Group

#### *0*

#### *DEPARTMENT OF COMPUTER SCIENCE ENGINEERING, SCHOOL OF ENGINEERING AND TECHNOLOGY, SHARDA UNIVERSITY, GREATER NOIDA*

**Analysis Of Public Perception During Lockdown Due To Coronavirus**

#### *A project submitted*

#### *In partial fulfillment of the requirements for the degree of Bachelor of Technology in Computer Science and Engineering*

#### *By:*

**JIGYASA CHAUDHARY(2018015575)**

**PIYUSH CHHABRA(2019003228)**

**ABHIRAJ(2018014721)**

**AYUSHYA NIRWAN(2018003532)**

**Supervised by:**

**Dr. Hoor Fatima, Assoc.Prof(CSE)**

**May,20022**

i

**CERTIFICATE**

This is to certify that the report entitled “ANALYSIS OF USER TWEET DURING COVID-19 LOCKDOWN IN INDIA” submitted by “Jigyasa Chaudhary(2018015575), Piyush Chhabra (2019003228),Ayushya Nirwan (2018003532), Abhiraj(2018014721)” to Sharda University, towards the fulfillment of requirements of the degree of “Bachelor of Technology” is record of bonafide final year Project work carried out by him in the “Department of Computer Science and Engineering, School of Engineering and Technology, Sharda University”.

##### The results/findings contained in this Project have not been submitted in part or full to any other University/Institute for award of any other Degree/Diploma.

##### Signature of Supervisor

##### Name: Dr.Hoor Fatima

##### Designation: Assoc. Prof (CSE)

##### Signature of Head of Department Name: Prof. (Dr.) Nitin Rakesh Place:

##### Date:

**Signature of External Examiner**

**Date:**

ii

**Acknowledgement**

A major project is a golden opportunity for learning and self-development. We consider our self very lucky and honored to have so many wonderful people lead us through in completion of this project.

First and foremost we would like to thank Dr. Nitin Rakesh, HOD, CSE who gave us an opportunity to undertake this project.

My grateful thanks to **Dr. Hoor Fatima**for his guidance in my project work.

**Dr. Hoor Fatima,** who in spite of being extraordinarily busy with academics, took time out to hear, guide and keep us on the correct path. We do not know where we would have been without his help.

CSE department monitored our progress and arranged all facilities to make life easier. We choose this moment to acknowledge their contribution gratefully.

Name and signature of Students

**JIGYASA CHAUDHARY(2018015575)**

**PIYUSH CHHABRA(2019003228)**

**ABHIRAJ(2018014721)**

**AYUSHYA NIRWAN(2018003532)**

iii

**ABSTRACT**

With the fast augmentation inside the usage of the web, feeling examination has become one in all the first famous fields of language process (NLP). abuse assessment examination, the inferred tendency inside the text will be burrowed effectively for various occasions. individuals square measure abuse electronic media to get and confer changing sorts of data for an enormous extension all through COVID-19 blast. Mining such content to guage people's sentiments will accept an indispensable part in making conclusions to remain things in line. The focal point of this survey is to mine the sensations of Indian residents concerning the crosscountry internment executed by the Indian government to scale down the speed of spreading of Coronavirus.

In this work, the inclination assessment of tweets has been performed double-dealing human language development and Computer based intelligence classifiers. information are eliminated from Twitter, remarked on double-dealing TextBlob and preprocessed double-dealing the language specialist's unit given by the Python. RNN is used for feeling plan during this endeavor. This audit assumes that the weight of Indian voters square measure supporting the choice of the internment maintained by the Indian government all through crown emission.

***Keywords*** **-**Convolutional Neural Network, Deep Learning, Character, Segmentation.

iv

**Contents**

Title Page… i

[CERTIFICATE ii](#bookmark=id.26in1rg)

[ACKNOWLEDGEMENT iii](#bookmark=id.lnxbz9)

[Abstract iv](#bookmark=id.35nkun2)

[Chapter1: INTRODUCTION 6](#bookmark=id.1ksv4uv)

* + [Problem Definition 7](#bookmark=id.44sinio)
  + [Project Overview/ Requirement Specifications 8](#bookmark=id.2jxsxqh)
  1. [Hardware Specifications 9](#bookmark=id.z337ya)
  2. [Software Specifications 9](#bookmark=id.3j2qqm3)

[Chapter2: Literature Survey 10](#bookmark=id.1y810tw)

* + [Existing System: 12](#bookmark=id.4i7ojhp)
  + [Proposed System 14](#bookmark=id.2xcytpi)
  + [Feasibility Study 14](#bookmark=id.1ci93xb)
  + [Technical Specification: 16](#bookmark=id.3whwml4)

Chapter 3: Data Collection and Preprocessing…………………………………………………………………..20

1. Data Collection…………………………………………………………………………………………….20

2. Data processing…………………………………………………………………………………………………….. 21

[Chapter 4: System Analysis and Design 24](#bookmark=id.2bn6wsx)

* + System Design 27
  + [Flowcharts/DFDs/ERDs/Architecture 29](#bookmark=id.qsh70q)
  + [Methodology 32](#bookmark=id.3as4poj)
  1. [Testing Process 36](#bookmark=id.1pxezwc)

[Chapter 5: RESULTS / OUTPUTS 3](#bookmark=id.49x2ik5)9

[Chapter 6: Conclusion 44](#bookmark=id.2p2csry)

[Chapter 7: References 45](#bookmark=id.147n2zr)

5

**Chapter1: INTRODUCTION**

**1.1.Problem Definition**

Begun in Wuhan, China, the emotional spread of Coronavirus sickness (COVID-19) has caused general prosperity crisis commonly and universally moreover. The CDC actuated its EOC, and the WHO conveyed its first report concerning situation about COVID-19 on 20 January2020. The World Health Organization saw and gave the name "2019-nCOV" to novel Coronavirus. The troublesome situation of the COVID-19 pandemic had been misjudged till the National Health Commission (NHC) gathered it as a B-type communicable disease formally and took measures to fight for this pandemic. The most reasonable methods of becoming secured are defend and social eliminating without even a hint of inoculation for this powerful disease.Eliminating social contacts is crucial to limiting the spread of the COVID-19 pandemic, where particular physical isolating constraints are applied to fight against COVID-19. People are urged to stay at their homes and to ensure social isolating basically all of the countries approved lockdown to ensure social eliminating at public spots. All through the whole presence of human development, these events were never fit where the whole world has been residing under lockdown.Lockdown prevents people from moving unhinderedly without hindrance.

According to Canadian Mental Health Association,42% of Albertans report that their top mental health concern during the pandemic has been isolation throughout, 44% say a lack of socialization has been a mental health challenge during this time. 57% of Albertans say staying connected with loved ones has helped their mental health and wellbeing during the pandemic [3]. A lot of conversations and emotional expressions on the virus and its effect on peoples’ lives has becoming a very popular topic on social media. In this research work, we will be analyzing these conversions and expressions in order to determine the areas where people are most affected by the virus and help the medical practitioners and the government narrow down to the specific areas to look into in order to reduce the effect of the virus on peoples’ lives. To immune human beings from this virus, different companies came up with vaccines and some of the approved vaccines by World Health Organization are AstraZeneca manufactured by Oxford, BioNTech manufactured by Pfizer, Beijing manufactured by Sinopharm and Sputnik V [4]. About two hundred and twelve (212) countries commenced the vaccination of people against coronavirus. For a person to be considered fully

6

vaccinated, It is required that 2 shot of the vaccine is administered with six weeks intervals. The first 10 countries with Canada ranked 15th with the most administered doses (at least one dosage) of approved vaccines [5] as at displayed in the Table 1.1.

Vaccination campaigns for Covid-19 will start in December 2020 based on the country.As vaccine supplies are inadequate to meet the needs of the entire population at once, people at risk and those directly involved in the fight against the pandemic are the first to receive the vaccine.

The following table ranks countries by the number of doses administered as of 07 December 2002.

The data are taken from country-specific health agencies and ministries in the absence of a global database.

| **Rank** | **Country** | **Doses \*** | **Rank** | **Country** | **Doses \*** | **Rank** | **Country** | **Doses \*** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | **China** | 2.55 billion | 11 | **Pakistan** | 126.59 million | 21 | **Spain** | 77.16 million |
| 2 | **India** | 1.28 billion | 12 | **Turkey** | 120.91 million | 22 | **Argentina** | 69.89 million |
| 3 | **United States** | 470.30 million | 13 | **United Kingdom** | 117.88 million | 23 | **Canada** | 62.03 million |
| 4 | **Brazil** | 314.61 million | 14 | **Iran** | 107.53 million | 24 | **Colombia** | 58.59 million |
| 5 | **Indonesia** | 241.44 million | 15 | **France** | 107.04 million | 25 | **Malaysia** | 53.88 million |
| 6 | **Japan** | 197.62 million | 16 | **Bangladesh** | 102.11 million | 26 | **Morocco** | 48.85 million |
| 7 | **Mexico** | 134.37 million | 17 | **Italy** | 98.73 million | 27 | **Saudi Arabia** | 47.70 million |
| 8 | **Russia** | 130.63 million | 18 | **Thailand** | 93.75 million | 28 | **Poland** | 42.67 million |
| 9 | **Germany** | 127.66 million | 19 | **Philippines** | 91.78 million | 29 | **Peru** | 42.11 million |

Table 1.1: Vaccinaction against COVID 19

7

**1.2.Project Overview/ Requirement Specification**

With this virus, there has been an increase in mental health issues. The various problems that came along with the virus, such as job loss, isolation, depression, loss of income, homeschooling, and a lot more have had quite a lot of mental stress and this illness has increased drastically. The government does carry out surveys from time to time to understand people’s perception of the virus as it relates to mental health but this method is tedious, time-consuming, and expensive. Even with the survey, only a minute percentage of the population can be reached and this sample is too small to predict and help health care practitioners make proper diagnoses and create solutions for the eradication of this pandemic. With Twitter data, there is access to free large data. Although Twitter data is not reliable as there may be false tweets, false tweets can be detected and government can rely on this data from initial findings. These initial findings can help the government takes quick steps to avoid social and economic damage which could be enormous if time and adequate data are not put into perspective. This research work takes into account people’s opinions on Twitter (as a large community is created on social media which is limitless to race, region, or country) as people err their views and opinions on the effect of covid 19 on mental health. With machining learning, this research work tables different models to analyze peoples’ sentiments and deduced the different avenues that health care workers can utilize for more accurate diagnosis and cure which eventually leads to a decrease in mental health.

