Reflective Statement

First of all, I would like to thank our Lecturer: Mr. Abubakr Siddig for giving us this assignment. The assignment has given me chance to explore the research work on the different algorithms for Machine learning to gain some meaningful insight. During the process of research, I was facing a challenge in finding the appropriate topic for which I can understand the algorithm. Initially, I selected a mental health prediction topic but not much paper was available on that topic then I researched a few more topics for data mining and prediction and shortlisted 2 of them after which we found our current topic more relevant and I was confident enough to understand the different concept and write a research paper on it. I was also able to find relevant research work on the topic in the past from authentic publishers so we moved forward with it.

The most frustrating part was in the bibliography section to arrange the references. At the end of the writing process, we can say that we have knowledge of different types of algorithms. And it was good learning for us also we came to know about the importance of the optimum information retrieval system as a must in today's world of big data.

While doing the assignment, we were not having in-depth knowledge of python library such as pandas, NumPy, seaborn. So we started to learn python first and gain some knowledge on it. After that, we started to find a resource for above stated lib python. We have also. Gained knowledge on the library of sklearn which Is used for training and testing of the different ML learning model. Because of the sklearn training was just like cake walk.

We have select the data from Kaggle website on heart disease prediction the data had 14 attributes Along with the predictor attribute in the binary form. We have decided to have 3 hypotheses and we have used different types of the algorithm for test their accuracy and comparing it with each other. We also have split the data into. 80:20 ratio. Of training and testing set. Initially there were 1000 plus rows but almost the were 700 duplicate row so we decided to remove the duplicate rows. And then proceed with the other activity like scaling the data continuous data for linear model such. As LR, SVM and KNN also. Used the dummies function provided by the Sklearn library to get the dummy value for the different categorical attributes.

Overall I can say that this assignment was a good learning experience apart from the machine learning models. I got to learn python programming seaborn visualization which is clear to me now. The most important learning for me was to co-ordinate in the team and get the work done which I think me and my partner did.