



STATISTICAL ANALYSIS OF POST COVID SYMPTOMS ON RECOVERED PATIENTS WITH ML

Submitted by:

Bibi Zoynob

ID: 213-44-223

Course: Thesis

Batch: 16th

Supervised by:

Afsana Begum

Assistant Professor

Department of Software Engineering

Daffodil International University

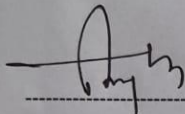
**The Thesis report has been submitted in fulfillment of the requirements for the Degree of Masters
of Science in Data Science.**

APPROVAL

APPROVAL

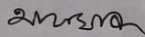
This thesis/project/internship titled on "Statistical Analysis of Post-covid Syndrome on Recovered Patients", submitted by Bibi Zoynob Tumpa, ID: 213-44-223 to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Masters of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS



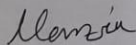
Chairman

Dr. Abdul Kadar Muhammad Masum
Professor
Department of Software Engineering
Daffodil International University



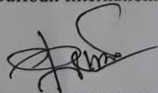
Internal Examiner 1

Afsana Begum
Assistant Professor
Department of Software Engineering
Daffodil International University



Internal Examiner 2

Dr. Marzia Ahmed
Assistant Professor
Department of Software Engineering
Daffodil International University



External Examiner

Md. Tanvir Quader
Senior Software Engineer,
Technology Team,
a2i Program

DECLARATION

DECLARATION

I hereby state that I have taken this thesis under the supervision of **Afsana Begum, Assistant Professor, Department of Software Engineering, Daffodil International University**. I also acknowledge that neither this thesis nor any part of this has been submitted elsewhere for the award of any degree previously by others.



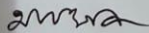
Bibi Zoynob

ID: 213-44-223

Department of Software Engineering

Faculty of Science & Information Technology Daffodil International University

Certified by:



Afsana Begum

Assistant Professor

Department of Software Engineering

Faculty of Science and Information Technology

Daffodil International University

© All right Reserved by Daffodil International University

21

ACKNOWLEDGEMENT

I'm very appreciative to the Almighty Allah (SWT) for permitting me to finish this Masters of Science study with this examination.

I'm appreciative to my parents who have given me support for pursuing my study all the time.

I need to thank my well-wisher and my interviewees. Also great thanks to my doctors who helps me to validation my questionnaires for this research. Without their contribution and time, it wouldn't be a completion of my paper.

I would like to want to show my regard, respect and appreciation to my Supervisor Afsana Begum Madam for supporting me in finishing this examination and empowering me in research and execution of this work.

Furthermore, I might want to express gratitude toward Dr. Imran Mahmud, Associate Professor and Head in Charge of the Department of Software Engineering, Daffodil International University for spreading his knowledge to us to explore more. I likewise need to thank my instructor again who has been ceaselessly supporting us all through this postgrad. Finally yet importantly, I need to thank all the staff of the Department of Software Engineering for consistently endeavoring to furnish us with the best offices in study halls and research facilities.

TABLE OF CONTENTS

APPROVAL	ii
DECLARATION.....	iii
ACKNOWLEDGEMENT.....	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	vii
ABSTRACT.....	1
CHAPTER 01.....	2
1.1 BACKGROUND	2
1.2 MOTIVATION OF THE RESEARCH	4
1.3 RESEARCH QUESTIONS	4
1.4 RESEARCH OBJECTIVES	4
1.5 RESEARCH SCOPE	4
1.6 RESEARCH ORGANIZATION	5
CHAPTER 02.....	6
2.1 INTRODUCTION.....	6
2.2 PREVIOUS LITERATURE	6
2.3 LITERATURE GAP	7
CHAPTER 03.....	8
3.1 DATA COLLECTION & OVERVIEW	8
3.2 SURVEY QUESTIONNAIRE	9
3.3 DATA PRE-PROCESSING.....	11
Algorithm Model (Technique):	16
3.4 ALGORITHM USED	17
3.3.1 LINEAR REGRESSION	17
3.3.2 NAÏVE BAYES CLASSIFIER.....	17
3.3.3 RANDOM FOREST	17
3.3.4 CLASSIFICATION & REGRESSION TREE (CART).....	17
3.3.5 SUPPORT VECTOR MACHINE (SVM)	17
ANOVA TEST	18
CHAPTER 04.....	21
4.1 RESULT ANALYSIS	21

TRAIN DATASET ACCURACY TEST:	22
CHAPTER 05: CONCLUSION & RECOMENDATION	27
5.1 FINDINGS & CONTRIBUTIONS	27
5.2 LIMITATIONS	28
5.3 RECOMMENDATION FOR FUTURE WORK	28
REFERENCES	29
PLAGIARISM REPORT	30
ACCOUNTS CLEARANCE	32

LIST OF FIGURES

Figure-1: Long term effects of COVID-19.....	10
Figure 2: Dataset of covid 19 patients.....	15
Figure: 3: Survey form (Questionnaire).....	17
Figure 4: Data collection methodology.....	18
Figure 5: Descriptive Statistics of Dataset.....	19
Figure 6: outliers.....	20
Figure-7: Histogram of the dataset.....	21
Table-1: Descriptive Statistics of the Dataset.....	22
Table-2: Pearson's correlation.....	22
Figure 8: Algorithm technique.....	23
Figure 9: ANOVA Two-way Test Result.....	25
Figure 10: K-means Algorithm and clustering.....	27
Figure 11: OLS result from regression analysis.....	29
Figure 12: Train dataset accuracy test.....	29
Figure 13: Test dataset accuracy.....	30
Figure 14: Vaccination wise Post Syndrome in Boxplot.....	31
Figure 15: Vaccination wise post covid syndrome in Distribution plot.....	32
Figure 16: Vaccination wise post syndrome in Heat map.....	33

ABSTRACT

As the COVID-19 pandemic elapses, too many attention is being given to long COVID: it is a set of symptoms that people experience after covid infection. Since it is so new, we are still learning about it. The more we understand post-covid syndromes, the more it can help people with both long COVID and everybody else. Those with long COVID is called 'long haulers', with the research community it is still settling on a formal term that's accurate.

The study aims to investigate the clinical manifestations and vaccination which were received by the patients that, examining that the patients has post Covid syndrome even if they received Covid vaccine or not.

By integrating insights from diverse fields, this thesis endeavors to provide a holistic understanding of Post Covid Syndrome's and foster targeted interventions for patient care and health management. So that, patients can be awarded about taking vaccination.

Covid-19 vaccinations currently authorized for Bangladesh are Pfizer and Moderna. This thesis research will try to resolve the complexity of post Covid syndrome for these vaccination in Bangladesh.

Keywords

World Health Organization, Blood Pressure, Low Pressure, Post Covid Syndrome, Vaccine, Odorless, Dyspnea, Low Vision

CHAPTER 01

INTRODUCTION

1.1 BACKGROUND

There were 2 Million Covid cases in Bangladesh where deaths were 29,493 till 2023. Not only in Bangladesh, has the Covid-19 pandemic had fell impact on global health, millions of people infected by it. While many people recover from this illness, some continue to have post covid syndrome even though they are vaccinated and some cases they are not. Most people who develop COVID-19 fully recover, but currently we see they are approximately 50%-60% affected by the post covid syndrome which is mid and long-term effects. That's why it increases attention of researchers to signify the challenge for healthcare sector and has become an area of increasing concern for public health authorities.

It's important to remember that our understanding of post COVID-19 syndrome may help affected individuals to survive.

The list of symptoms is too random and isn't fully defined yet. Researchers are working with patients who develop post COVID-19 syndrome. However to better understand more about the syndromes as in vaccinated or not also its symptoms and effects, this research will try to find out them while data collecting.

The main purpose of this thesis is to find out the most affected patients when they are vaccinated or not.

In data collecting, patients were too helpful to assist while surveying. They provided their information about covid affection. Patients were in different ages. But they all faces post covid syndrome. Some of the patients suffered from low vision, memory loss which they never experienced before covid. Also, High blood pressure, Low blood pressure, pain in the body, lungs infection, and uric acidic problem has been found in this research.

