Machine learning with big data to solve real-world problems

Maryam Rahmaty[1]

[1] *Department of Management, Chalous Branch, Islamic Azad University, Chalous, Iran;*

ARTICLE INFO	ABSTRACT
	Machine learning algorithms use big data to learn future trends and predict them for businesses. Machine learning can be very efficient for deciphering data in industries where understanding consumer patterns can lead to big improvements. The use of machine learning can be a giant leap for businesses and cannot simply be integrated as the top layer. This requires redefining workflow, architecture, data collection and storage, analytics, and other modules. The magnitude of the system overhaul should be assessed and clearly communicated to the appropriate stakeholders. The main focus of machine learning is to develop computer programs that can access data and use it to learn. The learning process starts with observations or data, to find a pattern in the data and make better decisions. The main goal of data analysis using machine learning is that it allows the computer to learn automatically without human intervention and help and can adjust its actions accordingly. Considering the many applications that data analysis has found in the real world, therefore, in this article, a review of the basic applications of machine learning as one of the tools of artificial intelligence has been done with an emphasis on big data analysis. The purpose of this article is to understand the dimensions, components and applications, and challenges of using machine learning in the real world.

1. Introduction

Machine learning is a science that forces computers to learn and behave like humans. As these systems interact with data, networks, and people, they automatically become smarter so that they can eventually solve or predict a practical issue in the world for us. Machine learning is a type of artificial intelligence that uses an algorithm to extract patterns from raw data (Nahr et al., 2021). The main focus of machine learning is to enable computer systems to learn from experience without human intervention or being explicitly programmed. There are many different algorithms for machine learning, and hundreds of new algorithms are produced every day, and they are usually classified by learning style (such as supervised learning, unsupervised learning, and semi-supervised learning) or according to their similarity (Nozari et al., 2021). They are grouped in form and function (such as classification, regression, decision tree,

[1] Corresponding Author: Rahmaty.Maryam61@iauc.ac.ir

grouping, deep learning, etc.). Regardless of learning style or performance, all machine learning algorithms are:

- Representation: A set of classifiers or a language that the computer understands.
- Evaluation: Also known as objective performance/scoring.
- Optimization: search method; Often the classifier with the highest score.

The fundamental goal of machine learning algorithms is to generalize learnings beyond the trained examples, that is, to successfully interpret the data.

In machine learning, instead of programming everything, data is fed to a generic algorithm, and it is this algorithm that builds its own logic based on the data it is fed. Machine learning has various methods, including supervised learning, unsupervised learning, and reinforcement learning. Algorithms used in machine learning are among these three categories. Therefore, machine learning emerged as a new capability for computers. Today, this science is used in various areas of technology, and its use has increased so much that people are often unaware of its existence in their daily tools and equipment (Tavakkoli-Moghaddam et al., 2022).

Machine learning can help companies understand their customers on a deeper level. By collecting customer data and relating it to behaviors over time, machine learning algorithms can help teams tailor product development and marketing initiatives to customer demand (Nozari and Nahr, 2022). The thing to note here is that machine learning algorithms can learn from their past experiences just like humans. When new data is fed to them, these algorithms learn, change, and grow without ever having to change the code. The results obtained may not be very accurate at first, but the machine learning algorithm can use its output data to improve the results in the future (Nozari et al., 2022).

It is very difficult to imagine any industrial activity without the use of machine learning or artificial intelligence. Machine learning is important because of its wide range of applications and its incredible ability to adapt and provide fast, effective, and optimal solutions to complex problems (Shayan Nia and Mirataollahi Olya, 2021). Machine learning has made it possible to quickly and automatically generate models that can analyze larger and more complex data and provide faster and more accurate results, even at a very large scale. By creating accurate models, there is a greater chance of identifying profitable opportunities or avoiding unknown risks (Aliahmadi et al., 2022).

Machine learning algorithms use various techniques to manage large volumes of complex data for decision-making. These algorithms do the job of learning from data with specific inputs given to the machine. Understanding how these algorithms and machine learning systems work is very important to know how to use them in the future. For these reasons, in this book, various types of machine learning methods and their applications in the real world are examined and analyzed.

