till date

As we have seen the total number of cases that have been reported by COVID-19 and I have also seen the total number of who have been recovered covid-19 positive disease now we will be looking forward to seeing how many death till date on 23rd 2020 if we look at our data more accurately that 477584 deaths have been reported till 23rd June 2020 around the world. The number of death that have been reported in a very short time is itself a shocking thing to see because of this World Health Organisation has stated this disease of COVID-19 as a pandemic disease in nature. This data can be more accurately viewed in figure below.



	Date	Deaths
0	2020-01-22	17
1	2020-01-23	18
2	2020-01-24	26
3	2020-01-25	42
4	2020-01-26	56
...	...	...
149	2020-06-19	460268
150	2020-06-20	464522
151	2020-06-21	468583
152	2020-06-22	472171
153	2020-06-23	477584

154 rows × 2 columns

Figure 3.7 Number of deaths till date

Now we will be making our data More clean and accurate so that we can get a good accuracy in our result for making our data clean we will and some columns as per our requirement. The data that which we have observed in the form of output after removing useless data and the data which is not required by us in the figure below we can observe that now our table and data set is more clean and we can expect a good accuracy in our result.

For analysing sentiment of people on social media platform search engine such as Facebook Instagram YouTube Twitter Google Hangout. We will be majorly focusing Twitter we will be analysing sweet that is of our use. In Twitter tweets there is so much of information which will be of less use for our requirement so we will be cleaning all those information which is having les importance in our project. At Tweets on Twitter consist of various symbol and various data such as at the@ username retweet like button, etc etc.

For every machine learning algorithm to work appropriately and accurately, data is the crucial requirement. Information is the heart of the Machine learning algorithm. On twitter, data is present in any form. It can be audio, video, image, and text. In this paper, we are concentrated on textual data, which is a tweet done by users on twitter during the coronavirus period of coronavirus. Data, knowledge, and information present on twitter is used in the various way here we are using it to analyze the sentiment of people on twitter regarding coronavirus. We are using data from multiple sources, and for data to be in good amount, we are using various external sources and library of python. We can install tweepy in python by using the command "pip install tweepy" The primary source remains twitter only. To gather the data, we have followed a simple approach to using hash-tags. We used python library Tweepy and twitter developer application programming interface. For using twitter API, one should have a twitter developer account and permissions from twitter for fetching data. We have tried to use the unique dataset for this particular sentimental analysis. Many tweets were extracted using tweepy and textblob. Textblob is a python API for getting more knowledge and information about the textual data in natural language processing. Hashtags used for extracting tweets are as #covid19, #covid-19, #positive #negative #corona, #virus, #COVID19, #coronaviras, #corona-virus, #covid19-virus, #sarscov2, #China, #china_virus, #Chinese_virus #Wuhan, #who etc. For variety in the dataset, we have also considered top trending hashtag during May and June. The top 20 hashtags during this period are as follows shown in the figure below.

