International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal

Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

A Comprehensive Review of AI-Enhanced Unsupervised Classification Frameworks Utilizing CART Algorithms

Shazia Sultan

Research Scholar, Department of Computer Science, Malwanchal University, Indore

Dr. Sharad Patil

Supervisor, Department of Computer Science, Malwanchal University, Indore

Abstract

This review examines the evolution and advancements of AI-enhanced unsupervised classification frameworks built upon Classification and Regression Trees (CART), highlighting their growing relevance in handling complex, high-dimensional, and unlabeled datasets. While traditional CART has been predominantly applied in supervised contexts, recent innovations integrate deep learning–based feature extraction, reinforcement learning for adaptive split selection, and hybrid clustering-tree architectures to enhance performance, scalability, and interpretability. The review synthesizes contemporary research from 2018 onward, offering a comprehensive analysis of theoretical foundations, algorithmic extensions, and major application areas including healthcare analytics, cybersecurity, remote sensing, and market segmentation. Additionally, it identifies key limitations such as computational overhead, sensitivity to data sparsity, and challenges in optimizing tree structures for unsupervised tasks. The study concludes by outlining future research opportunities aimed at improving transparency, robustness, and integration with semi-supervised and explainable AI frameworks.

Keywords: AI-enhanced CART; Unsupervised learning; Clustering trees; Hybrid algorithms; Explainable AI

Introduction

The rapid advancement of artificial intelligence (AI) has significantly transformed the landscape of unsupervised learning, particularly in the domain of decision-tree–based models such as Classification and Regression Trees (CART). Traditionally, CART has been widely recognized for its interpretability, hierarchical structure, and robustness in supervised tasks; however, recent developments have extended its applicability to unlabeled datasets through innovative unsupervised adaptations. These advancements leverage AI-driven optimizations—including deep feature extraction, reinforcement learning strategies for dynamic tree formation, and hybrid clustering-tree frameworks—that collectively enhance CART’s capability to discover hidden patterns in high-dimensional and complex data. As industries increasingly rely on large-scale, unstructured, and mixed-attribute datasets, the demand for unsupervised classification methods

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

that provide both accuracy and interpretability has intensified, positioning AI-enhanced CART models as compelling alternatives to opaque deep unsupervised systems. Despite the proliferation of clustering algorithms such as K-Means, DBSCAN, hierarchical models, and density-estimation techniques, they often suffer from limitations like rigid assumptions, sensitivity to noise, or inability to offer transparent decision boundaries. In contrast, AI-driven unsupervised CART frameworks integrate symbolic reasoning with machine-intelligent feature learning, enabling more adaptive, explainable, and computationally efficient classification. Moreover, research from 2018 onward indicates a growing trend toward hybrid architectures where CART acts as a structural backbone while AI components optimize split selection, reduce overfitting, and enhance generalization across diverse application domains such as healthcare analytics, cybersecurity threat detection, remote sensing, financial segmentation, and industrial IoT data analysis. Yet, the literature remains fragmented, with limited comprehensive assessments of algorithmic modifications, comparative performance, and emerging challenges. This review therefore aims to synthesize existing research, analyze methodological advances, and highlight the evolving synergy between AI mechanisms and CART-based unsupervised learning. By presenting an integrated perspective on the theoretical foundations, algorithmic innovations, real-world applications, and existing gaps, the review provides a critical foundation for researchers seeking to advance interpretable and intelligent unsupervised classification models.

Significance of the Study

The significance of this study lies in its comprehensive exploration of how AI-enhanced methodologies are transforming the capabilities of Classification and Regression Trees (CART) within unsupervised learning environments, where the absence of labeled data traditionally constrains model performance, interpretability, and adaptability. As industries increasingly deal with massive, heterogeneous datasets generated from IoT systems, social media, biomedical sensors, cybersecurity logs, and remote sensing platforms, there is a growing need for clustering and pattern-discovery methods that not only provide accurate insights but also remain transparent and computationally efficient. This study is particularly important because it bridges a critical gap in the literature by unifying dispersed advancements—such as deep feature extraction, hybrid clustering mechanisms, reinforcement-learning-driven tree optimization, and evolutionary algorithms—into a coherent framework that demonstrates how CART can be successfully reengineered for unsupervised applications. Through synthesizing empirical findings and methodological innovations from recent research, the study enhances understanding of how AI-driven enhancements address longstanding limitations of traditional unsupervised techniques, including sensitivity to noise, inability to capture nonlinear structures, and lack of interpretability. Furthermore, by examining applications across areas like healthcare