For this research work, there are six (6) chapters. Chapter 1 is the Introduction which provides a detailed background of the research work, explaining the reasons for research work, the problem statement, the prediction of Covid 19 data in Canada, and how the introduction of vaccines had helped calm the mental health of people. Chapter 2 carries the Literature Review and is sectioned into three. The first section discusses related works of sentimental analysis of Twitter data. Section two discusses the related work of sentimental analysis of other social media data asides from Twitter data while section three discusses related works that pertain to the sentimental analysis of Covid-19 Twitter data. Chapter 3, which is the methodology discusses in detail the three main methods for the research work which include feature selection, supervised learning, and semi-supervised learning. Chapter 4 is the Data collection and preprocessing which discussed in detail how data was collected, preprocessed to ensure it is cleaned enough for machine learning analysis. Chapter 5 talked about the results and discussions of the analysis. Every step of the analysis is broken down here to understand different insights that were extracted from the data. Chapter 6 concludes this research work

8

**1.3. HARDWARE REQUIREMENTS**

* Processor - I3/Intel Processor
* RAM - 4GB (min)
* Hard Disk - 128 GB
* Key Board - Standard Windows Keyboard
* Mouse - Two or Three Button Mouse
* Monitor - LCD

**1.4. SOFTWARE REQUIREMENTS**

* Operating System : Windows 7+
* Server side Script : Python 3.6+
* IDE : PyCharm IDE or Colab
* Libraries Used : Pandas, Numpy, tensorflow, textblob, nltk.

9

**Chapter2: Literature Survey**

**2.1. Existing System:**

This study examines the use of machine convolutional neural networks in the work related to covid-19 and lockdowns. The model achieved the best accuracy of using Recurrent Neural Networks. The training and testing dataset were captured from the system or uploaded to the system.

Recent developments in machine learning-based RNN have allowed researchers to greatly improve the accuracy of the work. In this a machine learning-based approach to detect the sentiment using the feature extraction from the powerful libraries captured in real-life scenarios with heterogeneous backgrounds and experimented with convolutional neural networks on our large dataset. Machine learning offers a number of possible uses in clinical trials and studies. As anyone in the telecommunications business will tell you, using microblog data to understand citizens' sentiments is challenging. Using ML-based predictive analytics to anticipate probable outcomes might assist governments in drawing a better pool from a diverse set of tweets from social media, for example.Machine learning systems have also been used to ensure real-time monitoring and data access to trial participants, determine the appropriate sample size for testing, and eliminate data-based errors by using electronic records.. Here is an algorithm for sentiment extraction and technique which is used in the Twitter dataset and classified for the best. Where the AI sector is something on which the economy of our country is highly dependent. Here the government needs to understand the feelings and to control the pandemic as per the citizens. If proper care is not taken in this area then it causes serious effects on the citizens of the country.

Based on the above-related works, we have implemented a system using Machine Learning. The next sections will describe our proposed work.

10

| **Author & Year** | **Proposed** | **Finding/Outcomes** |
| --- | --- | --- |
| Wu, Jianhong & Leung, Kathy & Leung, Gabriel[10] | Nowcasting and Forecasting the Potential Domestic and International Spread of the 2019-nCoV Outbreak Originating in Wuhan, China | We assess the scale of the outbreak in Wuhan depending on the number of instances transmitted from Wuhan to places from outside mainland China, and we project the magnitude of both the national and worldwide public health impacts for outbreaks, taking into consideration social and non-pharmaceutical preventative strategies[10]. |
| Medford, Richard & Saleh, Sameh & Sumarsono[2] | Leveraging High-Volume Twitter Data to Understand Public Sentiment for the COVID-19 Outbreak | Tweets with bad opinions and emotions correspond to the number of instances reported mostly in the COVID-19 epidemic. Twitter is a useful channel for understanding public opinion throughout actual time and targeting public health messaging depending on user interest and sentiment. |
| Li, Sijia & Wang, Yilin & Xue, Jia & Zhao[3] | The Impact of COVID-19 Epidemic Declaration on Psychological Consequences | The findings revealed that negative emotions (including anxiety, despair, and anger) and susceptibility to social hazards increase, whereas good emotions (such as Oxford happiness) and life satisfaction are reduced[3]. |

* **Other Social Media Data Sentiments**

Li et al collected data from Weibo, China second largest social media platform and We analyzed the datasets using Natural Language Processing (NLP) in order to classify the information into seven different types, such as help, emotional support, donations, cautions.This type of information helps governments, healthcare personnel, and individuals to diagnose and respond appropriately to the mental and personal health issues relating to the pandemic [9]. Wang et al carried out emotional analysis method on COVID-19 data collected from Sina

11

Weibo ,a microblogging site in China using machining learning models such as Support Vector Machine, naïve

Bayes and Random Forest classifiers. Extracting reviews of 184 hospitals from mouthshut.com, Ankita Bansal et al analysed the sentiment of these hospital reviews to provide relevant and important information about the operating conditions and current status of hospitals to the general public. Sentiment analysis applied to patient’s reviews to quantify the direction and/or magnitude of the emotive content. Patient comments were segregated into different sections and analysed to quantify the positive or negative aspects of the reviews. With this analysis, the overall rating of the hospital, based on key parameters will help people to understand the hospital’s current condition. This analysis births information which provides great value to the patients, giving them the power to make the best options of available hospitals [10]. 2.3 Covid-19 Tweets Sentiments With a lot of researches carried out on twitter data analysis, only a few of these reseaches focused on sentimental analysis of social media datasets on the COVID-19 pandemic. Koustuv Saha et al carried out a research on the psychosocial effects of the COVID-19 crisis by using social media data (Twitter) from 2020. In their research, they found out that people’s mental health symptomatic and support expressions increased significantly during the COVID-19 period as compared to similar data from 2019 [11]. Jelodar et al with NLP automated the extraction of COVID-19–related discussions from social media to discover various issues related to COVID-19 from public opinions. They also investigated the use the DL-based Long Short-Term.

* **COVID-19 TWEETS SENTIMENTS**

In analyzing COVID-19 comments, we developed a memory-based (LSTM) approach for sentiment classification and discovered it produced better results than other well-known algorithms[12]. Lyu, Chen, Wang, and Luo analyze millions of twitter data to identify two groups of users who use term “coronavirus” and the other terms “Chinese virus” or “Wuhan virus”, and predict the number of users who are more likely to use later words than the former using machine learning models [13]. Amrita Mathur et al extracted information from twitter and classified the tweets into positive and negative sentiments. They further classified these tweets into six basic emotions given by Ekman i.e joy, sadness, anger, fear, disgust and surprise [14]. While analyzing twitter data, Man Hung et al identified different themes that relate to COVID-19 which are health care environment, emotional support, business economy, social change, and psychological stress. These themes

12

dictate the influences of the pandemic on human mental and overall health [15]. Sohini Sengupta et al carried out an overview on the discussions about