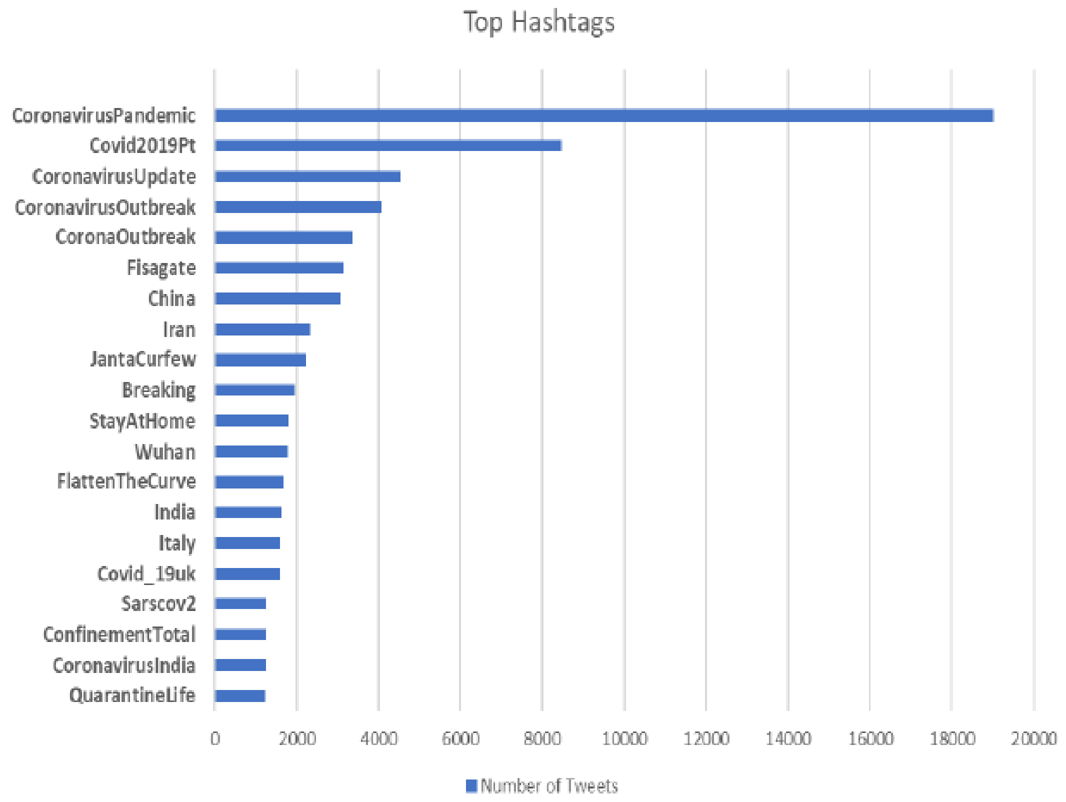


Figure 3.8 Top hashtag during the period

We have also tried to find out the sources of tweets done by users of twitter. By analyzing this, we have found that most number of tweets done in May and June is by android users. Sign in, and tweet done by android users are more in number as compared to other users such as i-phone and web ap of twitter. 37% of the total population on twitter use twitter from the android phone during the COVID-19 period, whereas 34% of users have used i-phone for signing in to twitter and doing tweets related to the novel coronavirus and COVID-19. 20% of the total population has used twitter and done tweets related to pandemic COVID-19 by using their desktop or laptop by web app of twitter. After data acquisition, the most crucial step is to clean the data. Data should be clean in such a way that our need's data should only be left for more analysis. All the data which is not of our use is being thrown in the garbage, and further necessary operations are performed on the data of our purpose. The data which we have gather creates a lot of noise and is not in a structured manner. Hence there is a need to remove noise and useless items present in the data. Since our data contain tweets, there is a high tendency that it will contain unnecessary and random symbols. Therefore, we are applying some cleaning processes on this data obtained. First of all, we are removing URL, also known as Uniform Resource Locator, from our dataset as URL does not contain any sentiment, and they are of no use for our machine learning model. After removing URL with the help of the NLTK library inbuild in python. We will now move a little forward and clean the dataset in a manner that it does not contain any email. Cleaning of email and username is done using the NLTK library included in python. After deleting this, we will be deleting and removing more unstructured data with more noise. This includes deleting of mention and tagging. In twitter, users generally tag and mention by using a particular symbol "@." Eliminating naming's "@" special symbol and the word after @ symbol, which only means another name of user present on the tweet or tagging the other user. As the name of the user does not contain any sentiment or emotions in it, we will be removing this by using a regular expression. After removing all this, then comes the turn of removing hashtags. In our dataset, we have used a variety of hashtags. Hashtags do not specify any emotions or opinions or sentiments. We cannot detect any feelings by using the hashtag; therefore, we will be removing all the hashtags by using a regular expression. Hashtags in Twitter are being used by mentioning a special symbol "#"(hash). Hashtag starts with "#" using a regular expression, we will be deleting and eliminating all the hashtags. As of now, much or more noise has been removed from data, and data is more cleaned and structured as compared to the information we gathered in our first stage. After this, we have tried to remove the newline and whitespace. We have removed newlines and extra whitespaces from tweets collected in our dataset. Newline and whitespace do not signify and convey any emotions, or opinions or any sentiments. Hence, we have removed this and made our data cleaner and less noisy

We have also tried to split the multiple tweets into a new list of words by applying the wonderful Python NLTK Tweet Tokenizer and tried to signify diverse classes of information to a particular name utilizing unique Python NLTK WordNetLemmatizer. In this phase, we have also neglected the retweet done by the various user because they signify the same emotion or feeling or opinion or sentiment. The various phases of tweet sentimental analysis can be seen in below figure.

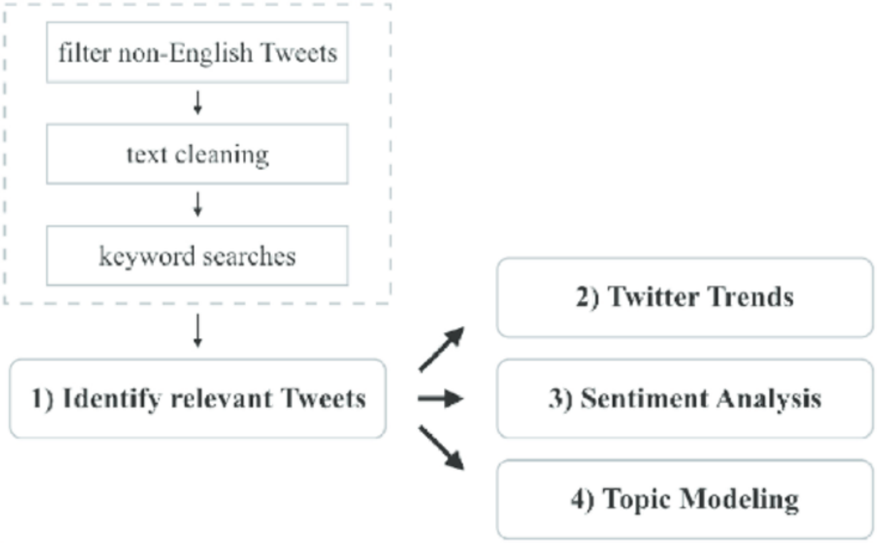


Figure 3.9 Sentimental analysis of tweets

3.3 METHODOLOGY

After pre-processing, we now move on to our transformer model architecture.

3.3.1 Model Architecture for sentiment analysis

Performing sentimental and on we will be following one at torch we will be doing with scrapping of tweet that has been related to coronavirus we will be using various # hashtags then we will scrap all those tweets using web scrapper and tweepy library. Now we will call and store them in a data set after storing them in our data set we will process and clean tweet after doing the people says and cleaning all the tweet we will then exact list of word which are of our use after doing this we will mean we will maintain in which we will be storing the word which are showing negative emotions and negative sentiments.

Various # hashtags then we will scrap all those tweets using web scrapper and tweepy library.

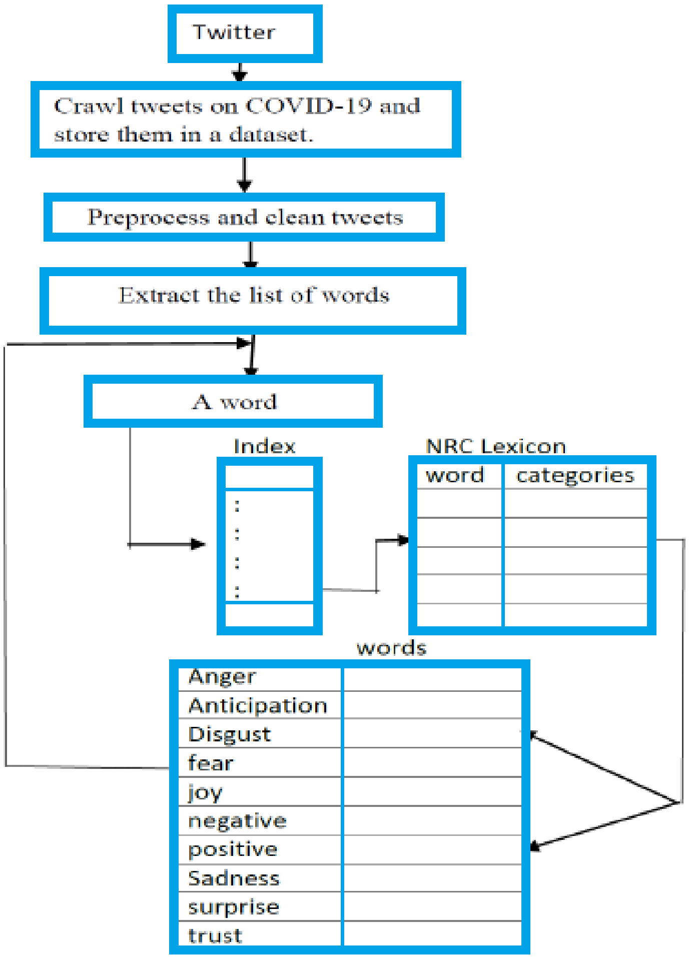


Figure 3.10 COVID tweet analysis

Performing sentimental and on we will be following one at torch we will be doing with scrapping of tweet that has been related to coronavirus we will be using various # hashtags then we will scrap all those tweets using web scrapper and tweepy library. Now we will call and store them in a data set after storing them in our data set we will process and clean tweet after doing the people says and cleaning all the tweet we will then exact list of word which are of our use after doing this we will mean we will maintain in which we will be storing the word which are showing negative emotions and negative sentiments such as anticipation fear sadness surprise Now we will generate the word cloud which is being generated by our tweet after cleaning them The word cloud by our data set is shown below.

