iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** III **Month of publication:** March 2023

DOI: https://doi.org/10.22214/ijraset.2023.49376

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue III Mar 2023- Available at www.ijraset.com

Performance Analysis by Using the Knime Analytical Platform to Forecast Heart Failure Using Several Machine Learning Methods

Lakshmi K.C[1], Dr. Arjun B. C[2]
[1]*M.Tech Student,* [2]*HOD, ISE, Rajeev Institute of Technology, Hassan*

Abstract: ***Using a privately available dataset from kaggle.com, this research compares the performance of six well-known machine-learning approaches for predicting heart failure. which include Logistic Regression, Gradient Boosted Trees (GBT), Naive Bayes, Random Forest (RF), and Tree Ensemble. Heart failure is a major public health problem and it is necessary to improve the treatment of heart disease patients to increase the rate of survival. Delicacy was used to assess the performance of machine learning methods. RF produced the highest performance score of 80% when compared to Decision Tree Classifier and Tree Ensemble, Gradient BoostedTrees (GBT), Naive Bayes, and Logistic Regressions.***
Keywords: Machine Learning, KNIME, Classifier, Analytics, Heart Failure

I. INTRODUCTION

The heart is a muscular organ that pumps blood into the body, this process is known as circulation. Apart from this, the blood vessels and blood as a unit is a part of the cardiovascular system.

Heart conditions are the main reason for death worldwide. According to the World Health Organization (WHO), heart attacks cause 17.9 million people to die in 2019.

Using Machine learning algorithms, we can predict heart disease and also many disorders in the medical industry. So that we can save many lives and very easy to deliver successful treatments.

Symptoms of heart disease: coronary artery disease is a common heart condition, and it is very different for men and women. Chest pain, chest tightness, chest pressure, shortness of breath, pain in the upper belly area, weakness or coldness in legs or arms. Heart disease can also cause by irregular heartbeats.

II. LITERATURE SURVEY

Avinash Golande and colleagues [1] studied different ML techniques to predict heart disease. They studied Decision Tree, KNN, and K- Means. Decision Tree was the highest gained accuracy than other algorithms.

B.Gomathy et al. suggested a technique [2] that included data mining methods. According to this, the accuracy of 45% of the testing set was less accurate than the typical fuzzy artificial neural network.

Fahd Saleh Alotaibi [3] studied different machine learning algorithms such as Decision trees, Logistic Regression, Random timber, Naive Bayes, and SVM. The decision tree algorithm had the highest accuracy.

Umair Shafique et al. [4] employed data mining methods, decision trees, Nave Bayes, and Neural Network algorithms, and obtained a delicacy of 82% for Naive Bayes and 78% for the Decision tree.

Sabarinathan Vachiravel et al. [5] suggested a decision tree method to predict heart disease and obtained 85% accuracy.

N.Komal Kumar, G. Sarika Sindhu et.al. [6] used Random Forest, Logistic Regression, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). and obtained an accuracy of 85% for the random forest, 74% for logistic regression, 77% for SVM, and 68% for K-NN.

Malkari Bhargav et al. [7] used ANN, Regression Technique, Random Forest, Decision tree, SVM and KNN. Obtained the accuracy of 96% for ANN, 88% for Regression technique, 83% using SVM, and 68% using KNN.

Gayatri Ramamoorthy et al. [8] suggested different machine learning algorithms such as K-NN and Naïve Bayes, SVM and received the highest accuracy of 80% for K-NN, 65% for SVM, and 80% for Naïve Bayes.

Apurb Rajdhan et al. [9] suggested some machine learning techniques such as Decision tree, logistic retrogression, and naive Bayes and obtained an accuracy of 81% for Decision tree, 85% for Logistic regression, and 85% for Naïve Bayes.

