International Journal of Management Economics and Business, Vol. 19, No. 2, 2023
Uluslararası Yönetim İktisat ve İşletme Dergisi, Cilt 19, Sayı 2, 2023

Research Article / *Araştırma Makalesi*

SMART PHONE MARKET ANALYSIS WITH APRIORI ALGORITHM: THE CASE OF TURKEY

Mehmet ÖZÇALICI[1], Ayşe Nur SOYSAL BİLMİŞ[2]

ABSTRACT

With the disappearance of borders among countries, competition has been moved to an international level. With customer-oriented marketing understanding, customers have become the most important assets of enterprises. With the global continuation of competition, carrying out different marketing activities for their current and potential customers has become a necessity for enterprises to ensure continuity. At the point of being able to carry out correct delivery regarding the product and service preferences of the customers, the enterprises' need for further customer data has emerged. In the context of reaching bigger data and interpreting the data, data mining has become an important topic for enterprises Analyzing customer behavior with big data helps businesses in their product and service offerings. This study focuses on examining the technical features of the devices offered for sale in the smartphone market. With the help of the a-priori algorithm, which technical features appear together were separately analyzed for both iOS and Android operating system devices. For the applicability of the a-priori algorithm, discretization of the dataset should be ensured. Different number of categories are applied and the number of categories that has the highest support value is used to categorize each of the sets. It is found that the feature with the highest correlation to a phone's price is screen brightness. The achieved results will be helpful for a clearer understanding of the features of the products available in the mobile phone market.

Keywords: *Apriori Algorithm, Market Basket Analysis, Smart Phone Market.*

JEL Classification Codes: *C60, D12, M30*

APRIORI ALGORİTMASI İLE AKILLI TELEFON PİYASASI ANALİZİ: TÜRKİYE ÖRNEĞİ

ÖZET

Ülkeler arasındaki sınırların ortadan kalkması ile birlikte rekabet uluslararası düzeye taşınmıştır. Müşteri odaklı pazarlama anlayışı ile birlikte müşteriler işletmelerin en önemli varlıklarından olmuştur. Rekabetin küresel bağlamda devam etmesi ile birlikte işletmelerin devamlılığı sağlamak adına mevcut müşterileri ve potansiyel müşterileri için farklı pazarlama faaliyetlerinde bulunması zorunluluk haline gelmiştir. Müşterilerin ürün ve hizmet tercihlerine ilişkin doğru sunumlar gerçekleştirebilme noktasında işletmelerin daha fazla müşteri verisine ihtiyaç ortaya çıkmıştır. Daha fazla veriye ulaşma bağlamında ve o verileri yorumlama açısından veri madenciliği son yıllarda işletmeler tarafından önemli bir konu haline gelmiştir. Büyük veriler ile müşteri davranışlarını analiz etme, işletmelere ürün ve hizmet sunumlarında

[1] Assoc. Prof., Kilis 7 Aralık University, FEAS, Kilis, Turkey, *mozcalici@kilis.edu.tr*
[2] Teaching Asst. PhD., Akdeniz University, Korkuteli VS, Antalya, Turkey, *nursoysal@akdeniz.edu.tr*

www.ijmeb.org ISSN:2147-9208 E-ISSN:2147-9194
http://dx.doi.org/10.17130/ijmeb.1191680
Received: 19.10.2022, Accepted: 17.04.2023

yardımcı olmaktadır. Bu çalışma akıllı telefon piyasasında satışa sunulan cihazların teknik özelliklerini incelemeye odaklanmaktadır. A-priori algoritması yardımıyla hangi teknik özelliklerin bir arada bulunduğu hem IOS hem de Android işletim sistemine sahip cihazlar için ayrı ayrı incelenmiştir. A-priori algoritmasının uygulanabilirliği için veri setinin ayrıklaştırılması sağlanmalıdır.En yüksek destek değerine sahip kategori sayısı, kümelerin her birini kategorize etmek için kullanılmıştır. Bir telefonun fiyatı ile en yüksek korelasyona sahip özelliğin ekran parlaklığı olduğu tespit edilmiştir. Erişilen sonuçlar, cep telefonu pazarında bulunan ürünlerin özelliklerinin daha net anlaşılmasına yardımcı olacaktır.

Anahtar Kelimeler: *Apriori Algoritması, Pazar Sepet Analizi, Akıllı Telefon Piyasası.*

1. Introduction

Nowadays, the development of data production and collection ability ensures the reaching for a sheer volume of data. Nevertheless, to compete effectively, enterprises are required to benefit in time from the high-income opportunities. As a result, the enterprises' necessity for benefiting from the data stacks they produced or collected during their daily routine activities and for converting the processed data into beneficial information arises. However, the difficulty of understanding the value of the information prevents many enterprises from benefiting properly from the data richness available in their hands (Ha & Park, 1998:1). According to modern marketing understanding, customers have been perceived as important assets for enterprises. Therefore, enterprises must acquire new customers successfully and retain high-value customers. To reach these aims, many enterprises need a considerable amount of data, which will be analyzable and applicable for developing new business strategies and opportunities thereafter, arises. (Hsieh, 2004:623). In this regard, data mining has become a topic that attracts attention in relevant research fields such as database systems, statistics, artificial intelligence and machine learning.

Data mining, also called knowledge discovery in databases, is a science emerging with the application of modern statistical and artificial intelligence technologies to the extraction of valid, previously unknown, understandable and actionable knowledge from big databases and their usage for important tasks. Thus, databases serve as rich and reliable sources for knowledge production and verification (Ha & Park, 1998:1). Data mining is an important topic in producing beneficial knowledge from data, finding hidden and unknown customer information from a wealth of customer data and thus obtaining productive results. When the enterprises understand the demands and needs of the customers, they can develop correct marketing strategies for their customers, which may form repurchase intention by increasing customer satisfaction (Wei et al., 2013:7514).

Nowadays, communication has an important place across the world. Communication tools are of a great variety, therefore, the designers and the producers of these facilities constantly try to produce new products and keep them up-to-date. One of the communication tools commonly used is mobile phones (Mostafavi et al., 2013:202). All around the world, mobile phones have become indispensable for socio-economic development. In many distant regions in developing countries, access to communication means, health services, postal services, financial services etc., is limited. However, the mobile phone represents the first modern infrastructure available in these inner regions. The mobile phone market has undergone a rapid change from its establishment till today. In the beginning, acceptance of these mobile phones showed an annual increase at the ratio of 25-35 percent in the period of 1983-1997. (Ahmad et al., 2019; Hausman, 1999). Also, severe competition emerged between the mobile phone producers whose market shares significantly changed month by month. (Dewenter et al., 2007:4). With the developments in the technology field, enterprises can obtain and store a lot of information with regard to their customers, at reasonable costs. Besides this easiness, one of the difficulties faced by the companies making huge investments in customer data collection is how they are going to extract important information from the large customer databases to achieve a competitive advantage.

