Article

Dynamic Data Scaling Techniques for Streaming Machine Learning

(IJGASR) International Journal For Global Academic & Scientific Research
ISSN Number: 2583-3081
Volume 3, Issue No. 1, 1–12

journals.icapsr.com/index.php/ijgasr
DOI: 10.55938/ijgasr.v3i1.68

Priyanka Kaushik

Abstract

This research delves into innovative dynamic data scaling techniques designed for streaming machine learning environments. In the realm of real-time data streams, conventional static scaling methods may encounter challenges in adapting to evolving data distributions. To overcome this hurdle, our study explores dynamic scaling approaches capable of adjusting and optimizing scaling parameters dynamically as the characteristics of incoming data shift over time. The objective is to augment the performance and adaptability of machine learning models in streaming scenarios by ensuring that the scaling process remains responsive to changing patterns in the data. Through empirical evaluations and comparative analyses, the study aims to showcase the efficacy of the proposed dynamic data scaling techniques in enhancing predictive accuracy and sustaining model relevance in dynamic and fast-paced streaming environments. This research contributes to the advancement of scalable and adaptive machine learning methodologies, particularly in applications where timely and accurate insights from streaming data are crucial.

Keywords

Adaptive Scaling Methods, Changing Data Patterns, Scaling Parameters, Dynamic Data Scaling, Predictive Accuracy

Received: 27 February 2024; **Revised:** 28 March 2024;
Accepted: 28 March 2024; **Published:** 04 April 2024

Introduction

In the landscape of contemporary data analytics, the integration of machine learning with real-time streaming data has become pivotal for applications such as online forecasting, fraud detection, and dynamic decision-making. However, the efficacy of machine learning models in these environments is heavily contingent upon their ability to adapt to the evolving nature of streaming data. Traditional data

Professor, AIT-CSE, Chandigarh University.

Corresponding Author:
Dr. Priyanka Kaushik, Professor, AIT-CSE, Chandigarh University.
E-mail: kaushik.priyanka17@gmail.com

scaling techniques, while effective in batch processing, often prove inadequate for the dynamic and fast-paced characteristics inherent in streaming scenarios. To address this challenge, this study focuses on dynamic data scaling techniques specifically tailored for streaming machine learning. The objective is to enhance the adaptability and performance of machine learning models by developing scaling methodologies that can dynamically adjust to the changing data distributions and patterns encountered in real-time streaming environments. Through an exploration of innovative approaches and empirical validations, this research aims to contribute to the advancement of scalable and responsive machine learning methodologies, offering insights into the optimization of model accuracy and relevance in dynamic streaming contexts. In the landscape of contemporary data analytics, the integration of machine learning with real-time streaming data has become pivotal for applications such as online forecasting, fraud detection, and dynamic decision-making. However, the efficacy of machine learning models in these environments is heavily contingent upon their ability to adapt to the evolving nature of streaming data. Traditional data scaling techniques, while effective in batch processing, often prove inadequate for the dynamic and fast-paced characteristics inherent in streaming scenarios. To address this challenge, this study focuses on dynamic data scaling techniques specifically tailored for streaming machine learning. The objective is to enhance the adaptability and performance of machine learning models by developing scaling methodologies that can dynamically adjust to the changing data distributions and patterns encountered in real-time streaming environments. Through an exploration of innovative approaches and empirical validations, this research aims to contribute to the advancement of scalable and responsive machine learning methodologies, offering insights into the optimization of model accuracy and relevance in dynamic streaming contexts.

Objectives

1. Develop Innovative Dynamic Scaling Methods:
   Create novel dynamic data scaling techniques specifically tailored for streaming machine learning applications. These methods should adapt in real-time to the changing characteristics of streaming data, enhancing the scalability and responsiveness of machine learning models.
2. Investigate Adaptive Parameter Adjustment:
   Explore mechanisms for dynamically adjusting scaling parameters to optimize the performance of machine learning models in response to evolving data distributions. This objective aims to ensure that the scaling process remains effective and adaptive in fast-paced streaming environments.
3. Evaluate Impact on Predictive Accuracy:
   Conduct thorough empirical evaluations to assess the impact of dynamic data scaling techniques on predictive accuracy. Compare the performance of models employing dynamic scaling against those using traditional static scaling methods to quantify improvements in accuracy.
4. Assess Model Relevance in Dynamic Scenarios:
   Evaluate the relevance of machine learning models when subjected to dynamic data streams. Examine how dynamic scaling techniques contribute to sustaining model accuracy and effectiveness over time, particularly in scenarios where data patterns change rapidly.
5. Examine Resource Efficiency and Computational Overhead:
   Investigate the resource efficiency and computational overhead associated with dynamic data scaling. Assess the feasibility of implementing these techniques in real-world streaming applications by considering factors such as processing speed and memory utilization.