III. .PROPOSED METHODOLOGY

A. Data Collection

The gathered dataset contains 700 records of case data and 11 characteristics. A dataset is information or a tool that is required to do any type of research or design. We gathered information from the dataset provider- Kaggle.com. [10] Fedesoriano is the author of the dataset.(September 20, 2021). Dataset for Predicting Heart Failure.
Date recovered fromhttps://www.kaggle.com/fedesoriano/heart-failure- prediction.

B. About Dataset

Attribute Information

1) *Chest Pain:* chest pain not related to the heart (ATA: Atypical Angina), Non-heart-related (NAP: Non- AnginalPain), Chest pain not showing signs of disease (ASY: Asymptomatic)
2) *Resting BP:* resting blood pressure (mm Hg)
3) *Cholesterol:* cholesterol (mm/ dl)
4) *FastingBS:* fasting blood sugar (1 if FastingBS>120 mg/ dl, 0 else)
5) *Resting ECG:* resting electrocardiogram.
6) *Max HR:* maximum heart rate.
7) *Exercise Angina:* Y- yes, N- no.
8) *Oldpeak:* Number representing the depression score
9) *ST_Slope:* [Up: exercising raises the heart rate (uncommon), Flat: hardly any change, down: signs of a heart illness.]
10) *HeartDisease:* 1- heart disease, 0-Common

C. About KNIME

KNIME gives users the ability to easily create data flows that run either some or all of the analysis processes and then afterward utilize interactive widgets and views to evaluate the results and models.
Steps for creating a KNIME workflw

1) Drag the CSV reader node into KNIME workflow. Drag and drop the .csv file into CSV reader node.
2) Partition Node: Make a training set out of 70% of the data rows and a test set out of the remaining 30%.
3) The decision tree has a learner node to train the model on the training set. This algorithm also has a predictor node to apply the model to other input data. In the test phase, the predictor node is used to apply the trained model to the test data.
4) The scorer node compares the original classes with the predicted classes in the test records and measures the model performance based on the true and false positive, true and false negative.

D. Applying Algorithms

Comparing 6 machine learning algorithms like Decision tree, Random Forest, and Tree Ensemble, Gradient Boosted Trees (GBT), Naive Bayes, Logistic Regression.

1) Decision Tree Classifier

Decision tree are applicable to two different data mining techniques i.e., classification and prediction. It is used to visually define the rules for simple interpretations and understandings. Pre-processing is done in this approach by separating data into training and test data. This algorithm achieves 65% accuracy. Decision tree classifier using KNIME as shown in Figure 1 and Figure 2 shows decision tree.

