ISSN: 3123-948X (Media Online)
Journal of Data Science and Informatics Engineering
Vol. 1 No. 1, December 2025 | pp: 1-9
DOI: https://doi.org/10.64803/jodsie.v1i1.12

Enhancing Data Processing Efficiency through Machine Learning Algorithms: A Comprehensive Study

Muhammad Hasanuddin[1]

[1]Master's Degree in Information Technology, Universitas Pembangunan Panca Budi, Medan, Indonesia
Email: [1*]muhammadhasanuddin.skom@gmail.com
(*Email Correspondence: muhammadhasanuddin.skom@gmail.com)
Received: October 10, 2025| Accepted: December 5, 2025 | Published: December 30, 2025

Abstract

This study explores the transformative potential of machine learning algorithms to optimize data processing efficiency across diverse applications and address the growing challenges posed by big data. Specifically, machine learning can significantly enhance railway operations by optimizing maintenance schedules, reducing service interruptions, and improving overall network velocity. By applying advanced analytical techniques to railway data, it is possible to predict potential failures and proactively schedule maintenance, thereby minimizing costly downtime and enhancing the reliability of rail transportation infrastructure. This approach enables the transition from reactive to predictive maintenance strategies, leading to more efficient resource allocation and improved operational safety. This shift towards predictive maintenance, driven by machine learning, is crucial for mitigating risks and extending the lifespan of critical railway assets. This is particularly evident in power systems, where continuous monitoring and fault detection are paramount for maintaining stability and preventing disruptive outages, highlighting the broad applicability of these methodologies.

Keywords: machine learning, data processing, railway operations, predictive maintenance, power systems

1. INTRODUCTION

The escalating volume and complexity of data necessitate advanced methodologies for efficient processing, analysis, and interpretation. Machine learning algorithms have emerged as powerful tools to address these challenges, offering sophisticated approaches to extract meaningful insights and optimize various data-centric operations [1]. This is particularly crucial given that modern applications often generate vast datasets, some of which may contain irrelevant information, making the conversion of raw data into valuable insights a highly relevant research topic [2]. The integration of machine learning into data processing pipelines enables enhanced scalability, real-time processing, and accuracy in handling multi-dimensional and varied data types [3], [4]. These algorithms facilitate the transformation of raw data into a more usable and desirable format, thereby rendering it more meaningful and informative for subsequent analysis or decision-making [5], [6]. Machine learning, a subfield of artificial intelligence, uses statistical techniques to enable computational systems to learn from data without explicit programming, significantly reducing the human effort required for tasks such as categorization [7], [8].

This capacity for autonomous learning from vast, high-dimensional datasets allows machine learning models to uncover actionable insights that might otherwise remain hidden, fundamentally transforming big data analytics and management [9]. Consequently, ML algorithms are pivotal for analyzing, understanding, and evaluating collected data to detect trends and anomalies in large datasets [10]. This capability is particularly beneficial given the challenges of data availability and quality, as well as the potential for inherited biases in human-generated data, which can lead to flawed decision-making [3]. Therefore, machine learning plays a critical role in addressing these challenges by providing methods for data cleansing, feature engineering, and pattern recognition, which are essential for deriving accurate and reliable conclusions from complex datasets [11], [12]. Moreover, ML techniques can automate and streamline the data cleaning process by identifying and handling missing values, outliers, duplicates, and inconsistencies, thereby improving overall data quality [13]. This automated refinement process is critical for ensuring the integrity of datasets, as models trained on biased or low-quality data can perpetuate and even amplify existing inaccuracies [3].

Furthermore, machine learning algorithms can analyze complex datasets that exceed human analytical capabilities, continuously updating their interpretation of data in response to emerging trends [14]. This

JoDSIE © 2025 Author. Published by CV. Raskha Media Group.
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

ISSN: 3123-948X (Media Online)
Journal of Data Science and Informatics Engineering
Vol. 1 No. 1, December 2025 | pp: 1-9
DOI: https://doi.org/10.64803/jodsie.v1i1.12

continuous improvement distinguishes machine learning from traditional programming paradigms, which require explicit rule sets for every possible scenario [3]. This adaptability allows ML models to effectively process diverse data formats, including structured, semi-structured, and unstructured data, enabling comprehensive data integration and analysis [13]. The increasing accessibility of datasets further fuels the application of machine learning across diverse industries, enabling the extraction of valuable insights from data and advancing beyond explicit programming [15].

