



# Düzce Üniversitesi Bilim ve Teknoloji Dergisi

*Araştırma Makalesi*

## Extracting Association Rules of Turkish Retail Company from Online Transactions: Case Study

 Elif Şafak SİVRİ<sup>a</sup>,  Mustafa Cem KASAPBAŞI<sup>a\*</sup>

<sup>a</sup> *Computer Engineering, Faculty of Engineering, Istanbul Commerce University, İstanbul TURKEY*

*\* Corresponding author's e-mail address: mckasapbasi@ticaret.edu.tr*

DOI : 10.29130/dubited.510529

### ABSTRACT

The extracting association rules of inter-user-product relations used by companies in decision-making processes have been popular for some time, especially for market basket analysis. In this study it is aimed to discover association rules from original online store transaction of a Turkish retail company, in order to help administrator and decision maker also Customer Relationship Management department to initiate campaigns. The main objective is to find out which product item sets are bought together. In order to better compare the results the data are analyzed with and without clustering according to range of ages and gender. Data mining Association analysis methods such as Apriori Algorithm, FP-Growth (Frequent Pattern) then applied which are used to extract association rules. Moreover some of the collaborative filtering metrics namely Jaccard, Pearson, and Cosine function are used to understand the association between products to obtain a recommendation system. The proposed recommendation methods successfully recommended the associated product for the obtained original dataset as high as %65 accuracy. Obtained association rules are shared with the marketing department to initiate and direct forthcoming marketing campaigns.

**Keywords:** *Data mining; Associative analysis; apriori algorithm, FP-Growth; e-commerce*

## Türk Perakende Şirketindeki Çevrimiçi Alış Verişler için İlişkililik Kurallarını Çıkarılması: Durum Çalışması

### ÖZET

Şirketlerin karar verme süreçlerinde kullandıkları ürünler-kullanıcılar arası ilişkililik kuralları özellikle market sepet analizleri için bir süreden beri popülerliğini korumaktadır. Bu çalışmada Türk perakende şirketinin online alışveriş sitesine ait orijinal veri hareketleri incelenerek ürünler arasında ilişkililik kuralları çıkarılması hedeflenmiştir. Bu şekilde yönetici ve karar vericilere aynı zamanda Müşteri İlişkileri Yönetimi biriminin yeni kampanyalarına yardım etmesi hedeflenmiştir. Ana hedef hangi ürün kümelerinin beraber alındığının keşfedilmesidir. Veriler sonuçların daha iyi kıyaslanabilmesi için, kümelenmeden ayrıca yaş aralığı ve cinsiyete

göre kümelenmiş olarak analiz edilmiştir. Apriori ve FP-Growth gibi veri madenciliği analiz metotları kullanılarak ilişkililik kuralları çıkartılmıştır. Ayrıca bazı işbirlikli filitreleme ölçütleri olan Jaccard, Pearson ve Cosine fonksiyonları ile ilişkili ürünler için bir tavsiye sistemi geliştirilmek için kullanılmıştır. Önerilen tavsiye sistemi veri setindeki ilişkili ürünleri başarı %65 gibi yüksek bir oran ile tavsiye etmiştir. Elde edilen ilişkililik kuralları pazarlama birimi ile paylaşılıp gelecekteki kampanyalarda kullanılması sağlanmıştır.

*Anahtar Kelimeler: Veri madenciliği; ilişkililik analizi; apriori algoritması, FP-Growth; e-Ticaret*

## I. INTRODUCTION

Nowadays almost every company using databases to store not only sale transaction but also deep knowledge about products relations waiting to be extracted. Extracting meaningful knowledge from such information is valuable since they not only give insights about customers' buying habits but also allow decision makers take more accurate decisions. It is desired to extract meaningful knowledge from such a day by day accumulating gigantic data [1], however finding similar pattern and meaning is not possible for human by himself alone [2, 3]. Data mining techniques are used to fill this gap and are developed to find associations, classifications, clusters or estimations from such gigantic data. Data mining techniques can be categorized in three main groups namely clustering, classification and association.

The main focus of this study is to discover association rules from transactions of online database of Turkish retail company using associative analysis, which are developed for discovering frequent patterns. Following sections are organized like in the next section a brief literature review is given, the third section is dedicated for explanation of data set, materials and applied methods in the study, in fourth section results and discussion is presented.

