**Dhirajlal Gandhi**

**College of Technology**

# 

# 

# 

# ***Department of AI&DS***

# ***Subject Name: Machine Learning***

# ***Subject Code: AL3451***

# ***Assignment No: 1***

# ***Topic: Train & Test In Machine Learning***

Submitted by : S.VIGNESH

## **What is Train/Test**

To measure if the model is good enough to measure the accuracy of your model

the method called Train/Test

It is called Train/Test because you split the data set into two sets: a training set and a testing set.

* 80 % for Training and 20 % for Testing
* You *train* the model using the training set.
* You *test* the model using the testing set.

## **Why Training & Testing is Important**

### **Training Set**

* **Purpose**: Used to train the machine learning model by enabling it to learn patterns and relationships from the data.
* **Details**: The model adjusts its parameters (e.g., weights in a neural network) based on the data in this set to minimize error or loss.
* **Outcome**: The model becomes specialized in recognizing patterns in the training data.

### **Test Set**

* **Purpose**: Used to evaluate the model's performance on unseen data to measure its ability to generalize.
* **Details**: After training, the model is applied to this independent dataset to check how well it performs on data it has not seen before.
* **Outcome**: Helps identify whether the model overfits (too closely fits the training data, leading to poor generalization) or underfits (fails to learn enough from the training data).

### **Why Both Are Important**

1. **Avoid Overfitting**: If you only evaluate your model on the training set, it may appear to perform very well because it has "memorized" the data, but it may fail on new, unseen data.
2. **Measure Generalization**: The test set helps gauge how well the model can make predictions on data outside of the training distribution.
3. **Performance Metrics**: By using a separate test set, you can calculate performance metrics like accuracy, precision, recall, or mean squared error, which provide an objective measure of your model's effectiveness.
4. **Bias and Variance Analysis**: By comparing performance on training and test sets, you can analyze bias (underfitting) and variance (overfitting) issues in your model

**Use case of train and test**

### **!. Medical Diagnosis with ML**

* **Objective: Predict the presence of a disease based on patient data (e.g., symptoms, test results).**
* **Train Set:**
  + **Includes labeled patient records (e.g., whether they were diagnosed with the disease or not).**
  + **Used to train a classification model to distinguish between patients with and without the disease.**
* **Test Set:**
  + **Contains a separate set of patient records.**
  + **Used to validate the model's performance before deploying it in real-world diagnostics.**
* **Importance: Helps ensure the model provides accurate diagnoses for patients it hasn't seen before.**

### **2. Spam Email Detection**

* **Objective: Classify emails as spam or not spam.**
* **Train Set:**
  + **Contains labeled emails where spam and non-spam emails are identified.**
  + **Used to train a text classification model to recognize patterns in spam emails (e.g., specific keywords, formatting).**
* **Test Set:**
  + **Contains unseen emails.**
  + **Used to evaluate the model’s ability to detect spam in new email data.**
* **Importance: Ensures the model can filter spam emails accurately without flagging legitimate ones.**

.