**ENSEMBLE MACHINE LEARNING-BASED HEART DISEASE PREDICTION WITH HYPER-TUNING PARAMETERS**

**Abstract**

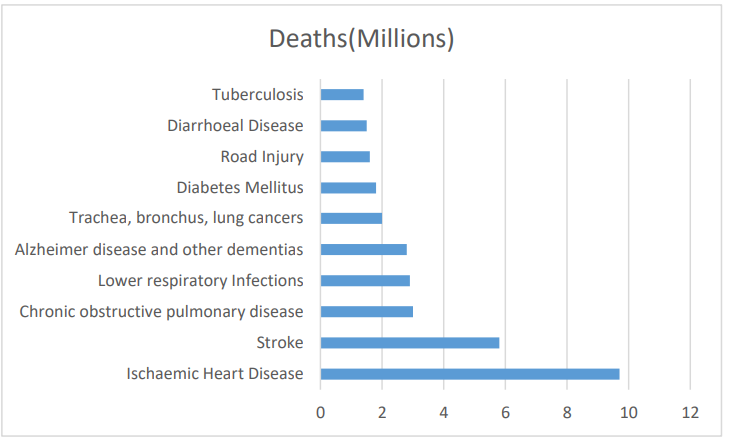
Heart disease is the most dangerous issue in the world. The health sector also faces problems with this disease in the world. A serious investigation avoids the issue in a short time. Optimal prediction generates accurate results for this disease. This type of investigation helps doctors identify diseases and cures for health. Machine learning techniques play a crucial role in this scenario. In health industry utilizes such types of engineering techniques for speedy identification and recovery. Here we use ensemble learning with hyper-tuning parameters of the dataset. Our research observes that if I use the different machine learning models individually, then the accuracy for the decision tree is 70.37%, the Random Forest tree is 79.63%, the Support Vector Machine is 75.93%, and the Logistic Regression predicts 81.48%. But ensemble models of decision tree and Random Forest tree generate an accuracy rate is 71. 29%, and the SVM and LR accuracy rate is 76.85%.

1. **INTRODUCTION**

Our heart is the most crucial organ in our body. The heart consists of the entire coronary arteries. If any one thing is damaged, the entire system of the body is affected immediately. Enduring situations encompass cardiovascular diseases [1]. The World Health Organization has formed certain rules regarding this disease for prevention and treatment. Health sector people also face problems with this disease in the world. A serious investigation avoids the issue in a short time [2]. Optimal prediction generates accurate results for this disease. This type of investigation helps to doctors identify diseases and cures for health. Machine learning techniques play a crucial role in this scenario. In health industry utilizes such type of engineering techniques for speedy identification and recovery [3].

So many emerging technology techniques are used for the critical evaluation of healthcare records and data extraction for easy understanding. Using these techniques, finding the intensity of severity of the disease and maintaining of connection between organs with the heart [4]. Clinical decision-making plays a key role in this context for generating the best disease prevention procedures. In modern lifestyles, due to public eating habits cause cardiac problems [5]. Some other medical conditions are affecting this cardiac disease. In this context, patient health information is most important for handling heart disease patients. Information extraction is the most important and challenging task for medical science [6].

Heart disease is the most life-threatening health issue in the real-world scenario. That means the cardiovascular not pumping the proper amount of blood to the body [13]. It is essential to produce the insusceptible arrangement of the human body. A lot of symptoms include cardiovascular symptoms like breathing problems, balance exhaustion, increased moist pulse rate, and swollen feet [7].

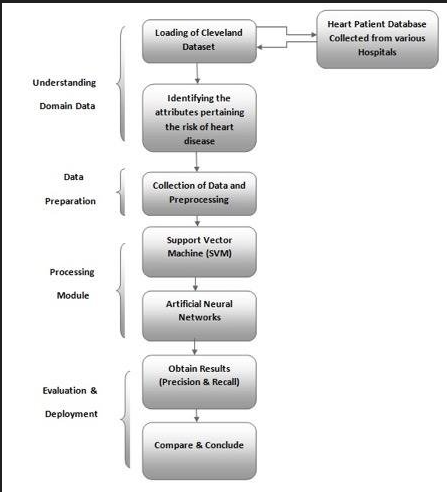


**Figure 1:** 10 Global issues of deaths 2023

The following section discusses the proposed work and architecture of heart disease prediction. Section three states the results and analysis, and the final section concludes the paper.

