



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 Issue: VI Month of publication: June 2019

DOI: <http://doi.org/10.22214/ijraset.2019.6295>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Prediction Analysis of Cardiac Disease using Classification

Darshani Mohite¹, Suhasini Vijaykumar²

¹MCA Student, ²Professor, BVIMIT (University Of Mumbai)

Abstract: Data mining is process of extraction of useful information from dataset.it explore the hidden patterns and recognizes the future trends.it used to finds anomalies and discovers the patterns from large datasets. In This paper the technique used for data analysis is Classification algorithm that is Naïve Bayes algorithm.The prediction analysis of heart disease system is developed for predicting the presence or absence of cardiac disease when medical details of patient is given like age, gender, chest pain, resting blood pressure, cholesterol, fasting blood sugar, thalach etc. that are used to predict the likelihood of having the heart disease present or absent in patient. As the cardiac disease is one of the major cause of death throughout the world, this paper suggest some solution for identifying heart disease using data mining technique and statistics.

Keywords: data mining, naïve bayes, analysis, dataset, Heart Disease, Statistics.

I. INTRODUCTION

cardiac disease is caused due to narrowing or blockage of coronary arteries. This is caused due to insufficient flow of oxygen towards the blood vessels. There are some of major heart disease factors which include Diabetes, high blood pressure, obesity, smoking, eating habits, alcohol that affects our whole body. As shown in the figure.1 the blood gets clot due to layers of cholesterol within the arteries that prevents flow of oxygen and get cardiac arrest.

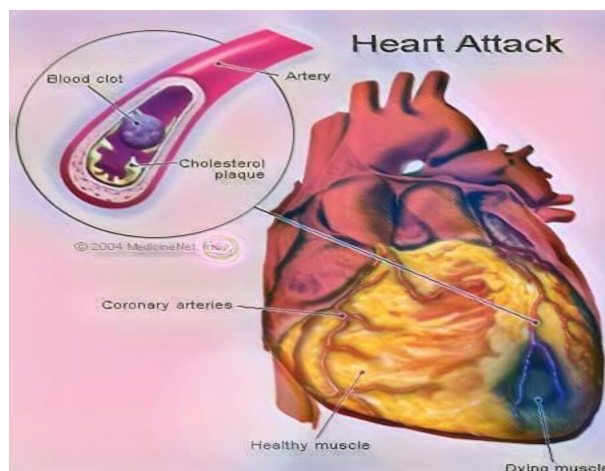


Fig. 1 heart attack

In the current scenario there is money and time lost as the number of test conducted by hospital. sometimes result get inappropriate.so this paper focus on a system that generate particular volume of a data on a daily basis. the system collect data about the person including attributes such as person's age, sex, blood pressure, chest pain type, fasting blood sugar, restecg, serum-cholesterol, exang (exercise induced angina) etc. For diagnosis of Heart disease.

The number of tools are used for data analysis, so out of which **Anaconda** is a freemium open source distribution of python language for predictive analysis and also scientific computing.the system produce accurate result according to values which are given for the attribute.the result will show that particular person having heart disease or not in term of result like absent or present. the result shall produce on the basis of range of value attribute. this paper focuses on the system that compares accurate prediction when the number of medical attributes used for prediction is decreased. here we use Naïve Bayes algorithm of classification which used to predict if person have affected with cardio vascular disease and any heart related disease.the paper will discuss algorithm and techniques that is used for prediction.

II. LITERATURE REVIEW

Cardiac disease is a term that assigns to a large number of medical conditions related to heart. These medical conditions describe the abnormal health conditions that directly influence the heart and all its parts. cardiac arrest/heart attack is the major cause of death in today's time. According to World Health Organization it has estimated that around **12 million** of death occur worldwide, every year because of the cardiac disease.

Although, many computational tools have been designed and developed for to improve and enhance the ability of physicians for making decision for patient about the condition of a disease, low performance by current detection model is remain matter of concern till time. Artificial Intelligence is widely spread branch of computer science its uses is also hugely increasing among field such as medical, data analysis etc. firstly Artificial Intelligence enable machine or device act like human that it include knowledge, ability to learn new things, perception, problem solving skills etc. so it perform all the work as like the human. it is emphasizes on developing the intelligent machine. there are various Algorithm created in Artificial Intelligence which have best potential for identifying hidden patterns in the large datasets. so such patterns are utilize for clinical diagnosis and medical data analysis. this is use to create Expert System.

