Ensemble Machine Learning Approach to Heart Disease Classification

Rabbani Nur Kumoro
(21/472599/PA/20310)
Department of Computer Science and
Electronics, Universitas Gadjah Mada
Building C, 4th Floor North
Yogyakarta, Indonesia
rabbani.nur.kumoro@mail.ugm.ac.id

Abstract — **This proposed approach focuses on addressing the heart disease classification problem using ensemble machine learning techniques. The dataset used in this study, known as the Cleveland Dataset from the UCI Machine Learning Repository, consists of a training set with 14 attributes. To begin, a series of data preprocessing steps were carried out, followed by exploratory data analysis to gain valuable insights for feature engineering. The engineered features played a crucial role in the subsequent modeling process. In order to develop predictive models, ensemble methods such as Random Forest and XGBoost algorithms were employed. Additionally, a novel Stacking Ensemble method was proposed and implemented. A comparison of the performance of these models revealed that the Stacking Ensemble model achieved a higher recall score of 0.86%, while the Random Forest and XGBoost algorithms achieved lower scores of 0.84 and 0.82 respectively. These results emphasize the effectiveness of the proposed ensemble approach in accurately predicting heart disease and demonstrate its potential in addressing the heart disease classification problem.**

Keywords — Ensemble, Heart Disease, Machine Learning, Random Forest, Stacking, XGBoost

I. Introduction

Heart disease is a significant global health concern, accounting for a substantial number of deaths annually, with one-third of these occurring before the age of 70 [**1**]. The heart, being a vital organ, plays a crucial role in the overall well-being of individuals [**2**]. Heart disease encompasses various conditions characterized by abnormal heart function, often resulting from factors such as blood clots, arterial issues, and other related causes. Unhealthy habits, including high cholesterol, obesity, elevated triglyceride levels, and hypertension, contribute to an increased risk of heart disease [**1**].

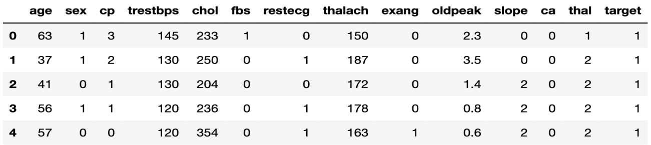
Heart disease involves several factors, including age, cholesterol levels, weight, height, gender, blood pressure, resting electrocardiogram (ECG) results, chest pain, smoking, obesity, and dietary habits [**3**]. Additionally, heart diseases fall under the broader category of cardiovascular diseases [**4**], which encompass complications of the blood vessels and heart, such as cerebrovascular disease, rheumatic heart disease, and other cardiac conditions [**5**]. Early prediction of heart disease and adopting a healthy lifestyle are crucial in preventing and managing this condition.

Within the field of medical science, machine learning techniques have been widely applied with ongoing efforts to optimize and improve their effectiveness [**6**]. Ensemble learning, a subfield of machine learning, has proven to be a promising approach to enhancing machine learning tasks. Ensemble classifiers combine multiple individual classifiers, often using majority voting and stacking, to improve predictive performance. The main objective of this project is to develop an ensemble model capable of determining whether a patient has heart disease based on their attributes.

Machine learning holds significant potential in the healthcare industry, enabling more accurate predictions for doctors in faster processing and analysis of medical data. Predictive analytics algorithms, powered by machine learning, can efficiently be trained on large datasets and perform in-depth analyses of numerous variables with minimal adjustments [**6**]. This report aims to leverage machine learning techniques, specifically ensemble methods, to enhance heart disease classification, contributing to improved patient care and outcomes in the field of healthcare.

II. Dataset

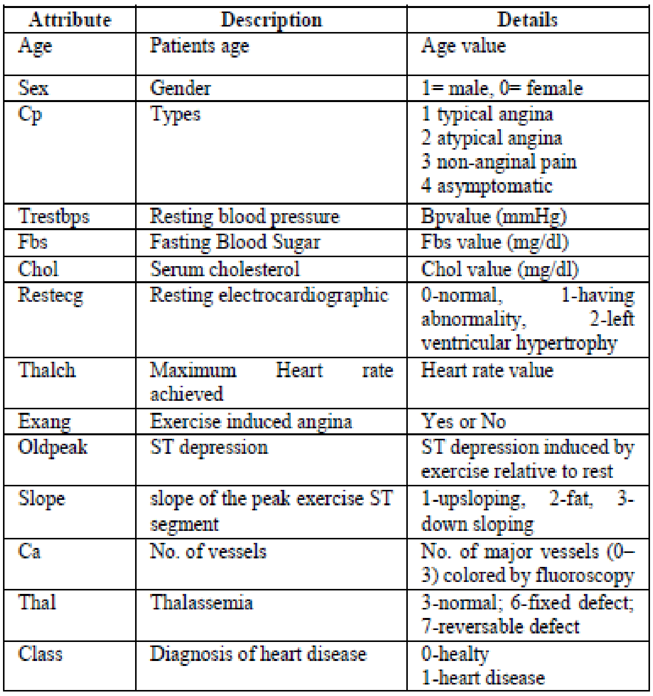
The dataset utilized in this problem is the Heart Disease Dataset obtained from the UCI Machine Learning Repository [**7**]. This dataset is a compilation of four different datasets, with the Cleveland dataset being the most commonly utilized. Specifically, the dataset comprises 297 instances and encompasses 14 features. Initially, the dataset consisted of 76 attributes; however, for the purposes of these studies, only a subset of 14 attributes has been employed. The dataset is designed in such a way that it contains one dependent variable, "Diagnosis," while the remaining 13 attributes serve as independent variables. These attributes include factors such as age, sex, chest pain type, blood pressure, serum cholesterol, and several others. As a result, the dataset exhibits a multivariate nature, enabling the exploration of relationships between multiple features and the diagnosis of heart disease. **Figure 1** provides an overview of the dataset, revealing the presence of two distinct types of values: categorical data and numerical data.



