European Journal of Business and Management
ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online)
Vol.6, No.22, 2014

The Use of Rough Set Theory in Determining the Preferences of the Customers of an Insurance Agency

Eylem Koç
Department of Econometrics, Dumlupınar University, Faculty of Economics and Administrative Sciences
Turkey
* E-mail: eyacar@gmail.com

Abstract
In today's market environment a fierce competition is being experienced. It can be clearly stated that the businesses that determine the customer profiles well and manufacture related products in accordance with the requests/needs of the customers gain superiority over their rivals. Within this scope, this fact is also an important issue for the companies that are trying to keep up with other competitors in the insurance sector. In this study, this critical problem of EPD which is an agency of Allianz Insurance was solved by using Rough Set Theory (RST) method. Ten condition attributes (i.e. age, gender, etc.) were examined in the study. Decision attribute is the variable of the insurance type which includes individual retirement, health and life insurances. With the method of RST, a set of rules were identified which may help in developing strategies that will bring in new customers to EPD while keeping present ones. The attained results were presented to the executives of EPD. The executives have re-determined their marketing strategies in compliance with these results and exercised these strategies accordingly. Feedbacks from the executives indicated that the RST helps in facilitating the development of marketing strategies based on the characteristics of the customers and determining their profiles.
Keywords: Rough set theory, customer's profile, insurance, decision rules

1. Introduction
In this day and age, the borders among the countries have disappeared and a global market environment has occurred. Therefore, the businesses tend to manufacture products which meet the requests/needs of their customers completely in order to be the preferred one among a large variety of options. In other words, it can be said that the key factor of gaining superiority in this competitive environment is related to the customer perception. For this reason, companies should determine customer profiles well and target appropriate customer groups. In the insurance branch, product designing concerns a large group of customers because it takes into consideration different kinds of demands (Çildağ 2007). The competition in this sector has increased more in recent years especially together with the application of state incentives to insurance of individual retirement. For this reason, it can be stated that the superiority of insurance companies over their rivals succeeds by means of recognizing the customers appropriately.

It is a very significant problem for the insurance companies to produce information related to the customer profiles and detect the suitable insurance product that will be preferred by the customers. In this study, the problem of determining the insurance product preferences of the customers of EPD was solved by Rough Set Theory (RST). As a result, characteristics which are efficient in the insurance product preferences were identified. By taking advantage of this information, specification of the decision rules is also aimed.

This study consists of four sections. In the first section, brief information is given regarding the study area. In the next section, a review of RST is presented. The third section includes the application by which the characteristics that are efficient on the insurance type preferences of the customers of EPD were determined and with this information the decision rules were acquired by using RST. Last section includes the evaluation of the outcomes and recommendations to the executives of EPD.

2. Rough Set Theory
Rough Set Theory (RST) which was developed by Pawlak in 1982 is a mathematical method based on sets theory (Pawlak 1982). The philosophy of RST is based on the assumption that there is some information which are related to the characteristics of each object although these information have not been revealed. In other words, in contrast with the classical set theory in which the set is defined as single with its elements without another additional data, RST is based on the assumption that at the beginning there is a particular need for some data regarding the elements of the set in order to define it (Ağırgün 2009; Jaaman *et.al.* 2009).

The main objective of RST method is to degrade the data sets, acquire meaningful decision rules by discovering the hidden data patterns and derive new information by means of these rules. In addition, RST is also used for the purposes such as deleting the unnecessary data from the set and predicting the deficient observations (Chen *et al.* 2010; Uyar & Kaya 2011). In the literature; other methods like decision trees, artificial neural networks, particle swarm optimization, bayes/naive bayes, k-nearest neighborhood, event based causing, genetic algorithms and genetic programming are also used for these purposes (For detailed examination, see Jagielska *et al.* 1999;

