

Mini review Article

A model for Heart Disease Prediction using Data Mining Classification Techniques (Decision Trees, Naive Bayes, and KNN)

Comment [m1]: heart disease differs. I think it is necessary to specify which heart disease the prediction is applied to?

ABSTRACT

Nowadays the guts malady is one amongst the foremost causes of death within the world. Thus it's early prediction and diagnosing is vital in medical field, which might facilitate in on time treatment, decreasing health prices and decreasing death caused by it. The treatment value the disease isn't cheap by most of the patients and Clinical choices are usually raised supported by doctors' intuition and skill instead of on the knowledge-rich information hidden within the stored data. The model for prediction of heart disease using a classification techniques in data mining reduce medical errors, decreases unwanted exercise variation, enhance patient well-being and improves patient results. The model has been developed to support decision making in heart disease prediction based on data mining techniques. The experiments were performed using the model, based on the three techniques, and their accuracy in prediction noted. The decision tree, naïve Bayes, KNN and WEKA API were the various data mining methods that were used. The model can predict the like hood of getting a heart disease using medical attributes such as blood pressure, sex age cholesterol and blood sugar. 740 Record sets with medical attributes was obtained from a publicly available database for heart disease from machine learning repository with the help of the datasets, and the patterns significant to the heart attack prediction was extracted and divided into two data sets, one was used for training which consisted of 296 records & another for testing consisted of 444 records, and the fraction of accuracy of every data mining classification that was applied was used as standard for performance measure. The performance was compared by calculating the confusion matrix that assists to find the precision recall and accuracy. High performance and accuracy was provided by the complete system model. Comparison between the proposed techniques and the existing one in the prediction capability was presented. The model system assists clinicians in survival rate prediction of an individual patient and future medication is planned for. Consequently, the families, relatives, and their patients can plan for treatment preferences and plan for their budget consequently.

Comment [m2]: Other factors may influence the process such as genetic factors, sedentary or socio-economic status and race

Keywords: WEKA API; Decision Tree; Naive Bayes; KNN.

1. INTRODUCTION

The Heart being a strong organ, situated close to the middle of the chest; it is duty is pumping blood to different parts of the body and together with system of vessels and blood from the human body's cardiovascular framework; interferences to this dissemination of blood can result in serious medical issue including death [5]. People have been affected by dangerous sicknesses all through the past. The system for prediction can assist to lower the dangers of the disease. Prediction is done dependent on the present data fed to the framework model Using WEKA API which is open source information mining programming in Java. The model is being created dependent on three distinct information mining strategy that is Nave Bayes, KNN, decision tree with WEKA API. The input dataset is analyzed using different classification algorithms and comparison is done for accuracy

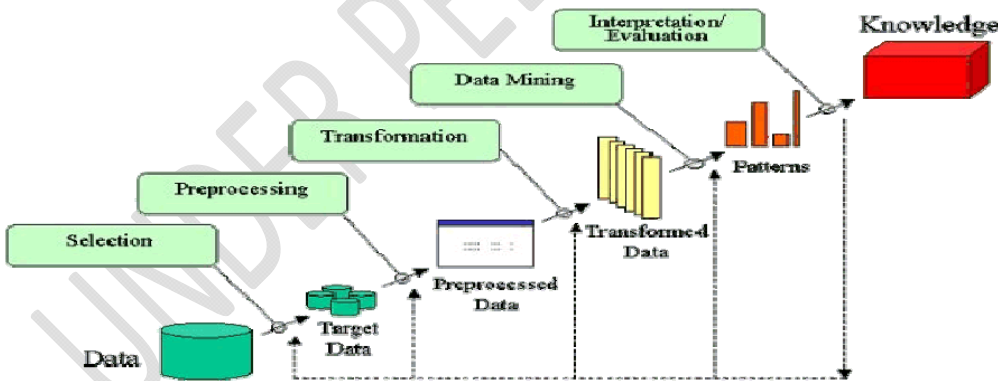
Nowadays an immense measure of information is gathered and kept in a daily basis. There is a significant need to break down this information yet with no scientific device, this appears to be

45 unimaginable. This has prompted the improvement of Knowledge Discovery in Databases (KDD) which
 46 changes the low dimension information to a top state learning. KDD comprises of different procedures at
 47 various advances and Data mining is one of those procedures. Information mining is the way toward
 48 finding fascinating learning from huge measure of information kept in databases, information stockrooms
 49 or other data vaults. The fundamental point of information mining procedure is to separate data from a
 50 dataset and change it into a reasonable structure so as to help basic conclusions [45]. A tremendous
 51 measure of information is accessible in healthcare industry however the mining of this information is poor.
 52 In this way, the investigation of the medicinal services information is a must. Information Discovery in
 53 databases is getting to be famous research instrument for open human services information. In this study,
 54 we will do the exhibition investigation of various information mining grouping strategies on medicinal
 55 services information. This work will help discovering the best information mining arrangement method as
 56 far as precision on the specific dataset. The examined characterization systems are K-closest neighbor
 57 (KNN), Naive Bayes, Decision tree. The exhibition of these procedures is estimated dependent on their
 58 exactness. This investigation will assist the future scientists with getting proficient outcomes in the wake
 59 of realizing best information mining grouping method for specific dataset.

Comment [m3]: It is necessary to specify which department of which hospital

60 Information Mining is the nontrivial procedure of recognizing substantial, novel, conceivably valuable and
 61 at last reasonable example in information with the wide utilization of databases and the touchy
 62 development in their sizes. Information mining refers to removing or "mining" learning from a lot of
 63 information. Information digging is the quest for the connections and worldwide examples that exist in
 64 enormous databases however are tucked away among a lot of information [17]. The fundamental
 65 procedure of Knowledge Discovery is the change of information into learning so as to help in making
 66 judgments is known as information mining. Information Discovery procedure comprises of an iterative
 67 grouping of information cleaning, information coordination, information determination, information mining
 68 design acknowledgment and learning introduction. Information digging is the quest for the connections
 69 and worldwide examples that exist in enormous databases bramble are tucked away among a lot of
 70 information.

71 The figure beneath illustrates Steps of the Knowledge Discovery in Databases process on the most
 72 proficient method to separate learning from information with regards to enormous databases. [14]



73

74 **Figure1.0 steps of Knowledge Discovery in Databases process** ||

Comment [m4]: The source of the figure must be specified unless it is personal

75

76 Various health industry information systems are structured to help patient charging, stock organization
 77 and making some simple calculation. A couple of health sectors utilize decision model systems yet are,
 78 as it were, limited. They can address simple inquiries like "What is the ordinary time of patients who have
 79 coronary disease?" "What number of therapeutic techniques had achieved crisis facility stays longer than
 80 10 days?", "Recognize the female patients who are single, more than 30 years old, and who have been
 81 treated for coronary sickness." However they can't respond to complex inquiries like "Given patient

82 records, foresee the probability of patients getting a coronary disease." Clinical decisions are as often as
83 possible made subject to experts' impulse and experience rather than on the learning rich data concealed
84 in the database.

85 This preparation prompts bothersome tendencies, botches and super helpful costs which impacts the
86 idea of care provided for patients. The proposed structure that coordinates the clinical decision help with
87 PC based patient records could reduce therapeutic errors, overhaul tolerant security, decrease
88 bothersome practice assortment, and improve getting result. This suggestion is promising as data
89 modeling and analysis tool like data mining can make a learning rich condition which can help to in a
90 general sense improve the idea of clinical decisions.

