

Analysis of the Pandemic and Mortality Rate

Mohammed Omar Khan, Akshat Jain and Ankit Sati

INTRODUCTION

This project's major goal is to determine the pandemic's nature and then predict how it will develop. The global impact of Covid 19's pandemic-scale spread cannot be overstated. As a result, the healthcare systems in several nations are in disarray. The decision-makers and healthcare professionals would benefit from the forecast and analysis of the number of cases in order to improve patient care. Earlier studies for forecasting COVID cases included Time Series Forecasting of New Cases and New Deaths Rate for COVID-19 using Deep Learning Methods and Arko Barman's Time Series Analysis and Forecasting of COVID-19 Cases Using LSTM and ARIMA Models [3] where a comparison between LSTM and ARIMA models is made. Bidirectional modelling for prediction is evaluated in [4]. We intend to use the Arima Model to forecast the cases in our project after analyzing the various aspects that are connected to the instances.

HYPOTHESIS

To determine whether different time periods have noticeably different observations, we would forecast and analyze the cases in various states and counties. We wish to examine many elements that influence the death rate. We must take into account the variables that influence this statistic in various regions of the USA in order to understand the death rate there. These variables could include the population of each county, the quantity of tests administered, the number of testing facilities established, the varying climatic conditions across the nation, etc. Gaining a comprehensive grasp of the Covid 19 situation in the USA will be made easier with the help of this analysis and visualization combined with a forecast of the number of instances in the near future.

PROJECT DESIGN PLAN

We have explained the steps to begin the project however the timelines will vary on the datasets and hence are not shared below.

- Data Collection: We'll use Kaggle to gather data. Daily Covid 19 cases for each county in the USA make up the dataset. The initial stage in pre-processing would be to look for null values in each column, after which the data would be changed from daily cases to monthly cases for each county. The time series model would afterwards use this for visualizations.
- Understanding various factors: In this stage, we'll compare the number of fatalities in different regions of the nation to things like
 - Population in each county
 - Number of testing facilities
 - Impact of temperature on cases
 - Number of tests done
- Visualization and Data Wrangling: After pre-processing the data, we would assess its uniformity. Then, to gain important information about the data, we would perform visualizations. Several examples of the visuals include:
 - Exploring the counties with highest covid cases and number of deaths
 - Number of deaths with respect to each county

- *Model:* We intend to forecast using the ARIMA model since we will learn about the various factors influencing the number of instances in each county. To comprehend the impact of seasonality on COVID situations, we also want to use the Seasonal ARIMA model. As the measurement for our model's evaluation, we will utilize root mean squared error.

REFERENCES

1. Singh, S., Chowdhury, C., Panja, A. K., & Neogy, S. (2021). Time Series Analysis of COVID-19 Data <https://doi.org/10.1007/s40031-021-00585-7>
2. Fuad Ahmed Chyon, Md. Nazmul Hasan Suman, Md. Rafiul Islam Fahim, Md. Sazol Ahmmed, <https://doi.org/10.1016/j.jviromet.2021.114433>
3. <https://arxiv.org/ftp/arxiv/papers/2006/2006.13852.pdf>
4. <https://arxiv.org/ftp/arxiv/papers/2104/2104.15007.pdf>