Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/10.47709/cnahpc.v8i2.7942

Submitted :2026-01-31
Accepted : 2026-02-05
Published : 2026-04-30

Analysis of Segmentation and Client Target Market Business Decisions in Construction Service Company Using K-Means and Decision Tree Algorithms : Case Study at CV Jowon Solusindo

wondho Setia[1)], Mardiani[2)]
[1,2)] Universitas Multi Data Palembang,South Sumatera, Indonesia
[1)]wondho.setia.st@gmail.com ,[2)] mardiani@mdp.ac.id

ABSTRACT

High competition in the construction service industry requires companies to adopt efficient marketing strategies to reduce Customer Acquisition Costs (CAC). CV Jowon Solusindo faces challenges regarding marketing inefficiency due to the implementation of a one size fits all strategy and a high number of unconverted leads (Lost Prospects). This study aims to classify customer characteristics and discover decision rules to formulate personalized marketing strategies. This research employs a quantitative approach with Data Mining methods based on the CRISP-DM framework. The dataset consists of 576 historical transaction records that have undergone data cleaning processes. The method used is a hybrid approach, combining the K-Means Clustering algorithm for customer segmentation and the Decision Tree (C4.5) for rule pattern extraction. The results indicate that the K-Means algorithm with k=3 successfully mapped customers into three distinctive segments, Young Emerging Clients (Average age 33 years with the highest project value), Established Senior Clients (Average age 54 years with stable frequency), and Lost Prospects (Average age 42 years with the lowest offer value). The Decision Tree analysis yielded an accuracy of 67% and identified Age as the primary determinant factor with a split point at 43.5 years. Based on these findings, it is recommended to differentiate marketing strategies into digital visual approaches for customers under 43.5 years and personal approaches for those above that age, as well as pricing strategy adjustments to minimize failure in the Lost Prospects segment.

Keywords: Data Mining, Customer Segmentation, K-Means Clustering, Decision Tree, Marketing Strategy, Construction Industry.

INTRODUCTION

Information technology developments and the availability of big data have fundamentally changed the global business competition landscape (Sáez-Ortuño et al., 2023). The use of Artificial Intelligence (AI) and Data Mining techniques is no longer just an additional option, but a strategic necessity to transform raw data into precise business decisions, especially in predicting purchasing behavior and improving marketing efficiency (Owoade, 2025). One crucial aspect of modern business strategy is the implementation of Customer Relationship Management (CRM) supported by sharp customer segmentation analysis. Research shows that identifying key customers through the Recency, Frequency, and Monetary (RFM) model is far more effective than simply relying on traditional demographic segmentation in allocating marketing resources (Imani et al., 2022).

This gap is clearly reflected in the operational conditions of CV Jowon Solusindo, a construction services company that is currently undergoing a crucial transition from a Micro, Small and Medium Enterprise (MSME) to a Small and Medium Enterprise (SME). In this transformation effort, the company faces serious marketing inefficiencies, where out of an average of 30 prospects acquired, the company is only able to convert 1-2 clients per month (low conversion rate of around 3-7%). This situation results in high customer acquisition costs (CAC) reaching 5% per client. The inability to accurately target potential clients leads to wasted promotional resources on unproductive market segments, thereby hindering the company's targeted revenue growth.

An approach to data mining that is proposed to solve this problem is the integration of the K-Means algorithm for clustering and Decision Tree for classification and business rule extraction. The K-Means algorithm was chosen for its superiority in efficiently grouping RFM numerical data to find customer profiles with similar behaviors, while Decision Tree was used to overcome the weakness of K-Means in terms of interpretation by generating transparent If-Then rules that are easy for business decision makers to understand (Budilaksono et al., 2021). (Mohamad Anas

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/10.47709/cnahpc.v8i2.7942

Submitted :2026-01-31
Accepted : 2026-02-05
Published : 2026-04-30

Sobarnas & Iskandar, 2020).

This study was designed to fill a gap in the literature by applying the Hybrid K-Means and Decision Tree methods specifically to construction project (B2B) transactional data, which has different characteristics from the high-frequency retail data that dominated previous studies. Unlike retail research that focuses on daily transactions, this study adapts segmentation variables to the construction context, which has low transaction frequency but very high monetary value (Shirole et al., 2021)(Hartanto et al., 2025). Through an in-depth analysis of CV Jowon Solusindo's historical data, this study aims to produce a segmentation model and decision tree that can improve market targeting effectiveness, increase conversion rates, and serve as a strategic foundation for the company to achieve its business growth targets.

