HEART DISEASE PREDICTION USING ML ALGORITHM

Satyam Gaur
Dept. B.Tech CSE
Vellore Institute of Technology
Chennai, India
satyam.gaur2020@vitstudent.ac.in

ABSTRACT. Many patients don't get proper treatment due to shortage of doctors. Proper treatment is necessary for the well-being of the people. Thus, predicting a disease using the patient's symptoms has become an important task these days. To solve this acute shortage of doctors there must be a predicting system for predicting the general diseases which would help in proper utilization of the resources. Data analysis and Machine learning can help in deciding the line of treatment to be followed by extracting knowledge from suitable databases. Healthcare facilities need to be advanced so that better decisions for patient diagnosis and treatment options can be made. In this paper, a model is proposed for predicting the disease suffered by a person by knowing the symptoms. The model uses the Logistic Regression algorithm, which assigns observations to a discrete set of classes and provides a good level of accuracy. It collects the dataof a person's symptoms and suggests a suitable disease accordingly.It will help in assisting healthcare practitioners by reducing the pressure on overcrowded clinics.To showcase the accuracy of the proposed model, it has been implemented on a heart disease dataset to predict the occurrence of heart disease in the next 10 years. The implementation willillustrate the effectiveness of the proposed model which can help in the development of an intelligent healthcare system and reduce the cost of treatment. This system will improve the efficiency of the healthcare department and reduce the time wasted in finding out which diseasethe patient is suffering from.

Keywords: Disease prediction, Logistic regression, Naïve Bayes, Random forest classifier, KNN, Stochastic Gradient classifier

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 OVERALL DESCRIPTION

Numerous patients go untreated or are not treated precisely by the specialists, legitimate treatment is important for the prosperity of the individuals. Consequently, foreseeing a sickness utilizing the patient's manifestations has become a significant undertaking nowadays. There is an absence of specialists in India, there is 1 specialist for each 10,198 people in India (WHO suggests the proportion of 1:100). To address this intense deficiency of specialists there should be a foreseeing framework for anticipating the overall sicknesses which would help in legitimate use of the assets.

The initial phase in treating a patient is the right location of the wellbeing of a person by utilizing the given side effects. The expectation of the illness has become an indispensable errand of late anyway the right forecast of sicknesses has gotten excessively extreme for a specialist. The framework proposed in this paper is intended to build up a sickness forecast framework by using ML. The order inside the forecast framework is finished with the assistance of the logistic regression algorithm. This may encourage right expectation of health and furthermore encourage in the right treatment of illness.

The fundamental spotlight is on to utilize ML in medical services to enhance understanding consideration for better outcomes. ML has made simpler to recognize various sicknesses and finding accurately. Prescient investigation with the assistance of effective numerous ML calculations assists with foreseeing the infection all the more accurately and help treat patients. The medical care industry creates a lot of medical services information day by day that can be utilized to separate data for anticipating sickness that can happen to a patient in future while utilizing the therapy history and wellbeing information. This shrouded data in the medical care information will be later utilized for full of feeling dynamic for patient's wellbeing. Likewise, this zone need improvement by utilizing the educational information in medical care.

Information volume is a huge test in any industry however especially in medical care where information will in general sit inert in information bases oversaw by dated EHR(Electronic Health Record) frameworks. Numerous organizations assemble their business on getting enormous volumes of information from these frameworks to make them accessible and significant as they power prescient investigation, choice help, imaging, activity streamlining, and different applications. Different associations utilize protection claims information that have as of late become accessible through state governments. In any case, rules for obtaining entrance for business objects are incipient using complex cycles, so fruitful applications have been rare.

ML in medical services has as of late stood out as truly newsworthy. Google has built up an ML calculation to help distinguish malignant tumors on mammograms. Stanford is utilizing a profound learning calculation to distinguish skin malignant growth. A new JAMA article announced the consequences of a profound ML calculation that had the option to analyze diabetic retinopathy in retinal pictures. Obviously ML places another bolt in the bunch of clinical dynamic.

All things considered, ML fits a few cycles in a way that is better than others. Calculations can furnish prompt advantage to disciplines with measures that are reproducible or normalized. Likewise, those with enormous picture datasets, for example, radiology, cardiology, and pathology, are solid competitors. ML can be prepared to see pictures, recognize anomalies, and highlight territories that need consideration, accordingly improving the precision of every one of these cycles. Long haul,

Machine Learning will profit the family expert or internist at the bedside. ML can offer a target assessment to improve effectiveness, unwavering quality, and precision.

Human insight can scarcely be contrasted with some other marvel. ML knowledge in medical

services has a great deal of potential outcomes to improve the savvy choices made by people. The particular advantages of including ML into medication incorporate precise information can educate experts about normal examples, ML can proceed just as a human does and invalidates pressure and fatigue factors, informational collections can prepare ML calculations and models to address key medication creation issues which would help in relieving more individuals under lower cost and saving the customized approach.

Computer based intelligence utilizes modern calculations to remove, learn, anticipate, and predict from gigantic measures of clinical information while additionally offering proficient help and help. Concerning sicknesses, disease, cardiovascular, and sensory system issues are the most much of the time explored including ML instruments. Self-prepared frameworks can follow administered and unaided getting the hang of, encouraging early discovery and analysis significantly. To perform well, self-prepared frameworks ought to cooperate continually with the clinical examinations information, so clearly human movement is interconnected with machine learning.

