

# Implementation of Various Machine Learning Algorithms for Heart Attack Prediction

Rajesh Rahangdale<sup>1</sup>, Mr. Kranti Kumar Dewangan<sup>2</sup>

<sup>1,2</sup> *Department of Computer Science and Engineering, Rawatpura Sarkar University, Raipur (C.G)*

**Abstract -** The Heart Disease as exhibited by the diagram is the guideline wellspring of death any place on the world. The flourishing locale has an immense heap of information, yet despicably, this information is not overall around used. This is an immediate consequence of nonattendance of persuading evaluation mechanical congregations to find striking models in information. Information Mining can assist with recovering critical information from open information. It assists with arranging model to foresee patients' flourishing which will be quicker showed up contrastingly according to clinical trial and error. An enormous heap of examination has been done utilizing the Different Heart datasets. Different Implementation of AI calculations like K-Nearest Neighbor, Support Vector Machine, Logistic Regression, and soon have been applied. This assessment is on a very basic level looking for the capable computations that will work totally on coronary ailment. We will examine different computations on a given dataset and notice the entire preliminary achieve our work.

**Index Terms -** Heart Disease, Machine Learning Algorithms, KNN, Random Forest method, and Logistic Regression.

## I. INTRODUCTION

As per estimations open till 2018, an ordinary 17.9 million enormous number of passings happen overall reliably as a result of cardiovascular sickness (CVD) which checks to 31% of entire passings all over the planet. On account of existing examples proceed, the yearly figure of passings from CVD will mount to 22.2 million by 2030 [<http://www.who.int>]. An all-out estimate by using data mining techniques might provide us with an early accurate completion of this disease. A combination of data mining approaches like Decision tree [1], Neural Network, Naive Bayes, KNN estimation and moreover some cross variety system called brain association outfit for instance mix of brain association and gathering based techniques are used to arrange, expect and bundle data to choose right or

precise dynamic for the risk of coronary ailment. The term CVD contains different kinds of disturbance that might hurt the heart.

As a result of specific perils connected with clinical treatments like the deferral in the result and the non-availability of the clinical workplaces to people, the assumption model is proposed. So we will apply the unmistakable machine estimations in our dataset to get the specific assumption and moreover to see which features is more co related with the disease. The place of this investigation is to ponder different AI computation like decision tree, Support vector machine, backslide estimations to understand the association between the qualities of coronary ailment dataset.

## II. LITERATURE SURVEY

A Hybrid Machine Learning Approach for Prediction of Heart Diseases

The place of this paper is to present a capable system of anticipating heart diseases using AI moves close. Thusly we proposed a crossbreed approach for heart estimate using Random Forest classifier and essential k-suggests computation AI techniques. The dataset is furthermore evaluated using two other assorted AI estimations, to be explicit, J48 tree classifier and Naive Bayes classifier and results are contemplated. Results accomplished through Random boondocks classifier and the relating disorder network shows life of the framework.

A Survey on Predicting Heart Disease utilizing Data Mining Techniques

The mark of this paper is to present a capable technique of predicting heart ailments using AI moves close. In this manner we proposed a mutt approach for heart estimate using Random Forest classifier and fundamental k-infers computation AI strategies. The dataset is furthermore surveyed using two other

assorted AI computations, to be explicit, J48 tree classifier and Naive Bayes classifier and results are pondered. Results accomplished through Random woodlands classifier and the relating chaos network shows life of the framework.

#### Forecast of Heart Disease Using Machine Learning

With the wild development in the heart stroke rates at juvenile ages, we really want to set up a system to have the choice to perceive the signs of a heart stroke at a starting stage and therefore hinder it. It is impractical for a normal individual to regularly go through extreme tests like the ECG and consequently there should be a structure set up which is useful and at the same time strong, in expecting the chances of a coronary sickness. Consequently we propose to cultivate an application which can expect the shortcoming of a coronary ailment given central aftereffects like age, sex, beat rate, etc The AI estimation brain associations has shown to be the most precise and strong computation and accordingly used in the proposed system.

