International Journal of Society Reviews (INJOSER) e-ISSN: 3030-802X
Vol. 2 No. 5 2025, pages: 3181~3189

AI ALGORITHMS AND MACHINE LEARNING FOR BIG DATA PROCESSING: TRENDS AND PROGRESS

Loso Judijanto
IPOSS Jakarta, Indonesia
losojudijantobumn@gmail.com

Abstract

AI and machine learning algorithms play an important role in big data processing, a field that is constantly evolving and innovating. The application of these technologies enables the analysis of huge volumes of data, increasing the ability to discover complex patterns and generate insights that were previously hard to reach. Current trends focus on the adoption of more automated and efficient solutions such as AutoML, as well as improvements in tools and frameworks that facilitate integration on a broader level. While challenges such as interpretability and transparency remain, developments in explainable AI seek to address them. All of these advancements demonstrate the huge potential of AI and machine learning in transforming big data processing, providing significant benefits to various industry sectors through faster and more effective decision-making.
Keywords: AI Algorithm, Machine Learning, Big Data Processing.

Introduction

In today's digital age, the volume of data generated by various sources, such as social media, IoT sensors, online transactions, and mobile devices, is increasing rapidly. This phenomenon is known as big data. Big data has unique characteristics, often summarised in "3V": Volume, Variety, and Velocity. Volume refers to the large amount of data, Variety refers to the different types of data that exist, while Velocity describes the speed at which new data is generated and must be processed. (Madhuri et al., 2022)..

Big data is a collection of very large, diverse, and complex data that cannot be processed using conventional methods and tools. The main characteristics of big data are referred to as "3V": Volume (huge amount of data), Variety (diversity of data types such as text, images, videos, and sensor data), and Velocity (high speed in acquiring and processing data). (Moorthy & Gandhi, 2022). In this context, big data requires advanced technologies and specialised algorithms, such as artificial intelligence (AI) and machine learning (ML), to extract valuable information and support more efficient and effective decision-making in various industry sectors. (Samuel et al., 2023).

Processing big data requires technology that can analyse it quickly and efficiently to extract valuable information from it. This is what makes Artificial Intelligence (AI) and Machine Learning (ML) algorithms increasingly relevant. (Wang, 2024).

Artificial Intelligence (AI) and Machine Learning (ML) algorithms are sets of instructions or procedures designed to enable computers to perform certain tasks that would normally require human intelligence. AI algorithms include various techniques and models that allow computers to perform functions such as reasoning, natural language understanding, and pattern recognition. (Mittal et al., 2021).. Machine Learning is a sub-

field of AI that focuses on developing algorithms that allow systems to learn from data and improve their performance over time without being explicitly programmed. ML algorithms utilise the data to build predictive or decision-making models that can be applied in various domains, such as speech recognition, product recommendation, and anomaly detection. (Soujanya & Tembhurne, 2023)..

AI and ML offer a variety of methods to address big data processing challenges, such as pattern recognition, prediction, and automated decision-making. In recent years, there have been significant developments in AI and ML algorithms that enable big data processing to be more efficient and effective. The latest algorithms not only improve performance in terms of speed and accuracy but also their ability to handle highly diverse and complex data. However, challenges in big data processing still remain. Issues such as scalability, interpretability of results, the need for high computation, and data preservation remain major concerns. (Sharma et al., 2020). Therefore, research on trends and advancements in AI and ML algorithms is essential to identify the latest innovations that can mitigate these challenges.

This research aims to explore the current trends in AI and ML algorithm development, and identify technological advances that have a significant impact on big data processing. As such, this research is expected to contribute to technological advancements in dealing with big data challenges and maximising its potential benefits for various industry sectors.

Research Methods

The study in this research uses the literature method. The literature research method is a systematic and structured approach to identifying, evaluating, and synthesising a large amount of literature relevant to a particular research topic. The steps in this method involve searching for scientific sources, such as journals, books, reports, and other publications, related to the research question or hypothesis. (Firman, 2018); (Suyitno, 2021). The researcher then assesses the quality and credibility of these sources, categorises key findings, and analyses patterns, trends and gaps in current knowledge. The main objectives of the literature research method are to provide a strong theoretical foundation, identify areas that require further research, and place research findings in a broader context to support valid arguments and conclusions. (Jelahut, 2022).