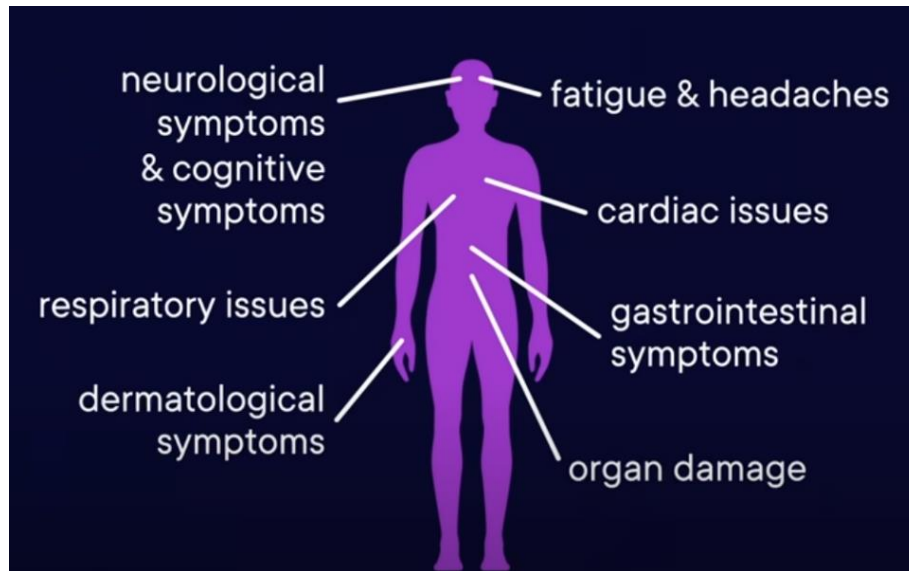


Figure-1: Long term effects of COVID-19 (Google).

1.2 MOTIVATION OF THE RESEARCH

Researchers from other countries are trying to define post Covid syndrome or long Covid. However, researchers should find out more specification. Surroundings that I have found some patients who developed several symptoms that they did not have before. In this research, we will use random method and questionnaires to find out best information regarding post-covid syndrome. By using machine learning algorithm, this research will try to specify the symptoms of post covid if patients are vaccinated or not.

1.3 RESEARCH QUESTIONS

1. Does covid recovered patients suffers from post covid syndrome ?
2. What are the symptoms they are suffering ?
3. Are they vaccinated who suffered from post covid syndrome ?
4. What are the ratio of suffering post covid syndrome if vaccinated ?

1.4 RESEARCH OBJECTIVES

Objective 1: To find out the ratio of patients who are vaccinated.

Objective 2: To find out at the most syndrome into covid patients.

Objective 3: This research will show Bangladesh's situation about covid patients.

1.5 RESEARCH SCOPE

- This research will help people to find out more about post covid syndrome.
- People will acknowledge if they are affected via vaccination or not.
- People will acknowledge about the most affected syndrome so that, they can remain aware.

1.6 RESEARCH ORGANIZATION

Chapter 1: It presents the introduction of this thesis paper.

Chapter 2: This chapter represents literature review, literature gap and previous literature that researcher works on this topic previously.

Chapter 3: This chapter is providing about the data collection method, pre-processing, survey, questionnaire, model architecture.

Chapter 4: This chapter analyzed result that found while proceeding experiment, evaluation of the study.

Chapter 5: This chapter highlights the contribution, limitations and future work.

CHAPTER 02

LITERATURE REVIEW

2.1 INTRODUCTION

In literature review, a researcher should review the previous research, article, journals, publications etc. to see if the relevant work has been done till now on that topic. In a research paper, a researcher can use literature review as an establishment and as help for another knowledge that you contribute to get better results.

2.2 PREVIOUS LITERATURE

In this research, a large population-based prospective cohort study has been done, a significant proportion of patients with COVID-19 developed long-term health problems after the acute infection. It says, population-based prospective cohort of 431 patients in Switzerland.

Eligibility criteria included a positive RT-PCR test result and age of 18 years or older. Positive RT-PCR test result between Oct. 6, 2020 and Jan. 26, 2021, at a median of 7.2 months after their diagnosis. [5]

In this study, they summarized post covid syndrome's affects and 10-35% of patients suffering from covid. However, only very few common symptoms and primary precautions were discussed here. [6]

In this research, researcher studied only on 153 patients who suffered from covid-19 at a specific care center in Sri Lanka. The study conducted particularly on South Asian people. It is a prospective cohort study to follow up patients with moderate to severe disease or mild disease with multi chronic conditions at 2 and 6 weeks, 3 and 6 months and 1 year from discharge. [7]

Basically, the after health care of covid-19 is discussed here. It says pathologic entity, persistent physical, medical and cognitive sequelae are referred to post covid syndrome. This research discussed about pathologic fibrosis of organs and vasculature which is worsened quality of life. It also worked on inhibiting transforming growth factor beta (TGF- β), fibrosis modulator. [8]

In this research, they discussed about those patients who discharged from the hospital but readmitted again. It is showing after acute covid-19 patients were readmitted 14060 out of 47780 and more than 1 in 10 died

after discharge. It also discussed about to quantify rates of organ specific dysfunction in individuals with covid-19 after discharge from NHS hospitals in England. [9]

This research focuses on organ damaging if it is caused by covid-19. It says, after the initial acute infection, like many other viral disorders, a multi conditional long-lasting symptoms have been described. [10]

This research shows only two symptoms which are: fatigue and dyspnea. It studied on only 43 data in a specific area and on who are lingering and debilitating symptoms following acute covid-19 infection. It studies on National Institute for Health and Care Excellence patients who suffering from post covid symptoms three weeks beyond from initial stage. [11]

This research had discussed on Dhaka Medical College Hospital between June 01 2020 to August 10, 2020. It studied only 3 months data of that hospital. Among 400 patients, they analyzed 355 patients. It founds 46% patients developed post-covid symptoms and 70% cases, post viral fatigue being the most appeared symptom. [12]

This research have been done on 14392 participants who were suffering from covid in Bangladesh. It discussed between Jun 2020 to Nov 2020. It defined 8 long covid symptom they have found. The duration of the post covid symptom is 5.2 weeks. Researchers focused on specifying post covid syndrome. [13]

2.3 LITERATURE GAP

All of the research says about post covid syndrome and its duration. But my proposal is to be confirmed that patients have post covid syndrome if they are vaccinated most. None of the research paper mentioned about vaccination.

CHAPTER 03

METHODOLOGY

3.1 DATA COLLECTION & OVERVIEW

The dataset was collected from 2 district **Dhaka & Feni** from Bangladesh. The method used for data collection is **Simple Random Sampling** model. It was randomly collected from those two districts and these data are **real data** collected via live interview. There are 12 columns including patient name, gender, age, year, month, location, vaccination, post covid syndrome, length of syndrome, weakness, cured or not, previous medical history if any covid patient have and 8 rows according to 2020 to 2022.

[] df.head(10)

	Patient_Name	Gender	Age	Covid_Year	Location	Vaccinated	Post_Covid_Syndrome	Has_Post_Syndrome	Length_months	Length_days	...	Odorless	Tasteless	Pain in the body	Allergy on cold	Lungs Infection
0	MD. MAHMUDUL HASAN	1	30	2021	Mohammadpur	1	High BP	1	7	214	...	2	2	2	2	2
1	KAMRUN NAHAR ONNI	2	30	2021	Mohammadpur	1	High BP	1	12	365	...	2	2	2	2	2
2	MD. NURUL HOQUE	1	66	2022	Dhaka Cantt.	1	Low BP	1	21	638	...	2	2	2	2	2
3	MANOARA BEGUM	2	55	2022	Dhaka Cantt.	1	High BP	1	14	426	...	2	2	2	2	2
4	BIBI ZOYNOB TUMPA	2	33	2021	Dhaka Cantt.	1	Low BP, Dyspnea	1	7	213	...	2	2	2	2	2
5	AFROZA AKTAR	2	39	2021	Feni	2	Odorless, Tasteless	2	18	549	...	1	1	2	2	2
6	SHAKUR SHAMS	1	35	2021	Dhanmondi	1	Dyspnea	1	20	609	...	2	2	2	2	2
7	MRS. NISHITA	2	33	2021	Dhanmondi	1	Pain in the body	1	23	700	...	2	2	1	2	2
8	MRS KULSUM	2	28	2021	Mohammadpur	1	Weakness	1	6	182	...	2	2	2	2	2
9	MRS ZAMILA KHATUN	2	35	2021	Mohammadpur	1	Weakness	1	13	396	...	2	2	2	2	2

Figure 2: Dataset of covid 19 patients.

