RESEARCH ARTICLE OPEN ACCESS

Recent Advances in Data Mining: A Systematic Review

Ashwinee Patil[1], Onkar Nath Thakur[2], Rakesh Kumar Tiwari[3], Dr. Vikas Gupta[4]
MTech Scholar[1], Assistant Professor[2&3], Professor[4]
Department of Computer Science & Engineering[1,2&3],
Department of Electronics & Communication Engineering[4]
Technocrats Institute of Technology & Science, Bhopal, India

Abstract-
In data-centric applications, data mining has become a vital pillar enabling the finding of hidden patterns and insightful analysis from large and complex datasets. Traditional analytic techniques are becoming more and more insufficient as data floods across sectors like healthcare, banking, cybersecurity, and e-commerce. With an eye on creative algorithms, hybrid models, optimization strategies, and specific domain applications, this systematic study seeks to combine current advances in data mining. Starting with the selection of 500 research papers, the evaluation process included 250 studies from other relevant fields, bringing the total to 750 records. Fifty-six records were left out following 150 duplication removals depending on relevancy and quality evaluation. At last, 44 studies were thoroughly examined and 44 reports included to the last analysis to guarantee a concentrated and high-quality synthesis. The examined literature shows a paradigm change from traditional approaches like decision trees and k-means clustering to sophisticated methodologies such as ensemble learning, deep learning, and graph-based mining. Strategies for optimization—including Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), and the Bat Algorithm (BAT)—have been effectively used with classifiers to enhance performance measures as accuracy and recall. The assessment also assesses how big data systems and real-time analytics work together to handle issues such high-dimensionality, feature sparsity, and computational overhead. This paper provides a brief and perceptive road map to changing data mining techniques by means of present trends, research gaps, and future possibilities, hence guiding researchers.

Keywords-Advanced Data Mining Techniques, Hybrid Classification Models, Optimization Strategies, Big Data Processing, Intelligent Decision-Making Systems.

Date of Submission: 06-05-2025 Date of acceptance: 18-05-2025

I. INTRODUCTION

All sectors of business apply data as their most valuable asset during the digital transformation era.[1] The massive quantity of data production continues at a rapid pace from sectors that include social media and e-commerce as well as banking and healthcare[2]. The unpacking of useful insights from extensive complex datasets represents both an imperative and a demanding challenge because data mining provides the necessary solution. Data mining operates through database analysis by statistical and machine learning and artificial intelligence tools to discover hidden relationship patterns and trends. Modern decision-making depends more heavily on it to make strategic choices while optimizing operations and identify forthcoming market patterns[3], and Essential Stages of Data Mining is shown in Fig. 1.

Figure 1: Essential Stages of Data Mining

Data mining represents an organized strategy for extracting essential knowledge from extensive databases. The process divides into five major putative phases. The start of data mining involves Problem Definition that facilitates the effective guidance of decision making through business objectives and problem definition. The second stage, Data Collection, focuses on gathering relevant and high-quality data from various sources[4]. During the Data Analysis stage researchers employ statistical analysis along with machine learning to detect hidden patterns inside the gathered data. The fourth section consists of Evaluation where researchers verify model performance levels and validate accuracy rates along with assessing if obtained insights fulfill designated business goals. The Deployment phase integrates the developed results and models with real-world decisions systems for practical implementation. Organizations benefit from this structured framework to make data-based choices and enhance operations while securing competitive superiority in their particular markets[5].

1.1 Background and importance of data mining

Data mining methods require more expertise and operational efficiency because data volumes continue to grow as well as the diversity and speed of data collection[6]. Decision trees along with Naïve Bayes combined with clustering algorithms form a basic foundation yet they experience challenges when used with high-dimensional data and real-time scalability demands. Researchers along with practitioners are compelled to develop hybrid optimization models that increase data mining performance accuracy levels[7].

1.2 Motivation for the review

A review has emerged due to the necessity to examine modern data mining techniques and their solutions for existing difficulties. The research goal of this paper is to close the knowledge gap by examining contemporary advancements especially through the combination of optimization algorithms PSO, ACO and BAT with SVM and Naïve Bayes classifiers.

1.3 Objectives and scope

This paper aims to give a thorough summary of the latest developments in data mining, underline hybrid methods that enhance feature selection and classification accuracy, and assess their performance on large-scale datasets. The range covers theoretical enhancements, algorithmic inventions, practical uses, and model comparison studies on benchmark data sets. This study intends to be a useful tool for researchers, data scientists, and business professionals interested in the evolution and future directions of data mining by means of a summary of important results and trends..

