Analyzing GPU Efficiency in Cryptocurrency Mining: A Comparative Study Using K-Means Clustering on Algorithm Performance Metrics

Joe Khosa[1,*], Ayorinde Olanipekun[2]

[1]Department of Engineering Management, Faculty of Engineering and the Built Environment, University of Johannesburg, South Africa
[2]Data Science Across Disciplines Research Group, Faculty of Engineering and the Built Environment, University of Johannesburg, South Africa

ABSTRACT

This study employs clustering analysis to evaluate the efficiency of GPUs used in cryptocurrency mining, categorizing them into distinct groups based on computational output and power consumption. Using K-Means clustering, GPUs were grouped into three clusters: low-efficiency, moderate-efficiency, and high-efficiency. High-efficiency GPUs demonstrated superior hash rates (e.g., 104.79 Mh/s for AbelHash and 218.35 Mh/s for Autolykos2) despite higher power consumption, making them ideal for high-performance mining operations. Conversely, low-efficiency GPUs exhibited lower computational output and modest energy use, highlighting opportunities for hardware upgrades or repurposing. Visualization techniques, including scatter plots and pair plots, provided clear distinctions between clusters, while a silhouette score of 0.35 indicated moderate cluster separation, suggesting areas for further refinement. The findings offer actionable insights for optimizing hardware selection, reducing operational costs, and improving energy efficiency in mining operations. Additionally, this study underscores the importance of sustainability in cryptocurrency mining and provides a foundation for future research, including the integration of additional performance metrics, exploration of alternative clustering algorithms, and development of energy-efficient mining practices. These insights contribute to the broader goal of fostering a more sustainable and data-driven approach to cryptocurrency mining.

Keywords GPU Efficiency, Cryptocurrency Mining, K-Means Clustering, Energy Consumption, Computational Output, Hardware Optimization

INTRODUCTION

Cryptocurrency mining is a critical process that underpins the functionality of blockchain networks, driving the validation of transactions and the generation of new coins. This intricate process involves the solving of complex mathematical problems, necessitating substantial computational power and robust energy resources. Mining operations predominantly rely on specialized hardware, such as Application-Specific Integrated Circuits (ASICs) and Graphics Processing Units (GPUs), which are pivotal in determining the operational efficiency of mining endeavors [1], [2]. The selection of mining technology critically influences outcomes, dictating hash rates and sculpting energy consumption profiles [1], [3].

Driven by consensus mechanisms, mining activity thrives within frameworks such as Proof of Work (PoW). Here, miners engage in a computational melee, vying to be the first to solve cryptographic puzzles—a conquest rewarded by

Submitted 20 January 2025
Accepted 28 April 2025
Published 1 June 2025

Corresponding author
Joe Khosa,
joe.khosa@gmail.com

Additional Information and Declarations can be found on page 131

DOI: 10.47738/jcrb.v2i2.34

How to cite this article: J. Khosa, A. Olanipekun, "Analyzing GPU Efficiency in Cryptocurrency Mining: A Comparative Study Using K-Means Clustering on Algorithm Performance Metrics," *J. Curr. Res. Blockchain,* vol. 2, no. 2, pp. 118-134, 2025.

cryptocurrency [4], [5]. This model not only stabilizes networks, safeguarding against double-spending, but also incentivizes vigilance among miners, fortifying blockchain integrity. Yet, the high energy demands that accompany PoW mining pose environmental dilemmas, especially in terms of the gargantuan carbon footprint left by cryptocurrencies like Bitcoin [5], [6]. Therein lies a quest for balance, as studies advocate for renewable energy adoption and the deployment of energy-efficient hardware to mitigate these impacts [6].

The profitability of mining, inhibited or propelled by multifarious factors, dances with market conditions, mining difficulty nuances, and operational costs, encompassing electricity and equipment maintenance [3], [7]. Price volatility in cryptocurrency markets can precipitate speculative bubbles, where the allure of soaring prices magnetically draws additional miners, amplifying energy consumption and subsequent environmental impact [8]. This volatile interplay accentuates the profound connections within market dynamics and mining ventures, with shifts in one reverberating through the other.

As we journey further into this digital realm, the specter of cryptojacking unfurls, introducing pronounced cybersecurity threats where malicious actors commandeer computational resources for cryptomining without consent [9], [10]. Such acts exploit system vulnerabilities, raising intricate ethical and legal issues within the cryptocurrency domain [11]. As this landscape continues to evolve, the implementation of robust detection and preventive frameworks is paramount in safeguarding user interests and preserving network sanctity [12], [13].

Efficient Graphics Processing Units (GPUs) are indispensable in cryptocurrency mining, where maximizing profitability dovetails with the imperative of reducing energy consumption. At the core of mining is the brute force of parallel processing, a hallmark of GPU capabilities that become especially potent when handling the expansive datasets and intricate algorithms characteristic of this domain. The ingenuity of GPUs lies in their prowess at executing multiple calculations simultaneously, a trait that has underscored their essential role in driving higher hash rates and enhancing the probability of transaction validation for rewards [14].

Yet, beyond performance metrics, the gargantuan energy appetite of mining hardware presents a formidable environmental challenge. Optimizing GPU performance emerges as a pivotal strategy in curbing this energy voracity. Recent research posits that by channeling advancements in GPU technology towards energy efficiency, significant reductions in operational costs can be achieved, as energy-efficient models amplify computational power while tempering electricity consumption [15], [16]. The delicate calculus between performance and efficiency is critical; mining operations mired in inefficiencies risk eroding the very profitability they aspire to bolster through exorbitant electricity expenses [17].