	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Figure 1. Sample Training Dataset

Figure 2 displays 14 variables in the dataset, each explained with its description and details.



Attribute	Description	Details
Age	Patients age	Age value
Sex	Gender	1= male, 0= female
Cp	Types	1 typical angina 2 atypical angina 3 non-anginal pain 4 asymptomatic
Trestbps	Resting blood pressure	Bpvalue (mmHg)
Fbs	Fasting Blood Sugar	Fbs value (mg/dl)
Chol	Serum cholesterol	Chol value (mg/dl)
Restecg	Resting electrocardiographic	0-normal, 1-having abnormality, 2-left ventricular hypertrophy
Thalch	Maximum Heart rate achieved	Heart rate value
Exang	Exercise induced angina	Yes or No
Oldpeak	ST depression	ST depression induced by exercise relative to rest
Slope	slope of the peak exercise ST segment	1-upsloping, 2-fat, 3-down sloping
Ca	No. of vessels	No. of major vessels (0–3) colored by fluoroscopy
Thal	Thalassemia	3-normal; 6-fixed defect; 7-reversable defect
Class	Diagnosis of heart disease	0-healty 1-heart disease

Figure 2. Dataset Description

III. Methodology

This section presents a methodology for developing an effective machine learning model. It includes stages such as dataset preparation, pre-processing, feature engineering, model selection, stacking, and evaluation. The methodology serves as a guide for implementing the proposed solution and is supported by a flowchart shown in **Figure 4** that illustrates the process.

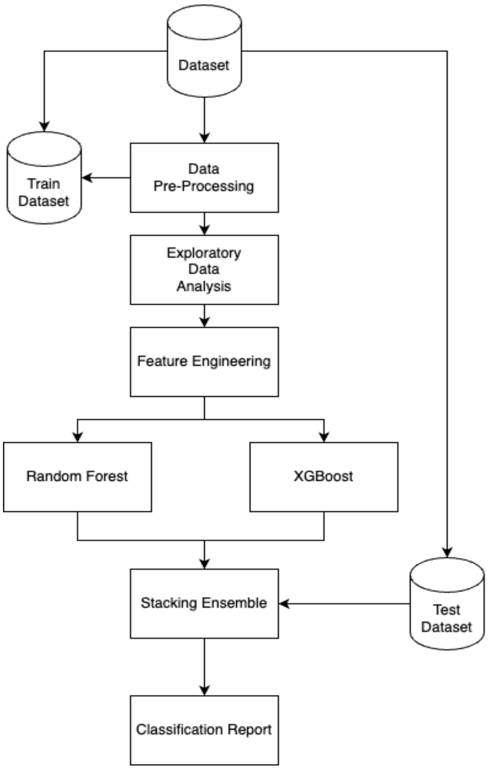
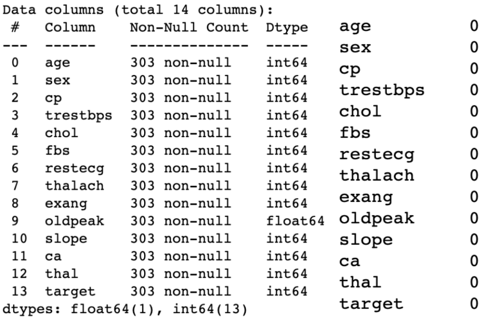


Figure 4. Flowchart Process

A. *Data Pre-Processing*

The subsequent step involves data preprocessing, which encompasses checking for missing values and skewness. **Figure 3** demonstrates that this dataset contains 0 null values, indicating the absence of missing data.

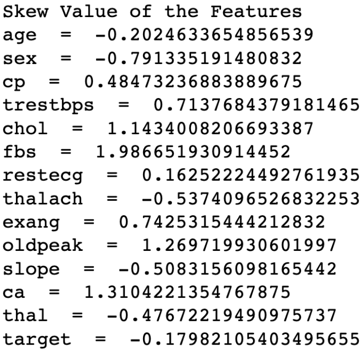


```
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   age       303 non-null    int64
 1   sex       303 non-null    int64
 2   cp        303 non-null    int64
 3   trestbps  303 non-null    int64
 4   chol      303 non-null    int64
 5   fbs       303 non-null    int64
 6   restecg   303 non-null    int64
 7   thalach   303 non-null    int64
 8   exang     303 non-null    int64
 9   oldpeak   303 non-null    float64
 10  slope     303 non-null    int64
 11  ca        303 non-null    int64
 12  thal      303 non-null    int64
 13  target    303 non-null    int64
dtypes: float64(1), int64(13)
```

```
age         0
sex         0
cp          0
trestbps    0
chol        0
fbs         0
restecg     0
thalach     0
exang       0
oldpeak     0
slope       0
ca          0
thal        0
target      0
```