91 In this fast moving world people need to continue with an extravagant life so they work like a machine to
92 win some portion of money and continue with a pleasant life appropriately in this race they disregard to
93 manage themselves, because of this there sustenance affinities change in their entire lifestyle change, in
94 this sort of lifestyle they are logically stressed they have heartbeat, sugar at a young age and they don't
95 give enough rest for themselves and eat what they get and they even don't overemphasize the idea of the
96 sustenance whenever cleared out the go for their own special prescription in light of all these little
97 indiscretion it prompts a significant threat that is the coronary disease [7]. On account of this people go to
98 therapeutic administrations experts but the prediction made by them isn't 100% definite [25].

99 Quality facility proposes diagnosing patients precisely and controlling medications that are convincing.
100 Poor clinical decisions can incite tragic outcomes which are along these lines unsatisfactory. Medicinal
101 centers ought to in like manner limit the cost of clinical tests. They can achieve these results by using
102 fitting PC based information or decision support system.

103 The treatment cost of heart disease is not affordable by most of the patients, and the Clinical decisions
104 are often made based on doctors' intuition and experience rather than on the knowledge-rich data hidden
105 in the database. This practice leads to unwanted biases, errors and excessive medical costs which
106 affects the quality of service provided to patients. The proposed model for Heart Disease Prediction using
107 Data Mining Classification Techniques reduces medical errors, enhances patient safety, decrease
108 unwanted practice variation, reduce treatment cost and improves patient outcome. This suggestion is
109 promising as data modeling and analysis tools have the potential to generate a knowledge-rich
110 environment which can help to significantly improve the quality of clinical decisions [32].

111 2. LITERATURE REVIEW

112 This part goes for investigating the different information mining methods presented as of late for coronary
113 illness expectation. The man-made brainpower methods centering K-closest neighbor (KNN), Naive
114 Bayes and Decision tree will be presented. Recently distributed papers in displaying survival will be talked
115 about and the recommendations for another strategy are introduced

116 2.1 Theoretical and Empirical Review

117 Various information mining systems have been utilized in the analysis of CVD over various Heart illness
118 datasets. A few papers utilize just a single method for conclusion of coronary illness and different
119 scientists utilize more than one information mining technique for the finding of coronary illness.

120 In [23,27] Jyoti et.al presented three classifiers Decision Tree, Naïve Bayes and Classification by
121 methods for gathering to break down the proximity of coronary sickness in patients. Request by methods
122 for bundling: Clustering is the route toward social occasion relative segments. This framework may be
123 used as a preprocessing adventure before urging the data to the portraying model. Preliminaries were
124 driven with WEKA 3.6.0 gadget Enlightening list of 909 records with 13 particular properties. All properties
125 were made supreme and anomalies were made due with straightforwardness. To update the desire for
126 classifiers, innate request was joined. Observations show that the Decision Tree data mining technique
127 beats other two data mining methods in the wake of intertwining feature subset assurance yet with high
128 model improvement time.

129
130 Jyoti Soni et.al in [23] Showed use of Data Mining Technique in Healthcare and Prediction of Heart
131 Attacks. The potential use of request based data mining strategies, for instance, rule based Decision
132 Tree, Naïve Bayes and Artificial Neural Network to the gigantic volume of social protection data. Tanagra
133 data burrowing instrument was used for exploratory data examination, AI and quantifiable learning
134 estimations. The arrangement enlightening record included 3000 events with 14 unmistakable
135 characteristics. The cases in the dataset were addressing the eventual outcomes of different sorts of
136 testing to envision the precision of coronary disease. The introduction of the classifiers was surveyed and
137 their results were bankrupt down. The delayed consequences of examination relied upon 10 ten times
138 cross-endorsements. The relationship made among these request computations out of which the Naive
139 Bayes figuring demonstrated better execution.
140
141 [27] Nidhi et.al discernments revealed that the Neural Networks with 15 characteristics improved in
142 examination with other data mining frameworks [27]. The investigation concentrate assumed that
143 Decision Tree technique showed better execution with the help of innate figurings using included subset
144 assurance. This examination work furthermore proposed a model of Intelligent Heart Disease Prediction
145 structure using data mining frameworks explicitly Decision Tree, Naïve Bayes and Neural Network. An
146 aggregate of 909 records were obtained from the Cleveland Heart Disease database. The results
147 declared in the investigation work guarded the better execution of Decision Tree methodology with 99.6%
148 accuracy using 15 qualities. In any case, Decision tree technique in mix with inherited estimation the
149 introduction declared was 99.2% using 06 qualities.
150
151 In [8,9] Chaitrali et.al exhibited that Artificial Neural Network outmaneuvers other data mining
152 methodology, for instance, Decision Tree and Naïve Bayes. In this investigation work, Heart disorder
153 desire system was made using 15 characteristics [8,9]. The investigation work included two extra
154 properties weight and smoking for capable finish of coronary sickness in making convincing coronary
155 disease desire system.
156
157 [2] Abhishek et.al ask about work was intended to structure a judicious model for coronary disease
158 recognizable proof using data mining methodologies from Transthoracic Echocardiography Report
159 dataset that is fit for improving the relentless nature of coronary sickness examination using
160 echocardiography. The models depended on the preprocessed Transthoracic Echocardiography dataset
161 with three assorted coordinated AI figurings J48 Classifier, Naïve Bayes and Multilayer Perception using
162 WEKA 3.6.4 AI programming. The display of the models was surveyed using the standard estimations of
163 precision, exactness, audit and F-measure. The best model to foresee patients with coronary sickness
164 radiated an impression of being a J48 classifier realized on picked qualities with a course of action
165 exactness of 95.56%. From a total of 15 properties that were available, 8 characteristics that were
166 extraordinarily appropriate in foreseeing coronary ailment from Transthoracic Echocardiography dataset
167 were picked in the examination work.
168
169 [31] Researchers in year 2013 showed Hybrid Intelligent Techniques for the figure of coronary ailment.
170 Some Heart Disease gathering system was researched in this examination and shut with legitimization
171 noteworthiness of data mining in coronary sickness end and course of action. Neural Network with
172 separated getting ready is helpful for sickness conjecture in starting time and the extraordinary execution
173 of the structure can be gotten by preprocessed and institutionalized dataset. The game plan precision can
174 be improved by decline in features.
175
176 [47] Vikas et.al, in their investigation work used three standard data mining figuring's CART (Classification
177 and Regression Tree), ID3 (Iterative Dichotomized 3) and decision table (DT) removed from a decision
178 tree or rule based classifier to develop the conjecture models using a greater dataset. Discernment
179 showed that presentation of CART computation was better when differentiated and other two course of
180 action procedures.
181
182 V. Manikandan et.al in [46] recommended that association standard mining is used to remove the thing
183 set relations. The data game plan relied upon MAFIA counts which achieved better precision. The data
184 was surveyed using entropy based cross endorsement and bundle strategies and the results were

185 considered. MAFIA (Maximal Frequent Item set Algorithm) used a dataset with 19 characteristics and the
186 goal of the examination work was to have exceedingly definite audit estimations with bigger measures of
187 precision.