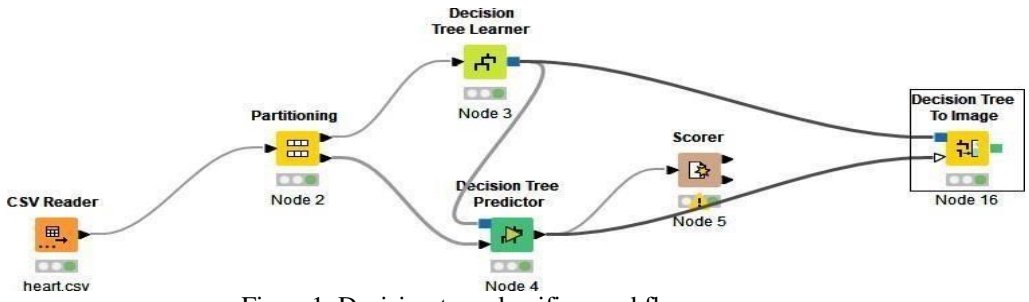


Figure1. Decision tree classifier workflow

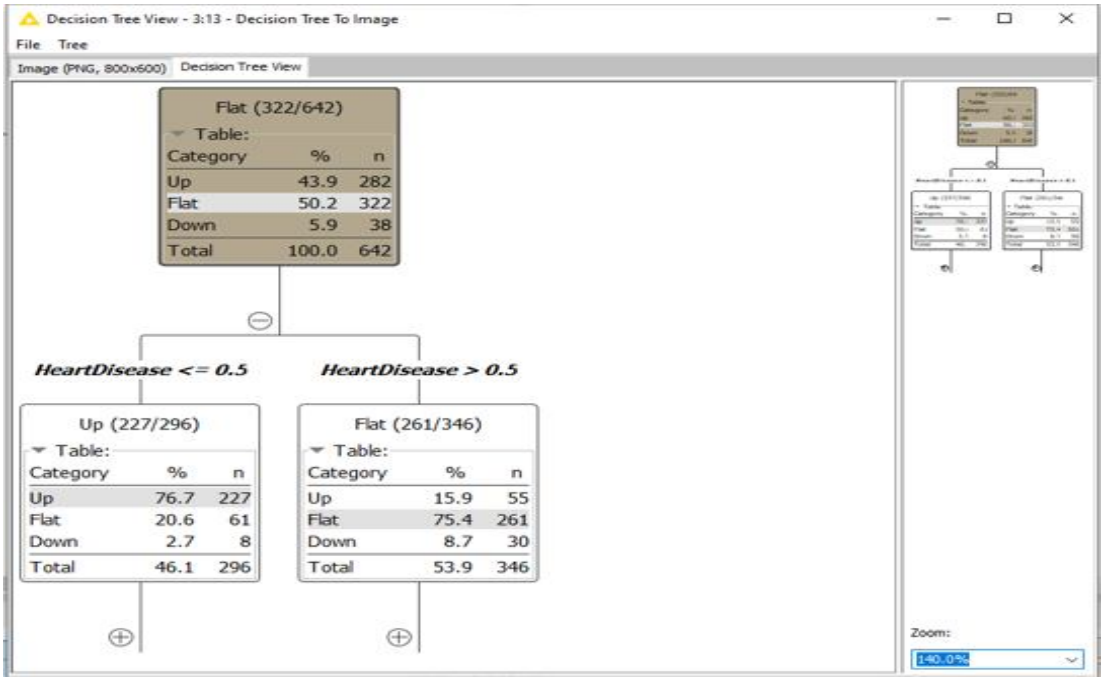


Figure2. Decision Tree

2) Random Forest

A technique called random forest builds many decision trees during the training stage. The ultimate choice is made by the random forest using the majority vote of the trees. The accuracy predicted by the random forest classifier is 80.4%. Random Forest classifier using KNIME as shown in Figure 3.

