

Heart Disease Prediction Using Machine Learning Algorithm

Prof. Waghmode P.S., Kharat Dipali ,Wayase Dhanashri

Assistant Professor, Computer Department, SPCOET Someshwarnagar College, Baramati, India Student, Computer Department, SPCOET Someshwarnagar College, Baramati, India Student, Computer Department, SPCOET Someshwarnagar College, Baramati, India

Abstract -

In recent times, Heart Disease prediction is one of the most complicated tasks in medical field. In the modern era, approximately one person dies per minute due to heart disease. Data science plays a crucial role in processing huge amount of data in the field of healthcare. The proposed work predicts the chances of Heart Disease and classifies patient's risk level by implementing different data mining techniques such as Naive Bayes, Decision Tree, Logistic Regression and Random Forest. Support Vector Machine. Thus, this paper presents a comparative study by analysing the performance of different machine learning algorithms. The trial results verify that Support Vector Machine has achieved the highest accuracy of 80.32% compared to other ML algorithms implemented.

Key Words: Machine Learning, Support vector machine Algorithm, Jupiter, Flask framework, vs code.

INTRODUCTION

Human heart is the principal part of the human body. Basically, it regulates blood flow throughout our body. Any irregularity to heart can cause distress in other parts of body. Any sort of disturbance to normal functioning of the heart can be classified as a Heart disease. In today's contemporary world, heart disease is one of the primary reasons for occurrence of most deaths. Heart disease may occur due to unhealthy lifestyle, smoking, alcohol and high intake of fat which may cause hypertension. According to the World Health Organization more than 10 million die due to Heart diseases every single year around the world.

MOTIVATION

The motivation to do this problem comes from World Health Organization estimation. According to the World Health Organization estimation till 2030, very nearly 23.6 million individuals will pass on because of Heart malady. So to minimize the danger, expectation of coronary illness ought to be finished.

OBJECTIVE

The main goal of this paper is to provide a tool for doctors to detect heart disease as early stage. This in turn will help to provide effective treatment to patients and avoid severe consequences. ML plays a very important role

to detect the hidden discrete patterns and thereby analyse the given data

PROBLEM STATEMENT

Mainly our problem statement is classification and we use algorithms like Logistic Regression, Decision Trees, Naïve- Bayes, KNN, Random Forest, and SVM to predict the heart disease.

LITERATURE SURVEY

A literature review is a text of a scholarly paper, which includes the current knowledge including substantive findings, as well as theoretical and methodological contributions to a particular topic. Literature reviews use secondary sources, and do not report new or original experimental work

1. Paper name: "Predicting the Risk of Heart Failure With EHR Sequential Data Modeling"

Author: Bo Jin, Chao Che et al.

model designed by applying neural network. This paper used the electronic health record (EHR) data from real-world datasets related to congestive heart disease to perform the experiment and predict the heart disease before itself. We tend to use one-hot encoding and word vectors to model the diagnosing events and forecast coronary failure events. The essential principles of an extended memory network model. By analyzing the results, we tend to reveal the importance of respecting the sequential nature of clinical records.

2. paper name: "Heart Disease Prediction using Evolutionary Rule Learning"

Author. Aakash Chauhan et al.

This study eliminates the manual task that additionally helps in extracting the information (data) directly from the electronic records. To generate strong association rules, we have applied frequent pattern growth association mining on patient's dataset. This will facilitate (help) in decreasing the amount of services and shown that overwhelming majority of the rules helps within the best prediction of coronary sickness

3. paper name: "An Intelligent Learning System based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection"

Author: Ashir Javeed, Shijie Zhou et al

This paper uses random search algorithm (RSA) for factor selection and random forest model for diagnosing

the cardiovascular disease. This model is principally optimized for using grid search algorithmic program. Two forms of experiments are used for cardiovascular disease prediction. In the first form, only random forest model is developed and within the second experiment the proposed Random Search Algorithm based random forest model is developed.

4. paper name: “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques

Author: Senthilkumar Mohan, Chandrasegar Thirumalai et al

.efficient technique using hybrid machine learning methodology. The hybrid approach is combination of random forest and linear method. The dataset and subsets of attributes were collected for prediction. The subset of some attributes were chosen from the pre- processed knowledge(data) set of cardiovascular disease .After prep-processing ,the hybrid techniques were applied and diagnosis the cardiovascular disease

5. paper name: “Fast Rule-Based Heart Disease Prediction using Associative Classification Mining”.