The stakes are further compounded by the intricate design and architecture of GPUs, which underpin their utility in mining. Optimal GPU performance hinges on the meticulous management of memory resources and the strategic allocation of tasks to minimize latency and drive throughput [18]. This domain necessitates a nuanced comprehension of hardware-software synergies and the deployment of tailor-made algorithms that harness the architectural advantages of GPUs in cryptocurrency mining [19]. Specific algorithms can be precisely engineered to exploit these capabilities, thereby accelerating computational processes and enhancing efficiency [20].

Concomitant with hardware innovation, software enhancements hold equal weight in the quest for efficiency. Methods such as overclocking and undervolting offer pathways to optimize performance while balancing energy costs, suggesting that strategic tweaks can yield substantial improvements [1]. The profound implications of these optimizations extend into realms of profitability and sustainability, addressing increasingly vocal environmental concerns tethered to cryptocurrency mining operations [16].

The objective of this study transcends mere examination—it ventures into the analytical realm by employing K-Means clustering to unravel GPU efficiency through performance metrics across varied mining algorithms. This endeavor is pivotal for delineating optimal configurations and strategies that could amplify the performance of GPUs in the inherently resource-intensive milieu of cryptocurrency mining.

K-Means clustering, lauded for its adeptness at organizing data into distinct groups based on feature likeness, serves as the methodological backbone of this analysis. Within the context of GPU efficiency, the deployment of K-Means offers an organized lens through which to examine diverse performance metrics, such as hash rates and energy consumption, alongside processing times linked to disparate mining algorithms research [21], [22]. Through methodical clustering, this study seeks to extract patterns that illuminate which configurations parade the optimal performance, subsequently serving as a strategic guide for miners aiming to fine-tune their operations.

The criticality of GPU efficiency in the mining landscape is accentuated by its dual impact on profitability and energy usage, the latter a pressing issue given the ecological footprint of cryptocurrency mining [16]. This study leans on a robust corpus of literature surrounding GPU performance and clustering methodologies, aspiring to construct a cogent framework for evaluating GPU efficiency in mining scenarios. Prior inquiries have showcased that GPU-accelerated K-Means implementations notably outperform traditional CPU-driven approaches concerning computational speed and overall efficiency [23], [24].

Furthermore, the study delves into the ramifications that various mining algorithms impose on GPU efficacy. The interplay between algorithmic demands and GPU architectural nuances can lead to divergent efficiency profiles [25], [26]. By leveraging K-Means clustering, this research methodically dissects these discrepancies, offering insights to miners on selecting algorithms that best resonate with their hardware configurations.

Equally compelling is the exploration of parallelization techniques and their bearing on GPU efficiency. Studies contend that optimizing K-Means clustering through parallel applications yields considerable performance enhancements, thereby asserting its relevance as a technique for scrutinizing GPU efficiency within the mining domain [27]. The insights gleaned from this investigation might inform the crafting of best practice protocols for miners, aiming to bolster operational efficiency and mitigate energy expenditures.

In essence, the study is steadfast in its pursuit to analyze GPU efficiency via K-Means clustering across multiple mining algorithms. It endeavours to yield profound insights into GPU configuration optimization and the discernment of suitable mining algorithms, charting a course toward more sustainable and lucratively successful mining ventures.

Literature Review

Existing Studies on Mining Efficiency

Previous research has delved deeply into GPU performance metrics within the realm of cryptocurrency mining, revealing a spectrum of insights yet stopping short of embracing clustering-based methodologies for efficiency analysis. This oversight delineates a fertile ground for innovation, urging a reevaluation of how GPU performance can be optimized across various mining algorithms.

Jiang et al. stand as a cornerstone in this scholarly domain, mapping the substantial energy consumption tethered to Bitcoin mining and the attendant sustainability dilemmas [15]. Despite the pivotal insights into energy demand dynamics, their inquiry does not traverse the terrain of GPU performance efficiency via clustering techniques. In parallel, Zadé et al. navigate the power demands intrinsic to blockchains, focusing acutely on mining machine efficiency through historical lens—yet they omit the clustering frameworks that might unearth nuanced performance subtleties hidden within mining algorithms [17].

A different perspective emerges in the work of Sapra and Shaikh, who spotlight governmental regulation as a catalyst for energy-efficient mining practices [28]. Their discourse on energy stewardship subtly weaves the narrative of mining's ecological impact without harnessing clustering methods that hold potential for scrutinizing GPU efficiency. Aligning in thematic concerns, Zheng et al. elucidate the synergy between cryptocurrency transactions and electricity usage, tracing environmental ramifications [29]. Yet again, the absence of clustering-based explorations leaves a gap in translating these insights into actionable GPU performance strategies.

In contrast, Shuaib et al. chart a path centered on optimizing GPU mining through techniques like overclocking and undervolting, probing avenues to bolster profitability and energy savings [1]. While commendable, this discourse doesn't employ clustering to systematically appraise the efficiency of varying mining algorithms. Wilson's excavation into GPU pricing vis-à-vis cryptocurrency returns casts light on economic influences but similarly omits a clustering framework that might categorize and elevate performance metrics into actionable tiers [30].

The clear absence of clustering-based efficiency scrutiny is further underscored by Dzyuba et al., who attend to electricity cost management in mining but overlook clustering's capability to unearth optimal GPU configurations [31]. Mashuri's proposal of a decision support system for mining machine selection, while pragmatic, also skirts the potential of clustering techniques to illuminate performance across algorithms [32].

K-Means Clustering in Data Mining

K-Means clustering, a stalwart in the realm of data mining, proficiently partitions datasets into K distinct clusters based on similarity. Its essence lies in repetitively assigning data points to nearby cluster centroids, subsequently recalibrating these centroids to reflect the mean of the assigned points, culminating in a clustering solution that minimizes intra-cluster variance [33]. This elegance and computational grace render K-Means a favored method, especially adept at navigating the vast expanses of large datasets with alacrity.

