# Executive Summary of Life Expectancy Data Set

This Project is highly useful for decision-makers (individuals or a company) who want to predict the fetal health of a new patient, or an individual given the information on the respective predictors used in the project.

The two key questions that might be important to the decision maker are, i) Are the variables predicted important in the data set key predictors for predicting the fetal health before in hand or do we need some other criteria to alleviate the risks for pregnant women? ii) Is the data balanced across the globe i.e., does it have equal amount of data related to all the races and all the ages?

Three key findings relevant to the decision maker from the project results include i) The precision and accuracy are higher for sophisticated and computationally complex modeling techniques than the base models having simple computations, for example, boosted tree has accuracy of 92.83019, and Random Forest has the accuracy of 92.83019. ii) There are some non-zero or zero variance producing predictors that are removed form data set to increase the accuracy of the models and its predicting precision, the variables are severe\_decelerations, prolongued\_decelerations, and percentage\_of\_time\_with\_abnormal\_long\_term\_variability. However, the decision maker should decide whether it would make sense to include those relevant to their research question. iii) The dimensionality reduction can be done if necessary if the data is huge, and the code can be run in parallel threads. However, including Principal components as predictors might further reduce the interpretability of the sophisticated models such as Random Forest, Boosted Tree, Support Vector Machines, etc.

The variable importance as predicted by the models though helpful, sometimes have been un-consistent, in the sense that it has rated some variables lower which might be good predictors. So, I would recommend taking those predictors too into consideration based on individual relevancies, also I think some predictors such as ones related to nutrition, previous abortions, family history of certain problems related to PCODs, and other gynecological factors too into consideration while predicting the fetal health. The data does not contain any information on the age of the women, and the factors that would inform us that the data is taken from across the globe. Thus, one should take mindful analysis of what is needed for their research question and the sufficiency of the dataset.