**INTERNSHIP**

**PROGRAM REPORT**

[AI & ML]

**Industry Partner – Tevatron Technologies Pvt Ltd**

****

**Hexnbit Online Internship**

**www.hexnbit.com**

*Intern Detail*

Artificial intelligence and Machine Learning

**Name :**………………………………………………………………………………………………………………….

Kollipara Naga Prasanna Suparnika

**Name of Intern:** ………………………………………………………………………………………………………………….

9491768505

**Intern’s Ph. no.:** ………………………………………………………………………………………………………………….

5-469,Geetha Apartments Colony,NRT Centre, Chilakaluripet, Guntur(Dt),Andhra Pradesh

**Intern’s Add.:** ………………………………………………………………………………………………………………….

SRM UNIVERSITY AP

**Name of College:**………………………………………………………………………………………………………………….

Mangalagiri, Neerukonda, Tadikonda Rd, Mangalagiri,Mandal, Andhra Pradesh 522502

**College’s Add.: ……..**………………………………………………………………………………………………………….

CSE

**Branch:** ………………………………………………………………………………………………………………….

**Industry Mentor: Mr. Gagan Preet Singh**

**Designation: R&D Head**

**Company: Hexnbit EdTech Pvt. Ltd**

**Email ID: [training@hexnbit.com](mailto:training@hexnbit.com)**

**Contact Number: +91-9818894299**

*Table of content*

|  |  |  |
| --- | --- | --- |
| **SNO** | **Headings** | **Page no** |
| **1** | About company | 1 |
| **2** | Fundamentals of Python | 5 |
| **3** | Introduction to AI & ML | 7 |
| **4** | Scientific Toolkit(Numpy & Pandas) | 9 |
| **5** | Data visualization Toolkit | 12 |
| **6** | Supervised machine learning | 14 |
| **7** | Unsupervised machine learning | 17 |
| **8** | Computer vision | 20 |

