COVID-19 Risk Prediction

Anonymous

1. INTRODUCTION

In this proposal, we plan on creating a model that will predict the risk of COVID 19 in a country based on the past trend in that country. This model will interpolate COVID data-sets from existing sources and extract knowledge related to predicting countries that have a high risk of COVID 19. First, we will clean the data and then load it into a database management system(DBMS). We will be using MySQL as our DBMS. Then we will extract the attributes that will help us classify the countries while discarding the attributes that introduce noise to the prediction. We then plan to use a Recurrent Neural Network with LSTM cells to achieve the task of classifying countries based on the risk level as it will be able to use the temporal dependencies in the data to make the predictions more accurate. We also plan on using hierarchical clustering of high-risk countries predicted by our model to find out clusters of countries that are risky to travel to.

2. DATA-SET DESCRIPTION

Our primary dataset is from John Hopkins University. This data set provides details about the spread of COVID-19 in a country on a daily basis with attributes such as daily new COVID-19 cases, daily new confirmed deaths due to COVID-19. We extended our research to include our second data-set. This not only adds more data about the number of fatalities per country but also provides weather data to determine the potency of COVID-19 in different environmental conditions.

3. DBMS

The dataset that is obtained from multiple sources will combined and stored in MySQL. MySQL is our primary choice for performing all the database management systems' operations. We would like to take advantage of all the tools that are offered alongwith MySQL. However, we will be also researching other database management tools as we move forward with the project so that if needed we can transfer the data from MySQL or original sources to other tools too. Hence, we would like to keep MySQL as our primary choice and then change it if needed.

4. DELIVERABLE

phase 1 - In this phase we brainstormed about choosing a well relatable domain to current worldly affairs. After shortlisting a couple of domains we finalized on the healthcare and emergency domain. We indexed through various sources related to COVID-19 data and finalized the given data-sets.[2] [1]

phase 2 - In this phase, we want to learn about the existing techniques that are used for classification tasks in which there are temporal dependencies in the data. After a critical examination of these techniques, we will decide on a technique that is best suited for the problem at hand. Simultaneously we also plan on working on cleaning the data and loading it to our DBMS system as well as identifying the attributes that will help us fit our model to the data.

phase 3 - By this phase, we are determined to be able to prototype a working model that satisfies all our key takeaways. We shall extend our research and branch outwards by adding new features to the model.

5. REFERENCES

[1] R. Pal, A. A. Sekh, S. Kar, and D. K. Prasad. Neural network based country wise risk prediction of covid-19. *Appl. Sci.*, 2020.

[2] N. Rochmawati, H. B. Hidayati, Y. Yamasari, W. Yustanti, L. Rakhmawati, H. P. Tjahyaningtijas, and Y. Anistyasari. Covid symptom severity using decision tree. In *2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE)*, pages 1–5. IEEE, 2020.