COVID-19 Vaccine Data Analysis and Prediction using Machine Learning

A Perspective in Bangladesh

BY

Sheikh Fardin Oyon
ID: 191-15-12071

Md Redwan Hussain
ID: 191-15-12109

This Report Presented in Partial Fulfillment of the Requirements for

The Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Mr. Abdus Sattar

Assistant Professor

Department of CSE

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

TUESDAY, 24th JANUARY 2023

APPROVAL

This research titled **"COVID-19 Vaccine Data Analysis and Prediction using Machine Learning, A perspective in Bangladesh"**, submitted by **Sheikh Fardin Oyon, ID: 191-15-12071 and Md Redwan Hussain, ID: 191-15-12109** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on ***24th January, 2023***.

BOARD OF EXAMINERS

28/01/23

Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

24.01.23

Internal Examiner

Dr. Mohammad Shamsul Arefin
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

24.01.23

Internal Examiner

Md. Sabab Zulfiker
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

24.1.23

External Examiner

Dr. Ahmed Wasif Reza
Associate Professor
Department of Computer Science and Engineering
East West University

DECLARATION

We hereby declare that, this thesis has been done by us under the supervision of **Abdus Sattar, Assistant Professor, and Department of CSE** Daffodil International University. We also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Abdus Sattar
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Sheikh Fardin Oyon
ID: 191-15-12071
Department of CSE
Daffodil International University

Md Redwan Hussain
ID: 191-15-12109
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final thesis successfully.

We really grateful and wish our profound our indebtedness to **Abdus Sattar**, Assistant Professor, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of **"COVID-19 Vaccine data analysis and prediction using machine learning"** to carry out this thesis. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice , reading many inferior draft and correcting them at all stage have made it possible to complete thisthesis.

We would like to express our heartiest gratitude to **Prof. Dr. Touhid Bhuiyan, Professor, and Head,** Department of CSE, for his kind help to finish our thesis and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and passion of our parents.

ABSTRACT

Coronavirus disease (COVID-19), a highly infectious and transmitted disease that was first discovered in Wuhan city in China in December 2019. Every person in the world is suffering from the coronavirus, directly or indirectly. Someone is confronted directly, when the virus attacks them and some are indirectly affected because of the closure of their businesses, work, and everyday work. Today, the global economy is also slowing down day in and day out. Prediction that is based on machine learning (ML) is regarded as a crucial method for enhancing planned decision-making. Prediction, prioritizing, and decision-making leverage modern ML models. After forecasting, several ML algorithms are employed to enhance decision-making at various levels. Predicting the geographic distribution of the active virus at the moment of vaccination is one of the key requirements for fair vaccine distribution. This study focuses on the potential prediction of the COVID-19 vaccine's efficacy, which has been described as a light in the otherwise dark. People have a variety of issues, including doubts regarding the COVID-19 vaccine's effectiveness. Based on these ideas, the COVID-19 vaccine would either have side effects that would worsen the patient's current health or reduce their likelihood of getting the condition following injection. People have openly voiced their concerns about vaccination in this regard. Some insignificant and ferocious deaths has also occurred by inquiring different types of vaccine and their doses. This article details Bangladeshi people's readiness to pay as well as their intention to obtain a COVID-19 vaccination with proper data analysis and visualization. The current study verified the safety of the licensed COVID-19 vaccines and showed that vaccinations increase people's sense of security. The majority of the bad reactions following vaccination are mild to moderate, which are indications that the body's immune system is developing defenses. Based on the input data, ML may also be used to predict the degree of side effects; anticipated severe situations may call for additional medical care or possibly hospitalization. So prevention of covid19 is a mandatory now to create a better healthy world and safe environment to all of the people.

TABLE OF CONTENTS

CONTENTS

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Introduction

COVID-19 is a novel type of coronavirus that emerged from Wuhan, China. From the first case reported in late 2019 to 12th February, of this year,103,391,516 confirmed cases and approximately 23,80,436 deaths were identified over 213 countries around the world This large number of infected patients with a global mortality rate of 4.15% demonstrates that the coronavirus disease is extremely contagious. In March 2020, World Health Organization (WHO) announced a COVID19 pandemic. Application for machine learning are frequently used for forecasting to enhance decision-making processes in a variety of fields, such as medical treatment and self-driving vehicles. In addition, machine learning algorithms are crucial to the processing of images, videos, voices, and natural language in robotics and other fields. In the beginning, the COVID-19 vaccine supply will be limited. Therefore, the basic question of what the optimized vaccine distribution might be must be answered. [1] This study sought to examine the reactions and side effects that individuals in Jordan reported after having various COVID-19 vaccinations. An online survey was created with that goal in mind, and statistical analysis was done on the collected data. Additionally, machine learning (ML) technologies were used to create prediction models that incorporate statistical methods for learning and forecasting events like the severity of post-vaccination side effects based on participant answers, demographic information, and medical data. [1] In this global health emergency, the health sector is actively looking for new technologies and strategies to track and control the spread of the coronavirus pandemic. Machine learning (ML), which can track the pace and detect the growth rate of the corona virus and identify the danger and severity of Corona virus patients, is one of the largest applications of global technology at the time. Undoubtedly, politics has played a role in people's hesitation to get the COVID-19 vaccine. In order to increase the use of life-saving vaccines, it is urgent for scientists and health professionals to better understand the causes of politicization as they relate to COVID-19 vaccines, the factors that affect people's sensitivity to scientific misinformation in politicized contexts, and how to combat the politicization of science.

1.2 Motivation

The fact that we could finally realize how all the math we had studied in college was used in real life and that it was not only interesting but also incredibly practical was what inspired us to study. However, simply the idea that people can gain something beneficial from the data is already incredibly motivating. Learning by Machines, Prediction these sectors are very helpful in these situations because they can sequence vaccines, perform speedier diagnosis, perform scanner analyses, or, less typically, manage maintenance and delivery robots. Because of this, the majority of nations are developing and seek assistance from machine learning.[2]

Working on machine learning all across the world to help with this terrible problem Machine learning is a very useful field in this pandemic. To stay up with the times, we must battle the coronavirus. We must underline that machine learning is an excellent tool for us to use in this case. It's essential to comprehend the fundamental reasons why people choose to be vaccinated in order to devise winning methods. In individualistic society, vaccination requirements are unlikely to be a workable option. If there is a high rate of COVID-19 vaccine resistance, it is crucial to pinpoint the underlying causes of the resistance and devise strategies for dealing with them.

As a result, many weeks after making immunization available to all employees, the goal of the current study was to map the demand for vaccination against COVID-19 across all employees of a tertiary care hospital with respect to their fundamental features. The study also sought to examine the drivers of acceptance among the immunized and the causes of resistance within the workforce that is still anti-vaccine. We propose a machine learning model that can be continuously run for accurate spread prediction and proactive development of strategic response by the government and citizens.

By doing so, we will be able to forecast the number of COVID-19 cases and the potential end dates of the pandemic in various nations, as well as learn more about how COVID-19 is affecting the global population.

1.3 Problem Definition

The term "machine learning" is essential in the current ICT era. Our technological sector will progress with the application of ML. To propose an effective solution, it is crucial to recognize the problems and pressing demands in this industry. Understanding the needs of the software business, governmental regulations, and instructional strategies is also necessary to implement ML in the technical sector. Gather information on the problems with vaccines and machine learning that hinder the diagnosis of the corona virus by conducting a quick survey of developers and people. Vaccines differ from typical prescription drugs in two important ways.

