

# KABA KÜME YAKLAŞIMIYLA TÜKETİCİ DAVRANIŞLARININ ÖNGÖRÜLMESİ\*

## PREDICTION OF CONSUMER BEHAVIOUR WITH ROUGH SET APPROACH

Nilgün GÜLER BAYAZIT\*\*  
Yasemen UÇAN\*\*\*

### Özet

Bu çalışmada, tüketici davranışlarının kaba küme teorisine dayanan kural türetme algoritmalarıyla öngörülmesi amaçlanmıştır. Romanya Akademisi Dünya Ekonomisi Enstitüsü tarafından toplanan “Romanya - finansal okuryazarlık ve finansal hizmetler anketi” verileri kullanılmıştır. Çalışmada ön işlem olarak kaba küme teorisinin alt yaklaşım ve üst yaklaşım kümeleri kullanılarak veri kümesindeki gereksiz nitelikler elenmiş sadece vazgeçilemeyen niteliklerden oluşan çekirdek veri kümesi oluşturulmuştur. Çekirdek veri kümesine kural türetme algoritması olan LEM2 algoritması uygulanarak tüketici davranışlarını yüksek başarıyla öngörebilen kurallar elde edilmiştir. Elde edilen sonuçlara göre kaba küme teorisine dayanan kural türetme algoritmaları tüketicilerin karar verme biçimleri ve davranış biçimlerinin öngörülmesini sağlayabilecek uygun bir araçtır.

**Anahtar Kelimeler:** Pazarlama, tüketici davranışı, kaba kümeler, kural tümevarım, veri madenciliği.

**JEL Sınıflandırması:** C45, C53, C63, D10

### Abstract

The aim of this work is the prediction of consumer behaviour by induction algorithms based on rough set theory. To demonstrate this point, “Romania - financial literacy and financial services survey” data collected by the Institute for World Economy of Romanian Academy has been used as the data set. As a preprocessing step in the current work, the redundant features in the data set have been filtered out by employing the lower and upper approximation of the rough set theory to yield the core reduct data set consisting of only the essential features. By applying the LEM2 rule induction algorithm to the

---

\* Makalenin farklı bir versiyonu IV. European Congress on Economic Issues(2018) de sunulmuş ve özeti konferans özet kitabında basılmıştır.

\*\* Dr. Öğretim Üyesi, Yıldız Teknik Üniversitesi, Kimya-Metalürji Fakültesi, Matematik Mühendisliği Bölümü, guler@yildiz.edu.tr, Orcid Id: 0000-0003-0221-294

\*\*\* Dr. Öğretim Üyesi, Yıldız Teknik Üniversitesi, Kimya-Metalürji Fakültesi, Matematik Mühendisliği Bölümü, ucan@yildiz.edu.tr, Orcid Id: 0000-0001-7634-7869

core reduct data set, rules that can predict consumer behaviour with high performance (success rates) have been obtained. It can be concluded that rule induction algorithms based on rough set theory are suitable tools for predicting the decision making ways and behaviour patterns of consumers.

**Keywords:** Marketing, consumer behaviour, rough sets, rule induction, datamining

**JEL Classification:** C45, C53, C63, D10

## I. Giriş

Tam ve kesin olmayan bilgi modellemesi için küme teorisini kullanan yöntemlerden biri de kaba kümelerdir (Rough sets). Kaba kümeler cebirsel özellikleri nedeniyle matematiğin yanı sıra tıp, ekonomi ve mühendislik alanlarında da kullanılmaktadır. Bu makalede de tüketici davranışı, kaba küme yöntemi ile analiz edilecektir. Özellikle küreseleşmenin etkisiyle artan rekabet koşulları altında firmaların ürün pazarlama stratejilerinin bir sonucu olarak farklılaşan tüketici davranışlarının kaba küme yöntemi ile analizi, tüketim kararlarını etkileyen faktörlerin neler olduğunun belirlenmesi ve uygun stratejilerin geliştirilebilmesi bağlamında önemlidir.

Tüketicilerin davranış bilgisi, firmalar için çok önemlidir. Bu nedenle, firmalar, yalnızca yeni müşteriler elde etmek için değil aynı zamanda müşterileriyle olan mevcut ilişkilerini geliştirmek ve diğer firmalar ile rekabet edebilirliklerini geliştirmek için tüketicilerin satın alma eğilimlerini öngörebilmelerine olanak sağlayan modellere ihtiyaç duymaktadırlar. Tüketici davranışlarının tahmin edilmesinde en yaygın kullanılan veri toplama yöntemi anket yöntemidir. Analiz sürecinde, anketlerden elde edilen verilerin analizi için kullanılan regresyon ve korelasyon analizi, makina öğrenmesi (örneğin YSA) yöntemleri değişkenler arasındaki ilişkiyi sayısal olarak belirlemede bu yöntemler, değişkenler arasındaki neden ve sonuç ilişkisini tam olarak açıklayamazlar. Kaba küme yaklaşımı mevcut yöntemler içinde en yeni ve etkin olanıdır.

Bazı çalışmalar ise anketlerde olası belirsizliğe ve muğlaklığa karşı bulanık küme teorisini önermişlerdir. Zadeh (1965)<sup>1</sup> tarafından önerilen bulanık küme, bilginin kademeli oluşu ile ilgilenir ve belirsizlik bulanık üyelik ile ifade edilir. Ancak anket çalışmaları bağlamında önemli bir sorun olan eksik veya tam olmayan bilgi ile başa çıkabilme yeteneği Pawlak (1982)<sup>2</sup> tarafından geliştirilen kaba küme teorisi ile çözülmüştür. Bu yaklaşım veri hakkında önsel yada ek bilgiye ihtiyaç duymamaktadır. Bilginin parçacıklı oluşu ile ilgilenir ve ayırt edilemezlik bağıntısı ile ifade edilir. Ayırt edilemezlik söylem evreninin tanımını etkileyen bilgi granülleri<sup>3</sup> anlamına gelir<sup>4</sup>.

