TITLE: ATTRIBUTES SELECTION
I. INTRODUCTION

Studies have shown the side effect of having abnormal body fat on people’s health [1]. This meant higher risks of heart disease, high blood pressure, gallstones, type 2 diabetes, osteoarthritis, and certain cancers. Hence, concerned people needed to keep track of their body fat, mostly nowadays by using one of the electronics in the market that measure the fat percentage [2]. It’s true that body fat can be alternatively estimated using a scale and measuring tape [3], yet this project is intended to reduce these variables used in the estimation. The reduction is hoped to lead to the minimum set of features that still produce accurate estimate of body fat. In other words, this helps people to focus more on improving these features in their bodies to accelerate the reduction of body fat.

II. STATEMENT OF PROBLEM

Having people take into concern many characteristics to look after in order to maintain normal body fat can be frustrating and misleading. This project is aimed to reduce these attributes into a minimal set which lead to two major benefits. First, it helps people better focus on specific attributes to maintain their body fat. Secondly, show a clearer view of the major factors contributing to body fat which will further guide medical researches into more studies about these attributes. Twenty years ago, body fat took into concern 19 factors for fat estimation, yet these days, home tests can be simply made based on small number of attributes – as low as 7 characteristics.

III. PROPOSED PROJECT AND PURPOSE

This project is based on data collect by Dr. A. Garth Fisher, Human Performance Research Center, Brigham Young University, Provo, Utah (1985) [4]. Each record present a value of body fat, age, weight, height, and ten body circumference measurements (e.g., thigh) for 252 men. The study took in concern one gender due to the major difference of the fat percentage for females than males. The goal is to reduce these 19 attributes to the least set using methods taught in data mining course.

Basically, attributes elimination is approached in three steps. First, due to the great effect of missing values on this work, records with missing values will be deleted using na.omit() function. Secondly, variance value is used to delete data with low standard deviation. This is due to the fact that data with low variance indicate more difficulty on the prediction process. Lastly, the culsvarsel package in R is used as a final step of attributes’ selection [5]. If there is time, we might even validate our final minimal set of attributes by comparing it to other possible combinations of attributes using multiple (linear) regression.

Certainly, there are many other techniques followed by researchers for attribute selection. This includes, exhaustive best first, simulated annealing, genetic algorithm, greedy forward selection, greedy backward elimination. Yet, our choice of greedy backward elimination is due to its global clustering property and efficiency. [6]

The final report of this project will include the minimal set of attributes used in predicting the body fat as well as some regression plots used in this study.
VI. REFERENCES

[2] Bodytronics.bodytronics.com/CTGY/Body_Fat_Measurement?s=g&kw=body+fat+calculator&gclid=CMOUjq3Fq5ICFQijPAodeS5wSQ