International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 4
ISSN: 2088-5334

Challenges in Supervised and Unsupervised Learning: A Comprehensive Overview

Mohammed Tuays Almuqati [a, b], Fatimah Sidi [a,*], Siti Nurulain Mohd Rum [a], Maslina Zolkepli [a], Iskandar Ishak [a]

[a] *Department of Computer Science, Faculty of Computer Science and Information Technology, University Putra Malaysia, Serdang, Malaysia*
[b] *Deanship of Electronic Transaction and Communication at King Saud University, Riyadh, Saudi Arabia*
*Corresponding author: *fatimah@upm.edu.my*

***Abstract*— Data science and machine learning are at the forefront of modern technological advancements, promising automated insights, predictions, and decision-making. Supervised and unsupervised learning are pivotal paradigms within this dynamic landscape, each presenting its unique challenges. This article provides a comprehensive overview of the multifaceted challenges inherent to both supervised and unsupervised learning. This article reviews research studies published between 2019 and 2023. This article discusses the challenges of supervised and unsupervised learning. In supervised learning, challenges include data labeling, overfitting, limited generalization, and balancing mistake equivalence and decision-making goals. In unsupervised learning, difficulties encompass issues like overfitting, choosing the appropriate algorithm, and interpreting results. This includes evaluating the quality of clustering, deciding the optimal number of clusters, and managing noise and outliers. The article aims to provide insights into these challenges, enhancing the understanding of machine learning for both novices and experts. Researchers and practitioners constantly evolve their methods and tools to overcome these complexities. This article is a valuable reference for researchers and experts in the field, empowering them to navigate these challenges confidently. As technology advances, a thorough understanding of these challenges is essential for unlocking the full potential of these powerful tools. Finally, several recommendations were given to guide future researchers in applying machine learning in the journey of data-driven discovery and automation, offering challenges and opportunities for those who embark on it.**

***Keywords*— Data science; machine learning; supervised learning; unsupervised learning; challenges in machine learning.**

Manuscript received 7 Apr. 2024; revised 22 Jun. 2024; accepted 12 Jul. 2024. Date of publication 31 Aug. 2024.

I. INTRODUCTION

Data science is an emerging multidisciplinary area that combines computing, statistics, and subject-matter expertise. It centers on collecting, analyzing, and interpreting data to extract valuable insights, make informed decisions, and solve complex problems. Data science leverages various computational techniques and tools to uncover hidden patterns, trends, and relationships within datasets. It has widespread applications across multiple industries and domains, making it a critical area of study and practice in the modern era. Several definitions of data science have emerged in recent years due to the field's sudden surge in prominence.

Data science is focused on analyzing data to derive meaningful insights from it [1]. Data science is a file focused on mining massive datasets for meaningful patterns and providing accompanying confidence and error intervals. Data science is a burgeoning field that leverages statistical methodologies and computer science expertise to generate meaningful predictions and insights across various conventional academic domains. Data science is a broad field that draws on computer science, statistics, domain expertise, and other fields to derive meaning from large amounts of raw data. Data science plays a significant role in computer science and involves various aspects, including data collection, storage, processing, analysis, and interpretation.

Given these concepts and issues, this study focuses on the research questions such as:

- What is Data Science? And What is Machine Learning?
- What is the difference between machine learning and artificial intelligence?
- What are the key components of the Machine Learning Process?
- What are the Research Methods in Machine Learning?
- In machine learning, what sets supervised learning apart from unsupervised learning?

- What are the challenges of machine learning methods?

The organization of this paper is as follows. Section 1 introduces the fundamental concepts of data science and machine learning, providing the foundation for the subsequent discussion. It presents a comprehensive literature review, exploring the critical challenges associated with supervised and unsupervised learning based on theoretical and practical research. Section 2 details the research methods employed in this study, covering both supervised and unsupervised learning methods. In Section 3, we delve into the specific challenges encountered in supervised and unsupervised learning, analyzing their impact on the effectiveness of machine learning models. Finally, Section 4 concludes the paper by summarizing the main findings and discussing their significance.

