

TEAM ID	PNT2022TMID03943
PROJECT NAME	University Admit Eligibility Predictor

PRIOR KNOWLEDGE:

Prior knowledge required for completing this project are discussed below:

- Machine Learning
- Supervised and Unsupervised Learning
- Clustering, Classification, Regression
- Python Flask

MACHINE LEARNING:

Machine Learning is the field of study that gives computer the capability to learn without being explicitly programmed. It focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

Machine learning classifiers fall into three primary categories. They are:

- Supervised Learning
- Unsupervised Learning
- Semi-Supervised Learning

SUPERVISED MACHINE LEARNING:

Supervised machine learning is defined by its use of labeled datasets to train algorithms that to classify data or predict outcomes accurately. As input data is fed into the model, it adjusts its weights until the model has been fitted appropriately. This occurs as part of the cross validation process to ensure that the model avoids over fitting/ under fitting. Supervised learning helps organizations solve for a variety of real-world problems at scale, such as classifying spam in a separate folder from your inbox.

Methods used in supervised learning includes

- Neural Networks
- Naive Bayes
- Linear regression
- Logistic regression
- Random forest
- Support Vector Machine (SVM), and more.

UNSUPERVISED MACHINE LEARNING:

Unsupervised machine learning uses machine learning algorithms to analyze and cluster unlabeled datasets. These algorithms discover hidden patterns or data groupings without the need for human intervention. Its ability to discover similarities and differences in information make it the ideal solution for exploratory data analysis, cross-selling strategies, customer segmentation, image and pattern recognition. It's also used to reduce the number of features in a model through the process of dimensionality reduction; principal component analysis (PCA) and singular value decomposition (SVD) are two common approaches for this.

Other algorithms used in unsupervised learning includes

- Neural Networks
- K-Means Clustering
- Probabilistic clustering methods, and more.

SEMI-SUPERVISED LEARNING:

Semi-supervised learning offers a happy medium between supervised and unsupervised learning. During training, it uses a smaller labeled data set to guide classification and feature extraction from a larger, unlabeled data set. Semi-supervised learning can solve the problem of having not enough labeled data (or not being able to afford to label enough data) to train a supervised learning algorithm.

CLUSTERING, CLASSIFICATION, REGRESSION:

Machine Learning algorithms are generally categorized based upon the type of output variable and the type of problem that needs to be addressed.

These algorithms are broadly divided into three types

- Regression
- Clustering
- Classification

Regression and Classification are types of supervised learning algorithms while **Clustering** is a type of unsupervised algorithm. . Clustering algorithms are generally used when we need to create the clusters based on the characteristics of the data points.

When the output variable is continuous, then it is a **regression** problem whereas when it contains discrete values, it is a **classification** problem.

CLASSIFICATION:

Classification is a type of supervised machine learning algorithm. For any given input, the classification algorithms help in the prediction of the class of the output variable.

For instance, you can label a pineapple as a fruit or vegetable in a database or categorize products based on department, segment, category, or subcategory.

TYPES:

Logistic Regression: – It is one of the linear models which can be used for classification. It uses the sigmoid function to calculate the probability of a certain event occurring. It is an ideal method for the classification of binary variables.

K-Nearest Neighbours (kNN): – It uses distance metrics like Euclidean distance, Manhattan distance, etc. to calculate the distance of one data point from every other data point. To classify the output, it takes a majority vote from k nearest neighbors of each data point.

Decision Trees: – It is a non-linear model that overcomes a few of the drawbacks of linear algorithms like Logistic regression. It builds the classification model in the form of a tree structure that includes nodes and leaves. This algorithm involves multiple if-else statements which help in breaking down the structure into smaller structures and eventually providing the final outcome. It can be used for regression as well as classification problems.

Random Forest: – It is an ensemble learning method that involves multiple decision trees to predict the outcome of the target variable. Each decision tree provides its own outcome. In the case of the classification problem, it takes the majority vote of these multiple decision trees to classify the final outcome. In the case of the regression problem, it takes the average of the values predicted by the decision trees.

Naive Bayes: – It is an algorithm that is based upon Bayes' theorem. It assumes that any particular feature is independent of the inclusion of other features.

i.e. They are not correlated to one another. It generally does not work well with complex data due to this assumption as in most of the data sets there exists some kind of relationship between the features.