3.2 SURVEY QUESTIONNAIRE

The questionnaire I used for collecting data is real life live question and it is verified by **Dr. Sultan Abu Sadat Md. Sayem**, MBBS(CU), BMDC.

1. Patient Name
2. Gender
3. Age
4. Profession
5. Covid affected year
6. Did you take covid-19 vaccine ?
7. Did you take vaccine before corona disease ?
8. After recovery, do you feel any post syndromes ?
9. Do you have any previous medical history ?
10. Do you feel musculoskeletal symptom (পেশীর উপসর্গ) like pain in the body, joint etc. ?
11. Do you have blood pressure (High/Low) before covid infected ?

Survey form:

Figure 3 is showing the survey form which used for data collection given below:

QUESTIONNAIRES TO BE ANSWERED (SURVEY)		
Name		
Gender		
Age		
Profession		
Covid Date	Month - year	

Q-Set: 01

SL	Questionnaires	Y/N
1	Did you take covid-19 vaccine ?	
2	Did you take vaccine before corona disease ?	
3	After recovery, do you feel any post syndromes ?	
4	Do you have any previous medical history ?	
5	Do you feel musculoskeletal symptom (শেখীয়া উপসর্গ) like pain in the body, joint etc. ?	
6	Do you have blood pressure (High/Low) before covid infected ?	
7	Other:	

Q-Set: 02

Do you think you have these symptom after recovered from covid: (tick ✓ if you feel)

Weakness (দুর্বলতা)		How Long
Dyspnea (শ্বাসকষ্ট)		
Pain in the Body		
High Blood Pressure		
Low Blood Pressure		
Odorless (গন্ধহীন)		
Tasteless (স্বাদহীন)		
Other:		

Your Comment: (if you feel any other way that aren't in above questionnaire)

Figure: 3: Survey form (Questionnaire)

3.3 DATA PRE-PROCESSING

The detailed block diagram of proposed methodology has given below. I utilized that I worked with 1 real dataset, so I didn't need to merge it. I conduct the preprocessing, interviewing, EDA, standardization and k-means clustering part.

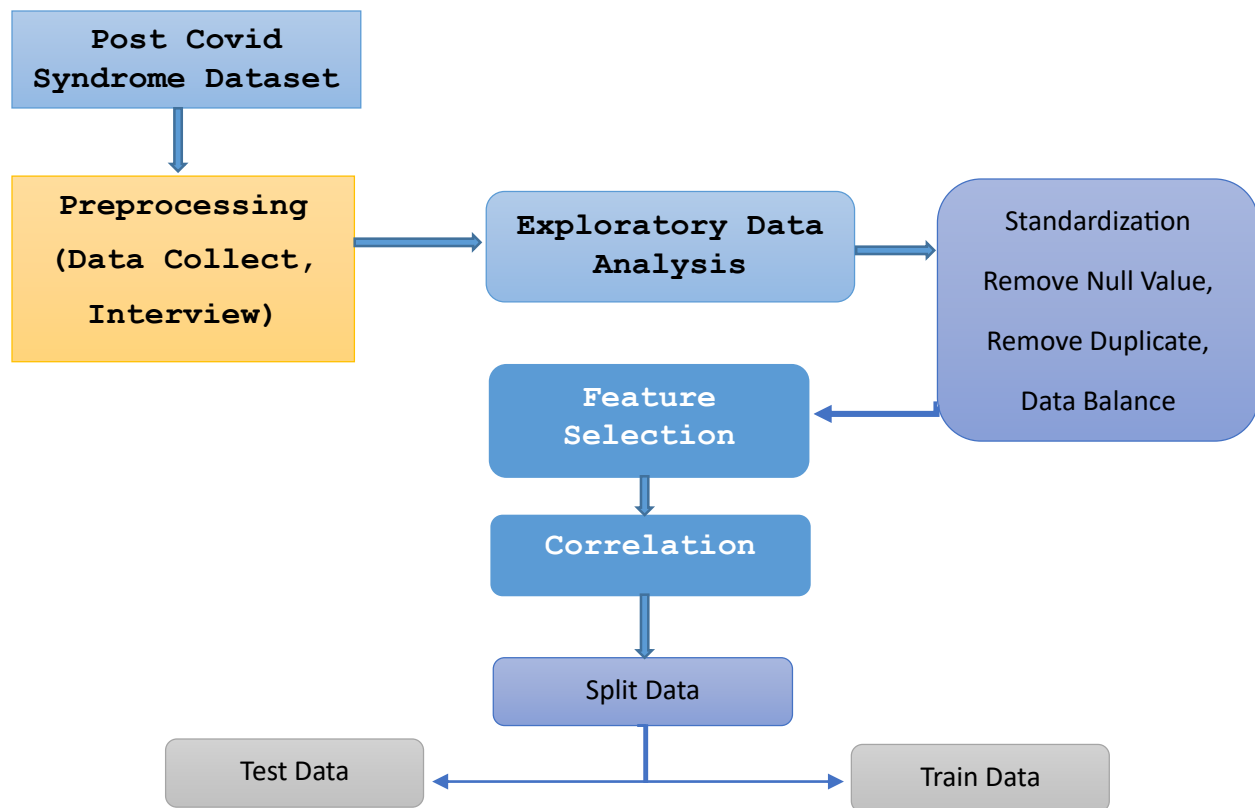


Figure 4: Data collection methodology.

MERGING AND PROCESSING

Since the data is about COVID-19 patients post syndromes, so there were no needed to merging the data. The first step is the data pre-processing. For the experimental set up, I used python programming language. In this step, I imported all necessary libraries and the proposed dataset. By removing null values, duplicate values; dataset is ready for EDA.

EXPLORATORY DATA ANALYSIS (EDA)

Exploratory Data Analysis (EDA) is a process to explore the unseen clear-sightedness from the data. Through EDA analysis, we can find various information from the data set, which is helpful in making various decisions as well as in selecting the machine-learning model. EDA is done through various types of statistical methods, data visualization etc.

DESCRIPTIVE STATISTICS

Table 1 shows the descriptive statistics of the variables. The AGE and LENGTH_DAYS have the highest standard deviation that means the most variability of these variables.

	Patient_Name	Gender	Age	Covid_Year	Location	Vaccinated	Post_Covid_Syndrome	Has_Post_Syndrome	Length_months	Length_days	...	Odorless	Tasteless	Pain in the body
count	260	260.000000	260.000000	260.000000	260	260.000000	260	260.000000	260.000000	260.000000	...	260.000000	260.000000	260.000000
unique	248	NaN	NaN	NaN	14	NaN	23	NaN	NaN	NaN	...	NaN	NaN	NaN
top	MRS JANNATUL FERDOUS	NaN	NaN	NaN	Mohammadpur	NaN	Allergy on cold	NaN	NaN	NaN	...	NaN	NaN	NaN
freq	4	NaN	NaN	NaN	83	NaN	67	NaN	NaN	NaN	...	NaN	NaN	NaN
mean	NaN	1.711538	42.557692	2021.234615	NaN	1.123077	NaN	1.130769	13.880769	409.415385	...	1.969231	1.969231	1.965385
std	NaN	0.453921	13.620921	0.499762	NaN	0.340687	NaN	0.337798	5.456498	167.561746	...	0.173025	0.173025	0.183156
min	NaN	1.000000	17.000000	2020.000000	NaN	0.000000	NaN	1.000000	6.000000	161.000000	...	1.000000	1.000000	1.000000
25%	NaN	1.000000	32.000000	2021.000000	NaN	1.000000	NaN	1.000000	9.000000	244.000000	...	2.000000	2.000000	2.000000
50%	NaN	2.000000	43.000000	2021.000000	NaN	1.000000	NaN	1.000000	14.000000	397.000000	...	2.000000	2.000000	2.000000
75%	NaN	2.000000	53.250000	2022.000000	NaN	1.000000	NaN	1.000000	19.000000	568.250000	...	2.000000	2.000000	2.000000
max	NaN	2.000000	80.000000	2023.000000	NaN	2.000000	NaN	2.000000	24.000000	731.000000	...	2.000000	2.000000	2.000000

Figure 5: Descriptive Statistics of Dataset.