This enables the development of sophisticated predictive models, anomaly detection systems, and optimization routines, all crucial for various applications [16]. The advantages of machine learning, including its ease of identifying trends and patterns, minimal human intervention, and continuous improvement capabilities, make it exceptionally efficient at handling multidimensional and multivariate data [3]. This ability stems from their capacity for feature extraction and dimensionality reduction, which simplifies complex inputs without losing critical information [17]. These models excel at processing and analyzing vast, intricate datasets with remarkable speed and precision, offering valuable insights and predictive capabilities [18]. Furthermore, machine learning models exhibit significant flexibility, enabling them to effectively capture non-linear relationships within diverse data types and scale efficiently to manage substantial amounts of information [19]. This comprehensive overview demonstrates that machine learning encompasses a range of methodologies, including supervised, unsupervised, and reinforcement learning, each suited to distinct data processing challenges [3]. Supervised learning, for instance, involves training models on labeled datasets to predict outcomes or classify data, while unsupervised learning focuses on discovering hidden patterns and structures within unlabeled data [20]. Reinforcement learning, conversely, allows an agent to learn through trial and error in an interactive environment, utilizing feedback from its actions to optimize decision-making over time [3].

2. RESEARCH METHOD

2.1 Literature Review

The escalating volume and complexity of data necessitate advanced methodologies for efficient processing, analysis, and interpretation. Machine learning algorithms have emerged as powerful tools to address these challenges, offering sophisticated approaches for extracting meaningful insights and optimizing various data-centric operations [1]. This is particularly crucial given that modern applications often generate vast datasets, some of which may contain irrelevant information, making the conversion of raw data into valuable insights a highly relevant research topic [2]. The integration of machine learning into data processing pipelines enables enhanced scalability, real-time processing, and accuracy in handling multi-dimensional and varied data types [3], [4]. These algorithms facilitate the transformation of raw data into a more usable and desirable format, thereby rendering it more meaningful and informative for subsequent analysis or decision-making [5], [6]. Machine learning, a subfield of artificial intelligence, uses statistical techniques to enable computational systems to learn from data without explicit programming, significantly reducing the human effort required for tasks such as categorization [7], [8]. This capacity for autonomous learning from vast, high-dimensional datasets allows machine learning models to uncover actionable insights that might otherwise remain hidden, fundamentally transforming big data analytics and management [9]. Consequently, ML algorithms are pivotal for analyzing, understanding, and evaluating collected data to detect trends and anomalies in large datasets [10]. This capability is particularly beneficial given the challenges of data availability and quality, as well as the potential for inherited biases in human-generated data, which can lead to flawed decision-making [3]. Therefore, machine learning plays a critical role in addressing these challenges by providing methods for data cleansing, feature engineering, and pattern recognition, which are essential for deriving accurate and reliable conclusions from complex datasets [11], [12].

Moreover, ML techniques can automate and streamline the data cleaning process by identifying and handling missing values, outliers, duplicates, and inconsistencies, thereby improving overall data quality [13]. This automated refinement process is critical for ensuring the integrity of datasets, as models trained on biased or low-quality data can perpetuate and even amplify existing inaccuracies [3]. Furthermore, machine learning algorithms can analyze complex datasets that exceed human analytical capabilities, continuously updating their interpretation of data in response to emerging trends [14]. This continuous improvement distinguishes machine learning from traditional programming paradigms, which require explicit rule sets for every possible scenario [3]. This adaptability allows ML models to effectively process diverse data formats, including structured, semi-structured, and unstructured data, enabling comprehensive

JoDSIE © 2025 Author. Published by CV. Raskha Media Group.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

ISSN: 3123-948X (Media Online)
Journal of Data Science and Informatics Engineering
Vol. 1 No. 1, December 2025 | pp: 1-9
DOI: https://doi.org/10.64803/jodsie.v1i1.12

data integration and analysis [13]. The increasing accessibility of datasets further fuels the application of machine learning across diverse industries, enabling the extraction of valuable insights from data and advancing beyond explicit programming [15]. This enables the development of sophisticated predictive models, anomaly detection systems, and optimization routines, all crucial for various applications [16]. The advantages of machine learning, including its ease of identifying trends and patterns, minimal human intervention, and continuous improvement capabilities, make it exceptionally efficient at handling multidimensional and multivariate data [3].

This ability stems from their capacity for feature extraction and dimensionality reduction, which simplifies complex inputs without losing critical information [17]. These models excel at processing and analyzing vast, intricate datasets with remarkable speed and precision, offering valuable insights and predictive capabilities [18]. Furthermore, machine learning models exhibit significant flexibility, enabling them to effectively capture non-linear relationships within diverse data types and scale efficiently to manage substantial amounts of information [19]. This comprehensive overview demonstrates that machine learning encompasses a range of methodologies, including supervised, unsupervised, and reinforcement learning, each suited to distinct data processing challenges [3]. Supervised learning, for instance, involves training models on labeled datasets to predict outcomes or classify data, while unsupervised learning focuses on discovering hidden patterns and structures within unlabeled data [20]. Reinforcement learning, conversely, allows an agent to learn through trial and error in an interactive environment, utilizing feedback from its actions to optimize decision-making over time [3].