The first is that they are designed to prevent disease rather than treat it. They achieve this by training the patient's immune system to identify a specific pathogen, such as a virus or bacteria that causes an illness. Vaccinations can be very helpful in avoiding sickness rather than treating it once it has already started since this "memory" might persist for years or even longer in some situations. The second is that vaccines are biological products by definition, as opposed to chemical ones, like the bulk of drugs. Because of this, they are frequently less stable than chemicals and more susceptible to temperature changes, which not only makes the production processes more difficult and costly. A moral trade-off must be made between delivering a vaccine sooner and preventing more COVID-19 deaths at the expense of potential side effects or ineffectiveness, and waiting for longer and more thorough testing to ensure safety and efficacy at the expense of more COVID-19 deaths. There is no magic solution, but there will be significant pressure to launch a vaccine sooner given the devastating economic, social, and health effects of different anti-COVID-19 tactics like lockdown. The fragility of an individual makes immunization difficult. [3]

Providing vaccinations becomes significantly more difficult in regions where people are displaced because of conflict, calamities, or humanitarian crises. This is particularly concerning in light of the fact that illness frequency might increase with frailty. That’s why the problem comes that to fight the COVID every variant vaccination is mandatory and to proof its importance a perfect data visualization is also mandatory. This will help people get encouraged in vaccination.

1.4 Research Objectives

Vaccines are used to protect against serious infection-related illness, avoid hospital, and avert death. When a pandemic strikes, the objective is to protect those who are more vulnerable, save as many lives as possible, and guarantee the best possible outcome for society by limiting social disturbance. Not only that, vaccines are globally mandatory for every person for prevention of different deadliest viruses such as Ebola, Flu etc. There are some benefits of using ML in vaccine prediction. There are some technical and theory objectives of using ML.[3]

Some of the technical objectives are given below:

- Create a reliable prediction model to identify the vaccination process.
- To motivate software developers to use the model while working with ML.
- Include the model in webpages and applications for mobile devices.

Some of the theory objectives are given below:

- Help the student to know about this prediction.
- Make the student to motive this framework.
- To analyze vaccination process in Bangladesh using current authentic real life reports collected from many hospitals in Dhaka and around Bangladesh.
- To clarify and predict the total number of vaccinated people from different companies vaccine as like Pfizer, Moderna, Sinopharm etc.
- Perfect death record model after vaccination
- To showcase vaccination outcomes and further possibility of vaccines with an attractive graphic model using the concept of machine learning and data mining methods.
- Showcasing the aftermath of vaccination process.
- Analyze a prediction model of a specific significant vaccine of a vaccine company

1.5 Research Questions

Following are the key inquiries on which this thesis is focused:

- What is the current situation of Bangladesh's population and how is ML used in the vaccine industry?
- What are the restrictions on using ML to discover corona virus in the vaccine industry?
- How might the challenges associated with corona virus detection be overcome?
- What are the benefits of getting vaccinated?
- How do COVID-19 vaccines work?
- How long does protection from COVID-19 vaccines last?
- Can COVID-19 vaccines affect fertility?
- How were COVID-19 vaccines developed so quickly?

1.6 Research Layout

Chapter 1: Will discuss about introduction, motivation, Problem Definition, Research Question, Research Methodology and the expected outcome of our project.

Chapter 2: Will discuss about background of this research and the related work and current status based on Bangladesh perspective and government goals and regulations.
Chapter 3: Will describe dataset describe, pre-processing, architecture of the model and some machine learning models.

Chapter 4: Will discuss about result comparison and analysis.

Chapter 5: It is focus to the research outcomes such as research limitations, conclusion and future work.

At last we discuss about all research paper which is help our research work.

CHAPTER 2

BACKGROUND

2.1 Related Works

Bangladesh has a solid history of implementing large vaccination campaigns, particularly for illnesses like measles and rubella. But there have been particular difficulties in distributing COVID-19 dosages to communities all around the country. COVID-19 has flaked an inclusive terror which causes illness ranging in intensity. The main problem the world is facing in controlling the COVID 19 cases is, lack of awareness about the virus and vaccine doses. The objective of this paper is to increase the proper accuracy by using other prediction models and by constructing a data-driven argument support system using Data Mining and machine learning models. We also want to know which features are relevant in predicting vaccine functionality. The structure of the paper is divided into three steps. Firstly, delivering an overview of the dataset and how the dataset is modified for better proficiency. Secondly presenting the experiments performed to evaluate the patterns and algorithms to determine the vaccine functionality. Finally, the paper is concluded with a proper discussion and explanation about the results.

A quite small amount of research has been conducted on Bangladeshi vaccination process but globally In the paper "predicting vaccination hesitancy of COVID-19 patients using logistic regression"[1] Damian et al, has used several prediction models to identify the core base of Pfizer vaccine stem in people with an accuracy of 80%. Also Li R, Liu J. [2] Presented in their paper "Effectiveness of a single dose and two doses of SARS-CoV-2 vaccines. Prof. J Bing Han [3] Presented in their paper "expectations to the vaccines! Are they really boosting immune system of human body?". Lastly Lyn Ellet use a prediction model which will target and predict vaccination "Taking a machine learning approach to optimize prediction of vaccine hesitancy in high income countries". Muhammad Lutfar Rahman showed the acceptance in vaccine sector in Bangladesh in "Acceptance of COVID-19 vaccine and its determinants: evidence from a large sample study in Bangladesh". Joe Han use an analysis on "Vaccine booster dose a better implementation to prevent COVID19" to showcase the booster dose efficiency.

2.2 Bangladesh Perspective

Bangladesh began the administration of COVID-19 vaccines on 27 January 2021, focusing initially on a pilot program of 500 health workers, while mass vaccination started on 7 February 2021. But Vaccines do not give you the antibody of the virus, rather it teaches your immune system to recognize and fight the infection. Those impacted most severely by the pandemic's effects are women. Women are particularly vulnerable to economic shocks in general since they often earn less, have fewer savings, and have less secure positions to begin with. The pandemic has ravaged historically female-dominated sectors including hospitality, tourism, and retail, depriving many women of their means of subsistence. Women have been more likely than males to leave the workforce during the epidemic in every region. With little to no worker protections like paid sick leave and unemployment, the majority of working women (58%) are engaged in the informal economy. The COVID-19 vaccine presents the body with instructions to build with immunity and does not alter human cells. Bangladesh is getting worse COVID19 cases like other countries such as India facing a mass Delta, Delta Plus variant which cause 266,310 deaths in just 5 months. Bangladesh also faced almost 14,000 deaths between this 5 months and the rate was above 65%. To prevent this high infection rate the government of the people's republic of Bangladesh pilot vaccine campaigns all over the realm and start a vast vaccination process. The following figure shows District wise 1st, 2nd, 3rd & 4th Doses Administered.