1. **PROPOSED WORK AND ARCHITECTURE**

We used classification methods, virtuously for construction relations amongst massive databases by simply expecting the results by bearing in mind the type of connotation [14]. This kind of method plays an important role in each facet of science and engineering [8]. The health sector also faces problems with this disease in the world [9]. A serious investigation avoids the issue in a short time. Optimal prediction generates accurate results for this disease [11]. This type of investigation helps doctors for the identification of diseases and improve health. Machine learning techniques play a crucial role in this scenario. In health industry utilizes such type of engineering techniques for speedy identification and recovery [12].

****

**Figure 2:** Architecture of Proposed Work[16]

The following figure 2 displays the building of our projected work. It starts from loading data to predicting results, with multiple phases available for data flow. Finally, get the accurate result.

**III. RESULTS AND ANALYSIS**

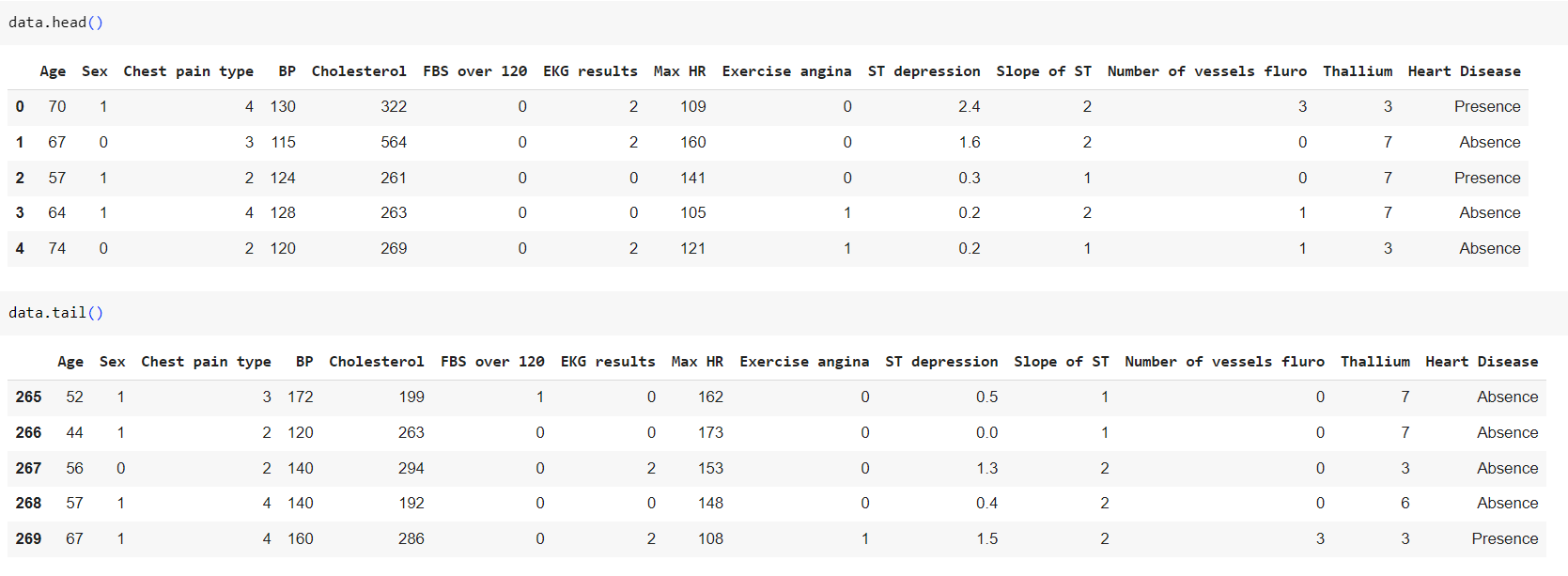
**3.1 Data and Libraries of Python**

After installing the numpy and pandas packages, we are ready to fetch data using the pandas package. Before we use it, we need to know where our dataset is located. This means what is the path of our dataset?

# 3.1 Data Collection

# Data is a collection of attributes. Different types of attributes provide information regarding disease. Load the dataset into the system.