A. What is Machine Learning?

Machine learning is a branch of data science in which computers learn from data and use that knowledge to make inferences and judgments. This encompasses a range of techniques, from basic linear regression to advanced deep learning models. [2]. Machine learning (ML) is a division of artificial intelligence (AI) that instructs computers to analyze data and draw conclusions from large amounts without being specifically programmed. In essence, machine learning empowers computers to improve their performance on a specific task through data analysis rather than relying on explicit instructions [3]. Computer scientists develop and implement machine learning models for diverse purposes, including functions like images.

B. What is the difference between Machine Learning and Artificial Intelligence?

ML and AI are closely related fields, but they are not the same. AI is a broader concept encompassing various technologies and techniques, while ML is a subset of AI [4]. Here's a breakdown of the differences between the two:

1) Artificial Intelligence (AI): AI is the study and development of computational systems and machines that can perform activities generally associated with human intelligence [5]. Some examples of these activities are problem-solving, making decisions, grasping the meaning of natural language, spotting patterns, and gaining knowledge through experience. Artificial intelligence aims to develop systems that can mimic human intellect and carry out tasks that generally require human-level cognition and reasoning.

2) Machine learning (ML): ML, a subset of artificial intelligence, focuses on crafting algorithms and models that empower computers to learn from data and enhance their task performance progressively [6]. ML is a technique used to achieve AI. In other words, it's a tool or approach used within the field of AI to enable systems to learn and adapt. The aim of machine learning is the formulation of algorithms and models that can improve their performance on specific tasks through the analysis of data. For example, ML focuses on specific tasks like image recognition, recommendation systems, natural language processing (NLP), healthcare, financial services, language translation, autonomous vehicles, and more [7].

C. What are the Critical Components of the Machine Learning Process?

Machine learning is a powerful force behind many modern technologies, but what fundamental elements drive it? In this study, we will review the essential components that underpin the machine learning process. Whether you're new to the field or a seasoned expert, understanding these key building blocks is vital for harnessing the potential of machine learning. To grasp the inner workings of machine learning, let's break it down into its core components:

1) Data: Machine learning heavily relies on data. This data can be in text, numbers, images, or any other information a computer can process. The quality and quantity of data play a critical role in the success of a machine learning project. These data types are used for analytical model building and institution decision support. Image data is used for tasks such as object recognition or object counting. Text data is used for sentiment analysis, machine-driven translation, and document summarization [8].

2) Algorithms: Machine learning algorithms are the heart of the process. These algorithms are responsible for learning patterns and relationships in the data. Supervised, unsupervised, and reinforcement learning are just a few examples of the machine learning algorithms at your disposal, each tailored for different types of problems [9],[10].

3) Training: An algorithm is trained using labeled data in supervised learning. This approach allows the algorithm to learn from historical data where the correct answers or labels are known. The primary objective is to enable the algorithm to make accurate predictions when presented with new, unseen data. During training, the algorithm fine-tunes its parameters by comparing its predictions to the actual labels, gradually improving its accuracy.

4) Model: During training, the machine learning algorithm builds a model that encapsulates the patterns and relationships it has learned from the data. This model can then be used for making predictions or decisions [11].

5) Testing and Validation: After training, assessing the model's performance on a distinct dataset it has not encountered previously is essential. This process ensures the model's ability to apply its knowledge to novel, unseen examples [12].

6) Deployment: Once a machine learning model has undergone training and validation, it can be implemented in real-world applications. This might involve integrating it into software, systems, or devices to automate tasks, make predictions, or assist in decision-making [13].

D. Machine Learning Workflow

Developing machine learning models can be complex, involving numerous tasks that require careful planning and execution. The Machine Learning Workflow provides a structured approach to handle these tasks efficiently [14]. This workflow typically includes the following key steps as depicted in Fig. 1:

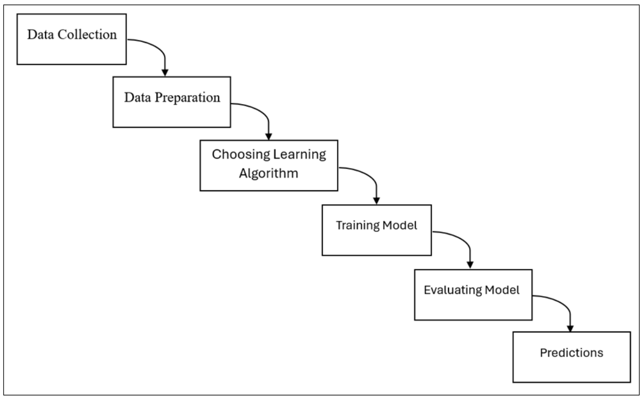

Fig. 1 Diagram Machine Learning Workflow

1) Data Collection: Collecting data is essential because the quality and quantity of data significantly influence the accuracy, reliability, and fairness of machine learning models. High-quality data improves model performance, while poor-quality data can lead to biased or inaccurate outcomes. However, several challenges exist in data collection. Data Quality is dealing with incomplete, noisy, or irrelevant data. Data Quantity comprises the risk of too little or too much data can impact model effectiveness. Access to Data is Privacy concerns and data ownership issues that may restrict access to crucial data. Technical Challenges: Integrating data from various sources and ensuring its relevance to the task is complex [15].

2) Data Preparation: Data preparation in machine learning involves cleaning, transforming, and organizing raw data into a format suitable for effective model training. This process includes tasks such as addressing missing values, normalizing or scaling data, encoding categorical variables, and selecting relevant features. The primary goal of data preparation is to ensure that the data used by the machine learning model is accurate, consistent, and ready for analysis. Challenges in data preparation include handling missing or noisy data, managing data imbalances, selecting appropriate features, and processing large datasets that demand substantial computational resources [16].

3) Choosing Learning Algorithm: Selecting a learning algorithm in machine learning involves choosing the most suitable algorithm to train the model on the available data. This decision depends on the type of problem (e.g., classification, regression), the characteristics of the data, and the desired outcome. Common algorithms include decision trees, support vector machines, neural networks, and k-means clustering. Choosing the right algorithm is crucial because it determines how effectively the model will learn from the data and how accurately it will make predictions. A well-chosen algorithm can lead to improved performance and efficiency [17].

4) Training Model: In the machine learning workflow, training a model is the phase where the selected algorithm is applied to the prepared data to learn patterns and relationships. During this process, the model adjusts its internal parameters based on the input data and corresponding output (in supervised learning) to minimize the error between the predicted and actual values.

Training the model is crucial because it is the step where the model learns how to perform its task, whether it's classifying data, predicting outcomes, or identifying patterns. The effectiveness of this training determines how well the model can generalize from the training data to make accurate predictions on new, unseen data, which is essential for its real-world application [18].

5) Evaluating Model: Evaluating a model within the machine learning workflow entails assessing its performance using a test dataset to gauge how well it generalizes to new, unseen data. This step is essential because it helps verify that the model is not overfitting to the training data and is suitable for real-world deployment. Challenges in model evaluation include managing overfitting, selecting appropriate evaluation metrics, addressing imbalanced data, and applying cross-validation to measure the model's predictive capabilities on new data accurately [18].

6) Predictions: In the machine learning workflow, predictions refer to the final step where the trained model is used to make forecasts or decisions based on new, unseen data. After the model has been trained and evaluated, it is deployed to predict outcomes for new data inputs, which can be anything from classifying an email as spam or not to forecasting stock prices or diagnosing medical conditions. Predictions are important because they are the ultimate purpose of building and training a machine learning model. The model's value is realized when it can accurately and reliably predict new data. These predictions can drive business decisions, automate processes, improve efficiency, and provide insights that were impossible through traditional methods. However, there are several challenges in making predictions, including ensuring that the model generalizes well to new, unseen data (generalization), dealing with changes in underlying data patterns over time that can decrease prediction accuracy (model drift), managing the computational demands of real-time predictions, especially with large datasets or complex models, and handling uncertainty in predictions, which is particularly important in critical fields like healthcare or finance [19].

II. MATERIAL AND METHOD

The rapid advancement of machine learning has brought about significant developments in supervised and unsupervised learning, two fundamental paradigms underpinning a wide range of applications. However, the application and effectiveness of these learning methods are often hindered by various challenges, ranging from the complexities of data labeling and overfitting in supervised learning to the intricacies of clustering and noise management in unsupervised learning. In this study, the comprehensive overview focuses on the key challenges associated with supervised and unsupervised learning in the context of machine learning. It covers both theoretical and practical difficulties that have been identified in previous research studies.

