DEÜ FMD 25(74), 491-503, 2023

Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi
Dokuz Eylul University Faculty of Engineering Journal of Science and Engineering

Basılı/Printed ISSN: 1302-9304. Elektronik/Online ISSN: 2547-958X

Customer Segmentation Using K-Means Clustering Algorithm and RFM Model

K-Means Kümeleme Algoritması ve RFM Modeli Kullanarak Müşteri Segmentasyonu

Gözde Aslantaş [1*], Mustafacan Gençgül [1], Merve Rumelli [1], Mustafa Özsaraç [1], Gözde Bakırlı [1]

[1] VESTEL Electronics, TÜRKİYE
Sorumlu Yazar / Corresponding Author *: gozde.aslantas@vestel.com.tr

Geliş Tarihi / Received: 03.10.2022 *Araştırma Makalesi/Research Article*
Kabul Tarihi / Accepted: 24.11.2022 *DOI:10.21205/deufmd.2023257418*
Atıf şekli/ How to cite: ASLANTAŞ, G., GENÇGÜL, M., RUMELLİ, M., ÖZSARAÇ, M., BAKIRLI, G.(2023). Customer Segmentation Using K-Means Clustering Algorithm and RFM Model. DEUFMD, 25(74), 491-503.

Abstract

The key points in customer segmentation are determining target customer groups and satisfying their needs. Recency-Frequency-Monetary (RFM) analysis and K-Means clustering algorithm are the popular methods for customer segmentation when analyzing customer behavior. In our study, we adapt the K-means clustering algorithm to RFM model by extracting features that represent RFM aspects of home appliances. Customers with similar RFM-oriented features are assigned to the same clusters, while customers with non-similar RFM-oriented features are assigned to different clusters. In the experiments, clustering achieved the determined threshold for Silhouette Score. The resulting clusters were ranked and named by Customer Lifetime Value (CLV) metric, which measures how valuable a customer is to the business.

***Keywords:** Customer Segmentation, RFM Model, K-Means Algorithm, Customer Lifetime Value*

Öz

Hedef müşterinin belirlenmesi ve ihtiyaçlarının karşılanması, müşteri segmentasyonunda önemli noktalardır. Yenilik-Sıklık-Tutar (RFM) Analizi ve K-Means kümeleme algoritması, müşteri davranışını analiz eden müşteri segmentasyonu için kullanılan popüler yöntemlerdir. Çalışmamızda, ev cihazlarının, RFM bileşenlerini temsil edecek şekilde özelliklerini çıkararak K-Means kümeleme algoritmasını RFM modeline uyarladık. Böylece, benzer RFM özelliklerine sahip müşteriler aynı kümelere atanırken, benzer olmayan RFM özelliklerine sahip müşteriler farklı kümelere atanmıştır. Deneylerde, kümeleme çalışmasının, belirlenen Silhouette Skorunu geçerek başarılı olduğu gözlenmiştir. Ortaya çıkan kümeler, bir müşterinin işletme için ne kadar değerli olduğunu ölçen Müşteri Yaşam Boyu Değeri (CLV) metriğine göre sıralanmış ve adlandırılmıştır.

***Anahtar Kelimeler:** Müşteri Segmentasyonu, RFM Model, K-Means Algoritması, Müşteri Yaşam Boyu Değeri*

1. Introduction

With the advent of the age of technology, expectations from home appliances have shifted to being about how smart the appliances are. Developments in this field have led to the need to make new definitions in the customer profile to specify the the target audience's characteristics and find a way to satisfy the customers' needs. In this context, the purpose of customer segmentation comes into prominence to foresee the trends in the market, which assist in evaluating customer behaviors and hence determine the actions needed to preserve and enhance the current situation in the market.

Meet the customers' needs is vital to preserve the existing customers and attract potential customers. The way to understand the customer's needs is to know their habits, identify their interests, and interpret their tendencies correctly. A well-defined customer segmentation, which divides customers into segments based on the common behaviors around the prioritized components, ensures recognizing of the target audience and identifying the needs to realize the compatibility with the marketing strategies.

Recency, Frequency, Monetary (RFM) analysis is one of the most popular methods used in customer segmentation based on customer behavior analysis, and many studies conducted RFM models for customer segmentation [1-6]. The RFM model was first proposed by Arthur Hughes in 1994 [7], and the usage of the model became prevalent in this area. In RFM Model, recency (R) refers to the recent time when the purchase occurred, frequency (F) refers to how frequently the purchase is made, and monetary (M) indicates the amount of money that is spent on the purchase in general. In this study, we modified the meaning of these dimensions and assigned more than one feature to each RFM dimension. In this context, R represents the recent time that the customer has used the device, F indicates the frequency of uses of the device, such as average daily usage, average duration, etc., and M includes usage habits of the device that reflect as cost and the price of the device.

One of the most popular clustering algorithms, K-Means, is used for clustering customers with similar behaviors. In the literature, significant applications of K-Means are seen as customer segmentation which assists in specifying customer segments, such as most expensive spenders, inexpensive spenders, average spenders, etc., in improving the marketing strategies [8-14]. The k-means algorithm is an unsupervised learning algorithm used to group data into the optimal number of clusters, where similar data is located in the same cluster, whereas different data are assigned to different clusters. In this context, similarity refers to customers with similar characteristics according to the features determined by the study.

In this paper, our main contribution is to adapt the K-means clustering algorithm to the RFM model by extracting features that represent RFM aspects of home appliances. In this regard, we utilize RFM-oriented features for segmenting customers using the K-means algorithm. Thence, customers with similar RFM-oriented features are assigned to the same segments, while customers with different RFM-oriented features are located in different segments. The resulting clusters are ranked using Customer Lifetime Value (CLV) metric, which is based on the weights of the RFM dimensions and hence, helps interpret the customer segments according to the customers' value to the business. The definition of the customers' value is determined by the RFM model created.