II. FOUNDATIONS OF DATA MINING

A fundamental part of knowledge discovery in databases (KDD), data mining is the process of identifying actionable, previously undiscovered, and relevant patterns from vast datasets. By using methods from statistics, machine learning, artificial intelligence, and database systems, it helps to convert raw data into useful insights[8]. Fundamentally, data mining includes several activities like classification, clustering, regression, association rule mining, anomaly detection, and forecasting.

Important ideas in data mining are data pretreatment (cleaning, integration, selection, transformation), feature selection, model development, assessment criteria, and validation approaches. The efficiency of mining algorithms is strongly influenced by the quality of data, hence efficient data pretreatment is absolutely vital[9].

Approaches-wise, conventional data mining emphasized structured data with well-defined schema and relied mostly on statistical models and rule-based systems. Due to their simplicity and interpretability, algorithms such as Decision Trees (C4.5), Naïve Bayes, k-Nearest Neighbors (k-NN), and k-Means clustering were commonly employed[10]. These techniques, however, frequently have problems with unstructured, noisy, high-dimensional data, and an Overview of Key Studies Utilizing Data Mining Techniques are shown in Table 1.

Table 1. Overview of Key Studies Utilizing Data Mining Techniques

Author	Main Focus	Techniques	Dataset	Limitations
[11]	Survey on predicting students' academic performance using educational data mining techniques and influential attributes.	ANN, Random Forest, WEKA tool	269 studies reviewed	Lacks empirical validation; scope limited to supervised methods.
[12]	Clustering aggregate electrical loads using unsupervised pretraining methods to improve demand-side management without relying on	CBHG, GRU, Cosine Similarity, Embedding Centroids	Electricity Load Diagrams (2011–2014)	Requires extensive training data; minimum data requirement unclear.

	demographic or appliance data.			
[13]	Proposing an automated clustering framework for classifying electricity customer loads with optimized clustering algorithms and feature selection methods.	Hierarchical clustering, LOESS, CART, PAM, k-means	218 customers from Seville, Spain	Small dataset; limited to specific customer tariff
[14]	Enhancing accuracy of non-invasive blood-glucose testing using upgraded FTIR spectroscopy and advanced machine learning methods.	FTIR, MATR, QCL, 2D correlation spectroscopy, Random Forest	Glucose-albumin solutions (70–300 mg/dL)	Experimental setup complexity; not yet tested clinically
[15]	Identifying determinants of household electricity consumption through clustering to target effective demand-side management.	K-means, Silhouette, Feature Selection, Cluster Validity	310 households, Botswana	Limited dataset; lacks smart-meter consumption data
[16]	Exploring key predictors of school performance in PISA 2018 using educational data mining on a global scale.	K-means, Decision Trees (C4.5), J48, Weka	PISA 2018 (78 countries, 20,663 schools)	Limited by cross-sectional data; generalizability hindered by SES overemphasis
[17]	Support game for dyslexia intervention using data mining and Orton-Gillingham method.	J48 algorithm, ADHD screening, Orton-Gillingham	Children with ADHD symptoms	Limited generalization; lacks broad clinical validation
[18]	Guest editorial overview on integrating blockchain and AI in 5G mobile edge computing environments.	Blockchain, 5G, MEC, Stackelberg Game, RFID	6 published articles in special section	Lacks experimental depth; mainly conceptual/editorial
[19]	Predicting housing prices using spatio-temporal data and clustering in Tehran.	OLS, GWR, GTWR, SKATER, PCA	Tehran Real Estate (SABAA)	Regional focus only; data not universally applicable
[20]	Hybrid ACO-GA for rule induction and classification accuracy improvement.	ACO, GA, Ant-Miner, Entropy, Pruning	12 UCI ML benchmark datasets	Computationally intensive; domain-specific tuning needed
[21]	Enhancing initial clustering using swarm intelligence for data density computation.	FCM, PSO, Leapfrog algorithm, Swarm intelligence	Simulated data (unspecified)	Sensitivity to initial parameters; not real-world validated
[22]	Integrating Bluetooth, traffic counts, and model data to analyze and predict highway traffic flows.	PredictiveApriori, Association Rules, Traffic Modeling	SGC Fi-Pi-Li Highway data	Limited by Bluetooth coverage and support values
[23]	Enhancing immunotherapy decision support via datamining in CBR retrieval phase using feature selection.	CBR, Decision Tree, KNN, jCOLIBRI	Immunotherapy patients (warts disease)	Focused on retrieval only; manual adaptation still needed
[24]	Load estimation for domestic smart meters using K-means and segmented clustering profiles.	K-means, Canberra, Manhattan, Pearson distances	UK smart meter data (segmented profiles)	Sensitive to cluster size and initial centroids

