Machine learning models for Prediction of the need for future Covid-19 vaccine booster

Ahmad Al Marzook
aalmarz6@gmu.edu

Ge Xu
gxu4@gmu.edu

Prajna Shetty Jagannath
pjaganna@gmu.edu

Abstract—About 4 million Virginia citizens are fully vaccinated against COVID-19. According to medical data from nations like Israel and the United Kingdom, all those people may require another shot in the near future, at least when keeping transmissible diseases at bay. Health officials are looking at whether a booster shot is needed to ensure the vaccine's effectiveness. To prepare it for predictions, we will develop a new simple logistic regression model. Make prediction models using classification and probability. With the vaccine immune response lifetime taken into account, we will run our model to forecast the completely vaccinated number timeline and observe when it achieves the herd immunity percentage of the overall Virginia population. The percentage of people that need to be vaccinated to attain COVID-19 herd immunity is still up for dispute. The goal of this study was to forecast when everyone would be fully immunized. COVID-19 vaccination loses nearly half of its defense antibodies every 108 days, according to Elie Dolgin. As a result, immunizations that initially provided 90% protection against mild episodes of the disease may only provide 70% protection after 6 or 7 months.

Index Terms—Machine learning, Vaccine, Booster, Prediction

I. INTRODUCTION

Since the pandemic outbreak, pharmaceutical companies in the United States have been working around the clock to discover, test, and manufacture COVID-19 treatments and vaccines. There are now three COVID-19 are presently available, one of which was recently licensed for use in specific populations by the U.S. Food and Drug Administration (FDA [1]).

COVID-19 vaccinations are safe; over 370 million doses have been given in the United States alone, with billions more throughout the world [1]. As of today, the FDA has approved one antiviral for use in COVID-19 patients aged 12 and up, with the potential to reduce recovery time by up to 5 days. Patients are able to stay out of the hospital and recuperate thanks to monoclonal antibodies. As time passes, every 108 days or so, people who have been immunized against COVID-19 are thought to lose half of their protective antibodies. As a result, vaccines that initially gave 90% protection against mild episodes of the disease may only provide 70% protection after 6 or 7 months [2].

Vaccine booster may be appropriate for some patients whose primary vaccination, defined as the original one-dose or two-dose series of each vaccine, did not provide significant protection—for example, those who received vaccines with low efficacy or those who are immune-compromised to (although people who did not respond robustly to the primary vaccination might also not respond well to a booster). This is unclear if an additional dose of the same vaccine or a different vaccine that complements the first immune response would be more beneficial to such impaired people [3].

II. BACKGROUND

As the epidemic progresses, the world faces new obstacles, including variations and low vaccination rates, to name a few. As new information becomes available, the medical community continues to examine data and collaborates with health groups to provide public health recommendations. This is particularly true in the case of the COVID-19 booster shot [4].

Those in Virginia who have been waiting for a Moderna or Johnson & Johnson booster shot may soon be able to get one. Meanwhile, health officials debated the efficacy and safety of mixing and matching COVID vaccine brands for booster shots. While booster doses of the Covid-19 vaccine are not required at this time, more evidence is needed to determine if patients will need them in the future, and a spike in so-called breakthrough cases could provide a clue, according to federal vaccine experts. Also machine learning models and data analytic tools show this phenomenon [4]–[9] in June.

BOOSTER DOES FOR PERSONS OVER THE AGE OF 18 [10]: label=""

1) VACCINES FROM PFIZER-BIONTECH AND MODERNA: Adults 18 years and older who received a Pfizer-BioNTech, or Moderna COVID-19 vaccine should receive a booster shot six months or more after their original series.

2) JOHNSON AND JOHNSON VACCINES: For anyone who got the Johnson & Johnson (Janssen) COVID-19 vaccine, a single booster shot is suggested two months or more after the initial vaccination for anyone 18 years of age or older.

3) MIX AND MATCH: Eligible people can choose which vaccine they want as a booster shot. Some people prefer the vaccination type they had when they were first immunized, while others prefer to obtain a different booster. The brand to use may be determined by the availability of the brand and an individual benefit-risk assessment. If you're unsure about the brand of booster to take, talk to your doctor.
III. RELATED WORK

It had been almost two years since the new crown epidemic in 2020, and there have been many papers published on the epidemic and expert analysis and research. Our main area of research is directed at the effectiveness of vaccination and the need for future vaccination boosters.

