



Journal of Artificial intelligence and Machine Learning

Journal homepage: www.sciforce.org

Machine Learning Algorithms: A Comparative Study on Efficiency, Effectiveness, and Practical Applications Using the GRA Method

Akhilesh Reddy Eppa*

*Sr. AI Engineer, master's in Big Data Systems at Arizona State University, Maryland, USA

ARTICLE INFO

Article history:

Received: 20250227

Received in revised form: 20250303

Accepted: 20250303

Available online: 20250311

Keywords:

Unsupervised Learning;

Reinforcement Learning;

Clustering;

Generalization & Over fitting;

Decision Trees & Random Forests;

Neural Networks & Deep Learning;

Ensemble Methods;

Medical Imaging & Cyber security.

ABSTRACT

Machine learning has emerged as a powerful tool for data analysis, enabling systems to identify patterns and make predictions without explicit programming. Among its various approaches, unsupervised learning plays a crucial role in discovering hidden structures within data, especially in scenarios where labeled examples are scarce or costly to obtain. This study provides a comprehensive analysis of unsupervised learning techniques, with a particular focus on clustering and reinforcement learning.

This study provides an in-depth exploration of unsupervised learning techniques, emphasizing their ability to identify patterns in data without the need for labeled training examples. This is particularly significant in domains where acquiring labeled data is costly or impractical. By highlighting the role of reinforcement learning in unsupervised systems, the research advances the understanding of how agents improve behavior through rewards and penalties, which has implications for robotics and strategic game-playing applications.

Other options include K-Nearest Neighbors (KNN), Neural Networks, Support Vector Machines (SVM), and Decision Trees. Assessment Criteria: Memory Usage, Accuracy, Training Speed, and Error Rate.

According to the results, K-Nearest Neighbors (KNN) had the lowest quality, while neural networks had the highest quality.

According to the GRA approach, neural networks are the most valuable datasets for machine learning algorithms.

© Akhilesh Reddy Eppa.

*Corresponding author. e-mail: akhieppa@gmail.com

Introduction

This work provides a thorough analysis of unsupervised learning methods, emphasizing how they can detect structures and patterns in data without the need for labeled events. This is particularly helpful when obtaining labeled data is difficult or too expensive. The discussion highlights the operation of reinforcement learning in unsupervised systems, which shows how agents can improve their behavior in response to rewards and punishments. This perspective advances our understanding of how unsupervised learning can be applied to complex domains such as robotics and game playing. A major focus is on clustering as a method of unsupervised learning, which demonstrates its ability to group similar data points. It has useful applications in domains such as social media and marketing, where it is necessary to explore user behavior or customer groups. The paper also explores the challenges of over fitting

and generalization in machine learning models, underscoring the importance of developing robust algorithms that maintain robust performance on unobserved data. Generalization is crucial for applying machine learning to real-world applications, ensuring that it is used effectively. In addition, the success of unsupervised learning in creating AI systems that can outperform human expertise in games like backgammon highlights its potential for improved strategic decision-making. Finally, the paper explores how unsupervised learning algorithms extract meaningful structures from data, a skill essential for applications in information retrieval and computer vision. This combination expands our knowledge of how data can be used to make more informed decisions and gain better decision-making capabilities. [1,20] Reinforcement learning

teaches models to make consistent decisions by rewarding desired behaviors and punishing undesirable ones.

Cyber security: Improves security systems through real-time threat detection and anomaly detection. Healthcare: Supporting clinical decision-making and clinical data analysis for improved patient care. E-commerce: Personalizing shopping experiences by recommending products based on consumer behavior. Agriculture: Improving crop yield predictions and resource management using data-driven insights. [2,19] A popular technique for classification tasks is a decision tree, which divides data into subsets according to input feature values and then makes choices in the form of a tree. The branches show possible outcomes, and each node represents a choice based on an attribute. The process continues until the final classification result - a leaf node - is reached. Information gain, Gini index, chi-square, and entropy are important statistical measures in decision trees that help identify the best feature for data partitioning at each node. Binary trees are created by partitioning the data at each node according to the feature that best separates the target variable in a particular type of decision tree used for classification and regression tasks. The ID3 algorithm follows a top-down, greedy approach, selecting features that maximize information gain at each step. Advanced versions of ID3 improve on it by accommodating both categorical and continuous data while using more refined feature selection methods. [3,18] They determine which hyper plane in the feature space best separates different classes. Decision trees create a decision model that resembles a tree by partitioning the data according to the input feature values.

They are often used for classification tasks due to their intuitiveness and ease of understanding. The K-Nearest Neighbors (KNN) approach is useful for both classification and regression, as it classifies a data point according to the majority class of its nearest neighbors. Naive Bayes makes the assumption that features are fairly independent of each other, based on Bayes' theorem. Adari et al. has been published for their study on building a connected banking ecosystem through interoperability and data modernization in the International Journal of Computer Engineering and Technology, highlighting frameworks essential for data-driven transformation in financial services. [4,17] In large datasets, a priori methods find common item sets and association rules. A Markov decision process is a mathematical model for decision making that incorporates both randomness and controllable factors. A rule-based learning approach that uses logic programming to represent training data and input examples. Very useful in pattern recognition and computer vision applications. Techniques that reduce the need for human interaction by assisting models in automatically learning representations for feature selection and classification. Natural selection serves as an inspiration for optimization methods used to solve constrained and unconstrained problems.