To solve this problem, we propose a data mining approach that integrates the K-Means algorithm for clustering and Decision Tree for classification and business rule extraction. The K-Means algorithm was chosen for its ability to group numerical data on business growth during its transition to a well-established SME.

Based on the background described above, it was identified that the main problems faced by CV Jowon Solusindo were low sales conversion rates and high customer acquisition costs due to a marketing strategy that was not data-driven (blind marketing). The absence of a systematic method for mapping client characteristics made it difficult for the company to distinguish between potential and non-potential prospects. Data and Attribute Scope (RFM Model)

The data analyzed was sourced from CV Jowon Solusindo's internal transaction history over a 10-year period (January 2014 – December 2024). The analysis focused on transaction behavior variables using the RFM (Recency, Frequency, Monetary) model as the main attribute in cluster formation. External macroeconomic variables (such as inflation, interest rates) or competitor data were not included in this model, as the research focused on optimizing internal data for customer retention strategies, as an approach validated as effective for B2B segmentation (Imani et al., 2022).

LITERATURE REVIEW

Customer Relationship Management (CRM) and Business Metrics

In the dynamics of modern industry competition, the marketing paradigm has fundamentally shifted from product-centric to customer-centric. Customer Relationship Management (CRM) is no longer defined merely as software for recording data, but as a comprehensive business strategy that combines processes, people, and technology to attract sales prospects, convert them into customers, and retain existing customers (Maghfirah et al., 2015).

Segmentation, Targeting, and Positioning (STP)

At the core of CRM-based marketing strategy is the concept of STP (Segmenting, Targeting, Positioning). Market segmentation is the process of grouping a heterogeneous market into homogeneous market units, where each group has similar needs, characteristics, or behaviors (Shirole et al., 2021).

Imani et al. (2022): This study was used to validate that identifying key customers through the RFM model is far more effective than traditional demographic segmentation in allocating marketing resources. Dawane et al. (2021) and Sarkar et al. (2023) also confirm that RFM attributes have a strong correlation in identifying customer lifetime value. Utilization of Artificial Intelligence (AI) and Data Mining for Business Strategy: Sáez-Ortuño et al. (2023): Mention that the development of information technology and the availability of big data have changed the competitive landscape, making the use of AI and Data Mining techniques a strategic necessity for precise business decisions. Owoade (2025): Used to emphasize the importance of AI in predicting purchasing behavior and improving marketing efficiency.

Integration of Hybrid Algorithms (K-Means and Decision Tree):

Budilaksono et al. (2021) & Mohamad Anas Sobarnas & Iskandar (2020): This literature supports the selection of the K-Means and Decision Tree hybrid algorithm. K-Means excels in RFM numerical data clustering, while Decision Tree is used to generate transparent and easy-to-understand If-Then rules, overcoming the interpretive weaknesses of K-Means. Rahmadhan & Wasesa (2022): Emphasizes that converting cluster results into If-Then rules (Explainable K-Means Clustering) significantly improves the adoption of analytical results by business users.

Construction Industry Context Data mining is the process of extracting previously unknown but useful information from available data so that it can be used to support decision making. In the context of business, government, and research, data mining is often used for customer behavior analysis, fraud detection, trend prediction, and data classification (Amna, et al., 2023).

Shirole et al. (2021) & Hartanto et al. (2025): This literature was used to identify research gaps, namely the dominance of studies in the retail sector (high frequency). This study was specifically designed to fill the gap in the application of hybrid methods to transactional data from construction projects (B2B), which have low frequency but very high monetary value.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/10.47709/cnahpc.v8i2.7942

Submitted :2026-01-31
Accepted : 2026-02-05
Published : 2026-04-30

In summary, the literature serves as the basis for justifying:
1. The use of RFM as the primary segmentation variable.
2. The use of K-Means for clustering.
3. The use of Decision Tree for interpretation and extraction of explicit business rules.
4. The novelty of this research in applying these methods to the Construction Services sector (B2B).

State of The Art

To map the novelty of this research, the focus is not on new algorithms, but on the industrial context, hybrid approaches, and high-value data-driven business decision-making. This research conducted a systematic review of 40 relevant scientific articles, covering the last five years of publications (2021-2025) as well as several fundamental classical literature. The review focused on the application of Customer Relationship Management (CRM), RFM analysis, and hybrid algorithms (K-Means and Decision Tree) in various industrial domains Based on the literature reviewed, there is a noticeable trend shifting from simple descriptive analysis to predictive analysis based on machine learning.

METHOD

This study uses a quantitative approach with a computational experiment method. The research framework is structured based on the CRISP-DM standard, which is cyclical and adaptive. This approach was chosen to ensure that the resulting data mining model is not only mathematically accurate but also relevant to the business objectives of CV Jowon Solusindo, namely marketing efficiency and customer segmentation.In general, the research flow integrates two machine learning methods (hybrid method), namely Unsupervised Learning using the K-Means algorithm for the segmentation process, followed by Supervised Learning using the Decision Tree algorithm for decision rule extraction.

The Research The research framework is based on a synthesis of CV Jowon Solusindo's business needs and data mining technical capabilities. The research logic begins with the identification of business problems, namely high acquisition costs and suboptimal utilization of historical data, These problems are then approached using CRM theory and the RFM model as the basis for variables. Raw transaction data is processed through the Knowledge Discovery in Database (KDD) stage, which refers to the CRISP-DM standard. The core process involves combining two algorithms: K-Means Clustering for segmentation and Decision Tree for profiling.in Figure 1

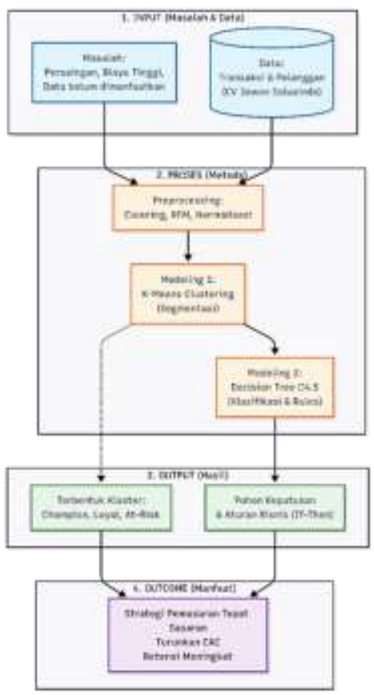

Fig. 1 Framework of Research Flow

Data Collection and Preparation

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/10.47709/cnahpc.v8i2.7942

Submitted :2026-01-31
Accepted : 2026-02-05
Published : 2026-04-30

Research Period
The research was conducted over an effective period of six months, starting from the stage of problem identification, data collection, to the preparation of the final report. The research method used in this paper is a quantitative approach with a computational experiment method.
The following are details of the methodology applied:
Main Approach Quantitative and Computational Experiment: This study focuses on processing historical company data using data mining algorithms to produce analytical models, rather than qualitative field studies.
CRISP-DM Framework: The research process is structured based on the Cross-Industry Standard Process for Data Mining (CRISP-DM) standard to ensure that the resulting models are not only mathematically accurate but also relevant to business objectives (marketing efficiency and segmentation).
Hybrid Algorithm Methodology: This research integrates two different machine learning methods:
Segmentation (Unsupervised Learning) Algorithm: K-Means Clustering. Function: To segment customers based on RFM (Recency, Frequency, Monetary) transaction behavior variables. Optimization: Determination of the optimal number of clusters (k) is limited and evaluated using the Elbow Method and Silhouette Coefficient.
Classification and Rule Extraction (Supervised Learning) Algorithm: Decision Tree (C4.5 or CART). Function: To classify and extract decision rules (rule generation). The goal is to convert cluster results into explicit business rules (If-Then Rules) so that they are easy for management to interpret (explainability).
Research Data and Objects:
Data Focus: Transaction behavior variables using the RFM model as the main attribute. External variables (macroeconomics or competitors) are not included. Computing Tools: The entire computing process is carried out using the Python programming language with the help of the Pandas and Numpy libraries for data processing.
Study Outputs: Outputs are limited to Analytical Models and Strategic Knowledge (in the form of segment profiles and If-Then rules) that are ready for adoption by management, not full CRM software development.

Table 1 shows a wide range of project values (from 3 million to 779 million), indicating the need for segmentation because clients with 3 million projects will certainly behave differently from clients with 700 million projects. Additionally.

Table.1 Research Data Set Descriptive Statistics (N=576)

Atribute	Statistic	value
Demografi	Average	43 age
	Minimum	22 age
	Maximum	64 age
Finansial	Rata-rata Nilai Proyek	Rp 120.388.900,-
	Nilai Terendah	Rp 3.627.688,-
	Nilai Tertinggi	Rp 779.949.500,-
Distribusion	Status Deal (Ya)	391 (67,9%)
	Status Tidak Deal (Tidak)	185 (32,1%)

Classification Model Evaluation (Supervised Evaluation)
For the Decision Tree model, evaluation is carried out by comparing the model's prediction results with the actual data in the Testing Set (20% of the data). The measurement tool used is the Confusion Matrix, which produces four performance metrics:

a. Accuracy: The percentage of total correct predictions (both Deal and No Deal predictions).
b. Precision: The accuracy rate when the model predicts that a client will "Deal." High precision is important for CV Jowon Solusindo to minimize marketing costs wasted on the wrong prospects.
c. Recall (Sensitivity): The model's ability to find all existing potential clients. High recall is important so that the company does not miss out on large project opportunities.
d. F1-Score: The harmonic mean between Precision and Recall, used as a balance if there is an imbalance in the amount of data between classes.

RESULT

Research results the discussion and output obtained from this study—which integrates K-Means Clustering and Decision Tree on 576 historical transaction data from CV Jowon Solusindo have produced optimal customer segmentation and actionable decision rules. Optimization of Customer Segmentation (K-Means Output) , This study validates that three clusters (k=3) are the most optimal and stable configuration for CV Jowon Solusindo customer data, supported by Elbow

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/10.47709/cnahpc.v8i2.7942

Submitted :2026-01-31
Accepted : 2026-02-05
Published : 2026-04-30

Method and Silhouette Score evaluations. The three distinctive clusters that were successfully mapped are in table 2.

Table. 2 . The Three distinctive clusters

Cluster	Segmen	Key Features	Key Strategic Implications
Cluster 0	**Young Emerging Clients**	The average age is 33 years old. They have the highest average transaction value (Rp 127.6 million).	The most promising market segment. Suitable for market nurturing and long-term relationship building strategies due to future growth potential.
Cluster 2	**Established Senior Clients**	The average age is 54 years old. Transaction frequency is stable with an average project value of Rp 120.2 million.	It is worth targeting as a core market for large, long-term projects, as transaction patterns are more consistent.
Cluster 1	**Lost Prospects**	No Deal group with an average age of 42 years old. Lowest average offer value (IDR 113.1 million), indicating high price sensitivity.	Risky segment that requires re-evaluation of the bidding process and business communication.

Decision Rules and Distinguishing Factors (Output Decision Tree)
Decision Tree (C4.5 Algorithm) is used to provide interpretation (explainable profiling) of the clusters formed, with a global model accuracy of 67%.
Dominant Variable: The main result of the Decision Tree shows that the Age variable is found to be the most dominant attribute (root node) in distinguishing the behavior of construction customers.
Effectiveness of Explainable Profiling: This model is highly effective in predicting the Young Clients and Senior Clients segments (with a recall value of 1.00), indicating that demographics (age) are the main determining factors in this B2B construction segmentation.
Model Weaknesses: However, the model's weakness in predicting the Lost Prospects segment confirms that customer conversion failure cannot be explained solely through demographic variables, suggesting the need for variable enrichment for ambiguous segments.
Implications for Data-Driven Marketing Strategies
The integration of these two algorithms provides comprehensive insights for formulating marketing strategies:
Increased Target Effectiveness: Companies can now tailor their promotions (customized marketing strategy) to each segment, for example, targeting Established Senior Clients for large projects.
Decreased CAC: By accurately targeting potential clients, companies can reduce wasteful promotion resources in unproductive market segments, thereby lowering Customer Acquisition Cost (CAC) and supporting the business transition of CV Jowon Solusindo.
Managerial Guidelines: The extracted decision rules (If-Then Rules) serve as objective guidelines for management to predict the potential of new clients, reduce dependence on subjective intuition, and minimize business risks.
Generated Rules
Based on the decision tree visualization, simple logic rules are formed that the marketing team can use to quickly profile new potential customers:

RULE 1 (Young Segment):
IF Age $\leq$ 43.5 Years THEN Predict Entry into the Young Emerging Clients Segment (Cluster 0).

RULE 2 (Senior Segment):
IF Age > 43.5 Years THEN Predict Entry into the Established Senior Clients Segment (Cluster 2).

This rule confirms that the age limit of 43-44 years is a psychological turning point for CV Jowon Solusindo customers in determining their construction needs.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/10.47709/cnahpc.v8i2.7942

Submitted :2026-01-31
Accepted : 2026-02-05
Published : 2026-04-30

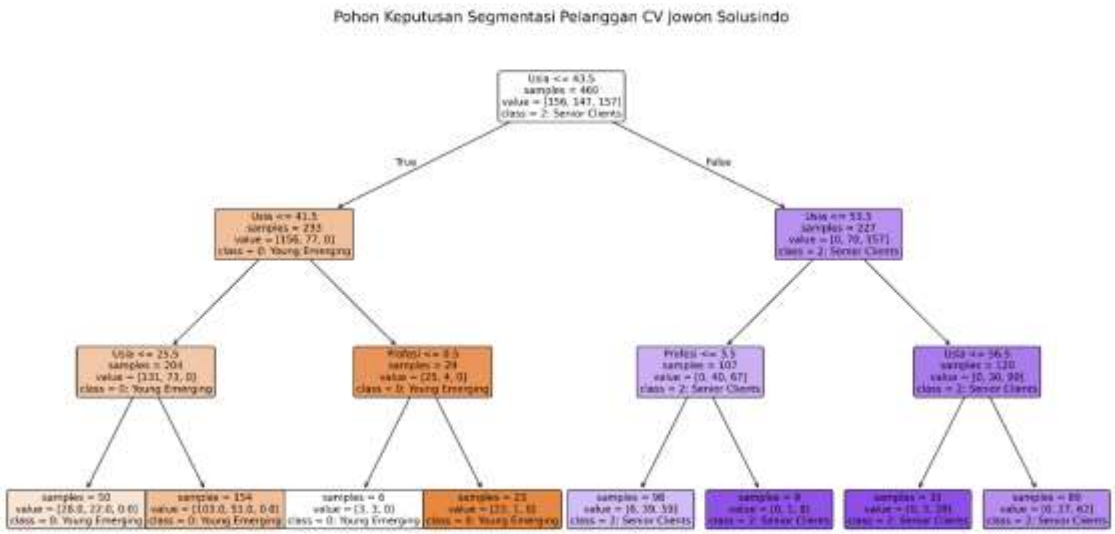

Fig 2 The Customer Segmentation Decision Tree

Figure 2 visualizes how the algorithm breaks down the customer population.

1. Root Node: The top node shows that age ≤ 43.5 years is the primary splitter. This confirms that age is the strongest determinant of construction customer behavior.
2. Left Branch: If age is below 43.5 years, the probability of entering class 0 (Young Clients) is very high (marked with a colored box [specify color in the image]).
3. Right Branch: If age is above 43.5 years, the algorithm performs further checks, but the majority of data leads to class 2 (Senior Clients).Based on the visualization of the CV Jowon Solusindo customer segmentation decision tree, the age variable emerges as the most dominant attribute in distinguishing customer segments.

Discussion of the relationship between methods and managerial implications
Based on the clustering results using the Elbow Method and Silhouette Score, it can be concluded that the formation of three customer clusters is the most optimal and stable configuration for CV Jowon Solusindo data. The Elbow Method shows a clear elbow point in the range of k = 3–4, while the Silhouette Score value at k = 3 is still at a relatively good level and is methodologically acceptable**.**
From the above research results, the integration of the Elbow Method, Silhouette Score, and Decision Tree shows that CV Jowon Solusindo's customer segmentation has formed three optimal clusters with age as the main distinguishing factor, where senior customers are the main high-value target market, young customers have long-term development potential, and the failed prospect segment requires a different non-demographic approach strategy in the future.

DISCUSSIONS

In broad terms, this research flow integrates two machine learning methods (hybrid method), namely Unsupervised Learning using the K-Means algorithm for the segmentation process, followed by Supervised Learning using the Decision Tree algorithm for decision rule extraction. The entire computational process was carried out using the Python programming language with the help of the Pandas and Numpy libraries for data processing. The systematic stages of the research are illustrated in the flowchart in Figure 3.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/10.47709/cnahpc.v8i2.7942

Submitted :2026-01-31
Accepted : 2026-02-05
Published : 2026-04-30

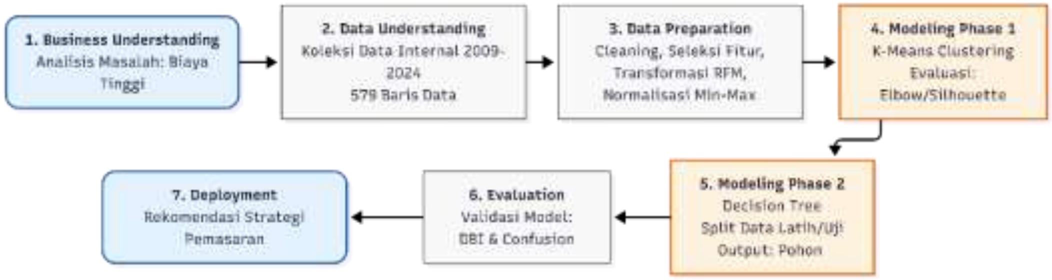

Fig. 3 CRISP-DM Research Flow

Interpretation of Customer Segmentation Results (K-Means RFM)

The clustering results using K-Means validated with k=3 (supported by the Elbow Method and Silhouette Score) prove that there is significant heterogeneity in the behavior of CV Jowon Solusindo customers. The formation of three clusters—Young Emerging Clients, Established Senior Clients, and Lost Prospects—provides a transformative strategic map, changing the company's blind marketing into a data-driven approach.

This finding specifically highlights that in the context of B2B construction, age and monetary transaction value are the main distinguishing variables, unlike the high-frequency retail sector, which often emphasizes Recency and Frequency. The Young Emerging Clients cluster (Cluster 0), despite its youth, has the highest average transaction value, indicating that young clients in this construction sector are willing to take on high-value projects, possibly driven by expansive business needs or innovation. Conversely, Established Senior Clients (Cluster 2) show more stable and consistent transaction patterns, reflecting the nature of stability and more measurable risks in property investment or project development. This segmentation validates the adaptation of the RFM model, which is sensitive to the low-frequency but high-value characteristics of construction data.

However, the Confusion Matrix analysis found limitations in the Lost Prospects cluster (Cluster 1), where the precision value was 0.00. This limitation provides critical insight: conversion failure (prospects that did not result in a deal) at CV Jowon Solusindo was not dominated by demographic factors (age or profession). Instead, it indicates that external factors or non-RFM variables—such as uncompetitive bid prices, unsuitable technical specifications, or external macroeconomic issues—are the causes of transaction failures. This finding is important because it limits the use of Decision Tree solely as an effective profiling tool and suggests the need for variable enrichment in the future, such as including bid price variables or sentiment analysis.

CONCLUSION

Based on the conclusion of the research and limitations found during the analysis process, the author offers the following constructive suggestions:

Suggestions for CV Jowon Solusindo (Managerial), Implementation of Age Rules: The sales team is advised to use the “43.5-year-old age limit” rule as a rule of thumb when dealing with new potential clients. If the client is below that age, offer a modern design package with flexible payment options. If they are above that age, focus communication on quality assurance and after-sales service

Price Structure Evaluation: Considering that the Lost Prospects Segment (Cluster 1) has the lowest average project value, the company is advised to create an “Economy” or “Starter Pack” package to capture this market opportunity so that it does not switch to competitors. And Digital Data Transformation: The company needs to start recording non-demographic data more diligently, such as “Customer Information Source” (from Instagram/Friends/Web) and “Reason for Rejection” (Price/Design). This data is important for improving the accuracy of future predictions.

REFERENCES

Abdul-Rahman, S., Arifin, N. F. K., Hanafiah, M., & Mutalib, S. (2021). Customer Segmentation and Profiling for Life Insurance using K-Modes Clustering and Decision Tree Classifier. International Journal of Advanced Computer Science and Applications, 12(9), 434–444. https://doi.org/10.14569/IJACSA.2021.0120950

Al-Ali, A. S., Haris, R. M., Akbari, Y., Saleh, M., Al-Maadeed, S., & Rajesh Kumar, M. (2025). Integrating binary classification and clustering for multi-class dysarthria severity level classification: a two-stage approach. Cluster Computing, 28(2), 1–19. https://doi.org/10.1007/s10586-024-04748-1

Anitha, P., & Patil, M. M. (2022). RFM model for customer purchase behavior using K-Means algorithm. Journal of

* Corresponding author

Submitted :2026-01-31
Accepted : 2026-02-05
Published : 2026-04-30

King Saud University - Computer and Information Sciences, 34(5), 1785–1792. https://doi.org/10.1016/j.jksuci.2019.12.011

Anitsal, I. (2011). Volume 15, Special Issue, Number 2 ACADEMY OF MARKETING STUDIES JOURNAL. 15(2).

Aulia, S., Muhammad, A. N., & Wibowo, A. (2025). Optimization of Customer Segmentation with RFM, K-Means, and FP-Growth for Marketing Strategy. SINTECH Journal, 8(2), 163–177. https://doi.org/10.31598

B, A. (2018). Comparative Evaluation of SOM-Ward Clustering and Decision Tree for Conducting Customer-Portfolio Analysis. Advances in Multidisciplinary and Scientific Research Journal Publication, 8(1), 1161–128. https://doi.org/10.22624/aims/cisdi/v8n1p11

Budilaksono, S., Jupriyanto, J., Suwarno, M. A. S., Suwartane, I. G. A., Azhari, L., Fauzi, A., Mahpud, M., Mariana, N., & Effendi, M. S. (2021). Customer Profilling for Precision Marketing using RFM Method, K-MEANS algorithm and Decision Tree. SinkrOn, 6(1), 191–200. https://doi.org/10.33395/sinkron.v6i1.11225

Chang, C. C., & Chen, S. H. (2015). A comparative analysis on artificial neural network-based two-stage clustering. Cogent Engineering, 2(1). https://doi.org/10.1080/23311916.2014.995785

Chen, A. H. L., Liang, Y. C., Chang, W. J., Siauw, H. Y., & Minanda, V. (2022). RFM Model and K -Means Clustering Analysis of Transit Traveller Profiles: A Case Study. Journal of Advanced Transportation, 2022. https://doi.org/10.1155/2022/1108105

Chen, D., Sain, S. L., & Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. Journal of Database Marketing and Customer Strategy Management, 19(3), 197–208. https://doi.org/10.1057/dbm.2012.17

Dawane, V., Waghodekar, P., & Pagare, J. (2021). RFM Analysis Using K-Means Clustering to Improve Revenue and Customer Retention. SSRN Electronic Journal, Icsmdi. https://doi.org/10.2139/ssrn.3852887

Deldadehasl, M., Karahroodi, H. H., & Haddadian Nekah, P. (2025). Customer Clustering and Marketing Optimization in Hospitality: A Hybrid Data Mining and Decision-Making Approach from an Emerging Economy. Tourism and Hospitality, 6(2), 1–19. https://doi.org/10.3390/tourhosp6020080

Gustriansyah, R., Suhandi, N., & Antony, F. (2019). Clustering optimization in RFM analysis based on k-means. Indonesian Journal of Electrical Engineering and Computer Science, 18(1), 470–477. https://doi.org/10.11591/ijeecs.v18.i1.pp470-477

Harris, A. F., & Austin, V. (n.d.). Comparative Study of Supervised and Unsupervised Machine Learning Approaches in Banking Applications. Ml.

Hartanto, U. I., Buditjahjanto, I. G. P. A., & Yustanti, W. (2025). Hybrid Clustering and Classification of At-Risk Customer Segments in Network Marketing. Journal of Information Engineering and Educational Technology, 9(1), 42–50. https://doi.org/10.26740/jieet.v9n1.p42-50

Hassan, M. M., & M, T. (2018). Customer Profiling and Segmentation in Retail Banks Using Data Mining Techniques. International Journal of Advanced Research in Computer Science, 9(4), 24–29. http://dx.doi.org/10.26483/ijarcs.v9i4.6172

Imani, A., Abbasi, M., Ahang, F., Ghaffari, H., & Mehdi, M. (2022). 1. RIEJ_Volume 11_Issue 1_Pages 62-76. International Journal of Research in Industrial Engineering, 11(1), 62–76. http://www.riejournal.com/article_138379.html

Kasem, M. S. E., Hamada, M., & Taj-Eddin, I. (2024). Customer profiling, segmentation, and sales prediction using AI in direct marketing. Neural Computing and Applications, 36(9), 4995–5005. https://doi.org/10.1007/s00521-023-09339-6

Lee, C. (2021). Predicting land prices and measuring uncertainty by combining supervised and unsupervised learning. International Journal of Strategic Property Management, 25(2), 169–178. https://doi.org/10.3846/ijspm.2021.14293

Lee, J. W., & Harel, O. (2025). A Two-Stage Classification for Dealing with Unseen Clusters in the Testing Data. Journal of Data Science, 23(1), 188–207. https://doi.org/10.6339/24-JDS1140

Lee, Z. J., Lee, C. Y., Chang, L. Y., & Sano, N. (2021). Clustering and classification based on distributed automatic feature engineering for customer segmentation. Symmetry, 13(9), 1–14. https://doi.org/10.3390/sym13091557

Maghfirah, Adji, T. B., & Setiawan, N. A. (2015). Appropriate Data mining Technique and Algorithm for Using in Analysis of Customer Relationship Management (CRM) in Bank Industry. Seminar Nasional Aplikasi Teknologi Informasi (SNATI), 7–10.

Mohamad Anas Sobarnas, & Iskandar. (2020). Komparasi Akurasi Metode K-Nearest Neighbor Dan C4.5 Untuk Customer Relationship Management Pada Perusahaan Pembiayaan. INFOTECH : Jurnal Informatika & Teknologi, 1(1), 1–14. https://doi.org/10.37373/infotech.v1i1.33

Ogwoka, T. M., Cheruiyot, W., & Okeyo, G. (2015). A Model for Predicting Students' Academic Performance using

* Corresponding author

Submitted :2026-01-31
Accepted : 2026-02-05
Published : 2026-04-30

a Hybrid of K-means and Decision tree Algorithms. International Journal of Computer Applications Technology and Research, 4(9), 693–697. https://doi.org/10.7753/ijcatr0409.1009

Owoade, A. A. (2025). Development of customer segmentation system using supervised and unsupervised machine learning algorithms. Science World Journal, 20(2), 554–565. https://doi.org/10.4314/swj.v20i2.16

Palaniappan, S., Mustapha, A., Foozy, C. F. M., & Atan, R. (2017). Customer profiling using classification approach for bank telemarketing. International Journal on Informatics Visualization, 1(4–2), 214–217. https://doi.org/10.30630/joiv.1.4-2.68

Palma, G. R., Skoczeń, M., & Maguire, P. (2024). Combining supervised and unsupervised learning methods to predict financial market movements. 1–22. http://arxiv.org/abs/2409.03762

Park, H. G., Shin, K. S., & Kim, J. C. (2025). Efficient Clustering Method for Graph Images Using Two-Stage Clustering Technique. Electronics (Switzerland), 14(6), 1–27. https://doi.org/10.3390/electronics14061232

Perdana, S. H. (2023). Customer Relationship Management. EAI/Springer Innovations in Communication and Computing, Part F1354, 237–264. https://doi.org/10.1007/978-3-031-39626-7_10

Pertiwi, N. K. P. A., Sutramiani, N. P., & Wibawa, K. S. (2025). Design and Development of Customer Relationship Management in a Construction Company. INOVTEK Polbeng - Seri Informatika, 10(2), 659–669. https://doi.org/10.35314/qhkc3j28

Rahmadhan, R., & Wasesa, M. (2022). Segmentation using Customers Lifetime Value: Hybrid K-means Clustering and Analytic Hierarchy Process. Journal of Information Systems Engineering and Business Intelligence, 8(2), 130–141. https://doi.org/10.20473/jisebi.8.2.130-141

Sáez-Ortuño, L., Huertas-Garcia, R., Forgas-Coll, S., & Puertas-Prats, E. (2023). How can entrepreneurs improve digital market segmentation? A comparative analysis of supervised and unsupervised learning algorithms. International Entrepreneurship and Management Journal, 19(4), 1893–1920. https://doi.org/10.1007/s11365-023-00882-1

Saputra, A. D., & Yustanti, W. (2025). A Hybrid Clustering – Classification Framework for SMEs Success Level Prediction. 9(2), 89–100.

Sarkar, M., Puja, A. R., & Chowdhury, F. R. (2023). Journal of Business and Management Studies Determinants of Employee Retention in Pharmaceutical Companies: Case of Saudi Arabia. Determinants of Employee Retention in Pharmaceutical Company: Case of Saudi Arabia, 2709–0876, 8–22. https://doi.org/10.32996/jbms

Shirole, R., Salokhe, L., & Jadhav, S. (2021). Customer Segmentation using RFM Model and K-Means Clustering. International Journal of Scientific Research in Science and Technology, 591–597. https://doi.org/10.32628/ijsrst2183118

Ufeli, C. P., Sattar, M. U., Hasan, R., & Mahmood, S. (2025). Enhancing Customer Segmentation Through Factor Analysis of Mixed Data (FAMD)-Based Approach Using K-Means and Hierarchical Clustering Algorithms. Information (Switzerland), 16(6), 1–25. https://doi.org/10.3390/info16060441

Widiputra, H., Kisworo, M., Novita, A., & Pardede, T. (2015). Two-stage graph-clustering algorithm and localised classification model to identify apt business locale. International Journal of Business Information Systems, 20(2), 195–218. https://doi.org/10.1504/IJBIS.2015.071537

Zhang, Y., & Fan, Y. (2025). POSTER PRESENTATION Characterizing heterogeneous atrophy patterns of Alzheimer ' s disease dementia with mixture-of-experts (MOE) for joint classification and clustering. 21, 2–4. https://doi.org/10.1002/alz70856

* Corresponding author