Project Proposal

Project Title: Predicting Covid-19 Transmission Risk At Certain Establishments Through Analyzation of Various Indicators

Author: Daiwik Pal

Date: 11/10/2020

Project Definition:

The aim of this project is to provide people with an easy tool to gain insight on Covid-19 and its risks on a local, state, national, and international level. This will be done through a webapp that utilizes a chatbot to quickly and effectively disperse vital information about COVID -19 on demand. Also, the project aims to design a device and system that is capable of warning users of potential COVID-19 transmission risk at specific establishments by analyzing a number of indicators. Overall, the project's goal is to reduce the spread of COVID-19 by providing the public with vital information on a specified local scale, so they can take the necessary precautions to eliminate it .

Background:

Beginning in December of 2019, the novel Coronavius spread across the world rapidly. Specifically, cases in the United States increased swiftly in mid-March of 2020. Over five months, due to effective quarantining and quick action taken by officials, cases were slowly lowering. However, now, as the country opens back up the threat of contracting the virus still seems to grow day by day. In fact, as of October 30th, 2020, more than 9 million Americans have contracted the Coronavirus, and that figure grows by the thousands each and every day (*COVID-19 Map*, 2020). On top of that, with a mortality rate of 70.47 per 100,000 population in the United States, Coronavirus has caused the death toll to rise to 231,000 people, mainly elders. (*Mortality Analyses*, 2020) Unfortunately, many hospitals in COVID-19 hotspots across the United States may not be able to treat future cases because hospitalizations have increased by 40% in one month of reopening alone. Many hospitals in the southern US - especially in states that are growing hotspots - are projected to approach 90% capacity by October, 2020, which only worsens the condition (Rio & Bogel-Burroughs, 2020). On top of that, growth in cases affects the country's economy because of travel and import bans implemented to reduce spread of the virus. This October, the decline can easily be seen when looking at DOW's decreasing by 650 points due to the coronavirus (Oh, 2020). Moreover, on a more personal level, a 2016 paper found that when individuals who have contracted a viral disease were more likely to lose money due to not being as productive as possible after recovering, not only losing money themselves, but also losing money for their company (Adda, 938). To prevent these horrible cultural and economical effects, COVID-19 cases need to be lowered and eventually eliminated. Without a viable vaccine upon the horizon, the main way to prevent cases is through effectively dispersing all information relating to COVID-19 and urging take the necessary precautions to eliminate preventable cases.

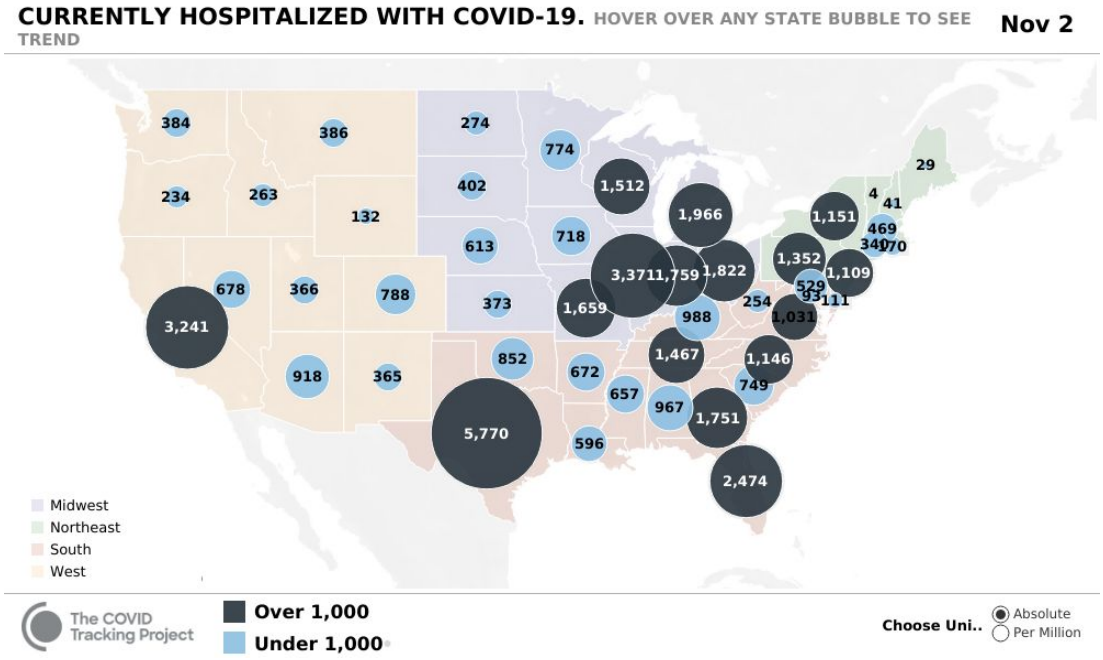
The goal of this project is to create a webapp that enables people to easily get specific information relating to Covid-19 in their locale. This will be achieved with an AI powered chatbot and a system that monitors public establishments(supermarkets, malls, theaters, universities, etc) and gives users an estimated transmission risk. Because most public establishments are accessed through the use of cars, population density at these establishments will be estimated through analyzing the parking lots.