*About Tevatron Technologies Pvt. Ltd*

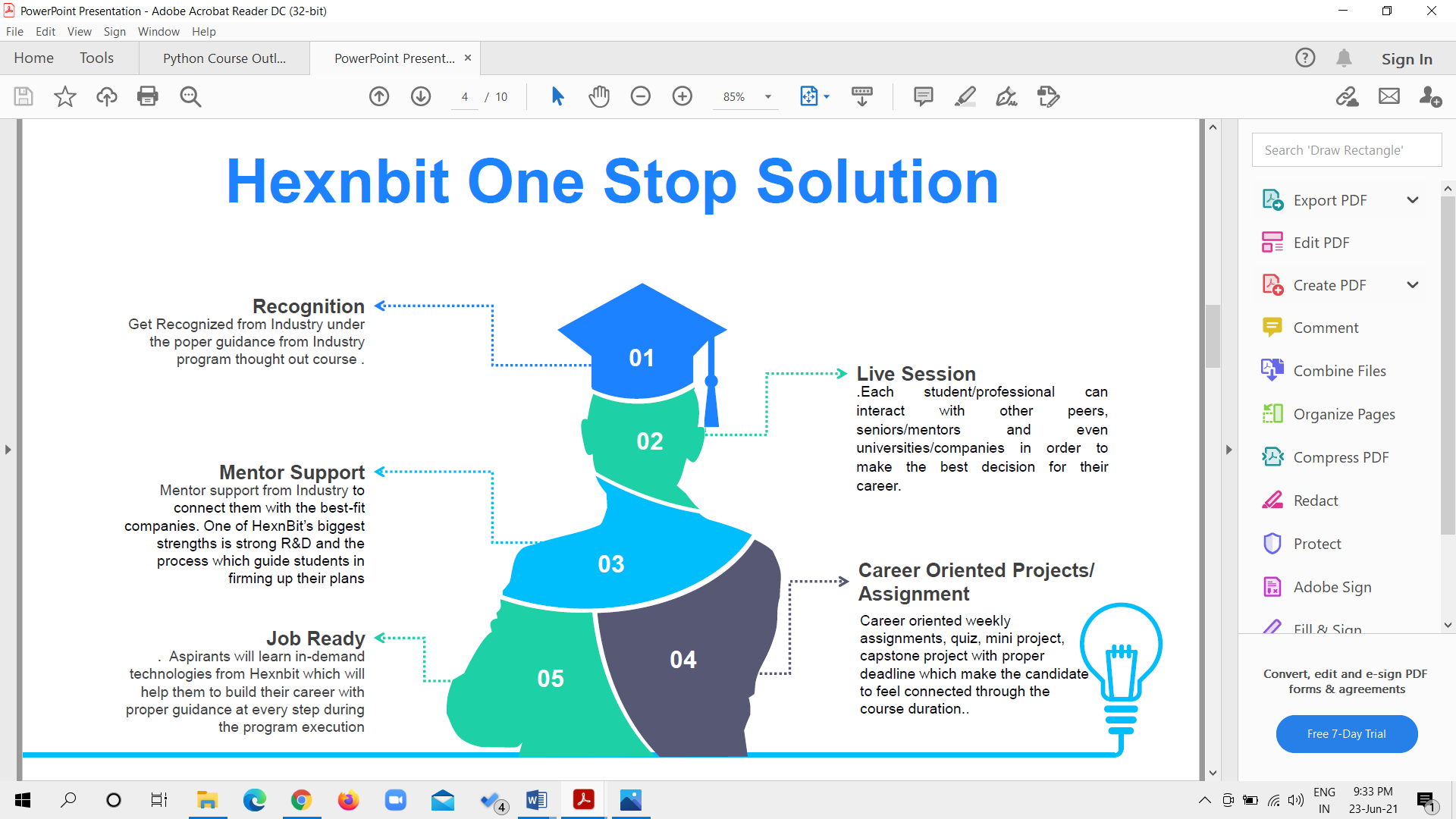
**Tevatron Technologies Pvt Ltd** is a **R&D Design & Services** Company focused on **Artificial Intelligence & Machine learning, Internet of Things (IoT), Embedded Hardware and Software Systems, Sensor and MCUs, VLSI Chip Design  and PCB Design covering entire ESDM** space from concept to Productization. We are also a member of **IESA( Indian Electronics Semiconductor Association)**. We are actively working towards **#Make in India** as well as **#Design In India** based initiatives

We have in collaboration with our training division **“Hex N Bit”** to provide Online Industrial Internship program. We have added few out of the box features in our programs like: **Project Tracking using Project Management Tools, Access to Faculty for Monitoring, Mentorship during Project, Live Sessions for Understanding of Concepts, Weekly Assignments to keep track of progress, Industry Expert Webinars, eLearning Modules and Assessments**.



*About Hexnbit EdTech Pvt. Ltd*

* Hexnbit (an ISO certified company) is India’s first Ed-tech company which provides one-stop solutions for the students/Professionals in Industry-connect Skill development courses. The platform not only provides subject expertise to the candidates but also, give them industry exposure to apply their learnings analytically in a practical real-world.
* Having registered over 60,000+ candidates, tied up with 250+ universities, 60+ mentors and10+ Industry,the company aims to bridge the gap between the academics & Industry by providing them with practical knowledge & analytical skills under one umbrella.
* To reach out to the candidates all over the globe, the company has formed Labs in many institutions so any candidate can learn & grow technically with the mentor support (On-Site support as well as virtual support)
* **Hexnbit is now recognised worldwide recognised by STMicroelectronics (<https://www.st.com/en/support-and-applications/technical-training-on-stm32.html>)**



*List of Software& Modules used*

**List of Software:**

* Anaconda Navigator
* Jupyter Notebook

**List ofModules/Libraries:**

* Numpy
* Pandas
* Matplotlib
* Seaborn
* Sckit-Learn
* Os Module

*Modules List*

*Fundamentals of Python*

**Description:**

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

**Characteristics of Python Programming:**

* It supports functional and structured programming methods as well as OOP.
* It can be used as a scripting language or can be compiled to byte-code for building large applications.
* It provides very high-level dynamic data types and supports dynamic type checking.
* It supports automatic garbage collection.
* It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

**Applications:**

* Python can be used on a server to create web applications.
* Python can be used alongside software to create workflows.
* Python can connect to database systems. It can also read and modify files.
* Python can be used to handle big data and perform complex mathematics.
* Python can be used for rapid prototyping, or for production-ready software development.