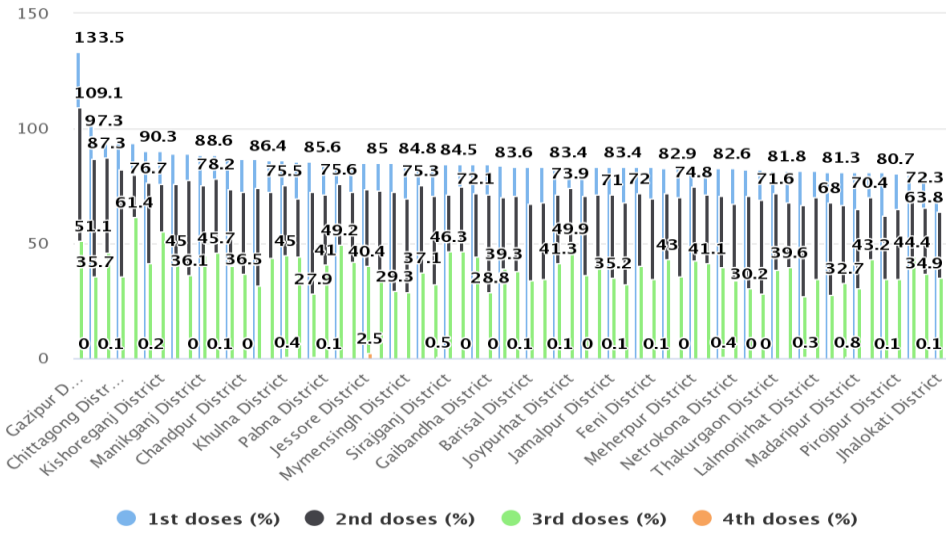


Fig 2.2.1: District wise 1st, 2nd, 3rd & 4th Doses Administered

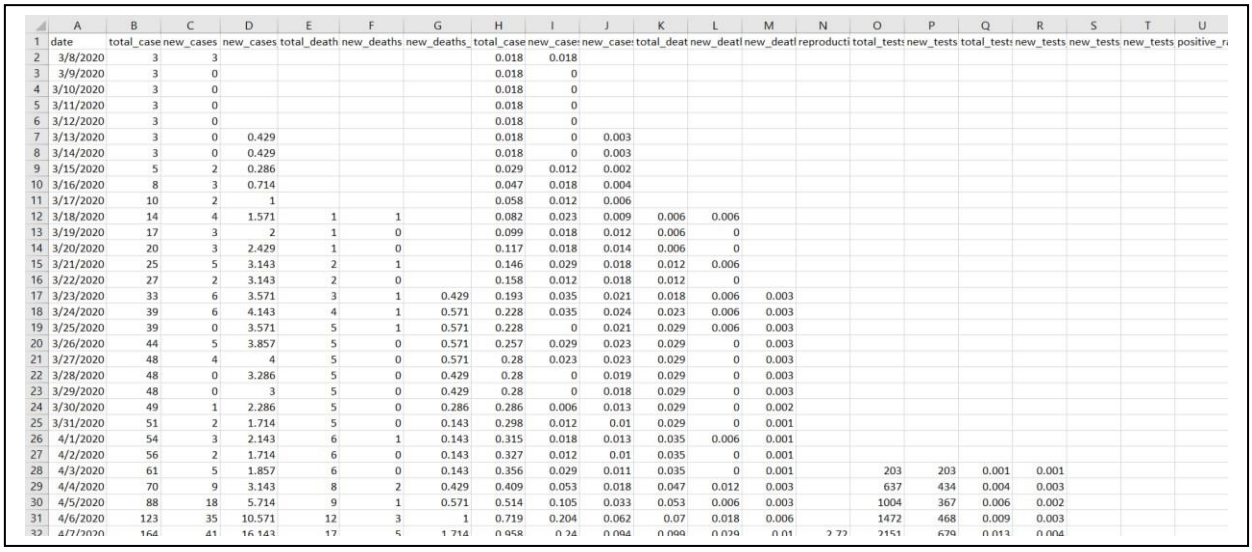
CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

This research focuses on the extraction of significant and usable information from sizable data collections that cannot be found in any other way (using conventional querying methods). The resulting tools enable the discovery of intriguing patterns that are deeply buried in the data. The methodology provides a framework that aim to review and refine the forecasting model such as Business Understanding, Data Understanding, Data Preparation and Modelling. Research methodology is divided into Types of Data and Data visualization part. The first segment is discussion about the data. How to collect data's and their types. The Data visualization the construction of prediction models involves the application of supervised machine learning methods. By establishing a connection between the input characteristics and the appropriate outputs, these models are trained. Data has been collected from many places. But some newspapers and websites have done a lot of help for data collection. The model will forecast the result based on the dependencies it was trained with when given a new input feature. Numerous supervised learning techniques are available that may be applied to prediction. Machine learning applications need fitting the data, selecting the best algorithms to fit the data, and providing solutions for the algorithms in order to solve a particular problem. It won't provide the right answer to the issue if the algorithms were chosen incorrectly or if they weren't appropriate for the solution. Algorithms that will be used in the job must be chosen. This research may work more quickly while using algorithms to analyze the data. As a result, this will create more time to focus on other initiatives. As a result, data mining techniques is employed to address these issues. All of the data on these vaccine analysis may be discovered using mining techniques. The mapping of input and output information is the main application of applied algorithms. As these algorithms are most suited for categorical outcomes, decision trees, random forests, and support vector machines are implemented in this study.[1]

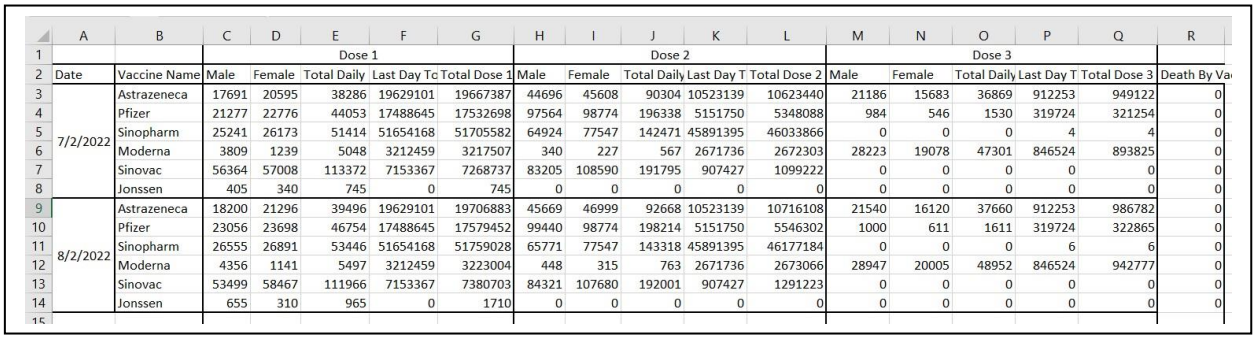
3.2 Dataset Description

Fetching two datasets and the dataset collected by the Directorate General of Health Services (DGHS) and Bangladesh COVID19 dashboard collects daily basis with total vaccination of each vaccine company such as AstraZeneca, Pfizer, Moderna, Sinopharm, Sinovac, Janssen, Comirnaty and percentage data from February-2021 to February-2022. Dataset tracks both first, second and third doses of these vaccines and average percentage. The dataset containing large vaccine dose information with every vaccine's 1-3 doses and also tracks any death records. The number of vaccinations is rising daily. The vaccination doses, the number of vaccine-related deaths, and the total number of COVID vaccine cases are the data categories. The other includes all good instances, all cases of recovery, The number of fatalities per day, the number of positive cases per day, the number of test cases per day, and the daily recovery rate are all statistics. Daily data is chosen based on the date. Daily data is being selected according to date.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	date	total_case	new_cases	new_cases	total_death	new_deaths	new_deaths_	total_case	new_case	new_case	total_deat	new_deatl	new_deatl	reproducti	total_tests	new_tests	total_tests	new_tests	new_tests	new_tests	positive_r
2	3/8/2020	3	3					0.018	0.018												
3	3/9/2020	3	0					0.018	0												
4	3/10/2020	3	0					0.018	0												
5	3/11/2020	3	0					0.018	0												
6	3/12/2020	3	0					0.018	0												
7	3/13/2020	3	0	0.429				0.018	0	0.003											
8	3/14/2020	3	0	0.429				0.018	0	0.003											
9	3/15/2020	5	2	0.286				0.029	0.012	0.002											
10	3/16/2020	8	3	0.714				0.047	0.018	0.004											
11	3/17/2020	10	2	1				0.058	0.012	0.006											
12	3/18/2020	14	4	1.571	1	1		0.082	0.023	0.009	0.006	0.006									
13	3/19/2020	17	3	2	1	0		0.099	0.018	0.012	0.006	0									
14	3/20/2020	20	3	2.429	1	0		0.117	0.018	0.014	0.006	0									
15	3/21/2020	25	5	3.143	2	1		0.146	0.029	0.018	0.012	0.006									
16	3/22/2020	27	2	3.143	2	0		0.158	0.012	0.018	0.012	0									
17	3/23/2020	33	6	3.571	3	1	0.429	0.193	0.035	0.021	0.018	0.006	0.003								
18	3/24/2020	39	6	4.143	4	1	0.571	0.228	0.035	0.024	0.023	0.006	0.003								
19	3/25/2020	39	0	3.571	5	1	0.571	0.228	0	0.021	0.029	0.006	0.003								
20	3/26/2020	44	5	3.857	5	0	0.571	0.257	0.029	0.023	0.029	0	0.003								
21	3/27/2020	48	4	4	5	0	0.571	0.28	0.023	0.023	0.029	0	0.003								
22	3/28/2020	48	0	3.286	5	0	0.429	0.28	0	0.019	0.029	0	0.003								
23	3/29/2020	48	0	3	5	0	0.429	0.28	0	0.018	0.029	0	0.003								
24	3/30/2020	49	1	2.286	5	0	0.286	0.286	0.006	0.013	0.029	0	0.002								
25	3/31/2020	51	2	1.714	5	0	0.143	0.298	0.012	0.01	0.029	0	0.001								
26	4/1/2020	54	3	2.143	6	1	0.143	0.315	0.018	0.013	0.035	0.006	0.001								
27	4/2/2020	56	2	1.714	6	0	0.143	0.327	0.012	0.01	0.035	0	0.001								
28	4/3/2020	61	5	1.857	6	0	0.143	0.356	0.029	0.011	0.035	0	0.001		203	203	0.001	0.001			
29	4/4/2020	70	9	3.143	8	2	0.429	0.409	0.053	0.018	0.047	0.012	0.003		637	434	0.004	0.003			
30	4/5/2020	88	18	5.714	9	1	0.571	0.514	0.105	0.033	0.053	0.006	0.003		1004	367	0.006	0.002			
31	4/6/2020	123	35	10.571	12	3	1	0.719	0.204	0.062	0.07	0.018	0.006		1472	468	0.009	0.003			
32	4/7/2020	164	41	16.143	17	5	1.714	0.958	0.24	0.094	0.099	0.029	0.01	2.72	2151	679	0.013	0.004			