Medical care industry has become enormous business. The medical services industry delivers a lot of medical care information every day that can be utilized to extricate data for foreseeing illness that can happen to a patient in future while utilizing the therapy history and wellbeing information. This shrouded data in the medical care information will be later utilized for full of feeling dynamic for patient's wellbeing. Additionally, this zone need improvement by utilizing the educational information in medical care. Significant test is the means by which to remove the data from these information on the grounds that the sum is huge so some information mining and machine learning strategies can be utilized. Likewise, the normal result and extent of this venture is that if sickness can be anticipated than early therapy can be given to the patients which can decrease the danger of life and save life of patients and cost to get therapy of illnesses can be diminished up somewhat by early acknowledgment. The fast selection of electronic wellbeing records has made an abundance of new information about patients, which is a goldmine for improving the comprehension of human wellbeing.

At the point when machine learning is utilized in guide to enhance dealing with patients, high outcomes are accomplished. It has made it simpler to spot different sorts of infections and perform analyze precisely. Performing prescient examination with the help of numerous productive ML calculations may encourage foreseeing any infection with extraordinary precision and help to treat patients. The gigantic measure of clinical data containing treatment history and wellbeing information are regularly used to extricate information for anticipating infections that may happen to a patient inside what's to come. The concealed information inside the clinical data are frequently later utilized for a viable dynamic cycle for the patient's wellbeing.

One of the chief crucial utilizations of ML is inside the field of medical care. The medical care offices must be constrained to be progressed with the goal that better choices for quiet therapy are regularly made. When ML is utilized in medical care, it causes people to handle immense and complex illness datasets to examine them into accommodating clinical bits of knowledge. At that point this will be additionally utilized by clinical specialists to give exact treatment to patients. Thus, ML, when upheld in medical care will bring about high patient fulfillment. In this paper, the calculated relapse calculation will be utilized to foresee sicknesses utilizing the patient's therapy history and wellbeing information.

Logistic Regression algorithm

The logistic regression is likewise alluded to as the sigmoid function that helps simple portrayal of charts. It also gives high precision. In this calculation, the information ought to be first imported and afterward it ought to be prepared. It is a kind of regression examination calculation, which is utilized for expectation of the result of a categorical dependent variable from a bunch of independent or predictor variables.

The vital portrayal in logistic regression are the coefficients, much the same as linear regression. The coefficients in logistic regression are assessed utilizing a cycle called maximum-likelihood estimation. In logistical regression, the variable amount is typically binary. It is chiefly utilized for prediction and furthermore ascertaining the probability. Logistic regression is utilized in light of the fact that it is an effective regression predictive analysis algorithm that doesn't dispose off any information and uses all information productively.

Naïve Bayes

Naive Bayes algorithm based on Bayes' theorem with the assumption of independence between every pair of features. Naive Bayes classifiers work well in many real-world situations such as document classification and spam filtering.

Decision Tree

Given a data of attributes together with its classes, a decision tree produces a sequence of rules that can be used to classify the data.

Stochastic gradient descent

Stochastic gradient descent is a simple and very efficient approach to fit linear models. It is particularly useful when the number of samples is very large. It supports different loss functions and penalties for classification.

K-Nearest Neighbours

Neighbours based classification is a type of lazy learning as it does not attempt to construct a general internal model, but simply stores instances of the training data. Classification is computed from a simple majority vote of the k nearest neighbours of each point.

Random Forest

Random forest classifier is a meta-estimator that fits a number of decision trees on various sub-samples of datasets and uses average to improve the predictive accuracy of the model and controls over-fitting. The sub-sample size is always the same as the original input sample size but the samples are drawn with replacement.

Support Vector Machine

Support vector machine is a representation of the training data as points in space separated into categories by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

1.2 LITERATURE SURVEY

Sana Bharti[12] and Monika Gandhi[14] worked on deep learning and data mining to predict the disease. Their approach was to check the medical history of the patient and observe the new symptoms in the patient and depending on both of them predict the possibility of the patient to develope heart disease. Using this model it was believed that the accuracy to predict the heart attack would increase. Model proposed by Sarath Babu[15] uses deep learning algorithms but even after using these models the accuracy was still very less.

A system proposed by Chen et al.[16] for health monitoring using smart clothing. He was able to achieve cost minimization for heterogeneous systems. The disease history of patient is recorded which helps to give solutions which minimize the cost of medical treatment. But the drawback of the system is that it could predict the disease but not the sub-type of disease.

The system proposed by M.Akhil and Ms. Chandra[17] uses ANN(Artificial Neural Network) to classify heart disease. Their model has an accuracy level of 92.8%.

Model proposed by [18]A. Sheik Abdullah uses Data mining to detect Coronary Heart Disease(CHD). Using feature selection the number of attributes were reduced. This model has an accuracy of 60.74%.

According to model proposed in[24], Support Vector Machine(SVM) is used to practically implement the theories which were theoretically laid.

Paper[25] aims at finding out the death rate by using ML tool and implementing it on the data set.

Paper by B.Dhomse and M.Mahale[26] proposed a system that uses lesser number of characteristics to predict coronary illness by using advanced machine learning calculations.

[2] Akash C. Jamgade and Prof. S. D. Zade.One such application of machine learning algorithms is in the area of healthcare. Machine learning in healthcare aids the humans to process big and complex medical datasets and then examine them into clinical intuitions. This can farther be used by physicians in giving medical care.

[3] Vinitha S, Sweetlin S, Vinusha H and Sajini S,In the proposed structure, it provides machine learning algorithms for effectual prediction of various disease occurrences in disease-frequent societies. It experiment the altered estimate models over real-life hospital data collected.

[4] Paper By; S. Patel and H. Patel, in this article many types of Data Mining techniques such as classification, clustering, association and also highlights related work to examine and foresee human disease.

