**HEART DISEASE PREDICTION EMPLOYING VARIOUS MACHINE LEARNING TECHNIQUES**

Heart disease has become the major cause of death throughout the earth during the latest period. As a result, early discovery and regular monitoring can help to lower the death rate. With the rise in population and illness, diagnosing illness and providing required medication immediately has become a difficulty. However, there is hope that recent technological advancements have hastened the development of improved beneficial medical specialty solutions in the broader public health sector. And the combination of streaming big data analytics and machine learning is a game-changing technology that has the potential to revolutionize the healthcare industry, particularly in terms of early identification of heart disease. This technology has the potential to be more powerful while also being less expensive. Our project seeks to investigate several data mining approaches, such as Probabilistic models like Naïve Bayes, Bagging techniques like Random Forest Classification, Decision tree, and complex Machine learning techniques such as Support Vector Machine, using a qualifying data for cardiopathy prediction that includes parameters such as gender, age, pain level, blood glucose, and so on. The project's study entails employing different data processing techniques to uncover correlations between the dataset's many features and then using the attributes appropriately to forecast the likelihood of cardiopathy. These machine learning algorithms require less time to anticipate sickness with a high degree of accuracy, which can help us save lives all across the world.

**Literature survey**

Numerous studies have been conducted in medical camps employing various data mining based approaches and ML-based algorithms to develop disease prediction systems.

K. Polaraju et al. [1] suggested a Multiple Regression Model for Heart Disease Prediction. It suggests, Multiple Linear Regression is effective in predicting the likelihood of developing heart disease. This study is carried out utilizing training data of 3000 data points and 13 different attributes indicated before. The used data is separated into 2 portions, with Seventy per cent being used for training the model and thirty per cent being used for testing purposes. Following `the results, it is obvious that the accuracy of the Regression algorithm's classification is superior to that of the rest of the methods.

Marjia et al. [2] used different methods like j48, KStar, Bayes Net, SMO and also WEKA, to produce predictions of the heart-disease. Using cross-validation techniques, SMO and Naive Bayes model outperformed the KStar-model, Multilayer Perception, and J48 approaches in terms of performance from many factors. The levels of accuracy achieved by those algorithms are still significantly lower. Consequently, the accuracy of the system is being improved, which will allow for improved and more accurate disease diagnosis in the near future. ,+However, even with those algorithms, the accuracy levels achieved are still lower. As a result, the accuracy of the system is being improved, which will allow for improved and more accurate disease diagnosis to be achieved.

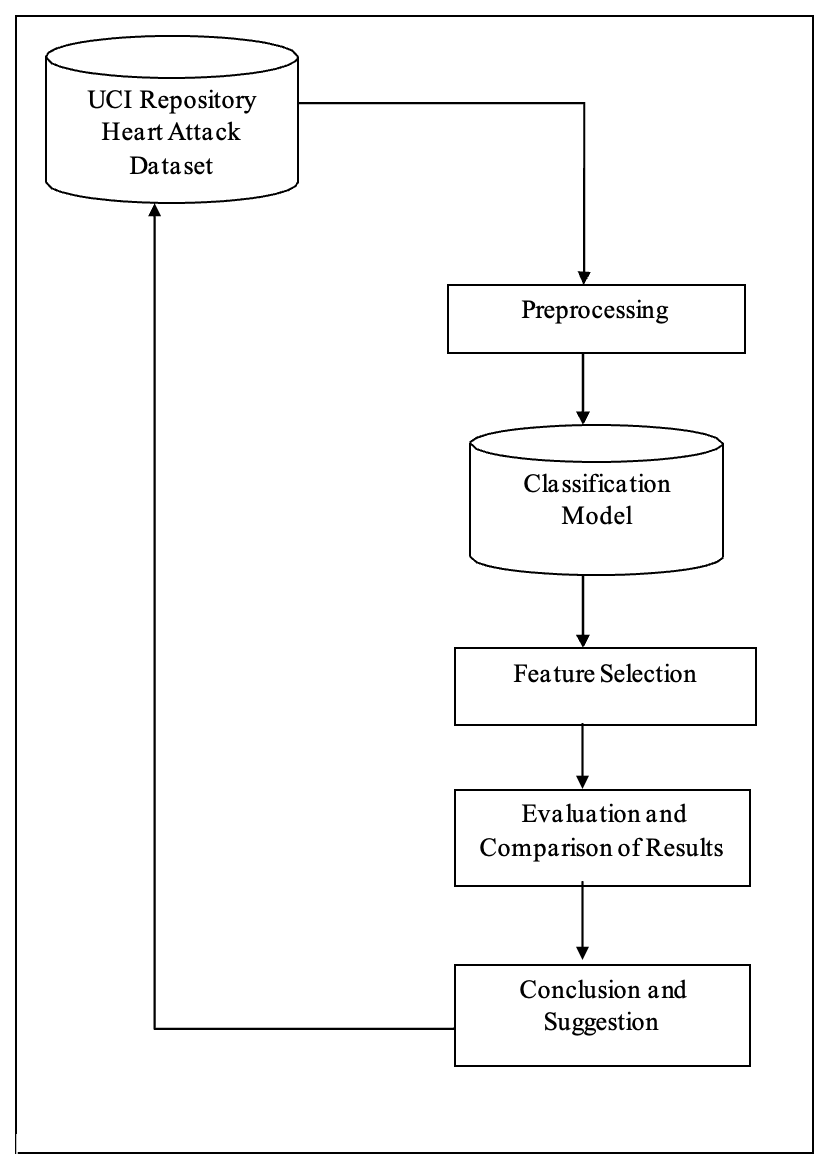
S.Seema et al. [3] utilize Naive Bayes, Decision Trees, Support Vector Machines, and also employee deep learning techniques like Neural Networks to predict chronic disease using samples collected from previous health tests. In order to measure increased performance at a more accurate rate, a comparative study of classifiers is performed. As a result of this analysis, SVM has proven to be having the greatest accuracy rate.

