International Transactions on Artificial Intelligence (ITALIC)
Vol. 1 No. 2, April 2023, pp. 205~212x
P-ISSN: 2963-6086, E-ISSN: 2963-1939, DOI: 10.34306

Technology Integration in Data Analysis Using Data Science

Alwiyah[1]
[1]University of Wiraraja
e-mail: alwiyah@wiraraja.ac.id

Article Info

Article history:

Received February 10, 2023
Revised May 2, 2023
Accepted May 5, 2023

Keywords:

Data Science
Machine Learning
Deep Learning
Big Data

ABSTRACT (10 PT)

Nowadays, data has become a very important thing for an entity. Many entities are competing to utilize the information and data they have. Because of this, data analysis has become very important. However, with the increasing amount of data available, managing and analyzing data has become increasingly difficult with the methods used previously. With the development of technology, new data analysis methods can be used to overcome this problem. Nowadays, we have to cope with not only structured data but also unstructured data.

Corresponding Author:

Alwiyah
Universitas of Wiraraja
Email: alwiyah@wiraraja.ac.id

1. Introduction

Economic actors interpret Revolution 4.0 as the ability to make decisions quickly and accurately. Data helps in making quick decisions, known as Data-Driven Decision Making. The problem is that data analysis requires a lot of time and effort, so urgent decisions often depend on intuition [1].

The ease of delivery of digital media such as audio, video, and text over the internet has increased. Once you have access to the data, the next step is to find a way to utilize it. If data is only stored historically, it will only result in a decrease in storage capacity. According to an article in Forbes, 90% of data was created in the last two years, indicating the rapid growth of data [2].

From the facts presented, it can be concluded that entities that are able to utilize the data they have appropriately and use appropriate methods, will gain significant benefits from managing the data [3].

Data Science is a branch of science that specializes in understanding data, especially numerical data, both structured and unstructured. The fields in Data Science cover all processes related to data, from collecting, analyzing, processing, managing, archiving, categorizing, presenting, distributing, and transforming data into information that can be understood by everyone [4].

Data science is supported by various disciplines such as mathematics, statistics, computer science, information systems, management, information science, communication and literature, and is also important for economics, especially business science [5].

A professional who works in the field of Data Science is known as a data scientist, which used to be called a statistician. Therefore, it is not surprising that data scientists today more often use algorithms to create computer programs that are able to process data programmatically [6].

Data Science is a rapidly growing science that is applied by organizations that utilize the principles of Data Driven Decision Making. Some of the technologies that support the development of Information Analysis include Data Enterprise, Automated Learning, and the IoT. When these technologies can be combined and processed to meet the needs of the organization, the potential for Data Science to provide benefits to the organization can be very large [7].

2. Research Methodology

There are five phases to go through in the Data Science implementation process, namely:

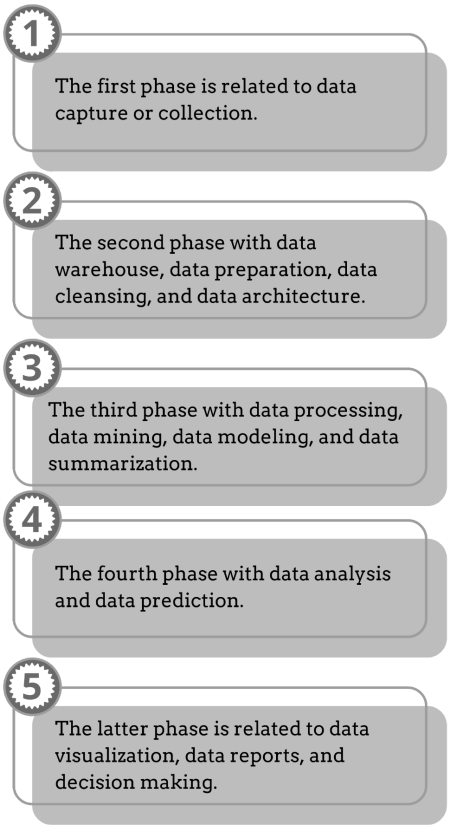

Figure 1. Data Science Lifecycle

3. Analysis and Discussion

As discovered earlier, there are several technologies that strongly support the implementation of Information Analysis include Data Enterprise, Automated Learning, and the IoT [8].