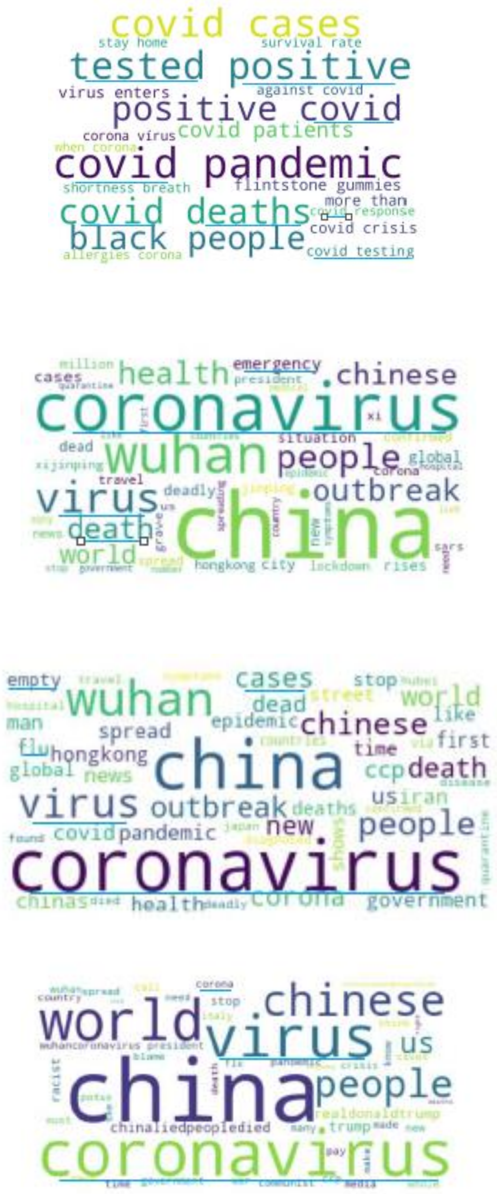


Figure 3.11 Word cloud of tweet

A word cloud is a typical representation of textual data left after cleaning. It gives a visualizing idea regarding each word in the cloud. The word cloud observed by our data is as follows shown in the figure. We can find that, the cloud consists of word such as Coronavirus, COVID-19, china, virus, deaths, etc etc. The tweets continued designed within word clouds to examine which terms have remained continuously applied by Twitter users throughout the world. As word cloud visualization consists of the measurement and visible importance of information being weighted through their frequency regarding existence within the textual corpus, we can further procure insights from the most frequent words occurring. Most frequent words are 'COVID-19', 'china', 'corona', or 'virus' words like 'corona, 'positive' made a massive quantity of mentions. As the cases around the world are rising day by day, there will be a significant increase in those numbers in the world cloud also, as they are directly related to feelings or emotions of users using twitter. In the above figure, we can clearly see that on June 1, the word "test" has been used more than 2300 times in tweets. We can analyze the emotion or sentiment clearly after this step and by using the word cloud.

3.3.2 Logistic Regression

Logistics regression (LR) is a famous and more advanced classifier for classification. David Cox first developed Logistic Regression in 1958. In this machine learning classifier model, the probabilities representing the likely results concerning an individual action are shaped using a logistic function. Using a log function, the possibility regarding those consequences transpires converted within binary states (0 and 1). Most likelihood estimate classifications are usually practiced to reduce the chance of a fault within that model. Furthermore, the resistance concerning the logistic regression classifier remains cheaper than that of another classifier due upon the extensive spread of the costs of average distribution efficiency. Logistic regression (LR) remains one probabilistic distribution system that may be applied for supervised machine learning models.

For analysis, a machine learning model consists typically concerning subsequent elements such as

1) A feature description of the information

2) A function for classification

3) An objective function

4) An algorithm used for the optimization.

For logistic regression, we are using a particular function called the sigmoid function to build the binary classifier. Let us consider an input comment or tweet 'x', which is a tweet or comment in our case and further is denoted by a vector of features $[X1\ X2\ X3\ X4 \ldots \ldots\ Xn]$. In logistic regression, the classifier's output contains two values only as it is discrete in nature, and that will be either y = 1 or y = 0. The purpose of this classifier is to understand $P(y = 1|x)$, which signifies the probability concerning sentiment that is positive within the classification regarding Tweets related to Coronavirus, whereas $P(y = 0|x)$, that correspondingly signifies the probability concerning sentiment that is negative within the classification regarding Tweets related to Coronavirus.

wi is denoting the mass of input taken, characteristic xi from a set that is used for training furthermore b is meaning the intercept or bias term, we will perceive the outcome weighted amount toward a particular class as: -.

$$z \equiv \sum_{i=1}^{n} w_i x_i + b \tag{3.1}$$

The dot product of vector w and vector x will be as: -

$$z = w.x + b \tag{3.2}$$

Now we will be using a sigmoid function to map our value in the interval range of [0,1]. It is described as below: -

$$y = \frac{1}{1+e^{-z}} \tag{3.3}$$

After doing sigmoid function this, we will get our desired result for both positive and negative value:

$$p(y = 1 / \varkappa) = \frac{1}{1+e^{-w_x-b}} \tag{3.4}$$

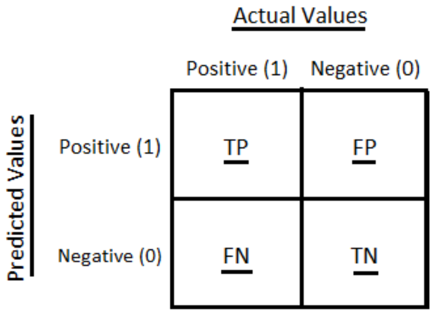
$$P\left(y = {}^{0}/_{x}\right) = \frac{e^{-\omega x-b}}{1+e^{-wx-b}} \tag{3.5}$$

Gradient descent method is being used for minimizing the loss function: -

$$L(\bar{Y}, y) = -[y \log\left(\frac{1}{1+e^{-w_x-b}}\right) + (1-y)\log(\frac{e^{-w_x-b}}{1+e^{-w_x-b}})] \tag{3.6}$$

Now we will perform partial derivative of the equation and result will be as follows: -

$$\frac{dL(w,b)}{dw} = [\sigma(w.x+b) - y].x\imath \tag{3.7}$$



Predicted Values \ Actual Values	Positive (1)	Negative (0)
Positive (1)	TP	FP
Negative (0)	FN	TN

Figure 3.12 Accuracy flow

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{3.8}$$

3.3.3 Naïve Bayes

The crux of Naive Bayes classifier is based on the Bayes algorithm, which is on foundation of conditional probability. Conditional probability Bayes theorem is:-

$$P\left(\frac{x}{y}\right) = \frac{P\left(\frac{y}{x}\right).P(x)}{P(y)} \tag{3.9}$$