**Table 1:** Dataset [10]

****

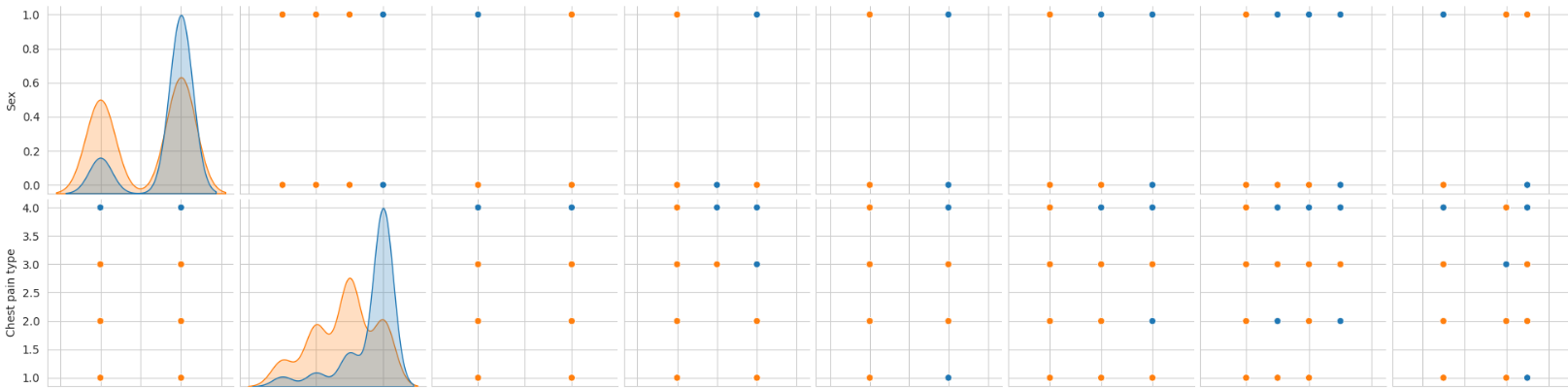
# Data Analysis

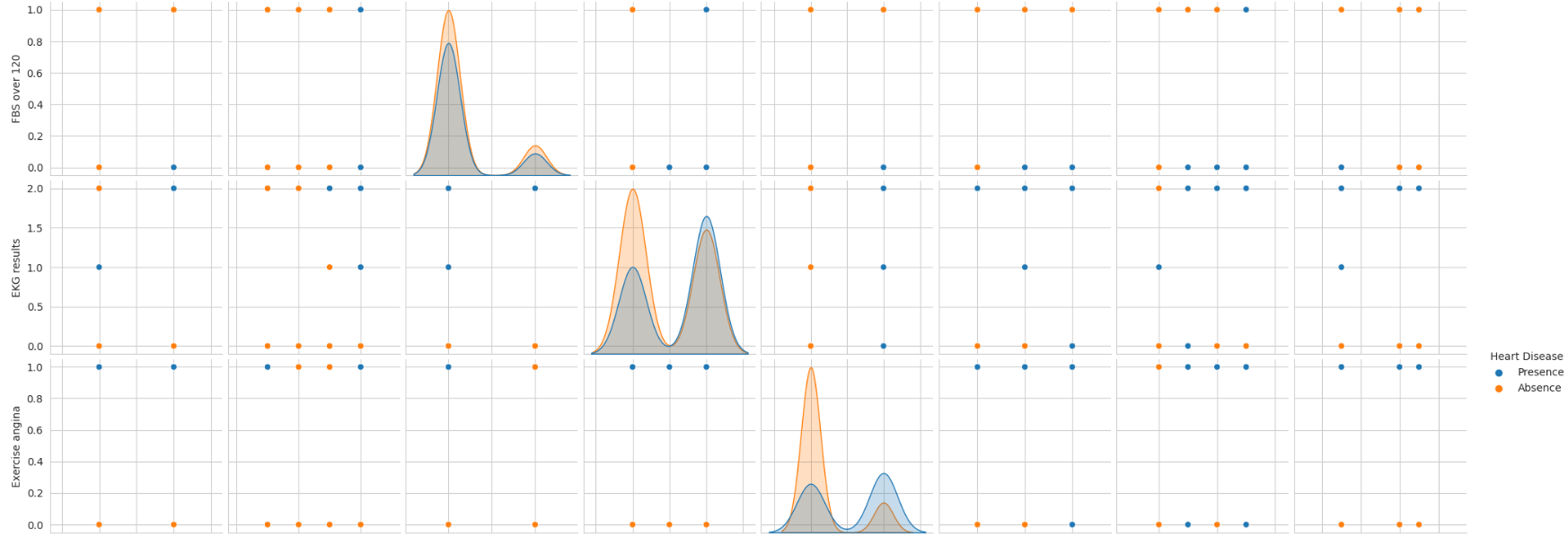
# Table 2: Attribute information

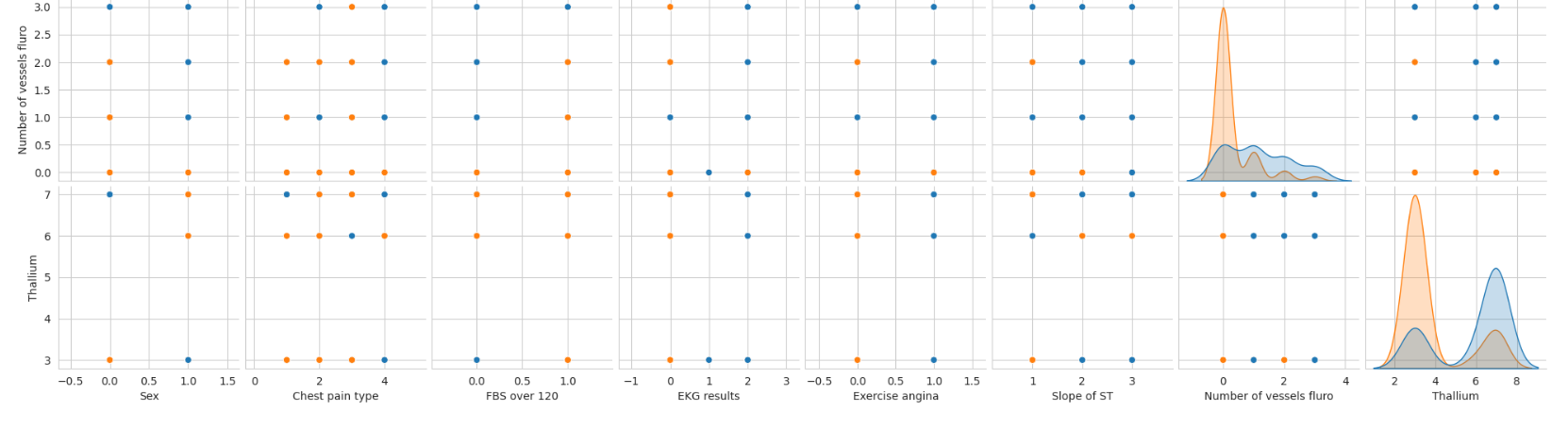
# 

# Data Visualization

# In our research, multiple attributes are used to find results, the follows. Figure 3 shows the data visualization of the attributes.