[6] Paper By;Min Chen, Yixue Hao, Kai Hwang, Lu Wang and Lin Wang, they organize machine learning algorithms for efficacious divination of chronic disease outbreak in disease-frequent communities.

[7] Paper By:Tarigoppula V.S Sriram, M. Venkateswara Rao, G V Satya Narayana, DSVGK Kaladhar and T Pandu Ranga Vital, Diagnosis of the Parkinson disease with the help of machine learning approach provides better understanding from PD dataset in the present decennium.

[11]M.A. Jabbar, Hidden Naïve Bayes is a data mining model that lessens the traditional Naïve Bayes conditional independence supposition. This model gives that the Hidden Naïve Bayes (HNB) can be applied to heart disease classification (prediction).

[19] Reddy Prasad, Pidaparthi Anjali, S. Adil, N. Deepa, In this paper the logistic regression algorithms is used and the health care data which gives the details of the patients whatever they are having heart diseases or not in accordance to the information in the database record. Also they will try to use this data a model which predicts the patient if they are having heart disease or not.

Paper By:[20]Nikhil Gawande, In this ECG(electrocardiography) signals are used. ECG is likely to the show condition of the patient and for the diagnosis and treatment of all the types of cardiac diseases.

1.3 MOTIVATIONS AND SCOPE

At present to stay sound, normal body finding is important. Today, there are different sources accessible as individual forecast or proposal framework yet the need of great importance is to have a coordinated model involving both. Likewise, it would be more proper and helpful if individuals could get fundamental determination online 24x7 instead of visiting medical clinics and centers often. Subsequently, lessening cost and saving time. On the off chance that specific abnormalities found in the analysis, at that point suggestion of close by subject matter expert and medical clinics as per client's inclination would encourage in fast and suitable therapy. Medical care being an area advancing persistently and producing an enormous measure of information builds up a need to utilize the information for valuable information which pulls in huge associations to put vigorously in this field.

Here the extent of the venture is that coordination of clinical choice help with PC based patient records could diminish clinical blunders, upgrade quiet security, decline undesirable practice variety, and improve understanding result. The application grants client to share their heart associated issues. It at that point measures client explicit subtleties to determine for differed ailment that may be identified with it. Here we will in general utilize some astute information mining methods to figure the preeminent right disease that may be identified with patient's subtleties. In view of result, framework consequently shows the outcome explicit specialists for greater treatment. The framework grants client to see specialist's subtleties and can likewise be utilized if there should be an occurrence of crisis.

2. PURPOSE

Illness forecast utilizing understanding treatment history and wellbeing information by applying data mining and ML procedures is progressing battle for as long as many years. Numerous works have been applied data mining strategies to obsessive information or clinical profiles for expectation of explicit illnesses. These methodologies attempted to anticipate the reoccurrence of illness. Additionally, a few methodologies attempt to do expectation on control and movement of sickness. The new achievement of profound learning in unique territories of ML has driven a move towards ML models that can learn rich, various leveled portrayals of crude information with minimal pre-preparing and produce more precise outcomes. Quantities of papers have been distributed on a few information digging procedures for analysis of coronary illness, for example, Decision Tree, Naive Bayes, neural organization, part thickness, consequently characterized gatherings, sacking calculation and backing vector machine indicating various degrees of exactness in infections forecast.

3. EXISTING SYSTEM

In the current framework, useful utilization of different gathered information is tedious, machine can anticipate illnesses yet can't foresee the sub sorts of the infections brought about by event of one sickness. It neglects to foresee all potential states of the individuals. Existing framework handles just organized information. A machine can distinguish an illness yet can't expect the sub sorts of the infections and sicknesses brought about by the presence of one bug. The expectations of illnesses have been vague and inconclusive. For event, if a gathering of individuals are predicted with Diabetes, without a doubt some of them may have complex danger for Heart infections because of the reality of Diabetes. Determination of the condition exclusively relies on the specialist's instinct and patient's records. Discovery is absurd at a previous stage that may later possibly hurt the patient.

4. PROPOSED SYSTEM

The proposed framework has been created to order individuals, who are blasted by illness and solid individuals. The presentation of the prescient model with chosen highlights is tried to anticipate the probabilities of experiencing coronary illness. Highlight determination calculation was utilized to choose significant highlights, and on these chose highlights, the exhibition of the classifiers was tried. The Framingham heart condition dataset is taken from Kaggle and has been utilized in our examination. The mainstream ML classifier logistic regression is utilized inside the framework. The model's approval and execution assessment measurements are processed. It is adaptable and can be generally utilized for different sicknesses with high paces of accomplishment. The strategy of the proposed framework is organized into five phases which include:

(1) pre-processing of a dataset, (2) feature selection, (3) cross-validation method, (4) machine learning classifiers, and (5) classifiers' performance evaluation methods.