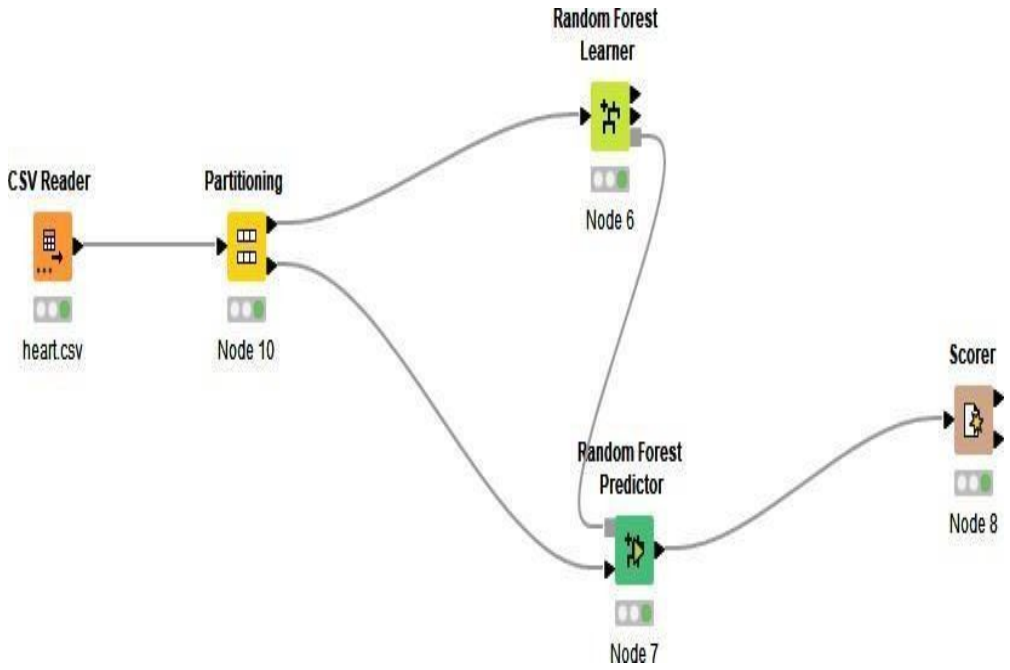


Figure 3. Random Forest Classifier

3) Tree Ensemble

The name ensemble means combining multiple models. We are trying to utilize multiple models and will train to particular datasets and finally will get the output. This algorithm gains 74% accuracy. Tree Ensemble using KNIME as shown in figure 4.

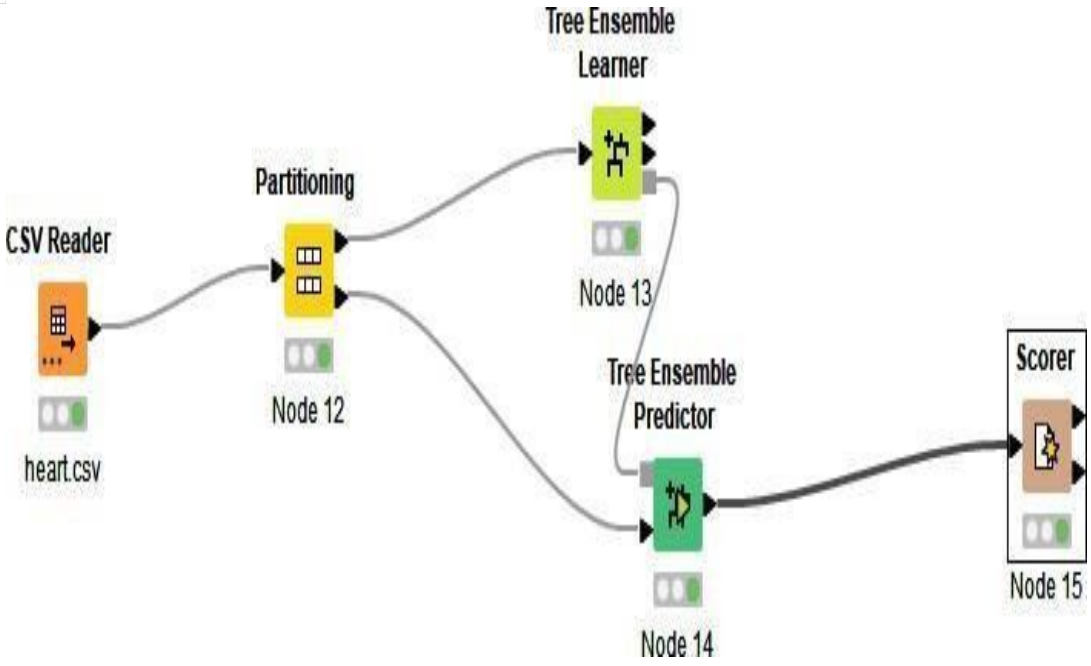


Figure 4. Tree Ensemble model workflow

4) Gradient Boosted Trees (GBT)

Gradient boosted trees are an ensemble approach that combine the results from several trees to accomplish classification or regression. Boosting is a technique used by GBT. Boosting repeatedly combines weak learners such that each new tree fixes the mistakes of the preceding one.

This algorithm achieves 76% accuracy. GBT using KNIME as shown in Figure 5.