Results and Discussion

Challenges in Big Data Processing

Big data, or big data, is a collection of very large, diverse, and complex data that cannot be managed, processed, or analysed with traditional data processing tools and techniques. The main characteristics of big data can be summarised by the concept of "3V": Volume (very large amounts of data, often measured in petabytes or even more), Variety (diversity of data types including structured, semi-structured, and unstructured data such as text, images, video, and sensor data), and Velocity (high speed of data

generation and flow). (Yacobovitch, 2024). In addition to the 3Vs, some experts also add other characteristics such as Veracity (the accuracy and reliability of the data) and Value (the value that can be generated from processing and analysing this data). Big data requires advanced technologies and specialised analytical methods, such as artificial intelligence and machine learning, to extract valuable insights that support more informed and efficient decision-making in various industries. (Suseela & Parekh, 2021).

Challenges in Big Data Processing consist of;

Firstly, technical challenges are the most significant in big data processing. Managing huge volumes of data requires a robust computer infrastructure and adequate storage. In addition, highly diverse data from various sources such as social media, IoT sensors, and business transactions create additional complexity in the integration, cleaning, and transformation of data into formats suitable for analysis. Data processing speed is also a critical issue as many big data applications require near-real-time processing to provide maximum added value. (Wong, 2021).

Secondly, data quality and security issues pose a major challenge. Veracity, or data accuracy, is critical in data-driven decision-making; yet in many cases, data may be incomplete, inconsistent, or contain errors. Maintaining data cleanliness and reliability requires great effort in the extraction, transformation, and loading (ETL) process. (Solanki et al., 2022).. In addition, data security and privacy are serious issues given the volume and sensitivity of data that often includes important personal and business information. Protection against data breaches, unauthorised access and cyber-attacks is a high priority in big data management. (Saxena et al., 2021)..

Thirdly, the aspect of human resources and the expertise required adds another layer of complexity to big data processing. There is a high need for professionals skilled in the fields of data science, data analytics, and artificial intelligence. Identifying, recruiting and retaining talent with a deep understanding of computational techniques, statistics and relevant business domains can be challenging. In addition, an organisational culture that supports the use of big data in decision-making must also be built and sustained. (Chinnaiyan, 2020).

As such, effectively managing and utilising big data requires a comprehensive and holistic approach that includes technical challenges, data quality and security, and the availability of qualified human resources. While these challenges are significant, the potential benefits of careful and accurate data analysis can provide great competitive advantage, innovation and improved operational efficiency. Therefore, organisations must be prepared to invest in advanced technologies, robust processes, and human capability development to fully exploit the potential of big data.

AI and Machine Learning Algorithms

Artificial Intelligence (AI) is a branch of computer science that focuses on developing systems that can perform tasks that typically require human intelligence, such as speech recognition, image recognition, natural language understanding, and decision-

making. The basic principles of AI involve the use of algorithms and mathematical models to process data and make decisions similar to the way humans think (Lehmann et al., 2020). Meanwhile, Machine Learning (ML) is a sub-field of AI that specifically focuses on developing algorithms and techniques that enable computers to learn from and make predictions based on data. By utilising statistical and computational algorithms, ML allows systems to incrementally improve their performance by identifying patterns and insights from such data without having to be explicitly programmed for each task. (Thomas & Anderson, 2023).

Machine Learning algorithms can generally be categorised into three main types: supervised learning, unsupervised learning, and reinforcement learning. Supervised Learning involves training a model using a dataset that is already labelled, meaning each input has a correct output pair. The purpose of the algorithm is to learn relationships or patterns from the labelled data in order to make predictions or decisions on new, unlabelled data. (Ullman, 2023). Examples of supervised learning applications include linear regression to forecast house prices based on their features or classification of emails as spam or not spam using algorithms such as decision trees and support vector machines.

