Classification VERSUS Prediction

Classification and prediction are two main methods used to mine the data. We use these two techniques to analyze the data, to explore more about unknown data.

What is classification?

- Classification is too familiar to the level or the class note of a new observation or information. In classification, without a label data or the information is given to the model, it should find the class in a specified place.
- Classification and Prediction are two forms of data mining that can be used to abstract models describing significant data classes or to predict future data direction.

- Example: Bank loan approval to customer based on the customers age.

What is a prediction?

- The second way to operate data mining is Prediction. It is repeatedly used to detect several data. Same thing as over in classification, the behaviour of data set holds the inputs and similar numerical output values. Compatible with the behaviour of the dataset, the algorithm (division) gets the model or a predictor.
- When the new information is given, the model should detect a numerical output. Despite the classification, this procedure does not have the class label or notes. The model estimates the current valued action or command value.
- Regression (Growth) in most cases is used for Predication. Predicting the price of a house rely on cases such as the number of the apartment, the total region, and so on is an illustration for predication. An organization has the power to find the amount of bank notes payout by the person during a negotiation.

- Example: Fraud detection, medical diagnosis.

Difference between Prediction and Classification:

S.No.	Prediction	Classification
1	Prediction is about predicting a Missing/ unknown element(continuous value) of a dataset	Classification is about determining a (categorial) class (or label) for an element in a dataset
2	Eg. We can think of prediction as predicting the correct treatment for a particular disease for an individual person.	Eg. Whereas the grouping of patients based on their medical records can be considered classification.
3	Eg. Whereas the grouping of patients based on their medical records can be considered classification.	The model used to classify the unknown value is called a classifier.
4	The predictor is constructed from a training set and its accuracy refers to how well it can estimate the value of new data.	A classifier is also constructed from a training set composed of the records of databases and their corresponding class names

Comparison of Classification and Prediction Methods:

Here are the few criteria that we will be used for comparing the methods of Classification and Prediction:

- Accuracy
- Speed
- Robustness
- Scalability
- Interpretability

Accuracy: Accuracy of the classifier can be referred to as the ability of the classifier to predicts the class label correctly, and the accuracy of the predictor can be referred to as how well a given predictor can estimate the unknown value.

- Speed: The speed of the method depends on the computational cost of generating and using the classifier/predictor

.

- Robustness: Robustness is the ability to make correct predictions or classifications, in the context of data mining robustness is the ability of the classifier or predictor to make correct predictions from incoming unknown data.

- Scalability: Scalability is referring to an increase or decrease in performance of the classifier or predictor based on the given data.

- Interpretability: Interpretability can be referred to as how readily we can understand the reasoning behind predictions or classification made by the predictor or classifier.

Process of Classification

Classification is a two step process —

1. Training the classifier / Model construction
2. Testing the classifier / Model usage

Step 1: Training

•The model has to be trained by describing a set of predetermined classes in the training dataset.

•Each tuple/sample from the training dataset is assumed to belong to a predefined class, as determined by the class label attribute.

•The set of tuples used for model construction is training set.

•The model is represented as classification rules, decision trees, or mathematical formulae.

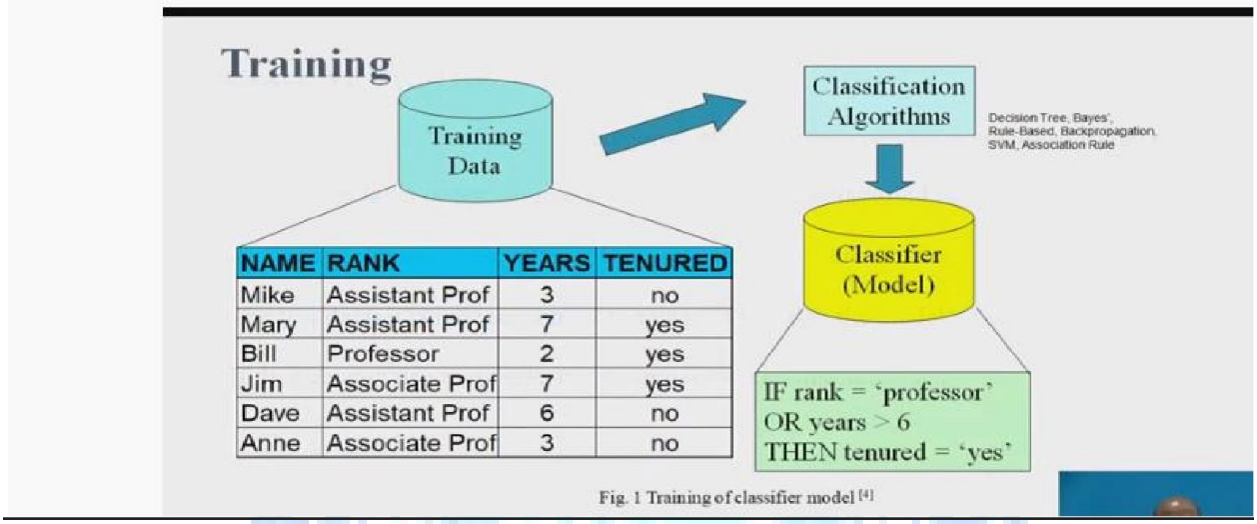

Fig. 1 Training of classifier model [4]

Step 2: Testing

- Testing is done to predict class categories of data tuples which are unknown to the model.
- The known label of test sample is compared with the classified result from the model.
- Accuracy rate is the percentage of test set samples that are correctly classified by the model.
- The test dataset is independent of training set to judge the model's accuracy.
- If the accuracy is acceptable, use the model to classify new data.

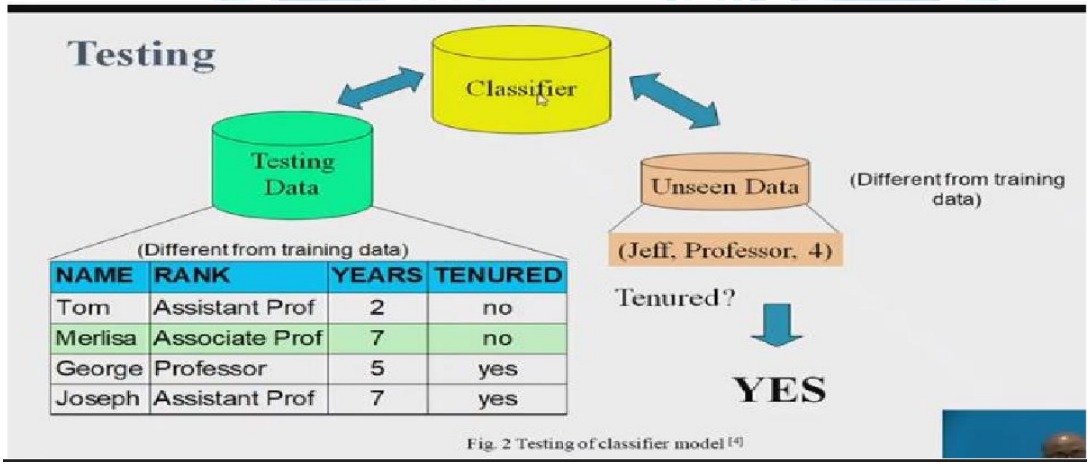

Fig. 2 Testing of classifier model [4]