

# Covid-19 Diagnosis Meets Machine Learning

## CSC-493 Capstone Project, Fall 2020 Covid-19 Diagnosis Tool

Introduction Data Predictive Models Decision Tree

### Objectives/Deliverables

We are aiming to build a tool that provides covid-19 diagnosis to individuals based on symptoms and can be used by people with no medical experience or doctors that cannot test all of their patients for Covid-19.

- GUI that will serve individuals that want to get quick diagnosis or orientation for their medical condition.

- Various ML classification algorithms that provide a well rounded picture of the individual condition with data regarding the forecast.

- We use a reliable dataset with over million observations that has more than 6 features.



Samridh Gupta



Senior At American University  
Data Scientist

Yon Garber



Senior At American University  
Data Scientist

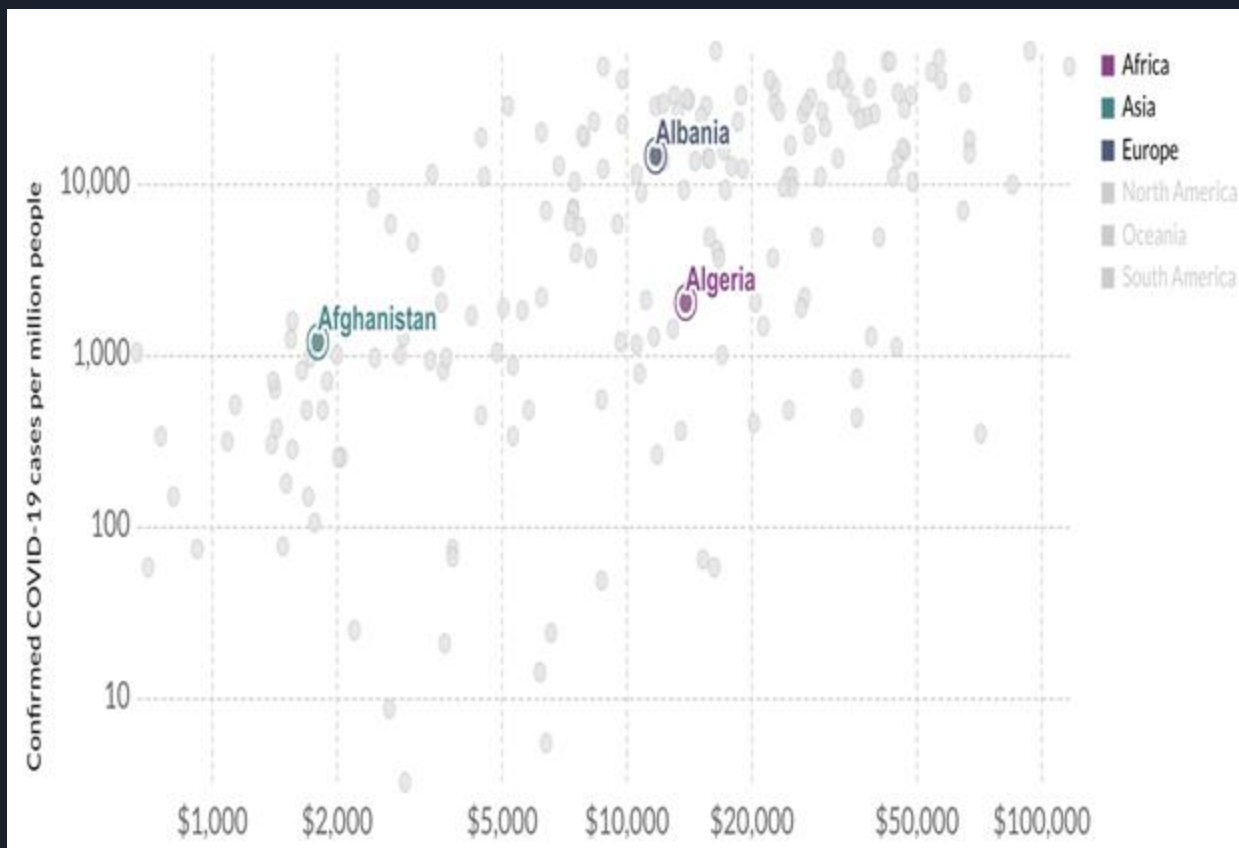
Zhenming Yang



Senior At American University  
Data Scientist

# Problem

The COVID-19 is a serious global health threat. We are living during a pandemic now. There is an urgent need to be able to diagnose infected individuals in order to stop the infection chain.



# Objective/Deliverables

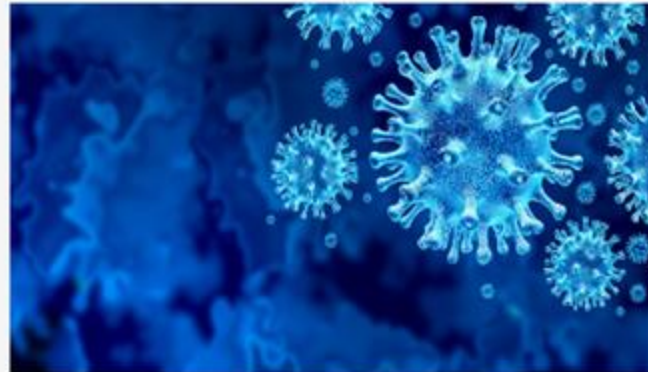
## Objectives/Deliverables

We are aiming to build a tool that provides covid-19 diagnosis to individuals based on symptoms and can be used by people with no medical experience or doctors that cannot test all of their patients for Covid-19.

~GUI that will serve individuals that want to get quick diagnosis or orientation for their medical condition.

~Various ML classification algorithms that provide a well rounded picture of the individual condition with data regarding the forecast.

~We use a reliable dataset with over million observations that has more than 6 features.



# Data

We currently have an official governmental categorical dataset with more than 1 million observations and 10 columns (see below). About 950K negative and around 50K positive.

**CSC-493 Capstone Project, Fall 2020**  
**Covid-19 Diagnosis Tool**

Introduction **Data** Predictive Models Decision Tree

Show 10 entries Search:

	X	cough	fever	sore_throat	shortness_of_breath	head_ache	age_60_and_above	result	gender	X.1
1	1	0	0	0	0	0	0	negative	0	
2	2	0	0	0	0	0	0	negative	0	
3	3	0	0	0	0	0	1	negative	0	
4	4	0	0	0	0	0	0	negative	1	
5	5	0	0	0	0	0	0	negative	0	
6	6	0	0	0	0	0	0	negative	0	
7	7	0	0	0	0	0	0	negative	0	
8	8	0	0	0	0	0	0	negative	0	
9	9	0	0	0	0	0	0	negative	1	
10	10	0	0	0	0	0	0	negative	1	

Showing 1 to 10 of 1,035,503 entries

Previous 1 2 3 4 5 ... 103551 Next



# Data Processing

The dataset we obtained was extremely unbalanced. Unbalanced data when implementing ML models especially with a categorical dataset can generate biased predictions.

For example, 95% of the data being negative and 5% Positive means that for a negative result being guessed randomly, we will be right 95% of the time.

Solution- we will randomly pick equal size of samples from the positive and negative results to overcome the unbalanced issue.