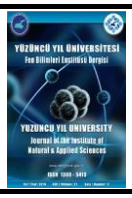
Yuzuncu Yil University
Journal of the Institute of Natural & Applied Sciences
https://dergipark.org.tr/en/pub/yyufbed



Research Article

Improving Residential Marketing Campaigns via Customer Data Clustering

Mohammed MUQBEL, Selçuk ÖZCAN*, Çağrı SEL

Karabük University, Faculty of Engineering, Department of Industrial Engineering, 78050, Karabük, Türkiye
*Corresponding author e-mail: selcukozcan@karabuk.edu.tr

Abstract: As the construction industry struggles to develop effective marketing plans for residential projects, using rich datasets to understand customer demand helps builders of residential complexes with complex use cases. Decision-makers often struggle to understand big data. Solving this problem begins with relevant data being mined and collected. Multi-criteria decision-making (MCDM) models are used to rank the data or alternative options according to their importance to decision-makers. The weights of the criteria, obtained according to their importance, are essential to reveal the relative value of different criteria. To understand and analyze big data, cluster analysis within the data mining discipline is used to segment and score data. This analysis is an effective tool for determining marketing strategies and understanding customer behavior. This study was conducted to determine the marketing strategy appropriate for customer segmentation. For this, the Rank Order Centroid (ROC) criterion weight method gives weights to the criteria according to their relative importance. The K-means cluster analysis algorithm uses the values obtained from the ROC method. By combining ROC and K-means methods, this study will contribute to extracting information from large data sets and simplifying decision-making processes in the residential sector. As a result of the study, customers were divided into groups, and it was concluded that the groups with the highest scores should be prioritized in marketing strategies.

Keywords: Clustering, Data mining, Marketing, Multicriteria decision-making, Resident

Müşteri Verilerinin Kümelenmesi Yoluyla Konut Pazarlama Kampanyalarının İyileştirilmesi

Öz: İnşaat sektörü, konut projeleri için etkili pazarlama planları geliştirmekle uğraşırken, müşteri taleplerini anlamak için zengin veri kümelerini kullanmak, karmaşık kullanım alanlarına sahip konut komplekslerinin inşa edicilerine yardımcı olmaktadır. Karar vericiler genellikle büyük verileri anlama konusunda zorluklarla karşılaşırlar. Bu sorunu çözmek, ilgili verilerin madenciliği ve toplanmasıyla başlar. Çok kriterli karar verme (MCDM) modelleri, verileri veya alternatif seçenekleri karar vericiler için önemlerine göre sıralamak için kullanılır. Önem derecelerine göre elde edilen kriterlerin ağırlıkları, farklı kriterlerin göreceli değerini ortaya çıkarmak için esastır. Büyük veriyi anlamak ve analiz etmek için, veri madenciliği disiplini içinde kümeleme analizi kullanılır; bu analiz, pazarlama stratejilerini belirleme ve müşteri davranışını anlama konusunda etkili bir araçtır. Bu çalışma, müşteri segmentasyonu için uygun pazarlama stratejisini belirlemek amacıyla yapılmıştır. Bunun için Rank Order Centroid (ROC) kriter ağırlığı yöntemi kriterlere göreceli önemlerine göre ağırlıklar vermektedir. K-ortalama kümeleme analizi algoritması ROC yönteminden elde edilen değerleri kullanır. Bu çalışma, ROC ve K-ortalama metotlarını bir arada kullanarak, büyük veri setlerinden bilgi çıkarılmasına ve konut sektöründe karar alma süreçlerinin basitleştirilmesine katkı sağlayacaktır. Çalışma sonucunda müşteriler gruplara ayrılmış ve en yüksek puanı alan grupların pazarlama stratejilerinde önceliklendirilmesi gerektiği sonucuna varılmıştır.

Anahtar Kelimeler: Çok kriterli karar verme, Konut, Kümeleme, Pazarlama, Veri madenciliği

Received: 04.04.2024
Accepted: 02.12.2024

How to cite: Muqbel, M., Özcan, S., & Sel, Ç. (2025). Improving residential marketing campaigns via customer data clustering. *Yuzuncu Yil University Journal of the Institute of Natural and Applied Sciences*, *30*(1), 129-144. https://doi.org/10.53433/yyufbed.1463691

1. Introduction

In the business world and the construction industry, creating the right marketing strategies is not only important for the success of housing projects but also necessary to understand and fully meet customer needs. In this context, having rich and diverse information, such as bank customers' data, can be a great advantage in building mixed-use residential complexes.

With the growth of civilization and the development of technology, housing construction and basic consumer goods keep pace with this and increase the quality of living space with additional features. The housing need arising from rapid population growth, structural changes in agriculture, social and cultural development, and migration from rural to urban areas has made the housing sector particularly useful (Tacoli et al., 2015). In the business world and the construction industry, decision-makers often face difficulties understanding big data. Solving this problem begins with relevant data mined and collected. Data mining is a discipline that aims to extract meaningful information from large data sets. This process includes collecting, cleaning, analyzing, and finalizing data. Data mining helps businesses, organizations, and researchers better understand data and make more informed decisions. For this reason, data mining has become a useful research and business field today (Bilici & Özdemir, 2022). After data is collected, decision-makers may encounter differences in the weights of the collected data. In this case, Multicriteria Decision-Making (MCDM) methods come into play to resolve these weight differences.

Multiple criteria are included in many real-world decision-making situations. The criteria weights play a critical role in determining the relative importance of various criteria. In this way, information is given to decision-makers regarding which factors, when assessing the data, they should give more weight to. One of the MCDM methods is Rank Order Centroid (ROC) weight. This method is a straightforward way to assign weights to criteria based on their relative relevance (Roszkowska, 2013). To understand and analyze big data, cluster analysis within the data mining discipline is used to segment and score data. This analysis is a useful tool for understanding the structure and relationships in data sets, and one of the most widely used clustering algorithms is the K-means clustering algorithm (Pande et al., 2012). K-means cluster analysis divides the data into several clusters and calculates the centroids that define these clusters. Data are assigned to relevant clusters, considering their distance to these centers. This way, data with similar traits are grouped in one cluster to make different clusters. K-means analysis helps us better understand patterns and clustering in the dataset (Singh et al., 2011). K-means clustering analysis, which is used especially in marketing strategy, geographical analysis, and customer segmentation, plays a useful role in data mining projects. This clustering analysis simplifies the decision-making process by reducing the complexity of large data sets and providing valuable information to businesses (Ali & Kadhum, 2017).