Over the years, the research on predicting ,heart disease has been tremendously increased. Many researcher uses the data mining technique including data sets and algorithm to achieve the accurate and efficient result. the aim of this paper is to analyze data from different dataset and predict the accurate outcome. various technique has been used for data analysis and here we are using naïve bayes algorithm.

Naive Bayes Algorithm: this algorithm has been studied extensively form 1960 and it is introduced in text retrieval community and it has been remain popular method for text categorization. judging the particular text, document or result belonging to specific category with word frequencies. it is probabilistic study of particular data where it recognize one belongs to which class. naive bayes classifier is highly scalable, reliable and are able to find solution for particular problem. naive bayes is a technique for constructing the models, classifier and assign class labels to the problem instances.

Probabilistic Model: naive Bayes is a conditional probability model.

$$P(C_k | x) = p(C_k)p(x | C_k)/p(x).$$

Where it can be also writes as posterior=prior*likelihood/evidence.

A. Functionality

Bayesian Classifier can discover the knowledge from hidden patterns that is associated with disease from records of the patients having the cardiac disease. posterior probability include hypothesis of whether the patient having heart disease or not when certain amount of attributes are given. conditional probability depends upon the condition that is specified. classifier predict the probability and gives the result whether given record belong to target class or not. Below Figure 1 shows the implementation of Naïve Bayes Classification Algorithm on patient data.

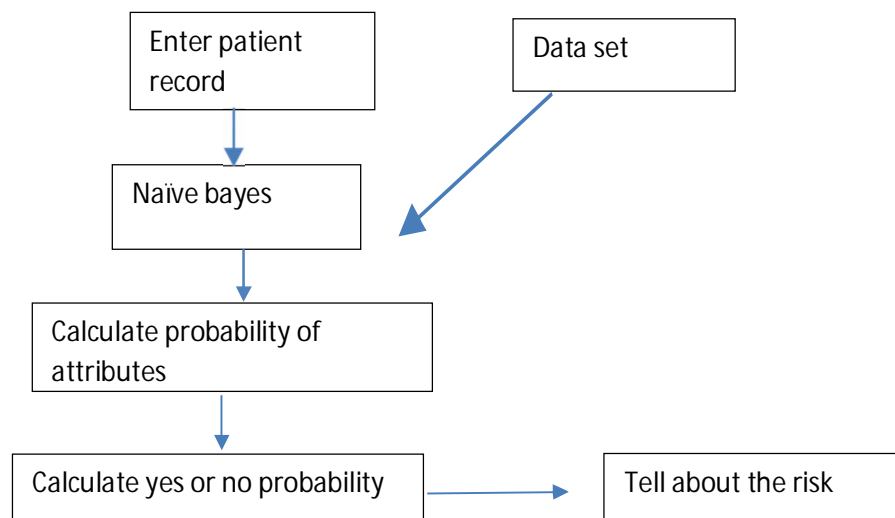
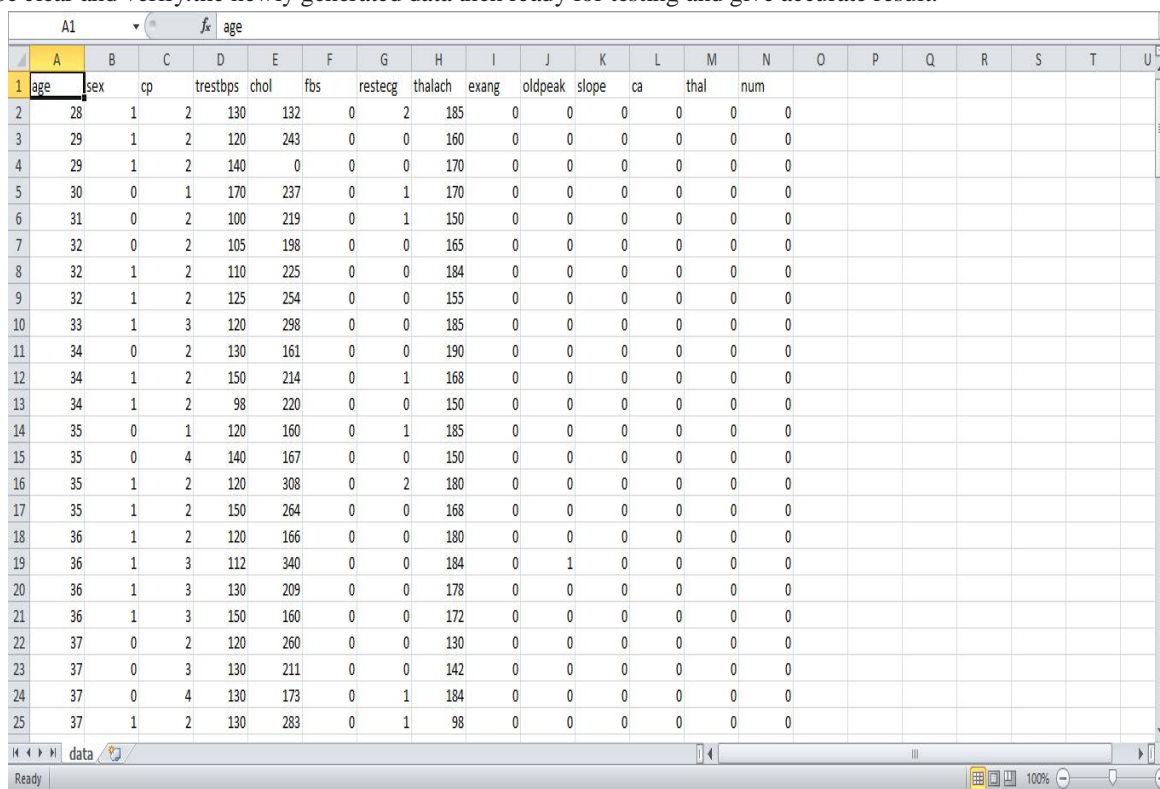


