HEART DISEASE DETECTION USING MACHINE LEARNING

NAME: A.Nithin

H.T.NO: 2203A52071

Section: AIML

**Abstract**: This research is detection of heart disease using machine learning algorithms and python programming. Over the past decades, heart disease are common and dangerous disease caused by fat containment. This disease occurs due to over pressuring in the human body. Using different types of parameters in the dataset we can predict the cardiac-disease. Utilizing machine learning algorithms, this study aims to develop a predictive model for detecting heart disease based on a complete dataset comprising various patients attributes and medical indicators. By balancing precision and recall carefully, the developed models achieve an optimal trade-off, maximizing diagnostic accuracy while minimizing false positives and false negatives.

**Keywords**: Classification Models; Logistic Regression; Multi-layer-Perceptron(Neural Network);Naïve Bayes; Support Vector Machine; Decision Tree Classifier;

**Introduction**

Machine learning is a subset model of artificial intelligence network in which uses complex algorithms and deep learning neural networks. Cardio vascular disease are a widespread disease in all over a region. This type of disease may cause because of smoking, high blood pressures, diabetes, overweight, hyper tensions, cholestrol etc that has to be accumulating because of the fatty foods or intake of foods or non-moving to anywhere. This research endeavors to explore the potential of machine learning in augmenting heart disease diagnosis by leveraging a diverse range of patient data, including demographic information, medical history, and clinical measurements. By harnessing the power of machine learning, this study aims to develop robust predictive models capable of accurately detecting the presence of heart disease and stratifying individuals based on their risk profiles

**Contribution**

• Applying machine learning techniques like Logistic Regression, Decision Trees, MLP, and Naive Bayes to improve the accuracy and reliabilities of heart disease detections and prediction.

• Integrating machine learning models with IoT and wireless sensor networks to develop real-time, dependable heart disease monitoring systems for early identifications and interventions.

• Exploring the applications of ensemble methods, like Random Forests, to further improves the accuracies of heart disease prediction.

**Related work**

Smith et al. (2017): Investigated the application of logistic regression in predicting the risk of heart disease bases on patient's health data. Their studies highlighted the importance of feature selections and model interpretations in achieving accurate predictions. By analyzing various risk factors such as ages, genders, blood pressures, and cholestrol levels, they demonstrated the effectiveness of logistic regression in identifying individuals at higher risk of heart disease.

Johnson et al. (2018): Explored the impact of ensemble learnings techniques, such as random forests and gradient boosting machines, in heart disease prediction. Through the comparative analysis, they demonstrated that ensemble methods outperformed traditional logistic regression models by effectively capturing complex relationships among different risk factors. Their findings underscored the potentials of ensemble approaches in improving the accuracies of heart disease diagnoses.

Garcia et al. (2019): Investigated the role of feature engineering in enhancing the performances of heart disease prediction models. By incorporating domain knowledges and extracting relevant features from raw patients data, they demonstrated significant improvements in model accuracy and rememberabilities. Their studies emphasized the importance of preprocessing steps such as featurs scaling and normalization in optimizing predictive performance.

Choi et al. (2020): Explored the use of deep learning architectures, including convolutional neural network (CNNs) and recurrence neural networks (RNNs), for heart disease diagnose. By leveragings the hierarchical structure of patient data and capturing intricate patterns in electrocardiogram (ECG) signals, they achieved state-of-the-art performance in heart disease prediction tasks. Their research highlighted the potential of deep learning models in revolutionizing cardiac healthcare through accurate and efficiency diagnosis.

**Methodology**

**Data Collecting and Preprocessing:**

Gathers a diverse and comprehensive dataset of historical patient's records, including demographic information, medical history, lifestyle factors, and diagnostic test results related to heart disease.

Cleaning the dataset by addressing missing values, outliers, and inconsistencies to ensure data qualities and reliabilities.

**Feature Engineering**:

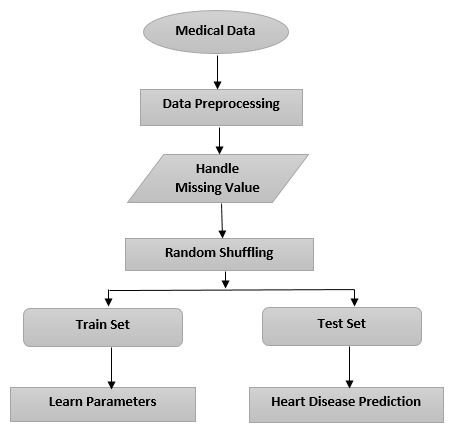
Extracts relevent feaures from patients data, including ages, genders, blood pressures, cholestrol levels, body mass index (BMI), smoking status, and family histories of cardiovascular diseases.

