

COVID-19 Prediction with Symptoms

Guankai Zhai, Weiyan Wu, Jianyang Zhou

October 2021

Introduction

As the world began to mount an unprecedented worldwide response to COVID-19 in early 2020, it lacked a standardized, global way to measure COVID-19 illness and track the pandemic that would help guide decision-making (e.g., implementation of social distancing measures). Thus, a model to identify infected patients from the symptoms they display is urgently needed.

The Data

In response, the Delphi Group at Carnegie Mellon University (CMU), the University of Maryland (UMD) Joint Program in Survey Methodology, and Facebook launched an opt-in, off-platform survey about COVID19-like symptoms to a daily sample of the 2B+ monthly users of Facebook, starting on April 6, 2020 in the United States (US) and May 1, 2020 globally.

To date, over 30M total responses have been recorded across 200+ countries and territories, and in 55+ languages. The global dataset is made available by UMD, while data in the US is made available by the CMU. The dataset includes features including but not limited to the time of submission, the geographic data of submission, demographic information about the respondent, and reported symptoms.

Our Efforts

We, a team of three undergraduate students at Cornell University, believe that the above dataset would be sufficient for us to build a model that predicts COVID-19 infections from symptoms. We will utilize knowledge and skills learned through ORIE 4741 to effectively clean our data, choose and fit among possible existing models such as decision tree and linear regression, and test our model's performance on a test dataset. We believe that this is a worthwhile effort given the global harms caused by this pandemic and the promising prospects of this project to diagnose patients earlier, easier, and more accurately.