User Profiling Based on Financial Transaction Patterns: A Clustering Approach for User Segmentation

Satrya Fajri Pratama[1,*], Nadya Awali Putri[2,]

[1]*Department of Computer Science, School of Physics, Engineering and Computer Science, University of Hertfordshire, Hatfield AL10 9AB, United Kingdom*

[2]*Magister of Computer Science, Universitas Amikom Purwokerto, Jawa Tengah, Indonesia*

(Received: June 1, 2024; Revised: July 10, 2024; Accepted: September 15, 2024; Available online: December 1, 2024)

Abstract

User profiling based on financial transaction patterns is crucial for improving customer segmentation and personalizing financial services. This study uses clustering techniques, specifically K-means, to analyze transaction data and segment users based on transaction amounts, times, and types. Three clusters were identified, each demonstrating distinct transaction behaviors: Cluster 0, primarily focused on purchases and occurring early in the week; Cluster 1, which emphasizes transfers and higher transaction amounts, typically occurring mid-week; and Cluster 2, similar to Cluster 0 but with a preference for later-week transactions. The analysis demonstrates that transaction patterns, including amount, time, and type, provide valuable insights for targeting specific user groups with personalized marketing strategies and financial products. The study also highlights the importance of improving clustering accuracy, as indicated by the moderate Silhouette Score of 0.33, suggesting that further refinement in the clustering methodology could lead to more distinct user segments. The findings of this study emphasize the potential for clustering techniques to enhance user profiling, ultimately improving business strategies, customer satisfaction, and loyalty. Limitations of the study, including the dataset's single-month duration, suggest that further research incorporating larger and more diverse datasets, as well as alternative clustering techniques, could offer deeper insights into user behavior and refine segmentation strategies.

Keywords: Profiling, Financial Transactions, Clustering Techniques, K-Means, Customer Segmentation

1. Introduction

In the digital financial landscape, analyzing transaction data is crucial for optimizing services and detecting fraud. As financial technology evolves rapidly, identifying risks associated with digital transactions becomes essential to mitigate fraud and enhance stability [1]. Integrating data from diverse online platforms can complicate financial reporting for Micro, Small, and Medium Enterprises (MSMEs), highlighting the need for effective data integration strategies. Machine learning is increasingly deployed to detect financial crimes and improve cybersecurity, ensuring consumer data protection. This comprehensive approach, encompassing data governance, fraud detection, and cybersecurity, is essential for financial institutions navigating the complexities of digital transactions.

User profiling is key to enhancing customer experience and personalizing financial services. By leveraging detailed user data, institutions can tailor their products to meet specific client needs, fostering loyalty and financial well-being [2]. Effective user profiling also supports personalized recommendation systems using AI and machine learning, which suggest financial products based on individual behavior [3]. Furthermore, institutions that understand their customers' diverse needs can promote financial inclusion, particularly for marginalized groups such as individuals with disabilities [4]. This leads to improved customer satisfaction and trust, while also contributing to broader economic goals of accessibility and participation in the financial ecosystem [5].

Transaction-based user segmentation is essential for better understanding user behavior. By analyzing transaction patterns using clustering methods like K-Means, financial institutions can segment customers effectively, enabling

***Corresponding author: Satrya Fajri Pratama (sf.pratama@herts.ac.uk)**
DOI: https://doi.org/10.47738/ijaim.v4i4.92

personalized services and targeted marketing [6]. Traditional fraud detection methods often focus on broad, common patterns, limiting their ability to detect unique behavioral traits crucial for identifying sophisticated fraud schemes. Transaction data from diverse platforms, such as e-commerce and cryptocurrency exchanges, shows that users possess varied interaction patterns influenced by factors such as subjective norms and transaction conditions [7]. These complexities make it difficult to model individual behavior effectively, necessitating advanced techniques for managing the high dimensionality and heterogeneity of transaction datasets.

Traditional user segmentation methods often rely heavily on historical transaction data, which may not fully capture evolving user behaviors. This leads to inaccuracies in predicting future behaviors, especially when users have insufficient historical data [8]. Moreover, segmentation practices frequently overlook diverse investor profiles and their distinct impacts on market dynamics, limiting the effectiveness of market prediction techniques [9]. As transaction data grows more complex, advanced methodologies are required to account for heterogeneous user behavior [10]. Failing to integrate real-time data and user-specific variables results in missed opportunities and overlooked patterns, limiting the insights derived from the data [9].

The need for machine learning techniques, such as clustering, arises from the increasing complexity and diversity of transaction data. Traditional methods often miss the nuances inherent in transaction patterns, leading to incomplete insights. Clustering techniques, like K-means, enable financial institutions to segment users based on behaviors, rather than relying on predefined categories. This allows the discovery of hidden patterns within vast datasets [11]. These methods help institutions better understand customer needs and preferences, leading to more targeted marketing strategies and personalized financial products [12]. Moreover, machine learning clustering can identify segments that traditional methods may overlook, providing a deeper understanding of user attitudes and predictive behaviors [12]. Integrating machine learning into user segmentation enhances analytical capabilities, supporting strategic decision-making in increasingly dynamic financial environments.

User profiling based on financial transaction patterns requires advanced techniques to capture the diversity of user behaviors. Traditional profiling methods often lack the granularity to differentiate user actions across different contexts [13]. Machine learning algorithms, especially clustering, allow for the identification of subtle patterns that improve classification accuracy. Predictive analytics can also create dynamic user profiles that evolve as behavior changes, enhancing the accuracy of financial forecasts and recommendations. Techniques like ensemble learning and Gaussian mixture models can refine segmentation by accounting for diverse transaction characteristics, further enhancing the personalization of financial services [14], [15].