Used in recommender systems to recommend related items, based on learned similarity criteria. Mishra, et al. apply the COPRAS method to optimize IoV system selection, presenting a detailed case study for real-world application. [5,16] By combining multiple models, ensemble methods increase prediction accuracy. This makes them very useful when individual models, such as regression and classification trees, show inconsistency. Compared to any single model in the ensemble, the combined model is more reliable and generally produces lower error rates. Many machine learning techniques originally developed for binary classification have been adapted to solve multi-class situations. For example, multi-class classification is broken down into multiple binary problems using SVM and boosting algorithms. These problems are then solved and combined to produce comprehensive predictions. In ecological research, machine learning (ML) approaches are increasingly used to help with tasks including analyzing ecological dynamics, predicting species distribution, and determining habitat suitability. These applications demonstrate how flexible and successful machine learning is in addressing difficult environmental problems. While machine learning (ML) provides robust solutions, it cannot solve all environmental data problems. Model performance can be affected by environmental complexity and potential data quality constraints. Therefore, rigorous validation and careful model selection are essential to generate reliable insights. [6,15] A powerful machine learning technique called Random Forest combines multiple decision trees to increase classification accuracy. The most frequently selected class becomes the final prediction in this "forest" of decision trees, each of which provides a vote. This method reduces over fitting, a problem that traditional models often face. Random Forest is popular for classification and regression tasks due to its excellent classification accuracy.

It is a reliable choice for practical applications because it excels at handling noisy data and outliers. Due to its adaptability and efficiency in handling complex information, Random Forest is widely used in many fields, including data mining and biological research. [7,14] These complex algorithms help machines mimic human learning processes and continually improve their understanding to perform better. They play a significant role in research in various domains, including financial forecasting, medical diagnosis, and weather forecasting. BP Neural Network: This type of artificial neural network is good at collecting complex data relationships because it is trained through back propagation. They are frequently used for classification and regression tasks as supervised learning models, and they perform particularly well in high-dimensional spaces. Neural Network Learning Algorithms Single-hidden-layer feed-forward neural network learning algorithms are known for their high generalization ability and fast learning speed. This performance makes it well suited for handling a variety of data-intensive applications. Benefits of Model Combination: Research suggests that

combining multiple models often leads to improved forecasting performance over using a single model. Empirical studies have shown that hybrid models improve accuracy in applications such as container performance analysis and rainfall forecasting. Performance Evaluation: Machine learning algorithms are evaluated through simulation experiments on various datasets. Findings consistently indicate that combined models outperform individual models in both prediction accuracy and overall performance. [8,13] It performs exceptionally well in predictive modeling and pattern recognition, especially on complex datasets. In medical imaging, deep learning improves prediction accuracy by processing the data through multiple layers of abstraction.[9, 11] The process begins with medical images being fed into machine learning algorithms, followed by segmentation to isolate regions of interest. Features are extracted, noise is filtered, and classifiers are used to make predictions.

An alternative approach, pixel-based analysis, directly estimates pixel values instead of relying on feature extraction, simplifying the process and reducing potential errors. Machine learning algorithms play a key role in diagnosing diseases through medical imaging. They help in detecting lesions and organs that are challenging to analyze using conventional methods. Vagvala, et al. conducted a comparative analysis of ERP system separation strategies, introducing the EDAS method as a tool for automation performance benchmarking during enterprise migration. [10,12]

MATERIALS AND METHOD

The goal of multi-attribute decision making (MADM) is to select the best option from a variety of options that are evaluated using various quantitative and qualitative criteria. Researchers from various fields have paid increasing attention to MADM in recent years.

Gray system theory in decision making

Mathematical modeling of systems with partial information greatly benefits from the use of gray system theory, a powerful method for understanding uncertainty. A system is classified according to the amount of information it has access to as follows:

- A black system is when no information is accessible, a white system is when all information is known.
- A gray system is when some details are known but others are unknown.

Grey prediction, grey correlation analysis (GRA), grey decision, grey programming and grey control are the five primary subfields of grey systems theory. Of these, grey correlation analysis (GRA) is most helpful in managing complex relationships between a large number of variables and factors. It is often used to solve problems involving uncertainty where there is no unique and sufficient data.

Applications and Benefits of GRA

GRA is one of the most commonly used techniques to analyze relationships within discrete datasets and to aid in multi-attribute decision making. Its main advantages are as follows:

- Direct use of original data without transformation.
- Simple and efficient calculations.
- High performance in business decision-making environments.

Gray structure theory has been widely used in many domains since it was first introduced by Deng in 1982 (Lin, Chang, & Chen, 2006). It has shown excellent performance in dealing with insufficient, ambiguous, and partial information. GRA has been effectively used in many MADM applications, for example;

- Choosing who to hire (Olson & Wu, 2006).
- Reorganizing the power distribution system (Chen, 2005).
- Jiang, Daci, and Wang (2002) conducted an investigation of the integrated circuit coding process.
- Hierarchical modeling of quality functions (Wu, 2002).
- Silicon wafer slicing defect detection (Lin et al., 2006).

GRA process for MADM

GRA simplifies MADM problems by integrating the performance attribute values of different alternatives into a single comparative value, effectively reducing the multi-attribute problem to a straightforward decision-making task. The process involves:

- In other words, gray correlation is the process of generating a comparative assessment using the performance data of all options.
- The definition of a reference ranking is the process of generating a criterion ranking that reflects the best goal.
- The calculation of the gray correlation coefficient measures the relationship between the reference ranking and the performance ranking of the alternatives.
- When making a decision, the option that most closely matches the reference ranking or has the highest gray correlation score is selected as the best option.

Alternative

Decision Tree

A decision tree is a machine learning model used for classification and regression tasks. It organizes decisions and their possible outcomes in a hierarchical, tree-like structure,

where internal nodes represent features, branches represent decision rules, and leaf nodes represent final predictions. This model repeatedly divides data into smaller subsets based on feature values, improving prediction accuracy. Decision trees are widely favored for their interpretability and visual clarity, making them useful in fields such as healthcare and finance. They efficiently handle both numerical and categorical data, providing a logical representation of decision-making processes.