Fig 3.2.2: Test Dataset



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1			Dose 1					Dose 2					Dose 3					
2	Date	Vaccine Name	Male	Female	Total Daily	Last Day Tc	Total Dose 1	Male	Female	Total Daily	Last Day T	Total Dose 2	Male	Female	Total Daily	Last Day T	Total Dose 3	Death By Va
3	7/2/2022	Astrazeneca	17691	20595	38286	19629101	19667387	44696	45608	90304	10523139	10623440	21186	15683	36869	912253	949122	0
4		Pfizer	21277	22776	44053	17488645	17532698	97564	98774	196338	5151750	5348088	984	546	1530	319724	321254	0
5		Sinopharm	25241	26173	51414	51654168	51705582	64924	77547	142471	45891395	46033866	0	0	0	4	4	0
6		Moderna	3809	1239	5048	3212459	3217507	340	227	567	2671736	2672303	28223	19078	47301	846524	893825	0
7		Sinovac	56364	57008	113372	7153367	7268737	83205	108590	191795	907427	1099222	0	0	0	0	0	0
8		Jonssen	405	340	745	0	745	0	0	0	0	0	0	0	0	0	0	0
9	8/2/2022	Astrazeneca	18200	21296	39496	19629101	19706883	45669	46999	92668	10523139	10716108	21540	16120	37660	912253	986782	0
10		Pfizer	23056	23698	46754	17488645	17579452	99440	98774	198214	5151750	5546302	1000	611	1611	319724	322865	0
11		Sinopharm	26555	26891	53446	51654168	51759028	65771	77547	143318	45891395	46177184	0	0	0	6	6	0
12		Moderna	4356	1141	5497	3212459	3223004	448	315	763	2671736	2673066	28947	20005	48952	846524	942777	0
13		Sinovac	53499	58467	111966	7153367	7380703	84321	107680	192001	907427	1291223	0	0	0	0	0	0
14		Jonssen	655	310	965	0	1710	0	0	0	0	0	0	0	0	0	0	0

Fig 3.2.1: Real Vaccine Dataset

3.3 Data Pre-Processing

The first step of implementation was checking for the missing values in the dataset and then process the values. A certain amount of missing values was noted such as specific COVID vaccine was not available at that time also some stored vaccine got finished and new slot of the vaccine was not available. A minor amount of people's death was registered. Processing the missing values will have an effect on the prediction of vaccine. The algorithms used here will fetch only the big data segments and find the exact result values of vaccine doses to predict the accuracy of vaccine companies and any other effects of it. Data is that the most significant for analysis work. This research work is predicated on data. Vaccination is increasing day by day. Vaccine forecasting a method of using applied machine learning is one that we often employ for a particular issue. The applied machine learning algorithm and series issue solution make use of the correlation between the dataset and algorithms.