Author : K.Prasanna Lakshmi, Dr. C.R.K.Reddy

In the proposed Stream Associative Classification Heart Disease Prediction (SACHDP), we used associative classification mining over landmark window of data streams. This paper contains two phases: one is generating rules from associative classification mining and next one is pruning the rules using chi-square testing and arranging the rules in an order to form a classifier. Using these phase to predict the heart disease easily

METHODOLOGY

The used heart disease data set was taken and data processing has done to filter the data and to remove some unnecessary data. The model was trained with the processed data using the support vector machine algorithm to predict the heart disease of used with higher accuracy. Fig 1shows the structured outline for proposed Methodology.

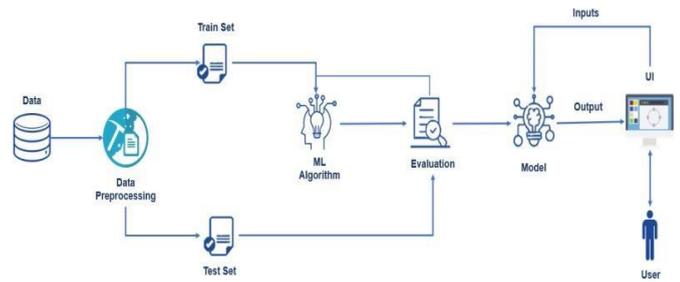


Fig 1: Structured outline of Proposed Methodology

A) Dataset Collection

It is the process of gathering the information from the source for the evaluation. The Used Heart data set is collected from a website Kaggle which is in a CSV format. The data set contains 14 variables as shown in Fig 2.

B) Data Preprocessing

This step is one of the important steps in supervised machine learning. It includes the following.

i) Removal of Non-numerical part from numerical features :

This step removes the non-numerical words from the features like slope, sex, exang ,restcgp,chol and cp for data processing. Step1: Converting the data frame into the list Step2: Splits the list based on a delimiter. Step3: Store the required data back to the data frame

ii)Converting Categorical values into numerical

Here, the categorical values like slope, sex, exang ,restcgp,chol ,fbs and cp, are converted to numerical because machine learning deal with numerical values easily because of the machine-readable form. This is done by using Label Encoder which is a python package.

Step1: We have to select categorical values based on its datatype.

Step2: Converting the categorical values into numericalvalues by using Label encoder API in python.

iii)Separate the target variable :

Here, we have to separate the target feature which is we are going to predict.

Step1: The target variable is assigned to the variable ‘y’.

Step2: The preprocessed data set except the target variable is assigned to the variable ‘X’.

After all preprocessing steps have done, the data was shown As in a fig. 3

```
In [49]: M df.head()

Out[49]:
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0

Fig 2: Sample data before data preprocessing

Here,
 y =Dependent variable w = hyperplane
 x = independent variable b =constant

The values of x and y variables are training datasets of svm model representation.

RESULT

Fig 4: Prediction page

```
In [16]: M cate_val

Out[16]: ['sex', 'cp', 'fbs', 'restecg', 'exang', 'slope', 'ca', 'thal', 'target']

In [17]: M df[['cp']].unique()

Out[17]: array([0, 1, 2, 3], dtype=int64)

In [18]: M cate_val.remove('sex')
M cate_val.remove('target')
df=pd.get_dummies(df,columns=cate_val,drop_first=True)

In [19]: M df.head()