Profitability of customer value which is defined as the present value of the future profit over a given time contacted with the customer [15], is essential for an increase in market sharing. CLV metric plays a key role in measuring customer value, especially in terms of customer retention and migration in marketing. According to the literature, CLV is commonly used in the applications such as customer segmentation [16-19], sales and marketing strategy decisions [20, 21], customer retention [22], and product recommendation [23].

The organization of the paper is as follows: In Section 2, related work on customer segmentation that uses RFM analysis and K-means algorithms in the literature has been presented. The methods and the materials with the experimental study have been detailed in Section 3. The experimental results have been presented in Section 4. The conclusion of the study and future work are discussed in Section 5.

2. Related Work

In the literature, RFM Model and the K-means algorithm are widely used for customer segmentation [1-6, 8-14]. Zhao et al. [1] combined both methods and the additional Apriori algorithm to segment customers and provided a recommendation system using historical sales data. RFM model was used in various sectors, such as Business-to-customer (B2C) systems [1, 6], e-commerce applications [2, 4, 5, 8, 11-13], online retails [3, 10], and banking [14].

Mensouri et al. [4] proposed an extended version of the RFM model by adding new dimensions to analyze and scale customer satisfaction over time. Adding a new dimension to the model assisted in the identification of potential customers. Wu et al. [6] presented an empirical study by applying the RFM model and K-means algorithm to online purchasing data. Principal Component Analysis (PCA) has been used to add weights to each RFM dimension and achieved good results based on the increase in customers. Huang et al. [11] improved the RFM model by extending the community (C) dimension to represent community relations to introduce the value of social interaction to the educational e-commerce system and achieved accuracy by the modification. Chen et al. [9] investigated the travel patterns of using public transport via subway and bike-sharing using the RFM model and K-means clustering.

3. Materials and Methods

In this study, first, data is preprocessed to comply with RFM dimensions, and following the feature extraction and selection, clusters are specified with RFM Analysis and K-means algorithm. Based on the customer's contribution to business profits, clusters are named using Cluster Lifetime Value (CLV). The general flow of the methodology is presented in Figure 1.

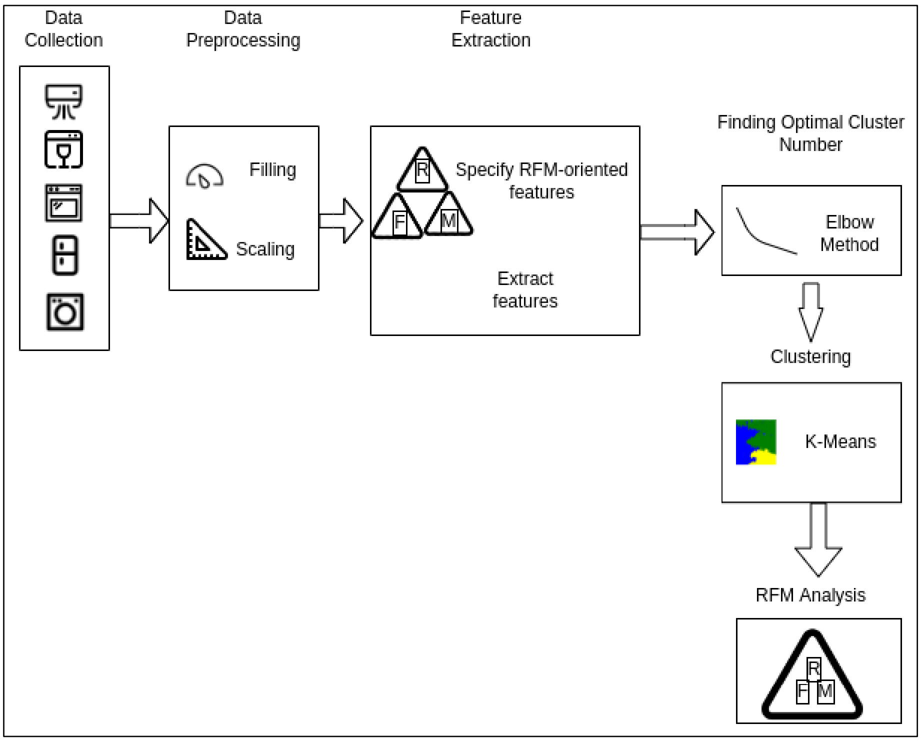


Figure 1. The general flow of the methodology

Şekil 1. Metodolojinin genel akış diyagramı

3.1. Dataset description

Data was collected from connected home appliance devices consisting of air conditioners (AC), dishwashers (DW), ovens (OV), refrigerators (RF), and washing machines (WM) between June 1st, 2021, and April 30th, 2022. These devices with WiFi modules were produced in a white goods factory, and we collected data from these devices via wireless network when the purchased devices were connected to the WiFi by the customer. The number of devices for which data was collected is 18K for AC, 5.6K for DW, 79 for OV, 7.6 K for RF, and 5K for DW. The total device count in the dataset is ~41K. The raw dataset consists of device statuses and sensor data sent as a result of human interaction with the device for each of them.

Device statuses and human interactions with the devices that are collected from the device sensors are transmitted over the wireless network to the cloud. The raw data is aggregated and stored in the cloud. The collection process of the data is presented in Figure 2. Data collected from each device is processed to comply with the RFM dimensions.