Market basket analysis (association rule) is the method of discovery for the customer purchasing models through extracting association formations from the transaction databases of the stores. For instance; the presence of the likelihood of the fact that the supermarket customers purchase milk, bread and cheese together or the likelihood of the fact that the bank customers jointly use a series of service, may help the managers in designing store arrangement, websites, product mix, packing and other marketing strategies (Chen et al., 2005:339). Regarding the phone market, market sales for each phone vary depending on the preferences of the concerned consumers, with a series of features available within them. The issue of how the demand for a phone model depends on the value that the consumers attach to each feature is on the agenda (Montenegro & Torres, 2016:3). Mobile phones are currently seen as a necessity, their industry has been rapidly developing and characterised by rapid technological developments. Mobile phones are not simple voice communication devices anymore, and they contain lots of high technology features within a single hand device. Apart from their basic function, phones can also be used as a video recorder, camera, video game console, radio, television, GPS device etc. Many mobile phone brands are available in the market, and each brand has various models. Furthermore, there is severe competition among mobile phone producers. Information on consumers' preferences for various qualities is crucial for the investments intended for the future in the rapidly changing world (Ahmad et al., 2019:299-300).

This study focuses on analyzing the technical features of smartphones available in the market using association rules. By doing so, it aims to reveal important data about the features present in the phones available. What sets this study apart from others is that, according to the literature review conducted, association rules have not been used for this purpose before. This is considered to be a significant gap because identifying co-existing technical features could help users make more informed decisions when purchasing a smartphone. The study primarily included the findings of the similar research conducted with respect to the market basket analysis (association rule). After that, it included the topic explanation and the analysis findings. The study is composed of 4 sections. Following this introduction section, section 2 summarizes the studies about smart phone and association rules in the literature. Section 3 summarizes the working principles of the a-priori algorithm. Section 4 introduces the dataset and includes the analysis results.

2. Literature Review

Smart phones are high technology products and they have a great number of technical and performance features. These features render them into products which are differentiated with a series of alternative design and sales price in the market. Market prices of phones vary depending on consumer preferences for a series of features available in them (Ahmad et al., 2019:300). Vendors should know about consumer preferences and their interest areas to be successful in the business. Mobile phones are designed based on consumer preferences and interests (Mostafavi et al., 2013:202). Retaining current customers and attracting unsatisfied customers of the rivals through increasing satisfaction are strategical necessities for the companies at the maturity stage. Apart from customer satisfaction, other factors in the industry also play an essential role as the identifiers of customer loyalty.

For this reason, it is crucially important to understand customer loyalty in the smartphone market and to identify the factors (including identifiers and moderators) influencing customer loyalty (Kim et al., 2016:937). Data mining is commonly used by various companies and institutions in the scientific fields such as medicine, retail, marketing, finance and communication in particular. This situation enables the companies to obtain information on diseases, sales behaviors, customer satisfaction and corporate profits. Through data mining, institutions can increase the profitability of their interactions with the customers and improve risk management in marketing. Furthermore, the patterns extracted using data mining help the institutions make better decisions. (Ziafat & Shakeri, 2014:71). On the other hand, data mining techniques can discover patterns that the most experienced business people might miss out. As a result, the combination of the expertise in the sector with the power of the data mining techniques can be helpful for the institutions to gain a competitive advantage in their endeavors of optimizing customer management (Ziafat & Shakeri, 2014:71).

Market basket analysis determines the relations among various customers' products add to their shopping baskets and examines the customer purchasing models. Determination of such associations may be helpful for retailers in expanding their marketing strategies by gaining insight into which products are frequently jointly purchased by the customers (Gupta & Mamtora, 2014:409-410). The main idea in market basket analysis is finding the item sets commonly observed in big basket examples, based on the assumption that purchasing one or more item inside a dataset will lead to purchasing the rest. This situation provides easiness for cross-sales, packing, product positioning etc. The basic assumption in market basket analysis is that the co-existence of two or more items in most of the baskets implies that these items are complementary in purchasing (if not in consumption) and thus, purchasing one of them will lead to purchasing the others (Kamakura, 2012:505). In addition to this, market basket analysis has been used in many fields. This section of the study included the research conducted about the topic to identify in which different fields the market basket analysis is applied.

A-priori algorithm is used for different purposes. For example, Fagerlind et al. (2022) used the Market basket analysis to look at the relationships between injuries in traffic accidents. In this study, it was anticipated that some injuries would be connected. To ascertain the statistical association between injury data gathered from various sources, it was examined using the apriori technique. Samboteng et al. (2022) used the market basket analysis to identify the most essential plant-based food inventory for food processing. Wang et al. (2022) discuss researches that use association rules to address the issue of the relationship between production rules and products in the vehicle manufacturing sector. The outcomes demonstrate that over 87% accuracy is achieved when data mining association rules are applied. Administrators may make better judgments and cut down on the amount of time it takes to produce in the IoT environment by using the production rules provided by the data mining application. Pradhan et al. (2022) used the Market basket analysis in order to increase sales to merchants and maximize profitability (efficient and less efficient customers). Sudirman et al. (2021) used the market basket analysis for businesses in the tourism sector in their study.

To determine the variations in decision-making in the selection of tourist destinations, the data gathered by the questionnaire were evaluated. With the use of the survey results, it was determined that the Association Rule can be used to choose tourism destinations. Unvan (2021) used association rules to perform a market basket analysis for his research. The author examined information on a store's website. The analyses produced findings about the positioning of product shelves in supermarkets. In their study, Leote et al. (2020) suggested using Market Basket Analysis (MBA), a data mining method designed to analyse huge amounts of data and mathematically comparable to Indicator Value Analysis (IndVal), which is utilized in ecological monitoring research.As predicted, market basket analysis was successful in selecting ecological indicators in datasets where IndVal failed, as well as ecological indicators without species preselection. The authors draw the conclusion that under such circumstances and using objective association criteria, market basket analysis can be utilized to choose ecological indicators that can subsequently be processed and summarized using the IndVal technique. Additionally, it briefly describes Market basket analysis' potential to supplement IndVal and addresses the benefits and drawbacks of applying it to ecological indicators.

In their study, Santarcangelo et al. (2018) used the visual market basket analysis to observe the customers' shopping behaviors in the store. Visual market basket analysis is based on the analysis of the images obtained by the cameras mounted to the shopping baskets. They collected the data through the cameras mounted to the baskets, during real shopping sessions in a retail store. Videos were tagged according to a hierarchy formed of 14 different customer behaviors from the beginning till the end of the shopping. Different representation and classification techniques were compared, and a multimode method benefiting from visual, motion and audio identifiers was suggested. Moreover, it was suggested that the obtained results may be combined with the classical market basket analysis knowledge to help the retailers improve field management and marketing strategies.