This study uses the ROC method to give criteria weights according to their relative relevance using the Microsoft Excel program and the K-means algorithm for cluster analysis using the IBM SPSS Statistics program.

The main objective of this study is to develop a methodology that will assist decision-makers in the construction industry in the more effective analysis of large data sets and the subsequent development of marketing strategies. The following section presents a summary of the contributions made by this study.

1) Presents a novel approach that combines Multi-Criteria Decision-Making (MCDM) with data mining techniques, with the aim of assisting decision-makers in the construction industry in the analysis of large datasets and the formulation of marketing strategies.

2) Illustrates the practical integration of the ROC (Rank Order Centroid) method for the weighting of criteria and K-means clustering for the segmentation of customers, thereby enhancing the effectiveness of marketing strategies.

3) Demonstrates how the ROC method can enhance decision-making by prioritising pivotal factors in the marketing processes of housing projects, thereby offering more transparent guidance to stakeholders.

4) Demonstrates the efficacy of K-means clustering for customer segmentation, thereby facilitating the development of more targeted and efficient marketing strategies within the construction industry.

2. Literature Review

Cluster analysis techniques are valuable in marketing research, particularly for market segmentation. These techniques have been successfully applied in the marketing sector, particularly in data mining, where it has been used to segment markets or segment customers and predict future trends (Chiang, 2011; Tleis et al., 2017; Maciejewski et al., 2019; Kuo et al., 2020). For instance, Heryati and Herdiansyah used student data and were divided into five groups using the K-means clustering method according to regional origin and study programs; these groups were formed based on different promotional strategies (Heryati & Herdiansyah, 2020). Hutagalung used the K-means clustering method to determine the market segmentation of Indonesia Digital Home (Indihome), a leading digital fiber optic service package (Hutagalung et al., 2022). In customer segmentation, it has been found that K-means clustering is the most efficient algorithm (Maulina et al., 2019). Bai used the Fuzzy C-means algorithm; this study states that this methodology is applied to managerial insight and decision-making processes. Using a two-stage technique, companies were first divided into four clusters and then ranked within these clusters using MCDM (Bai et al., 2014).

The importance of criteria weighting in marketing decision-making is highlighted in several studies. Gürbüz and Hung both emphasize the use of MCDM methods, with Gürbüz specifically using a fuzzy metric distance and analytic hierarchy process (AHP) hybrid method (Hung et al., 2012; Gürbüz et al., 2014). Singh and Pant further review three weighing methods, highlighting their function in establishing the relative weight of the criteria in MCDM (Singh & Pant, 2021). These studies collectively underscore the usefulness of criteria weighting in marketing decision-making. The ROC method, a popular model for criteria weight estimation in MCDM in marketing, has been the focus of many studies. Pandiangan, in determining the location of minimarkets, the ROC method allows for obtaining an objective and precise result by focusing on various criteria (Pandiangan et al., 2023). Lastly, Wijaya used the ROC method to determine the most appropriate online sales platform. It evaluated the relationships and weights, determined the min and max values, and analyzed the effect of the criterion weights on the result (Wijaya et al., 2022). Studies underscore the versatility and effectiveness of the ROC method in various decision-making and marketing applications.

Table 1 lists the articles examined during the literature review and the methods they used. The table was made so readers could see the resolved problems and methods applied to this study and other studies.

The topic has attracted more attention recently than in past years, and the literature review conducted under the direction of the relevant subject and its subheadings has revealed techniques that could be applied to many sectors. This study aims to make clustering methods and criteria weighting methods accessible to a broader audience and serve as a resource for analyzing customer data and determining the right marketing strategy. There are a limited number of studies on integrating data clustering techniques, particularly the K-means algorithm, with Multi-Criteria Decision Making (MCDM) methods for developing marketing strategies. Existing studies address these methods separately but do not focus enough on the deeper and more precise marketing insights that can be gained by using these two methods together. This study, which determines criteria weights using the ROC method and performs customer segmentation with the K-means algorithm, aims to fill this gap. The integration of these two methods contributes to the development of more targeted and effective marketing strategies through the analysis of large data sets.

Table 1. Methods used in literature

Study	Problem	Criteria Weight Methods					Cluster Analysis Methods			
		ROC	AHP	ANP	BWM	DANP	K-means	Ward's	K-medoids	Fuzzy C-means
(Tleis et al., 2017)	Market segmentation						✓			
(Maciejewski et al., 2019)	Customer segmentation							✓		
(Kuo et al., 2020)	Market segmentation									✓
(Chiang, 2011)	Market segmentation							✓		
(Heryati & Herdiansyah, 2020)	Determine the promotion strategy						✓			
(Hutagalung et al., 2022)	Market segmentation								✓	
(Maulina et al., 2019)	Customer segmentation						✓			✓
(Bai et al., 2014)	Evaluating the performance of organizations									✓
(Gürbüz et al., 2014)	Marketing strategies and marketing decision processes analysis		✓							
(Hung et al., 2012)	Evaluating performances and improving professional services					✓				
(Singh & Pant, 2021)	The performance of weighing methods in MCDM		✓	✓	✓					
(Pandiangan et al., 2023)	Determination of mini market location	✓								
(Wijaya et al., 2022)	Selection of online sales platforms	✓								
This study	Customer segmentation	✓					✓			

3. Material and Methods

In this section, the methods used in the study are explained. The dataset used is introduced and how the methods are applied is explained step by step.