**Topics Covered**:

* Getting around Anaconda and Jupyter Notebook
* Python Basics
* Data Types
* Conditional Statements, Loops and Control Statements
* Functions
* Lambda Functions and other built-in functions
* File Handling

*Fundamentals of Python*

**Learning Outcome:**

In this module, exposure was given around the fundamentals of Python Programming to build up a strong foundation. As strong programming foundation will be helpful in data preparation, data preprocessing, analyzing, etc.

*Introduction to Artificial intelligence and Machine learning*

**What is AI & ML?**

**Artificial intelligence:**

A computer system able to perform tasks normally require human intelligence such as visual perception, speech recognition, decision making & translation between languages.

Machine learning:

ML is the ability to learn without being explicity programmed consists more advanced techniques and models that enable computers to figure things out from data and make decisions.

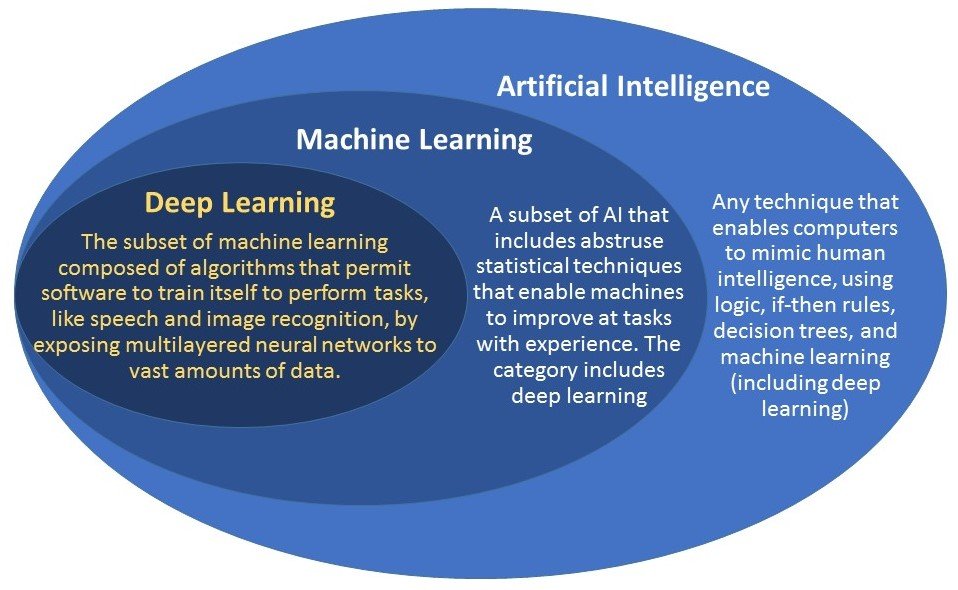
Deep learning:

Human brain, neutral net consists 1000s and millions of processing nodes are inter connected these multilayered artificial neural networks deliver high accuracy such as object detection, speech recognization etc

Now that you know what is AI & ML, let’s see why AI & ML is essential in the current scenario.

**Why** AI & ML**?**

Artificial Intelligence and Machine Learning  always interests and surprises with their innovations. AI and Ml have reached industries like Customer Service, E-commerce, Finance and where not. By 2020, 85% of the customer interactions will be managed without a human. There are certain implications of AI and ML to incorporate data analysis like Descriptive analytics, Prescriptive analytics and Predictive analytics.