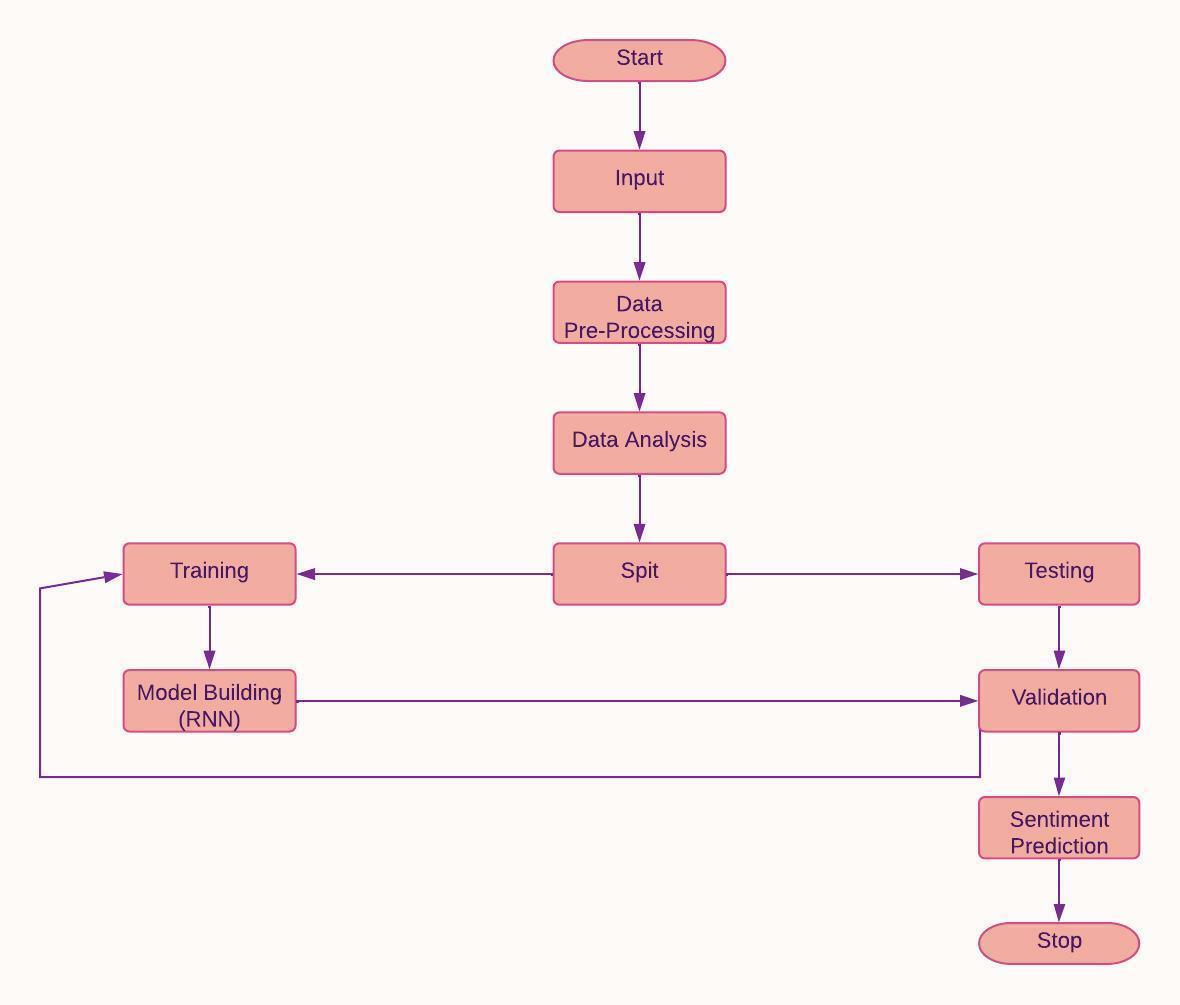
mental health as at June 2020 and the influences of covid-19 on the health issue. They also carried out an overal sentiment analysis to further understand the emotions of twitter users [16]. Jim Samuel et al present textual analyses of Twitter data to identify public sentiment, specifically, tracking the progress of fear, which has been associated with the rapid spread of Coronavirus and COVID-19 infections.A methodological approach is proposed to analyze Twitter data in relation to the current COVID-19 pandemic to identify sentiments, keyword associations, and trends.. Their discussion centers around the search for insights with descriptive textual analytics and data visualization [17]. Anna Kruspe et al analyze Twitter messages (tweets) restricted to only European countries that were collected during the first months of the COVID-19 pandemic. These data were analyzed with a neural network for sentiment analysis using multilingual sentence embeddings. The results were separated by country of origin, and their temporal development was correlated with the events in those countries. The moods of the citizens were studied and they found out that the lockdown announcements correlated with a deterioration of mood in almost all surveyed countries, which recovers within a short time span [18]. Furqan Rustam et al used a proposed feature set developed by concatenating bag-of-words and term frequency-inverse document frequency to examine the performance of various machine learning classifiers. Twitter data were classified as positive, neutral, or negative sentiment, The performance of classifiers was evaluated on the accuracy, precision, recall, and F1 score. Further investigation was made on the dataset using the Long Short-Term Memory (LSTM) architecture of the deep learning model. With comparison with the machine learning classifiers, LSTM achieved lower accucary as Extra Trees Classifiers outperform all other models by 9 2.3. COVID-19 TWEETS SENTIMENTS CHAPTER 2. LITERATURE REVIEW achieving a 0.93 accuracy score [19]. Harleen Kaur et al conducted a reseach work by designing an algorithm called Hybrid Heterogeneous Support Vector Machine (H-SVM) which was used to perform sentimental analysis of twitter data. This twitter data was collected based on hashtag keywords, including COVID-19, coronavirus, deaths, new case, recovered.The sentiment was classified into positive, negative and neutral sentiment scores. A comparison was made between the proposed algorithm and Recurrent Neural Network (RNN) and Support Vector Machine (SVM) on specific parameters, such as precision, recall, F1 score, and accuracy.[20].

13

**2.2.** **Proposed System:**

The data subsequently assembled can assist us in giving the reaction of the occupants regarding the lockdown did the country over in light of a legitimate concern for COVID-19. Data can be aggregated and refined through the man-made intellectual ability (Computer based insight) model, which can finally survey the ID of people reaction whether they are positive or adverse to the lockdown did the country over using Sentimental assessment. We will use neural associations for portrayal purposes.

**FLOW CHART:**



14

ADVANTAGES:

* High accuracy.
* Time Saving.
* Low complexities.

**2.3.Feasibility Study**

The practicality of a task is surveyed at this stage and the strategic plan depends on the most well-known project plan and explicit quotes. During the examination of the arrangement a plausibility investigation of the proposed framework will be conducted. This is to affirm that the arranged framework isn't a weight to the corporate In request to investigate the possibility. Three key considerations are involved in possible analysis.

* ECONOMICAL FEASIBILITY
* TECHNICAL FEASIBILITY
* SOCIAL FEASIBILITY
* **ECONOMICAL FEASIBILITY**

This audit was planned to study the financial impact of the program on the affiliation. The proportion of resource that an association can place assets into investigation and program improvement is limited. Expenses ought to have a clarification. In this manner the structure improved and inside spending plan and this was cultivated in light of the fact that by far most of the advancement used is open in vain. Just extraordinarily planned things should be purchased.

* **TECHNICAL FEASIBILITY**

This audit was coordinated to study the feasibility of the advancement, that is, the specific essentials of the system. Any refreshed system should not have an interest for open particular resources. This will incite more interest for open particular resources. This will incite more prominent degrees of prominence on the client. A further evolved system should have a limited need, as minor or minor changes are required in executing the program.

15

* **SOCIAL FEASIBILITY**

A component of the audit is to review the level of structure affirmation by the customer. This consolidates the course of setting up the customer to use the structure properly. The level of customer affirmation depends just upon the systems used to show the customer the structure and get that individual familiar with it. Her certainty ought to be raised so she can in like manner make supportive investigation, which is OK, as she is the individual who uses the last structure.

**2.4.Technical Specification**

Analyzing data to extract using insights requires some methods and this research work is not an exception. The methodologies utilized can have a huge impact on the outcome or performance of the project work. The proposed architecture for mental health prediction with machine learning that depicts these methods are written below

* **Text Features**

There are techniques to convert text data into numeric features without losing the meaning of the data, since machine learning models are only compatible with numeric values, and most text data types aren't compatible with machine learning models. Some of the feature extraction techniques are mentioned below:

* **N-gram**

A N-gram is a sequence of words for a sentence, and it is probably the simplest machine learning concept. There are quite some varieties of the usefulness of N-gram. It offers autocorrection of words, autospell check, and grammar checking[21]. The relationship between words can also help in determining a person's emotions or sentiments by determining what the person is most likely to say based on the word's meaning.An N-gram is any combination of words used together, where a monogram is any N-gram with N = 1, a bigram is an N-gram with N = 2, and a trigram is an N-gram with N = 3. N-grams help to identify the next possible word to be used after a phrase.

16

* **The Word2Vec Model**

Word2vec is a method to efficiently create word embeddings. Created by Google in 2013, it is a predictive deep learning-based model to compute and generate high quality, distributed and continuous dense vector representations of words, which capture contextual and semantic similarity [22]. It is a type of unsupervised model that takes a large corpus of words, creates a vocabulary of words from it, and generates dense word embeddings for each word. The words are transformed into vectors to allow machine learning algorithms to perform algebra operations on numbers as against words. This transformation is referred to as word embedding [23]. With distributed Hypothesis in Word2Vec, the lexicon for a word is found in its neighboring words. A word can be predicted by looking at these close words.

* **Continuous Bag of Words**

In the CBOW model, the architecture tries to predict the current target word (usually the word in the middle) based on the source context words which are the surrounding words [22], [24]. The corpus is built such that extraction of each unique word from the dictionary can be done and mapped to a unique numeric identifier. The major python module used is Kera preprocessing. After this, the CBOW generator is built with two variables: context and target. Keras and Tensorflow are used to build the deep learning architecture for the CBOW model. This model tends to do better with smaller datasets. It is very fast to train the model and provide better accuracy. 3.1.2.2 Continuous Skip-Gram Model This is the inverse of CBOW as it predicts the surrounding words from the current target words. This model does better with a larger dataset. The objective is to predict the contexts of a given word. To implement this model, the corpus dictionary is built such that each unique word can be extracted from the dictionary and assigned a unique identifier. The mappings to transform words to and from their unique identifiers are also maintained. The skip-gram generator is built next which will provide the pair of words and their relevance. To build the skip-gram model, Keras on top of TensorFlow is taken advantage of to build it. The model is trained and the embedded words are retrieved [22].

* **Bag of Words**

It is a type of feature extraction or feature encoding used to extract features from the text. It is called Bag of words because the order with which the text appears is irrelevant and discarded. It is only concerned with whether or not the known word occurs in the text, not the position in the text. Each document of free text is converted into a vector which can be used as input or output for a machine learning model. In a python

17

programming language, the module for BOW is referred to as CountVectorizer. It transforms a given text into a vector based on the frequency (count) of each word that occurs in the entire text. See documentation here. Before generating vector representation of words, CountVectorizer pre-processes the text data. CountVectorizer creates a matrix in which each unique word is represented by a column of the matrix, and each text sample from the document is a row in the matrix [25].

* **Supervised Learning**

This is a type of Machine learning where a dependent variable can be predicted based on one or more independent variables. For example, suppose we want to predict whether a bank customer will pay a loan or not (which is represented as a variable called status) based on its loan amount, duration, and demography, the status variable here is the dependent variable whilst loan amount, duration and demography are the independent variables. The independent variables, usually represented with X are also known as input variables while the dependent variable, usually represented by Y is also known as output variables. As an illustration, there is a dataset with different animals which are dogs, cats, horses, and lions. A subset of the data that are properly labeled is given to the machine learning model to understand.

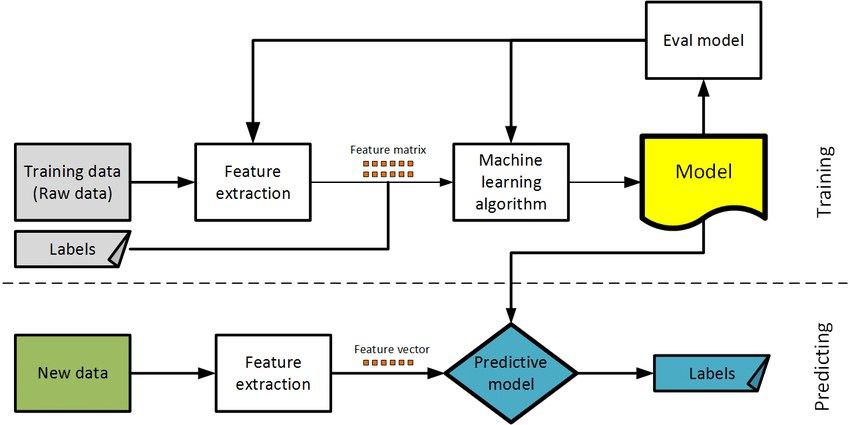


Figure 3.3: Supervised Learning

* **SEMI-SUPERVISED LEARNING**

This, the remaining data (test data) is given to the model to sort. Because the model already understands the features of the different animals, it can sort out which animal is which using the test data accurately. Based on the independent variables, supervised machine learning can predict the dependent variable.