Monteserin & Armentano (2018) examined the market basket analysis with an approach suggestion based on the social influence concept. From the point of view that the customers who purchased the products exposed to social influence would be influenced in purchasing the other products as well, they analyzed the data with the dataset obtained from online shopping and the dataset collected from a supermarket. According to the approach they suggested, they concluded that the items they determined increased the influence expansion as per different bases (Best-selling products, most promoted products, etc.).

By explaining the association rule in their study, Kaur & Kang (2016) provided a new algorithm that could aid in analyzing consumer behavior and boosting sales.

In their study, Gupta & Mamtora (2014) presented a survey on market basket analysis's current data mining algorithm. Analysis of the current algorithms demonstrated that usage of association rule algorithms for market basket analysis will be helpful for a better classification of a tremendous amount of data. They asserted that the a-priori algorithm can be efficiently modified to decrease time complexity and increase accuracy.

In their study, Raorane et al. (2012) used the market basket analysis to interpret consumer behaviors and give the correct decision that will provide advantage against the rivals by analyzing a huge amount of data. Experimental analysis was carried out by using association rules. As a result of the analysis, it was observed that data mining tools can be efficiently used for optimizing the patterns related to the dynamic behaviors of the transactions that customers do while purchasing specific products. It was observed that a specific relation exists among the products purchased by customers. Moreover, it was reached to the conclusion that association analysis can be used in terms of managing product placement on supermarket shelves.

Setiabudi et al. (2011) used market basket analysis in their studies to analyse purchasing habits of the customers. Analysis was conducted with the data obtained from a mini market. As a result of the analysis, they showed that the application could produce the data of what kind of products are frequently purchased by the customers at the same time according to hybrid-dimension association rules criteria. It was reached to the conclusion that there exists a correlation between the association rules. Besides, it was stated that the findings obtained from the research would provide additional benefit to the decision-making processes of the market owners.

In their study, Pande & Abdel-Aty (2009) applied market basket analysis to the data they obtained regarding the accident rates. To determine the interdependence between the specialties of the accidents, the accidents were analyzed as supermarket transactions. Analysis results include the basic rules showing which crash specialties are associated with each other. The application data are the data received from the state of Florida in 2004. In the study, as per the association rules, it was reached that there is a meaningful relation between lack of illumination and highness of crash intensity. Moreover, it was found out that straight sections with vertical curves tend to crash in rainy weather.

From the point of view that the current methods used in market basket analysis might be unsuccessful in the discovery of the critical purchasing models in a multiple store environment due to the assumption that the products in question are always available on the shelf in all the stores; in their study, Chen et al. (2005) suggested a new method to overcome this weakness. They concluded that the method they suggested is productive in terms of calculation and that it is more advantageous than the traditional method when the sizes of the stores are different, the product mix changes rapidly in the course of time, and more stores and periods are taken into consideration.

In the literature review, it was determined that the a priori algorithm was successfully used for different purposes. However, it has not yet been used to analyze the technical specifications of the phones. This study aimed to apply the apriori algorithm to a data set obtained from the mobile phone market.

3. Methodology

3.1. Data Mining

Data mining is the process of extracting interesting patterns from a huge amount of data during the course of knowledge discovery. It is related to solving problems by analyzing existing data in the data clusters. In addition, it provides tools for developing models in terms of forecasting the tendencies and behaviors in the future from the historical data (Ziafat & Shakeri, 2014:71). Data mining is commonly used to analyse customer behaviours and support business decisions in the context of marketing (Santarcangelo et al., 2018:84). However, the usage of data mining in marketing is quite recent. Even though this field is rapidly expanding, data mining is still an unfamiliar topic for many marketers who only trust their own experiences

(Ziafat & Shakeri, 2014:70). Association analysis is a part of data mining analysis. The analysis includes looking at the data registered in the supermarket record to identify the product groups purchased together. Association analysis is also known as the market basket analysis. All the specialties of the transactions are being analyzed to search whether certain specialties have tendency to be together or not. Regarding understanding the results, association rules are preferred compared to the clustering analysis because relations that are specific and easy to identify are provided among the transaction specialties. An essential specialty of the technique is that it is not assigned dependently to or independently from a variable (Pande & Abdel-Aty, 2009:146).

3.2. Apriori Algorithm (Association Rule or Market Basket Analysis)

Association rules are developed in the computer science field and are often used in critical applications such as measuring the association between products purchased by a particular consumer or measuring the associations between pages viewed sequentially by a visitor (Guidici & Figin, 2009). The objective of association rules is to uncover hidden patterns in large data sets by detecting relationships or associations between specific values of categorical variables (Nisbet et al., 2018:124). An association rule is a statement between two item sets and can be written as where both and are item sets. The rule indicates that and occur together; if occurs, then also occurs.

The strength and accuracy of the association rules are measured using the following three metrics:

• **Support:** Support of and together is represented as the number of times they appear together divided by the total number of transactions in data set .

$$S(x \to y) = \frac{Number\ of\ times\ x\ y\ appear\ together}{Number\ of\ transactions} = \frac{n_{x \to y}}{n} \quad (1)$$

Support values will always be in the range [0,1]. Support value of zero indicates that rule has never occurred in the transaction dataset. Support value of one indicates that rule is observed in each and every record.

• **Confidence:** Conditional probability of when already occurred plays a significant role when evaluating the strength of the rule. Confidence for is defined as the ratio of the support for and together to the support for (Bhatia, 2019). In other words it is the frequency of occurrence of , conditionally on being true (Guidici & Figin, 2009).

$$C(x \to y) = \frac{support(xy)}{support(x)} = \frac{n_{x \to y}}{n_x} \quad (2)$$

Confidence values will be in the range of [0,1]. A confidence value equal to one indicates that for all of the transactions where is occurred, has also occurred. Similarly, a confidence value equal to zero indicates that none of the transactions includes , also includes .

• **Lift:** Lift is a ratio between the relative probability of both items occurring together and the relative probability of the same event but assuming the two items are independent. A lift value greater than 1 indicates a positive association, and a value less than 1 a negative association.

$$L(a \to b) = \frac{C(x \to y)}{s(y)} \tag{3}$$

A lift ratio greater than 1.0 suggests that the association level between the antecedent and consequent item sets is higher than expected if they were independent (Shmueli et al., 2018:317). In other words, if the result is greater than one, then and are positively correlated (when one is present, the other is likely to be present) (Li et al, 2017). If the lift value is less than one, then the appearance of and are negatively correlated (when one is present, the other is likely to be absent). The larger the lift ratio, the greater the strength of the association. If the result of the formula is equal to one, then and are independent, and there is no correlation between the two (Li et al, 2017).

The task in association rule discovery is to find all rules fulfilling given pre-specified frequency and accuracy criteria (Hand et al., 2001:433). Apriori algorithm developed by Agrawal can be used for extraction rules from large datasets (Agrawal & Srikant, 1994).