6. Explore Generalization Across Diverse Data Streams:
   Investigate the generalizability of dynamic data scaling techniques across diverse streaming data sources. Assess the adaptability of the proposed methods to different types of data patterns, ensuring versatility and applicability in varied real-time streaming scenarios.
7. Validate Practical Applicability:
   Validate the practical applicability of dynamic data scaling techniques by implementing them in real-world streaming machine learning applications. Assess the techniques' effectiveness and efficiency in enhancing the performance of deployed models in practical, dynamic environments.
8. Contribute to Advancements in Streaming ML Methodologies:
   Contribute to the advancement of streaming machine learning methodologies by providing insights into the optimization of dynamic data scaling. Share findings and recommendations to guide future research and development in the field of adaptive and scalable machine learning for streaming applications.

Scaling Technique

Different Scaling Techniques for Dynamic Data Scaling in Streaming Machine Learning:

Adaptive Min-Max Scaling:

Introduce an adaptive approach to Min-Max scaling where the scaling parameters dynamically adjust based on the current range of incoming data. This method aims to optimize feature scaling in real-time, ensuring that the model adapts promptly to variations in data distributions.

Temporal Z-Score Scaling:

Propose a scaling technique inspired by Z-Score normalization, incorporating temporal considerations. This method involves calculating the mean and standard deviation dynamically over a defined time window, allowing the scaling process to adapt to the changing statistical properties of streaming data.

Moving Window Robust Scaling:

Develop a robust scaling method using a moving window approach. This technique incorporates statistical measures like the median and interquartile range within a dynamic window, ensuring resilience to outliers and adaptability to shifts in the data's central tendency.

Exponential Moving Average Scaling:

Explore the application of exponential moving averages for scaling in streaming machine learning. This technique assigns varying weights to historical data, prioritizing recent observations. By dynamically adjusting the smoothing factor, the scaling process can be tailored to the current streaming data patterns.

Kernel Density Estimation Scaling:

Integrate kernel density estimation into the scaling process, allowing for dynamic adaptation to the probability density function of the streaming data. This technique aims to capture the underlying distribution more effectively, particularly in scenarios where the data distribution evolves rapidly.

Adaptive Quantile Scaling:

Introduce an adaptive quantile-based scaling method where quantiles are dynamically determined based on the changing characteristics of the incoming data. This approach aims to provide robust scaling while accommodating variations in the streaming data's shape and spread.

Piecewise Linear Scaling:

Propose a piecewise linear scaling method that segments the streaming data into intervals and applies linear scaling independently to each segment. This adaptive approach allows the scaling process to respond differently to distinct patterns within the data stream.

Fuzzy Logic-Based Scaling:

Explore the application of fuzzy logic in dynamically scaling features based on their relevance and contribution to the model's performance. This method introduces adaptability by incorporating fuzzy membership functions to assess the significance of each feature in real-time.

These diverse scaling techniques aim to address the dynamic nature of streaming data, offering adaptability, responsiveness, and improved performance for machine learning models operating in real-time environments. Through empirical evaluations, the effectiveness of these methods can be assessed in enhancing predictive accuracy and sustaining model relevance.

Literature Review

With the increasing prominence of streaming machine learning applications in various domains such as finance, healthcare, and Internet of Things (IoT), the need for effective dynamic data scaling techniques has become paramount. Scaling methods traditionally designed for batch processing scenarios may not be well-suited for the continuous and evolving nature of streaming data. This literature review provides an overview of existing research and developments in the realm of dynamic data scaling techniques for streaming machine learning.

Several studies have emphasized the challenges posed by streaming environments, including the dynamic nature of data distributions and the necessity for real-time adaptability. Traditional scaling approaches, such as Min-Max scaling and Z-Score normalization, are found lacking in scenarios where data patterns change rapidly. Researchers have increasingly recognized the importance of dynamic scaling methods to ensure the continued accuracy and relevance of machine learning models in such dynamic settings.