3. RESEARCH METHODS

These diverse learning paradigms contribute to the versatility of machine learning in addressing complex data processing tasks, from predictive analytics to pattern recognition and autonomous system control [21], [22]. The choice of a specific machine learning model depends on various factors, including the task type, data characteristics, and desired outcome [3]. For instance, supervised models are typically employed when historical data with known outcomes are available, enabling the algorithms to learn the mapping from inputs to outputs [3], [8]. Conversely, unsupervised techniques are more appropriate for exploratory data analysis or when labels are scarce, aiming to uncover inherent groupings or structures within the data [23], [24]. Popular supervised learning algorithms include linear regression and neural networks, while unsupervised methods such as K-means and principal component analysis are widely used to discover latent structures [20].

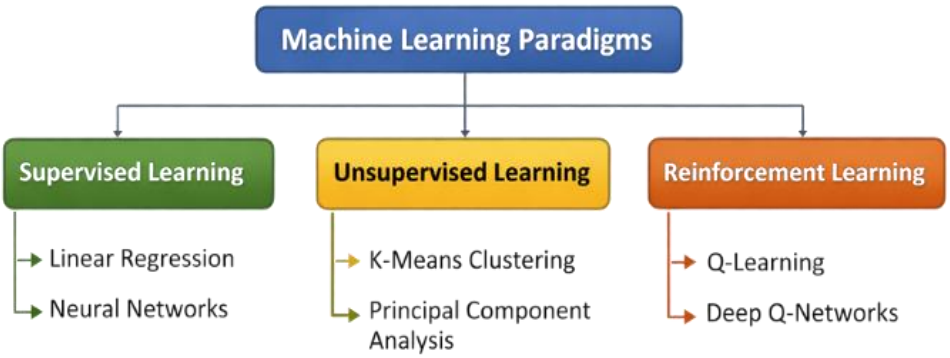


Figure 1. Machine Learning Paradigms [3], [21], [24].

Moreover, reinforcement learning algorithms, such as Q-learning and policy gradients, are particularly effective in dynamic environments where sequential decision-making is crucial [3], [25]. Each of these learning paradigms addresses distinct data-processing challenges, highlighting the adaptability of machine learning across various applications [20], [26]. Specifically, supervised learning models are widely used for tasks such as classification and regression, where the objective is to predict discrete categories or continuous values from labeled input data [27], [28]. These models learn from input-output pairs to generalize patterns and make predictions on new, unseen data points [29]. Conversely, unsupervised learning algorithms are employed for tasks such as clustering and dimensionality reduction, where the aim is to uncover intrinsic structures or representations within unlabeled data without prior knowledge of the

JoDSIE © 2025 Author. Published by CV. Raskha Media Group.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

ISSN: 3123-948X (Media Online)
Journal of Data Science and Informatics Engineering
Vol. 1 No. 1, December 2025 | pp: 1-9
DOI: https://doi.org/10.64803/jodsie.v1i1.12

output variables [30], [31]. This type of learning is especially beneficial when labeled data is scarce or expensive to acquire, enabling algorithms to extract meaningful insights through methods such as clustering similar data points or reducing data dimensionality [32].

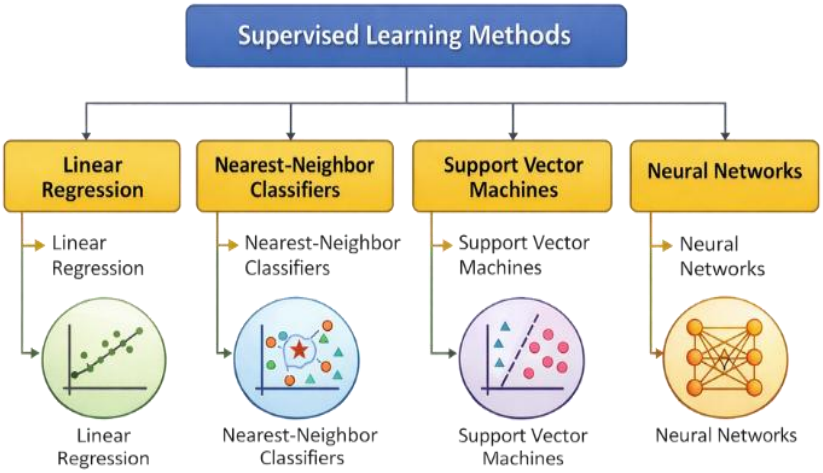


Figure 2. Overview of supervised learning methods [20], [33].

