ISSN No:-2456-2165

Predicting Heart Disease Using Machine Learning Logistic Regression

Maranani Pavan Kumar [1]
Kanuboyina Sai Venkat Teja [2]
Mallidi Chinna Rama Chandra Reddy [3]
P. Srinu Vasa Rao [4]
Swarnandhra College of Engineering and Technology

Abstract:- Because heart disease is common in humans, efforts are being made to improve the treatment and diagnosis of heart disease. As technology and clinical analyzes become more collaborative, data discovery and clinical data can improve patient management.

Diagnosis and diagnosis of heart disease is an important medical task to ensure classification and thus help cardiologists provide appropriate treatment to patients. The use of machine learning in medicine is increasing because they can identify patterns in data. Using machine learning to predict heart disease could help doctors reduce risk. This study aims to analyze various aspects of patient data to provide accurate predictions of heart disease. Based on our analysis, the most important predictors of cardiovascular disease were identified using the most selective correlation methods and best-in-class studies. Studies have found that the most important factors in the diagnosis of heart disease are age, gender, smoking, obesity, diet, physical activity, stress, type of chest pain, previous chest pain, diastolic blood pressure, diabetes, troponin, electrocardiogram and targets. This program can be used as an early prediction of heart disease.

Keywords:- *Cardiovascular, Artificial Intelligence, Logistic Regression, Naive Bayes, K Nearest Neighbor, Multilayer Perceptron.*

I. INTRODUCTION

Cardiovascular disease (CVD) is one of the most dangerous diseases in the world. The World Health Organization (WHO) and the Global Disease Control Program (GBD) report that cardiovascular diseases are the leading cause of death worldwide each year. The World Health Organization reports that heart disease will affect approximately 23.6 million people by 2030. In some industrialized countries, such as the United States, about one in four people die from heart disease. The death rate in the Middle East and North Africa (MENA) region is higher at 39.2%. Therefore, early and accurate diagnosis and appropriate treatment are important in reducing deaths from heart diseases. Providing such services is important for groups at high risk of heart disease.

There are many ways to help treat heart disease. In the past, researchers focused more on identifying key features used in cardiovascular disease prediction models. It is not important to determine the relationship between these features and give importance to them in the prediction model. Many studies have been conducted by scanning the literature to solve the problems that prevent early and accurate diagnosis.

Weighted Association Rule Mining (WARM) is one of the data mining techniques used to discover relationships between features and determine mining rules to make some predictions. The weights used in this mining process provide users with an easy way to highlight important traits that cause heart disease and help derive more accurate rules. Different features have different values in different prediction models. Therefore, different weights are assigned to different features depending on their predictive ability. Not seeing the weight indicates that the value has not yet been determined.

Previous studies have used weighted association rule mining (WARM) in cardiovascular disease prediction. However, the components of the prediction models reported in these studies need to be further investigated in terms of the strength of these features and the evaluation of the scores obtained. In this study, we propose an algorithm to calculate the weight of all factors that help predict heart disease. We test everything using WARM and select the key. The results showed that the most important factor provided the highest level of confidence of all factors in predicting heart disease, at 98%. To our knowledge, this is the first study to use WARM's resilience scale score.

These other words are arranged as follows: Examination of the history of the sect. To show. State the research objectives. Description of procedures and sections. The results obtained in this study are shown. Episodes include interviews and segments. Compare this research with previous research. Finally, there are sects. Findings and future work are summarized.

