**Project Design Phase-I**

**Solution Architecture**

|  |  |
| --- | --- |
| Date | 20 November 2023 |
| Team ID | Team-592802 |
| Project Name | Disease Prediction Using Machine Learning |
| Maximum Marks | 4 Marks |

**Solution Architecture:**

The Disease Prediction Using Machine Learning project employs a comprehensive solution architecture that seamlessly integrates business requirements, data processing, machine learning models, and a user-friendly web application. The solution architecture addresses the following key goals:

1. Problem Specification:

The project aims to predict diseases based on symptoms using machine learning. This involves identifying individuals at risk of specific diseases, enabling personalized preventive care and early intervention.

2. Business Requirements:

- Accurate and Reliable Information: The model must provide precise predictions to ensure correct identification of diseases.

- Trust: Establishing user trust is crucial, considering the sensitive nature of healthcare predictions.

- Compliance: Adherence to relevant laws and regulations, ensuring the model aligns with healthcare standards.

- User-Friendly Interface: Designing an intuitive interface that simplifies user interactions and avoids unnecessary input requests.

3. Literature Survey:

A thorough exploration of existing studies and projects related to disease prediction. This includes understanding current classification systems, methodologies, and gaps that the project can address.

4. Social or Business Impact:

- Social Impact: Improved preventive diagnosis, enhanced online consulting for doctors, and reduced hospital rush.

- Business Impact: Expanded online consulting opportunities for doctors, better care for critical patients, and optimized hospital workflows.

5. Data Collection and Preparation:

- Dataset Collection: Using Kaggle, the project acquires disease-related data in CSV format.

- Data Preparation: Involves cleaning the dataset, handling redundant columns, and ensuring there are no missing values.

6. Exploratory Data Analysis:

- Descriptive Statistical Analysis: Utilizes pandas to understand data features, including unique values, top values, and statistical measures.

- Visual Analysis: Employs matplotlib for univariate and bivariate analysis, exploring relationships between symptoms and disease prognosis.

7. Model Building:

- Multiple Algorithms: Constructs machine learning models using K Nearest Neighbors, SVM, Decision Tree, and Random Forest.

- Evaluation Function: Develops a standardized function for model evaluation, assessing accuracy on training, validation, and testing data.

8. Performance Testing & Hyperparameter Tuning:

- Model Testing: Assesses model performance using various evaluation metrics and compares accuracies.

- Hyperparameter Tuning: Determines optimal hyperparameters to enhance model performance.

9. Model Deployment:

- Saving the Best Model: Selects the K Nearest Neighbors model with 45 features and saves it using the pickle library.

- Integration with Web Framework: Develops a Flask web application with HTML pages (index, details, results) for user interaction.

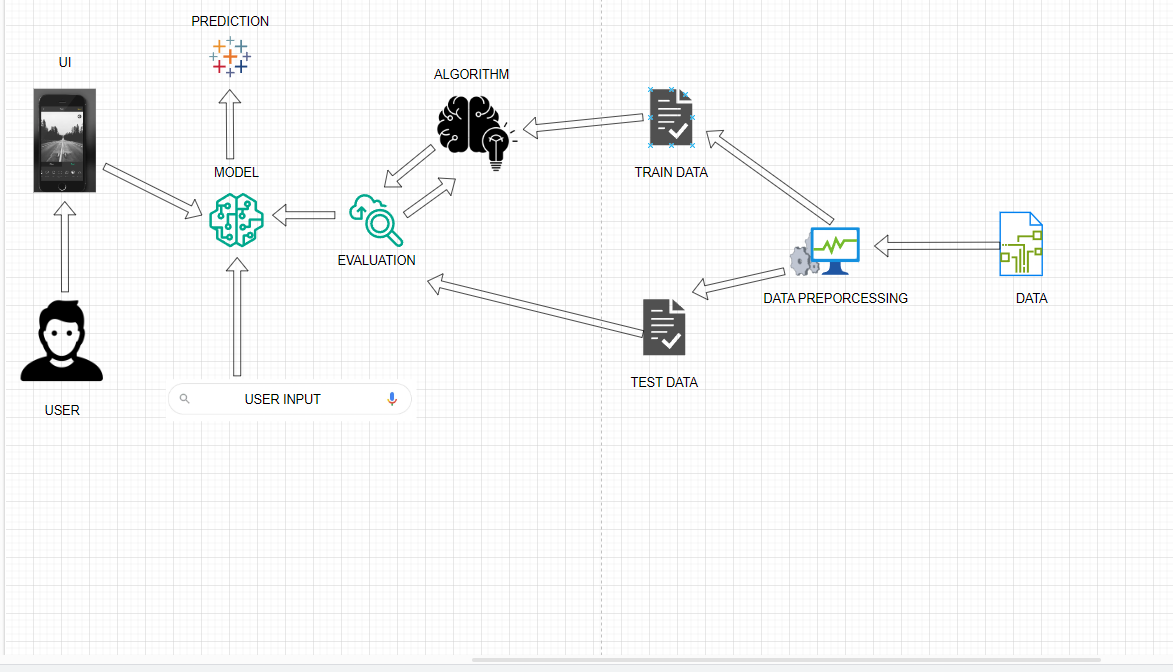
10. User Interface:

- Home Page: Provides information about the project and navigation options.

- Predict Page: Allows users to input symptoms for disease prediction.

- Results Page: Displays the predicted disease based on user inputs.

**Solution Architecture Diagram:**



*Figure 1: Architecture and data flow of the voice patient diary sample application*

**Reference: https://aws.amazon.com/blogs/industries/voice-applications-in-clinical-research powered-by-ai-on-aws-part-1-architecture-and-design-considerations/**