18

Supervised learning is generally a classification problem where the target variable (also the dependent variable) is a categorical data type to determine which category a data is. Some of the algorithms to create supervised learning models under supervised machine learning are LogisticRegression, Random Forest, Decision Tree, KNN, Linear Support Vector Machines, Non-Linear Support Vector Machines, Naive Bayes Theorem, and many more.

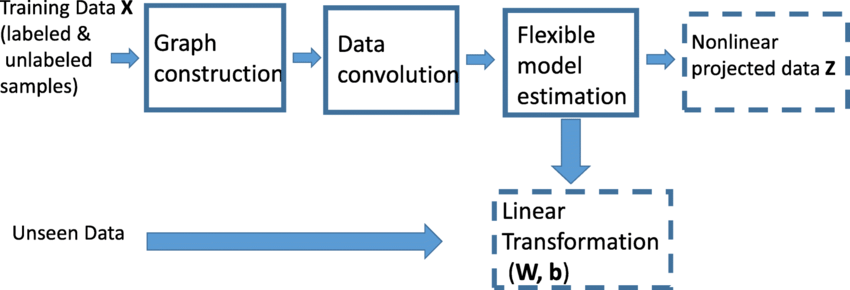


Figure 3.4: Semi-Supervised Learning

19

**Chapter 3: Data Collection and Preprocessing**

**3.1. Data Collection**

In recent times, human interaction is centered around the web and as such, individuals tend to communicate how they feel over social media such as Facebook, Twitter, TikTok, and other social platforms. Twitter information tweeted by individuals pertaining to the pandemic was extracted for this research work to dissect the feelings of each individual. Several types of research were done on the collection of Twitter data. To access Twitter data, the user must have a Twitter developer account. This is done by applying for this account and providing the required information for approval. Upon approval, which usually takes twenty-four to forty-eight hours, the user’s profile is created, and API access is given. From the Twitter developer account profile, the user can retrieve the access token, access token secret key, consumer key, and consumer secret key. Without these keys, it is impossible to access Twitter data. The data extracted is usually in a JSON file. The JSON file is usually a mix of ‘root-level attributes and child objects. Some of the root-level attributes are stated in the table below: Other attributes can be found in this link. There are several ways of retrieving Twitter data. For this research work, the focus is on python modules that can be used for tweets extraction some of which are:

* **Snscrape**

snscrape provides scraping for social networking services.Using scraping, it finds items, such as user profiles and hashtags, and returns them..

The following services are currently supported:

* Facebook: user profiles, groups, and communities (aka visitor posts)
* Instagram: user profiles, hashtags, and locations
* Mastodon: user-profiles and toots (single or thread)
* Reddit: users, subreddits, and searches (via Pushshift)
* Telegram: channels
* Twitter: users, user profiles, hashtags, searches, tweets (single or surrounding thread), list posts, and trends
* VKontakte: user profiles
* Weibo (Sina Weibo): user profiles

20

* **Requirements**

snscrape requires Python 3.8 or higher.Installing Snscrape automatically installs Python package dependencies.

Note that one of the dependencies, lxml, also requires libxml2 and libxslt to be installed.

* **Installation**

pip3 install snscrape

!pip install snscrape

import snscrape.modules.twitter as sntwitter

import pandas as pd

*# Creating list to append tweet data to*

tweets\_list2 = []

*# Using TwitterSearchScraper to scrape data and append tweets to list*

for i,tweet **in** enumerate(sntwitter.TwitterSearchScraper('COVID-19 since:2020-01-01 until:2020-12-31').get\_items()):

if i>10:

break

tweets\_list2.append([tweet.date, tweet.id, tweet.content])

*# Creating a dataframe from the tweets list above*

tweets\_df2 = pd.DataFrame(tweets\_list2, columns=['Datetime', 'Tweet Id', 'Text'])

tweets\_df2.to\_csv('vaccine.csv')

**3.2 .Data Preprocessing**

In data mining, preprocessing refers to the process of transforming datasets for knowledge extraction and preparing them for evaluation[reference 2 data preprocessing].As part of the preprocessing process, several techniques are used.Cleansing, integrating, transforming, and reducing the dataset are a few of them. Modeling can be done using structured/clean data as a result.It is important to first clean the raw data obtained from different sources before analyzing it, since the raw data is of no use for analysis as it is unsuitable for analysis.. For all analysis projects, data cleaning takes about 70% of project work. It is cumbersome but extremely unavoidable. For this project work, data was extracted from January 2020 to April 2021. The data extraction was limited to the tweet id, the date of tweets, and the text. The focus is on the text as that is the conversations to be analyzed. The uncleaned data was preprocessed to ensure the data is clean enough for model acceptance. The data contains three columns; the id, text, and the month of the tweet. A new column is created where the

21

preprocessed data is stored named tidy\_text. The preprocessing 19 is done of column text. Some of the preprocessing done on our raw data are stated below:

* **Remove the urls from the text**

To remove the urls from a text, a function as shown below is written in python and applied to the text. def remove\_url(row): txt = str(row[’tidy\_text’]).split(’https’)[0] return txt data[’tidy\_text’] = data.apply(remove\_url, axis = 1)

* **Remove the special characters**

Some regular expressions were expressed to remove the special characters as stated in the code below data[’tidy\_text’] = data.tidy\_text.str.replace("?!,\ &:;%()", " ", regex=True)

* **Remove all usernames with @**

The following line of code can be used to remove and replace usernames. data[’tidy\_text’] = data[’text’].str.replace(’@[\w:]\*’,”)

* **Convert to lowercase**

All the characters are converted to lowercase.The line of code below convert the characters into lowercase to avoid dulpication of words with different cases. data[’tidy\_text’] = data[’tidy\_text’].apply(lambda x: x.lower())

* **Remove numbers from characters**

Here, all numeric values are removed with the regular expression (regex) pattern \d+. The addition sign ensures multiple numbers such as 10 are interpreted as such and not interpreted as two separate numbers. The line of code below removed all numeric values from the text. data[’tidy\_text’] = data[’tidy\_text’].str.replace(’\d+’,"").

* **Drop Null Values**

Some unstructured/raw data, there may be missing values. These values are sometimes called null values. They are usually filled up with the most common words or are 20 dropped completely.If null values are not dealt with, it will affect our models are machine models don’t accept null values. In our preprocessing, we dropped the null values with the line of code stated below: data[’tidy\_text] = data[’tidy\_text].dropna(inplace = True)

22

* **Removal of stopwords**

It is possible to remove stopwords without altering the meaning of a sentence because they add no value to the passage.[reference stopwords]. Stopwords are found in most languages but for the purpose of this project work, stopwords in English are utilized. For sentimental analysis, Stop words are to be removed from the data to keep

only the root words. Common stopwords include[i,me,my,myself,we,our,ours,ourselves,you,your]. A list of common stopwords (in english) can be found here. With these data, stopwords are removed by importing stopwords from the corpus of nltk, the python module for natural language processing. import nltk from nltk.corpus import stopwords stop\_words = set(stopwords.words(’english’)) Using the lambda function, apply the stopwords to the text to remove them. Also, extract words that have 4 characters and above in order to filter words that are meaningful to the research work. data[’tidy\_text’] = data[’tidy\_text’].apply(lambda x: ’ ’.join([w for w in x.split() if w not in stop\_words])) data[’tidy\_text’] = data[’tidy\_text’].apply(lambda x: ’ ’.join([w for w in x.split() if len(w)> 4])) Below is the top 10 data showing the id, month, the uncleaned text and the tidy text The complete python code for data preprocessing.



**Fig 8.** Organizing and cleaning data.

23

**Chapter4:SYSTEM ANALYSIS & DESIGN**

**4.1.System Design**

* **UML DIAGRAMS**

In the area of programming-based programming, UML serves as a standard language for demonstrating general concepts.

A standard must be set, and it must be made, by the Asset Management Team.

In order to achieve this goal, UML must become a standard language for making PC programming models that are object-based.In its current manifestation, UML consists of two important parts: the Metamodel and the Documentation..A few other types of method or process may be added later; or they may be associated with UML.

Incorporated Language Modeling is a typical language for enunciation, Visualization, Creating and interpreting programming framework work of art, just as business demonstrating and other non-programming programs.

UML is a vital piece of programming based turn of events and programming advancement processes. UML frequently utilizes realistic texts to outline the plan of programming projects.

**Goals**

The essential standards for building an UML are as per the following:

•Give customers the tools they need to make and exchange sensible models with an easy-to-use visual model.

• Give flexibility and particular means to develop pertinent considerations.

• Give a solid foundation to comprehension showing language.

• Enable the improvement of OO instrument market.

• Consolidate endorsed systems.

