**Project Title:**

**Heart Disease Prediction Using Machine Learning**

**1.Defining the Subject**

An Introduction:

Cardiovascular diseases have been considered the major problem of morbidity and mortality worldwide, and thus, early detection is paramount for the most effective treatment. This leaves key opportunities for innovation with ML (machine learning) models to enable accurate prediction of heart disease based on clinically available parameters.

Action Items: List of tasks that can be executed to reap the results of statistical learning **Abstract:** This project builds predictive models that aim to diagnose heart disease using the UCI Heart Disease Dataset. We will determine different machine learning algorithms for the development of efficient and accurate system to assist the healthcare professionals in early diagnosis and treatment.

**2.Business Need**

Heart disease constitutes nearly a third of deaths worldwide. Early diagnosis leads to lower mortality and better patient outcomes. Conventional diagnostic approaches, on the other hand, can be tedious and subjective. A data-driven, automated approach can provide quick, consistent, and accurate diagnoses, which can help health care systems optimize resource utilization.

**Societal Impact:**

An ML based heart disease prediction system can be implemented to:

Sentence 1: In the area of early detection of diseases, you can enhance early diagnosis.

Reduce healthcare costs.

Deliver affordable diagnostics in low-resource settings.

**3.Problem Statement**

Current methods of diagnosing heart disease are expensive and not available in remote locations. They also need a great deal of clinical expertise. The objective of this project is to create an accurate prediction system for ML that predicts the probability of heart disease based on patient data (such as age, cholesterol levels, and blood pressure).

**4. High-Level Architecture**

**Components of Architecture Diagram:**

**Dataset:** Patient data (14 attributes) from UCI dataset.

Data Preprocessing:

Data cleaning.

Feature Scaling/Normalization

Categorical encoding.

EDA (Exploratory Data Analysis):

Correlations and distribution visualizations

Model Training:

Algorithms: Logistic Regressor, Random Forest, SVM, Neural Nets

Evaluation:

Metrics: ROC-AUC, F1-Score, Auc Score.

Deployment:

Flask API for backend.

Frontend visualization with Streamlit or Dash

Monitoring and Maintenance — Performance Monitoring for Model in Production

**5.Minimum Viable Product (MVP)**

**Definition:**  
The MVP is a functioning heart disease prediction model with basic user interaction through a simple UI. It will allow users to input clinical data and receive real-time predictions.

**Core Features:**

1. Data preprocessing pipeline.
2. Implementation of 4 ML models.
3. Basic accuracy comparison.
4. Simple web interface for input/output.

**6.Gantt Chart**

**Gantt Chart: 8-Week Project Timeline**

| **Week** | **Task** |
| --- | --- |
| 1 | Data exploration, preprocessing setup |
| 2 | Feature engineering, EDA, and visualization |
| 3 | Model selection: Implement Logistic Regression |
| 4 | Train and tune Random Forest |
| 5 | Develop and train SVM and Neural Network models |
| 6 | Evaluate and compare model performance |
| 7 | Develop web UI (Flask/Streamlit integration) |
| 8 | Final testing, deployment, and documentation |

**7.Team Identification**

**SURENDER** – Data Preprocessing, EDA and Model Development

**GOWTHAMI –** Evaluation, Deployment and API Integration

**8.Peer-Reviewed Articles**

peer-reviewed articles relevant to heart disease prediction using machine learning:

* **IEEE Xplore:** Access cutting-edge research in engineering and technology.
* **Scopus:** Offers a wide range of scientific journals, ensuring comprehensive literature coverage.
* **Springer:** Provides access to high-impact journals in data science and healthcare.
* **PubMed:** Ideal for biomedical research, particularly heart disease-related studies.
* **ResearchGate:** A platform for accessing papers and collaborating with researchers.

1. **Heart Disease Prediction Using Machine Learning** - This article details various machine learning models applied to predict heart disease, focusing on accuracy and model comparison ([IEEE Xplore](https://ieeexplore.ieee.org/document/9734880))
2. **Prediction of Heart Disease Using Machine Learning Techniques** - Discusses various classification algorithms like Support Vector Machines (SVMs) and Decision Trees in heart disease diagnosis ([IEEE Xplore](https://ieeexplore.ieee.org/document/9702625))
3. **Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques** - Explores hybrid models combining multiple ML techniques to enhance predictive accuracy ([IEEE Xplore](https://ieeexplore.ieee.org/document/8740989))
4. **A Comprehensive Review on Heart Disease Risk Prediction** - Provides a detailed review of risk factors and ML methodologies used in clinical settings ([Springer](https://link.springer.com/article/10.1007/s10237-021-01511-4)).
5. **Advancements in Heart Disease Prediction: A Comprehensive Review of ML Algorithms** - Examines traditional and modern ML algorithms and their performance metrics ([IEEE Xplore](https://ieeexplore.ieee.org/document/9449578))
6. **Deep Learning Approaches for Heart Disease Prediction** - Evaluates various deep learning frameworks and their effectiveness ([Springer](https://link.springer.com/article/10.1007/s10489-021-02258-2)).
7. **Machine Learning for Real-Time Heart Disease Prediction** - Focuses on real-time data processing and deployment of ML models in clinical applications ([IEEE Xplore](https://ieeexplore.ieee.org/document/8820427)).
8. **Predictive Analytics in Healthcare: A Case Study of Heart Disease** - Discusses the use of healthcare data and predictive models to forecast cardiac conditions ([ResearchGate](https://www.researchgate.net/publication/333815334)).
9. **Machine Learning Techniques for Cardiovascular Disease Prediction: A Review** - Comprehensive review of feature extraction methods and predictive modeling in CVD detection ([IEEE Xplore](https://ieeexplore.ieee.org/document/8925487)).

**9.Reference Management Tools:**

**Top Tools:**

* **Zotero:**
  + Free and open-source.
  + Helps collect, organize, and share research.
  + Integrates with Microsoft Word and Google Docs.
* **Mendeley:**
  + Provides PDF annotation features.
  + Offers social collaboration with researchers globally.
  + Automatically generates bibliographies.
* **EndNote:**
  + Advanced citation and bibliography creation.
  + Ideal for managing large volumes of references.
* **Evernote:**
  + Useful for note-taking and research organization.
  + Syncs across devices for easy access.
* **arXiv and PubMed:**
  + arXiv: Preprint archive for research in data science and AI.
  + PubMed: Essential for accessing biomedical and healthcare studies.

**10.Dataset Access**

* **Source:** UCI Repository- <https://github.com/Ruohan-Yang/Heart-Disease-Data-Set/blob/main/UCI%20Heart%20Disease%20Dataset.csv>
* **Description:** Contains 14 key attributes such as age, cholesterol, and blood pressure.