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

classification, anomaly detection, customer segmentation, and satellite image analysis, the study highlights CART's evolving role as a scalable, interpretable alternative to opaque deep learning models. This is particularly valuable for fields requiring explainable decision flows, regulatory compliance, and trust in automated systems. Additionally, the insights derived from this review can guide the development of next-generation unsupervised CART models, promote standardization of evaluation metrics, and inspire further innovation in hybrid AI-tree architectures. Overall, the study provides both academic and practical significance by advancing theoretical knowledge, improving real-world implementation potential, and laying a strong foundation for future research in explainable and intelligent unsupervised classification.

Background of Unsupervised Learning

Unsupervised learning represents a foundational branch of machine learning focused on discovering hidden structures, intrinsic patterns, and meaningful relationships within unlabeled data, making it indispensable in domains where manual annotation is impractical, costly, or impossible. Unlike supervised learning, which relies on predefined input–output mappings, unsupervised techniques operate autonomously, clustering similar data points, reducing dimensionality, identifying anomalies, or uncovering latent variable structures without explicit guidance. Historically, methods such as K-Means clustering, hierarchical clustering, principal component analysis (PCA), and self-organizing maps (SOM) have dominated the field, offering valuable insights but often constrained by simplistic assumptions, sensitivity to initialization, difficulty in handling nonlinear or high-dimensional relationships, and limited interpretability. The exponential growth of big data—characterized by volume, velocity, and variety—has further highlighted these limitations, prompting a shift toward more intelligent and adaptive unsupervised frameworks that integrate machine learning, statistical modeling, and computational intelligence. The emergence of deep unsupervised models, including autoencoders, generative adversarial networks (GANs), and deep clustering architectures, has significantly expanded the capability to represent complex data distributions; however, these approaches frequently sacrifice transparency and interpretability, which are critical for high-stakes applications such as healthcare diagnostics, financial risk assessment, industrial automation, and cybersecurity. In this context, decision-tree–based methods, particularly Classification and Regression Trees (CART), have gained renewed interest due to their structural clarity, rule-based decision flows, and potential for hybridization with advanced AI mechanisms. Extending CART into unsupervised learning allows researchers to merge the interpretability of tree structures with the adaptability of modern AI, resulting in models capable of performing clustering, density estimation, and feature grouping while maintaining explainability. As unsupervised learning continues to evolve, its role in pattern recognition, anomaly detection, recommendation systems, natural language processing, and image segmentation underscores its

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

growing importance across scientific and industrial domains. The increasing integration of AI-driven enhancements—such as reinforcement learning, evolutionary optimization, and deep feature extraction—signals a new era where unsupervised methods become more robust, scalable, and capable of handling diverse, complex datasets. This expanding landscape establishes the essential background for understanding how AI-enhanced unsupervised CART frameworks are reshaping data-driven discovery.

Evolution of Classification and Regression Trees (CART)