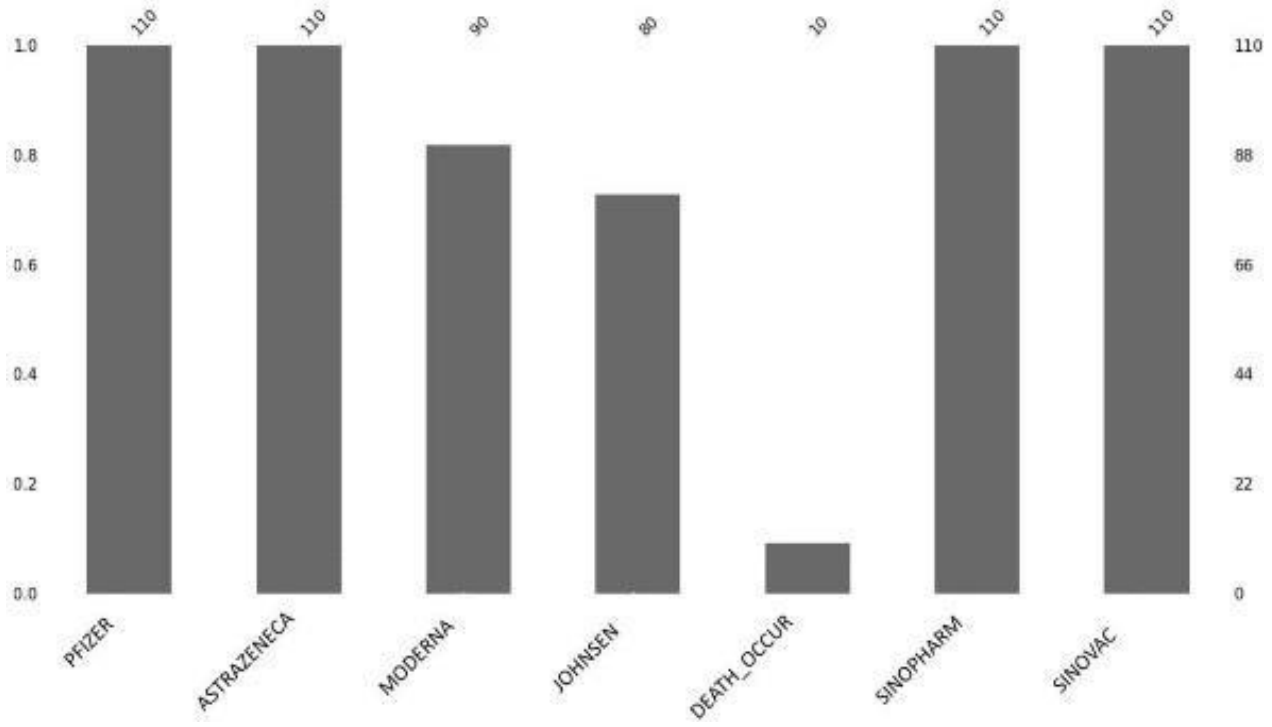


Fig 3.3.1: Processed Data

3.4 Architecture of the Model and Questions

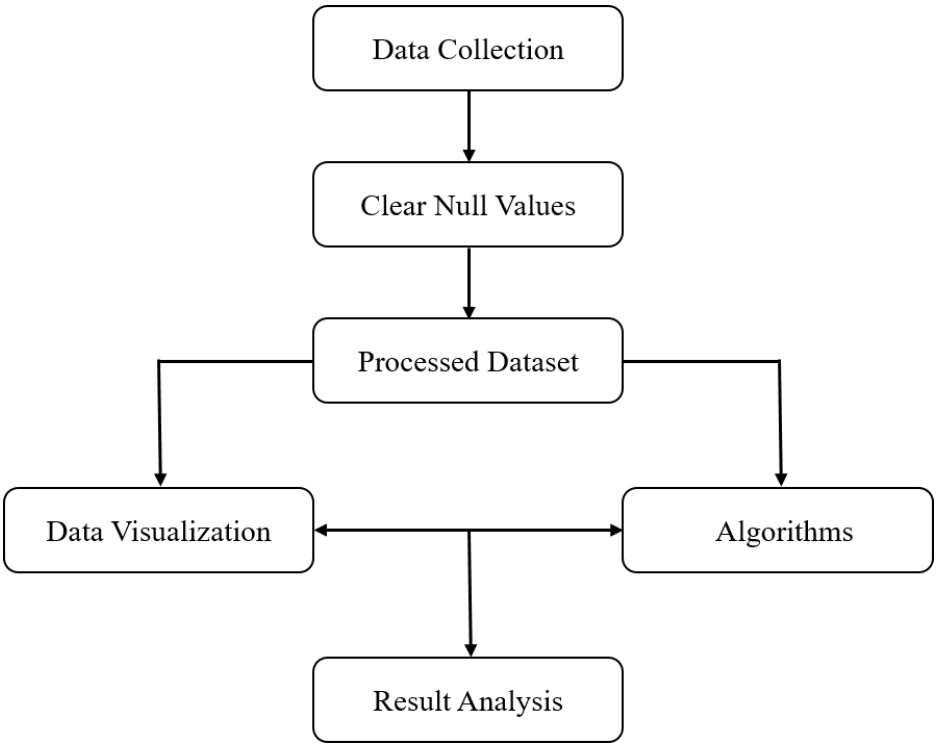


Fig. 3.4.1: Process of Data Collection

The first research question requires tracking down the given vaccines of different companies and their further results. These vaccine internal formula contains polyethylene glycol, ditetradecylacetamide, distearoyl-sn-glycero, phosphocholine cholesterol and dibasic sodium phosphate dehydrate. These are the core formula and common in every vaccine which creates antibody against COVID. The male and female vaccination rate and finding the significant accuracy of the prediction model. Also track any death occurred by any vaccine. This shows the ratio of first and second dose of each vaccine both in male and female and any side effects of vaccine. The training data and test data are already given so the splitting of training data can be skipped. The performance of the evaluation is measured using accuracy of different algorithms. Accuracy declares to the final to predict the specific vaccine and the benefit of the vaccine. Lastly when someone thinks that he has been vaccinated properly, his safety purposes decreases to a lower level which can cost him a serious consequences by again infected by COVID. The risk level still continues.

The following two figures portraits the vaccine 1st and 2nd doses of vaccine where the first dose has a medium rage and the second dose is in a higher rage. The average ratio of vaccine has increased a higher level and more people are taking vaccine daily because of the functionality of these vaccines are helping people to prevent COVID infections.

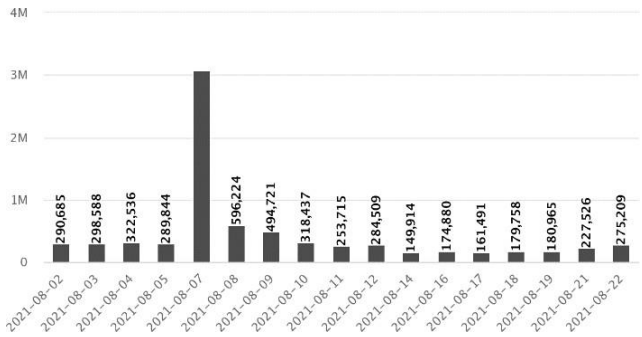


Fig 3.4.2: 1st dose

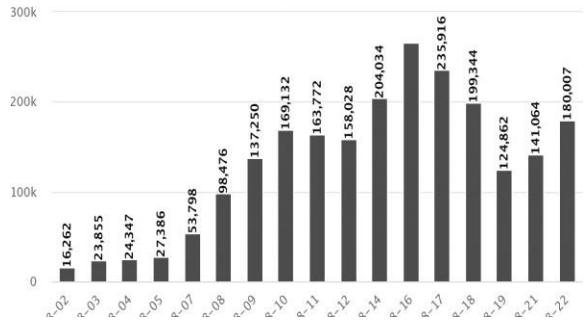


Fig 3.4.3: 2nd dose

The second research question is to determine specific vaccine functionally that which vaccine is the most effective and adjusted with the human immune system and create a COVID antibody. Also is vaccine has occur any death regarding vaccination process or the best non side effect vaccine dose. A frequent case registered as a proper vaccinated person is again infected by COVID. So, to predict the functionality of vaccines. As the vaccines has almost 99% of accuracy fighting against covid-19 severe situations. Some cases is also registered that a vaccinated person died after proper vaccination. So a slight chances of lack accuracy can occur. By using different algorithms to prevent this issues and visualize a perfect model of vaccination.

3.5 Machine Learning Models

3.5.1 Decision Tree

A prediction model for categorical variables is the decision tree. Until the leaf node has homogeneous data, the algorithm divides the data from the root node into subgroups. Based on criteria, the nodes are divided.[1] Quinlan proposes a related splitting criterion, called Gain Ratio or Uncertainty Coefficient. This serves to normalize information gain on an attribute Xi relative how much entropy this attribute has. For Gini value:

Gini(P) = $\sum_{i=1}^{n} p_i(1-p_i) = 1 - \sum_{i=1}^{n}(p_i)^2$ where, p = $(p_1, p_2, \ldots)$

3.5.2 Random Forest

A number of decision trees make up a random forest algorithm, and all of the decision trees' votes are used to forecast the outcome. Because the outcome relies on the amount of votes from each decision tree, random forest avoids overfitting. By adjusting the value of the option 'n' which is set to 100, the number of trees created may be customized.[1] The mathematical formulation:

MSE $= \frac{1}{N}\sum_{i=1}^{1}(fi - yi)^2$ where, N is the number of data points

3.5.3 Support Vector Machine

A machine learning model called the support vector machine algorithm (SVM) can be utilized for both classification and regression. It can be applied to variables that are categorical or continuous. A hyperplane that divides the two classes is made via SVM. [1] The formula is:

Value = $max \sum_{i=1}^{m} a_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j a_k$

The linear karnel will be $K(x_i, x_j) = x_i * x_j$

More accurately, should be KKT (Karush-Kuhn-Tucker) multipliers because we are dealing with inequality constraints here. But we will use the term Varangian multipliers for continuity.

CHAPTER 4

RESULT COMPARISON AND ANALYSIS

The accuracy of most efficient vaccine are as follows and 3 algorithms are applied in dataset:

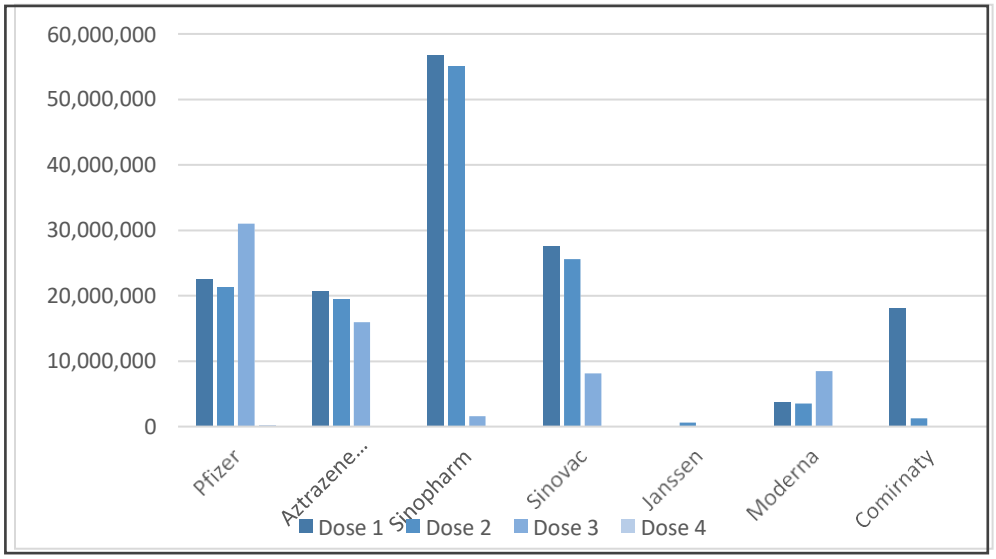
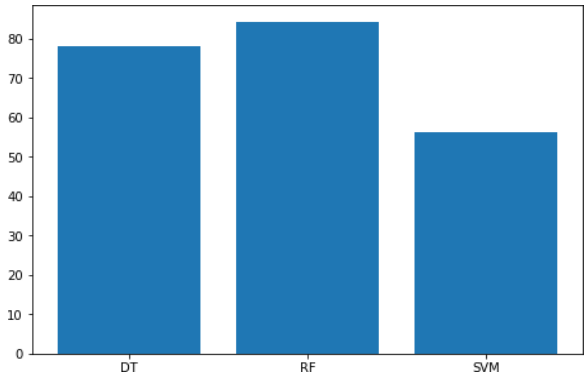