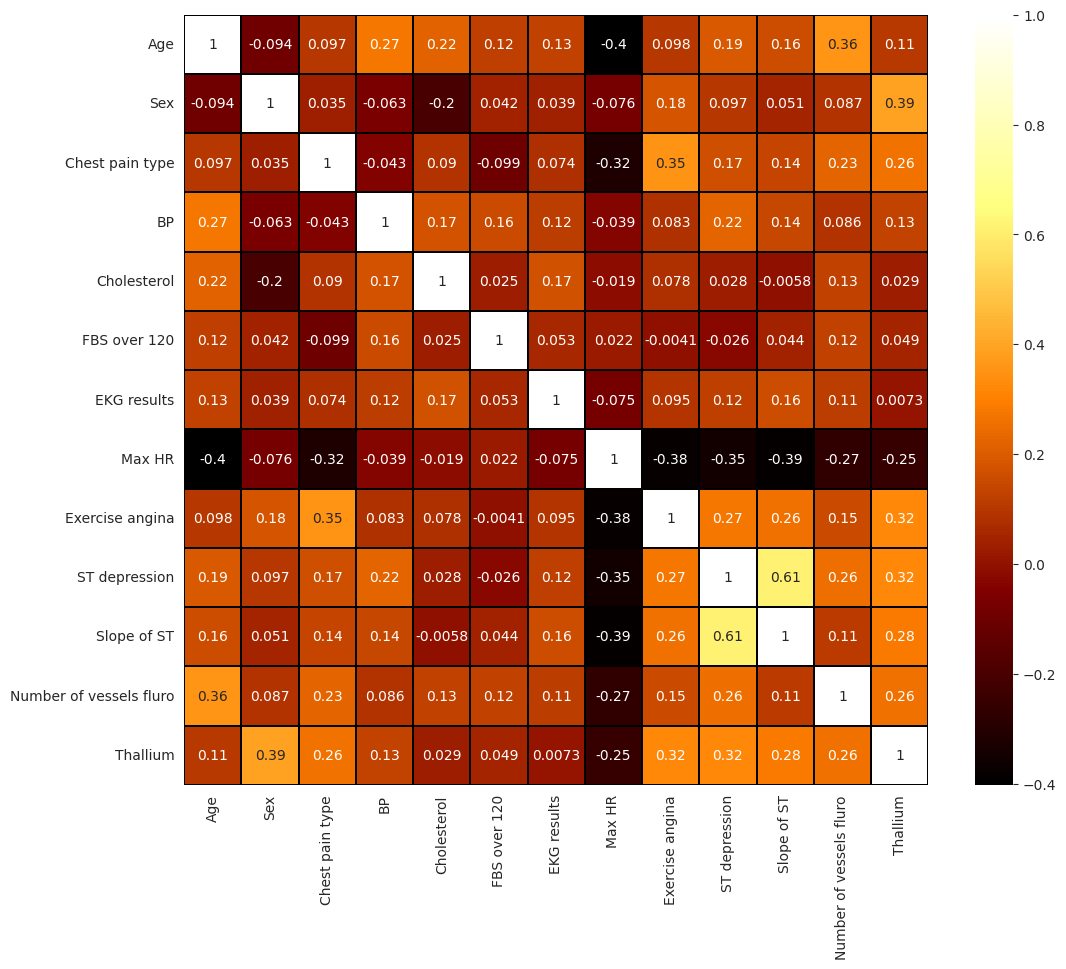
****

****

****

**Figure 3:** Visualization of the given input dataset

Heat Maps are graphical depictions of data. The primary determination of Heat Maps is to improve the visualization of the capacity of locations within a dataset and support directional addresses to areas on data visualizations. The following Figure 4 is for a multiple attributes representation of a heat map.



**Figure 4:** Heatmap of the Given input dataset

# 3.4 Model Implementation

# Model implementation with multiple libraries for starting our research with Python.

### **3.5 Feature Selection**

# Select a feature for accurate predictions from the given input data.

# Table 3: Selection of features

# 

### **3.6 Data split and scaling**

After selection of features data can be split and scaling of data(preprocessing).

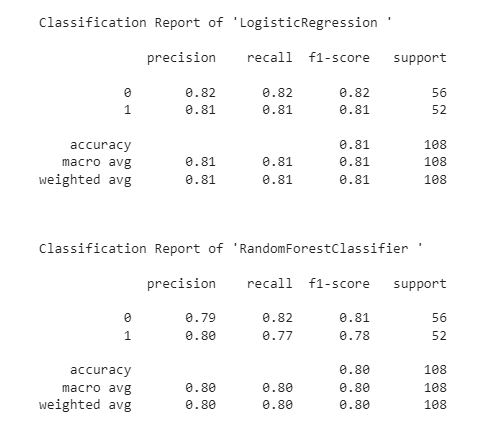


# 3.7 ML Model Selection and Model Prediction

Model selection and prediction are most important for the accuracy of research. It affects the performance.