****

In the above picture, DL is subset of ML and both are subsets of AI.

*Scientific Toolkit (Numpy & Pandas)*

**What is Numpy?**

NumPy is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more.

At the core of the NumPy package, is the ndarray object. This encapsulates n-dimensional arrays of homogeneous data types, with many operations being performed in compiled code for performance. There are several important differences between NumPy arrays and the standard Python sequences:

* NumPy arrays have a fixed size at creation, unlike Python lists (which can grow dynamically). Changing the size of an ndarray will create a new array and delete the original.
* The elements in a NumPy array are all required to be of the same data type, and thus will be the same size in memory. The exception: one can have arrays of (Python, including NumPy) objects, thereby allowing for arrays of different sized elements.
* NumPy arrays facilitate advanced mathematical and other types of operations on large numbers of data. Typically, such operations are executed more efficiently and with less code than is possible using Python’s built-in sequences.
* A growing plethora of scientific and mathematical Python-based packages are using NumPy arrays; though these typically support Python-sequence input, they convert such input to NumPy arrays prior to processing, and they often output NumPy arrays. In other words, in order to efficiently use much (perhaps even most) of today’s scientific/mathematical Python-based software, just knowing how to use Python’s built-in sequence types is insufficient - one also needs to know how to use NumPy arrays.
* The points about sequence size and speed are particularly important in scientific computing. As a simple example, consider the case of multiplying each element in a 1-D sequence with the corresponding element in another sequence of the same length.

*Scientific Toolkit (Numpy & Pandas)*

**What is Pandas?**

Pandas is an open source Python package that is most widely used for data science/data analysis and machine learning tasks. It is built on top of another package named Numpy, which provides support for multi-dimensional arrays. As one of the most popular data wrangling packages, Pandas works well with many other data science modules inside the Python ecosystem, and is typically included in every Python distribution, from those that come with your operating system to commercial vendor distributions like ActiveState’s ActivePython.

**Library Highlights:**

* A fast and efficient DataFrame object for data manipulation with integrated indexing.
* Tools for reading and writing data between in-memory data structures and different formats: CSV and text files, Microsoft Excel, SQL databases, and the fast HDF5 format.
* Intelligent data alignment and integrated handling of missing data: gain automatic label-based alignment in computations and easily manipulate messy data into an orderly form.
* Flexible reshaping and pivoting of data sets.
* Intelligent label-based slicing, fancy indexing, and subsetting of large data sets.
* Columns can be inserted and deleted from data structures for size mutability.
* Aggregating or transforming data with a powerful group by engine allowing split-apply-combine operations on data sets.
* High performance merging and joining of data sets.
* Hierarchical axis indexing provides an intuitive way of working with high-dimensional data in a lower-dimensional data structure.
* Time series-functionality: date range generation and frequency conversion, moving window statistics, date shifting and lagging. Even create domain-specific time offsets and join time series without losing data.
* Python with pandas is in use in a wide variety of academic and commercial domains, including Finance, Neuroscience, Economics, Statistics, Advertising, Web Analytics, and more.

**Topics Covered:**

* **NumPy**  
  NumPy Basics  
  Operations  
  Indexing, Slicing and Copies

*Scientific Toolkit (Numpy & Pandas)*

* **Pandas**Series  
  DataFrames  
  Fix Missing Data  
  GroupBy  
  Merge  
  Operations  
  File Reading and Writing

**Learning Outcome:**

In this module, Introduction to Numpy and Pandas library was given that how these libraries can be helpful in preparing and cleaning data, processing the data, bringing in the insights about the data, data analysis, understanding important features related to business domain, etc.

*Data Visualization Toolkit*

**What is Data Visualization?**

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

In the world of Big Data, data visualization tools and technologies are essential to analyze massive amounts of information and make data-driven decisions.

**Why is Data Visualization important?**

Because of the way the human brain processes information, using charts or graphs to visualize large amounts of complex data is easier than poring over spreadsheets or reports. Data visualization is a quick, easy way to convey concepts in a universal manner – and you can experiment with different scenarios by making slight adjustments.