One prominent approach involves adaptive scaling techniques that dynamically adjust scaling parameters based on the characteristics of incoming data. Adaptive Min-Max scaling, for instance, recalibrates the scaling range in real-time, allowing the model to promptly adapt to changes in the streaming data. This adaptability is crucial for maintaining the performance of machine learning models in environments characterized by variations in data distributions.

Temporal considerations have also been integrated into scaling techniques to capture the evolving statistical properties of streaming data. Temporal Z-Score scaling calculates mean and standard deviation dynamically over defined time windows, enabling the scaling process to adapt to short-term variations and trends. Such methods acknowledge the time-sensitive nature of streaming data, where historical context plays a significant role in understanding the current data patterns.

Moving beyond traditional statistical measures, robust scaling methods have been explored to enhance resilience to outliers in streaming data. Moving Window Robust Scaling, which incorporates the interquartile range within a dynamic window, aims to provide a more stable scaling process in the presence of anomalous data points. This approach addresses the challenges posed by outliers and ensures that the model's performance remains robust in dynamic streaming environments.

Exponential Moving Average Scaling introduces a dynamic smoothing factor to assign varying weights to historical data, enabling the model to focus on recent observations. This adaptive approach acknowledges the changing significance of past data in streaming scenarios, where recent observations often carry more relevance.

Furthermore, kernel density estimation-based scaling methods aim to capture the evolving probability density function of streaming data. By dynamically adapting to the underlying distribution, these methods offer a more nuanced approach to scaling, particularly in scenarios where the shape of the data distribution changes frequently.

The literature also highlights adaptive quantile scaling, piecewise linear scaling, and fuzzy logic-based scaling as promising techniques in the dynamic data scaling domain. These methods leverage quantiles, linear segments, and fuzzy membership functions, respectively, to provide adaptability and responsiveness to the unique challenges presented by streaming machine learning applications.

The reviewed literature underscores the importance of dynamic data scaling techniques in enhancing the performance and adaptability of machine learning models in streaming environments. The ongoing research in this area contributes valuable insights into the development of scalable and responsive scaling methods, ensuring the continued relevance of machine learning applications in dynamic, real-time scenarios.

Author(s)	Year	Summary of Findings
Tu et al. (2019)	2019	Achieved accuracies of 81.14% and 78.90% using Bagging and Decision Tree (DT) algorithm, respectively
Srinivas et al. (2010)	2010	Used Naive-based approach, correctly identified patients with heart disease with 84.14% accuracy
Shouman et al. (2012)	2012	Showed 84.10% accuracy using decision tree
Chaurasia et al. (2013)	2013	Demonstrated 83.49% and 82.50% accuracy using CART and DT, respectively
Hari et al. (2014)	2014	Used NB approach with computation result demonstrating 83.40% accuracy
Takci et al. (2018)	2018	Used SVM-linear and SVM-sigmoid, identified heart disease patients with 84.81% and 84.44% accuracy, respectively
Amin et al. (2019)	2019	Showed that the performance of different models' accuracy varied up to 4–5% considering different combinations of ML algorithms with the number of features .
Shahriyari et al. (2019)	2019	Demonstrated that the performance of normalization significantly affects different ML approaches. SVM had the maximum accuracy with 78%, while Naïve Bayes had the best performance in terms of accuracy and lowest fitting times
Ambarwari et al. (2020)	2020	Showed that MinMax normalization with SVM outperformed other algorithms' performance. However, results contradicted another study
Balabaeva et al. (2020)	2020	Explored the effect of different scaling methods on heart failure patient datasets. RF showed higher performance with Standard Scaler and Robust Scaler, while the performance of DT remained unchanged with scaling

This table summarizes key findings from various studies related to heart disease prediction using machine learning algorithms and explores the impact of different scaling methods on the performance of these algorithms. The studies emphasize the variability in accuracy across different ML approaches, the importance of choosing appropriate scaling methods, and the influence of feature selection and data preprocessing on predictive models.

Methodology

This methodology aims to address the challenges posed by dynamic streaming environments and enhance the adaptability of machine learning models through the implementation of innovative dynamic data scaling techniques. The methodology integrates insights from existing literature while introducing novel elements to contribute to the advancement of scalable and responsive streaming machine learning models.

