Project 2 Overview Document

Python Programming (CSEG1021)

Project Title

Heart Disease Prediction Using Machine Learning

Objective

To develop a machine learning model that predicts the presence of heart disease in patients based on clinical data. The project leverages a dataset from the UCI Machine Learning Repository and includes steps from data cleaning to model deployment via a web interface.

Tools & Libraries Used

- Python 3.x
- Pandas, Numpy
- Matplotlib, Seaborn
- Scikit-learn
- Flask or Streamlit
- Pickle/Joblib

Steps Involved

1. Problem & Dataset Selection:
   - Chosen dataset: Heart Disease UCI Dataset from UCI Repository.

2. Data Cleaning and Transformation:
   - Addressed missing values.

- Converted categorical variables.
- Normalized/standardized numerical features.

3. Exploratory Data Analysis (EDA):

- Visualized data distribution and correlations.
- Used histograms, heatmaps, pair plots, etc.

4. Model Building:

- Applied classification models (e.g., Logistic Regression, Random Forest).
- Evaluated performance using metrics like accuracy, precision, recall, and F1 score.

5. Model Export:

- Saved the trained model using pickle for later use.

6. Web Application Development:

- Created a user-friendly interface using Flask/Streamlit.
- Users can input health parameters and get prediction output.

7. Integration and Deployment:

- Integrated the model with the web app for real-time predictions.

8. Submission:

- Uploaded source code and documentation to GitHub.
- Recorded a demo video showcasing the app and its features.

Project 2 Overview Document

Outcome

Built a machine learning-based web application that predicts heart disease risk based on patient data inputs.

The solution is scalable and user-friendly.

Links

- GitHub Repository: [Insert GitHub link here]

- Demo Video: [Insert YouTube or Drive link here]