Figure 3. Dataset Information

Nonetheless, there were certain features that exhibited noticeable right-skewness in the dataset, namely: chol, fbs, oldpeak, ca as shown in **Figure 4**. It is important to note that features in the dataset displayed a normal distribution, as indicated by skewness values of less than one.



```
Skew Value of the Features
age  =  -0.2024633654856539
sex  =  -0.791335191480832
cp  =  0.48473236883889675
trestbps  =  0.7137684379181465
chol  =  1.1434008206693387
fbs  =  1.986651930914452
restecg  =  0.16252224492761935
thalach  =  -0.5374096526832253
exang  =  0.7425315444212832
oldpeak  =  1.269719930601997
slope  =  -0.5083156098165442
ca  =  1.3104221354767875
thal  =  -0.47672219490975737
target  =  -0.17982105403495655
```

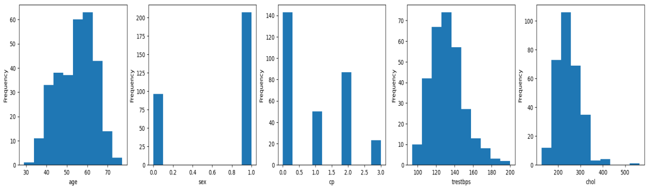
Figure 4. Skew Data

B. *Exploratory Data Analysis*

Exploratory Data Analysis (EDA) is a crucial step in understanding and analyzing the dataset before applying classification machine learning algorithms. Several EDA techniques for classification tasks including histograms, grouped bar charts, and box plots are going to be explained further below.

To better understand the distribution of each feature, a Univariate Analysis using histograms from **Figure 5** was performed on all features in the dataset. This approach was chosen because all features had been encoded as numeric values, eliminating the need for categorical encoding during the feature engineering stage.

By utilizing histograms, valuable insights were gained into the frequency and distribution patterns of the variables. This analysis not only saved time typically spent on categorical encoding but also provided a comprehensive understanding of the dataset's features.



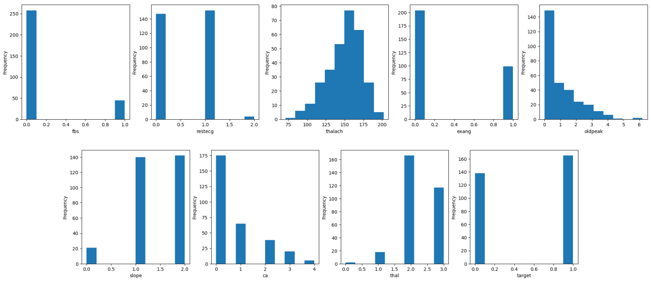


Figure 5. Histograms

A grouped bar chart was created in **Figure 6** to illustrate the influence of categorical values on the target value. This visualization provided a clear representation of how different categorical values contributed to the prediction of the target. The grouped bar chart served as a valuable tool to analyze the relationship between categorical variables and the target, enabling insights into their predictive power in the context of the dataset.

For instance, when comparing the target distribution between different values of the 'sex' feature, noticeable variations were observed, suggesting a significant correlation between this feature and the target. Conversely, if the target distribution remained consistent across various categorical features, it indicated a lack of correlation between them.

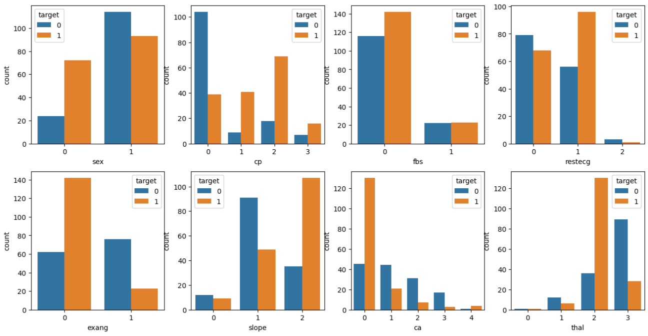
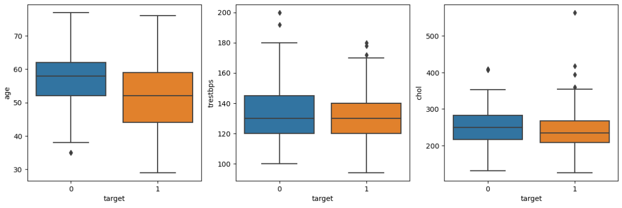


Figure 6. Grouped Bar Chart

The box plot was employed to analyze the relationship between numerical features and the target variable by illustrating how the values of these features varied across different target groups. **Figure 7** presents the box plot visualization, showcasing the variations in numerical feature values with respect to the target.

Notably, the box plot for "oldpeak" feature demonstrated a pronounced disparity between the target values of 0 and 1, indicating its significance as a predictor. Conversely, the box plots for "trestbps" and "chol" exhibited less distinct differences, suggesting a more similar distribution across the target groups for these features. By utilizing box plots, we gained valuable insights into the impact and predictive power of numerical features in relation to the target variable.