**Data visualization can also:**

* Identify areas that need attention or improvement.
* Clarify which factors influence customer behavior.
* Help you understand which products to place where.
* Predict sales volumes.

**Common general types of data visualization:**

* Charts
* Tables
* Graphs
* Maps
* Infographics
* Dashboards

**Data visualization can be used for:**

* Making data engaging and easily digestible.
* Identifying trends and outliers within a set of data.
* Telling a story found within the data.
* Reinforcing an argument or opinion.
* Highlighting the important parts of a set of data.

*Data Visualization: Matplotlib*

**What is Matplotlib?**

Matplotlib is a cross-platform, data visualization and graphical plotting library for Python and its numerical extension NumPy. As such, it offers a viable open source alternative to MATLAB. Developers can also use matplotlib’s APIs (Application Programming Interfaces) to embed plots in GUI applications.

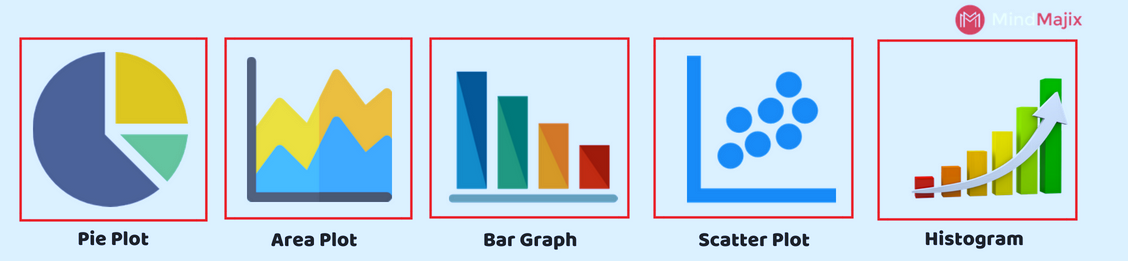
A Python matplotlib script is structured so that a few lines of code are all that is required in most instances to generate a visual data plot. The matplotlib scripting layer overlays two APIs:

* The pyplot API is a hierarchy of Python code objects topped by matplotlib.pyplot
* An OO (Object-Oriented) API collection of objects that can be assembled with greater flexibility than pyplot. This API provides direct access to Matplotlib’s backend layers.

One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals. Matplotlib consists of several plots like line, bar, scatter, histogram etc.

**Python Matplotlib : Types of Plots**

There are various plots which can be created using python matplotlib. Some of them are listed below:



**Learning Outcome:**

In this module, Introduction to Data Visualization was given that how these libraries can be helpful in better understanding of data, bringing in the insights about the data, understanding important features related to business domain, representing facts in compact manner to the stakeholders, etc.

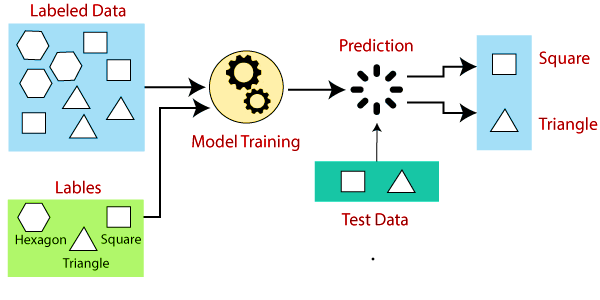
*Supervised Machine Learning*

**What is Supervised Machine Learning?**

* Supervised learning is the types of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output. The labelled data means some input data is already tagged with the correct output.
* In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly. It applies the same concept as a student learns in the supervision of the teacher.
* Supervised learning is a process of providing input data as well as correct output data to the machine learning model. The aim of a supervised learning algorithm is to find a mapping function to map the input variable(x) with the output variable(y).
* In the real-world, supervised learning can be used for Risk Assessment, Image classification, Fraud Detection, spam filtering, etc.

