

# Human Heart Disease Prediction System using Enhanced Decision Tree Algorithm in Data Mining

**Er.. Pooja Verma**

M.Tech (CSE), Department of Computer Sc. & Engg. R.I.E.T Phagwara

poojaverma05@gmail.com

**Er. Parminder Singh**

A.P in Department of Computer Sc. & Engg. R.I.E.T Phagwara

hod\_cse@riet.ac.in

**ABSTRACT:** The statistical characteristics of patients have a great influence on determining the likelihood of heart disease. To determine the disease in medical diagnostics, statistical methods are most often used Data Mining, which with large amounts of information and complex relationships can give more accurate estimates, especially with a large number of similar characteristics. In this paper, we consider the task of classification data to determine the likelihood of heart disease for patients with similar characteristics. To solve the problem, the Enhanced Decision Tree Algorithm with normalization was used, and NetBeans and Weka software were chosen to implement the algorithm. The developed software module allows visualization of the algorithm.

**Keywords: EDTA, DT, WEKA**

**I. INTRODUCTION:** An analytical process which is developed to examine data inform of patterns which are consistent is known as data mining. From ages, only the physical exception of patterns from data is going. The data collection, storage, and manipulations have increased by the accretion and prevalence of computer technology. Due to the grown size as well as the complexity of datasets the direct manual analysis has amplified with the indirect and automatic processing of data. These are various methods as clustering, neural networks, other genetic algorithms, vector support machines, and decision trees which are applied to data with aim of not hiding the patterns which are hidden. Data mining is sometimes turning as knowledge discovery and its tools are here to predict behaviors and future trends making proactive business and knowledge-driven decisions.

## II. CORONARY ANGIOPLASTY

Angioplasty, an adaptable cylinder called a catheter with a little inflatable is gone through your course. The inflatable is swelled to augment your vein, a work tube known as a stent, may then be embedded into your coronary corridor. The stent pushes against your supply route dividers, squashing greasy stores against the conduit divider so blood can course through it all the more openly. The system more often than not takes around 30 minutes, yet it can take longer relying upon what number of segments of your conduit require treatment.

## III. CLASSIFICATION

Classification is the system of making a gathering of objects into lessons of analogous objects. A Classification of facts objects can be handled as one institution. While doing the Classification exam, we first separation the position of statistics into groups based totally on facts resemblance after which allow the label to the corporations. Classification divides data into tremendous or beneficial organizations (Classifications).

In this research, the various technologies of data mining (DM) models for forecast of heart disease are discussed. Data mining plays an important role in building an intelligent model for medical systems to detect heart disease (HD) using data sets of Angioplasty and Stents patients, which involves risk factor associated with heart disease. Medical practitioners can help the patients by predicting the heart disease before occurring. The large data available from medical diagnosis is analyzed by using data mining tools and useful information known as knowledge is extracted. Mining is a method of exploring massive sets of data to take out patterns which are hidden and previously unknown relationships and knowledge detection to help the better understanding of medical data to prevent heart disease. There are many DM techniques available namely Classification techniques involving Naïve bayes (NB), Decision tree (DT), Neural network (NN), Genetic algorithm (GA), Artificial intelligence (AI) and Clustering algorithms like KNN, and Support vector machine (SVM). Several studies have been carried out for developing prediction model using individual technique and also by combining two or more techniques. This research provides a quick and easy review and understanding of available prediction models using data mining. The comparison shows the accuracy level of each model given by different researchers.

#### **IV. OBJECTIVES OF THE STUDY**

The purpose of this research work to increase the efficiency of decision tree technique dataset. The main purpose of this goal is to increase the efficiency and accuracy of the decision tree dataset.

1. To propose enhanced decision algorithm which will work on enormous scope high dimensional dataset.
2. To increase the efficiency of correct classified instances with a new algorithm based on decision tree algorithm.
3. To reduces the error to the same dimensions as the quantity being predicted by using sum of square error.

#### **V. THE RANDOM FOREST ALGORITHM**

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction. The fundamental concept behind random forest is a simple but powerful one — the wisdom of crowds. In data science speak, the reason that the random forest model works so well is:

A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models.