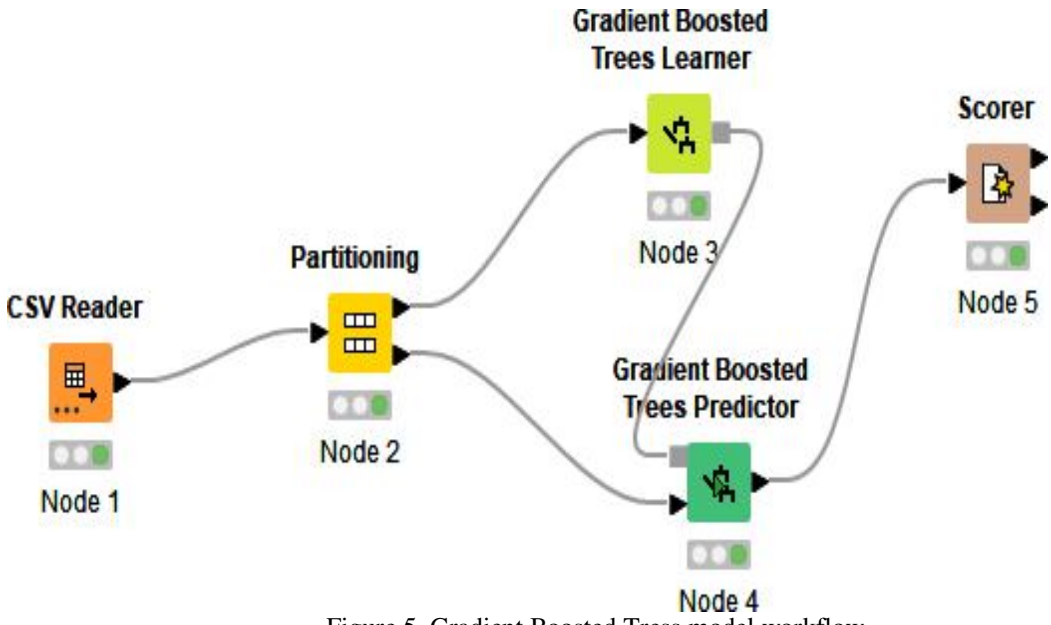
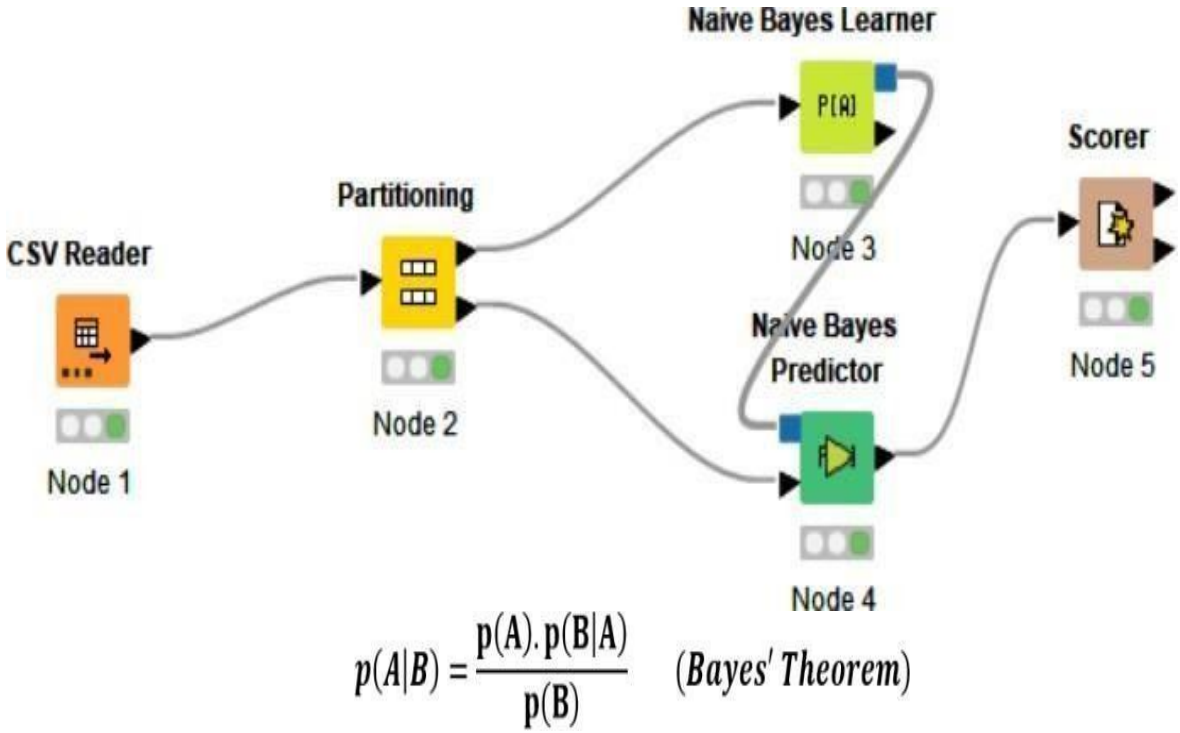