Supervised learning, which relies on labeled data for training, faces significant challenges such as overfitting and the high resource requirement for data labeling. Algorithms like SVM and Random Forest have been widely used, but they also come with their challenges, particularly in terms of scalability and accuracy [20].

Unsupervised learning, which works with raw, unlabeled data, is often challenged by issues such as determining the quality of clustering and dealing with noise in the data. Techniques like K-Means and PCA are commonly employed, yet they struggle with scalability and interpretability [20]. Supervised learning models, while powerful, face significant challenges, such as the necessity for large, labeled datasets and the risk of overfitting, where models perform well on training data but fail to generalize to new data [21]. Unsupervised learning, which deals with raw, unlabeled data, presents unique challenges, such as determining clustering quality and ensuring model interpretability, mainly when dealing with large datasets [21].

To sum up, supervised learning struggles with the need for large, labeled datasets and overfitting, limiting its scalability and generalization, while unsupervised learning encounters challenges in clustering quality, noise management, and interpretability, especially with large, complex datasets. Machine learning involves a structured and systematic approach to conducting research and experiments in the field. It encompasses designing, implementing, and evaluating machine learning models and algorithms. Here is an overview of the research methodology in machine learning:

A. Supervised Learning Method

Supervised learning is a machine learning method where algorithms are trained using labeled data, meaning the input data comes with corresponding correct outputs. The goal is to enable the algorithm to classify new data or make predictions based on patterns it learned during training, as shown in Fig. 2. It is not a methodology but an approach for solving specific types of machine learning problems. Supervised learning can be incorporated into various machine learning methodologies or workflows, where it plays a central role when the project's objective is to predict or classify based on labeled data [21].

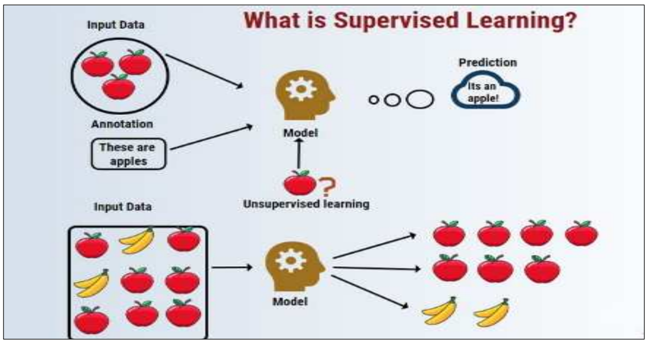

Fig. 2 Supervised Learning Diagram

B. Supervised Learning Algorithms

1) Linear Regression: Linear regression is a foundational algorithm in supervised machine learning used to model the relationship between a dependent variable (target) and one or more independent variables (features). It is primarily applied to solve regression problems, where the objective is to predict a continuous numeric output. Linear regression is commonly used to forecast future outcomes by establishing the relationship between the dependent and independent variables. When there is only one independent variable and one response variable, the model is known as "simple linear regression." In contrast, "multiple linear regression" is used when multiple independent variables are involved. All linear regression techniques utilize the least-squares method to determine the optimal regression line. Unlike other regression models, linear regression produces a straight line when plotted on a graph [21].

2) Logistic regression: Logistic regression is used instead of linear regression because the latter is better suited for continuous dependent variables, while the former is more suited for categorical dependent variables with binary outputs like "true" and "false" or "yes" and "no." Logistic regression is typically used to tackle binary classification issues, such as spam identification, while both graphical and non-graphical regression models strive to discover correlations between data inputs [22].

3) Decision trees: Decision trees are widely used in supervised machine learning for tasks such as classification and regression. The algorithm works by recursively splitting the data into smaller subsets based on the values of input features. Each internal node in the tree represents a decision based on a specific feature, branches illustrate the possible outcomes of those decisions, and the leaf nodes provide the final prediction or classification [23].