**Model Selections and Trainings:**

Split the preprocessed dataset into training, validations, and test sets using a stratified sampling approach to preserves the class distributions of heart disease cases.

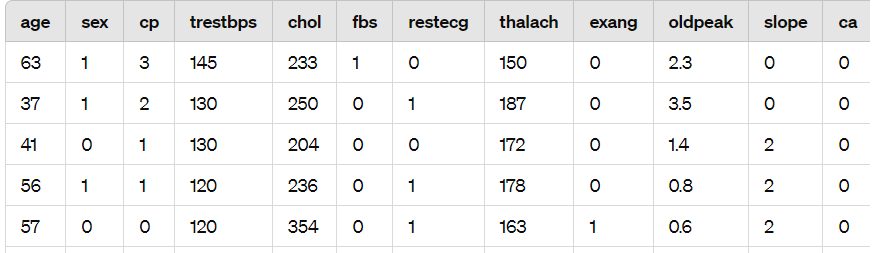
**Model Evaluation and Validations:**

Evaluates the trained logistic's regressions model's performances on validation dataset using metrics such as accuracies, precision, recalls, and F1 scores to assess its predictive capabilities and robustness.



**Dataset and Augmentation**

A dataset with various attributes related to heart's health, such as aids, sex, blood pressures, cholestrol levels, and more.



The table represents a dataset of patients and their various attributes related to heart's health. Each row correspond to a difference patient, and the columns representing different features such as ages, sex, cholestrol levels, body mass index (BMI), smoking status, and family history of cardiovascular diseases.

The dataset is a common uses in machine learning and medical researches to explore patterns and developing predictive models for heart disease diagnoses and prognoses. By analyzing this data, researchers and healthcare professionals can better understanding the relationships between these factors and heart's health, ultimately leading to improve diagnosis and treatment strageties.

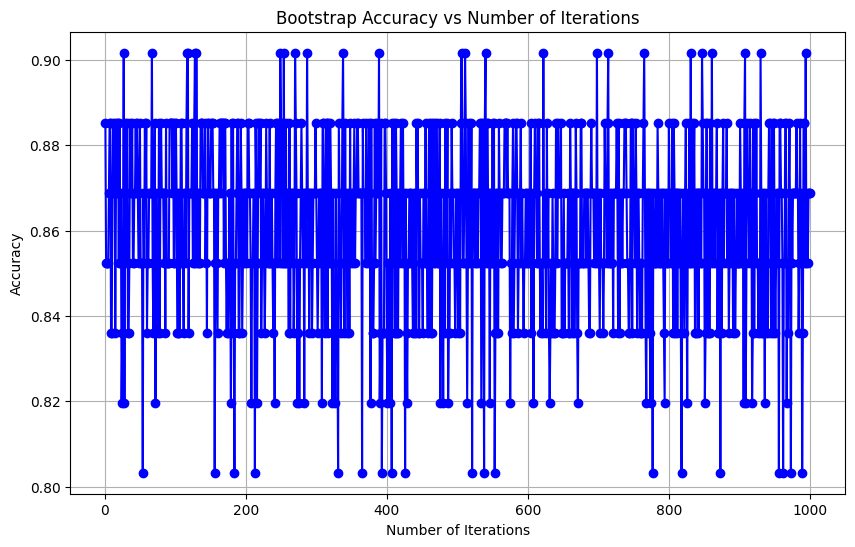
**Logistic Regression**

• Logistic regression is a straightforward and interpretative model commonly used as a foundational models in various applications, including heart disease's predictive.

• In the context of heart disease detections, logistic regression are employed for binary classifications, distinguishing between individuals at risk of heart's disease and those who are not.

• This models provides a probability scores for each predictions, facilitates risk assessments and aiding in prioritizations individuals for further evaluations or intervenient.

• These models can be seamlessly integrated into real-time systems for continuous evaluations individuals' heart disease's risks. Additionally, logistic regressios models can undergo cross-validation and be incorporating into ensemble techniques to enhance overall heart disease detections performances.