Fig a. Implementation of Naïve Bayes Classification algorithm on patient data.

III. METHODOLOGY

We are here using raw data or data that has been extracted from large data set to generate the knowledge discovery. we preprocess data in such a way that it can give accurate outcome. there are lots of anomalies, irregularity, divergence some errors found in data so it has to be clear and verify. the newly generated data then ready for testing and give accurate result.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num							
2	28	1	2	130	132	0	2	185	0	0	0	0	0	0							
3	29	1	2	120	243	0	0	160	0	0	0	0	0	0							
4	29	1	2	140	0	0	0	170	0	0	0	0	0	0							
5	30	0	1	170	237	0	1	170	0	0	0	0	0	0							
6	31	0	2	100	219	0	1	150	0	0	0	0	0	0							
7	32	0	2	105	198	0	0	165	0	0	0	0	0	0							
8	32	1	2	110	225	0	0	184	0	0	0	0	0	0							
9	32	1	2	125	254	0	0	155	0	0	0	0	0	0							
10	33	1	3	120	298	0	0	185	0	0	0	0	0	0							
11	34	0	2	130	161	0	0	190	0	0	0	0	0	0							
12	34	1	2	150	214	0	1	168	0	0	0	0	0	0							
13	34	1	2	98	220	0	0	150	0	0	0	0	0	0							
14	35	0	1	120	160	0	1	185	0	0	0	0	0	0							
15	35	0	4	140	167	0	0	150	0	0	0	0	0	0							
16	35	1	2	120	308	0	2	180	0	0	0	0	0	0							
17	35	1	2	150	264	0	0	168	0	0	0	0	0	0							
18	36	1	2	120	166	0	0	180	0	0	0	0	0	0							
19	36	1	3	112	340	0	0	184	0	1	0	0	0	0							
20	36	1	3	130	209	0	0	178	0	0	0	0	0	0							
21	36	1	3	150	160	0	0	172	0	0	0	0	0	0							
22	37	0	2	120	260	0	0	130	0	0	0	0	0	0							
23	37	0	3	130	211	0	0	142	0	0	0	0	0	0							
24	37	0	4	130	173	0	1	184	0	0	0	0	0	0							
25	37	1	2	130	283	0	1	98	0	0	0	0	0	0							

A. Dataset

As shown in above picture there is dataset where it contain the record of the patients it displays the various attributes such as sex, chest pain, fbs (fasting blood sugar), exang etc. so by analyzing those attribute for a particular patient we examined that if that person is having heart disease or not. we have collected different patients record/data from hospitals. There are various phase for conducting these process. as it is shown in **flowchart**.

Here is the snippet of importing dataset:

```
# Importing the dataset
dataset = pd.read_csv('data.csv')
print("\n \t The data frame has {0[0]} rows and {0[1]} columns. \n".format(dataset.shape))
dataset.info()
dataset.head(3)
X = dataset.iloc[:, [1,3]].values
y = dataset.iloc[:, 6].values
```

From above snippet it shows that it include file name data.csv and print the data.iloc in above is used to select data by row number.

```
# Splitting the dataset into the Training set and Test set
from sklearn.cross_validation import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25, random_state = 0)
```

From above snippet the dataset is further splitted into training and testing state.