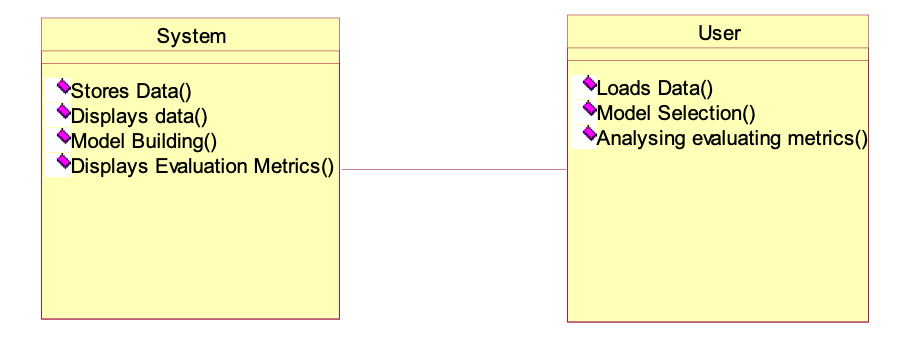
24

* **USE CASE DIAGRAM**

Its inspiration is to present a diagram of the introduction of the structure given by the characters, their terms (tended to as terms of use), and any conditions between those terms of usage.The purpose of the case diagram is to show which characters perform which system limits.

* **CLASS DIAGRAM**

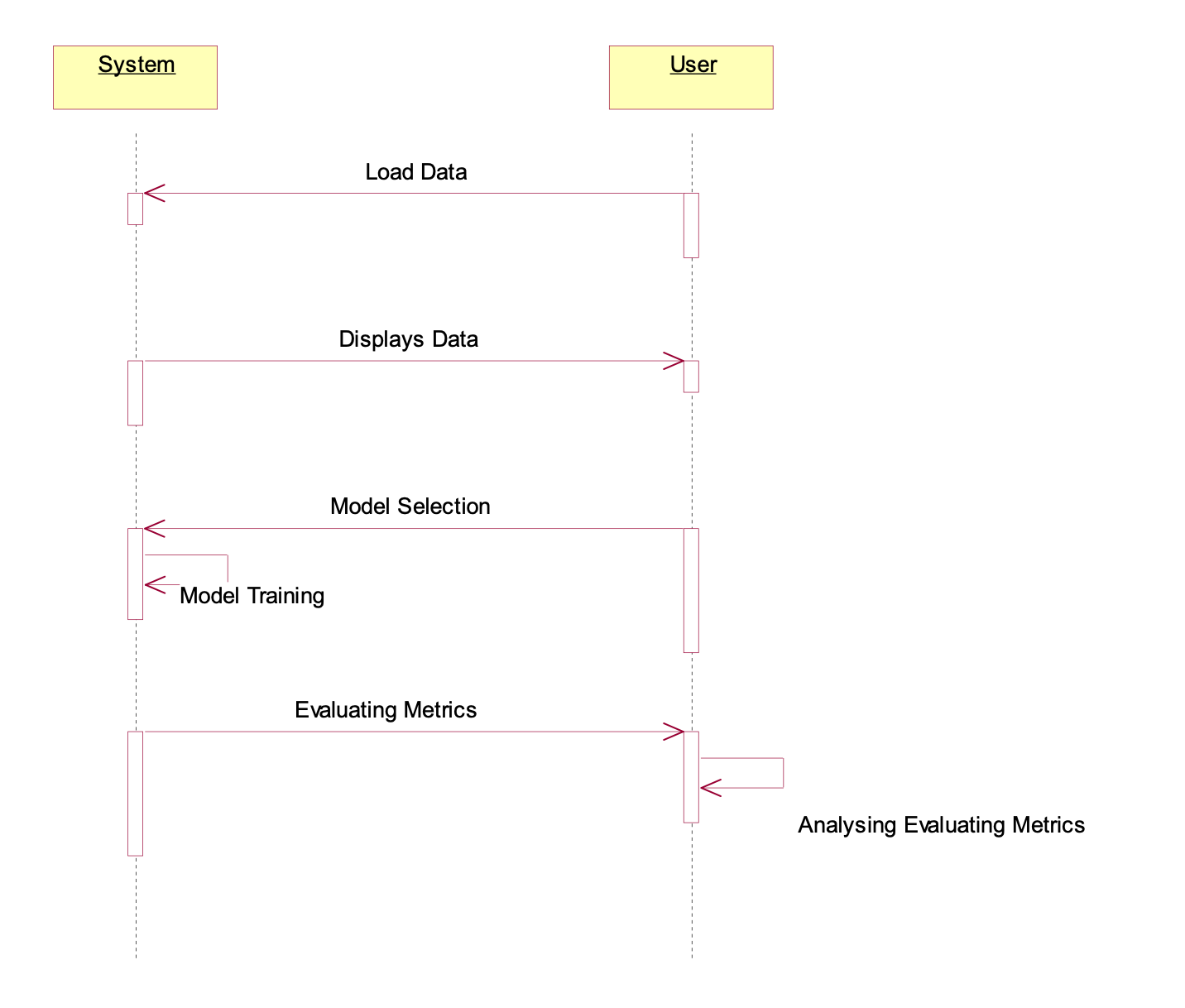
A UML diagram is a static shape that displays a system's state with the aid of showing the machine's setup, its attributes, exercises (or techniques), and the links between each block of instructions. It describes which classes contain different types of information.



25

**SEQUENCE DIAGRAM**

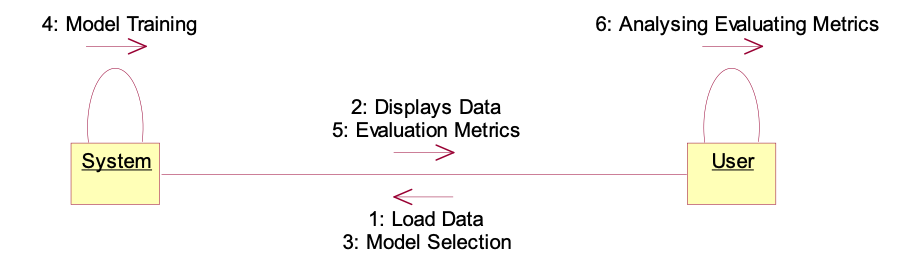
A succession outline in Unified Modeling Language is a sort of intuitive chart showing how processes work with one another and in what request to develop of a Message Sequence Chart. Succession charts region unit by and large alluded to as occasion charts, occasion circumstances and worldly request graphs.



26

* **COLLABORATION DIAGRAM**

In joint exertion diagram the approach decision gathering is exhibited by some determination system as shown underneath. the sum shows in any case the habits in what area unit implied as one when another. in any case the separation is that the plan diagram doesn't depict the article affiliation however the collaboration chart shows the article affiliation.



* **DEPLOYMENT DIAGRAM**

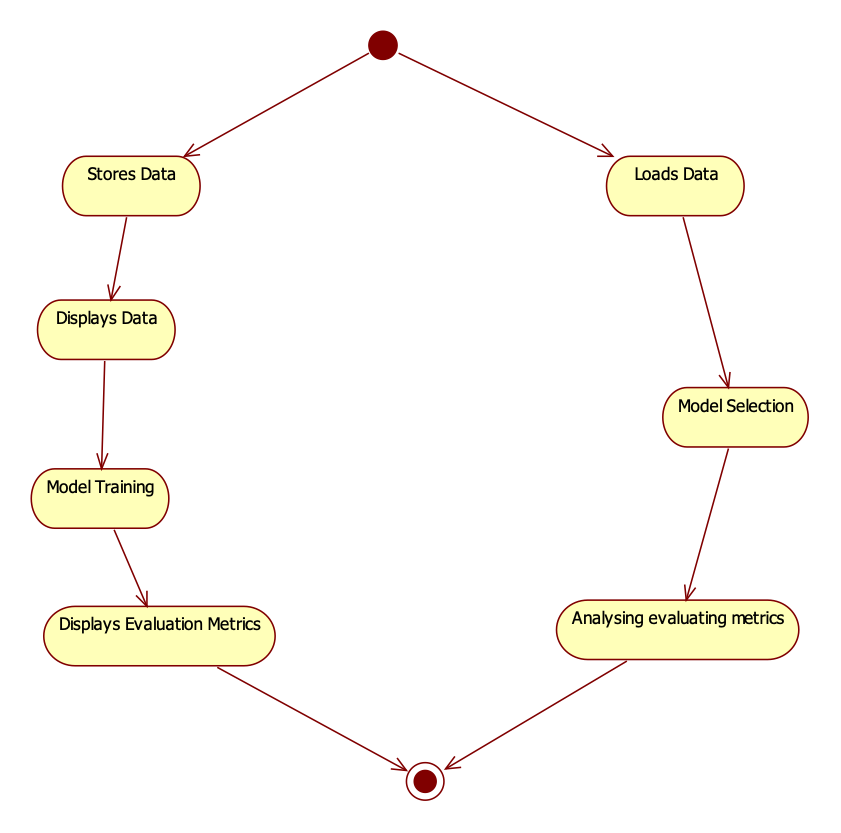
Association layout tends to the sending viewpoint on a system. It is related to the part graph. Since the parts are used using game plan charts. A course of action diagram involves hubs. Hubs are just virtual PC gear used to run the program.



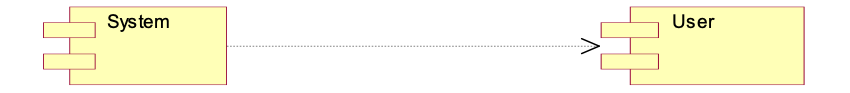
* **ACTIVITY DIAGRAM**

Action chart is a portrayals of work processes of stepwise games and developments with help for option, age and simultaneousness. in the Unified Modeling Language, interest charts can be wont to clarify the business undertaking and functional step by step work processes of variables in an absolutely framework. An interest chart proposes the general float of the executives.

27



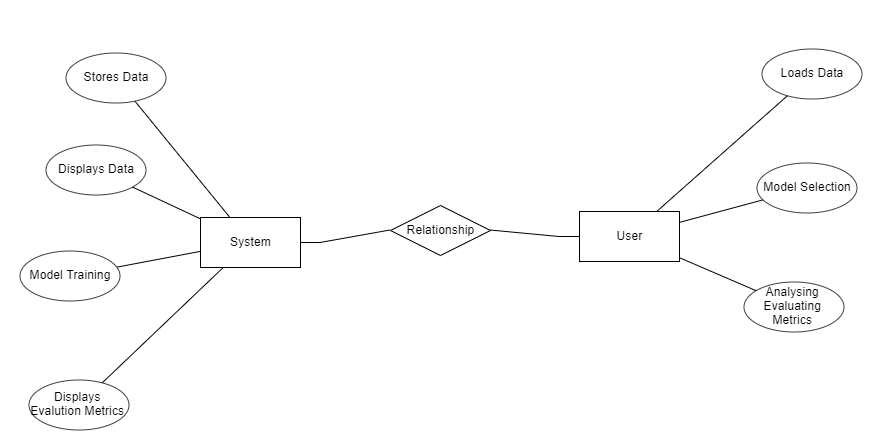
* **COMPONENT DIAGRAM**

It otherwise called an UML part diagram. Component charts are regularly attracted to assist with demonstrating execution subtleties and twofold check that each part of the framework's necessary capacities is covered by arranged turn of events.

