Real-World Application Of Data Mining By Using Artificial Intelligence Techniques

1Dr. CH N Santhosh Kumar,

Associate Professor, Dept. Of Computer Science and Engineering

GITAM (Deemed to be University), Hyderabad, India

[santhosh.ph10@gmail.com](mailto:santhosh.ph10@gmail.com)

2Mrs. N Lakshmi Deepthi,

Assistant Professor, Dept. Of Computer Science and Engineering

RISHI M.S Institute of Engineering and Technology for Women, Hyderabad, India

[lakshmideepthi16@gmail.com](mailto:lakshmideepthi16@gmail.com)

**ABSTRACT- Data mining and artificial intelligence have been utilized in various industries to handle classification, segmentation, association, diagnostic, and prediction challenges. It is the general goal of this special issue to foster a dialogue among scholars actively engaged in developing algorithms and applications. Several machine learning techniques may improve an application's intelligence and capabilities. In addition, the artificial neural network, which may be used to evaluate data intelligently, is part of a more prominent family of machine learning technologies. Therefore, a fundamental contribution of this study is to describe the concepts of numerous machine learning techniques and their use in diverse real-world application fields like cyber security systems, smart cities, healthcare, e-commerce, and many more. We also discuss the problems and other research avenues that may arise from our work.**

**Keywords -- Data mining, Artificial intelligence, cyber security.**

1. INTRODUCTION

Modern society is data-driven, with virtually every aspect of our lives documented in digital format. To give one example, there is a wide range of data types in the electronic world, including IoT and Cyber Security data as well as corporate and social media information, and health and COVID-19 datasets. Structured, semi-structured, or unstructured data are all examples of real-world data that can be used to train machine learning models. Insights gained from these data can be applied to various innovative applications in the relevant fields. Examples include building an automated and intelligent Cyber Security system, creating tailored, contextually aware, innovative mobile applications, etc. The applicable data can be used in each of these examples. As a result, real-world applications rely on data management tools and processes that can quickly and intelligently extract insights or usable knowledge from data.

In recent years, there has been a dramatic increase in the use of artificial intelligence (AI) and machine learning (ML) in data analysis and computing. It is generally accepted that machine learning (ML) is one of the most popular new technologies in the fourth industrial revolution since it allows systems to learn and improve from experience without being explicitly programmed (4IR or Industry 4.0). An example of "Industry 4.0" would be new smart technologies, such as machine learning automation, continuously automating traditional manufacturing and industrial procedures, including exploratory data processing. Thus, machine learning algorithms are essential to analyze these data and generate the appropriate real-world applications. Section "Types of Real-World Data and Machine Learning Techniques" briefly discusses the four primary types of learning algorithms: supervised, unsupervised, semi-supervised, and reinforcement learning. According to Google Trends data collected over the past five years, these learning methods have become increasingly popular. The X-axis shows the dates, and the y-axis shows the associated popularity score from 0 (lowest) to 100 (highest) within the period. Popularity indication values for these learning styles were down in 2015, but they're steadily rising for the year. Because of these figures, we decided to do research on machine learning in this paper, which has the potential to play a crucial role in the actual world as part of Industry 4.0 automation.

There are numerous applications of artificial intelligence to solve classification, planning

, prediction, optimization, diagnosis, and computation issues in a wide range of industries, as well as collecting and analyzing customer information, gaining insights into what customers want or need, and then taking action on those insights. For academics working on algorithms and artificial intelligence applications, this special issue aims to reflect the most recent developments in this field and provide advanced knowledge. As a result, data mining and big data are ubiquitous, and it is critical to safeguard the massive amounts of data generated to ensure that no vital information is lost. In many cases, this type of data is processed utilizing artificial intelligence. Algorithms are at the heart of Artificial Intelligence and its various subfields, such as Machine Learning, Deep Learning, and Neutral Networks. These algorithms are applied to massive amounts of data (known as "Big Data") to uncover patterns, trends, and predictions. When it comes to big data analytics, artificial intelligence (AI) does jobs faster than anyone could have imagined. Using a stepwise scientific method, an artificial neural network was built to optimize a criterion known as the learning rule. For these networks, input/output training information is critical, as it communicates the data necessary for optimal operation. Additionally, the nonlinear nature of neural networks results in component portions of a flexible system that can be used in many ways. According to this definition, computers' intelligence is compared to the intelligence of people in general, concerned about computers' performance, usually seen as intelligent. Because of this, we can use artificial intelligence to mine large datasets and analyze them. It is used to analyze and evaluate information. This applies to any data with a structure to be used in an algorithm. The rule will be divided into ordered or parallel processes and executed on the computer's processor or graphics processing unit. The speed of the rule is affected by the choice of which of them to use. As a general rule, the end outcome is associated with several outputs. Multiple outputs may result in a new definition. This is an entirely mechanical and computerized performance. It's the human interpretation of the context, the rule procedure, and the outcomes that are of value, not the information itself. A report, an analysis, or a prediction can all be classified as an associate in analysis. Whether or if a data miner is successful is entirely up to them. The pinnacle of artificial intelligence (AI) is the creation of a fake agent capable of acting or thinking independently, much like a human. Big Data is all about sifting through massive amounts of data to find patterns, trends, etc. They're connected because one can be used to enhance the other. For example, AI might be used to make these decisions instead of people when interpreting, refining, and acting on significant data analysis. As an alternative, AI could exploit Big Data for self-learning and making decisions. (Turban, Sharda, and colleagues, 2010)