CLUSTERING:

Clustering is a type of unsupervised machine learning algorithm. It is used to group data points having similar characteristics as clusters. Ideally, the data points in the same cluster should exhibit similar properties and the points in different clusters should be as dissimilar as possible.

Clustering is divided into two groups ;

- Hard clustering
- Soft clustering

In hard clustering, the data point is assigned to one of the clusters only whereas in soft clustering, it provides a probability likelihood of a data point to be in each of the clusters.

K-Means Clustering: – It initializes a pre-defined number of k clusters and uses distance metrics to calculate the distance of each data point from the centroid of each cluster. It assigns the data points into one of the k clusters based on its distance.

REGRESSION:

A regression model provides a function that describes the relationship between one or more independent variables and a response, dependent, or target variable. For example, the relationship between height and weight may be described by a linear regression model.

LINEAR REGRESSION:

Linear regression is a type of model where the relationship between an independent variable and a dependent variable is assumed to be linear.

There are two kinds of Linear Regression Model:-

- **Simple Linear Regression:** A linear regression model with one independent and one dependent variable.
- **Multiple Linear Regression:** A linear regression model with more than one independent variable and one dependent variable.

Other regression models are;

Polynomial regression, Logistic regression, Quantile regression, Ridge regression, Lasso regression, Support vector regression.

FLASK:

Flask is a Python-based micro framework used for developing small-scale websites. Flask is very easy to make Restful APIs using python.

Flask is used for **developing web applications using python**, implemented on (WSGI) web server gateway interface which is a standard for python web application development. It is considered as the specification for the universal interface between the web server and web application & Jinja2 is a web template engine which combines a template with a certain data source to render the dynamic web pages. Advantages of using Flask framework are: There is a built-in development server and a fast debugger provided.

Flask is a web framework that provides libraries to build lightweight web applications in python. It is developed by **Armin Ronacher** who leads an international group of python enthusiasts (POCCO).

Machine Learning

A cartoon character asks, "Hey SIRI, how far is the nearest subway?". The audio is converted into binary code (01100, 10110, 11110) and decoded using ML and neural networks. The diagram shows the flow from User to Siri to Apple Server. A video player interface at the bottom shows the video is at 0:37 / 9:39. A thumbnail for "Applications of Unsupervised Learning" is visible, showing a tree diagram and a robot.

Hey SIRI, how far is the nearest subway?

01100
10110
11110

Decoding with the help of ML and neural network

Applications of Unsupervised Learning 8:02

User Siri Apple Server

0:37 / 9:39 • What is Machine Learning? >

Supervised Learning

A robot character is shown next to a diagram of the supervised learning process. The process involves feeding "Known Data" (represented by icons of spoons and knives) into a "Model" to produce a "Response" (a cloud containing the word "Spoon"). "Known Response" (Spoon, Knife) is also fed into the "Model". "New Data" (a single spoon icon) is then fed into the "Model" to produce a prediction.