It is possible to create an enormous number of rules. Some minimal support (minsup) and minimal confidence (minconf) thresholds can be considered to limit the number of rules. Thus only the rules that satisfy the following inequalities are valid.

$$\begin{aligned} S(x \to y) &\geq minsupp \\ C(x \to y) &\geq minconf \end{aligned} \tag{4}$$

The advantages and disadvantages of apriori algorithm can be summarized as follow (Cavique, 2007:402).

- The outputs of the algorithm are easy to understand
- Many new patterns can be identified
- The sheer number of association rules may make the interpretation of the results difficult.
- Computational time can be a problem for large datasets.

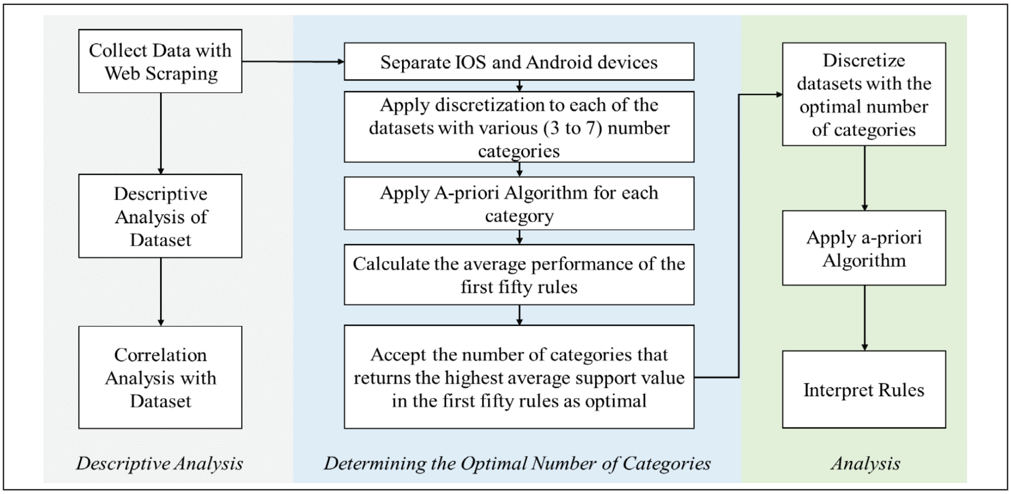
Algorithms sorting the association rules productively are essential in terms of data discovery. Understanding association rules has become easier with the common usage of algorithms. However, there is no method for selecting the appropriate support and confidence thresholds, but when they are selected very low, the user may encounter with thousands of weak rules which do not represent meaningful associations (Raeder & Chawla, 2011:98-99).

In the study, the codes prepared by Loren on the MATLAB platform were used (Loren, 2015).

4. Analysis and Results

The model of the study is presented in Figure 1. The study consists of three stages. In the first stage, descriptive statistics of the data set is examined and correlation analysis is performed. In the second stage, the optimal number of categories are determined separately for each device type. In the last stage, the analysis run with the number of categories determined as optimal and the results are interpreted.

Figure 1: Model of the Study



4.1. Data Collection

On 17 November 2021, data on 200 mobile phones were scraped from the website of one of Turkey's leading electronics stores. Incomplete and incorrect entries were removed from the collected data set. Finally, a dataset of technical features of 197 smartphones was collected. The distribution of phones by brands is given in the Table 1. About 20% of the phones consist of the phones using the iOS system. The remaining 80% use Android system. Most of the phones offered for sale on the website are Android-based. One of the reasons of this unbalanced dataset can be the price of the devices.

Table 1: Brand Frequency Distribution of the Dataset

Brand	Frequency	Percent (%)
Apple	40	20.30
Casper	11	5.58
General Mobile	8	4.06
Huawei	4	2.03
Oppo	15	7.61
Realme	11	5.58
Reeder	4	2.03
Samsung	43	21.83
Tcl	11	5.58
Tecno	9	4.57
Vestel	1	0.51
Vivo	9	4.57
Xiaomi	31	15.74
Total	197	100

The descriptive statistics of the technical features of the phones using the iOS system (minimum value, maximum value, arithmetic mean, median, standard deviation, skewness, kurtosis and for how many phones the feature information is not available) are included in Table 2. 10 technical features have been selected. Technical variables were chosen to create a data set with as few missing values as possible. Some specifications are not available for every device. For example, the CPU performance of each device is not listed on the site. Therefore, it was decided to exclude variables that were not available for each device from the analysis. Using such variables will cause too many missing values to be included in the data set and this will affect the results of the study. The chosen technical features are, pixel density, screen brightness, internal memory, screen size, rear camera resolution, front camera resolution, height, width, depth and price. When the values in the Table 2 are examined, the lowest priced phone using the iOS system is 5699 TRY, and the highest is 18499 TRY. Therefore, the average price was calculated as 11899 TRY.

Table 2: Descriptive Statistics of IOS

Technical Features	Min	Max	Mean	Median	Std	Skewness	Kurtosis	Number of Missing
Pixel Density	326	476	441.65	460	54.50	-1.63	3.80	3
Screen Brightness	625	800	708.13	625	88.50	0.10	1.01	3
Internal Memory	64	512	208.37	128	161.72	1.03	2.63	0
Screen Size	4.70	6.70	5.84	6.10	0.69	-0.22	1.83	0
Rear Cam. Resolution	12	64	14.98	12	11.14	3.53	13.92	0
Front Cam. Resolution	7	44	11.62	12	6.62	4.16	21.25	14
Height	131.50	164.30	145.13	146.70	12.32	0.25	1.52	0
Width	64.20	78.10	70.45	71.50	5.58	0.20	1.49	0
Depth	7.30	8.30	7.45	7.40	0.20	3.03	11.84	0
Price	5699	18499	11899.70	10999	4750.22	-0.11	1.71	0

The descriptive statistics of the technical features of the phones using the Android system are presented in Table 3. The lowest-priced product is 897 TRY and the highest priced product is 15999 TRY. The average price of the phones using the Android system is calculated as 4130 TRY. In the last column of the table, the number of phones do not have the relevant technical feature. It has been determined that more missing values occur in devices with the Android system.

Table 3: Descriptive Statistics of Android

Feature No	Feature Name	Min	Max	Mean	Median	Std	Skewness	Kurtosis	Number of Missing
1	Pixel Density	263	425	368.43	395	57.29	-1.20	2.55	115
2	Screen Brightness	400	1200	599.11	500	205.76	0.82	2.87	112
3	Internal Memory	8	256	114.19	128	60.76	0.77	3.68	7
4	Screen Size	5	6.90	6.53	6.52	0.20	-3.49	25.17	0
5	Rear Cam. Resolution	8	108	36.25	48	25.68	0.72	2.99	0
6	Front Cam. Resolution	5	48	14.13	10	9.68	1.28	3.74	3
7	Height	7.83	172.10	137.82	162.30	57.03	-1.80	4.29	19
8	Width	7.54	165.20	87.04	75.80	30.71	1.88	5.44	18
9	Depth	6.81	164.40	21.75	8.90	33.40	2.76	10.63	18
10	Price	897	15999	4130.66	3099	2853.45	2.42	8.90	0

4.2. Correlation Analysis Results

Correlation analysis was used to examine the degree of correlation of technical features with each other. The Pearson correlation coefficient of each technical feature with the other technical features in the analysis set was calculated. The results were presented in the form of a heatmap in Figure 2. In panel (a) of Figure 2, there are the correlation coefficients of the phones related to the iOS system. The heatmap of the correlation coefficients of the phones using the Android system is shown in panel (b).