Concepts:

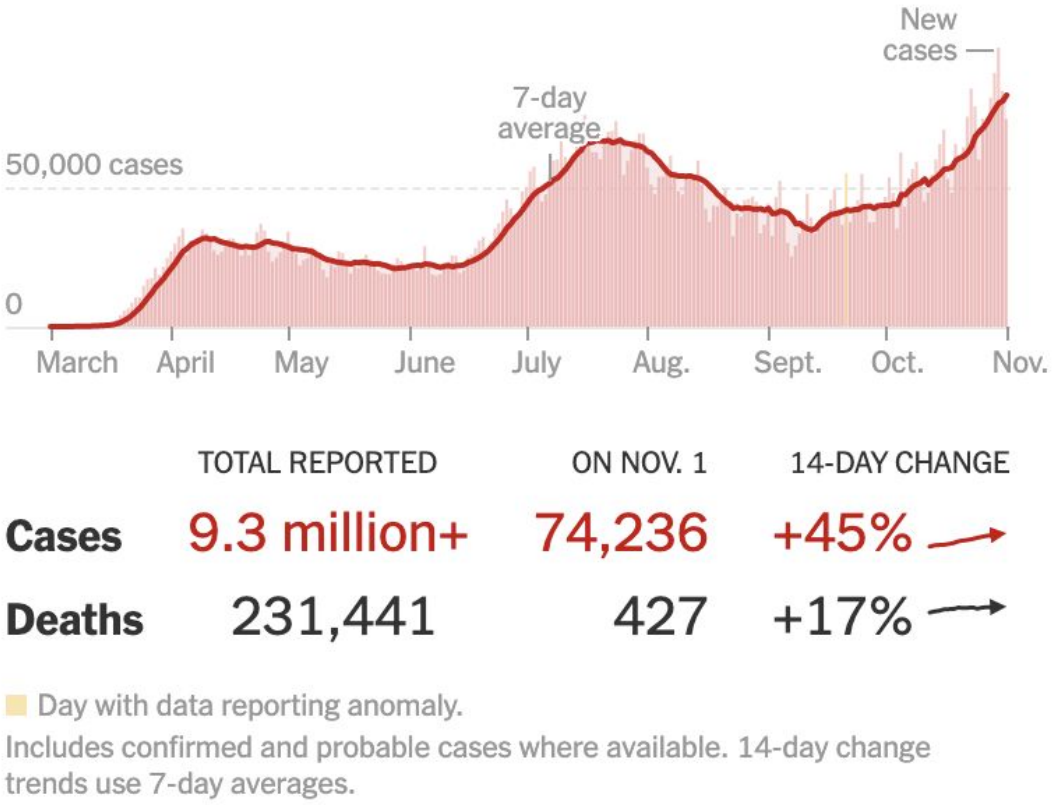
To create this system, I plan using a multitude of machine learning and deep learning concepts. Firstly, this project will analyze parking lots to estimate human density in an establishment. To do this, I would need object detection software that runs on a decentralized image processing system which will use deep learning and neural networks to identify all the cars in the image. Specifically, a You Only Look Once (YOLO) based detection system or MobileNet. Second, the other component of this system will predict potential transmission risks at particular establishments. This will use concepts of machine learning and regression to learn off of a data set that I plan on building. The information dispersal system, which includes the webapp and chat bot will again use concepts of web design/ graphic design and deep learning, respectively.