Supervised learning, being one of the most widely used categories, encompasses methods ranging from linear regression and nearest-neighbor classifiers to sophisticated support vector machines and neural networks, often termed artificial neural networks due to their loose biological inspiration [20], [33]. In contrast, unsupervised learning operates without explicit labels, focusing instead on identifying inherent patterns, structures, or anomalies within the dataset [20]. Prominent unsupervised learning techniques include K-means for clustering and Principal Component Analysis for dimensionality reduction, both critical for tasks such as customer segmentation and genetic data analysis [34], [35]. More recent unsupervised methods, such as t-SNE and dictionary learning, further enhance the capabilities for data visualization and feature learning [20]. This can be particularly useful in settings where data labeling is expensive or impractical [20], [36].

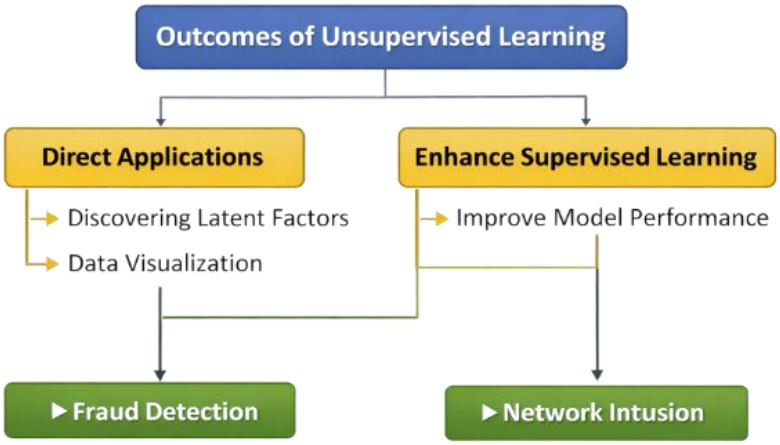


Figure 3. Unsupervised learning outcomes and applications [20], [37], [38].

The outcomes of unsupervised methods can be used either directly to discover latent factors, for data visualization, or integrated into a supervised learning framework to enhance performance [20]. Furthermore, unsupervised learning excels at revealing underlying structures in data and identifying anomalies, which is crucial for applications such as fraud detection and network intrusion [37], [38].

4. RESULTS AND DISCUSSION

JoDSIE © 2025 Author. Published by CV. Raskha Media Group.
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

ISSN: 3123-948X (Media Online)
Journal of Data Science and Informatics Engineering
Vol. 1 No. 1, December 2025 | pp: 1-9
DOI: https://doi.org/10.64803/jodsie.v1i1.12

The subsequent sections provide a detailed analysis of the performance metrics for these machine learning paradigms, offering insights into their strengths and limitations across various data processing scenarios. For instance, model evaluation is crucial for assessing how well a trained model performs on unseen data, using different performance metrics to understand its efficacy [39]. These metrics, such as accuracy, precision, recall, F1-score, and ROC curves, provide a quantitative basis for comparing models and ensuring their robustness and generalizability.

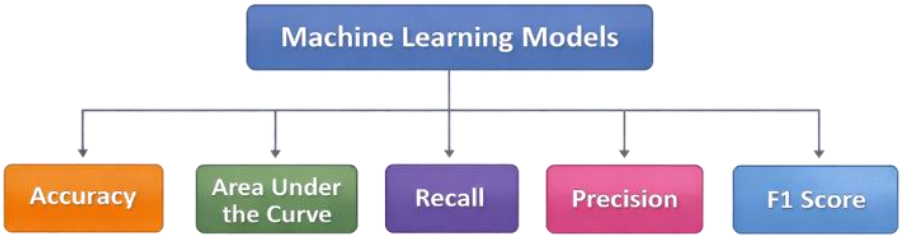


Figure 4. Key machine learning evaluation metrics [45], [46].

The learning phase, fundamental to all machine learning models, involves a systematic process of data preprocessing, model selection, training, validation, and hyperparameter tuning, where data quality and appropriate algorithm selection are paramount for optimal performance [40]. Model training involves iteratively adjusting model parameters using the chosen optimization algorithm and loss function, while validation techniques such as cross-validation ensure the model generalizes effectively to new datasets [41]. This systematic approach ensures the trained model is not only accurate but also reliable and adaptable to varied real-world scenarios [42]. Comparative analysis often integrates supervised and unsupervised techniques to provide a comprehensive evaluation, considering factors such as accuracy, precision, recall, and computational efficiency [42]. Supervised algorithms, such as decision trees and support vector machines, generally require labeled datasets for training and excel at tasks like fraud detection and medical diagnosis [42]. Conversely, unsupervised methods, including clustering and anomaly detection algorithms, are adept at identifying anomalies and discerning inherent patterns in unlabeled data, achieving moderate accuracy rates of 75%-80% [43]. The evaluation of these models necessitates a comprehensive approach that encompasses threshold testing, comparative analysis, robustness against adversarial attacks, scalability, and efficiency, to provide a holistic understanding of their performance across diverse operational contexts [44].

