International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

HEART DISEASE PREDICTION USING DATA MINING CLASSIFICATION TECHNIQUES

RISHI PREETHAM ALLU[1], SITHA RAMANAJNEYULU THOTA[2]

[1]*Student SRM Institute of Science and Technology, Dep. Electronics and Communications Engineering, Chennai, TamilNadu, India*
[2]*Student Koneru Lakshmaiah Education Foundation, Dept. computer science, Vijayawada, Andhra Pradesh, India*

---

1. Introduction

In day to day life many factors affect a human heart. Many problems are occurring at a rapid pace and new heart diseases are rapidly being identified.In this day and age of pressure Heart, is a fundamental organ to banter for solid living. The soundness of a human heart depends on the encounters in an individual's life and is totally subject to the expert and individual practices of an individual. There may likewise be a few hereditary components through which a kind of coronary illness is passed down from ages. As indicated by the World Health Organization, consistently in excess of 12 million passings are happening worldwide because of the different sorts of heart sicknesses which are likewise known by the term cardiovascular illness. The term Heart infection incorporates numerous illnesses that are assorted and explicitly influence the heart and the veins of an individual. Indeed, even youthful matured individuals around their 20-30 years of life expectancy are getting influenced by heart sicknesses. The expansion in the chance of coronary illness among youthful might be because of the awful dietary patterns, absence of rest, fretful nature, sadness and various different factors like weight, horrible eating routine, family ancestry, hypertension, high blood cholesterol, inactive conduct, family ancestry, smoking, and hypertension. The determination of heart infections is vital and is itself the most confounded assignment in the clinical field. Every one of the referenced variables are mulled over when examining and understanding the patients by the specialist through manual registration at ordinary time periods. The indications of coronary illness incredibly rely on which of the distress felt by a person. A few indications are not typically distinguished by the ordinary citizens. Nonetheless, basic side effects incorporate chest torment, shortness of breath, and heart palpitations. As of late, the medical services industry has been creating colossal measures of information about patients and their sickness conclusion reports are as a rule particularly taken for the forecast of coronary failures around the world. At the point when the information about coronary illness is gigantic, AI procedures can be carried out for the examination Recently, the healthcare industry has been generating huge amounts of data about patients and their disease diagnosis reports are being especially taken for the prediction of heart attacks worldwide. When the data about heart disease is huge, machine learning techniques can be implemented for the analysis. Data Mining is the task of extracting the vital decision making information from a collective of records for future analysis or prediction. The order is one information mining procedure through which the future result or forecasts can be made dependent on the accessible authentic information. The clinical information mining made a potential answer for coordinate the grouping methods and give electronic preparing on the dataset that further prompts investigating the secret examples in the clinical informational indexes which are utilized for the forecast of the patient's future state. Subsequently, by utilizing cli patients, these examples are a lot of fundamental. In basic English, clinical information mining utilizes order calculations that are a nical information mining it is feasible to give bits of knowledge on a patient's set of experiences and can offer clinical help through the investigation. For clinical investigation of the crucial piece of distinguishing the chance of a coronary failure before the event. The arrangement calculations can be prepared and tried to make the forecasts that decide the individual's idea of being influenced by coronary illness.

2. LITERATURE SURVEY

There are various works have been done identified with infection expectation frameworks utilizing various information mining strategies and AI calculations in clinical focuses.The expectation of Heart sickness, Blood Pressure, and Sugar with the guide of neural systems was proposed by Niti Guru et al. [4]. The dataset contains records with 13 traits in each record. The administered systems for example Neural Network with back spread calculation is utilized for preparing and testing information.

S. Seema et al,[9] centers around methods that can foresee constant ailment by mining the information containing in chronicled wellbeing records utilizing Naïve Bayes, Decision tree, Support Vector Machine(SVM) and Artificial Neural Network(ANN). A near report is performed on classifiers to quantify the better execution at a precise rate. From this investigation, SVM gives the most elevated precision rate, while for diabetes Naïve Bayes gives the most noteworthy exactness.

MeghaSashi et al, [11] proposed Heart Disease Prediction System utilizing Data Mining Techniques. WEKA programming utilized for programmed analysis of sickness and to give characteristics of administrations in medicinal services places. The paper utilized different calculations like SVM, Naïve Bayes, Association rule, KNN, ANN, and Decision Tree. The paper suggested SVM is compelling and furnishes more precision as contrasted and other information mining calculations.

R. Sharmila et al, [13] proposed to utilize non-straight grouping calculation for coronary illness forecast. It is proposed to utilize bigdata instruments, for example, Hadoop Distributed File System (HDFS), MapReduce alongside SVM for a forecast of coronary illness with improved property set. This work examined the utilization of various information-digging methods for foreseeing heart maladies. It proposes to utilize HDFS for putting away huge information in various hubs and executing the forecast calculation utilizing SVM in more than one hub at the same time utilizing SVM. SVM is utilized in equal design which yielded preferable calculation time over consecutive SVM.