---

1 Zadeh, L. A. (1965). Fuzzy sets, Information and control, 8(3): 338-353.

2 Pawlak, Z. (1982). Rough Sets, International Journal of Computer and Information Sciences, 11: 341-356.

3 Ayırt edilemezlik denklik bağıntısına göre oluşturulan denklik sınıflarına granül (parçacık) denir. Tripathy, B.K. (2009). "Rough Sets on Fuzzy Approximation Spaces and Intuitionistic Fuzzy Approximation Spaces"; Rough Set Theory: A True Landmark in Data Analysis, Derleyen: Abraham A., Falcón R., Bello R., Series: Studies in Computational Intelligence, (174), Springer, Berlin, Heidelberg: s.6.

4 Tripathy, 2009, s.6

Bu çalışmanın iki amacı vardır, ilki tüketici davranışlarını kaba küme yöntemi ile modellemek iken ikincisi kaba küme yaklaşımının yukarıda bahsedilen yöntemlerden özellikle Yapay sinir ağları (YSA) dan üstünlüğünü göstermektir. Yöntem YSA gibi makina öğrenmesi algoritmalarına önemli bir alternatiftir<sup>5</sup>.

Makale takip eden şekilde organize edilmiştir. İkinci bölümde literatür taraması verilmiştir. Üçüncü bölümde kaba küme teorisi, LEM2 tümevarım kural algoritması ve YSA tanıtılıp, dördüncü bölümde veri seti tanımlanmıştır. Beşinci bölümde kaba küme ve YSA yöntemleri kullanılarak sonuçlar elde edilmiştir. Son bölümde sonuç ve tartışmalar verilmiştir.

## 2. Literatür Taraması

Kaba küme teorisi Pawlak (1982) tarafından önerildikten sonra birçok alanda kullanılmıştır. Bazı çalışmalar yöntemi tüketici davranışlarını analiz etmek için kullanmıştır. Cui ve Cui (2009)<sup>6</sup> büyük ölçekli bir anketten toplanan verileri kullanarak, tüketicilerin satın alma kararlarında yurt içi ve yurt dışı marka tercihlerini öngörmek için kaba küme yaklaşımını kullanmıştır. Huang ve diğerleri (2010)<sup>7</sup> Tayvanlı cep telefonu kullanıcıları üzerinde, 4G mobil cihazlarının hangi özellikleri olmalı konulu anket çalışmasının sonuçlarını değerlendirmek için kaba küme teorisinden yararlanmışlardır. Çalışmalarında, kullanıcı davranışlarının kural tabanlı ifade edilmesi için Pawlak (2005) akış çizgesini<sup>8</sup> kullanmışlardır. Liou ve Tzeng (2010)<sup>9</sup> çalışmasında, baskınlık-ilkesine dayanan kaba küme<sup>10</sup> yaklaşımını kullanarak hava yolları müşterilerinin davranışlarını ve bağlılıklarını başarıyla öngörmüşlerdir. Emilio Celotto ve diğerleri (2012)<sup>11</sup> çalışmalarında, İtalyan turistlerin davranışlarını belirleyen kuralları çıkarmak (analiz etmek) için kaba küme teorisini ve veri madenciliği tekniklerini kullanmışlardır. Furaji ve diğerleri (2013)<sup>12</sup> reklam çekiciliğinin kadın ve erkek satın alma kararlarına olan etkisini inceledikleri çalışmalarında tüketici davranış kurallarını belirlemek için kaba küme teorisini kullanmışlardır. Bu amaçla ev aletleri satın alma tercihleriyle ilgili 44 sorudan oluşan anketi 174 kişiye uygulamışlar, cinsiyet faktörünün satın alma kararlarında etkin bir faktör olmadığını bulmuşlardır. Zhang ve diğerleri

5 YSA gibi makina öğrenmesi algoritmaları çok yüksek başarımları vermelerine rağmen, verdikleri sonuçların açıklanabilirliği düşüktür

6 Cui, N., Cui, Q. (2009). A Rough-Set Based Approach to Predict Consumers' Brand Preference, International Workshop on Intelligent Systems and Applications, Wuhan, 1-4 .

7 Huang, C., Yang, Y., Tzeng, G., Cheng, S., Lee, H. (2010). 4G Mobile Phone Consumer Preference Predictions by Using the Rough Set Theory and Flow Graphs, PICMET 2010 Technology Management for Global Economic Growth, Phuket: 1-10.

8 Pawlak, Z. (2005). Rough Sets and Flow Graphs, Lecture Notes in Computer Science, 3641: 1-11.

9 Liou, J.J.H, Tzeng, G., (2010). A Dominance-based Rough Set Approach to Customer Behavior in the Airline Market, Information Sciences, 180 (11): 2230-2238.

10 Blaszczynski, ve diğerleri (2007). Multi-criteria Classification – a New Scheme for Application of Dominance-based Decision Rules, European Journal of Operational Research 181: 1030–1044.

11 Celotto, E., ve diğerleri (2012). Short-medium term tourist services demand forecasting with rough set theory. Procedia Economics and Finance, 3, 62-67.

12 Furaji ve diğerleri (2013). Study on the influence of advertising attractiveness on the purchase decisions of women and men., Journal of International Studies, 6(2): 20-32.