RANGE AND OUTLIERS OF THE VARIABLES

The Box Plot in Figure 6 for visualizing, YEAR variable of post covid syndrome has been found as the largest range among other variables. At the same time, AGE, YEAR and LENGTH_DAYS shows some outliers. Especially, LENGTH_DAYS contains most of the extreme values. MONTH, VACCINATED, WEAKNESS, IS_CURED variables have the lowest range without any outliers.

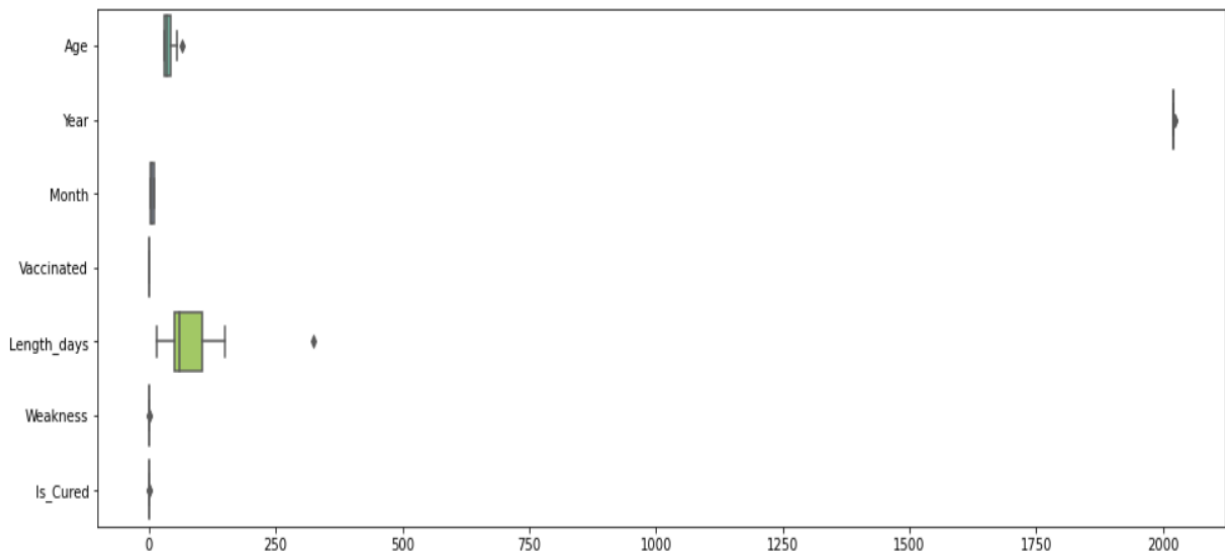


Figure 6: outliers

NORMALITY CHECK (HISTOGRAM)

Figure 7 describes the distribution of the variables. Visually the distribution shows histogram plot to show the dataset to check normality.

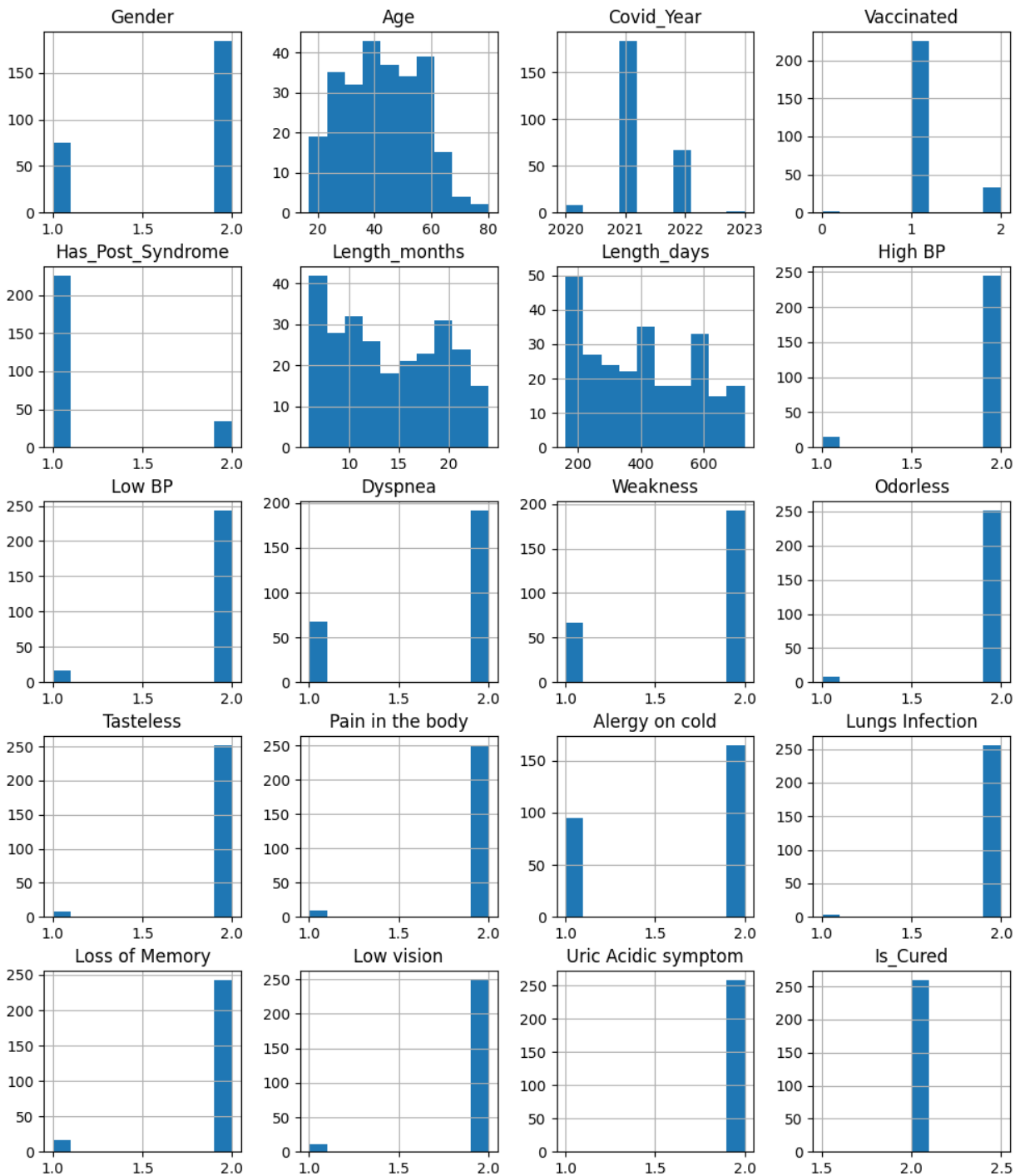


Figure-7: Histogram of the dataset.

After performing the **Anderson Darling Normality Test** to check normality, we found out that no variables following normality. The result is given in Table 1.

Variable	Result	Decision
LENGTH_DAYS	stat=11.395901, p=0.003353	Not a normal distribution
AGE	stat=11.395901, p=0.003353	Not a normal distribution
YEAR	stat=11.395901, p=0.003353	Not a normal distribution
MONTH	stat=11.395901, p=0.003353	Not a normal distribution

Table-1: Descriptive Statistics of the Dataset.

CORRELATION CHECKING

After performing the Hypothesis test for adherence checking using Pearson's correlation, Table 2 shows the weak correlation among the variables except one decision. The result shows that there are 3 independent variables and 1 dependent which is most important for this paper. From this analysis, we seem to understand that if a covid patient is vaccinated, the post covid symptoms appears most.