2. Types of machine learning

Machine learning is often classified based on how the algorithm learns. There are four main approaches: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Each of these methods has a specific purpose. The type of algorithm that data scientists choose depends on the type of data they want to predict. Among these, two categories of supervised and unsupervised learning are considered the main categories (Aliahmadi and Nozari , 2023). These two main types are shown in Figure 1.
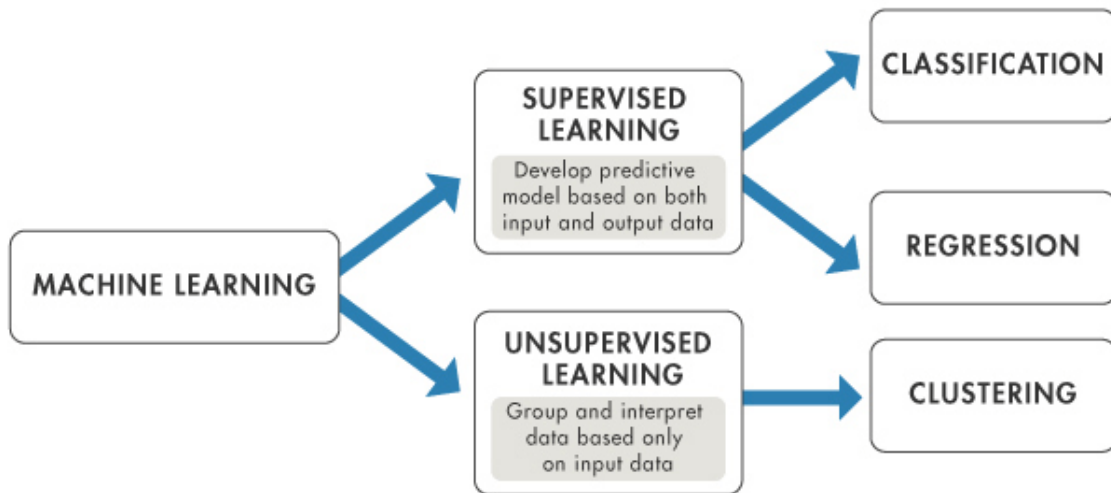
Fig. 1. Two main types of machine learning.

- **Supervised learning**

In supervised learning, data scientists train the algorithm with labeled data and the variables that they want the algorithm to evaluate. Once the model is trained with a set of known (labeled) data, the model is fed unknown (unlabeled) data to get a new response. Supervised machine learning requires data scientists to train the algorithm with both labeled inputs and desired outputs (Ghahremani Nahr and Nozari, 2021). Figure 2 shows the supervised machine learning schematically.
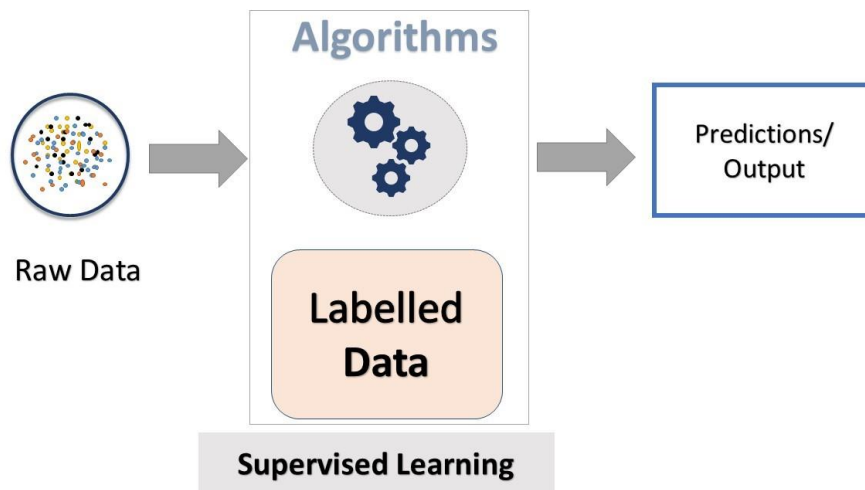

Fig. 2. Supervised machine learning (Nozari and sadeghi 2021)

Supervised learning algorithms are suitable for the following tasks:

- Binary classification: dividing data into two categories
- Multiclass classification: choosing between more than two classes
- Regression Modeling: Predicting Continuous Values
- Ensemble classification: combining predictions from multiple machine learning models to produce an accurate prediction

- *Unsupervised learning*

This type of machine learning involves algorithms that learn using unlabeled data. Unsupervised machine learning algorithms examine unlabeled data in search of patterns that can be used to group data points into subsets. One thing to note is that unsupervised learning is not able to add labels to the data (Aliahmadi et al., 2022).

