www.ijecs.in
International Journal Of Engineering And Computer Science
Volume 13 Issue 09 September 2024, Page No. 26459-26462
ISSN: 2319-7242 DOI: 10.18535/ijecs/v13i09.4900

Evaluation of Machine Learning Model's Accuracy for Heart Attack Prediction

Paras Negi.[1], Manoj Kumar Bisht [2]

[1] Research Scholar, Department of Computer Science, S.S.J. University, Campus, Almora, Uttarakhand, India

[2]Assistant Professor, Department of Computer Science, S.S.J. University, Campus, Almora, Uttarakhand, India

Abstract
Machine learning is a type of artificial intelligence that allows applications to become more accurate at predicting results without being explicitly programmed to do so. Machine learning techniques such as Decision tree, K-Nearest Neighbors (KNN), Naïve Bayes, Support Vector Machine (SVM), Logistic Regression, Multi-layer perceptron etc, are used to predict the heart attack. These studies offer an analysis of the existing machine learning algorithm and provides a comprehensive overview of the previous research and evaluate the accuracy of the machine learning modals. Both low and high-risk patients for a heart attack were evaluated for the study. The results indicate that methods Logistic Regression and Support Vector Machine algorithms outperform other traditional classifiers in terms of prediction accuracy and generalization.

Keywords: Machine learning (ML), Logistic Regression, SVM, Hyperparameter Tuning, Confusion matrix, Accuracy.

1. Introduction

The heart is an essential organ that pumps oxygen-rich blood throughout the body, supplying tissues and organs with nutrition and eliminating waste, all the while preserving important processes and promoting general health. The heart maintains blood circulation through arteries and veins by means of a periodic cycle of contraction and relaxation. There are many risk factors for these diseases, like an unhealthy diet, physical inactivity, tobacco use, alcohol, increased blood pressure, hypertension, exercise induced angina, raised blood sugar, lipids, being overweight, high cholesterol, smoking, etc. The number of heart attack related deaths worldwide rose from 12.4 million in 1990 to 19.8 million in 2022. Additionally, the study calculates that India has a higher age-standardized CVD death rate than the rest of the world, with 272 deaths per 100,000 people, compared to 235 deaths per 100,000 people worldwide. [1]

Machine learning models are useful for the analysis and prediction of heart attack occurrences. In the process of machine learning, the primary factor is the quality of information that the external environment provides to the system. System learns this information as knowledge, and this knowledge is put into the repository. The repository is the second factor that impacts the design of the learning system [2]. The goal of machine learning algorithm is to create models that can identify the prediction accuracy of the dataset. Machine learning (ML) focuses on the study of algorithms and statistical models that enable computers to perform tasks without explicit instructions. Instead, these systems learn from data and progressively improve their capabilities. Machine learning makes it possible for programs

to predict outcomes more accurately. To generate predictions, machine learning (ML) uses a variety of algorithms such as linear regression, support vector machine (SVM), logistic regression, random forest, decision tree, naïve bayes, k-nearest neighbors and so on. The use of machine learning algorithm is to work on different models that learn from a training set and define the accuracy. In this study, we used support vector machine (SVM), Logistic Regression, Decision Tree, Random Forest for prediction of heart attacks.

2. Literature Review

K. Polaraju et al, [3] proposed Prediction of Heart Disease using Multiple Regression Model and they find that Multiple Linear Regression is appropriate for predicting heart disease chance. Based on the results, it is clear that the classification accuracy of Regression algorithm is better compared to other algorithms.

Marjia et al, [4] developed heart disease prediction using j48, SMO, kStar and Multilayer perception using WEKA software. Based on performance from different factor SMO achieve optimum performance than KStar, Multilayer perception and J48 techniques.

