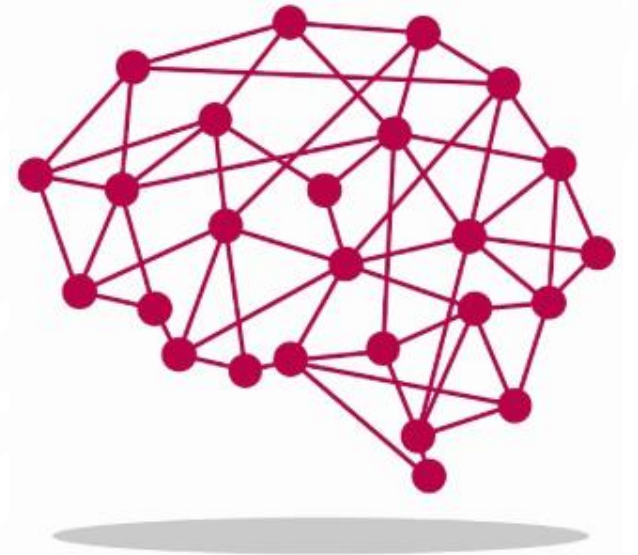# C379 EMERGING TECHNOLOGIES
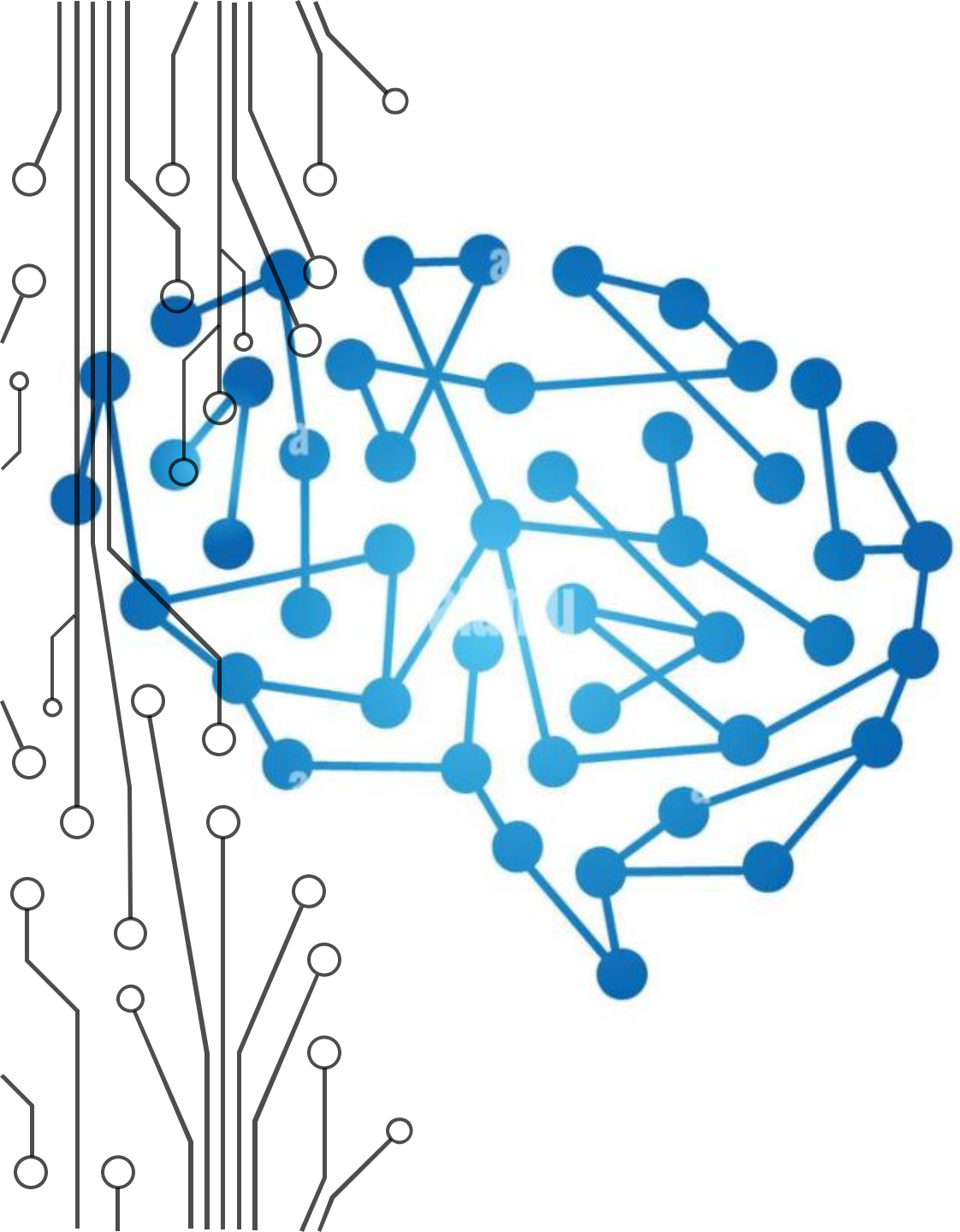
LESSON 19: RANDOM FOREST