Skills:

I believe that my knowledge with computer science, data science, and calculus (main components of machine learning and AI) will allow me to understand and engineer the various machine learning and deep learning algorithms that this system requires. Also, to create the information dispersal system, I already have a background with HTML/CSS to build the web app and data retrieval using javascript. I also have experience working with linux computers and Raspberry Pi Computers so setting up the servers and related components would be very similar to that. Overall, the skills required for this project would be programming (with Python and Javascript); familiarity with deep learning, machine learning, and AI; web design with HTML and CSS; and technical IT skills to create/set up the cameras and servers.



Source: (*Currently Hospitalized by State*, n.d.)



Source: (Times, n.d.)

Experimental Design:

The information dispersal system for the Novel Covid-19 Virus has three main parts: data gathering section, data processing, and data dispersal.

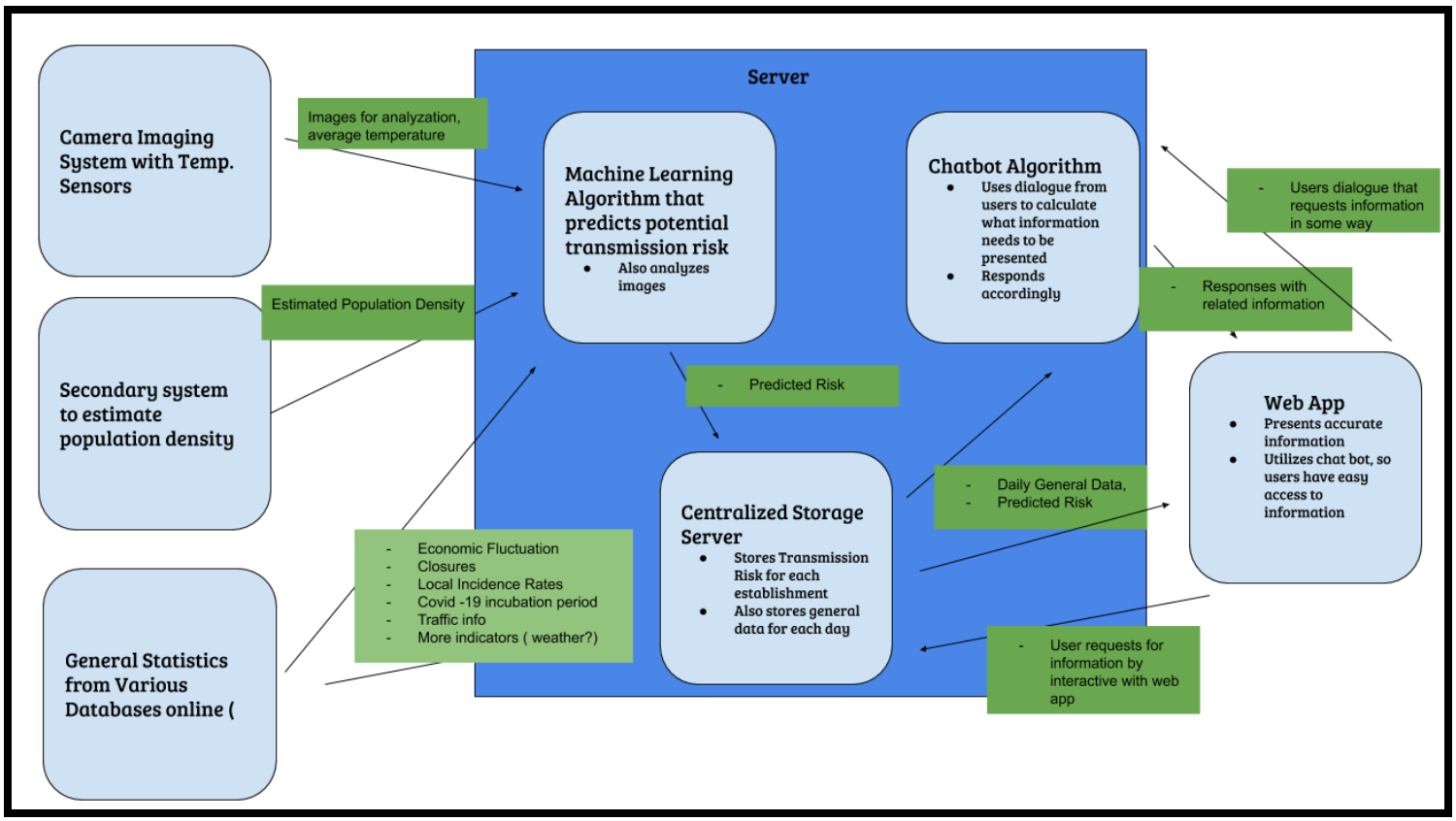
The data gathering section refers to the process where the system gathers all information it needs to reliably predict transmission risks at different establishments. This system collects information directly from the establishment's parking using cameras. These cameras are powered by Raspberry Pi computers and will analyze the images by utilizing an object detection algorithm. This algorithm will be run using Python and will most likely be similar to the You Only Look Once object detection algorithms. The images will be analyzed to count all the parked cars in the parking lot, and then an algorithm onboard the Raspberry Pi will estimate the population density. Also, the cameras will be fitted with a temperature sensor. Since these cameras will be independently powered, they will also be equipped with Solar Panels. Overall, the camera system will provide the algorithm that predicts transmission risk with the temperature and population density of the establishment.