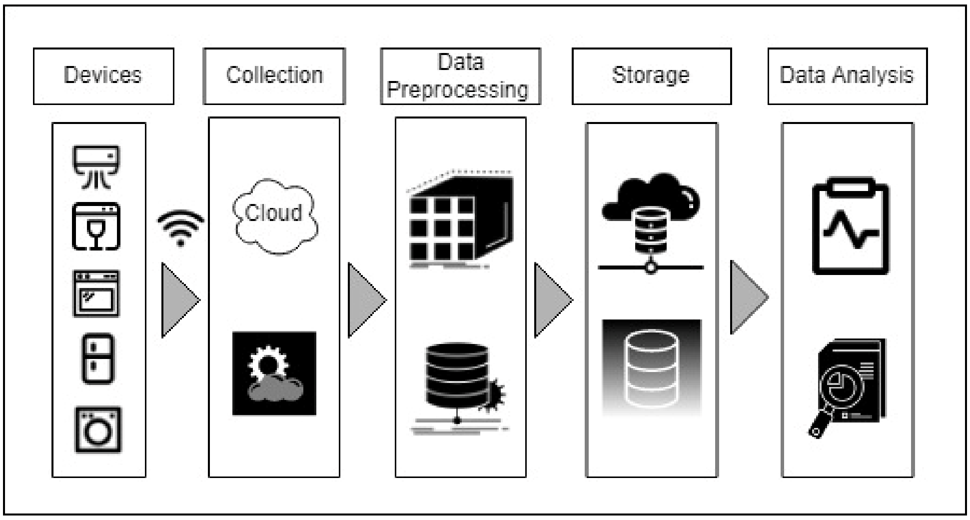


Figure 2. The general flow of data collection and preprocessing

Şekil 2. Veri toplama ve ön işlemenin genel akışı

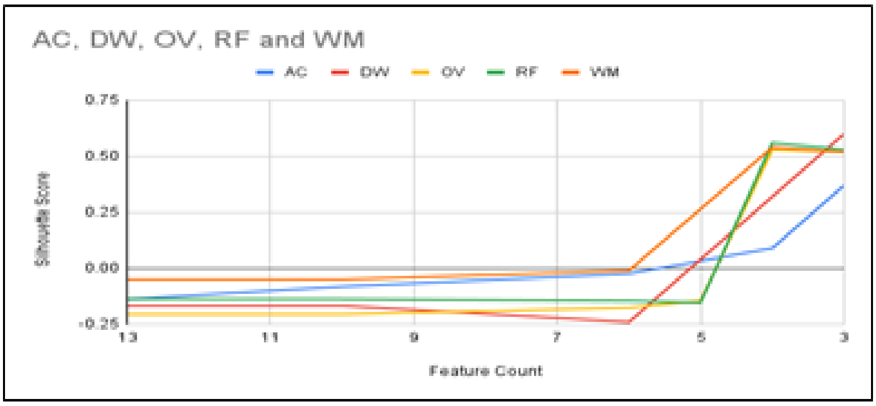


Figure 3. Feature count vs. silhouette score

Şekil 3. Özellik sayısı ve silhouette skoru

At the beginning of the experiments, we started with 10-15 features for the devices. Based on the silhouette score, the number of features is empirically reduced as shown in Figure 3.

Features are comprised of the data sent by the devices. Each device sends device statuses and human interactions with the device to the cloud. Device statuses and activities are logged in real-time and stored. Activity durations are calculated by subtracting the start time of the activity from the end time. Each activity change according to the devices.

The average duration for AC is calculated by averaging, in days, the difference between the start time when the AC is first opened and the end time when the AC is turned off in number of days.

The average daily usage of DW is calculated by dividing the daily usage count to the number of days when the DW is turned on. Daily usage count is computed by counting the statuses sent by the DW during the time when it is turned on.

The average daily usage of OV is calculated by taking the average number of times when the OV is turned on a daily basis.

RF sends data when the RF door is opened and closed. The average count of the door open statuses on daily basis is taken as the average daily door open count.

The average duration for WM is calculated by taking the average of the difference between the start and the end time when WM is turned on and off, respectively, in days. The average daily usage for WM is obtained by dividing the count of the statuses sent by the WM while it is being used to the number of days when it is first turned on.

Activities related to the usage of the devices are named as usage for each device in the dataset.

In the experiments, 11-month data was used for each device. The price data was retrieved on April 2022.

Features selected from devices based on RFM dimensions are summarized in Table 1.

Table 1. Features used in the final RFM model.

Tablo 1. Nihai RFM modelinde kullanılan özellikler.

Device	Recency (R)	Frequency (F)	Monetary (M)
AC	Last Usage	Average Duration	Price
DW	Last Usage	Average Daily Usage	Price
OV	Last Usage	Average Daily Usage	Price & Full Grill-Usage Rate
RF	Last Usage	Average Daily Door Open Count	Price
WM	Last Usage	Average Daily Usage & Average Duration	Price

- "*Last Usage*" refers to the recent time that the customer has used this device. It is converted to the number of days in the time range in which data was collected.
- "*Average Duration*" denotes the average duration that AC is turned on in days, and is calculated as follows: Total Duration in days / Total Number of days.
- "*Average Daily Usage*" refers to the average days that the device is used and is calculated as follows: Total Usage in days / Total Days in month.
- "*Full Grill-Usage Rate*" is the ratio that the OV is used for the full-grill program, and calculated as follows: Total Number of Full Grill Program Usage Count / Total Program Usage Count.

3.2. Data preprocessing

Following the feature selection and extraction stages, feature values were normalized using the standard scale to deal with the different measurements and magnitudes of the features that could cause undesirable effects. In this context, each feature was normalized by subtracting the mean of each feature and dividing by the standard deviation of the feature. The intuition here is to move the mean of each feature to 0 and the standard deviation to 1.

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

where x represents a feature value, μ is the mean of the attribute, and σ indicates the standard deviation. Missing values lead to erroneous results, and dealing with missing values is important. Therefore, in this study, missing values were replaced with median values.