(2014)<sup>13</sup> e-ticaretteki tüketici davranışlarını tahminlemede kullanılmak üzere ayırt edilebilirlik matrisini kullanarak nitelik azaltma algoritması önermişlerdir. Liao ve Chang (2016)<sup>14</sup> çevrimiçi tüketicilerin olası davranış değişikliklerini belirlemek için kaba küme temelli ilişki kuralı yaklaşımı önermişlerdir.

### 3. Yöntem

#### 3.1 Kaba Küme Teorisi

Kaba kümeler tanıtılırken genellikle matematiksel terimlerin yerine uygulama alanına yönelik terimler kullanılır. Bu çalışmada, boş kümeden farklı evren kümesi  $E$  sonlu bir küme,  $\mathbf{P}(E)$  kuvvet kümesi,  $A$  niteliklerin kümesi olmak üzere  $BS=(E, A)$  bilgi sistemi olarak tanımlanır. Bu çalışmada kullanılan tanımlar Pawlak (1982), Walzack ve Massart (1999)'dan hareketle aşağıda verilmiştir<sup>1516</sup>:

##### Tanım 1:

$V_a$  niteliklerin aldığı değerlerin kümesi olmak üzere  $f_a$  bilgi fonksiyonu ;

$\forall a \in A$  için  $f_a : E \rightarrow V_a$  ile tanımlanır.

##### Tanım 2:

$\forall B \subseteq A$  alt kümesinde ayırt edilemezlik bağıntısı  $\beta$  olmak üzere  $\forall x_i, x_j \in E$  ve  $\forall a \in B$  için  $a(x_i) = a(x_j)$  şeklinde tanımlanan  $\beta$  bağıntısı bir denklik bağıntısıdır. Herhangi bir  $x_i \in E$  için  $\beta$  bağıntısına göre  $x_i$ 'nin denklik sınıfı  $[x_i]_\beta = \{x_j \in E : x_i \beta x_j\}$  şeklinde tanımlanır. Bu denklik sınıfları uygulamalarda elementer küme veya granül olarak adlandırılır ve kaba kümelerle sınıflandırmada ilk adımdır.

##### Tanım 3:

$X$  kümesi  $E$  evreninin bir alt kümesi olsun.  $X$  kümesinin  $B \subseteq A$  alt kümesinde, alt ve üst yaklaşımları sırasıyla

$BX_* = \{x_i \in E : [x_i]_\beta \subseteq X\}$  ve  $BX^* = \{x_i \in E : [x_i]_\beta \cap X \neq \emptyset\}$  şeklinde tanımlanır.

13 Zhang, ve diğerleri. (2014). Application of rough sets in E-commerce consumer behavior prediction, Advanced Science and Technology Letters Vol.53 (ICM 2014): 255-260

14 Liao , S. H., Chang, H. K (2016). A rough set-based association rule approach for a recommendation system for online consumers, Information Processing & Management, 52(6):1142-1160.

15 Pawlak, Z. (1982). Rough Sets, International Journal of Computer and Information Sciences, 11: 341-356.

16 Walczak, B., Massart, D. L. (1999). Rough Sets Theory, Chemometrics and Intelligent Laboratory Systems, 47 (1): 1-16

$E$  de  $X$  kümesinin sınırı ise  $BX_s = BX^* - BX$ , şeklinde tanımlanır. Eğer alt ve üst yaklaşımlar özdeşse bu durumda  $X$  tanımlanabilir, aksi durumda  $X$  kümesi  $E$ 'de tanımlanamaz.

#### Tanım 4:

Bir  $X$  kümesinin  $B \subseteq A$  alt kümesinde ölçümü;  $s(BX_s)$  ve  $s(BX^*)$  sırasıyla alt ve üst yaklaşım kümelerinin elaman sayısı olmak üzere ,

$$\mu_B(X) = \frac{s(BX_s)}{s(BX^*)} \text{ şeklinde tanımlanır.}$$

Eğer  $X$ ,  $E$  de tanımlanabilirse  $\mu_B(X) = 1$ ,

eğer  $X$ ,  $E$  de tanımlanamaz ise  $\mu_B(X) < 1$  dir.

Dolayısıyla ölçüm olma koşulundan  $0 \leq \mu_B \leq 1$  dir.

#### Tanım 5:

Eğer  $\beta(A) = \beta(A - a_i)$  koşulu sağlanıyorsa  $a_i$  niteliğine gereksiz nitelik denir. Aksi durumda  $A$  kümesinde vazgeçilemez nitelik olarak adlandırılır<sup>17</sup>.

Bilgi sistemine yukarıdaki tanımlar uygulandıktan sonra, kaba kümelere gereksiz nitelikler elenerek bilgi sistemi basitleştirilir. Bu indirgeme yöntemi ile bilgi sisteminde sadece vazgeçilemeyen niteliklerden oluşan çekirdek kümesi elde edilir. Çekirdek ve indirgemeyi hesaplayabilmek için ayırt edilebilirlik matrisi kullanılır. Böylece tanımlı işlemler kullanılarak karar tablosu oluşturulur.

### 3.2 Kural Tümevarım Algoritmaları

Verilerdeki gizli kalıpların kurallar ile ifade edildiği, kural tümevarımı makine öğrenmesinin en önemli tekniklerinden biridir. Türetilen kurallar verideki bilginin ifade edilmesinde kullanılırlar. Kolayca anlaşılabilir ve konunun uzmanları tarafından incelenerek geçerlilikleri kontrol edilebilir. Kurallar aşağıdaki formdadır:

Eğer (Nitelik<sub>1</sub>, değer<sub>1</sub>) ... .. ve (Nitelik<sub>n</sub>, değer<sub>n</sub>) ise (Karar, değer)

Burada nitelikler bağımsız değişken iken karar ise bağımlı değişkendir. AQ<sup>18</sup>, CN2<sup>19</sup>, ID3<sup>20</sup> ve C4.5<sup>21</sup> algoritmaları kural türetmede kullanılan algoritmalarından bazılarıdır.