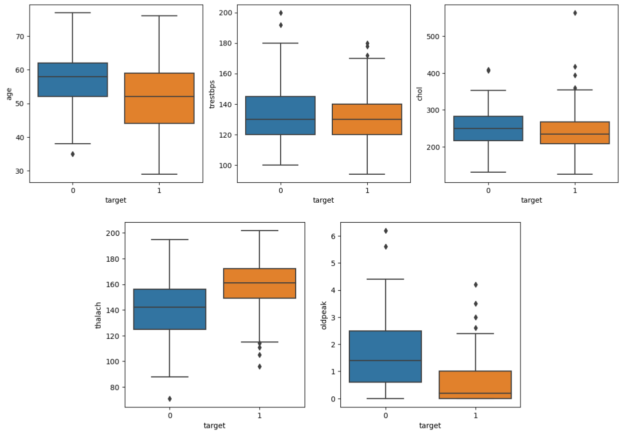
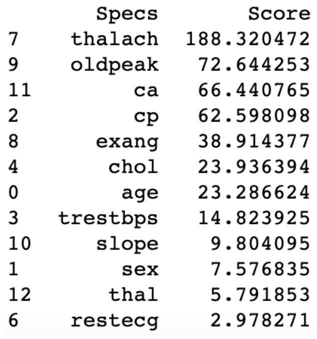


Figure 7. Box Plot

C. Feature Engineering

Univariate Selection, Feature Importance, and Correlation Matrix are used for the Feature Engineering phase. Univariate Selection is a statistical test used to select features with the strongest relationship to the target variable. Feature Importance ranks features based on their relevance and significance in predicting the target variable. The Correlation Matrix examines the relationships between features, revealing the strength and direction of linear relationships. These techniques provide valuable insights into the individual and collective contributions of features, aiding in informed decisions during feature selection and modeling.

Univariate Selection is a statistical test utilized to identify specific features that exhibit the strongest relationship with the performance variable. This test helps in selecting the features that have the best predictive power or influence on the target variable. The SelectKBest class, which utilizes a suite of different statistical tests, including the chi-squared test for non-negative features, is employed in Univariate Selection to identify the features that exhibit the strongest relationship with the performance variable. In the following **Figure 8**, the chi-squared test is utilized to select the top 10 features that demonstrate the highest association with the performance variable.



```
       Specs       Score
7    thalach  188.320472
9    oldpeak   72.644253
11        ca   66.440765
2         cp   62.598098
8      exang   38.914377
4       chol   23.936394
0        age   23.286624
3   trestbps   14.823925
10     slope    9.804095
1        sex    7.576835
12      thal    5.791853
6    restecg    2.978271
```

Figure 8. Univariate Score

Utilizing Feature Importance will enable us to assess the significance of each feature in the dataset. It assigns a score to each feature based on its contribution to the model's performance, with higher scores indicating greater importance. This analysis is facilitated by the built-in class called Feature Importance, commonly employed with Tree-Based Classifiers. The Extra Tree Classifier will be utilized to extract the top 10 features from the dataset as shown in **Figure 8**.

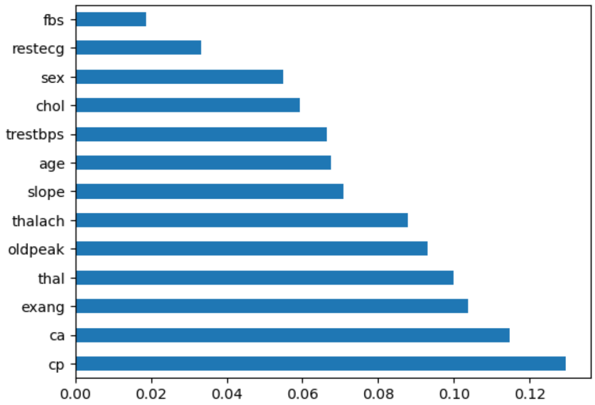


Figure 8. Graph Plot

The Correlation Matrix provides insights into the relationships between features and the target variable. It reveals whether the correlation is positive, indicating that an increase in one feature value corresponds to an increase in the target variable, or negative, indicating that an increase in one feature value corresponds to a decrease in the target variable. By utilizing a heatmap in **Figure 9**, it becomes easier to identify the features that are most relevant to the target variable.

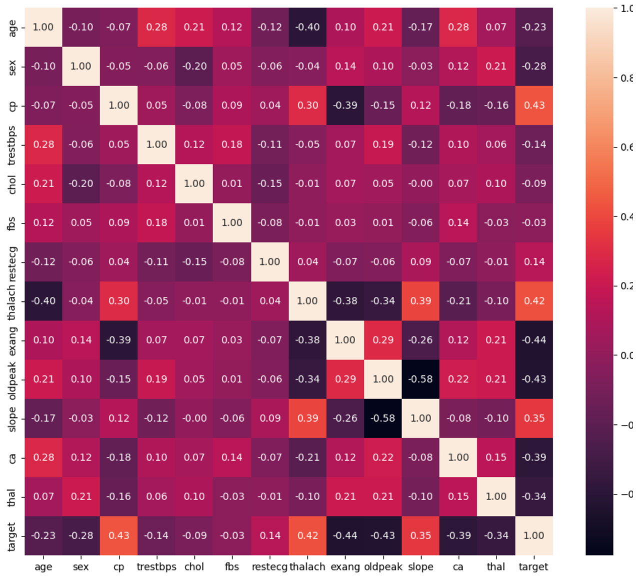


Figure 9. Heatmap

D. Modeling

In the modeling phase, the power of the Stacking Ensemble, a robust ensemble-based approach, will be utilized. It involves employing a meta-learner to train a model by merging the predictions of multiple weak learners. The base models are trained using the entire dataset, and their outputs serve as inputs for the meta-learner. By extracting features from the base models, the meta-learner maximizes performance and enhances the effectiveness of the model [**8**].

