BRACT's Vishwakarma Institute of Information Technology

PROJECT REPORT OF

DATA SCIENCE & MACHINE LEARNING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
(ARTIFICIAL INTELLIGENCE)

Sr No.	Name	Email Id
1	Uday Hese	uday.22210639@viit.ac.in
2	Harsh Mane	harsh.22211525@viit.ac.in
3	Nimish Marathe	nimish.22211391@viit.ac.in
4	Siddhesh Mundhe	siddhesh.22210559@viit.ac.in

Under Guidance of

Dr. Anuradha Yenkikar

HEART DISEASE PREDICTION

- **PROBLEM STATEMENT:**

Many people are unaware of their risk of heart disease until it becomes a significant health concern. Visiting doctors for regular check-ups and heart disease screenings can be costly, especially for individuals with limited financial resources. This leads to a lack of early detection and preventive measures, resulting in higher healthcare expenses due to advanced treatments and hospitalizations. Therefore, there is a critical need for an affordable and accurate heart disease prediction system that can identify individuals at risk early on, allowing for timely interventions and reducing long-term medical expenses.

- **Liabrary:**

    sklearn.model_selection, sklearn.preprocessing, sklearn.tree, sklearn.ensemble, sklearn.metrics, sklearn.linear_model,sklearn.base, sklearn.neighbors, sklearn.neural_network.

- **Introduction:**

    In today's schools, effective management of computer laboratories is essential to support the diverse needs of students and teachers.
    However, bookkeeping management
    systems often lack the flexibility, accessibility, and instant tracking capabilities nee ded tobe effective. Therefore, a new credit system needs to be used to solve these p roblems and improve the general management of computerized examination centers.

- **Theory:**

    Adaptive Boosting Ensemble:
    AdaBoost algorithm, short for Adaptive Boosting, is a Boosting technique used as an Ensemble Method in Machine Learning. It is called Adaptive Boosting as the weights are re-assigned to each instance, with higher weights assigned to

incorrectly classified instances. What this algorithm does is that it builds a model and gives equal weights to all the data points. It then assigns higher weights to points that are wrongly classified. Now all the points with higher weights are given more importance in the next model. It will keep training models until and unless a lower error is received.

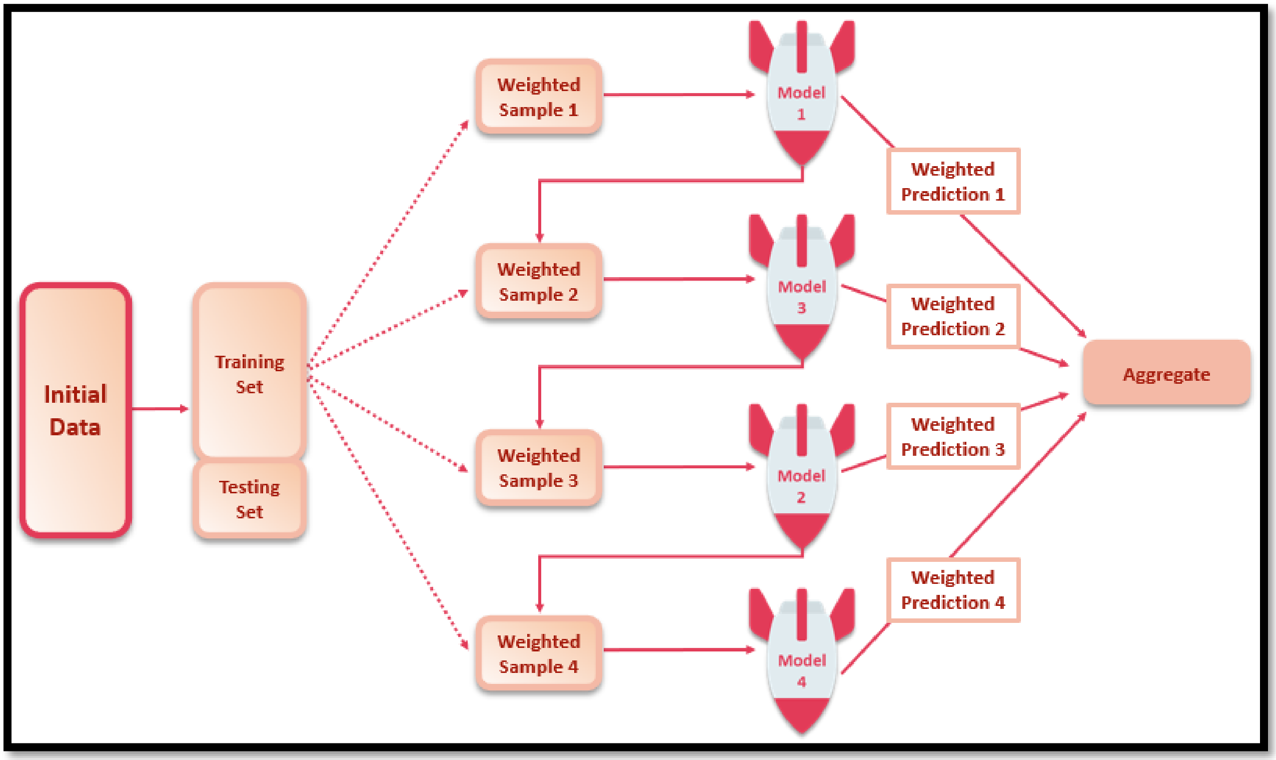
BAGGING Ensemble: Bagging, also known as Bootstrap aggregating, is an ensemble learning technique that helps to improve the performance and accuracy of machine learning algorithms. It is used to deal with bias-variance trade-offs and reduces the variance of a prediction model. Bagging avoids overfitting of data and is used for both regression and classification models, specifically for decision tree algorithms. It is a homogeneous weak learners' model that learns from each other independently in parallel and combines them for determining the model average.