Based on the article [11], we learned that to date, current estimates of effective population immunization levels for interruption of transmission are about 60-70%, but the vaccination coverage needed to interrupt transmission with partially effective vaccines is likely to be higher. With 8 billion people requiring two vaccine doses, people may need 1-11 billion doses to interrupt transmission [11]. However, in the absence of long-term data, little is known about the duration of their confer immunity and online analytic can provide more tools [12]–[18].

In the case of the COVID-19 vaccine, it remains unknown how long immune protection lasts, but vaccine developers and health officials know it may not last forever, and those emerging variants of the virus may evade immunization, machine learning graphics also show these results [19]–[27]. While a typical booster dose uses a previously administered vaccine to remind the immune system of its immunity to the pathogen, any future booster dose for the COVID-19 vaccine could use a completely different vaccine. Regardless of whether the booster dose of the coronavirus vaccine is modified, missing a booster dose could leave someone with reduced protection against COVID-19 if a booster dose is recommended in the future. Three companies currently licensed in the United States - Pfizer, Moderna, and Johnson & Johnson - are investigating the potential use of boosters.

The World Health Organization (WHO) has called for a moratorium on COVID-19 boosters until the end of September, intending to ensure that at least 10 percent of people in all countries and territories are vaccinated before additional doses are distributed. The status of boosters has not been confirmed [28]. Most COVID-19 vaccines currently in use remain highly effective several months after administration, especially against severe disease and death. To date, there is little evidence that booster immunization is needed to protect fully vaccinated individuals.

However, faced with a surge in infections caused by the highly infectious SARS-CoV-2 Delta variant and suggesting that the immunity triggered by the COVID-19 vaccine may fade over time, some countries consider providing additional doses to those who have already been vaccinated [29]. Germany and Israel have now announced booster vaccination programs, and a growing number of countries, including the United Arab Emirates, China, and Russia, have begun administering additional injections. However, evidence for the effectiveness of boosters still lacks worldwide, so providing additional doses in the coming months would be a poor use of resources [7], [29]–[33].

All these literatures will help us find the link between the problem questions and data that need to be addressed and help us predict the future demand for COVID-19 vaccine boosters, thus helping us to develop protocols.

IV. PROBLEM DESCRIPTION

COVID-19 is fully vaccinated in about 4 million Virginia residents [34]. All of those patients may need another shot in the near future. To ensure the vaccine’s effectiveness, health officials are investigating if a booster shot is required.

Antibody levels in vaccinated people have been steadily declining, according to immunological studies [32]. An increasing risk of breakthrough infection has been discovered during the long-term monitoring of immunization trial participants [35]. COVID-19 vaccines are losing their effectiveness, at least when it comes to keeping transmissible diseases at a distance, according to healthcare records from countries such as Israel, the United Kingdom, and others. That’s without taking into consideration the Delta danger because it’s apparent that vaccine-induced antibodies are less effective in recognizing SARS-CoV-2 variations than the virus’s ancestral strain. However, it’s unknown to what extent the immune system’s defenses against serious disease, hospitalization, and death are eroding as well.

V. APPROACH

The dataset of vaccinations was obtained from the Virginia Department of Health for Virginia residents.

1) The data was cleaned to obtain the data we needed.
2) The cleaned data were visualized and analyzed to obtain the necessary information to study.
3) The processed data classified according to Elie Dolgin [36] estimates that people who receive the COVID-19 vaccine lose about half of their defensive antibodies every 108 days or so, and we will classify the time of vaccination of the population. In this step, we will also continue with an in-depth analysis of future booster vaccinations.