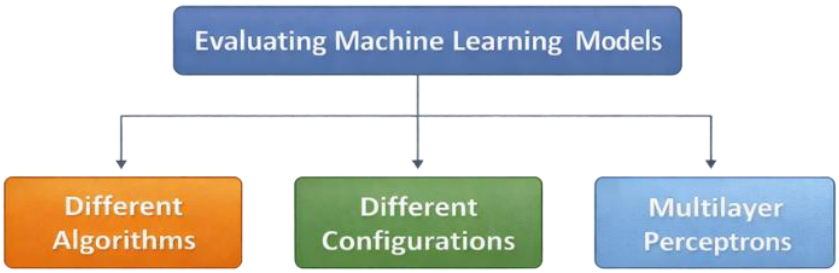


Figure 5. Evaluating machine learning models infographic [45], [46].

This thorough evaluation of machine learning models involves assessing a diverse array of algorithms and configurations, including different Multilayer Perceptrons, across metrics such as Accuracy, Area Under the Curve, Recall, Precision, and F1 Score [45], [46]. These metrics are critical for understanding the model's predictive accuracy, its ability to correctly identify positive instances, and its overall balance between precision and recall, especially in imbalanced datasets [47], [48]. For example, Logistic Regression, Support Vector Machines, Random Forests, and Neural Networks are commonly employed supervised learning models, and their performance is rigorously evaluated using these metrics [49].

5. CONCLUSION

This comprehensive evaluation includes assessing the models' training and testing stages: a labeled training set is used to train the model, followed by rigorous testing to validate its predictive accuracy on unseen instances. This involves a systematic process of tuning parameters and evaluating outputs during the

JoDSIE © 2025 Author. Published by CV. Raskha Media Group.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

ISSN: 3123-948X (Media Online)
Journal of Data Science and Informatics Engineering
Vol. 1 No. 1, December 2025 | pp: 1-9
DOI: https://doi.org/10.64803/jodsie.v1i1.12

training phase to achieve optimal classification results. All six classifiers are implemented utilizing the Python scikit-learn package, ensuring standardized API access and computational efficiency. The evaluation encompasses application-specific preprocessing techniques and uses a range of performance metrics to thoroughly assess each model's efficacy. Customizations are enabled through hyperparameter tuning, such as configuring ensemble tree counts, kernel functions, and k-neighbors, enabling diverse model applications. The consistent performance of algorithms such as Decision Trees and Random Forests across datasets, including ORIGA and G1020, underscores their robustness in classification tasks. Deep learning and gradient boosting machines often achieve superior performance in regression tasks compared to classification, whereas deep learning models frequently outperform GBM in classification across various evaluation metrics. The implementation of these models, particularly decision trees and random forests, often leverages powerful Python libraries such as scikit-learn, TensorFlow, and Keras, facilitating development and robust evaluation on unseen data.

REFERENCES

[1]. Adugna, T. D., Ramu, A., & Haldorai, A. (2024). A Review of Pattern Recognition and Machine Learning [Review of A Review of Pattern Recognition and Machine Learning]. Journal of Machine and Computing, 210. https://doi.org/10.53759/7669/jmc202404020

[2]. Alsajri, A. (2023). Review on Machine Learning Strategies for Real-World Engineering Applications. Deleted Journal, 2023, 1. https://doi.org/10.58496/bjml/2023/001

[3]. Amato, A., & Lecce, V. D. (2023). Data preprocessing impact on machine learning algorithm performance. Open Computer Science, 13(1). https://doi.org/10.1515/comp-2022-0278

[4]. Bezuidenhout, M., Jooste, J. L., Lücke, D., & Fourie, C. J. (2023). Leveraging digitilisation and machine learning for improved railway operations and maintenance. Procedia CIRP, 120, 702. https://doi.org/10.1016/j.procir.2023.09.062

[5]. Bharadiya, J. P. (2023). The role of machine learning in transforming business intelligence. International Journal of Computing and Artificial Intelligence, 4(1), 16. https://doi.org/10.33545/27076571.2023.v4.i1a.60

[6]. Bhaskar, R. (2025). Survey on Personality Detection on Multilingual Dataset using Machine Learning and Explainable AI. International Journal for Research in Applied Science and Engineering Technology, 13(5), 6930. https://doi.org/10.22214/ijraset.2025.71767