Jayami Patel et al, [14] recommended coronary illness expectation utilizing information mining and AI calculation. The objective of this examination is to remove concealed examples by applying information mining methods. The best calculation J48 dependent on UCI information has the most elevated precision rate contrasted with Logistic Model Tree (LMT).

3. DATA SOURCE

The freely accessible coronary illness database is utilized. The Cleveland Heart Disease database [11] comprises of 303 records and Starlog Heart Disease database comprises of 270 records [12]. The informational index comprises of 3 sorts of properties: Input, Key, and Predictable quality which are recorded beneath.

3.1. INPUT ATTRIBUTES

Table 1. Description of 13 input attributes

S.No	Attribute	Description	Values
1	age	Age in years	Continuous
2	sex	Male or female	1 = male 0 = female
3	cp	Chest pain Type	1 = typical Type 1 2 = typical Type agina 3=non-agina Pain 4 = asymptmatic
4	thestbps	Resting Blood Pressure	Continuous value in mm hg
5	chol	Serum cholesterol	Continuous value in mm/dl
6	restecg	Resting electrographic results	0 = normal 1=having_ST_T wave abnormal 2=left ventricular hypertrophy
7	fbs	Fasting blood sugar	1 ≥ 120 mg/dl 0 ≤ 120 mg/dl
8	thalach	Maximum heart rate achieved	Continuous value
9	Exang	Exercise induced agina	0 = no 1 = yes
10	oldpeak	ST depression induced by exercise relative to rest	Continuous value
11	slope	Slope of the peak exercise ST segment	1 = unsloping 2 = flat 3=downslopin g
12	ca	Number of major vessels colored by fluoroscopy	0-3 value
13	thal	Defect type	3 = normal 6 = fixed 7 = reversible defect

All the exploration papers alluded above have utilized 13 information characteristics for the expectation of Heart disease. To get more proper outcomes two significant characteristics i.e., weight also, smoking are added to enter qualities.

4. DATA MINING TECHNIQUES USED FOR PREDICTIONS

Oflate, heap of crisis facilities supervise social protection data using the human administrations information system; as the structure contains a gigantic proportion of data, used to eliminate covered information for making a savvy clinical assurance. The essential objective of this assessment is to build an Intelligent Heart Disease Prediction System that gives discoveries of coronary sickness using an irrefutable heart information base. To develop this system, clinical terms, for instance, sex, heartbeat, and cholesterol like 13 data qualities are used. To get progressively legitimate results, two extra attributes for instance weight and smoking are used, as these qualities are considered as huge qualities for coronary sickness. The data mining plan techniques Decision Trees, and Naive Bayes are used.

4.1. NAVIE BAYES

Naive Bayes relies upon AI and data mining techniques. It progressively refined the request strategy. It is used to make the conjecture model. It shows the expected state for the probability of every quality. This model creates a logically powerful yield contrast and other yield. The essential favored posi-tion of credulous Bayes,require a restricted amount of planning data to evaluate the limit for order. It is good for figuring the most conceivable yield depends upon the info. It is definitely not hard to incorporate new data at runtime for a prevalent classifier. It is anything but difficult to include new information at runtime for a superior classifier.

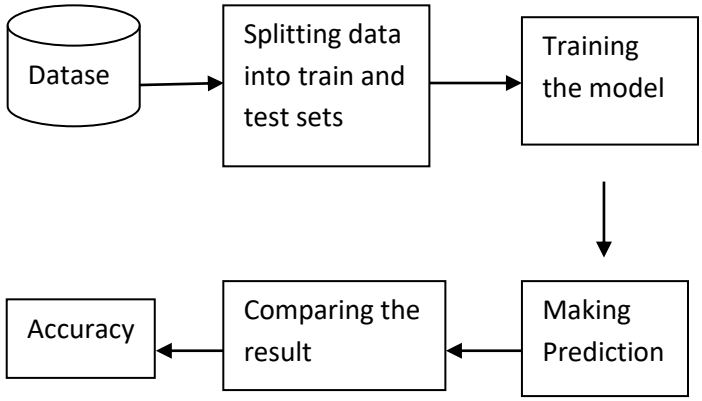
4.2. DECISION TREE

A decision tree is a different evened out structure involving centers and composed edges. A Decision tree is a construction that incorporates a root center, branches, and leaf center points. Every inside center methods a test on a trademark, each branch implies the consequence of a test, and each leaf center holds a class name. The most noteworthy center point in the tree is the root hub. In a decision tree, each leaf is given out a class mark. The non-terminal hubs, which consolidate the root and other internal hubs, contain credit test conditions to isolate records that have different qualities.

4.3. STEPS

- Importing dataset into the python environment
- Splitting data into test and train sets
- Training the model on a train set
- Making predictions on the testing set
- Comparing actual responses with predicted responses
- Finding accuracy.