STACKING Ensemble Technique: Stacking (sometimes called *Stacked Generalization*) is a different paradigm. The point of stacking is to explore a space of different models for the same problem. The idea is that you can attack a learning problem with different types of models which are capable to learn some part of the problem, but not the whole space of the problem. So, you can build multiple different learners and you use them to build an intermediate prediction, one prediction for each learned model. Then you add a new model which learns from the intermediate predictions the same target.
This final model is said to be stacked on the top of the others, hence the name. Thus, you might improve your overall performance, and often you end up with a model which is better than any individual intermediate model. Notice however, that it does not give you any guarantee, as is often the case with any machine learning technique.

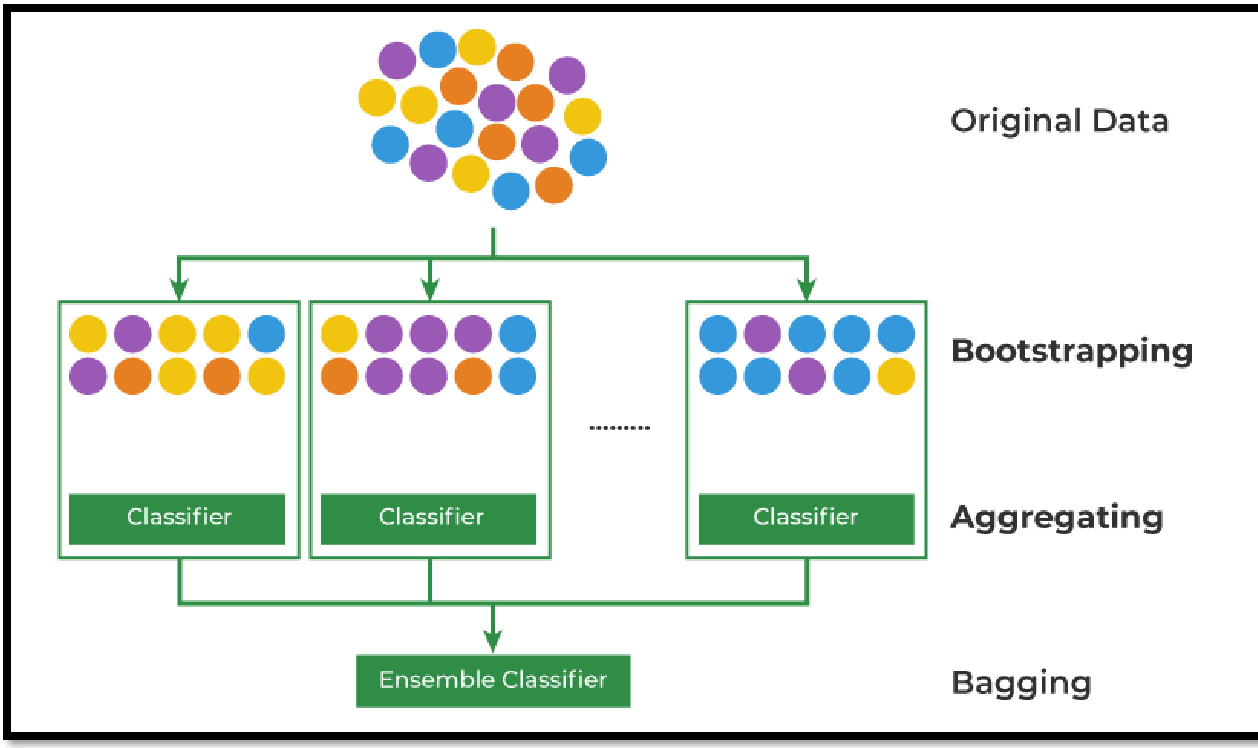
Random Forest Ensemble: Random Forest algorithm is a powerful tree learning technique in Machine Learning. It works by creating a number of Decision Trees during the training phase. Each tree is constructed using a random subset of the data set to measure a random subset of features in each partition. This randomness introduces variability among individual trees, reducing the risk of overfitting and improving overall prediction performance. In prediction, the algorithm aggregates the results of all trees, either by voting (for classification tasks) or by averaging (for regression tasks) This collaborative decision-making process, supported by multiple trees with their insights, provides an example stable and precise results. Random forests are widely used for classification and regression functions, which are known for their ability to handle complex data, reduce overfitting, and provide reliable forecasts in different environments.