Where $P(x/y)$ means probability of occurrence of event x given that event y has already been occurred.
This can also be further written as

$$P\left(\frac{x}{y}\right) = \frac{P(x\cap y)}{P(y)} \tag{3.10}$$

Where they can be easily understand and read in figure below as shown

$$\underbrace{P(c\,|\,x)}_{\text{Probability}} = \frac{\overbrace{P(x\,|\,c)}^{\text{Likelihood}}\,\overbrace{P(c)}^{\text{Prior Probabilty}}}{\underbrace{P(x)}_{\text{Predictor Probability}}}$$

$$P(c\,|\,\mathrm{X}) = P(x_1\,|\,c)\times P(x_2\,|\,c)\times\cdots\times P(x_n\,|\,c)\times P(c)$$

Training of classifier: -
In the aforementioned process, we will be finding the value of $P(c)$

$$P(C) = \frac{N_C}{N_T} \tag{3.11}$$

$$P(w) = \frac{count(t,c)+1}{\sum w\in V(count(t,c)+|T|)} \tag{3.12}$$

The difference between Logistic Regression and Naïve Bayes classifier is shown in the table below. It is differentiated between various strength weakness and characteristics.

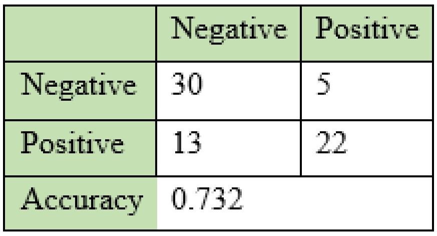
Classifier	Characteristic	Strength	Weakness
Logistic regression	Probability of an outcome is based on a sigmoid or logistic function	Transparent and easy to understand; Regularized to avoid over-fitting	Expensive training phase; Assumption of linearity
Naïve Bayes classifier	Depends upon guess concerning independence between predictor variables	Effective with real world data; Efficient and can deal with dimensionality	Over-simplified assumptions; Limited by data scarcity

Table 3.1 Difference of classifiers

CHAPTER 4 RESULTS

Accuracy of Logistic Regression

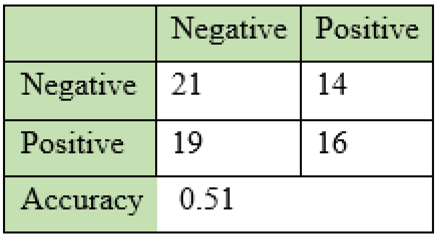
We have applied logistic regression classifier on our cleaned data set in tow form. First, we have applied logistic regression to the tweets of less than 77 characters around 30 percent of tweets that have character less than 77. We have used python and Jupyter notebook to train our model by using the library sci-kit learn we have trained our model and test the data. Confusion matrix will convey the accuracy of logistic regression model as: 73.2%, Confusion matrix for logistic regression is as shown in figure



	Negative	Positive
Negative	30	5
Positive	13	22
Accuracy	0.732	

Figure 4.1 Logistic regression accuracy when word length is less

After this we have find accuracy of our model and run our model on tweets that are having 125 characters tweets that are having less than 125 characters, Accuracy of that model is as :- 51.9 %. The confusion matrix is as shown below in figure.

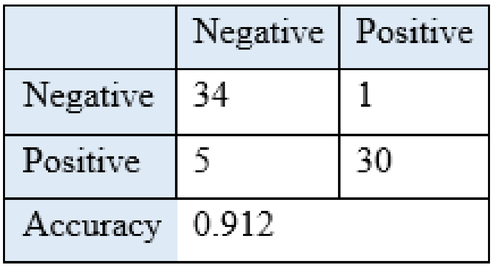


	Negative	Positive
Negative	21	14
Positive	19	16
Accuracy	0.51	

Figure 4.2 Logistic regression Accuracy when word length is more

Accuracy of Naive Bayes

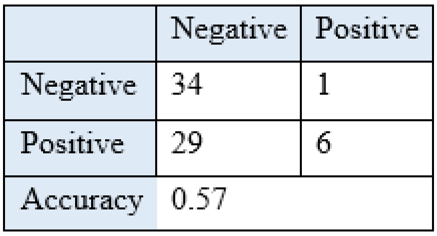
Similar to logistic regression in naive Bayes also we have divided and tested our model on two horizontals, the first one being tweets having character less than 70 and other being tweets having character less than 125. By making confusion matrix for the first case, we have observed an accuracy of 91 % as shown in figure below: -



	Negative	Positive
Negative	34	1
Positive	5	30
Accuracy	0.912	

Figure 4.3 Naive Bayes accuracy when word length is less

The second case where tweets have character less than 125 an accuracy observed as 57 %. Confusion matrix for both the cases are shown below

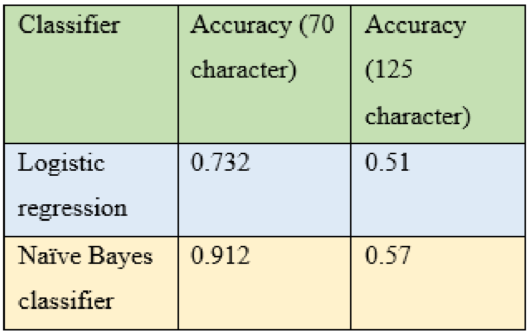


	Negative	Positive
Negative	34	1
Positive	29	6
Accuracy	0.57	

Figure 4.4 Naive Bayes accuracy when word length is more

Observation of Logistic Regression and Naïve Bayes

The comparison of naïve bayes and logistic regression is done below.



Classifier	Accuracy (70 character)	Accuracy (125 character)
Logistic regression	0.732	0.51
Naïve Bayes classifier	0.912	0.57

Figure 4.5 Complete accuracy comparison

Number of Tweets done related to COVID: -

With the help of the below figure, we can clearly see that there is around a very a smaller number of tweets on three May 2020. can clearly see that there are more number of tweets near about 7000 which have been related with our topic coronavirus on 20 may 2020.

