EXPERIMENT - 0

AIM: To study about various tools and techniques in Machine Learning.

THEORY:

**Machine Learning**

Arthur Samuel first introduced the term "machine learning" in 1959. Machine learning is a branch of artificial intelligence (AI) and computer science that focuses on using data and algorithms to imitate the way that humans learn, gradually improving its accuracy. It evolved from the study of pattern recognition in Artificial Intelligence.

Machine learning, deep learning, and neural networks are all sub-fields of artificial intelligence. However, deep learning is a sub-field of machine learning, and neural networks are a sub-field of deep learning.

**Features of Machine Learning:**

1. Machine learning uses data to detect various patterns in a given dataset.
2. It can learn from past data and improve automatically.
3. It is a data-driven technology.
4. Machine learning is much similar to data mining as it also deals with the huge amount of the data.

**Need for Machine Learning**

The need for machine learning is increasing day by day. The reason behind the need for machine learning is that it is capable of doing tasks that are too complex for a person to implement directly. As a human, we have some limitations as we cannot access the huge amount of data manually, so for this, we need some computer systems and here comes the machine learning to make things easy for us.

We can train machine learning algorithms by providing them the huge amount of data and let them explore the data, construct the models, and predict the required output automatically. The performance of the machine learning algorithm depends on the amount of data.

Currently, machine learning is used in self-driving cars, cyber fraud detection, face recognition, and friend suggestion by Facebook, etc. Various top companies such as Netflix and Amazon have build machine learning models that are using a vast amount of data to analyze the user interest and recommend product accordingly.

**Machine Learning Algorithms (Techniques)**

**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.

**Types: -**

Supervised learning can be further divided into two types:

* **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
* **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

**Unsupervised Machine Learning**

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. 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.

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.

**Types: -**

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

* **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.

* **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.

**Semi-Supervised Machine Learning**

The most basic disadvantage of any Supervised Learning algorithm is that the dataset has to be hand-labeled either by a Machine Learning Engineer or a Data Scientist. This is a very costly process, especially when dealing with large volumes of data. The most basic disadvantage of any Unsupervised Learning is that its application spectrum is limited.

To counter these disadvantages, the concept of Semi-Supervised Learning was introduced. In this type of learning, the algorithm is trained upon a combination of labeled and unlabelled data. Typically, this combination will contain a very small amount of labeled data and a very large amount of unlabelled data. The basic procedure involved is that first, the programmer will cluster similar data using an unsupervised learning algorithm and then use the existing labeled data to label the rest of the unlabelled data.

[

**Reinforcement Machine Learning**

Reinforcement Learning is a feedback-based Machine learning technique in which an agent learns to behave in an environment by performing the actions and seeing the results of actions. For each good action, the agent gets positive feedback, and for each bad action, the agent gets negative feedback or penalty.

In Reinforcement Learning, the agent learns automatically using feedbacks without any labeled data, unlike supervised learning. Since there is no labeled data, so the agent is bound to learn by its experience only.

The agent interacts with the environment and explores it by itself. The primary goal of an agent in reinforcement learning is to improve the performance by getting the maximum positive rewards.

The agent learns with the process of hit and trial, and based on the experience, it learns to perform the task in a better way.

It is a core part of Artificial intelligence, and all AI agent works on the concept of reinforcement learning. Here we do not need to pre-program the agent, as it learns from its own experience without any human intervention.

**Machine Learning Tools**

**Popular languages for machine learning**

**Python**

Python is a general-purpose programming language, so it can be used for many things. Python is used for web development, AI, machine learning, operating systems, mobile application development, and video games.

Python is one of the most popular ML languages. It is flexible and easy to learn. Python is an old language, and it has a rich set of libraries and frameworks that are regularly updated. These resources help to develop machine learning solutions faster thanks to sets of pre-programmed elements.

**R**

R language was created in order to work with statistical analysis. It has powerful visualization capabilities. If you want to work with R, you will need special packages. Ubuntu Pit has collected 20 best R packages for you to use in ML.

**Machine learning frameworks and tools**

**TensorFlow**

TensorFlow is an open-source framework developed by Google that comes in handy for large-scale as well as numerical ML. It is a blender of machine learning as well as neural network models. Moreover, it is also a good friend of Python.

The most prominent feature of TensorFlow is, it runs on CPU and GPU as well. Natural language processing, Image classification are the ones who implement this tool.

**Pytorch**

Pytorch is a deep learning framework. It is very fast as well as flexible to use. This is because Pytorch has a good command over the GPU. It is one of the most important tools of machine learning because it is used in the most vital aspects of ML which includes building deep neural networks and tensor calculations. Pytorch is completely based on Python. Along with this, it is the best alternative to NumPy.

**Jupyter Notebook**

Jupyter notebook is one of the most widely used machine learning tools among all. It is a very fast processing as well as an efficient platform. Moreover, it supports three languages viz. Julia, R, Python.

Thus the name of Jupyter is formed by the combination of these three programming languages. Jupyter Notebook allows the user to store and share the live code in the form of notebooks. One can also access it through a GUI. For example, winpython navigator, anaconda navigator, etc.

**Scikit-Learn**

Scikit-Learn is an open-source machine learning package. It is a unified platform as it is used for multiple purposes. It assists in regression, clustering, classification, dimensionality reduction, and preprocessing. Scikit-Learn is built on top of the three main Python libraries viz. NumPy, Matplotlib, and SciPy. Along with this, it will also help you with testing as well as training your models.

**Machine Learning Platforms**

**Google Colab**

Google Colaboratory is a free online cloud-based Jupyter notebook environment that allows us to train our machine learning and deep learning models on CPUs, GPUs, and TPUs.

It does not matter which computer you have, what it’s configuration is, and how ancient it might be. We can still use Google Colab! All we need is a Google account and a web browser. We also get access to GPUs like Tesla K80 and even a TPU, for free. TPUs are much more expensive than a GPU, and we can use it for free on Colab.

**Amazon Web Services**

AWS is a platform that provides artificial intelligence and machine learning services to developers. It is possible to choose one of the pre-trained AI services to work with computer vision, language recognition, speech generation, build recommender system and prediction models.

Using Amazon SageMaker, you can quickly create, train, and deploy scalable machine learning models, or create custom models that support all the popular open-source ML platforms.

**Microsoft Azure**

Azure Machine Learning Studio allows developers who don’t have experience in machine learning to use the drag-and-drop functionality. This platform allows one to build solutions directly “on the cloud” and easily create BI applications regardless of the quality of the data.

Microsoft also offers Cortana Intelligence, a tool that allows you to fully manage big data and analytics and transform data into meaningful information and subsequent actions.

RESULT:

We studied about various tools and techniques in Machine Learning.