Random Forest

A random forest is an ensemble learning technique designed for classification and regression tasks. It builds multiple decision trees during training and combines their predictions to improve accuracy and reduce overfitting. By training each tree on a randomly selected subset of data, the model improves diversity and robustness. Average predictions reduce variance, often leading to better performance compared to individual decision trees. Additionally, Random Forest estimates feature importance, aiding in model interpretation. It is widely used in various domains, including healthcare, where it helps in diagnosing conditions such as glaucoma and diabetic retinopathy.

Support Vector Machine (SVM)

SVM is a supervised machine learning algorithm used primarily for classification tasks. It determines the optimal hyper plane that divides data points into different classes in a high-dimensional space, increasing the margin between classes for improved generalization.

K-Nearest Neighbors (KNN)

KNN is a simple but effective supervised machine learning algorithm used for classification and regression tasks. It classifies new data points based on the majority class of their nearest neighbors, which is very intuitive but computationally intensive for large datasets.

Neural Network

Neural networks are computational models inspired by the human brain, designed to recognize patterns and solve complex problems efficiently. They are particularly useful in deep learning applications, enabling advances in image recognition, natural language processing, and autonomous systems.

Beneficial parameters (higher values are better):

Accuracy (%) – Measures the prediction accuracy of the model.

Training speed (events/second) – Indicates how fast the model is trained.

Non-beneficial parameters (lower values are better):

Error rate (%) – Measures the percentage of incorrect predictions.

Memory usage (MB) – Indicates the computational cost in terms of memory.

ANALYSIS AND DISSECTION

TABLE 1. Machine Learning Algorithms

Algorithm	DATA SET			
	Accuracy	Training Speed	Error Rate	Memory Usage
Decision Tree	85	1500	15	50
Random Forest	92	1200	8	100
Support Vector Machine (SVM)	89	800	11	200
K-Nearest Neighbors (KNN)	86	500	14	150
Neural Network	94	1000	6	300
	zeta	0.5		

Table 1 compares several machine learning methods based on four important performance metrics: accuracy, training speed, error rate, and memory usage. These metrics help you choose the best model for different applications by highlighting

the tradeoffs of each approach. The neural network stands out with its very high accuracy (94), making it very effective at making correct predictions. It also has a very low error rate (6), which reinforces its strong reliability.

However, its training speed (1000) is slower than some models, and it requires high memory usage (300), making it computationally demanding. Random Forest offers a balance between accuracy (92) and performance. With a relatively low error rate (8) and fast training speed (1200), it is a strong performer for many tasks. Its memory usage (100) is modest—less than a neural network but more than some other models—making it an efficient choice. The Support Vector Machine (SVM) provides solid accuracy (89) and a moderate error rate (11). However, its training speed (800) is slower than that of Random Forest and Neural Network, which may make it less ideal for large datasets. In addition, it has a relatively high memory requirement (200), which means that it may not be the most resource-efficient option.

The Decision Tree is the fastest model to train (1500), making it more efficient for applications that require rapid decision-making. However, its accuracy (85) is lower than other models, and it has a very high error rate (15), which indicates a trade-off between speed and prediction reliability. On the other hand, its memory usage (50) is very low; making it suitable for resource-constrained environments-Nearest Neighbors (KNN) achieves slightly higher accuracy (86) than Decision Tree, but has a similar error rate (14). However, it has the lowest training speed (500) and relatively high memory usage (150) of all models, which may limit its performance for large-scale tasks. The zeta value (0.5), although not explicitly defined, may represent a weighting factor or parameter that affects the performance evaluation in these metrics.

