CONCLUSION:

The overall objective of our project is to predict accurately with less number of tests and attributes the presence of heart disease. In this project, fourteen attributes are considered which form the primary basis for tests and give accurate results more or less. Many more input attributes can be taken but our goal is to predict with less number of attributes and faster efficiency to predict the risk of having heart disease at a particular age span. Five data mining classification techniques were applied namely K-Nearest Neighbor, Naive Bayes, Decision Tree, Random Forest & Logistic Regression. It is shown that Random Forest has better accuracy than the other techniques.

This is the most effective model to predict patients with heart disease. This project could answer complex queries, each with its own strength with respect to ease of model interpretation, access to detailed information and accuracy.

This project can be further enhanced and expanded. For example, it can incorporate other medical attributes besides the 14 attributes we used. It can also incorporate other data mining techniques, e.g., Time Series, Clustering and Association Rules. Continuous data can also be used instead of just categorical data. Another area is to use Text Mining to mine the vast amount of unstructured data available.

This project is presented using data mining techniques. From logistic regression,

KNN, Naive Bayes, Decision Tree, Random forest are used to develop the system. Random Forest proves the better results and assists the domain experts and even the person related to the medical field to plan for a better and early diagnosis for the patient. This system performs realistically well even without retraining