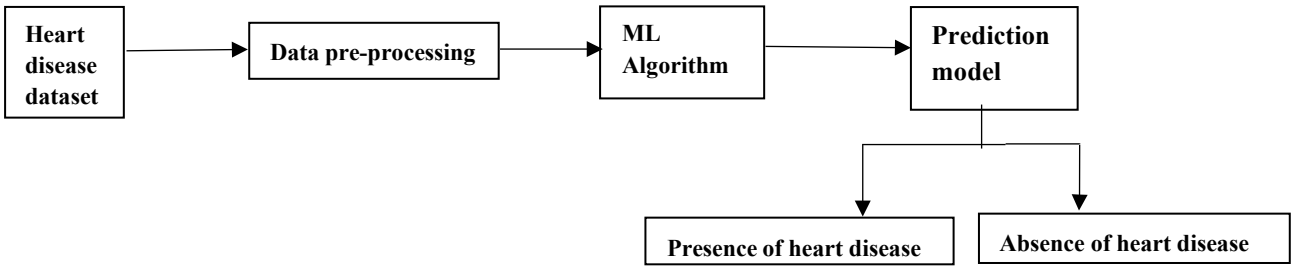

Figure 1: Disease predicting model framework for predicting heart disease

5. IMPLEMENTATION

The proposed framework is actualized on a coronary illness dataset, which is accessible on the Kaggle site and it contains a cardiovascular report on inhabitants of the town of Framingham, Massachusetts. The grouping objective is to anticipate whether the patient has a danger of future heart disease(HD) in 10 years or not. The dataset gives the patients' data like segment, social and clinical danger factors. The dataset contains more than 4,000 records and around 15 credits.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	male	age	education	CurrentSm	CigsPerDa	BPMeds	PrevalentS	PrevalentH	Diabetes	TotChol	SysBP	DiaBP	BMI	HeartRate	Glucose	TenYearHD	
2	1	39	4	0	0	0	0	0	0	195	106	70	26.97	80	77	0	
3	0	46	2	0	0	0	0	0	0	250	121	81	28.73	95	76	0	
4	1	48	1	1	20	0	0	0	0	245	127.5	80	25.34	75	70	0	
5	0	61	3	1	30	0	0	1	0	225	150	95	28.58	65	103	1	
6	0	46	3	1	23	0	0	0	0	285	130	84	23.1	85	85	0	
7	0	43	2	0	0	0	0	1	0	228	180	110	30.3	77	99	0	
8	0	63	1	0	0	0	0	0	0	205	138	71	33.11	60	85	1	
9	0	45	2	1	20	0	0	0	0	313	100	71	21.68	79	78	0	
10	1	52	1	0	0	0	0	1	0	260	141.5	89	26.36	76	79	0	
11	1	43	1	1	30	0	0	1	0	225	162	107	23.61	93	88	0	
12	0	50	1	0	0	0	0	0	0	254	133	76	22.91	75	76	0	
13	0	43	2	0	0	0	0	0	0	247	131	88	27.64	72	61	0	
14	1	46	1	1	15	0	0	1	0	294	142	94	26.31	98	64	0	
15	0	41	3	0	0	1	0	1	0	332	124	88	31.31	65	84	0	
16	0	39	2	1	9	0	0	0	0	226	114	64	22.35	85	NA	0	
17	0	38	2	1	20	0	0	1	0	221	140	90	21.35	95	70	1	
18	1	48	3	1	10	0	0	1	0	232	138	90	22.37	64	72	0	
19	0	46	2	1	20	0	0	0	0	291	112	78	23.38	80	89	1	
20	0	38	2	1	5	0	0	0	0	195	122	84.5	23.24	75	78	0	
21	1	41	2	0	0	0	0	0	0	195	139	88	26.88	85	65	0	
22	0	42	2	1	30	0	0	0	0	190	108	70.5	21.59	72	85	0	
23	0	43	1	0	0	0	0	0	0	185	123.5	77.5	29.89	70	NA	0	
24	0	52	1	0	0	0	0	0	0	234	148	78	34.17	70	113	0	
25	0	52	3	1	20	0	0	0	0	215	132	82	25.11	71	75	0	
26	1	44	2	1	30	0	0	1	0	270	137.5	90	21.96	75	83	0	
27	1	47	4	1	20	0	0	0	0	294	102	68	24.18	62	66	1	
28	0	60	1	0	0	0	0	0	0	260	110	72.5	26.59	65	NA	0	
29	1	35	2	1	20	0	0	1	0	225	132	91	26.09	73	83	0	
30	0	61	3	0	0	0	0	1	0	272	182	121	32.8	85	65	1	
31	0	60	1	0	0	0	0	0	0	247	130	88	30.36	72	74	0	
32	1	36	4	1	35	0	0	0	0	295	102	68	28.15	60	63	0	
33	1	43	4	1	43	0	0	0	0	226	115	85.5	27.57	75	75	0	
34	0	59	1	0	0	0	0	1	0	209	150	85	20.77	90	88	1	

Figure 2: an imported dataset from Kaggle

At that point we settle on a decision for picking the variable amount and variable amount, where each quality is considered as a potential danger factor. There are a few segment, social and clinical danger factors included.

Demographic:

sex: male or female(Nominal)

age: age of the patient(Continuous)

Behavioral:

CurrentSmoker- whether or not the patient may be a current smoker (Nominal)

CigsPerDay: the number of cigarettes that the person smoked on the average in one day(can be considered continuous as one can have any number of cigarettes, even half a cigarette.)

Medical(history):

BPMeds: whether or not the patient was on vital sign medication (Nominal)

PrevalentStroke: whether or not the patient had a stroke before (Nominal)

PrevalentHyp: whether the patient was hypertensive or not (Nominal)

Diabetes: whether the patient had diabetes or not (Nominal)

Medical(current):

TotChol: Total Cholesterol level (Continuous)

SysBP: Systolic Blood Pressure (Continuous)

DiaBP: Diastolic Blood Pressure (Continuous)

BMI: Body Mass Index (Continuous)

HeartRate: pulse rate (Continuous - In medical research, variables like pulse rate though after all discrete, yet are considered continuous thanks to an outsized number of possible values.)