Quick convergence characterizes its operational landscape, rendering it efficiently powerful when initial centroids are aptly chosen [34].

Yet, it is this facile simplicity that harbors a lurking complexity—the sensitivity to the choice of K vastly influences the caliber of clustering outcomes. Determining the optimal K has led to the evolution of various evaluative techniques such as the silhouette and elbow methods, each scrutinizing facets like compactness and separation of clusters [35], [36]. However, amidst its widespread acclaim, K-Means is not exempt from limitations. Its predilection for spherical clusters and homogeneous variance across clusters can betray it in datasets characterized by irregular distributions or divergent densities [37], [38]. In pursuit of transcending these confines, scholars have devised adaptations of the K-Means algorithm, including alternative distance metrics and ensemble methods that bolster clustering resilience [20].

The integration of advanced machine learning techniques into K-Means epitomizes the evolution of this trusted algorithm. By adopting adaptive feature weighting, K-Means now tailors itself to acknowledge the varying significance of diverse data attributes, a testament to its burgeoning adaptability [39]. Such advancements herald its application across distinct domains, from image processing to bioinformatics, highlighting its versatility in extracting substantive patterns from complex datasets [40], [41].

The K-Means clustering algorithm in data mining seeks to partition datasets into K distinct clusters based on the similarity of data points. Its core objective is embodied in minimizing the sum of squared distances between data points and their corresponding cluster centroids, articulated mathematically as:

$$\sum_{i=1}^{K} \sum_{x \in C_i} \left| |x - \mu_i| \right|^2 \quad (1)$$

In this eloquent formula, K denotes the cluster number, C_i represents the conglomeration of points within cluster i , x signifies an individual data point therein, and μ_i captures the centroid of that cluster. The squared Euclidean distance $\left| |x - \mu_i| \right|^2$ crystallizes the essence of clustering, gauging the cohesion of points within their clusters [42].

Operationally, K-Means embarks on its clustering journey through a bifurcated iterative dance—assignment and update. Initially, it allocates data points to their proximal centroids. This is followed by recalibrating the centroids to reflect the aggregated mean of assigned points, iterating until assignments stabilize or centroids achieve permanence [43]. The swiftness of K-Means, particularly with aplomb in handling voluminous datasets, stands as a testament to its formidable efficiency, albeit being intriguingly susceptible to the initial centroids' positioning, which can precipitate varied clustering transformations [44].

Responses to such challenges have spurred innovations like the K-Means++ initialization strategy, enhancing the selection of initial centroids and fortifying the algorithm against inconsistency [45]. Nonetheless, K-Means' presumption of spherical cluster configuration and homogenous variance remains a caveat, particularly when choosing the number of clusters K is uncertain [46]. To combat these challenges, alternatives like K-Medoids and density-centric clustering variants have emerged, offering flexibility and superior results where K-Means may falter [47].

Method

Exploratory Data Analysis (EDA)

The first step in the methodology involved Exploratory Data Analysis (EDA), a crucial process for understanding the dataset's structure, distribution, and overall quality. This initial phase is vital for identifying patterns, spotting anomalies, and generating hypotheses that could lead to further data exploration or modeling.

To begin, the Pandas library in Python was utilized, a powerful tool for data analysis, which facilitated the loading of the dataset from dataset.csv into a DataFrame. This structured and flexible format allowed for seamless data manipulation and exploration, providing a robust foundation for comprehensive analysis. The DataFrame's tabular format is ideal for handling large datasets, enabling efficient data processing and transformation.

One of the first tasks was to obtain a statistical overview of the dataset, achieved through the .describe() method in Pandas. This method generated a summary of the dataset's distributional properties, providing insights into key measures of central tendency such as mean and median, as well as dispersion metrics like standard deviation, minimum and maximum values, and interquartile ranges. These statistical insights are indispensable for assessing the fundamental characteristics of each GPU performance metric.

Key metrics under consideration included hash rates like AbelHash (Mh/s), Autolykos2 (Mh/s), and zkSNARK (Mproof/s), along with their respective power consumptions: AbelHashPower, Autolykos2Power (Watt), and zkSNARKPower (Watt). These metrics are essential as they directly influence the efficiency and performance evaluation of GPUs, crucial for determining the energy efficiency and computational power of the graphics processing units under examination. Understanding these metrics helps in comparing different GPU models and configurations for tasks such as cryptocurrency mining or computational tasks requiring high performance.

To ensure data integrity, missing value assessments were conducted through the .isnull().sum() function. This function helped identify any potential gaps in the data, ensuring that no entries were missing values, which could affect the accuracy of the analysis. Confirming that no imputation was required was a critical verification step, assuring those subsequent data analyses were based on complete datasets. This enhances the reliability and validity of the analysis, providing confidence in the findings and conclusions drawn from the data.

Furthermore, EDA included visualizations to aid in understanding the data. Histograms and box plots were employed to visualize the distributions of key metrics, providing a clear graphical representation of data spread and outliers. Scatter plots were also used to investigate relationships between different metrics, such as hash rates and their corresponding power consumptions, offering insights into performance efficiency trends. These visual tools are invaluable for making data-driven decisions and guiding further analysis steps.

Overall, this comprehensive EDA process laid a solid groundwork for more advanced analysis and modeling, ensuring that all subsequent steps are informed by a thorough understanding of the initial dataset characteristics.

Data Preprocessing

Preprocessing the data was imperative to standardize the variables for clustering, a critical step in data analysis to ensure consistent scaling across different metrics. Initially, specific features were meticulously curated for analysis, encompassing both performance outcomes and associated energy costs. These selected features were: AbelHash (Mh/s), Autolykos2 (Mh/s), zkSNARK (Mproof/s), along with their relevant power consumption metrics. The choice of these features was driven by their relevance in determining the efficiency and computational capacity of GPUs, which are pivotal in various high-performance tasks such as cryptocurrency mining and scientific computations.