17 Walczak, B., Massart, D. L. (1999), s.1-16.

18 Michalski, ve diğerleri (1986). The AQ15 Inductive Learning System: an Overview and Experiments, Department of Computer Science, University of Illinois.

19 Clark, P., Niblett, T. (1989). The CN2 Induction Algorithm, Machine Learning, 3 (4): 261-283.

20 Quinlan, J.R. (1986). Induction of Decision Trees, Machine. Learning, 1:81-106.

21 Quinlan, J.R. (1993). C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers Inc., San Francisco, CA

### 3.2.1 LEM2 Algoritması (Learning from Example Module)

LEM2 algoritması, kaba küme teoresinin alt yaklaşım kümesi yada üst yaklaşım kümesine dayanan, veriden kural oluşturmada kullanılan gözetimli bir öğrenme algoritmasıdır.<sup>22</sup> Algoritma, (nitelik, değer) ikilileri seviyesinde çalışır. Her bir kavram için yerel bir kapsama hesaplanır. Bu algoritmayı tanımlayan kavramlar Grzymala-Busse (2009)<sup>23</sup>'da verilmiştir:

#### Tanım 6:

$(a, v)=t$  her bir (nitelik, değer) ikilileri için  $t$ 'nin bir bloğu  $[t]$  ile gösterilir. Öyleki  $a$  niteliği için  $v$  değerine sahip olan  $E$ 'deki tüm durumların kümesidir.

#### Tanım 7:

$F$  kümesi,  $(d, w)$  (karar, değer) ikilileri ile temsil edilen alt veya üst yaklaşım kümesi olsun. Bu durumda

$F$  kümesi  $t=(a, v)$  ikililerinin oluşturduğu bir  $T$  kümesine bağlıdır  $\Leftrightarrow \emptyset \neq [T] = \bigcap_{t \in T} [t] \subseteq F$ .

$T$  kümesi,  $F$ 'nin minimal kompleksidir ancak ve ancak  $F, T$ 'ye bağlıdır ve  $T$ 'nin  $T'$  diye bir özalt kümesi yoktur.

#### Tanım 8:

$\mathcal{T}$ , (nitelik, değer) ikililerinin oluşturduğu boş kümeden farklı kümelerin bir ailesi olsun. Bu durumda

$\mathcal{T}, F$ 'nin yerel örtüsüdür  $\Leftrightarrow$  aşağıdaki koşullar sağlanıyorsa

(1)  $\forall T \in \mathcal{T}$  üyesi  $F$ 'nin minimal kompleksidir,

(2)  $\bigcup_{T \in \mathcal{T}} [T] = F$ ,

$\mathcal{T}$  minimaldir yani  $\mathcal{T}$  mümkün olan en az sayıda üyeye sahiptir.

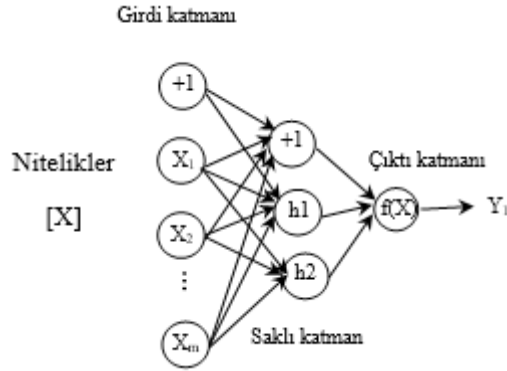
### 3.3 Yapay Sinir Ağları

YSA, bir veri kümesi üzerinde eğitilerek bir işlevi  $f(X):R^m \rightarrow R^n$  öğrenen denetimli bir öğrenme algoritmasıdır. Sınıflandırma ve regresyonda kullanılan bir YSA yapısı olan çok katmanlı algılayıcılar (ÇKA) doğrusal olmayan bir fonksiyon tahmincisini öğrenebilirler. ÇKA yapısı girdi, çıktı ve bir yada daha fazla saklı katmandan oluşmaktadır. Şekil 1'de girdi, çıktı ve bir saklı katmandan oluşan ÇKA yapısı görülmektedir. Katmanlardaki her bir düğüm ile bir sonraki katmandaki düğüm arasında  $w_i \in \mathfrak{R}$  bağlantı ağırlığı tanımlıdır<sup>24</sup>. Her bir düğümün çıktı değeri, girdilerinin ağırlıklı toplamına doğrusal olmayan bir aktivasyon işlevi uygulanarak hesaplanır.

22 Grzymala-Busse J.W. (2009). Rule Induction, In: Maimon O., Rokach L. (eds) Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA:1-19.

23 Grzymala-Busse (2009), 9.

24 Alpaydın, E. (2011). Yapay Öğrenme. İstanbul, Boğaziçi Üniversitesi Yayınevi, s.200.



Şekil 1: 1 saklı katmanlı ÇKA.

ÇKA'lar da öğrenme, eğitim algoritmasının veri kümesindeki her bir girdi için doğru çıktı değerini veren  $w$  ağırlık değerlerinin belirlenmesidir. Bunun için Rumelhart ve diğerleri tarafından (1986)<sup>25</sup> geliştirilen hata geri yayma algoritması kullanılır.

