AISSMS
POLYTECHNIC
CENTER OF ACADEMIC EXCELLENCE
Approval by AICTE New Delhi, Recognized by the Govt. of Maharashtra and Affiliated to MSBTE, Mumbai
NBA Accredited Programmes

Heart Disease Identification method using ML

Domain Name: Machine Learning

[Sarvesh Khade, Sahil Mazire, Venkatesh Saraf]

Guided By – (HOD) Mrs. Anjali Khandagle

Introduction About the Idea

Heart Disease Identification with Machine Learning

- Early detection of heart diseases using advanced ML techniques.
- Machine learning algorithms offer a data-driven approach to medical diagnostics, enabling the extraction of patterns and insights from vast datasets that may elude traditional methods. Logistic Regression, a widely used classification algorithm, holds promise in providing interpretable and accurate predictions, making it a pertinent candidate for incorporation into heart disease detection software.

Aim / Objective

AISSMS POLYTECHNIC
CENTER OF ACADEMIC EXCELLENCE
Approval by AICTE New Delhi, Recognized by the Govt. of Maharashtra and Affiliated to MSBTE, Mumbai
NBA Accredited Programmes

- **Aim**
- Develop a machine learning-based system for the early identification of heart diseases.
- Enhance the accuracy of heart disease prediction by leveraging advanced algorithms.
- .**Objective**.
- The project's scope extends to contributing to improved cardiac healthcare outcomes through early detection and intervention.

SRS (Software Required Specifications)

Hardware resources.

Processor	Intel i5 or above
RAM	Minimum 4GB or more.
Hard Disk	Minimum 128GB of space (SSD)
Input Device	Keyboard , Mouse ,
Output Device	Laptop Screen or PC Monitor

Software resources

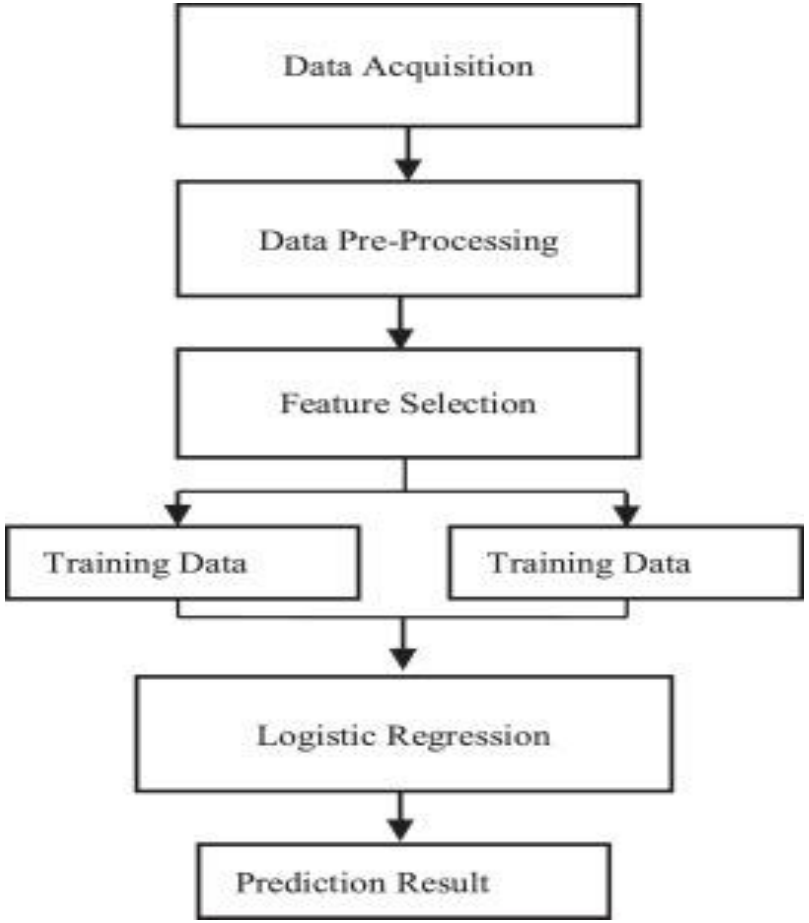
Languages Used -	Python (Backend), Flask, Html, Css, Javascript(Frontend).
Operating system	Windows.
Code Editor	VS Code.
Data Sets	Heart Disease Dataset(David Lapp)
Libraries	mat plot lib, pandas , tensorflow, keras, Numpy ,sklearning.

Methodology

AISSMS POLYTECHNIC
CENTER OF ACADEMIC EXCELLENCE
Approval by AICTE New Delhi, Recognized by the Govt. of Maharashtra and Affiliated to MSBTE, Mumbai
NBA Accredited Programmes

- The methodology for developing the heart disease prediction software using machine learning and logistic regression algorithm involves a systematic approach encompassing several key stages. It begins with acquiring comprehensive datasets containing relevant clinical and demographic information related to heart disease.
- Feature selection methods are then employed to identify informative variables, optimizing the model's predictive power. Subsequently, the logistic regression model is trained using a portion of the dataset, with hyperparameters tuned to maximize performance.