Clustering techniques, including K-means, hierarchical clustering, and spectral clustering, are essential for segmenting users based on transaction behavior. These methods provide a more precise grouping of users based on their transactional characteristics and behaviors [16]. By identifying unique segments, financial institutions can design targeted marketing campaigns and offer personalized financial products [17]. The ability to analyze transaction data in real time provides institutions with insights into current user needs, enabling better engagement and service delivery [18]. Additionally, combining big data analytics with clustering algorithms strengthens segmentation, helping institutions adapt to the dynamic financial landscape and respond effectively to user demands [19]. This integration fosters customer satisfaction and loyalty by aligning services with actual user behaviors.

Identifying key transaction features such as transaction amount, time, and transaction type plays a crucial role in refining user segmentation. The amount of a transaction is a key indicator of user behavior, influencing both spending habits and fraud detection efforts. Clustering techniques can categorize users based on transaction amounts, helping institutions identify distinct behavior patterns [20]. Time features, like transactional timestamps, reveal patterns related to the time of day and transaction frequency, enabling tailored services [21]. Transaction types purchases, transfers, or withdrawals further contribute to understanding user preferences and enhancing segmentation. By integrating these features into clustering models, institutions can improve user classification accuracy and develop more effective marketing strategies.

This study emphasizes the importance of analyzing transaction features for effective user segmentation, enabling financial institutions to offer personalized services. By leveraging transaction characteristics such as amount, time, and type, institutions can better understand user behavior and tailor products to meet customer needs. Enhanced

segmentation leads to customized marketing strategies and improved service delivery based on each user segment's unique requirements, maximizing relevance and satisfaction. Advanced clustering techniques also provide valuable insights that inform product development and service improvements, addressing the distinct challenges of each segment. A deeper understanding of transaction behaviors facilitates the creation of targeted financial products, promoting financial inclusion and strengthening customer loyalty, which is essential for growth in a competitive marketplace.

Implementing user segmentation based on transaction behaviors leads to better customer retention and engagement. Personalized strategies enhance satisfaction and foster customer loyalty, as users feel recognized and valued. Financial institutions can optimize customer service by anticipating needs and providing proactive solutions, improving overall satisfaction [22]. Tailored communication and services reduce friction during interactions and enhance service efficiency [23]. Personalized financial products boost user engagement and loyalty, driving business growth and promoting financial accessibility [3]. By integrating user segmentation into strategic planning, financial institutions can adapt to dynamic user needs, fostering long-term growth and stability [24].

Ultimately, the application of advanced machine learning techniques for user segmentation in financial transactions offers a powerful framework for improving services, personalization, and marketing. By addressing the complexities of user behavior and transaction patterns, institutions can gain deeper insights into their customers, foster engagement, and create tailored solutions that enhance customer satisfaction and drive business success.

2. Literature Review

Clustering techniques are essential for categorizing data into meaningful groups based on similarity, with K-Means and DBSCAN being two of the most widely utilized methods. K-Means clustering aims to partition observations into k distinct clusters based on their mean distances, making it particularly effective in fields like marketing to segment customers and enhance business strategies [25]. This method's robustness is enhanced through iterative refinements of initial centroids.

On the other hand, density-based clustering methods like DBSCAN identify clusters based on the density of data points, which helps in detecting arbitrarily shaped clusters and outliers, enabling more nuanced data analysis. The flexibility of K-Means and density-based methods makes them suitable for diverse applications, ranging from social media analysis to healthcare studies [26], thus demonstrating their importance in extracting insights from complex datasets.

User profiling and segmentation have garnered significant attention in the realm of data analysis, particularly utilizing transaction data to refine marketing and customer engagement strategies. A variety of methodologies, particularly clustering techniques, have been employed for effective user segmentation. Ge et al. demonstrate that user behaviors within online brand communities can be categorized into distinct segments information-oriented, entertainment-oriented, and multi-motivation users highlighting how behavior-oriented segmentation can enhance targeted marketing efforts. In a similar vein, Martinovska et al. [27] emphasize the importance of selecting appropriate clustering algorithms tailored to specific business needs to improve decision-making and user profiling [27].

The analysis of financial transaction data for user segmentation has become prevalent in understanding consumer behavior and optimizing business strategies. Zhao et al. [28] propose an extended regularized K-means approach that focuses on transaction records, emphasizing its significance in online customer segmentation by incorporating transaction amounts, times, and types to identify distinct customer groups [28]. This aligns with the study by Zhang et al. [8] which examines user behaviors through historical transaction records and employs a DBSCAN clustering algorithm to distinguish between normal and fraudulent transactions, thus enhancing the understanding of transaction patterns [8]. The integration of transaction features such as amounts and timings enables more nuanced strategies, which are supported by Bilgiç et al. [29], who demonstrated successful store segmentation using comprehensive transaction data from a global retailer [29]. Collectively, these studies underscore the diverse methodologies and insights achieved through the analysis of financial transaction data, emphasizing its critical role in user segmentation and business intelligence.

Machine learning (ML) has emerged as a transformative force within the financial services sector, enhancing market segmentation and enabling personalized service delivery. As highlighted by Komati, ML architectures facilitate real-time financial decision-making, allowing institutions to process transactional data for improved customer segmentation and engagement strategies through predictive modeling [30]. This capability not only enhances marketing efforts but also aids in maintaining compliance with regulatory requirements.

Chukwukaelo et al. [31] elaborate on the variety of ML techniques available, including predictive analytics, which improve corporate financial planning by enabling personalized services, such as tailored product recommendations and dynamic pricing [31]. Similarly, Olowe et al. provide an overview of ML's impact on customer analytics and risk management, emphasizing its role in transforming data into actionable insights, thereby optimizing market strategies [32]. Furthermore, Feng discusses the integration of advanced ML methodologies, which leverage diverse data sources for enhanced risk assessment. This exemplifies how financial institutions can refine products to better align with the preferences of individual consumers [33]. Ultimately, these applications of machine learning not only enhance operational efficiency but also significantly improve customer experiences within the financial services landscape.

