**Classification Models**

Classification models predicts categorical values. The dependent or response variable in the training dataset is a categorical variable, that is, the label is a categorical variable.

It is a data analysis technique to extract models describing important data classes and predict future values (Sangeetha, 2015).

A classification model can be a **binary, multi–class, or multi–label classifiers**. A binary classifier classifies an input sample or observation into two exhaustive categories. A multiclass classifier classifies or predicts a label (observation/input sample) that can have more than two classes (Guller, 2015). A multi–label classification is a classification task where each input sample can be assigned multiple labels.

**Steps involved in Classification Process**

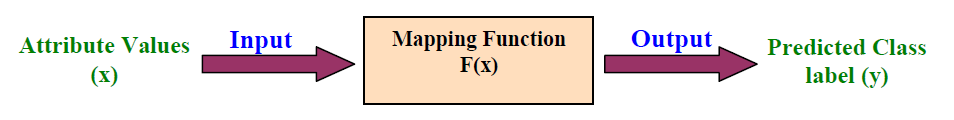
The steps involved in classification process are as follows:

Deriving a classifier model

Testing the derived model

**Deriving a Classifier Model**

A classifier is derived using the predetermined set of data classes or labels. This is also known as training step or learning step, where a classifier model is built using the training dataset (Sangeetha, 2015). It can also be viewed as a mapping function to predict the associated class labels which is depicted in the Figure 1.1 given below.



**Figure 1.1:** Classification models as a Mapping Function

**Testing the derived model**

In this stage, the estimated model is used for classification of test data, and the predictive accuracy of the classifier model is estimated (Sangeetha, 2015).

**K–Nearest Neighbors (K–NN)**

It is a classification algorithm that takes a bunch of a labelled points and uses them to learn how to label other points. This algorithm classifies cases based on their similarity to other cases. Cases that are near to each other are called **“neighbors”.**

Advantages

Disadvantages

**Procedure**

1. Pick a value for
2. Calculate the distance of unknown case from all cases
3. Select the dimensional in the training data that are ***“nearest”*** to the unknown data point
4. Predict the response of the unknown data point using the most popular response value from the