Project Proposal Report

❖ **Team:**
202018009 - Drumil Shah
202018030 - Himani Desai
202018053 - Mansi Thakkar
202018058 - Dhaval Vachheta

❖ **Problem Statement**
Machine Learning based prediction of covid-19 diagnosis based on symptoms.

❖ **Background**
The recent Coronavirus Disease 2019 (COVID-19) pandemic has placed severe stress on healthcare systems worldwide, which is amplified by the critical shortage of COVID-19 tests.We believe that the ability of predicting COVID-19 patients based on clinical variables and using an easily accessible computational model would be extremely useful to address the widespread lack of testing capabilities for COVID-19 worldwide. Because many countries and hospitals are not able to allocate sufficient testing resources, healthcare systems are deprived of one of their most effective tools for containing a pandemic: identification of case hotspots and targeted action towards regions and specific individuals with the disease.

❖ **DataSources**
The links for the datasets that we are going to use in our project are as follow:

1. https://www.kaggle.com/hemanthhari/symptoms-and-covid-presence
2. https://www.kaggle.com/saichandrareddy/covid-dataset-symptoms-own-preparation
3. https://www.kaggle.com/zhiruo19/covid19-symptoms-classification

❖ **Description of the Tools We Plan to Use**

We have planned to use Google Colab as the platform for the execution of code. The language to be used is Python and the packages that we are going to use are pandas, scikit learn, sklearn, numpy, scipy, matplotlib etc.

❖ **Evaluation Strategy**

To get proper prediction we have to use multiple models against the same dataset and each of them will have their respective accuracies by which we could confirm a proper Machine Learning model. Normally we can apply classification algorithms in which we could train the model for a patient to be COVID-19 positive or negative based on the symptom but there are few cases where patients turn positive even when they are asymptomatic so, we had thought of applying clustering algorithms if required or we can treat that data as outliers. So the basic idea is to apply various machine learning algorithms and observe the results.