1. TYPES OF REAL-WORLD DATA AND MACHINE LEARNING TECHNIQUES

These algorithms use data to learn patterns about individuals, corporate processes, transactional events, and so forth. Real-world data and machine-learning algorithms are covered in the following sections.

## 2.1 Types of Real-World Data

Data availability is typically seen as an essential factor in building a machine-learning model or data-driven real-world systems. Structured, semi-structured, and

Unstructured data are all types of data. The "metadata" is another sort of data that often contains information about the data itself. Data of this type is discussed in the following paragraphs.

Structured: An entity or a computer program can access and use it because it has a well-defined structure that adheres to a data model that follows a standard order. Structured data is often stored in a tabular manner in well-defined schemes, such as relational databases. Structured data includes names, dates, addresses, credit card numbers, stock information, and geolocation.

There is no predefined format or organization for unstructured data, making it challenging to capture, process, and analyze because it primarily consists of textual and multimedia content, making it more challenging to process and analyze. Unstructured data includes, but is not limited to, sensor data, emails, blogs, wikis, word processing documents, PDF files, audio files, videos, photos, presentations, web pages, and many other business documents.

For example, a semi-structured data set doesn't have to be kept in a relational database, but it has some organizational characteristics that make it easier to examine. Semi-structured data examples are HTML, XML, JSON, and NoSQL databases.

## 2.2 Types of Machine Learning Techniques

Supervised, unsupervised, semi-supervised, and reinforcement learning are the primary categories of machine-learning algorithms. As depicted in Fig. 1. In the following, we'll take a quick look at the various learning methods and how they could be used to address real-world issues.

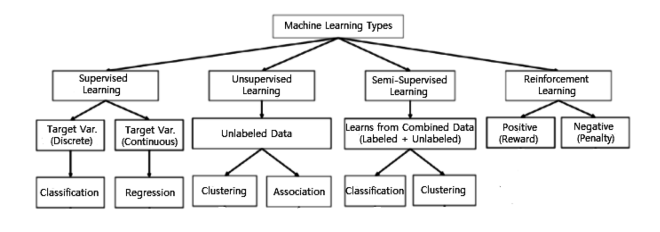


Fig. 1 Various types of machine learning techniques.

Supervised: An input-output mapping function is often learned using examples of input-output pairs in machine learning. To infer a process, it uses labeled training data and a database of training samples. It's a task-driven technique that uses supervised learning when certain goals are identified from a specific set of inputs. "Classification" and "regression" are the two most typical supervised jobs since they segregate the data and apply a mathematical function. Text classification is an example of supervised learning. For example, predicting a piece of text's class label or emotion is an example of supervised learning.

Unsupervised: Unsupervised learning is a data-driven technique that analyses datasets that a human expert has not labeled. This is commonly used to find relevant trends and structures, groups in results, and exploratory objectives. Unsupervised learning problems include clustering, density estimation, feature learning, dimensionality reduction, finding association rules, anomaly detection, etc.

– Semi-supervised: It is possible to think of semi-supervised learning as a combination of the supervised and unsupervised approaches discussed above because it uses both labels and unlabeled data. As a result, it's somewhere between "without supervision" and "with supervision" learning. Semi-supervised learning is effective when labeled data is few and unlabeled data is plentiful. The ultimate goal of a semi-supervised learning model is to deliver a better prediction output than that obtained by the model alone. Examples of semi-supervised learning applications are machine translation, fraud detection, data labeling, and text classification.

1. ARTIFICIAL NEURAL NETWORK AND DEEP LEARNING

Artificial neural network (ANN)-based machine learning approaches with representation learning are all part of a more prominent family of techniques known as deep learning (DL). A computational architecture is provided for deep learning by mixing input, hidden, and output layers to learn from data. When it comes to learning from a vast dataset, deep learning outperforms traditional machine learning algorithms in several ways. As more data is collected, deep understanding exceeds machine learning on a general basis, as seen in Figure 2. Data features and experimental setup can, however, influence this. Multilayer Perceptron (MLP), Convolutional Neural Networks (CNN, or ConvNet), and Long Short-Term Memory Recurrent Neural Networks are the most prevalent deep learning methods (LSTM-RNN).

