**IOSR Journal of Engineering (IOSRJEN)** **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 11, Issue 12, December. 2021, || Series -I || PP 56-58

# Machine Learning Types and Techniques

**OBINNAYA OMANKWU;**
**NGENE, CHIGOZIE CHIDIMMA**
**AND**
**UBAH, VALENTINE**

**Department of Computer Science,**
**Michael Okpara University of Agriculture,Umudike,Abia State**

**Department of Computer Science,**
**Nnamdi Azikiwe University,Awka, Anambra State**

**Department of Computer Science,**
**Michael Okpara University of Agriculture,Umudike,Abia State**

*Received 04 December 2021; Accepted 18 December 2021*

**ABSTRACT**
In the current age of the Fourth Industrial Revolution, the digital world has a wealth of data, such as Internet of Things (IoT) data, cyber security data, mobile data, business data, social media data, health data, etc. To intelligently analyze these data and develop the corresponding smart and automated applications, the knowledge of artificial intelligence (AI), particularly, machine learning (ML) is the key. Various types of machine learning algorithms such as supervised, unsupervised, semi-supervised, and reinforcement learning exist in the area. Besides, the deep learning, which is part of a broader family of machine learning methods, can intelligently analyze the data on a large scale. In this paper, we present a comprehensive view on these machine learning algorithms that can be applied to enhance the intelligence and the capabilities of an application. Thus, this study's key contribution is explaining the principles of different machine learning techniques and their applicability in various real-world application domains, such as cyber security systems, smart cities, healthcare, e -commerce, agriculture, and many more.

**KEYWORDS:** Machine learning · Deep learning · Artificial intelligence · Data science ·

## I. INTRODUCTION

We live in the age of data, where everything around us is connected to a data source, and everything in our lives is digitally recorded (Cao, 2017). For instance, the current electronic world has a wealth of various kinds of data, such as the Internet of Things (IoT) data, cyber security data, smart city data, business data, smart phone data, social media data, health data,, and many more. The data can be structured, semi-structured, or unstructured, discussed briefly in Sect. "Types of Real -World Data and Machine Learning Techniques", which is increasing day -by-day. Extracting insights from these data can be used to build various intelligent applications in the relevant domains. For instance, to build a data-driven automated and intelligent cyber security system, the relevant cyber security data can be used (Sarker etal.2020); to build personalized context-aware smart mobile applications, the relevant mobile data can be used (Sarker etal.2020) and so on. Thus, the data management tools and techniques having the capability of extracting insights or useful knowledge from the data in a timely and intelligent way is urgently needed, on which the real-world applications are based.

Artificial intelligence (AI), particularly, machine learning (ML) have grown rapidly in recent years in the context of data analysis and computing that typically allows the applications to function in an intelligent manner (Sarker etal.2021),."ML usually provides systems with the ability to learn and enhance from experience automatically without being specifically programmed and is generally referred to as the most popular latest technologies in the fourth industrial revolution (4IR or Industry 4.0) (Sarker etal.2020),."Industry 4.0" [114] is typically the ongoing automation of conventional manufacturing and industrial practices, including exploratory data processing, using new smart technologies such as machine learning automation. Thus, to intelligently analyze these data and to develop the corresponding real-world applications, machine learning algorithms are the key. The learning algorithms can be categorized into four major types, such as supervised, unsupervised, semi-supervised, and reinforcement learning in the area (Mohammed, 2016) discussed briefly in Sect. "Types of Real-World Data and Machine Learning Techniques". The popularity of these approaches to learning is increasing day-by-day, based on data collected from Google Trends over the last five years.

**Types of Real World Data**

Machine learning algorithms typically consume and process data to learn the related patterns about individuals, business processes, transactions, events, and so on. In the following, we discuss various types of real-world data as well as categories of machine learning algorithms.

Usually, the availability of data is considered as the key to construct a machine learning model or data-driven real-world systems (Sarker etal.2021). Data can be of various forms, such as structured, semi-structured, or unstructured (Sarker etal.2021), Besides, the "metadata" is another type that typically represents data about the data. In the following, we briefly discuss these types of data.

–*Structured:* It has a well -defined structure, conforms to a data model following a standard order, which is highly organized and easily accessed, and used by an entity or a computer program. In well-defined schemes, such as relational databases, structured data are typically stored, i.e., in a tabular format. For instance, names, dates, addresses, credit card numbers, stock information, geolocation, etc. are examples of structured data.

–*Unstructured:* On the other hand, there is no pre-defined format or organization for unstructured data, making it much more difficult to capture, process, and analyze, mostly containing text and multimedia material. For example, sensor data, emails, blog entries, wikis, and word processing documents, PDF files, audio files, videos, images, presentations, web pages, and many other types of business documents can be considered as unstructured data.

–*Semi-structured:* Semi-structured data are not stored in a relational database like the structured data mentioned above, but it does have certain organizational properties that make it easier to analyze. HTML, XML, JSON documents, NoSQL databases, etc., are some examples of semi-structured data.