3.1. Marketing strategies

Developing a sound marketing strategy is beneficial to succeed in today's competitive business environment. Marketing strategies are plans and tactics that enable a product or service to reach its target audience, increase demand, and achieve long-term business goals. A business's marketing approach is key in determining its future, expanding its customer base, differentiating from competitors, and ensuring profitability.

There are numerous things to consider when developing a marketing strategy. The first thing to do is to determine who our target market is. Understanding who our customers are and how we can meet their needs and preferences is the foundation of a successful marketing strategy. Market research, customer feedback, and data analysis are useful (Rohmawati & Winata, 2021). The next action is to make a unique offer. Consider how to differentiate our business's products or services from competitors. We must offer our customers a value proposition explaining why they should choose us (Phillips-Wren et al., 2016). A marketing strategy can include many different tactics, including communication. Advertising, social media, content marketing, promotions, events, and many other tools can help our business engage with our target audience. When deciding which communication channels and tactics to use, we must consider where our target audience is and their preferred communication methods (Bajpai et al., 2012). Marketing strategies should be designed not only for short-term benefits but also for long-term growth. We should consider our business's future growth potential and adjust our marketing plan according to these goals. Finally, we should regularly review and adjust our marketing strategies to adapt to customer needs and market conditions (Wirtz & Daiser, 2018).

In addition to increasing a company's profits, marketing strategies also increase customer satisfaction, increase brand value, and provide a competitive advantage. Therefore, creating and implementing marketing strategies for businesses is extremely useful.

3.2. Data mining

The development of technology produces copious amounts of data, making data mining even more useful. Data mining aims to obtain meaningful information from copious amounts of data by combining statistics, artificial intelligence, and computer science. Contribute to strategic decision-making processes in business and science by identifying patterns in data using analytical techniques such as classification, regression, and clustering. Data mining is used in many fields, such as finance, health, and marketing (Madni et al., 2017).

3.2.1. Concept of data mining

Data mining is a procedure that combines fields such as statistics, mathematics, artificial intelligence, and computer science to find patterns, information, and relationships in big data sets (Sharma, 2014). In this field, various techniques are used to analyze complex and large data sets where information is inaccessible. Data mining analyzes data for purposes such as extracting information, making predictions, classifying, clustering, and identifying patterns of relationships. As such, it is used in many application areas in the business world, such as strategic decision-making, understanding customer behavior, and optimizing marketing strategies (Guo & Qin, 2017). Data mining is also a useful research area for developing artificial intelligence and machine learning. The stages of data mining are as follows (Wu et al., 2021):

1. Identification of the problem and definition of the goal: The aims and objectives of the project are determined.
2. Data Collection: Required data is collected from various sources.
3. Data preprocessing: Data is organized, and missing information is completed.

4. Data Discovery and Exploratory Analysis: Patterns and relationships are discovered within a data set.
5. Modeling: Create learning algorithms or statistical models.
6. Evaluation: Evaluate the created model using test data.
7. Deployment: Success models are integrated into business processes.
8. Monitoring and Maintenance: Model performance is monitored, updated, or maintained as necessary.

The development of artificial intelligence and machine learning, as well as well-informed decision-making processes, are facilitated by this organized approach to data mining.

3.3. Data mining methods

Data mining techniques refer to techniques and methods used to extract patterns from copious amounts of data and obtain meaningful information. These techniques can be applied to various fields, such as classification, regression, clustering, relational analysis, time series analysis, and anomaly detection. Data mining techniques aim to reveal hidden information in data sets by using fields such as mathematics, statistics, machine learning, and artificial intelligence (Papakyriakou & Barbounakis, 2022).

3.3.1. Clustering analysis

Understanding the patterns hidden in data sets is critical for the business and scientific worlds. Cluster analysis combines data points with similar characteristics in a data set to identify homogeneous groups. This analysis method is an excellent tool in data mining and is used in many areas (Zou, 2020).

3.3.1.1. Concept of clustering analysis

Cluster analysis is a statistical technique that divides similar data points in a data set into specific groups, or "clusters." These groups contain data points with similar characteristics. Cluster analysis aims for homogeneity within groups, but heterogeneity between different groups should be maximized (Zou, 2020).

3.3.1.2. Process order of clustering analysis

Cluster analysis typically follows the following sequence of operations (Oyelade et al., 2019):

1. Collecting and preparing data: The first step is to collect the data set for analysis and prepare this data using preprocessing steps. Operations such as filling in missing data, processing outliers, and cleaning unnecessary features are performed at this stage.
2. Feature Selection: Selecting the features to be used for analysis reduces the complexity of the dataset and allows the clustering algorithm to work more efficiently.
3. Data standardization: The data set is standardized to compensate for different scales between features and to obtain values of similar size.
4. Selection of clustering algorithm: The clustering algorithm appropriate to the data set is selected. K-means, Hierarchical clustering, DBSCAN, etc. Appropriate algorithms, such as popular algorithms, are preferred.
5. Determine the ideal number of clusters: Knowing the ideal number of clusters for k-means and similar algorithms is useful.
6. Clustering process: According to the selected algorithm, the data set is divided into clusters. Each cluster contains data points with similar characteristics.
7. Evaluation of the results: Clustering results are visualized and analyzed. At this stage, we must ensure that the clusters created are meaningful and consistent.
8. Applications of business results: The clustering results obtained can be used in various areas, such as optimizing business strategies, segmenting customers, and developing marketing strategies.

This process provides general steps for successfully performing cluster analysis. However, these steps can be changed or adjusted depending on the standard of the data set and the analysis goal.