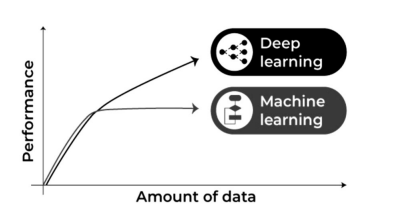
****

Fig. 2 Machine learning and deep learning performance generally with the amount of data.

Several deep learning methods are discussed below to develop effective data-driven models for various applications.

– Multilayer perceptrons (MLPs) are the underlying architecture of deep learning, also known as the feed-forward artificial neural network (MLP). MLPs are typically made up of a single input layer, one or more hidden layers, and a single output layer. All the nodes in a layer link to each other with a specific weight. It uses "Backpropagation," a neural network technique, to modify the weight values internally as the model is being built. The number of hidden layers, neurons, and iterations that can be tweaked can result in a computationally expensive model with the MLP, which is sensitive to scaling features.

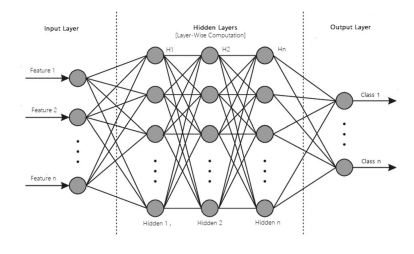


Fig. 2 A structure of an artificial neural network modeling with multiple processing layers

– CNN or ConvNet: The convolution neural network (CNN) improves on an essential ANN, including convolutional and pooled layers, as shown in Figure 4. Typical applications include image and video recognition, image processing and classification, medical image analysis, natural language processing, etc., since it takes advantage of the input data's two-dimensional (2D) structure. This means that while CNN has a higher processing overhead than ordinary neural networks, it is more potent since it automatically detects the most relevant properties. Advanced CNN-based deep learning models like AlexNet, Xception, Inception, Visual Geometry Group (VGG), and ResNet can be applied in the field.

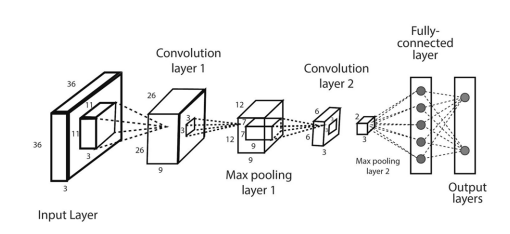
****

Fig. 3 An example of a convolutional neural network (CNN or ConvNet) including multiple convolutions and pooling layers.

Additionally, several alternative deep-learning techniques can be used for other purposes. For instance, unsupervised learning is used in the self-organizing map (SOM) to reduce the dimensionality of high-dimensional data. In addition to dimensionality reduction and feature extraction in unsupervised learning problems, the autoencoder (AE) is a popular learning tool. For example, RBMs can be used for dimensionality reduction, classification, regression, collaborative filtering, and feature learning. Backpropagation neural networks (BPN) and simple unsupervised networks like restricted Boltzmann machines (RBMs) form the core of a deep belief network (DBN) (BPNN). An adversarial network (GAN) is a deep learning network that can generate data closely resembling its source.

CONCLUSION

Machine learning techniques for intelligent data analysis and applications have been thoroughly analyzed in this study. We have briefly explored how various machine-learning approaches can be utilized to solve various real-world problems in accordance with our goal. Performance must be considered to build an excellent machine-learning model, data, and algorithm. Before the system can aid with intelligent decision-making, the system must be educated using the obtained real-world data and information about the target application. We also reviewed several common application areas to show how machine-learning techniques may be used to solve real-world problems. Lastly, we've highlighted and analyzed the obstacles, research possibilities, and prospects in the field.

##### REFERENCES:

1. Baker, R. S. and P. S. Inventado (2014). Educational data mining and learning analytics. Learning analytics, Springer: 61-75.
2. Hall, M., et al. (2019). "The WEKA data mining software: an update." ACM SIGKDD explorations newsletter11(1): 10-18.
3. Kitchin, R. (2014). "Big Data, new epistemologies, and paradigm shifts." Big data & society1(1): 2053951714528481.
4. Mobasher, B. (2017). Data mining for web personalization. The adaptive web, Springer: 90-135.
5. Provost, F. and T. Fawcett (2013). "Data science and its relationship to big data and data-driven decision making." Big data1(1): 51-59.
6. Purohit, H., et al. (2015). "Gender-based violence in 140 characters or fewer: A# BigData case study of Twitter." arXiv preprint arXiv:1503.02086.
7. Srinivas, K., et al. (2010). "Applications of data mining techniques in healthcare and prediction of heart attacks." International Journal on Computer Science and Engineering (IJCSE)2(02): 250-255.
8. Turban, E., et al. (2010). "Decision Support and Business Intelligence Systems (required)." Google Scholar.