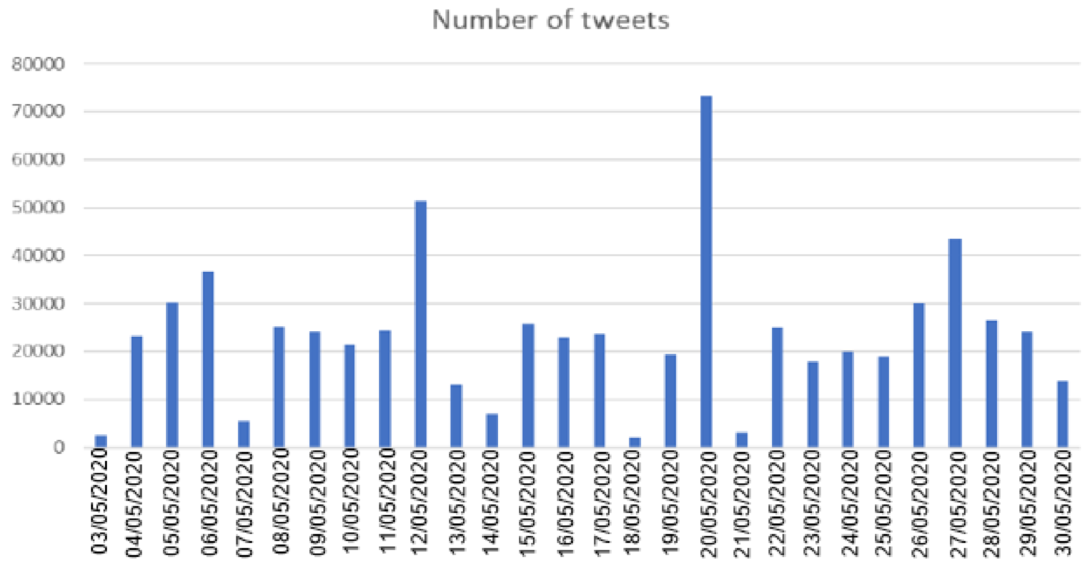


Figure 4.6 Total number of tweets.

Furthermore in other figure we can see that sentiment of tweets that have been released to a topic are as follows The topic which have been widely used during the time period of March 15, to June 23 are as covid-19, patients, positive, President, Donald Trump, news, ppe, etc etc. This can be seen clearly with the help of below figure. We can also see the number of tweets that have been done during the given period

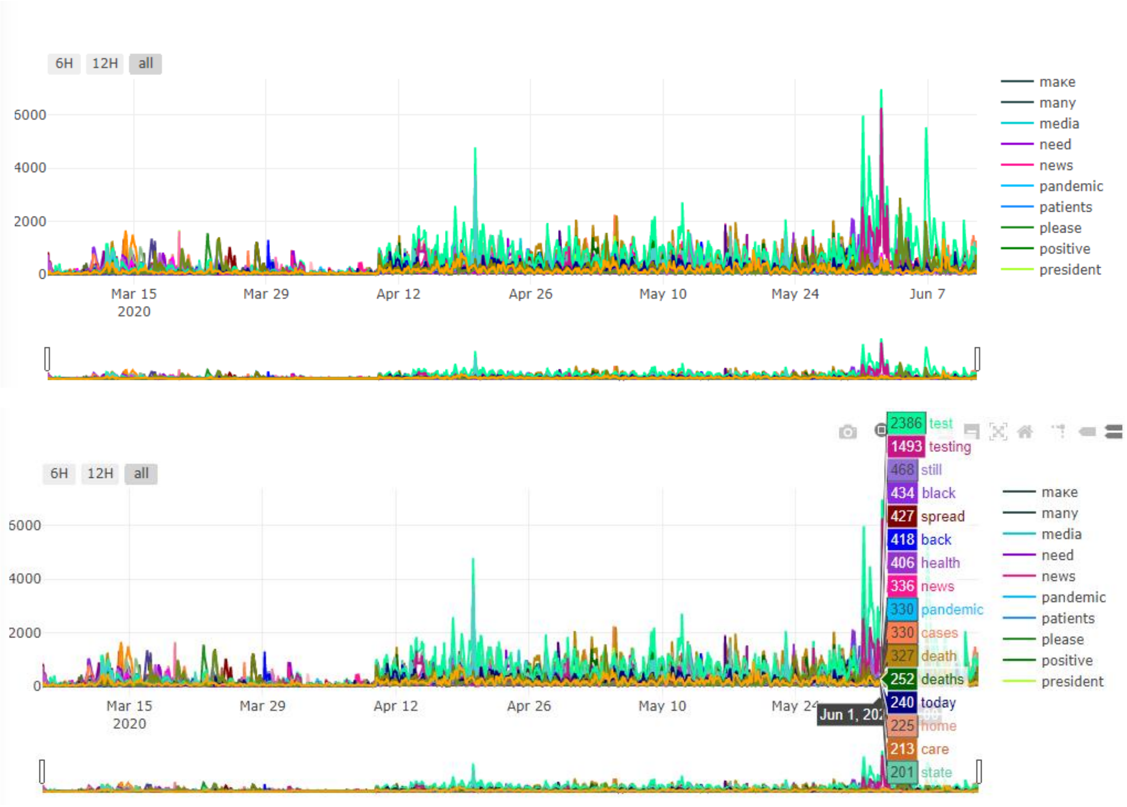


Figure 4.7 Number of time word has been used.

Comparison of Confirmed, Recovered and Death Cases:-

The comparison of Confirmed, Recovered and Number of Death that has been reported till 23 June 2020 around the world because of COVID-19 is as shown below, Number of cases are on Y scale and date is on X scale.

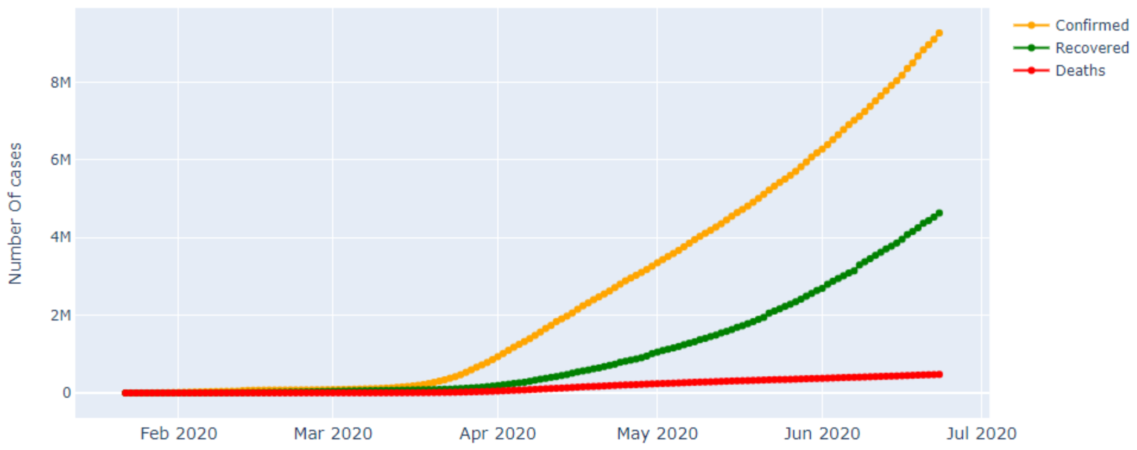


Figure 4.8 Count of confirmed recovered and death cases

Cases over the time with area plot:-

The number of cases over the time with area plot is shown in figure below we can see that there are 4155491 Active cases, 477584 Deaths and People recovered 4630391 so far.