Fig 4.1.1: Total Vaccine Data

The following figure (4.1.1) describes the total numbers of vaccine given in Bangladesh and each vaccine has 4 doses. From the dataset sinopharm has the highest 1st and 2nd dose but the 3rd dose has a medium range. Secondly Pfizer has the most vaccine rage in Bangladesh and the highest 3rd dose range. Janssen has a minor vaccine range here with lowest registration.

Accuracy in seven different bio markers algorithms are:



Accuracy DT: 78.07%

Accuracy RF: 84.4%

Accuracy SVM: 63.01

Fig 4.1.2: Accuracy using 3 different algorithms

Figure (4.1.2) denotes the three algorithms DT, RF, SVM and the result of three biomarkers from the vaccine dataset. Here we can see the highest accuracy is in Random Forest algorithm with 84.8% of accuracy rate and SVM shows a range of 63.01% accuracy. A large portion (84.4%) of people who had positive vaccine intentions believed that vaccination should be made mandatory for everyone and has taken PFIZER as the 3rd booster dose to prevent covid-19.

The Total vaccine administered over time in a range chart:

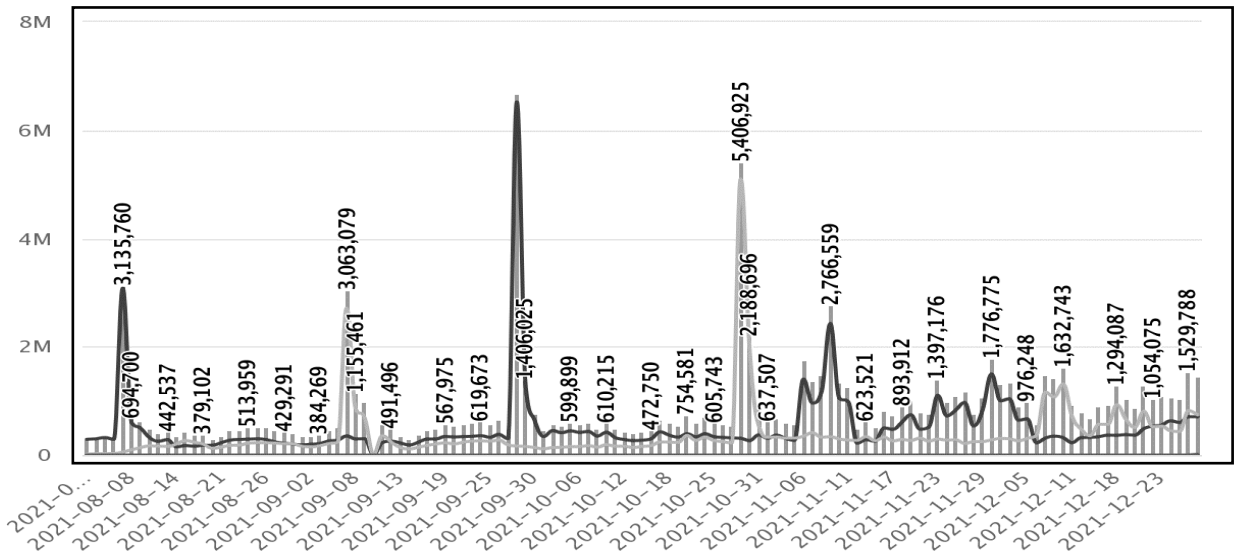
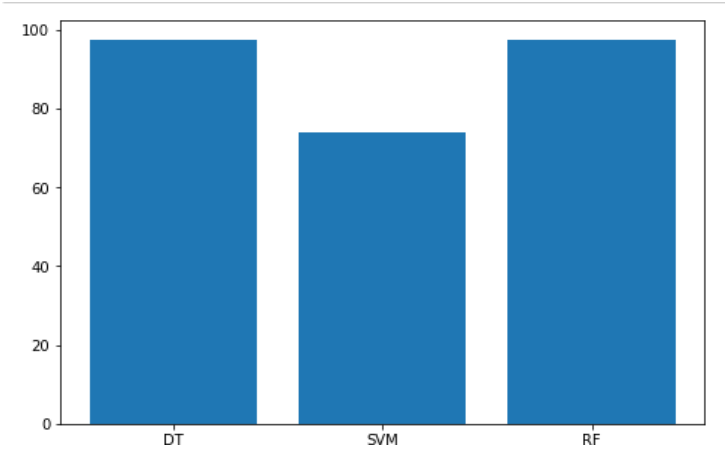


Fig 4.1.3: Vaccine Administered

In figure (4.1.3) the administered vaccine information from August-21 to December-21 a four months of vaccine first and second dose data analysis where each vaccine administered with its availability and higher limit range of vaccine doses over time. The second figure (4.1.4) is used to show the accuracy of the provided data analysis where three algorithms are used to predict a secure accuracy in vaccine dose scaling. The accuracy of the algorithm chart has higher prophecy and has almost 92.63% accuracy in Random Forest. The SVM accuracy is also 74.1% a higher accuracy result. Now the result accuracy provides a higher rage PFIZER efficiency in vaccine doses.

Accuracy in 4 different bio markers algorithms are:



Accuracy DT: 92.53%

Accuracy RF: 92.63%

Accuracy SVM: 74.1%

Fig 4.1.4: Vaccine Administered algorithm accuracy

Death Record over any vaccine doses with Accuracy in 4 different bio markers algorithms:

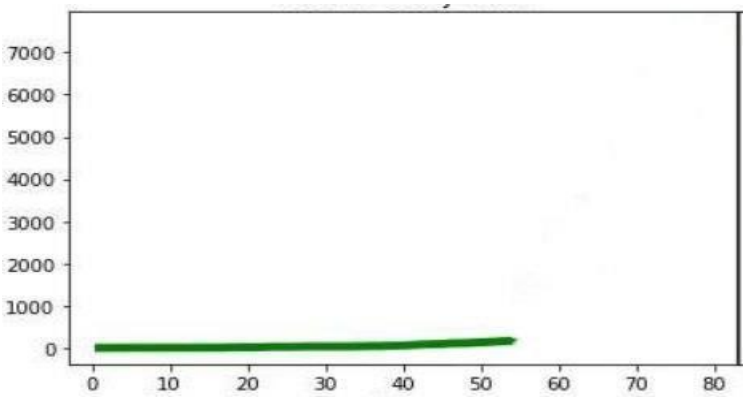
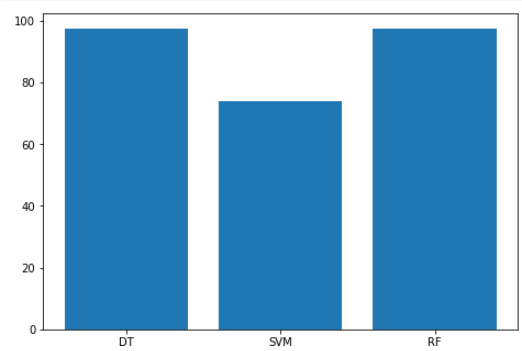


Fig 4.1.5: Vaccine Death Cases



Accuracy DT: 89.11%

Accuracy RF: 94.13%

Accuracy SVM: 72.1%

Fig 4.1.6: Death Accuracy in 3 algorithms

As many people has taken vaccine and different companies vaccine, many recent cases of side effects such as fever,cough,coldness,inside neck pain has recorded. Not only that a minor amount of people got affected by COVID in proper vaccine time and also died. But the death record over the time was a minor value. From the algoritms used 94.13% of accuracy has just faced minor symtoms are get well very soon and the death minor death rate accuracy in SVM was 72.1%.