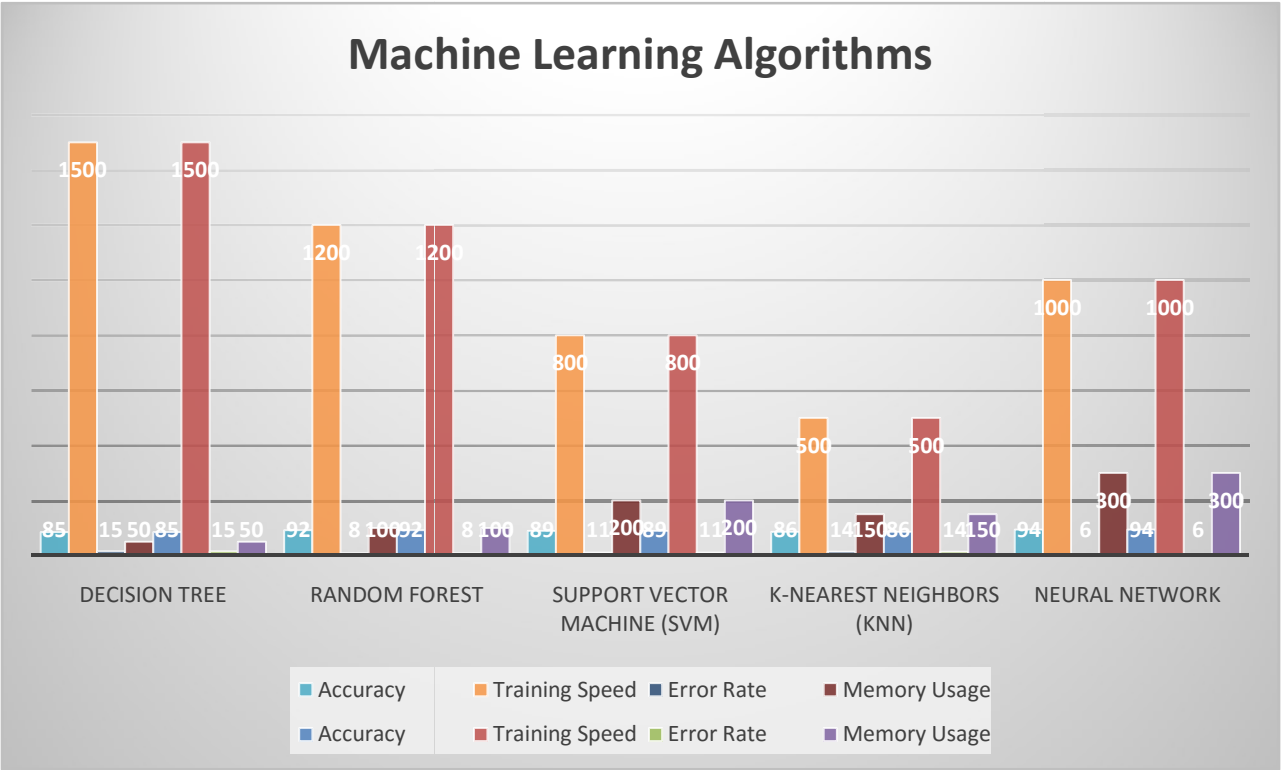


FIGURE 1 . Machine Learning Algorithms

Five machine learning algorithms—decision tree, random forest, support vector machine (SVM), K-nearest neighbors (KNN), and neural network—are compared in Figure 1 based on four important performance metrics: accuracy, training speed, error rate, and memory usage. A bar chart illustrates the performance of each algorithm on these metrics, clearly showing its relative advantages and disadvantages.

The graph illustrates the performance of decision trees and random forests on model training, showing that they achieve the highest training speed with a value of 1500. In addition, they

maintain low error rates (approximately 85–92), demonstrating reliable prediction performance. However, their memory usage is relatively high, which may affect their feasibility in environments with limited computational resources. The Support Vector Machine (SVM) exhibits a moderate training speed (800), which places it between the fast decision tree and the random forest and the slow models. However, its error rate (109) is higher than the best performing algorithms, indicating lower prediction accuracy. Furthermore, its memory usage (800) is considerable, suggesting significant computational

requirements during both training and inference. Of the five models, K-Nearest Neighbors (KNN) has the slowest training speed (500), making it very slow to train. It also records the highest error rate (145), indicating weak prediction performance. However, its memory usage (500) is lower than many other algorithms, which may make it a viable option when computational efficiency is a priority.

The neural network stands out as the most effective in terms of accuracy (94), reflecting its strong prediction ability.

However, this advantage comes with a high memory usage (1000), which is computationally intensive. Also, due to its moderate training speed (1000), it takes longer to train than Decision Tree and Random Forest. While neural networks offer improved accuracy with reduced memory usage, Decision Trees and Random Forests offer faster training speed with respectable accuracy. However, SVM and KNN perform poorly and have large error rates, which makes them less desirable in this evaluation.

TABLE 2. Normalized Data

	Normalized Data			
	Accuracy	Training Speed	Error Rate	Memory Usage
Decision Tree	0.0000	1.0000	0.0000	1.0000
Random Forest	0.7778	0.7000	0.7778	0.8000
Support Vector Machine (SVM)	0.4444	0.3000	0.4444	0.4000
K-Nearest Neighbors (KNN)	0.1111	0.0000	0.1111	0.6000
Neural Network	1.0000	0.5000	1.0000	0.0000

Normalized statistics for four key performance metrics for many machine learning algorithms – accuracy, training speed, error rate, and memory usage – are shown in Table 2. The relative efficiency and effectiveness of each method are normalized between 0 and 1 thanks to these values. The neural network stands out with the highest accuracy (1.0000), which indicates its strong predictive performance. It also has the lowest error rate (1.0000), reinforcing its reliability. However, its training speed (0.5000) is moderate, meaning it takes longer to train compared to some other models.

On the other hand, its memory usage is very low (0.0000), making it very efficient in terms of storage requirements. Random Forest strikes a balance between performance and efficiency. It achieves high accuracy (0.7778) and maintains a reasonable training speed (0.7000), allowing for relatively fast training while preserving predictive power. Its error rate (0.7778) is also very low, demonstrating its reliability. Although its memory usage (0.8000) is higher than some models, this trade-off may be acceptable given its overall performance. On

each metric, the support vector machine (SVM) performs reasonably well. Its 0.4444 accuracy and 0.4444 error rate place it in the middle of the predicted confidence range. But compared to many other models, its training speed (0.3000) is slow, which can be a problem when working with large datasets. Its memory usage (0.4000) is moderate, indicating that it requires a fair number of computational resources. Decision Tree is a fast model to train (1.0000), making it suitable for situations where speed is a priority.

However, it has very low accuracy (0.0000) and a high error rate (0.0000), meaning that it prioritizes performance over accuracy. In addition, its memory usage (1.0000) is very high, which can be a drawback in environments with limited resources. Finally, K-Nearest Neighbors (KNN) has relatively low accuracy (0.1111) and slow training speed (0.0000). Its error rate is also low (0.1111), indicating that while it is not the most efficient model, it is still valuable in cases where simplicity and interpretability are more important than computational efficiency.