L19 LEARNING OBJECTIVES

- Apply Random Forest algorithm to train predictive model

- Test and evaluate the Random Forest model

- Explain and effectively present results to end-user devices

MODEL BUILDING

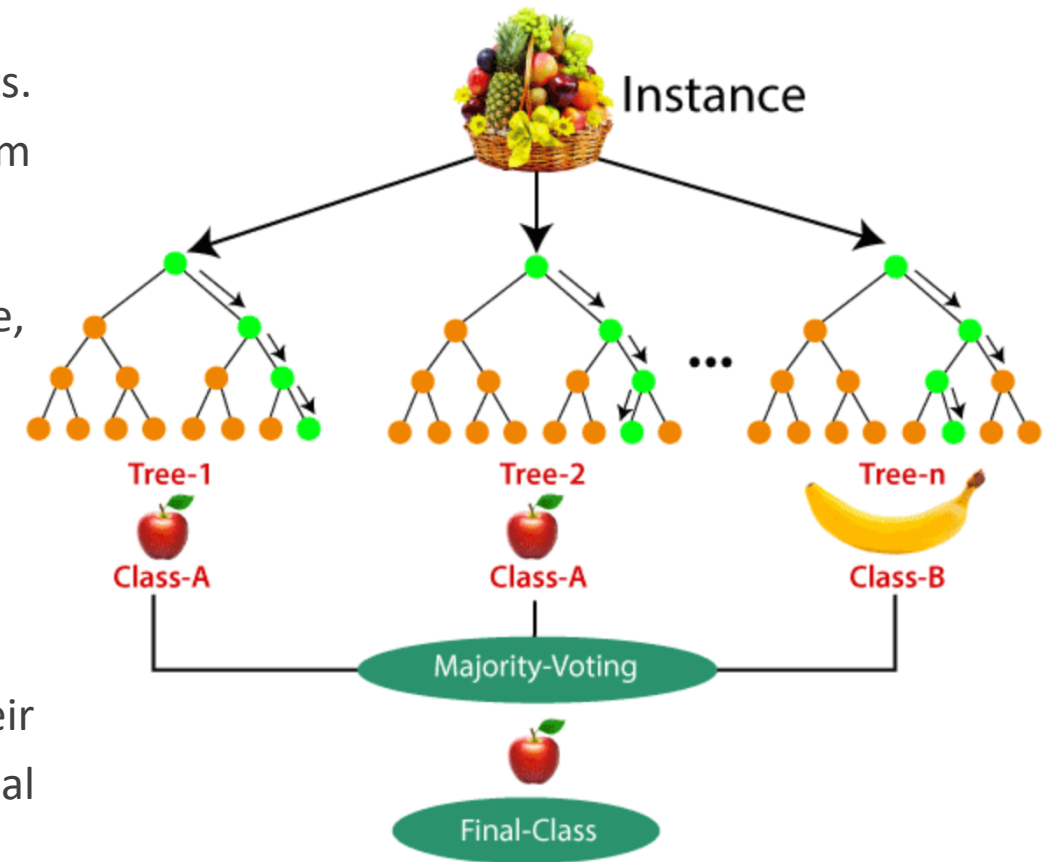RANDOM FOREST

RANDOM FOREST

- Random Forest model is a kind of **non-parametric models** that can be used both for regression and classification. It is one of the most popular ensemble methods.

- E*nsemble methods* involve using **many learners** to enhance the performance of any single one of them individually. These methods can be described as techniques that use a **group of *weak learners*** (*those who on average achieve only slightly better results than a random model*) together, in order **to create a stronger, aggregated one**.