[25]	Hybrid IoT intrusion detection using cloud-fog ML framework for early threat mitigation.	ML, Data Mining, Honeypot, IDS, Cloud-Fog	IoT traffic (unspecified architecture)	High complexity, lacks large-scale real-world validation
[26]	Addressing unknowns, interdependency, and distinguishability in high-dimensional medical data for decision support.	Bayes Net, Inference Net, Q-UEL	Real-world structured medical data	Computationally intensive; abstract validation of coherence
[27]	Type 2 diabetes prediction using K-means and Random Forest classifiers on medical data.	K-means, Random Forest, MATLAB, Weka	UCI Diabetes Dataset	Overfitting risk; lacks comparison with deep learning
[28]	Evaluating K-means clustering performance with different distance metrics and datasets.	K-means, Euclidean, Manhattan, Minkowski	Synthetic datasets (100k–500k points)	Only distance metric focused; lacks real-world validation

Modern data mining methods have developed to manage real-time analytics and huge data. They combine sophisticated machine learning and deep learning models like Random Forests, Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Deep Neural Networks (DNN). Hybrid and ensemble models are also becoming more popular because of their strength and better accuracy[29]. Classifiers are being progressively integrated with optimizations employing nature-inspired algorithms such as PSO, ACO, and BAT to improve efficiency.

1. Recent Trends and Emerging Techniques

The combination of sophisticated machine learning and deep learning models has helped the area of data mining to expand quickly, hence enhancing the capacity to extract important patterns from complicated, high-dimensional information[30]. Deep learning architectures like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are increasingly supplementing machine learning techniques such as ensemble methods (Random Forest, Gradient Boosting) and support vector machines (SVM). Opening new horizons in fields including sentiment analysis, fraud detection, and healthcare analytics, these models shine at learning from unstructured data like photos, videos, and text[31].

Data mining research follows the process represented in the figure. The initial 500 plus 250 articles reduced to 150 duplicates before exclusion of 56 irrelevant records left 44 studies for review. A total of 44 studies survived the detailed review process after researchers excluded 56 records that proved to be irrelevant. A systematic methodology enables the production of focused synthesis containing high-quality information about current data mining technical developments and applications, and Systematic Review Architecture for Advancements in Data Mining Techniques is shown in Fig. 2.

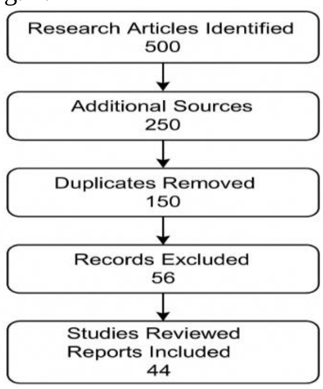

Figure 2: Systematic Review Architecture for Advancements in Data Mining Techniques

The inclusion of large data into data mining procedures is another important development. Distributed computing platforms like Hadoop and Apache Spark have made data mining systems capable of effectively analyzing petabytes of data. Big data integration enables thorough analysis, hence helping companies to glean insights from many data sources like transactional logs, IoT devices, and social media[32].

Real-time and stream mining approaches have evolved in response to the rising need for real-time decision-making[33]. These techniques let data streams be continuously analyzed, therefore allowing immediate anomaly detection, recommendation systems, and fraud protection. Algorithms are increasingly being created to be flexible and scalable, able to update their models on the fly as new data comes. These developing trends show a movement toward more smart, scalable, and responsive data mining tools, enabling businesses to obtain quick and actionable insights in a dynamic data environment [33], [34].

2. Optimization Techniques in Data Mining

Especially in the age of machine learning, deep learning, and big data, optimization is essential for improving the performance and accuracy of data mining algorithms. Optimization strategies include gradient descent, stochastic gradient descent, and adaptive methods like Adam and RMSprop are increasingly employed to fine-tune model parameters for quicker convergence and better accuracy as machine learning (ML) and deep learning (DL) continue to expand[35]. In deep learning, particularly in training sophisticated neural networks, such techniques lower computing load while improving generalization performance.

The amount, speed, and diversity of information create new difficulties for data mining by means of big data integration. Parallel and distributed optimization techniques are used to effectively handle large datasets, hence addressing these issues[36]. Frameworks like as Apache Spark and TensorFlow enable scalable optimization algorithms capable of managing data splitting, parallel computing, and in-memory processing, therefore facilitating real-time analytics and model upgrades.

A horizontal bar chart illustrates what percentage of organizations utilizes and considers important different data mining methods[37]. The results indicate Classification takes the top position with 8 before Clustering and Regression and Text Mining share the second position with 6. The place of Association Rules is midway in the ranking system with a 5 score. The application frequency of Anomaly Detection and Summarization remains low because survey participants rated them at 4 and 3 points respectively. The distribution shows a major preference for pattern recognition and predictive methods including classification and clustering which serve as essential elements in current data-driven systems. Primary focus is placed away from anomaly detection and summarization by authors in the surveyed publications, and Data Mining Techniques are shown in Fig. 3.