3. Methodology

Figure 1 illustrates the step-by-step process involved in cluster analysis, beginning with data collection, followed by data preprocessing, and then applying the clustering approach.

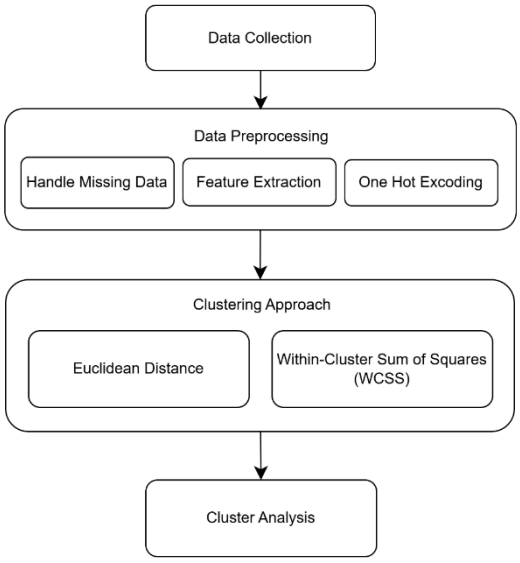

Figure 1. Research Methodology

3.1.Data Collection

The dataset utilized in this study is sourced from Kaggle, a reputable platform for data science and machine learning resources. Specifically, it is derived from the "Financial Transactions Dataset" available on Kaggle, which contains simulated financial transaction records for a fictional financial institution. The dataset was generated using the Python Faker library, ensuring a diverse and realistic range of financial activities. It includes key fields such as transaction ID, amount, transaction type, customer ID, and transaction time, providing a comprehensive overview of customer financial behaviors. This dataset serves as an ideal resource for various analytical tasks, including customer behavior analysis, fraud detection, and the development of personalized financial services.

3.2. Data Preprocessing

In the preprocessing of the dataset, several important steps were taken to ensure data quality and suitability for analysis. Missing data was identified and addressed by imputing numerical values using the mean or median (for columns like amount) and filling categorical columns (such as transaction_type and customer_id) with the most frequent value. If the missing data was too extensive and could not be reliably imputed, those rows were removed to avoid bias in the

analysis. Techniques such as deep learning and multiple imputation algorithms are effective in filling missing data gaps, especially in time-series datasets [34], [35].

Feature extraction was performed by deriving additional features from the transaction_time field, such as extracting the day of the week, month, and hour, enabling better time-based analysis and helping to identify trends and patterns based on different times. Data transformations were applied to ensure the data was suitable for modeling. Numerical features like amount were scaled to a standard range using techniques like Min-Max scaling or standardization to prevent them from dominating the model.

Additionally, categorical variables such as transaction_type and customer_id were converted into numerical values using techniques like one-hot encoding or label encoding. Feature extraction often involves transforming raw data into a more analyzable format, such as using wavelet transforms or convolutional neural networks (CNNs) for automatic feature learning [36]. Efficient preprocessing strategies are essential to ensure optimal model performance and derive meaningful insights from the data [37].

3.3. Clustering Approach

The K-Means algorithm is one of the most widely used clustering techniques in unsupervised machine learning. It partitions data into K clusters based on their proximity to the cluster centers or centroids. The process starts by selecting the desired number of clusters, K. The initial centroids are selected randomly or using methods like K-Means++ for better centroid distribution. Each data point x_i is then assigned to the nearest centroid C_j using a distance metric, typically Euclidean distance:

$$d(x_i, C_j) = \sqrt{\sum_{k=1}^{n}(x_{i,k} - c_{j,k})^2} \quad (1)$$

Each data point is assigned to the cluster whose centroid is closest. Next, the centroid of each cluster is recalculated by taking the mean position of all data points within that cluster. This process repeats the assignment step and the update step until the centroids no longer change significantly, indicating that the algorithm has converged. The primary objective of K-Means is to minimize the within-cluster sum of squares (WCSS), which is the sum of squared distances between each data point and the centroid of its respective cluster:

$$WCSS = \sum_{i=1}^{K} \sum_{x_i \in C_k} (x_i, c_k)^2 \quad (2)$$

K-Means has advantages such as simplicity, computational efficiency, and its ability to handle large datasets. However, it has limitations, including dependence on the selection of K, sensitivity to the initialization of centroids, and the assumption that clusters are spherical and of equal size. The algorithm also struggles with data containing many outliers or clusters with varying density and shapes. Despite these drawbacks, K-Means remains a popular choice for customer segmentation, anomaly detection, and document clustering, especially when the number of clusters is known or can be estimated.

In the clustering process, the selection of features plays a vital role in achieving meaningful segmentation. For this dataset, several key features were chosen to capture the most relevant aspects of user behavior. Amount, representing the monetary value of each transaction, is an essential feature as it helps group users based on their spending habits, distinguishing between high spenders and low spenders. This feature is particularly useful for identifying spending patterns and detecting anomalies such as unusually large or small transactions. Transaction Time provides insights into temporal patterns, such as peak transaction periods throughout the day, week, or year. By extracting components like the day of the week, hour, and month, clustering can reveal trends in when users are most active, highlighting possible seasonality or time-dependent behaviors. Lastly, transaction type, which categorizes transactions as purchases, transfers, or withdrawals, allows for segmentation based on the purpose of the transaction. This feature helps identify user groups with distinct behaviors, such as those who predominantly make purchases versus those who engage in transfers or withdrawals. By combining these features, the clustering algorithm can uncover underlying patterns in the data, enabling financial institutions to segment their users effectively and tailor their offerings to meet specific customer needs.

