FACULTY OF
COMPUTER SCIENCE

Data Science with R - Project Proposal
"ANALYSIS AND PREDICTIONS OF THE IMPACT OF COVID-19"

Anne Rother, Hetti Perera, Mohammed Farhaan Shaikh, Poornima Venkatesha, Shivani Hegde

May 20th, 2020

DATA SCIENCE WITH R - PROJECT PROPOSAL

TEAM MEMBERS

- Anne Rother
- Hetti Maneendra Perera
- Mohammed Farhaan Shaikh
- Poornima Venkatesha
- Shivani Hegde

ANALYSIS AND PREDICTIONS OF THE IMPACT OF COVID-19

BACKGROUND AND MOTIVATION

The novel coronavirus outbreak which originated in Wuhan,China has grasped the entire world. In fact, it has reached every corner of the globe and has brought the world to a standstilll. As it has been confirmed in large number of countries the World Health Organization (WHO) has declared COVID-19 as pandemic.

Right now, being the only topic that has captured everyone's heart, mind and soul our team would like to understand the impact of it on different domains. Undoubtedly, it is much more than a health crisis and is creating devastating social, economic and political crisis. This inspires us to know when the countries are likely to be safe again. Also, understanding the cause and drift of the virus on different people is also one of the major concern that everybody is curious to know about. Lastly, we would like to understand the trend of different tweets due to the spread of coronavirus in Twitter to see the opinions and experience of people around the world.

PROJECT OBJECTIVES

Objective 1: Visualizing and analyzing the patient medical data

Objective 2: Prediction when the country is likely to go out of danger

Objective 3: Understanding the trend of COVID-19 in Twitter

DATASETS

1. Patient Medical Data for Novel Coronavirus COVID-19
(Medical records of patients infected with novel coronavirus COVID-19.This data was imported and made computable on May 11, 2020.)
Dataset can be downloaded here: *[Patient Medical Data for Novel Coronavirus COVID-19] (https://datarepository.wolframcloud.com/resources/Patient-Medical-Data-for-Novel-Coronavirus-COVID-19)*

2. Novel Coronavirus 2019 Time Series Data on Cases
(Sourced from the upstream repository maintained at Johns Hopkins University Center for Systems Science and Engineering (CSSE))
Dataset can be downloaded here: *[Novel Coronavirus 2019 time series data on cases] (https://github.com/datasets/covid-19/blob/master/data/countries-aggregated.csv)*

3. Twitter Data
We will work on the most recent dataset aggregated from *Twitter* using tritteR and rtweet libraries within a particular timeframe.

DESIGN OVERVIEW

Objective 1 We will compare the age distribution between male and female patients using one of the Chart types. Plot the sequence of disease caused events taken place between different age groups. Map visualizations of the locations the person resides.Use regression with categorical variables to predict the likelihood of death of a patient with the history of chronic diseases.
Objective 2 Granger causality tests for confirmed, recovered, deaths. Forecasting some countries with the help of naive method or ARIMA.
Objective 3 Perform Sentimental analysis on the retrieved tweets and plot the trends from the specific locations.

TIME PLAN

TASKS	RESPONSIBILITIES
Brainstorming	All team members
Objective Formulation	All team members
Implementation of Objective 1	Shivani, Poornima
Implementation of Objective 2	Anne, Maneendra, Farhaan
Implementation of Objective 3	Poornima, Shivani
Create Project Website	Farhaan, Maneendra
Create Project Screencast	Shivani, Anne, Poornima
Final Project Presentation	All team members