4) Support Vector Machine (SVM): The support vector machine, created by Vladimir Vapnik, is a well-known supervised learning model with applications in both data classification and regression. However, its primary application is in classification issues, where it creates a hyperplane in which the gap between two sets of data is the largest. The decision boundary is the hyperplane that divides two sets of data (oranges and apples, for example) into distinct groups [23].

5) Naive Bayes: Naive Bayes is a classification approach founded on the assumption of conditional independence among classes, as posited by the Bayes Theorem. This indicates that the influence of each predictor is equal and that the existence of one trait does not affect the presence of another in the likelihood of a particular event. Three variations of Naïve Bayes classifiers include Bernoulli Naïve Bayes, Multinomial Naïve Bayes, and Gaussian Naïve Bayes. This method is commonly employed in tasks such as text classification, recommendation engines, and spam identification [23].

C. Unsupervised Learning Method

Unsupervised learning depicted in Fig. 3 is another machine learning paradigm that focuses on uncovering patterns, structures, or relationships within unlabeled data. Unsupervised learning can be part of machine learning workflows, especially when the goal is to discover insights from data, and perform clustering, dimensionality reduction, or anomaly detection [24]. Its capability to identify similarities and distinctions in data positions is the perfect solution for tasks like exploratory data analysis, strategies for cross-selling, customer segmentation, and image recognition [25].

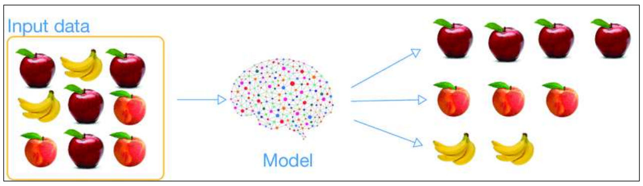

Fig. 3 Unsupervised Learning Diagram

D. *Unsupervised Learning Algorithms*

1) Clustering: Clustering is an approach to organizing data wherein items sharing similar attributes are grouped together within a cluster, while those with distinct characteristics are assigned to separate clusters. Cluster analysis categorizes data objects into groups according to whether or not they share characteristics [26].

2) Association*:* An association rule is a form of unsupervised learning that can be used to uncover hidden correlations within a dataset. It identifies the groups of data points that frequently appear together. Adhering to the standards established by an organization can enhance promotional effectiveness. Customers who buy X (let's say bread) are also likely to buy Y (let's say butter and jam). It is common practice to use market basket analysis to illustrate the association rule [26].

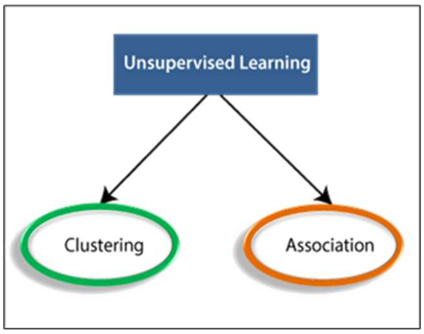

Fig. 4 Unsupervised Learning Algorithms

The most often used unsupervised learning algorithms depicted in Fig. 4 are as follows:

- *K-Means:* K-Means is a clustering algorithm used in unsupervised machine learning. It groups data points into "k" clusters, where "k" is a number you specify. The goal is to assign each data point to the cluster with the closest center (or centroid), thereby minimizing the distance between the data points and their respective centroids. The algorithm iteratively adjusts the centroids and reassigns data points until the clusters are as tight and distinct as possible [27].
- *Hierarchical Clustering: It divides* data into a hierarchy of clusters, often represented as a tree-like structure (dendrogram), where similar data points are grouped. Initially, every data point is designated as an individual, distinct cluster. Through processes such as combining two datasets, assigning data to an existing cluster, or merging two clusters iteratively, a novel cluster can be created [28].
- *DBSCAN (Density-Based Spatial Clustering of Applications with Noise):* In machine learning, DBSCAN is a clustering algorithm that distinguishes between high-density and low-density clusters. DBSCAN is a density-based clustering technique, a powerful tool for locating clusters of similar observations in large datasets. The primary aim of the DBSCAN algorithm is to identify clusters of densely packed points separated by regions of low-density points [29].