[7]. Bianco, M. J., Gerstoft, P., Traer, J., Ozanich, E., Roch, M. A., Gannot, S., & Deledalle, C.-A. (2019). Machine learning in acoustics: Theory and applications. The Journal of the Acoustical Society of America, 146(5), 3590. https://doi.org/10.1121/1.5133944

[8]. Coronado, E., Thomas, A., & Riggio, R. (2020). Adaptive ML-Based Frame Length Optimisation in Enterprise SD-WLANs. Journal of Network and Systems Management, 28(4), 850. https://doi.org/10.1007/s10922-020-09527-y

[9]. Cui, Y., Li, X., Sun, Y., Liu, Y., Zhang, W., Jiang, Y., Zhou, Y., Liu, J., Gong, B. P., Wu, Y., Li, S., Zhuang, L., Cong, W., & Zhang, J. (2025). Separation, characterization, AI screening, and bioactivities of marine bioactive peptides: A review [Review of Separation, characterization, AI screening, and bioactivities of marine bioactive peptides: A review]. Food Chemistry, 495, 146312. Elsevier BV. https://doi.org/10.1016/j.foodchem.2025.146312

[10]. Czech, S. (2023). Evaluating the Role of Machine Learning in Economics: A Cutting-Edge Addition or Rhetorical Device? Studies in Logic Grammar and Rhetoric, 68(1), 279. https://doi.org/10.2478/slgr-2023-0014

[11]. Dang, S., Wei, F., Wu, M., Xie, R., & Wu, J. (2024). Research on Power Load Data Acquisition and Integrated Transmission Systems in Electric Energy Calculation and Detection. EAI Endorsed Transactions on Energy Web, 11. https://doi.org/10.4108/ew.5521

[12]. Dutt, A., & Karuna, G. (2024). Machine learning approaches for fault detection in renewable microgrids. MATEC Web of Conferences, 392, 1192. https://doi.org/10.1051/matecconf/202439201192

[13]. Eshaghi, M. S., Anitescu, C., & Rabczuk, T. (2024). Methods for enabling real-time analysis in digital twins: A literature review [Review of Methods for enabling real-time analysis in digital twins:

JoDSIE © 2025 Author. Published by CV. Raskha Media Group.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

A literature review]. Computers & Structures, 297, 107342. Elsevier BV. https://doi.org/10.1016/j.compstruc.2024.107342

[14]. Feng, P., Bi, Z., Wen, Y., Pan, X., Peng, B., Liu, M., Xu, J., Chen, K., Liu, J., Yin, C. H., Zhang, S., Wang, J., Niu, Q., Li, M., & Wang, T. (2024). Deep Learning and Machine Learning, Advancing Big Data Analytics and

[15]. Management: Unveiling AI's Potential Through Tools, Techniques, and Applications. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2410.01268

[16]. Feng, P., Bi, Z., Wen, Y., Peng, B., Liu, J., Yin, C. H., Wang, T., Chen, K., Zhang, S., Li, M., Xu, J., Liu, M., Pan, X., Wang, J., & Niu, Q. (2024). Mastering AI: Big Data, Deep Learning, and the Evolution of Large

[17]. Language Models -- AutoML from Basics to State-of-the-Art Techniques. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2410.09596

[18]. Hsieh, W. C., Bi, Z., Chen, K., Peng, B., Zhang, S., Xu, J., Wang, J., Yin, C. H., Zhang, Y., Feng, P., Wen, Y., Wang, T., Li, M., Liang, C. X., Ren, J., Niu, Q., Chen, S., Yan, L., Xu, H., … Liu, M. (2024). Deep Learning, Machine Learning, Advancing Big Data Analytics and Management. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2412.02187

[19]. Huertas-García, Á., Martí-González, C., Maezo, R. G., & Rey, A. E. (2023). A Comparative Study of Machine Learning Algorithms for Anomaly Detection in Industrial Environments: Performance and Environmental Impact. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2307.00361

[20]. Huertas-García, Á., Martí-González, C., Maezo, R. G., & Rey, A. E. (2024). A Comparative Study of Machine Learning Algorithms for Anomaly Detection in Industrial Environments: Performance and Environmental Impact. In Algorithms for intelligent systems (p. 373). Springer Nature. https://doi.org/10.1007/978-981-99-9436-6_26

[21]. Islam, A., & Rashid, Md. M. (2024). Cyberattack Detection Using Unsupervised Learning Techniques. Research Square (Research Square). https://doi.org/10.21203/rs.3.rs-4328744/v2

[22]. Islam, M. M., Sharma, A., & Khilji, N. (2025). Utilising Machine Learning Algorithms to Address Computational Challenges in Big Data Analytics. International Journal of Computational and Experimental Science and Engineering, 11(3). https://doi.org/10.22399/ijcesen.3165