Stepaniuk & Kierzkowska 2003; Wang *et al*. 2006; Zhai *et al*. 2002). However, RST has significant advantages compared to other methods. One of these advantages is that it only uses the information within data. In RST, there is no need for additional information and assumptions related to data set such as fitness for the probability distribution, function of membership etc. (Ahmady 2010; Aydoğan 2008; Liou & Tzeng 2010; Pawlak *et al*. 1995). The other one is that RST can work with uncertain and inaccurate data and can be used in forming "if-then" rules despite deficient observations (Huang 2013; Kaya *et.al.* 2011; Stokic *et.al.* 2010). For this reason, RST is a method which is efficiently used in data mining, knowledge discovery, pattern recognition, machine learning, expert systems, decision analysis and similar ones both alone and with heuristic methods like genetic algorithms, ant colony, particle swarm optimization, fuzzy set theory, decision trees (Ahn *et al*. 2000; Chen & Zhang 2011; He *et al*. 2008; Wang *et al*. 2006).

The process related to RST method basically consists of six steps (Lazim *et al.* 2012; Rahman *et al.* 2013). These steps can be sequenced as (i) definition of the problem, (ii) attaining data, (iii) forming the information and decision systems, (iv) determining the relations of indiscernibility and set approximations, (v) determining the reduced sets and (vi) forming decision rules. Each step related to this process is explained in the following sub-parts.

2.1 Definition of the problem

The first step of the process is the definition of the problem. In this step, decision problem is explained in all its dimensions in a clear and obvious way. In addition, the attributes to be handled in the problem is listed and defined without deficiency at this stage. Attributes in RST are separated into two as the decision and condition attributes. Decision attribute is the class label of the unit. In other words, the class that the unit takes part in is expressed by decision attribute. For this reason, decision attribute can also be named as class label. The other attributes that indicates the characteristics of the units are called condition attributes (Wu 2009). The purpose in RST is to explain the decision attribute with the help of condition attributes and form decision rules for the purpose of using them in classification of the units.

2.2 Collecting and preparing the data

After defining the problem, the values of the units' condition and decision attributes which are taken into consideration in the problem are collected. At this stage, these values related to the units can be obtained for the first time with the use of various data collection methods while they can also be acquired by compiling previously recorded data. In RST, it is important for all the data to be categorical. In other words, the data should be measured at the level of nominal measurement or they should be reduced to this level. Therefore if there is continuous data in the data set, these should be converted to the discrete form by using various transformation methods (Namdeo & Jayakumar 2014).

2.3 Forming the information and decision systems

At this stage of the process, information and decision systems are formed by making use of collected data. An information system is a table in which units and the condition attributes of these units are formed. The data collected with the use of RST are shown with the help of this table named as information system.

In an information system, units are in the rows and the condition attributes are in the columns (Leung *et al.* 2008) The information system is shown as S=(U,A). Here, U is a non-empty finite set which consists of n units; and it is shown as U=$\{x_1, x_2, \ldots, x_n\}$ (Wan *et al.*, 2008). The set of A is a non-empty finite set including m pieces of condition attributes; and expressed as A= $\{a_1, a_2, \ldots, a_m\}$. a: U → V_a for any attribute of a. Here, V_a is the value set of the attribute of a (Lee & Vatchsevanos 2002).

The decision system is the table formed with the addition of decision attribute to the last column of the information system. Namely, there are attributes related to the units in the columns and units in the rows and also in the decision system as it is in the information system. However, the last column of the table related to the decision system is the column of class label, namely decision attribute. The value (namely the data of decision attribute) of the units depending on the decision attribute takes place in this column. Decision system is shown as D= (U, A∪{d},V). Here, d is the decision attribute and the set of condition attributes is not the element of A (Ayhan & Erdoğmuş 2010).