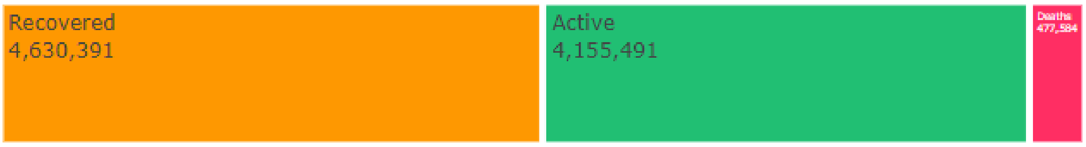


Figure 4.9 Cases over the time

In the below figure which tells about cases over the time we can observe that, As the date is increasing there is an increase in number of we can also see that in the beginning number of cases are less as number of days pass the number of cases reported as also increase number of death has also increased and number of cases that have been recovered has also increased with respect to time and date hence we can say that in the entire world if we compare the date of the world and each country number of cases are increasing day by day with respect to number of increase in days with respect to date and time.

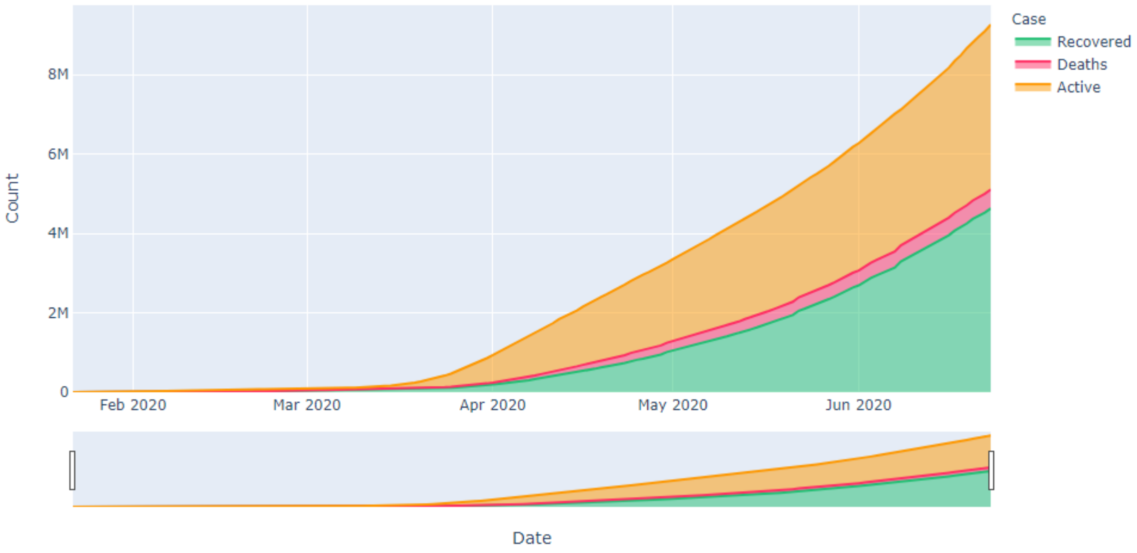


Figure 4.10 Detailed cases count.

Death and recovery per 100 cases:-

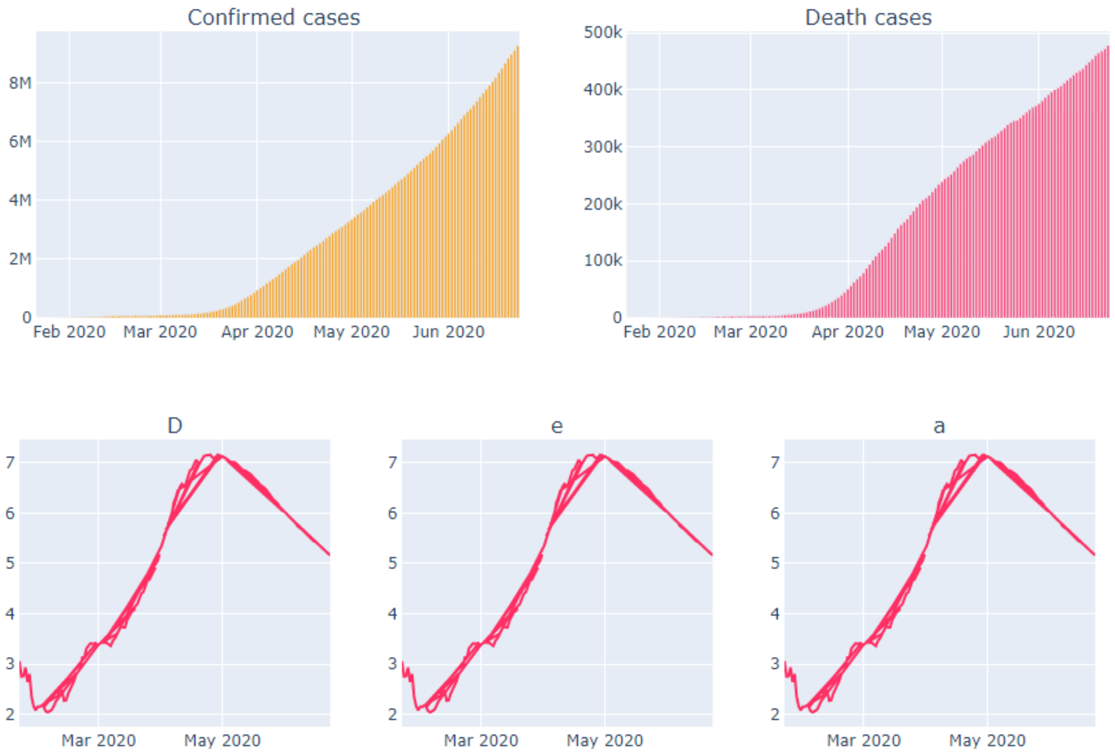


Figure 4.11 Death and recovery cases.

In the above figure we can observe that there is a peak in the graph once the number of cases has been increased there will come a peak in graph and after that the number of cases who are being tested positive will get decrease.

Confirmed Cases and Number of Death

Now we will be saying number of confirmed cases and number of deaths reported in the below graph and figure we can see that if we look at confirmed cases country like United State Russia India Italy and France have reported highest number of compounds aces for covid-19 that have been reported positive while testing now let us look at the number of deaths reported country wise find that there is Amazon number of death reported in country majority of number of deaths reported in country like US, Brazil, France, Mexico, etc.

