COVID-19 PROACTIVE WEB APP

Machine learning to assess risk factors and predict future U.S. county hotspots

# Overview

Our team consists of Beth Emborsky, Brodie Armstrong and Ellen Rud Gentile. We propose the creation of a full stack, interactive web application, on the cloud, which will create visualizations of risk factors and predicted future covid-19 hotspots based on machine learning models.

## The Objective

* Need #1: Understand which factors are most important to the spread of covid-19
* Need #2: Create a model to assess counties and their risk factors to predict future hotspots over time.
* Need #3: Enable government officials and the public to act proactively to mitigate covid-19 spread in predicted at risk zones.

# Our Proposal

Current covid-19 predictive models are graphical and at the state level. County level, geographical data is available, but only with historical data. Our team will create machine learning models to determine the relative importance of a variety of possible covid-19 transmission and mortality risk factors. We will use the outcome of this study to create a second model that will take the risk factors and historical data to predict the timing and location of future hotspots (at the county level). A cloud-based web application will allow the models inputs and outputs to be visually communicated to users with interactive visualizations. This predictive visualization will allow government officials and individuals to take a more pro-active approach as they decide what measures are appropriate for mitigating covid-19 risks.

## Execution Strategy

Our execution strategy involves proven methodologies, extremely qualified personnel and a highly organized approach to managing deliverables. Following is a description of our project methods, including how the project will be developed, a proposed timeline of events, and reasons for why we suggest developing the project as described.

## Technical/Project Approach

* Source Gathering: A variety of data sources will be assessed for accuracy, timeliness, and completeness/ cleanness of data.
* Data Cleansing and Pre-processing: Source data will be cleansed using Python/ pandas code modules that can easily be re-run when updated source data is obtained. Only pertinent information will be maintained. An ERD (entity relationship diagram) will be created to visually organize the database we will create. Tables will be created based on the ERD using pandas.
* ETL (Extract, Transform, Load) using PostgresSQL
* Flask App to Access Relevant Data
* Machine learning model to assess relative importance of a number of factors relating to how a county becomes a covid-19 hotspot.
* Machine learning model to predict upcoming covid-19 hotspots over a certain timespan
* Interactive web-application hosted on the web, using html, CSS/ bootstrap, javascript, plotly, geopandas and d3.js.

## Timeline for Execution

Key project dates are estimated below. These dates are subject to change, except for the presentation.

|  |  |  |  |
| --- | --- | --- | --- |
| Description | Start Date | End Date | Duration |
| Project Start | 10/08/2020 | 10/20/2020 | 13 days |
| Cleanse & pre-process data | 10/08/2020 | 10/10/2020 | 3 days |
| Create cloud SQL database | 10/10/2020 | 10/12/2020 | 3 days |
| Create flask app layer | 10/10/2020 | 10/12/2020 | 3 days |
| Create front end dummy web app with placeholders for input elements | 10/10/2020 | 10/12/2020 | 3 days |
| Create interactive visualization modules | 10/13/2020 | 10/14/2020 | 2 days |
| Connect application layers | 10/15/2020 | 10/15/2020 | 1 days |
| Deploy & test application | 10/16/2020 | 10/16/2020 | 1 day |
| Revise and improve application as fit | 10/17/2020 | 10/19/2020 | 2 days |
| Prepare presentation | 10/19/2020 | 10/20/2020 | 2 days |
| Project presentation | 10/20/2020 | 10/20/2020 | 10 minutes |