#### Prescient Data Mining to Support Clinical Decisions: An Overview of Heart Disease Prediction Systems

Clinical benefits affiliations are gone up against with hardships to give monetarily clever and incredible patient thought. The two chiefs and clinicians need to research an overflow of data open in the informational indexes of clinical consideration information systems to track down data and to make instructed decisions. This is essential explicitly to work on the sufficiency of sickness therapy and contraventions. It happens to more huge assuming there should be an event of coronary ailment (HD) that is considered the fundamental clarification for death in adults. Data mining fills in as an assessment mechanical assembly to track down concealed associations and models in HD clinical data. This paper reviews five models worked of single and united data mining methodologies to help clinical decisions in (HD) assurance and estimate. The five structures give customized plan affirmation and attempts to reveal associations among different limits and signs of HD. Each structure shows set of characteristics and obstructions to the extent that the sort of data it handles, precision, effortlessness of interpretation, trustworthiness and theory limit. Powerless hypothesis limit is at this point a huge open issue for data mining

in clinical benefits essentially because of the shortfall of data and cost of re-taking care of.

### III. ML ALGORITHMS

There are different extents of supervisor systems utilized for data mining that are made in the latest years and used in data mining practical applications that fuse alliance, grouping, assumption and model appraisals, etc

#### Classification:

Ordering is among the principal methods of data mining that have a spot with area of AI. It is considered as a procedure to arrange all of the things present in a lot of data. Portrayal moreover incorporates abuse of different frameworks and techniques of science and experiences like direct programming, decision tree, and brain association.

#### Clustering:

Gathering is one of the data mining systems which are helpful for batching substances having similar features using mechanical methods. Bundling is totally not quite the same as gathering. Here the classes are described by gathering techniques and things are set on them. In course of action methodology, objects are shared with predefined classes. Through gathering thick and additional regions in object space can be seen and find scattering plans and intriguing associations among the characteristics of data. It suggests data division [10].

#### Naïve based Bayes:

Gullible Bayes is one of the AI estimations that keep an eye on the gathering issue, which relies upon Bayes probability theory. Earlier it was popular for text gathering that charms high layered planning enlightening files. The Naive Bayes portrayal is a probabilistic classifier. It relies upon probability models which rely upon strong opportunity assumption. For example, a disease may be considered a heart hardship on the off chance that an individual encase chest torture, circulatory strain and cholesterol. A blameless bayes classifier contemplated all of these features to contribute in relating to the probability that this disease is a coronary ailment or not. The condition for guileless bayes is given underneath:

$$P(Y) = (P(Y|C) * P(C)) / (P(Y)) \quad (1)$$

Where Y is the case to be anticipated and C is the class an incentive for example. The above-given recipe or conditions used to decide the class where element expected to order.

**Decision Tree:**

A decision tree is a coordinated learning computation classifier that is not difficult to understand and unravel. It oversees both numerical and absolutely enlightening files. Decision tree looks as similar as the tree structure looks where internal center points, branches and leaf center points are accessible and all of those branches implies quality potential gains of given dataset. A test is explained by internal center points on a given game plan of attributes. On the other hand, the classes which are considered or recommend the end results are showed up by the leaf centers. In view of farsighted quality and the given rules, game plan of request begins from the root center point to leaf centers. The most regularly utilized decision tree approaches consolidates CART, ID3, C4.5, J48, and CHAID are indispensable in the gauge of diseases.

**K-means Algorithm:**

K-suggests is a vector quantization computation that produces k pack from given objects of issue region so as objects of each gathering are more like. Despite recognizing evidence of the bundle numbers, k-suggests also "learns" the gathering on its have without additional information concerning a discernment should which pack, which is the essential clarification that k-infers procedure is considered as semi-managed learning. K-suggests is especially effective for gigantic enlightening lists.

**Support Vector Machine:**

Support vector machines from time to time in like manner called help vector network are one of the controlled learning models. Recently, SVM is one of the for the most part used learning estimations that recognize data for gathering. In this estimation, we plan each data thing as a point in n-layered space where n is number of features you have, with the value of every part being the value of a particular orchestrate. By then, we perform portrayal by finding the hyper-plane that isolates the two classes well in general. Support Vectors are only the co-ordinates of individual insight. SVMs perform non-straight request as well as performing direct gathering. Backing vector machines are significant in text and hypertext request,

course of action of pictures and much more locales now day to day. Backing vector machine is fitting for silly cases and showed the best show [11].