Unsupervised learning algorithms are suitable for the following tasks:

- Clustering: Dividing data sets into groups based on similarity.
- Anomaly detection: identifying unusual data points in a data set.
- Dependency exploration: Identifying sets of items in a data set that often occur together.
- Dimensionality reduction: reducing the number of variables in a data set.

- *semi-supervised learning*

This approach is a combination of supervised learning (with labeled training data) and unsupervised learning (without labeled training data). In semi-supervised learning, only a small number of input data are labeled (Rafierad et al., 2022).

In semi-supervised machine learning, the model is first trained using labeled data. Unlabeled data is then fed to the model. The model labels unlabeled data with some accuracy. These data are called pseudo-labeled data. Finally, the combination of pseudo-labeled data and original labeled data is used to improve the accuracy of the model.

Today, there is a huge amount of data in various industries. Many large companies have already collected millions of terabytes of data and are still collecting it. But tagging the collected data requires a lot of manpower and resources and is therefore very expensive. So many real databases fall into this category (Aliahmadi et al., 2022).

Some of the fields in which semi-supervised learning is used are:

- Machine translation: teaching language translation algorithms.
- Fraud detection: Identifying fraud cases when there are only a few positive samples.
- Data Labeling: Algorithms trained on small data sets can learn to automatically apply data labels to larger data sets.

- *Reinforcement learning*

Data scientists typically use Reinforcement Machine Learning to perform a multi-step process that has specific rules. Data scientists program the algorithm to do a task, and while the algorithm is trying to do it, they give it positive or negative signals. When the model predicts or produces an outcome, it is penalized if the prediction is wrong and rewarded if the prediction is correct and trains its model accordingly (Nozari et al., 2021).

Some of the fields in which reinforcement learning is used are:

- Robotics: Robots can learn to perform real-world tasks using this technique.

- Video games: Reinforcement learning has been used to train robots to play a number of video games.
- Resource management: Given limited resources and a specific goal, reinforcement learning can help companies plan how to allocate resources.

3. Machine learning processes

Explaining how a complex machine-learning model works can be very challenging. In some industries, data scientists must use simple models (Eisapour et al., 2013). Because explaining how decisions are made for that business is important. Sophisticated models can provide accurate predictions, but explaining how the output is determined is difficult for a layperson. In general, machine learning processes are:

✓ Data collection and preparation

The first step in the machine learning process is to provide the knowledge and data required for a machine. These data are divided into two groups one group is used for training the system and the other group is used for testing the system (Chen et al., 2022).

✓ Choosing and training a model

The second and next step in the principles of machine learning is choosing a model and training it. We have different types of algorithms and machine learning models that have already been developed and modified to solve a specific type of problem or problem. Therefore, depending on the need and suitability of the model to solve the problem in question, choose a model and train it (Moeini et al., 2013).

✓ Evaluating a model

A machine learns different patterns and characteristics from the data it is taught and trains itself to make decisions in different areas such as identifying, classifying or predicting new data. To check exactly how well the machine is able to make these decisions, they test the predictions on the training data (Toloie-Eshlaghy et al., 2013).

4. Applications of machine learning

Machine learning is everywhere. You may use it many times in your daily life and not even know about it. The most important applications of machine learning are (Pishkar et al., 2021):

✓ *Image recognition and processing*

One of the common applications of machine learning is image recognition. There are many situations for classifying objects in digital images and for this purpose machine learning can be used.

✓ *Speech recognition*

Another application of machine learning is speech-to-text conversion. This problem is also known as computer speech recognition with automatic speech recognition.