#### 4. Veri Seti

Çalışmada Dünya Bankası'nın sponsorluğunda 2008-2009 yıllarında tüketici korunması ve finansal okuryazarlık konusunda Romanya'da yapılan anketin verileri kullanılmıştır.<sup>26</sup> Anket çalışmaları Romanya Akademisi Dünya Ekonomisi Enstitüsü tarafından yürütülmüştür. Anket

- Finansal piyasaya ilişkin tutum ve görüşler
- Finansal hizmetler hakkında bilgi
- Finansal okur-yazarlık
- Hane halkı maliyesi yönetimi
- Hane halkı üyelerinin demografik bilgileri
- Finansal servislerin kullanımı

hakkında soruların sorulduğu çoktan seçmeli 60 sorudan oluşmaktadır. Anket 2389 katılımcı tarafından cevaplanmıştır.

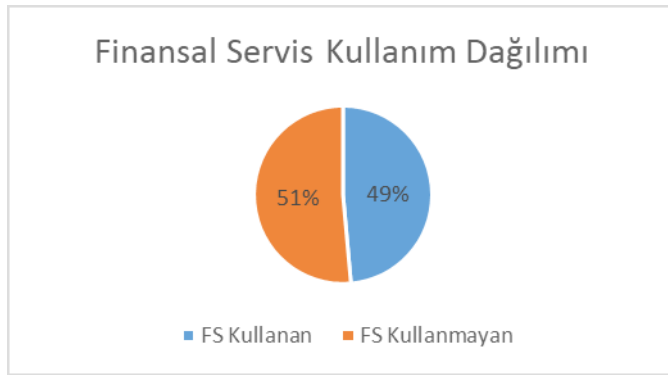
25 Rumelhart ve diğerleri (1986), Learning Representation by Back-propagating Errors, Nature, 323, 533-536.

26 <http://microdata.worldbank.org/index.php/catalog/1027/datafiles> url adresinden veri setine ve anket sorularına erişmek mümkündür (Erişim Tarihi: 30.10.2018).

## 5. Analiz Sonuçları

Amacımız anketi cevaplandıran katılımcıların finansal servis „(FS) kullanma alışkanlıklarını belirlemek için karar kurallarını belirlemektir. Çalışma, açık kaynak kodlu R programı<sup>27</sup> üzerinde RoughSets<sup>28</sup> kütüphanesi kullanılarak gerçekleştirilmiştir.

Bu amaçla katılımcıların, demografik bilgileri (cinsiyet, yaş, meslek, mezun olunan okul, oturulan bölge, uyruk vb), finansal verileri (aylık gelir, eve gelen toplam gelir miktarı vb.), günlük alışkanlıkları (ne sıklıkta gazete okunuyor, radyo dinleniyor, TV programı izleniyor vb) veri olarak kullanılmıştır. Veri içerisindeki eksik değerler içeren katılımcıların cevapları silindiğinde geriye 1486 örnek kalmıştır. Anketteki hangi FS'lerin kullanıldığı (SF1-SF6) sorularına verilen cevaplar bağlamında “FS kullananlar” ve “FS kullanmayanlar” diye sınıf etiketleri oluşturulmuştur. Şekil 2 ankete katılan katılımcıların FS kullanma dağılımlarını göstermektedir.



**Şekil 2:** Finansal servis kullanım dağılımı

İlk olarak, kaba küme kapsamında ayırt edilebilirlik matrisi oluşturulmuş ve bu matris yardımıyla çekirdek küme belirlenmiştir. Çekirdek küme, tüketicilerin davranışlarını etkileyen en önemli (vazgeçilemeyen) niteliklerden oluşmaktadır. Tablo 1’de tüketicilerin davranışlarını etkileyen en önemli nitelikler listelenmiştir. Buna göre oturulan bölge, tüketicinin eğitim durumu gibi nitelikler FS kullanma alışkanlıkları üzerinde etkin rol oynamaktadır.

27 R Core Team (2018). R :A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. R version 3.5.0. <https://www.R-project.org/> (Erişim Tarihi: 1.9.2018).

28 Rıza, L. S., Janusz, A. (2015). RoughSets: Data Analysis Using Rough Set and Fuzzy Rough Set Theories. R package version 1.3-0. <https://CRAN.R-project.org/package=RoughSets> (Erişim Tarihi: 1.9.2018).



**Tablo 1:** FS Kullanan Tüketicilerin Davranışlarını Etkileyen Nitelikler

Oturulan bölge(region)	Oturulan bölgenin refah düzeyi
Meslek	Kredi kartı sahibi olma
Aylık gelir	Yaş
Oturulan bölgenin özelliği	Oturulan yerin köy/kasaba/şehir olması
Mezun olunan son okul	Tüketici kredisi kullanımı
Mezuniyetten sonra uzmanlaşma eğitimi alınması	Cinsiyet

Daha sonra LEM2 algoritması kullanılarak karar kuralları oluşturulmuş, bu kurallara ait destek ve güvenilirlik ( $\ell$  : laplace) değerleri hesaplanmıştır. 133' tane kural çıkarılmıştır. Tablo 2'de FS kullanan ve kullanmayan tüketiciler için algoritma tarafından üretilen 3'er tane örnek kural gösterilmiştir.

$$laplace(A \rightarrow C) = \frac{\#(AUC)+1}{\#(A)+2} \quad [1]$$

**Tablo 2:** Örnek Karar Kuralları ( $S$ : kuralla örtüşen örnek sayısını,  $\ell$ : kuralın güvenilirlik derecesini göstermektedir  $0 \leq \ell \leq 1$ ).

