Analysis of the Pandemic and Mortality Rate

Mohammed Omar Khan, Akshat Jain and Ankit Sati

INTRODUCTION

As a species, we lost over 6.5 million deaths due to covid 19 without any idea about how it can impact the future. This project's major goal is to determine the pandemic's nature and then predict how it will develop[1]. In the times where the healthcare systems of various countries are in jeopardy, this analysis can help the professionals and decision makers to make educated choices. Earlier studies for forecasting COVID cases included Time Series Forecasting of New Cases and New Deaths Rate for COVID-19[2] using Deep Learning Methods and 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 modeling for prediction is evaluated in [4]. In our case we will show why we will be using the Arima model[5] and show our analysis with various aspects that are connected to this instances.

OBJECTIVE

We would forecast and examine the cases in different states and counties to see if different time periods have observably different observations. We want to look at a variety of factors that affect the death rate. To comprehend the death rate in the USA, we must consider the factors that affect this number in various parts of the country. These factors might include the amount of people living in each county, the number of tests given, the number of testing centers built, the various climatic conditions across the country, etc. With the aid of this analysis, visualization, and prediction of the number of incidents in the near future, gaining a thorough understanding of the Covid 19 situation in the USA will be made easier. We are even planning to go a step beyond and measure the financial impact that the pandemic has had on the people lives throughout the year of 2022 [6].

DATASETS

For all the different datasets used in this study, data will be collected using Kaggle. The collection consists of Daily Covid 19 instances for each county in the United States. A collaborative model can be used to demonstrate a time series dependence using just the first dataset. Pre-processing would begin by scanning each column for null values before switching the data from daily cases to monthly cases for each county. Following that, this would be used for visualizations by the time series model. To create the visualizations on other aspects, we must combine the second dataset with the first one.

PROJECT DESIGN PLAN

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

- **Fields in the Datasets :** 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 per country
  - ➜ Impact of temperature on cases
  - ➜ Number of tests done
  - ➜ Time span before testing

- **<u>Visualization and Data Wrangling Operations</u>:** Once we are through 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
  - ➔ Deaths post vaccinations
  - ➔ Number of deaths with respect to each county
  - ➔ Region with the net highest spend

- **<u>Models:</u>** Since we will discover the numerous factors affecting the number of incidents in each county, we plan to forecast using the ARIMA model. We also wish to apply the Seasonal ARIMA model to understand how seasonality affects COVID situations. We will use root mean squared error as the metric for measuring the performance of our model.
  - ➔ LSTM
  - ➔ SARIMA
  - ➔ ARIMA

<u>REFERENCES</u>

1. https://www.kaggle.com/code/stubbornrival/time-series-analysis-for-covid-19-dataset/log
2. 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
5. https://www.kaggle.com/code/ssismasterchief/knnbaseline-recommendation-systems-surprise/notebook
6. https://www.kaggle.com/datasets/mirichoi0218/insurance