

MACHINE LEARNING APPLICATION IN CARDIOLOGY: EMPLOYABILITY OF SUPPORT VECTOR MACHINE AND LOGISTIC REGRESSION IN THE EARLY STAGE DETECTION OF CARDIAC DISEASE

Raj Verma

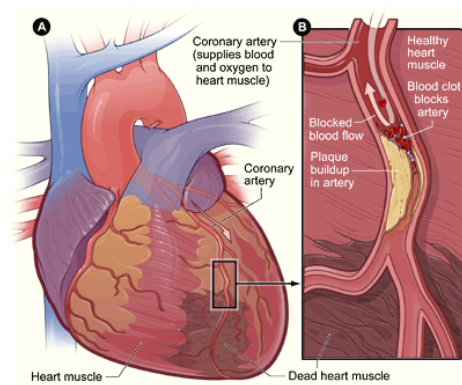
ABSTRACT

A typical term heart disease is only a cardiovascular infection or a coronary illness which lessens the proficiency and legitimate working of heart by blocking veins, supply route or veins around it. Coronary illness causes handicap, for example, harm to the mind bringing about death. Given Statistics [10], it demonstrates that scope of age amass from 25 to 69 have 25% danger of having heart maladies. Some indispensable reasons for cardiovascular sickness are, physical idleness, smoking, expending more shoddy nourishment and dependence of liquor which are real foundations for stroke, chest agony, and heart assault. Anyway as a result of the mindfulness about components and indications that are in charge of the heart issue, it is conceivable to anticipate any heart issue dependent on a measurable examination of medical records. Anyway, Data mining, a cutting-edge strategy has given a programmed method for investigating information utilizing standard arrangement techniques. Although many classifiers are accessible in information mining that can be utilized to foresee the heart issues, this paper accentuates on finding the fitting classifier that can give better exactness by applying information mining systems viz: gullible Bayes, Support Vector machine and Logistic Regression.

Keywords: *Coronary, Naive Bayes, Support Vector Machine, Logistic Regression*

1. INTRODUCTION

Heart infections particularly coronary illness is an extremely lethal and risky malady provided that tolerant overlooks its prior indications, which is by all accounts a notice sign, it gives no opportunity to understanding for recuperation and in the end, may prompt demise on the spot. This is called heart assault. It happens because the courses can supply oxygen-rich blood to the heart yet because of the greasy and other substance the plaque is framed which transforms ordinary coronary corridor into a narrowing of the coronary conduit. Coronary illness is a turmoil in which a waxy component called plaque develops inside the coronary supply routes. Along these lines, the stream of the blood to the heart can either back off or stop as appeared in the Fig 1.



The principal factors which are in charge of a coronary issue are named controllable hazard factors and wild hazard factors. A portion of the controllable hazard factors incorporates diabetes, smoking and heftiness or abundance weight, cholesterol, High circulatory strain, the absence of activity and physical movement. Wild hazard factors incorporate family ancestry, age, sex, and past medicinal scatter. Presently a days Electronic Medical Records(EMR) and Electronic Health Records (EHR) has streamlined and systematized the examination part to distinguish the patients issues, yet forecast of conclusion with exactness is as yet a test among present analysts who are contributing in proposing and creating diverse techniques in the field of basic Human ailments, for example, Cancer, coronary illness, diabetes and so forth. A portion of the commitments is referenced in the following area which gives a short portrayal of exercises going on by and by.

2. PROPOSED METHODOLOGY

There has been loads of work done on Coronary Heart ailment order and expectation utilizing following information mining procedures.

- Naïve Bayes
- Decision tree
- K closest neighbour
- Support vector machine

From the writing survey it's clear that despite the fact that innocent Bayes and bolster vector machines are better calculations the exactness has yet than be checked with bigger informational collections. So the proposed arrangement will work in three noteworthy modules.

First by expanding the span of the informational collection the exactness of expectation would be registered. Proposal dependent on the exactness among the two calculations with expanded informational indexes would be made.