The Silhouette Score is an effective method for determining the optimal number of clusters in clustering algorithms like K-Means. It evaluates how well each data point fits into its assigned cluster by comparing its similarity to points within the same cluster and those in other clusters. The score ranges from -1 to +1, with a value closer to +1 indicating that the point is well-matched to its own cluster and poorly matched to neighboring clusters, while a value near 0 suggests that the point is near the boundary between two clusters. A score close to -1 implies that the point is likely assigned to the wrong cluster. To find the optimal number of clusters, the Silhouette Score is calculated for different values of K (the number of clusters), and the average score is computed for each. The optimal K is the one that maximizes the average Silhouette Score, indicating the best separation between clusters and the most accurate assignment of data points. This method helps ensure that the clustering results are both meaningful and well-structured, providing clear and distinct groupings in the data.

4. Results and Discussion

4.1. Model Evaluation Result

Table 1 presents the key features of three distinct clusters based on transaction amounts, times, and days. Each cluster exhibits unique patterns in terms of average transaction size, time of day, and frequency of transactions across the week.

Table 1. Cluster Characteristics Summary

cluster	mean_amount	std_amount	mean_hour	mean_day_of_week	mean_month
0	1876.92	1143.21	11.15	0.77	1
1	4147.06	964.44	13.59	3.12	1
2	1970	913.7	11.2	4.85	1

Cluster 0 has a mean transaction amount of 1876.92, with a standard deviation of 1143.21, suggesting moderate variation in transaction sizes ranging from 500 to 4500. Transactions in this cluster tend to occur around 11:15 AM, and they mostly happen early in the week, with an average day of the week value of 0.77, likely indicating Monday. All transactions in Cluster 0 are recorded in the first month.

Cluster 1 stands out with the highest average transaction amount of 4147.06, with less variation in amounts (standard deviation of 964.44) and a range of 2000 to 6000. Transactions here tend to occur around 1:35 PM (mean hour of 13.59) and are concentrated around mid-week, with an average day of the week value of 3.12, likely pointing to Wednesday. Like Cluster 0, these transactions are also from the first month.

Cluster 2 has a mean amount of 1970.00, with a standard deviation of 913.70, indicating moderate variation in transaction sizes, ranging from 700 to 3500. The average transaction time is around 11:20 AM, similar to Cluster 0. However, these transactions are more common later in the week, with a mean day of the week value of 4.85, likely indicating Friday. Transactions in Cluster 2 also occur in the first month. Table 2 presents the statistical summary for each cluster, highlighting key characteristics such as average transaction amounts, transaction times, and transaction types.

Table 2. Cluster Statistical Summary

cluster	mean_amount	std_amount	mean_hour	mean_day_of_week	transaction_type
0	1876.92	1143.21	11.15	0.77	Purchase
1	4147.06	964.44	13.59	3.12	Transfer
2	1970	913.7	11.2	4.85	Purchase

Cluster 0 has a mean transaction amount of 1876.92 with a standard deviation of 1143.21, indicating moderate variability in transaction sizes, and the transactions primarily occur around 11:15 AM. The average day of the week for transactions in this cluster is 0.77, which likely corresponds to Monday. The transaction type for this cluster is predominantly Purchase.

Cluster 1, on the other hand, has the highest average transaction amount at 4147.06, with a standard deviation of 964.44, indicating a higher variation in transaction sizes. Transactions in this cluster mostly occur around 1:35 PM, and the

 ISSN 2776-8007

average day of the week for these transactions is 3.12, likely representing Wednesday. The transaction type here is Transfer. Cluster 2 shows a mean transaction amount of 1970.00, with a standard deviation of 913.70. The transactions typically happen around 11:20 AM, similar to Cluster 0, but occur later in the week, with an average day of the week of 4.85, which suggests these transactions are likely happening on Friday. Like Cluster 0, the transaction type is predominantly Purchase.

Figure 2 illustrates the 3D clustering of users based on their transaction patterns, with the Hour of Transaction on the x-axis, Transaction Amount on the y-axis, and Cluster ID on the z-axis. The plot shows three distinct clusters: Cluster 0 (yellow points), characterized by moderate transaction amounts (around 1876.92) and early transaction times around 11:00 AM, Cluster 1 (blue points), with higher transaction amounts (averaging 4147.06) and transactions occurring around 1:35 PM, and Cluster 2 (purple points), which shares similar transaction times to Cluster 0 (around 11:20 AM) but with slightly higher transaction amounts (averaging 1970.00).

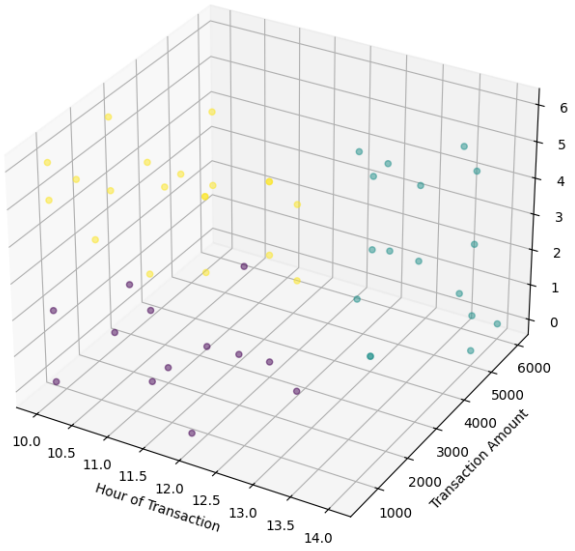

Figure 2. 3D Clustering of Users Based on Transaction Patterns

Figure 3 presents a bar chart displaying the distribution of transactions across the three clusters. The y-axis represents the Number of Transactions, while the x-axis shows the Cluster. Cluster 2 has the highest number of transactions, with just over 20 transactions. Cluster 1 follows with a slightly lower count, around 17 transactions, and Cluster 0 has the least, with just under 15 transactions. This chart visually emphasizes the differences in transaction volumes across the clusters, highlighting that Cluster 2 has the most activity, while Cluster 0 shows the least.