1. IoT

   IoT, It is a revolutionary school of thought that aims to capitalize on the potential of the constantly evolving internet connectivity. IoT makes concepts such as information sharing, remote control, etc. a reality for real-world objects, such as groceries, electronic appliances, collectibles, home appliances, and even living

organisms that are linked to local and global networks through installed and always-on sensors.

Today, the internet infrastructure is advancing and making a huge impact on the world. Not only smartphones and computers are connected to the internet, but various real objects are also connected. For example, production machinery, cars, electronic equipment, and human wearables. These objects have been tied to domestic and foreign internet ties through the use of smart mechanical indicators and devices that can convert electrical signals into mechanical actions. Items built with Machine to Machine linkage capabilities are often referred to as "advanced systems" such as advanced labels, advanced meters, and advanced interconnected sensors.

IoT comes from a school of thought where the word "internet" means an internet connection. IoT devices do not necessarily need to be connected to the internet, but they do need to be connected to a specific network. There are several protocols that can facilitate this connection, such as TCP/IP. Specifically for wireless networks, the WiFi protocol is commonly used. Some applications also utilize the Bluetooth protocol, especially those that require a short-distance connection. As for applications that require long distances, LoRa is becoming a popular protocol.

The development board is a very important part of IoT implementation. This board is the foundation where sketches or programs will be implemented. The program will allow the board to perform various activities as desired.

The Arduino board is a mainstay for IoT. The advantage of this board is the open design of both hardware and software.
and software, so that many parties can produce this board at affordable prices and easily accessible to the public.

The ESP8266 board is one of the most popular. Its advantage is that it already has a WiFi module that makes it easy to connect to the network. There are many implementation versions of this board, such as Espressif and NodeMCU, which makes it very affordable and popular.

The viability and efficiency of an IoT system depends on the price of the board used. If the quantity of points in the procedure is not too much, the price difference does not matter much. But if the quantity of points in the procedure reaches 1000, then the price aspect becomes a very important factor. Even a price difference of Rp. 10,000,- can make a difference of ten million rupiah. There are boards that have better quality, but the price is also much higher, such as the is also much higher, such as Puck.js with a price per unit of Rp. 300,000.

The name for the program or set of code used to control the development board is sketch. Sketches can be written using various IDEs or text editors, such as ArduinoIDE, Visual Studio Code (with the help of the platformIO plugin). One of the most widely used IDEs is ArduinoIDE.

When opening the ArduinoIDE application, the user is greeted by a default view that shows two important parts of the program. These are the void setup and void loop functions. The void setup function will run only once when the board is turned on, while the void loop function will repeat endlessly as long as the board is still running the inputted program.

Apart from programming Arduino boards, ArduinoIDE can also be used to upload sketches to other boards such as WeMos. However, for boards other than Arduino, there are some steps that need to be taken to configure the IDE environment so that it can run properly together with ArduinoIDE.

Sensors are one way to connect the IoT world with the surrounding environment. It has the task of converting the information received into data that can be processed and processed. For example, a temperature sensor can monitor temperature and convert it into temperature data that can be analyzed and taken action.

In IoT applications, there are various protocols that can be used, one of the most widely used is the Wifi protocol. These protocols become the connection method for devices in the IoT system to communicate with each other. For example,

to build a weather sensor, a server is needed to store the data received from the sensor, so that it can be processed and analyzed. Communication between the sensor, board, and network, be it the internet or local, can use the Wifi protocol.

Message Queue Telemetry Transport protocol, which is a lightweight publisher and subscriber protocol, is often used in implementations. It is especially useful in low-power devices, but can be applied in a variety of situations. The basics of this protocol involve publishing and subscribing to messages within a specific to topic. Some clients connect to the broker and receive messages based on that topic, while others can connect as subscribers and send messages using the same topic [9].

IoT performs an important function in the capture stage of the data science lifecycle. Through the utilization of development boards and connected sensors, IoT devices can collect data and information at a high speed and frequency. This helps speed up the information collection process and strengthens the final results in the data science process.

2. Big Data

The definition of Big Data terminology still does not have definite and standardized boundaries in its concept. According to Diebold, the definition of Big Data today is different from the definition that existed about 15 years ago. References to Big Data before the year 2000 are interesting, but do not yet have sufficient evidence. The need for big data management and analysis was recognized in the mid-1990s, when Silicon Graphics (SGI) began to realize its importance [10].

