Project Proposal
CS 6140:Machine Learning
Group 3
Arpit Semwal, Himanshu Sharma

Background:

Machine learning is still far from replacing doctors, but it can certainly help the government make better health care decisions. Being intrigued by the subject of finance and economy and considering the health care sector is one of the fastest growing sectors regardless of the current economic turmoil, the health care system directly contributing to public health which directly affects the economy is the prime motivation behind the following problem statement.

Health care facilities being increased rapidly has led to certain problems like being more decisive about what facilities each health care center needs. This being a manual task is cumbersome, time consuming and not very accurate. I believe this task can be simplified and automated using machine learning algorithms. Machine learning algorithms can analyze the data and help health care workers make better decisions about distribution of medical facilities over the country wide health care centers. This would make the health care system more efficient and robust to uncertain events like pandemic.

Many Canadians face issues while accessing healthcare. Usually there are long wait times when trying to access healthcare due to limited healthcare facilities. The problem is more prevalent in remote areas which are difficult to reach. Due to these reasons many Canadians forgo seeking healthcare at all. The healthcare demand in many areas is greater than the regional supply to find care. Data visualization and machine learning are powerful means that can be used to identify underlying patterns and provide insights. These insights can be used to identify hard to serve areas and provide hospitals with resources such as medical specialists and beds.

Data:

We are planning to use healthcare data from various sources to do our analysis. Some of the dataset that we are planning to use are:
1. The open dataset of healthcare facilities which is a public dataset provided by Statistics Canada: https://www.statcan.gc.ca/en/lode/databases/odhf.

2. Hospital Beds staffed and in operation provided by Canadian Institute for Health Information:
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ah UKEwjrsNOe8pX7AhULATQIHZ3GATEQFnoECBcQAQ&url=https%3A%2F%2F www.cihi.ca%2Fsites%2Fdefault%2Ffiles%2Fdocument%2Fbeds-staffed-and-in-operation-2020-2021-en.xlsx&usg=AOvVaw1MA_XDmKP5eZLf_jSkRX6V
3. Wait times for priority procedures in Canada provided by Canadian Institute for Health Information:
https://www.cihi.ca/en/wait-times-for-priority-procedures-in-canada

All of these datasets represent various aspects of healthcare needs in Canada like the location, number of hospital beds and wait times at healthcare facilities. This information will help us to come up with actionable insights to improve the overall healthcare in Canada.

Objectives:

We are planning to build a clustering model to automatically cluster the healthcare facilities into groups and identify the locations where we need more beds, doctors, etc. This will help healthcare workers and the government to be more decisive about distribution of facilities countrywide and meet the growing demand of healthcare needs, which further leads to economic growth.

Action Plan:

Till now we have identified the public datasets that we will be using to train the clustering model. We will start with the initial data analysis including cleaning of the data, merging all the relevant datasets together and other preprocessing tasks. We will start working on the modeling part after the preprocessing stage. We are planning to experiment with different clustering algorithms like k-means and Gaussian Mixture Models and select the one that provides best results. The final goal is to group the healthcare facilities into different clusters depending on the current healthcare facilities data.