VI. DATASET DESCRIPTION

The targeted population in this research is Virginia residents. There is only one official Health Department in the state of Virginia. The main source of the dataset is The Virginia Department of Health. This dataset includes the number of COVID-19 vaccine doses administered for each locality in Virginia by administration date and by facility type [37]. The data set is published publicly in an online dashboard. The data set is also accessible via The Socrata Open Data API (SODA), which gives the ability to access and iterate the data within the programming language. The first date recorded in the data of the first vaccine administered was 12/14/2020. The data set is updated daily, which means it is also increasing in size daily. (See I)

VII. DATASET CLEANSING

The way the dataset is built is somehow accurate, but there is manual reporting in the data, making errors occur. The health facility reports the counts of the doses administrated to
TABLE I
DATASET METADATA

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>administration_date</td>
<td>The date when the vaccine was administered to the person.</td>
<td>Date</td>
</tr>
<tr>
<td>fips</td>
<td>5 unique digits code to specify the locality</td>
<td>Text</td>
</tr>
<tr>
<td>locality</td>
<td>The city or the county where the person who took the vaccine lives</td>
<td>Text</td>
</tr>
<tr>
<td>health_district</td>
<td>The health district name is assigned by The Virginia Department of Health.</td>
<td>Text</td>
</tr>
<tr>
<td>facility_type</td>
<td>Facility type of the provider that performed the vaccine administration.</td>
<td>Text</td>
</tr>
<tr>
<td>vaccine_manufacturer</td>
<td>The manufacture name of the vaccine administered</td>
<td>Text</td>
</tr>
<tr>
<td>dose_number</td>
<td>The dose number for the person who took the vaccine.</td>
<td>Number</td>
</tr>
<tr>
<td>vaccine_doses_administered</td>
<td>The total number of vaccine doses administered.</td>
<td>Number</td>
</tr>
</tbody>
</table>

The Virginia Department of Health updates the online dashboard daily, and it is expected to be updated by 12:00 noon daily.

The U.S. Food and Drug Administration (FDA) currently authorized three COVID-19 vaccines for emergency use [38]. These three vaccines are Pfizer-BioNTech, Moderna, and Janssen & Janssen. The first analysis of the data showed us a fourth vaccine administered in Virginia that the FDA does not authorize. The fourth vaccine is AstraZeneca. We excluded AstraZeneca from the analysis since FDA does not approve it.

We also noticed that there is a quite number of vaccines that are Non-specified (See Figure 2). Most of the Non-specified vaccines are administered at federal facilities. We assume that these doses are either Moderna or Pfizer.

We also found Pfizer has two types of vaccines one type entered with the following vaccine manufacturer name “Pfizer 5-11” instead of entering “Pfizer” only. We assume that this is Pfizer dose for children from age 5 to 11 years old.

VIII. DATA MATERIALS

As of November 6, 2021, the current dataset contains 458,103 entries from 35 health districts across Virginia. Each entry may contain more than one vaccine administered in the same facility. The breakdown of the doses administered within Virginia is as follows: Pfizer has 6,666,840 doses administered, including Dose 1, Dose 2, and Dose 3. Moderna has 4,043,755 doses administered, including Dose 1, Dose 2, and Dose 3. J&J has 377,481 doses administered, including Dose 1 only. Also, there are 631,930 non-specified doses administered, including Dose 1 and Dose 2 only.

Fig. 1. All Doses Total by Manufacturer
Fig. 2. First Dose Total by Manufacturer
Fig. 3. Second Dose Total by Manufacturer
Fig. 4. Third Dose Total by Manufacturer
IX. Methodology

Our approach includes using visual analytic models in Python and scikit-learn models. We will first use a pie chart to represent vaccine usage in Virginia and then a line graph visualization to represent the number of vaccines administered daily and the percentage of people fully vaccinated in Virginia. In addition, we will create a new simple logistic regression model to make it ready for prediction. After finalizing our model, we will train the model on the existing data and create the classification and probability models for prediction. In summary, we will run our model to predict the schedule of the number of people fully vaccinated and see when it reaches the percentage of the total population in Virginia that is herd immunized, taking into account the life span of the vaccine immune response.

X. Preliminary Results

The Virginia population in 2021 is estimated to be 8.67 million [39], and currently, 5.21 million are fully vaccinated, which means 61% of the Virginia population are fully vaccinated [40]. The count of fully vaccinated people is increasing daily (See Source: Our World in Data Figure 5), but we do not know how fast the increment is occurring, and because of that, this research was initiated. This research aimed to predict the fully vaccinated timeline in the future. Based on Elie Dolgin [36], every 108 days or so, the COVID-19 vaccine loses about half of its defensive antibodies. As a result, vaccines that initially offered, say, 90% protection against mild cases of disease might only be 70% effective after 6 or 7 months [36]. The percentage required to achieve COVID-19 herd immunity using the vaccine is still being debated. Still, we found an article in the American Medical Association newspaper containing the following: “What we need to do is we need to get 85% of the U.S. population vaccinated—we need all of the adolescents and all of the adults,” said Dr. Hotez [41]. Based on the 85% herd immunity, our research will conclude the results. Upon our online searching and investigation, we could not find research similar to our research. We assume that our research will be unique on the aimed goals that the research was initiated for.

