



Machine Learning: Life Cycle and its Techniques

B. Anusuya Devi¹, K. Pavithra²

¹Department of Computer Science, R.B. Gothi Jain College for Women, Redhills, Chennai- 600 052

ABSTRACT

Machine learning has arisen in the multidisciplinary agriculture technologies sector, together with big data technologies and high-performance computing, to create new prospects for data intensive science. The scientific study of algorithms and statistical models that computer systems employ to complete a certain task without being explicitly taught is known as machine learning (ML). Many applications that we use on a daily basis involve learning algorithms. When you use a web search engine like Google to search the internet, one of the reasons it works so effectively is because it uses a learning algorithm to rank web sites. A quick assessment and future outlook on the broad uses of machine learning algorithms are presented in this study.

Keywords: ML, Features, Supervised, unsupervised, reinforcement

1. Introduction

We begin with a few instances of machine teaching in order to contrast machine teaching and machine learning, namely passive and active learning in supervised learning. Artificial intelligence has an area called machine learning (AI). The goal of machine learning is to comprehend the structure of data and fit that data into models that people can comprehend and use. Machine learning is a branch of computer science that is distinct from standard computational methods. Algorithms are sets of clearly programmed instructions used by computers to calculate or solve problems in traditional computing. Machine learning is a field that is constantly evolving. As a result, there are some things to keep in mind when working with machine learning approaches or analyzing data. Tasks are often categorized into major groups in machine learning. These classifications are based on how learning is received and how the system is given feedback on the learning. Unsupervised learning, which provides the algorithm with no labelled data in order to allow it to find structure within its input data, and supervised learning, which trains algorithms based on example input and output data that is labelled by humans, are two of the most widely used machine learning methods..

2. Types of Machine learning

Supervised Learning: In supervised learning, the computer is given example inputs and their expected outputs are labelled. The goal of this method is for the algorithm to "learn" by comparing its real output to the "learned" outputs in order to identify faults and adjust the model accordingly. As a result, supervised learning employs patterns to predict label values on unlabeled data.

Based on the labelled data it was trained on, a classification algorithm seeks to sort inputs into a set of categories or classes. Binary classifications, such as filtering email into spam or non-spam and categorising customer feedback as good or negative, can be done with classification algorithms. Another classification challenge tackled by supervised learning is feature recognition, such as recognising handwritten letters and numbers or classifying medications into multiple

categories. In regression tasks, the model is expected to provide a numerical connection between the input and output data. Regression models can be used to estimate real estate prices based on zip code, predict click rates in online ads depending on time of day, or figure out how much customers are willing to pay.

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.

Unsupervised learning

This sort of algorithm restructures data into new features that may indicate a class or a new set of uncorrelated values. They're great for giving humans fresh insights into the meaning of data and giving supervised machine learning algorithms new inputs. Clustering is a common topic of unsupervised learning. Clustering is the process of grouping objects or data points that are similar but not identical to those in other clusters. Distinct methods for clustering can be used by machine learning engineers and data scientists, with the algorithms themselves falling into different categories based on how they work.

Advantages of Unsupervised Learning:

- Unsupervised learning is used for more multifaceted tasks as compared to supervised learning because, in unsupervised learning, we don't have labelled input data.
- Unsupervised learning is preferable as it is easy to get unlabelled data in comparison to labelled data.

Disadvantages of Unsupervised Learning:

- Unsupervised learning is essentially 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 labelled, and algorithms do not know the exact output in advance.

3. Machine learning Life cycle

Machine learning has given computers the ability to learn on their own without having to be explicitly programmed. However, how does a machine learning system function? As a result, the machine learning life cycle can be used to explain it. Machine learning life cycle is a cyclic process to build an efficient machine learning project. The life cycle's primary goal is to find a solution to the problem or project.

3.1. Gathering Data:

The first phase in the machine learning life cycle is data collection. This step's purpose is to identify and collect all data-related issues. We must first identify the various data sources, as data might come from a variety of places, including files, databases, the internet, and mobile devices.

3.2. Data preparation:

After gathering data, we must prepare it for future processing. Data preparation is the stage in which we organise and prepare our data for use in machine learning training. This step can be further broken down into two steps:

- **Data exploration**
- **Data pre-processing**

3.3. Data Wrangling:

The process of cleaning and turning raw data into a usable format is known as data wrangling. It is the process of cleaning the data, selecting the variable to utilise, and changing the data into a suitable format for analysis in the following phase. It is one of the most crucial steps in the entire procedure. To overcome the quality issues, data must be cleaned.

3.4. Data Analysis:

Now the cleaned and prepared data is passed on to the analysis step. This step involves:

- **Selection of analytical techniques**
- **Building models**
- **Review the result**

3.5. Train Model:

The following stage is to train the model, in which we increase our model's performance in order to achieve a better solution to the problem. Various machine learning techniques are used to train the model utilising datasets. A model must be trained in order for it to comprehend the numerous patterns, rules, and features.

3.6. Test Model:

We test our machine learning model once it has been trained on a specific dataset. We check the correctness of our model in this stage by feeding it a test dataset. The % correctness of the model is determined by testing it against the project or problem requirements.

3.7. Deployment:

The final phase in the machine learning life cycle is deployment, which involves putting the model into action in a real-world system. We deploy the model in the real system if the above-prepared model produces an accurate output that meets our requirements at a reasonable pace. However, before launching the project, we'll see if it improves performance by utilising accessible data.

4. Conclusion

There are two types of machine learning: supervised and unsupervised. Choose Supervised Learning if you have less data and clearly labelled data for training. For huge data sets, unsupervised learning would provide greater performance and results. Deep learning techniques should be used if you have a large data collection readily available. Finally, when it came to creating your own machine learning models, you investigated the various programming languages, IDEs, and platforms available. The next step is to begin learning and practising each machine learning approach.

REFERENCES

- 1) <https://www.digitalocean.com/community/tutorials/an-introduction-to-machine-learning>
- 2) <https://www.techtarget.com/searchenterpriseai/definition/supervised-learning>
- 3) <https://www.techtarget.com/searchenterpriseai/definition/unsupervised-learning>
- 4) <https://www.javatpoint.com/machine-learning>

-
- 5) M Praneesh and Jaya R Kumar. Article: Novel Approach for Color based Comic Image Segmentation for Extraction of Text using Modify Fuzzy Possiblistic C-Means Clustering Algorithm. IJCA Special Issue on Information Processing and Remote Computing IPRC(1):16-18, August 2012. Published by Foundation of Computer Science, New York, USA.
 - 6) <https://data-flair.training/blogs/advantages-and-disadvantages-of-machine-learning/>