3.3. Methods

RFM Analysis [7] is one of the most popular marketing techniques for analyzing customer behavior to improve customer segmentation based on recency, frequency, and monetary dimensions. In our study, recency refers to the recent time that the customer used the device; frequency indicates how frequently the customer uses the device based on daily usage, duration, or door open counts, which can differ from device to device; and monetary represents the price of the device and usage habits of the device that reflects the bill.

Features are extracted from data and processed to refer to the RFM dimensions. The number of features is determined according to the silhouette score empirically.

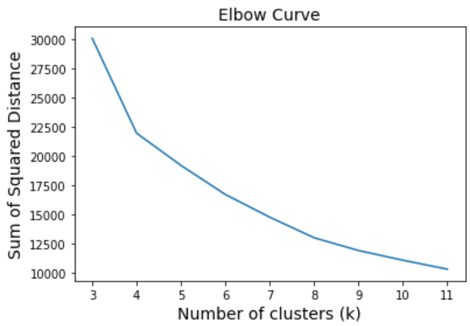
K–Means clustering algorithm helps us discriminate customers according to the features represented by the RFM model into different segments. Given RFM-based features as input to the K-Means clustering algorithm, the optimal number of clusters, which is denoted by *k*, is found according to the Elbow curve. As seen in Figure 4, the optimal cluster count was found as 4.

Based on the optimal number of clusters, the K-Means algorithm was run for each device. The performance of the clustering is measured by the Silhouette score, which evaluates the quality of the clusters and ranges between -1.0 and 1.0. Positive scores indicate that cluster assignments have well cohesion and separation, while negative scores indicate that cluster assignments are less appropriate. Silhouette score in the range of 0.51-0.7 is interpreted as a reasonable structure found in cluster assignment, which is just below the maximal score range [24, 25]. Based on the literature [3, 24 ,25], sthe threshold for the Silhouette score is applied as 0.5, which is in the range of the scores for a reasonable clustering [25], and is commonly used.

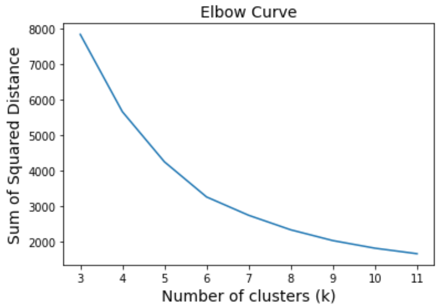
The Silhouette score is calculated as follows:

$$Score_i = \frac{b_i - a_i}{\max(a_i,\, b_i)} \qquad (2)$$

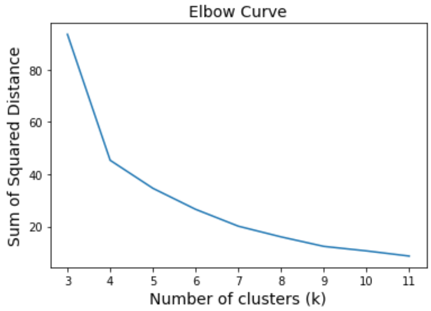
where a_i denotes the mean of the distance between the data points in the same cluster, and b_i represents the mean of the nearest different cluster for data point *i*.



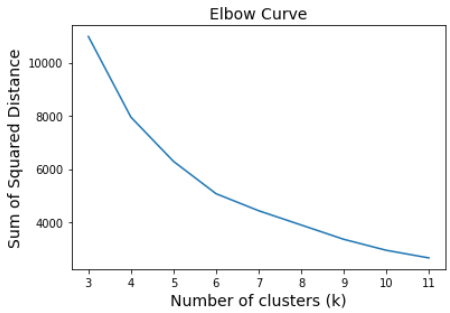
(a) AC, *k*=4



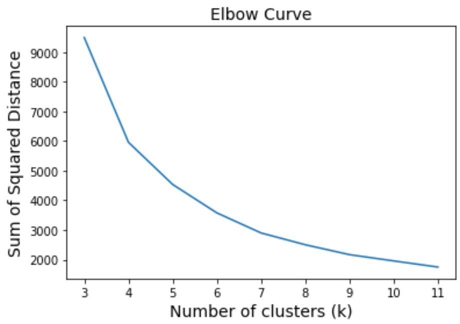
(b) DW, *k*=4



(c) OV, *k*=4



(d) RF, *k*=4



(e) WM, *k*=4

Figure 4. The Elbow method for the optimal number of clusters (*k*)

Şekil 4. Optimum küme sayısı için Elbow yöntemi (k)

As a result of the K-means algorithm, given the features based on the RFM dimensions and the optimal number of clusters, we obtain four segments. The pseudo-code for the clustering algorithm with RFM dimensions is given in Figure 5.

In our study, we aim to identify the customer segments based on device usage habits. In terms of the spending trend, features related to the frequency and monetary dimensions are evaluated, while the loyalty of the customer is mostly associated with its recent usage, which refers to the recency dimension. According to these specified criteria, weights are assigned to corresponding RFM dimensions, and Customer Lifetime Value (CLV) metric is applied given the weighted values for each feature. As recent work [15-23] suggests, the CLV metric is commonly used to measure how valuable a customer is to the business according to the cluster they are.

$$R' = \frac{R - R_{min}}{R_{max} - R_{min}} \quad (3)$$

$$F' = \frac{F - F_{min}}{F_{max} - F_{min}} \quad (4)$$

$$M' = \frac{M - M_{min}}{M_{max} - M_{min}} \quad (5)$$

$$CLV = \frac{(R_w \times R' + F_w \times F' + M_w \times M')}{R_w + F_w + M_w} \quad (6)$$

where $R_w = 1$, $F_w = 5$, $and\ M_w = 3$ are weighted values for the features. The pseudo-code for CLV application is presented in Figure 6.

For each device, the CLV values of each segment are calculated and the segments are ranked according to these CLV values. These ranks correspond to a level where the segments meet the criteria. With this in mind, each segment was named Best Customer, Good Customer, Average, and Worst Customer.

