**Disease Prediction System using Machine Learning**

* *Shubham Chakraborty*

**Project Overview**

The "Disease Prediction System using Machine Learning" aims to develop an intelligent system that predicts the likelihood of a person having a particular disease based on various health-related features. This system utilizes machine learning algorithms to analyze historical health data and make predictions, contributing to early disease detection and proactive healthcare management.

**Table of Contents**

1. Project Overview
2. Project Details
3. Installation
4. Usage
5. Data Collection
6. Data Preprocessing
7. Feature Selection
8. Model Development
9. Cross-Validation
10. Hyperparameter Tuning
11. Model Interpretability
12. User Interface
13. Integration with Electronic Health Records (EHR)
14. Validation and Testing

**Project Details**

**Domain:** Artificial Intelligence and Machine Learning  
**Project Name:** Disease Prediction System  
**Level:** Intermediate

**Installation**

1. Clone the repository:

sh

Copy code

git clone https://github.com/yourusername/medical-ML.git

cd medical-ML

1. Create a virtual environment:

sh

Copy code

python -m venv venv

source venv/bin/activate # On Windows: venv\Scripts\activate

1. Install the required packages:

sh

Copy code

pip install -r requirements.txt

**Usage**

**Training and Saving Models:**

* Ensure your dataset is available in the dataset directory.
* Run the Jupyter Notebook to train models and save them:

sh

Copy code

jupyter notebook

# Open and run the notebook to train and save the models

**Running the Streamlit App:**

* Run the Streamlit application:

sh

Copy code

streamlit run app.py

* Open your browser and go to http://localhost:8501.

**Data Collection**

Gathered a diverse dataset containing relevant health features, including but not limited to age, gender, BMI, blood pressure, cholesterol levels, and family medical history.

**Data Preprocessing**

Performed thorough data cleaning and preprocessing to handle missing values, outliers, and ensure data quality. Normalized or standardized features to bring them to a consistent scale.

**Feature Selection**

Employed feature selection techniques to identify the most influential variables for disease prediction. Ensured that selected features contribute significantly to the accuracy of the machine learning models.

**Model Development**

Explored and implemented various machine learning algorithms such as logistic regression, decision trees, random forests, and support vector machines for disease prediction. Evaluated and compared the performance of different models using metrics like accuracy, precision, recall, and F1-score.

**Cross-Validation**

Implemented cross-validation techniques to assess the generalization performance of the models and mitigate overfitting.

**Hyperparameter Tuning**

Fine-tuned the hyperparameters of selected machine learning models to optimize their performance.

**Model Interpretability**

Enhanced the interpretability of the models to provide insights into the factors influencing the predictions. Used techniques such as SHAP (SHapley Additive exPlanations) values or feature importance plots.

**User Interface**

Developed a user-friendly interface using Streamlit that allows users to input their health-related data and receive predictions about the likelihood of having a particular disease.

**Integration with Electronic Health Records (EHR)**

Explored the integration of the disease prediction system with electronic health records, facilitating seamless information flow between healthcare providers and the system.

**Documentation**

Provided comprehensive documentation covering data sources, methodology, model architecture, and instructions for using the prediction system.

**Validation and Testing**

Conducted extensive testing and validation to ensure the accuracy, reliability, and robustness of the disease prediction system.

**Screenshots**