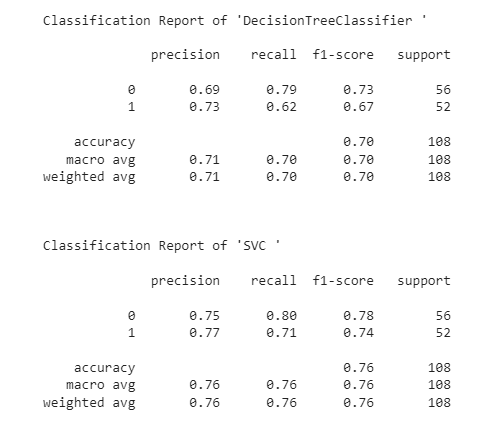
### **3.8 Model Implementing**

**Table 4:** Classification report of Logistic Regression

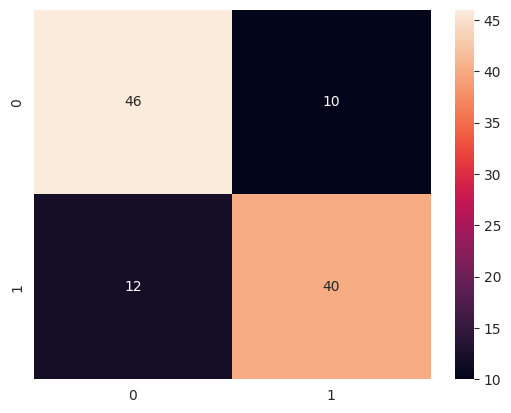


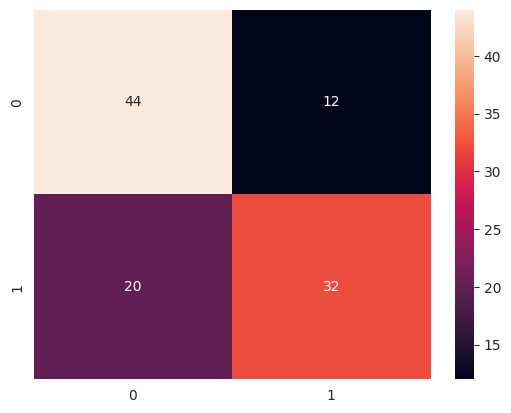
Model implementation purpose we calculate Precision, recall, F1 Score, and accuracy parameters for the prediction of results.

**Table 5:** Classification report of Decision Tree



A confusion matrix is a performance assessment tool in the artificial intelligence domain. Model implementation purpose we calculate Precision, recall, F1 Score, and accuracy parameters for the prediction of results.



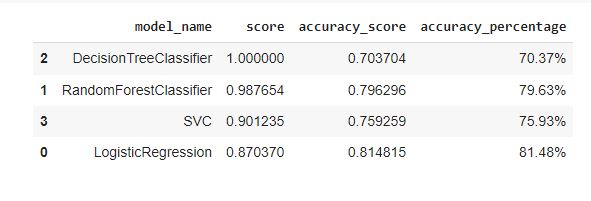




**Figure 5:** Confusion matrix of our models

While predicting, we can store the model’s score and prediction values in a newly generated data frame. The following Table 6 shows the different machine learning models' accuracy as the decision tree predicts 70.37%, the Random Forest tree is 79.63%, the Support Vector Machine is 75.93& and the Logistic Regression predicts 81.48%. If you observe this table, Logistic regression performs more accurately compared to other models.

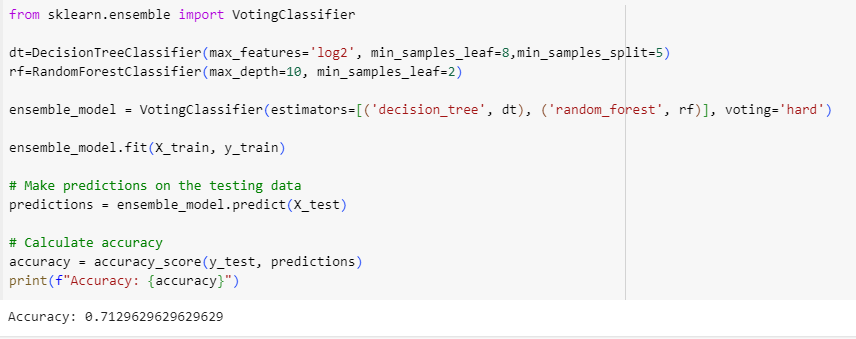
**Table 6:** Accuracy table



### **3.9 Hyper-tuning the ML Model**

For Hyper Tuning, we can use Grid Search CV to know the finest performing parameters.

## **3.10 Ensemble model (Decision Tree + Random Forest)**

****

## **3.11 Ensemble model (SVC + Logistic Regression)**

****

Our proposed ensemble model, decision tree, and Random Forest generate an accuracy rate is 71. 29% and Support Vector Machine and Logistic Regression accuracy rate is 76.85%. Finally, our research observes that if I use the different machine learning models individually, then the accuracy for the decision tree is 70.37%, the Random Forest tree is 79.63%, the Support Vector Machine is 75.93%, and Logistic Regression predicts 81.48%. But ensemble models of decision tree and Random Forest tree generate an accuracy rate is 71. 29%, and the Support Vector Machine and Logistic Regression accuracy rate is 76.85%.

In this research, except for the decision tree model remaining three models perform the same or have high performance compared to ensemble learning.