Segment names can also be referred to as Customer Types that we need to identify.

The customer types and the RFM pattern that reflects the CLV ranking and the descriptions are presented in Table 2.

Table 2. Customer types in the RFM model.

Tablo 2. RFM modelindeki müşteri türleri.

Customer Type	RFM Pattern	Description
Best Customer	R↑F↑M↑	• Customer that spends the most and frequently • Promising customer • Follow up big spending customer
Good Customer	R↑F↑M↓ or R↑F↓M↑ or R↓F↑M↑	• Good customer • Follow up moderate customer • Loyal cheap spending customer
Average Customer	R↑F↓M↓ or R↓F↓M↑ or R↓F↑M↓	• Need to take action for the moderate customer • Customers need to be encouraged
Worst Customer	R↓F↓M↓ or R↓F↔M↓ or R↓F↓ M↔	• Customers that spend the least • One timer Customers

Input:

k: Optimal number of clusters

D: Dataset with n instances

I: Dataset instance with the following features:

F_R: Feature for Recency

F_F: Feature for Frequency

F_M: Feature for Monetary

Output:

C_1... C_k: k clusters

Algorithm:

1. Select k random instances from dataset D, as cluster center
2. Assign each instance, I to the cluster where the distance of the instance to the cluster center is the smallest
3. Re-organize the clusters by recalculating mean of each cluster
4. Repeat steps 2-3 until no change happens in the clusters
5. Evaluate each cluster by using the Silhouette score as below:

$$Score_i = \frac{b_i - a_i}{\max(a_i, b_i)}$$

Where,

$Score_i$: Silhouette score

a_i: represents the mean of the distance between the i^{th} instance and all other instances in the same cluster

b_i: represents the mean of the nearest different cluster to the i^{th} instance

Figure 5. Clustering algorithm with RFM dimensions

Şekil 5. RFM boyutlarıyla kümeleme algoritması

Input:

F_R^i: Average recency feature values for cluster i

F_F^i: Average frequency feature values for cluster i

F_M^i: Average monetary feature values for cluster i

R_W , F_W , M_W: Weights for each RFM dimension

Output:

CLV_1... CLV_k: CLV values for each clusters

Algorithm:

1. Apply following formula to calculate CLV values:

$$R' = \frac{R - R_{min}}{R_{max} - R_{min}}$$

$$F' = \frac{F - F_{min}}{F_{max} - F_{min}}$$

$$M' = \frac{M - M_{min}}{M_{max} - M_{min}}$$

$$CLV = \frac{(R_w \times R' + F_w \times F' + M_w \times M')}{R_w + F_w + M_w}$$

Where R, F, and M represent F_R^i, F_F^i and F_M^i respectively

2. Repeat the calculation for each cluster
3. Sort the clusters based on the CLV values

Figure 6. Algorithm for Customer Lifetime Value (CLV)

Şekil 6. Müşteri Yaşam Boyu Değeri Algoritması (CLV)

Clusters that were produced as the result of the algorithm are presented in Figures 7-11.