(1) Eğer «Köy/kasaba/şehir merkezinde oturuyorsa» ve «yaştan dolayı emekliyse» ve «Mezun olduktan sonra uzmanlaşma / yeniden yeterlilik için başka ders aldıysa» → “FS kullanmaz”	( $S=8; \ell=0.90$ )
(2) Eğer «Eğitimine devam ediyorsa» ve «geliri [42, 450] aralığında ise» ve «Oltenya bölgesinde ikamet ediyorsa» → “FS kullanmaz”	( $S=8; \ell=0.90$ )
(3) Eğer « Mezun olduktan sonra uzmanlaşma / yeniden yeterlilik için başka ders almadıysa » ve « köy/kasaba/şehir'in merkezi dışında oturuyorsa » ve «eğitim seviyesi lisenin altındayda » → “FS kullanmaz”	( $S=8; \ell=0.90$ )
(4) Eğer « Eğer iyi bil bölgede oturuyorsa » ve « çalışıyor » ve « geliri [1000, 20000] aralığında ise » ve « 200.000 bin nüfuslu bir kasabada yaşıyorsa » → “FS kullanır”	( $S=20; \ell=0.95$ )
(5) Eğer « Mezun olduktan sonra uzmanlaşma / yeniden yeterlilik için başka ders almadıysa » ve « [41, 54] yaş aralığında » ve « eğitim seviyesi üniversite/yüksek lisans ise » → “FS kullanır”	( $S=13; \ell=0.93$ )
(6) Eğer «Bayansa» ve «Köy/kasaba/şehir merkezinde oturuyorsa» ve «gelişmiş kırsal bölgede yaşıyorsa» ve «Muntenia bölgesinde ikamet ediyorsa» → “FS kullanır”	( $S=8; \ell=0.9$ )

Örnek karar kuralı 4'e göre eğer kişi 200.000 nüfuslu bir kasabanın iyi bir bölgesinde oturup, gelir seviyesi [1000, 20000] aralığında ise FS kullanmaktadır. Bu karar kuralının sıklığı 20, güven değeri ise 0,95'tir.

Elde edilen kurallar incelendiğinde şu sonuçlar gözlenmiştir:

- yaştan dolayı emekli olan
- eğitim hayatı devam eden
- eğitim seviyeleri lise ve altı olan

- ülkenin az gelişmiş bölgelerinde, yada kırsal kesiminde oturanlar
- gelir seviyeleri düşük olan
- nüfus yoğunluğu düşük olan bölgelerde yaşayan tüketicilerin FS kullanmadıkları, buna karşın

- eğitim seviyeleri en az üniversite olan
- gelir seviyeleri yüksek olan
- ülkenin gelişmiş bölgelerinde yaşayan
- nüfus yoğunluğu yüksek şehir merkezlerinde yaşayan

tüketicilerin FS kullandıkları gözlenmiştir.

Badea (2014)'te aynı anket veri kümesini kullanarak katılımcıların FS kullanım davranışlarını ÇKA yöntemiyle öngörecekle bir model üzerinde çalışmıştır. Badea çalışmasında bilgi değeri metriğini kullanarak belirlediği 16 nitelik'ten oluşan veri kümesini kullanmıştır. Bu niteliklerin listesi Tablo 3'de verilmiştir.

**Tablo 3:** Badea (2014)<sup>29</sup> Çalışmasında Kullanılan Nitelikler

Oturulan bölge(region)	Oturulan bölgenin özelliği II
Meslek	Oturulan bölgenin refah düzeyi
Yaşanılan şehir tipi	Kredi kartı var mı
Oturulan bölgenin özelliği I	Yaş
Mezun olunan son okul	Oturulan yerin köy/kasaba/şehir
Mezuniyetten sonra uzmanlaşma eğitimi alınıp alınmadığı	Tüketici kredisi kullanımı
Aile arabası var mı	Cinsiyet
Uyruğu	İpotek/araba kredisi kullanımı

Kaba küme yaklaşımıyla elde edilen karar kurallarının tüketici davranışlarını öngörmedeki başarımını karşılaştırmak için çekirdek kümesindeki niteliklerden oluşan veri kümesine ÇKA uygulanarak elde edilen sonuçlar, doğruluk oranı (DO)[2] ölçütünde karşılaştırılmıştır.

$$DO = \frac{\text{Doğru öngörülen örnek sayısı}}{\text{Örnek sayısı}} \quad [2]$$

Tablo 4 LEM2 ve ÇKA ile elde edilen başarım sonuçlarını göstermektedir. Veri küme 1, kaba küme teorisinin veri indirgeme özelliği kullanılarak elde edilen 12 nitelikten, Veri küme 2 ise Badea'nın belirlediği 16 nitelikten oluşmaktadır.

<sup>29</sup> Bilgi değeri metriği kullanılarak, etkili olduğu belirlenen nitelikler (Şekil 1). Badea (Stroie), L. M. (2014). Predicting Consumer Behavior with Artificial Neural Networks, Emerging Market Queries in Finance and Business EMQFB2013, Procedia Economics and Finance 15, s. 242.

Tablo 4’te de görüldüğü gibi 12 nitelikten elde edilen öngörü sonuçları çok daha başarılıdır. Ayrıca kaba küme teorisine dayanan yöntemin YSA’ya üstünlüğü açık şekilde görülmektedir. Yöntemin, YSA’ya üstünlüğünün nedeni YSA’nın verdiği sonuçların açıklayıcı kurallara dönüştürülmesindeki zorluktur. Üstelik öğrenme katsayısı, saklı katmandaki düğüm sayısı vb. parametreler algoritma sonucunu etkilemektedir. Diğer bir neden ise YSA’ların girdi bilgisi olarak sadece nümerik değerler almasıdır. Veri kümesini YSA’da modelleyebilmek için tüm kategorik nitelikler gölge değişkene çevrilmelidir. Oysa ki anket verileri kategoriktir. Bu da veri kümesindeki nitelik sayısını çok arttırmaktadır. Örneğin 16 kategorik nitelikten oluşan bir veri kümesi gölge değişkenlerle 70 niteliğe çıkabilmektedir. Diğer yandan anketlerin maliyetler yüzünden sadece sınırlı sayıda katılımcıya uygulanması nedeniyle örnek sayısının az, nitelik sayısının çok olduğu durumlar boyut laneti olgusunu ortaya çıkarmaktadır. YSA ile veri eğitilirken hatayı minimum yapmak için kullanılan geriye dönük induksiyon (backward induction) algoritmasında, objektif fonksiyonunun, her bir değer kombinasyonu için hesaplanması gerekmektedir. Veri kümesindeki nitelik sayısının büyük olduğu durumlarda bu önemli bir engeldir. Bu nedenle anket çalışmalarından elde edilen veri kümelerini kullanarak tüketici davranışlarının modellenmesinde kaba küme teorisine dayanan yöntem daha başarılıdır.