Glucose: Glucose level (Continuous)

Predict variable (desired target): risk of future heart disease(HD) in 10 years (binary: "1" means "Yes" and "0" means "No")

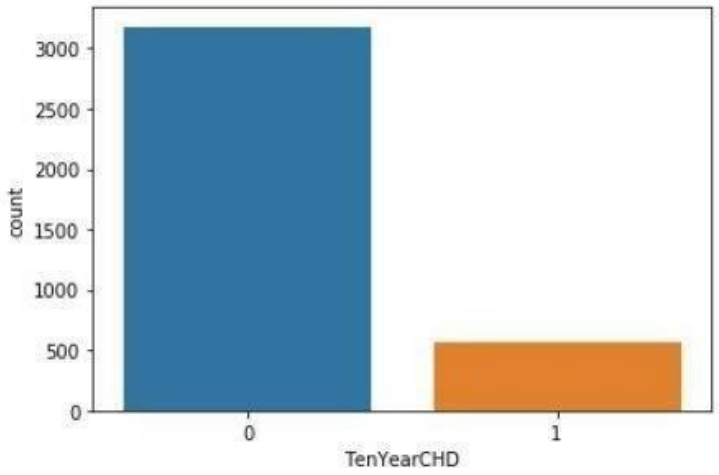

Figure 3. Number of people with history of heart disease

The above figure shows the clinical record of 4,000 individuals out of which around 3,500 individuals have not experienced cardiovascular sickness before though 5,000 individuals have experienced cardiovascular infection.

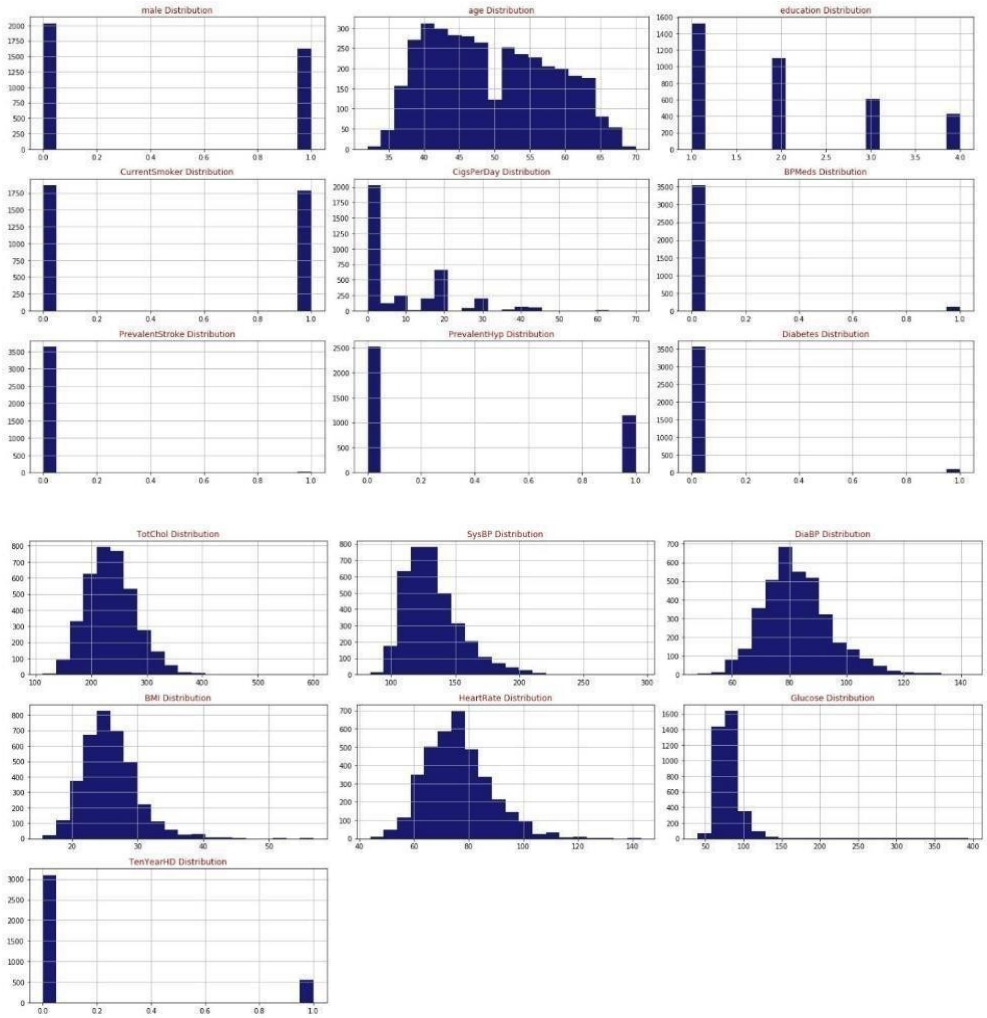

Figure 4: independent and dependent variables

The coronary illness dataset is then part into two subsets for example preparing information and testing information and that we fit our model on train information to frame expectations on the test information After that two things can wind up occurring, we may overfit our model or we may underfit our model. Any of those things happening would influence the consistency of our model, so we may end up utilizing a model with lower exactness. For the coronary illness dataset, the training set is taken as 80% of the genuine information and test set as 20% of the information.