Ashwini shetty et al, [5] recommended to develop the prediction system which will diagnosis the heart disease from patient's medical dataset. They used 13 risk factors of input attributes have considered to build the system. After analysis of the data from the dataset, data cleaning and data integration was performed Noura Ajam [6] recommended artificial neural network for heart attack predication. After considering appropriate function, classification accuracy reached to 88. ANN shows result significantly for heart disease prediction. Ashok Kumar Dwivedi et al, [7] used different algorithms like Naive Bayes, Classification Tree, KNN, Logistic Regression, SVM and ANN. The Logistic Regression gives better accuracy compared to other algorithms Sairabi H. Mujawar et al, [8] used k-means and naïve bayes to predict heart disease. This paper contains 13 independent variables for building the system. To extract knowledge from database, data mining techniques such as classification, clustering, classification methods can be used. 13 attributes with total of 300 records were used from the Cleveland Heart Database. This model is to predict whether the patient have heart disease or not based on the values of 13 attributes. Megha Shahi et al, [9] suggested Heart Disease Prediction System using Data Mining Techniques. They used WEKA software for automatic diagnosis of disease and to give qualities of services in healthcare centres. The paper used various algorithms like SVM, Naïve Bayes, Association rule, KNN, ANN, and Decision Tree. The paper recommended SVM is effective and provides more accuracy as compared with other data mining algorithms. Boshra Brahmi et al, [10] developed different classification techniques like J48, Decision Tree, KNN, SMO and Naïve Bayes to evaluate the prediction and diagnosis of heart disease. After this, evaluating some performance in measures of accuracy, precision, sensitivity, specificity is evaluated and compared. J48 and decision tree gives the best technique for heart disease prediction. Kumar Dwivedi [11] used SVM and KNN for heart disease prediction and found that SVM obtained 82% accuracy. Similar for that and M. Kumari et al. [12] used SVM with 84.12% accuracy and R. Sharmila et al, [13] used SVM and gave 85% accuracy. [14] G.Purusothaman and P.Kirshnakumari (2015)[10] had cited various data mining prediction models namely Decision Table, Association Rule, KNN, Artificial Multilayer Perceptron, Naïve Bayes and Hybrid models with accuracies as 76%, 55%, 58%, 85%, 69%, 86% and 96%. Hybrid data mining has outperformed other data mining heart disease diagnosing techniques.

3. Methodology

Machine learning algorithms are used to predict the categorical outcomes. They can recognize patterns in labeled data and predicted new unseen data. The selection of algorithm depends on factors like the nature of the data, the complexity of the problem, and the desired performance metrics.

I. Dataset Source

This paper will use a CSV file dataset from the online repository Kaggle. The shape of the dataset is (310, 14) that contain attribute such as age, gender, blood pressure, chest pain etc.

II. Dataset and Model Description

Gather or collect the heart attack dataset from the specified repository and then comprehend the variables, missing values, and structure of the dataset. After performing a preliminary analysis, select appropriate classification algorithms such as support vector machines, decision trees, random forests, and logistic regression, k-nearest neighbors or etc. For prediction we need to train the model to learn patterns and relationships within the data and then test the modal accuracy by using confusion metrics.

III. Use Hyperparameter Tuning to Get the Maximum Accuracy

4. Confusion Matrix & Measures Derived

Confusion matrix is a binary classification predicts test data sets as positive and negative, and with the help of predicted and Actual output they produce four outcomes: True positive, true negative, false positive, false negative (Fig 1).

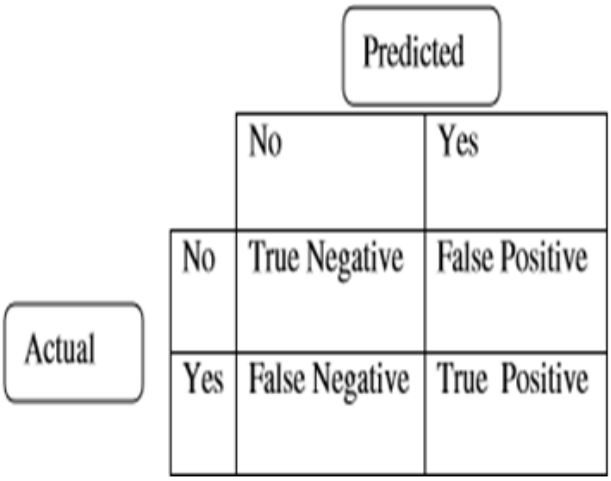

Fig 1: Confusion matrix [15]

Accuracy: The ratio of all accurate predictions to the entire number of true/false, positive/negative predictions is known as accuracy. One represents the best accuracy, and zero represents the lowest.

$$\text{Accuracy} = \frac{(\text{True Positive} + \text{True Negative})}{(\text{True Positive} + \text{True Negative} + \text{False Negative} + \text{False Pos}}$$

5. Conclusion

This paper examined different machine learning algorithms for the prediction of heart attacks. The objectives of our paper were to analyze the Heart Attack dataset by evaluating Machine Learning modals accuracy. For this research paper, we used Decision Tree, Logistic Regression, Random Forest and Support vector machine and we concluded that Logistic Regression and support vector machine has the highest accuracy rate for the prediction of heart attacks, with an Accuracy of 88 %. and 87% respectively. (Below result Table. 1 and 2).