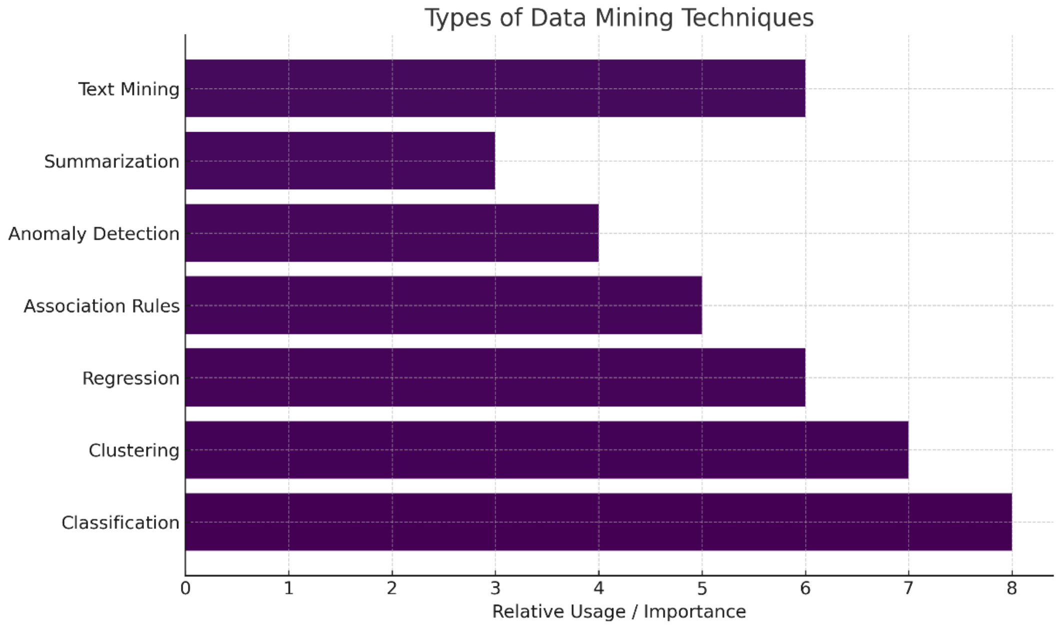

Figure 3: Data Mining Techniques

Furthermore, the rise of stream mining and real-time data calls for not just quick but also flexible optimization strategies. Models are updated constantly as new data streams in using online learning algorithms and incremental improvement techniques[38]. Applications like fraud detection, recommendation engines, and network monitoring—where quick response is required—benefit most from these techniques[39].

Modern data mining has made optimization methods absolutely essential. Especially in complicated settings defined by high-dimensional, dynamic, and streaming data, they guarantee that mining systems stay scalable, accurate, and responsive.

3. Data Mining Applications in Social Media Analytics

Particularly Twitter, Facebook, Instagram, and Reddit among social media channels have become strong providers of real-time, user-generated information. These sites provide large amounts of both organized and unstructured data that may be mined to find significant trends, user behavior, and public opinion[40]. Of all of them, Twitter is exceptional because of its open API, quick content flow, and brief message structure, making it perfect dataset for opinion mining and sentiment analysis.

A major data mining tool, sentiment analysis identifies positive, negative, or neutral sentiments spoken in posts or comments using natural language processing and machine learning methods. It clarifies public opinion on services, events, political concerns, and products. Opinion mining offers useful economic and societal insights by means of identification of important elements, entities, and emotional tones underlying user opinions[41].

Many case studies show how well data mining is in social media analytics. Mining Twitter data, for example, has been used to forecast results, identify false information, or examine public sentiment during elections other worldwide events. Companies track brand emotion in real time in marketing to change their plans. In public health, tracking disease outbreaks or public reaction to health measures has been aided by tweet analysis[42].

Experimental results have indicated that hybrid models—combining machine learning algorithms with optimization strategies—greatly increase accuracy in classification and sentiment detection. These methods improve the accuracy of social media analytics as well as provide more thorough analysis[43].

4. Evaluation Metrics and Performance Analysis

In data mining, evaluating model performance is essential to determine its reliability and predictive power[36]. Four key evaluation metrics are commonly used: Accuracy, Precision, Recall, and F1-score, all derived from the confusion matrix, which consists of:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1} - \text{score} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Where, A prediction system labels positive cases correctly as True Positives while negative cases correctly get classified as True Negatives. A false positive scenario occurs when the algorithm misidentifies negative cases as positive instances and false negative events develop when positive cases get mistakenly classified as negative instances. Model performance evaluation depends on these performance metrics.