<u>Working Algorithm/Diagram:</u>

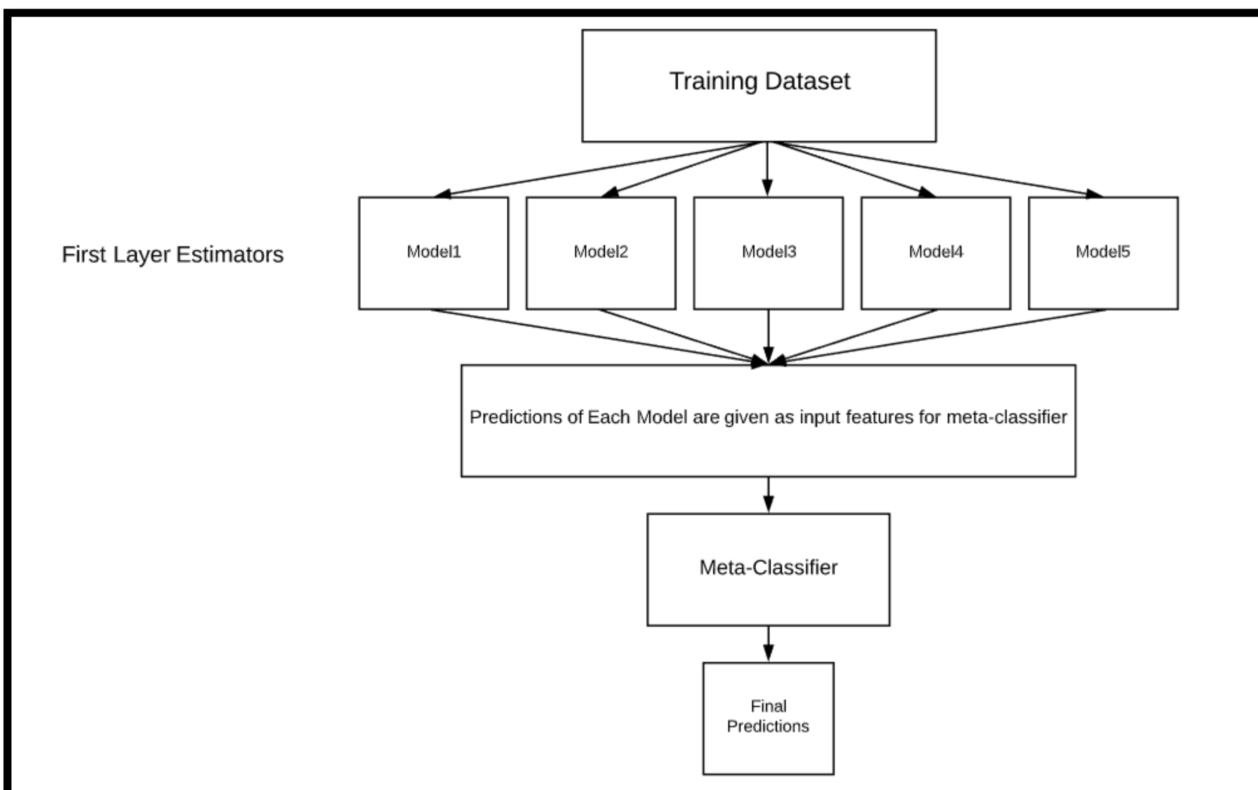
Adaptive Boosting Algorithm:



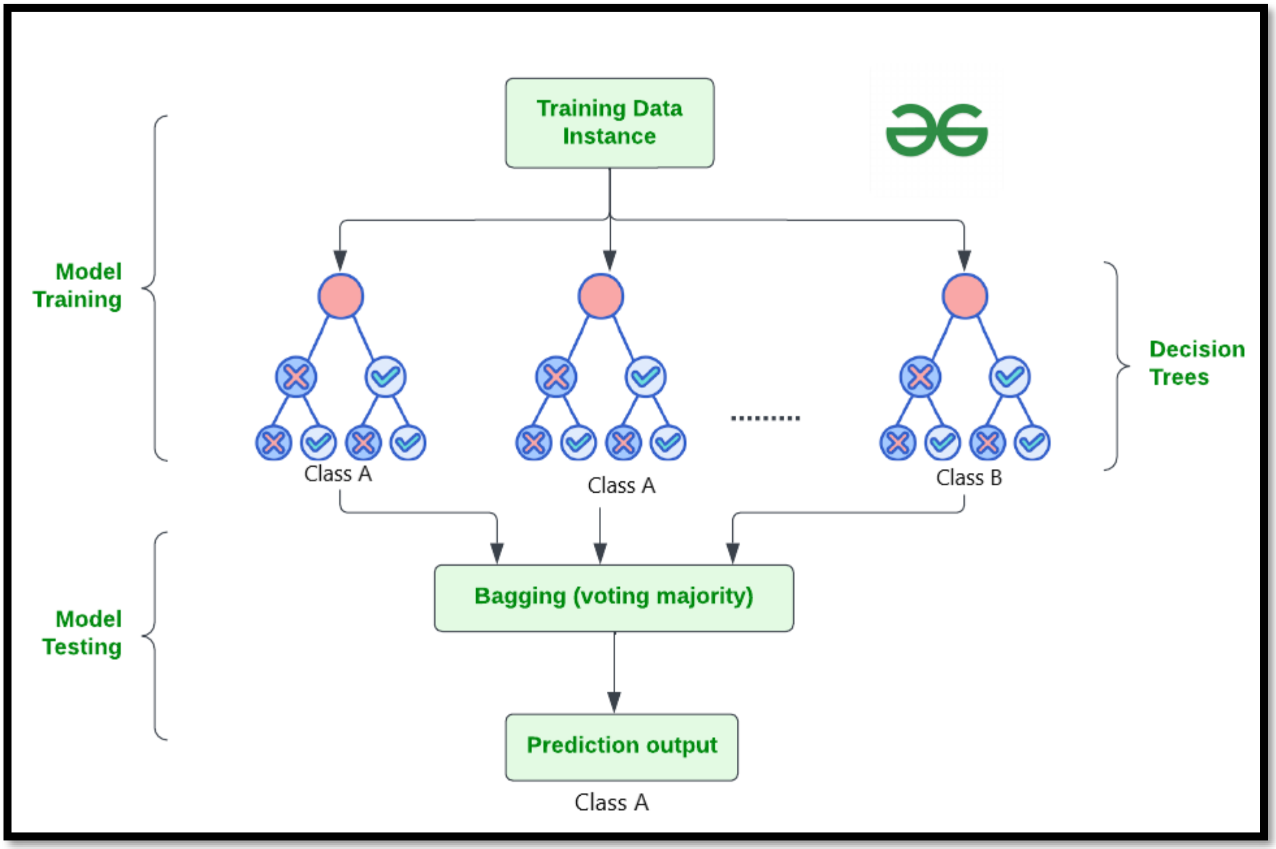
BAGGING Ensemble Technique:



STACKING Ensemble Technique:



Random Forest Ensemble Technique:



- ## <u>Conclusion:</u>

The heart disease prediction project employed a diverse range of ensemble techniques, including adaptive boosting, bagging, and random forest, to develop robust predictive models. Adaptive boosting (AdaBoost) effectively adjusted the weights of misclassified samples during training, improving the overall model's accuracy and generalization. Bagging, through the use of multiple base estimators trained on random subsets of the data, reduced variance and enhanced model stability, leading to reliable predictions. Random forest, leveraging an ensemble of decision trees with randomized feature selection, provided high accuracy and resilience against overfitting, making it a valuable addition to the predictive modeling arsenal. Collectively, these ensemble techniques demonstrated their efficacy in heart disease prediction by combining the strengths of individual models, achieving superior performance, and contributing to more accurate risk assessment and early intervention strategies in cardiovascular health management.