Variable 1	Variable 2	Pearson's Correlation	Result	Decision
Has_Post_Syndrome	Age	-0.019	stat=-0.019268, p= 0.757154	Independent
Vaccinated	Age	0.001	stat=0.001792, p= 0.977058	Independent
Has_Post_Syndrome	Vaccinated	0.933	stat=0.933195, p= 0.000000	Dependent
Vaccinated	Gender	-0.094	stat=-0.094106, p= 0.130162	Independent

Table-2: Pearson's correlation.

Algorithm Model (Technique):

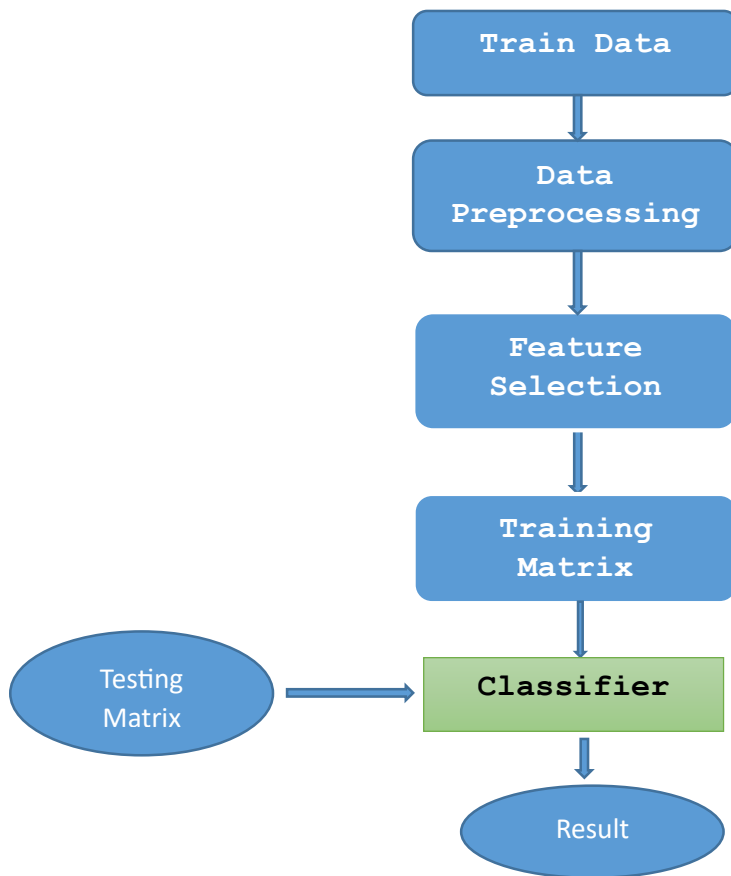


Figure 8: Algorithm technique.

3.4 ALGORITHM USED

3.3.1 LINEAR REGRESSION

Linear regression is a supervised machine learning technique. It is a statistical method used to model the relationship between a dependent variable and one or more independent variables by conducting into a linear equation to observed data. Our primary goal is to predict the value of the dependent variable based on the values of the independent variables. It predicts numerical value.

3.3.2 NAÏVE BAYES CLASSIFIER

The Naïve Bayes Classifier is also a supervised machine learning algorithm that is used for classification tasks such as text classification. They use principles of probability to perform classification tasks. Naïve Bayes is part of a family of generative learning algorithms, meaning that it seeks to model the distribution of inputs of a given class or category. Unlike discriminative classifiers, like logistic regression, it does not learn which features are most important to differentiate between classes.

3.3.3 RANDOM FOREST

The random forest algorithm is an extension of the bagging method as it utilized both bagging and feature randomness to create an uncorrelated forest of decision trees. Random forest is a commonly-used machine learning algorithm that combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fueled its adoption, as it handles both classification and regression problems.

3.3.4 CLASSIFICATION & REGRESSION TREE (CART)

A classification and regression tree is a predictive model, which explains how an outcome variable's values can be predicted based on other values. A CART output is a decision tree where each fork is a split in a predictor variable and each end node contains a prediction for the outcome variable.

3.3.5 SUPPORT VECTOR MACHINE (SVM)

A support vector machine is a type of supervised learning algorithm used in machine learning to solve classification and regression tasks; SVMs are particularly good at solving binary classification problems, which require classifying the elements of a data set into two groups.

ANOVA TEST

To test the difference of post covid syndrome, I performed ANOVA two-way test. Figure-9 shows the test result. As per the test result of each separate variable, I found some facts. That are,

- Covid patients whether male or female both may have post covid syndrome. It is not gender dependent.
- Appearing post covid syndromes does depend on vaccination. If vaccinated, the patient may suffer from post covid syndrome even without experiencing covid-19.
- Every patient is suffering from massive weakness and it became a common syndrome for all covid patients.

	Patient_Name	Gender	Age	Year	Month	Location	Vaccinated	Post_Covid_Syndrome	Length_days	Weakness	Is_Cured	Previous_Med_His
0	Md. Mahmudul Hasan	Male	30	2021	9	Mohammadpur	0	High BP	90	1	1	No
1	Kamrun Nahar Onni	Female	30	2021	9	Mohammadpur	0	High BP	14	1	1	New mom with 5 months baby
2	Md. Nurul Hoque	Male	66	2022	2	Dhaka Cantt.	1	Low BP	60	1	1	Diabetes, High BP
3	Manoara Begum	Female	55	2022	2	Dhaka Cantt.	1	High BP	150	0	0	High BP, Cholesterol
4	Bibi Zoynob Tumpa	Female	30	2021	7	Dhaka Cantt.	0	Low BP, Shortness of Breath	325	1	0	Tuberculosis with Tumour Cell

Figure 9: ANOVA Two-way Test Result.

STANDARDIZATION

Using the maximum and minimum values of this variable which scales Min – Max Normalizer, so that all data we would found within a certain range to the normalized state. In this method each data is subtracted from its minimum value (of that variable), the subtraction is divided by the range. This is how each data is scaled.

Range: Range is the difference between the maximum and minimum values of a variable.

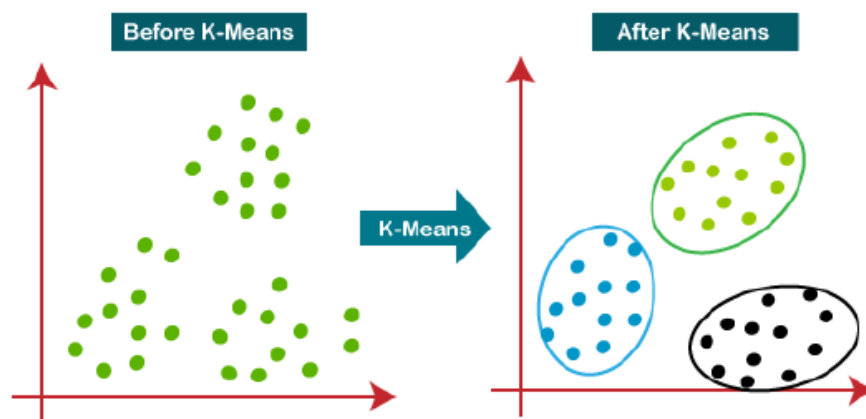
$$X_{new} = \frac{Xi - \min(X)}{\max(X) - \min(X)} \dots \dots \dots (1)$$

K MEANS CLUSTERING ALGORITHM

Figure 10 depicts K-means Clustering, an Unsupervised Learning algorithm that divides the random dataset into different clusters. In this case, K denotes the number of pre-defined clusters that must be created during the process; for example, if K=2, there will be two clusters; for K=3, there will be three clusters, and so on.

It shows us to cluster the data into different groups like post covid syndrome, location which it appears most, so that we can understand the data and research.

The algorithm starts with a random dataset, divides it into k clusters, and then repeats the process until it finds its utmost best cluster. In this algorithm, the value of k should be predetermined.