Logit Regression Results

Dep. Variable:	TenYearHD	**No. Observations:**	3751
Model:	Logit	**Df Residuals:**	3744
Method:	MLE	**Df Model:**	6
Date:	Sat, 30 May 2020	**Pseudo R-squ.:**	0.1149
Time:	23:18:09	**Log-Likelihood:**	-1417.7
converged:	True	**LL-Null:**	-1601.7
Covariance Type:	nonrobust	**LLR p-value:**	2.127e-76

	coef	std err	z	P>|z|	[0.025	0.975]
const	-9.1264	0.468	-19.504	0.000	-10.043	-8.209
sex_male	0.5815	0.105	5.524	0.000	0.375	0.788
age	0.0655	0.006	10.343	0.000	0.053	0.078
CigsPerDay	0.0197	0.004	4.805	0.000	0.012	0.028
TotChol	0.0023	0.001	2.106	0.035	0.000	0.004
SysBP	0.0174	0.002	8.162	0.000	0.013	0.022
Glucose	0.0076	0.002	4.574	0.000	0.004	0.011

Figure 5: Logistic regression results using backward elimination (P value approach)

The results show that there are some attributes with P value higher than the favored alpha (5%) and subsequently indicating low genuinely critical relationship with the likelihood of coronary illness. We use backward elimination way to deal with and eliminate those ascribes with highest P value each in turn, at that point the regression is run over and again until all attributes have P Values under 0.05. A property having P value under 0.05 shows that the adjustment in its value will cause change in the chances of having a coronary illness.

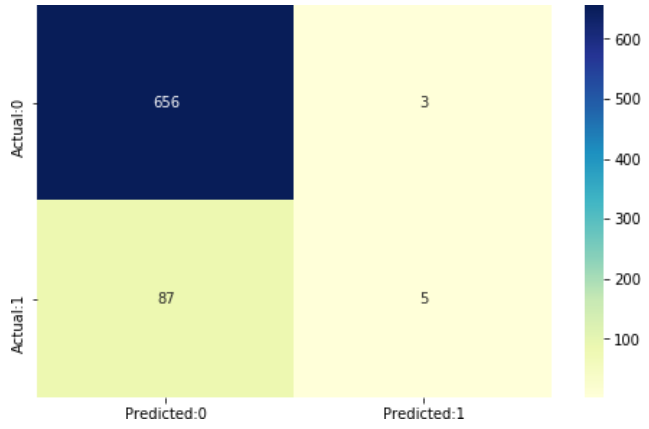

Figure 5: Confusion matrix for model evaluation using Logistic Regression

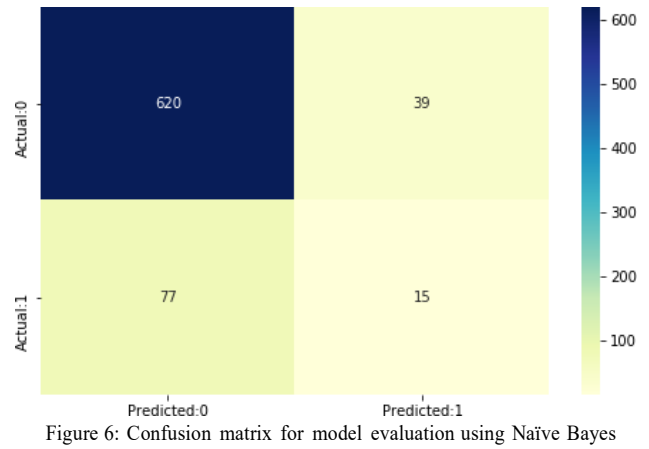

Figure 6: Confusion matrix for model evaluation using Naïve Bayes

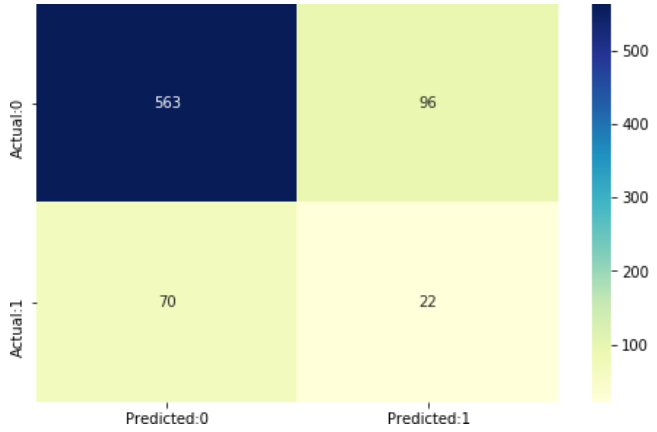

Figure 7: Confusion matrix for model evaluation using Decision Tree Classifier

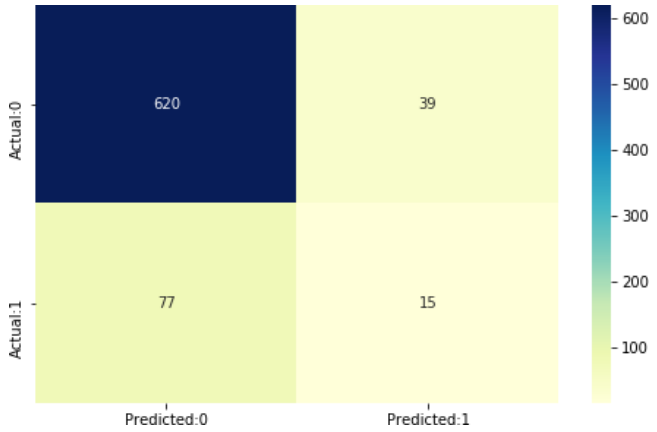

Figure 8: Confusion matrix for model evaluation using Stochastic Gradient Descent Classifier