Model performance across several categorization thresholds is also frequently assessed using Receiver Operating Characteristic (ROC) curves and the Area Under the Curve (AUC). A higher AUC shows more discriminative capacity. Data volume and resolution have also important effects. As dataset size increases, models generally improve due to more learning instances, but may also face overfitting or increased computation[44]. Higher data resolution improves feature richness, which results in more exact classifications. All things considered, integrating these measures lets one fully grasp the strengths, shortcomings, and appropriateness of models for practical data mining uses.

5. Challenges in Modern Data Mining

Many key problems current data mining runs into greatly affect its effectiveness and practical use. One major challenge is handling noisy, high-dimensional data. When datasets get more complex, including hundreds of features, conventional algorithms may struggle to identify important trends. Noise and irrelevant features might degrade model performance and need rigorous preprocessing, feature selection, and dimensionality reduction techniques as PCA or autoencoders. Scalability and computational complexity provide another challenge. Given the growth of big data in many industries, data mining systems must manage massive datasets efficiently. Algorithms must be tuned for speed and resource usage, especially when running in real-time environments or over distant systems. Keeping accuracy while ensuring scalability is a delicate balance.

Furthermore, ethical and privacy concerns are becoming more and more relevant. Often, data mining contains sensitive personal information, particularly in sectors like banking, healthcare, and social media. Data anonymization, adherence of legal requirements (such as GDPR), and transparent data usage policies help to maintain public confidence and legal compliance. Overcoming these challenges requires interdisciplinary approaches combining innovative concepts in algorithm design, data management, and ethics. Managing them releases the entire power of present data mining in practical applications.

6. Future Research Possibilities and Directions

Rapidly changing, the data mining sector is being shaped by several developing trends and research possibilities. The improvement in automation and AutoML (Automated Machine Learning) is one important path forward. From data preparation and model selection to hyperparameter tweaking and deployment, AutoML solutions seek to automate the end-to-end process of applying machine learning to practical issues. This not only speeds up the data mining process and makes it more accessible and efficient across sectors but also lessens the need for profound technical knowledge.

The use of Explainable artificial intelligence (XAI) into data mining methods is another important future development. Although black-box models like deep neural networks can provide great performance, their lack of openness is concerning in sensitive sectors such healthcare, banking, and law. By means of insights into how and why judgments are produced, explainable artificial intelligence methods seek to make data mining models more interpretable. This increases user confidence, helps ethical decision-making, and makes debugging easier.

Furthermore, cross-domain integration and multidisciplinary research will shape data mining in the future. The increasing need to integrate knowledge and techniques across disciplines is driven by the exponential expansion of data in areas such genetics, social sciences, economics, and environmental research. More context-aware, adaptable, and smart data mining systems may be produced via cooperative research including computer scientists, domain specialists, and data analysts.

The bar chart shows how research articles grew annually between 2016 to 2025. The data reveals rising publication numbers during 2016 to 2019 until 2020 brought a steep surge that reached its highest point at 13. Between 2021 and 2022 the publication numbers remained at 11 and 7 respectively. The annual publication numbers decrease considerably between 2023 and 2025 where each subsequent year registers between 1 to 2 articles. The research activity reached its highest level during 2020–2021 and this period probably enhanced both research interest and advancements which later led to decreased publication levels, and Annual Publication Trends in Data Mining Research is shown in Fig. 4.

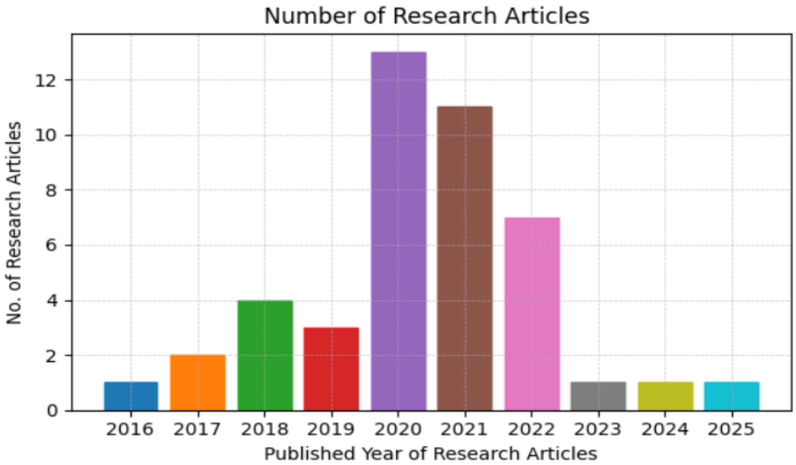

Figure 4: Annual Publication Trends in Data Mining Research

As society becomes increasingly data-driven but privacy-conscious and ecologically aware, research on privacy-preserving mining, federated learning, and sustainable computing is gaining steam. Real-time analytics, edge computing, and the application of quantum computing for managing complicated and large-scale mining challenges are also drawing more attention.