## II. LITERATURE REVIEW

Common classical application field of association rule analysis is called market basket analysis. This analysis aims to discover of customer buying habits by emerging the correlations between what had been bought[1]. Result of the analysis is evaluated by the metrics namely Support, Confidence and Lift. For example; Assume that there is an association rule between chips-snacks. Association rule is "Customers who bought chips also bought snack with a %15 probability. %10 of all transaction contains chips and snack". %15 value corresponds to Support value, %10 percent corresponds to confidence value [3-5]. In the example The Support value can be interpreted as in %15 of all transactions chips and snack were bought together. Also for the confidence value it can be said that %10 percent of all the chip buyers also bought snack. Association analysis let us give an answer to question "What is most frequently bought with itemset A?".

Other than Apriori another most frequently used algorithm is FP-Growth which is used to find out frequent patterns of object sets in association analysis. This Algorithm holds the statistics in a data structure called FP tree which reduces the cost of operation. Despite Apriori oriented algorithms it

only traverse the transaction twice. At first, support values of objects are calculated; secondly tree structure is formed [6].

Association algorithms are not only used for Customers sales transections. In [7] it is used for locational temporal data mining applications on Van Lake water level association among rainfall level evaporation on the surface. The goal of the study is to determine the parts of the flooding lake shores and recommending required precautions.

Text Mining of frequently used following Turkish letters study is done in [8] with comparison of Apriori, AIS (Agrawal, Imielinski, and Swami) and SETM (Set Oriented Mining). Apriori performed better then the other algorithms in that study.

An other study is carried out for performance measure of different traverse strategies in terms of the I/O, memory, and computational time requirements and proposed a Tree projection algorithm for more efficient [9].

One of the highly cited study about Market basket analysis is presented in [10] which basically propose an optimized version of classical analysis.

CRM (Customer Relationship Management) is an other application area of data mining. Mainly it is used to cluster, classify or understand the behavior and buying habits of customers. Google scholar search gives many academic studies related to CRM and data mining two of book [11, 12] get the highest number of citation.

Dixit and his friend [13] proposed a personalized recommender agent for e-commerce using user-item matrix generated utilizing clickstream behavior of the users. To asses the similarity of the user collaborative filtering techniques is employed.

In [14] The proposed study analyzes content-based recommendations for the E-commerce site, where association rule mining and Apriori algorithm are used for product prediction and recommendation.

Current and future researches in recommendation systems encompass social network association and personalization which consist of implicit local and personal information gathered from internet of things. However for companies which does not have such social network framework or equipment infrastructure to gather implicit personal information for their customers still relay on conventional transactions data for constructing recommendation systems.

### III. MATERIALS AND METHODS

#### *A. ASSOCIATION RULES*

Association rules are developed to discover meaningful correlation between what customers had bought so far and what they may buy in the future. Therefore first of all it is used to extract association is like customer who bought X itemset also bought Y itemset, secondly if a customer only bought X item set so far it becomes a recommendation system. This is of course very useful tool for CRM, marketing department and administrative purposes.

For association analysis many algorithm have been proposed so far. Some common association algorithm used for association analysis are Apriori, AprioriTID, DIC (Dynamic Itemset Counting), Apriori Hybrid, SETM Partition, FP-Growth and Eclat [4]. In the survey study these algorithm are systematized according to strategy to traverse the search space and strategy to determine support values of itemsets. AprioriTID and DIC are derivation of the Apriori algorithms. Apriori algorithm is the one of the most frequently used association analysis algorithm.

These techniques have three important metrics support, confidence and lift:

**Support:** One of the important property of itemset is support number given in Eq. 1 which represents number of transaction for a specific itemset.  $T$  represents all transactions, while  $t_i$  itemset  $X$  is the item.

$$\sigma(X) = |\{t_i \mid X \subseteq t_i, t_i \in T\}| \quad (1)$$

Support for a rule is denoted as  $\text{Sup}(X \rightarrow Y)$ , and calculated as given in Eq.2 [15, 16]

$$\text{Support}(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \quad (2)$$

**Confidence:** Confidence of a rule is  $X \rightarrow Y$ , and formulation is given in Eq. 3 [15, 16]

$$\text{Confidence}(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (3)$$

**Lift :** Lift is used to asses the performance of the model and it is calculated as given in Eq. 4 [16]

$$\text{Lift}(X \rightarrow Y) = \frac{\text{Confidence}(X \rightarrow Y)}{\text{Support}(Y)} \quad (4)$$

## B. COLLABORATIVE FILTERING

At present, a recommendation system includes, the rules of association, Collaborative Filtering content based recommendation, and combinations of those approaches. The Collaborative Filtering algorithm, which is a traditional algorithm of recommended systems, is the best known and accepted algorithm due to its many advantages [17].