To prepare the data for the clustering process, normalization was employed using StandardScaler from the sklearn.preprocessing module. This technique transformed each feature to a standard format with a mean of zero and a standard deviation of one. This transformation process is not merely a formality but a critical step in ensuring that the clustering algorithm treats each feature equally, preventing features with larger scales from disproportionately affecting the clustering process. This disproportionate influence could skew the results and lead to erroneous conclusions about GPU efficiency, which could, in turn, impact decision-making processes related to hardware investments and operational optimizations.

The normalization process involves computing the mean and standard deviation of each feature and using these statistics to scale the data. By doing so, every data point is transformed relative to the entire dataset's mean and variation, which enhances the robustness and reliability of the clustering analysis. This step is especially imperative when dealing with datasets that combine performance metrics and power consumption measures, as these can vary significantly in their scales and units of measurement.

Overall, the preprocessing step forms the backbone of any meaningful data analysis, ensuring that the insights drawn from clustering are both accurate and actionable. It helps in aligning the data with the assumptions of many machine learning models, including clustering algorithms, thereby improving the interpretability and validity of the classification results.

Clustering and Visualization

The heart of the clustering process was the application of the K-Means algorithm. Before determining the K value (number of clusters), the elbow method was employed to identify the optimal cluster count. This involved plotting the within-cluster sum of squares (WCSS) for cluster numbers ranging from 1 to 10. The WCSS metric, indicative of clusters' compactness, was plotted using Matplotlib to visualize the 'elbow point'—a sharp transition where adding more clusters yields diminishing returns in variance reduction. For this dataset, the elbow method suggested that an optimal K value was 3, balancing comprehensiveness and parsimony.

Subsequently, the K-Means clustering algorithm was executed with these parameters (`n_clusters=3`, `random_state=42`) to ensure reproducibility. This step assigned each GPU configuration to one of three clusters based on similarity across the standardized features. Results were appended to the DataFrame with a 'Cluster' label, enabling enhanced interpretability of the data.

The visualization of clustering outcomes was conducted using Seaborn, lending vibrant graphical clarity to the data. A scatter plot was crafted to depict the relation between GPU hash rates and power consumption, annotated by cluster, using distinct hues to encode membership (e.g., `AbelHash (Mh/s)` vs. `AbelHashPower`). This visual representation not only highlighted the concentration of efficient configurations within certain clusters but also delineated outliers that deviate from cluster norms. Additionally, a comprehensive pairplot provided a multidimensional overview of feature interrelationships, integrating all selected metrics. This intricate depiction aided in recognizing emergent patterns—revealing clusters characterized by high computational output with reduced energy input as potentially optimal configurations.

Cluster centroids, extracted and denormalized, were also illustrated, symbolizing the 'average' performance within each group. These centroids provided tangible benchmarks for comparison among the groups.

Through this meticulous methodical approach—merging EDA, data preprocessing, and clustering—the study surfaced nuanced patterns of GPU efficiency. The clarity and precision from the clustering insights equip decision-makers with data-driven strategies to optimize hardware selections and operational parameters, potentially reducing energy consumption and elevating mining profits. These findings lay the groundwork for sustainable best practices in energy use and configuration strategies in cryptocurrency mining operations.

Result and Discussion

Findings of K-Means Clustering

The application of K-Means clustering on the GPU performance dataset has revealed distinct patterns that uncover varying categories of GPU efficiency, each characterized by unique performance metrics. Upon analysis, three primary clusters emerged, reflective of differing computational capabilities and energy consumption profiles, as shown in figure 1.

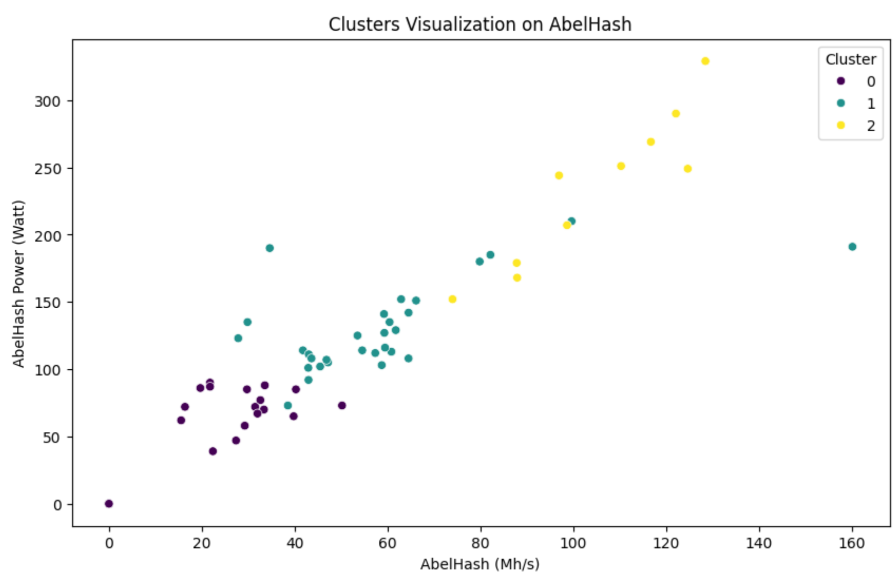

Figure 1 Clusters Visualization

Cluster 0, identified as the "Low Efficiency Group," contains GPUs that demonstrate relatively low hash rates across key metrics, such as 11.30 Mh/s for `AbelHash` and 57.07 Mh/s for `Autolykos2`. These GPUs also record modest power consumption figures, with 27.80 Watts for `AbelHashPower` and 64.52 Watts for `Autolykos2Power`. The interpretation of this cluster suggests that these are likely older GPU models or systems not optimized for high-performance output, offering less computational power relative to their energy input. This group's existence highlights potential inefficiencies that could be addressed through technology updates or operational adjustments.