Unsupervised Learning is a machine learning algorithm that works with unlabelled datasets, so the algorithm must find hidden structures or patterns in the data. The goal of unsupervised learning is to identify clusters, segmentation, or dimensionality reduction without the guidance of the correct output. Examples of unsupervised learning include k-means clustering which groups data based on similarity or Principal Component Analysis (PCA) which is used to reduce the number of features in the data while retaining its main variance. (Panesar, 2020).

Reinforcement Learning differs from the previous two types in that it involves an agent that learns through interaction with the environment to achieve a specific goal. This agent attempts various actions to maximise a reward obtained from the environment in response to each action performed. (Hossen, 2020). The basic principle of reinforcement learning is trial and error, where the agent adjusts its action strategy based on the feedback received. Examples of applications of reinforcement learning include the game of chess or Go, where a computer agent learns to play and defeat a human opponent, or an autonomous navigation system that learns to drive a vehicle optimally in a dynamic environment. (Kumari & Mrunalini, 2022)..

Furthermore, in a practical context, these three categories of machine learning algorithms have varying applications and advantages according to the problem at hand. Supervised learning is commonly used in applications where large amounts of labelled data are available, such as in financial fraud detection or medical diagnosis based on test results. Unsupervised learning is useful in situations where we seek to uncover previously unknown data structures, such as in market analysis to identify different customer segments or in social network analysis. On the other hand, reinforcement learning shows tremendous power in situations involving sequential decisions and dynamic environments, such as in robotics, gaming, and investment management. (Djafri, 2021).

In conclusion, the general categories of machine learning algorithms-supervised, unsupervised, and reinforcement learning-represent different approaches to tackle various problems of learning from data. Supervised learning is effective when we have correctly labelled data and want to predict the output for new data. Unsupervised learning is useful for finding hidden patterns in unlabelled data, while reinforcement learning is perfect for problems that involve dynamic interactions with the environment. A deep understanding of these three categories and their implementation can greatly improve the efficiency and effectiveness of machine learning-based solutions built for specific needs.

Recent Trends in Big Data Processing

In recent years, big data processing has undergone rapid development influenced by the latest technologies and trends. One of the main trends is the use of artificial intelligence (AI) and machine learning (ML) to analyse big data. Artificial intelligence enables the processing and analysis of large amounts of data in a faster and more efficient way than traditional methods. (Kumar et al., 2022).. Machine learning algorithms, in particular deep learning, enable the discovery of more complex patterns in very large data, such as in medical image analysis or market trend prediction (Plesničar et al., 2020)..

Another trend is the use of cloud computing technologies that enable large-scale processing of big data without requiring a large initial investment in physical infrastructure. Cloud service providers such as Amazon Web Services (AWS), Microsoft Azure and Google Cloud Platform offer integrated big data solutions, including data storage, analytics tools and computing capabilities. Cloud technology not only increases scalability and flexibility but also enables better collaboration between teams and organisations in different locations. (Dulhare & Gouse, 2020).

Subsequently, edge computing started to gain attention as the latest big data processing trend. Edge computing involves processing data near the data source (such as IoT devices) to reduce latency and increase response speed. With edge computing, data can be processed locally before being sent to the data centre or cloud for further analysis. This is particularly useful in real-time applications such as autonomous vehicles, remote healthcare, or industrial sensors, where the speed and efficiency of data processing is critical (Koul et al., 2022).

Finally, there is an increasing focus on data privacy and security in big data processing. With ever-growing volumes of data, including personal and sensitive data, there is an urgent need to ensure that it is protected from breaches and unauthorised use. Methods such as data anonymisation, sophisticated encryption techniques, and strict policies and regulations, such as GDPR in Europe, are more important than ever. Organisations must ensure that they adhere to best practices in managing data privacy and security to maintain user trust and comply with applicable regulations. (YILDIRIM et al., 2021).

Overall, recent trends in big data processing reflect the ever-evolving technological advancements and the need to manage and analyse data in a more efficient, secure and

innovative way. By continuously monitoring and adopting these trends, organisations can gain deeper insights and make better decisions based on big data.