The technical feature with the highest correlation with phone price using the iOS system was determined as screen brightness ($r = 0.9125$, $p = 2.47e\text{-}16$). This variable was followed by internal memory ($r=0.6597$, $p = 1.49e\text{-}06$) and screen size ($r=06535$, $p = 2.03e\text{-}06$). It is possible to state that devices with iOS systems with high technical features will be more expensive.

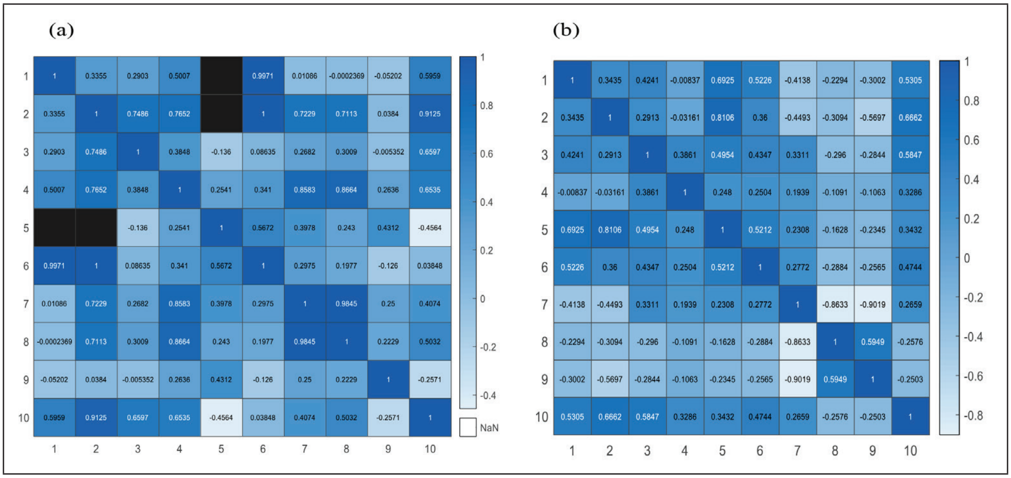
The technical feature with the lowest correlation coefficient with the phone price using the iOS system was determined as the rear camera resolution ($r = -04564$, $p = 0.002$). However, when the distribution of the rear camera resolution is examined, it is understood that this value is 12 megapixels in all other devices, except for three devices. The technical feature with the second-lowest correlation coefficient with the price was determined as depth, but it was determined that this correlation coefficient could not differ significantly from zero ($r = -0.2571$, $p = 0.0959$).

The technical feature with the highest correlation coefficient with price in Android system devices was determined as screen brightness ($r=0.6662$, $p = 5.87e\text{-}07$). This is followed by internal memory ($r = 0.5847$, $r = 3.989e\text{-}15$) and pixel density ($r = 0.5305$, $p = 0.00030$).

The technical features that have a negative correlation coefficient with price in devices with Android system are width ($r = -0.2576$, $p = 0.0022$) and depth ($r = -0.2503$, $p = 0.00296$).

When the devices with iOS and Android systems are compared, it is possible to express that the technical features related to price exhibit similarities.

Figure 2. Heatmap of IOS (panel a) and Android (panel b) Correlation Data



4.3. Determining Optimal Number of Categories

The collected data set is in continuous form. Association rules can only be applied to a categorical data set. However, in the literature, there is no clear procedure for how many categories should be divided in order to make a continuous form data set categorical to apply a-priori algorithm. For this reason, each variable in the data set was divided into categories between 3 and 7 with help of a function (namely discretize) available on the MATLAB platform. The average of the first 50 rules' support, confidence, and lift values was calculated for each of the categories. The number of categories with the highest average support value was considered as the optimal number of category. The results are presented in Table 4.

When the data set for iOS devices is divided into three categories, the average of the support values reaches its highest value (0.7326). The data set for Android devices, on the other hand, is divided into five categories and the average of the support values reaches the highest value (0.5460). That is why, the iOS data set was divided into 3 categories and the Android data set into 5 categories, and the study continued with this categorical data set.

Table 4: Descriptive Statistics of First 50 Rules for Different Number of Categories

	IOS Dataset			**Android Dataset**		
Number of Categories	**Support**	**Confidence**	**Lift**	**Support**	**Confidence**	**Lift**
3	***0.7326***	***0.3790***	***1.5376***	0.5423	0.1660	1.5026
4	0.6941	0.3310	1.5673	0.4984	0.1375	1.6202
5	0.6604	0.2720	1.7545	***0.5460***	***0.1765***	***1.5419***
6	0.6177	0.1969	1.4959	0.5247	0.1386	1.5491
7	0.6450	0.2810	1.5389	0.5291	0.1715	1.6761

4.4. Applying Apriori Algorithm

By using the optimal number of categories determined in the previous step, the continuous form data set was converted to categorical form. The lower and upper edges of the categories are presented in Table 5. They are obtained with the discretize function available on MATLAB language. According to this table, concerning the iOS devices, the devices with a price range between 4000 TRY and 9000 TRY are evaluated in the first category, the devices with a price range between 9000 TRY and 14000 TRY are evaluated in the second category, and the devices with a price range between 14000 TRY and 19000 TRY are evaluated in the third category. A similar interpretation may be done for Android devices as well. The devices with a price lower than 3200 TRY are evaluated in the first category, the devices with a price range between 3200 TRY and 6400 TRY are evaluated in the second category, the devices with a price range between 6400 TRY and 9600 TRY are evaluated in the third category, the devices with a price range between 9600 TRY and 12800 TRY are evaluated in the fourth category. The devices with a price range between 12800 TRY and 16000 TRY are evaluated in the fifth category.