3.4. Clustering analysis methods

Clustering involves dividing information into a data set and dividing it into groups based on certain proximity criteria. Each group is called a "cluster" (Gülagiz & Sahin, 2017). The methods of clustering analysis are as follows:

3.4.1. Hierarchical clustering methods

Hierarchical clustering best suits small samples (typically n < 250). Researchers can use these methods to identify similarities and decide to merge/split clusters. The methods of hierarchical clustering analysis are as follows (Dabhi & Patel, 2016):

1. Associative/Additive methods:
- Connection techniques: Single, Full, Average connection
- Variance techniques: Ward's method
- Centralization techniques: Median, Centroid
2. Discriminative/Divided methods:
- Split averages
- Automatic interference identification

Hierarchical clustering methods include addition (joining) and division (separation). Aggregation combines similar clustering techniques; partitioning, on the other hand, is based on large clusters representing all data points.

3.4.2. Non-hierarchical clustering methods

When the number of clusters is known, non-hierarchical clustering is employed. The methods of Non-hierarchical clustering analysis are as follows (Gülagiz & Sahin, 2017):

- K-means clustering
- K-medoids clustering
- Partitioning around medoids (PAM)
- Fuzzy C-means (FCM) method

These clustering techniques should be selected based on the specific problem context and data set characteristics. Knowing the number of clusters in advance can increase the accuracy of our analysis and give us more meaningful clustering results.

3.4.2.1. K-means algorithm

K-means analysis is a clustering algorithm that groups data points into a certain number of clusters. This algorithm aims to achieve uniform clustering by assigning data points with similar characteristics in the dataset to the same cluster. Since K-means is a grouping process performed without using labeled data, it falls into the unsupervised learning category (Ali & Kadhum, 2017). Implementation of this algorithm includes the following steps (Singh et al., 2011):

1. Initial estimate of cluster centers: A certain number of K cluster centers are randomly selected.
2. Each data point should be assigned to the closest cluster center: Based on the distance between K cluster centers, each data point is assigned to the closest cluster center.
3. Calculation of new cluster centers: New centers are calculated for each cluster. The cluster's center represents the mean of all the data points.
4. New centers are calculated: The second and third steps are repeated until the cluster's center is fixed. These steps are repeated until the data point is permanently assigned to the last assigned cluster.
5. Evaluation of the results: Clustering results should create meaningful clusters compatible with the determined K number.

This analysis is an effective research method used in many areas, especially data mining and segmentation.

3.5. ROC weight method

Using the rank order centroid (ROC) weigh method is a straightforward way to assign weights to criteria depending on their relative relevance. It is frequently easier for decision-makers to rank criteria than to weigh them. These ranking ranks are the input for this algorithm, which uses them to create weights for each criterion (Tamba et al., 2021). The following are the steps involved in this method (Roszkowska, 2013):

Step 1. All factors are ranked by decision-makers in order of relative relevance, from most to least useful.

Step 2. Equation (1) is employed to calculate the anticipated value of the weights when there are more than two criteria. The formula considers the rank positions of the criteria. For each criterion j, the weight is calculated using the formula:

$$wj(ROC) = \frac{1}{n}\sum_{k=j}^{n}\frac{1}{r_k} \quad (1)$$

Where n is the total number of criteria, r_k represents the rank position of the criteria, and $wj\ (ROC)$ is the weight assigned to the j-th criterion. The criteria weights are as follows: $w_1 \geq w_2 \geq ... \geq w_n$. The total of the criterion weights is 1.

Step 3. The arithmetic mean of the predicted values is calculated to determine the final criterion weights.

In this study, ROC and K-means algorithms were used synthetically. How these two methods are used together is explained with the flow chart in Figure 1.

4. Experimental Studies

On the " UC Irvine Machine Learning Repository" site, there is data related to direct marketing campaigns of a Portuguese banking institution. Telemarketing initiatives depended on calls. Often, multiple contacts with the same customer were required to assess whether to subscribe (yes) or not (no) to the product (term deposit at a bank) (Moro et al., 2014). Based on these data, the weight and importance of the criteria were calculated using the ROC method and the Microsoft Excel program. Then, data analysis was performed, and customers were divided into groups with the K-means clustering algorithm using the IBM SPSS Statistics program. These groups were created to conduct a residential complex marketing campaign. The study aims to analyze customer behavior and divide customers into groups for marketing purposes.

Input Variables, bank customer's data:

1. No: Customer
2. Age: (Numeric)
3. Job: Kind of work (admin., unknown, unemployed, management, housemaid, entrepreneur, student, blue-collar, self-employed, retired, technician, services)
4. Marital: Marital status (married, divorced, single; note: "divorced" denotes a widow or divorce)
5. Education: (Unknown, secondary, primary, tertiary)
6. Default: Possesses defaulted credit (yes, no)
7. Balance: Average annual sum, expressed in euros (numeric)
8. Housing: Possesses a mortgage (yes, no)
9. Loan: Possesses a personal loan (yes, no)
10. Previous: Quantity of connections made both for this client and before this campaign (numeric)
11. Poutcome: The outcome of the previous advertising effort (unknown, other, failure, success)

The original data for the customers are shown in Table 2 before they have been amended.