The second part of the data gathering system is online databases. From these databases, the machine learning algorithm that predicts transmission risk will gather information on local, state, and national incidence rates, traffic info (on average from where people are coming into the general area), school closures, COVID-19 incubation period, and average weather data. These are the main indicators I am going to use because a 2016 paper analyzing the economic effects of various viral diseases also used similar indicators to this, and their results showed how these indicators were clearly related to transmission risk.

After all the information is gathered, the algorithm that predicts transmission risk at a specified establishment will be able to give the user an estimated risk they have of catching the virus. In the beginning, this may just output a rating of low, medium, and high, but as the algorithm gets more refined it may be able to estimate a percentage.

This predicted risk will be stored in a server and updated every minute. It is from this server, when the webapp and chat bot will gain their data to disperse to the public. The webapp will give users an easy to use system where they simply enter their desired location that is supported by the system, meaning they have the camera system to estimate population density and temperature, and they will get that establishment's predicted risk. The chatbot on the other hand serves the purpose of allowing users to easily find information about the virus. Most websites are convoluted and hard to navigate due to the plethora of information. This chat bat will simplify the users experience to obtain information and allow them to make more informed travel decisions.

Software System Diagram



Materials List

	Quantity	Component	Price
1)	1	Raspberry Pi 3 Model B	Already Owned
2)	1	Computer (Mac-Book Pro 2020)	Already Owned
3)	1	Pycharm or other Python IDE	Already Owned (School License from WPI)
4)	1	Temperature sensor	Already Owned (Compatible with raspberry pi)
5)	1	Camera Sensor (Webcam)	$30 to $60
6)	1	Rechargeable Batteries	$80
7)	1-2	Solar Panel(s)	($27 to $60)*2
Totals:	7-8	n/a	$137-260

Risk/Safety Concerns:

Even though this project relates to the coronavirus, it is solely an engineering and computer science project. As presented in the Experimental Design section, this project only analyzes data and predicts the transmission risk without any physical interaction with the human user. The only risks that could arise with this project is if the algorithm predicts an erroneous transmission risk and gives users a false sense of security. Luckily, the aim of this project is to provide the most accurate prediction as possible so something like that happening would be unlikely. Overall, this project has extremely low risks because it is an engineering/computer science project that predicts transmission risk solely through data analysis without any biological factors.