TABLE 3. Deviation sequence

	Deviation sequence
--	--------------------

	Accuracy	Training Speed	Error Rate	Memory Usage
Decision Tree	1.0000	0.0000	1.0000	0.0000
Random Forest	0.2222	0.3000	0.2222	0.2000
Support Vector Machine (SVM)	0.5556	0.7000	0.5556	0.6000
K-Nearest Neighbors (KNN)	0.8889	1.0000	0.8889	0.4000
Neural Network	0.0000	0.5000	0.0000	1.0000

Five machine learning algorithms—decision tree, random forest, support vector machine (SVM), K-nearest neighbors (KNN), and neural network—are evaluated on four important performance metrics: accuracy, training speed, error rate, and memory usage. Table 3 shows the deviation order values for each approach. The deviation order of each algorithm indicates how much it deviates from a correct reference point; lower values correspond to higher performance on the relevant criteria. The neural network approach has the lowest deviation (0.0000) in accuracy and error rate among the algorithms, indicating that it closely matches the best benchmark in these areas. This demonstrates its exceptional ability to handle complex patterns and high-dimensional data, making it the go-to choice for challenging machine learning applications. However, its high deviation in memory usage (1.0000) indicates that it requires significant computational resources, which can be a drawback. In addition, its moderate deviation in training speed (0.5000) indicates that the training process is relatively slow.

The decision tree exhibits strong performance in accuracy and error rate, with a bias of 1.0000 in both, meaning it meets the best criteria in these areas. Furthermore, it has very low bias (0.0000) in training speed and memory usage, indicating that it

is computationally efficient and fast to train. Despite potential limitations in prediction accuracy, this makes decision trees a practical choice for situations where interpretability and efficiency are prioritized. The Random Forest algorithm provides well-balanced performance, maintaining low biases across all criteria. Its accuracy (0.2222) and error rate (0.2222) biases indicate strong predictive capabilities. However, its training speed (0.3000) and memory usage (0.2000) biases indicate slightly higher computational requirements compared to decision trees. The SVM algorithm shows high deviations in training speed (0.7000) and memory usage (0.6000), highlighting its computational intensity and slow processing time.

However, its moderate deviations in accuracy (0.5556) and error rate (0.5556) indicate that it performs reasonably well in prediction tasks. Finally, KNN shows the highest deviation in training speed (1.0000) and accuracy (0.8889), indicating long training times and low accuracy. While its memory usage deviation (0.4000) is moderate, its high error rate deviation (0.8889) makes it the least favorable of the evaluated algorithms.

TABLE 4. Grey relation coefficient

	Grey relation coefficient			
	Accuracy	Training Speed	Error Rate	Memory Usage
Decision Tree	0.3333	1.0000	0.3333	1.0000
Random Forest	0.6923	0.6250	0.6923	0.7143
Support Vector Machine (SVM)	0.4737	0.4167	0.4737	0.4545
K-Nearest Neighbors (KNN)	0.3600	0.3333	0.3600	0.5556
Neural Network	1.0000	0.5000	1.0000	0.3333

Gray correlation coefficient values for several machine learning algorithms are shown in Table 4 for four important performance metrics: memory usage, error rate, training speed, and accuracy. The relative importance of each component in affecting the overall performance of these algorithms is assessed

with the help of these coefficients. Neural Network achieves a very high accuracy (1.0000), which makes it very reliable for correct predictions. However, its training speed (0.5000) is lower than some other models, which indicates that it requires more computational resources.

With a low error rate (1.0000), it produces fewer incorrect predictions. In addition, its memory usage (0.3333) is relatively low, which means that it does not demand excessive storage space. Random Forest provides well-balanced performance across multiple metrics. It has strong accuracy (0.6923) and moderate training speed (0.6250), which allows it to train faster than some models while maintaining high predictive performance. It is a reasonable option for many applications due to its relatively low error rate (0.6923) and effective memory usage (0.7143). The Support Vector Machine (SVM) provides moderate accuracy (0.4737), but is one of the slower models in terms of training speed (0.4167). Its error rate (0.4737) indicates that it performs better than some models, but is not always the most optimal choice. In addition, its memory usage (0.4545)

indicates that it requires a significant number of computational resources.

The decision tree stands out for its high training speed (1.0000), which makes it suitable for fast decision making. However, it has a relatively low accuracy (0.3333) and a high error rate (0.3333), meaning that it sacrifices accuracy for speed. Furthermore, it has a high memory usage (1.0000), which can be a drawback in resource-constrained environments. Finally, K-Nearest Neighbors (KNN) has low accuracy (0.3600) and slow training speed (0.3333). Its error rate (0.3600) and memory usage (0.5556) indicate that while it is not the most efficient model, it is useful in situations where simplicity and ease of interpretation are essential.

TABLE 5. GRG

	GRG
Decision Tree	0.6667
Random Forest	0.6810
Support Vector Machine (SVM)	0.4546
K-Nearest Neighbors (KNN)	0.4022
Neural Network	0.7083

Table 5 presents the Gray Relationship Grade (GRG) values for five machine learning algorithms: Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Neural Network. These values reflect the overall performance of each algorithm, with higher GRG scores indicating better performance based on the evaluation criteria. Among the models studied, the Neural Network achieves the highest GRG value of 0.7083, making it the best performer.

This is consistent with its well-known strengths in handling complex patterns, adapting to high-dimensional data, and excelling in deep learning applications. Following closely, Random Forest is in second place with a GRG value of 0.6810. As an ensemble learning method that combines multiple decision trees, Random Forest improves accuracy and reduces overfitting, demonstrating strong performance in classification

and regression tasks. The Decision Tree algorithm has a middling ranking with a GRG value of 0.6667. Although easy to explain and implement, its tendency to over fit on complex datasets may contribute to its slightly lower GRG compared to Random Forest.

At the lower end of the ranking, the Support Vector Machine (SVM) has a GRG value of 0.4546, indicating weak performance compared to the top three models. While SVM is effective in high-dimensional spaces, its computational demands and sensitivity to parameter tuning may have affected its score. Finally, K-Nearest Neighbors (KNN) receives the lowest GRG value of 0.4022, making it the least favorable model in this evaluation. Although KNN is a simple and intuitive algorithm, its high computational cost for large datasets and sensitivity to irrelevant features may contribute to its poor ranking.