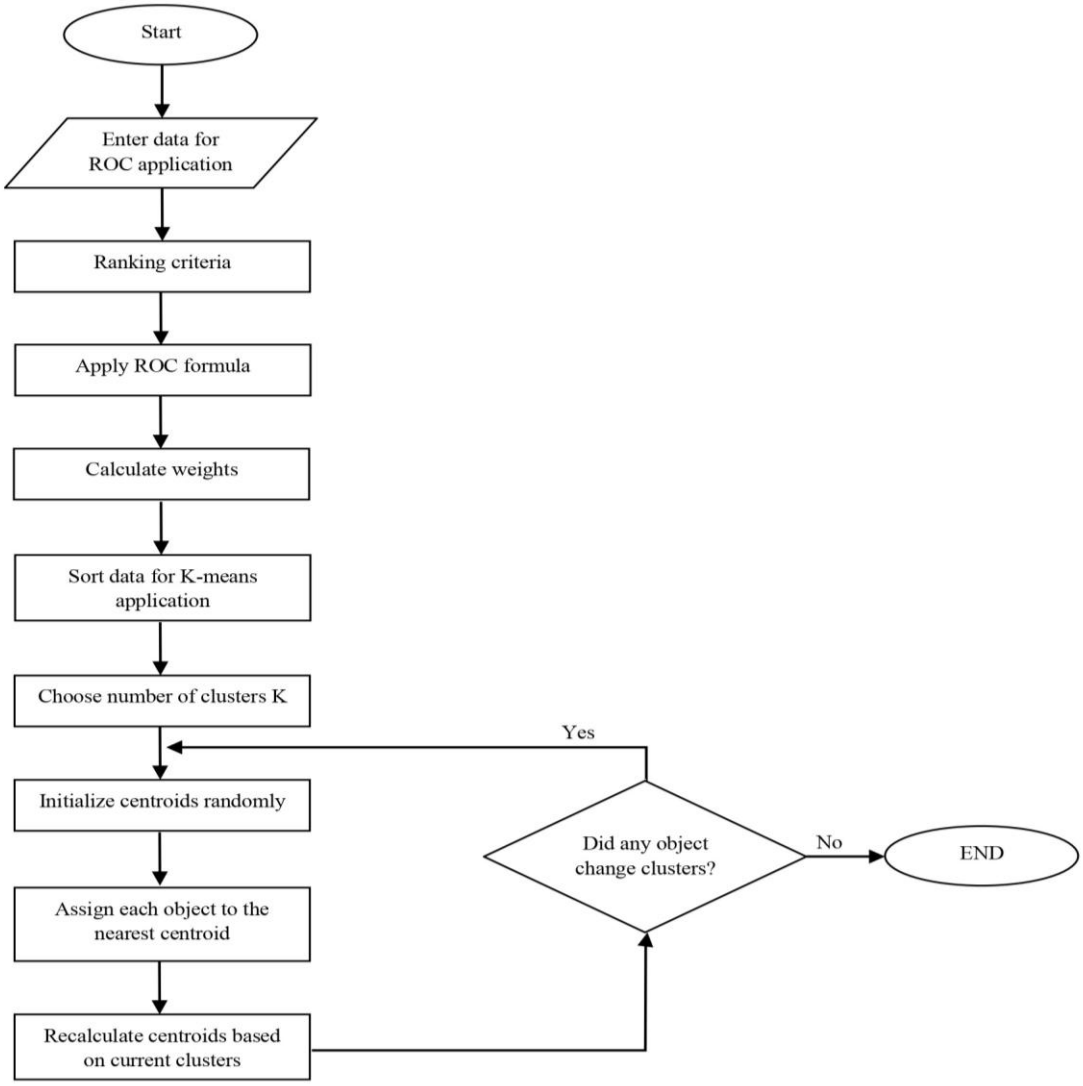


Figure 1. Flow chart of ROC method and K-means algorithm.

Table 2. Original customer data

No	Age	Job	Marital	Education	Default	Balance	Housing	Loan	Previous	Poutcome
1	30	unemployed	married	primary	no	1 787	no	no	0	unknown
2	33	services	married	secondary	no	4 789	yes	yes	4	failure
3	35	management	single	tertiary	no	1 350	yes	no	1	failure
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
4 519	57	technician	married	secondary	no	295	no	no	0	unknown
4 520	28	blue-collar	married	secondary	no	1 137	no	no	3	other
4 521	44	entrepreneur	single	tertiary	no	1 136	yes	yes	7	other

The verbal concepts applied in this analysis were converted into numerical values to begin the analysis. The variable numbers for which this change was made are 6, 8, 9, and 11. “1” was used for “No” and “0” for “Yes” responses. As for the 11th variable, Poutcome, has been calculated as 1 for "success" responses and 0 for other cases. Table 3 shows a list of analyzable customer data.

Table 3. Analyzable customer data

No	Age	Job	Marital	Education	Default	Balance	Housing	Loan	Previous	Poutcome
1	30	unemployed	married	primary	1	1 787	1	1	0	0
2	33	services	married	secondary	1	4 789	0	0	4	0
3	35	management	single	tertiary	1	1 350	0	1	1	0
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
4 519	57	technician	married	secondary	1	295	1	1	0	0
4 520	28	blue-collar	married	secondary	1	1 137	1	1	3	0
4 521	44	entrepreneur	single	tertiary	1	1 136	0	0	7	0
		MAX			1	71 188	1	1	25	1

4.1. ROC method application

The ROC method calculates the importance and weight of the criteria. It can be summarized in three steps:

Step 1. Using the linear normalization matrix, the decision matrix is normalized by considering the benefit or cost-oriented nature of the criterion. All values entered in this study were applied as benefits by dividing each value by the highest value among them. The result of the process is shown in Table 4 below.

Table 4. Normalized decision matrix

No	Age	Job	Marital	Education	Default	Balance	Housing	Loan	Previous	Poutcome
1	30	unemployed	married	primary	1	0.025103	1	1	0	0
2	33	services	married	secondary	1	0.067273	0	0	0.16	0
3	35	management	single	tertiary	1	0.018964	0	1	0.04	0
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
4 519	57	technician	married	secondary	1	0.004144	1	1	0	0
4 520	28	blue-collar	married	secondary	1	0.015972	1	1	0.12	0
4 521	44	entrepreneur	single	tertiary	1	0.015958	0	0	0.28	0

Step 2. In this step, the importance of the criteria was determined by reviewing the literature and choosing the best ranking. The arrangement is as follows. 1. Balance, 2. Poutcome, 3. Default, 4. Housing, 5. Loan, 6. Previous.

The weights are analyzed using the ROC method, and the results are sorted in Table 5.

Table 5. ROC weighting results

Criteria(C)	C1	C2	C3	C4	C5	C6	Total
Alignment	1	2	3	4	5	6	
Weight (W)	0.408333	0.241667	0.158333	0.102778	0.061111	0.027778	1

Step 3. The criterion weights acquired by the ROC method are multiplied by the normalized matrix to calculate the weight of the matrix in Table 6.