Technological Advancements in Machine Learning

Technological advancements in machine learning have seen a significant surge in the last decade. One of the biggest developments has been in deep learning algorithms, which allow computers to learn from unstructured data such as images, text, and video in a way that mimics human neural networks. (Acharya et al., 2022).. Algorithms such as convolutional neural networks (CNN) have revolutionised image recognition, while recurrent neural networks (RNN) and transformers have improved natural language processing (NLP) capabilities. The use of these models has enabled the achievement of very high accuracy in tasks such as object detection, speech transcription, and automatic machine translation. (Luchs et al., 2023)..

In addition to advances in algorithms, computing GPUs (graphics processing units) and TPUs (tensor processing units) have become an important foundation in accelerating the training of complex machine learning models. GPUs, originally designed for graphics rendering, are now being utilised for the parallel computing required in the training process of AI models. Meanwhile, TPUs were specially created by Google to accelerate tensor operations which are a key component in deep learning algorithms. It enables the training of large models in less time and with more efficient power consumption. (Ghanbari & Najafzadeh, 2020)..

Another advancement can be seen in the emergence of open-source machine learning platforms and frameworks, such as TensorFlow, PyTorch, and Keras. These platforms provide powerful and easy-to-use tools and libraries for the development of machine learning models. They also support the collaboration and innovation of a worldwide community of data scientists and software engineers. (Shahare et al., 2023). For example, TensorFlow and PyTorch have enabled rapid research and implementation in production of a wide range of applications, from healthcare to autonomous vehicles. These frameworks provide extensive documentation, tutorials, and pre-training models that simplify and accelerate the development of machine learning solutions. (Malik et al., 2022).

Finally, the adoption of automation in machine learning or AutoML has changed the way data scientists and engineers work. AutoML provides tools to automatically select the best model, set hyperparameters, and process data with minimal manual intervention. This not only increases efficiency, but also allows individuals with limited technical expertise to utilise machine learning at scale. AutoML has been applied in areas such as market demand prediction, disease detection, and sentiment analysis, providing new capabilities to optimally utilise data.

Overall, technological advancements in machine learning have improved computational capabilities, accelerated development processes, and expanded

accessibility to advanced tools and techniques. This trend looks set to continue, promising greater innovation and expansion of applications in various industries and sectors.

Conclusion

AI and machine learning algorithms have brought significant changes in big data processing. With their ability to handle massive volumes of diverse data, these algorithms enable more in-depth analyses and the discovery of hidden patterns that were previously undetectable. Technologies such as deep learning and natural language processing provide advancements in practical applications such as consumer behaviour analysis, market trend prediction, and anomaly detection in systems. The ability to process big data quickly and accurately gives organisations a competitive advantage, facilitating better and faster decision-making.

On the other hand, current trends show an increased focus on automation and efficiency through AutoML, as well as the adoption of tools and frameworks designed to ease the implementation of AI algorithms on a wider scale. While challenges remain in terms of interpretability and transparency, developments in explainable AI (XAI) continue to address these issues. All of these advancements represent a promising roadmap towards more effective big data processing, enabling industries to be more responsive and adaptive to rapid market changes and dynamics.

References

Acharya, V., Ghosh, A., Kang, I., Munasinghe, T., & Binita, K. C. (2022). Landslide Likelihood Prediction using Machine Learning Algorithms. *2022 IEEE International Conference on Big Data*, *Query date: 2024-12-06 20:45:39*, 5395-5403. https://doi.org/10.1109/bigdata55660.2022.10020433

Chinnaiyan, R. (2020). Big Data. *Machine Learning and Big Data*, *Query date: 2024-12-06 20:45:39*, 105-130. https://doi.org/10.1002/9781119654834.ch5

Djafri, L. (2021). Dynamic Distributed and Parallel Machine Learning algorithms for big data mining processing. *Data Technologies and Applications*, *56*(4), 558-601. https://doi.org/10.1108/dta-06-2021-0153

Dulhare, U. N., & Gouse, S. (2020). Hands on MAHOUT-Machine Learning Tool. *Machine Learning and Big Data*, *Query date: 2024-12-06 20:45:39*, 361-421. https://doi.org/10.1002/9781119654834.ch14