TABLE 6. Rank

	Rank
Decision Tree	3
Random Forest	2

Support Vector Machine (SVM)	4
K-Nearest Neighbors (KNN)	5
Neural Network	1

The five machine learning algorithms are ranked based on the evaluation process presented in Table 6. The algorithms assessed include Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Neural Network. Each technique is assigned a numerical ranking, where 1 represents the highest (best) performance and 5 signifies the lowest (worst) performance. Neural networks achieve the top ranking of 1, outperforming the other algorithms due to their high accuracy, ability to capture complex patterns, and effectiveness in deep learning applications.

The Random Forest algorithm follows in second place, making it the second-best option. As an ensemble learning method, Random Forest enhances generalization, minimizes overfitting, and improves predictive accuracy by aggregating multiple decision trees. This ranking suggests that Random Forest excels in accuracy, consistency, and reliability. Ranked

third, the Decision Tree algorithm falls in the middle of the ranking, indicating solid yet suboptimal performance. While decision trees are easy to interpret and implement, their moderate placement is influenced by their tendency to over fit training data.

The Support Vector Machine (SVM) is placed fourth, making it less favored than the preceding methods. Despite its strong performance in high-dimensional spaces and suitability for classification tasks, SVM's lower ranking may stem from its computational demands and sensitivity to parameter tuning. Lastly, the K-Nearest Neighbors (KNN) algorithm is ranked the lowest, receiving a rank of 5. While KNN is a simple yet effective technique, its high computational cost for large datasets and sensitivity to irrelevant features likely contribute to its poor evaluation score.

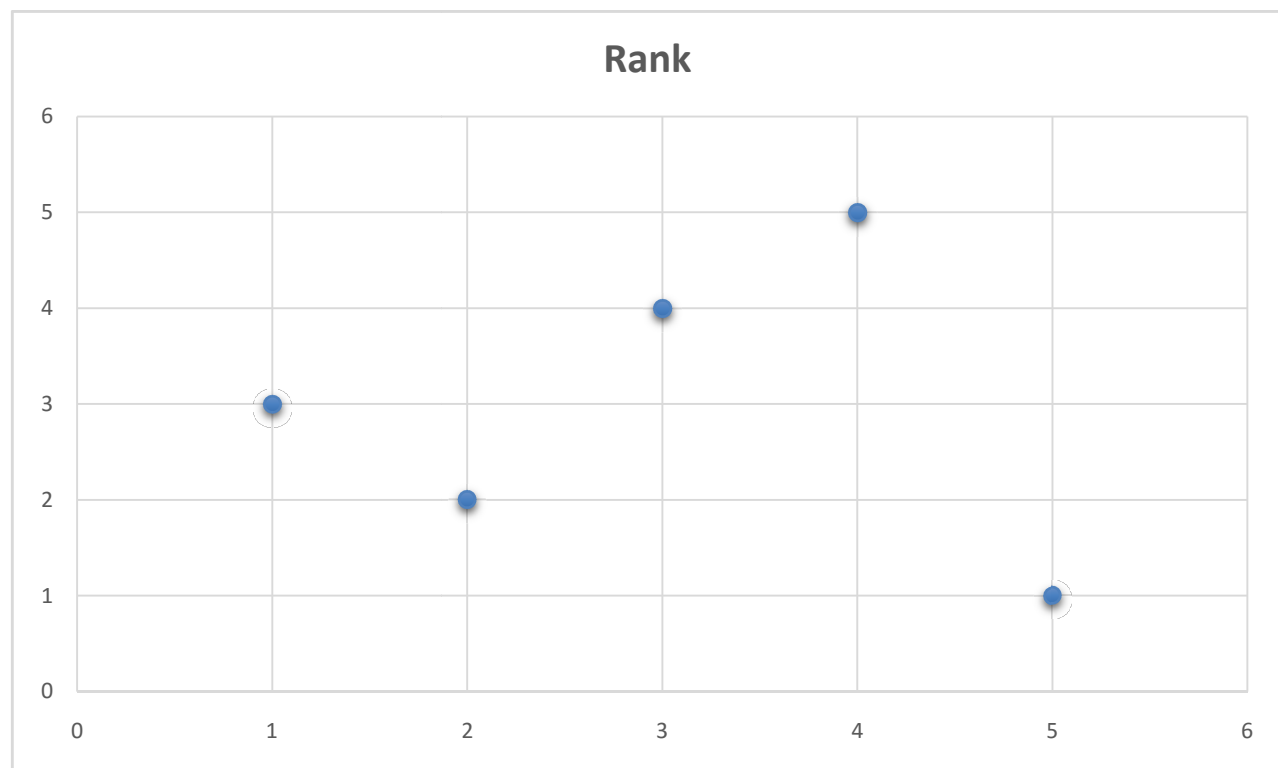


FIGURE 2. Rank

The scatter plot in Figure 2 shows how different options are ranked according to the evaluation procedure. The choices

numbered 1 to 5 are represented by the x-axis, and their corresponding rankings from 1 to 5 are shown by the y-axis.