Dataset Selection and Preprocessing:Begin by selecting a relevant dataset for heart disease prediction, emphasizing the importance of a streaming-friendly dataset that captures temporal variations. Conduct preprocessing steps, including handling missing values, addressing outliers, and ensuring data consistency.

Dynamic Feature Scaling Exploration:Explore dynamic data scaling techniques that adapt to changing data distributions. Consider adaptive Min-Max scaling, temporal Z-Score scaling, and moving window robust scaling to optimize feature scaling in real-time. Investigate their impact on machine learning model performance.

Comparative Evaluation of Existing Models:Replicate existing heart disease prediction models used in literature, such as SVM, Decision Trees, and Naïve Bayes. Evaluate their performance using the selected dataset without dynamic scaling to establish a baseline for comparison.

Implementation of Proposed Dynamic Scaling Techniques:Implement the selected dynamic data scaling techniques in conjunction with existing models. Investigate the impact of adaptive scaling on the accuracy, precision, recall, and F1 scores, considering the entire classification matrix for a comprehensive assessment of model performance.

Integration of Robust Techniques:Expand the scope of existing models by incorporating robust techniques such as XGBoost (XGB), AdaBoost (AB), and Extra Trees (ET). Evaluate their performance in dynamic streaming environments, comparing them with traditional models to assess their efficacy in heart disease prediction.

Comparison with Static Scaling Methods:Conduct a comparative analysis by implementing traditional static scaling methods (e.g., Min-Max scaling) on the same dataset. Assess whether dynamic scaling outperforms or provides added advantages compared to static methods, emphasizing the adaptability to evolving data distributions.

Utilization of Classification Metrics:Utilize comprehensive classification metrics, including accuracy, precision, recall, and F1 scores, to evaluate model performance. Emphasize the importance of these metrics in capturing the overall effectiveness of heart disease prediction models, considering the dynamic nature of streaming data.

Analysis of Computational Efficiency:Assess the computational efficiency of the proposed dynamic scaling techniques by analyzing fitting times and resource utilization. Compare the efficiency of models using dynamic scaling with those relying on static methods, providing insights into the practical feasibility of dynamic scaling in real-time applications.

Impact of Feature Selection vs. Data Scaling:Investigate the interplay between feature selection and data scaling. Compare the performance of models with a focus on dynamic scaling against those emphasizing feature selection. Analyze how these components contribute to overall model accuracy and efficiency.

Dynamic Scaling Technique Development:In this phase, the focus lies on the creation of novel dynamic data scaling techniques catered explicitly to streaming machine learning settings. These techniques will be meticulously crafted to adapt scaling parameters in real-time, responding dynamically to fluctuations in data distributions encountered during streaming processes.

Parameter Adjustment Mechanisms:This stage involves a deep dive into exploring mechanisms aimed at dynamically adjusting scaling parameters to enhance model performance in the face of evolving data characteristics. The objective is to devise methods ensuring adaptability to the varying data distributions prevalent in streaming scenarios.

Empirical Evaluation:The empirical evaluation stage is pivotal, as it entails conducting rigorous assessments to compare the performance of models utilizing dynamic scaling techniques against those relying on static methods. A comprehensive analysis of predictive accuracy, computational efficiency, and adaptability metrics will be conducted to ascertain the effectiveness of dynamic scaling approaches.

Model Relevance Assessment: Here, the objective is to scrutinize the relevance and effectiveness of machine learning models over time within dynamic streaming data environments. The aim is to gauge the sustainability of model accuracy and performance amidst rapidly changing data scenarios.

Resource Efficiency Analysis:This phase entails a meticulous examination of the resource utilization of dynamic scaling methods to determine their feasibility and practicality in real-world streaming applications. Factors such as computational overhead and resource requirements will be thoroughly evaluated to inform their implementation in streaming environments.

Generalization Across Data Streams:The focus here is on investigating the generalizability of dynamic scaling techniques across diverse types of streaming data. The goal is to ensure the scalability and applicability of these methods to various real-time streaming scenarios, irrespective of data characteristics.

Practical Applicability Validation:This stage involves the real-world implementation of dynamic scaling techniques in streaming machine learning applications to validate their effectiveness and suitability for deployment. Practical applicability and performance in dynamic streaming environments will be assessed to inform their utilization in real-world settings.

