Heart Disease Machine learning

Data Science – Part Time Final project

By:Mariam Alsaeed

Overview

- Heart disease describes a range of conditions that affect your heart.
- Cardiovascular disease symptoms may be different for men and women. For instance, men are more likely to have chest pain; women are more likely to have other symptoms along with chest discomfort, such as shortness of breath, nausea and extreme fatigue.
- heart disease is the leading cause of death in the United Kingdom, United States, Canada, and Australia. One in every four deaths in the U.S. occurs as a result of heart disease.

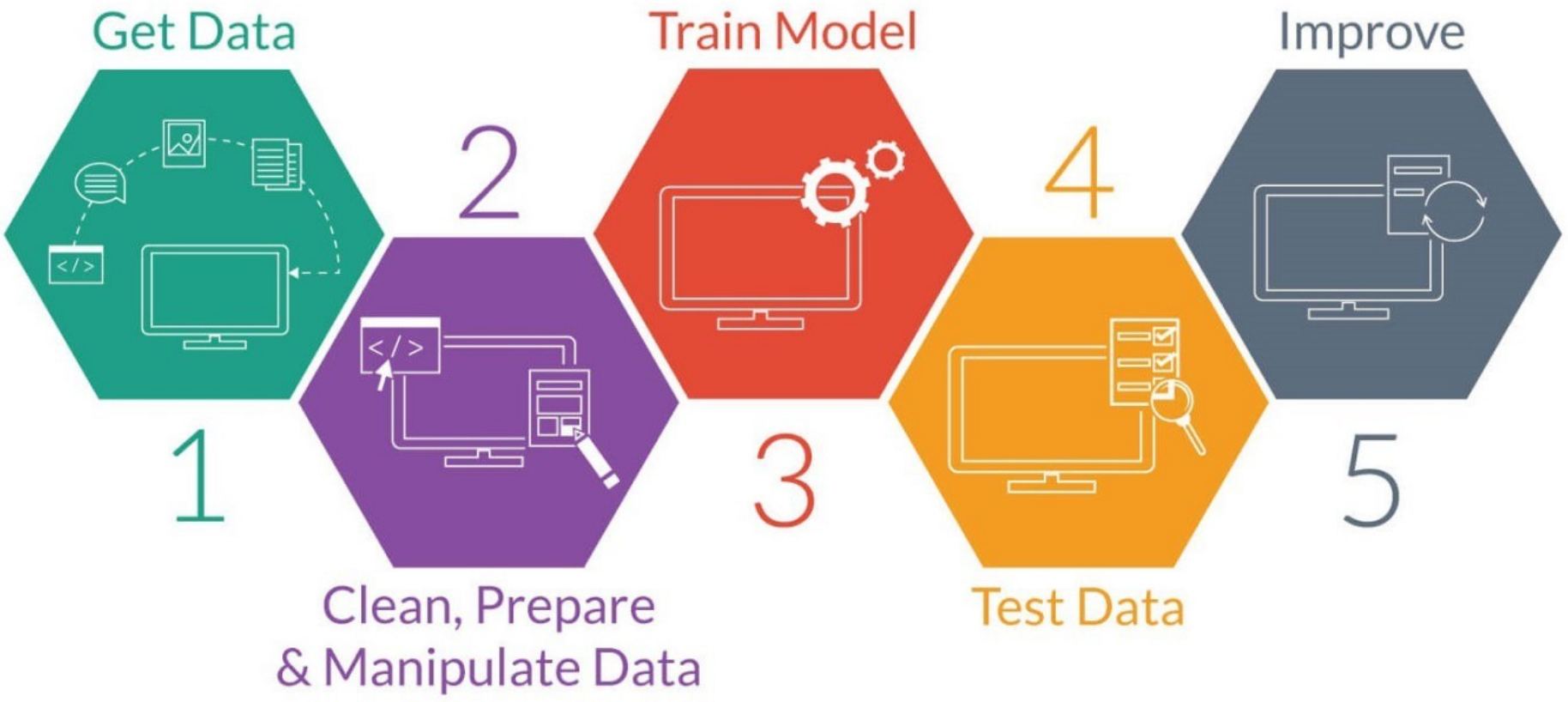
Machine learning

- Machine learning (ML) proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry,hard to predicted manually.
- We will applying Machine Learning approaches(and eventually comparing them) for classifying whether a person is suffering from heart disease or not, using Cleveland Heart Disease dataset from the UCI Repository.

What is our data?

- Our data set consist from 303 rows and 14 columns
- Age, sex, chest pain type, resting blood pressure, cholesterol, fasting blood sugar, rest ecg, 'max heart rate achieved,exercise induced angina, 'st depression, st slope, number major vessels, thalassemia and our target thet person have heart disease or not.

The process



- Source: https://machinelearning-blog.com/2017/11/19/fsgdhfju/

ML Algorithm I used

- Logistic Regression with default parameters.

  For all of below we apply grid search to find the best parameters to our data.
- Decision Trees.
- Random Forests.
- K-Nearest Neighbors.

Results

- 1-Logistic Regression Classifier score for Heart Disease data set is 0.88 With accuracy score equal to 0.86
- 2-K-Neighbors Classifier with 10 Neighbors score for Heart Disease data set is 0.70 with accuracy equal to 0.60
- 3-Decision Tree Classifier score for Heart Disease data set is 0.81 with accuracy equal to 0.78
- 4- Random Forest Classifier score for Heart Disease data set is 0.87with accuracy equal to 0.79

Conclusion

- The Logistic Regression classifier have the best score with higher accuracy than the other classifier for our data set.

- Hope to every one good heart and good health.

- Thanks for your interest.