By leveraging the collective intelligence of multiple models, an ensemble technique aims to enhance performance compared to individual models [6]. This approach combines the strengths of different algorithms and mitigates their weaknesses, resulting in a more robust and accurate predictive model. In our case, we will utilize two prominent ensemble methods, namely Random Forest and eXtreme Gradient Boosting (XGBoost), and employ stacking to combine their predictions. Stacking further improves the overall predictive power by leveraging the complementary strengths of these algorithms.

The model is trained on the training set and evaluated using a separate testing set. In this particular approach, the dataset is divided into two subsets: a training set, which constitutes 67% of the data, and a testing set, which accounts for 33% of the data. This division ensures an unbiased assessment of the model's performance on unseen data, providing a reliable estimation of its effectiveness.

1. Random Forest

The Random Forest (RF) algorithm is an ensemble learning technique that combines the strength of multiple decision trees. Its purpose, as developed by [**9**], is to achieve high accuracy in various learning tasks, and it has demonstrated effectiveness across multiple datasets [**10**]. RF creates a collection of decision trees by randomly selecting subsets of features for splitting, using a technique known as bootstrapping with replacement [**11**]. This incorporation of randomness enhances the algorithm's ability to handle missing data while preserving accuracy. By employing bagging, RF selects the optimal split at each node from a random subset of predictors, which helps prevent overfitting and ensures robustness [**12**].

The Random Forest algorithm creates a forest of randomly constructed decision trees to make predictions as depicted in **Figure 19**. Unlike regular decision trees that select the best split from all variables at each node, Random Forest introduces an extra level of randomness [**13**]. By combining the results of individual trees, Random Forest leverages the diverse perspectives provided by different subsets of features and training data. This approach enhances accuracy and generalization compared to relying on a single decision tree.

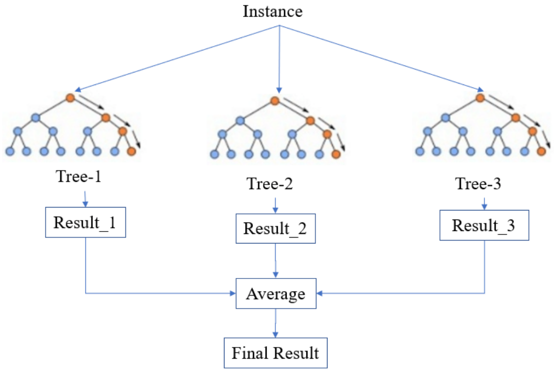


Figure 19. Random Forest Illustration

To mitigate overfitting and enhance robustness, decision trees in Random Forests randomly select a subset of features at each node [**9**]. The predictions of all decision trees are then aggregated, typically through averaging or majority voting, to obtain the final prediction. By employing bootstrapping, where training data samples are randomly chosen with replacement, Random Forest constructs a collection of randomized decisions and makes the final decision based on the majority vote. These techniques contribute to the effectiveness of Random Forest as a powerful ensemble learning algorithm for predictive modeling.

2. eXtreme Gradient Boosting

XGBoost is an ensemble decision tree-based algorithm that operates on the principles of the gradient boosting framework [**14**]. It is a highly accurate and scalable algorithm commonly used for classification tasks. Unlike Random Forest, each tree model in XGBoost aims to minimize the residual error from the previous tree model. While traditional gradient-boosted decision trees use only the first derivative of error information, XGBoost employs the second-order Taylor expansion of the loss function, utilizing both the first and second derivatives. This process can be visualized in **Figure 20**. Notably, the residual from the first decision tree is passed to the second decision tree, and this iterative process continues, progressively reducing the residual error [**15**].

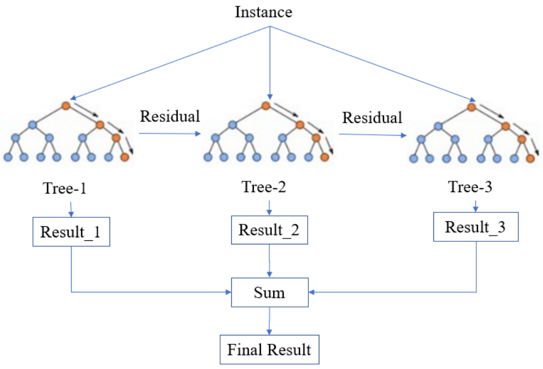


Figure 20. eXtreme Gradient Boosting Illustration

One distinguishing feature of XGBoost is its support for customized loss functions, setting it apart from other gradient-boosting algorithms. XGBoost incorporates a regularization model within its loss function, effectively controlling overfitting [**16**]. This regularization term strikes a balance between fitting the training data well and avoiding overfitting, resulting in a more robust and accurate model. By integrating this regularization term, XGBoost mitigates overfitting, leading to improved generalization and predictive performance [**17**].