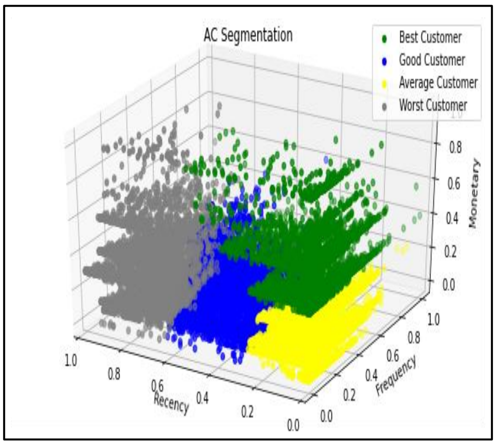


Figure 7. AC segmentation

Şekil 7. AC segmentasyonu

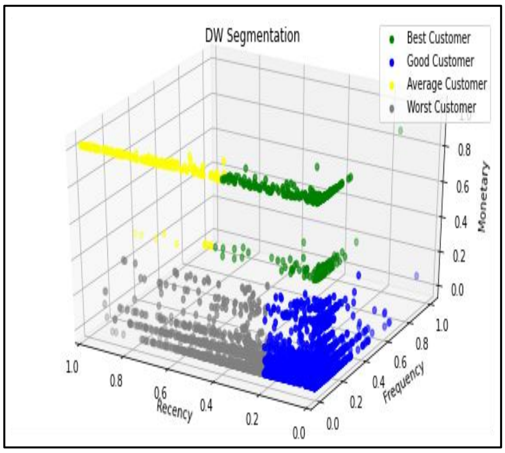


Figure 8. DW segmentation

Şekil 8. DW segmentasyonu

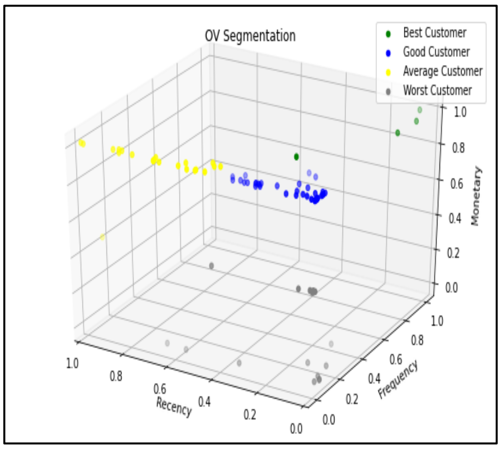


Figure 9. OV segmentation

Şekil 9. OV segmentasyonu

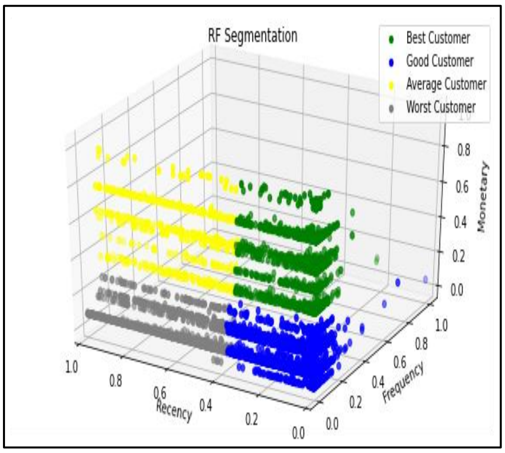


Figure 10. RF segmentation

Şekil 10. RF segmentasyonu

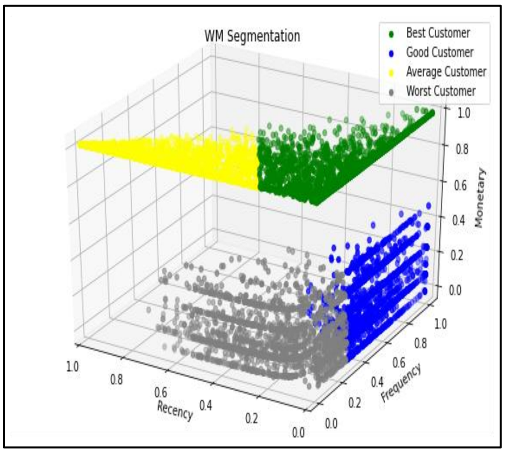


Figure 11. WM segmentation

Şekil 11. WM segmentasyonu

4. Results

Table 3 represents the final scores of the K-Means algorithm with the features that are involved in each device. As can be seen, the algorithm has passed the threshold, which is 0.5, except for AC. The score for AC is close to the threshold.

Table 3. Performance results of final clusters.

Tablo 3. Nihai kümelerin performans sonuçları.

Device	Silhouette Score	Features
AC	0.37	R: last seen, F: avg. duration, M: price
DW	0.60	R: last seen, F: avg. usage per day, M: price
OV	0.53	R: last seen, F: avg. usage per day, M: price & full grill usage rate
RF	0.56	R: last seen, F: avg. daily door open count, M: price
WM	0.54	R: last seen, F: avg. duration & avg. usage per day, M: price

In order to interpret the resulting clusters in terms of business value, we used Customer Lifetime Value (CLV) as the naming approach. This metric is used to measure how valuable a customer is to the business according to the cluster they are involved in.

Based on the CLV values, each cluster is ranked and named. The clusters for each device with the rank are summarized in Tables 4-8. In the tables, abbreviations within the parentheses are the units.

Table 4. Cluster summary for AC.

Tablo 4. AC için kümeleme özeti.

Cluster	Avg. Price (TL)	Average Duration (Hours)	Avg. Last Seen (Days)	Device Count	CLV
Best	14,109.35	3.90	24.52	5267	0.91
Good	11,192.33	4.08	160.88	2937	0.66
Avg.	10,234.14	3.88	18.74	7100	0.57
Worst	11,867.01	3.08	285.83	3379	0.14

Table 5. Cluster summary for DW.

Tablo 5. DW için kümeleme özeti.

Cluster	Avg. Price (TL)	Average Usage Per Day (Hours)	Avg. Last Seen (Days)	Device Count	CLV
Best	10,817.61	0.78	29.03	430	0.87
Good	6,335.86	0.92	16.96	3939	0.67
Avg.	11,285.34	0.15	271.25	183	0.34
Worst	6,339.67	0.14	154.14	1077	0.08

Table 6. Cluster summary for OV.

Tablo 6. OV için kümeleme özeti.

Cluster	Avg. Price (TL)	Avg. Usage Per Day (Hours)	Full Grill Usage Rate (Count)	Avg. Last Seen (Days)	Device Count	CLV
Best	11,399	0.37	0.14	32.75	4	1
Good	11,399	0.024	0.01	51.33	36	0.3
Avg.	11,262	0.015	0.05	242.46	22	0.32
Worst	6,528	0.033	0.02	53.24	17	0.15

Table 7. Cluster summary for RF.

Tablo 7. RF için kümeleme özeti.

Cluster	Avg. Price (TL)	Avg. Daily Door Open (Count)	Avg. Last Seen (Days)	Device Count	CLV
Best	14,734.69	0.25	25.71	2051	0.99
Good	11,671.21	0.24	26.96	3696	0.65
Avg.	14,834.36	0.03	225.25	560	0.35
Worst	11,592.66	0.05	252.32	1373	0.05

Table 8. Cluster summary for WM.

Tablo 8. WM için kümeleme özeti.

Cluster	Avg. Price (TL)	Avg. Usage Per Day (Hours)	Average Duration (Hours)	Avg. Last Seen (Days)	Device Count	CLV
Best	7,225	0.55	2.49	10.35	2032	0.78
Good	11,399	0.40	1.43	21.97	3202	0.74
Avg.	11,399	0.06	0.24	243.84	1713	0.22
Worst	7,098	0.10	0.31	78.16	1917	0.13

Best clusters represent the customers with relatively high spending customer behavior in terms of usage habits and purchasing. On the other hand, the worst clusters represent the customers with low spending and one timer attitude. In our CLV formula, Frequency and Monetary play a severe role in covering the continual behavior within the scope of the business values.