New Vaccine dose rage and accuracy in algorithms :

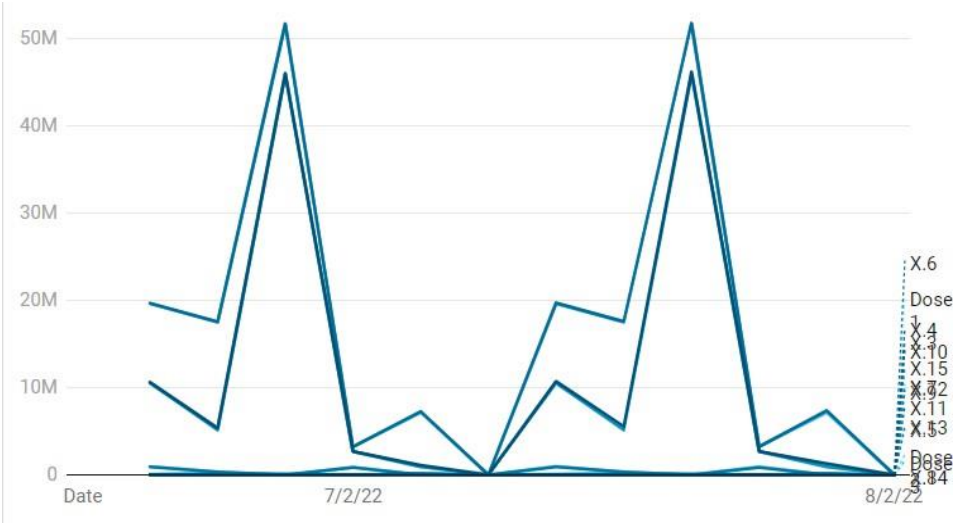
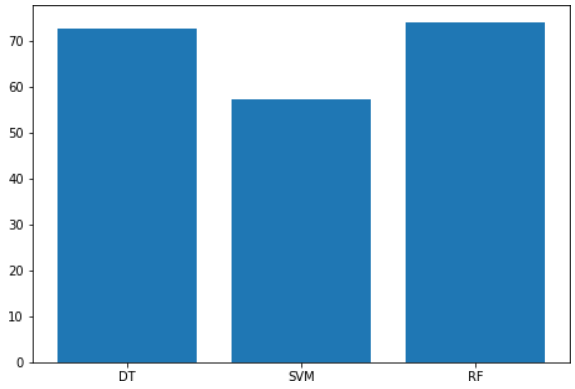


Fig 4.1.7: New vaccine dose rage

Almost administering vaccine for 2 years a broad people is under the vaccine process. Figure (4.1.7) is the data analysis shows just a rage of vaccine in just 1 months. From1-3 all doses analysis provides a higher vaccine rage of every vaccine.



Accuracy DT: 72.6%

Accuracy RF: 74.2%

Accuracy SVM: 57.4%

Fig 4.1.8: Accuracy of rage in vaccine

In the figure 4.1.8 the algorithms DT, SVM, RF shows the accuracy of vaccine rage almost 74.6% in Random forest and in SVM the accuracy rate is 57.4%. This describes that the people are more interested getting the vaccine and also the minor side effects of vaccine accuracy is helping the prevention of COVID. From now on 119,221,953 people is administered under vaccine process which is almost the 70% of the total population and the male and female ration is 50/50%. The vaccine booster dose and the 4th booster dost has also ranging day by day.

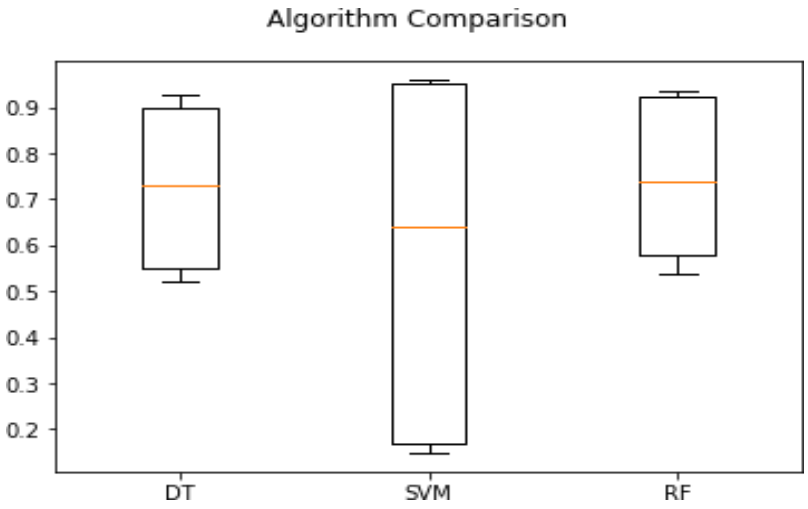


Fig 4.1.9: Comparison of 3 algorithms

Figure (4.1.9) show the difference between three algorithms that we applied in the research. The first dose against the population is 87.85% and second dose against the population is almost 74.67%. The booster dose has 51.65% of rate and finally the 4th booster has 10.1% of the population.

CHAPTER 5

Limitation, Conclusion & Future Work

5.1 Research Limitations

The worldwide epidemic fear is COVID-19. We ought to put in the effort from where we are. We use extreme caution and endeavor to lessen its effects in all areas. The entire industry will benefit the most from AI in this epidemic condition. Research will go on as long as this virus remains in Bangladesh. Future projects could use this one as a better example. This pandemic may be used to implement this concept in several nations. Future applications of the model and analytical process are possible. Alternately, the algorithm that isn't delivering accurate forecasts has to be improved in order to achieve this goal. Algorithms allow for the creation of more models. [1]

From the result obtained in chapter 5, all the models performed very well in predicting the vaccine administered of COVID 19 cases. With respect to the first research question, the accuracies from the models are 78.07% for the decision tree, 84.4% for the random forest, and 63.01% for the support vector machine. Here random forest performed very well than other models. Then from the research, the three selected biomarkers were used on the same models and the accuracies were 94.13% for the Decision tree, 92.63% for the random forest, and 74.1% for SVM which has increased the performance of prediction compared to using three biomarkers.

There are several advantages to our study. As far as we are aware, this study is unique since it is the first to specifically assess the COVID-19 vaccine coverage rate among older persons in Bangladesh and around the globe. Second, the research location and demographic were unique since Bangladesh has the lowest percentage of COVID-19 immunization coverage in South Asia and elderly people are one of the populations who receive the most attention internationally. Despite these advantages, a number of limitations should be thought about when evaluating the results of the current study. First, because our study was cross-sectional in nature, causation could not be shown.

5.2 Conclusion

The worldwide epidemic fear is COVID-19. We ought to put in the effort from where we are. We use extreme caution and endeavor to lessen its effects in all areas. The entire industry will benefit the most from AI in this epidemic condition. Research will go on as long as this virus remains in Bangladesh. Future projects could use this one as a better example. This pandemic may be used to implement this concept in several nations. The objectives of the research have, in our opinion, been met. Different algorithms can be used to predict vaccine administered more accurately. Various data mining techniques were used to determine the accuracy models.