III. Conclusion

Data mining has become a must in data-centric applications for finding hidden patterns and deriving useful insights from huge, complicated datasets. Traditional analytical techniques are increasingly insufficient as data creation across areas like healthcare, banking, cybersecurity, and e-commerce grows. Evaluating 750 records—comprising 500 core research articles and 250 from allied domains—this methodical analysis combines present developments in data mining by deleting 150 duplicates and discarding 56 irrelevant studies. To guarantee a concentrated synthesis, a last group of 44 high-quality research was exhaustively examined. The results show a clear transition from traditional approaches like decision trees and k-means clustering to more sophisticated ones including ensemble learning, deep learning, and graph-based mining. Especially for high-dimensional data, hybrid models combining classical classifiers like Naïve Bayes and Support Vector Machines with metaheuristic optimizations—such as Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), and Bat Algorithm (BAT)—show significant gains in performance measures like accuracy and recall. Furthermore, tackling contemporary issues including feature sparsity and computational complexity requires the combination of big data platforms and real-time stream mining. This paper not only describes the present scene of data mining but also highlights future trends, current gaps, and possible research paths. Enriched by automation, explainable artificial intelligence, and multidisciplinary cooperation—ultimately allowing scalable, transparent, and efficient data-driven solutions—the future of data mining is in the evolution of intelligent, adaptable, and morally grounded systems.

References

[1] "A comparative study of clustering techniques for electrical load."

[2] D. Bari, M. Ameksa, and A. Ouagabi, "A comparison of datamining tools for geo-spatial estimation of visibility from AROME-Morocco model outputs in regression framework," in *2020 IEEE International conference of Moroccan Geomatics (Morgeo)*, Casablanca, Morocco: IEEE, May 2020, pp. 1–7. doi: 10.1109/Morgeo49228.2020.9121909.

[3] M. K. Gupta and P. Chandra, "A comprehensive survey of data mining," *Int. j. inf. tecnol.*, vol. 12, no. 4, pp. 1243–1257, Dec. 2020, doi: 10.1007/s41870-020-00427-7.

[4] "A data mining-based framework for the identification of daily electricity."

[5] A. Djellouli *et al.*, "A datamining approach to classify, select and predict the formation enthalpy for intermetallic compound hydrides," *International Journal of Hydrogen Energy*, vol. 43, no. 41, pp. 19111–19120, Oct. 2018, doi: 10.1016/j.ijhydene.2018.08.122.

[6] F. Es-Sabery *et al.*, "A MapReduce Opinion Mining for COVID-19-Related Tweets Classification Using Enhanced ID3 Decision Tree Classifier," *IEEE Access*, vol. 9, pp. 58706–58739, 2021, doi: 10.1109/ACCESS.2021.3073215.

[7] Y. Ma, Y. Lei, and T. Wang, "A Natural Scene Recognition Learning Based on Label Correlation," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 6, no. 1, pp. 150–158, Feb. 2022, doi: 10.1109/TETCI.2020.3034900.

[8] C. Savaglio and G. Fortino, "A Simulation-driven Methodology for IoT Data Mining Based on Edge Computing," *ACM Trans. Internet Technol.*, vol. 21, no. 2, pp. 1–22, Jun. 2021, doi: 10.1145/3402444.

[9] M. Taktak and S. Triki, "A spatiotemporal datamining approach for road profile estimation using low-cost device," *Procedia Computer Science*, vol. 207, pp. 2767–2781, 2022, doi: 10.1016/j.procs.2022.09.335.

[10] A. Gamazo and F. Martínez-Abad, "An Exploration of Factors Linked to Academic Performance in PISA 2018 Through Data Mining Techniques," *Front. Psychol.*, vol. 11, p. 575167, Nov. 2020, doi: 10.3389/fpsyg.2020.575167.

[11] S. Batool, J. Rashid, M. W. Nisar, J. Kim, H.-Y. Kwon, and A. Hussain, "Educational data mining to predict students' academic performance: A survey study," *Educ Inf Technol*, vol. 28, no. 1, pp. 905–971, Jan. 2023, doi: 10.1007/s10639-022-11152-y.

[12] X. Ruhang, "Efficient clustering for aggregate loads: An unsupervised pretraining based method," *Energy*, vol. 210, p. 118617, Nov. 2020, doi: 10.1016/j.energy.2020.118617.

[13] F. Biscarri, I. Monedero, A. García, J. I. Guerrero, and C. León, "Electricity clustering framework for automatic classification of customer loads," *Expert Systems with Applications*, vol. 86, pp. 54–63, Nov. 2017, doi: 10.1016/j.eswa.2017.05.049.