**IV. METHODOLOGY AND EXPERIMENTAL RESULTS**

In this paper, to estimate the different machine learning algorithms that be capable to envisage heart diseases based on measurements and datasets. The accompanying area gives data about the information that is utilized in my exploration. This information originates from an online website dataworld.com show. Our dataset is having total 14 columns in which all related numerical values of essential features are given. The dataset contain the 1000 rows and 14 columns. The some of the column information is given below:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Age	sex	cp	restingcp	chol	fbs	resting	maxhr	maxwp	oldpeak	slope	ca	thal	target						
2	52	1	0	125	222	0	1	146	0	1	2	2	3	0						
3	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0						
4	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0						
5	41	1	0	146	203	0	1	145	0	0	2	1	3	0						
6	62	0	0	138	204	1	1	166	0	1.6	0	1	2	0						
7	58	0	0	100	143	0	0	122	0	1	1	0	2	1						
8	58	1	0	114	193	0	2	140	0	4.4	0	3	1	0						
9	55	1	0	160	209	0	0	145	1	6.8	1	1	3	0						
10	48	1	0	130	149	0	0	144	0	6.8	2	0	3	0						
11	54	1	0	122	186	0	0	118	1	3.2	1	2	2	0						
12	71	0	0	112	149	0	1	125	0	1.6	1	0	2	1						
13	43	0	0	122	142	1	0	136	1	3	1	0	3	0						
14	54	0	1	118	188	0	1	152	0	6.7	2	0	2	1						
15	51	1	0	140	198	0	1	122	1	4.2	1	3	3	0						
16	52	1	0	128	204	1	1	156	1	1	1	0	0	0						
17	54	0	1	118	199	0	1	192	0	6.7	2	0	2	1						
18	51	0	2	140	188	0	0	142	0	1.5	2	1	2	1						
19	54	1	0	124	166	0	0	109	1	2.2	1	1	3	0						
20	50	0	1	120	144	0	1	162	0	1.1	2	0	2	1						
21	58	1	2	140	211	1	0	165	0	0	2	0	2	1						
22	49	1	2	140	185	0	0	155	0	3	1	0	2	0						
23	67	0	0	106	223	0	1	142	0	0.3	2	2	2	1						
24	45	1	0	104	108	0	0	148	1	3	1	0	2	1						
25	63	0	2	135	252	0	0	172	0	0	2	0	2	1						

Fig.1 Heart Disease Dataset

AI and Machine Learning (ML) field is a huge example dull of the IT business. While conversations over the security of its improvement continue to rise, fashions grow cutoff points and cutoff of fake sharpness. Today Artificial Intelligence went far past sci-fi thought. It changed into a need. Being generally utilized for preparing and taking a gander at epic volumes of information, AI assists with managing the work that is unfathomable truly any additional thinking about its all around expanded volumes and power.

For example, AI is applied in evaluation to make suppositions that can assist individuals with making solid structures and journey for logically powerful plans. FinTech applies AI in principle stages to do genuine research and anticipate where to contribute

assets for more conspicuous benefits. The meandering out industry utilizes AI to pass on changed ideas or dispatch Chatbots, despite update the general client experience. These models show that AI and ML are utilized technique heaps of information to offer better client experience, constantly precious and definite one.

As AI and ML are being applied across different channels and adventures, colossal associations put resources into these fields, and the interest for specialists in ML and AI makes as necessities be. Jean Francois Puget, from IBM's AI office, conveyed his assessment that Python is the most standard language for AI and ML and set up it in regards to a model recorded records on indeed.com.

### Pandas

A panda is an open-source Python Library giving top notch information control and appraisal gadget utilizing its notable information structures. The name Pandas is gotten from the word Panel Data an Econometrics from Multidimensional information. In 2008, maker Wes McKinney began making pandas while requiring dominating, flexible contraption for assessment of information. Before Pandas, Python was essentially utilized for information munging and status. It had near no liability towards information appraisal. Pandas took care of this issue. Utilizing Pandas, we can achieve five common strides in the managing and assessment of information, paying little mind to the underpinning of information - load, plan, control, model, and check out. Python with Pandas is utilized in a wide degree of fields including scholarly and business areas including store, financial issue, Statistics, assessment, and so forth

### Scikit Learn

The scikit-learn experience began as scikits. learn, a Google Summer of Code experience by David Cournapeau. Its name comes from the likelihood that it is a "SciKit" (SciPy Toolkit), a uninhibitedly made and given outsider extension to SciPy. The first codebase was in this manner changed by different modelers. In 2010 Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort and Vincent Michel, all from the French Institute for Research in Computer Science and Automation in Rocquencourt, France, took authority of the undertaking and made the fundamental open delivery on February the underlying 2010. Of the

assorted scikits, scikit-advance in basically the same manner as scikit-picture were portrayed as "by and large around kept up and outstanding" in November 2012. Scikit-learn is perhaps the most observed AI libraries on GitHub.