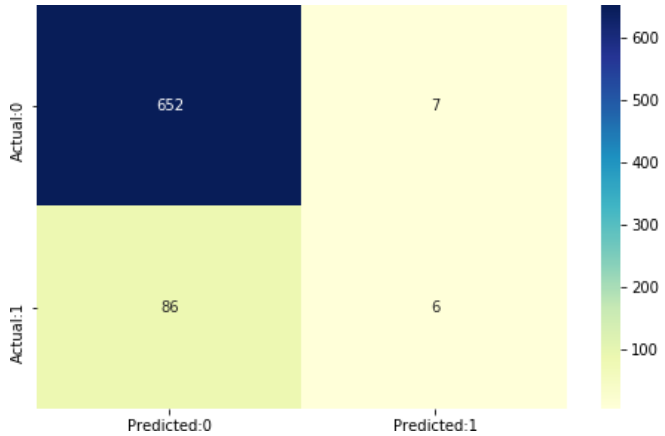

Figure 9: Confusion matrix for model evaluation using KNN

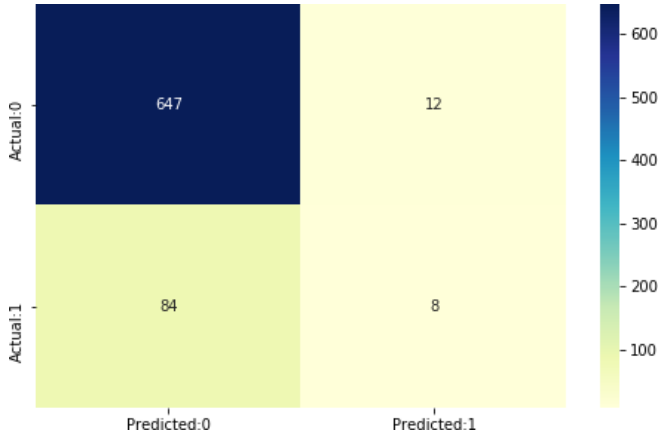

Figure 10: Confusion matrix for model evaluation using Random Forest Classifier

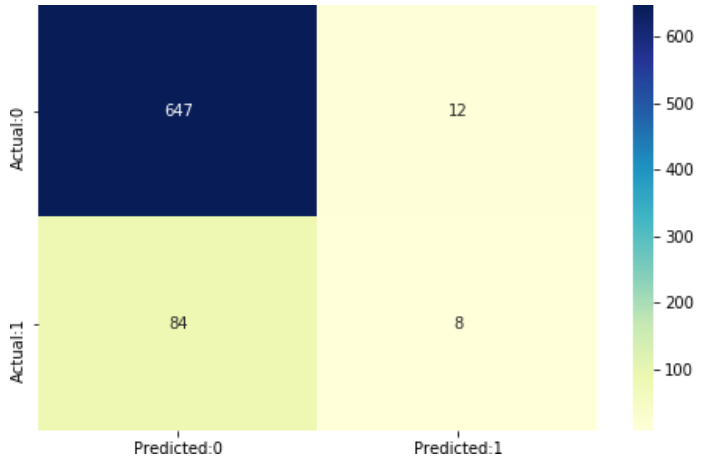

Figure 11: Confusion matrix for model evaluation using Support Vector Machine

Algorithm:

Step1: Import all the important and relevant libraries

Step2: Import the dataset

Step3: Remove the rows and columns that are irrelevant

Step4: Remove all the null values from the dataset

Step5: Perform data exploratory analysis of clean data

Step6: Make TenYear CHD dependent variable and others independent variable

Step7: Eliminate the features that are irrelevant using backward elimination (P value) approach
Step8: Split data into 80:20 and train the model using

8.1: Logistic Regression

8.2: Naïve Bayes

8.3: Decision Tree Classifier

8.4: Stochastic Gradient Descent

8.5: K-nearest neighbor

8.6: Random Forest classifier

8.7: Support vector machine

Step9: Evaluate the correctness of the model and using confusion matrix calculate the amount of correct and incorrect prediction Step10: Stop

The confusion matrix shows that the model evaluation using Logistic Regression made 661 right expectations which are the highest of all the algorithms tested and 90 inaccurate ones. From the above insights it very well may be clarified that the model is profoundly specific than sensitive. The negative values are anticipated more precisely than the positives. Likewise, the model accomplishes a precision pace of 88%.

Algorithms	Accuracy
Logistic Regression	0.8801597869507324
Naive Bayes	0.8455392809587217
Decision Tree Classifier	0.7789613848202397
Stochastic Gradient Descent Classifier	0.8455392809587217
KNN	0.8761651131824234
Random Forest Classifier	0.8721704394141145

Support Vector Machine	0.8721704394141145

6. OUTPUT

	CI 95%(2.5%)	CI 95%(97.5%)	Odds Ratio	pvalue
const	0.000043	0.000272	0.000109	0.000
sex_male	1.455242	2.198536	1.788687	0.000
age	1.054483	1.080969	1.067644	0.000
CigsPerDay	1.011733	1.028128	1.019897	0.000
TotChol	1.000158	1.004394	1.002273	0.035
SysBP	1.013292	1.021784	1.017529	0.000
Glucose	1.004346	1.010898	1.007617	0.000

Figure 12: Effect in odds of heart disease

The fitted model shows that, holding any remaining highlights steady, the chances of experiencing cardiovascular illness for guys are 78.8% higher than the chances for females. The coefficient for age says that, holding all others consistent, we will see 6.76% expansion in the chances of experiencing cardiovascular sickness with one year increment in age. Additionally, with each additional cigarette one smokes there is a 2% expansion in the chances of getting cardiovascular illness. For Total cholesterol level and glucose level there is no critical change. Likewise, there is a 1.7% expansion in chances for each unit increment in systolic Blood Pressure.

7. CONCLUSION

It has been presumed that men are more inclined to coronary illness than ladies. An expansion in age, alongside the quantity of cigarettes smoked every day and systolic pulse likewise show expanded chances of getting a cardiovascular infection. Absolute cholesterol shows no huge change in the chances of Heart Disease (HD), this could be because of the presence of good cholesterol in the cholesterol perusing. Essentially, glucose also causes a truly immaterial change in chances (0.2%).