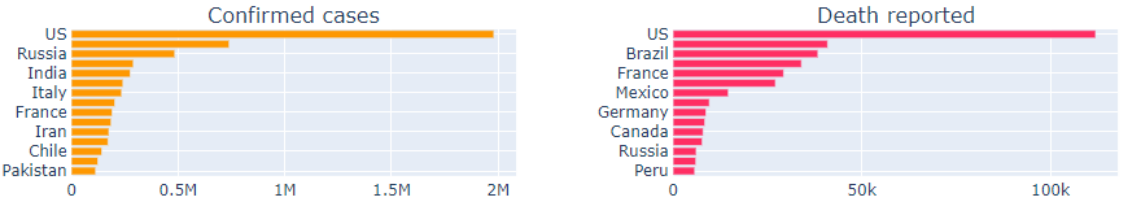
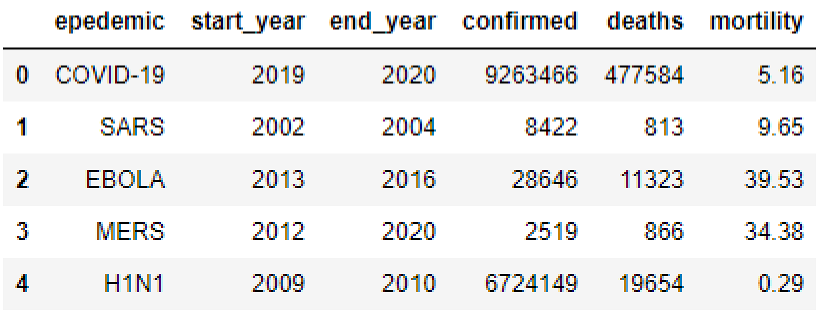


Figure 4.12 Confirmed and death cases count

Comparison with other disease



	epedemic	start_year	end_year	confirmed	deaths	mortility
0	COVID-19	2019	2020	9263466	477584	5.16
1	SARS	2002	2004	8422	813	9.65
2	EBOLA	2013	2016	28646	11323	39.53
3	MERS	2012	2020	2519	866	34.38
4	H1N1	2009	2010	6724149	19654	0.29

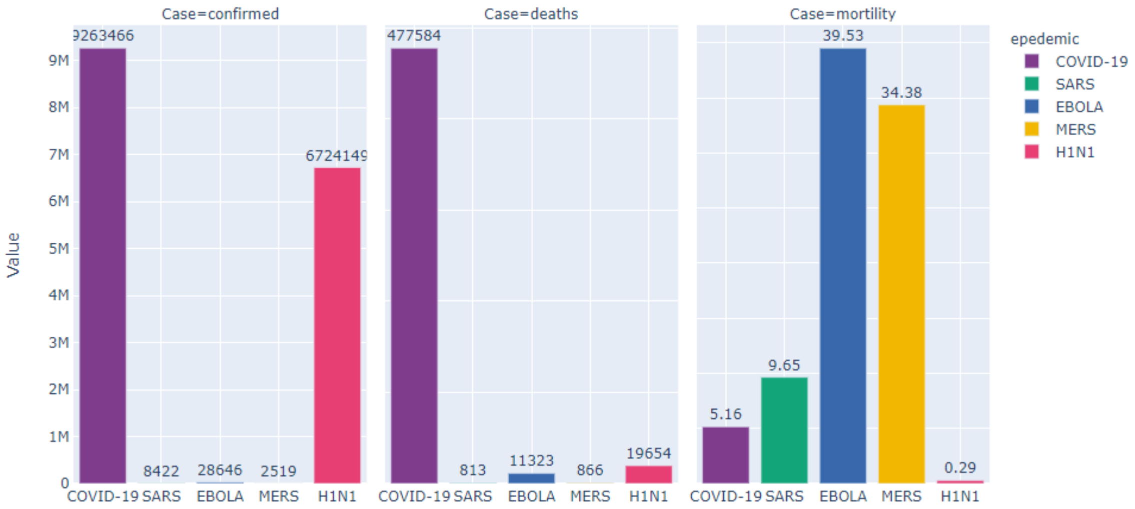


Figure 4.13 Comparison with other diseases.

We can clearly see that COVID-19 is a pandemic disease that has reported a huge number of cases in a very short time. If compared with other disease it has also taken the most number of death in this comparison. Hence there is a much need and demand for it's vaccine and remedy.

CHAPTER 5 FUTURE WORK AND CONCLUSION

We ought tried to examine tweets and comments of user. We have seen the result in both the cases of logistic regression and the Naive Bayes classifier algorithm. Classifier that is based on Bayes theorem and is particularly known as Naive Bayes has performed well if we compared it with logistic regression classifier when the number of characters is less than 75. In contrast, in the other case, also the Naive Bayes classifier has performed well when the count of character in the tweet is less than 125. Below explore the more advantage and disadvantage of logistics regression and Naive Bayes classifier when applied upon the textual data of twitter for analysing sentiments. The comparison of the accuracy of logistic regression and the Naive Bayes classifier on various character length of a tweet is compared. We have tried to analyse tweets and comments of a user during a pandemic COVID-19, we have cleaned the data as per our requirement so that our model should get simple and clear. Checking sentiment of the user using tweet itself shows good result in the various incident earlier as well. We have tried to approach the issues which users are facing during COVID-19 pandemic, one such problem in the diverse country is lack of testing kits, shortage of mask, shortage of personal protective equipment kits, one major issue and concern is some country is stuck of people during lockdown and shortage of food for the people who have got stuck in lockdown. Another major issue remains in some state is for the lack of beds in hospitals. This all has clearly shown that people are having a negative sentiment regarding this COVID-19 disease. This has led to a growth in fear and negative emotion in users around the world. As future work, we will be trying to use more sentimental analysis algorithm on the data. Good data collection will lead to good accuracy of models and good result in future. More techniques to clean and structure the data can be used. Training the model on the various algorithm can result in more accuracy with different training and test split. Another possible future work can be to explore the way the virus has spread in the entire world. One more approach is to analyse the data by comparing the number of populations of the country with the number of cases by taking into account the number of people recovered and mortality rate. As social distance is one of the ways to make this virus spread less taking area of each country in the account can give an exciting and functional result.

The deep analysis of COVID-19 with respect to various parameters can help in understanding the need of it’s vaccine and solution. It can also help in understanding the impact of COVID-19 on various industry and other sectors. This analysis can further tells that as of now wearing mask and making it as a habit can make the cases less and there will be a downfall in cases becoming positive. Social Distance is one of the best solutions for avoiding the risk of getting COVID-19 positive. Graph of COVID-19 become flat once the cases have reach to it’s peak. It can be seen that once there is a peak in graph of COVID-19 cases the case will get decrease.

The techniques implemented in this study can further be extended by more work and research on advanced upcoming methods. The proposed model can be implemented with other upcoming technique so that it can operate in lesser time.

References

[1] References

[2] Mitsui, Takehiro, Keiko Iwano, Kazuo Masuko, Chikao Yamazaki, Hiroaki Okamoto, Fumio Tsuda, Takeshi Tanaka, and Shunji Mishiro. "Hepatitis C virus infection in medical personnel after needlestick accident." Hepatology 16, no. 5 (1992): 1109-1114.

[3] Black, Craig Patrick. "Systematic review of the biology and medical management of respiratory syncytial virus infection." Respiratory care 48, no. 3 (2003): 209-233.