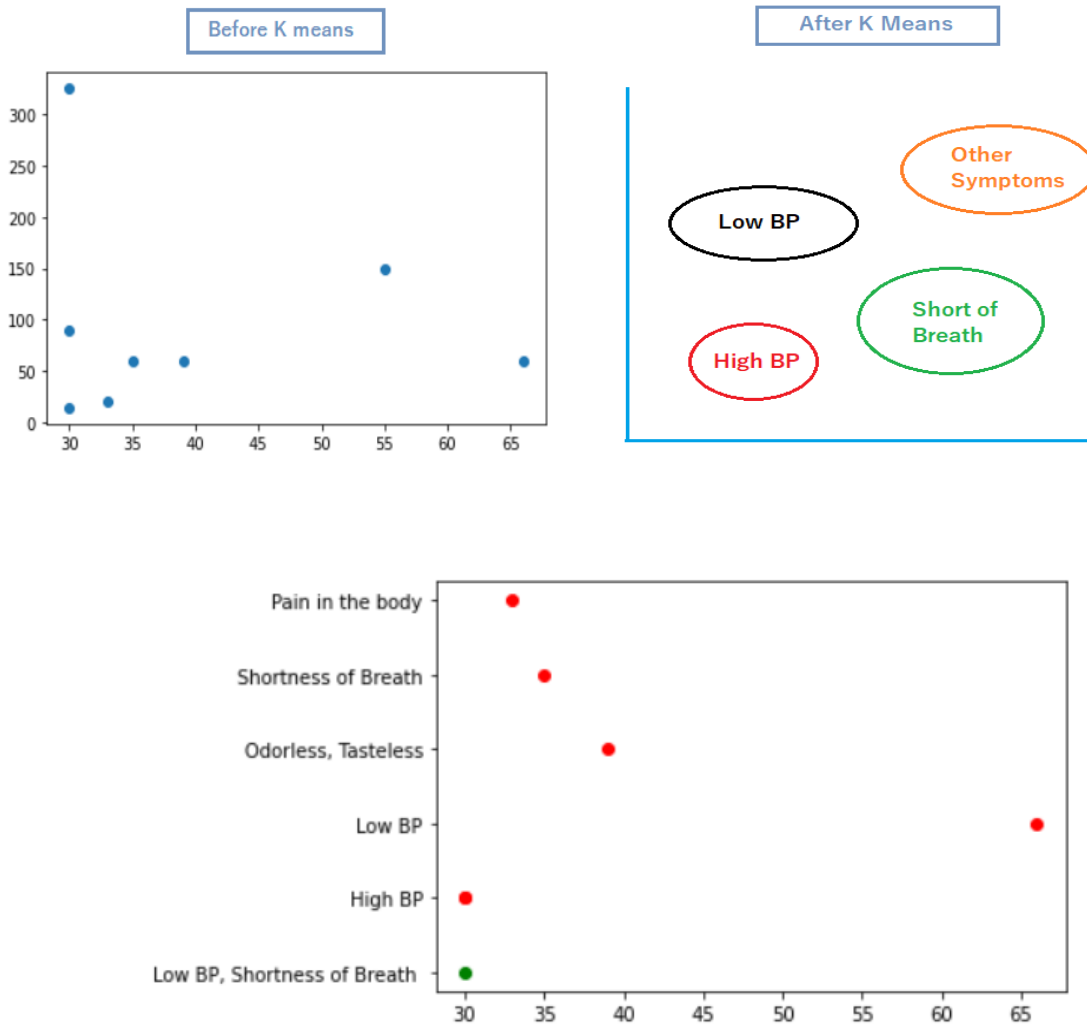


Figure 10: K-means Algorithm and clustering.

The k-means clustering algorithm mainly conduct two tasks:

Determines the best value for K Centre points and it continues with iterative process. Assigns each data point to its closest k-Centre. Those data points which are near to the particular k-Centre, create a cluster.

In K-means clustering, we use Euclidean distance to calculate the distance. Equation 2 depicts the Euclidean distance.

$$d(x, x') = \sqrt{(x_1 - x'_1)^2 + \dots + (x_n - x'_n)^2} \quad (2)$$

CHAPTER 04

We have done some experiment with the data set to see what the situation of patients if they are vaccinate or not. Does they have post covid syndrome even if they are vaccinated. The dataset is containing 260 data where

4.1 RESULT ANALYSIS

We observed from the dataset that, either vaccinated or not vaccinated both patients can have post covid syndrome. What we are looking for is for which one patients suffered from post covid. So that, we split data from the dataset and 75% data has been taken for Train data and 25% data for Test data.

We applied Linear Regression model and we get 83% accuracy of the data. Then we applied SVM and we got 88%. Next we applied Decision Tree and got 92%. Then we applied Random forest and got 98%. And lastly, we applied Naïve Bayes algorithm and we get 99% accuracy.

Linear Regression	83%
Support Vector Machine (SVM)	88%
Decision Tree	92%
Random Forest	98%
Naïve Bayes Classifier	99%

OLS Regression Results						
=====						
Dep. Variable:	Vaccinated	R-squared:	0.872			
Model:	OLS	Adj. R-squared:	0.870			
Method:	Least Squares	F-statistic:	434.3			
Date:	Tue, 08 Oct 2024	Prob (F-statistic):	1.66e-112			
Time:	06:29:53	Log-Likelihood:	178.78			
No. Observations:	260	AIC:	-347.6			
Df Residuals:	255	BIC:	-329.8			
Df Model:	4					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	0.0442	0.053	0.842	0.400	-0.059	0.148
Gender	-0.0138	0.017	-0.814	0.416	-0.047	0.020
Age	0.0005	0.001	0.896	0.371	-0.001	0.002
Has_Post_Syndrome	0.9413	0.023	41.414	0.000	0.897	0.986
Length_months	0.0012	0.001	0.851	0.395	-0.002	0.004
=====						
Omnibus:	576.562	Durbin-Watson:	1.972			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	621703.203			
Skew:	-15.199	Prob(JB):	0.00			
Kurtosis:	240.621	Cond. No.	338.			
=====						

Figure 11: OLS result from regression analysis.

TRAIN DATASET ACCURACY TEST:

From the dataset, 75% of data has been taken for the train dataset except 25% of data. The result is showing below:

Here, we can see in the below Figure 12, the Accuracy is 99%, so that, we can say our dataset model is performing accurately.

```

➡ Coefficients: [6.40606131e-04 9.27494818e-01]
Intercept: 0.04450535587905202
MAE: 0.014764986591513459
MSE: 0.0006429587747656019
RMSE: 0.025356631770911568
AccuracyII: 99.33091112724516 %
R2: 0.9933091112724516

```

Figure 12: Train dataset accuracy test.

TEST DATASET ACCURACY:

Figure 13 shows the accuracy of test dataset which is: 0.99% working accurately.

Train set Accuracy:	0.8387288494806799
Test set Accuracy:	0.9933091112724516

Figure 13: Test dataset accuracy.

VACCINATION WISE POST COVID SYNDROME

As we can see in Figure 14, it describes that, the patients who haven't taken covid vaccine got the less symptoms than who are vaccinated. So, we can pictured that, vaccinated patients are mostly affected by the post covid syndrome.