28

**4.2.ER DIAGRAM**

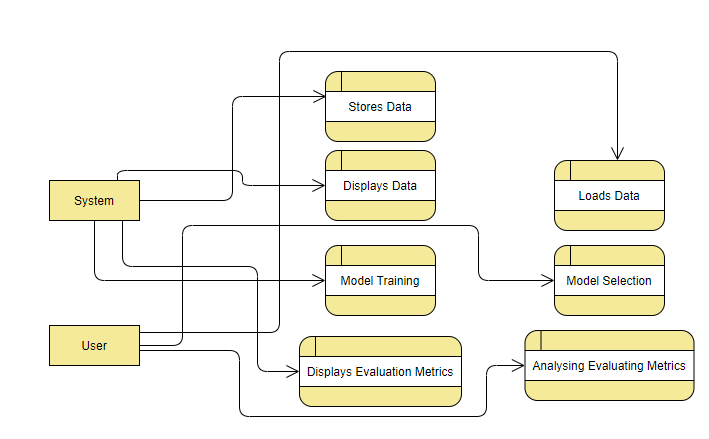
ER model may be a style or blueprint of a info that may later be enforced as a info. the most elements of E-R model are: entity set and relationship set.



29

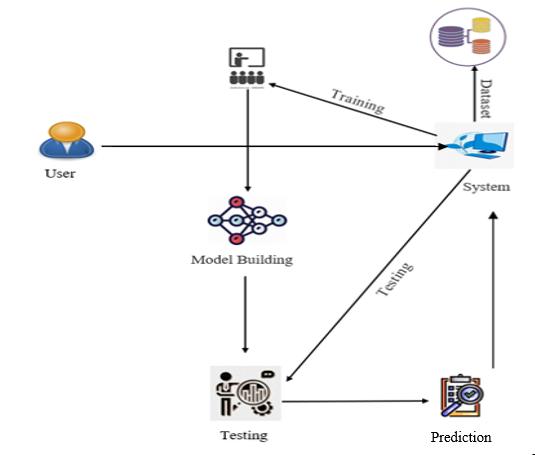
* **DFD DIAGRAM**

A Data flowchart (DFD) can be a memorable method for imagining the expertise streams among a device. A perfect and clean DFD will portray a good measure of the device wants diagrammatically. It could be manual, computerized, or a blend of each. It shows anyway information enters and leaves the framework, what changes {the information the information the data} and any place data is hang on. the point of a DFD is to demonstrate the degree and cutoff points of a framework as a whole.



30

* **Architecture**

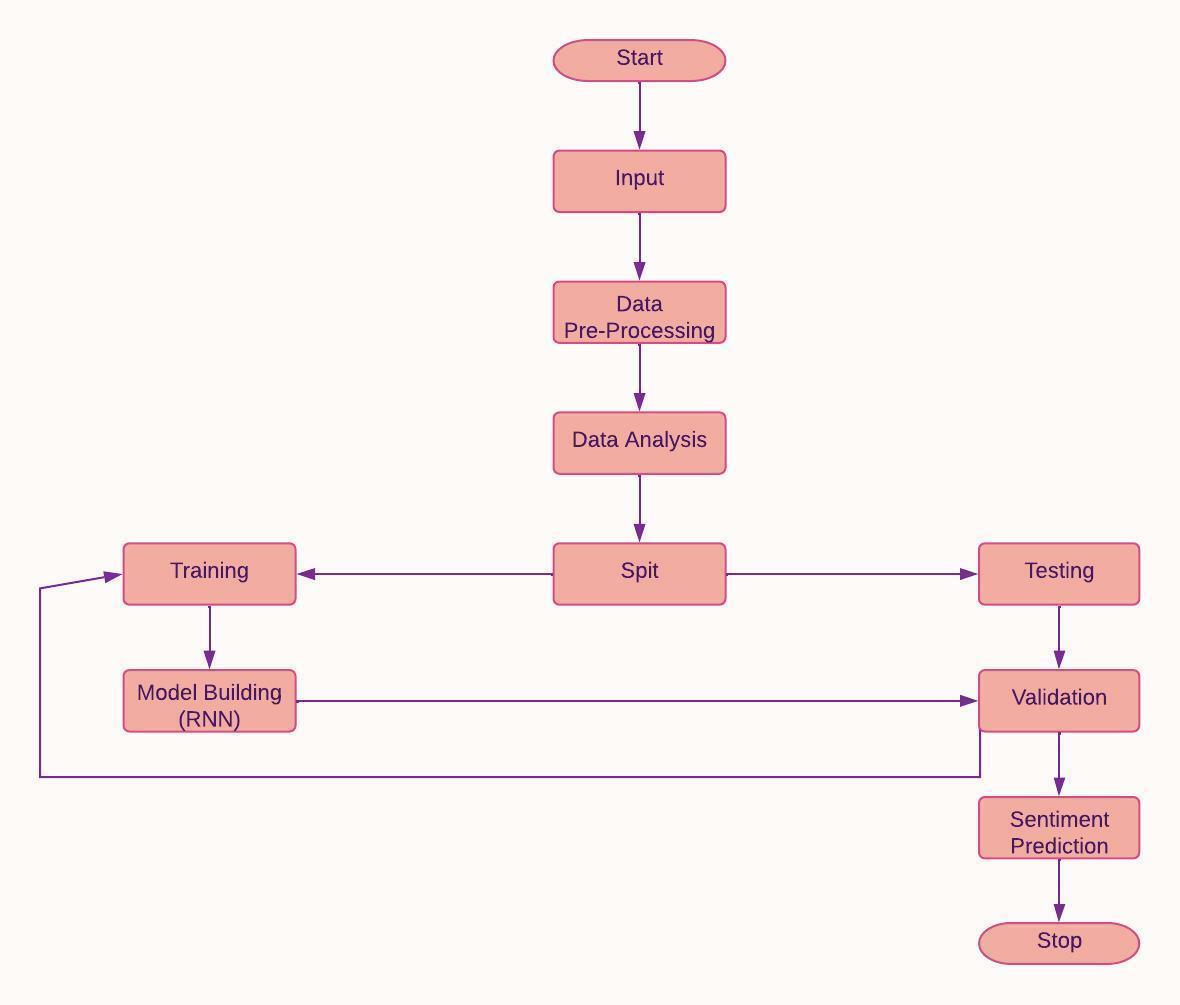
****

31

**4.3. Methodology**

This section describes clearly how our system was developed.

* + Firstly, to perform the operation we need a dataset related to tweets. We collect a Twitter dataset related to the covid-19 lockdown imposed across India. After preparing the dataset, we will perform the preprocessing on the prepared dataset.
  + Once the dataset is prepared and preprocessed, we upload our dataset into the model.
  + After uploading data, we perform necessary feature engineering techniques on the considered dataset. In this step, we remove noise from the data and split the dataset into train data and test data.
  + We will apply a few NLP techniques in order to convert the dataset into machine understandable format and then we will obtain the sentiments for the particulars.
  + In the training process, we built our proposed network with the help of Neural Networks from TensorFlow. Here, we are using a machine learning model.
  + The proposed RNN architecture is based on Neural Networks. The proposed model is built using neurons from TensorFlow.
  + Here the proposed model mainly classifies the tweets as positive, negative, or neutral to the lockdown based on this proposed model classification.



**Fig 1.** Block diagram of the proposed method

32

**4.3.1 Natural language processing (NLP)**

Natural language processing allows computer systems to talk with human beings in their local language even as additionally automating diverse language-associated processes. NLP, for instance, allows computer systems to study textual content, listen to the voice, examine it, gauge sentiment, and perceive which bits are significant. Computers can now interpret extra dialect facts than humans, without turning into fatigued and in a consistent, unbiased manner. Given the huge quantity of unorganized records created daily, from healthcare notes to media platforms, automation can be critical for efficiently studying textual content and audio records. Language studying is awe-inspiringly complex and nuanced. We can explicit ourselves orally and in writing in countless type of ways. There are numerous languages and dialects, each with its own personal set of grammatical and syntactic rules, vocabulary, and terminology. We mechanically misspell, truncate, or pass over punctuation whilst we write. When we speak, we've got local accents, stutter, stammer, and use vocabulary from many a couple of languages.

While supervised and unsupervised learning, particularly profound learning, are now used for modeling human language, semantic and syntactic interpretation, as well as domain knowledge, will continue to be required, which current machine learning approaches cannot always supply. NLP is important because it allows for the resolution of linguistic ambiguity and provides quantitative order to data in a range of application areas such as speech recognition and text analytics.

Natural language processing encompasses a wide range of tools for detecting human language, including statistical and machine learning approaches, as well as rule-based and algorithm-based approaches. We require a diverse set of techniques since text- and voice-based data, as well as actual applications, vary greatly.

Tokenization and parsing, lemmatization, stemming, part-of-speech, language recognition, and linking signals are all basic NLP activities. You've done these chores manually previously if you've ever diagrammed sentences in grade school.

NLP activities, in general, decompose language into smaller, essential components, attempt to understand their connections, and examine how those components interact to represent meaning.

33

**4.3.2 Neural Networks:**

• A Neural Network is created of multiple layers that are coupled along to figure out the structure and performance of the human brain. It is learning from vast amounts of knowledge and using difficult algorithms to be coached.

• As an example, a neural network will acknowledge a dog breed supporting its characteristics.