2.4 Determining the indiscernibility relations and set approximations

2.4.1 Determining the indiscernibility relations

All the data related to the decision problem take place in a decision system. This system which is shown as a table may present a large scale because of the inclusion of some unnecessary and excessive attributes and the repetition of the similar units. This situation negatively affects the ending time of the method, namely the performance. The data should be reduced for the purpose of terminating this negative situation. In RST, indiscernibility relations are used for the purpose of reducing the data (Ayhan 2014). While the information system is S=(U,A), indiscernibility relation is shown as the formula of $IND_A(B)$ for any B⊆A. This formula is expressed as the indiscernibility formula or equivalence relation. (x_1,x_2) being the unit pairs in U, $IND_A(B)$ is equal to $IND_A(B)=\{(x_1,x_2)\in UxU : \forall a\in B,\ a(x_1)=a(x_2)\}$. According to this formula, for the attribute of a, if the values taken by the units of x_1 and x_2 are equal, the units x_1 and x_2 are indiscernible from each other for the attribute of a; and it is said that they belong to the same class of equivalence (Zhao *et al.* 2007).

2.4.2 Determining the set approximations

Determining the inclusion of the units with their present characteristics to classes with the use of indiscernibility relations may not be certain. In this situation, the units that definitely belong to a set and the units which probably belong to this set are determined with the lower and upper approximations. The units definitely belong to X set are determined by taking the values of units with respect to attributes into consideration with lower approximation; and X set is named as lower approximation set. Lower approximation set is shown with the equation $\underline{BX} = \{x | [x]_B \subseteq X\}$ (Yao & Herbert 2007). However, upper approximation provides the determination of the units that are probable to be the element of X set by taking the attribute values of the units into consideration. This set which is named as upper approximation set is expressed as $\overline{BX} = \{x | [x]_B \cap X \neq \quad \}$. From this point, it is certain that each element of BX which is the lower approximation set belongs to X set. It is not certain but probable that each element of $\overline{BX}$ which is the upper approximation set belongs to X set. The set consisting of the units which cannot be decided whether they are the elements of X set or not is named as the border region. In other words, it may not be explained whether the unit gets in a set by using existing data or not. The border region is shown as $BN_B(X)$, and expressed by the equation $BN_B(X) = \overline{BX} - \underline{BX}$. If $BN_B(X)$ is not an empty set, the existence of a rough set is the matter (Telçeken & Doğan 2004).

2.5 Determining the reduced sets

Some attributes in an information system can be extracted from the table without losing the basic characteristics of the table. Assume that an attribute of a belongs to B and $B \subseteq A$. If I(B) = I (B $-\{a\}$), the attribution of a is unnecessary in B set. In a condition in which this equation is not provided, the attribute of a is necessary in B set (Ayhan & Erdoğmuş 2010). The subsets which are formed in this way and do not require unnecessary attributes are called reduced sets (Narlı & Sinan 2011). The intersection of all reduced sets is called as the core attribute set. Core attribute set can be an empty set. The combination of all reduced attribute sets related to an information system S is shown as RED(A). RED(A) includes all sets of A which have been reduced (Aydoğan & Gencer 2007; Tsang *et al.* 2008).

2.6 Forming decision rules

Decision rules are acquired by taking the reduced attribute sets into consideration (Uyar & Kaya 2011). The decision rules formed according to these sets are the propositions dependent on the values of decision attributes against condition attributes. A decision rule can generally be expressed as "If a unit takes the value of a_i against the condition attribute of a, the decision attribute of that unit will take the value of d_i".

Each decision rule has a percentage of accuracy. The percentage of accuracy which is a scale of the accuracy of decision rules is called as the rate of explanation. This rate is shown as $\alpha_B\,(x)$ (Ayhan & Erdoğmuş 2010; Lee & Vatchsevanos 2002). S=(U, A) being an information system and $B \subseteq A$, $X \subseteq U$; the prediction accuracy is attained with the formula $\alpha_B\,(x) = \frac{|\underline{BX}|}{|\overline{BX}|}$ (Erden & Tüysüz 2014).