Cluster 1, described as the "Moderate Efficiency Group," showcases an intermediate level of computational power and energy balance. This group features GPUs with modest hash rates, including 51.48 Mh/s for `AbelHash` and 131.70 Mh/s for `Autolykos2`, coupled with respective power consumption rates of 114.56 Watts and 122.82 Watts. They signify a balanced trade-off between computational output and energy use, making them viable for situations where budgetary constraints or energy considerations are priorities. This balance between performance and cost-effectiveness may appeal to operations seeking prudent investments in technology that offer reliable yet moderate returns.

The "High Efficiency Group," represented by Cluster 2, signifies top-tier GPU configurations optimized for maximum output with high hash rates, recording 104.79 Mh/s for `AbelHash` and 218.35 Mh/s for `Autolykos2`. Correspondingly, these GPUs show higher power consumption values of 233.80 Watts and 184.40 Watts, respectively, yet their superior performance represents a compelling case for high-demand mining operations where the energy costs are justified by the substantial computational returns. This group highlights the pinnacle of current GPU efficiency in the dataset, providing insights into potential hardware benchmarks for advanced mining frameworks.

Visualization of these results was achieved using scatter plots and pair plots, providing a nuanced view of relationships between hash rates and power consumption across clusters. The scatter plot, specifically analyzing `AbelHash (Mh/s)` against `AbelHashPower`, color-coded by cluster, clearly delineated each group’s position within the efficiency spectrum. This visualization illuminated the spectrum of performance categories, further corroborated by the pair plots showing broader metric interactions, thus delivering an enhanced interpretative layer to the raw data.

The implications of these findings stretch far into operational and strategic domains within cryptocurrency mining. Operators looking to optimize resource allocation and economic return would do well to invest in high-performance GPUs identified in Cluster 2, aligning their hardware choices with operational demands for high throughput and energy efficiency. Conversely, GPUs situated in the lower efficiency clusters suggest avenues for enhancement, whether through system upgrades or efficiency-focused strategies such as better cooling solutions or software optimizations that could elevate their performance to better align with industry needs.

Moreover, the silhouette score of 0.35, reflecting moderate cluster separation, suggests that while distinct efficiency categories exist, some overlaps point to potential areas for further investigation. This score encourages a nuanced review of clustering boundaries to refine understanding and identify subtle performance differentiators within the clusters. Altogether, these analytical insights inform strategic decision-making regarding technology deployment and energy management in mining operations, ultimately guiding firms towards more sustainable and profitable practices.

Discussion

The findings from the clustering analysis of GPU performance metrics have significant implications for cryptocurrency mining operations, hardware procurement strategies, and energy efficiency optimization. By categorizing GPUs into distinct clusters based on their computational output and power consumption, this study provides actionable insights that can guide decision-making processes in the mining industry. Below, we discuss the key implications of these findings in detail.

The identification of high-efficiency GPUs (Cluster 2) offers a clear pathway for mining operators to prioritize hardware investments. These GPUs, characterized by high hash rates (e.g., 104.79 Mh/s for `AbelHash` and 218.35 Mh/s for `Autolykos2`) and relatively high-power consumption (233.80 Watts and 184.40 Watts, respectively), are ideal for operations requiring maximum computational throughput. By focusing on GPUs within this cluster, mining operators can maximize their return on investment (ROI) by achieving higher mining yields while maintaining energy costs at manageable levels. Conversely, GPUs in the low-efficiency cluster (Cluster 0) should be phased out or repurposed for less demanding tasks, as their lower hash rates and modest power consumption make them less suitable for high-performance mining.

The clustering results highlight the importance of balancing computational output with energy consumption. GPUs in the moderate-efficiency cluster (Cluster 1) represent a middle ground, offering a balance between performance and energy use. For mining operations with limited energy budgets or those operating in regions with high electricity costs, these GPUs may provide a cost-effective solution. Additionally, the findings suggest that energy efficiency can be further improved by optimizing the configurations of GPUs in the low-efficiency cluster. For example, implementing advanced cooling systems, undervolting, or using energy-efficient mining algorithms could help reduce power consumption without significantly compromising performance.

The insights derived from the clustering analysis can directly contribute to reducing operational costs in mining operations. By identifying and deploying GPUs from the high-efficiency cluster, operators can minimize the number of GPUs required to achieve a given level of computational output, thereby reducing both hardware acquisition and maintenance costs. Furthermore, the ability to categorize GPUs based on their efficiency allows operators to allocate resources more effectively, ensuring that high-performance GPUs are utilized for critical tasks while lower-performance GPUs are reserved for less intensive operations.

The findings also have important implications for sustainability in cryptocurrency mining. The high energy consumption associated with mining operations has raised concerns about their environmental impact. By prioritizing GPUs in the high-efficiency cluster, mining operations can reduce their overall energy consumption per unit of computational output, thereby lowering their carbon footprint. Additionally, the insights gained from this study can inform the development of more energy-efficient mining algorithms and hardware designs, contributing to the broader goal of sustainable mining practices.

The clustering results provide a foundation for future research aimed at further optimizing GPU performance and energy efficiency. For instance, the moderate silhouette score of 0.35 indicates that while the clusters are distinct, there is room for refinement in the clustering process. Future studies could explore alternative clustering algorithms, such as DBSCAN or hierarchical clustering, to achieve more precise groupings. Additionally, the integration of additional features, such as thermal performance or cost per unit, could enhance the clustering model's ability to identify optimal GPU configurations.