The future improvement of the proposed framework will bring about forecast illnesses by utilizing progressed procedures and calculations in less time unpredictability. An insightful framework might be created utilizing the proposed model that can prompt the choice of legitimate treatment techniques. Information investigation and Machine learning can be of awesome assistance in choosing the line of treatment to be trailed by removing information from such appropriate information bases.

8. REFERENCES

[1] Akash C. Jamgade and Prof. S. D. Zade, "Disease Prediction Using Machine Learning", International Research Journal of Engineering and Technology, Vol. 5, Issue 6, May2019.

[2] Vinitha S, Sweetlin S, Vinusha H and Sajini S, "Disease prediction using machine learning over Big Data", Computer Science & Engineering: An International Journal, Vol. 8, No. 1, February 2018.

[3] S. Patel and H. Patel, "Survey of data mining techniques used in healthcare domain", Int. J. of Inform. Sci. and Tech., Vol. 6, pp. 53-60, March 2016.

[4] M. Abinaya, M. Marimuthu, K.S. Hariesh, K. Madhankumar and V. Pavithra, "A Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach", International Journal of Computer Applications, Vol. 181, No. 18, September 2018.

[5] Min Chen, Yixue Hao, Kai Hwang, Lu Wang and Lin Wang, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities", IEEE, Vol. 5, April 2017.

[6] Tarigoppula V.S Sriram, M. Venkateswara Rao, G V Satya Narayana, DSVGK Kaladhar and T Pandu Ranga Vital, "Intelligent Parkinson Disease Prediction Using Machine Learning Algorithms", International Journal of Engineering and Innovative Technology, Vol. 3, Issue 3, September 2013.

[7] B. Chen, M. Li, J. Wang and F. Wu, "A logistic regression based algorithm for identifying human disease genes," 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Belfast, 2014, pp. 197-200, doi: 10.1109/BIBM.2014.6999153.

[8] Florian Privé, Hugues Aschard, Michael G. B. Blum, Efficient Implementation of Penalized Regression for Genetic Risk Prediction, Genetics, 10.1534/genetics.119.302019, 212, 1, (65-74), (2019).

[9] A.A. Alizadeh, M.B. Eisen, R.E. Davis, C. Ma, I.S. Lossos, A. Rosenwald, *et al.***Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling** Nature, 403 (2000), pp. 503-511

[10] M.A. Jabbar ,2016;Heart disease prediction system based on hidden naviebayes classifier;2018 International conference on circuits , controls, communications and computing(14C)

[11] Sana Bharti,2015 .Analytical study of heart disease prediction comparing with different algorithms; International conference on computing, communication andautomation(ICCA2015).

[12] Nimai Chand Das Adhikari, Arpana Alka, and rajat Garg, "HPPS: Heart Problem Prediction System using Machine Learning"

[13] Monika Gandhi,2015. Prediction in heart disease using techniques of data mining, International conference on futuristic trend in computational analysis and knowledge management(ABLAZE- 2015)

[14] Sarath Babu,2017.Heart disease diagnosis using data mining technique,international conference on electronics,communication and aerospace technology ICECA2017

[15] Chen, M., Ma, Y., Song, J. et al. Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring. Mobile Netw Appl 21, 825–845 (2016).

[16] M. Akhil Jabbar, B.L Deekshatulu & Priti Chandra, "Classification of Heart Disease using Artificial Neural Network and Feature Subset Selection", Global Journal of Computer Science and Technology Neural & Artificial Intelligence, Volume 13 Issue 3 Version 1.0 Year 2013

[17] Sheik Abdullah, " A Data Mining Model to Predict and Analyze the Events Related to Coronary Heart Disease using Decision Trees with Particle Swarm Optimization for Feature Selection", International Journal of Computer Applications Volume 55– No.8, October 2012.

[18] Reddy Prasad, Pidaparthi Anjali, S. Adil, N. Deepa, "Heart Disease Prediction using Logistic Regression Algorithm using Machine Learning", International Journal of Engineering and Advanced Technology, Vol. 8, Issue 3S, February 2019.

[19] Nikhil Gawande ,2017. Heart diseases classification using convolutional neural network; 2012 2ndInternational conference on communication and electronics systems(ICCES)

[20] P. de Chazal and R.B.Reilly, "A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features," IEEE transaction biomedical engineering, vol. 53, no. 12, pp. 2535–2543, Dec. 2006.

[21] AditiGavhane , 2018 . Prediction of heart disease using machine learning, ISBN:978-1-5386-0965-1.

[22] Dinesh Kumar G, 2018 . Prediction of cardiovascular disease using machine learning algorithms, procedingof 2018 IEEE International Conference on Current Trends toward Converging Technologies, Coimbatore, India.

[23] Geert Meyfroidt, Fabian Guiza, Jan Ramon, Maurice Brynooghe" Machine learning techniques to examine large patient databases"-Best practice &ReasearchClinicalAnaesthesiology, ElsevierVolume 23 (1) – Mar 1, 2009.

[24] Matjaz Kuka, Igor Kononenko, Cyril Groselj, Katrina Kalif, JureFettich" Analysing and improving the diagnosis of ischaemic heart disease with machine learning" Elsevier -Artificial intelligence in Medicine, Volume23, May 1999.

[25] B.Dhomse Kanchan, M.Mahale Kishore "Study of Machine learning algorithms for special disease prediction using principal of component analysis" Global Trends in Signal Processing, Information Computing and Communication(ICGTSPICC),2016InternationalConference