4.4. IMPLEMENTATION DIAGRAM



5. RESULTS

From the outset, the dataset contained a couple of fields, wherein some impetus in the records was missing. These were perceived and sup-planted with the most legitimate characteristics using ReplaceMissingValues. The ReplaceMissingValues channel checks all records and replaces missing characteristics with a mean mode methodology. This strategy is known as Data PreProcessing. After pre-dealing with the data, data mining request methodologies, for instance, Decision Trees, and Naive Bayes were ap-used

A disarray network is gotten to figure the precision of the grouping. A disarray network shows the number of occurrences have been appointed to each class. In our investigation, we have two classes, and subsequently we have a 2x2 disarray lattice. Class a = YES (has coronary illness)Class b = NO (no heart disease)

Table 2. A confusion matrix

	a(has heart disease)	b(no heart disease)
a(has heart disease)	TP	FN
b(no heart disease)	FP	TN

TP (True Positive): It denotes the number of records classified as true while they were actually true.

FN (False Negative): It denotes the number of records classified as false while they were actually true.

FP (False Positive): It denotes the number of records classified as true while they were actually false.

TN (True Negative): It denotes the number of records classified as false while they were actually false

Table 3. Confusion matrix obtained for Naive Bayes (Train set) :

	a	b
a	109	18
b	16	38

Table 4. Confusion matrix obtained for Naive Bayes(Test set) :

	a	b
a	57	24
b	22	19

Table 5. Confusion matrix obtained for Decision tree (Train set)

	a	b
a	125	0
b	0	56

Table 6. Confusion matrix obtained for Decision tree (Test set)

	a	b
a	56	25
b	23	18

The following table provides the result of two classifications

Table 7. Result analysis of two algorithms

Techniques	Accuracy(Train set)	Accuracy(Test set)
Naive Bayes	81%	51%
Decision Tree	100%	65%

From the table that the decision tree algorithm gave the highest accuracy values when compared with the accuracy of the Naive Bayes algorithm.

6. CONCLUSION

The general target of our work is to anticipate more accurately the nearness of coronary illness. In this paper, two information mining grouping strategies were applied to be specific Decision trees and Naive Bayes. From results, it has been seen that Decision trees give exact outcomes in contrast with Naive Bayes.

This framework can be further expanded. Other information mining procedures can likewise be utilized for prediction for example Bunching, Time arrangement, Association rules. Content mining can be utilized to mine gigantic measures of unstructured information accessible in the medicinal services industry database.

7. REFERENCES

[1] Sushmita Manikandan, "Heart attack prediction system", Institute of Electrical and Electronics Engineers(IEEE),21 June 2018

[2] Aditi Gavhane; Gouthami Kokkula; Isha Pandya; Kailas Devadkar, "Prediction of Heart Disease Using Machine Learning", Institute of Electrical and Electronics Engineers(IEEE),01 October 2018

[3] Monika Gandhi; Shailendra Narayan Sing, "Predictions in heart disease using techniques of data mining", Institute of Electrical and Electronics Engineers(IEEE),13 July 2015

[4] J. Thomas; R Theresa Princy, "Human heart disease prediction system using data mining techniques", Institute of Electrical and Electronics Engineers(IEEE),04 August 2016

[5] Chaitanya Suvarna ; Abhishek Sali ; Sakina Salmani, "Efficient heart disease prediction system using optimization technique", Institute of Electrical and Electronics Engineers(IEEE),08 February 2018

[6] A H Chen ; S Y Huang ; P S Hong ; C H Cheng ; E J Lin, "HDPS: Heart disease prediction system", Institute of Electrical and Electronics Engineers(IEEE),09 March 2012

[7] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques", 2008 IEEE/ACS International Conference on Computer Systems and Applications, pp. 108-115, 2008.

[8] K. Kwon, H. Hwang, H. Kang, K. G. Woo, and K. Shim, "A remote cardiac monitoring system for preventive care", 2013 IEEE International Conference on Consumer Electronics (ICCE), pp. 197-200, 2013.

[9] Salam Ismaeel, Ali Miri, et al., "Using the Extreme Learning Machine (ELM) technique for heart disease diagnosis", IEEE Canada International Humanitarian TechnologyConference, 03 September 2015.

[10] N. Oliver and F. F. Mangas, "HealthGear: a real-time wearable system for monitoring and analyzing physiological signals", IEEE International Workshop

on Wearable and Implantable Body Sensor Networks, 2006.

[11] C. Ordonez, "Association rule discovery with the train and test approach for heart disease prediction", IEEE transactions on Information technology in biomedicine: a publication of the IEEE Engineering in Medicine and Biology Society, vol. 10, no. 2, pp. 334-43, Apr. 2006.

[12] V. Krishnaiah, G. Narsimha, N. Subhash Chandra, "Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review", International Journal of Computer Applications, February 2016.

[13] Megha Shahi, R. Kaur Gurm, "Heart Disease Prediction System using Data Mining Techniques", Orient J. Computer Science Technology, vol.6 2017, pp.457-466

[14] Mr. Chala Beyene, Prof. Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, 2018

[15] R. Sharmila, S. Chellammal, "A conceptual method to enhance the prediction of heart diseases using the data techniques", International Journal of Computer Science and Engineering, May 2018

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