The evolution of Classification and Regression Trees (CART) represents a pivotal trajectory in the history of machine learning, beginning with its formal introduction by Breiman, Friedman, Olshen, and Stone in 1984 as a unified framework for building interpretable decision-tree models capable of handling both categorical and continuous outcomes. Unlike earlier rule-based or statistical classification methods, CART introduced a binary recursive partitioning approach that systematically split data into homogeneous subsets based on impurity measures such as Gini index, entropy, or variance reduction. Its ability to model nonlinear relationships, manage missing values, accommodate mixed data types, and produce transparent, rule-based decision pathways made CART a revolutionary tool across statistics, data mining, and pattern recognition. In the subsequent decades, enhancements emerged to address overfitting and stability, leading to the development of pruning techniques, surrogate splits, and ensemble variants such as Random Forests and Gradient Boosted Trees, which significantly improved predictive accuracy but reduced interpretability. As computational power increased and datasets grew more complex, CART began evolving beyond its traditional supervised framework, inspiring the creation of unsupervised extensions such as clustering trees, model-based trees, and density-estimation trees, allowing it to operate on unlabeled data. With the integration of artificial intelligence and machine learning innovations post-2010, CART experienced a new wave of development: deep learning–driven feature extraction enabled CART to handle high-dimensional representations; evolutionary algorithms optimized tree structure and split selection; and reinforcement learning provided adaptive policies for dynamic tree growth. Contemporary research has also focused on hybrid architectures, where CART is fused with clustering algorithms (e.g., K-Means, DBSCAN), probabilistic models, or neural embeddings to enhance performance in tasks such as anomaly detection, pattern discovery, and segmentation. Moreover, the rise of explainable AI (XAI) has renewed interest in CART, positioning it as a powerful model that balances interpretability with advanced analytical capabilities. Today, CART continues to evolve as a foundational yet adaptable method, bridging classical decision theory with modern AI technologies, and enabling interpretable solutions to increasingly complex supervised and unsupervised classification challenges.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

Need for AI-Enhanced Unsupervised Classification

The need for AI-enhanced unsupervised classification has become increasingly critical as modern data ecosystems produce massive volumes of complex, heterogeneous, and unlabeled information across domains such as healthcare, cybersecurity, finance, remote sensing, and industrial IoT. Traditional unsupervised learning methods—including K-Means, hierarchical clustering, PCA, DBSCAN, and SOM—provide foundational mechanisms for pattern discovery but often struggle with challenges such as sensitivity to noise, rigid assumptions about cluster shapes, poor scalability, and difficulty handling nonlinear relationships or high-dimensional feature spaces. Moreover, these algorithms generally lack interpretability, making them unsuitable for applications requiring transparent decision-making or regulatory compliance. In this context, integrating artificial intelligence enhancements into unsupervised models, particularly Classification and Regression Trees (CART), offers a transformative solution by combining the interpretability of tree-based structures with the adaptive learning capabilities of modern AI. AI-driven techniques such as deep feature extraction allow CART to operate effectively on complex datasets by generating rich latent representations, while reinforcement learning introduces dynamic decision-making that optimizes tree growth and splitting policies. Similarly, evolutionary algorithms can refine tree architecture, improving accuracy and reducing computational complexity. Hybrid clustering-tree frameworks merge the strengths of classical clustering algorithms with CART's rule-based interpretability, enabling more stable and context-aware classification outcomes. As data diversity increases, the ability to automatically uncover meaningful substructures without human-labeled input becomes essential for tasks like anomaly detection, segmentation, fraud detection, medical image grouping, and customer behavior analysis. Furthermore, organizations increasingly demand models that not only perform well but also explain their logic, a requirement amplified in industries governed by ethical guidelines and transparency standards. AI-enhanced unsupervised classification directly addresses these needs by producing more accurate, robust, scalable, and interpretable outcomes compared to conventional clustering methods. Additionally, the integration of AI mechanisms enables continuous adaptation as data evolves, supporting real-time decision-making in dynamic environments. Overall, the escalating complexity of data analytics, combined with the limitations of traditional unsupervised methods, underscores the urgent necessity for AI-enhanced frameworks that elevate both the performance and the interpretability of unsupervised classification systems.

Literature Review

The evolution of AI-enhanced unsupervised learning and decision-tree methodologies has been shaped by a diverse body of research that spans algorithmic development, hybrid intelligence models, and application-driven innovation. Foundational work in tree-based learning continues to influence emerging CART variants, particularly those designed for unlabeled or semi-labeled data domains. Significantly contribute to this trajectory through their development of the semi-CART algorithm, designed explicitly to operate in semi-supervised environments where partial labeling exists. Their work demonstrates how iterative optimization strategies and hybrid impurity measures can overcome the limitations of traditional CART, which has historically relied on fully labeled datasets. This innovation sets the groundwork for extending decision trees into unsupervised spaces by showing that tree structures can incorporate heuristic knowledge, statistical inference, and reinforcing signals during the learning process. Such studies highlight the adaptability of decision trees and reinforce the importance of structural interpretability in modern machine learning systems.