–*Metadata:* It is not the normal form of data, but "data about data". The primary difference between "data" and "metadata" is that data are simply the material that can classify, measure, or even document something relative to an organization's data properties. On the other hand, metadata describes the relevant data information, giving it more significance for data users. A basic example of a document's metadata might be the author, file size, date generated by the document, keywords to define the document, etc. other types of business documents can be considered as unstructured data.

–*Semi-structured:* Semi-structured data are not stored in a relational database like the structured data mentioned above, but it does have certain organizational proper-ties that make it easier to analyze. HTML, XML, JSON documents, NoSQL databases, etc., are some examples of semi-structured data.

–*Metadata:* It is not the normal form of data, but "data about data". The primary difference between "data" and "metadata" is that data are simply the material that can classify, measure, or even document something relative to an organization's data properties. On the other hand, metadata describes the relevant data information, giving it more significance for data users. A basic example of a document's metadata might be the author, file size, date generated by the document, keywords to define the document, etc.

**Types of Machine Learning Techniques**

Machine Learning algorithms are mainly divided into four categories: Supervised learning, unsupervised learning, Semi-supervised learning, and Reinforcement learning (Sarker etal.2021), In the following, we briefly discuss each type of learning technique with the scope of their applicability to solve real-world problems.

–*Supervised:* Supervised learning is typically the task of machine learning to learn a function that maps an input to an output based on sample input-output pairs (Han,2017). It uses labeled training data and a collection of training examples to infer a function. Supervised learning is carried out when certain goals are identified to be accomplished from a certain set of inputs (Sarker etal.2021), i.e., a *task-driven approach*. The most common supervised tasks are "classification" that separates the data, and "regression" that fits the data. For instance, predicting the class label or sentiment of a piece of text, like a tweet or a product review, i.e., text classification is an example of supervised learning.

– *Unsupervised:* Unsupervised learning analyzes unlabeled datasets without the need for human interference, i.e., a *data -driven process* (Han,2017). This is widely used for extracting generative features, identifying meaningful trends and structures, groupings in results, and exploratory purposes. The most common

unsupervised learning tasks are clustering, density estimation, feature learning, dimensionality reduction, finding association rules, anomaly detection, etc.

– *Semi-supervised:* Semi-supervised learning can be defined as a *hybridization* of the above-mentioned supervised and unsupervised methods, as it operates on both labeled and unlabeled data (Sarker etal.2021), Thus, it falls between learning "without supervision" and learning "with supervision". In the real world, labeled data could be rare in several contexts, and unlabeled data are numerous, where semi-supervised learning is useful (Mohammed, 2016). The ultimate goal of a semi-supervised learning model is to provide a better outcome for prediction than that produced using the labeled data alone from the model. Some application areas where semi-supervised learning is used include machine translation, fraud detection, labeling data and text classification.

– *Reinforcement:* Reinforcement learning is a type of machine learning algorithm that enables software agents and machines to automatically evaluate the opti-mal behavior in a particular context or environment to improve its efficiency (Mohammed, 2016), i.e., an *environment-driven approach*. This type of learning is based on reward or penalty, and its ultimate goal is to use insights obtained from environmental activists to take action to increase the reward or minimize the risk (Mohammed, 2016). It is a powerful tool for training AI models that can help increase automation or optimize the operational efficiency of sophisticated systems such as robotics, autonomous driving tasks, manufacturing and supply chain logistics, however, not preferable to use it for solving the basic or straightforward problems.

## II. CONCLUSION

In this paper, we have conducted a comprehensive overview of machine learning algorithms for intelligent data analysis and applications. According to our goal, we have briefly discussed how various types of machine learning methods can be used for making solutions to various real-world issues. A successful machine learning model depends on both the data and the performance of the learning algorithms. The sophisticated learning algorithms then need to be trained through the collected real-world data and knowledge related to the target application before the system can assist with intelligent decision-making. We also discussed several popular application areas based on machine learning techniques to highlight their applicability in various real-world issues. Finally, we have summarized and discussed the challenges faced and the potential research opportunities and future directions in the area. Therefore, the challenges that are identified create promising research opportunities in the field which must be addressed with effective solutions in various application areas. Overall, we believe that our study on machine learning -based solutions opens up a promising direction and can be used as a reference guide for potential research and applications for both academia and industry professionals as well as for decision-makers, from a technical point of view.

## REFERENCES

[1]. Cao L. Data science: a comprehensive overview. ACM Comput Surv (CSUR). 2017;50(3):43.
[2]. Google trends. In https://trends.google.com/trends/, 2019.
[3]. Mohammed M, Khan MB, Bashier Mohammed BE. Machine learning: algorithms and applications. CRC Press; 2016.
[4]. Sarker IH. Ai-driven cyber security: an overview, security intelligence modeling and research directions. SN Comput Sci. 2021.
[5]. Sarker IH, Hoque MM, MdK Uddin, Tawfeeq A. Mobile data science and intelligent apps: concepts, ai-based modeling and research directions. Mob Netw Appl, pages 1–19, 2020.
[6]. Sarker IH, Kayes ASM, Badsha S, Alqahtani H, Watters P, Ng A. Cyber security data science: an overview from machine learning perspective. J Big Data. 2020;7(1):1–29.
[7]. Han J, Pei J, Kamber M. Data mining: concepts and techniques. Amsterdam: Elsevier; 2011.