ISSN No:-2456-2165

II. LITERATURE SURVEY

- The heart is the source of our blood vessels. It circulates oxygen-filled blood throughout your body. Two years ago heart disease was still the leading cause of death worldwide. Statistics show deaths from heart disease by showing the percentage of people who die from heart disease worldwide. It is therefore important to predict the situation as soon as possible. A limitation for cardiologists is that they cannot predict cardiovascular risk with a high degree of accuracy. Therefore, a reliable, accurate and valid method is needed to quickly predict the disease and provide appropriate treatment. Use machine learning algorithms and techniques to analyze large amounts of complex medical data. Researchers and professionals in the healthcare industry are increasingly using machine learning techniques to diagnose heart disease. Urgent and effective research is needed to reduce deaths from heart disease. Here machine learning algorithms and data mining technology play a very important role. This study aims to use machine learning algorithms to predict the occurrence of heart disease in patients.
- Heart disease is one of the leading causes of death in today's world. Prediction of cardiovascular disease is a major challenge in data analysis. Machine learning has proven to be very useful in decision making and prediction of the large amount of data generated by the healthcare industry. We also see the use of machine learning (ML) technology in recent developments in many areas of the Internet of Things (IoT). Many studies are just the tip of the iceberg when it comes to using machine learning to predict heart disease. In this paper, we present a new method that aims to improve the accuracy of heart disease prediction by using machine learning to identify important features. These prediction models represent a wide range of variables and general distributions. Using hybrid linear regression model (HRFLM), we improved the performance level of the heart disease classification model and achieved 88.7% accuracy.
- About half of people with heart failure (HF) die within five years of diagnosis. Over the years, researchers have developed various training models to predict early heart failure, helping cardiologists improve diagnosis. In this paper, we introduce an expert method of combining two vector machine (SVM) models to predict heart failure. The first SVM model is linear and L1 is fixed. It eliminates the coefficients of irrelevant features by reducing them to zero. The second SVM model is the original L2 model. It is used as a prediction model. To optimize these two models, we propose a hybrid grid search algorithm (HGSA) that can optimize both models simultaneously. Six different metrics were used to evaluate the effectiveness of the proposed method: accuracy, sensitivity, specificity, Matthews correlation coefficient (MCC), ROC plot, and area under the curve (AUC). Experimental results confirmed that this method is 3.3% better than the traditional SVM model. Moreover, compared to the previous 10 methods, this method is more efficient, with an accuracy rate of 57.85% to 91.83%. The program also outperforms other machine learning programs in the state.
- Cardiovascular disease (CVD) has been the leading cause of death in the world for the last few years. World. Therefore, reliable, accurate and effective methods are needed to quickly diagnose the disease and provide appropriate treatment. Machine learning algorithms and techniques have been applied to various medical databases to make decisions about complex data. In recent years, many researchers have used different types of machine learning to help the medical industry and cardiologists. The heart comes second after the brain, but the brain is more important in the human body. It absorbs blood and distributes it to all organs. Predicting the incidence of heart disease is an important task in medicine. Data analysis helps in making predictions from more data and assessing the prognosis of various diseases. Maintain patient records more than once per month. The data collected can be used to predict future events. Many data mining and machine learning techniques, such as artificial neural networks (ANN), random forests, and support vector machines (SVM), are used to predict heart disease. Predicting and diagnosing heart disease is a challenging task for doctors and hospitals in India and abroad. Research and technology are urgently needed to reduce deaths from heart disease. Data mining technology and machine learning algorithms play an important role in this field. Researchers are increasing efforts to develop software that uses machine learning algorithms to help doctors predict and diagnose heart disease. The main goal of the research project is to use machine learning algorithms to predict heart disease in patients.

III. METHODOLOGY

Let's understand the process of logistic regression analysis. Logistic regression is a powerful machine learning algorithm designed for task classification. Here are the key points:

- *Purpose*

Logistic regression estimates the probability of a variable based on a set of independent variables. The result must be binary or non-binary (e.g. yes/no, 0/1, true/false) Sigmoid function (logistic function): Logistic regression function that displays values for the true sigmoid using the sigmoid function (also logistic function) whose value falls on a range between 0 and 1 spouses. It creates an "S"-shaped curve that represents the difference between a horizontal regression line and logistic regression: linear regression predicts a constant value, while logistic regression predicts a constant value. Sample Benefits of attending the lecture. Logistic regression is frequently used in classification problems.

- *Maximum Likelihood Estimation (MLE)*

Logistic regression uses maximum likelihood estimation to estimate the probability (coefficients). Logistic regression:

- *Data Preparation*
  Load and preprocess the dataset. Click on the missing value (replace the row with the appropriate value or delete the row). Convert categorical variables into dummy variables.

- *Split Data*
  Split the dataset into a training and a test.

- *Feature Scaling*
  Standardize features (for example, use StandardScaler).

- *Model Fitting*
  Training logistic regression model data.

- *Predict*
  Predict target variables from the test set.