188
189 Beant et.al in [6] circulated an investigation paper in IJRITCC "Review on Heart Disease using Data
190 Mining Techniques". The maker referenced created by gigantic number of experts and investigated
191 diverse data mining strategies reliant on execution and accuracy.

192
193 Methaila et.al [3] in their examination work focused on using different counts and mixes of a couple of
194 target qualities for amazing heart ambush figure using data mining. Decision Tree has beaten with
195 99.62% precision by using 15 characteristics. Moreover the exactness of the Decision Tree and Bayesian
196 Classification further improves in the wake of applying inherited computation to diminish the genuine data
197 size to get the perfect subset of value satisfactory for coronary disease estimate.

198
199 The experts [19] proposed a model for desire for coronary ailment using J48, Bayes Net, and Naïve
200 Bayes, Simple CART and REPTREE Algorithms using understanding educational accumulation from
201 Medical Practitioners.

202
203 Appraisal of the disorder matrix showed that J48, REPTREE and SIMPLE CART exhibit a figure model of
204 89 cases with a peril factor positive for heart attacks. The strategies immovably prescribed that data
205 mining counts can foresee a class for judgments.

206
207 B.Venkatalakshmi et.al [5] played out an examination on coronary disease finding using data mining
208 methodology Naïve bayes and Decision Tree techniques. Different sessions of examinations were
209 coordinated with the proportional datasets in WEKA 3.6.0 contraption. Instructive gathering of 294
210 records with 13 attributes was used and the results revealed that the Naïve Bayes beat the Decision tree
211 frameworks.

212
213 The synopsis of looked into writing alongside the quantity of properties utilized for the forecast of
214 Cardiovascular Disease (CVD) is given in table beneath

215
216 **Table 1.0: Table shows different data mining techniques used in the diagnosis of Heart**
217 **disease.**

Author/Researcher	Data Mining Technique used	Year	Number of Attributes Selected
Jyoti Sonia, et.al.	Naïve Bayes, Decision Tree, KNN	2011	13
K.Srinivas et.al.	Naïve Bayes, knn and D.L.	2011	14
Nidhi Bhatla et.al.	Naïve Bayes, Decision Tree, Neural Network	2012	15 and 13
Chaitrali S.Dangare & Sulabha S.Apte	Naïve Bayes, Decision Tree, Neural Network	2012	13 and 15

Abhishek Taneja	Naïve Bayes, J48 unpruned tree, Neural Network	2013	15 and 8
R. Chitra et. al.	Hybrid Intelligent Techniques	2013	15
Vikas Chaurasia, et.al.	CART, ID3, Decision Table	2013	Not mentioned
V. Manikandan et al.	K-Mean based on MAFIA, K-Mean based on MAFIA with ID3, K-Mean based on MAFIA with ID3 and C4.5	2013	19
Beant Kaur & Williamjeet	Papers Reviewed	2014	Nil
Aditya Methaila et. al.	Decision Tree, Naive Bayes, Neural Network, Genetic Algorithm	2014	15 and 16
Hlaudi Daniel Masethe, Mosima Anna Masethe	J48, REPTREE, Naïve Bayes, Bayesnet, Simple CART	2014	15
B.Venkatalakshmi and M.V Shivsankar	Decision Tree and Naïve Bayes	2014	13

218

219 2.2 Artificial Intelligence Techniques in Heart Disease Prediction

220 Information mining has been generally connected in the therapeutic field as this give enormous measure
221 of information. Different scientists had connected the various information mining procedures on social
222 insurance information [11]. connected 5 arrangement calculations for example choice tree, fake neural
223 system, strategic relapse, Bayesian systems and credulous Bayes and stacking-sacking technique for
224 structure arrangement models and thought about the precision of the plain and outfit model to foresee
225 whether a patient will return to a medicinal services Center or not. From results, the best order model
226 relies upon informational collection for example ANN in 3M informational index, choice tree in 6M and
227 strategic relapse in 12M informational collection [23, 26] contrasted the information mining and

228 conventional insights and expresses a few focal points of mechanized information framework. This paper
229 gives an outline of how information mining is utilized in social insurance and medication. [29] decides if an
230 individual is fit or unfit dependent on authentic and constant information utilizing grouping calculations that
231 is K-means and D-stream are connected. The presentation and precision of D-stream calculation is more
232 than K-implies [4] utilized choice tree to construct an arrangement model for anticipating representative's
233 exhibition. To manufacture a characterization model CRISP-DM was received.

234 In light of execution, work title is most grounded trait then college pursued by different qualities. [22] Ho
235 et.al proposed a choice emotionally supportive network to recognize a hazard score for foreseeing the
236 coronary illness. A cooperative arrangement calculation utilizing hereditary methodology is proposed for
237 forecast. Exploratory outcomes demonstrate that the majority of the classifier standards help in best
238 forecast of coronary illness. Garchchopogh et al [15] clarified the usage of therapeutic information mining
239 in deciding when we ought to perform medical procedure. The choice tree calculation intended for this
240 investigation produces right expectation for over 86.25% tests cases [12], connected choice tree J48 to
241 locate the shrouded examples for Classification of ladies wellbeing illness (Fibroid). Choice tree J48
242 calculation is executed utilizing WEKA 3.75 information excavator. It ordered the information into
243 effectively and erroneously occasion. Jabbar et.al [22,32] assessed the ease of use of administered
244 information mining to anticipate dietary quality. Fake Neural Networks and Decision trees were utilized.
245 The ANN had a marginally higher precision than the choice tree. Sundar et.al in [41, 45] examined the
246 exhibition of the Naive Bayes and WAC (weighted affiliated classifier) to foresee the probability of patients
247 getting a coronary illness. This system uses CRISP-DM procedure to fabricate the mining models.
248 These techniques delineate that the WAC gives most elevated level of right forecasts for diagnosing
249 patients with a coronary illness. Al-Radaideh [4], inspected and thought about the adequacy of neural
250 systems, choice tree, strategic relapse, memory based thinking and the troupe model in assessing
251 whether the awful obligation is probably going to be reimbursed. They utilized SAS Enterprise Miner to
252 manufacture beginning and last model.

253 PC reproduction demonstrates that the strategic relapse, neural system model and troupe model
254 delivered best generally speaking grouping precision. Koç et al [24] connected ANN and strategic relapse
255 to anticipate if the customer will buy in a term store or not subsequent to promoting effort. ANN orders
256 84.4% information accurately while calculated relapse characterizes 83.63% information effectively
257 however LR takes 54 seconds and ANN takes 11 seconds to run. Along these lines, with more
258 information and higher dimensional element space, utilizing ANN will be progressively productive. Fartash
259 et.al [13] contrasted the different order calculations with anticipate the transmission capacity use design in
260 various time interims among various gatherings of clients in the system correlation of various
261 characterization calculations including. Choice Tree and Naïve Bayesian utilizing Orange is finished. The
262 Decision Tree calculation accomplished 97% exactness and effectiveness in foreseeing the required data
263 transfer capacity inside the system. [35] gave a total examination of various information mining
264 characterization procedures that incorporates choice tree, Bayesian systems, k-closest neighbor classifier
265 and fake neural system.