Out[19]:
```

	age	sex	trestbps	chol	thalach	oldpeak	target	cp_1	cp_2	cp_3	restecg_1	slope_1	slope_2	ca_1	ca_2	ca_3	ca_4	thal_1	thal_2	thal_3			
0	52	1	125	212	168	1.0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	
1	53	1	140	203	155	3.1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
2	70	1	145	174	125	2.6	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
3	61	1	148	203	161	0.0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1
4	62	0	138	294	106	1.9	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0

Fig 3: Sample data after data preprocessing



Fig 6: Prediction page

ALGORITHM

Support Vector Machine Algorithm

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n- dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

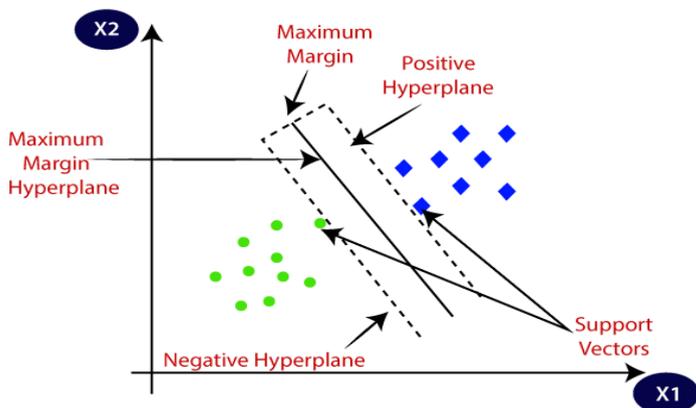


Fig 3: SVM Algorithm

The final prediction system has been incorporated into the HTML CSS GUI application for the Heart Disease prediction. The simple application GUI, shown in Fig. 5 and fig.6 enables to understand whether the particular person is heart disease or not.

CONCLUSIONS

In this paper, we proposed a method for heart disease prediction using machine learning techniques, these results showed a great accuracy standard for producing a better estimation result. By introducing

new proposed Random forest classification, we find the problem of prediction rate without equipment and propose an approach to estimate the heart rate and condition. Sample results of heart rate are to be taken at different stages of the same subjects, we find the information from the above input via ML Techniques. Firstly, we introduced a support vector classifier based on datasets.

FUTURE WORK

The future scope of this system aims at giving more sophisticated prediction models, risk calculation tools and feature extraction tools for other clinical risks.

ACKNOWLEDGEMENT

First and foremost, we would like to thank All, for giving us the strength to complete the research project we started. I would like to take this opportunity to thank my guides **Prof. Waghmode P.S.** the instructor for coaching & sharing ideas & real-life examples throughout the course. Without their guidance and support, this report could not be completed. We are also grateful to **Prof. Shah.S.N** Head of Computer Engineering Department, **SSPM's SHARADCHANDRA PAWAR COLLEGE OF ENGINEERING AND TECHNOLOGY** for her indispensable support, suggestions. We are grateful to Principal **Dr. Deokar** for his encouragement and for giving us guidance and valuable time for the discussion. We express our sincere thanks to all the staffs, parents and colleagues who have helped us directly in completing this project. We are grateful for the many useful comments and suggestions provided by reviewers, which have resulted in significant improvements in project.

REFERENCES

- [1] A. S. ABDULLAH AND R. R. RAJALAXMI, "A DATA MINING MODEL FOR PREDICTING THE CORONARY HEART DISEASE USING RANDOM FOREST CLASSIFIER," IN PROC. INT. CONF. RECENT TRENDS COMPUT. METHODS, COMMUN. CONTROLS, APR. 2012, pp. 22–25.
- [2] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, "Using PSO algorithm for producing best rules in diagnosis of heart disease," in Proc. Int. Conf. Comput. Appl. (ICCA), Sep. 2017, pp. 306–311.
- [3] N. Al-milli, "Backpropagation neural network for prediction of heart disease," J. Theor. Appl. Inf.

Technol., vol.56, no. 1, pp. 131–135, 2013.

- [4] C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika, "Analysis of neural networks based heart disease prediction system," in Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI), Gdansk, Poland, Jul. 2018, pp. 233–239.
- [5] P. K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," J. King Saud Univ.- Comput. Inf. Sci., vol.24, no.1, pp.27–40, Jan.2012. doi:10.1016/j.jksuci.2011.09.002.
- [6] L. Baccour, "Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets," Expert Syst. Appl., vol. 99, pp. 115–125, Jun. 2018. doi: 10.1016/j.eswa.2018.01.025.
- C.-A. Cheng and H.-W. Chiu, "An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide database," in Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2017, pp. 2566–2569.
- [8] H. A. Esfahani and M. Ghazanfari, "Cardiovascular disease detection using a new ensemble classifier," in Proc. IEEE 4th Int. Conf. Knowl. Based Eng. Innov. (KBEI), Dec. 2017, pp. 1011–1014.
- [9] F. Dammak, L. Baccour, and A. M. Alimi, "The impact of criterion weights techniques in TOPSIS method of multi-criteria decision making in crisp and intuitionistic fuzzy domains," in Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE), vol. 9, Aug. 2015, pp. 1–8.
- [10] R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," Expert Syst. Appl., vol. 36, no. 4, pp. 7675–7680, May 2009. doi: 10.1016/j.eswa.2008.09.013.