Similarity calculation between items or users is fundamental step in collaborative filtering (CF). For item based collaborative filtering main philosophy of calculating similarity between item  $i$  and  $j$  is working for those users that has associated with these records then calculate  $w_{i,j}$  similarity. For user based collaborative filtering algorithm,  $w_{u,v}$  the first similarity of users having same itemsets is calculated for  $u$  and  $u$  users [18]. There are numerous methods for calculating similarities for both between users and items. In this study Cosine, Jaccard, and Pearson similarities are used and they are described in the following subsection.

### Cosine Similarity

Cosine similarity is one of the mostly used collaborative filtering calculation for assessing similarities that of two documents. The similarity between documents, the calculation of angle between the word frequency vectors of those documents. Result will be between 0 and 1 representing no similarity and maximum similarity (identical) respectively. Calculation equation of similarities of vector A and vector B is given in Eq. 5 [19, 16].

$$w_{A,B} = \cos(A, B) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (5)$$

“.” represents dot multiplication of two matrices.

### Jaccard Similarity

It is also known as the similarity coefficient since it is a similarity measures between objects that contain only binary attributes and generally have values between 0 and 1. A value 1 means that the two objects are totally similar, while a value of 0 indicates that the objects are in no way similar [20, 16]. Jaccard similarity equation is given in Eq. 6.

$$J(A, B) = \frac{\# \text{ of matching items in } A \text{ and } B}{\# \text{ of the rest except non matching}} = \frac{f_{11}}{f_{01} + f_{10} + f_{11}} \quad (6)$$

$f_{11}$  number of all matching items in A and B

$f_{10}$  number of items that are in A but not in B

$f_{01}$  number of item that are in B but not in A

### Pearson Similarity

It is the measure of linear correlation between X and Y variable having the range of values between +1 and -1. +1 means there is linear correlation, 0 means no correlation at all and -1 negative correlation. Calculation of the value is given in Eq. 7 [16].

$$Corr(X, Y) = \frac{cov(X, Y)}{\sigma_x \sigma_y} \quad (7)$$

cov represents the covariance of X and Y, also  $\sigma$  is the standard deviation of denoted variable.

## C. INTRODUCTION TO DATA SET AND PREPROCESSING

Among 11555 data age of only 304 of them were unknown. Ages are grouped in five; 21-30, 31-40, 41-50, 51-60 and 60 above. When missing data are omitted it has been observed that %20 of the customers fall in age group of 21-30; %38 of customer fall in age group of 31-40; %28 of customer fall in age group of 41-50, %9 of customers fall in age group of 51-60 and lastly %4 of customer were above age 60. Marital state of 9208 records was missing after they are ignored, %56 of the records were for married %44 of the records were belong to singles. When these records investigated according to their segment there were 6910 missing data.

## D. METHODOLOGY

The analyses are carried out in two phases at first data handled without clustering. Apriori, FP-Growth are used to extract association rules then Jaccard, Cosine and Pearson are used to compare and find recommendation for the results. In the second phase data is clustered using K-Means algorithm with regarding age and gender. Weka software is utilized for Apriori and FP-Growth while a custom program implemented in C# is used for collaborative filtering [21]. In order to evaluate the methodology transaction data for November, December and January month are used to find recommendation and data of February month is used as test data applied for 100 customers in

February. There are more than 11000 transaction data for a single season. General steps of how association rules are extracted is depicted in Fig.1.

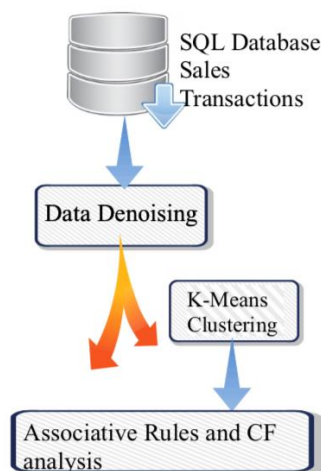


Figure 1. Steps for extracting Associative rules

### III. RESULTS & DISCUSSIONS

In [22] an open source program called WEKA is analyzed which has been gaining appreciation and acceptance in both academia and business environment. WEKA is a software tool that is used for fundamental data mining applications such as clustering, association and classification. In this study WEKA is used to analyze online transactions of Turkish Retail Company to extract association rules between sold itemsets.