5. Conclusion and Future Work

In this paper, we applied the K-means clustering algorithm to the features of home appliances extracted based on RFM Analysis in order to determine the target customers and improve the marketing strategies according to the findings. At the beginning of the experiments, the initial feature size was 10-15, which included specific programs and options for the devices. The feature set was reduced empirically based on the clustering performance measured by the silhouette score. The threshold for the silhouette score was defined as 0.5, which is in the range of the scores for reasonable clustering based on the literature [3, 24, 25]. The scores for each device were 0.37, 0.60, 0.53, 056, and 0.54 for devices AC, DW, OV, RF, and WM, respectively. Given the features with the best silhouette score, the K-means algorithm was applied to find the optimal cluster number for each device. As a result, 4 clusters were found, and CLV was applied to measure the quality of the clusters regarding the business values with the weighted RFM method. The RFM weights may vary with the type of industries and business expertise, and RFM analysis is the key to determining the weights that comply with the intended strategy. In our study, device usage habit that affects the spending trend was aimed to identify, and therefore, more weights were given to frequency and monetary than the recency. For each device, CLV values were obtained for each cluster, and each cluster was sorted according to its own CLV values. The name of the clusters was determined based on their rank, corresponding to the level where the clusters meet the criteria. With this consideration, each cluster was named Best Customer, Good Customer, Average, and Worst Customer.

The use of customer segmentation in the industry is a key field for identification of the target customers and hence, developing marketing strategies to grow up on the market share. CLV-focused applications help analyze the segments in terms of the RFM model, where the weights of RFM dimensions can be customized according to the type of industry. With the motivation of the obtained results, we plan to expand the use of the analysis for customer segment-based product recommendations and develop a web service that lists the segments of the target customers and integrates the solution into the system.

5. Tartışma ve Sonuç

Bu makalede, hedef müşterileri belirlemek ve bulgulara göre pazarlama stratejilerini geliştirmek için RFM Analizine dayalı olarak çıkarılan özelliklere K-Means kümeleme algoritmasını uyguladık. Başlangıçta belirlenen özellik sayısı, cihaz programları ve program seçenekleriyle birlikte 10-15 civarındaydı. Bu özellik seti, silhouette skoruyla ölçülen kümeleme performansına göre deneysel olarak azaltıldı. Silhouette skoru eşik değeri, literatürdeki kümeleme skor aralığına göre belirlenmiş olup 0.5 olarak tanımlandı [3, 24, 25]. AC, DW, OV, RF, ve WM cihazları için sırasıyla 0.37, 0.60, 0.53, 0.56, ve 0.54 skorları elde edildi. En iyi silhouette skoruna sahip özellikler göz önüne alınarak, her cihaz için en uygun küme sayısını bulmak için K-Means algoritması uygulandı. Sonuçta 4 küme bulundu ve kümelerin iş değerleri açısından kalitesini ölçmek için ağırlıklı RFM yöntemi ile CLV uygulandı. RFM ağırlık katsayıları, endüstri türüne ve iş uzmanlığına göre değişmekte ve RFM analizi, amaçlanan stratejiyle uyumlu ağırlıkları belirlemekte anahtar rol oynamaktadır. Çalışmamızda, müşterilerin para harcama eğilimlerini etkileyen cihaz kullanım alışkanlıklarının belirlenmesi amaçlandı ve bu nedenle sıklık (F) ve tutar (M) bileşenlerin ağırlık katsayıları, cihazın en son kullanıldığı zamanı temsil eden yenilik (R) bileşenine göre daha fazla olacak şekilde belirlendi. Her cihaz için elde edilen her bir kümenin CLV değeri hesaplandı ve kümeler, bu hesaplanan CLV değerlerine göre sıralandı. Küme isimleri, kümelerin kriterleri karşıladığı düzeydeki sıralamasına uygun olarak belirlendi. Buna göre, en yüksek CLV değerine sahip ilk sıradaki küme En İyi Müşteri, bir alt sıradaki küme İyi Müşteri ve sırayla Ortalama ve En Kötü Müşteri isimleri seçildi.

Müşteri segmentasyonunun endüstride kullanımı, hedef müşterilerin belirlenmesi ve dolayısıyla pazar payında büyümek için pazarlama stratejilerinin geliştirilmesi için kilit rol oynar. CLV odaklı uygulamalar, RFM bileşenlerinin ağırlıklarının endüstri türüne göre özelleştirildiği RFM modeli ile küme analizine yardımcı olur. Elde edilen sonuçların motivasyonu ile müşteri segmentine dayalı ürün önerileri için analizleri genişletmeyi ve hedef müşterilerin segmentlerini listeleyen çözümleri sisteme entegre eden bir web servis geliştirmeyi planlıyoruz.

6. Ethics committee approval and conflict of interest statement

There is no need to obtain permission from the ethics committee for the article prepared.

There is no conflict of interest with any person / institution in the article prepared.

References

[1] Xian, Z., Keikhosroiani, P., XinYing, C., Li, Z. 2022. An RFM Model Using K-Means Clustering to Improve Customer Segmentation and Product Recommendation. Keikhosroiani, P., ed. Handbook of Research on Consumer Behavior Change and Data Analytics in the Socio-Digital Era, IGI Global, Malaysia, pp. 124-145. DOI: 10.4018/978-1-6684-4168-8.ch006

[2] Li, P., Wang, C., Wu, J., Madleňák, R. 2022. An E-commerce Customer Segmentation Method Based on RFM Weighted K-means. International Conference on Management Engineering, Software Engineering and Service Sciences (ICMSS), 61-68. DOI: 10.1109/ICMSS55574.2022.00017

[3] Anitha, P., Malini M.P. 2022. RFM Model for Customer Purchase Behavior Using K-Means Algorithm, Journal of King Saud University - Computer and Information Sciences, Vol. 34, No. 5, p. 1785–1792. DOI: 10.1016/j.jksuci.2019.12.011

[4] Mensouri D., Azmani A., Azmani M., 2022. K-Means Customers Clustering by Their RFMT and Score Satisfaction Analysis, International Journal of Advanced Computer Science and Applications, Vol. 13, No. 6, p. 469–476. DOI: 10.14569/ IJACSA.2022.0130658

[5] Wan, S., Chen, J., Qi, Z., Gan, W., Tang, L. 2022. Fast RFM Model for Customer Segmentation. In Companion Proceedings of the Web Conference (WWW '22), Association for Computing Machinery, New York, NY, USA, 965–972. DOI: 10.1145/3487553.3524707

[6] Wu, J., Shi, L., Lin, W-P., Tsai, S-B., Li, Y., Yang, L., Xu, G. 2020. An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K-Means Algorithm, Mathematical Problems in Engineering, Vol. 2020, p. 1-7. DOI: 10.1155/2020/8884227

[7] Hushes, A. M. 1994. Strategic Database Marketing 4e: The Masterplan for Starting and Managing a Profitable, Customer-Based Marketing Program. Mc Graw-Hill, 608p.