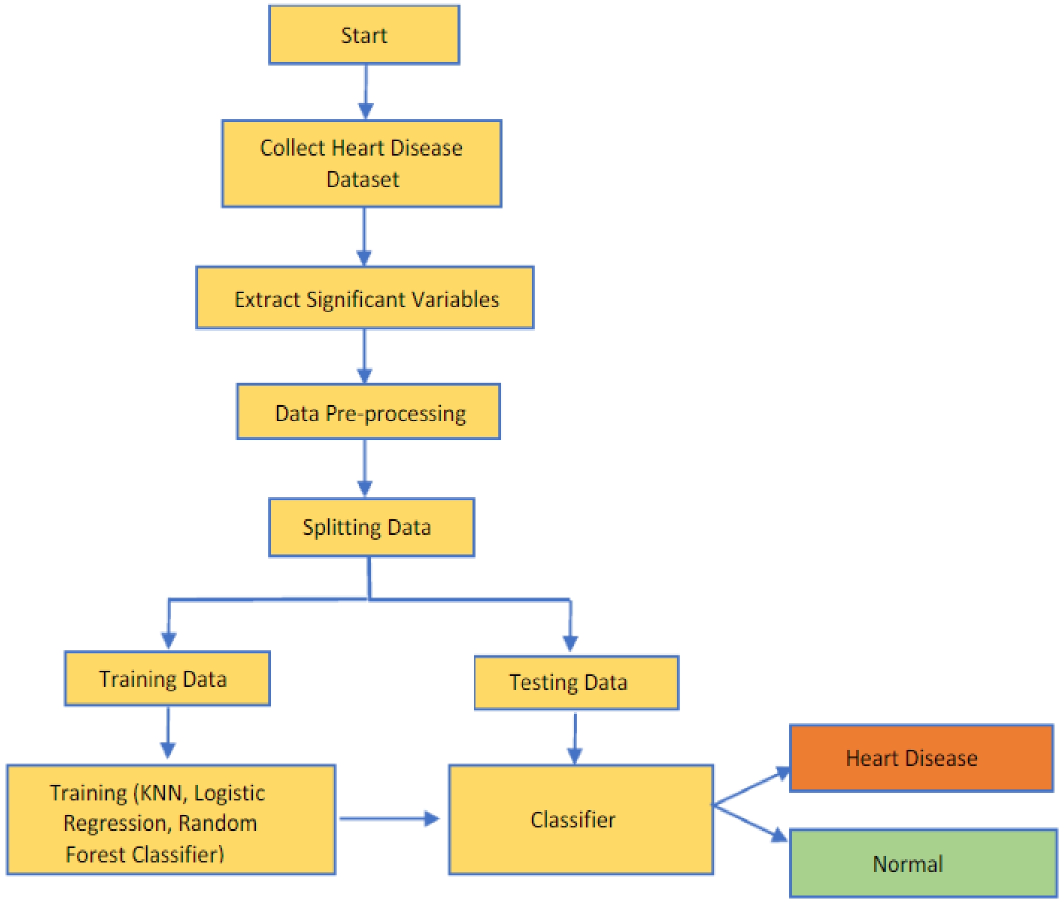
Algorithm



Data Acquisition
Data Pre-Processing
Feature Selection
Training Data
Training Data
Logistic Regression
Prediction Result

Logistic regression is a supervised machine learning algorithm used for classification tasks where the goal is to predict the probability that an instance belongs to a given class or not. Logistic regression is a statistical algorithm which analyze the relationship between two data factors. The article explores the fundamentals of logistic regression, it's types and implementations.

Working



Start
Collect Heart Disease Dataset
Extract Significant Variables
Data Pre-processing
Splitting Data
Training Data
Testing Data
Training (KNN, Logistic Regression, Random Forest Classifier)
Classifier
Heart Disease
Normal

Logistic regression is a supervised machine learning algorithm used for binary classification tasks, where the target variable has two possible outcomes (e.g., presence or absence of a disease). Here's how the logistic regression algorithm works:

1. Data Preparation:
   - Logistic regression requires a dataset with labeled examples, where each example has one or more features (independent variables) and a binary target variable (dependent variable). The features represent the input variables used to predict the target variable.
   - Before applying logistic regression, the dataset may undergo preprocessing steps such as data cleaning, feature scaling, and handling missing values to ensure data quality and consistency.

