**A**

**MINI PROJECT ON**

**“Analytics on Covid\_vaccine\_statewise dataset”**

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE

## COMPUTER ENGINEERING

## As prescribed by

## SAVITRIBAI PHULE PUNE UNIVERSITY



SUBMITTED BY:

SHRUTI SANTOSH GAIKWAD T190434259

DIKSHA DINESH GIRI T190434265

RUTUJA SUNIL KHUTWAD T190434302

PRATIKSHA GANPAT KOLEKAR T190434305

SUBMITTED TO: -PROF. D. L. FALAK

**DEPARTMENT OF COMPUTER ENGINEERING**

**STES’S SINHGAD ACADEMY OF ENGINEERING**

KONDHWA (BK), PUNE 411048

2023-2024



**CERTIFICATE**

This is to certify that the project report entitled

**“Analytics on Covid\_vaccine\_statewise dataset”**

**Submitted By**

Shruti Santosh Gaikwad T190434259

Diksha Dinesh Giri T190434265

Rutuja Sunil Khutwad T190434302

Pratiksha Ganpat Kolekar T190434305

Is a bonafide work carried out by them under the supervision by

Prof. D.L.Falak and it is approved for the partial fulfilment of the requirement of Savitribai Phule Pune University for the Project in the Third Year of Computer Engineering.

|  |  |  |
| --- | --- | --- |
| **Prof. D.L.Falak**  **Guide**  **Dept. of Computer**  **Engg.** | **Prof. S. N. Shelke**  **H.O.D**  **Dept. of Computer**  **Engg.** | **Dr. K. P. Patil Principal SAOE, Pune** |
| Place: Pune  Date: / /2024 |  |  |

**CONTENTS**

* Abstract
* Introduction
* Problem statement
* Hardware and Software used
* About dataset
* Code Implementation
* Conclusion
* References

**ABSTRACT**

The COVID-19 outbreak has created effects on [everyday life](https://www.sciencedirect.com/topics/medicine-and-dentistry/activities-of-daily-living) worldwide. Many research teams at major [pharmaceutical companies](https://www.sciencedirect.com/topics/computer-science/pharmaceutical-company) and research institutes in various countries have been producing vaccines since the beginning of the outbreak. There is an impact of gender on vaccine responses, acceptance, and outcomes. Worldwide promotion of the COVID-19 vaccine additionally generates a huge amount of discussions on [social media platforms](https://www.sciencedirect.com/topics/computer-science/social-medium-platform) about diverse factors of vaccines including protection and efficacy. Twitter is considered one of the most well-known social media platforms which have been widely used to share a public opinion on vaccine-related problems in the COVID-19 pandemic. However, there is a lack of research work to analyze the public perception of COVID-19 vaccines systematically from a gender perspective. In this paper, we perform an in-depth analysis of the [coronavirus](https://www.sciencedirect.com/topics/medicine-and-dentistry/coronavirinae) vaccine-related tweets to understand the people’s sentiment towards various vaccine brands corresponding to the gender level. The proposed method focuses on the effect of COVID-19 vaccines on gender by taking into account descriptive, diagnostic, predictive, and prescriptive analytics on the Twitter dataset. We also conduct experiments with [deep learning models](https://www.sciencedirect.com/topics/computer-science/deep-learning-model) to determine the sentiment polarities of tweets, which are positive, neutral, and negative. The results reveal that [LSTM](https://www.sciencedirect.com/topics/medicine-and-dentistry/short-term-memory) performs better compared to other models with an accuracy rate of 85.7%.

## Keywords: Data analytics, Covid-19 vaccine, Tweet, Sentiment analysis, Deep learning

.

**INTRODUCTION**

The outbreak of the novel [Coronavirus](https://www.sciencedirect.com/topics/medicine-and-dentistry/coronavirinae) Disease (COVID-19) has hit the earth drastically and billions of people are affected worldwide. According to a recent report by the World Health Organization (WHO), more than 260 million people have been infected with the virus, and over 5 million deaths are caused by the COVID-19.[1](https://www.sciencedirect.com/science/article/pii/S2352914822001149" \l "fn1) COVID-19 has affected public life extensively and compared to past epidemics such as the [Spanish Flu](https://www.sciencedirect.com/topics/medicine-and-dentistry/spanish-influenza) of 1918 and the Black Death of the 13th century, it is considered to be the most serious epidemic of this century. [[1]](https://www.sciencedirect.com/science/article/pii/S2352914822001149" \l "b1). Now it has become a common goal for all countries to eliminate COVID-19 and return to normal life activities. Many countries officially have followed a specific series of measures to lessen the spread of COVID-19 including closing borders, reducing work in public locations (e.g., gyms, restaurants, shopping malls), learning from homes, limiting travel, and wearing masks, maintaining social distance, and personal hygiene. These measures have had a positive effect on controlling the spread of the virus. However, the latest COVID-19 [Omicron](https://www.sciencedirect.com/topics/medicine-and-dentistry/omicron-coronavirus-variant) variant put the countries around the world again in concern [[2]](https://www.sciencedirect.com/science/article/pii/S2352914822001149" \l "b2). The presence of COVID-19 may be durable in the future and vaccination is the most effective long time way to handle the COVID-19 pandemic situation. Thus, the countries and [pharmaceutical companies](https://www.sciencedirect.com/topics/computer-science/pharmaceutical-company) around the world have begun the development of vaccines and [clinical trials](https://www.sciencedirect.com/topics/medicine-and-dentistry/clinical-trial) from the starting of the pandemic outbreak.