[23]. Kamencay, P., Hockicko, P., & Hudec, R. (2024). Sensors Data Processing Using Machine Learning. Sensors, 24(5), 1694. https://doi.org/10.3390/s24051694

[24]. Katyare, P., Joshi, S., & Kulkarni, M. (2024). Utilizing Machine Learning Approach to Forecast Fuel Consumption of Backhoe Loader Equipment. International Journal of Advanced Computer Science and Applications, 15(5). https://doi.org/10.14569/ijacsa.2024.01505121

[25]. Kazbekova, G., Ismagulova, Z., Zhussipbek, B., Abdrazakh, Y., Iskendirova, G., & Toilybayeva, N. (2024). Machine Learning Enhanced Framework for Big Data Modeling with Application in Industry 4.0. International Journal of Advanced Computer Science and Applications, 15(3). https://doi.org/10.14569/ijacsa.2024.0150332

[26]. Kommisetty, P. D. N. K., vijay, A., & rao, M. bhasker. (2024). From Big Data to Actionable Insights: The Role of AI in Data Interpretation. IARJSET, 11(8). https://doi.org/10.17148/iarjset.2024.11831

[27]. Li, J., Wu, A., Liu, L., Qu, A., Xu, C., Kuang, H., & Xu, L. (2025). Analysis of food safety based on machine learning: A comprehensive review and future prospects [Review of Analysis of food safety based on machine learning: A comprehensive review and future prospects]. Food Chemistry, 490, 145170. Elsevier BV. https://doi.org/10.1016/j.foodchem.2025.145170

[28]. Li, L., Wang, J., Wang, X., Peng, P., Shen, J., Zhu, H., & Zhang, Z. (2025). Big data and data science in global governance: anticipating future needs and applications in the UN and beyond. Frontiers in Political Science, 7. https://doi.org/10.3389/fpos.2025.1583772

[29]. Lohit, V. S., Mujahid, M. M., & Sai, G. K. (2022). Use of Machine Learning for Continuous Improvement and Handling Multi-Dimensional Data in Service Sector. Computational Intelligence and Machine Learning, 3(2), 39. https://doi.org/10.36647/ciml/03.02.a006

[30]. Mitrovic, M. (2024). Data-Driven Stochastic AC-OPF using Gaussian Processes. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2402.11365

JoDSIE © 2025 Author. Published by CV. Raskha Media Group.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

[31]. Naren, R., & Subhashini, J. (2020). Comparison of deep learning models for predictive maintenance. IOP Conference Series Materials Science and Engineering, 912(2), 22029. https://doi.org/10.1088/1757-899x/912/2/022029

[32]. Obuse, E., Ajayi, J. O., Oladimeji, O., Erigha, E. D., Akindemowo, A. O., Tafirenyika, S., & Moyo, T. (2023). Comparative Analysis of Supervised and Unsupervised Machine Learning for Predictive Analytics. International Journal of Management and Organizational Research, 2(3), 70. https://doi.org/10.54660/ijmor.2023.2.3.70-86

[33]. Osman, A. I., Abd-Elaziem, W., Nasr, M., Farghali, M., Rashwan, A. K., Hamada, A., Wang, Y., Darwish, M. A., Sebaey, T. A., Khatab, A., & Elsheikh, A. H. (2024). Enhanced hydrogen storage efficiency with sorbents and machine learning: a review [Review of Enhanced hydrogen storage efficiency with sorbents and machine learning: a review]. Environmental Chemistry Letters, 22(4), 1703. Springer Science+Business Media. https://doi.org/10.1007/s10311-024-01741-3

[34]. Pandey, Mrs. A. (2023). Machine Learning. International Journal for Research in Applied Science and Engineering Technology, 11(8), 864. https://doi.org/10.22214/ijraset.2023.55224

[35]. Patel, M., Magre, N., Motwani, H., & Brown, N. B. (2024). Advances in Machine Learning, Statistical Methods, and AI for

[36]. Single-Cell RNA Annotation Using Raw Count Matrices in scRNA-seq Data. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2406.05258

[37]. Pathak, U., & Piyush, Er. (2023). Sentiment Analysis: Methods, Applications, and Future Directions. International Journal for Research in Applied Science and Engineering Technology, 11(2), 1453. https://doi.org/10.22214/ijraset.2023.49165

[38]. Pathmanaban, P., Gnanavel, B. K., Anandan, S. S., & Sathiyamurthy, S. (2023). Advancing post-harvest fruit handling through AI-based thermal imaging: applications, challenges, and future trends. Discover Food, 3(1). https://doi.org/10.1007/s44187-023-00068-2