The accuracy of the random forest approach was 92% for the dataset that was provided. The great accuracy may have been achieved by substituting the median value of the corresponding characteristics for the missing data. If there were less or no missing data, the research may be enhanced. The model and analysis procedure can be applied in the future. Furthermore, elderly persons in Bangladesh have a greater level of Vaccination reluctance due to widespread misunderstandings about COVID-19. Therefore, Policymakers and public health professionals should think about distributing age-appropriate COVID-19 vaccination-related communications. This is why the policy level should adopt a Proactive strategy that emphasizes the importance of herd immunity. In addition, our suggestion is to employ the mass media to address the concerns about the vaccine's efficacy and accurate public health information.[4] The current study found that the authorized COVID-19 vaccines are safe and that becoming vaccinated makes people feel more secure, despite significant Differences in the existence and severity of adverse effects between these immunizations. The majority of adverse reactions following vaccination are mild to moderate, a sign that the immune system of the body is fortifying its defenses. Additionally, machine learning algorithms were employed to forecast the severity of side effects, and it was discovered that, based on the input data, RF, XGBoost, and MLP all offered very reliable predictions of the severity of side effects. Exceptional cases might necessitate hospitalization or additional medical care care.[1]

COVID-19 vaccination helps protect you by creating an antibody response without you having to experience potentially severe illness or post-COVID conditions. Getting sick with COVID-19 can cause severe illness or death, even in children, and we can't reliably predict who will have mild or severe illness.

5.3 Future Work

If you receive the COVID-19 virus, a vaccination may be able to keep you from getting it or keep you from getting really ill if you do. Vaccine upgrades your immune system and get familiar with the virus sells and build antibody to create a self-destruction wall to keep the immune system awake. Following vaccination, you can resume a variety of activities that you might not have been able to undertake due to the pandemic with greater safety. However, whether or not you have received the vaccination, the CDC advises wearing a mask inside in public if you live in a region where there are a lot of COVID-19 patients in hospitals and recent incidents of the disease. The CDC advises using the most protective mask you can find that fits properly and is pleasant to wear. Local authorities and recognized and esteemed Imams (religious leaders) can play a crucial role in disseminating this knowledge to the elderly population.

Due to the restricted reading and knowledge of older people, it may be more feasible to employ audio-visual materials for educating the elderly population about the COVID-19 immunization than brochures or written documentation. Furthermore, elderly Bangladeshis have a high degree of misconceptions about COVID-19, which contributes to their resistance to vaccination. Therefore, policymakers and professionals in public health should think about delivering age-appropriate COVID-19 vaccination-related information.[5] By generating an immune response, the COVID-19 vaccination serves to protect you without requiring you to suffer from a potentially serious sickness or post-COVID problems. Even in children, COVID-19 can cause serious sickness or death, and it is impossible to anticipate with any level of certainty whether an individual will experience moderate or severe symptoms.

REFERENCES

[1] R. L. A. C. U. o. D. B. J. R. N. D. 1. I. Department of Microbiology, "COVID-19 Vaccine: A comprehensive status report".

[2] S. G. Rao and D. J. Bramhanandareddy, “Predicting the Mortality of COVID-19 Patients ∗,” 2020.

[3] A. Abdullha and S. Abujar, “COVID-19: Data Analysis and the situation Prediction Using Machine Learning Based on Bangladesh perspective,” *Proc. - 2020 15th Int. Jt. Symp. Artif. Intell. Nat. Lang. Process. iSAI-NLP 2020*, 2020, doi: 10.1109/iSAI-NLP51646.2020.9376812.

[4] I. Haq, M. I. Hossain, A. A. S. Saleheen, M. I. H. Nayan, and M. S. Mila, “Prediction of COVID-19 Pandemic in Bangladesh: Dual Application of Susceptible-Infective-Recovered (SIR) and Machine Learning Approach,” *Interdiscip. Perspect. Infect. Dis.*, vol. 2022, 2022, doi: 10.1155/2022/8570089.

[5] M. M. Akiful Haque *et al.*, “Acceptance of COVID-19 vaccine and its determinants: evidence from a large sample study in Bangladesh,” *Heliyon*, vol. 7, no. 6, 2021, doi: 10.1016/j.heliyon.2021.e07376.

[6] S. K. Mistry *et al.*, “COVID-19 vaccination coverage is extremely low among older population in Bangladesh: findings from a cross-sectional study,” *Hum. Vaccines Immunother.*, vol. 18, no. 1, 2022, doi: 10.1080/21645515.2022.2030624.

[7] S. G. Rao and D. J. Bramhanandareddy, “Predicting the Mortality of COVID-19 Patients ∗,” 2020.

[8] A. Abdullha and S. Abujar, “COVID-19: Data Analysis and the situation Prediction Using Machine Learning Based on Bangladesh perspective,” *Proc. - 2020 15th Int. Jt. Symp. Artif. Intell. Nat. Lang. Process. iSAI-NLP 2020*, 2020, doi: 10.1109/iSAI-NLP51646.2020.9376812.

[9] I. Haq, M. I. Hossain, A. A. S. Saleheen, M. I. H. Nayan, and M. S. Mila, “Prediction of COVID-19 Pandemic in Bangladesh: Dual Application of Susceptible-Infective-Recovered (SIR) and Machine Learning Approach,” *Interdiscip. Perspect. Infect. Dis.*, vol. 2022, 2022, doi: 10.1155/2022/8570089.

[10] M. M. Akiful Haque *et al.*, “Acceptance of COVID-19 vaccine and its determinants: evidence from a large sample study in Bangladesh,” *Heliyon*, vol. 7, no. 6, 2021, doi: 10.1016/j.heliyon.2021.e07376.

[11] S. K. Mistry *et al.*, “COVID-19 vaccination coverage is extremely low among older population in Bangladesh: findings from a cross-sectional study,” *Hum. Vaccines Immunother.*, vol. 18, no. 1, 2022, doi: 10.1080/21645515.2022.2030624.

[12] S. G. Rao and D. J. Bramhanand, “Predicting the Mortality of COVID-19 ∗,” 2020.

[13] A. Abdullha and S. Abujar, “COVID-19: Data Analysis and the situation Prediction Using Machine Learning Based on Bangladesh perspective,” *Proc. - 2020 15th Int. Jt. Symp. Artif. Intell. Nat. Lang. Process. iSAI-NLP 2020*, 2020, doi: 10.1109/iSAI-NLP51646.2020.9376812.

[14] I. Haq, M. I. Hossain, A. A. S. Saleheen, M. I. H. Nayan, and M. S. Mila, “Prediction of COVID-19 Pandemic in Bangladesh: Dual Application of Susceptible-Infective-Recovered (SIR) and Machine Learning Approach,” *Interdiscip. Perspect. Infect. Dis.*, vol. 2022, 2022, doi: 10.1155/2022/8570089.

[15] M. M. Akiful Haque *et al.*, “Acceptance of COVID-19 vaccine and its determinants: evidence from a large sample study in Bangladesh,” *Heliyon*, vol. 7, no. 6, 2021, doi: 10.1016/j.heliyon.2021.e07376.

[16] S. K. Mistry *et al.*, “COVID-19 vaccination coverage is extremely low among older population in Bangladesh: findings from a cross-sectional study,” *Hum. Vaccines Immunother.*, vol. 18, no. 1, 2022, doi: 10.1080/21645515.2022.2030624.

PLAGARISM REPORT

23/10/23

COVID-19 Vaccine Data Analysis and Prediction using Machine Learning A Perspective in Bangladesh

ORIGINALITY REPORT

19%	10%	6%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
2	Abir Abdullha, Sheikh Abujar. "COVID-19: Data Analysis and the situation Prediction Using Machine Learning Based on Bangladesh perspective", 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), 2020. Publication	2%
3	Submitted to De Montfort University Student Paper	2%
4	Submitted to AUT University Student Paper	2%
5	Submitted to Florida International University Student Paper	1%
6	www.cell.com Internet Source	1%
7	www.cdc.gov Internet Source	1%