Second since calculated relapse can likewise be utilized for forecast, an expectation display utilizing strategic relapse would be produced. Third an examination dependent on the over three modules dependent on its exactness would result in a best model for coronary illness forecast.

The qualities which are significantly utilized are taken from the referenced paper [13] where properties and its use are referenced in detail, those are as per the following:

- Age
- Gender
- Chest torment
- Resting circulatory strain
- Cholesterol
- Fasting glucose
- Resting electrocardiographic outcomes
- Maximum pulse accomplished
- Exercise incited angina
- ST dejection actuated by exercise with respect to rest
- Slope of the pinnacle
- Number of significant vessels shaded by fluoroscopy
- Thalassemia

Ventures of usage are as per the following: -

1. Collect Dataset that can be utilized for the testing reason (utilization of informational index from UCI Repository)
2. Implement Naïve Bayes Algorithm so that it ought to have the capacity to take the informational collection as an information set of characteristics esteems.
3. Implement Support Vector Machine correspondingly that can likewise take the informational collection as info esteems.
4. Build a forecast Model utilizing Logistic Regression Approach
5. Compare the Accuracy dependent on information and result investigation.

3. DATA SET

Informational index from Data mining store of University of California, Irvine (UCI) has been gathered for testing reason, which comprises of accumulation of informational index from Cleveland, Hungary, Switzerland and long shoreline and Stat log. Some arrangement of Data from UCI Repository.

3.1 Naïve Bayes Algorithm

Guileless Bayes calculation is a decent apparatus in therapeutic conclusion. For example, given a rundown of properties named side effects, it can hypothesize likelihood of a sickness. Innocent Bayes consider restrictively autonomous properties. The classifier forms each characteristic's likelihood in a class. The most elevated back Probability class will be considered as consequence of the order. Innocent Bayes is basic, productive and great execution in characterization. It likewise gives great exactness to universally useful examination. Because of its great exactness it very well may be utilized in restorative analysis. The essential methodology that can be utilized is delineated in following figure 3.

3.2 Support Vector Machine (SVM):

A help vector machine is an order kind of strategy used to examine information and perceive patters in a relapse and characterization examination. Bolster vector machine (SVM) is utilized when your information is delegated two classes. A SVM perceives and isolates comparative information by finding the best hyper plane that isolates all information purposes of one class from those of alternate class. Display turns out to be better when edges are bigger between classes. An edge ought not have focuses in its inside part. The help vectors are the information organizes that are on the limit of the edge. Scientific capacities are engaged with SVM plan which is every now and again used to show certifiable issues. Its execution amplify with number of traits [12].

3.3 Logistic Regression

It's a kind of factual relapse investigation strategy utilized for estimate and forecast of aftereffect of a distinct ward traits. Subordinate means it can take just some arrangement of values for instance twofold qualities, for example, genuine or false, great or terrible, on or off in like manner. Strategic relapse is for the most part utilized for expectation other than that it can likewise be utilized in figuring the likelihood of progress. Essentially Logistic Regression includes fitting a condition of the shape to the information:

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n - \text{eq. 1}$$

The relapse coefficients are normally evaluated utilizing most extreme probability estimation. The most extreme probability proportion decides the measurable importance of autonomous factors on the needy factors. The probability proportion test evaluates the commitment of individual indicators (free factors). At that point the likelihood (p) of each case is determined utilizing chances proportion,

$$P/(1-P) = e^Y \text{ -- eq. 2}$$

From this p-value is found out. This gives the probability or chance for the individual to have coronary heart disease[12].

4. RESULT ANALYSIS

4.1 Performance Analysis Metrics:

Performance will be evaluated based on following parameters: Accuracy = $(TP + TN) / (TP + FP + TN + FN)$

Where TP = true positives are the number of positive cases correctly classified

TN = true negatives is the number of negative cases correctly classified

FP = false positives is the number of negative cases incorrectly classified as positive

FN = false negatives is the number of positive cases incorrectly classified as negative.

