Project Proposal

1) What problem are you going to be tackling on your project?

Predict Covid-19 vaccine adverse reactions to help people make decisions.

People hesitate if they should get covid-19 vaccine. On the one hand, they want normal life back. With vaccine, they may travel with family and hang out with friends. On the other hand, they are worried about the side effect of vaccine.

We want to design a service that let users input a list of features, like medical history, age, sex…, then we will be able to:

A. Recommend the most suitable vaccine for him/her.
B. Predict possible symptoms/side effects after getting this vaccine.

2) Why is that an interesting/useful application of data science?

We all want to get back to normal life and vaccines are the most helpful way to achieve this goal. Through analyzing World Vaccination Progress Report, people can get insights of various vaccines. Also, our service will provide users with useful information for them fully preparing for possible side effects and symptoms, so they can decide if they want to get the vaccines.

Data science is a great tool to make this happen. Firstly, we can perform extensive data analysis and visualization by using various tools, such as Pandas and Matlab. Instead of the raw data, this visualization can demonstrate a clear idea of what the data means by providing visual context through graphs or plots. Users can better comprehend what current vaccine progress is and easier to identify trends and patterns.

Secondly, by training the dataset, we can get a model to predict a possible outcome given a specific instance. For example, Tom would like to know what possible symptoms he may get after getting Pfizer. He just needs to input some of his personal information, and we will be able to generate the results.

Thus, I think data science can help us mining the data and analyzing the data, and finally find patterns as well as predict certain results. It can also help us visualize huge

amounts of data to get insights of the dataset, which could provide more information than raw data.

3) What data and models are you envisioning training to address that (e.g., classification, regression, clustering)?

Three datasets we are going to work on:

Data on COVID-19 (coronavirus) cases, deaths, hospitalizations, tests --- Part 1
https://github.com/owid/covid-19-data/tree/master/public/data

COVID-19 Vaccine Adverse Event Report -- Part 2
https://www.kaggle.com/ayushggarg/covid19-vaccine-adverse-reactions

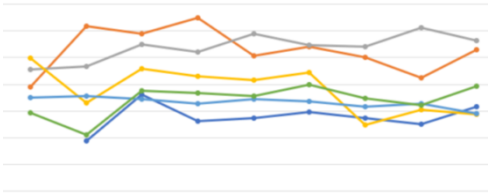
COVID-19 World Vaccination Progress -- Part 3
https://www.kaggle.com/gpreda/covid-world-vaccination-progress

COVID-19 World Vaccination Progress dataset is used for visualization purpose, which could give users insights of what current vaccination progress is.

COVID-19 Vaccine Adverse Event Report is used for training purpose, which will generate a model to predict possible vaccine adverse reactions.

We may consider using clustering models to train the data as we think it should be a unsupervised learning.

4) What will a user interface that packages your model(s) look like and how will you make it user-friendly for someone to leverage your work?



What vaccine are you interested ▼
Johnson and Johnson
BioNTech
Moderna

Please select any past medical history
Diabetes
Hear problem

Current medication
___

Allergies to medications/food/other
___

Generate report for me

Report for Tom	
Recommend vaccine	Moderna
Possible reactions	Nausea
	wheezing
	headache