The ranking given to a particular option is represented by each point on the graph. The graph exhibits a nonlinear ranking distribution, which indicates variations in the performance of the alternatives. For example, the first alternative has a rank of 3, which indicates average performance. This study by Ramancha applies the COPRAS method to evaluate SAP implementations at Union Pacific Railroad, providing a multi-criteria decision-making model to optimize enterprise resource planning (ERP) adoption. The research contributes to improving strategic alignment and operational efficiency in large-scale transportation enterprises.[24]

The second alternative, ranked 2, indicates that it performs better than some but does not hold the first place. The third alternative ranks 4, showing a slightly lower position than the first two. Of all the alternatives, the fourth one has the highest rank of 5, which indicates the least preferred option. On the other hand, the fifth alternative, ranked 1, emerges as the most favorable or optimal choice. These rankings are based on specific evaluation criteria, which may include factors such as

CONCLUSION

Machine learning techniques have revolutionized artificial intelligence, allowing computers to analyze data and make informed decisions without the need for explicit programming. Their growing use in industries such as marketing, finance, healthcare, and automation highlights their importance in modern technology. A wide variety of algorithms are available to handle a variety of data sources and analysis tasks, including supervised learning models, decision trees, support vector machines, neural networks, and unsupervised techniques such as clustering and dimensionality reduction. One of the primary benefits of machine learning is its ability to efficiently process large amounts of data and discover hidden patterns that are difficult for human analysts to detect. This ability has played a key role in advancing fields such as autonomous systems, image recognition, natural language processing, and predictive analytics. Supervised learning methods, such as regression and classification, improve prediction accuracy and decision-making, while reinforcement learning helps systems continuously refine their performance based on experience.

Unsupervised techniques such as clustering help uncover underlying structures in data without the need for labeled information, which is proving particularly useful in customer segmentation and anomaly detection. Despite its progress, machine learning still faces several hurdles that need to be addressed. The quality of input data is a key factor determining

efficiency, cost, performance metrics, or other qualitative and quantitative attributes. This scatter diagram effectively visualizes the ranking system, providing an intuitive way to assess the relative standing of each alternative. By analyzing the point distribution, one can easily determine which options perform better and which are less desirable. The absence of a clear linear trend indicates variations in performance among the evaluated alternatives. Such ranking methods are commonly used in multi-attribute decision making (MADM), where multiple factors influence the decision-making process.

Techniques such as TOPSIS, AHP, or Gray Relational Analysis (GRA) are often used in such contexts to determine the most suitable alternative. Figure 2 serves as a clear representation of the rankings of five alternatives based on the evaluation process. The visualization helps to compare different alternatives, allowing for better informed decision-making. The varying ranking distribution highlights differences in performance, helping decision-makers focus on improving lower-ranked options or selecting the most favorable alternative.

model performance, and challenges such as biased datasets, over fitting, and privacy concerns can affect accuracy and fairness. Furthermore, machine learning models often require a large amount of processing power, and their complexity can make interpretation challenging. Ethical considerations, including transparency, accountability, and fairness in AI decision-making, are essential for building trust and ensuring responsible usage. The capabilities of AI have been significantly enhanced by advancements in deep learning, a specialized area of machine learning. Deep learning architectures such as convolution neural networks (CNNs) and recurrent neural networks (RNNs) have demonstrated exceptional performance in applications like natural language processing, autonomous vehicle navigation, and image and audio recognition.

However, these sophisticated models present challenges, as they require large amounts of labeled data and substantial computational resources. Ongoing research continues to focus on improving algorithm efficiency, interpretability, and scalability, paving the way for further advancements in machine learning. Emerging technologies like federated learning and quantum machine learning are shaping the future of AI by enhancing computational power, preserving data privacy, and enabling decentralized model training. As the field progresses, collaboration among data scientists, engineers, and domain experts will be vital in fostering responsible AI development and maximizing the impact of machine learning across industries.