266 Execution of these calculations is investigated dependent on precision, capacity to deal with undermined
267 information and speed. [43], in this paper, the learning is recovered from a tremendous measure of
268 information about understudies utilizing a productive strategy of information mining to assist the
269 organization with making a brisk choice. S.Asha Rani and Dr.S.Hari Ganesh [33] investigated the
270 proficiency of various grouping calculation in information mining utilizing blood transfusion dataset. The
271 examination of different calculations in order is done .The calculation Random tree has indicated 93.18%
272 precision inside brief length when contrasted and different calculations in characterization. Pushpalata
273 Pujari [30] portrayed the exhibition examination of various information mining classifiers, for example,

274 classifiers Logistic Regression, SVM and Neural Network when highlight determination on binomial
275 informational index. The characterization execution of all classifiers depends on different factual
276 execution estimates like precision, particularity and affectability. Increase diagram and R.O.C graph are
277 additionally used to gauge the exhibitions of the classifiers.

278 Clinical databases have gathered enormous amounts of data about patients and their ailments. The term
279 Heart illness includes the assorted sicknesses that influence the heart. Coronary illness is the real reason
280 for setbacks on the planet. The term Heart illness includes the assorted ailments that influence the heart.
281 Coronary illness kills one individual at regular intervals in the United States [48]

282 **2.3 Data Mining Review**

283 Notwithstanding the way that data burrowing has been around for more than two decades, its potential is
284 simply being recognized now. Data mining solidifies quantifiable examination, AI and database
285 advancement to think hid models and associations from gigantic databases Fayyad portrays data mining
286 as "a method of nontrivial extraction of saw, in advance darken and possibly profitable information from
287 the data set away in a database" [44] describes it as "a method of assurance, examination and showing
288 of colossal measures of data to discover regularities or relations that are at first cloud with the purpose of
289 getting clear and accommodating results for the owner of database" [17]

290 Data mining uses two systems: oversight and unsupervised learning. In oversight learning, a planning set
291 is used to learn model parameters however in unsupervised adjusting no arrangement set is used (e.g., k
292 means grouping is unsupervised) [28]. Each datum mining methodology fills another need dependent
293 upon the exhibiting objective. The two most ordinary showing goals are gathering and figure. Game plan
294 models predict full scale names (discrete, unordered) while estimate models envision steady regarded
295 limits Decision Trees and Neural Networks use portrayal counts while Regression, Association Rules and
296 Clustering use desire figurings [10]. Decision Tree figurings consolidate CART (Classification and
297 Regression Tree), ID3 (Iterative Dichotomized [10] and C4.5. These computations shift in selection of
298 parts, when to keep a center point from part, and undertaking of class to a non-split center [11] CART
299 uses Gini rundown to check the dirtying impact of a package or set of getting ready tuples [17]. It can
300 manage high dimensional unmitigated data.

301 Decision Trees can moreover manage constant data (as in backslide) yet they ought to be changed over
302 to straight out data. Gullible Bayes or Bayes' Rule is the explanation behind a few, AI and data mining
303 methods [42]. The standard (estimation) is used to make models with insightful capacities. It gives better
304 methodologies for researching and getting data. It gains from the "evidence" by figuring the association
305 between the goal (i.e., subordinate) and other (i.e., independent components. Neural Networks includes
306 three layers: input, concealed and yield units (factors). Relationship between data units and concealed
307 and yield units rely upon centrality of the doled out worth (weight) of that particular data unit. The higher
308 the weight the more huge it is. Neural Network computations use Linear and Sigmoid trade limits. Neural
309 Networks are sensible for setting up a ton of data with few wellsprings of information. It is used when
310 various systems are unacceptable.

311 **3. RESEARCH DESIGN**

312 In this examination, three information digging procedures for prescient information mining assignment
313 were utilized, that incorporates Decision tree, K-NN and Naïve Bayes. These strategies were utilized for
314 producing learning to settle on it valuable for basic leadership. Every strategy delivered various outcomes
315 to arrange the locale into centered or non-centered states involving the accessible factors in dataset .The
316 experimentation was performed utilizing WEKA programming interface.

317 **3.1. Proposed Model**

318 The proposed engineering of coronary illness forecast framework is given beneath

319

320

321

322

323

324

325

326

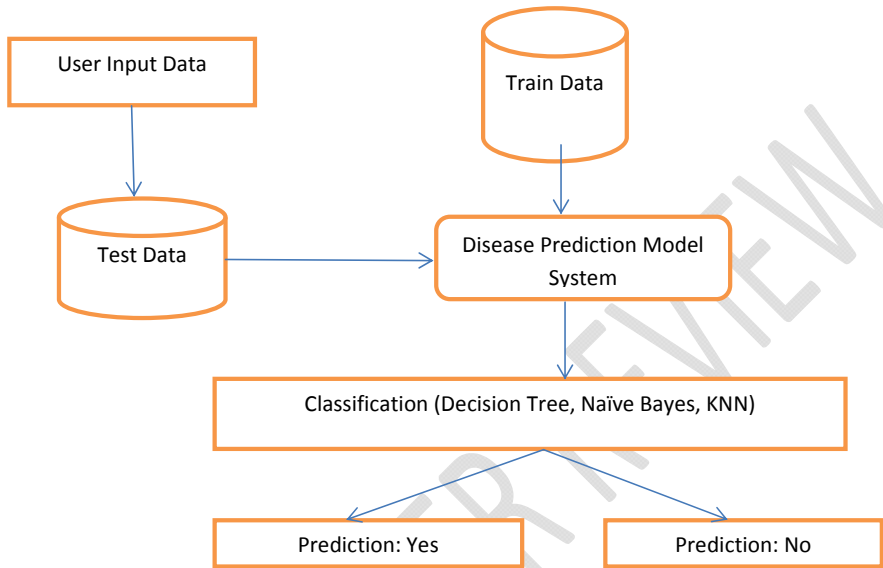
327

328

329

330

331



331 **Figure 2.0: The System Model**

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

It comprises of preparing dataset and client contribution as the test dataset. Weka information mining apparatus with programming interface was utilized to actualize the coronary illness forecast framework. The source code of Weka is in java. The framework is planned with java swing and use Weka programming interface to call the various techniques for Weka. The segments utilized are cases, various classifiers and strategies for assessment. Administered learning strategy is utilized here. A directed learning calculation examinations the preparation information and derives a capacity from the named preparing set. It tends to be utilized for mapping new models. The preparation information got from ucl repository coronary illness database is the preparation model. This preparation information comprise of the class name and its comparing esteem. Credulous Bayes, KNN and choice tree classifiers are administered learning calculations. They gain from the given preparing models. At the point when another case with same characteristics as in preparing information with various qualities other than those in the preparation model comes, these calculations effectively characterize the new case dependent on the speculation made from the preparation set. Gullible Bayes, KNN and choice tree classifiers are order the new perception into two classifications based on preparing dataset. The preparation dataset is in the ARFF group. The preparation set comprises of 296 traits including the class characteristic. Coronary illness forecast framework acknowledges contribution from the client through a graphical UI. Every one of the traits required for grouping is gotten from a content field. The graphical UI is fabricated utilizing swing. The following procedure is to move the client information acquired from graphical UI into a record of CSV (Comma isolated Value) augmentation. At that point the CSV record is changed over into ARFF document. Weka programming interface give local strategies to changing over from CSV to ARFF. The changed over client information is treated as test information. The test informational index will contain every one of the characteristics of preparing dataset. In the event that the client did not enter a property estimation a '?' will be relegated at the estimation of that comparing trait. Weka will deal with this missing worth. This test information is kept running on Naive Bayes, KNN and choice tree calculation. These

356 calculations order the occasions got from the client and foresee the opportunity to have coronary illness.
357 Netbeans IDE is utilized to code in Java.