In data cleaning step customers' informations are reorganized so that they reflect what has been bought by customers. Since these data were to analyze in WEKA software a special effort had to given to transform data into suitable form. Most purchased products was determined. In the cleaning stage also products which are bought less than 10 times are erased from dataset. Using WEKA Apriori and FP Growth association algorithms are used to find the frequently together bought itemsets.

#### A. EVALUATION OF THE ASSOCIATION RULES

After applying Apriori algorithm first 20 most meaningful results according to LIFT value are given in Table 1. LIFT values show the interestingness of the association rules. If the more bigger then one more stronger and more interesting the rule is. First 15 association rules are listed in Table 1

Table 1. Apriori Algorithm Results

First Shopping (X)	Second Shopping (Y)	Confidence	Lift
shirt, tshirt, pant (108)	knitwear, jacket(23)	0.21	11.41
Shirt, tshirt, knitwear(76)	thsirt, jacket (23)	0.3	10.25
tshirt, pant, knitwear (73)	shirt, jacket (23)	0.32	10.08
pant, knitwear (121)	shirt, thsirt jacket(23),	0.19	9.95

shirt, tshirt, knitwear (91)	pant, jacket (23)	0.25	9.87
tshirt, knitwear(143)	shirt, pant, jacket(23)	0,16	9,26
thsirt, pant,knitwear (73)	shirt, sweatshirt(25)	0.34	8.76
shirt, knitwear(155)	tshirt, pant, jacket(23)	0.15	8.76
shirt, tshirt,sneakers (81)	pant, knitwear(37)	0.46	8.69
tshirt, knitwear(143)	shirt, pant,sweatshirt (25)	0.17	8.57
pant, knitwear (121)	shirt, tshirt,sweatshirt (25)	0.21	8.35
tshirt,shoes, pant(75)	sneakers, knitwear(26)	0.35	8.32
tshirt, pant (194)	shirt, sneakers,sweatshirt (23)	0.12	8.27
Tshirt, pant (194)	Shirt, knitwear,jacket (23)	0.34	8.27
Shirt,tshirt, pant(108)	Sneakers,knitwear (37)	0.34	8.22

As an example we can interpret Table 1 first row as itemset X{shirt, tshirt, pant} 108 times has been bought together and itemset Y{knitwear, jacket} 23 times bought together. If one customer bought X itemset, s/he also bought Y itemset. Moreover Confidence of %21 percent means that all member of itemset X and itemset Y were in the %21 percent of all sale transection. Rest of the rows can be interpreted likewise to construct the associative rules. After investigating results with Apriori algorithm, an other experiment is also carried out with FP Growth algorithm on same the dataset, results are tabulated in Table 2.

**Table 2.** FP-Growth Algorithm Results

First Shopping (X)	Second Shop (Y)	Confidence	Lift
tshirt, pant, knitwear(73)	shirt, sweatshirt (25)	0.34	8.76
shirt, tshirt, sneakes (81)	pant, knitwear (37)	0.46	8.69
tshirt, knitwear (143)	shirt, pant, sweatshirt(25)	0.17	8.57
pant, knitwear (121)	shirt, tshirt, sweatshirt(25)	0.21	8.35
shoes, tshirt, pant (75)	sneakers, knitwear (26)	0.35	8.32
shirt, tshirt, pant (108)	sneakers, knitwear (37)	0.37	8.22
tshirt, sneakers, pant (78)	shoes, knitwear (26)	0.33	8.17
shoes, tshirt, sneakers (61)	pant, knitwear (26)	0.43	8.11
pant, sweatshirt (80)	sneakers, knitwear (27)	0.34	8.1
sneakers, knitwear (96)	shirt, jacket (24)	0.25	8
pant, sweatshirt (80)	shirt, tshirt, knitwear (25)	0.31	7.91
shirt, tshirt, sneakers (81)	shoes, knitwear (26)	0.32	7.86

shirt, sneakers (166)	knitwear, jacket (24)	0.14	7.74
shirt, sneakers, pant (88)	shoes, knitwear (27)	0.31	7.52

As an example we can interpret Table 2 first row as itemset X{knitwear, tshirt, pant} 73 times has been bought together and itemset Y{tshirt, sweatshirt} 25 times bought together. If one customer bought X itemset, s/he also bought Y itemset. Moreover Confidence of %34 percent means that all member of itemset X and Itemset Y were in the %34 percent of all sale transection. Rest of the rows can be interpreted like wise.