✓ *Diagnosis*

Machine learning can be used in techniques and tools used to diagnose diseases. This technology can be used to analyze clinical parameters and combine them to predict disease progression, medical information extraction, research to achieve results, treatment planning, and patient monitoring.

✓ *Statistical Analysis*

In economic affairs, one of the important issues is obtaining short-term strategies for buying and selling securities. To obtain these strategies, the user uses trading algorithms to buy and sell securities based on factors such as historical correlations and general economic variables. Machine learning can be very useful to get these short-term strategy algorithms.

✓ *Learning Communities*

Learning communities mean the process of developing insights between different products. One of the applications of machine learning is to study the relationship between the products that people buy.

✓ *Categorization*

The meaning of categorization is to place each person, object, etc. in different categories under study. Classification helps to analyze the measurements of an object to identify the category to which it belongs. To create an efficient relationship, analysts use data (Tootian et al., 2022).

✓ *Financial Services*

Machine learning has many capacities to be used in the financial and banking fields. Popular financial services can be provided using machine learning and artificial intelligence. Machine learning can help banks and financial institutions make more accurate decisions, help financial services providers detect account closures before they happen, track customer spending patterns, perform market analysis, track patterns Teach the cost to smart machines, and finally, machine learning algorithms can easily identify upcoming trends and react in real time (Nozari & Szmelter-Jarosz, 2022).

While machine learning algorithms have been around for decades, they have gained new popularity with the rise of artificial intelligence. Deep learning models in particular power today's most advanced AI applications. As machine learning becomes increasingly important to business operations and artificial intelligence becomes more applicable in enterprise settings, the machine learning platform war is intensifying (Nozari et al., 2023).

5. Conclusion

As we mentioned, machine learning is a subset of artificial intelligence. By using machine learning techniques, the computer system learns the patterns in the processed information or data and can use them. Also pay attention to the fact that in these techniques, computer system learning takes place without explicit programming.

Machine learning solutions continue to make significant changes to core business processes and become more common in our daily lives. Many companies have already started using machine learning because of its potential for more accurate predictions and business decisions. Ongoing research in deep learning and artificial intelligence is increasingly focused on developing more general applications. Today's AI models require extensive training to produce a machine learning algorithm that is highly optimized to perform a task. But some researchers are exploring ways to make the models more flexible, looking at techniques that allow a machine to apply context learned from one task to different future tasks.

The field of machine learning is a tool to improve and apply to the immediate problems of today's world, which makes our life easier and more modern. The main difference between the old software and the software that is based on machine learning is that in the new method, the system is trained to use a large amount of information, and this method is also based on experience. As a result, the machine learning approach is much more effective than the old approaches for solving problems. For this reason, the most

important applications of machine learning have been evaluated in this research. Understanding the most important applications of machine learning with an emphasis on big data can be a valuable guide for researchers in this field.

References

Aliahmadi, A., & Nozari, H. (2023, January). Evaluation of security metrics in AIoT and blockchain-based supply chain by Neutrosophic decision-making method. In Supply Chain Forum: An International Journal (Vol. 24, No. 1, pp. 31-42). Taylor & Francis.

Aliahmadi, A., Nozari, H., & Ghahremani-Nahr, J. (2022). A framework for IoT and Blockchain Based on Marketing Systems with an Emphasis on Big Data Analysis. International journal of Innovation in Marketing Elements, 2(1), 25-34.

Aliahmadi, A., Nozari, H., & Ghahremani-Nahr, J. (2022). AIoT-based sustainable smart supply chain framework. International journal of innovation in management, economics and social sciences, 2(2), 28-38.

Aliahmadi, A., Nozari, H., & Ghahremani-Nahr, J. (2022). Big Data IoT-based agile-lean logistic in pharmaceutical industries. International Journal of Innovation in Management, Economics and Social Sciences, 2(3), 70-81.

Bayanati, M., Peivandizadeh, A., Heidari, M. R., Foroutan Mofrad, S., Sasouli, M. R., & Pourghader Chobar, A. (2022). Prioritize Strategies to Address the Sustainable Supply Chain Innovation Using Multicriteria Decision-Making Methods. Complexity, 2022.