Firman, F.-. (2018). *QUALITATIVE AND QUANTITATIVE RESEARCH*. *Query date: 2024-05-25 20:59:55*. https://doi.org/10.31227/osf.io/4nq5e

Ghanbari, E., & Najafzadeh, S. (2020). Machine Learning. *Machine Learning and Big Data*, *Query date: 2024-12-06 20:45:39*, 153-207. https://doi.org/10.1002/9781119654834.ch7

Hossen, Md. S. (2020). Data Preprocess. *Machine Learning and Big Data*, *Query date: 2024-12-06 20:45:39*, 71-103. https://doi.org/10.1002/9781119654834.ch4

Jelahut, F. E. (2022). *Various Theories and Types of Qualitative Research*. *Query date: 2024-05-25 20:59:55*. https://doi.org/10.31219/osf.io/ymzqp

Koul, S., Koul, B., & Bakshi, B. (2022). Influence of AI and Machine Learning to Empower the Healthcare Sector. *Machine Learning, Deep Learning, Big Data, and Internet of*

Things for Healthcare, *Query date: 2024-12-06 20:45:39*, 55-76. https://doi.org/10.1201/9781003227595-4

Kumar, M., Singh, A. J., Sharma, B., & Cengiz, K. (2022). Evaluation of machine learning algorithms on academic big dataset by using feature selection techniques. *Intelligent Network Design Driven by Big Data Analytics, IoT, AI and Cloud Computing*, *Query date: 2024-12-06 20:45:39*, 61-91. https://doi.org/10.1049/pbpc054e_ch4

Kumari, K., & Mrunalini, M. (2022). Detecting Denial of Service attacks using machine learning algorithms. *Journal of Big Data*, *9*(1). https://doi.org/10.1186/s40537-022-00616-0

Lehmann, C., Huber, L. G., Horisberger, T., Scheiba, G., Sima, A. C., & Stockinger, K. (2020). Big Data architecture for intelligent maintenance: A focus on query processing and machine learning algorithms. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/s40537-020-00340-7

Luchs, I., Apprich, C., & Broersma, M. (2023). Learning machine learning: On the political economy of big tech's online AI courses. *Big Data & Society*, *10*(1). https://doi.org/10.1177/20539517231153806

Madhuri, G. S., Mahesh, T. R., & Vivek, V. (2022). A Novel Approach for Automatic Brain Tumour Detection Using Machine Learning Algorithms. *Big Data Management in Sensing*, *Query date: 2024-12-06 20:45:39*, 87-101. https://doi.org/10.1201/9781003337355-7

Malik, S., Tyagi, A. K., & Sahoo, R. (2022). Machine learning algorithms for Big Data analytics including deep learning. *Machine Learning, Blockchain Technologies and Big Data Analytics for IoTs: Methods, Technologies and Applications*, *Query date: 2024-12-06 20:45:39*, 75-98. https://doi.org/10.1049/pbse016e_ch4

Mittal, R., Arora, S., Kuchhal, P., & Bhatia, M. P. S. (2021). An Insight into Tools and Software Used in AI, Machine Learning and Data Analytics. *Studies in Big Data*, *Query date: 2024-12-06 20:45:39*, 45-64. https://doi.org/10.1007/978-981-33-4412-9_2

Moorthy, U., & Gandhi, U. D. (2022). A Survey of Big Data Analytics Using Machine Learning Algorithms. *Research Anthology on Big Data Analytics, Architectures, and Applications*, *Query date: 2024-12-06 20:45:39*, 655-677. https://doi.org/10.4018/978-1-6684-3662-2.ch031

Panesar, A. (2020). Machine Learning Algorithms. *Machine Learning and AI for Healthcare*, *Query date: 2024-12-06 20:45:39*, 85-144. https://doi.org/10.1007/978-1-4842-6537-6_4

Plesničar, M. M., Završnik, A., & Šarf, P. (2020). Fighting Impunity with New Tools: How Big Data, Algorithms, Machine Learning and AI Shape the New Era of Criminal Justice. *The Fight Against Impunity in EU Law*, *Query date: 2024-12-06 20:45:39*. https://doi.org/10.5040/9781509926909.ch-014