Data Analysis:

The first way I plan on testing the device is by using the parking lot analysis and risk predicting systems on an actual shopping plaza in my town. The object detection algorithm used to detect cars in parking lots will be tested using the CNRPark-Patches parking lot images used by researchers at (inert college) to test their decentralized vs centralized parking lot analysis systems. This test set will be the most beneficial because it has images of vehicles from multiple orientation and heights, allowing for greater accuracy. Second, the algorithm that predicts how many people are at different establishments will be tested by manually counting how people go into each building from the parking lot and comparing it to what the algorithm initially predicts. The algorithm can then be tweaked from this information to make it as accurate as possible. After both the object detection algorithm and population density algorithms are perfected, the main machine learning algorithm that predicts potential risk of visiting a specified establishment will need to be tested. Unfortunately, this algorithm will not have a clear margin of error to measure how accurate the algorithm is. Instead, this algorithm will be tested on historical data

collected from the World Health Organization. (add specifics) Finally, the information dispersal system, a web app that utilizes a chat bot, will be tested through extensive usage.

Potential Roadblocks:

One potential roadblock is that I do not get approval from the town to test my system on an actual parking lot/ shopping plaza. If that happens, I can test the device using simulations and virtual parking lots. However, this can only be utilized to test the machine learning models and algorithms, as the physical components (cameras) of this project cannot be tested virtually. To test the models virtually, I would need to simulate an establishment with a parking lot. it can be created literally through 3D-Modelling software, or it can be done by just emulating the information(images and temperature) that the sensors capture at the parking lot. Other data utilized by the models will be found in databases, so that part does not need to be simulated. After acquiring this information from the simulated parking lot, the models will function normally and can be refined upon as outlined in the data analysis section. In this method, I would still be able to test the website and servers since all of the information being presented will be the same; the only difference is that the establishment is emulated.

References: (In APA Format with in-text citations):

Adda, J. (2016). Economic activity and the spread of viral diseases: Evidence from high frequency data. *The Quarterly Journal of Economics*, *131*(2), 891–941. https://doi.org/10.1093/qje/qjw005

COVID-19 Map. (n.d.). Johns Hopkins Coronavirus Resource Center. Retrieved November 4, 2020, from https://coronavirus.jhu.edu/map.html

Currently hospitalized by state. (n.d.). The COVID Tracking Project. Retrieved November 4, 2020, from https://covidtracking.com/data/charts/currently-hospitalized-by-state

Mortality analyses. (n.d.). Johns Hopkins Coronavirus Resource Center. Retrieved November 4, 2020, from https://coronavirus.jhu.edu/data/mortality

Mortality risk of covid-19—Statistics and research. (n.d.). Our World in Data. Retrieved November 4, 2020, from https://ourworldindata.org/mortality-risk-covid

Oh, W. W., , Sunny. (n.d.). *Dow ends 650 points lower as rising COVID-19 cases, stalled stimulus efforts highlight recovery jitters*. MarketWatch. Retrieved November 4, 2020, from

https://www.marketwatch.com/story/dow-futures-fall-260-points-as-covid-19-cases-surge-stimulus-remains-stalled-11603712231

Rio, G. M. N. del, & Bogel-Burroughs, N. (2020, October 30). 'At capacity': Covid-19 patients push u. S. Hospitals to brink. *The New York Times*. https://www.nytimes.com/2020/10/23/us/covid-hospitalizations.html

Times, T. N. Y. (n.d.). Massachusetts covid map and case count. *The New York Times*. Retrieved November 4, 2020, from https://www.nytimes.com/interactive/2020/us/massachusetts-coronavirus-cases.html

Timeline: (with action steps identified- sub-deadlines will continue to evolve):

Rough timeline of major phases. As these phases get established, specific tasks under these phases will be defined further.

Task #	Task	Time
1	Research on Models about viruses, object detection system, and data dispersal systems specifically (2 -3 papers each)	9 Nov - 22 Nov
2	Create Data for Machine Learning Model for Transmission risk (assuming that I can already obtain Population density information)	23 Nov - 29 Nov
3	Create Model Transmission risk model	30 Nov - 13 Dec
4	Refine Model as much as possible	14 Dec - 20 Dec
5	Create Object Detection system for Parking lots	21 Dec - 3 Jan (december break)
6	Refine the algorithm	4 Jan - 13 Jan
7	Implement into Website	13 Jan - 27 Jan