Parallel advancements in AI-driven anomaly detection and cybersecurity applications also shed light on the practical relevance of enhanced CART-based unsupervised systems. Ahmad, Han, and Mahmood (2023) emphasize the importance of integrating AI modules with risk assessment processes for connected and autonomous vehicles (CAVs). Their research identifies how machine learning systems can detect latent safety risks by analyzing vast quantities of unlabeled sensor and environmental data. Although their study is not exclusively CART-focused, the principles underlying their "AI-enhanced threat analysis" highlight the pressing need for interpretable models in safety-critical domains. Similar work by Alghanmi et al. (2019) strengthens this argument through their hybrid learning anomaly detection model (HLMCC), which combines clustering with machine learning classifiers to detect abnormalities in IoT networks. Their use of hybrid mechanisms closely mirrors the logic of unsupervised CART variants, showcasing the efficacy of integrating clustering algorithms with rule-based learning for clearer, more reliable anomaly detection. These studies collectively demonstrate the increasing demand for explainable hybrid models capable of adapting to dynamic, large-scale environments.

Within the healthcare domain, decision tree methods and hybrid learning systems have shown substantial potential. Ahmed and Husien (2023) offer a review of hybrid machine learning models for heart disease prediction, illustrating how different algorithmic combinations—such as decision trees with ensemble or optimization techniques—enhance predictive accuracy and stability. Their analysis underscores the importance of combining interpretable models with advanced AI optimizers to improve clinical decision support systems. Although focused on supervised prediction, their insights are broadly applicable to unsupervised CART systems, especially in how hybrid approaches address nonlinearity, noise, and high-dimensional medical data challenges. Meanwhile, Almuqati et al. (2021) provide a broad overview of supervised and unsupervised learning challenges, noting that unsupervised models often lack interpretability and require significant enhancements to manage large, unstructured datasets. Their work strengthens the argument for CART-based unsupervised frameworks, which inherently offer transparent decision-making while being adaptable to hybrid AI improvements.

A growing segment of the literature focuses on hybrid optimization algorithms and their synergy with learning models. Azevedo et al. (2023) present a systematic review exploring hybrid optimization and machine learning approaches, emphasizing how evolutionary algorithms, swarm intelligence, and metaheuristics enhance the performance, convergence, and robustness of classification systems. Their findings directly align with the recent movement toward evolutionary-enhanced CART models, where genetic algorithms optimize node splits, tree depth, and structural efficiency. This interaction between optimization and learning demonstrates how unsupervised decision trees can become more accurate and scalable when fused with intelligent search mechanisms. Earlier foundational work by Bala et al. (1995) also supports this trajectory, illustrating how genetic algorithms can refine decision tree learning, laying early groundwork for today's hybrid CART systems. Their research demonstrates that combining evolutionary processes with symbolic structures like decision trees yields models that balance interpretability with computational power—an essential requirement for modern unsupervised AI systems.

Finally, broader analytical perspectives, such as those provided by Bensi and Esquivel (2023), highlight the enduring significance of classification trees across both traditional and modern AI contexts. Their literature review synthesizes decades of research on classification tree performance, interpretability, and adaptability. They emphasize that decision trees remain central to machine learning because they offer clarity, transparency, and structural flexibility—qualities that are even more critical when transitioning into unsupervised or semi-supervised frameworks. Their insights reinforce the growing acceptance that CART-based architectures, enhanced with AI mechanisms, provide an effective middle ground between interpretability and advanced analytical capacity. Collectively, the reviewed literature reveals a consistent progression toward hybrid, interpretable, and scalable decision-tree models, illustrating how AI-enhanced CART systems are becoming essential tools for handling the complexity, volume, and unlabeled nature of modern data environments.

Theoretical Foundations of CART Algorithms

- **CART Structure and Mathematical Basis**

The theoretical foundation of CART is built upon binary recursive partitioning, a process that divides the dataset into progressively purer subsets by generating binary splits at each decision node. Mathematically, CART seeks to minimize impurity within nodes by selecting optimal features and thresholds that maximize homogeneity in the resulting partitions. Each internal node represents a decision rule, while terminal nodes represent final class labels or predicted values. This hierarchical structure enables CART to model nonlinear relationships, capture complex interactions, and handle both categorical and numerical features without requiring prior assumptions about data distribution.