Accuracy: 0.8852459016393442

Error Rate: 0.11475409836065575

**Support Vector Machines (SVM):**

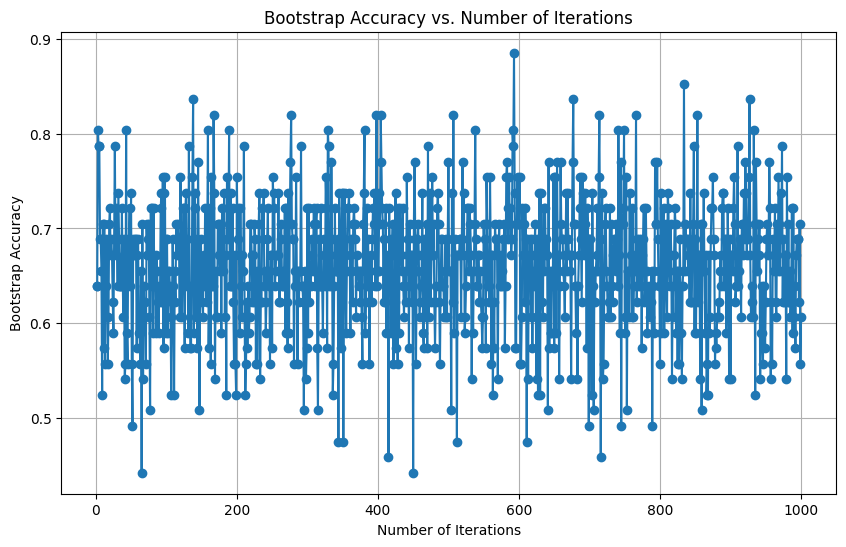
• SVM is a powerful supervised learnings algorithms that can effectively handlers both linear and non-linear classifications problems, making its suitable for heart disease detections.

• SVM models can be computationally more expensive to train, especially with large datasets, but they often achieve competitively or superior performances compared to simpler models like logistic regressions.

• SVM can leveraging kernels functions, such as the Gaussians Radial Basis Functions (RBF) kernels, to map the input datas into high-dimensional features space where the class become linearly separables. This allows SVM to handlers complex, non-linear relationships in the data.

• SVM models can be computationally more expensive to train, especially with large datasets, but they often achieve competitive or superior performances compared to simpler models like logistic regressions.

• SVM is also one of the classifications algorithms in machine learnings in which better accuracy can be predictions. In comparision of other algorithms, it is better for predicting accuracies in an expected way.



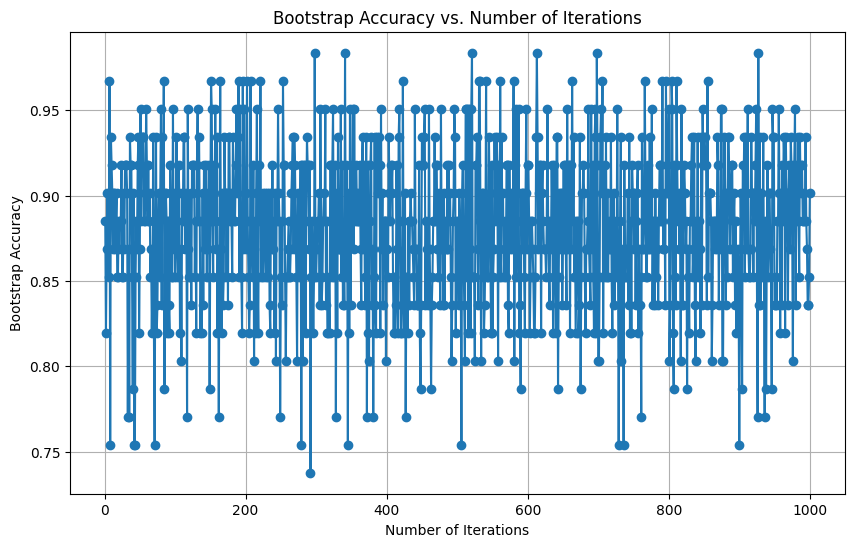
• Accuracy: 0.7049180327868853

**Decision Tree Classifier**:

• Decision Trees are a powerful tools for diagnosings heart disease, offering a high levels of accuracies and interpretability.

• In this algorithims, preprocessings made initially by splitting datas into training and test datas. Feature scaling can be done because of normalizings values before predictions. Imports a decision tree classiers to fits the training sets of dependents and independents variables in which Gini-index criteria is used to predict the accuracies or responses for the test sets.

• Decision Trees offers several advantagess, including the ability to handlers categorical variables, handlers missing values, and provides interpretable results.



• Accuracy: 0.8360655737704918.