Press reports even suggest that awareness of this has been around for a while. Gandomi and Haider, also referring to Diebold, state that the term Big Data started to be talked about in scientific publications in 2012 and 2013, with the number of publications containing Big Data increasing 2-3 times compared to 2011.

In general, people often use data size as a reference to determine whether data is big data or not. However, data size (volume) is not the only factor to consider. Gandomi and Haider refer to Laney and use three parameters namely Volume, Variety, and Velocity (also known as Three-V) as a reference in the definition of Big Data. There are several other definitions that emphasize on these three parameters as the main factors in the definition of Big Data.

Volume refers to the size of the data being managed (in units of MB, GB, TB, PB, or ZB). The boundaries for the size of data referred to as Big Data can still vary. Furht and Villanustre, as well as Driscoll, state that the minimum size of a Big Data system is between Terabytes (TB) to Petabytes (PB). Variety refers to the degree of diversity in the data structure, which can be structured (such as tables), semi-structured (such as XML documents), and unstructured (such as documents, emails, text messages, audio, video, images, graphs, etc.). Velocity refers to the way and speed of receiving data, including batch, near-time, real-time, or stream processing.

3. Machine Learning

Lately, many people have started to pursue knowledge about Machine Learning. There are several things that are often discussed and asked, such as:

a. What exactly is Machine Learning and how does it relate to Big Data and Business Analytics? What are the differences between the three?
b. What is the difference between Machine Learning, Data Analytics, Data Mining, Data Science and Artificial Intelligence? What are the fundamental differences between the three?

By definition, machine learning is the study of algorithms and statistical models that are used by computer systems to perform certain tasks without instruction. algorithms and statistical models used by computer systems to perform certain tasks without explicit instruction. Explicit instructions.

AI or Artificial Intelligence is a method of programming computers to perform tasks in a rational manner. For example, a machine can trigger an alarm if a parameter exceeds a specified limit and impacts the final outcome of the process [11].

Machine Learning is a subset of AI that allows machines to improve their performance through experience and data. By combining collected data and algorithms such as Support Vector Machine (SVM), machines can make more informed and intelligent decisions [12].

Machine Learning is a subset of AI that allows machines to improve their performance through experience and data. By combining collected data and algorithms such as Support Vector Machine (SVM), machines can make more informed and intelligent decisions.

Data Mining hunts for specific information, while Machine Learning prioritizes specific tasks. For example, teaching someone how to dance is an example of Machine Learning, while finding the best dance center in a city is an example of Data Mining [13].

The Machine Learning process involves building structures, each stage improving the machine to make it better [14]. For ease of understanding, the Machine Learning process can be divide into 3 parts.

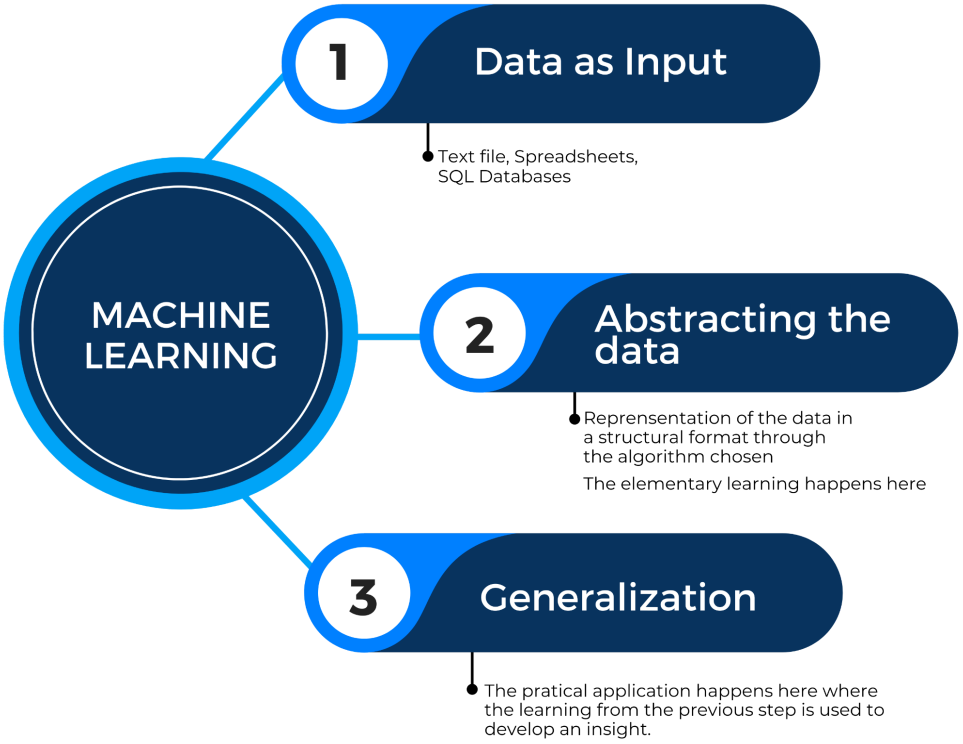

Figure 2. Three main aspects of machine learning.