[14] J. K. Mandal and D. Bhattacharya, Eds., *Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018*, vol. 937. in Advances in Intelligent Systems and Computing, vol. 937. Singapore: Springer Singapore, 2020. doi: 10.1007/978-981-13-7403-6.

[15] L. Song, Z. Han, P.-W. Shum, and W.-M. Lau, "Enhancing the accuracy of blood-glucose tests by upgrading FTIR with multiple-reflections, quantum cascade laser, two-dimensional correlation spectroscopy and machine learning," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 327, p. 125400, Feb. 2025, doi: 10.1016/j.saa.2024.125400.

[16] E. L. Ofetotse, E. A. Essah, and R. Yao, "Evaluating the determinants of household electricity consumption using cluster analysis," *Journal of Building Engineering*, vol. 43, p. 102487, Nov. 2021, doi: 10.1016/j.jobe.2021.102487.

[17] A.-C. Paola *et al.*, "GlyphReader App: A support game for the application of the Orton-Gillingham Method with DataMining Techniques.," *Procedia Computer Science*, vol. 191, pp. 373–378, 2021, doi:

10.1016/j.procs.2021.07.071.

[18] "Guest Editorial: Blockchain and AI Enabled 5G Mobile Edge Computing," *IEEE Trans. Ind. Inf.*, vol. 16, no. 11, pp. 7067–7069, Nov. 2020, doi: 10.1109/TII.2020.2983764.

[19] A. Soltani, C. J. Pettit, M. Heydari, and F. Aghaei, "Housing price variations using spatio-temporal data mining techniques," *J Hous and the Built Environ*, vol. 36, no. 3, pp. 1199–1227, Sep. 2021, doi: 10.1007/s10901-020-09811-y.

[20] H. N. K. AL-Behadili, K. R. Ku-Mahamud, and R. Sagban, "Hybrid Ant Colony Optimization and Genetic Algorithm for Rule Induction," *Journal of Computer Science*, vol. 16, no. 7, pp. 1019–1028, Jul. 2020, doi: 10.3844/jcssp.2020.1019.1028.

[21] W. Xiong, "Initial Clustering Based on the Swarm Intelligence Algorithm for Computing a Data Density Parameter," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–8, Jun. 2022, doi: 10.1155/2022/6408949.

[22] A. Pratelli, M. Petri, M. Ierpi, and M. Di Matteo, "Integration of Bluetooth, Vehicle Count Data and Trasport Model Results by Means of Datamining Techniques," in *2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)*, Palermo: IEEE, Jun. 2018, pp. 1–6. doi: 10.1109/EEEIC.2018.8493997.

[23] F. Saadi, B. Atmani, and F. Henni, "Integration of datamining techniques into the CBR cycle to predict the result of immunotherapy treatment," in *2019 International Conference on Computer and Information Sciences (ICCIS)*, Sakaka, Saudi Arabia: IEEE, Apr. 2019, pp. 1–5. doi: 10.1109/ICCISci.2019.8716415.

[24] A. Al-Wakeel, J. Wu, and N. Jenkins, "k - means based load estimation of domestic smart meter measurements," *Applied Energy*, vol. 194, pp. 333–342, May 2017, doi: 10.1016/j.apenergy.2016.06.046.

[25] A. E. Ghazi and A. Moulay Rachid, "Machine learning and datamining methods for hybrid IoT intrusion detection," in *2020 5th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications (CloudTech)*, Marrakesh, Morocco: IEEE, Nov. 2020, pp. 1–6. doi: 10.1109/CloudTech49835.2020.9365895.

[26] B. Robson, S. Boray, and J. Weisman, "Mining real-world high dimensional structured data in medicine and its use in decision support. Some different perspectives on unknowns, interdependency, and distinguishability," *Computers in Biology and Medicine*, vol. 141, p. 105118, Feb. 2022, doi: 10.1016/j.compbiomed.2021.105118.

[27] M. S. Geetha Devasena, R. Kingsy Grace, and G. Gopu, "PDD: Predictive Diabetes Diagnosis using Datamining Algorithms," in *2020 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India: IEEE, Jan. 2020, pp. 1–4. doi: 10.1109/ICCCI48352.2020.9104108.

[28] T. M. Ghazal *et al.*, "Performances of K-Means Clustering Algorithm with Different Distance Metrics," *Intelligent Automation & Soft Computing*, vol. 29, no. 3, pp. 735–742, 2021, doi: 10.32604/iasc.2021.019067.

[29] S. Umamaheswari and K. Harikumar, "Analyzing Product Usage Based on Twitter Users Based on Datamining Process," in *2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, Dubai, United Arab Emirates: IEEE, Jan. 2020, pp. 426–430. doi: 10.1109/ICCAKM46823.2020.9051488.