In recommendation systems, association rules like Apriori and FP-growth algorithms have been used, but in this project it is applied with a different aspect and clustering is used on similar customers. Clustering was made using data samples belonging to November, December and January months; about gender (man and woman) and age (10-20, 20-30, 30-40, 40-50, 50-60+). Clustering was applied with K-Means technique and clusters were specified. While finding similarities of two samples, Jaccard, Cosine and Pearson algorithms were used. In Table 4 detailed comparison of the studied datasets are given. After clustering the dataset in Table 3 summary of the mostly bought products according to age and gender is given.

**Table 3.** Top Products by Age and Gender

	Male	Female
10-20	Tshirt	Tshirt, Shoe, knitwear
20-30	Tshirt	Sneakers
30-40	Shirt	Shoe
40-50	Shirt	Shoe, Gmlek
50-60+	Shoe, Shirt	Shoe, Trousers
60-60+	Shirt	Shirt, Shoe

**Table 4.** Comparison of clustered and whole data analysis

	Clustered Data	Whole Data
Jaccard	Total: 274 products Recommended : <b>178</b> products not Recommended: <b>96</b> products <b>%64.96</b> correctly recommending statistics.	Total: 274 products Recommended: <b>159</b> products not Recommended: <b>115</b> products <b>%58.03</b> correctly recommending statistics.
Cosine	Total: 274 products Recommended: <b>178</b> products not Recommended: <b>96</b> products <b>%64.96</b> correctly recommending statistics.	Total: 274 products Recommended: <b>150</b> products not Recommended: <b>124</b> products <b>%54.74</b> correctly recommending statistics.



	Total: 274 products Recommended: <b>178</b> products not Recommended: <b>96</b> products <b>%64.96</b> correctly recommending statistics.	Total: 274 products Recommended: <b>157</b> products not Recommended: <b>117</b> products <b>%57.30</b> correctly recommending statistics.
Pearson	Statistics vary depending on the parameters.	Statistics vary depending on the parameters.
Apriori	Statistics vary depending on the parameters.	Statistics vary depending on the parameters.
FP-Growth	Statistics vary depending on the parameters.	Statistics vary depending on the parameters.

After Pearson and clustering methods are used in tests; between 274 products 180 products have correctly been found, 94 products could not correctly predicted. The accuracy of testing between 100 shopping transaction is resulted with 65.7% ratio.

In clustered set, For Jaccard similarities function 178 out of 270 were correctly recommended with a ratio of 64.96%. Same result is also obtained for Cosine similarity function.

When Jaccard, Pearson and Cosine similarity functions are applied to whole data the accurately recommending performance percentages are 58.03%, 64.96%, and 54.74% respectively.

## V. CONCLUSIONS

In this study original data for online sale transaction of a Turkish retail company has been analyzed in order to find out associative rules for itemsets. First of all data cleared and filtered to get demographical information about data set in hand. WEKA software is utilized for extracting Association rules using two methods namely Apriori and FP-Growth. For collaborative filtering metrics a custom software developed using C# programming language so that better recommendations could be carried out. After data are clustered using K-Means according to ages and gender same analysis is carried out for both clustered and unclustered data to compare the recommendation performance. A performance shift from %54.74 to %64,97 is achieved in recommendation performance. These results contribute in understanding the customers' buying habits and behaviors. These results illustrates that the proposed recommendation system generates valuable knowledge for further processed by relevant departments of the e-commerce site. Since this is a promising result, these association rules then going to be validated with real customers who bought X itemset but not bought Y itemset therefore hoping to promote total sales amount. These association rules are going to be shared with marketing department of the company to initiate a customer relationship management campaign after more association rules are extracted.