358 3.1.1 Decision Tree

359 J48 choice tree is an open source java execution of regularly known C4.5 regulated arrangement
360 calculation in WEKA. It is an advancement and expansion of ID3 calculation created by Quinlan. It is a
361 portion between data increase and its part data.

362 3.1.2 Naïve Bayes

363 This technique depends on probabilistic information. The gullible Bayes principle yields probabilities for
364 the anticipated class of every individual from the arrangement of test example. Gullible Bayes depends on
365 administered learning. The objective is to foresee the class of the experiments with class data that is
366 given in the preparation information.

367 The quality "Analysis" is distinguished as the anticipated characteristic with worth "1" for patients with
368 coronary illness and worth "0" for patients with no coronary illness. "Quiet Id" is utilized as the key; the
369 rest are info traits. It is expected that issues, for example, missing information, conflicting information, and
370 copy information have all been settled.

371 Predictable attribute

372 1. Diagnosis (value 0: <50% diameter narrowing (no heart disease); value 1: >50% diameter narrowing (has heart
373 disease))

374 Key attribute

375 Patient Id – Patient's identification number

376 Input attributes (Description of attributes)

- 377 a. Age in Year
- 378 b. Sex (value 1: Male; value 0: Female)
- 379 c. Chest Pain Type (value 1: typical type 1 angina, value 2: typical type angina, value 3: non angina pain; value 4:
380 asymptomatic)
- 381 d. Fasting Blood Sugar (value 1: >120 mg/dl; value 0: <120 mg/dl)
- 382 e. Restecg – resting electrographic results (value 0: normal; value 1: having ST-T wave abnormality; value 2:
383 showing probable or definite left ventricular hypertrophy)
- 384 f. Exang - exercise induced angina (value 1: yes; value 0: no)
- 385 g. Slope – the slope of the peak exercise ST segment (value 1: unslping; value 2: flat; value 3: downsloping)
- 386 h. CA – number of major vessels colored by floursopy (value 0-3)
- 387 i. Thal (value 3: normal; value 6: fixed defect; value 7: reversible defect)
- 388 j. Trest Blood Pressure (mm Hg on admission to the hospital)
- 389 k. Serum Cholestrol (mg/dl)
- 390 l. Thalach – maximum heart rate achieved
- 391 m. Oldpeak – ST depression induced by exercise
- 392 n. Smoking – (value 1: past; value 2: current; value 3: never)
- 393 o. Obesity – (value 1: yes; value 0: no)Execution of Bayesian Classification

394 Implementation of Bayesian Classification

Comment [m5]: These parameters are only quoted with others in the abstract

395 The Naïve Bayes Classifier strategy is especially fit when the dimensionality of the sources of info is high.
396 In spite of its effortlessness, Naive Bayes can frequently outflank increasingly advanced grouping
397 techniques. Gullible Bayes model recognizes the attributes of patients with coronary illness. It
398 demonstrates the likelihood of each information trait for the anticipated state.

399 Why favored Naive Bayes calculation

400 Credulous Bayes or Bayes' Rule is the reason for some, AI and information mining techniques. The
401 standard (calculation) is utilized to make models with prescient abilities. It gives better approaches for
402 investigating and getting information.

403 Why preferred naive Bayes implementation:

- 404 a. At the point when the information is high.
- 405 b. At the point when the properties are free of one another.
- 406 c. When we need increasingly proficient yield, when contrasted with different strategies yield

407 Bayes Rule

408 A restrictive likelihood is the probability of some end, C, given some proof/perception, E, where a reliance
409 relationship exists among C and E.

410 This likelihood is meant as $P(C|E)$ where

$$411 P(C/E) = P(E/C) P(C)/p(E)$$

412 3.1.3 K-NN – k-Nearest Neighbors

413 K-NN is a kind of occasion based learning, or apathetic realizing, where the capacity is just approximated
414 locally and all calculation is conceded until characterization. K-NN arrangement, the yield is a class
415 participation. An article is ordered by a dominant part vote of its neighbors, with the item being doled out
416 to the class most basic among its k closest neighbors (k is a positive whole number, normally little). In the
417 event that $k = 1$, at that point the item is just appointed to the class of that solitary closest neighbor.

418 3.2 Data Source

419 Clinical databases have aggregated enormous amounts of data about patients and their ailments. The
420 term Heart infection includes the assorted illnesses that influence the heart. Coronary illness is the real
421 reason for setbacks on the planet. Coronary illness kills one individual at regular intervals in the United
422 States. Coronary illness, Cardiomyopathy and Cardiovascular infection are a few classifications of heart
423 ailments. The expression "cardiovascular malady" incorporates a wide scope of conditions that influence
424 the heart and the veins and the way where blood is siphoned and coursed through the body.
425 Cardiovascular ailment (CVD) results in extreme disease, incapacity, and passing.

426

427 740 Record sets with therapeutic qualities will be gotten from a freely accessible database for coronary
428 illness from AI archive will be utilized, that is Cleveland, Hungary, Switzerland and the VA Long Beach
429 Heart Disease databases with the assistance of the datasets, and the examples noteworthy to the heart
430 assault forecast are separated.

431 3.3. Processing and Analysis

432 The records will be split into two datasets: training dataset and testing dataset. To avoid bias, the records
433 for each set will be selected randomly. The table below shows the description of dataset selected for this
434 work.

435

Comment [m6]: In the neural network it is necessary to specify the input variables and the output variable as a table

Comment [m7]: These data you're referring to are missing

436 **Table 2.0 Dataset Description**

437

Dataset	No. Of Attributes		Instances	Classes
	A	B		
Health Services Data	13	15	740	2

438

439 The model was developed and the first 13 input attributes were used then two more other attributes which
 440 are **obesity and smoking** were added, as these attributes are considered as important attributes for
 441 heart disease.

442 Also the deaths due to heart disease in many countries occur due to: work overload, mental stress and
 443 many other problems, these are the other factor attributes we had considered in observing the prediction
 444 change.

445 Most of the research papers referred upon have used 13 input attributes for prediction of Heart disease,
 446 to get more appropriate results two more important attributes were added that is obesity and smoking.