The findings from this study can serve as a benchmark for the cryptocurrency mining industry, providing a standardized framework for evaluating GPU performance and efficiency. Mining operators can use these insights to establish best practices for hardware selection, energy management, and operational optimization. By adopting a data-driven approach to decision-making, the industry can move towards more efficient and sustainable mining practices, ultimately benefiting both operators and the environment.

Conclusion

The clustering analysis of GPU performance metrics has yielded valuable insights into the efficiency and operational characteristics of GPUs used in cryptocurrency mining. By categorizing GPUs into three distinct clusters—low-efficiency, moderate-efficiency, and high-efficiency—we have identified clear patterns in computational output and energy consumption. These findings provide a robust framework for optimizing mining operations, enabling operators to make data-driven decisions that enhance performance, reduce costs, and promote sustainability.

The high-efficiency cluster (Cluster 2) stands out as the most promising group, with GPUs delivering exceptional hash rates (e.g., 104.79 Mh/s for AbelHash and 218.35 Mh/s for Autolykos2) despite their higher power consumption. These GPUs are ideal for high-performance mining operations, offering the best return on investment and aligning with the industry's need for maximizing computational throughput. Conversely, the low-efficiency cluster (Cluster 0) highlights the limitations of older or less optimized hardware, suggesting opportunities for upgrades or repurposing to improve overall efficiency.

The moderate-efficiency cluster (Cluster 1) represents a balanced option for operations with constrained energy budgets or those seeking cost-effective solutions. These GPUs provide a viable middle ground, offering reasonable computational output without excessive energy consumption. This cluster underscores the importance of tailoring hardware choices to specific operational needs and resource availability.

The visualization of clustering outcomes, through scatter plots and pair plots, has further enhanced the interpretability of the results, enabling operators to clearly distinguish between efficiency categories and identify outliers or anomalies. The silhouette score of 0.35, while indicating moderate cluster separation, also points to areas for further refinement and exploration, such as the integration of additional performance metrics or the application of alternative clustering algorithms.

From a broader perspective, these findings have significant implications for the cryptocurrency mining industry. By prioritizing high-efficiency GPUs, operators can reduce operational costs, minimize energy consumption, and lower their environmental impact. The insights gained from this study also pave the way for future research, encouraging the development of more energy-efficient hardware and mining algorithms. Furthermore, the clustering framework established here can serve as a benchmark for industry best practices, fostering a more sustainable and data-driven approach to mining operations.

The findings from this study open several promising avenues for future research that could deepen our understanding of GPU efficiency and its implications for cryptocurrency mining. One key area for exploration is the integration of additional performance metrics, such as thermal performance, cost per unit, and hardware lifespan, into the clustering model. This would provide a more comprehensive evaluation of GPU efficiency, accounting for factors beyond computational output and power consumption. Additionally, future studies could investigate alternative clustering algorithms, such as DBSCAN, hierarchical clustering, or Gaussian Mixture Models (GMM), which may offer more precise or nuanced groupings, particularly in datasets with overlapping clusters or varying densities. Another important direction is the development of dynamic clustering approaches that adapt to changes in GPU performance over time, such as hardware degradation or software updates, enabling real-time insights into efficiency trends. Research into energy-efficient mining algorithms, including proof-of-stake (PoS) or other low-energy consensus mechanisms, could also significantly enhance the sustainability of mining operations. Furthermore, the impact of environmental factors, such as temperature, humidity, and cooling systems, on GPU performance warrants further investigation, as these variables can influence efficiency and should be incorporated into clustering models for more accurate recommendations. A detailed analysis of the economic and environmental impact of adopting high-efficiency GPUs could provide valuable insights for policymakers and industry stakeholders, including cost-benefit analyses, carbon footprint assessments, and lifecycle evaluations of hardware. Additionally, the application of machine learning for predictive maintenance could help mining operators proactively address hardware failures or performance degradation, reducing downtime and extending the lifespan of GPUs. Finally, expanding the dataset to include GPUs from different manufacturers or platforms would enable cross-platform comparisons of efficiency and performance, offering a broader perspective on optimal hardware choices for mining operations. These future research directions hold the potential to refine the insights from this study and drive the cryptocurrency mining industry toward greater efficiency, sustainability, and innovation.

Declarations

Author Contributions

Conceptualization: J.K., A.O.; Methodology: J.K., A.O.; Software: J.K.; Validation: J.K.; Formal Analysis: J.K.; Investigation: J.K.; Resources: J.K.; Data Curation: A.O.; Writing Original Draft Preparation: J.K.; Writing Review and Editing: J.K.; Visualization: A.O.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Shuaib et al., “A Novel Optimization for GPU Mining Using Overclocking and Undervolting,” Sustainability, vol. 14, no. 14, p. 8708, 2022, doi: 10.3390/su14148708.

[2] M. Khazraee, I. Magaki, L. V. Gutierrez, and M. Taylor, “ASIC Clouds: Specializing the Datacenter,” Ieee Micro, pp. 1–1, 2017, doi: 10.1109/mm.2017.265085658.

[3] Ü. Hacıoğlu, D. Chlyeh, M. K. Yılmaz, E. Tatoğlu, and D. Delen, “Crafting Performance-Based Cryptocurrency Mining Strategies Using a Hybrid Analytics Approach,” Decision Support Systems, vol. 142, p. 113473, 2021, doi: 10.1016/j.dss.2020.113473.

[4] M. Saad, A. Khormali, and A. Mohaisen, “Dine and Dash: Static, Dynamic, and Economic Analysis of in-Browser Cryptojacking,” pp. 1–12, 2019, doi: 10.1109/ecrime47957.2019.9037576.

[5] C. Stoll, L. Klaaßen, and U. Gallersdörfer, “The Carbon Footprint of Bitcoin,” Joule, vol. 3, no. 7, pp. 1647–1661, 2019, doi: 10.1016/j.joule.2019.05.012.