**How Supervised Learning Works?**

* In supervised learning, models are trained using labelled dataset, where the model learns about each type of data. Once the training process is completed, the model is tested on the basis of test data (a subset of the training set), and then it predicts the output.



Suppose we have a dataset of different types of shapes which includes square, rectangle, triangle, and Polygon. Now the first step is that we need to train the model for each shape.

* If the given shape has four sides, and all the sides are equal, then it will be labelled as a Square.
* If the given shape has three sides, then it will be labelled as a triangle.
* If the given shape has six equal sides then it will be labelled as hexagon.

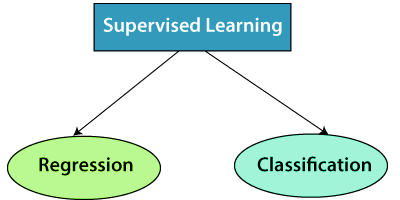
*Supervised Machine Learning*

Now, after training, we test our model using the test set, and the task of the model is to identify the shape.

The machine is already trained on all types of shapes, and when it finds a new shape, it classifies the shape on the bases of a number of sides, and predicts the output.

**Types of supervised Machine learning Algorithms:**

Supervised learning can be further divided into two types of problems:



**1. Regression**

Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc. Below are some popular Regression algorithms which come under supervised learning:

* Linear Regression
* Regression Trees
* Non-Linear Regression
* Bayesian Linear Regression
* Polynomial Regression

**2. Classification**

Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc.

* Spam Filtering,
* Random Forest
* Decision Trees
* Logistic Regression
* Support vector Machines

*Supervised Machine Learning*

**Advantages of Supervised learning:**

* With the help of supervised learning, the model can predict the output on the basis of prior experiences.
* In supervised learning, we can have an exact idea about the classes of objects.
* Supervised learning model helps us to solve various real-world problems such as fraud detection, spam filtering, etc.

**Disadvantages of supervised learning:**

* Supervised learning models are not suitable for handling the complex tasks.
* Supervised learning cannot predict the correct output if the test data is different from the training dataset.
* Training required lots of computation times.
* In supervised learning, we need enough knowledge about the classes of object.

**Topics Covered:**

* Machine Learning with Python
* Supervised- Linear Regression
* Supervised- Logistic Regression
* Supervised- Decision Tree
* Supervised- Support Vector Machine
* Supervised– K-Nearest Neighbours

**Learning Outcome:**

In this module, Introduction to supervised learning was given that how these libraries can be helpful in building up the prediction models using different algorithms for different business use cases, improving the models, etc.

*Un-Supervised Machine Learning*

**What is Un-Supervised Machine Learning?**

* As the name suggests, unsupervised learning is a machine learning technique in which models are not supervised using training dataset. Instead, models itself find the hidden patterns and insights from the given data. It can be compared to learning which takes place in the human brain while learning new things. It can be defined as:
* Unsupervised learning is a type of machine learning in which models are trained using unlabeled dataset and are allowed to act on that data without any supervision.
* Unsupervised learning cannot be directly applied to a regression or classification problem because unlike supervised learning, we have the input data but no corresponding output data. The goal of unsupervised learning is to find the underlying structure of dataset, group that data according to similarities, and represent that dataset in a compressed format.

**How Un-Supervised Learning Works?**

Working of unsupervised learning can be understood by the below diagram:



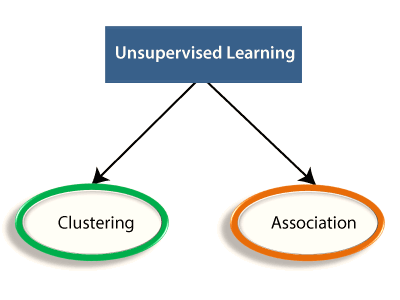
Here, we have taken an unlabeled input data, which means it is not categorized and corresponding outputs are also not given. Now, this unlabeled input data is fed to the machine learning model in order to train it. Firstly, it will interpret the raw data to find the hidden patterns from the data and then will apply suitable algorithms such as k-means clustering, Decision tree, etc.