III. Results and Discussions

Table I concisely compares supervised and unsupervised learning, highlighting the key distinctions in terms of data, goals, objectives, algorithms, typical applications, and examples.

TABLE I
THE DIFFERENCE BETWEEN SUPERVISED LEARNING AND UNSUPERVISED LEARNING IN MACHINE LEARNING

Aspect	Supervised Learning	Unsupervised Learning
Input Data	The algorithm is learned using labeled data, where input data is associated with corresponding output labels. It employs these labels to learn and make generalizations [22].	The algorithm is trained on unlabeled data, where there are no predefined output labels. Focuses on exploring data characteristics [22].
Goal	To learn a mapping function that can predict the output labels for new, unseen input data [22].	To identify patterns, relationships, and structures in the data without relying on prior knowledge [22].
Objective	Make predictions or classifications.	Discover patterns, structures, or relationships.
Algorithm	It gains insights from labeled data by minimizing the disparity between the predicted output and the actual output labels [22].	It learns from the data by finding inherent structures or clusters in the data [22].
Common Applications	Image classification, sentiment analysis.	Customer segmentation and data visualization.
Examples	Classification and regression tasks.	Clustering, dimensionality reduction, anomaly detection.

In machine learning, the challenges in supervised learning, where models rely on labeled data for predictions, and unsupervised learning, which uncovers hidden patterns within unlabeled data, are critical and multifaceted. This study delves into the intricacies of these challenges, offering insights into the complexities that practitioners and researchers encounter in their quest to harness the full potential of these approaches. Here is a compilation of several challenges encountered in supervised and unsupervised learning domains.

A. *In Supervised Learning*

- Challenges in supervised learning include the assumption of mistake equivalence (ME) and one-sided decision (OD) at the model construction level, which may not always align with decision-making goals.

- Supervised learning presupposes an impartial and just definition of success, which may not be possible in some instances.
- Supervised learning algorithms can be computationally expensive, requiring significant computational resources [30].
- Data Labeling: The most significant challenge in supervised learning is the need for labeled data. Obtaining and annotating a large dataset can be time-consuming and expensive.
- Overfitting: Supervised models are prone to overfitting, where the model learns noise in the training data rather than the underlying patterns. This requires careful regularization and model selection.
- Limited Generalization: Models in supervised learning tend to perform well on data similar to the training set but may struggle to generalize to new or unseen data.
- Imbalanced Data: Dealing with imbalanced datasets, where one class significantly outweighs the other, can lead to biased models and difficulty making accurate predictions.

B. *In Unsupervised Learning*

- Unsupervised learning faces challenges related to overfitting, as different unsupervised approaches have varying degrees of proneness to overfitting.
- The choice of algorithm and decision-making assumptions can influence the performance of unsupervised learning compared to supervised learning.
- Unsupervised learning could carry a lower risk of overfitting in contrast to certain types of supervised learning; however, the results are inconclusive [31].
- Clustering Quality: Evaluating the quality of clusters in unsupervised learning can be subjective, and there may be no ground truth to compare against.
- Determining the Number of Clusters: Selecting the optimal number of clusters is an unsolved problem and can greatly impact the quality of clustering results.
- Interpreting Results: Unsupervised learning often produces results that lack clear interpretations. Understanding the meaning and significance of discovered patterns can be challenging.
- Scalability: Some unsupervised learning algorithms, such as hierarchical clustering or spectral clustering, can be computationally expensive and may not scale well to large datasets.
- Noise and Outliers: Unsupervised learning methods are sensitive to noisy data and outliers, which can lead to incorrect or misleading results.
- Feature Extraction: In dimensionality reduction tasks, determining the optimal representation of the data can be challenging, and the chosen representation may not always be suitable for the specific task.

IV. Conclusion

This study explores the challenges in supervised and unsupervised learning, which are fundamental paradigms within the realm of data science and machine learning. Supervised learning faces difficulties related to data labeling, overfitting, balancing mistake equivalence with decision-making goals, and computational demands, among other complexities. On the other hand, unsupervised learning grapples with challenges like overfitting, algorithm selection, subjective cluster quality evaluation, determining the optimal number of clusters, and interpreting results.