### Matplotlib

Matplotlib is a shocking portrayal library in Python for 2D plots of groups. Matplotlib is a multi-stage data portrayal library in light of NumPy displays and planned to work with the more broad SciPy stack. It was introduced by John Hunter in the year 2002. Maybe the greatest benefit of portrayal is that it licenses us visual induction to enormous proportions of data in actually absorbable visuals. Matplotlib includes a couple of plots like line, bar, scatter, histogram, etc

We have used unmistakable machine computations for instance Innocent Bias, KNN, SVM, Logistic backslide and Decesion tree. Our models work fine anyway best of them are KNN and Random Forest with 88.52% of accuracy.

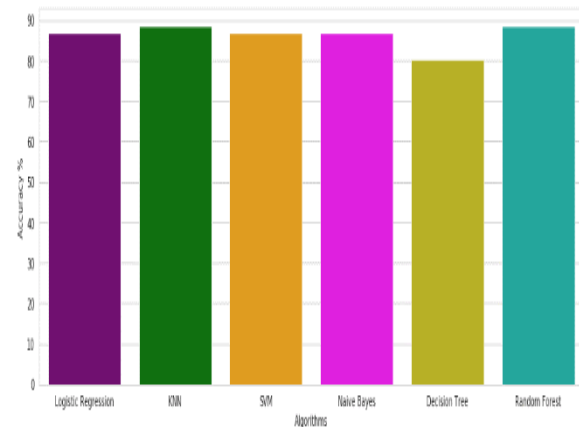


Fig.2 Comparison of all machines learning algorithm in heart disease dataset



Fig 3: Confusion Matrix for all applied algorithms

S. No.	Applied Algorithms	Accuracy (in percentage)
1	Logistic Regression	78.50
2	Support Vector Machine	86.32
3	Decision Tree	78.88
4	Random Forest	88.50
5	KNN	89.52
6	Naïve bias	88.29

Table 1

#### IV. CONCLUSION

The different infection forecast techniques are analyzed and separated in this work. The AI techniques used to predict heart diseases are analyzed here. Coronary ailment is a human contamination by its personality. This disease makes a couple of issues, for instance, coronary episode and passing. In the clinical space, the significance of counterfeit information and AI strategies is seen. Various advances are taken to apply important methodologies in the disease conjecture. The assessment works with strong methodology that are done by different examiners were gathered in this work. From the general assessment we can gather that KNN and Random Forest system method is a capable procedure for predicting Heart Disease. It gives incredible precision by seeing different assessment works.

#### REFERENCE

[1] S. Palaniappan and R. Awang, “Intelligent heart disease prediction system using data mining techniques,” in IEEE/ACS International Conference on Computer Systems and Applications. IEEE, 2008, pp. 108–115.

[2] D. P. Mandic and J. Chambers, *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*, 2001.

[3] Dangare, C. S. & Apte, S.S. “A Data mining approach for prediction of heart disease using neural network’s”, *International Journal of Computer Engineering & Technology(IJCET)*, Volume 3, Issue 3, October – December (2012), pp. 30-40.

[4] Indhumathi S, & Vijaybaskar G., “Web based health care detection using naive Bayes algorithm”, *International Journal of*

[5] Chitra R., & Seenivasagam, V. “Review of heart disease prediction system using data mining and hybrid intelligent techniques”, *ICTACT JOURNAL ON SOFT COMPUTING*, July 2013, volume: 03, issue: 04 pp.605-09.

[6] Sudhakar K. and Manimekalai M., “Study of Heart Disease Prediction using Data Mining”, *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 4, Issue 1, pp.1157-60, January 2014.

[7] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*, 2016.

[8] Anooj, P .K., “Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules,” *Journal of King Saud University – Computer and Information Sciences* (2012) 24, 27–40.