REFERENCE

1. Mahesh, Batta. "Machine learning algorithms-a review." *International Journal of Science and Research (IJSR)*. [Internet] 9, no. 1 (2020): 381-386.
2. Sarker, Iqbal H. "Machine learning: Algorithms, real-world applications and research directions." *SN computer science* 2, no. 3 (2021): 160.
3. Alzubi, Jafar, Anand Nayyar, and Akshi Kumar. "Machine learning from theory to algorithms: an overview." In *Journal of physics: conference series*, vol. 1142, p. 012012. IOP Publishing, 2018.
4. Adari, V. K. (2024). *Interoperability and Data Modernization: Building a Connected Banking Ecosystem*. *International Journal of Computer Engineering and Technology*, 15(6), 653–662.
5. Balaji, Thoguru K., Chandra Sekhara Rao Annavarapu, and AnnushreeBablani. "Machine learning algorithms for social media analysis: A survey." *Computer Science Review* 40 (2021): 100395.
6. Crisci, Carolina, Badih Ghattas, and Ghattas Perera. "A review of supervised machine learning algorithms and their applications to ecological data." *Ecological Modelling* 240 (2012): 113-122.
7. Nagababu. K, "Machine Learning Techniques in Fracture Mechanics a Comparative Study of Linear Regression, Random Forest, and Ada Boost Model" *Journal of Artificial intelligence and Machine Learning*., 2024, vol. 2, no. 2, pp. 1–13. doi: <http://dx.doi.org/10.55124/jaim.v2i2.257>
8. Liu, Yanli, Yourong Wang, and Jian Zhang. "New machine learning algorithm: Random forest." In *International conference on information computing and applications*, pp. 246-252. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.
9. Huang, Jui-Chan, Kuo-Min Ko, Ming-Hung Shu, and Bi-Min Hsu. "Application and comparison of several machine learning algorithms and their integration models in regression problems." *Neural Computing and Applications* 32, no. 10 (2020): 5461-5469.
10. Latif, Jahanzaib, Chuangbai Xiao, Azhar Imran, and Shanshan Tu. "Medical imaging using machine learning and deep learning algorithms: a review." In *2019 2nd International conference on computing, mathematics and engineering technologies (iCoMET)*, pp. 1-5. IEEE, 2019.
11. Vagvala, Prabhakar. "Comparative Analysis of ERP Systems Separation and Migration Strategies: Evaluating Automation Approaches Using the EDAS Methodology." *Journal of Computer Science Applications and Information Technology*, vol. 8, no. 1, 2023, pp. 1–10.
12. Balyen, Lokman, and Tunde Peto. "Promising artificial intelligence-machine learning-deep learning algorithms in ophthalmology." *The Asia-Pacific Journal of Ophthalmology* 8, no. 3 (2019): 264-272.
13. Kumar, D. Praveen, Tarachand Amgoth, and Chandra Sekhara Rao Annavarapu. "Machine learning algorithms for wireless sensor networks: A survey." *Information Fusion* 49 (2019): 1-25.
14. Sridhar Kakulavaram. (2022). *Life Insurance Customer Prediction and Sustainability Analysis Using Machine Learning Techniques*. *International Journal of Intelligent Systems and Applications in Engineering*, 10(3s), 390 –. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/7649>
15. Ngiam, Kee Yuan, and Wei Khor. "Big data and machine learning algorithms for health-care delivery." *The Lancet Oncology* 20, no. 5 (2019): e262-e273.
16. Khanum, Memoona, Tahira Mahboob, Warda Imtiaz, Humaraia Abdul Ghafoor, and Rabeea Sehar. "A survey on unsupervised machine learning algorithms for automation, classification and maintenance." *International Journal of Computer Applications* 119, no. 13 (2015).
17. Yang, Li, and Abdallah Shami. "On hyperparameter optimization of machine learning algorithms: Theory and practice." *Neurocomputing* 415 (2020): 295-316.
18. Mishra, Mahesh Kumar. "Enhancing Internet of Vehicles (IoV) System Selection through the Complex Proportionality Assessment (COPRAS) Method: A Case Study." *Journal of Business Intelligence and Data Analytics*, vol. 1, no. 2, 2024, pp. 17-29.
19. Vabalas, Andrius, Emma Gowen, Ellen Poliakoff, and Alexander J. Casson. "Machine learning algorithm validation with a limited sample size." *PloS one* 14, no. 11 (2019): e0224365.
20. Ballamudi, S. "Interleaved Feature Extraction Model Bridging Multiple Techniques for Enhanced Object Identification" *Journal of Artificial Intelligence and Machine Learning*., 2023, vol. 1, no. 2, pp. 1-7. doi: <https://doi.org/10.55124/jaim.v1i2.253>
21. Mittapally R (2024). *Intelligent Framework Selection: Leveraging MCDM in Web Technology Decisions*. *J Comp Sci Appl Inform Technol*. 9(1): 1-9.
22. McClendon, Lawrence, and Natarajan Meghanathan. "Using machine learning algorithms to analyze crime data." *Machine Learning and Applications: An International Journal (MLAIJ)* 2, no. 1 (2015): 1-12.
23. Quan, Huafeng, Shaobo Li, Hongjing Wei, and Jianjun Hu. "Personalized product evaluation based on GRA-TOPSIS and Kansei engineering." *Symmetry* 11, no. 7 (2019): 867.
24. Hatefi, Seyed Morteza, and Jolanta Tamošaitienė. "Construction projects assessment based on the sustainable development criteria by an integrated fuzzy AHP and improved GRA model." *Sustainability* 10, no. 4 (2018): 991.
25. Olabanji, Olayinka Mohammed, and Khumbulani Mpofu. "Appraisal of conceptual designs: Coalescing fuzzy analytic hierarchy process (F-AHP) and fuzzy grey relational

- analysis (F-GRA)." *Results in Engineering* 9 (2021): 100194.
26. Wang, Ting-Kwei, Qian Zhang, Heap-Yih Chong, and Xiangyu Wang. "Integrated supplier selection framework in a resilient construction supply chain: An approach via analytic hierarchy process (AHP) and grey relational analysis (GRA)." *Sustainability* 9, no. 2 (2017): 289.
27. Taguchijeve, U. G. A. G. I., and VRTILNO-TORNIM PROCESOM FSW. "Application of grey relation analysis (GRA) and Taguchi method for the parametric optimization of friction stir welding (FSW) process." *Mater Tehnol* 44 (2010): 205.
28. Ramancha, Nitesh Kumar. "Implementing and Evaluating SAP Solutions at the Union Pacific Railroad: Multi-Criteria Decision Analysis Using the COPRAS Method." *Journal of Business Intelligence and Data Analytics*, vol. 1, no. 2, 2025. DOI: <https://10.55124/jbid.v1i2.242>.
29. Javed, Saad Ahmed, and Sifeng Liu. "Bidirectional absolute GRA/GIA model for uncertain systems: application in project management." *IEEE Access* 7 (2019): 60885-60896.
30. Song, Xin, Wenmin Liu, Minglei Zhang, and Feng Liu. "A network selection algorithm based on FAHP/GRA in heterogeneous wireless networks." In *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*, pp. 1445-1449. IEEE, 2016.
31. Mohammed Zabeeulla, A. N., and Chandrasekar Shastry. "Automation of Leaf Disease Prediction Framework based on Machine Learning and Deep Learning in different Crop Species." *JOURNAL OF ALGEBRAIC STATISTICS* 13, no. 3 (2022): 3098-3113.
32. Watson, David S., Jenny Krutzinna, Ian N. Bruce, Christopher EM Griffiths, Iain B. McInnes, Michael R. Barnes, and Luciano Floridi. "Clinical applications of machine learning algorithms: beyond the black box." *Bmj* 364 (2019).
33. Portugal, Ivens, Paulo Alencar, and Donald Cowan. "The use of machine learning algorithms in recommender systems: A systematic review." *Expert Systems with Applications* 97 (2018): 205-227.