[39]. Rane, N. L., Mallick, S. K., Kaya, Ö., & Rane, J. (2024). Techniques and optimization algorithms in machine learning: A review [Review of Techniques and optimization algorithms in machine learning: A review]. https://doi.org/10.70593/978-81-981271-4-3_2

[40]. Rayyan, M., Sharifah, N., & Kuswati, R. (2024). Revolutionizing Talent Acquisition in Indonesia's E-Commerce Industry: The Transformative Impact of AI and Machine Learning. Journal of Humanities and Social Sciences Studies, 6(4), 1. https://doi.org/10.32996/jhsss.2024.6.4.1

[41]. Rehan, H. (2024). AI-Driven Cloud Security: The Future of Safeguarding Sensitive Data in the Digital Age. Deleted Journal, 1(1), 47. https://doi.org/10.60087/jaigs.v1i1.p66

[42]. Shoaib, A. S. M. (2024). MACHINE LEARNING AND THE STUDY OF LANGUAGE CHANGE: A REVIEW OF METHODOLOGIES AND APPLICATION [Review of MACHINE LEARNING AND THE STUDY OF LANGUAGE CHANGE: A REVIEW OF METHODOLOGIES AND APPLICATION]. Deleted Journal, 1(2), 48. https://doi.org/10.62304/ijmisds.v1i2.144

[43]. Shuford, J. (2024). Unveiling the Power of Deep Learning: Insights into Advanced Neural n Network Architectures. Deleted Journal, 3(1), 1. https://doi.org/10.60087/jaigs.v3i1.60

[44]. Sivalingam, S. M., & Thisin, S. (2024). A new framework to enhance healthcare monitoring using patient-centric predictive analysis. International Journal of Power Electronics and Drive Systems/International Journal of Electrical and Computer Engineering, 14(3), 3295. https://doi.org/10.11591/ijece.v14i3.pp3295-3302

[45]. Soltaninejad, M., Aghazadeh, R., Shaghaghi, S., & Zarei, M. F. E.-. (2024). Using Machine Learning Techniques to Forecast Mehram Company's Sales: A Case Study. Journal of Business and Management Studies, 6(2), 42. https://doi.org/10.32996/jbms.2024.6.2.4

[46]. Šprem, Š., Tomažin, N., Matečić, J., & Horvat, M. (2024). Building Advanced Web Applications Using Data Ingestion and Data Processing Tools. Electronics, 13(4), 709. https://doi.org/10.3390/electronics13040709

JoDSIE © 2025 Author. Published by CV. Raskha Media Group.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

ISSN: 3123-948X (Media Online)
Journal of Data Science and Informatics Engineering
Vol. 1 No. 1, December 2025 | pp: 1-9
DOI: https://doi.org/10.64803/jodsie.v1i1.12

[47]. Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2014). Machine Learning for Predictive Maintenance: A Multiple Classifier Approach. IEEE Transactions on Industrial Informatics, 11(3), 812. https://doi.org/10.1109/tii.2014.2349359

[48]. Tamascelli, N., Campari, A., Parhizkar, T., & Paltrinieri, N. (2024). Artificial Intelligence for safety and reliability: A descriptive, bibliometric and interpretative review on machine learning. Journal of Loss Prevention in the Process Industries, 90, 105343. https://doi.org/10.1016/j.jlp.2024.105343

[49]. Trinh, V. D. (2025). A Comprehensive Review: Applicability of Deep Neural Networks in Business Decision Making and Market Prediction Investment. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2502.00151

[50]. Yadav, D. K., Kaushik, A., & Yadav, N. (2024). Predicting machine failures using machine learning and deep learning algorithms. Sustainable Manufacturing and Service Economics, 3, 100029. https://doi.org/10.1016/j.smse.2024.100029

[51]. Yan, Y., Borhani, T. N., Subraveti, S. G., Pai, K. N., Prasad, V., Rajendran, A., Nkulikiyinka, P., Asibor, J. O., Zhang, Z., Shao, D., Wang, L., Zhang, W., Yan, Y., Ampomah, W., You, J., Wang, M., Anthony, E. J., Manović, V., & Clough, P. T. (2021). Harnessing the power of machine learning for carbon capture, utilisation, and storage (CCUS) – a state-of-the-art review [Review of Harnessing the power of machine learning for carbon capture, utilisation, and storage (CCUS) – a state-of-the-art review]. Energy & Environmental Science, 14(12), 6122. Royal Society of Chemistry. https://doi.org/10.1039/d1ee02395k

[52]. Zhang, Y., Wang, Q., & Liu, Y. (2021). Adaptive Intelligent Welding Manufacturing. Welding Journal, 100(1), 63. https://doi.org/10.29391/2021.100.006

JoDSIE © 2025 Author. Published by CV. Raskha Media Group.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.