[6] M. Polemis and E. G. Tsionas, “The Environmental Consequences of Blockchain Technology: A Bayesian Quantile Cointegration Analysis for Bitcoin,” International Journal of Finance & Economics, vol. 28, no. 2, pp. 1602–1621, 2021, doi: 10.1002/ijfe.2496.

[7] A. R. Kommuru, A. Gupta, and M. Chauhan, “Cryptocurrency: The Present and the Future Scenario,” International Journal for Research in Applied Science and Engineering Technology, vol. 10, no. 6, pp. 675–686, 2022, doi: 10.22214/ijraset.2022.43812.

[8] Ö. Haykır and İ. Yağlı, “Speculative Bubbles and Herding in Cryptocurrencies,” Financial Innovation, vol. 8, no. 1, 2022, doi: 10.1186/s40854-022-00383-0.

[9] M. H. K. Abbasi, S. Ullah, T. Ahmad, and A. Buriro, “A Real-Time Hybrid Approach to Combat in-Browser Cryptojacking Malware,” Applied Sciences, vol. 13, no. 4, p. 2039, 2023, doi: 10.3390/app13042039.

[10] S. Varlioglu, E. Nelly, Z. ElSayed, and M. Ozer, “The Dangerous Combo: Fileless Malware and Cryptojacking,” pp. 125–132, 2022, doi: 10.1109/southeastcon48659.2022.9764043.

[11] U. Widyatmoko, “Law Enforcement Against Cryptocurrency Abuse,” Journal of Social Research, vol. 3, no. 2, 2024, doi: 10.55324/josr.v3i2.1941.

[12] Y. Boshmaf, C. Elvitigala, H. A. Jawaheri, P. Wijesekera, and M. A. Sabah, “Investigating MMM Ponzi Scheme on Bitcoin,” pp. 519–530, 2020, doi: 10.1145/3320269.3384719.

[13] A. Zimba, C. Ngongola-Reinke, M. Chishimba, and T. F. Mbale, “Demystifying Cryptocurrency Mining Attacks: A Semi-Supervised Learning Approach Based on Digital Forensics and Dynamic Network Characteristics,” Zambia Ict Journal, vol. 5, no. 1, pp. 1–7, 2021, doi: 10.33260/zictjournal.v5i1.108.

[14] K. Iliakis, K. Koliogeorgi, A. Λίτκε, T. Varvarigou, and D. Soudris, “GPU Accelerated Blockchain Over Key-value Database Transactions,” Iet Blockchain, vol. 2, no. 1, pp. 1–12, 2022, doi: 10.1049/blc2.12011.

[15] S. Jiang et al., “Policy Assessments for the Carbon Emission Flows and Sustainability of Bitcoin Blockchain Operation in China,” Nature Communications, vol. 12, no. 1, 2021, doi: 10.1038/s41467-021-22256-3.

[16] M. Rukhiran, “Sustainable Optimizing Performance and Energy Efficiency in Proof of Work Blockchain: A Multilinear Regression Approach,” Sustainability, vol. 16, no. 4, p. 1519, 2024, doi: 10.3390/su16041519.

[17] M. Zadé, J. Myklebost, P. Tzscheutschler, and U. Wagner, “Is Bitcoin the Only Problem? A