Various Machine Learning algorithm options are available for use, such as :

1. Directed/Predictive Learning model
   This model is a type of Machine Learning algorithm that utilizes existing data to make predictions or classifications on new data [15]. In this model, the machine is "trained" by inputting data and corresponding labels, so that the machine can identify patterns and make accurate predictions. This model is tasked with guessing future outcomes based on historical data. Predictive models are usually received with clear instructions on what to learn and how to manage it. This algorithm is called Directed Learning. Example: when a marketing company tries to determine which customers are most likely to switch to another supplier, they use Directed Learning. These algorithms can also help predict risks, such as earthquakes, tornadoes, etc., in order to

accurately determine insurance rates [16]. The algorithms used include: Nearest Neighbor, Naive Bayes, Decision Tree, Regression, and many more.

2. Unsupervised/Explained Learning Model
   This model is used to train with no defined goals or priorities [17]. For example, a merchant could use it to find product matches that consumers buy most often. In the pharmaceutical industry, this model can be used to identify medical conditions that may co-occur with diabetes. Algorithms such as K-Means Clustering Algorithm is one of the methods used in this model.
3. Reinforcement Learning (RL)
   This model is a form of machine learning where machines are taught to make the right decisions to meet business needs and maximize efficiency [18]. Reinforcement learning relies on the machine/software to train continuously based on the environment that affects it, and apply the knowledge gained to solve business problems. This continuous learning process reduces human participation and saves time. Algorithms such as Markov Decision Process are examples used in Reinforcement learning. To distinguish between Supervised and Reinforcement learning, we can look at the example of a car. The car will use Reinforcement learning to decide which route to take, what speed to control, and make other decisions after interacting with the environment. While predicting taxi fares from one place to another is an example of Supervised learning. Large companies like Google and Facebook make extensive use of Machine learning to drive relevant ads to their users.
4. Banking & Financial Services
   Machine Learning provides the ability to project customers who are likely to default on loans or credit card bills. This is useful for banks to determine which customers are eligible for a loan or credit card facility. By utilizing the sophistication of Machine Learning, banks can optimize the decision-making process and minimize the risk of payment failure [19].
5. Healthcare
   Machine Learning helps in diagnosing deadly diseases like cancer through analyzing patient symptoms and comparing with similar patient data. This allows doctors to make more accurate and timely diagnoses, which is crucial in the treatment of cancer and similar diseases. Machine Learning algorithms process data and generate trusted recommendations, helping doctors to make more informed decisions [20].
6. Retailing
   Machine learning can help retailers identify products that sell more frequently and decide which products to display or remove from the shelves [21]. Machine learning algorithms can analyze sales data and forecast products that are likely to sell well in the future. This helps retailers to make informed and data-driven decisions about which products to display or remove from the shelves. In addition, machine learning can also help retailers find two or more products that are sold together, which can help retailers to increase customer loyalty and build loyal customers. This is done by analyzing sales data and forecasting products that are likely to sell together. Then, retailers can promote these products together to increase sales and build customer loyalty. Overall, machine learning can help retailers to make informed and data-driven business decisions, helping to increase sales and build loyalty.

In the data science life cycle, machine learning plays a key role in processing and deciphering information.

4. Conclusions

The phenomenal expansion of data in the 21st century has made data science the answer to data management problems. No wonder why the profession as a data scientist is considered as one of the most "exciting" jobs today. Thanks to the capabilities of data science and the technology behind it, including large and unstructured data, can be processed into better analysis than previous methods such as data warehouse and data mining.