Figure 5. Gradient Boosted Tress model workflow

5) Naïve Bayes Classifier

A classification method called Naive Bayes makes the premise that each characteristic is independent and is based on the Bayes theorem. This algorithm has a 76% percent accuracy. Naïve Bayesmodel using KNIME as shown in the figure 6.



$$p(A|B) = \frac{p(A).p(B|A)}{p(B)} \quad (Bayes' Theorem)$$

Figure 6 Naïve Bayes model workflow

6) Logistic Regression

It is used for binary classification such as spam email identification. It accepts any real-valued number and converts it to a number between 0 and 1. This algorithm achieves 54% accuracy. Logistic Regression using KNIME is shown in figure 7.

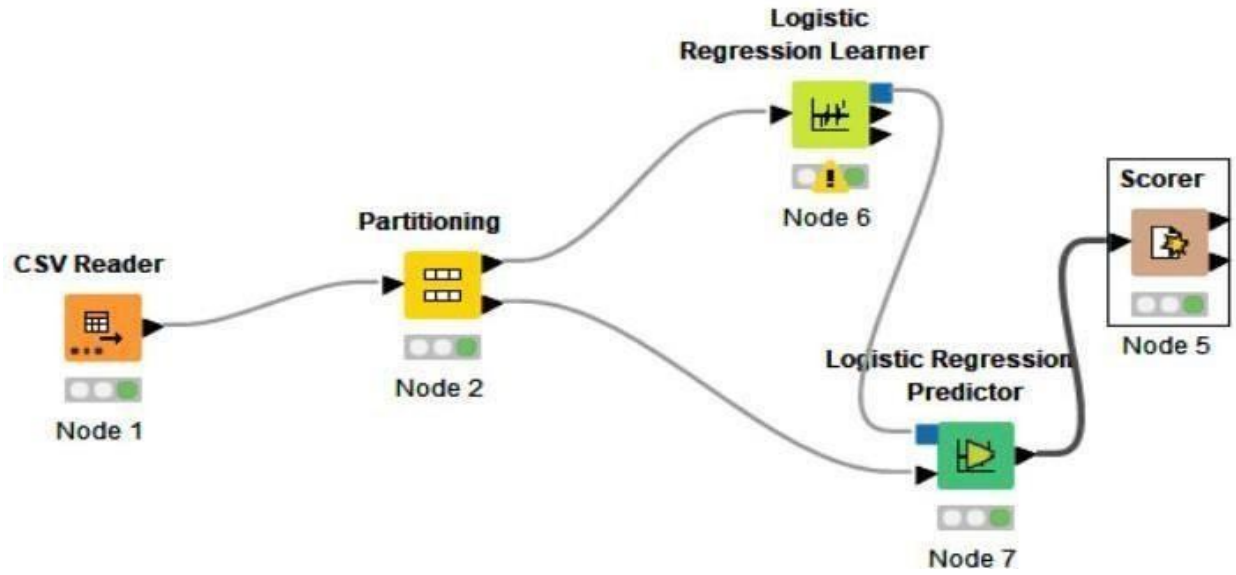


Figure 7. Logistic Regression model workflow

IV. RESULT ANALYSIS