Table 6. Weighted normalized ROC matrix

					W					
					0.1583	0.4083	0.1027	0.0611	0.0277	0.2416
No	Age	Job	Marital	Education	Default	Balance	Housing	Loan	Previous	Poutcome
1	30	unemployed	married	primary	0.1583	0.0102	0.1027	0.0611	0	0
2	33	services	married	secondary	0.1583	0.0274	0	0	0.0044	0
3	35	management	single	tertiary	0.1583	0.0077	0	0.0611	0.0011	0
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
4 519	57	technician	married	secondary	0.1583	0.0016	0.1027	0.0611	0	0
4 520	28	blue-collar	married	secondary	0.1583	0.0065	0.1027	0.0611	0.0033	0
4 521	44	entrepreneur	single	tertiary	0.1583	0.0065	0	0	0.0077	0

4.2. K-means algorithm application

K-means algorithm, a non-hierarchical clustering method, is used. The algorithm identifies a certain number of 'K' clusters depending on the problem. At the start of the algorithm, an object representing each cluster is randomly selected. Each object is assigned to the nearest cluster, and the cluster mean is calculated using the clustering criteria. These calculated averages are used as new cluster points, and each object is placed in the cluster most closely resembles it. This process continues until no further changes are observed in the cluster. Customers were divided into groups using the K-means clustering algorithm using the IBM SPSS Statistics program. Part of the dataset in the IBM SPSS Statistics program is shown in Table 7.

Table 7. Part of the dataset in the IBM SPSS Statistics program

No	Age	Job	Marital	Education	Default	Balance	Housing	Loan	Previous	Poutcome	Mean
1	30	unemployed	married	primary	0.1583	0.0102	0.1027	0.0611	0	0	0.06
2	33	services	married	secondary	0.1583	0.0274	0	0	0.0044	0	0.03
3	35	management	single	tertiary	0.1583	0.0077	0	0.0611	0.0011	0	0.04
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
4 519	57	technician	married	secondary	0.1583	0.0016	0.1027	0.0611	0	0	0.05
4 520	28	blue-collar	married	secondary	0.1583	0.0065	0.1027	0.0611	0.0033	0	0.06
4 521	44	entrepreneur	single	tertiary	0.1583	0.0065	0	0	0.0077	0	0.03

Each customer's score was averaged, and then the IBM SPSS Statistics program was used to distribute customers into four groups. Finding each cluster's initial center is one of the most important steps, and the program does this automatically. Depending on the method used, these initial centers may be predefined or determined according to specific criteria or random selection. Initial Cluster Centers are shown in Table 8.

Table 8. Initial cluster centers

	Cluster			
	1	2	3	4
Mean	0.00	0.12	0.01	0.11

These starting points represent the beginning of the cluster creation process within the clustering algorithm running on the data set. Algorithms such as K-means clustering determine these starting centers, update them repeatedly, assign data points to clusters, and perform the clustering process.

In Table 9, the distance to the cluster center determined by the clustering algorithm when creating each cluster is a useful measure of the elements of the cluster. This distance measure describes the location of a data point within a given cluster relative to the center of that cluster. This analysis is useful to understand the distribution of elements in each cluster and their relationship to the cluster center.

Table 9. Cluster membership

Case number	NO	Cluster	Distance
1	4 518.00	1	0.028
2	857.00	1	0.026
3	1 959.00	1	0.026
.	.	.	.
.	.	.	.
.	.	.	.
4 519	1 604.00	2	0.017
4 520	1 217.00	2	0.019
4 521	3 701.00	2	0.031

Table 10 refers to the final center of each cluster determined by the clustering algorithm. This is the stage when the clustering process ends and cluster centers stabilize.

Table 10. Final cluster centers

	Cluster			
	1	2	3	4
Mean	0.02	0.09	0.04	0.06

The final center of each cluster is an intermediate point determined by the applied clustering algorithm and fixed in the final stage of the clustering process. Especially when iterative algorithms such as K-means clustering are used, cluster centers are updated at each iteration, and these updates form the final centers when the clustering process reaches a stable result. The final centroid is useful for understanding each cluster's features, the distribution of data points, and the relationships within the

cluster. These final centers are used to analyze and interpret the clustering results after completing the clustering process, as they reflect the characteristics that represent each cluster.

The number of items in each cluster in Table 11 reflects the group size determined by the clustering algorithm. This information is important for understanding clustering results and assessing the representation of each cluster in the dataset. Item counts are used to evaluate the heterogeneity or homogeneity of clustering results, understand group spacing, and evaluate the performance of clustering algorithms.

Table 11. Number of cases in each cluster

Cluster	Number of groups	Groups sorted by score
1	452	4
2	133	1
3	2 308	3
4	1 628	2
Total	4 521	-

Figure 2 shows the distribution of customers into four diverse groups based on data analysis using the K-means algorithm. Each section in the circle represents a certain percentage of customers in each group.