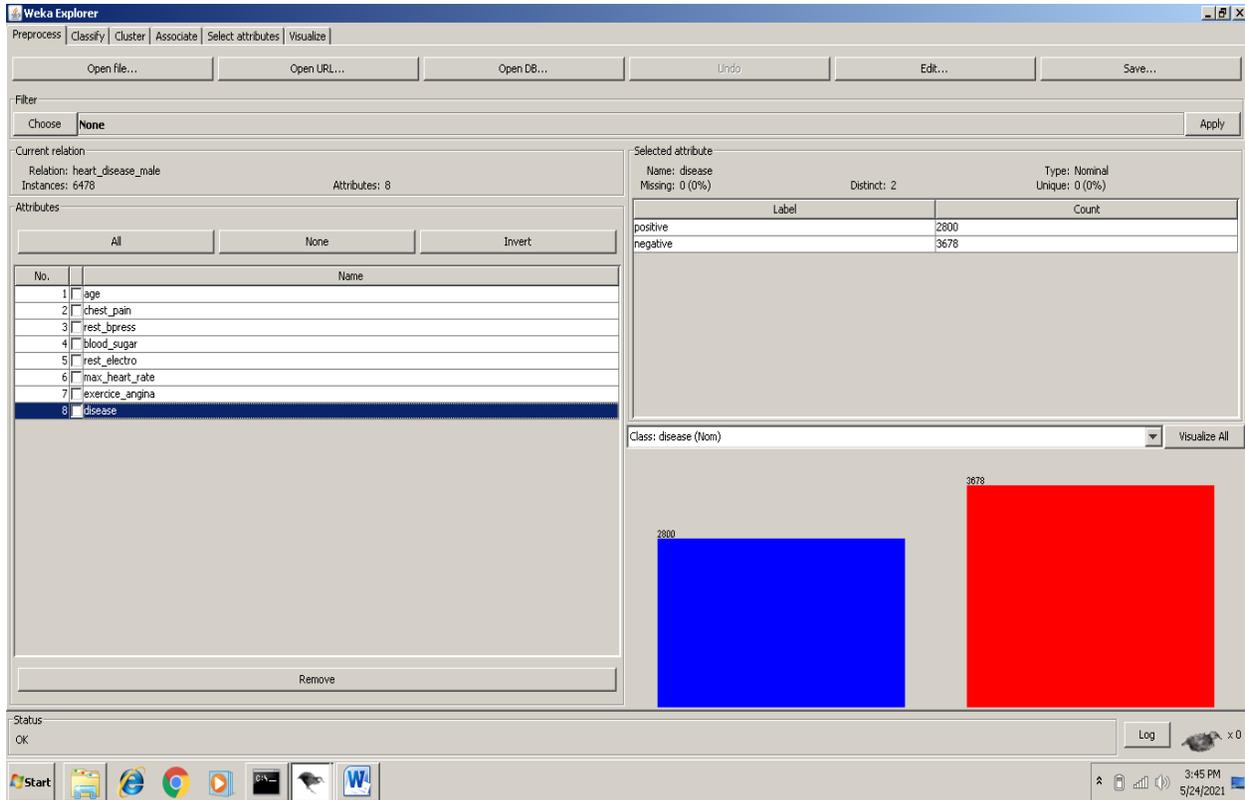
#### **VI. ENHANCED DECISION TREE ALGORITHM**

The tree starts as a isolated node as an alternative of the training samples .If the samples are all of the equivalent class, then the node becomes a piece of paper and is labeled with that class. Otherwise, the algorithm uses an entropy-based establish known as information attain as a heuristic for selecting the characteristic that will best undo the samples into human being classes.

Dataset Used

A diversity of heart disease data has to be collected from various hospitals of Jalandhar. The collected heart disease data is stored in heart disease database for further handling of data.