[8] Tabianan, K., Velu, S., Ravi, V. 2022. K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data, Sustainability, Vol. 14, No. 12. p. 1-15. DOI: 10.3390/su14127243

[9] Chen, A., 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, Vol. 2022, p. 1-14. DOI: 10.1155/2022/1108105

[10] Dewabharata, A. 2022. Customer Segmentation Using the K-Means Clustering as a Strategy to Avoid Overstock in Online Shop Inventory. Proceedings of the 1st International Conference on Contemporary Risk Studies, ICONIC-RS, 31 March-1 April, South Jakarta, DKI Jakarta, Indonesia, 1-12. DOI: 10.4108/eai.31-3-2022.2320688

[11] Huang, Y., Zhang, M., He, Y. 2020. Research on Improved RFM Customer Segmentation Model Based On K-Means Algorithm. 5th International Conference on Computational Intelligence and Applications (ICCIA), 19-21 June, 24-27. DOI: 10.1109/ICCIA49625.2020.00012

[12] Dedi, A., Dzulhaq, M.I., Sari, K.W., Ramdhan, S., Tullah, R., Sutarman 2019. Customer Segmentation Based on RFM Value Using K-Means Algorithm. Fourth International Conference on Informatics and Computing (ICIC), 1-7. DOI: 10.1109/ICIC47613.2019.8985726

[13] Vohra, R., Pahareeya, J., Hussain, A., Ghali, F., Lui, A. 2020. Using Self Organizing Maps and K-Means Clustering Based on RFM Model for Customer Segmentation in the Online Retail Business. Intelligent Computing Methodologies, 16th International Conference, Bari, Italy, October 2–5, Springer-Verlag, Berlin, Heidelberg, 484–497. DOI: 10.1007/978-3-030-60796-8_42

[14] Maryani, I., Riana, D., Astuti, R.D., Ishaq, A., Sutrisno, Pratama, E.A. 2018. Customer Segmentation based on RFM model and Clustering Techniques with K-Means Algorithm. Third International Conference on Informatics and Computing (ICIC), 1-6. DOI: 10.1109/IAC.2018.8780570

[15] Kotler, P. 1974. Marketing during Periods of Shortage, Journal of Marketing, Vol. 38, No. 3, p. 20-29. DOI: 10.1177/002224297403800305

[16] Shih, YY., Liu, CY. 2003. A Method for Customer Lifetime Value Ranking — Combining the Analytic Hierarchy Process and Clustering Analysis. Database Marketing & Customer Strategy Management,

Vol. 11, No. 2, p. 159–172. DOI: 10.1057/PALGRAVE.DBM.3240216

[17] Monalisa, S., Nadya, P., Novita, R., 2019. Analysis for Customer Lifetime Value Categorization with RFM Model, Procedia Computer Science, Vol. 161, p. 834-840. DOI: 10.1016/J.PROCS.2019.11.190

[18] Haenlein, M., Kaplan, AM., Beeser, Aj., 2007. A Model to Determine Customer Lifetime Value in a Retail Banking Context, European Management Journal, Vol. 25,No.3,p. 221-234. DOI: 10.1016/J.EMJ.2007.01.004

[19] He, X., Li, C., 2016. The Research and Application of Customer Segmentation on E-Commerce Websites, 6th International Conference on Digital Home (ICDH), 203-208. DOI: 10.1109/ICDH.2016.050

[20] Dwyer, FR. 1997. Customer Lifetime Valuation to Support Marketing Decision Making, Journal of Direct Marketing, Vol. 11, No. 4, p. 6-13. DOI: 10.1002/(SICI)1522-7138(199723)11:4<6::AID-DIR3>3.0.CO;2-T

[21] Khajvand, M., Zolfaghar, K., Ashoori, S., Alizadeh, S., 2011. Estimating Customer Lifetime Value based on RFM Analysis of Customer Purchase: Case Study, Procedia Computer Science, Vol. 3, p. 57-63. DOI: 10.1016/J.PROCS.2010.12.011

[22] Hidalgo, P., Manzur, EF., Olavarrieta, S., Farias, P. 2008. Customer Retention and Price Matching: The AFPs Case, Journal of Business Research, Vol. 61, No. 6, p. 691-696. DOI: 10.1016/J.JBUSRES.2007.06.046

[23] Shih, YY., Liu, DR. 2008. Product Recommendation Approaches: Collaborative Filtering via Customer Lifetime Value and Customer Demands, Expert Systems with Applications: An International Journal, Vol. 35, No. 1-2, p. 350-360. DOI: 10.1016/J.ESWA.2007.07.055

[24] De Amorim, RC, Hennig, C. 2014. Recovering the Number of Clusters in Data Sets with Noise Features using Feature Rescaling Factors, Information Sciences, Vol. 324, p. 126-145. DOI: 10.1016/J.INS.2015.06.039

[25] Leonard, K., Rousseeuw, P. 1990. Partioning Around Medoids (Program PAM). pp 68-125. Leonard, K., Rousseeuw, P., ed. 1990. Finding Groups In Data, Wiley, New York , 335p.