• The neural network’s input layer receives picture pixels from 2 completely different dog breeds. The image pixels are later processed for feature extraction within the hidden layers.

•Based on the output layer, the dog can either be a German shepherd or a Labrador.

•A network like this doesn't wish to store its previous output.

* Several neural networks can assist in the resolution of various business difficulties. Let us have a look at a few of them.

**4.3.3.Recurrent neural networks (RNN)**

Recurrent neural networks retain the result of a certain layer and use it to predict its output.

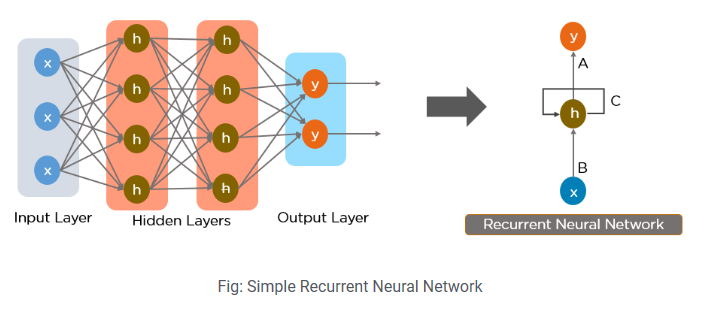


Fig 2.Simple recurrent neural network

In a feed-forward neural network, communication takes place only in a forward direction from input neurons to hidden layers and output nodes. The network is not cyclical or looped. A simple representation of a feed-forward neural network can be seen here:

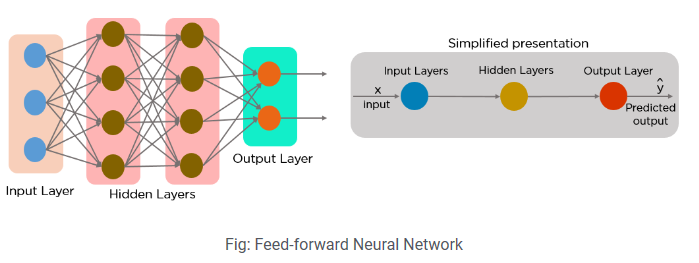


Fig 3.The feed forward neural network

34

Recurrent neural networks were developed as a result of a few flaws in the feed-forward neural network:

• Unable to process sequential data

• Analyzes only the current input

• Unable to remember earlier inputs

The Recurrent Neural Network is the solution to these problems (RNN).

An RNN may handle sequence information by accepting both current and previously received entries[12]. Because of their internal memory, RNNs can remember past inputs.

These RNNs are mostly used in the applications of NLP, Time series, Language Translation[13].

35

**4.4. TESTING PROCESS**

Testing is the technique engaged with endeavoring to perceive each possible flaw or weakness in the thing product. It is an item application to ensure that the item structure experiences its necessities and wants of customers and doesn't crash and burn in an unacceptable way. There are different sorts of tests. Every sort of test meets a specific test need.

**4.4.1TYPES OF TESTS**

* **Unit testing**

Unit testing incorporates the arrangement of test conditions that ensure that the inside system understanding works properly, and that structure inputs produce significant outcome. All pieces of decision and stream of inside code ought to be affirmed. Testing of each item unit of the application . This is a preliminary of the plan, considering data on its turn of events and attack. Unit testing performs fundamental tests at deficient level and tests a specific business cooperation, application, as well as structure plan. Unit testing ensures that each intriguing business process procedure works unequivocally in the formed detail and contains clearly portrayed information sources and expected outcomes.

* **Integration testing**

Tests run the event and are extraordinarily stressed over the major effect of screens or fields. Blending tests show that though the parts were only satisfied, as affirmed by the accomplishment of the unit test, the part parts are consistent and stable. Blending tests are unequivocally planned to perceive issues arising out of compound parts.

* **Functional testing**

Valuable testing offers organized hints that attempted endeavors are available as directed by business and advancement necessities, structure documentation, and customer manuals.

36

Functional assessment focuses on the following:

* Valid Input: Specified categories of valid entries should be accepted.
* Invalid Input: The specified categories for invalid input should be rejected.
* Tasks: The specified tasks should be used.
* Output: Specified categories of application results should be used.
* Systems: Processing systems or processes should be used.

The arrangement and change of execution assessments bases on exceptional requirements, endeavors, or phenomenal analyses. Additionally, conscious wellsprings of data related to perceiving the movement of Business process; data fields, pre-portrayed cycles, and progressive cycles should be considered for testing. Before the display test is done, additional tests are not actually settled forever the effective worth of the current test.

* **System test**

It truly investigates the plan to certify known and obvious results. An outline of a system test is a structure blend test zeroed in on fix. Structure testing relies upon process depictions and transfer, which highlights associations with pre-run cycles and compromise centers.

* **White Box testing**

The White Box is a test where the item analyzer is familiar with within working, development and language of the item, or on the other hand if nothing else its inspiration. It is purposeful. Used to examine closed off areas from the disclosure level.

* **Black Box testing**

Discovery Checking programming with close to no data on the inside limit, plan or language of the test module. The disclosure test, similarly as other various types of tests, should be written in an asserted source, still up in the air chronicles or requirements, for instance, a specific or necessities record. Testing gives data and responds to results dismissing how the item capacities.

37

* **Unit testing**

Unit testing is generally done as a component of the code coordinated with the product life cycle unit test segment, in spite of the fact that it is normal for coding and unit testing to be done as two separate stages.

**Test objectives**

• All field passages should work appropriately.

• Pages ought to be empowered from the predetermined connection.

• Login screen, messages and answers ought not be deferred.

**Features to be tested**

• Make specific the entrance is in the best configuration

• No copy sections ought to be permitted

* **Integration Assessmen**t

Programming compromise testing is the slow testing of the coordination of somewhere around two consolidated programming parts into one region to convey dissatisfactions coming about due to interoperability hurt.

A reconciliation test can check whether different parts of programming, for example a product framework or - in the future - the application of the product in its organization, work together without error.

* **Admission Test**

Client Recognition Testing is a significant period of any task and requires huge end client support. It additionally guarantees that the framework meets working prerequisites.

Test results: All of the experiments referenced above have been effective. There are no mistakes experienced.

38

**Chapter 5: Results/Output**

This session will discuss the results obtained by performing the above-proposed method and how they are extracted using the NLP and RNN architectures.

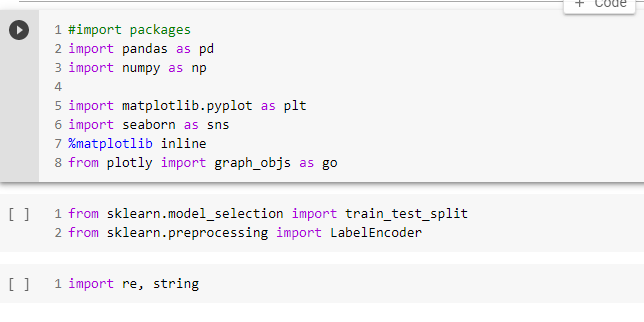


Fig 4. The Libraries You Will Need.

The above images represent the loading procedure of required libraries in order to implement the project.

39

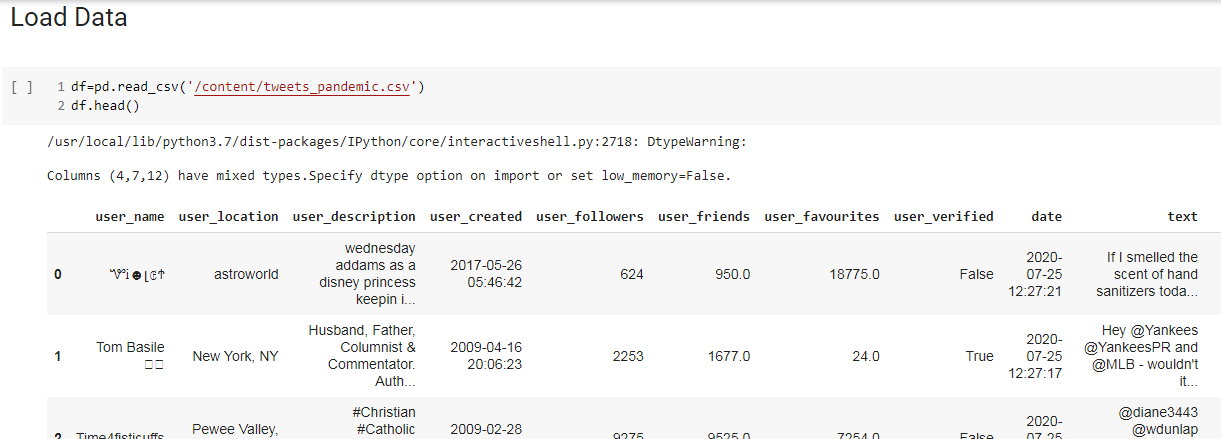


Fig 5. A data load is underway.

The above-shown figure is showing us how to load the data into the work environment i.e., python’s Jupyter notebook.



Fig 6. Plot of the Count.

The above-shown figure is the count plot which includes the count of positives, negatives, and neutrals in our particular column.

40

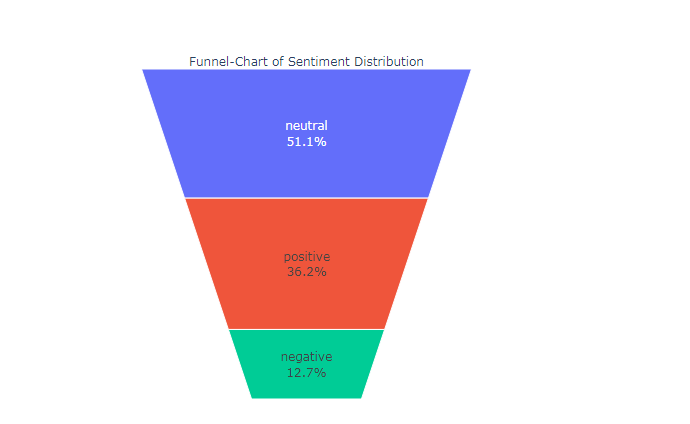


Fig 7.A funnel chart representing sentiment analysis

The above-shown figure is the Funnel count plot which includes the count of positives, negatives, and neutrals in our particular column with percentages.