XGBoost offers several advantages that differentiate it from other machine learning algorithms. It requires minimal feature engineering since it can handle tasks like data normalization and feature scaling. Additionally, XGBoost can handle missing values, providing flexibility in data preprocessing. Furthermore, XGBoost provides feature importance rankings, facilitating the understanding of input features and aiding in feature selection.

3. Stacking Ensemble

Stacking Ensemble is a powerful ensemble learning technique that combines the predictions of multiple base models to create a more accurate and robust prediction model [**18**]. It works by training several diverse base models on the same dataset and using their individual predictions as input features for a meta-learner. The meta-learner then learns to combine the predictions from the base models and make the final prediction [**19**].

The proposed approach to this problem is to stack Random Forest and XGBoost, the combination of these two algorithms can further enhance predictive performance. RF is known for its ability to handle complex relationships and capture feature interactions effectively [**20**], while XGBoost excels in capturing gradient-based patterns and optimizing the prediction model [**13**]. By stacking RF and XGBoost, it can leverage the complementary strengths of both algorithms and create a more robust and accurate prediction model. The base models, RF and XGBoost, will each contribute their unique perspectives and predictions, which will be combined by the meta-learner to make the final prediction [**21**].

The combination of RF and XGBoost in the stacking ensemble approach provides a synergistic effect, where the weaknesses of one algorithm can be compensated by the strengths of the other. Additionally, the stacking technique can help in mitigating the bias and variance issues often associated with individual models. By aggregating the predictions of RF and XGBoost, the meta-learner can learn to make more accurate predictions by considering the collective knowledge of both algorithms.

E. Performance Metrics

The evaluation metrics serve as quantitative measures to evaluate different aspects of the model's performance. The chosen evaluation metrics are carefully selected to assess the model's performance in accurately identifying positive and negative instances. These metrics provide a comprehensive understanding of the model's accuracy and effectiveness [**21**]. **Figure 22** illustrates the calculation methods for the four key metrics used.

$$\text{precision} = \frac{\text{tp}}{\text{tp} + \text{fp}}$$

$$\text{recall} = \frac{\text{tp}}{\text{tp} + \text{fn}}$$

$$\text{accuracy} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{tn} + \text{fp} + \text{fn}}$$

$$\text{F}_1 \text{ score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

Figure 22. Metrics Formula

The F1 score offers a balanced evaluation of the model's performance by taking into account both precision and recall. By considering both false positives (fp) and false negatives (fn), it provides a comprehensive assessment of the model's effectiveness, especially in situations with imbalanced class distribution.

Accuracy is a widely used metric that calculates the percentage of correctly classified instances. While it provides an overall measure of correctness, it may not be suitable for datasets with imbalanced classes, as it can be skewed by the dominant class.

Precision focuses on the accuracy of positive predictions among all predicted positive instances. It is particularly valuable in scenarios where minimizing false positives is critical, such as heart disease diagnosis, where the cost of false positives is high [**22**].

When it comes to diagnosing heart disease, the use of recall as an evaluation metric is pivotal [**22**]. Recall is chosen as a crucial metric because it emphasizes the importance of correctly identifying true positive cases of heart disease. Missing a positive case can have severe consequences, such as delayed treatment and failure to provide necessary interventions. Prioritizing recall aims to maximize the identification of individuals with heart disease, ensuring that a larger proportion of true positive cases are accurately classified [**23**].

In the context of disease classification, false negatives (fn) can have more significant implications than false positives (fp). In the case of heart disease problems, false negatives can result in undetected conditions and potential health risks. Therefore, a high recall rate is desirable to minimize false negatives and increase the detection rate of heart disease cases [**23**]. Focusing on recall reduces the chances of missing positive cases and improves the effectiveness of heart disease diagnosis.

IV. Result and Analysis

The outcomes of the model were obtained by applying it to the test dataset, and a comprehensive analysis of the results was conducted. The evaluation process involved assessing the model's performance using multiple metrics, including accuracy, precision, recall, and F1 score. These metrics were utilized to evaluate the model's predictive abilities and provide a thorough assessment of its performance.