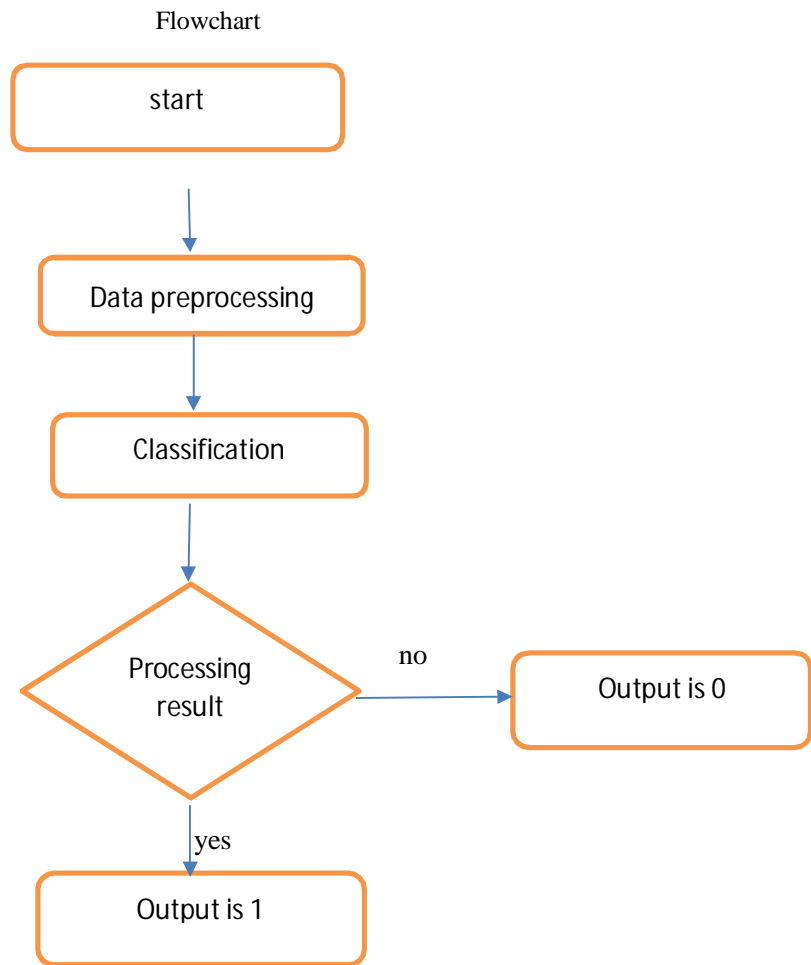


Fig b. flow chart

- 1) *Data Preprocessing*: It is process of preparing data for performing operation.it involves data selection, data transformation and filtering etc.it analyze dataset. The large amount of data is required to generate the data mining and this step makes it easy for further operation. Data will well trained and process. After data has been selected there is further process that is
 - a) Formatting the data in well structured manner. 2]Cleaning the data to remove the ambiguity.
 - b) Sampling of data to remove the running time for memory resources.
- 2) *Classification*: It is process of dividing and organizing data into categories and make the efficient and proper use of it. The goal of classification is to accurately predict the target class.e.g. here the disease will be classified as either present or absent.
- 3) *Output*: It shows the result of the analysis. Here it will sow whether the person having disease yes or not or if they have then how much amount or it will specifies the range.

IV. RESULT & DISCUSSION

We have implemented on the heart patient dataset it consists of various attributes and values.here we are using the python language on the file and generated the result which shows the outcome of person.

```

# Fitting Naive Bayes to the Training set
from sklearn.naive_bayes import GaussianNB
classifier = GaussianNB()
classifier.fit(X_train.astype(float), y_train.astype(float))
  
```

```
# Predicting the Test set results
y_pred = classifier.predict(X_test)

# Making the Confusion Matrix
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)
```

