MINI PROJECT REPORT

**Title –** Medical diagnosis of diseases based on adequate symptoms.

**Problem Statement –** Implement expert system to identify most probable disease patient is suffering from on basis of provided symptoms.

**Objective –**

1. To understand concept of Decision trees and Naïve Bayes classifier (MultinomialNB).
2. To compare performance of both algorithms by making them perform same task of prediction.

**Outcome -** After successfully completing this assignment, you should be able to

Display most probable disease patient is suffering on basis of passed symptoms.

**Software and Hardware requirements –**

1. Operating System : 64-bit Open source Linux or its derivative
2. Programming Languages: PYTHON/R

* **Algorithm :**

Step 1: Preprocess the Dataset (Disease Symptom).

Step 2: Calculate Feature Vector.

Step 3: Create Test Data.

Step 4: Use Mathematical model to Develop Predictive Model.

Step 5: Preprocess the Test Data.

Step 6: Calculate Feature Vector for Test Data.

Step 7: Use Predictive model to Test data and predict Disease.

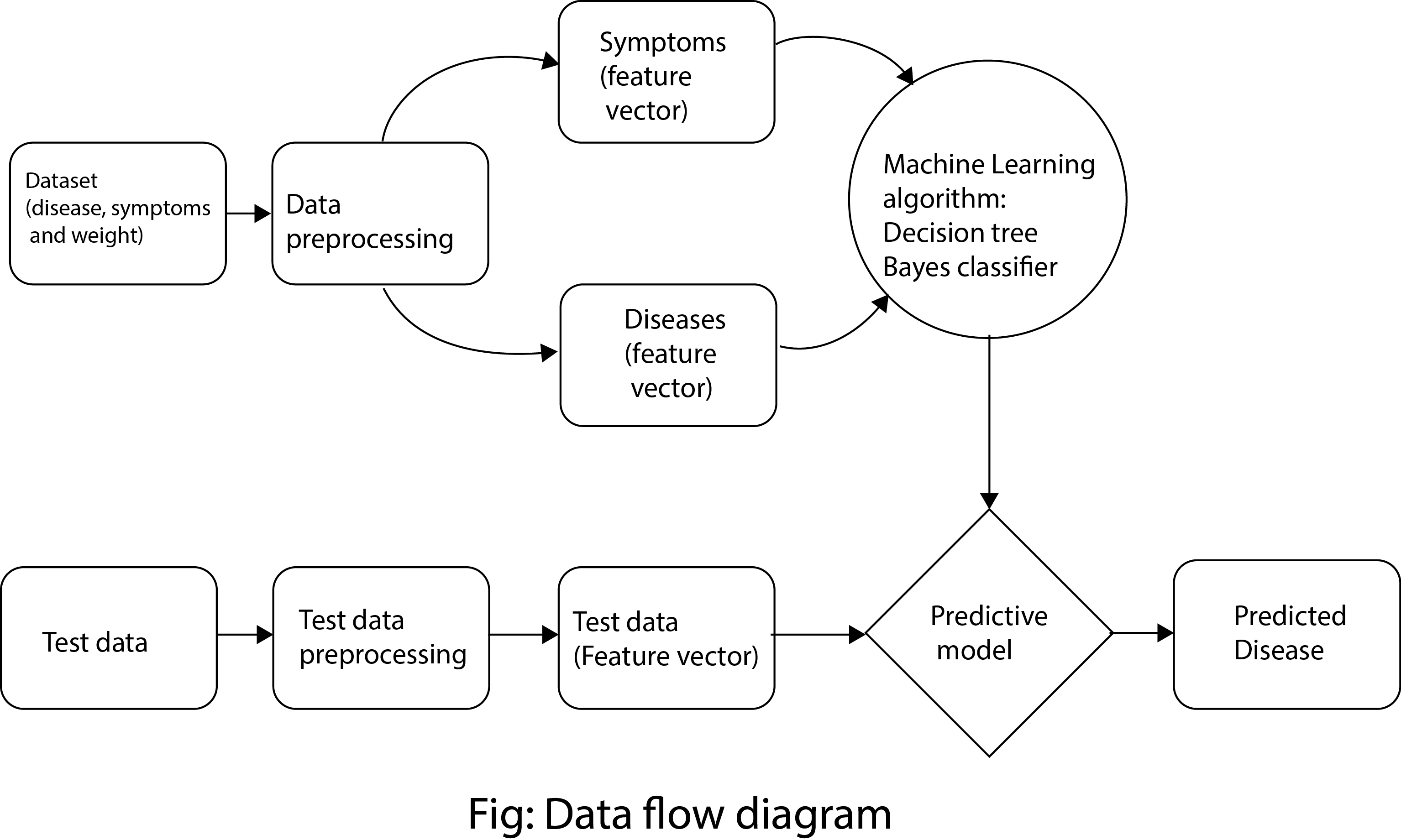
* **Concepts related Theory :-**

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. There is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, roundness, and diameter features.

For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods.

Decision tree learning is a method commonly used in data mining. The goal is to create a model that predicts the value of a target variable based on several input variables. An example is shown in the diagram at right. Each interior node corresponds to one of the input variables; there are edges to children for each of the possible values of that input variable. Each leaf represents a value of the target variable given the values of the input variables represented by the path from the root to the leaf.

A decision tree is a simple representation for classifying examples. For this section, assume that all of the input features have finite discrete domains, and there is a single target feature called the "classification". Each element of the domain of the classification is called a class. A decision tree or a classification tree is a tree in which each internal (non-leaf) node is labeled with an input feature. The arcs coming from a node labeled with an input feature are labeled with each of the possible values of the target or output feature or the arc leads to a subordinate decision node on a different input feature. Each leaf of the tree is labeled with a class or a probability distribution over the classes, signifying that the data set has been classified by the tree into either a specific class, or into a particular probability distribution (which, if the decision tree is well-constructed, is skewed towards certain subsets of classes).



**Conclusion –** Hence, we have implemented an expert system to detect most probable disease patient is suffering from on basis of passed symptoms.