Machine Learning (ML) methods are used on a data set with over 500 rows of data. We used 30% of the data for testing and 70% for training, 40% for testing and 60% for training, and 20% for testing and 80% for training, respectively. Table 1 shows a comparison of all the algorithms. Figure 8 shows a graph of the table 1.

Table 1. Result Comparison

Algorithms	Accuracy		
Partition	70/30	60/40	80/20
Decision Tree Classifier	69.1	72.6	65.2
Random Forest	80.4	78.1	76.1
Tree Ensemble	75	73.9	74.5
Gradient Boosted Trees	76.8	75	73.9
Naïve Bayes Classifier	76	79.1	73.4
Logistic Regression	54	66.3	44.6

Figure 8. Accuracy classification based on training and testing data

V. CONCLUSION

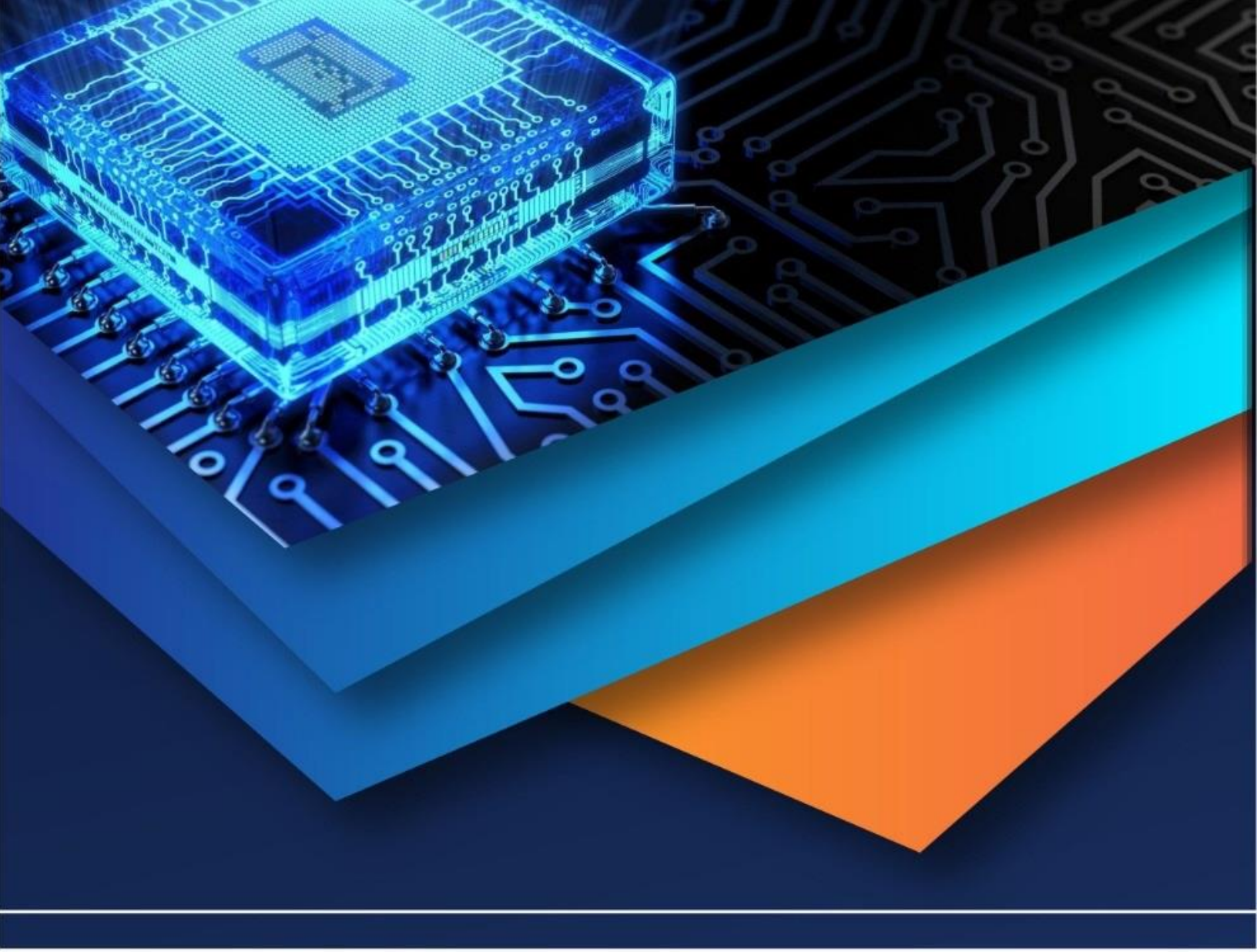
In this paper, we used six different Machine learning techniques to predict heart failure disease. Among these the Random Forest algorithm, which has an accuracy rating of 80.4% in this report's results, is the best algorithm for forecasting heart disease.