- Random Forest is an ensemble of many individual Decision Trees. It combines the **simplicity of Decision Trees** with the **flexibility** and **power** of an ensemble model.

RANDOM FOREST

- The characteristics of random forest model are as follows:
  - It can predict a category (classification) or a continuous variable (regression).
  - It can predict with both structured and unstructured data.
  - It does not generally overfit too badly, and it is very easy to stop it from overfitting.
  - You do not need a separate validation dataset. The model can tell you how well it generalizes even if you only have one dataset.
  - It requires very few pieces of feature engineering.

RANDOM FOREST – HOW IT WORK?

- Let's take an example of a training dataset consisting of various fruits such as bananas, apples, pineapples, and mangoes.

- The random forest classifier divides this dataset into subsets. These subsets are given to every decision tree in the random forest system.

- Each decision tree produces its specific output. For example, the prediction for trees 1 and 2 is *apple*.

- Another decision tree (*n*) has predicted *banana* as the outcome. The random forest classifier collects the majority voting to provide the final prediction.

- The majority of the decision trees have chosen *apple* as their prediction. This makes the classifier choose *apple* as the final prediction.

RANDOM FOREST CLASSIFIER

1. Importing decision tree method from scikit-learn framework

2. Calling RF method with the number of estimators in the model (= 99), the number of features for best split (=5), maximum depth of the tree (=4), minimum samples required to split (=100) and the seed used to generate random number (=85) as parameters

```python
from sklearn.ensemble import RandomForestClassifier

rf = RandomForestClassifier(n_estimators=99, max_features=5, max_depth=4,
                            min_samples_split=100, random_state=85)
rf.fit(X_train, y_train)
y_pred_rf = rf.predict(X_train)
```

3. From the training dataset and labels provided in the parameters, this method train the model to make prediction.

4. The predict method is used to classify incoming data point

- *X_train* matrix contains all the features of the training set
- *y_train* matrix contains the label from the training set

LAB DEMONSTRATION

LAB 19-1

USING A RANDOM FOREST MODEL
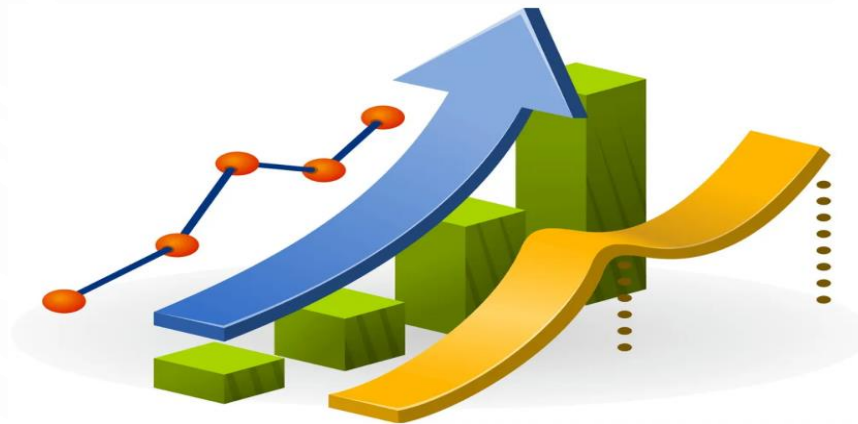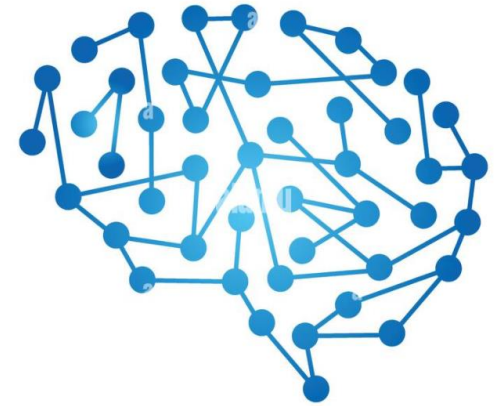
RANDOM FOREST

LAB 19-2

BUILD A RANDOM FOREST MODEL TO SOLVE

LAUNDROMAT PROBLEM

UNSCHEDULED ASYNCHRONOUS E-LEARNING

Watch the following videos before Lesson 20:

- Performance Assessment -

https://www.youtube.com/watch?v=CYy0TZ6OIDw

60 mins