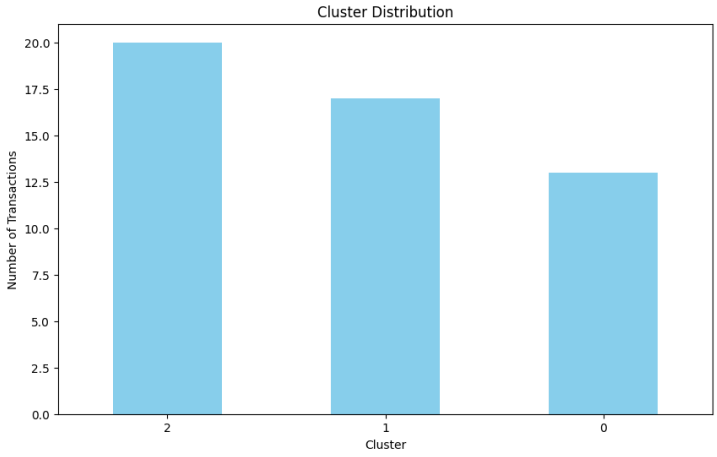

Figure 3. Cluster Distribution

Figure 4 shows the PCA (Principal Component Analysis) projection of the clustering results. The plot visualizes the distribution of clusters in a 2D space, where PCA Component 1 is on the x-axis and PCA Component 2 is on the y-axis. Each point represents a transaction, color-coded based on its cluster. The yellow points correspond to Cluster 0, the purple points represent Cluster 2, and the blue points correspond to Cluster 1. The color gradient on the right indicates the cluster assignment for each point, where higher values represent Cluster 1, medium values represent Cluster 2, and lower values represent Cluster 0. This visualization provides a clear separation of the clusters along the principal components, helping to observe how the clusters are distributed in the reduced-dimensional space.

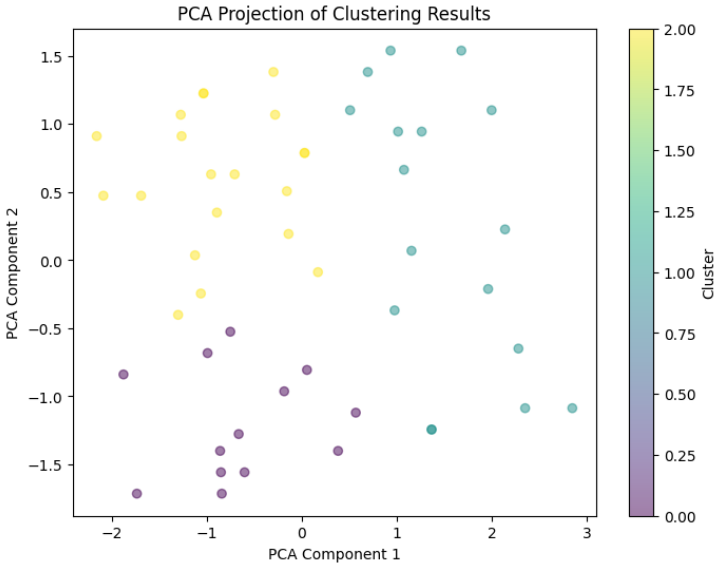

Figure 4. PCA Projection of Clustering Result

Figure 5 shows the t-SNE (t-Distributed Stochastic Neighbor Embedding) projection of the clustering results. The plot visualizes the distribution of the clusters in a 2D space, with t-SNE Component 1 on the x-axis and t-SNE Component 2 on the y-axis. Each point represents a transaction, color-coded based on its cluster. The yellow points represent Cluster 0, the purple points represent Cluster 2, and the blue points represent Cluster 1. The color gradient on the right indicates the cluster assignment for each point. This t-SNE visualization provides a clear separation between the clusters in a 2D space, showcasing the structure of the clusters in the reduced-dimensional representation.

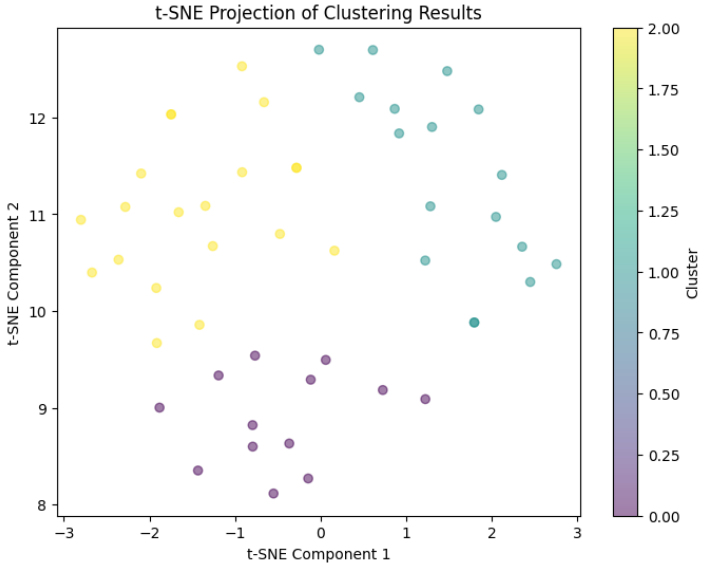

Figure 5. t-SNE Projection of Clustering Results

The Silhouette Score is a metric used to evaluate the quality of clustering results, with a value ranging from -1 to 1. A score close to 1 indicates well-separated and well-defined clusters, a score around 0 suggests that the clusters may be overlapping, and a score closer to -1 indicates that the data points are likely assigned to incorrect clusters. The score of 0.33 for the clustering results indicates a moderate level of cluster quality.

With a score of 0.33, the clustering is neither exceptionally good nor poor. It suggests that the clusters are somewhat distinguishable, but there is still some degree of overlap or ambiguity between them. This value indicates that while the clustering algorithm has successfully identified some patterns in the data, the separation between the clusters is not as strong as it could be. There might be instances where data points from different clusters are close to one another, which could potentially lead to misclassification or confusion about which cluster they belong to.

Improving this score could involve adjusting the clustering method, exploring alternative algorithms, or fine-tuning the model parameters to achieve better separation between the clusters. Additional dimensionality reduction techniques or more advanced clustering methods might help in achieving a more distinct clustering structure. Overall, while a score of 0.33 does indicate some structure in the data, it suggests that the clustering could be further optimized for better clarity and more meaningful separation of groups.

4.2. Discussion

The clustering analysis results presented in this study reveal distinct patterns in transaction amounts, times, and types across three clusters. Cluster 0 is characterized by a mean transaction amount of 1876.92, with moderate variation in transaction sizes (standard deviation of 1143.21). Transactions in this cluster predominantly occur around 11:15 AM on Mondays, and are mostly purchases. This aligns with findings from previous studies [27], which emphasized the importance of segmenting users based on time and frequency of transactions. Cluster 1, on the other hand, stands out with a higher mean transaction amount of 4147.06, showing less variation and a focus on transfer transactions that occur around 1:35 PM on Wednesdays. The distinct characteristics of Cluster 1 support the work of Zhao et al. [28] and Zhang et al. [8], who highlighted that larger transaction amounts and specific times can reveal behaviors such as transfers. Cluster 2, which shares similarities with Cluster 0 in terms of transaction times (around 11:20 AM), has slightly higher transaction amounts and more frequent transactions later in the week, particularly on Fridays. This indicates that transaction patterns vary not only by amount but also by the day of the week, as seen in the behaviors of Clusters 0 and 2.

The visualizations in Figure 2 and Figure 3 further confirm the distinct clusters. The 3D clustering plot and PCA projections show clear separation between the three clusters, with each cluster forming distinct groups based on the time of transaction and transaction amount. This separation helps to understand the different patterns of user behavior, supporting the findings of Komati [30], who emphasized the role of data visualization in interpreting clustering results. However, despite the clear patterns, the moderate Silhouette Score of 0.33 suggests that some overlap exists between the clusters. This indicates that while the clustering algorithm successfully identifies some patterns, there is still ambiguity between certain data points, which could lead to misclassification.

Additionally, the dataset used in this analysis is limited to a single month of transaction data, which may not capture long-term trends or seasonal variations in transaction behavior. This limitation restricts the generalizability of the findings. Future studies could include a more extended time period and incorporate other factors, such as geographic location, socio-economic status, or account types, which could further refine the clustering results and offer more comprehensive insights into user behavior. Moreover, a deeper exploration of the types of transactions beyond just purchases and transfers could provide a more nuanced understanding of how users interact with financial platforms.

Despite these limitations, the findings of this study have important implications for businesses looking to improve their customer segmentation and marketing strategies. By understanding the specific transaction behaviors of different clusters, businesses can tailor their products, services, and promotional campaigns more effectively. For instance, Cluster 1’s focus on high-value transactions and transfers could be targeted with premium financial products or services, while Clusters 0 and 2, which are more focused on purchases, could benefit from tailored promotions or discounts during peak transaction times. This segmentation approach can help businesses optimize their offerings and enhance customer satisfaction.

In conclusion, while this study presents valuable insights into transaction behavior and user segmentation, there is room for improvement in the clustering methodology. Future research could refine the clustering approach, incorporate additional variables, and extend the time frame to provide a more comprehensive understanding of user behaviors. The results of this study underline the importance of transaction data analysis in business strategy, highlighting how better segmentation can lead to more personalized and effective marketing efforts.

5. Conclusion

This study demonstrates the importance of user profiling based on financial transaction patterns using clustering techniques to segment users effectively. The analysis identifies three distinct clusters, each exhibiting unique patterns in transaction amounts, times, and types. Cluster 0, primarily focused on purchases and characterized by moderate variation in transaction amounts, tends to occur early in the week. Cluster 1, with a focus on transfers and higher transaction amounts, tends to occur mid-week, while Cluster 2, which also emphasizes purchases, shows a preference for later transactions in the week. These findings underscore the role of transaction data analysis in improving customer segmentation, enabling businesses to tailor their marketing strategies and financial products more effectively.

Despite the significant insights gained, the study also reveals certain limitations, including the moderate Silhouette Score, suggesting some overlap between clusters. This indicates that further refinement of the clustering methodology, such as using density-based techniques like DBSCAN, could improve the separation between clusters and the clarity of user behaviors. Additionally, the dataset, limited to a single month, restricts the ability to generalize the findings to broader time frames. Future research could address these limitations by incorporating larger datasets over extended periods, including additional user-specific variables, such as geographic and socio-economic factors, to refine segmentation.

This research highlights the potential of clustering techniques, particularly K-means, in understanding complex user behaviors in financial transactions. By enhancing user segmentation, businesses can improve personalization and customer satisfaction, leading to more effective and targeted marketing strategies. Future work will build on these findings by refining clustering techniques and exploring new methods to better capture evolving user behaviors and enhance predictive analytics in financial services.

6. Declarations

6.1. Author Contributions

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

6.2. Data Availability Statement

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

6.3. Funding

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

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. 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] H. Patria, "Predicting Fraudulence Transaction Under Data Imbalance Using Neural Network (Deep Learning)," *Data Sci. J. Comput. Appl. Informatics*, vol. 6, no. 2, pp. 67–80, 2022, doi: 10.32734/jocai.v6.i2-8309.

[2] D. P. Hoang, T. H. H. Nguyen, N. L. Vuong, and L. V Dat, "Linking Psychological Needs, Perceived Financial Well-Being and Loyalty: The Role of Commercial Banks," *J. Financ. Serv. Mark.*, vol. 28, no. 3, pp. 466–487, 2022, doi: 10.1057/s41264-022-00170-z.

[3] X. Yue, "Application of AI Technology in Personalized Recommendation System for Financial Services," *Appl. Math. Nonlinear Sci.*, vol. 9, no. 1, pp. 1–18, 2024, doi: 10.2478/amns-2024-1349.

[4] L. Puli, N. Layton, D. Bell, and A. Z. M. Shahriar, "Financial Inclusion for People With Disability: A Scoping Review," *Glob. Health Action*, vol. 17, no. 1, pp. 1–12, 2024, doi: 10.1080/16549716.2024.2342634.

[5] T. N. Odonkor, N. V Eziamaka, and A. A. Akinsulire, "Advancing Financial Inclusion and Technological Innovation Through Cutting-Edge Software Engineering," *Financ. Account. Res. J.*, vol. 6, no. 8, pp. 1320–1348, 2024, doi: 10.51594/farj.v6i8.1375.

[6] I. A. Adeniran, A. O. Abhulimen, A. N. Obiki-Osafiele, O. S. Osundare, E. E. Agu, and C. P. Efunniyi, "Data-Driven Approaches to Improve Customer Experience in Banking: Techniques and Outcomes," *Int. J. Manag. Entrep. Res.*, vol. 6, no. 8, pp. 2797–2818, 2024, doi: 10.51594/ijmer.v6i8.1467.

[7] G. Guo, "A Study on User Behavior of E-Commerce Platforms Based on Data Analysis," *Glob. Humanit. Soc. Sci.*, vol. 4, no. 04, pp. 181–185, 2023, doi: 10.61360/bonighss232014180806.

[8] Z. Zhang, L. Chen, Q. Liu, and P. Wang, "A Fraud Detection Method for Low-Frequency Transaction," *IEEE Access*, vol. 8, no. January, pp. 25210–25220, 2020, doi: 10.1109/access.2020.2970614.

[9] Y. Zhang, R. Garg, L. L. Golden, P. L. Brockett, and A. Sharma, "Segmenting Bitcoin Transactions for Price Movement Prediction," *J. Risk Financ. Manag.*, vol. 17, no. 3, pp. 1–17, 2024, doi: 10.3390/jrfm17030128.

[10] W. Wen and X. Han, "An Introduction of Transaction Session-induced Security Scheme Using Blockchain Technology: Understanding the Features of Internet of Things–based Financial Security Systems," *Manag. Decis. Econ.*, vol. 45, no. 4, pp. 1817–1834, 2024, doi: 10.1002/mde.4043.

[11] M. Alrahhal and F. Bozkurt, "Analyzing Big Social Data for Evaluating Environment-Friendly Tourism in Turkey," *J. Intell. Syst. Theory Appl.*, vol. 6, no. 2, pp. 130–142, 2023, doi: 10.38016/jista.1209415.

[12] P. O. Shoetan, A. T. Oyewole, C. C. Okoye, and O. C. Ofodile, "Reviewing the Role of Big Data Analytics in Financial Fraud Detection," *Financ. Account. Res. J.*, vol. 6, no. 3, pp. 384–394, 2024, doi: 10.51594/farj.v6i3.899.

[13] M. Kanwal, M. A. Khan, N. Ismat, N. A. Khan, and A. A. Khan, "Machine Learning Approach to Classification of Online Users by Exploiting Information Seeking Behavior," *IEEE Access*, vol. 12, no. April, pp. 53234–53249, 2024, doi: 10.1109/access.2024.3383444.

[14] A. Salamzadeh, P. Ebrahimi, M. Soleimani, and M. Fekete-Farkas, "Grocery Apps and Consumer Purchase Behavior: Application of Gaussian Mixture Model and Multi-Layer Perceptron Algorithm," *J. Risk Financ. Manag.*, vol. 15, no. 10, pp. 1–16, 2022, doi: 10.3390/jrfm15100424.

[15] W. Wang, W. Xiong, J. Wang, L. Tao, S. Li, Y. Yi, X. Zou, and C. Li, "A User Purchase Behavior Prediction Method Based on XGBoost," *Electronics*, vol. 12, no. 9, pp. 1–17, 2023, doi: 10.3390/electronics12092047.

[16] X. Han, J. Wang, X. Zhang, L. Wang, and D. Xu, "Mining Public Behavior Patterns From Social Media Data During Emergencies: A Multidimensional Analytical Framework Considering Spatial–temporal–semantic Features," *Trans. Gis*, vol. 28, no. 1, pp. 58–82, 2024, doi: 10.1111/tgis.13125.

[17] V. Chen, U. Bhatt, H. Heidari, A. Weller, and A. Talwalkar, "Perspectives on incorporating expert feedback into model updates," *Patterns*, vol. 4, no. 7, pp. 1–13, Jul 2023, doi: 10.1016/j.patter.2023.100780.

[18] Y. Hu, "User Behavior and Satisfaction in AI-Generated Video Tools: Insights From Surveys and Online Comments," *Appl. Comput. Eng.*, vol. 94, no. 1, pp. 136–145, 2024, doi: 10.54254/2755-2721/94/2024melb0065.

[19] L. Lv, J. Wang, R. Wu, H. Wang, and I. Lee, "Density Peaks Clustering Based on Geodetic Distance and Dynamic Neighbourhood," *Int. J. Bio-Inspired Comput.*, vol. 17, no. 1, pp. 24–33, 2021, doi: 10.1504/ijbic.2021.113363.

[20] Z. Huang, H. Zheng, C. Li, and C. Che, "Application of Machine Learning-Based K-Means Clustering for Financial Fraud Detection," *Acad. J. Sci. Technol.*, vol. 10, no. 1, pp. 33–39, 2024, doi: 10.54097/74414c90.

[21] C. Goh, B. Lee, G. Pan, and P. S. Seow, "Forensic Analytics Using Cluster Analysis: Detecting Anomalies in Data," *J. Corp. Account. Financ.*, vol. 32, no. 2, pp. 154–161, 2021, doi: 10.1002/jcaf.22486.

[22] B. J. Jansen, J. Salminen, and S. Jung, "Making Meaningful User Segments From Datasets Using Product Dissemination and Product Impact," *Data Inf. Manag.*, vol. 4, no. 4, pp. 237–249, 2020, doi: 10.2478/dim-2020-0048.

[23] A. M. Koziel and C. Shen, "Psychographic and Demographic Segmentation and Customer Profiling in Mobile Fintech Services," *Kybernetes*, vol. 54, no. 2, pp. 1262–1288, 2023, doi: 10.1108/k-07-2023-1251.

[24] O. C. Ojiaku, B. N. Chidirim, P. Abude, and A. Vincent, "Mobile Banking Service Quality and Customer Retention Among Commercial Banks' Customers: An Empirical Evidence From Southeast Nigeria," *Asian J. Econ. Bus. Account.*, vol. 23, no. 14, pp. 45–56, 2023, doi: 10.9734/ajeba/2023/v23i141004.

[25] Q. Chen, "Application of K-Means Algorithm in Marketing," *Adv. Econ. Manag. Polit. Sci.*, vol. 71, no. 1, pp. 178–184, 2024, doi: 10.54254/2754-1169/71/20241485.

[26] C. M. Eckhardt, S. J. Madjarova, R. J. Williams, M. Ollivier, J. Karlsson, A. Pareek, and B. U. Nwachukwu, "Unsupervised Machine Learning Methods and Emerging Applications in Healthcare," *Knee Surg. Sport. Traumatol. Arthrosc.*, vol. 31, no. 2, pp. 376–381, 2022, doi: 10.1007/s00167-022-07233-7.

[27] C. Martinovska, N. Stojkovic, E. Kosturanova, and M. Klekovska, "Data Mining in Client Oriented Businesses," *Int. J. Res. Stud. Publ.*, vol. 11, no. 12, pp. 378–383, 2021, doi: 10.29322/ijsrp.11.12.2021.p12054.

[28] H. Zhao, X. Luo, R. Ma, and L. Xi, "An Extended Regularized K-Means Clustering Approach for High-Dimensional Customer Segmentation With Correlated Variables," *IEEE Access*, vol. 9, no. March, pp. 48405–48412, 2021, doi: 10.1109/access.2021.3067499.

[29] E. Bilgiç, Ö. ÇAKIR, M. Kantardzic, Y. Duan, and G. Cao, "Retail Analytics: Store Segmentation Using Rule-Based Purchasing Behavior Analysis," *Int. Rev. Retail Distrib. Consum. Res.*, vol. 31, no. 4, pp. 457–480, 2021, doi: 10.1080/09593969.2021.1915847.

[30] D. Komati, "Machine Learning Architectures for RealTime Financial Decision Systems: Implementation and Impact," *Ijaem*, vol. 7, no. 3, pp. 744–752, 2025, doi: 10.35629/5252-0703744752.

[31] U. N. Chukwukaelo, R. P. Igbojioyibo, and J. K. Abu, "The Role of Machine Learning in Enhancing Corporate Financial Planning," *Asian J. Econ. Bus. Account.*, vol. 24, no. 11, pp. 22–32, 2024, doi: 10.9734/ajeba/2024/v24i111540.

[32] K. J. Olowe, N. L. Edoh, S. J. C. Zouo, and J. Olamijuwon, "Review of Predictive Modeling and Machine Learning Applications in Financial Service Analysis," *Comput. Sci. It Res. J.*, vol. 5, no. 11, pp. 2609–2626, 2024, doi: 10.51594/csitrj.v5i11.1731.

[33] S. Feng, "Integrating Artificial Intelligence in Financial Services: Enhancements, Applications, and Future Directions," *Appl. Comput. Eng.*, vol. 69, no. 1, pp. 19–24, 2024, doi: 10.54254/2755-2721/69/20241455.

[34] T.-S. Kim, J.-H. Kim, W. Yang, H. Lee, and J. Choo, "Missing Value Imputation of Time-Series Air-Quality Data via Deep Neural Networks," *Int. J. Environ. Res. Public Health*, vol. 18, no. 22, pp. 1–8, 2021, doi: 10.3390/ijerph182212213.

[35] C. Qiu, "A Method Using LSTM Networks to Impute Missing Temperatures in Temperature Datasets and to Predict Future Temperatures," *Highlights Sci. Eng. Technol.*, vol. 46, no. April, pp. 116–124, 2023, doi: 10.54097/hset.v46i.7691.

[36] J. Yang, Y. Zhang, K. Wang, Y. Tong, J. Liu, and G. Wang, "Coal–Rock Data Recognition Method Based on Spectral Dimension Transform and CBAM-VIT," *Appl. Sci.*, vol. 14, no. 2, pp. 1–13, 2024, doi: 10.3390/app14020593.

[37] P. Li, Z. Chen, X. Chu, and K. Rong, "DiffPrep: Differentiable Data Preprocessing Pipeline Search for Learning Over Tabular Data," *Proc. Acm Manag. Data*, vol. 1, no. 2, pp. 1–26, 2023, doi: 10.1145/3589328.