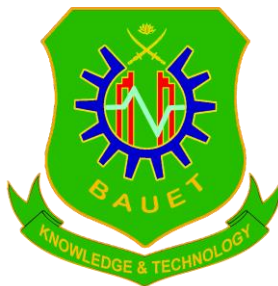# Bangladesh Army University of Engineering & Technology (BAUET)

Department of Computer Science and Engineering

Project Proposal

Course Code: CSE-4252
Course Title: Data Mining and Data Ware-housing Sessional

Project Name: Heart Disease Prediction using Machine Learning with Python

Submitted by	Submitted to
Sakib Haque Zisan ID: 18204008 Muzakkir Hossain Minhaz ID: 18204010 Md. Faisal Khan ID: 18204034	Subrata Kumer Paul Lecturer, Department of CSE, BAUET Rakhi Rani Paul Lecturer, Department of CSE, BAUET

Project Title: Heart Disease Prediction using Machine Learning with Python.

1. **Explanation of Some Existing Systems:** Now a days, heart disease prediction has been a major concept in recent world that is impacting the society towards health. The main concept is to identify the age group and heart rate using the Random Forest algorithm. Out project tell how the heart rate and condition is estimated based on the inputs such as blood pressure and many more being provided by the user to a system. This is being much better way when it comes with others algorithms the implementation of RFA gives the better experience and provide accurate result. This helps in early prediction of the disease and is used in many ways, where as it is being provided with the input, in order to find the heart rate based on the health condition.

   In this system, the input details are obtained from the patient. Then from the user inputs, using ML techniques heart disease is analyzed. Now, the obtained results are compared with the results of existing models within the same domain and found to be improved. The data of heart disease patients collected from the UCI laboratory is used to discover patterns with NN, DT, Support Vector Machines (SVM), and Naïve Bayes. The results are compared for performance and accuracy with these algorithms. The proposed hybrid method returns results of 87% for F-measure, competing with the other existing methods.

2. **Scope of Our Works:** After evaluating the results from the existing methodologies, we have used python and pandas operations to perform heart disease classification for the data obtained from the UCI repository. It provides and easy-to-use visual representation of the dataset, working environment and building the predictive analytics. ML process starts form a pre-processing data phase followed by feature selection based on data cleaning, classification of modelling performance evaluation. Random Forest technique id used to improve the accuracy of the result.

3. **Methodology:**
   Step 1: Firstly, loading the CSV file, and covert it to Pandas Data Frame.
   Step 2: Checking the distribution of 'target' variable.
   Step 3: Splitting the features and target.
   Step 4: Splitting the dataset into 'Training' and 'Testing' data.
   Step 5: Model training using 'Logistic Regression'.
   Step 6: Model evaluation with accuracy score.
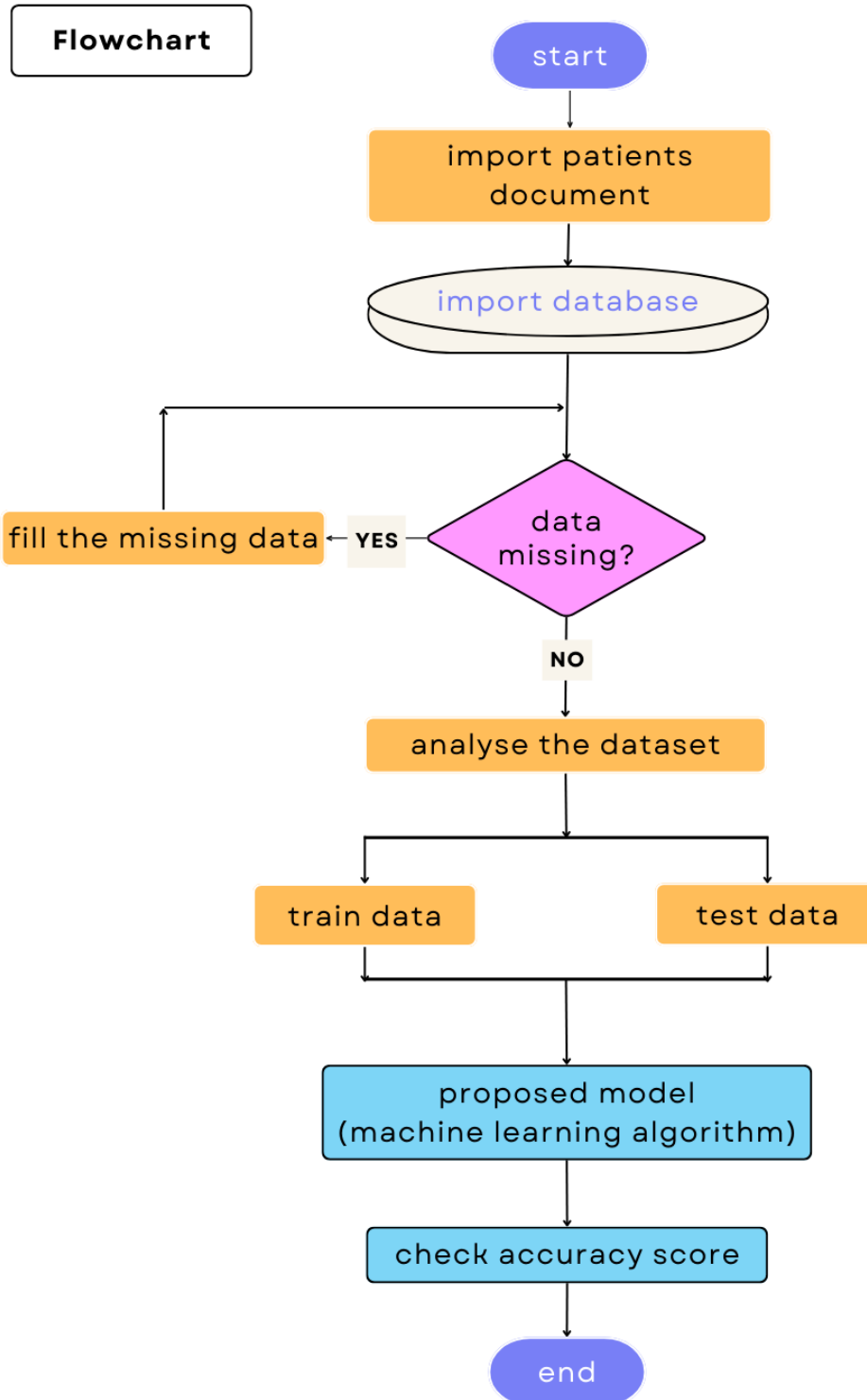   Step 7: Finally, 'Building a predictive system'.

Figure: Flowchart of Proposed System

4. Requirements for Our Proposed Model to Development:

i. Web IDE: Jupyter Notebook, Google Colab or Kaggle Notebook.
ii. Python Programming Language.
iii. Dataset (for Heart Disease Prediction)
iv. Personal Computer
v. Internet Connection

5. Discussion:

5.1. Advantages:

→ Increased accuracy for effective heart disease diagnosis.
→ Handles roughest (enormous) amount of data using random forest algorithm and feature selection.
→ Reduce the time complexity of doctors.
→ Cost effective for patients.

5.2. Disadvantages:

→ Prediction of cardiovascular disease results is not much more accurate.
→ Data mining techniques does not help to provide effective decision making.