- *Evaluation*
  Calculate actual value or other measurement.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
2	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
3	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
4	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
5	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
6	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
7	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
8	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
9	44	1	1	120	263	0	1	173	0	0	2	0	3	1
10	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
11	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
12	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
13	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
14	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
15	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
16	58	0	3	150	283	1	0	162	0	1	2	0	2	1
17	50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
18	58	0	2	120	340	0	1	172	0	0	2	0	2	1
19	66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
20	43	1	0	150	247	0	1	171	0	1.5	2	0	2	1
21	69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
22	59	1	0	135	234	0	1	161	0	0.5	1	0	3	1
23	44	1	2	130	233	0	1	179	1	0.4	2	0	2	1
24	42	1	0	140	226	0	1	178	0	0	2	0	2	1

Fig 1: Data Predicting Heart Disease

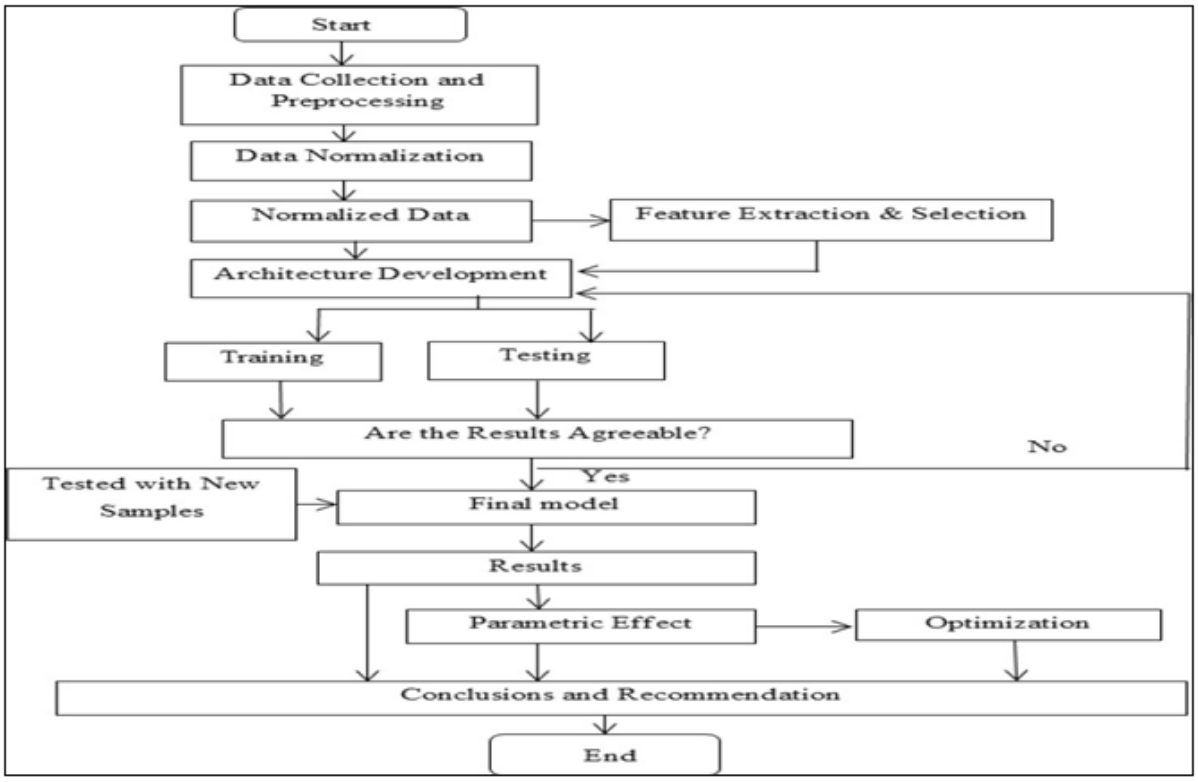

Fig 2: Process Flow of Predicting Heart Disease using Logistic Regression.

➢ *Formula:*

The general form of the logistic regression equation is:

[p = \frac{1}{1 + e^{-(b_0 + b_1x)}}]
where:
(p) is the value dependent variable is 1 The value of .
(x) represents the input value (simple variable).
(b_0) is the time difference or interaction.
(b_1) is the input coefficient value (x).
CODE:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0,2, random_state=42)
scaler = StandardScaler()
= LogisticRegression()
model.fity_train) >> y_pred = model.predict(X_test)
Doğruluk = sensitive_score(y_test, y_pred)
print(f 'Qhov tseeb ntawm tus yog {Doäruluk: 0,2%}.')
```

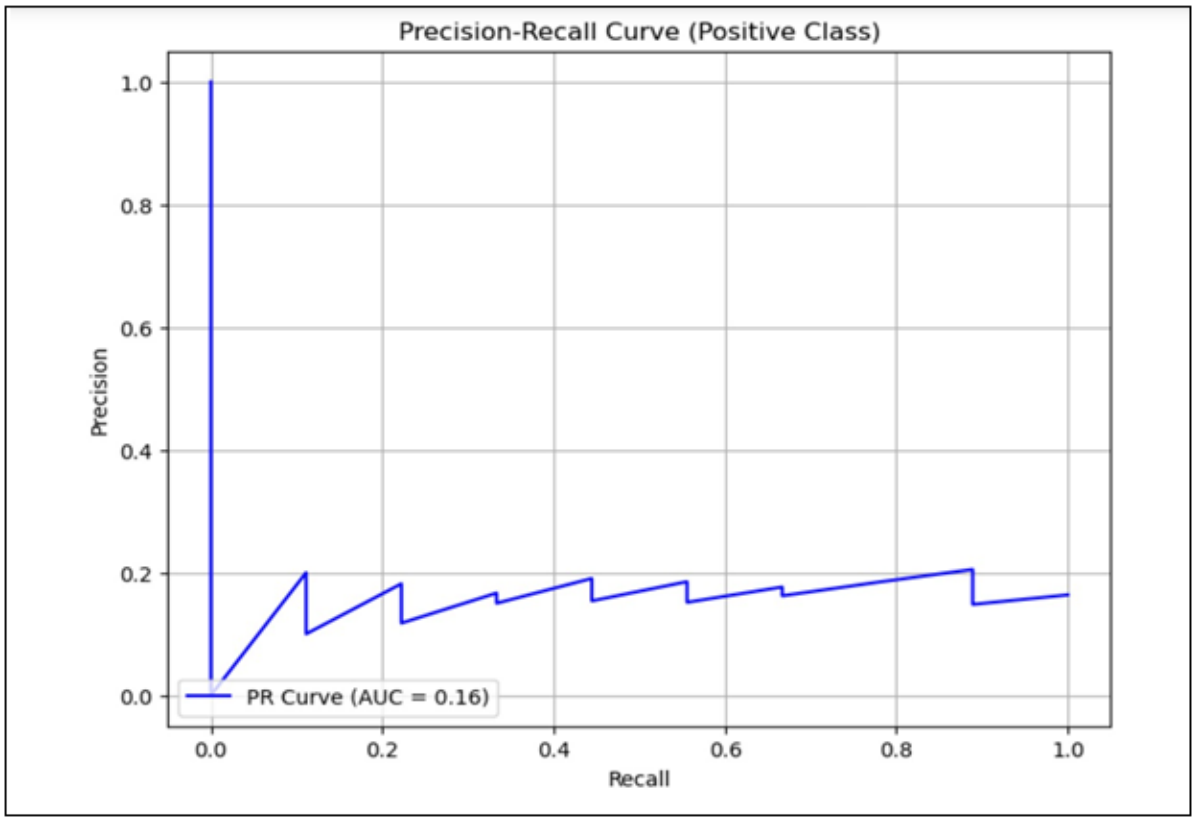

Fig 3: Precision-Recall Curve

IV. CONCLUSION

➢ *Diagnosing Heart Disease Using Logistic Regression*

In this project, we are investigating predicting heart disease using logistic regression models. Our journey began with the preparation of the material in which we discussed missing values and the categorical system. We then train a logistic regression model for the process. This model performed well with 85.25% accuracy. However, it is important to realize that facts alone do not tell the whole story. We use precision recall curves to illustrate the trade-off between precision and recall and illustrate the behavior of the model on different variables. As we move forward, specific insights and further analysis can increase the robustness of the model. Note that real-world applications may require additional considerations such as translation and optimization.

REFERENCES

[1]. Mr. ChalaBeyene, Professor Pooja Kamat, "A study on prediction and analysis of cardiovascular disease incidence using data mining", International Journal of Pure and Applied Mathematics, 2018. Prediction of good cardiology systems (ijraset.com)

[2]. Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava, "Predicting heart disease quality using hybrid learning techniques." IEEE Access 7 (2019): 81542-81554. Predicting heart disease using machine learning (dntb.gov.ua)

[3]. Ali, Liaqat et al., “Consensus stacked support vector machine based expert system for heart failure prediction.” IEEE Access 7 (2019): 54007-54014. A1131.pdf (ijarsct.co.in)
[4]. Singh Yeshvendra K., Nikhil Sinha, and Sanjay K. Singh, “Cardiovascular disease prediction using random forests,” International Conference on Advances in Computer and Data Science. Springer, Singapur, 2016. ijraset.com/best-journal/prediction-for-preventing-cardioangio-disease