- **Splitting Criteria**

CART determines the best splits through impurity minimization using statistical criteria. For classification tasks, Gini index evaluates the likelihood of incorrect classification, while entropy measures information disorder and aligns with information gain. Both aim to create partitions with maximum class purity. For regression tasks, variance reduction assesses how well a split decreases variability within subsets, enabling accurate prediction of continuous outcomes. These criteria ensure that CART generates highly discriminative and informative decision boundaries.

- **Pruning Strategies and Optimization**

To prevent overfitting—an inherent risk due to CART's flexibility—pruning strategies are applied to simplify tree structure without sacrificing predictive performance. Cost-complexity pruning evaluates the trade-off between tree size and classification error, removing branches that contribute minimal improvement. Optimization techniques such as cross-validation, surrogate splits for missing data, and regularization further enhance model robustness. Pruning ensures that the tree maintains generalizability, interpretability, and computational efficiency.

- **CART vs. Other Decision Tree Models**

CART distinguishes itself from other tree models such as C4.5, ID3, CHAID, and GUIDE through its strict binary tree structure and reliance on impurity-based splitting rather than multi-way partitions or chi-square statistics. Unlike C4.5, which uses information gain ratio, CART applies Gini or variance metrics, enabling more flexible handling of numerical data. CART's ability to perform both classification and regression, coupled with superior interpretability and robustness, positions it as a foundational model that continues to influence modern ensemble methods and AI-augmented decision frameworks.

AI Enhancements for Unsupervised Classification

- **Deep Learning–Driven Feature Representation**

Deep learning–driven feature representation plays a transformative role in enhancing unsupervised CART models by extracting rich, high-level abstractions from complex data that traditional manual or statistical feature engineering cannot capture. Autoencoders, convolutional neural networks (CNNs), and deep embedding models generate latent features that preserve intrinsic structures while reducing dimensionality, enabling CART to operate on more meaningful representations. This integration improves clustering stability, reduces noise sensitivity, and enhances the interpretability of final trees, making the hybrid deep learning–CART framework suitable for image grouping, anomaly detection, and text clustering.

- **Hybrid Clustering Methods (K-Means, DBSCAN + CART)**

Hybrid clustering methods further elevate unsupervised classification by combining CART with algorithms such as K-Means and DBSCAN to leverage their complementary strengths. In K-Means + CART hybrids, K-Means first partitions data into initial clusters, providing structure that CART refines using rule-based splits, thereby improving boundary detection and interpretability. DBSCAN + CART hybrids are particularly effective for irregular cluster shapes and noise-filled datasets, where DBSCAN identifies dense regions and outliers, and CART derives decision rules that clearly describe cluster membership. These hybrid approaches significantly overcome limitations of standalone clustering methods and provide scalable, interpretable, and accurate classification for heterogeneous datasets.

- **Reinforcement Learning for Tree Optimization**

Reinforcement learning (RL) introduces an adaptive and intelligent optimization layer to unsupervised CART by treating tree construction as a sequential decision-making process. RL agents learn optimal splitting policies by receiving rewards based on impurity reduction, cluster compactness, or density consistency. Unlike static greedy algorithms, RL-driven CART models iteratively refine decision boundaries, explore better feature–threshold combinations, and dynamically adjust tree depth, thereby producing more robust and generalizable structures. This enables CART to handle evolving data streams and non-stationary distributions common in real-time monitoring systems.

- **Evolutionary Algorithms for Node Splitting**

Evolutionary algorithms (EAs) such as genetic algorithms, genetic programming, and differential evolution provide powerful optimization strategies for node splitting in unsupervised CART. EAs evaluate multiple candidate trees simultaneously, evolving them over generations to optimize cluster purity, cohesion, and separation. Mutation, crossover, and selection operations help explore diverse tree structures, preventing the algorithm from getting trapped in local minima. This results in more globally optimal decision trees that outperform conventional greedy splitting methods. EA-enhanced CART is especially valuable for complex, high-dimensional datasets where exhaustive search is computationally infeasible. Overall, these AI-driven enhancements collectively improve the interpretability, adaptability, and accuracy of unsupervised CART models, enabling their successful deployment in diverse and challenging data environments.