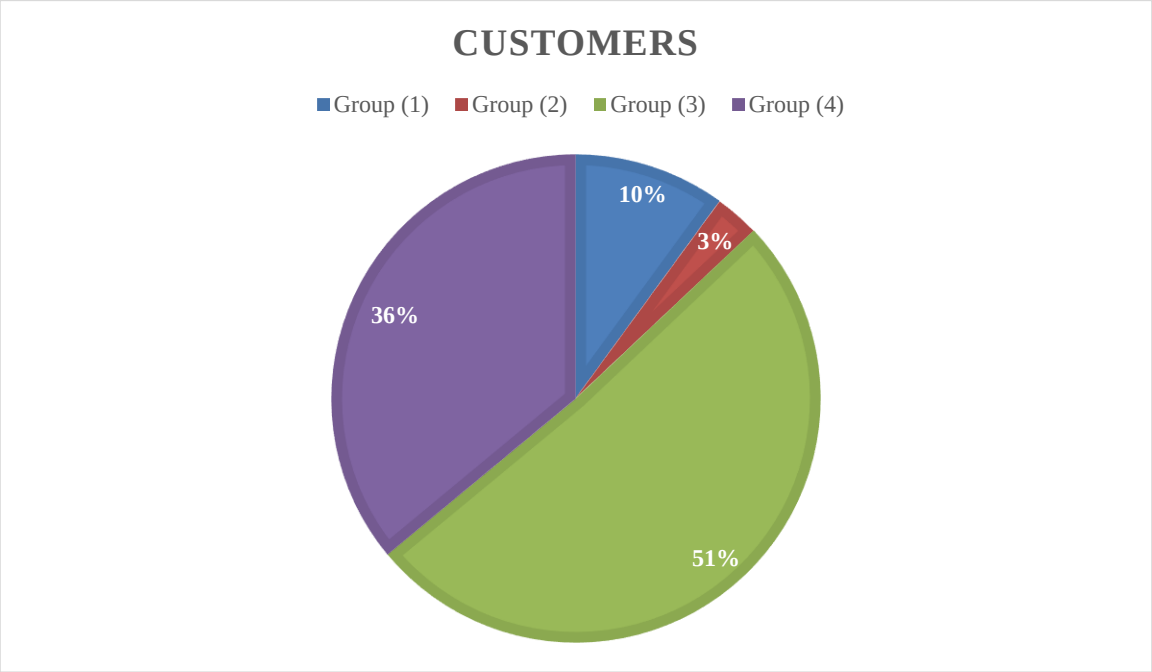


Figure 2. Percentage of customers in each group.

After analyzing the bank customers' data and dividing them into four diverse groups, all members of each group were identified. It has been observed that the members of Group (2), Group (4), Group (3) and Group (1) are ordered from highest to lowest in descending order according to the scores collected for each. In this context, the classification obtained by dividing customers into segments and recording each segment provides valuable information regarding understanding customer behavior and characteristics and executing and focusing marketing campaigns suitable for each group. High-scoring groups can be used specifically to develop specific and miscellaneous marketing strategies and manage Marketing campaigns management more effectively.

5. Conclusions and Recommendations

This study presents an integrated methodology utilizing the ROC (Rank Order Centroid) method for weighting decision criteria and the K-means clustering algorithm for customer segmentation, tailored specifically for the construction industry. By combining these two approaches, the research aims to simplify decision-making processes and optimize marketing strategies in residential projects. The results demonstrated that the most effective marketing strategies can be achieved by focusing on high-scoring customer segments identified through this integrated approach.

As a result of analyzing the data of banking customers and dividing them into four distinct groups, each group's members were identified. According to the scores obtained, Groups (2), (4), (3), and (1) were ranked from the highest to the lowest, respectively. High-scoring groups, particularly Group (2), with 133 customers, and Group (4), with 1 628 customers, represent high-value customer segments that can be prioritized in marketing strategies to achieve efficiency. By focusing on these groups, can conduct more precise and efficient marketing campaigns. This classification provides valuable insights into customer behavior, allowing for customized marketing campaigns tailored to the characteristics of each group.

The practical application of this methodology underscores the critical role of big data analysis in modern marketing. Using large datasets and advanced analytical techniques provides a deeper understanding of customer behaviors and preferences, paving the way for more precise and efficient marketing strategies. This approach not only enhances decision-making capabilities but also contributes to better financial performance and customer satisfaction in the construction sector.

Future research could further explore the potential of integrating additional data mining methods with other multi-criteria decision-making approaches to enhance the precision of customer segmentation. Additionally, applying this methodology to different industries or combining it with predictive analytics techniques could yield deeper insights and contribute to the development of dynamic marketing strategies that adapt in real-time to changes in customer behavior and market trends.

References

Ali, H. H., & Kadhum, L. E. (2017). K-means clustering algorithm applications in data mining and pattern recognition. *International Journal of Science and Research (IJSR)*, *6*(8), 1577–1584. https://doi.org/10.21275/ART20176024

Bai, C., Dhavale, D., & Sarkis, J. (2014). Integrating Fuzzy C-Means and TOPSIS for performance evaluation: An application and comparative analysis. *Expert Systems with Applications*, *41*(9), 4186–4196. https://doi.org/10.1016/j.eswa.2013.12.037

Bajpai, V., Pandey, S., & Shriwas, S. (2012). Social media marketing: Strategies & its impact. *International Journal of Social Science & Interdisciplinary Research*, *1*(7), 214–223.

Bilici, Z., & Özdemir, D. (2022). Data mining studies in education: Literature review for the years 2014-2020. *Bayburt Eğitim Fakültesi Dergisi*, *17*(33), 342–376. https://doi.org/10.35675/befdergi.849973

Chiang, W.-Y. (2011). To mine association rules of customer values via a data mining procedure with improved model: An empirical case study. *Expert Systems with Applications*, *38*(3), 1716–1722. https://doi.org/10.1016/j.eswa.2010.07.097

Dabhi, D. P., & Patel, M. R. (2016). Extensive survey on hierarchical clustering methods in data mining. *International Research Journal of Engineering and Technology (IRJET)*, *3*(11), 659–665.

Guo, F., & Qin, H. (2017). Data mining techniques for customer relationship management. *Journal of Physics: Conference Series*, *910*(1), 012021. https://doi.org/10.1088/1742-6596/910/1/012021

Gülagiz, F. K., & Sahin, S. (2017). Comparison of hierarchical and non-hierarchical clustering algorithms. *International Journal of Computer Engineering and Information Technology*, *9*(1), 6.

Gürbüz, T., Albayrak, Y. E., & Alaybeyoğlu, E. (2014). Criteria Weighting and 4P's Planning in Marketing Using a Fuzzy Metric Distance and AHP Hybrid Method: *International Journal of Computational Intelligence Systems*, *7*(Supplement 1), 94.