Samuel, P., K., R. A., Rajesh, S., M., K., & A., K. R. (2023). AI-Based Big Data Algorithms and Machine Learning Techniques for Managing Data in E-Governance. *Advances in Electronic Government, Digital Divide, and Regional Development*, *Query date: 2024-12-06 20:45:39*, 19-35. https://doi.org/10.4018/978-1-6684-7697-0.ch002

Saxena, A., Saxena, M., & Ilerena, A. R. (2021). Bias in Medical Big Data and Machine Learning Algorithms. *Artificial Intelligence and Machine Learning in Healthcare*, *Query date: 2024-12-06 20:45:39*, 217-228. https://doi.org/10.1007/978-981-16-0811-7_10

Shahare, V., Arora, N., Hayat, A., & Mouje, N. (2023). Machine Learning Algorithms for Big Data Analytics. *Demystifying Big Data Analytics for Industries and Smart Societies*, *Query date: 2024-12-06 20:45:39*, 47-63. https://doi.org/10.1201/9781003330875-4

Sharma, N., Gautam, S. K., Henry, A. A., & Kumar, A. (2020). Application of Big Data and Machine Learning. *Machine Learning and Big Data*, *Query date: 2024-12-06 20:45:39*, 305-333. https://doi.org/10.1002/9781119654834.ch12

Solanki, S. D., Solanki, A. D., & Borah, S. (2022). Assimilating Machine Learning Algorithms in Big Data Analytics: A Review. *Applications of Machine Learning in Big-Data Analytics and Cloud Computing*, *Query date: 2024-12-06 20:45:39*, 81-114. https://doi.org/10.1201/9781003337218-5

Soujanya, K. V., & Tembhurne, O. (2023). Analysis of Machine Learning Algorithms for Detection of Cyberbullying on Social Networks. *Springer Proceedings in Mathematics & Statistics*, *Query date: 2024-12-06 20:45:39*, 171-190. https://doi.org/10.1007/978-3-031-15175-0_14

Suseela, S., & Parekh, N. (2021). Applications of Machine Learning Algorithms to Cancer Data. *Big Data and Artificial Intelligence for Healthcare Applications*, *Query date: 2024-12-06 20:45:39*, 107-132. https://doi.org/10.1201/9781003093770-7

Suyitno. (2021). *QUALITATIVE RESEARCH METHODS CONCEPTS, PRINCIPLES AND OPERATIONS*. *Query date: 2024-05-25 20:59:55*. https://doi.org/10.31219/osf.io/auqfr

Thomas, R., & Anderson, J. (2023). *Big Data: The Fuel for Machine Learning and AI Advancement*. *Query date: 2024-12-06 20:45:39*. https://doi.org/10.31219/osf.io/36hzn

Ullman, J. (2023). Big-Data Algorithms That Are Not Machine Learning. *2023 IEEE International Conference on Big Data (BigData)*, *Query date: 2024-12-06 20:45:39*, 5-5. https://doi.org/10.1109/bigdata59044.2023.10386233

Wang, R. (2024). AI-Powered Predictive Cybersecurity in Identifying Emerging Threats through Machine Learning. *2024 IEEE 3rd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, *Query date: 2024-12-06 20:45:39*, 819-825. https://doi.org/10.1109/eebda60612.2024.10485789

Wong, Y. K. (2021). Applying AI and Big Data for Sensitive Operations and Disaster Management. *Advances in Machine Learning, Data Mining and Computing*, *Query date: 2024-12-06 20:45:39*, 349-356. https://doi.org/10.5121/csit.2021.111429

Yacobovitch, S. M. (2024). APPLICATION OF MACHINE LEARNING METHODS TO ENHANCE THE PERFORMANCE OF BIG DATA SORTING ALGORITHMS. *The American Journal of Engineering and Technology*, 6(10), 67-74. https://doi.org/10.37547/tajet/volume06issue10-08

YILDIRIM, M., ÇINAR, A., & CENGİL, E. (2021). Investigation of Cloud Computing Based Big Data on Machine Learning Algorithms. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, *10*(2), 670-682. https://doi.org/10.17798/bitlisfen.897573