Unsupervised Variants of CART

- **Clustering Trees (C-Trees)**

Clustering Trees (C-Trees) represent a foundational unsupervised extension of CART that focuses on partitioning unlabeled datasets into meaningful clusters through recursive binary splitting driven by similarity measures rather than class labels. Instead of minimizing classification error, C-Trees optimize cluster purity, cohesion, or variance reduction, making

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

them effective for exploratory data analysis and pattern discovery. These trees provide interpretable, rule-based cluster definitions and offer significant advantages over classical clustering methods by producing transparent decision boundaries that are easy to understand and replicate across applications such as market segmentation and medical subgroup identification.

- **Density-Based CART Models**

Density-based CART models extend traditional tree structures by integrating density estimation techniques to identify regions of varying data concentration. These models use density thresholds, local distribution patterns, and probabilistic criteria to form splits, enabling the detection of irregular cluster shapes, anomalies, and sparse regions that conventional clustering algorithms fail to capture. Density-enhanced CART variants excel in tasks involving noise-heavy datasets, fraud detection scenarios, or spatial analysis problems. By aligning tree construction with data density, these models create partitions that naturally adapt to the intrinsic geometry of the dataset.

- **Self-Organizing CART Structures**

Self-organizing CART structures draw inspiration from neural self-organizing maps (SOMs) by incorporating adaptive learning mechanisms that reorganize and refine tree formation over time. These structures dynamically adjust splitting rules based on evolving data patterns, enabling the creation of more coherent clusters through iterative refinement. Self-organizing CART is highly effective for streaming data, temporal datasets, and non-stationary environments, as it continuously self-corrects and improves cluster boundaries. The adaptability of these models makes them especially useful in industries like cybersecurity, sensor networks, and predictive maintenance.

- **Mixed-Attribute CART for Unlabeled Data**

Mixed-attribute CART variants address the challenge of clustering data that contain a combination of numerical, categorical, ordinal, and binary features. These models employ specialized impurity metrics and similarity functions tailored to heterogeneous data types, allowing more accurate and context-sensitive partitioning. By integrating algorithms such as Gower distance, heterogeneous entropy measures, or hybrid splitting rules, mixed-attribute CART enhances flexibility and applicability across diverse datasets commonly found in healthcare records, customer profiles, and socio-economic surveys. Overall, these unsupervised CART variants collectively expand the scope of tree-based learning, enabling structured, interpretable, and effective classification across increasingly complex and unlabeled data environments.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

Applications of AI-Enhanced Unsupervised CART

- **Healthcare Data Clustering**

AI-enhanced unsupervised CART models play a transformative role in healthcare by enabling refined patient clustering, disease subtype discovery, and anomaly detection in clinical datasets. Through deep feature extraction and hybrid clustering-tree integration, these models effectively handle high-dimensional medical records, genomic sequences, and imaging datasets while maintaining interpretability—an essential requirement for clinical decision-making. They support early diagnosis, personalized treatment planning, and stratification of patient risk groups, offering transparent rules that clinicians can easily validate.

- **Remote Sensing & Satellite Image Classification**

In remote sensing, AI-integrated unsupervised CART algorithms facilitate the classification of complex satellite imagery by combining deep embeddings with rule-based partitioning. These models detect land cover patterns, vegetation types, water bodies, and urban expansion with high precision, even under noise or varying illumination conditions. Their hierarchical structure allows for interpretable region-specific segmentation, supporting environmental monitoring, disaster assessment, and resource management.

- **Cybersecurity Threat Segmentation**

Cybersecurity greatly benefits from unsupervised CART frameworks that identify hidden threat patterns within massive network traffic logs and user-behavior data. AI enhancements enable the detection of zero-day attacks, anomalous access patterns, and malware signatures without requiring labeled datasets. Reinforcement learning–optimized CART structures adapt to evolving cyber threats, offering scalable, explainable decision paths critical for real-time defense and compliance.