**Tablo 4:** FS Kullanan Tüketici Davranışlarının Tahminlenmesi Doğruluk Oranları (Sonuçlar 10-Kat çapraz geçişleme kullanılarak elde edilmiştir)

Veri kümesi	LEM2	ÇKA
Veri küme 1 (12 nitelik)	%75,11	%72,0
Veri küme 2 (16 Nitelik)	%72,64	%73,28

Veri küme 1 ve Veri küme 2’nin nitelikleri (kullandıkları anket soruları) karşılaştırıldığında, Veri küme 2’nin daha çok finansal servis kullanımı ile ilgili anket cevaplarını kullandığı görülmektedir.

Badea (2014)’ün önerdiği ÇKA modeli, tüketici davranışlarının eğilimini belirlerken finansal servis kullanmayan tüketicilerin eğilimini belirlemede yetersiz kalmaktadır. Bizim, önerdiğimiz model tüketicilerin sadece demografik ve finansal bilgilerini kullanmakta, daha önceden finansal servis kullanıp kullanmadığı bilgisine ihtiyaç duymamaktadır. Bu sayede yeni müşterler için politika geliştirmeye olanak sağlamaktadır.

## 6. Tartışma ve Sonuç

Bu çalışmada tüketici davranışları kaba küme teorisi kullanılarak analiz edilmiştir. Çalışmanın sonuçlarına göre tüketici davranışlarının modellenmesinde, açıklanabilirlik seviyesi daha yüksek olan kural tabanlı LEM2 algoritması, YSA gibi makina öğrenmesi algoritmalarına önemli bir alternatiftir.

Kaba küme teorisi tabanlı algoritmalar eşikler veya uzman bilgisi gibi veriler hakkında ek bilgi gerektirmemesi, orijinal veriyi azaltarak, asıl verilerdeki bilgilerin aynısı olan asgari veri

kümelerini bulmaya imkan vermesi sebebiyle, anket verilerinin işlenerek, tüketicilerinin karar verme tarzları ve davranış biçimlerinin öngörülmesini sağlayabilecek uygun bir araçtır.

## **Kaynakça**

- ALPAYDIN, E. (2011). *Yapay Öğrenme*. İstanbul: Boğaziçi Üniversitesi Yayınevi.
- BADEA (STROIE), L. M. (2014). Predicting Consumer Behavior with Artificial Neural Networks, Emerging Market Queries in Finance and Business EMQFB2013, *Procedia Economics and Finance* 15: 238 – 246
- BLASZCZYNSKI, J., Greco, S., Slowinski, R. (2007). Multi-criteria Classification – a New Scheme for Application of Dominance-based Decision Rules, *European Journal of Operational Research* 181: 1030–1044.
- CLARK, P., Niblett, T. (1989). The CN2 Induction Algorithm, *Machine Learning*, 3 (4): 261-283.
- CELOTTO, E., Ellero, A., Ferretti, P. (2012). Short-medium term tourist services demand forecasting with rough set theory. *Procedia Economics and Finance*, 3:62-67.
- CUI, N., Cui, Q. (2009). A Rough-Set Based Approach to Predict Consumers' Brand Preference, *International Workshop on Intelligent Systems and Applications*, Wuhan, 1-4 .
- GRZYMALA-BUSSE J.W. (2009). Rule Induction, In: Maimon O., Rokach L. (eds) *Data Mining and Knowledge Discovery Handbook*. Springer, Boston, MA:1-19.
- FURAJI, F., Latuszyńska, M., Wawrzyniak, A., Wasikowska, B. (2013). Study on the influence of advertising attractiveness on the purchase decisions of women and men., *Journal of International Studies*, 6(2): 20-32.
- HUANG, C., Yang, Y., Tzeng, G., Cheng, S., Lee, H. (2010). 4G Mobile Phone Consumer Preference Predictions by Using the Rough Set Theory and Flow Graphs, *PICMET 2010 Technology Management for Global Economic Growth*, Phuket: 1-10.
- LLAO, S. H., Chang, H. K. (2016). A rough set-based association rule approach for a recommendation system for online consumers, *Information Processing & Management*, 52(6):1142-1160.
- LIOU, J.J.H, Tzeng, G., (2010). A Dominance-based Rough Set Approach to Customer Behavior in the Airline Market, *Information Sciences*, 180 (11): 2230-2238.
- MICHALSKI, R. S., Mozetic, I., Hong, J., Lavrac, N. (1986). The AQ15 Inductive Learning System: an Overview and Experiments, *Department of Computer Science, University of Illinois*.
- PAWLAK, Z. (1982). Rough Sets, *International Journal of Computer and Information Sciences*, 11: 341-356.
- PAWLAK, Z. (2005). Rough Sets and Flow Graphs, *Lecture Notes in Computer Science*, 3641: 1-11.
- R Core Team (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. R version 3.5.0. <https://www.R-project.org/> (Erişim Tarihi: 1.9.2018).
- RIZA, L. S., Janusz, A. (2015). RoughSets: Data Analysis Using Rough Set and Fuzzy Rough Set Theories. R package version 1.3-0. <https://CRAN.R-project.org/package=RoughSets> (Erişim Tarihi: 1.9.2018).
- RUMELHART, D. E., Hinton, G. E., Williams, R. J. (1986), Learning Representation by Back-propagating Errors, *Nature*, 323, 533-536.
- QUINLAN, J.R. (1986). Induction of Decision Trees, *Machine Learning*, 1:81–106.
- QUINLAN, J.R. (1993). *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers Inc., San Francisco, CA.