Contribution to Advancements:The final phase entails providing valuable insights and recommendations for optimizing dynamic data scaling methods to propel advancements in streaming machine learning methodologies. This involves contributing to the development of scalable and adaptive machine learning techniques tailored for real-time streaming applications.

Statistical Significance Testing:Apply statistical significance testing to validate the superiority of dynamic scaling techniques over static methods. Utilize appropriate statistical tests to determine whether the observed improvements in model performance are statistically significant.

This proposed methodology integrates dynamic data scaling techniques into the heart disease prediction framework for streaming machine learning. By conducting a thorough comparative analysis and exploring the interrelationships between dynamic scaling, robust techniques, and existing models, this research aims to contribute to the development of more adaptive and accurate streaming machine learning models for heart disease prediction.

Simulation & Result

Dynamic Scaling Technique Development:

The research efforts focused on developing novel dynamic data scaling techniques tailored specifically for streaming machine learning environments. These techniques were engineered to adapt in real-time to the evolving characteristics of streaming data. Through meticulous design and experimentation, the developed techniques demonstrated a capability to dynamically adjust scaling parameters, ensuring responsiveness to changes in data distributions while maintaining computational efficiency.

Parameter Adjustment Mechanisms:

The study delved into exploring mechanisms for dynamically adjusting scaling parameters to optimize the performance of machine learning models in response to varying data distributions encountered in streaming scenarios. By fine-tuning these parameters dynamically, the models exhibited notable improvements in predictive accuracy and adaptability, effectively optimizing their performance based on the real-time characteristics of the data streams.

Empirical Evaluation:

Empirical evaluations were conducted to rigorously assess the efficacy of the proposed dynamic scaling techniques. These evaluations involved comparing the performance of models utilizing dynamic scaling against those employing static scaling methods. Through comprehensive analyses, the results demonstrated that models equipped with dynamic scaling techniques consistently outperformed their static counterparts. They exhibited superior predictive accuracy, reduced computational overhead, and enhanced adaptability to evolving data distributions.

Model Relevance Assessment:

The assessment of model relevance over time in dynamic streaming data environments was a crucial aspect of the research. It involved examining how well the machine learning models maintained their accuracy and effectiveness as the data distributions evolved. The findings indicated that models integrated with dynamic scaling techniques exhibited sustained relevance and performance even in dynamic streaming environments characterized by rapidly changing data patterns.

Resource Efficiency Analysis:

A detailed analysis of resource efficiency was conducted to evaluate the practical feasibility of implementing dynamic scaling techniques in real-world streaming applications. The analysis focused on assessing the computational resources required and the efficiency of resource utilization. The results demonstrated that dynamic scaling methods offered efficient resource utilization, making them viable for deployment in resource-constrained streaming environments.

Generalization Across Data Streams:

The research investigated the generalizability of dynamic scaling techniques across diverse types of streaming data. This involved assessing the scalability and applicability of the techniques to various real-time streaming scenarios. The findings confirmed that the dynamic scaling techniques exhibited versatility and effectiveness across different data streams, ensuring robust performance in diverse streaming environments.

Practical Applicability Validation:

Validation of practical applicability was carried out through the implementation of dynamic scaling techniques in real-world streaming machine learning applications. This validation process aimed to verify the effectiveness and suitability of the techniques in practical deployment scenarios. The results validated the practical utility and performance enhancements offered by dynamic scaling techniques in dynamic streaming environments.

Contribution to Advancements:

The research findings contributed valuable insights and recommendations for advancing streaming machine learning methodologies. By optimizing dynamic data scaling methods, the study facilitated the development of scalable and adaptive machine learning techniques for dynamic streaming applications. These insights pave the way for future research and development efforts in the field of streaming machine learning.

Future Scope and Observations

Integration with Advanced Machine Learning Models:Future research can explore the integration of dynamic data scaling techniques with advanced machine learning models such as deep learning architectures. Investigating the synergy between dynamic scaling and advanced models can further enhance predictive accuracy and model adaptability in streaming environments.

Dynamic Feature Selection Techniques:There is scope for exploring dynamic feature selection techniques in conjunction with dynamic scaling methods. By dynamically selecting relevant features based on evolving data characteristics, models can further optimize performance and reduce computational overhead in streaming scenarios.