Scenario Model for the Power Demand of Blockchains," Frontiers in Energy Research, vol. 7, 2019, doi: 10.3389/fenrg.2019.00021.
[18] X. Chen and Arvind, "Efficient and Scalable Graph Pattern Mining on GPUs," 2021, doi: 10.48550/arxiv.2112.09761.
[19] Y. Gao, B. Wu, J. Luo, and H. Qiu, "GPU-based Arbitrary Polygon Intersection Area Algorithm," Destech Transactions on Engineering and Technology Research, no. ismii, 2017, doi: 10.12783/dtetr/ismii2017/16652.
[20] A. Cano, "A Survey on Graphic Processing Unit Computing for Large-scale Data Mining," Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery, vol. 8, no. 1, 2017, doi: 10.1002/widm.1232.
[21] G. He, S. Vialle, and M. Baboulin, "Parallelization of the K-Means Algorithm in a Spectral Clustering Chain on CPU-GPU Platforms," pp. 135–147, 2021, doi: 10.1007/978-3-030-71593-9_11.
[22] J. Bhimani, M. Leeser, and N. Mi, "Accelerating K-Means Clustering With Parallel Implementations and GPU Computing," 2015, doi: 10.1109/hpec.2015.7322467.
[23] S. Cuomo, V. D. Angelis, G. Farina, L. Marcellino, and G. Toraldo, "A GPU-accelerated Parallel K-Means Algorithm," Computers & Electrical Engineering, vol. 75, pp. 262–274, 2019, doi: 10.1016/j.compeleceng.2017.12.002.
[24] A. Prahara, D. P. Ismi, A. I. Kistijantoro, and M. L. Khodra, "Parallelized K-Means Clustering by Exploiting Instruction Level Parallelism at Low Occupancy," pp. 30–34, 2017, doi: 10.1109/icitisee.2017.8285516.
[25] A. Syberfeldt and T. Ekblom, "A Comparative Evaluation of the GPU vs the CPU for Parallelization of Evolutionary Algorithms Through Multiple Independent Runs," International Journal of Computer Science and Information Technology, vol. 9, no. 3, pp. 01–14, 2017, doi: 10.5121/ijcsit.2017.9301.
[26] M. Li, E. Frank, and B. Pfahringer, "Large Scale K-Means Clustering Using GPUs," Data Mining and Knowledge Discovery, vol. 37, no. 1, pp. 67–109, 2022, doi: 10.1007/s10618-022-00869-6.
[27] G. He, S. Vialle, and M. Baboulin, "Parallel and Accurate k-means Algorithm on CPU-GPU Architectures for Spectral Clustering," Concurrency and Computation Practice and Experience, vol. 34, no. 14, 2021, doi: 10.1002/cpe.6621.
[28] N. Sapra and I. Shaikh, "Impact of Bitcoin Mining and Crypto Market Determinants on Bitcoin-Based Energy Consumption," Managerial Finance, vol. 49, no. 11, pp. 1828–1846, 2023, doi: 10.1108/mf-03-2023-0179.
[29] M. Zheng, F. Guo, X. Zhang, and C. Chang, "The Transaction Behavior of Cryptocurrency and Electricity Consumption," Financial Innovation, vol. 9, no. 1, 2023, doi: 10.1186/s40854-023-00449-7.
[30] L. Wilson, "Gpu Prices and Cryptocurrency Returns," Applied Finance Letters, vol. 11, pp. 2–8, 2022, doi: 10.24135/afl.v11i.503.
[31] A. P. Dzyuba, I. Solovyeva, and D. Konopelko, "Managing Electricity Costs in Industrial Mining and Cryptocurrency Data Centers," International Journal of Energy Economics and Policy, vol. 13, no. 4, pp. 76–90, 2023, doi: 10.32479/ijeep.13955.
[32] C. Mashuri, "Decision Support System for Selecting the Best Cryptocurrency Mining Machine Using the Multifactor Evaluation Process Method," Sana, vol. 1, no. 1, pp. 30–36, 2023, doi: 10.58905/sana.v1i1.152.
[33] M. Ahmed, R. Seraj, and S. Islam, "The K-Means Algorithm: A Comprehensive Survey and Performance Evaluation," Electronics, vol. 9, no. 8, p. 1295, 2020, doi: 10.3390/electronics9081295.
[34] B. Rajitha and S. Agarwal, "Segmentation of Bone Pixels From EROI Image Using Clustering Method for Bone Age Assessment," vol. 9789, p. 97890H, 2016, doi: 10.1117/12.2225146.
[35] L. Luo, H. L. Shen, S. J. Shao, and J. H. Xin, "A Novel Method for Weft and Warp Yarn Segmentation in Multicolour Yarn-dyed Fabric Images," Coloration Technology, vol. 131, no. 2, pp. 165–171, 2015, doi: 10.1111/cote.12130.
[36] C. Y. Patil and I. Baidari, "Estimating the Optimal Number of Clusters K in a Dataset Using Data Depth," Data Science and Engineering, vol. 4, no. 2, pp. 132–140, 2019, doi: 10.1007/s41019-019-0091-y.
[37] D. W. Chen and Y. Jin, "An Active Learning Algorithm Based on Shannon Entropy for Constraint-Based Clustering," Ieee Access, vol. 8, pp. 171447–171456, 2020, doi: 10.1109/access.2020.3025036.
[38] Y. Resti, F. Nasution, I. Yani, A. S. Mohruni, and F. A. Almahdini, "A Cans Waste Classification System Based on RGB Images Using Different Distances of K-Means Clustering," Journal of Engineering and Scientific Research, vol. 2, no. 1, 2020, doi:

10.23960/jesr.v2i1.35.
[39] Z. Wu and Z. Wu, "An Enhanced Regularized K-Means Type Clustering Algorithm With Adaptive Weights," Ieee Access, vol. 8, pp. 31171–31179, 2020, doi: 10.1109/access.2020.2972333.
[40] J. Bennett, M. Pomaznoy, A. Singhania, and B. Peters, "A Metric for Evaluating Biological Information in Gene Sets and Its Application to Identify Co-Expressed Gene Clusters in PBMC," Plos Computational Biology, vol. 17, no. 10, p. e1009459, 2021, doi: 10.1371/journal.pcbi.1009459.
[41] B. Pandey, D. Anand, and D. K. Pandey, "Facioscapulohumeral Muscular Dystrophy Diagnosis Using Hierarchical Clustering Algorithm and K-Nearest Neighbor Based Methodology," International Journal of E-Health and Medical Communications, vol. 8, no. 2, pp. 33–46, 2017, doi: 10.4018/ijehmc.2017040103.
[42] M. Annas and S. N. Wahab, "Data Mining Methods: K-Means Clustering Algorithms," International Journal of Cyber and It Service Management, vol. 3, no. 1, pp. 40–47, 2023, doi: 10.34306/ijcitsm.v3i1.122.
[43] "Efficient K-Means Clustering Algorithm in Web Log Mining," Revista Tecnica De La Facultad De Ingenieria Universidad Del Zulia, 2016, doi: 10.21311/001.39.6.37.
[44] Z. D. Montero, "Customer Grouping for Customer Relationship Management Optimization With the K-Means Algorithm," Journal of Computer Science and Information Technology, pp. 98–105, 2022, doi: 10.35134/jcsitech.v8i4.46.
[45] H. Li, H. He, and Y. Wen, "Dynamic Particle Swarm Optimization and K-Means Clustering Algorithm for Image Segmentation," Optik, vol. 126, no. 24, pp. 4817–4822, 2015, doi: 10.1016/j.ijleo.2015.09.127.
[46] A. H. Nasyuha, Zulham, and I. Rusydi, "Implementation of K-Means Algorithm in Data Analysis," Telkomnika (Telecommunication Computing Electronics and Control), vol. 20, no. 2, p. 307, 2022, doi: 10.12928/telkomnika.v20i2.21986.
[47] E. L. Cahapin, "Clustering of Students Admission Data Using K-Means, Hierarchical, and DBSCAN Algorithms," Bulletin of Electrical Engineering and Informatics, vol. 12, no. 6, pp. 3647–3656, 2023, doi: 10.11591/eei.v12i6.4849.