These challenges in data science and machine learning are not roadblocks but stepping stones for innovation and progress. Researchers and practitioners constantly evolve their methods and tools to overcome these complexities. This study is a valuable reference for researchers and experts in the field, empowering them to navigate these challenges confidently. As technology advances, a thorough understanding of these challenges is essential for unlocking the full potential of these powerful tools. The journey of data-driven discovery and automation is ongoing, offering both challenges and opportunities for those who embark on it.

Acknowledgment

This study was conducted independently without funding.

References

[1] M. L. Brodie, "What Is Data Science?," *Applied Data Science*, pp. 101–130, 2019, doi: 10.1007/978-3-030-11821-1_8.

[2] A. Jain, Y. Sarma, K. Kishor , "Financial supervision and management system using Ml algorithm*". Solid State Technology,* vol 63 no.6, pp.18974–18984, 2020.

[3] P. B. Pires and J. D. Santos, "Artificial Intelligence and Marketing," Confronting Security and Privacy Challenges in Digital Marketing, pp. 95–118, Jun. 2023, doi: 10.4018/978-1-6684-8958-1.ch005.

[4] V. Bindu and C. Thomas, "Significance of Feature Selection and Pruning Algorithms in Machine Learning Classification of E-Mails," *Artificial Intelligence for Cyber Security: Methods, Issues and Possible Horizons or Opportunities*, pp. 29–50, 2021, doi:10.1007/978-3-030-72236-4_2.

[5] U. Cali, M. Kuzlu, M. Pipattanasomporn, J. Kempf, and L. Bai, "Foundations of Big Data, Machine Learning, and Artificial Intelligence and Explainable Artificial Intelligence," *Digitalization of Power Markets and Systems Using Energy Informatics*, pp. 115–137, 2021, doi: 10.1007/978-3-030-83301-5_6.

[6] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN computer science*, Mar. 2021, doi:10.20944/preprints202103.0216.v1.

[7] X. Wu, L. Xiao, Y. Sun, J. Zhang, T. Ma, and L. He, "A survey of human-in-the-loop for machine learning," Future Generation Computer Systems, vol. 135, pp. 364–381, Oct. 2022, doi:10.1016/j.future.2022.05.014.

[8] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," Electronic Markets, vol. 31, no. 3, pp. 685–695, Apr. 2021, doi: 10.1007/s12525-021-00475-2.

[9] G. Ciaburro and G. Iannace, "Machine Learning-Based Algorithms to Knowledge Extraction from Time Series Data: A Review," Data, vol. 6, no. 6, p. 55, May 2021, doi: 10.3390/data6060055.

[10] P. C. Sen, M. Hajra, and M. Ghosh, "Supervised Classification Algorithms in Machine Learning: A Survey and Review," Emerging Technology in Modelling and Graphics, pp. 99–111, Jul. 2019, doi:10.1007/978-981-13-7403-6_11.

[11] J. Wexler, M. Pushkarna, T. Bolukbasi, M. Wattenberg, F. Viegas, and J. Wilson, "The What-If Tool: Interactive Probing of Machine Learning Models," IEEE Transactions on Visualization and Computer Graphics, pp. 1–1, 2019, doi: 10.1109/tvcg.2019.2934619.

[12] X. Xie, J. W. K. Ho, C. Murphy, G. Kaiser, B. Xu, and T. Y. Chen, "Testing and validating machine learning classifiers by metamorphic testing," Journal of Systems and Software, vol. 84, no. 4, pp. 544–558, Apr. 2011, doi: 10.1016/j.jss.2010.11.920.

[13] A. Paleyes, R.-G. Urma, and N. D. Lawrence, "Challenges in Deploying Machine Learning: A Survey of Case Studies," ACM Computing Surveys, vol. 55, no. 6, pp. 1–29, Dec. 2022, doi:10.1145/3533378.