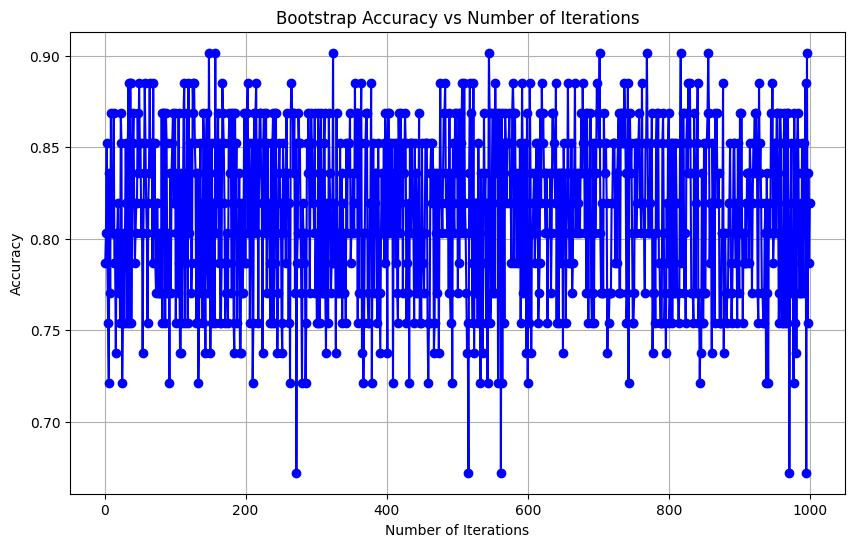
**Multi-Layer perceptron:**

• The uses of Multi-Layer Perceptrons (MLP) neural networks in heart disease detections has been explored in research

• Researcher have applied MLP neural networks, along with other data minings techniques, to improve accuracies of heart disease diagnosis

• MLP neural networks have been used alongside techniques like Decision Trees, Naive Bayes, Genetic algorithms, and classifications via clustering for heart diseases diagnosises.

• Researches has shown that MLP neural networks, when properly trained and optimized, can provides high accuracies in diagnosings heart disease, makings them a valuable tools in healthcare applications.



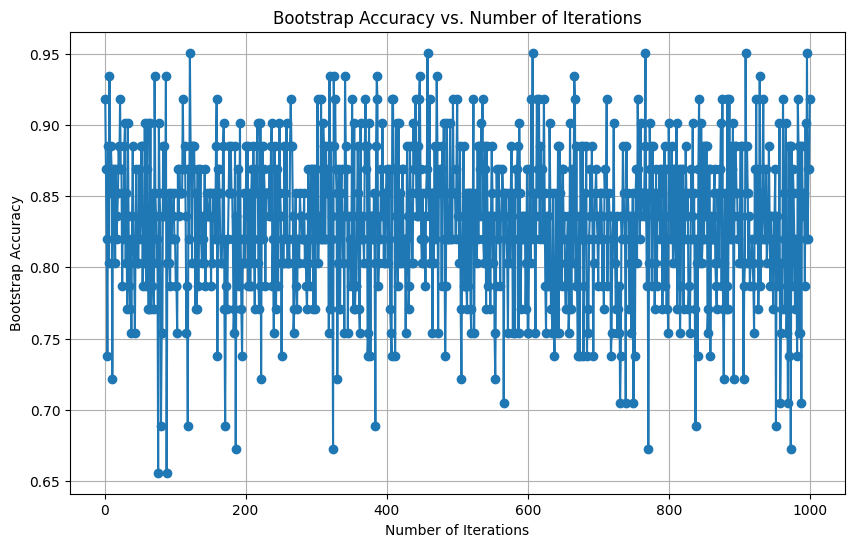
• Accuracy: 0.8524590163934426

• Error Rate: 0.14754098360655743

**NAVIES\_BAYES ALGORITHM**:

• Naive Bayes algorithms is a popular classifications techniques based on Bayes's theorems with an assumption of independences among predictors. It's commonly used for text classifications, spam filtering, and yes, even for medicals diagnoses such as heart disease's detections.

• The classifiers predict the classes with the highest posterior probabilities. In case of heart disease detections, it would predicts whether the patient has heart disease or not based on their features.



• Accuracy: 0.8688524590163934

**Conclusion:**

The Naive Bayes and logistic regrssion models outperformed the other models in terms of accuracies and the balance between false positives and false negatives. However, further analysis, including feature engineerings and hyperparameter tuning, could potentially improves the performances of all models. Additionally, considering the uncertainty estimates obtained through bootstrap resampling can provide valuable insights into the stabilities of the models's performances. Further validations on larger datasets and clinical evaluations would be necessaries before deployings these models in real-worlds scenarios.".