Table 1: Accuracy rate of different algorithms

Algorithms	Accuracy
Decision Tree	0.80
Logistic Regression	0.88
Support Vector Machine	0.87
Random Forest	0.83
Logistic Regression(After Hyperparameter Tuning)	0.90

Table 2: Confusion matrix stats of different Machine Learning Modals

Algorithms	True Negative	False Positive	False Negative	True Positive
Decision Tree	22	5	7	28
Logistic Regression	27	3	4	28
Support Vector Machine	25	4	4	29
Random Forest	24	4	6	28

Now, we can observe that out of all the four models, Logistic Regression gave the highest accuracy of 88%. So, we used use Hyperparameter tuning to get the maximum accuracy.

```
param_grid = {'C': [0.01, 0.1, 1, 10, 100]}
```

The Final Accuracy is: 0.9016 and Best params: {'C': 10}

These findings emphasize the importance of selecting appropriate machine learning modals for heart attack prediction, with Logistic Regression and SVM emerging as promising algorithms for further evaluation.

6. Future Scope

- Future research focus should be enhancing these models' accuracy.
- The accuracy of models can be improved by utilizing more data pre-processing techniques.

- ❖ Combining modals with other ML can create more robust and accurate predictive models.

References

1. https://economictimes.indiatimes.com/industry/healthcare/biotech/heartbeat-unplugged-addressing-the- silentepidemic-of-cardiovascular-diseases- through-an-indian perspective/articleshow/107849231.cms?from=mdr
2. Negi, P., & Bisht, M. (2022). Analysis and prediction of heart attack using machine learning models. In 7th International Conference on Computing, Communication and Security (ICCCS). https://doi.org/10.1109/ICCCS55188.2022.10079409
3. Polaraju, K., & Prasad, D. D. (2017). Prediction of heart disease using multiple linear regression model.International Journal of Engineering Development and Research, 5(2), 2321-9939..
4. Sultana, M., & Haider, A. (2017, March). Heart disease prediction using WEKA tool and 10-fold cross- validation. The Institute of Electrical and Electronics Engineers.
5. Shetty, A. A., & Naik, C. (2016). Different data mining approaches for predicting heart disease. International Journal of Innovative in Science Engineering and Technology, 5(5), 277–281.
6. Ajam, N. (2015). Heart disease diagnoses using artificial neural network. The International Institute of Science, Technology and Education, 5(4), 7–11.
7. Dwivedi, A. K. (2016). Evaluate the performance of different machine learning techniques for prediction of heart disease using ten-fold cross-validation. Springer.
8. Mujawar, S. H., & Devale, P. R. (2015). Prediction of heart disease using modified K-means and Naïve Bayes.International Journal of Innovative Research in Computer and Communication, 3, pages.
9. Shahi, M., & Gurm, R. K. (2017). Heart disease prediction system using data mining techniques. Oriental Journal of Computer Science and Technology, 6, 457-466.
10. Brahmi, B., & Hosseini Shirvani, M. (2015). Prediction and diagnosis of heart disease by data mining techniques. Journals of Multidisciplinary Engineering Science and Technology, 2, 164-168.
11. Dwivedi, A. K. (2018). Performance evaluation of different machine learning techniques for prediction of heart disease. Neural Computing and Applications, 29(10), 685-693.
12. Kumari, M., & Godara, S. (2011). Comparative study of data mining classification methods in cardiovascular disease prediction. International Journal of Computer Science and Technology, 2(2), 10-15.
13. Sharmila, R., & Chellammal, S. (2018). A conceptual method to enhance the prediction of heart diseases using data techniques. International Journal of Computer Science and Engineering.
14. Purusothaman, G., & Krishnakumari, P. (2015). A survey of data mining techniques on risk prediction: Heart disease. Indian Journal of Science and Technology, 12, June.
15. NBSHARE Notebooks. (n.d.). Confusion matrix diagram. NBSHARE. Available at https://www.nbshare.io/notebook/626706996/Learn-And-Code-Confusion-Matrix-WithPyton