Heryati, A., & Herdiansyah, M. I. (2020). The Application of Data Mining by using K-Means Clustering Method in Determining New Students' Admission Promotion Strategy. *International Journal of*

Engineering and Advanced Technology, *9*(3), 824–833. https://doi.org/10.35940/ijeat.C5414.029320

Hung, Y.-H., Huang, T.-L., Hsieh, J.-C., Tsuei, H.-J., Cheng, C.-C., & Tzeng, G.-H. (2012). Online reputation management for improving marketing by using a hybrid MCDM model. *Knowledge-Based Systems*, *35*, 87–93. https://doi.org/10.1016/j.knosys.2012.03.004

Hutagalung, J., Syahril, M., & Sobirin, S. (2022). Implementation of K-Medoids Clustering Method for Indihome Service Package Market Segmentation. *Journal of Computer Networks, Architecture and High Performance Computing*, *4*(2), 137–147. https://doi.org/10.47709/cnahpc.v4i2.1458

Kuo, R. J., Amornnikun, P., & Nguyen, T. P. Q. (2020). Metaheuristic-based possibilistic multivariate fuzzy weighted *c*-means algorithms for market segmentation. *Applied Soft Computing*, *96*, 106639. https://doi.org/10.1016/j.asoc.2020.106639

Maciejewski, G., Mokrysz, S., & Wróblewski, \Lukasz. (2019). Segmentation of coffee consumers using sustainable values: Cluster analysis on the Polish coffee market. *Sustainability*, *11*(3), 613. https://doi.org/10.3390/su11030613

Madni, H. A., Anwar, Z., & Shah, M. A. (2017). Data mining techniques and applications—A decade review. *2017 23rd International Conference on Automation and Computing (ICAC)*, 1–7.

Maulina, N. R., Surjandari, I., & Rus, A. M. M. (2019). Data mining approach for customer segmentation in B2B settings using centroid-based clustering. *2019 16th International Conference on Service Systems and Service Management (ICSSSM)*, 1–6.

Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, *62*. https://doi.org/10.1016/j.dss.2014.03.001

Oyelade, J., Isewon, I., Oladipupo, O., Emebo, O., Omogbadegun, Z., Aromolaran, O., Uwoghiren, E., Olaniyan, D., & Olawole, O. (2019). Data clustering: Algorithms and its applications. *2019 19th International Conference on Computational Science and Its Applications (ICCSA)*, 71–81. https://doi.org/10.1109/ICCSA.2019.000-1

Pande, S. R., Sambare, S. S., & Thakre, V. M. (2012). Data clustering using data mining techniques. *International Journal of Advanced Research in Computer and Communication Engineering*, *1*(8), 494–499.

Pandiangan, I. M., Mesran, M., Borman, R. I., Windarto, A. P., & Setiawansyah, S. (2023). Implementation of Operational Competitiveness Rating Analysis (OCRA) and Rank Order Centroid (ROC) to Determination of Minimarket Location. *Bulletin of Informatics and Data Science*, *2*(1), 1–8. http://dx.doi.org/10.61944/bids.v2i1.62

Papakyriakou, D., & Barbounakis, I. S. (2022). Data mining methods: A review. *International Journal of Computer Application*, *183*(48), 5–19. https://doi.org/10.5120/ijca2022921884

Phillips-Wren, G., Doran, R., & Merrill, K. (2016). Creating a value proposition with a social media strategy for talent acquisition. *Journal of Decision Systems*, *25*(sup1), 450–462. https://doi.org/10.1080/12460125.2016.1187398

Rohmawati, T., & Winata, H. (2021). Information technology for modern marketing. *International Journal of Research and Applied Technology (INJURATECH)*, *1*(1), 90–96.

Roszkowska, E. (2013). Rank ordering criteria weighting methods–a comparative overview. *Optimum. Studia Ekonomiczne*, *5 (65)*, 14–33.

Sharma, M. (2014). Data mining: A literature survey. *International Journal of Emerging Research in Management & Technology*, *3*(2).

Singh, K., Malik, D., & Sharma, N. (2011). Evolving limitations in K-means algorithm in data mining and their removal. *International Journal of Computational Engineering & Management*, *12*(1), 105–109.

Singh, M., & Pant, M. (2021). A review of selected weighing methods in MCDM with a case study. *International Journal of System Assurance Engineering and Management*, *12*, 126–144. https://doi.org/10.1007/s13198-020-01033-3

Tacoli, C., McGranahan, G., & Satterthwaite, D. (2015). *Urbanisation, rural-urban migration and urban poverty*. JSTOR.

Tamba, S. P., Purba, A., Kusuma, Y. E., Vidyastuti, M. A. S., & Dharma, S. (2021). Implementation of the rank order centroid (roc) method to determine the favorite betta fish. *INFOKUM*, *9*(2, June), 381–386.

Tleis, M., Callieris, R., & Roma, R. (2017). Segmenting the organic food market in Lebanon: An application of k-means cluster analysis. *British Food Journal*, *119*(7), 1423–1441. https://doi.org/10.1108/BFJ-08-2016-0354

Wijaya, B. K., Sudipa, I. G. I., Waas, D. V., & Santika, P. P. (2022). Selection of Online Sales Platforms for MSMEs using the OCRA Method with ROC Weighting. *Journal of Intelligent Decision Support System (IDSS)*, *5*(4), 146–152.

Wirtz, B. W., & Daiser, P. (2018). Business model development: A customer-oriented perspective. *Journal of Business Models*, *6*(3), 24–44.

Wu, W.-T., Li, Y.-J., Feng, A.-Z., Li, L., Huang, T., Xu, A.-D., & Lyu, J. (2021). Data mining in clinical big data: The frequently used databases, steps, and methodological models. *Military Medical Research*, *8*(1), 44. https://doi.org/10.1186/s40779-021-00338-z

Zou, H. (2020). Clustering Algorithm and Its Application in Data Mining. *Wireless Personal Communications*, *110*(1), 21–30. https://doi.org/10.1007/s11277-019-06709-z