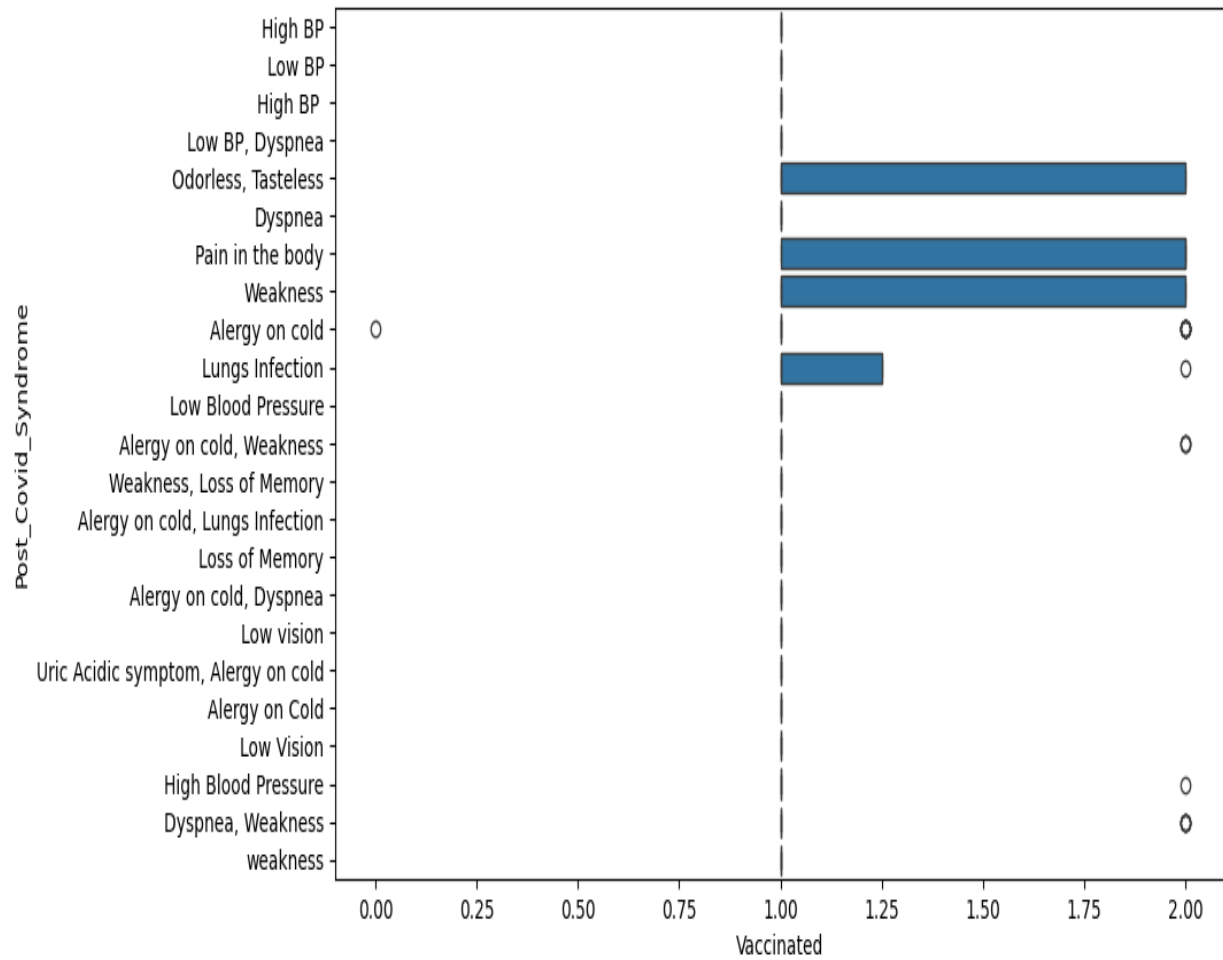


Figure 14: Vaccination wise Post Syndrome in Boxplot.

VACCINATION WISE POST COVID SYNDROME

Figure 15 and Figure 16 shows that if the patients are vaccinated, the possibility of suffering from post covid syndrome is too high.

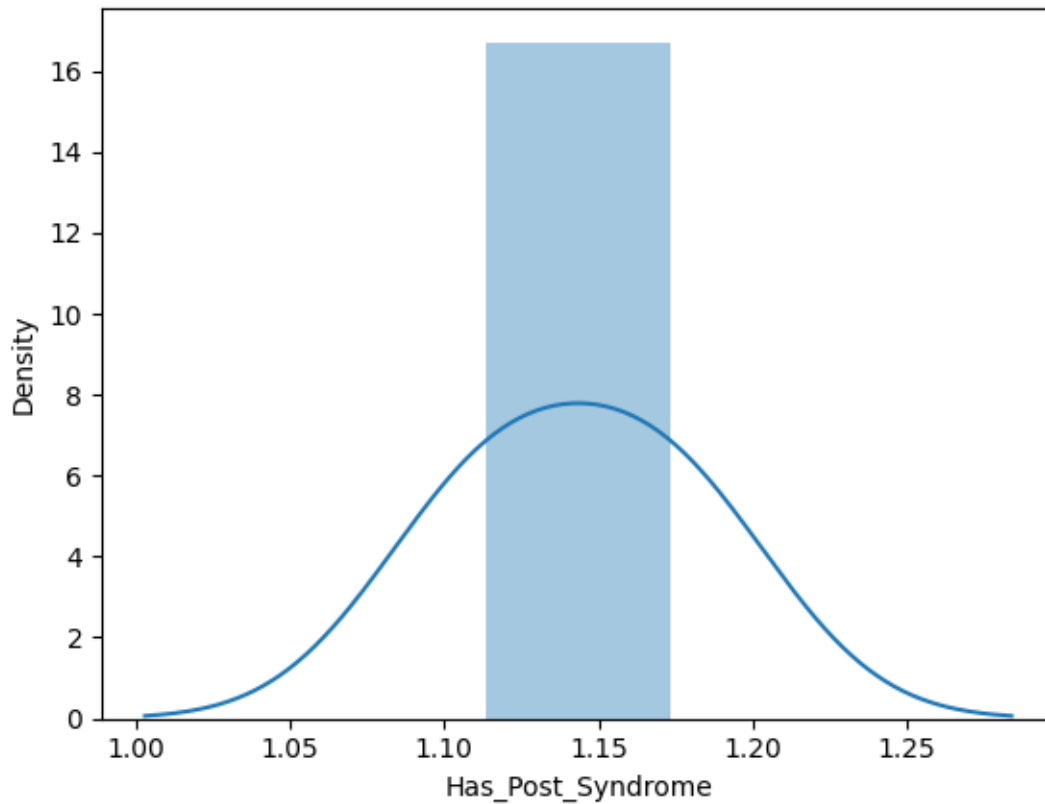


Figure 15: Vaccination wise post covid syndrome in Distribution plot.

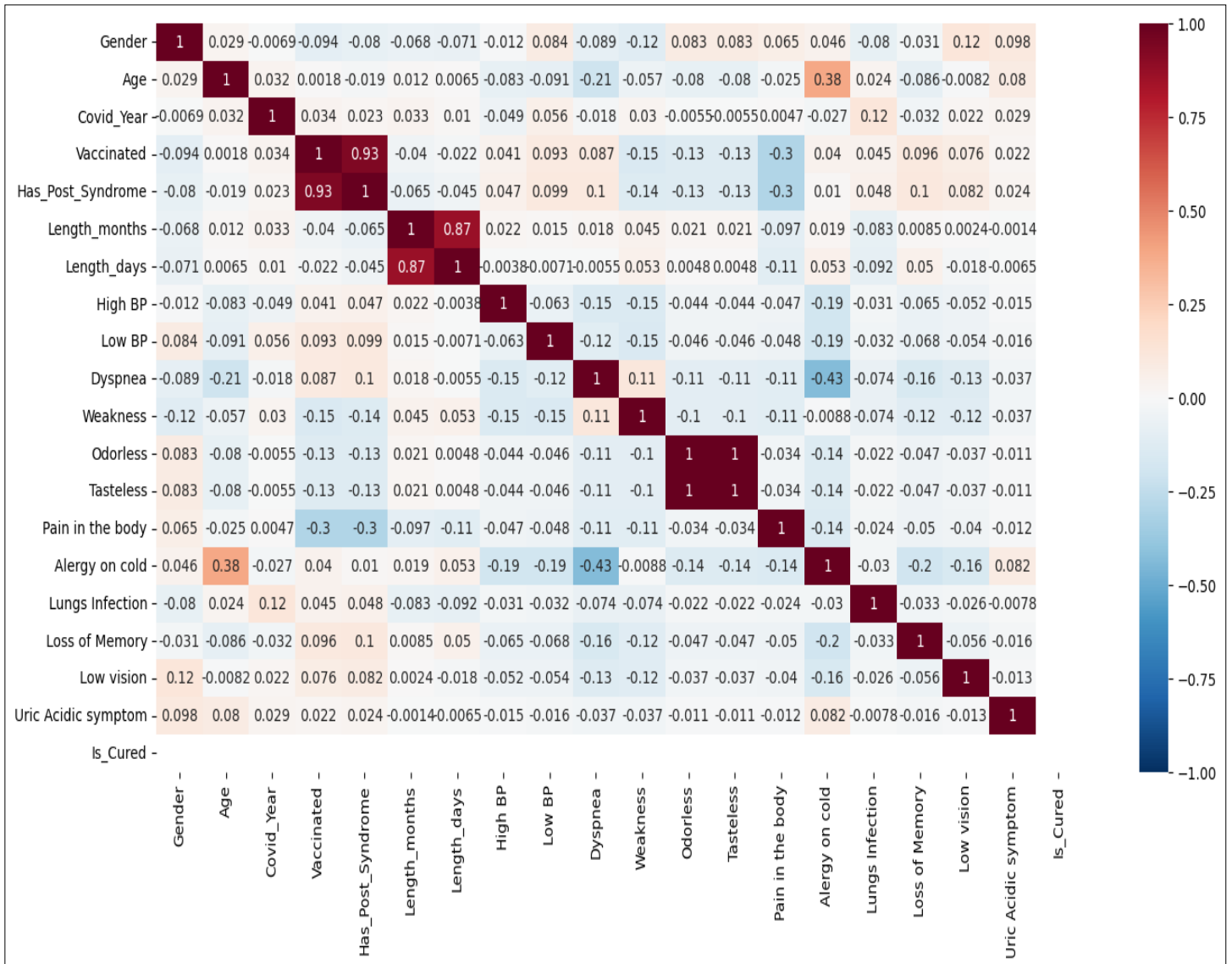


Figure 16: Vaccination wise post syndrome in Heat map.

