**Problem Identification - (Second Capstone Project)**

Hypothesis:

What are the possible ways for most people to identify their percentage of body fat, with approximately result accuracy of 70-88%, by avoiding the expensive and complicated procedures, through a few simple steps of measuring their body components and using a well-developed software application by the end of June 2021.

Context:

Around 70% of people in the US are facing health conditions due to obesity, based on the credited resources in 2021. Those people are trying to fight that nasty battle by finding the solution to control it. But what is the best way to control it? The most convenient one here is the knowledge of knowing the percentage of fat in your body. So, once you know it, you have the chance (choice) to balance it. Unfortunately, there are not many convenient, simple, and cheap ways to do so. That is why this project is targeting specifically those types of people. It's about giving them hope, a proper opportunity, and an option to fight this battle with appropriate tools (in our case it's software application). This software application should be able to predict body fat with an accuracy of 70-88% with a simple step of measuring your body components. Also, it will help to save a few hundred dollars (based on the regular medical procedures) on average and time (spent on transportation, documentation/application parts, waiting results, and other required things). The full functionality of this application should be ready by the end of June 2021.

Criteria for success:

The software application is fully functional by the end of June 2021. It predicts the percentage of body fat with an accuracy of 70-88%. It requires only the measurement of body components for its successful work (result).

The slope of solution space:

Most time the focus will be spent on optimizing the accuracy of the working model (application). It will include the reduction of the possibility of occurring the resulting errors based on the provided data within a specified scope of features (body measurement components). Moreover, the focus of work will be directed to the use of the relevant, most accurate equations of finding the percentage of body fat and other feature (variable) ratios.

Constraints within solution space:

Limited data - some information could be misrepresented, some could be missing. In other words, it could bring issues with the loss of important (necessary) information. The valid and reliable data could help to collaborate the working model (application) accordingly.

Time - there is a chance that there will not be enough time until the working version of this application is presented due to some not well-known future complications (circumstances).

Limited personal - there could be the need for advice of professionals in special working industries like health, biological, science departments.

Stakeholder to provide key insight:

In our scenario, the position of stakeholders will take regular people. They will decide the fate of the success of the developed application and its benefits on the existing problem. Unfortunately, there will be no insights on this project due to the limited personnel.

Key data sources:

The data will be provided by the well-known website "www.kaggle.com". The direct link to the dataset is "https://www.kaggle.com/fedesoriano/body-fat-prediction-dataset".