Table 5: Value Ranges of Features

	iOS			**Android**				
Feature Name	**1**	**2**	**3**	**1**	**2**	**3**	**4**	**5**
Pixel Density	300≤ x <360	360≤ x <420	420≤ x ≤480	240≤ x <277	277≤ x <314	314≤ x <351	351≤ x <388	388≤ x ≤425
Screen Brightness	600≤ x <670	670≤ x <740	740≤ x ≤810	400≤ x <560	560≤ x <720	720≤ x <880	880≤ x <1040	1040≤ x ≤1200
Internal Memory	0≤ x <180	180≤ x <360	360≤ x ≤540	0≤ x <60	60≤ x <120	120≤ x <180	180≤ x <240	240≤ x ≤300
Screen Size	4.2≤ x <5.1	5.1≤ x <6.0	6.0≤ x ≤6.9	4.8≤ x <5.3	5.3≤ x <5.8	5.8≤ x <6.3	6.3≤ x <6.8	6.8≤ x ≤7.3
Rear Cam. Resolution	11.50≤ x <11.83	11.83≤ x <12.17	12.17≤ x ≤12.50	0≤ x <22	22≤ x <44	44≤ x <66	66≤ x <88	88≤ x ≤110
Front Cam. Resolution	7≤ x <8.7	8.7≤ x <10.4	10.4≤ x ≤12.1	0≤ x <10	10≤ x <20	20≤ x <30	30≤ x <40	40≤ x ≤50
Height	126≤ x <138	138≤ x <150	150≤ x ≤162	0≤ x <35	35≤ x <70	70≤ x <105	105≤ x <140	140≤ x ≤175
Width	64≤ x <69	69≤ x <74	74≤ x ≤79	0≤ x <34	34≤ x <68	68≤ x <102	102≤ x <136	136≤ x ≤170
Depth	7.2≤ x <7.6	7.6≤ x <8.0	8.0≤ x ≤8.4	0≤ x <33	33≤ x <66	66≤ x <99	99≤ x <132	132≤ x ≤165
Price	4000≤ x <9000	9000≤ x <14000	14000≤ x ≤19000	0≤ x <3200	3200≤ x <6400	6400≤ x <9600	9600≤ x <12800	12800≤ x ≤16000

Information on how many values exist in each category is demonstrated in Table 6. According to Table 6, for iOS devices, the device number with pixel density being ($300 \le x < 360$) in the first category is 7. There is no device with pixel density being ($360 \le x < 420$) in the second category. Lastly, the device number with pixel density being ($420 \le x \le 480$) in the third category is 33.

It is possible to make similar interpretations for Android devices as well. For instance, the device number with internal memory value being ($0 \le x < 60$) in the first category is 25. The device number with internal memory value being ($60 \le x < 120$) in the second category is 7. The device number with internal memory value being ($120 \le x \le 180$) in the third category is 10. The device number with internal memory value being ($180 \le x < 240$) in the fourth category is 2 and lastly, the device number with internal memory value being ($240 \le x \le 300$) in the fifth category is 1.

Table 6: Number of Devices in Categories

		iOS			Android				
Feature Names	**Abbreviation**	**1**	**2**	**3**	**1**	**2**	**3**	**4**	**5**
Pixel Density	Pd	7	0	33	1	2	3	0	0
Screen Brightness	Sb	21	0	19	9	1	0	0	32
Internal Memory	Im	23	9	8	25	7	10	2	1
Screen Size	Ss	6	14	20	24	26	88	0	15
Rear Camera Resolution	Rcr	0	40	0	1	1	9	146	3
Front Camera Resolution	Fcr	7	0	19	76	0	78	0	6
Height	H	14	15	11	77	43	12	20	5
Width	W	20	9	11	22	0	1	0	118
Depth	D	37	2	1	1	0	122	0	19
Price	P	7	16	17	120	0	18	0	4

In Table 7, the rules generated with the data available in the iOS set is presented. There were no rules with two or more successors or predecessors. As previously stated, the variables were divided into three categories for the iOS set. It is required to identify the threshold value for support and confidence to begin the analysis. In the study, the lowest support and confidence value was identified as 0.05. Thus, the rules with support and confidence value over 5% were researched. For 82.5% of the iOS phones which were examined according to the rule in the first line, pixel density is between 420 and 480, and the rear camera resolution is between 11.50 and 11.83 MP. The fact that this rule's confidence value turned out to be 100 signifies that rear camera resolution is between 11.50 and 11.83 MP for all the phones with pixel density between 420 and 480. The fact that lift value turned out to be 1 indicates no relation among the features. Lastly, the rule count value turned out to be 33 shows that two features appear together in 33 phones in the set.

Table 7: Extracted Rules of IOS Set

	Precedent *(x)*	Antecedent *(y)*	Support	Confidence	Lift	Rule Count in Dataset
1	Pixel Density 420-480	Rear Camera Resolution 11.50-11.83	82.5	100	1	33
2	Pixel Density 420 - 480	Height 150 – 162	82.5	100	1	33
3	Pixel Density 420 - 480	Width 138-150	82.5	100	1	33
4	Pixel Density 420 - 480	Depth 7.20 – 7.60	82.5	100	1	33
5	Screen Brightness 300-360	Internal Memory 0-180	50	95.2381	1.6563	20
6	Pixel Density 420 - 480	Screen Brightness 670-740	47.5	57.5758	1.2121	19
7	Pixel Density 420 - 480	Screen Size 6.00-6.90	47.5	57.5758	1.1515	19
8	Pixel Density 420 - 480	Front Camera Resolution 11.50-11.83	47.5	57.5758	0.88578	19
9	Pixel Density 420 - 480	Price 1 4000 – 19000	42.5	51.5152	1.2121	17
10	Pixel Density 420 - 480	Internal Memory 0-180	40	48.4848	0.84321	16
11	Pixel Density 420 - 480	Price 9000-14000	40	48.4848	1.2121	16
12	Pixel Density 420 - 480	Screen Brightness 600-670	35	42.4242	0.80808	14
13	Pixel Density 420 - 480	Screen Size 5.10 – 6.00	35	42.4242	1.2121	14
14	Pixel Density 420 - 480	Front Camera Resolution Nan	35	42.4242	1.2121	14
15	Pixel Density 420 - 480	Internal Memory 180-360	22.5	27.2727	1.2121	9

In Table 8, some of the rules generated by means of using Android phones are listed. There were no rules with two or more successors or predecessors. Each variable was divided into five categories. Support and confidence threshold values were determined as 0.05. When the first rule is examined, in the 20% of the examined phones, pixel intensity is between 351 and 388 and screen size is between 6.80 and 7.3. The confidence value is turned out to be 100 indicates that screen size is between 6.80 and 7.30 for all the phones with pixel density between 351 and 388. The lift value for this rule turned out to be higher than one shows that screen size will most likely be between 6.80 and 7.30 for the phones with pixel density between 351 and