1. **CONCLUSION**

The health sector also faces problems with this disease in the world. A serious investigation avoids the issue in a short time. Optimal prediction generates accurate results for this disease. This type of investigation helps doctors for the identification of diseases and improve health. Machine learning techniques play a crucial role in this scenario. In health industry utilizes such type of engineering techniques for speedy identification and recovery. Our goal is to avoid such types of illnesses in people's lives. The major goal of this job effort extremely to forestall the incidence of coronary disease. Our research observes that if I use the different machine learning models individually, then the accuracy for the decision tree is 70.37%, the Random Forest tree is 79.63%, the Support Vector Machine is 75.93%, and the Logistic Regression predicts 81.48%. But ensemble models of decision tree and Random Forest tree generate an accuracy rate is 71. 29%, and the SVM and LR accuracy rate is 76.85%.

**Disclaimer (Artificial intelligence)**

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of manuscripts.

Option 2:

Author(s) hereby declare that generative AI technologies such as Large Language Models, etc have been used during writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

1.

2.

3.

**REFERENCES**

1. Arthy, S. (2018): A survey on heart disease prediction using data mining techniques. IJERCSE, 5(4), 280-287.
2. Australian Institute of Health and Welfare (AIHW). (2011). Cardiovascular Disease Series. AIHW, Canberra, Australia.
3. Ayon, S. I., Islam, M. M., & Hossain, M. R. (2020). Coronary artery heart disease prediction: A comparative study of computational intelligence techniques. Journal of the Institution of Electronics and Telecommunication Engineers, 1–20.
4. Bui, L., Horwich, T. B., & Fonarow, G. C. (2011). Epidemiology and risk profile of heart failure. Nature Reviews. Cardiology, 8(1), 30–41. doi:10.1038/nrcardio.2010.165 PMID:21060326.
5. Burse, K., Kirar, V. P. S., Burse, A., & Burse, R. (2019). Various preprocessing methods for neural network-based heart disease prediction. In Smart innovations in communication and computational sciences, (pp. 55–65). Springer. doi:10.1007/978-981-13-2414-7\_6
6. Dahiwade, D. (2019). Designing a disease prediction model using a machine learning approach. In the Third IEEE International Conference on Computing Methodologies and Communication (ICCMC). 10.1109/ ICCMC.2019.8819782
7. Dey, D., Slomka, P. J., Leeson, P., Comaniciu, D., Shrestha, S., Sengupta, P. P., & Marwick, T. H. (2019). Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. Journal of the American College of Cardiology, 73(11), 1317–1335.
8. Gavhane, A. (2018). Prediction of Heart Disease Using Machine Learning. Second International Conference on Electronics, Communication, and Aerospace Technology (ICECA), p. 1275–8.
9. B. Venkata Lakshmi, M.V. Shivsankar:” Heart Disease Diagnosis Using Predictive Data Mining”, TIFAC-CORE, Pervasive Computing Technologies, Velammal Engineering College, Chennai, India.
10. UCI Machine learning Repository <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>.
11. SantoshandSubhani Shaik,” Heart disease prediction with PCA and SRP”, International Journal of Engineering and Advanced Technology,” Volume-8, Issue-4, 2019.
12. Subhani shaik, Dr. Vijayalakshmi Kakulapati,” Prediction and Diagnosing of Heart Diseases using Data Mining and Artificial Intelligence based Algorithms”, ISBN13: 9781668457412, IGI Global Publication, Oct-2022.
13. KP Surya Teja, Vigneswara Reddy, and Subhani Shaik,” Flight Delay Prediction Using Machine Learning Algorithm XGBoost”, Journal of Advanced Research in Dynamical & Control Systems, Vol. 11, No. 5, 2019.
14. Dr. R. Vijaya Kumar Reddy, Dr. Shaik Subhani, Dr. G. Rajesh Chandra, Dr. B. Srinivasa Rao,” Breast Cancer Prediction using Classification Techniques”, International Journal of Emerging Trends in Engineering Research, Vol. 8, No.9,2020.
15. Mr. Sujan Reddy, Ms. Renu Sri, andSubhani Shaik,” Sentimental Analysis using Logistic Regression”, International Journal of Engineering Research and Applications (IJERA), Vol 11, Series-2, July 2021.
16. Ms. Mamatha, Srinivasa Datta and Subhani Shaik,” Fake Profile Identification using Machine Learning Algorithms”, International Journal of Engineering Research and Applications (IJERA), Vol 11, Series-2, July 2021.