XI. Model Evaluation

WHEN WILL WE Achieve A VACCINATION RATE OF 100% PER 100 PEOPLE?

This paper describes a simple time series analysis using the PyCaret regression module to forecast the COVID-19 Vaccination Rate in several counties across Virginia to address the above questions.

PyCaret is a low-code machine learning library and end-to-end model management solution for automating machine learning workflows that is created in Python. PyCaret Regression Module is a supervised machine learning module for estimating associations between a dependent variable (commonly referred to as the “outcome variable” or “target”) and one or more independent variables (often referred to as “features” or “predictors”).

The abstract information about the immunization in Virginia is shown in Figure 6.

Those who have been fully vaccinated: Total number of patients who received two doses of Pfizer or two doses of Moderna COVID-19 vaccination, or one dosage of Pfizer and one dose of Moderna COVID-19 vaccine, or one dose of Johnson & Johnson (J&J) COVID-19 vaccine.

Based on the data we utilized, the forecast of vaccination rate (blue line) in Virginia is shown below.

The vaccination rate in the state could be influenced by a variety of factors or characteristics. The predictors in this simple time series analysis using PyCaret forecasting equation are lags of the dependent feature (here vaccination rate) and lags of the forecasting mistakes.

The above prediction was made using the fully vaccinated plot’s training and test data, with test data assigned for the months of September through December.

The above graph depicts a fully vaccinated Virginian population with two doses of Pfizer, Moderna, and Johnson & Johnson vaccines (J&J). People who have had their booster shots but have not yet been immunized.

XII. Conclusion and Future Scope

COVID-19 is rapidly spreading throughout the world. COVID-19 immunization can be considered a safe haven in this pandemic to save a human life and prevent COVID-19 deaths and infections [42]. Since the COVID-19 vaccine was first released, a variety of studies have been conducted on it. Different strategies and approaches to forecast the outcomes of this immunization may be useful for future prediction [43]. The purpose of this research is to forecast people’s attitudes concerning COVID-19 immunizations, as well as dispel fallacies about vaccination and its health consequences.

The used dataset includes the percentage of people who have been vaccinated in each Virginia county, as well as their monthly booster injections. We used a time series dataset to extract the following features for this study:
1) Date Time Features: these are components of each observation’s time step. ‘Series’ refers to each time step (which you’ll see in the next cell) and ‘Target’ refers to the current time step’s target value.

2) Lag Features: values from previous time steps. ‘Shift1’ refers to the preceding time step’s goal value.

3) Window Features: these are a summary of values over a set of previous time steps in a specific window. We assumed that each window has ten samples. ‘Window mean’ is the target’s average over the past ten days.

Using PyCaret, we were able to estimate when the state will attain herd immunization. We divided the dataset into training and test sets and used 5-fold cross validation to compare the models. Forecasting the future is a little more difficult because we used lag and window features. For example, because we don’t know the target value for 2022-1-1, we don’t have the previous value for 2022-1-2. As a result, we’ll begin with the first future time step and generate forecasts as well as fill in the lag features for subsequent time steps. (Recursive functions). We may expect 85 percent of Virginians to be fully vaccinated on or around February 3, 2022, and 100 percent by April 2022. While the overall results with the given dataset demonstrate a favorable trend in people’s attitudes toward vaccination, there are still people who are hesitant to get vaccinated due to various myths and hazards, as well as other...
health concerns. Despite inadvertent errors of no relevance, the projections in this study should aid public-health practitioners and policymakers in better anticipating vaccine uptake habits and developing more appropriate policies. Furthermore, applicable forecast models can be used to predict future states and pandemic episodes. Because coronaviruses are likely to be weaker in the summer, vaccination willingness may also shift. The COVID-19 vaccination now has less than a year of history, making it difficult to evaluate seasonal aspects to date. With further data, we hope to improve the present model by incorporating the PyCaret technique, which takes seasonal aspects into account.

REFERENCES