Known Data

Model

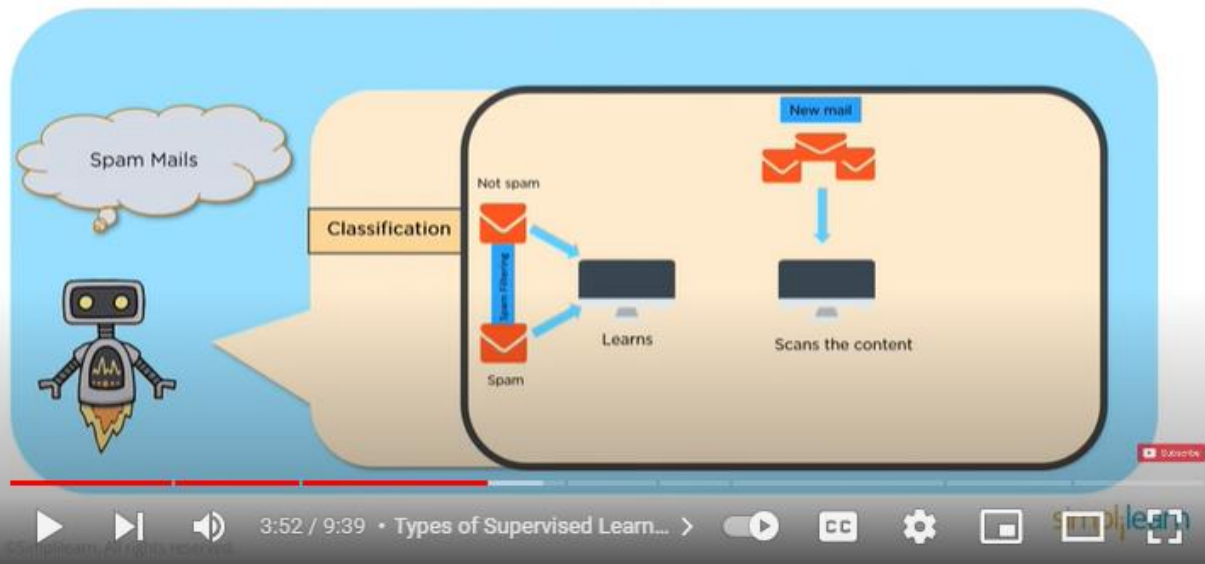
Response

Known Response

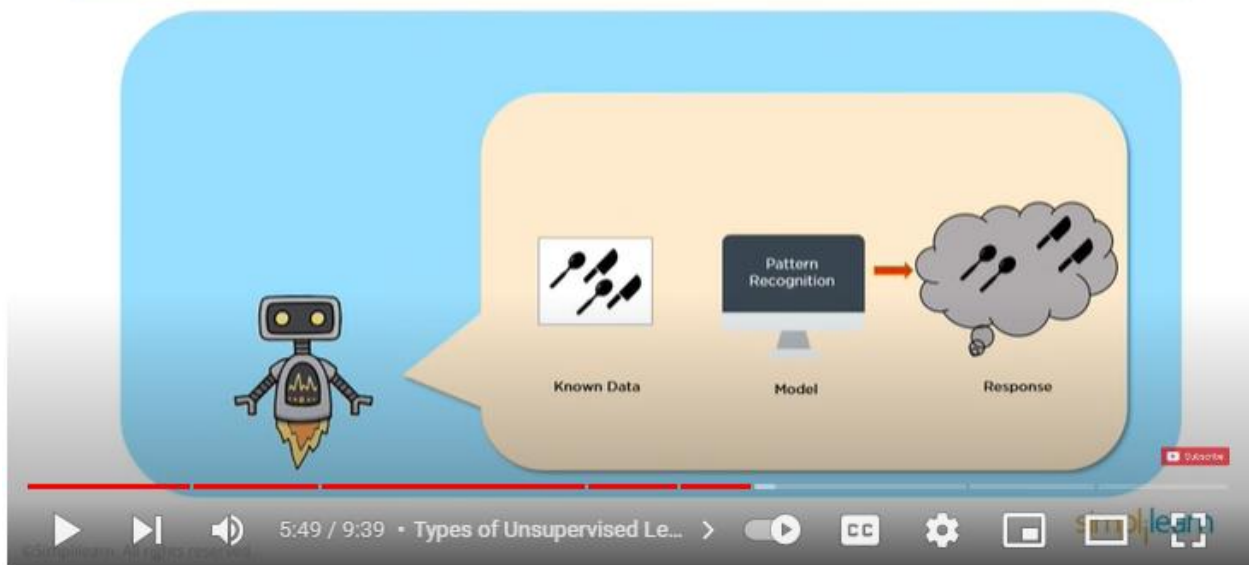
New Data

2:16 / 9:39 • Supervised Learning >

Types of Supervised Learning



Unsupervised Learning



Types of Unsupervised Learning

Suppose a telecom company wants to reduce its customer churn rate by providing personalized call and data plans

Clustering

Total Call duration

Internet Usage

Internet usage

Call duration

A

B

C

6:57 / 9:39 • Types of Unsupervised Le...

Flask – Static Files



Web application will require a static file such as JS or CSS file

```
from flask import Flask, render_template
app = Flask(__name__)

@app.route('/result')
def result():
    dict = {'phy':50,'che':60,'maths':70}
    return render_template('result.html', result = dict)

if __name__ == '__main__':
    app.run(debug = True)
```

Python

```
<!doctype html>
<html>
  <body>

    <table border = 1>
      (% for key, value in result.iteritems() %)

        <tr>
          <th> {{ key }} </th>
          <td> {{ value }} </td>
        </tr>

      (% endfor %)
    </table>

  </body>
</html>
```

SUBSCRIBE

25:51 / 35:49

Python Certification Training

www.edureka.co/python