Eduvest – Journal of Universal Studies
Volume 5 Number 8, Agustus, 2025
p- ISSN 2775-3735- e-ISSN 2775-3727

Customers Segmentation for Digital Signature Implementation: RFC Analysis Using KMeans Algorithms

Abdul Aziz Al Rasyid[1*], Rila Mandala[2]
[1]Faculty of Computing, President University, Cikarang, Indonesia
[2]Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung
Email: abdul.rasyid@student.president.ac.id

ABSTRACT

This research develops an innovative customer segmentation framework for Indonesia's national digital signature services. Analyzing 200 million transactions across 750 government institutions, we adapt the RFM model into an RFC (Recency, Frequency, Conversion Rate) framework suitable for non-commercial contexts. The K-Means algorithm identifies four distinct user segments with varying adoption patterns: High Performers (32%), Frequent Users (24%), Selective Adopters (28%), and Low Engagers (16%). Our RFC-KMeans model achieves superior clustering quality (Silhouette Score: 0.52; Davies-Bouldin Index: 1.15) compared to conventional approaches. Key findings reveal that Conversion Rate serves as the most significant differentiator ($p < 0.01$), explaining 42% of variance in adoption patterns. The segmentation enables targeted interventions, particularly for Low Engagers showing <30% active user rates despite adequate infrastructure. This research contributes both methodologically and practically by: (1) validating RFC as an effective alternative to RFM for public digital services; (2) demonstrating K-Means' scalability for government transaction analysis; and (3) providing an actionable roadmap for differentiated institution support strategies. The framework supports Indonesia's digital transformation goals while offering transferable insights for similar e-government initiatives globally.

KEYWORDS *Digital Signature, RFC Analysis, K-means, Clustering, Digital Transformation*

INTRODUCTION

Digital transformation has become an important aspect of modern governance today. The shift from manual to digital systems has been proven to increase efficiency, transparency, and security (Ayuandiani et al., 2023). Indonesia is currently developing rapidly to carry out digital transformation, as often conveyed by President Joko Widodo that digital transformation must be carried out gradually, with a focus on the integration of public services that are more easily accessible to the public. Many electronic systems currently support public services but have not accelerated existing business and service processes, this is because they have not been fully transformed, by still using manual signatures by related officials causing the process of issuing documents from government services to take a long time (Santiago & Nery, 2023). The

government adopts electronic certification and digital signatures as an important part of ensuring that the public service process runs well, safely and accountably.

Balai Besar Sertifikasi Elektronik (BSrE), as the Certificate Authority (CA) in Indonesia, is responsible for managing and issuing electronic certificates and digital signatures for all government agencies in Indonesia. With more than 750 agencies as stakeholders, over 750,000 users, and 200 million transactions, BSrE plays a vital role in supporting Indonesia's digital governance infrastructure (Sutopo et al., 2022). Digital signatures issued by BSrE help ensure data integrity, authentication, and non-repudiation in electronic transactions (Putra & Santosa, 2020). This contributes to enhancing public trust in online public services and reducing the risk of cyber fraud (Hidayat et al., 2021). Moreover, digital signature implementation is aligned with the national e-government roadmap, as mandated by the Ministry of Communication and Information Technology (Kominfo, 2023; Pratama et al., 2021). Studies also show that secure digital identity infrastructures, such as those enabled by BSrE, are essential to support interoperability and secure communication between government systems (Ramadhani & Nugroho, 2020). By implementing a robust digital signature system, BSrE helps accelerate digital transformation and paperless bureaucracy (Irawan et al., 2023), contributing to Indonesia’s commitment to a more transparent and efficient public sector (Rohmah & Widodo, 2021).

In this context, understanding how stakeholders are able to implement digital signatures is crucial to see the success of the digital transformation that has been carried out in Indonesia. Clustering stakeholders based on certain criteria can provide valuable insights into this progress, allowing BSrE to provide tailored support and resources for each stakeholder.

User segmentation is an important process to optimize business strategy in a company because with this segmentation can understand various user behaviors with various existing variables (Siyue et al., 2019). User segmentation can help companies manage each user group more effectively, for example, with certain special offers, services, or strategies, so that it can increase customer satisfaction and, of course, will increase the company's business value. User segmentation methods usually only rely on criteria in a small scope, while currently user segmentation is needed with very large user data and has various characteristics and determining variables, therefore machine learning is a more sophisticated method to understand and group users more deeply (Narayana et al., 2022).

One of the methods that is widely used today is RFM (Recency, Frequency, Monetary) analysis, this is the most effective method in segmenting or clustering customers. RFM is done to understand their behavior and value towards the business we have (Sabuncu et al., 2020). Several previous studies have proven that the use of machine learning for the RFM Analysis process can produce good results, as done by Wu et al. (2020) by clustering customer data in a company into four groups based on the behavior of each customer. The results of this clustering can then be used as a basis for making policies that can improve the company's performance index and increase customer satisfaction. In addition, there is also another study that discusses the use of RFM Analysis to obtain clustering of potential customers by preparing the right data for each cluster of potential customers. By looking at several variables such as Sales history and customer behavior itself (Anitha & Patil, 2022).

However, it is very difficult to do the clustering process if done manually, comparing several variables with very large values and interrelated (Mahfuza et al., 2022). As mentioned

in the previous paragraph one of them is with machine learning to make the grouping process much more effective and efficient and able to obtain valid data. K-Means is a machine learning algorithm with an unsupervised learning method that can be used for customer segmentation, targeted marketing, and recommendation systems. K-Means is suitable for clustering processes for very large datasets because it has a simple and quick process algorithm, and of course the clustering results obtained can provide clear information, especially when used with RFM analysis (Nirmala & Makzoom, 2023).

The KMeans algorithm divides the user dataset into several parts according to its criteria. The division of each cluster involves finding the minimum squared error between the various data points in the dataset and the cluster average, then assigning each data point to the closest cluster center. k means the number of clusters, while means is the average or midpoint of each cluster. If viewed mathematically, the following formula will be obtained:

$$d = \sqrt{\sum_{i=0}^{n} (xi - yi)^2}$$

Description:

d : Euclidean distance between two points in N-dimensional space.

n : Number of dimensions or features of the data being analyzed.

xi, yi : Components of two vectors representing the two points whose distances are being calculated (e.g., one vector is the data, and the other is the centroid).

Based on several previous studies, RFM is used when customers spend money to enjoy the Company's products or services so that the company has benefits from its customers which are used as one of the bases for the customer clustering analysis process (Sabuncu et al., 2020; Narayana et al., 2022). Then what if the Company does not take advantage of customers in other words customers do not spend money to enjoy services from the Company, this becomes a new problem when using RFM analysis.

BSrE is one of the Indonesian government agencies that provide digital signature services to all customers but cannot receive benefits from customers. Therefore, this study will try to adopt the RFM analysis by replacing the monetary factor with the conversion rate of digital signature service users in each customer agency, thus changing RFM to RFC analysis. The conversion rate referred to in this study is how evenly distributed users in an agency use digital signatures, while the agencies in this study are referred to as BSrE customers. In addition, in previous research, no one has used a dataset in the form of variables for lists of government agencies, digital signature transactions, and the number of users.

A significant problem in evaluating the implementation of digital signatures lies in the method of evaluation and monitoring to users. Traditional RFM analysis and tends to be random selection, which is currently used to see clusters of each customer to determine the evaluation and support of service implementation to the customer itself. In addition, BSrE does not directly benefit from users, so the monetary variable of the RFM analysis is irrelevant.

The purpose of this study is to cluster government agencies that utilize BSrE digital signature (TTE) services based on their activity variables, using the RFC base on RFM framework analysis. This segmentation is expected to provide valuable insights to develop a

more targeted and effective government policy strategy specifically in BSrE. The main objective is to increase the ratio of active TTE users nationally and evenly through tailored interventions based on the segmentation results.

METHOD

This section explains the methodological process used in this study, including the research approach, data collection and processing, and how to implement with existing tools and of course evaluate the results obtained. This study uses a quantitative approach focused on data-based methods of digital signature service users at BSrE which are then processed to obtain user segmentation. The main objective is to analyze customer behavior by applying machine learning techniques, especially the KMeans clustering algorithm with the RFM analysis approach by replacing Monetary values with Conversion Rates, from the results of this study will be obtained information that can be used to increase the value of digital signature implementation in Indonesia.

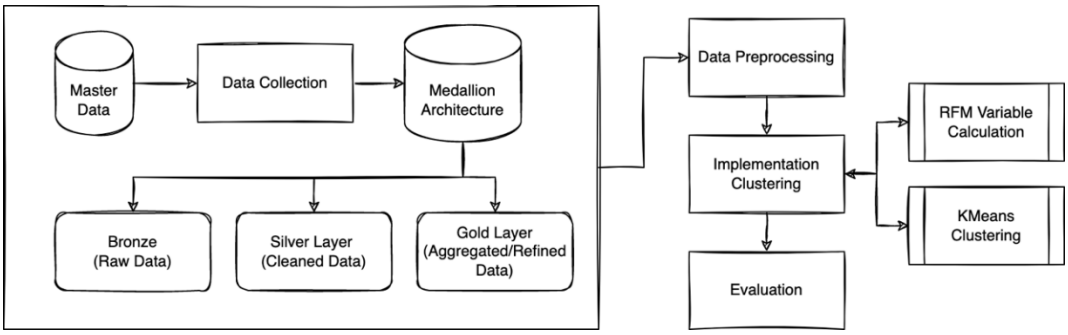

Figure 1. Research Design
Source: Researcher analysis, 2025

In figure 1, the data collection section starts from taking some data from the digital signature system master data, including user data with more than 850,000 registered users, electronic certificate data issued to users, often exceeding the number of users because individuals can have multiple certificates, agency data consisting of 750 government agencies, each representing a unit of analysis in the grouping process as a BSrE stakeholder, and the last is digital signature transaction data consisting of 200 million digital signature transactions and continues to increase in real-time. Then the data is orchestrated using medallion architecture, by dividing it into 3 data layers, namely brown, silver, and gold layers. there are several steps taken such as handling data with "null" values, incomplete data, outlier data, and data normalization.

Table 1. Dataset Table.

Attribute	Type	Description
institution_id	UUID	A unique random string owned by each institution
pengguna_id	UUID	A unique random string owned by each user
waktu	UTC	When the user performs a digital signature
times	Int	The number of digital signatures that have been carried out by each user

Source: BSrE primary data, 2025

Adopting RFM Analysis by changing the monetary variable with the conversion rate because BSrE is non-commercial. The components of the RFC model include:

- Recency (R): Calculation of the average time the user last made a digital signature transaction viewed at each institution, calculated for 30 days.
- Frequency (F): The average total number of digital signature transactions at an institution, calculated for 30 days.
- Conversion Rate (C): The ratio of active digital signature users to total registered users in an institution, which indicates the level of digital signature adoption at an institution.

After calculating the RFC Models, the data obtained is in table 2 below.

Table 2. RFC Result

nama_instansi	id_institusi	recency	frequency	conversion_rate
str	str	f64	f64	f64
"Arsip Nasional Republik Indone...	"b82606b8-80e9-4505-b9e2-b1a868...	1.096774	3818.191489	0.429972
"Badan Amil Zakat Nasional"	"2d995438-ff26-42a6-a25f-f8b4ea...	2.935484	60.565217	0.882353
"Badan Informasi Geospasial"	"90849341-b280-4e85-87a6-6d83ce...	1.032258	5861.191489	0.931559
"Badan Karantina Indonesia"	"4283817a-6ad5-4e97-aa18-b60f3a...	1.032258	2098.272727	0.339921
"Badan Keamanan Laut"	"05b4147e-599b-4078-be8d-c7d9d4...	1.83871	298.608696	0.584795
...	...	...	...	...

Source: RFC calculation results, 2025

Before carrying out the user segmentation process, the existing RFM data is first normalized to check whether there are any "null" values and transform the logoarithmic scale data to maintain data stability so that it has a good skewness value (normal distribution), so that the calculation process with the machine learning algorithm can be stable and maintain the value of outlier data. In research using the formula $log1p(x) = In(1 + x)$, Thus the value of log_RFC is obtained, as in table 3.

Table 3. Log_RFC Result

nama_instansi	id_institusi	recency	frequency	conversion_rate	log_frequency	log_recency	log_conversion_rate
str	str	f64	f64	f64	f64	f64	f64
"Arsip Nasional Republik Indone...	"b82606b8-80e9-4505-b9e2-b1a868...	1.0	3822.978723	0.429972	8.249047	0.693147	0.429972
"Badan Amil Zakat Nasional"	"2d995438-ff26-42a6-a25f-f8b4ea...	3.322581	60.565217	0.882353	4.120097	1.463853	0.882353
"Badan Informasi Geospasial"	"90849341-b280-4e85-87a6-6d83ce...	1.0	5863.170213	0.931559	8.676616	0.693147	0.931559
"Badan Karantina Indonesia"	"4283817a-6ad5-4e97-aa18-b60f3a...	1.0	2111.454545	0.339921	7.655606	0.693147	0.339921
"Badan Keamanan Laut"	"05b4147e-599b-4078-be8d-c7d9d4...	1.709677	298.73913	0.584795	5.702913	0.99683	0.584795
...	...	...	...	...	...	...	...

Source: Logarithmic data transformation, 2025

The user clustering process uses the KMeans algorithm because this algorithm is simple, scalable, and efficient in processing large data sets according to the existing dataset with 3 existing variables. The first step is to determine the optimal number of clusters (k) using the Elbow Method, this method evaluates the within-cluster sum of squares (WCSS) and plots it against the number of clusters to find the point where the reduction in WCSS becomes insignificant, which indicates the ideal value of k in a dataset, the value is also known as inertia.

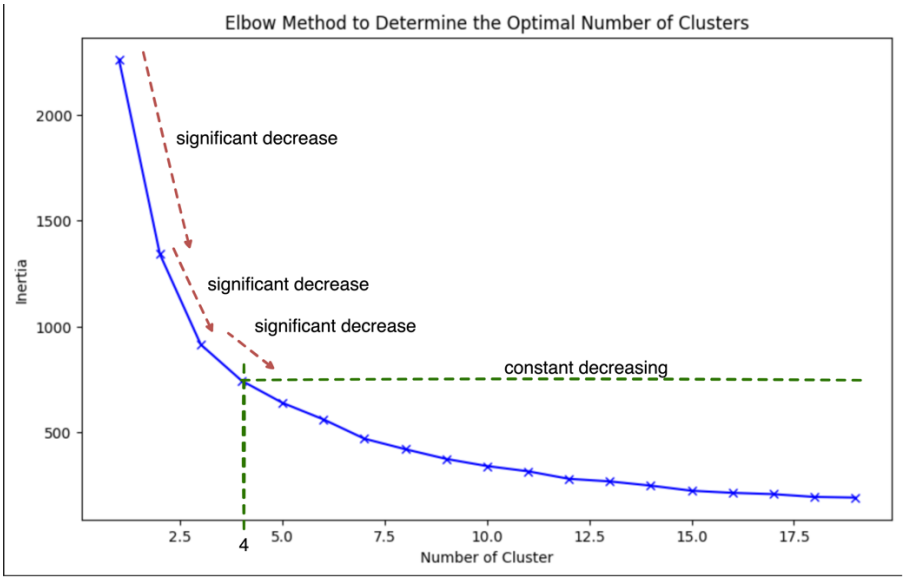

Figure 2. Elbow Method Result
Source: K-Means analysis using Python, 2025

From the calculation results in figure 2, the best value (k) is 4, then clustering will be carried out using the K-Means algorithm with 3 variables in the dataset, using the following formula:

$$di = \sqrt{\sum_{i=0}^{n} (x - Fi)^2 + (y - Ri)^2 + (z - Cri)^2}$$

From this formula, we can explore data from the calculation results. In Figure 3, we can see the exploration of cluster distribution data if connected between 2 of the 3 variables in the dataset, such as based on Frequency and Recency, Ratio and Recency, Frequency and Ratio.

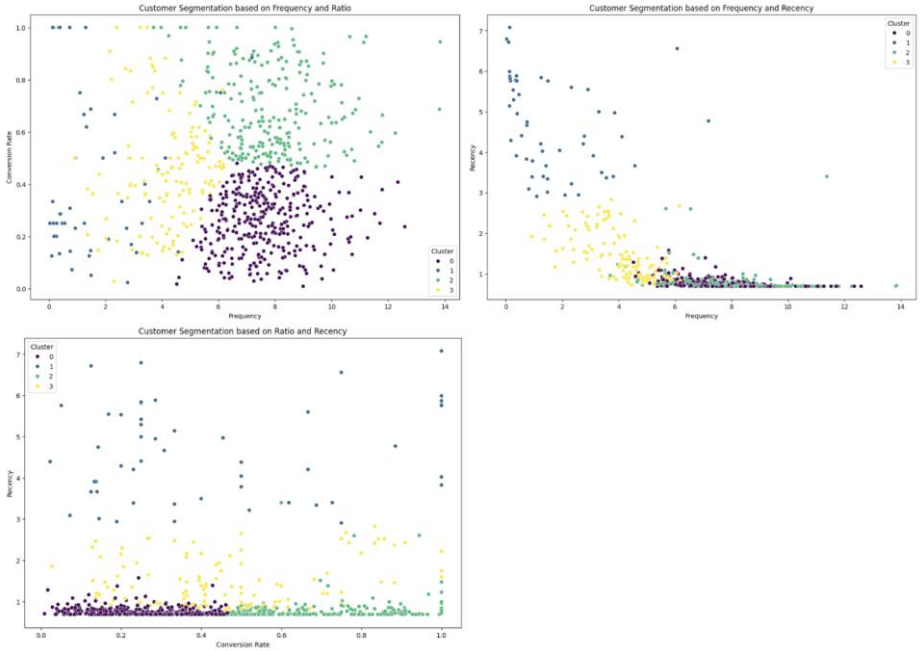

Figure 3. Data exploration
Source: Clustering results visualization, 2025

RESULT AND DISCUSSION

Furthermore, this study uses a boxplot graph to provide a clearer picture of the data distribution with the RFC approach within each cluster. The first graph (Frequency) in Figure 4 shows that cluster 1 and cluster 3 have high frequency values compared to other clusters, while cluster 2 has a lower frequency distribution with some outliers. This indicates that Cluster 1 and Cluster 3 tend to consist of institutions with a more active level of interaction on digital signature services, while Cluster 2 may include institutions with low activity.

In the second graph (Recency) Figure 5 and the third (Conversion Rate) Figure 6, there are significant differences in patterns between clusters. In the Recency dimension, Cluster 2 has a much higher value than other clusters, which means that institutions in this cluster rarely use digital signature services. Meanwhile, Cluster 4 has a relatively high Conversion Rate value, indicating that even though the frequency of activity is low, institutions in this cluster tend to have active users spread across each institution. So, from these 3 graphs, we can provide a clear understanding of the unique characteristics of each cluster, which in the future can be used to determine strategies in decision-making or institutional management based on their activity patterns.

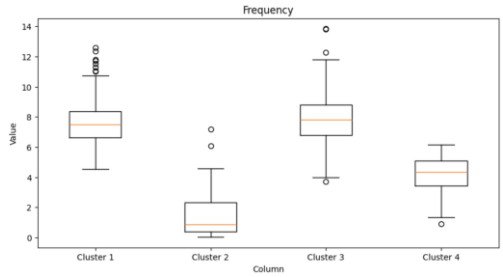

Figure 4. Boxplot Graph for Frequency
Source: Frequency distribution analysis, 2025

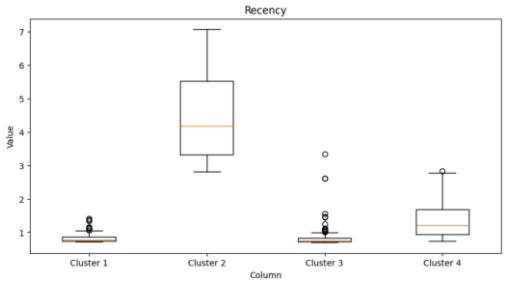

Figure 5. Boxplot Graph for Recency
Source: Recency distribution analysis, 2025

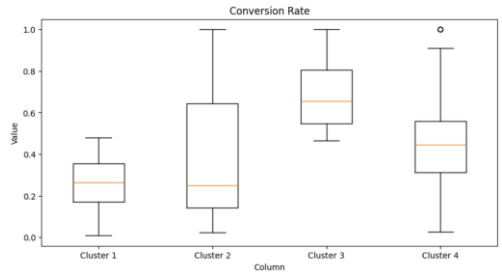

Figure 6. Boxplot Graph for Conversion Rate
Source: Conversion rate distribution analysis, 2025

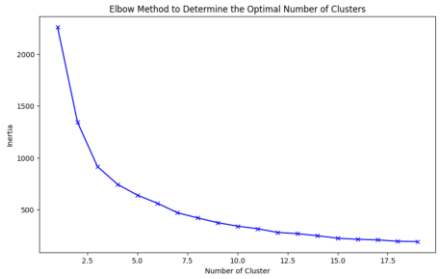

Figure 7. Elbow Method Result for Cluster Evaluation
Source: Optimal cluster number validation, 2025

Evaluation for k = 4 used using the elbow method approach in Figure 7, it can be seen that the WCSS inertia value experienced a significant decrease until the 4th point, after which the decrease slowed down and tended to be constant. This indicates that choosing the number of clusters to 4 provides more effective clustering. However, if the number of clusters is increased again, it only provides a very significant reduction in inertia, so it does not provide

an increase in cluster quality. Therefore, based on the Elbow Method, the optimal number of clusters for this dataset is 4 clusters.

CONCLUSION

This study demonstrates the effectiveness of employing the RFC analysis model combined with the K-Means algorithm to segment institutions using BSrE digital signature services. By substituting the monetary component in the traditional RFM framework with the conversion rate, the research successfully categorized institutions into meaningful groups that reflect distinct user behaviors and usage patterns. These insights provide a valuable foundation for informed decision-making in Indonesia's digital transformation efforts. For future research, it is suggested to explore the integration of additional behavioral or contextual variables, such as user satisfaction or technological readiness, to further enhance the precision and applicability of customer segmentation in e-government services.

REFERENCES

Anitha, P., & Patil, M. (2022). RFM model for customer purchase behavior using K-means algorithm. *Journal of King Saud University - Computer and Information Sciences, 34*(4), 1785–1792. https://doi.org/10.1016/j.jksuci.2019.12.011

Ayuandiani, W., Fausiah, Mukhram, M., Listiawati, N., & B, I. (2023). Digital transformation in financial management: Security and efficiency. *International Journal of Applied Research and Sustainable Sciences.* https://doi.org/10.59890/ijarss.v1i3.875

Hidayat, A., Firmansyah, A., & Ramadhan, R. (2021). Keamanan tanda tangan digital dalam e-government: Studi kasus di Indonesia. *Jurnal Teknologi Informasi dan Komunikasi, 9*(2), 101–112.

Irawan, Y., Suryadi, K., & Hapsari, R. (2023). Digital signature implementation for bureaucratic efficiency in Indonesia. *Procedia Computer Science, 219*, 332–338. https://doi.org/10.1016/j.procs.2023.01.045

Kominfo. (2023). *Roadmap SPBE Indonesia 2023–2028: Strategi transformasi digital pemerintah.* Kementerian Komunikasi dan Informatika Republik Indonesia. https://kominfo.go.id

Mahfuza, R., Islam, N., Toyeb, M., Emon, M. A. F., Chowdhury, S. A., & Alam, Md. G. R. (2022). LRFMV: An efficient customer segmentation model for superstores. *PLoS One, 17*(11), e0279262. https://doi.org/10.1371/journal.pone.0279262

Narayana, V., Sirisha, S., Divya, G., Pooja, N. L. S., & Nouf, Sk. A. (2022). Mall customer segmentation using machine learning. In *2022 International Conference on Electronics and Renewable Systems (ICEARS)* (pp. 1280–1288). IEEE. https://doi.org/10.1109/icears53579.2022.9752447

Nirmala, M., & Makzoom, M. S. (2023). Application development for customer segmentation using an unsupervised learning algorithm. *International Journal of Scientific Research in Science, Engineering and Technology.* https://doi.org/10.32628/ijsrset2310215

Pratama, A. P., Wibowo, R. A., & Dewi, R. (2021). Enhancing trust in digital government services through PKI-based signatures. *Journal of Cyber Security and Digital Forensics, 4*(3), 57–68. https://doi.org/10.1016/j.cys.2021.07.005

Putra, F. D., & Santosa, P. I. (2020). Public key infrastructure implementation and security issues in Indonesia. *Journal of Information Security Research, 8*(1), 11–23. https://doi.org/10.1016/j.jisr.2020.03.002

Ramadhani, A., & Nugroho, H. (2020). Interoperability challenges in digital identity for Indonesian e-government. *International Journal of Digital Governance, 5*(1), 44–59. https://doi.org/10.1016/j.ijdg.2020.05.007

Rohmah, N., & Widodo, A. (2021). Digital governance in Indonesia: Role of certificate authority in ensuring secure transactions. *Government Information Quarterly, 38*(4), 101614. https://doi.org/10.1016/j.giq.2021.101614

Sabuncu, I., Türkan, E., & Polat, H. (2020). Customer segmentation and profiling with RFM analysis. *TUJOM, 5*(1), 22–36. https://doi.org/10.30685/tujom.v5i1.84

Santiago, D., & Nery, I. X. (2023). Industry contribution: Digital signature as a method to strengthen enterprise risk management practices across the US government. *Digital Evidence and Electronic Signature Law Review, 20*, 143–158. https://doi.org/10.14296/deeslr.v20i.5605

Siyue, L., et al. (2019). Prediction of business user segmentation model based on customer value. In *2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)* (pp. 227–231). IEEE. https://doi.org/10.1109/ICCCBDA.2019.8725722

Wu, J., et al. (2020). An empirical study on customer segmentation by purchase behaviors using a RFM model and K-means algorithm. *Mathematical Problems in Engineering, 2020*, 8884227. https://doi.org/10.1155/2020/8884227