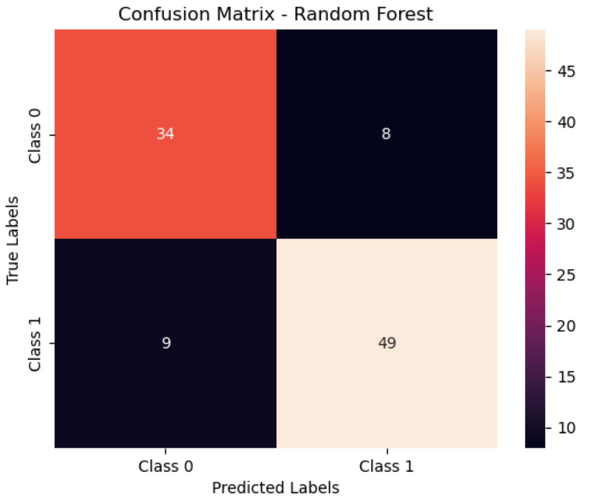


Figure 22. Random Forest Confusion Matrix

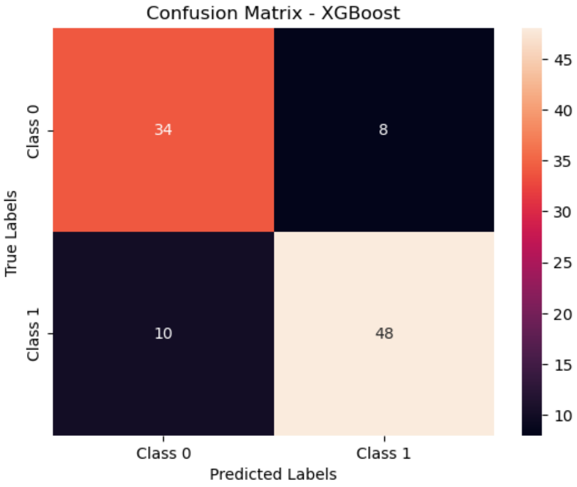


Figure 23. eXtreme Gradient Boosting Confusion Matrix

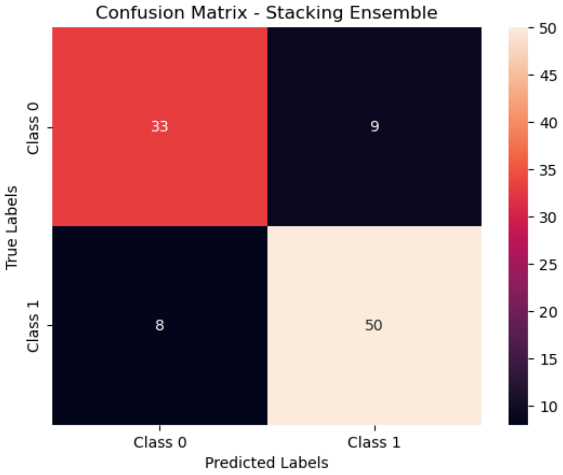


Figure 24. Stacking Ensemble Confusion Matrix

Figure 22 and **Figure 23** display the confusion matrices for the Random Forest (RF) model and XGBoost model. The y-axis represents the model's predictions, where 0 corresponds to a healthy patient and 1 indicates a patient with heart disease. The x-axis represents the actual results, with 0 indicating a healthy patient and 1 indicating a patient with heart disease.

In the case of the RF model, it accurately predicted 34 instances of healthy patients and correctly identified 49 patients with heart disease. On the other hand, the XGBoost model correctly predicted 34 instances of healthy patients and 48 patients diagnosed with heart disease.

By examining the confusion matrices, it is evident that the RF model performs better in terms of predicting both true negatives (healthy patients) and true positives (patients with heart disease) compared to XGBoost. This suggests that the RF model demonstrates stronger predictive capabilities for classifying both healthy and diseased individuals in the dataset.

Figure 24 presents the confusion matrix of the proposed Stacking Ensemble model. The matrix shows the model's predictions and the actual results for both healthy patients and patients with heart disease. The model accurately predicted 33 instances of healthy patients and correctly identified 50 patients with heart disease. This indicates that the Stacking Ensemble model demonstrated successful classification of both healthy and diseased individuals. It performed well in terms of accurately identifying true negatives (healthy patients) and true positives (patients with heart disease).

Table 2. Model Comparison

Model / Metrics	*Accuracy*	*Recall*	*Precision*	*F1-Score*
Random Forest	**83.0%**	0.84	**0.85**	**0.85**
Stacking Ensemble	**83.0%**	**0.86**	0.84	**0.85**
XGBoost	82.0%	0.82	**0.85**	0.84

Table 2 presents the evaluation of three models: Random Forest, XGBoost, and the proposed Stacking Ensemble model. The performance of these models is measured using various metrics, including accuracy, recall, precision, and F1-score. The stacking ensemble model achieved the highest recall score of 0.86. This indicates that the stacking model outperformed both Random Forest and XGBoost in correctly identifying positive cases of heart disease.

Although the accuracy scores were similar among the models about around 83.0%, the stacking ensemble model demonstrated superior performance in recall compared to the other models. This suggests that the stacking ensemble approach, which combines the predictions of multiple base models, enhances the model's ability to identify individuals with heart disease correctly.

Additionally, the stacking ensemble model achieved competitive precision and F1-score, with a precision of 0.84 and an F1-score of 0.85. These metrics indicate the model's ability to minimize false positives and maintain a balance between precision and recall.

V. Conclusion

Based on the evaluation of the proposed models, the Stacking Ensemble model demonstrated superior performance in predicting heart disease compared to the Random Forest and XGBoost models. The model achieved a high recall score of 0.86, indicating its ability to correctly identify a significant proportion of positive cases. Additionally, the model exhibited an accuracy of 83%, precision of 0.84, and F1-Score of 0.85, further supporting its effectiveness in heart disease diagnosis.

However, it is important to note that this project solely relied on machine learning models as per the project requirements, which may have limited the performance results. To achieve even better results, it is necessary to develop a more robust model and employ advanced feature engineering techniques. Additionally, future work could involve the integration of a broader range of machine learning algorithms and explore the potential of deep learning approaches to further enhance performance outcomes. In conclusion, the proposed stacking ensemble model has shown promising results in accurately diagnosing heart disease.