VI. ACKNOWLEDGMENT

Firstly, I'd like to express my gratitude to the Rajeev Institute of Technology's Computer Science and Engineering Department, as well as Dr. Arjun B.C., Head, department of Information Science and Engineering. Dr. Prakash H.N., Head of Department, CSE, Rajeev Institute of Technology, Hassan. deserves particular appreciation for his ongoing direction and assistance with our project work.

REFERENCES

[1] Pavan Kumar T and Avinash Golande, "Heart Disease Prediction Using Efficient Machine Learning Methods," International Journal of Current Technology and Engineering, Vol 8, pp.944-950, 2019.
[2] T. Nagamani, S.Logeswari, and B.Gomathy, "Heart Disease Prediction Using Data Mining and the MapReduce Algorithm," International Journal of Innovative Technology and Engineering, vol 8, pp.944-950,2019.
[3] Fahd Saleh Alotaibi, "Implementation of Machine Learning Model to Predict Heart Failure Disorder," (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.
[4] Umair Shafique, Fiaz Majeed, Haseeb Qaiser, and Irfan ul Mustafa "Data mining in healthcare for cardiac disorders," issn 2028-9324 vol. 10, number 4, (2015), pp. 1312-1322.
[5] V. Sabarinathan, "Diagnosis of heart illness using decision tree," international journal of research in computer applications and information technology, vol. 2, number 6, (2014), pp. 74-79.
[6] "Data mining technique to detect cardiac diseases," Vikas Chaurasia, international journal of advanced computer science and information technology (ijacsit), vol. 2, number 4, 2013, pp. 56-66, issn: 2296-1739.
[7] Malkari bhargav and j. Raghunath, "a study on risk prediction of cardiovascular disease using machine learning algorithms," international journal of emerging technologies and creative research (www.jstor.org), vol.7, issue 8, page no.683-688, august 2020.
[8] Gayathri ramamoorthy, "study of heart disease prediction using various machine learning algorithms," international conference on artificial intelligence, smart grid, and smart city applications, 2010. (ais gsc 2019).
[9] Apurb Rajdhan, "heart disease prediction using machine learning," issn: 2278-0181 vol. 9 issue 04, april-2020.
[10] Federico Fedesoriano (September 2021). Dataset on Heart Failure Vaticination. Recaptured (Date recaptured) retrieved from https://www.kaggle.com/fedesoriano/heart-failure- prediction.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY