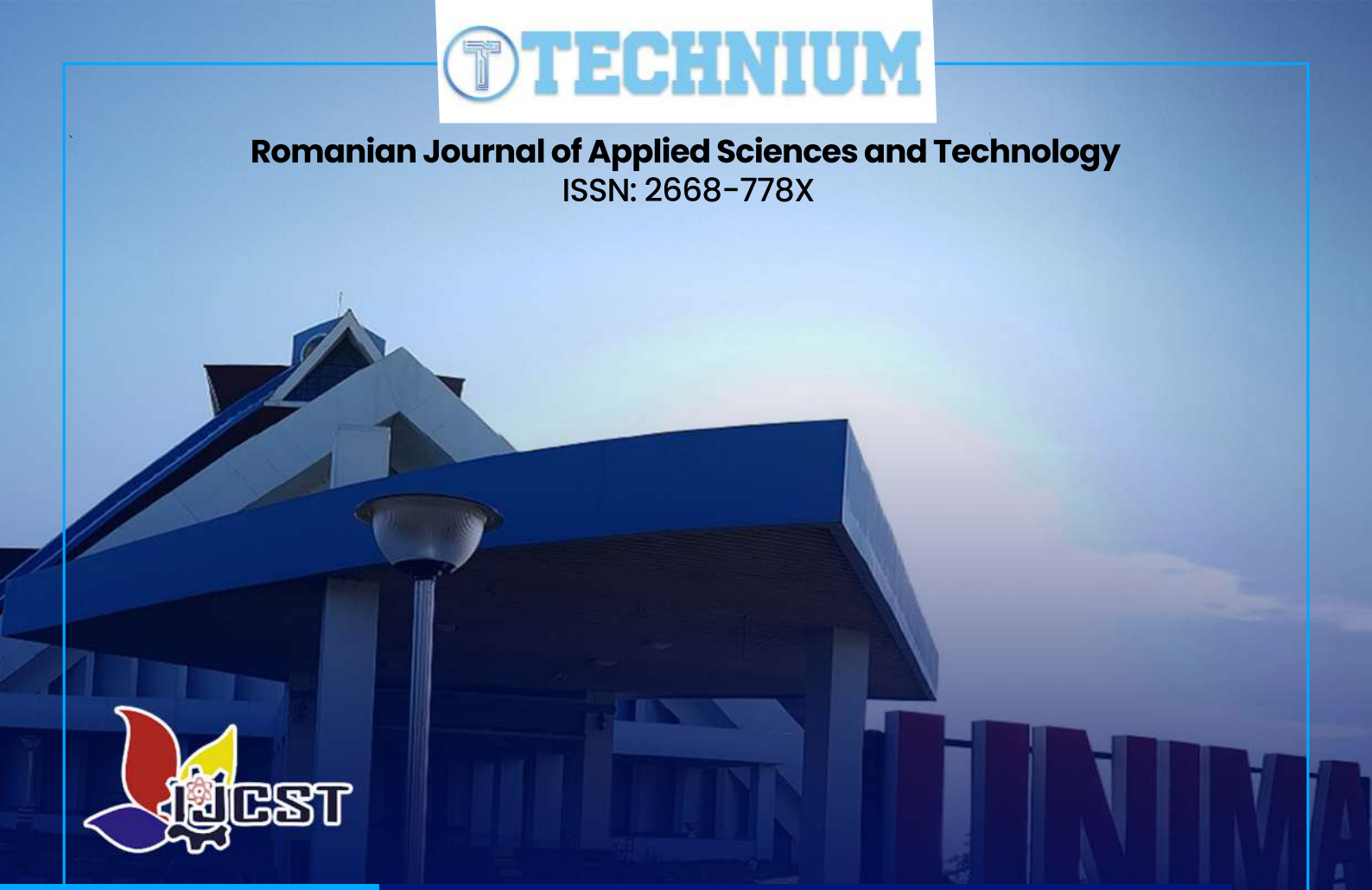
TECHNIUM

Romanian Journal of Applied Sciences and Technology
ISSN: 2668-778X



vol. 16 / 2023

The 7th International Conference on Science Technology

organized by
Faculty of Social Science and
Law Universitas Negeri Manado and
Consortium of International Conference
on Science and Technology

The Innovation Breakthrough in Digital and Disruptive Era

Technium: Romanian Journal of
Applied Sciences and Technology
Vol. 16, 176-182, October, 2023
ISSN: 2668-778X
www.techniumscience.com

Sales Product Clustering Using RFM Calculation Model And K-Means Algorithm on Primskystore

Agussalim[1], *Rahayu Kusumaningtyas Paramita Wardhani*[2], and *Amalia Anjani* [3]

[1] Master of Information Technology Department, Faculty of Computer Science, UPN Veteran Jawa Timur, Indonesia
[2,3]Information Systems Department, Faculty of Computer Science, UPN Veteran Jawa Timur, Indonesia

Abstract. The optimization of data processing procedures can yield high-quality information that is then utilized by business owners in the decision-making process. An effective company plan, particularly in the realm of promotion, holds significant importance for store proprietors in order to foster sustained business expansion. Primskystore is an e-commerce platform operating inside the retail industry, where the efficacy of product sales promotion is perceived to be suboptimal. Sales product advertising continues to revolve around a single sort of product. The objective of this work is to develop a data mining model through the implementation of a web-based application that utilizes the K-means clustering approach and the RFM model. The utilization of the K-Means clustering approach and the RFM model can facilitate the clustering of sales products at retailers. The data mining application employed on this website utilizes the Python programming language, MySQL as the database, and the Unified Modeling Language for system model creation. The findings of this research encompass the development of a web-based data mining tool designed to present the outcomes of product sales clustering within retail establishments. The system has the capability to present the outcomes of the RFM calculation model and KMeans clustering. Specifically, it reveals the presence of three distinct clusters, with cluster 0 accounting for 30% of the data, cluster 1 representing 7.5% of the data, and cluster 2 encompassing 62.5% of the data. This web-based data mining program has the potential to assist retailers in determining an optimal promotional business strategy.

[1] Corresponding author: rahayuwrdhn@gmail.com

TECHNIUM

Technium: Romanian Journal of Applied Sciences and Technology
Vol. 16, 176-182, October, 2023
ISSN: 2668-778X
www.techniumscience.com

1 Introduction

The expeditious advancement of information technology has yielded favorable outcomes for organizations in terms of their capacity to use data for the enhancement of performance and efficacy. Presently, e-commerce platforms have emerged as a burgeoning sector of the corporate landscape, experiencing exponential growth, particularly in the context of the ongoing global pandemic. Primskystore is an emerging e-commerce platform that exhibits substantial growth, offering a diverse range of fashion and cosmetic merchandise. The existing sales transaction database of Primskystore possesses considerable size; however, its utilization for the purpose of enhancing product sales has not been maximized. One potential approach for enhancing product sales involves the categorization of products based on consumer attributes [1-2].

The implementation of product grouping has been found to be a highly effective method in enhancing product sales within the context of online retailers. The clustering method is a data mining approach that is employed for the purpose of categorizing things into groups [3]. The utilization of clustering enables the grouping of products according to consumer attributes, encompassing shopping behavior, product preferences, and purchase habits. One of the often employed clustering algorithms in data analysis is K-Means, which partitions the dataset into many groups by utilizing the Euclidean distance metric [4].

Clustering is a widely used technique in data analysis that aims to locate groupings of data points that exhibit similar characteristics or patterns [5]. The K-Means algorithm is an iterative clustering technique that seeks to identify local maxima in each iteration, operating through a series of five distinct steps [6]. K-Means The clustering process involves the random initialization of centroids, followed by the assignment of data points to clusters based on their respective distances to the centroid. Subsequently, the centroid undergoes an update through the computation of the mean value of the data points within the corresponding cluster. This iterative procedure continues until the centroid remains unchanged or the predetermined number of iterations is reached.

The acronym RFM represents the three key metrics used in customer segmentation: Recency, Frequency, and Monetary value. The RFM calculation model is frequently employed in the field of business analytics to categorize products into various segments, including high value, medium value, and low value, among others [7]. The RFM approach operates by allocating weights to individual indicators, namely Recency, Frequency, and Monetary, and subsequently use these weight values to categorize consumers into many segments.

The Elbow technique offers a methodology for selecting an optimal cluster value, hence enhancing the value of the resulting cluster as a data model for finding the most suitable cluster. The Elbow technique offers a methodology for selecting an optimal cluster value, hence enhancing the value of the resulting cluster as a data model for finding the most suitable cluster [8]. The Silhouette coefficient is a widely employed technique for assessing the efficacy and integrity of a cluster, specifically in terms of the appropriateness of an object's placement within said cluster [9]. The Davies Bouldin Index (DBI) is a cluster evaluation or validation technique employed in clustering algorithms. It relies on the cohesion and separation matrices to determine the quality of the resulting clusters [10].

This project aims to employ a data mining model that utilizes the K-Means algorithm clustering approach and the RFM (Recency, Frequency, and Monetary) model to categorize products available at Primskystore. The RFM model is employed for assessing customer shopping behavior by considering three primary factors: recency, frequency, and monetary value. Recency refers to the most recent purchase of a product, frequency pertains to the number of products sold, and monetary value denotes the sales value associated with each product. The anticipated outcomes of this research endeavor are poised to assist Primskystore in ascertaining an optimal product promotion approach, hence fostering an augmentation in product sales.

2 Methods

1.1 Literature study

The literature review in this study was derived from a comprehensive selection of relevant prior papers and references sourced from reputable journals, books, and websites pertaining to clustering techniques employing the K-Means algorithm. The subsequent research titles have been utilized as references in this study: The RFM ranking method has been identified as an effective approach for customer segmentation [11]. In a case study conducted at PT. Herbal Penawar Alwahidah Indonesia Pekanbaru, the implementation of K-Means clustering based on RFM Mofek was found to be a valuable tool for mapping and supporting customer management strategy [12]. Additionally, the K-Means method was utilized for product RFM analysis, which involves evaluating recency, frequency, and monetary aspects [13]. Another study focused on data mining and employed the K-Means clustering algorithm for product grouping [14]. These are just a few examples of the research conducted in this area.

1.2 Interview and requirement analysis

The interview took place on Saturday, July 2, 2022, with Prima, the proprietor of Primskystore, located at Pasadena Puri Surya Jaya D-11 Housing Complex in Sidoarjo, East Java. The produced findings encompass data pertaining to the flow process of the sales

Technium: Romanian Journal of Applied Sciences and Technology
Vol. 16, 176-182, October, 2023
ISSN: 2668-778X
www.techniumscience.com

information system within the store. Specifically, these comprise the store's business processes, sales product data, sales transaction data, and information regarding store promotions.
A needs analysis is conducted in order to identify and propose solutions to the issues that have been presented. The analysis of the system implemented at Primskystore store revealed issues in the post-order or transaction phase of the business process. Specifically, there was a lack of follow-up from the store regarding the product acquisition process and the maintenance of sales levels. Consequently, a software application was developed to serve as a tool for Primskystore in streamlining the execution of operational procedures within the establishment.

1.3 Data Gathering

At the data collection stage, it is done by collecting sales transaction data at Primskystore stores. Fig. 1, The transaction data collected for this study is from January 2022 to June 2022 in the form of an excel file with 34 columns.

Table 1. Data Transaction

No	Data transaction		
	Attribute name	***information***	***Data***
1	Number	Transaction Number	16[a]
2	Invoice Number	Invoice Number on Tokopedia	INV/20220506/MPL/2293838637[a]
3	Payment Date	Payment Due	05/01/2022[a]

Sample of a Data based on Excel FIle

1.4 Pra-processing Data

During the data pre-processing phase, the task involves comprehending the significance of each attribute within the dataset. The sales transaction data at the Primskystore store is stored in an Excel file including 34 columns. The initial phase of data pre-processing entails two distinct procedures: attribute selection and data purification.

Attribute Selection

The attribute selection stage involves comprehending each attribute inside the dataset and subsequently choosing the suitable attribute or column for further utilization in research. A careful selection was made from the 34 columns present in the Primskystore sales transaction data excel file, focusing on those that were deemed relevant and significant. Following the completion of the selection process, a total of 15 columns were identified as suitable and significant for further analysis. Conversely, 19 columns from the Primskystore sales transaction data excel file that were not included in the selection were subsequently removed.

Data Cleansing

Once the attributes have been established, the subsequent step involves the execution of the data cleansing procedure, which is designed to eradicate any instances of null or incomplete data. In order to ascertain the presence of null or partial values inside the dataset.

2 Processing Data

During this stage of data processing, three distinct processes are executed, namely the calculation of the RFM model, the determination of the number of clusters, and the application of K-Means clustering [15].

2.1 Calculation of the RFM model

Currently, the RFM (Recency, Frequency, and Monetary) calculation model is utilized to process store sales transaction data. This is achieved through the implementation of the Flask framework, a Python programming language. Table. 2, The data is organized into product groups based on the RFM framework. Recency is determined by selecting the maximum transaction day. Frequency is calculated by counting the number of invoices. Monetary value is derived by multiplying the number of purchases by the selling price of the product.

Table 2. Testing Scenario

No	Data RFM Model			
	Product Name	***Recency***	***Freq.***	***Monetary***
1	Avoskin Miraculous Acne Solution Micro Serum 20ml[a]	101	7	1305000
2	Avoskin Miraculous Retinol Ampoule 30ml – Full Size[a]	5	131	24349000
3	Avoskin Miraculous Refining Toner 100ml Lilac (Anniversary Edition)[a]	1	12	1675000

Once the data has been subjected to the RFM model, the subsequent step involves performing data normalization due to the varying range of attribute values.

Table 3. Normalisasi RFM

No	Data RFM Model			
	Product Name	***Recency***	***Freq.***	***Monetary***
1	Avoskin Miraculous Acne Solution Micro Serum 20ml[a]	0,561	0,017	0,033
2	Avoskin Miraculous Retinol Ampoule 30ml – Full Size[a]	0,028	0,365	0,617
3	Avoskin Miraculous Refining Toner 100ml Lilac (Anniversary Edition)[a]	0,006	0,031	0,042

[a.] Sample of a Data based on the Normalisasi RFM

Technium: Romanian Journal of
Applied Sciences and Technology
Vol. 16, 176-182, October, 2023
ISSN: 2668-778X
www.techniumscience.com

2.2 Determine Clustering Number

In this phase, the determination of the number of clusters is conducted based on SATA transactions utilizing the elbow method and silhouette coefficient within the Python programming language. The process of determining the optimal number of clusters is undertaken in order to generate K values that are better suited to the dataset. Initially, it is important to ascertain the number of clusters by employing the elbow approach, as depicted in Figure 1.

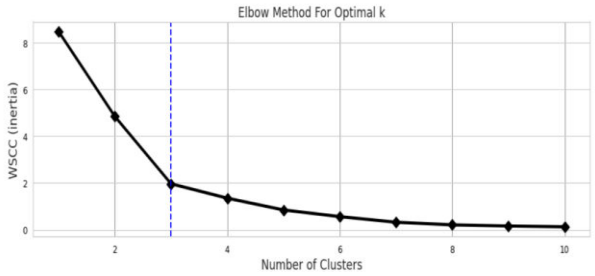


Fig. 1. Elbow Method

Based on the use of the elbow technique to determine the optimal value of K for the analyzed transaction data, it has been determined that K = 3 (representing the number of clusters) is deemed to be the most suitable or appropriate choice. Subsequently, ascertain the appropriate number of clusters for the given data by employing the second method, specifically the silhouette coefficient as depicted in Figure 2.

```
Number of clusters from 2 to 9:
 [2, 3, 4, 5, 6, 7, 8, 9]
For n_clusters = 2, silhouette score is 0.6535797282296946)
For n_clusters = 3, silhouette score is 0.6003190162515635)
For n_clusters = 4, silhouette score is 0.5704160333909495)
For n_clusters = 5, silhouette score is 0.5826003423820874)
For n_clusters = 6, silhouette score is 0.5985321718956056)
For n_clusters = 7, silhouette score is 0.6351842157474396)
For n_clusters = 8, silhouette score is 0.5806774993056509)
For n_clusters = 9, silhouette score is 0.5855108580344422)
```

Fig. 2. Silhouette Coeeficient Method

Based on the application of the silhouette coefficient method to determine the optimal number of clusters for the analyzed transaction data, the findings indicate that k = 3 (representing three clusters) yields a coefficient value of 0.60, while k = 8 (representing eight clusters) yields a coefficient value of 0.58, indicating a decrease of 0.05 compared to the preceding cluster. Consequently, the numbers of clusters 3 and 8 are deemed the most suitable or appropriate choices. The number of clusters determined using the elbow approach is k = 3. Consequently, the silhouette coefficient is computed for k = 3 clusters. The findings derived from the analysis of cluster determination utilizing two distinct methodologies, namely the elbow technique and the silhouette coefficient approach, indicate that the best number of clusters is three (k=3).

Clustering K-Means

After finding the value of K, then proceed with the clustering process using the k-means algorithm to be able to group products from sales transaction data at Primskystore according to the similarity of characteristics that are owned between each data in the cluster. Clustering is done using the k-means algorithm using the Python library Kmeans programming language.

```
kmeans_model = KMeans(n_clusters=3, init='k-means++', max_iter=300)
groups = kmeans_model.fit_predict(rfm_df_raw)
rfm_df_cutted['groups'] = groups
daf['Cluster'] = groups

[238] daf.head(10)
```

	Nama_Produk	Recency	Frequency	Monetary	Cluster
0	AVOSKIN MIRACULOUS ACNE SOLUTION MICRO SERUM 20ML	0.561111	0.016854	0.032869	2
1	AVOSKIN MIRACULOUS ACNE SOLUTION SPOT TREATMEN...	0.527778	0.022472	0.039083	2
2	AVOSKIN MIRACULOUS RETINOL AMPOULE 30ML - FULL...	0.027778	0.365169	0.617312	1
3	AVOSKIN MIRACULOUS RETINOL TONER 100ML - FULL ...	0.261111	0.016854	0.022851	0
4	AVOSKIN Miraculous Refining Serum 30ml FULL	0.016667	0.022472	0.036293	0
5	AVOSKIN Miraculous Refining Serum 30ml LILAC (...	0.144444	0.005618	0.012326	0
6	AVOSKIN Miraculous Refining Toner 100ml LILAC ...	0.005556	0.030899	0.042253	0
7	AVOSKIN Miraculous Retinol Ampoule 30ml LILAC ...	0.016667	0.025281	0.047199	0
8	AVOSKIN Miraculous Retinol Toner 100ml LILAC (...	0.055556	0.005618	0.010728	0
9	AVOSKIN NAUFA PURE OLIVE OIL BAR SOAP 50gr	0.105556	0.008427	0.004337	0

Fig. 3. Clustering K-Means

The procedural steps for clustering using the K-Means technique are illustrated in Figure 3. At Primskystore, the products are categorized into clusters based on sales transaction data, with the total number of clusters set at three. The centroid results for each cluster are presented. Cluster 0 comprises products that have a high recency value, while their frequency and monetary values are relatively low. The cluster comprises a total of 24 goods, which accounts for 30 percent (30%) of the whole sales transaction data. Cluster 1 comprises products that exhibit a low recency value alongside high frequency and monetary values. The cluster comprises either six products or 7.5 percent (7.5%) of the overall sales transaction data. Cluster 2 comprises products that exhibit intermediate levels of recency, frequency, and monetary values. The cluster under consideration comprises 50 goods, which accounts for 62.5 percent (62.5%) of the overall sales transaction data. The dominant cluster is being considered.

Based on the clustering results acquired, various recommendations can be derived that can be implemented by the store in its future promotional business strategy. For the products categorized under cluster 0, it is possible to implement a business plan that involves offering discounts on purchases above a specific threshold. In order to enhance consumer interest in products belonging to cluster 1, a viable business tactic would involve implementing a plan such as conducting flash sales or offering temporary discounts. In the case of products included inside cluster 2, a viable business strategy could involve implementing additional product promotions or introducing product bundling at a more competitive price point. These measures aim to enhance buyer engagement and stimulate interest in the respective products.

Technium: Romanian Journal of
Applied Sciences and Technology
Vol. 16, 176-182, October, 2023
ISSN: 2668-778X
www.techniumscience.com

2.3 Cluster Evaluation

Once the data processing stage has concluded, it becomes imperative to conduct testing in order to verify the validity of the required number of clusters. The study employed cluster testing, utilizing the Davies Bouldin Index, which was implemented in the Python programming language through the utilization of the davies_bouldin_index library function.

```
[141] from sklearn.metrics import davies_bouldin_score

[136] kmeans = KMeans(n_clusters=3, random_state=30)
      labels = kmeans.fit_predict(rfm_df_raw)

[137] db_index = davies_bouldin_score(rfm_df_raw, labels)
      print(db_index)

      0.4996054468482738
```

Fig. 4. Davies Bouldin Index (DBI)

Based on the calculations performed using the Davies-Bouldin Index (DBI) as shown in Figure 4, it can be concluded that the chosen number of clusters (k=3) is deemed appropriate. This conclusion is supported by the resulting DBI value of 0.49, which is non-negative and meets the criterion of being greater than or equal to zero.

2.4 System Development

Upon the conclusion of the cluster testing phase, the subsequent task involves the development of the database for the application. This process entails the utilization of the Conceptual Data Model (CDM) and the Physical Data Model (PDM) [16-17].

The Figure 5 illustrates the Conceptual Data Model (CDM) of the Primskystore store system, which serves as the foundational database structure for data storage in this research endeavor.

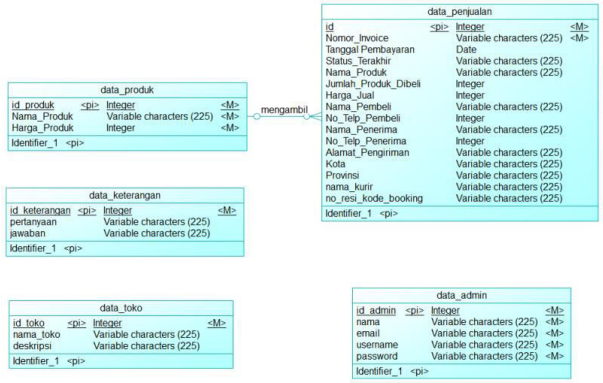


Fig. 5. Conceptual Data Model

Figure 6 illustrates the Physical Data Model (PDM) of the Primskystore store system, which serves as the database structure for data storage in this research endeavor.

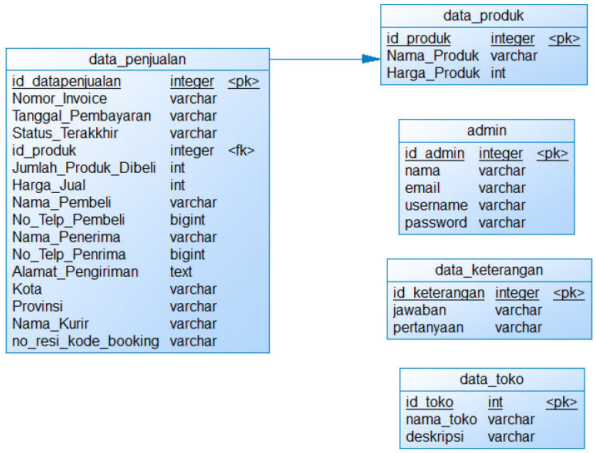


Fig. 6. Physical Data Model

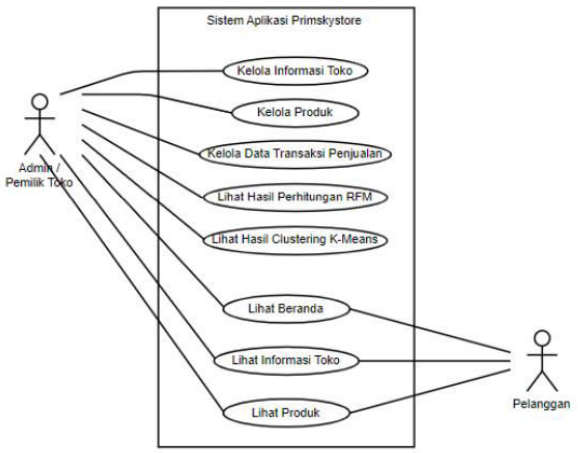


Fig. 7. Use Case Diagram

The subsequent steps involved in constructing a use case diagram entail a reciprocal exchange or discourse between the system and relevant actors, accompanied by the system's corresponding actions. Figure 7 depicts a use case diagram comprising two distinct players, namely the administrator or store owner, and the user. The user interface encompasses three distinct use cases: accessing the homepage, browsing products, and retrieving shop information. The administrative or shop ownership role encompasses eight distinct use cases: store information management, product management, sales transaction data management, viewing RFM calculation results, viewing K-Means clustering results, accessing the homepage, viewing store information, and viewing products.

2.5 Application Testing

The study phase was extended by implementing an application system utilizing the Python programming language within the Visual Studio Code application, aligning with the previously conducted system design. The subsequent section presents the construction of the display interface for the website-based product sales grouping application developed by Primskystore.

Technium: Romanian Journal of Applied Sciences and Technology
Vol. 16, 176-182, October, 2023
ISSN: 2668-778X
www.techniumscience.com

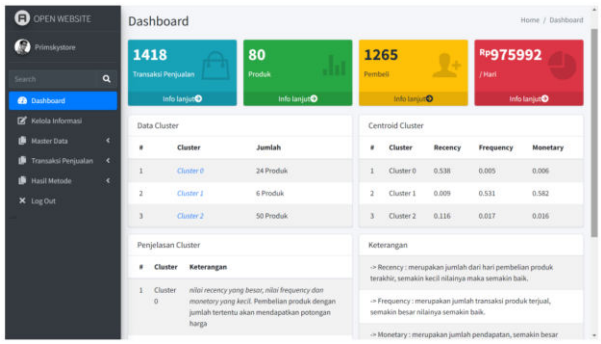


Fig. 8. Dashboard Systems

The Dashboard page view showcases the visual representation of the k-means clustering application and the RFM calculation methodology, exemplified by Figure 8. The displayed dashboard page, as depicted in the image, serves as a central hub or primary interface for presenting various information pertaining to sales transactions, product inventory, customer activity, and daily revenue at the Primskystore establishment. This includes the quantity of sales transactions, the number of products available in the store, the count of customers or buyers who have engaged in transactions, and the daily revenue generated. Additionally, the dashboard provides insights into the number of products within each cluster and their respective averages, accompanied by explanatory details for each cluster. The data presented on the dashboard page is organized in a box and table format. By clicking on the "More info" link within each box, users can navigate directly to the corresponding page. Similarly, clicking on a cluster within the cluster table will redirect users to the product information page associated with that particular cluster.

The study conducted testing of the website-based application system utilizing the blackbox testing methodology. The subsequent section presents the outcomes and rationales of the blackbox testing that was conducted. Figure 4 illustrates the implementation of blackbox testing, which serves the purpose of validating all functionalities inside the dashboard menu, in accordance with the pre-established system design.

Table 4. Blackbox testing

No	*Scenario*	*Result*	*Test Result*	*Summary*
1	The user is able to access and view the dashboard page.	The system has the capability to present the dashboard page view.	The system successfully displays the dashboard page	Valid
2	The total sales column can be programmed to generate show sales transaction data pages upon clicking.	The system has the capability to present sales data page.	The system successfully displays the sales page	Valid
3	The product data page can be displayed upon clicking the whole product icon.	The system has the capability to present sales data.	The system successfully displays the product page.	Valid

[a]Sample of a Data based on the Dashboard Blackbox Testing

3 Conclusion

The study presents a website-based application designed for the Primskystore store. This system enables the store to access a summary of sales, manually record sales, view sales transaction data, product data, and customer data. Additionally, it provides the functionality to calculate RFM (Recency, Frequency, and Monetary) metrics and perform product clustering analysis. These features assist the store in formulating an effective business strategy. The number of clusters is determined using two methods: the elbow method, which yields a result of k = 3 clusters, and the silhouette coefficient approach, which also determines k = 3 clusters with a value of 0.6. The findings from both of these methods indicate that the ideal number of clusters is three (k=3). In addition, the validity of the chosen number of clusters (k = 3) was assessed using the Davies Bouldin index. The resulting value of 0.49 indicates that the selected number of clusters is considered valid or satisfactory. This determination is based on the fact that the calculated DBI value is close to zero, as the index is non-negative and values closer to zero indicate better clustering performance.

The subsequent findings pertain to the sales product clustering procedure conducted at retail establishments, employing RFM (Recency, Frequency, and Monetary) computations and K-Means clustering. Cluster 0 comprises products that exhibit a high recency value and low frequency and monetary values when grouped together. Cluster 1 comprises products that exhibit a low recency value, yet possess high frequency and monetary values. Cluster 2 comprises products that exhibit intermediate levels of recency, frequency, and monetary values. Recommendations can be provided to facilitate the transformation of a static website into a dynamic one, enabling online buying and selling transactions through the store's own platform, provided that Primskystore possesses adequate human resources. There is an expectation that in subsequent developments, the data obtained through clustering results can be automatically and immediately put in the database.

References

[1] Zhang, M., Luo, M., Nie, R., & Zhang, Y. (2017). Technical attributes, health attribute, consumer attributes and their roles in adoption intention of healthcare wearable technology. International journal of medical informatics, 108, 97-109.

[2] Chen, K. K., & Zhang, J. J. (2011). Examining consumer attributes associated with collegiate athletic facility naming rights sponsorship: Development of a theoretical framework. Sport Management Review, 14(2), 103-116.

[3] Madhulatha, T. S. (2012). An overview on clustering methods. arXiv preprint arXiv:1205.1117.

[4] Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. Electronics, 9(8), 1295.Dhio Saputra. 2020. "Goods Stock Management Using the K-Means Algorithm Method." Jurnal Teknologi 10(1): 22–4.

[5] Imron, Mohammad, Uswatun Hasanah, and Bahrul Humaidi. (2020). "Analysis of Data Mining Using K-Means Clustering Algorithm for Product Grouping." IJIIS: International Journal of Informatics and Information Systems 3(1): 12–22.

[6] Gustriansyah, Rendra, Nazori Suhandi, and Fery Antony. (2019). "Clustering Optimization in RFM Analysis Based on K-Means." Indonesian Journal of Electrical Engineering and Computer Science 18(1): 470–77.

[7] Rahman, Aulia Tegar, Wiranto, and Anggrainingsih Rini. (2017). "Coal Trade Data Clustering Using K-Means (Case Study Pt. Global Bangkit Utama)." ITSMART: Jurnal Teknologi dan Informasi 6(1): 24–31.

[8] Hidayati, Rahmatina, Anis Zubair, Aditya Hidayat Pratama, and Luthfi Indana. (2021). "Analisis Silhouette Coefficient Pada 6 Perhitungan Jarak K-Means Clustering." Techno.com 20(2): 186–97.

[9] Hablum, Rofika, Amal Khairan, and Rosihan Rosihan. (2019). "Clustering Hasil Tangkap Ikan Di Pelabuhan Perikanan Nusantara (Ppn) Ternate Menggunakan Algoritma K-Means." JIKO (Jurnal Informatika dan Komputer) 2(1): 26–33.

[10] Yorozu, T., Hirano, M., Oka, K., & Tagawa, Y. (1987). Electron spectroscopy studies on magneto-optical media and plastic substrate interface. IEEE translation journal on magnetics in Japan, 2(8), 740-741.

[11] Young, M. (2002). Technical writer's handbook. Sausalito: University Science Books.

[12] Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2021). RFM ranking–An effective approach to customer segmentation. Journal of King Saud University-Computer and Information Sciences, 33(10), 1251-1257.

[13] Hadi, F., Mustakim, M., Rahmadia, D. O., Nugraha, F. H., Bulan, N. P., & Monalisa, S. (2017). Penerapan K-Means Clustering Berdasarkan RFM Mofek Sebagai Pemetaan dan Pendukung Strategi Pengelolaan Pelanggan (Studi Kasus: PT. Herbal Penawar Alwahidah Indonesia Pekanbaru). SITEKIN: Jurnal Sains, Teknologi dan Industri, 15(1), 69-76.

[14] Firmansah, R. Y., Irawan, J. D., & Vendyansyah, N. (2021). Analisis Rfm (Recency, Frequency And Monetary) Produk Menggunakan Metode K-Means. JATI (Jurnal Mahasiswa Teknik Informatika), 5(1), 334-341.

[15] Anitha, P., & Patil, M. M. (2022). RFM model for customer purchase behavior using K-Means algorithm. Journal of King Saud University-Computer and Information Sciences, 34(5), 1785-1792.

[16] Paulheim, H. (2019). Evaluating ontology matchers on real-world financial services data models.

[17] Mukaromah, S., Pratama, A., Ithriah, S. A., & Putra, A. B. (2020, July). Analysis and design student entrepreneurship information system. In Journal of Physics: Conference Series (Vol. 1569, No. 2, p. 022045). IOP Publishing.