447 Healthcare industry is generally "information rich", but unfortunately not all the data are mined which is
 448 required for discovering hidden patterns & effective decision making- that's why we looked for more other
 449 attributes which contribute to the heart disease

450 **4. EXPERIMENTS AND RESULTS**

451 The exhibition survey of a model for Heart Disease Prediction, utilizing Decision Trees, Naive Bayes, and
 452 KNN displaying strategies were assessed concerning AI calculations. The targets of the trials were: To
 453 break down the exhibition for the coronary illness expectation procedures, and portray how to improve
 454 their forecast power, Efficient and precise in coronary illness forecast; To examine the centrality of
 455 symptomatic highlights that best depict coronary illness information utilizing information mining strategies.
 456 The Experiments demonstrated that the proposed technique gives the exact conclusion of coronary
 457 illness than the current strategies

458 **4.2 Experimental Setup**

459 This exploration utilized classifiers given by Weka. The informational indexes were utilized as contribution
 460 to three AI calculations; Naive Bayes (NB), K-Nearest Neighbor (KNN) and Decision Trees (DT). The
 461 investigations began with 13 info properties and then 15 information traits esteems. Investigation results
 462 were then exhibited in tables, broke down and deciphered as definite

463 **4.3 Experimental Results and Analysis**

464 The test results and investigation accomplished for this examination was spoken to as in the tables
 465 beneath. The exploration system has been clarified in the past area. For the tests, different information
 466 mining grouping strategies were connected on the dataset. In this investigation, WEKA AI apparatus for
 467 information mining was utilized to achieve the goals. The level of precision rate and mistake rate of
 468 information mining Classification methods was utilized as the estimation parameters for investigation.
 469 These parameters recommend that the classifier having a higher exactness rate and lower estimation of
 470 blunder rate arrange the dataset in very amended way and the other way around. In this examination, the
 471 information was right off the bat isolated into preparing information and testing information. The
 472 preparation set was utilized to build the classifier and test set utilized for approval. In this examination, the
 473 level of dataset utilized for preparing and testing information were 40% and 60% individually. At that point,
 474 the 10 overlay cross approval technique was connected to create the classifiers utilizing recently
 475 referenced AI apparatuses. At last the outcomes were archived as far as precision rate and mistake rates.
 476 The outcomes were appeared in table beneath:

477 Displays the results for classification techniques applied on health facility services data in WEKA
 478 Considering accuracy and error rates as performance measure the classification techniques with highest
 479 accuracy are obtained for health facility Services data in given different techniques used.

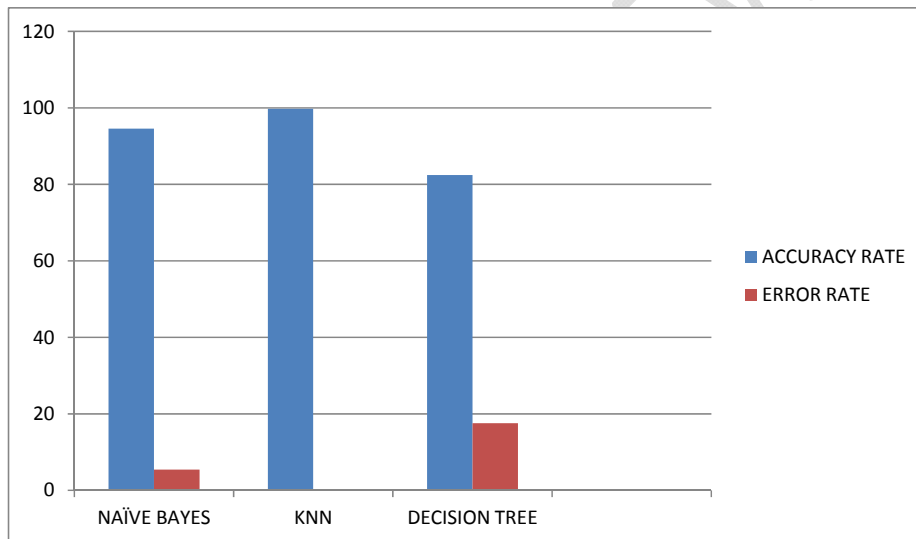
480
 481
 482

Table 3.1 Results Using WEKA API

Technique Used	Accuracy Rate		Error Rate	
	13 Attributes	15 Attributes	13 Attributes	15 Attributes
Naive Bayes	90.76	94.59	9.24	5.41
Decision Tree	97.07	99.77	2.93	0.23
KNN	79.28	82.43	20.72	17.57

483
 484
 485
 486

The graph below displays the performance analysis of classification techniques for 15 attributes using WEKA. The best classifier for this particular data set will then be chosen.



487
 488

Fig 3.1 Performance analysis of classification techniques using WEKA API

489
 490
 491
 492
 493
 494
 495
 496
 497
 498

4.4. Results

The dataset comprised of all 740 records in Heart illness database. The records were then divide into two, one utilized for preparing comprises of 296 records and another for testing comprises of 444 records. The information mining apparatus Weka 3.6.6 was utilized for trial. At first dataset contained a few fields, in which some incentive in the records was absent. These were recognized and supplanted with most fitting qualities utilizing Replace Missing Values channel from Weka 3.6.6. The Replace Missing Values channel checks all records and replaces missing qualities with mean mode technique. This procedure is known as Data Pre-Processing. After pre-handling the information, information mining order procedures, for example, KNN, Decision Trees, and Naive Bayes were connected. A disarray lattice is acquired to figure the exactness of arrangement. A perplexity grid demonstrates what number of occurrences has

499 been doled out to each class. In our analysis we have two classes, and in this manner we have a 2x2
500 perplexity network

501 Class A= YES (Has coronary illness)

502 Class B = (No coronary illness)

503 **Table 3.2 a Disarray Network**

	A(Has heart disease)	B(Has no heart disease)
A(has heart disease)	TP	FN
B(has no heart disease)	FP	TN

504

505 TP (True Positive): It indicates the quantity of records named genuine while they were in reality evident.
506 FN (False Negative): It signifies the quantity of records delegated false while they were in reality evident.
507 FP (False Positive): It indicates the quantity of records named genuine while they were in reality false. TN
508 (True Negative): It means the quantity of records named false while they were in reality false. Results got
509 with 13 properties are determined beneath

510 **Table 3.3 Confusion Networks Got For Three Arrangement Techniques with 13 Qualities**

511 **Confusion matrix for Naive Bayes:**

	A	B
A	182	13
B	28	221

512

513 **Confusion matrix for Decision Trees:**

	A	B
A	205	6
B	7	226

514

515 **Confusion matrix for KNN:**

	A	B
A	160	30
B	62	192

516

517 Results obtained by adding two more attributes i.e. total 15 attributes are specified below.

518 Table 3.4 Confusion matrixes obtained for three classification methods with 15 attributes

519 **Confusion matrix for Naive Bayes:**

	A	B
A	187	11
B	13	233

520

521 **Confusion matrix for Decision Trees:**

	A	B
--	---	---

A	168	0
B	1	275

522

523 **Confusion matrix for KNN**

	A	B
A	153	36
B	42	213

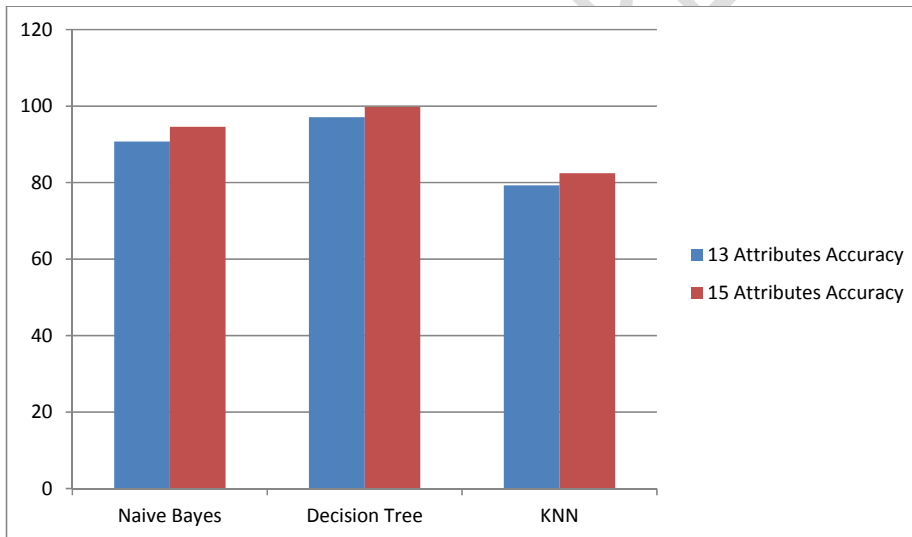
524

525 **Table 3.5 shows accuracy for different classification methods with 13 input attributes and 15 input**
 526 **attributes values.**

