SYNOPSIS ON
COVID DEATHS PREDICTOR

SUBMITTED BY- APARNA SINGH (191 300 155)

CHANCHAL (191 300 995)

SUBMITTED TO – AMIR KHAN SIR

SECTION – B

COURSE – Btech (III YEAR)

ABSTRACT

This work aims to forecast the number of deaths caused by the COVID-19 pandemic in the UK based on data about cases, deaths, government stringency measures etc. We train and compare two ML models for the same. We also attempt a “what if?” analysis that tries to examine what effect high stringency measures (or low stringency measures) right from the early days of the pandemic would have had on the number of deaths.

Project on github:-

https://github.com/chanchal231/MiniProject1

INTRODUCTION

COVID-19 is one of the worst pandemics to have ever occurred in human history. The pandemic started on a small scale in China, but within no time, it blew to international proportions. Even with the fatality rate of the virus being low, it has brought the world to a halt. This highlights how unprepared we, as a civilisation, are for such diseases. The natural questions that arise out of this pitiful situation are:

1) Whether something can be done to curb the spread of the virus now

2) How to better handle such situations in the future

Thus, it is very important to analyse the current situation around the world and to find out where we have gone wrong in managing the disease, and what can be done better. If good analysis of the situation is done, we can prevent the countries' healthcare systems from being so suddenly overwhelmed.

This pandemic is different from the ones before, because the general public has access to a lot of pandemic-data via the internet. Also, the data is being generated and logged in a much more systematic way, allowing for machine-assisted analysis. Machine Learning techniques can comb through enormous amounts of data and extract insights in a way that manual inspection simply cannot match.

Thus, we have decided to apply machine learning techniques to address two important questions about the Covid 19 crises. The questions we will try to answer are:

1) Can we predict the number of deaths due to Covid 19 in the next k days (k=14) given information for the last n days (n=30).

2) Can we predict the change in the number of deaths if the lockdown or stringency measures were different (more or less strict).

In this report, we focus specifically on one geographical location (i.e. UK) as a demonstration of our methods. Similar analysis can be done on other regions as well.

RELATED WORK

We surveyed past work on of COVID-19 modeling. There were many kinds of models including ones that predicted deaths, predicted cases etc. Based on the underlying principle of the models, we observed by-and-large two kinds of models. First, the typical ML models like linear regression, neural networks etc. Second, typical ML techniques combined with an underlying epidemiological model (like the SEIR epidemic model).

We observed that the models that use an epidemic model under the hood perform far better than simple ML models. Some examples of such hybrid models are [YYG] and [GOG]. We think these models perform so well because the underlying SEIR model is able to capture the disease mechanics effectively. In a sense the SEIR model captures the manner in which the data is being generated and provides a strong bias (which is empirically shown to be correct) of how an epidemic progresses hence helping the model make better predictions.

METHODOLOGY

The data we are using for our modeling comes from two reliable sources. Data on the number of new cases, new deaths, icu patients, hospital patients etc. comes from “Our World In Data” [OWID], while the Stringency Index, and its breakdown into different components comes from “Oxford COVID-19 Government Response Tracker” [OXCGRT].

The machine learning problem can be stated as a “Time-series regression problem”. ‘Time-Series’ because similar data is logged along a time axis, and information about past time units is required to predict variables in the future. ‘Regression’ because we will be attempting to predict the number of deaths for a given number of days in the future.

Software Specification

· Technology Implemented: Machine Learning

· Language Used: Python

· · Web Browser: Chrome / Firefox

Hardware Requirements

· Processor: intel i3

· Operating System: Windows 7/8/10

· RAM: 4+GB

· Hard disk: 64 GB

· Hardware Devices: Computer System

CONCLUSIONS

From all the discussions and results presented above, we can conclude that

1) An LSTM model is likely to do better than a Decision Tree model on predicting the number of deaths in the future given some information about the past. This may be because LSTMs can more naturally work on time-series data. On the other hand, a Decision Tree can tell us about feature importances in a human-interpretable way.

2) Both the models do much worse compared to mathematical models like SEIR.. A hybrid SEIR and ML model may be able to perform even better, with the parameters in the SEIR model being determined by the ML model, given some appropriate evaluation function.

3) There are several reasons why the ML models do not perform well. Presence of false reporting (accidental or intentional) and less data are two significant reasons.

4) Machine Learning, although a useful tool in such a situation, can do much better with the presence of effective biases to model the problem better.

REFERENCES

https://covid19-projections.com/about/

https://arxiv.org/abs/2008.00646

https://github.com/OxCGRT/covid-policy-tracker

https://www.who.int/emergencies/diseases/novel-coronavirus-2019