**Heart Disease Prediction Project: Problem Statement and Requirement Analysis**

**Problem Statement**

The objective of this project is to create a predictive model to diagnose heart disease using medical data. The project involves data preprocessing, exploratory data analysis, feature selection, and applying machine learning algorithms to identify the most accurate model for predicting heart disease. The developed model will assist healthcare professionals in early detection and treatment planning, ultimately improving patient outcomes.

**Requirement Analysis Document**

**1. Project Overview**

- **Title:** Heart Disease Prediction using Machine Learning

**- Objective:** Develop a predictive model to diagnose heart disease based on medical data.

- **Stakeholders:** Healthcare professionals, patients, data analysts, and developers.

**2. Functional Requirements**

- **Data Loading and Exploration:**

- Load the dataset from a CSV file.

- Display summary statistics and the first few records of the dataset.

- Identify and categorize features as categorical, integer, or float.

- **Data Visualization:**

- Generate visualizations (e.g., histograms, box plots) to explore the distribution of features.

- Create correlation matrices to identify relationships between features and the target variable.

**-Data Cleaning and Preparation:**

- Handle missing values by imputation or removal.

- Normalize or scale numerical features to ensure consistency.

- **Feature Selection:**

- Use statistical methods or feature importance techniques to select relevant features.

**- Model Training and Evaluation:**

- Split the dataset into training and testing sets.

- Train multiple machine learning models (e.g., Logistic Regression, Random Forest, SVM).

- Evaluate models using performance metrics such as accuracy, precision, recall, and F1-score.

**- Results Interpretation:**

- Identify the best-performing model based on evaluation metrics.

- Provide insights into feature importance and model predictions.

**3. Non-Functional Requirements**

**- Performance:**

- Ensure models are trained and evaluated efficiently.

- Models should provide reliable predictions within a reasonable time frame.

**- Usability:**

- Document the process clearly for easy understanding and replication.

**- Scalability:**

- Ensure the solution can handle larger datasets in the future.

**- Maintainability:**

- Write modular and well-commented code to facilitate maintenance and updates.

**4. Data Requirements**

- **Dataset Source:** The dataset should be a CSV file named "heart\_disease.csv".

- **Features:** The dataset should include relevant medical features for heart disease diagnosis.

- **Missing Data:** Ensure proper handling of missing values.

**5. Tools and Technologies**

- **Programming Language:** Python

- **Libraries:**

- Data manipulation: pandas

- Data visualization: matplotlib, seaborn

- Machine learning: scikit-learn

- Others: numpy

**6. Hardware Requirements**

- A computer with sufficient RAM and processing power to handle data manipulation and model training.

**7. Software Requirements**

- **Operating System:** Windows, macOS, or Linux

- **Development Environment:** Jupyter Notebook or any Python IDE

- **Python Version:** 3.7 or later

- **Dependencies:** Ensure installation of required Python libraries.

**8. Risk Analysis**

- **Data Quality:** Inaccurate or incomplete data may impact model performance.

- **Model Overfitting:** Ensure models generalize well to unseen data.

- **Performance Issues:** Handle computational resources for large datasets.

- **Maintenance:** Keep the codebase updated and address library deprecations.

**9. Milestones and Deliverables**

- Define milestones for data exploration, model development, and documentation.

- Deliverables include exploratory analysis, trained models, and a final project report.

**10. Conclusion**

- The Heart Disease Prediction project aims to develop a reliable model for diagnosing heart disease. By following the outlined requirements, the project will ensure a systematic approach to data preprocessing, model training, and evaluation, leading to accurate predictions and valuable insights for healthcare professionals.