[4] Sars, Georg Ossian. An Account of the Crustacea of Norway: with Short Descriptions and Figures of all the Species. Vol. 4. Bergen Museum, 1903.

[5] Leung, Wai K., Ka-fai To, Paul KS Chan, Henry LY Chan, Alan KL Wu, Nelson Lee, Kwok Y. Yuen, and Joseph JY Sung. "Enteric involvement of severe acute respiratory syndrome-associated coronavirus infection." Gastroenterology 125, no. 4 (2003): 1011-1017.

[6] "Home." World Health Organization. World Health Organization. Accessed July 3, 2020. https://www.who.int/.

[7] Novel, Coronavirus Pneumonia Emergency Response Epidemiology. "The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China." Zhonghua liu xing bing xue za zhi= Zhonghua liuxingbingxue zazhi 41, no. 2 (2020): 145..

[8] World Health Organization. Population-based age-stratified seroepidemiological investigation protocol for coronavirus 2019 (COVID-19) infection, 26 May 2020. No. WHO/2019-nCoV/Seroepidemiology/2020.2. World Health Organization, 2020.

[9] Zhu, Na, Dingyu Zhang, Wenling Wang, Xingwang Li, Bo Yang, Jingdong Song, Xiang Zhao et al. "A novel coronavirus from patients with pneumonia in China, 2019." New England Journal of Medicine (2020)."

[10] Cascella, Marco, Michael Rajnik, Arturo Cuomo, Scott C. Dulebohn, and Raffaela Di Napoli. "Features, evaluation and treatment coronavirus (COVID-19)." In Statpearls [internet]. StatPearls Publishing, 2020.

[11] Li, Qun, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren et al. "Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia." New England Journal of Medicine (2020).

[12] Wan, Yushun, Jian Shang, Rachel Graham, Ralph S. Baric, and Fang Li. "Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus." Journal of virology 94, no. 7 (2020).

[13] Mehta, Puja, Daniel F. McAuley, Michael Brown, Emilie Sanchez, Rachel S. Tattersall, Jessica J. Manson, and HLH Across Speciality Collaboration. "COVID-19: consider cytokine storm syndromes and immunosuppression." Lancet (London, England) 395, no. 10229 (2020): 1033.

[14] World Health Organization. "Coronavirus disease 2019 (COVID-19): situation report, 72." (2020).

[15] Fauci, Anthony S., H. Clifford Lane, and Robert R. Redfield. "Covid-19-navigating the uncharted." (2020): 1268-1269.

[16] Grundy, Scott M. "Metabolic syndrome pandemic." Arteriosclerosis, thrombosis, and vascular biology 28, no. 4 (2008): 629-636.

[17] Agarwal, Apoorv, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca J. Passonneau. "Sentiment analysis of twitter data." In Proceedings of the workshop on language in social media (LSM 2011), pp. 30-38. 2011.

[18] Bakshi, Rushlene Kaur, Navneet Kaur, Ravneet Kaur, and Gurpreet Kaur. "Opinion mining and sentiment analysis." In 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 452-455. IEEE, 2016.

[19] Rosenthal, Sara, Noura Farra, and Preslav Nakov. "SemEval-2017 task 4: Sentiment analysis in Twitter." In Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017), pp. 502-518. 2017.

[20] Alpaydin, Ethem. Introduction to machine learning. MIT press, 2020.

[21] Rahm, Erhard, and Hong Hai Do. "Data cleaning: Problems and current approaches." IEEE Data Eng. Bull. 23, no. 4 (2000): 3-13.

[22] Berkhin, Pavel. "A survey of clustering data mining techniques." In Grouping multidimensional data, pp. 25-71. Springer, Berlin, Heidelberg, 2006.

[23] CSSEGISandData. "CSSEGISandData/COVID-19." GitHub. Accessed July 4, 2020. https://github.com/CSSEGISandData/COVID-19.

[24] Kleinbaum, David G., K. Dietz, M. Gail, Mitchel Klein, and Mitchell Klein. Logistic regression. New York: Springer-Verlag, 2002.

[25] Menard, Scott. Applied logistic regression analysis. Vol. 106. Sage, 2002.

[26] Hosmer, David W., and Stanley Lemesbow. "Goodness of fit tests for the multiple logistic regression model." Communications in statistics-Theory and Methods 9, no. 10 (1980): 1043-1069.

[27] McCallum, Andrew, and Kamal Nigam. "A comparison of event models for naive bayes text classification." In AAAI-98 workshop on learning for text categorization, vol. 752, no. 1, pp. 41-48. 1998.

[28] Sokolova, Marina, Nathalie Japkowicz, and Stan Szpakowicz. "Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation." In Australasian joint conference on artificial intelligence, pp. 1015-1021. Springer, Berlin, Heidelberg, 2006.

[29] Edwin Sin, Lipo Wang, "Bitcoin Price Prediction Using Ensembles of Neural networks", 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNCFSKD), Nanyang Technological University, Singapore, 2017.

[30] N.I. Indera, I.M. Yassin, A. Zabidi, Z.I. Rizman, "Non-linear Autoregressive with Exogeneous Input (NARX)Bitcoin Price Prediction Model using PSO-Optimized Parameters and Moving average technical indicators", Thesis, Malaysia, September 2017.

[31] X. Li, L. Yang, F. Xue and H. Zhou, "Time series prediction of stock price using deep belief networks with intrinsic plasticity," 2017 29th Chinese Control and Decision Conference (CCDC), Chongqing, 2017, pp. 1237-1242.

[32] T. Le, J. Kim and H. Kim, "Classification performance using gated recurrent unit recurrent neural network on energy disaggregation," 2016 International Conference on Machine Learning and Cybernetics (ICMLC), Jeju, 2016, pp. 105-110.

LIST OF PUBLICATIONS OF THE CANDIDATE'S WORK

International Journal:-

[1] Sandeep Suri, Arushi Gupta, and Kapil Sharma. "Comparative Analysis of Ranking Algorithms Used On Web." Annals of Emerging Technologies in Computing (AETiC), no.4 (2020). 14-25. {Scopus Indexed}

International Conferences:-

[2] Sandeep Suri, Sachin Papneja, Kapil Sharma. "Frustration Detection on Reviews Using Machine Learning". *Internation Conference on Emerging Technology (INCET).* Belagavi: IEEE, June 2020. {Scopus Indexed}

[3] Sandeep Suri, Arushi Gupta, and Kapil Sharma. "Comparative Study of Ranking Algorithms." In *2019 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, pp. 73-77. IEEE, 2019. {Scopus Indexed}

[4] Arushi Gupta, Sandeep Suri, Kapil Sharma. "Ranking of Countries." *Internation Conference on Emerging Technology (INCET).* Belagavi: IEEE, June 2020. {Scopus Indexed}