As of December 01, 2021, there are 160 vaccine candidates and 24 of them have been approved by various countries around the world.[2](https://www.sciencedirect.com/science/article/pii/S2352914822001149" \l "fn2) At first, Pfizer/BioNTech vaccine received emergency validation by WHO on December 31, 2020, and then AstraZeneca, Covishield, Janssen, Moderna.[3](https://www.sciencedirect.com/science/article/pii/S2352914822001149" \l "fn3) Each country provides approval for multiple vaccinations and evolves precise guidelines to inspire all citizens to be vaccinated. In most countries, current vaccination rates have no longer yet reached the minimal standards for restricting the expansion of the pandemic. People’s lack of knowledge and belief in vaccines, and skeptical attitudes towards vaccines are the possible reasons for low vaccination rates in many countries. They fear that if the vaccine is not adequately tested, it could lead to chronic disease.

**PROBLEM STATEMENT**

Use the following covid\_vaccine\_statewise.csv dataset and perform following analytics on the given dataset <https://www.kaggle.com/sudalairajkumar/covid19-in-india?select=covid_vaccine_statewise.csv>

a. Describe the dataset

b. Number of persons state wise vaccinated for first dose in India

c. Number of persons state wise vaccinated for second dose in India

d. Number of Males vaccinated

e. Number of females vaccinated

**HARDWARE USED**

Device name: LAPTOP-3GE5PM13

Processor: 11th Gen Intel(R) Core(TM) i3-1115G4 @ 3.00GHz 2.90 GHz

Installed RAM:8.00 GB (7.65 GB usable)

System type 64-bit operating system, x64-based processor

**SOFTWARE USED**

Languages: python

Application: Jupyter Notebook

Dataset:Kaggle.com(<https://www.kaggle.com/sudalairajkumar/covid19-in-india?select=covid_vaccine_statewise.csv>

**ABOUT DATASET**

Coronaviruses are a large family of viruses which may cause illness in animals or humans. In humans, several coronaviruses are known to cause respiratory infections ranging from the common cold to more severe diseases such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). The most recently discovered coronavirus causes coronavirus disease COVID-19 - World Health Organization

The number of new cases are increasing day by day around the world. This dataset has information from the states and union territories of India at daily level.

State level data comes from [Ministry of Health & Family Welfare](https://www.mohfw.gov.in/)

Testing data and vaccination data comes from [covid19india](https://www.covid19india.org/). Huge thanks to them for their efforts!

Update on April 20, 2021: Thanks to the [Team at ISIBang](https://www.isibang.ac.in/~athreya/incovid19/), I was able to get the historical data for the periods that I missed to collect and updated the csv file.

### Content

COVID-19 cases at daily level is present in covid\_19\_india.csv file

Statewise testing details in StatewiseTestingDetails.csv fileTravel history dataset by [@dheerajmpai](https://www.kaggle.com/dheerajmpai) - <https://www.kaggle.com/dheerajmpai/covidindiatravelhistory>

### Acknowledgements

Thanks to Indian [Ministry of Health & Family Welfare](https://www.mohfw.gov.in/) for making the data available to general public.

Thanks to [covid19india.org](http://portal.covid19india.org/) for making the individual level details, testing details, vaccination details available to general public.

Thanks to [Wikipedia](https://en.wikipedia.org/wiki/List_of_states_and_union_territories_of_India_by_population) for population information.

Thanks to the [Team at ISIBang](https://www.isibang.ac.in/~athreya/incovid19/)

Photo Courtesy - <https://hgis.uw.edu/virus/>

### Inspiration

Looking for data based suggestions to stop / delay the spread of virus

**CONCLUSION**

The number of textual data in social media is increasing day by day. People express their needs, emotions, and opinions on social media platforms like Twitter on the outbreak of the COVID-19 pandemic. Thus in this research work, we present a systematic framework by taking the advantages of data analytics and sentiment analysis that might be beneficial to provide useful suggestions for appropriate vaccination by physicians, government officials, and policymakers. Our proposed approach might be helpful in the visualization of vaccine-related pandemic information, effectiveness assessment of various vaccine brands, prevention of side effects, and providing mental and physical satisfaction for both males and females. By implementing a data analytics strategy, we are enabled to perform descriptive, diagnostic, predictive, and prescriptive analyses. Through insightful data analysis, we conclude by providing helpful suggestions to experts to take necessary steps for the proper distribution of various vaccine brands amongst males and females. In the future, we will consider people’s age, location and other contextual information to assist the people for taking exact action in advance.

**REFERENCES**

**1.Kaggle.com-**

Kaggle allows users to find datasets they want to use in building AI models, publish datasets, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges.

**2.sciencedirect.com**

https://www.sciencedirect.com/science/article/pii/S2352914822001149