Forecasting daily COVID-19 spread in regions around the world.

Frank Lawrence Nii Adoquaye Acquaye*
Faculty of Computer Science
National Research University Higher School of Economics
fakvey@edu.hse.ru

Bense Valerie Caroline
Faculty of Computer Science
National Research University Higher School of Economics
vbense@edu.hse.ru

Abstract

1 Problem statement:

The year 2020 will forever be remembered as the year the earth stood still. This is primarily due to the spread of COVID-19. As Data Scientists we seek to provide solutions to problems facing humanity and the world at large. In this regard we seek to develop a forecasting model that will predict the daily spread of COVID-19 in regions around the world. Our model predicts the number of daily new cases in regions around the world in order to help policy makers plan and manage the COVID-19 pandemic.

2 Dataset summary and EDA:

2.1 Background of dataset:

The White House Office of Science and Technology Policy (OSTP) pulled together a coalition of research groups and companies (including Kaggle) to prepare the COVID-19 Open Research Dataset (CORD-19) to attempt to address key open scientific questions on COVID-19. Those questions are drawn from National Academies of Sciences, Engineering, and Medicine's (NASEM) and the World Health Organization (WHO).

2.2 Data sources:

The sources of data used in this project can be obtained from Kaggle Dataset

2.3 Actual data:

Since the accuracy of such a model is dependent on the freshness of the data, the most up to date data can be found here

*http://acquayefrank.github.io

2.4 Actual data used in project:

In this project we use frozen dataset i.e dataset that has been frozen in time and this dataset can be found here

2.5 Basic exploratory data analysis