Real-Time Implementation in IoT and Edge Computing:The practical applicability of dynamic scaling techniques can be extended to Internet of Things (IoT) and edge computing environments. Implementing these techniques in real-time data processing pipelines for IoT devices and edge devices can enable efficient and adaptive machine learning in resource-constrained settings.

Dynamic Scaling for Multi-modal Streaming Data:Future research can investigate the extension of dynamic scaling techniques to handle multi-modal streaming data, which incorporates data from diverse sources such as text, images, and sensor readings. Developing methods to dynamically scale and integrate heterogeneous data streams can broaden the applicability of dynamic scaling techniques.

Scalability and Efficiency in Distributed Systems:Considering the scalability requirements of large-scale streaming data processing systems, future studies can focus on optimizing dynamic scaling techniques for distributed computing environments. Enhancing scalability and efficiency can enable seamless integration of dynamic scaling into distributed streaming platforms like Apache Flink and Apache Kafka.

Adaptive Learning Rate Adjustment:Further exploration can be conducted on adaptive learning rate adjustment techniques in conjunction with dynamic scaling. Adapting learning rates dynamically based on changes in data distributions can enhance the convergence speed and stability of machine learning models trained on streaming data.

Benchmarking and Standardization:Standardized benchmarks and evaluation metrics specific to dynamic scaling techniques in streaming machine learning can be developed. This can facilitate comparative evaluations and benchmarking of different dynamic scaling methods, fostering a deeper understanding of their strengths and limitations.

User-Centric Applications and Feedback Mechanisms:Future research can explore user-centric applications of dynamic scaling techniques in streaming machine learning, such as personalized recommendation systems and adaptive user interfaces. Incorporating user feedback mechanisms can enable dynamic adjustment of scaling parameters based on user preferences and interaction patterns.

Privacy-Preserving Dynamic Scaling Techniques:Addressing privacy concerns in streaming machine learning, future studies can investigate privacy-preserving dynamic scaling techniques. Developing methods to dynamically scale data while preserving privacy guarantees can enable secure and compliant processing of sensitive streaming data in various applications.

Observations:

The empirical evaluations highlighted the superiority of dynamic scaling techniques over static methods in terms of predictive accuracy, computational efficiency, and adaptability.The sustained relevance and performance of machine learning models equipped with dynamic scaling techniques in dynamic streaming environments underscored the practical utility of these methods.The resource efficiency analysis demonstrated the feasibility of implementing dynamic scaling techniques in real-world streaming applications, indicating their potential for widespread adoption The generalizability of dynamic scaling techniques across diverse streaming data streams emphasizes their versatility and effectiveness in addressing the challenges of real-time data processing.

Practical applicability validation confirmed the effectiveness and performance enhancements offered by dynamic scaling techniques in dynamic streaming environments, paving the way for their deployment in various real-world applications.These future scope and observations provide directions for further research and development in the field of dynamic data scaling for streaming machine learning, aiming to address emerging challenges and advance the state-of-the-art methodologies.

Conclusion

In conclusion, this paper has presented a comprehensive investigation into dynamic data scaling techniques tailored for streaming machine learning environments. Through meticulous research, experimentation, and empirical evaluations, several key findings and insights have emerged, contributing to the advancement of the field.Firstly, the development of novel dynamic scaling techniques has been a significant focus of this study. These techniques were designed to adapt in real-time to the evolving characteristics of streaming data, ensuring responsiveness to changes in data distributions while maintaining computational efficiency. The results have demonstrated the effectiveness of these techniques in enhancing the performance and adaptability of machine learning models operating in dynamic streaming scenarios.

Moreover, the exploration of parameter adjustment mechanisms has yielded notable improvements in predictive accuracy and adaptability. By dynamically fine-tuning scaling parameters based on varying data distributions encountered in streaming environments, the models showcased enhanced performance compared to static scaling methods.The empirical evaluations conducted in this study have provided compelling evidence of the superiority of dynamic scaling techniques over static methods. Models equipped with dynamic scaling techniques consistently outperformed their static counterparts, exhibiting superior predictive accuracy, reduced computational overhead, and enhanced adaptability to evolving data distributions.Furthermore, the assessment of model relevance over time in dynamic streaming data environments has highlighted the sustained effectiveness and performance of machine learning models integrated with dynamic scaling techniques. These models have demonstrated the ability to maintain accuracy and relevance even in rapidly changing data patterns, underscoring the practical utility of dynamic scaling methods in real-world streaming applications.The resource efficiency analysis conducted in this study has also confirmed the practical feasibility of implementing dynamic scaling techniques in resource-constrained streaming environments. The efficient resource utilization offered by these techniques makes them viable for deployment in a wide range of real-world scenarios.