Classification Techniques	Accuracy with	
	13 Attributes	15 Attributes
Naive Bayes	90.76	94.59
Decision Tree	97.07	99.77
KNN	79.28	82.43

527

528 The accuracy of each of the method is plotted on a graph as below:



529

530 **Figure 3.2: Graphical representation of accuracy for each method.**

531 **5. CONCLUSION AND FUTURE WORK**

532 **5.1 Knowledge Contributions**

533 This research proposed that using more attributes values for training and testing data sets improves
 534 accuracy performance of classifiers.

Comment [m8]: The content does not match the title

535 This research improved results on accuracy with increase in number of attributes. One of the research
536 objectives was to analyze the performance for the heart disease prediction techniques, and describe how
537 to improve their prediction power, Efficient and accurate in heart disease prediction. The research
538 proposed the decision tree, the prediction techniques used for training and testing affect accuracy
539 measure.

540 Another objective was to develop a novel approach based on the 3 methods above and compare its
541 performance with the existing methods. The success of classifiers therefore depends on the relevance of
542 features between training and testing data. This research proposed use of 14 attributes.

543 Another research objective was to analyze the significance of diagnostic features that best describe heart
544 disease data using data mining techniques Most of the attributes used had high accuracy values.

545 5.2 Conclusion

546 This examination presents a coronary illness expectation framework model for diagnosing coronary
547 illness in prior stage. The framework model uses information mining systems, for example, Decision trees,
548 Naive Bayes and KNN alongside Weka programming interface to call various strategies for Weka.
549 Various information mining characterization procedures were connected on the particular dataset. The
550 order procedure inside the framework model is performed with traits like age, sex, heart beat rate,
551 cholesterol level and so on. The expectation is then made dependent on this arrangement results. Here
552 the AI ability of the PC framework can be stretched out into the medicinal field. The proposed framework
553 model is best for lessening the blunder event during the illness expectation. In this examination the
554 exactness and precision of three unique classifiers are estimated, the outcome demonstrates choice tree
555 arrangement has high precision and less mistake rate, Naïve Bayer characterization strategy creates
556 preferred outcome over KNN grouping. This investigation can assist scientists with getting productive
557 outcomes in the wake of knowing the best order strategy for this specific dataset. The general target of
558 the examination was to foresee all the more precisely the nearness of coronary illness. In this exploration,
559 more information characteristics weight and smoking were utilized to get progressively precise outcomes.

560

561 5.3 Future Work

562 Heart Disease Prediction using Data Mining Classification Techniques can be used largely in hospital
563 based sectors for disease prediction, However, there is need for more research to be done on contextual
564 knowledge being incorporated as part of feature selection and model creation for specific domains where
565 precise context, which does not depend on attributes needs to be used in learning and prediction is
566 required also. There is need to experiment the prediction models with real live testing of heart disease.
567 This research can also be enhanced by experiment with more attributes in training and testing data sets.
568

569 REFERENCES

- 570 1. Abdullah H. Wahbeh, "A Comparison Study between Data Mining Tools over some Classification
571 Methods" (IJACSA) International Journal of Advanced Computer Science and Applications, Special
572 Issue on Artificial Intelligence.
- 573 2. Abhishek Taneja, Heart Disease Prediction System Using Data Mining Techniques; Oriental Journal
574 of computer science & Technology ISSN: 0974-6471 December2013.
- 575 3. Aditya Methaila, Early Heart Disease Prediction Using Data Mining Techniques; CCSEIT, DMDB,
576 ICBB, MoWiN, AIAP pp. 53–59, 2014.
- 577 4. Al-Radaideh "Using data mining techniques to build a classification model for predicting employee's
578 performance", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol.
579 3, No. 2, 2014.
- 580 5. B.Venkatalakshmi, M.V Shivsankar, Heart Disease Diagnosis Using Predictive Data mining;
581 International Journal of Innovative Research in Science, Engineering and Technology Volume 3,
582 Special Issue 3, March 2014.
- 583 6. Beant Kaur and Williamjeet Singh., " Review on Heart Disease Prediction System using Data Mining
584 Techniques", IJRITCC ,October 2014.