## V. REFERENCES

- [1] G. Gürgen, “Birliktelik Kuralları ile Sepet Analizi ve Uygulaması” M.Sc. Thesis Marmara University, İstanbul-Turkey, 2008
- [2] M. E. Aras, “Birliktelik Kuralları ile web siteleri için tavsiye moturu uygulaması”, M.Sc. Thesis Marmara University, İstanbul- Turkey, 2010
- [3] A. Kalkov, “Veri Madenciliği ile bir e-ticaret uygulaması” M.Sc. Thesis Gazi University, Ankara-Turkey, 2006
- [4] J. Hipp, U. Güntzer, and G. Nakhaeizadeh. “Algorithms for association rule mining — a general survey and comparison”, *SIGKDD Explor. Newsl.* vol. 2, no. 1, pp. 58-64, 2000
- [5] U. Sezer, “Optimization of Decision Tree with Association Rules” M.Sc. Thesis, Kocaeli University, Kocaeli-Turkey, 2008
- [6] G. Özdoğan, “Paralel FP-Growth Application in Cluster Computers” M. Sc. Thesis, TOBB Economy and Technology University, Ankara-Turkey, 2010
- [7] M. F. Alaeddinoğlu, “Birliktelik Kuralları ile Van Gölü İçin Mekansal-Zamansal Veri Madenciliği” M.Sc. Thesis ATATÜRK University, Erzurum –Turkey, 2012
- [8] E. Çelikyay, “By the method of text mining, analyze of most frequently used and successive words in Turkish and cooccurrence rules”, M. Sc. Beykent University, İstanbul- Turkey, 2010
- [9] R. C. Agarwal, C. C. Aggarwal, V. V. V. Prasad, “A Tree Projection Algorithm for Generation of Frequent Item Sets”, *Journal of Parallel and Distributed Computing* vol.61, no. 3, pp. 350-371, 2001
- [10] S. Brin, R. Motwani, J. D. Ullman, S. Tsur. , “Dynamic itemset counting and implication rules for market basket data”, *ACM SIGMOD International conference on Management of data, Tucson AZ-USA, 1997*
- [11] B. A. Smith, *Building Data Mining Applications for CRM*, McGraw-Hill Inc., NY USA, 2002
- [12] Tsipstis KK. Chorianopoulos, A. , “Data Mining Techniques in CRM: Inside Customer Segmentation”, *Wiley Publication, West Sussex-UK, 2010*
- [13] Dixit V.S., Gupta S. “Personalized Recommender Agent for E-Commerce Products Based on Data Mining Techniques”, *Intelligent Systems, Technologies and Applications pp 77-90, 2019*
- [14] Bandyopadhyay S., Thakur S.S., Mandal J.K. , “Product Recommendation for E-Commerce Data Using Association Rule and Apriori Algorithm: Proceedings of the International Conference on Modelling and Simulation”, *Modelling and Simulation in Science, Technology and Engineering Mathematics , 2019*

- [15] Samaraweera, Wishma & Waduge, Chekaprabha & Meththananda, Uma. (2016). "Market Basket Analysis: A Profit Based Approach to Apriori Algorithm" , *9th International Research Conference - KDU, Rathmalana, Sri Lanka ,2016*
- [16] Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Pearson, 2013
- [17] Jianjiang Li, Kai Zhang, Xiaolei Yang, Peng Wei, Jie Wang, Karan Mitra, Rajiv Ranjan, "Category Preferred Canopy-K-means based Collaborative Filtering algorithm" *Future Generation Computer Systems*, vol. 93, pp. 1046-1054 , 2019
- [18] B. N. Miller, J. A. Konstan, and J. Riedl, "PocketLens: Toward a personal recommender system", *ACM Transactions on Information Systems*, vol. 22, no. 3, pp. 437-476, 2014
- [19] Zhu, X., Su, S., Fu, M., Liu, J., Zhu, L., Yang, W., Jing, G., Guo, Y. "A Cosine Similarity Algorithm Method for Fast and Accurate Monitoring of Dynamic Droplet Generation Processes", *Scientific Reports Jul* vol. 2, no. 8-1, pp. 9967, 2018
- [20] Seven Kosub, "A note on the triangle inequality for the Jaccard distance", Department of Computer & Information Science, M.Sc. Thesis, University of Konstanz, Konstanz- Germany (2016)
- [21] Sivri E. Ş., "Veri madenciliği/e-ticaret sitesi için ürün tavsiye sistemi geliştirilmesi", M.Sc. Thesis, İstanbul Ticaret Üniversitesi İstanbul-Turkey 2015
- [22] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009