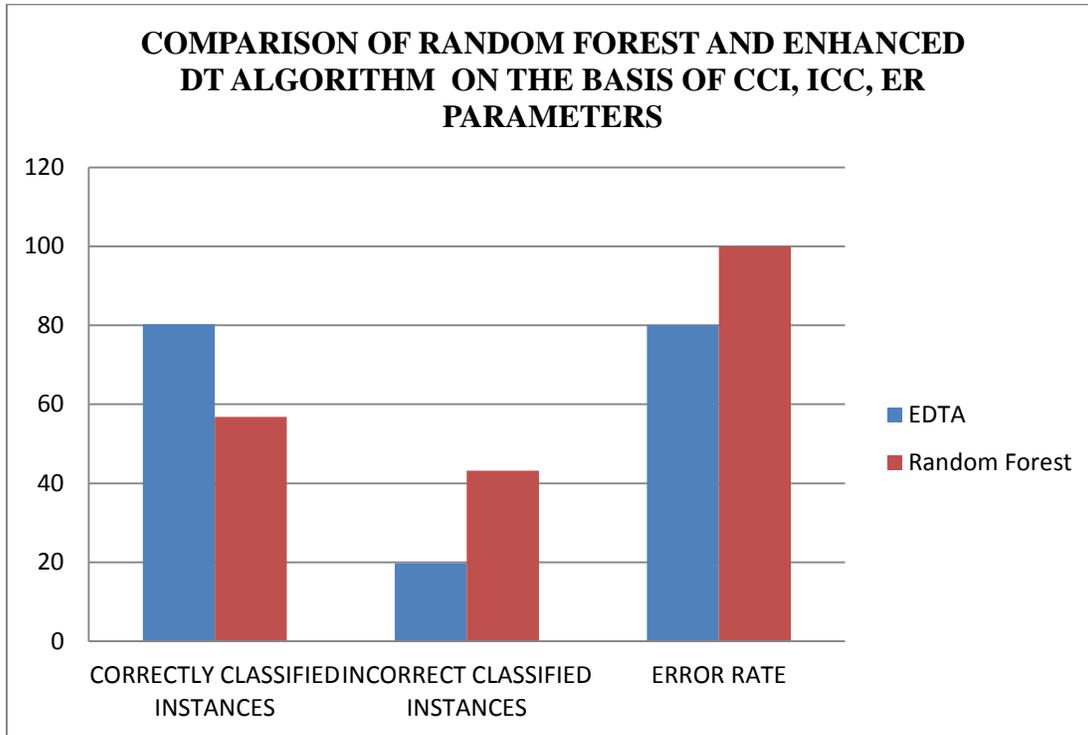
Results and Discussion



**Fig. 1.1 Shows the last attribute “Disease” is taken as a class attribute by the WEKA . This attribute contains two categories i.e. positive and negative.**

**Table 1.1 Comparison between Random Forest and EDTA on the basis of Correctly Classified Instances, Incorrectly Classified Instances and Error Rate**

<b>PARAMETERS</b>	<b>EDTA</b>	<b>Random Forest</b>
CORRECTLY CLASSIFIED INSTANCES	80.28	56.77
INCORRECT CLASSIFIED INSTANCES	19.71	43.22
ERROR RATE	80.12	100.00



**Fig. 1.2** Chart shows the comparison between Random Forest and Enhanced DT Algorithm on the basis of Correctly Classified Instances, Incorrectly Classified Instances and Error Rate.

The Correctly Classified Instances, Incorrectly Classified Instances and Error Rate esteems for Enhanced Decision Tree Algorithm is superior to other traditional Algorithm.

## VII. CONCLUSION

In this work, we presented a EDTA (Enhanced Decision Tree Algorithm) design based on machine learning technique and study is being done on EDTA classification and Random Forest algorithms.. The comparison of proposed algorithm is done with the existing algorithms Random Forest algorithm on heart disease dataset using WEKA data mining tool. The results by increasing the efficiency of the proposed method gives better performance than Random Forest algorithm and reducing the sum of square error which signifies that EDTA have high accuracy. Also the proposed algorithm can handle large datasets more effectively.

## REFERENCES

- [1] Ievgen Meniailov, Dmytro Chumachenko, Ksenia Bazilevych, “Determination of Heart Disease Based on Analysis of Patient Statistics using the Fuzzy C-means Clustering Algorithm”, August 21-25, 2020, Lviv, Ukraine.
- [2] Haolin Wang , Zhilin Huang, Danfeng Zhang, Johan Arief, Tiewei Lyu, And Jie Tian, “Integrating Co-Clustering and Interpretable Machine Learning for the Prediction of Intravenous Immunoglobulin Resistance in Kawasaki Disease” <https://creativecommons.org/licenses/by/4.0/> VOLUME 8, 2020, IEEE, 2020.
- [3] Jian Ping Li, Amin Ul Haq, Salah Ud Din, Jalaluddin Khan, Asif Khan, And Abdus Saboor, “Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare”, IEEE ACCESS, 2020.
- [4] Luiz Antonio da Ponte Junior, Rita de C´assia Alves<sup>3</sup>, Liana Catarina Lima Portuga, “Identifying Post-Traumatic Stress Symptoms Using Physiological Signals and Data Mining”, 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), 2020.
- [5] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, “Design And Implementing Heart Disease Prediction Using Naives Bayesian”, Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore, 2019.
- [6] Sarath Babu, Vivek EM, Famina KP, Fida K, Aswathi P, Shanid M, Hena M, “Heart Disease Diagnosis Using Data Mining Technique”, International Conference on Electronics, Communication and Aerospace Technology ICECA 2017.
- [7] Cincy Raju, Philipsy E, Siji Chacko, L Padma Suresh, Deepa Rajan S, “A Survey on Predicting Heart Disease using Data Mining Techniques”, Proc. IEEE Conference on Emerging Devices and Smart Systems (ICEDSS 2018).
- [8] Chaitanya Suvarna, Abhishek Sali, Sakina Salmani, “Efficient Heart Disease Prediction system using Optimization Technique”, Proceedings of the IEEE 2017 International Conference on Computing Methodologies and Communication (ICCMC), 2017.
- [9] Mafizur Rahman, Maryam Mehzabin Zahin, Linta Islam, “Effective Prediction On Heart Disease: Anticipating Heart Disease Using Data Mining Techniques”, Second International Conference on Smart Systems and Inventive Technology (ICSSIT 2019) IEEE Xplore, 2019.

- [10] Yukti Sharma, Rikku Veliyambara, Prof. Rajashree Shettar, “Hybrid Classifier for Identification of Heart Disease”, IEEE, 2019.
- [11] Radhanath Patra, Bonomali Khuntia, “ Predictive Analysis of Rapid Spread of Heart Disease with Data Mining”, IEEE, 2019.
- [12] Umar Sidiq Dr. Syed Mutahar Aaqib, “Disease diagnosis through data mining Techniques”, Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2019) IEEE Xplore, 2019.