REFERENCES

[1] Sriliasta, C., Wuisan, D. S. S., & Mariyanti, T. (2022). Functions of Artificial Intelligence, Income Investment Instrument, and Crypto Money in Era of The Fourth Revolution. International Transactions on Artificial Intelligence, 1(1), 117-128.
[2] Sinta, I., Ilham, R. N., ND, M. A., Subhan, M., & Usman, A. (2022). UTILIZATION OF DIGITAL MEDIA IN MARKETING GAYO ARABICA COFFEE. IRPITAGE JOURNAL, 2(3), 103-108.
[3] Rahardja, U., Hongsuchon, T., Hariguna, T., & Ruangkanjanases, A. (2021). Understanding the impact of continuance intent from s-commerce activity: The role of customer experience, perceived value, and mediation of relationship quality. Sustainability , 13 (20), 11492.
[4] Nambiar, A., & Mundra, D. (2022). An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management. Big Data and Cognitive Computing, 6(4), 132.
[5] Virkus, S., & Garoufallou, E. (2020). Data science and its relationship to library and information science: a content analysis. Data Technologies and Applications, 54(5), 643-663.
[6] Green, B. (2021). Data science as political action: grounding data science in a politics of justice. Journal of Social Computing, 2(3), 249-265.
[7] Troisi, O., Maione, G., Grimaldi, M., & Loia, F. (2020). Growth hacking: Insights on data-driven decision-making from three firms. Industrial Marketing Management, 90, 538-557.
[8] Guo, Y., Wang, N., Xu, Z. Y., & Wu, K. (2020). The internet of things-based decision support system for information processing in intelligent manufacturing using data mining technology. Mechanical Systems and Signal Processing, 142, 106630.
[9] Singh, D., Singh, R., Gupta, A., & Pawar, A. V. (2022). Message queue telemetry transport and lightweight machine-to-machine comparison based on performance efficiency under various scenarios. International Journal of Electrical and Computer Engineering, 12(6), 6293.
[10] Sestino, A., Prete, M. I., Piper, L., & Guido, G. (2020). Internet of Things and Big Data as enablers for business digitalization strategies. Technovation, 98, 102173.
[11] Oktradiksa, A., Bhakti, C. P., Kurniawan, S. J., & Rahman, F. A. (2021). Utilization artificial intelligence to improve creativity skills in society 5.0. In Journal of Physics: Conference Series (Vol. 1760, No. 1, p. 012032). IOP Publishing.
[12] Sunarya, P. A. (2022). Machine Learning and Artificial Intelligence as Educational Games. International Transactions on Artificial Intelligence, 1(1), 129-138.
[13] Jamalpur, B. (2020). Data Exploration As A Process Of Knowledge Finding And The Role Of Mining Data Towards Information Security. Journal Of Mechanics Of Continua And Mathematical Sciences, 15(6), 2020-2028.
[14] Akinosho, T. D., Oyedele, L. O., Bilal, M., Ajayi, A. O., Delgado, M. D., Akinade, O. O., & Ahmed, A. A. (2020). Deep learning in the construction industry: A review of present status and future innovations. Journal of Building Engineering, 32, 101827.
[15] Shah, D., Patel, S., & Bharti, S. K. (2020). Heart disease prediction using machine learning techniques. SN Computer Science, 1, 1-6.
[16] Wang, W. L., van De Lindt, J. W., Hartman, B., Cutler, H., Kruse, J. L., McAllister, T. P., & Hamideh, S. (2022). Determination of individual building performance targets to achieve community-level social and economic resilience metrics. Journal of Structural Engineering, 148(5), 04022045.
[17] Polu, S., & Sutskever, I. (2020). Generative language modeling for automated theorem proving. arXiv preprint arXiv:2009.03393.
[18] Lieder, M., Asif, F. M., & Rashid, A. (2020). A choice behavior experiment with circular business models using machine learning and simulation modeling. Journal of Cleaner Production, 258, 120894.
[19] Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. Business Horizons, 63(2), 157-170.
[20] Mijwil, M. M., & Abttan, R. A. (2021). Utilisation of machine learning techniques in testing and training of different medical datasets. Asian Journal of Computer and Information Systems (ISSN: 2321–5658), 9(4).
[21] Pfeiffer, J., Pfeiffer, T., Meißner, M., & Weiß, E. (2020). Eye-tracking-based classification of information search behavior using machine learning: evidence from experiments in physical shops and virtual reality shopping environments. Information Systems Research, 31(3), 675-691.

	**Alwiyah,** Universitas of Wiraraja, alwiyah@wiraraja.ac.id