Covid -19 Prediction Proposal

Anuj Tanwar

Bellevue University

Topic

COVID-19 outbreaks don’t only impact lives of people but disrupts economy and healthcare infrastructure of the county. So, it is important to study the pandemic and predict potential future infectious disease outbreaks. A comprehensive understanding of the spread and past count of cases will help in predicting recent future and enable administration by giving a heads-up to stay alert and can provide sufficient time to facilitate support interventions.

Keywords: Covid-19, Predictions

Covid -19 Prediction Proposal

India has seen an increase in COVID-19 case again during the start of year 2022. A simple google search can tell us that on April 4th, there were only 795 new cases, on Apr 14th there were 949 and in last week count of new cases have been on constant rise and going above 2000 per day now. This trend shows the cases have been rising. As per Ministry of Health and Family Welfare Government of India (https://www.mohfw.gov.in/) there had already been more than half a million documented deaths due to COVID and active cases are on rise. Government of India has also issued new guidelines and restrictions. Referring to the below google graph, we can see that COVID spread had increased in India during the summer months in last 2 years. 2022 summer season has already started in India, so the concerning questions that we have are:

1. Are we going to have another wave of COVID?
2. If so, then how severe it can be?
3. When will we see the peak?
4. How long did the previous waves last?
5. What was the trend in death toll every day?
6. What was the cured trend

I will use data mining techniques learnt in this course to study the data and make prediction on it.

Data Source

COVID-19 Data is collected from below sources:

1. **Datameet** website. It is community collected, cleaned and organized COVID-19 datasets about India, sourced from different government websites which are freely available.
   https://projects.datameet.org/covid19/
   https://github.com/datameet/covid19/tree/master/data

2. Ministry of Health and Family Welfare Government of India: https://www.mohfw.gov.in/

Methods EDA Summary

1. **Data Collection:** Data was collected from all_totals.json file from datameet git location https://github.com/datameet/covid19/tree/master/data.
2. **Data Cleaning:** Keys and values were read separately from JSON dataset and then combined into a dataframe. At this point, dataframe has multiple lines for each date, one row for each type(active cases, deaths, cured, total confirmed cases). This need dataframe to be pivoted on date field to convert types to individual columns.
3. **Data Preparation**

1. Joined case dataset with Vaccine data(mohfw_vaccination_status.json) to get additional information.
2. Dropped features not useful for analysis.
3. Transformed features and join case dataset with vaccine dataset.
4. Dropped records with invalid values.
5. Handle NaNs.
6. Used MinMaxScaler() to rescale each column.
7. Filled missing data with mode of respective column.

4. **Model building and evaluation**

- Model to be used: Long short-term memory (LSTM). It is an artificial neural network used in the fields of artificial intelligence and deep learning.
- I will use sequential model from tensorflow.keras package.
- Compile the model with adam optimizer
- Planning to use EarlyStopping class from keras callbacks for data fitting. This will enable model to stop training when a metric has stopped improving.

Ethical Considerations

1. Uneasiness around the sudden loss of personal liberties that had been taken for granted by citizens in many countries. Regulation and protocols must be in place to avoid panic and frustration in public.
2. Privacy is at risk with digital surveillance like tracing apps and enforcing people to report covid results.
3. Personal biasing due to scare of COVID or otherwise. Personal biasing during the analysis can influence the result drawn from analysis.

Challenges/New Opportunities that need to be explored?

Model can be enhanced to consider below factors in future predictions:

- Second and subsequent vaccine dosage
- Enforced Restrictions and guideline such as mask mandates, lock downs, number of people allows in public transportation such as buses and trains
- Weather conditions such as summer or winter counts

References

https://www.kff.org/global-health-policy/issue-brief/economic-impact-of-covid-19-on-pepfar-countries/