Comment [m9]: I do not see any tool clearly available to the clinicians

Comment [m10]: Missing pages, year and edition

Comment [m11]: Missing pages

Comment [m12]: Missing pages

Comment [m13]: Missing pages

- 585 7. Blake, C.L.,Mertz, C.J.“UCI Machine Learning Databases”
586 8. Chaitrali S. Dangare, Sulabha S. Apte, —Improved Study of Heart Disease Prediction System using
587 Data Mining Classification Techniques; International Journal of Computer Applications (0975 – 888)
588 Volume 47– No.10, June 2012
589 9. Chaitrali S.Danagre, Sulabha S.Apte, Ph.D, Improved Studyof Heart Disease Prediction Systemusing
590 Data mining Classification Techniques,IJCA,June 2012.
591 10. Charly, K.: “Data Mining for the Enterprise”, 31st Annual Hawaii Int. Conf. on System Sciences, IEEE
592 Computer, 7, 295-304, 2014.
593 11. Choi Keunho et al. “Classification and Sequential Pattern Analysis for Improving Managerial
594 Efficiency and Providing Better Medical Service in Public Healthcare Centers” health inform res,
595 pp.67-76, June 2014
596 12. D.K, “Classification of women health disease (Fibroid) using decision tree algorithm”, International
597 Journal of Computer Applications in Engineering Science Vol.2, Issue 3, September2016. |
598 13. Fartash. Haghanikhameneh “A Comparison Study between Data Mining Algorithms over
599 Classification Techniques in Squid Dataset” International Journal of Artificial Intelligence, Autumn
600 (October) 2015, Vol. 9, |
601 14. Fayyad, Piatetsky-Shapiro, Smyth, "From Data Mining to Knowledge Discovery: An Overview", in
602 Fayyad, Piatetsky-Shapiro, Smyth, Uthurusamy, *Advances in Knowledge Discovery and Data Mining*,
603 AAI Press / The MIT Press, Menlo Park, CA, 2014. |
604 15. Garchchopogh et al, “Application of decision tree algorithm for data mining in healthcare operations:
605 A case study”, International Journal of Computer Applications Vol 52 – No. 6, August 2014. |
606 16. Global Atlas on Cardiovascular Disease Prevention and Control (PDF). World Health Organization in
607 collaboration with the World Heart Federation and the World Stroke Organization. pp. 3–18. ISBN
608 978-92-4-156437-3. |
609 17. Han, J. and Kamber, M. (2014). Data Mining: Concepts and Techniques. fourth Edition, Morgan
610 Kaufmann Publishers, San Francisco. |
611 18. Hearty “Analysis of meal patterns with the use of supervised data mining techniques-Artificial Neural
612 Network and Decision Tree”, The American Journal of Clinical Nutrition |
613 19. Hlaudi Daniel Masethe, Mosima Anna Masethe-prediction of Heart Disease using Classification
614 Algorithms; Proceedings of the World Congress on Engineering and Computer Science 2014.
615 20. Ho, T. J.: Data Mining and Data Warehousing, Prentice Hall, 2016. |
616 21. Huang, Li, Su, Watts, & Chen, 2007; Ishibuchi, Kuwajima, Nojima, 2007; Karabatak & Ince, 2009;
617 Shin et al., 2010; Wang & Hoy, 2015) |
618 22. Jabbar et al “Heart disease prediction system using associative classification and Genetic Algorithm”,
619 International Conference on Emerging Trends in Electrical, Electronics and Communication
620 Technologies-ICECIT, 2015
621 23. Jyoti Soni et.al. Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease
622 Prediction; International Journal of Computer Applications (0975 – 8887) Volume 17– No.8, March
623 2011.
624 24. Koç et al, “A comparative study of artificial neural network and logistic regression for classification of
625 marketing campaign results”, Mathematical and Computational Applications, Vol. 18, No. 3, 2013, pp.
626 392-398
627 25. Mrs.G.Subbalakshmi, “Decision Support in Heart Disease Prediction System using Naive Bayes”,
628 Indian Journal of Computer Science and Engineering. Vol. 3, No. 5, May 2014. |
629 26. Nakul Soni, Chirag Gandhi, “Application of data mining to health care”, International Journal of
630 Computer Science and its Applications. |
631 27. Nidhi Bhatla, Kiran Jyoti, “ An Analysis of Heart Disease Prediction using Different Data Mining
632 Techniques” International Journal of Engineering and Technology Vol.1 issue 8 2012. |
633 28. Obenshain, M.K: “Application of Data Mining Techniques to Healthcare Data”, Infection Control and
634 Hospital Epidemiology, 25(8), 690–695, 2014
635 29. Patil Dipti “An adaptive parameter for data mining approach for healthcare applications” (IJACSA)
636 International Journal of Advanced Computer Science and Applications, Vol. 3, No. 1, 2014. |
637 30. Pushpalata Pujari “ Classification and comparative study of data mining classifiers with feature
638 selection on binomial data set” Journal of Global Research in Computer Science, Vol. 3, No. 5, May
639 2016. |

Comment [m14]: Missing pages

Comment [m15]: Missing pages

Comment [m16]: Missing pages

Comment [m17]: Missing pages

Comment [m18]: Missing year

Comment [m19]: to review the data relating to the volume, pages,

Comment [m20]: to review the data relating to the volume, pages, year

Comment [m21]: Missing pages

Comment [m22]: Missing pages

Comment [m23]: Missing pages

Comment [m24]: to review the data relating to the volume, pages, year ...

Comment [m25]: Missing pages

Comment [m26]: Missing pages

Comment [m27]: Missing pages

- 640 31. R. Chitra, Review Of Heart Disease Prediction System Using Data Mining And Hybrid Intelligent
641 Techniques; Ictact Journal On Soft Computing, July 2013, volume: 03, Issue: 04. |
- 642 32. R.Wu, W.Peters, M.W.Morgan, "The Next Generation Clinical Decision Support: Linking Evidence to
643 Best Practice", *Journal of Healthcare Information Management*. 16(4), pp. 50 55, 2016.
- 644 33. S.Asha Rani and Dr.S.Hari Ganesh, " A comparative study of classification algorithm on blood
645 transfusion" International Journal of Advancements in Research & Technology, Volume 3, Issue 6,
646 June-2014. |
- 647 34. Saichanma et al. "The Observation Report of Red Blood Cell Morphology in Thailand Teenager by
648 Using Data Mining Technique." *Advances in hematology*, 2014. |
- 649 35. Sakshi and Prof.Sunil Khare "A Comparative Analysis of Classification Techniques on Categorical
650 Data in Data Mining" International Journal on Recent and Innovation Trends in Computing and
651 Communication Vol. 3 Issue: 8, pp.5142 – 5147
- 652 36. Sayad AT, Halkarnikar PP. Diagnosis of heart disease using neural network approach. *Int J Adv Sci*
653 *Eng Technol*. 2014;2:88–92.
- 654 37. Setiawan, *et al*," A Comparative Study of Imputation Methods to Predict Missing Attribute Values in
655 Coronary Heart Disease Data Set", *Journal in Department of Electrical and Electronic*
656 *Engineering*, Vol.21, PP. 266–269, 2008
- 657 38. Shadab Adam Pattekari and Asma Parveen, prediction system for heart disease using naïve bayes,
658 *International Journal of Advanced Computer and Mathematical Sciences*, 2012. |
- 659 39. Shanthi Mendis; Pekka Puska; Bo Norrving; World Health Organization (2011).
- 660 40. Shelly Gupta et al. "Performance Analysis of Various Data Mining Classification Techniques on
661 Healthcare Data" International Journal of Computer Science & Information Technology (IJCSIT) Vol
662 3, No 4, August 2011. |
- 663 41. Sundar et al. "Performance analysis of classification data mining techniques over heart disease
664 database", [IJESAT] International Journal of Engineering Science and Advanced Technology,
665 Volume-2, Issue-3, pp. 470 – 478. |
- 666 42. Tang, Z. H., MacLennan, J.: *Data Mining with SQL Server 2005*, Indianapolis: Wiley, 2015. |
- 667 43. Tariq O. Fadl Elsid and Mergani. A. Eltahir "An Empirical Study of the Applications of Classification
668 Techniques in Students Database" *Int. Journal of Engineering Research and Applications* ISSN:
669 2248-9622, Vol. 4, Issue 10(Part - 6), pp.01-10, October 2014
- 670 44. Thuraisingham, B.: "A Primer for Understanding and Applying Data Mining", *IT Professional*, 28-31,
671 2015
- 672 45. Umadevi, D.Sundar, Dr.P.Alli, "A Study on Stock Market Analysis for Stock Selection – Naïve
673 Investors' Perspective using Data Mining Technique", *International Journal of Computer Applications*
674 (0975 – 8887), Vol 34– No.3,2011.
- 675 46. V. Manikandan and S. Latha, "Predicting the Analysis of Heart Disease Symptoms Using Medical
676 Data Mining Methods "International Journal of Advanced Computer Theory and Engineering", Vol. 2,
677 Issue. 2, 2013.
- 678 47. Vikas Chaurasia, et al, Early Prediction of Heart Diseases Using Data Mining Techniques; Caribbean
679 *Journal of Science and Technology* ISSN 0799-3757, Vol.1,208-217, 2013.
- 680 48. World Health Organization; Cardiovascular Diseases (CVDs) Fact Sheet Reviewed June 2016
- 681 49. David W. Aha (714) 856-8779
- 682

Comment [m28]: Missing pages

Comment [m29]: Missing pages

Comment [m30]: Missing pages

Comment [m31]: Missing pages

Comment [m32]: Missing pages

Comment [m33]: Missing year

Comment [m34]: Missing pages

Comment [m35]: Missing pages

Comment [m36]: Incomplete

683

684

685