Project: Milestone 2

06/16/2022

DSC 630-T301

The Algorhythms: Andrew Jordan & Jahedur Rahman

**Introduction**

Early in the year 2020 the world was struck by the rampant spread of the Covid-19 virus, a deadly pathogen that required an unprecedented global response to curtail. Chief among the world’s focus was the creation of a vaccination against Covid-19, a goal that was achieved-with initial distribution following shortly after-in 2021. As this pandemic has had a global effect, most countries have tracked and reported data regarding many aspects of Covid-19 response including hospitalizations, infection rate, severity, mortality, testing, and vaccination rate. With this amount of data, from a global sample size, potentially predicting the long-term effects Covid-19 vaccination will have on the global Covid-19 response is not only possible, but it has become a social responsibility to do so.

**The Data**

Our dataset comes from a combined Johns Hopkins University (JHU) Covid-19 dataset and global data reports compiled by *Our World in Data*. Together these two sets of data provide insight into each country’s case numbers, deaths, hospitalizations, tests, and vaccination rate at an ongoing pace from January 1, 2020 through today. Each of the above variables listed are further broken down within the dataset by continent, new and total reports, and reports per capita. As there is no lack of data available with which to work there appears to be little cause to split the data into training and test data, however the initial analysis may prove otherwise.

**Modeling**

There are two steps we must take to ensure we properly evaluate the data, exploratory data analysis and regression modeling. Through visualization and other Exploratory data analysis techniques we expect to identify any potential relationships and trends occurring within the data. In particular, we anticipate our visualizations will provide a great deal of insight into Covid-19 infection, mortality, hospitalization, and vaccination rate over time. This phase of modeling will require far less time and evaluation than the next, but is a necessary step that will provide the framework by which the next modeling phase will be guided.

To determine the effect the Covid-19 vaccine has had on the global pandemic response we will use a multiple linear regression model to evaluate several relationships, most notably between Covid-19 vaccination rate and reported case data variables. Though the initial Exploratory data analysis may provide unexpected potential relationships to examine more deeply, we expect the initial analysis to guide the focus of our regression modeling to Covid-19 vaccinations’ impact on the reported global case data.

**Evaluating Results**

Examining the predictive model’s results will be done so through the examination and comparison of R2 and p-values, as well as through the creation of visualizations to compare actual versus expected results. Comparing actual versus predicted results will be done so by means of a scatter plot, a simple and efficient tool that can be created to display results both at a singular point in time and throughout the entirety of the pandemic.

**Risks**

By not splitting the dataset into training and test data we run the risk of overfitting, compromising the accuracy of our model’s results. There is also risk contained in including too many variables, potentially creating an indication of statistical significance that does not actually exist.

**Contingency Plan**

The two portions of our contingency plan exist to counter the two biggest risks posed by our proposal. If we determine we have compromised the accuracy of our model through overfitting, we will split our data into training and test data before re-running our regression model. The second portion of our contingency plan relies upon risk mitigation through the evaluation of our initial exploratory data analysis to identify and include only the most relevant variables. Failure to do so will require a re-evaluation of the variables included and starting the modeling process over.