Fig 8. Organizing and cleaning data.

The above-shown figure is the cleaning part of data where we are removing unwanted attributes within the dataset.

41

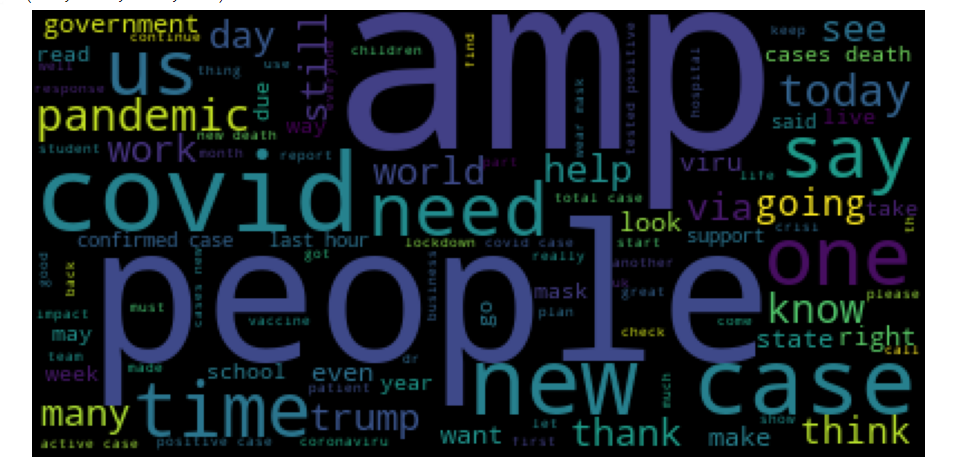


Fig 9. Cloud of tweets.

The above-shown figure is the Tweets cloud for our data which shows the data in a cloud representation.

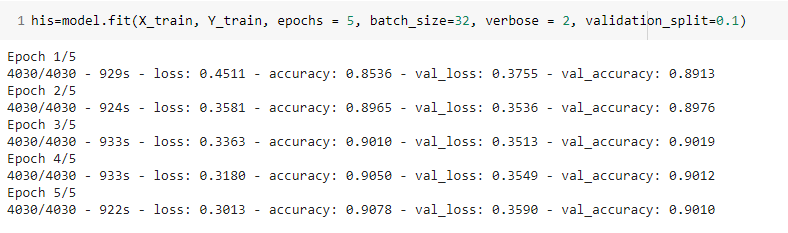


Fig 7. Construction of model buildings.

The above-shown figure represents the things related to model building using RNN architecture.

42

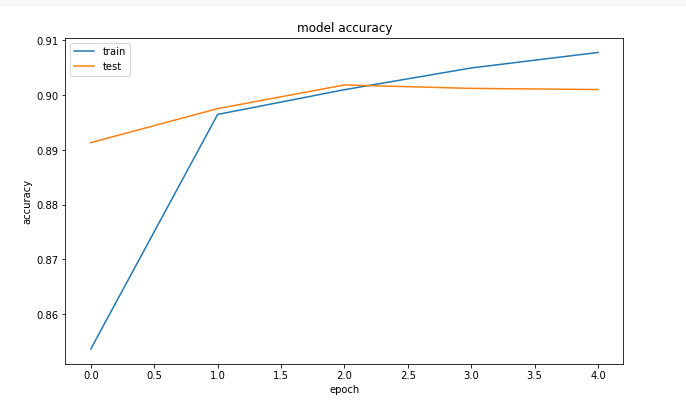


Fig 8.The Plot of Accuracy

The above-shown figure is the accuracy plot for a particular model built for the project.

43

**Chapter 6: Conclusion**

The study provides the classification of tweets using a machine learning technique. Here we used Recurrent Neural Network’s architecture (RNN) algorithm. The proposed RNN architecture is based on neural networks and is trained and tested on the dataset collected.

In this project, we will explore sentiment analysis for COVID-19 during lockdown among Indians. We used RNN for modeling purposes generating adequate emotion prediction accuracy. The emotions are classified into 3 groups positive, neutral, and negative with an accuracy of around 90%. In the future, we may consider new data for lockdown 2.0 as well and remodel it. It could act as a comparison between sentiments of lockdown 1.0 vs 2.0.

44

**Chapter 7: REFERENCES**

[1] Coronavirus cases, 2021. [Online]. Available: https://www.worldometers.info/coronavirus/ ?utm%5C\_campaign=homeAdvegas1?%5C#countries.

[2] medrxiv.org. [Online]. Available: [https://www.medrxiv.org/content/10.1101/2020.08.07. 20170548v1](https://www.medrxiv.org/content/10.1101/2020.08.07.%25252020170548v1).

[3] Canadian mental health association on reconciliation and mental health, 2021. [Online]. Available: https://cmha.ca/news/canadian- mental- health- association- on- reconciliationand-mental-health.

[4] Covid-19 vaccinations, 2021. [Online]. Available: https : / / ourworldindata . org / covid - vaccinations.

[5] Vaccine ranking of countries per doses administered, 2021. [Online]. Available: https://www. atlas- mag.net/en/article/covid- 19- vaccine- ranking- of- countries- per- dosesadministered.

[6] M. Motoyama, B. Meeder, K. Levchenko, G. M. Voelker, and S. Savage, “Measuring online service availability using twitter.,” WOSN, vol. 10, pp. 13–13, 2010.

[7] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, “Sentiment analysis of twitter data,” Proceedings of the Workshop on Languages in Social Media, Jan. 2011.

45

[8] D. H. Jin, L. Chae-Gyun, K. Y. Jin, and C. H. Jin, “Analyzing emotions in twitter during a crisis: A case study of the 2015 middle east respiratory syndrome outbreak in korea,” in 2016 international conference on big data and smart computing (BigComp), IEEE, 2016, pp. 415–418. [9] L. Lifang, Z. Qingpeng, W. Xiao, et al., “Characterizing the propagation of situational information in social media during covid-19 epidemic: A case study on weibo,” IEEE Transactions on Computational Social Systems, vol. 7, no. 2, pp. 556–562, 2020.

[10] B. Ankita, M. Manoj, K. Niranjan, and T. Siddharth, “Analysis of hospital reviews through sentiment analysis: An approach to aid patients in the times of covid-19 pandemic,” 2021.

[11] S. Koustuv, T. John, C. E. D, and D. C. Munmun, “Social media reveals psychosocial effects of the covid-19 pandemic,” medRxiv, 2020.

[12] J. Hamed, W. Yongli, O. Rita, and H. Shucheng, “Deep sentiment classification and topic discovery on novel coronavirus or covid-19 online discussions: Nlp using lstm recurrent neural network approach,” IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 10, pp. 2733–2742, 2020. 36 BIBLIOGRAPHY BIBLIOGRAPHY

[13] L. Hanjia, C. Long, W. Yu, and L. Jiebo, “Sense and sensibility: Characterizing social media users regarding the use of controversial terms for covid-19,” IEEE Transactions on Big Data, 2020.

[14] M. Amrita, K. Purnima, and V. Sonali, “Emotional analysis using twitter data during pandemic situation: Covid-19,” in 2020 5th International Conference on Communication and Electronics Systems (ICCES), IEEE, 2020, pp. 845–848.

[15] H. Man, L. Evelyn, H. E. S, et al., “Social network analysis of covid-19 sentiments: Application of artificial intelligence,” Journal of medical Internet research, vol. 22, no. 8, e22590, 2020.

46

[16] S. Sohini, M. Sareeta, and S. Garima, “An exploration of impact of covid 19 on mental health-analysis of tweets using natural language processing techniques,” medRxiv, 2020.

[17] S. Jim, A. GG, R. Md, E. Ek, S. Yana, et al., “Covid-19 public sentiment insights and machine learning for tweets classification,” Information, vol. 11, no. 6, p. 314, 2020.

[18] K. Anna, H. berle Matthias, K. Iona, and Z. X. Xiang, “Cross-language sentiment analysis

of european twitter messages duringthe covid-19 pandemic,” arXiv preprint arXiv:2008.12172, 2020.

[19] R. Furqan, K. Madiha, A. Waqar, R. Vaibhav, M. Arif, and C. G. Sang, “A performance comparison of supervised machine learning models for covid-19 tweets sentiment analysis,” Plos one, vol. 16, no. 2, e0245909, 2021.

[20] K. Harleen, A. S. Ul, A. Bhavya, and C. Victor, “A proposed sentiment analysis deep learning algorithm for analyzing covid-19 tweets,” Information Systems Frontiers, pp. 1–13, 2021.

[21] Understanding n-grams. [Online]. Available: <https://towardsdatascience.com/understandingword-n-grams-and-n-gram-probability-in-natural-language-processing>.

[22] Implementing deep learning methods. [Online]. Available: https : / / www . kdnuggets . com / 2018 / 04 / implementing - deep - learning - methods - feature - engineering - text - data - cbow.html.

[23] An implementation guide to word2vec using numpy and google sheets, 2018. [Online]. Available: https : / / towardsdatascience . com / an - implementation - guide - to - word2vec - using - numpy-and-google-sheets-13445eebd281.

47

[24] Implementing deep learning methods. [Online]. Available: https://towardsdatascience.com/ nlp-101-word2vec-skip-gram-and-cbow.

[25] Using countvectorizer to extracting features from text, 2020. [Online]. Available: https://www. geeksforgeeks.org/using-countvectorizer-to-extracting-features-from-text.

[26] Semi-supervised learning, 2020. [Online]. Available: <https://algorithmia.com/blog/semisupervised-learning>.

[27] Semi-supervised learning, 2021. [Online]. Available: https : / / www . geeksforgeeks . org / feature-selection-techniques-in-machine-learning/.

[28] N. Abbasi, What is tf-idf? [Online]. Available: https://www.educative.io/edpresso/whatis-tf-idf.