Different classification machine learning models such as Naive Bayes classifier, Decision Tree Classification model, K-Nearest neighbour, logistic regression, Support vector machines, and artificial neural networks were tried by Ashok Kumar Dwivedi et al, [4]. Logistic Regression outperforms other algorithms in terms of accuracy.

**​​Objectives of the study**

To determine whether or not a patient is suffering from heart disease. On a certified dataset, we utilize a variety of machine learning methods. Then you'll get to a conclusion about the relationships between completely different characteristics. We strive to create a clear plan for our predicted data, apply processing strategies, analyze the results, and compare the results of different data processing approaches.

**Research design and methodology:**



**Datasets to be used:**

The dataset was obtained from the UCI Repository, and it consists of data from 303 individuals who were subjected to different tests. These datasets, along with the 13 features, will be used to predict heart disease. The features are listed below.

* Age
* sex
* chest\_pain\_type
* resting\_blood\_pressure
* serum\_cholesterol
* fasting\_blood\_sugar
* number\_of\_major\_vessels
* oldpeak
* exercise\_induced\_angina
* resting\_electro\_cardio\_graphic\_results
* maximum\_heart\_rate\_achieved
* exercise\_induced\_anginA
* slope\_of\_peak\_exercise\_ST\_segment
* number\_of\_major\_vessels\_colored\_by\_flourosopy
* thal

**Conclusion**

The goal is to use several machine learning algorithms to forecast a person's heart-related disorders and then choose the best performing models that find the best patterns between the chosen features.

**References**

[1] Polaraju, K, et al. “Prediction of Heart Disease Using Multiple Linear Regression Model.” International Journal of Engineering Development and Research, vol. 5, no. 2321-9939, 2017, p. 1419, [www.ijedr.org/papers/IJEDR1704226.pdf](http://www.ijedr.org/papers/IJEDR1704226.pdf).

[2] Marjia Sultana, Afrin Haider, “Heart Disease Prediction using WEKA tool and 10-Fold cross-validation”, The Institute of Electrical and Electronics Engineers, March 2017.

[3] Dr.S.Seema Shedole, Kumari Deepika, “Predictive analytics to prevent and control chronic disease”, https://www.researchgate.net/punlication/316530782, January 2016

[4] Ashok kumar Dwivedi, “Evaluate the performance of different machine learning techniques for prediction of heart disease using ten-fold cross-validation”, Springer, 17 September 2016

[5] Jayami Patel, Prof. Tejal Upadhay, Dr. Samir Patel, “Heart disease Prediction using Machine Learning and Data mining Technique”, March 2017.