This precision fluctuates with the span of the informational collection, so here a correlation table has been made to examine the effect of size on every technique viz. Credulous, SVM and Logistic while giving exactness that is delineated.

5. CONCLUSION AND FUTURE WORK

In this paper, have talked about the original arrangement strategies utilized in information digging for the expectation of coronary illness is examined and with the assistance of exactness examination, we have demonstrated that SVM is superior to anything other two techniques when we have an extensive informational collection of passages.

The framework with SVM will help for early conclusion of the coronary illness for some random patient. This framework when sent can supplement conventional coronary illness recognition framework and can help the specialists as well as the patients. This framework can go about as a help for the medicinal ventures for the discovery of the coronary illness with better precision.

6. REFERENCES

- [1] Minas A. Karaolis, Joseph A. Moutiris, Demetra Hadjipanayi, Constantinos S. Pattichis, "Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining With Decision Trees", IEEE Transactions On Information Technology In Biomedicine, VOL. 14, NO. 3, MAY 2010.
- [2] Raghunath Nambiar, Adhiraaj Sethi, Ruchie Bhardwaj, Rajesh Vargheese, "A Look at Challenges and Opportunities of Big Data Analytics in Healthcare", 2013 IEEE International Conference on Big Data.

- [3] T.John Peter, K. Somasundaram,” An Empirical Study on Prediction of Heart Disease Using Classification Data Mining Techniques”, IEEE, International conference on Advances in engineering, science and management,pp.514-518, 2012.
- [4] EmanAbuKhoussa, Piers Campbell,” Predictive Data Mining to Support Clinical Decisions: An Overview of Heart Disease Prediction Systems”, IEEE, International Conference on Innovations in Information Technology, pp.267-272, 2012.
- [5] Aqueel Ahmed, Shaikh Abdul Hannan,” Data Mining Techniques to Find Out Heart Diseases: An Overview”, International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2012.
- [6] LamiaAbedNoor Muhammed,” Using Data Mining technique to diagnosis heart disease”, IEEE, International conference on statistics in science, Buiseness and Engineering, pp.1-3, 2012.
- [7] Sivagowry S, Dr.Durairaj. M and Persia. A & Research Scholar, “An Empirical Study on Applying Data Mining Techniques for the Analysis and Prediction Heart Disease”, IEEE, International Conference on Information Communication and embedded system, pp.265-270, 2013.
- [8] M.Akhiljabbar , Dr.Priti Chandra, Dr.B.LDeekshatulu,” Heart Disease Prediction System Using Associative Classification and Genetic Algorithm”, ICECIT, 2012.
- [9] Ranganatha S., Pooja Raj H.R., Anusha C., Vinay S.K.,” Medical Data Mining And Analysis For Heart Disease Dataset Using Classification Techniques”,IEEE, National conference on challenges in research and technology in the coming decades,pp.1-5,2013.
- [10] VikasChaurasia, Saurabh Pal, “Early Prediction of Heart Diseases Using Data Mining Techniques”, Carib.j.SciTech, 2013, Vol.1, 208-217.
- [11] Mamuna Fatima, IqraBasharat, Dr.Shoab Ahmed Khan, Ali Raza Anjum,, “Biomedical (Cardiac) Data Mining: Extraction of significant patterns for predicting heart condition”, IEEE conference on Computational Intelligence in bioinformatics and computational biology, pp.1-7, 2014.
- [12] Mythili T., Dev Mukherji, Nikita Padalia, and Abhiram Naidu “A Heart Disease Prediction Model using SVM- Decision Trees-Logistic Regression (SDL)”, IJCA, Vol.68- No.16 April 2013.
- [13] Carlos O., Edward O ,Levien de Braal, and team “Mining Constrained Association Rules to Predict Heart Disease”, IEEE, International Conference on Data Mining p.433- 440, 2001.