- **Market Segmentation & Customer Behavior Mapping**

In marketing and consumer analytics, AI-enhanced CART supports unsupervised segmentation by revealing latent customer groups based on purchase behavior, demographic patterns, online interactions, and psychographic traits. Hybrid clustering-tree models generate interpretable rules for each segment, empowering businesses to personalize recommendations, optimize campaigns, and predict customer churn. Their clarity and adaptability make them valuable tools for strategic decision-making.

- **Industrial IoT and Sensor Data Categorization**

Industrial IoT systems generate continuous streams of sensor data that require efficient, real-time classification. AI-enhanced unsupervised CART models categorize equipment states, detect anomalies, and identify operational patterns, contributing to predictive maintenance, quality control, and system optimization. Their self-organizing and density-based capabilities allow them to adapt to fluctuating environmental conditions and heterogeneous industrial signals.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

Collectively, these applications showcase the versatility, interpretability, and powerful analytical capabilities of AI-enhanced unsupervised CART across diverse, data-intensive environments.

Conclusion

This comprehensive review highlights the growing importance and transformative potential of AI-enhanced unsupervised CART frameworks in addressing the challenges posed by modern, high-dimensional, and unlabeled datasets across diverse domains. The synthesis of research demonstrates that traditional CART, while historically grounded in supervised learning, has evolved significantly through the integration of deep learning–based feature extraction, hybrid clustering-tree mechanisms, reinforcement learning strategies, and evolutionary optimization techniques. These enhancements collectively advance the interpretability, scalability, and robustness of unsupervised classification models, enabling them to outperform conventional clustering algorithms in domains requiring transparent and reliable decision-making. Applications across healthcare, cybersecurity, remote sensing, market analytics, and industrial IoT illustrate the versatility and real-world value of these advanced CART variants, especially in scenarios where explainability and adaptability are critical. Despite these advancements, the literature also reveals ongoing limitations, including computational complexity, optimization challenges, and sensitivity to noisy or sparse data, which necessitate continued innovation. Future research directions should focus on integrating semi-supervised elements, improving density-aware and mixed-attribute handling, and leveraging explainable AI frameworks to enhance trustworthiness and user comprehension. Furthermore, the development of standardized benchmarks and evaluation metrics will be essential for comparing unsupervised CART models and guiding their practical deployment. Overall, the review underscores that AI-enhanced unsupervised CART represents a powerful and evolving paradigm at the intersection of interpretable machine learning and advanced AI, offering promising avenues for future exploration and application.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

References

1. Ahmad, U., Han, M., & Mahmood, S. (2023). AI-enhanced threat analysis and risk assessment for connected and autonomous vehicles. Software Quality Journal, 31(6), 2457–2479.
2. Ahmed, M., & Husien, I. (2023). Heart disease prediction using hybrid machine learning: A brief review. Journal of Robotics and Control, 4(6), 285–295.
3. Alghanmi, N., Alotaibi, R., & Buhari, S. M. (2019). HLMCC: A hybrid learning anomaly detection model for unlabeled data in IoT. IEEE Access, 7, 18643–18658.
4. Almuqati, M. T., Sidi, F., & Mohd Rum, S. N. (2021). Challenges in supervised and unsupervised learning: A comprehensive overview. International Journal on Advanced Science, Engineering and Information Technology.
5. Azevedo, B. F., Rocha, A. M. A. C., & Pereira, A. I. (2023). Hybrid approaches to optimization and machine learning methods: A systematic literature review. Machine Learning, 112(9), 2789–2820.
6. Bala, J., Huang, J., Vafaie, H., DeJong, K., & Wechsler, H. (1995). Hybrid learning using genetic algorithms and decision trees. IJCAI, 719–724.
7. Bensi, M. E., & Esquivel, R. A. (2023). Significance of classification tree algorithm: A literature review. European Journal of Theoretical and Applied Sciences, 21(4), 65–80.
8. Bose, I., & Chen, X. (2009). Hybrid models using unsupervised clustering for prediction of customer churn. Journal of Organizational Computing and Electronic Commerce, 19(3), 133–150.
9. Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering, 40(1), 16–28.
10. Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm. Journal of Applied Science and Technology Trends, 2(1), 17–23.
11. Chen, D., Liu, J., & Song, Y. (2020). Semi-supervised hybrid CART ensemble for intrusion detection. Knowledge-Based Systems, 203, 106–123.