[14] A. Aldelemy and Raed A. Abd-Alhameed, "Binary Classification of Customer's Online Purchasing Behavior Using Machine Learning," Journal of Techniques, vol. 5, no. 2, pp. 163–186, Jun. 2023, doi:10.51173/jt.v5i2.1226.
[15] G. Vidhya, D. Nirmala, and T. Manju, "Quality challenges in Deep Learning Data Collection in perspective of Artificial Intelligence," Journal of Information Technology and Computing, vol. 4, no. 1, pp. 46–58, Jun. 2023, doi: 10.48185/jitc.v4i1.725.
[16] H. Nugroho, "A Review: Data Quality Problem in Predictive Analytics," IJAIT (International Journal of Applied Information Technology), vol. 7, no. 02, p. 79, Aug. 2023, doi:10.25124/ijait.v7i02.5980.
[17] Q. Ling, "Machine learning algorithms review," Applied and Computational Engineering, vol. 4, no. 1, pp. 91–98, May 2023, doi:10.54254/2755-2721/4/20230355.
[18] C. Chai, J. Wang, Y. Luo, Z. Niu, and G. Li, "Data Management for Machine Learning: A Survey," IEEE Transactions on Knowledge and Data Engineering, pp. 1–1, 2022, doi: 10.1109/tkde.2022.3148237.
[19] A. Bianchi et al., "Trustworthy Machine Learning Predictions to Support Clinical Research and Decisions," 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS), Jun. 2023, doi: 10.1109/cbms58004.2023.00222.
[20] A. Nagar and J. Sharma, ''A comparative evaluation of supervised learning and unsupervised learning,'' Resmilitaris, vol. 12, no. 5, pp. 1498–1507, 2022. [Online]. Available: https://resmilitaris.net/menuscript/index.php/resmilitaris/article/view/2285.
[21] H. Nugroho, "A Review: Data Quality Problem in Predictive Analytics," IJAIT (International Journal of Applied Information Technology), vol. 7, no. 02, p. 79, Aug. 2023, doi:10.25124/ijait.v7i02.5980.
[22] T. Talaei Khoei, H. Ould Slimane, and N. Kaabouch, "Deep learning: systematic review, models, challenges, and research directions," Neural Computing and Applications, vol. 35, no. 31, pp. 23103–23124, Sep. 2023, doi: 10.1007/s00521-023-08957-4.
[23] G. Villa, C. Tipantuña, D. S. Guamán, G. V. Arévalo, and B. Arguero, "Machine Learning Techniques in Optical Networks: A Systematic Mapping Study," IEEE Access, vol. 11, pp. 98714–98750, 2023, doi:10.1109/access.2023.3312387.
[24] IBM Cloud Education. "Supervised Learning". https://www.ibm.com/topics/supervised-learning, 19 August 2020.
[25] B. Mahesh, "Machine Learning Algorithms - A Review," International Journal of Science and Research (IJSR), vol. 9, no. 1, pp. 381–386, Jan. 2020, doi: 10.21275/art20203995.
[26] IBM. What is Unsupervised Learning?. Accessed: Mar. 15, 2022. [Online]. Available: https://www.ibm.com/topics/unsupervised-learning.
[27] SR. Jena. "Machine Learning Concepts for Beginners: Theory and Applications". Scholars' Press, 2023.
[28] E. J. Salazar, M. E. Samper, and H. D. Patiño, "Dynamic customer demand management: A reinforcement learning model based on real-time pricing and incentives," Renewable Energy Focus, vol. 46, pp. 39–56, Sep. 2023, doi: 10.1016/j.ref.2023.05.004.
[29] T. C. Phan, A. Pranata, J. Farragher, A. Bryant, H. T. Nguyen, and R. Chai, "Machine Learning Derived Lifting Techniques and Pain Self-Efficacy in People with Chronic Low Back Pain," Sensors, vol. 22, no. 17, p. 6694, Sep. 2022, doi: 10.3390/s22176694.
[30] S. Aslangul, "Detecting Tunnels for Border Security based on Fiber Optical Distributed Acoustic Sensor Data using DBSCAN," Proceedings of the 9th International Conference on Sensor Networks, 2020, doi: 10.5220/0008869600780084.
[31] Y. S. Aljarbou and F. Haron, "Determining the Presence of Metabolic Pathways using Machine Learning Approach," International Journal of Advanced Computer Science and Applications, vol. 11, no. 8, 2020, doi: 10.14569/ijacsa.2020.0110845.