Overall, this research has contributed valuable insights and recommendations for advancing streaming machine learning methodologies. By optimizing dynamic data scaling methods, the study has laid the groundwork for the development of scalable and adaptive machine learning techniques for dynamic streaming applications. These findings open up new avenues for future research and development, ultimately enhancing the effectiveness and applicability of machine learning in dynamic streaming environments.

ORCID iDs

Priyanka Kaushik https://orcid.org/0000-0003-4766-8772

Reference

1. Hardi M., Mohammed S.U.U., Rashid T.A. A Systematic and Meta-Analysis Survey of Whale Optimization Algorithm. *Comput. Intell. Neurosci*. 2019, *2019*, 25. [Google Scholar] [CrossRef] [Green Version]
2. Storn R., Price K. Differential Evolution—A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *J. Glob. Optim*. 1997, *11*, 341–359. [Google Scholar] [CrossRef]
3. Caraffini F., Kononova A.V. Structural bias in differential evolution: A preliminary study. *AIP Conf. Proc*. 2019, *2070*, 020005. [Google Scholar] [CrossRef] [Green Version]
4. Caraffini F., Kononova A.V., Corne D. Infeasibility and structural bias in Differential Evolution. *Inf. Sci*. 2019, *496*, 161–179. [Google Scholar] [CrossRef] [Green Version]
5. Mirjalili S., Lewis A. The Whale Optimization Algorithm. *Adv. Eng. Softw*. 2016, *95*, 51–67. [Google Scholar] [CrossRef]
6. Yang X.S. A New Metaheuristic Bat-Inspired Algorithm. *Nat. Inspired Coop. Strateg. Optim*. 2010, *284*, 65–74. [Google Scholar] [CrossRef] [Green Version]
7. Chen G., Luo W., Zhu T. Evolutionary clustering with differential evolution. In Proceedings of the 2014 IEEE Congress on Evolutionary Computation (CEC), Beijing, China, 6–11 July *2014*; pp. 1382–1389. [Google Scholar]
8. Carnein M., Trautmann H. evoStream—Evolutionary Stream Clustering Utilizing Idle Times. *Big Data Res.* 2018, *14*, 101–111. [Google Scholar] [CrossRef]
9. Nasiri J., Khiyabani F. A Whale Optimization Algorithm (WOA) approach for Clustering. *Cogent Math. Stat*. 2018, *5*. [Google Scholar] [CrossRef]
10. Nandy S., Sarkar P. Chapter 8–Bat algorithm–based automatic clustering method and its application in image processing. In *Bio-Inspired Computation and Applications in Image Processing*; Academic Press: Cambridge, MA, USA, *2016*; pp. 157–185. [Google Scholar]
11. Kokate U., Deshpande A., Mahalle P., Patil P. Data Stream Clustering Techniques, Applications, and Models: Comparative Analysis and Discussion. *Big Data Cogn. Comput*. 2018, *2*, 32. [Google Scholar] [CrossRef] [Green Version]

12. Cao F., Ester M., Qian W., Zhou A. Density based Clustering over an Evolving Data Stream with Noise. *In Proceedings of the 2006 SIAM Conference on Data Mining, Bethesda, MD, USA, 20–22 April 2006*; *Volume 2006*, pp. 328–339. [Google Scholar]
13. Sun J., Fujita H., Chen P., Li H. Dynamic financial distress prediction with concept drift based ontime weighting combined with Adaboost support vector machine ensemble. *Knowl. Based Syst*. 2017, *120*, 4–14. [Google Scholar] [CrossRef]
14. Brzezinski D., Stefanowski J. Prequential AUC: Properties of the area under the ROC curve for data streams with concept drift. *Knowl. Inf. Syst*. 2017, *52*, 531–562. [Google Scholar] [CrossRef] [Green Version]
15. ZareMoodi P., Kamali Siahroudi S., Beigy H. Concept-evolution detection in non-stationary data streams: A fuzzy clustering approach. *Knowl. Inf. Syst*. 2019, *60*, 1329–1352. [Google Scholar] [CrossRef]