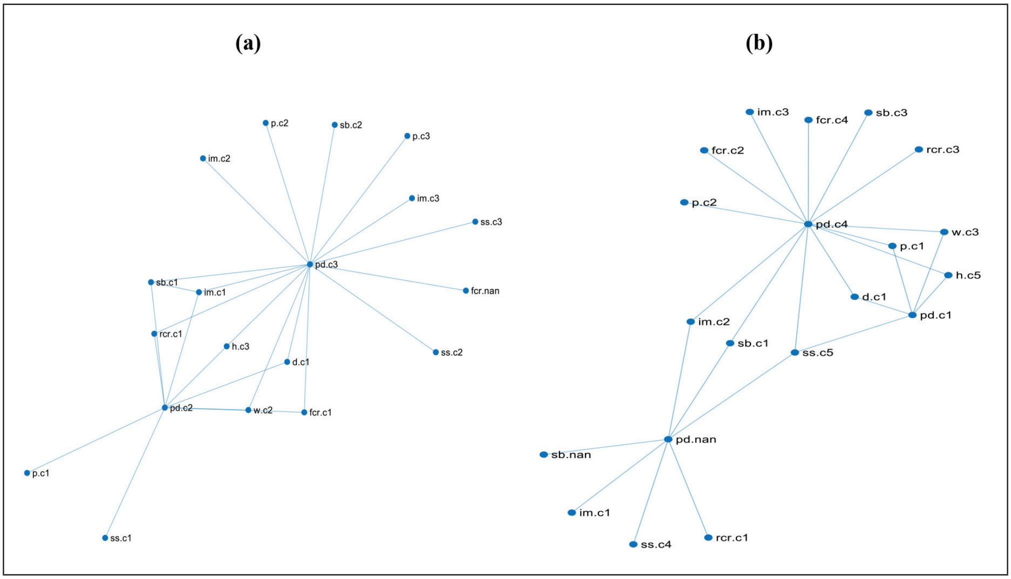
388. For 32 of the examined phones, pixel density is between 351, and 388 and screen size is between 6.80 and 7.30 inches.

Table 8: Extracted Rules of Android Set

	Precedent *(x)*	Antecedent *(y)*	Support	Confidence	Lift	Rule Count in Database
1	Pixel Density 351-388	Screen Size 6.80-7.30	20	100	1.0738	32
2	Pixel Density 351-388	Width 68-102	20	100	1.3115	32
3	Pixel Density 351-388	Depth 0-33	20	100	1.3333	32
4	Pixel Density 351-388	Height 140-175	19.375	96.875	1.3136	31
5	Pixel Density 351-388	Rear Camera Resolution 44-66	18.75	93.75	1.9231	30
6	Pixel Density 351-388	Internal Memory 60-120	12.5	62.5	1.1364	20
7	Pixel Density 351-388	Price 3200-6400	12.5	62.5	2.5641	20
8	Pixel Density 351-388	Screen Brightness 400-560	8.125	40.625	2.6	13
9	Pixel Density 351-388	Front Camera Resolution 22-44	7.5	37.5	1.3953	12
10	Pixel Density 351-388	Screen Brightness 720-880	6.25	31.25	5	10
11	Pixel Density 240-277	Screen size 6.80-7.30	5.625	100	1.0738	9
12	Pixel Density 240-277	Height 140-175	5.625	100	1.3559	9
13	Pixel Density 240-277	Width 68-102	5.625	100	1.3115	9
14	Pixel Density 240-277	Depth 0-33	5.625	100	1.3333	9
15	Pixel Density 240-277	Price 0-3200	5.625	100	1.5238	9

Network graphics of the rules are illustrated in Figure 3. Categories in the center of the figures are the categories that appear in the maximum number of rules. In other words, the most popular categories are in the center. For example, the most popular rules for IOS phones are pixel density category 3 and pixel density category 2. As for the Android phones, the most popular categories are pixel density category 4 and the category (NaN) in which value is not stated.

Figure 3. Visualizing IOS (a) and Android (b) Rule Relations with network graph



4.5. Comparisons With The Previous Studies

In their study, Dewerter et al. (2004) collected a dataset regarding the phones offered for sale in Germany and examined the variables influencing the price with the help of regression analysis. They expressed in their study that the phone volume has a negative impact on price, but this impact is not meaningful in terms of statistics. Furthermore, they reported that the age of the phone has a positive impact on price. Nazari et al. (2011) examined the features of the phones offered for sale in Iran, with the help of the hedonic pricing method. They expressed in their study that touchscreen, GPS, and camera resolution positively impact price. In their study, Ahmad et al. (2019) examined 324 phones offered for sale in Pakistan and as a result, they reported that the variable having the highest correlation with the price is the front camera resolution.

In consequence of the comparison made with the previous studies, it is possible to state that the technical features having the highest correlation with price show variability in different countries and in different time frames. Changes in customers' preferences lead to offering for sale at a higher price for phones with different features.

5. Conclusion

Data mining has recently become the focus of attention for researchers and enterprises. In this study, technical features of the smartphones available in the mobile phone market in Turkey were examined with the association rules, one of the data mining techniques. In the study, technical features of the smartphones with iOS and Android operating systems were analyzed with the association rules. According to the findings obtained from the analysis, it was found out that the highest correlation in the phones with iOS operating system is between price and screen brightness. The results reveal the descriptive characteristics of the market.

Moreover, it was concluded that the lowest correlation in the phones with iOS operating system is the rear camera resolution. Apart from the iOS phones, while there is a high correlation between price and screen brightness of the phones with Android operating system, it was found that there is a negative correlation between price and width and depth. The trial and error method (categories between 3 and 7) was employed on how many categories the data will be divided into. The dataset was discretized into the category number revealing the highest average support values. The optimal number of categories was determined as 3 for the IOS set and 5 for the Android set. For this reason, the data sets were divided into categories in optimal numbers and the apriori algorithm was applied in this way.

The most prominent rules can be summarized as follow. For IOS dataset, on smart phones where the pixel density is between 420 ppi and 480 ppi,

- rear camera resolution is between 11.50 megapixel tnd 11.83 megapixel
- the height of the phone is between 150 mm and 162 mm
- width is between 138 mm and 150 mm
- depth is between 7.2 mm and 7.6 mm.

For Android dataset, on smart phones where the pixel density is between 351 ppi and 388 ppi,

- screen size is between 6.8 inches and 7.3 inches
- width is between 68 mm and 102 mm
- depth is between 0 mm and 33 mm.

Based on the findings of the study, it has been considered that the association rules revealed in the study can help the customers to perform the telephone selection process rationally. Customers have the opportunity to evaluate the general features of the devices available in the market with a general overview. As for the technology stores may use the rules revealed with this study when they recommend phones to their customers. Thus, easiness can be ensured in the selection of the phones that will meet the needs of the customers. The limitation of the study is that there are relatively few phones in the data set used in the study. Another limitation of the study is that each variable is divided into the same number of categories. In further studies, when the variables are categorized in different numbers (for example, price is divided into three categories and pixel density can be divided into four categories), it can be investigated whether the quality of the rules created increases.