CHAPTER 05: CONCLUSION & RECOMENDATION

5.1 FINDINGS & CONTRIBUTIONS

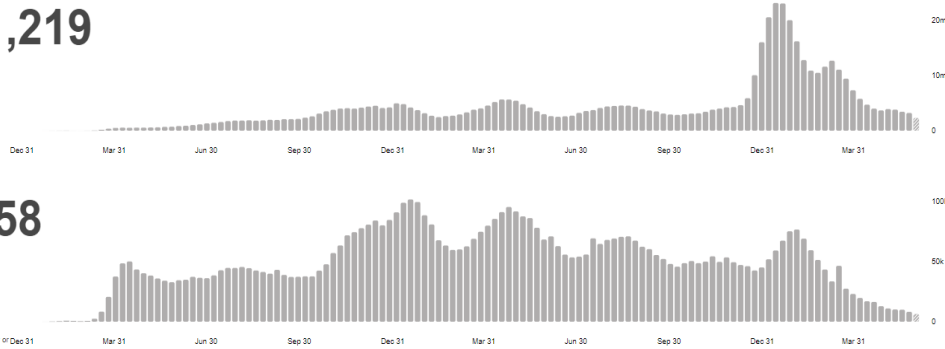
Internationally, there have been 532,201,219 confirmed cases of COVID-19, including deaths, reported to WHO as of 10 June 2022. COVID-19 is increasing day by day. Vaccine doses have been administered a total of 11,854,673,610 as of 6 June 2022. Since it is a new disease, lots of things do not clear to us till now.

Global Situation

532,201,219
confirmed cases

6,305,358
deaths

Source: World Health Organization
Data may be incomplete for the current day or Dec 31 week.



This report says, some specific post covid syndrome has been detected. Such as,

High Blood Pressure,

Low Blood Pressure, Tasteless

Short of Breath, Low Vision

Weakness, Loss of Memory

Odorless

Pain in the Body

We should be known about these post covid syndrome so that after effect can be diagnosed as early as possible.

5.2 LIMITATIONS

Collecting real time data was the main limitation we have faced. Not so much of hospitals wants to share their patients' information. Not so many people also wants to share their information. So, I have to collect data from my surroundings, acquaintances, friends and family. Since, it's a new disease, I had to prepare a questionnaire for collecting data but it needed to be verified by a Doctor. It would be better if we researchers has the facility to collect data easily so that, we can study further more.

5.3 RECOMMENDATION FOR FUTURE WORK

When vaccination will be produced for Covid-19 again, this research will help to diagnose about after covid symptoms. Therefore, we can also work as country wise post covid symptoms, district wise, and area wise. So, we can be so ensured about this post covid syndrome and will able to take action for it who suffers.

REFERENCES

- [1] <https://www.idsociety.org/covid-19-real-time-learning-network/disease-manifestations--complications/post-covid-syndrome/>
- [2] <https://covid19.who.int/>
- [3] https://www.who.int/emergencies/diseases/novel-coronavirus-2019?adgroupsurvey={adgroupsurvey}&gclid=Cj0KCQjwnNyUBhCZARIsAI9AYIHJ-MzkdsjpSkXWTIZVzNkWz9Rlv2ox1hJ4PqCxLUkXPiKZf_t8poaAteYEALw_wcB
- [4] <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0254523>
- [5] Menges, ‘Burden of Post-COVID-19 Syndrome and Implications for Health Care Service Planning: A Population-Based Cohort Study’. Plos.org July 2021
- [6] A. Pavli et al. 2021 ‘Post-COVID Syndrome: Incidence, Clinical Spectrum, and Challenges for Primary Healthcare Professionals.’
- [7] M. M. P. T. Jayasekera et al. ‘A prospective cohort study on post Covid syndrome from a tertiary care center in Sri Lanka.’ 2023
- [8] Bryan Oronsky, et al. ‘A Review of Persistent Post-COVID Syndrome (PPCS)’ springer.com. Volume 64, pages 66-74, (2023)
- [9] Daniel Ayoubkhani et al. ‘Post-covid syndrome in individuals admitted to hospital with covid-19: retrospective cohort study’ 31 Mar 2021
- [10] J. Lechner-Scott et al. ‘Long COVID or post COVID-19 syndrome’ ncbi.nlm.gov. 17 Sept. 2021.
- [11] Fahad M. Iqbal et al. ‘Characteristics and predictors of acute and chronic post-COVID syndrome: A systematic review and meta-analysis’ thelancet.com June 2021
- [12] Reaz Mahmud et al. ‘Post-COVID-19 syndrome among symptomatic COVID-19 patients: A prospective cohort study in a tertiary care center of Bangladesh’ journals.plos.org 18 Apr 2021
- [13] Mohammad Anwar Hossain et al. ‘Prevalence of Long COVID symptoms in Bangladesh: a prospective Inception Cohort Study of COVID-19 survivors’ Volume 6, Issue 12

PLAGIARISM REPORT

ORIGINALITY REPORT

11 %

SIMILARITY INDEX

10 %

INTERNET SOURCES

5 %

PUBLICATIONS

6 %

STUDENT PAPERS

PRIMARY SOURCES

1

vocal.media

Internet Source

2 %

2

Wilbert W. Ichwan, Edward A. Koesmanto, Dafa R. Syaidina, Nunung Nurul Qomariyah. "Finding the Correlation of Health Issues and Age with Stroke Using Machine Learning Algorithms", 2023 10th International Conference on ICT for Smart Society (ICISS), 2023

Publication

1 %

3

Submitted to Kaplan Professional

Student Paper

1 %

4

www.ncbi.nlm.nih.gov

Internet Source

1 %

5

www.techtarget.com

Internet Source

1 %

6

www.ibm.com

Internet Source

1 %

7

Submitted to University of Sussex

Student Paper

1 %

8	Submitted to Monash University Student Paper	1 %
9	docplayer.net Internet Source	<1 %
10	www.coursehero.com Internet Source	<1 %
11	docassas.u-paris2.fr Internet Source	<1 %
12	5dok.org Internet Source	<1 %
13	trepo.tuni.fi Internet Source	<1 %
14	www.aihw.gov.au Internet Source	<1 %
15	V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila. "Challenges in Information, Communication and Computing Technology", CRC Press, 2024 Publication	<1 %
16	orbi.uliege.be Internet Source	<1 %

ACCOUNTS CLEARANCE

Student Portal

Bibi Zoynob (213-44-223)

Payment Ledger

N.B: To enjoy scholarship and tuition fee waiver undergraduate students must take at least 12 credits and postgraduate students must take at least 9 credits in each semester.

For more details please visit DIU website or <https://daffodilvarsity.edu.bd/scholarship/diu-scholarship>.

Student Info

Name	Bibi Zoynob
ID	213-44-223
Email	joynob44-223@diu.edu.bd

Payment Ledger Summary

Total Payable	117750
Total Paid	117750
Total Due	0
Total Other	500
Waiver Semester	
Waiver Category	
Waiver Percentage	
Waiver / Scholarship calculation may vary the amounts	

Student Ledger

All