Once it applies the suitable algorithm, the algorithm divides the data objects into groups according to the similarities and difference between the objects.

*Un-Supervised Machine Learning*

**Types of Unsupervised Learning Algorithm:**

The unsupervised learning algorithm can be further categorized into two types of problems:



1. **Clustering:**

Clustering is a method of grouping the objects into clusters such that objects with most similarities remains into a group and has less or no similarities with the objects of another group. Cluster analysis finds the commonalities between the data objects and categorizes them as per the presence and absence of those commonalities.

1. **Association:**

An association rule is an unsupervised learning method which is used for finding the relationships between variables in the large database. It determines the set of items that occurs together in the dataset. Association rule makes marketing strategy more effective. Such as people who buy X item (suppose a bread) are also tend to purchase Y (Butter/Jam) item. A typical example of Association rule is Market Basket Analysis.

**Advantages of Unsupervised Learning**

* Unsupervised learning is used for more complex tasks as compared to supervised learning because, in unsupervised learning, we don't have labeled input data.
* Unsupervised learning is preferable as it is easy to get unlabeled data in comparison to labeled data.

**Disadvantages of Unsupervised Learning**

* Unsupervised learning is intrinsically more difficult than supervised learning as it does not have corresponding output.
* The result of the unsupervised learning algorithm might be less accurate as input data is not labeled, and algorithms do not know the exact output in advance.

*Supervised vs Un-Supervised Learning*

|  |  |
| --- | --- |
| Supervised Learning | Un-Supervised Learning |
| Supervised learning algorithms are trained using labeled data. | Unsupervised learning algorithms are trained using unlabeled data. |
| Supervised learning model takes direct feedback to check if it is predicting correct output or not. | Unsupervised learning model does not take any feedback. |
| Supervised learning model predicts the output. | Unsupervised learning model finds the hidden patterns in data. |
| In supervised learning, input data is provided to the model along with the output. | In unsupervised learning, only input data is provided to the model. |
| The goal of supervised learning is to train the model so that it can predict the output when it is given new data. | The goal of unsupervised learning is to find the hidden patterns and useful insights from the unknown dataset |
| Supervised learning can be categorized in Classification and Regression problems. | Unsupervised Learning can be classified in Clustering and Associations problems. |
| Supervised learning can be used for those cases where we know the input as well as corresponding outputs. | Unsupervised learning can be used for those cases where we have only input data and no corresponding output data. |
| It includes various algorithms such as Linear Regression, Logistic Regression, Support Vector Machine, Multi-class Classification, Decision tree, Bayesian Logic, etc. | It includes various algorithms such as Clustering, KNN, and Apriori algorithm. |

## **Computer Vision**

|  |
| --- |
| Computer vision is simply the process of perceiving the images and videos available in the digital formats. In Machine Learning (ML) and AI – Computer vision is used to train the model to recognize certain patterns and store the data into their artificial memory to utilize the same for predicting the results in real-life use.  The main purpose of using computer vision technology in ML and AI is to create a model that can work itself without human intervention. The whole process involves methods of acquiring the data, processing, analyzing, and understanding the digital images to utilize the same in the real-world scenario.  *CV_Venn.jpg* |

**Computer vision** is a field of **artificial intelligence** (**AI**) that enables **computers** and systems to derive meaningful information from digital images, videos and other visual inputs — and take actions or make recommendations based on that information.

*List of References*

* [https://matplotlib.org/stable/gallery/index.html#](https://matplotlib.org/stable/gallery/index.html)
* <https://www.tableau.com/learn/articles/data-visualization>
* <https://pandas.pydata.org/>
* <https://numpy.org/>
* <https://www.javatpoint.com/machine-learning>
* <https://www.python.org/>
* <https://www.hexnbit.com/>
* <https://www.tevatrontech.com/>
* <https://scikit-learn.org/stable/>