[30] M. Rahman, S. A. Khushbu, and A. K. Mohammad Masum, "Associative datamining Survey on Modern Era People's engagement of Gaming addiction," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kharagpur, India: IEEE, Jul. 2021, pp. 01–05. doi: 10.1109/ICCCNT51525.2021.9579980.

[31] J. Yang *et al.*, "Brief introduction of medical database and data mining technology in big data era," *J Evidence Based Medicine*, vol. 13, no. 1, pp. 57–69, Feb. 2020, doi: 10.1111/jebm.12373.

[32] M. Revathy and S. Kamalakkannan, "Collaborative learning for improving intellectual skills of dropout students using datamining techniques," in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, Coimbatore, India: IEEE, Mar. 2021, pp. 236–240. doi: 10.1109/ICAIS50930.2021.9395912.

[33] L. Liu and H. L. Zhuang, "Computational prediction and characterization of two-dimensional pentagonal arsenopyrite FeAsS," *Computational Materials Science*, vol. 166, pp. 105–110, Aug. 2019, doi: 10.1016/j.commatsci.2019.04.040.

[34] Y. Li, X. Chu, D. Tian, J. Feng, and W. Mu, "Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm," *Applied Soft*

Computing, vol. 113, p. 107924, Dec. 2021, doi: 10.1016/j.asoc.2021.107924.

[35] G. J. Oyewole and G. A. Thopil, “Data clustering: application and trends,” *Artif Intell Rev*, vol. 56, no. 7, pp. 6439–6475, Jul. 2023, doi: 10.1007/s10462-022-10325-y.

[36] S. Patil and R. J. Anandhi, “Diversity based self-adaptive clusters using PSO clustering for crime data,” *Int. j. inf. tecnol.*, vol. 12, no. 2, pp. 319–327, Jun. 2020, doi: 10.1007/s41870-019-00311-z.

[37] I. Benítez, J.-L. Díez, A. Quijano, and I. Delgado, “Dynamic clustering of residential electricity consumption time series data based on Hausdorff distance,” *Electric Power Systems Research*, vol. 140, pp. 517–526, Nov. 2016, doi: 10.1016/j.epsr.2016.05.023.

[38] A. Hamdi, K. Shaban, A. Erradi, A. Mohamed, S. K. Rumi, and F. D. Salim, “Spatiotemporal data mining: a survey on challenges and open problems,” *Artif Intell Rev*, vol. 55, no. 2, pp. 1441–1488, Feb. 2022, doi: 10.1007/s10462-021-09994-y.

[39] F. Saidi, N. Sebaa, A. Mahmoudi, H. Aourag, G. Merad, and M. Dergal, “Structural electronic and mechanical properties of YM2 (M=Mn, Fe, Co) laves phase compounds: First principle calculations analyzed with datamining approach,” *Solid State Communications*, vol. 274, pp. 9–20, Jun. 2018, doi: 10.1016/j.ssc.2018.02.013.

[40] A. Khan and S. K. Ghosh, “Student performance analysis and prediction in classroom learning: A review of educational data mining studies,” *Educ Inf Technol*, vol. 26, no. 1, pp. 205–240, Jan. 2021, doi: 10.1007/s10639-020-10230-3.

[41] M. Ye, “The Datamining Algorithm on Knowledge Dependence,” in *2018 International Conference on Smart Grid and Electrical Automation (ICSGEA)*, Changsha: IEEE, Jun. 2018, pp. 234–236. doi: 10.1109/ICSGEA.2018.00065.

[42] I. P. Karhiaho *et al.*, “The hidden epidemic: Uncovering incidental fatty liver disease and its metabolic comorbidities by datamining in a hospital data lake – A real-world cohort study,” *Diabetes Research and Clinical Practice*, vol. 210, p. 111609, Apr. 2024, doi: 10.1016/j.diabres.2024.111609.

[43] M. Chiny, O. Bencharef, and Y. Chihab, “Towards a Machine Learning and Datamining approach to identify customer satisfaction factors on Airbnb,” in *2021 7th International Conference on Optimization and Applications (ICOA)*, Wolfenbüttel, Germany: IEEE, May 2021, pp. 1–5. doi: 10.1109/ICOA51614.2021.9442657.

[44] Z. Majcen Rosker, M. Vodicar, and E. Kristjansson, “Video-oculographic measures of eye movement control in the smooth pursuit neck torsion test can classify idiopathic neck pain patients from healthy individuals: A datamining based diagnostic accuracy study,” *Musculoskeletal Science and Practice*, vol. 61, p. 102588, Oct. 2022, doi: 10.1016/j.msksp.2022.102588.