- TRIPATHY, B.K. (2009). "Rough Sets on Fuzzy Approximation Spaces and Intuitionistic Fuzzy Approximation Spaces", *Rough Set Theory: A True Landmark in Data Analysis*, Derleyen: Abraham A., Falcón R., Bello R., Series: Studies in Computational Intelligence, (174), Springer, Berlin, Heidelberg: 3-44.
- WALCZAK, B., Massart, D. L. (1999). *Rough Sets Theory, Chemometrics and Intelligent Laboratory Systems*, 47 (1): 1-16.
- ZADEH, L. A. (1965). Fuzzy sets, *Information and control*, 8(3): 338-353.
- ZHANG, Y., Zhao, Z., Yu, J., Wang, K. (2014). Application of rough sets in E-commerce consumer behavior prediction, *Advanced Science and Technology Letters Vol.53 (ICM 2014)*: 255-260.

## Extended Abstract

Under today's growing competition conditions, it is important to determine how consumers spend their time and which factors effect their decisions. To predict consumer behaviour with this goal in mind, a rough set method has been developed and used in this paper.

Although data collection through surveys is commonly used to predict consumer behaviours, data obtained through surveys can not be directly used as information in consumer behaviour analysis. In order to gain such information, methods like regression analysis, machine learning-based models, aimed at detecting the numerical relation among the variables, are usually employed. However, these methods do not allow the cause and effect relation among the variables to be examined.

The aim of this study is to demonstrate that the rule-based LEM2 algorithm, which can easily be interpreted in estimating consumer behaviors by using survey questions, may be an alternative to machine learning algorithms such as artificial neural networks (ANN). Although the algorithms such as the ANN give very high performance, interpretability of the their results is low. Furthermore, it is almost impossible to convert the results to interpretable rules. At the same time, the parameters such as learning rate, the number of nodes in the hidden layer impact the results of algorithm. The other problem is that the input information for the ANN needs to be only numerical values. In the ANN processing, all the categorical variables for modeling the data set should be converted to dummy variables. However, the survey data are categorical and this causes to increases in the number of attributes of the data set. On the other hand, when the survey is applied to a small number of participants due to cost considerations, the small number of samples gives rise to the problem of curse of dimensionality. In the backward induction algorithm which is used to make a minimum error when data is trained by ANN, the objective function should be calculated for each value combination. This is a significant barrier when the number of attributes in the data set is large. Therefore, in modeling of consumer behaviors by using data sets obtained from the survey studies, the methods based on rough set theory are superior than the other methods., Indeed, Cui and Cui (2009) suggested a model based on the rough set approach to forecast the domestic and foreign brand preferences of consumers in their purchasing decisions by using data collected through a large-scale survey. Huang et al. (2010) benefited from rough set theory to evaluate the results of a survey on Taiwanese mobile phone users on the topics of which

features that 4G mobile devices should have. In that study, they used the flowchart for rule-based expression of user behavior. Liou and Tzeng (2010) modelled the Dominance-based rough set approach and modelled the consumer behaviors by extracting a set of rules which was used to determine behavior and loyalty of the airline customers.

Rough Set, theory was first introduced by Pawlak in the early 1980s. In rough set theory, uncertainty is expressed by a restricted area. Fuzzy set theory refers to the uncertainty in the data with the degree of membership, while the rough set theory refers to the indiscernibility relations.

The LEM2 algorithm is a supervised learning algorithm that is used to construct rules from data and it is based on the lower or upper approximation of the rough set theorem (Grzymala-Busse, 2009). The algorithm works at the levels of qualifications and values. A local coverage is calculated for each concept.

The data obtained from the Financial Literacy Survey conducted by the Institute of World Economy of the Romanian Academy was used in this paper. The results of the survey as a data set of 310 attributes (variable) and 2389 observations were published at (<http://microdata.worldbank.org/index.php/catalog/1027/datafiles> url). The study was performed by using the Rough Sets library on the open source R program. Demographic information of participants (gender, age, occupation, graduated school, residence area, nationality, etc.), financial data (monthly income, total income from home, etc.) insurance policy, the daily habits (how often the newspaper is being read, the radio is being listened, the TV program is tracked etc.) were used as variable in the data. The answers of the participants with missing values in the data are deleted so that, 1486 samples and 60 qualifications remained. The class labels “FS used” and “FS not used” were created by using the answers given to the questions (SF1-SF6).

Firstly, the most important attributes were determined by using rough set based reduction method. Accordingly, attributes such as the residence area and the person’s educational status played an active role in predicting the behavior of the respondents. Then the rules of the core set data obtained by the rough set approach determined the rules of consumer behaviors by applying the LEM2 algorithm. In order to show the performance of the rules obtained according to a rough set theorem, the multi-layer perceptron which is one of the most popular algorithms of the ANN was applied and the results were compared by considering the accuracy measurement. The result of the proposed model is both more successful and more comprehensible in predicting consumer behaviors respect to the multi-layer perceptron.

As a result, for the reason of not requiring the additional information about thresholds or experting knowledge of such data, allowing to find minimum data sets by reducing the sample of the original data, the rough set theory based algorithms is a sufficient tool that enables to provide the prediction of the consumers’ decision-making and behavior patterns.