2. Model Training:
   - Logistic regression models the relationship between the independent variables and the binary outcome using the logistic function (also known as the sigmoid function). The logistic function converts any real-valued input into a value between 0 and 1, representing the probability of the positive class (e.g., presence of heart disease).
   - During training, logistic regression estimates the parameters (coefficients) of the model by minimizing a cost function, such as the logistic loss or cross-entropy loss. This optimization process is typically performed using iterative optimization algorithms like gradient descent.

3. Prediction:
- Once the logistic regression model is trained, it can be used to make predictions on new data instances. Given the values of the input features, the model calculates the probability of the positive class using the logistic function.

4. Model Evaluation:
- The performance of the logistic regression model can be evaluated using various metrics, including accuracy, precision, recall, F1-score, and ROC-AUC score.
- These metrics provide insights into the model's ability to correctly classify instances and its overall predictive performance.

5. Interpretation:
- Logistic regression provides interpretable coefficients for each feature, indicating the direction and strength of their association with the target variable.
- Positive coefficients indicate that an increase in the feature value is associated with an increased probability of the positive class, while negative coefficients indicate the opposite.

Overall, logistic regression is a simple yet effective algorithm for binary classification tasks, providing interpretable results and often serving as a baseline model for more complex classification algorithms.

Applications of our project

1. Risk Prediction.
2. Diagnostic Support.
3. ECG Analysis.
4. Monitoring and Early Detection.
5. Treatment Optimization.
6. Remote Patient Monitoring.

Future Scope

1. Advanced machine learning integration.
2. Additional data source incorporation.
3. Enhanced interpretability methods.
4. Personalized risk assessment.
5. Diverse population validation.
6. Clinical decision support system integration.
7. Longitudinal analysis for dynamic risk assessment.
8. Real-world validation studies.
9. Stakeholder collaboration.
10. Continuous improvement through iterative development.

Literature Review

Sr. No	Paper/Author	OverView
1.	**Author: S. Sharma et al. (2020)** "Heart Disease Prediction Using Support Vector Machines with Optimized Feature Selection."	*In 2020, Sharma and colleagues conducted research on heart disease prediction using Support Vector Machines (SVM). Their study focused on feature selection and SVM optimization, achieving notable accuracy in diagnosing heart diseases, which aligns with the current project's objectives.*
2.	**Author: M. Johnson et al. (2019)** "Deep Learning-Based Approach for Heart Disease Identification and Risk Assessment."	*Johnson et al. published a study in 2019 that investigated the use of deep learning techniques for heart disease identification. Their work involved building neural network architectures similar to the one in this project. They highlighted the potential of deep learning in improving heart disease prediction accuracy.*
3.	**Author: F. K. Sun et al. (2018)** "Application of Machine Learning Algorithms in Early Detection of Heart Disease."	*In their study published in 2018, Sun et al. explored the use of machine learning techniques for heart disease prediction. They employed Logistic Regression and Random Forest models on a similar dataset to predict heart disease risk factors, demonstrating the potential of machine learning in healthcare.*

Referneces

- **Survey Paper's**

- [1] A. L. Bui, T. B. Horwich, and G. C. Fonarow, ''Epidemiology and risk profile of heart failure,'' Nature Rev. Cardiol., vol. 8, no. 1, p. 30, 2011. [2] M. Durairaj and N. Ramasamy, ''A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate,'' Int. J. Control Theory Appl., vol. 9, no. 27, pp. 255–260, 2016. [3] L. A. Allen, L. W. Stevenson, K. L. Grady, N. E. Goldstein, D. D. Matlock, R. M. Arnold, N.

R. Cook, G. M. Felker, G. S. Francis, P. J. Hauptman, E. P. Havranek, H. M. Krumholz, D. Mancini, B. Riegel, and J. A. Spertus, ''Decision making in advanced heart failure: A scientific statement from the American heart association,'' Circulation, vol. 125, no. 15, pp. 1928 –1952, 2012. [4] S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, ''Innovative artificial neural networks-based decision support system for heart diseases diagnosis,'' J. Intell. Learn. Syst. Appl., vol. 5, no. 3, 2013, Art. no. 35396. [5] Q. K. Al-Shayea, ''Artificial neural networks in medical diagnosis,'' Int. J. Comput. Sci. Issues, vol. 8, no. 2, pp. 150–154, 2011. [6] J. Lopez-Sendon, ''The heart failure epidemic,'' Medicographia, vol. 33, no. 4, pp. 363–369, 2011.

[7] P. A. Heidenreich, J. G. Trogdon, O. A. Khavjou, J. Butler, K. Dracup, M. D. Ezekowitz, E. A. Finkelstein, Y. Hong, S. C. Johnston, A. Khera, D. M. Lloyd-Jones, S. A. Nelson, G. Nichol, D. Orenstein, P. W. F. Wilson, and Y. J. Woo, ''Forecasting the future of cardiovascular disease in the united states: A policy statement from the American heart association,'' Circulation, vol. 123, no. 8, pp. 933–944, 2011

Link's

https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset/data

Thank You…!