In the future studies, longitudinal studies may be performed. Data covering a long time period may be collected. In such a case, it may be examined whether different features in different time intervals are related to price or not. For instance; while the variable having the highest technical feature related to price can be sometimes identified as screen size; the variables such as screen brightness, battery power etc. may exhibit increase or decrease with the price, in different times, depending upon the change in the preference of the customers. In the progressive studies, segmentation process may be applied for phones available in the market. Technical features of the phones available in the smartphone market, which have become almost necessary for consumers, may be subjected to a more detailed examination with the help of different data mining techniques. Thus, a more efficient operation of the market may be ensured.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

The authors contributed to this study equally.

References

Agrawal, R. & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. Proceedings of the 20th International Conference on Very Large Data Bases (pp. 487- 499).

Ahmad, W., Ahmed, T. & Ahmad, B. (2019). Pricing of mobile phone attributes at the retail level in a developing country: Hedonic analysis. Telecommunications Policy, 43(4), 299-309.

Bhatia, P. (2019). Data mining and data warehousing. Principles and Practical Techniques. Cambridge Univesity Press.

Cavique, L. (2007). A scalable algorithm for the market basket analysis. Journal of Retailing and Consumer Services, 14, 400-407.

Chen, Y. L., Tang, K., Shen, R. J. & Hu, Y. H. (2005). Market basket analysis in a multiple store environment. Decision support systems, 40(2), 339-354.

Dewenter, R., Haucap, J., Luther, R. & Rötzel, P. (2007). Hedonic prices in the German market for mobile phones. Telecommunications Policy, 31(1), 4-13.

Fagerlind, H., Harvey, L., Humburg, P., Davidsson, J. & Brown, J. (2022). Identifying individual-based injury patterns in multi-trauma road users by using an association rule mining method. Accident Analysis & Prevention, 164, 106479.

Giudici, P. & Figin, S. (2009). Applied data mining for business and industry. West Sussex: John Wiley & Sons Ltd.

Gupta, S. & Mamtora, R. (2014). A survey on association rule mining in market basket analysis. International Journal of Information and Computation Technology, 4(4), 409-414.

Ha, S. H. & Park, S. C. (1998). Application of data mining tools to hotel data mart on the Intranet for database marketing. Expert Systems with Applications, 15(1), 1-31.

Hand, D., Mannila, H. & Smyth, P. (2001). Principles of Data Mining. Cambridge: The MIT Press.

Hausman, J. (1999). Cellular telephone, new products, and the CPI. Journal of Business & Economic Statistics, 17(2), 188-194.

Hsieh, N. C. (2004). An integrated data mining and behavioral scoring model for analyzing bank customers. Expert systems with applications, 27(4), 623-633.

Kamakura, W. A. (2012). Sequential market basket analysis. Marketing Letters, 23(3), 505-516.

Kaur, M. & Kang, S. (2016). Market basket analysis: Identify the changing trends of market data using association rule mining. Procedia computer science, 85, 78-85.

Kim, M. K., Wong, S. F., Chang, Y. & Park, J. H. (2016). Determinants of customer loyalty in the Korean smartphone market: Moderating effects of usage characteristics. Telematics and Informatics, 33(4), 936-949.

Leote, P., Cajaiba, R. L., Cabral, J. A., Brescovit, A. D. & Santos, M. (2020). Are data-mining techniques useful for selecting ecological indicators in biodiverse regions? Bridges between market basket analysis and indicator value analysis from a case study in the neotropics. Ecological Indicators, 109, 105833.

Li, Q., Zhang, Y., Kang, H., Xin, Y. & Shi, C. (2017). Mining association rules between stroke risk factors based on the apriori algorithm. Technology and Health Care, 25, 197-205.

Loren, S. (2015). Introduction to Market Basket Analysis. Erişim Tarihi: 15.12.2022, https://blogs.mathworks.com/loren/2015/01/29/introduction-to-market-basket-analysis/.

Montenegro, J. A. & Torres, J. L. (2016). Consumer preferences and implicit prices of smartphone characteristics. Málaga Economic Theory Research Center Working Papers.

Monteserin, A. & Armentano, M. G. (2018). Influence-based approach to market basket analysis. Information Systems, 78, 214-224.

Mostafavi, S. M., Roohbakhsh, S. S. & Behname, M. (2013). Hedonic price function estimation for mobile phone in Iran. International Journal of Economics and Financial Issues, 3(1), 202-205.

Nisbet, R, Miner, G. & Yale, K. (2018). Handbook of statistical analysis and data mining applications. Elsevier Academic Press.

Pande, A. & Abdel-Aty, M. (2009). Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. Safety science, 47(1), 145-154.

Pradhan, S., Priya, P. & Patel, G. (2022). Product bundling for 'efficient'vs 'non-efficient'customers: Market basket analysis employing genetic algorithm. The International Review of Retail, Distribution and Consumer Research, 32(3), 293-310.

Raeder, T. & Chawla, N. V. (2011). Market basket analysis with networks. Social Network Analysis and Mining, 1(2), 97-113.

Raorane, A. A., Kulkarni, R. V. & Jitkar, B. D. (2012). Association rule–extracting knowledge using market basket analysis. Research Journal of Recent Sciences.

Samboteng, L., Rulinawaty, R., Kasmad, M. R., Basit, M. & Rahim, R. (2022). Market basket analysis of administrative patterns data of consumer purchases using data mining technology. Journal of Applied Engineering Science, 20(2), 339-345.

Santarcangelo, V., Farinella, G. M., Furnari, A. & Battiato, S. (2018). Market basket analysis from egocentric videos. Pattern Recognition Letters, 112, 83-90.

Setiabudi, D. H., Budhi, G. S., Purnama, I. W. J. & Noertjahyana, A. (2011, August). Data mining market basket analysis' using hybrid-dimension association rules, case study in Minimarket X. In 2011 International Conference on Uncertainty Reasoning and Knowledge Engineering (pp. 196-199).

Shmueli, G., Bruce, P.C., Yahav, I., Patel, N. R. & Lichtendahl, K.C. (2018). Data mining for business analytics: concepts, techniques, and applications in R. wiley. New Jersey: John Wiley & Sons.

Sudirman, I. D., Nugraha, D. Y., Utama, I. D. & Tjoe, T. F. (2021). Comparing the relationship of tourist attraction by using association rule. IEOM Society International.

Ünvan, Y. A. (2021). Market basket analysis with association rules. Communications in Statistics-Theory and Methods, 50(7), 1615-1628.

Wang, L., Lin, B., Chen, R. & Lu, K. H. (2022). Using data mining methods to develop manufacturing production rule in IoT environment. The Journal of Supercomputing, 1-24.

Wei, J. T., Lee, M. C., Chen, H. K. & Wu, H. H. (2013). Customer relationship management in the hairdressing industry: An application of data mining techniques. Expert Systems with Applications, 40(18), 7513-7518.

Ziafat, H. & Shakeri, M. (2014). Using data mining techniques in customer segmentation. Journal of Engineering Research and Applications, 4(9), 70-79.