References

[1] "Cardiovascular diseases," World Health Organization, https://www.who.int/health-topics/cardiovascular-diseases/ (accessed Jun. 19, 2023).

[2] C. S.Dangare and S. S. Apte, "Improved study of heart disease prediction system using data mining classification techniques," International Journal of Computer Applications, vol. 47, no. 10, pp. 44–48, 2012. doi:10.5120/7228-0076

[3] M. T., D. Mukherji, N. Padalia, and A. Naidu, "A heart disease prediction model using SVM-decision trees-logistic regression (SDL)," International Journal of Computer Applications, vol. 68, no. 16, pp. 11–15, 2013. doi:10.5120/11662-7250

[4] M. Elhoseny et al., "A new multi-agent feature wrapper machine learning approach for heart disease diagnosis," Computers, Materials &amp; Continua, vol. 67, no. 1, pp. 51–71, 2021. doi:10.32604/cmc.2021.012632

[5] H. Rahid et al., "Heart Disease Prediction using Data Mining Classification Algorithms," Journal of Cardiovascular Disease Research, vol. 12, no. 03, 2021.

[6] N. S. Chandra Reddy, S. Shue Nee, L. Zhi Min, and C. Xin Ying, "Classification and feature selection approaches by Machine Learning Techniques: Heart Disease Prediction," International Journal of Innovative Computing, vol. 9, no. 1, 2019. doi:10.11113/ijic.v9n1.210

[7] A. Janosi, W. Steinbrunn, M. Pfisterer, and R. Detrano, "Heart Disease," UCI Machine Learning Repository, 1988. doi:10.24432/C52P4X

[8] A. Ashfaq et al., "Multi-model ensemble based approach for heart disease diagnosis," 2022 International Conference on Recent Advances in Electrical Engineering & Computer Sciences (RAEE & CS), 2022. doi:10.1109/raeecs56511.2022.9954490

[9] Breiman, L. "Random Forests". Machine Learning, 45, 5-32. 2001.

[10] S. Janitza and R. Hornung, "On the overestimation of Random Forest's out-of-bag error," PLOS ONE, vol. 13, no. 8, 2018. doi:10.1371/journal.pone.0201904

[11] R. Díaz-Uriarte and S. Alvarez de Andrés, "Gene selection and classification of microarray data using Random Forest," BMC Bioinformatics, vol. 7, no. 1, 2006. doi:10.1186/1471-2105-7-3

[12] R. Punnoose and P. Ajit, "Prediction of employee turnover in organizations using machine learning algorithms," International Journal of Advanced Research in Artificial Intelligence, vol. 5, no. 9, 2016. doi:10.14569/ijarai.2016.050904

[13] A. Liaw and M. Wiener, "Classification and Regression by randomForest," Northwestern University, vol. 2, no. 3, Dec. 2002.

[14] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," The Annals of Statistics, vol. 29, no. 5, 2001. doi:10.1214/aos/1013203451

[15] W. Wang, G. Chakraborty, and B. Chakraborty, "Predicting the risk of chronic kidney disease (CKD) using machine learning algorithm," Applied Sciences, vol. 11, no. 1, p. 202, 2020. doi:10.3390/app11010202

[16] R. Punnoose and P. Ajit, "Prediction of employee turnover in organizations using machine learning algorithms," International Journal of Advanced Research in Artificial Intelligence, vol. 5, no. 9, 2016. doi:10.14569/ijarai.2016.050904

[17] Y. Li and W. Chen, "A comparative performance assessment of ensemble learning for credit scoring," Mathematics, vol. 8, no. 10, p. 1756, 2020. doi:10.3390/math8101756

[18] D. H. Wolpert, "Stacked generalization," Neural Networks, vol. 5, no. 2, pp. 241–259, 1992. doi:10.1016/s0893-6080(05)80023-1

[19] J. B. Thomas, S. K. V., S. M. Sulthan, and A. Al-Jumaily, "Deep feature meta-learners ensemble models for COVID-19 CT Scan Classification," Electronics, vol. 12, no. 3, p. 684, 2023. doi:10.3390/electronics12030684

[20] A. U. Berliana and A. Bustamam, “Implementation of stacking ensemble learning for classification of COVID-19 using image dataset CT scan and Lung X-ray,” 2020 3rd International Conference on Information and Communications Technology (ICOIACT), 2020. doi:10.1109/icoiact50329.2020.9332112

[21] T. Velmurugan and J. Dhinakaran, “A novel ensemble stacking learning algorithm for parkinson’s disease prediction,” Mathematical Problems in Engineering, vol. 2022, pp. 1–10, 2022. doi:10.1155/2022/9209656

[22] S. A. Hicks et al., “On evaluation metrics for medical applications of Artificial Intelligence,” Scientific Reports, vol. 12, no. 1, 2022. doi:10.1038/s41598-022-09954-8

[23] M. M. Ahsan, S. A. Luna, and Z. Siddique, “Machine-learning-based disease diagnosis: A comprehensive review,” Healthcare, vol. 10, no. 3, p. 541, 2022. doi:10.3390/healthcare10030541