Above given snippet specifies about classifier that describe training vector and target vector. **fit** will ignore the outliers. Confusion matrix is used to evaluate the accuracy of the classification.

```
# Visualising the Training set results
from matplotlib.colors import ListedColormap
X_set, y_set = X_train, y_train
X1, X2 = np.meshgrid(np.arange(start = X_set[:, 0].min() - 1, stop = X_set[:, 0].max() + 1, step = 0.01),
                    np.arange(start = X_set[:, 1].min() - 1, stop = X_set[:, 1].max() + 1, step = 0.01)
)
plt.contourf(X1, X2, classifier.predict(np.array([X1.ravel(), X2.ravel()]).T).reshape(X1.shape),
             alpha = 0.75, cmap = ListedColormap(('red', 'green'))) # red= x1 and green =x2
plt.xlim(X1.min(), X1.max())
plt.ylim(X2.min(), X2.max())
```

Above code is small glimpse of how the visualising chart will generate the outcome of data. As shown below the chart shows the range in blue and red color the red color shows the extreme that is the person fall in that range has the heart disease where as other shows the minimum possibility of having the disease .

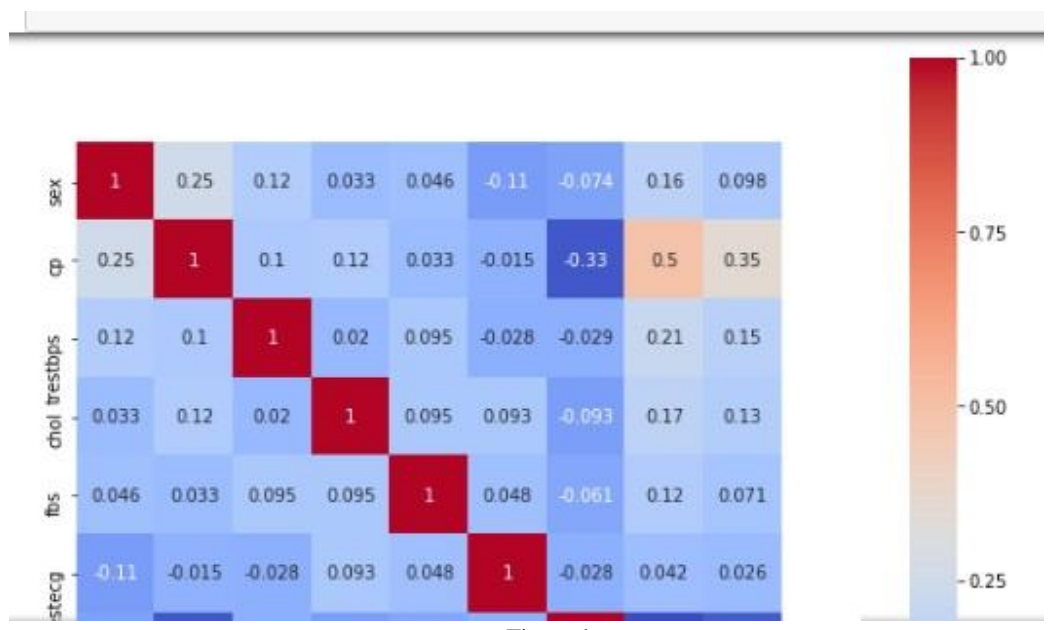


Fig. c chart

Hence the system will generate the result and calculate the result using naïve bayes formula.it is faster than compared to any other algorithm. which predict the accurate outcome.

V. CONCLUSION

As per research and review there are many algorithm by which we can predict the heart disease analysis but naïve bayes is faster and easily executable rather than other and its accuracy is also maximum.it gives accurate result.so its better to use the naive bayes classification method for data analysis.so in this paper we implemented the system which uses the naïve bayes methodology.



REFERENCES

- [1] Ramandeep Kaur, Er. Prabhsharn Kaur A Review - Heart Disease Forecasting Pattern using Various Data Mining Techniques
- [2] B.Venkatalakshmi, M.V Shivsankar Heart Disease Diagnosis Using Predictive Data mining
- [3] Himanshu Sharma, M A Rizvi, Prediction of Heart Disease using MachineLearning Algorithms: A Survey
- [4] Shadab Adam Pattekari and Asma Parveen *Shadab Adam Pattekari and AsmaParveen
- [5] Vincy Cherian, Bindu M.S Heart Disease Prediction Using Naïve Bayes Algorithm and Laplace Smoothing Technique
- [6] Syed Immamul Ansarullah, Pradeep Kumar Sharma, Abdul Wahid, Mudasir M Kirmani Heart Disease Prediction System using Data Mining Techniques: A study
- [7] Sohana Saiyed, Nikita Bhatt and Dr. Amit P. Ganatra IM A Survey on Naive Bayes Based Prediction of Heart Disease Using Risk Factors
- [8] Business Intelligence - Carlo Vercellis



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)