Chen, Y., Bayanati, M., Ebrahimi, M., & Khalijian, S. (2022). A Novel Optimization Approach for Educational Class Scheduling with considering the Students and Teachers' Preferences. Discrete Dynamics in Nature and Society, 2022.

Eisapour, K., Bayanati, M., & Yousefpour, J. (2013). A mathematical model for ranking R&D organizationsas a technology development factor. Advances in Environmental Biology, 717-721.

Ghahremani-Nahr, J., & Nozari, H. (2021). A Survey for Investigating Key Performance Indicators in Digital Marketing. International journal of Innovation in Marketing Elements, 1(1), 1-6.

Moeini, E., Bayanati, M., Givi, S. B., & Soheili, S. (2013). Individual social responsibility a platform for intellectual property rights: Iranian users of CDs case. Advances in Environmental Biology, 709-717.

Nahr, J. G., Nozari, H., & Sadeghi, M. E. (2021). Green supply chain based on artificial intelligence of things (AIoT). International Journal of Innovation in Management, Economics and Social Sciences, 1(2), 56-63.

Nozari, H., & Nahr, J. G. (2022). The Impact of Blockchain Technology and The Internet of Things on the Agile and Sustainable Supply Chain. International Journal of Innovation in Engineering, 2(2), 33-41.

Nozari, H., & Sadeghi, M. E. (2021). Artificial intelligence and Machine Learning for Real-world problems (A survey). International Journal of Innovation in Engineering, 1(3), 38-47.

Nozari, H., & Szmelter-Jarosz, A. (2022). IoT-based Supply Chain For Smart Business (Vol. 1). ISNET.

Nozari, H., Fallah, M., Kazemipoor, H., & Najafi, S. E. (2021). Big data analysis of IoT-based supply chain management considering FMCG industries. Бизнес-информатика, 15(1 (eng)), 78-96.

Nozari, H., Fallah, M., Szmelter-Jarosz, A., & Krzemiński, M. (2021). Analysis of security criteria for IoT-based supply chain: a case study of FMCG industries. Central European Management Journal, 29(4).

Nozari, H., Ghahremani-Nahr, J., & Szmelter-Jarosz, A. (2023). AI and machine learning for real-world problems. Advances In Computers, (online first).

Nozari, H., Szmelter-Jarosz, A., & Ghahremani-Nahr, J. (2022). Analysis of the Challenges of Artificial Intelligence of Things (AIoT) for the Smart Supply Chain (Case Study: FMCG Industries). Sensors, 22(8), 2931.

Pishkar, N., Nasimi, M. A., & RAHMATI, M. (2021). Developing a Conceptual Model of Green Supply Chain Antecedents and Consequences in the Qualification Approach.

Rafierad, S., Aghajani, H. A., Agha Ahmadi, G., & Rahmaty, M. (2022). Construction and Validation of Dimensions and Components of the Organizational Anomie Scale in order to provide a Native Model in Government Hospitals. Journal of System Management, 8(2), 57-73.

Shayan Nia, S. A., & Mirataollahi Olya, R. (2021). Provide a model for manpower scheduling using a hybrid meta-innovative algorithm in the Water and Sewerage Company. Journal of Industrial Strategic Management, 6(2), 18-32.

Shayannia, S. A. (2022). Designing a Multiobjective Human Resource Scheduling Model Using the Tabu Search Algorithm. Discrete Dynamics in Nature and Society, 2022.

Tavakkoli-Moghaddam, R., Ghahremani-Nahr, J., Samadi Parviznejad, P., Nozari, H., & Najafi, E. (2022). Application of internet of things in the food supply chain: a literature review. Journal of applied research on industrial engineering, 9(4), 475-492.

Toloie-Eshlaghy, A., & Bayanati, M. (2013). Ranking information system success factors in mobile banking systems with VIKOR. Middle-East Journal of Scientific Research, 13(11), 1515-25.

Tootian, S., Bayanaty, M., & Jalali, A. (2022). Identify The Dimensions And Components Of Communication Technology In Order To Empower Women Entrepreneurs In Business. An Approach to Business Management, 2(4), 32-49.