3. Determining the Insurance Preferences of the Customers with the Use of Rough Set Theory

3.1 Definition of the problem

EPD is an insurance agency serving since 2006. The company desires to recognize the customers better and present them suitable products in order to survive in a highly competitive environment by increasing its market share in the sector. Therefore, the executives of EPD would like to determine the profile of their customers and depending on this, the suitable type of insurance will be presented to the customers. This problem of EPD was solved by using RST. The characteristics of customers which are efficient on the insurance preferences were determined and decision rules depending on these characteristics were formed. For this purpose, firstly, the condition and decision attributes in the problem were identified. In order to determine above mentioned attributes a team was formed which includes a customer representative who is an officer in the company, an insurance expert and the manager of the company. An interactive study at all stages of the solution process was conducted with this team. As a result of the interviews conducted with the team, condition attributes/decision attribute and the levels of these attributes which is given in Table 1 was acquired.

Table 1. Condition attributes /decision attribute handled in the study and their levels

Name of attribute	Levels	Codes related to their levels	Type of attribute
Age	18-25, 26-33, 34-41, 42-49, 50-57, 58-65, 65+	1, 2, 3, 4, 5, 6, 7	Condition attribute
Gender	Male, Female	1, 2	Condition attribute
Marital status	Married, Single	1, 2	Condition attribute
Level of education	Illiterate, Primary school, Secondary school, High school, University, Master's degree/Doctorate	1, 2, 3, 4, 5, 6	Condition attribute
Occupation	Engineer, Teacher, Nurse, Officer, Worker, Doctor, Businessman	1, 2, 3, 4, 5, 6, 7	Condition attribute
Business sector	Private, Public	1, 2	Condition attribute
Level of income	1000 and -, 1001-2000, 2001-3000, 3001-4000, 4001-5000, 5000 and +	1, 2, 3, 4, 5, 6	Condition attribute
Number of children	0, 1, 2, 3 and +	0, 1, 2, 3	Condition attribute
Owning a house	Yes, No	1, 2	Condition attribute
Owning a car	Yes, No	1, 2	Condition attribute
Type of insurance	Individual retirement insurance, Life insurance, Health insurance	1, 2, 3	Decision attribute

As it can be seen in Table 1; it was decided that a total of 10 customer characteristics which are age, gender, marital status, level of education, occupation, business sector, level of income, number of children, owning a house and owning a car was taken as condition attributes. Decision attribute is the type of insurance. In this study, in accordance with the opinions of the company executives, it was decided that three types of insurance; individual retirement, life and health insurance which includes the best-selling products were handled as the levels of decision attribute.

3.2 Implementation

In this study, the data was collected from 500 people who are the customers of EPD by taking the variables listed in Table 1 and their related levels into consideration. After collecting the data, the information and decision systems were formed. As a result, an information system with the dimensions of (500x10) in which each customer takes place in the row and each condition attribute takes place in the column was acquired. The table of decision system is a table with the dimensions of (500x11) which is formed by the addition of decision attribute to the last column of the table of information system. A part of the decision system is shown in Table 2 as sample.

Table 2. A part of the table of decision system formed in this study

Customer no	Condition attributes										Decision attribute
	Age	Gender	Marital status	Level of education	Occupation	Business sector	Level of income	Number of children	Owning a house	Owning a car	Type of insurance
1	25	Male	Single	University	Worker	Private	3000	0	No	No	Individual retirement
2	42	Female	Married	High school	Nurse	Private	2500	2	Yes	Yes	Individual retirement
3	37	Male	Married	University	Engineer	Private	4000	2	Yes	Yes	Life
4	24	Female	Married	University	Nurse	Private	3500	0	No	Yes	Individual retirement
5	33	Female	Married	University	Worker	Private	2500	1	No	Yes	Individual retirement
6	32	Male	Married	University	Officer	Private	2500	1	No	Yes	Individual retirement
7	29	Male	Married	University	Teacher	Public	3000	1	Yes	Yes	Health
8	25	Female	Married	High school	Worker	Public	2100	0	No	Yes	Health
9	34	Female	Married	University	Engineer	Private	3000	1	No	Yes	Individual retirement
10	30	Female	Married	University	Worker	Private	2000	0	No	No	Individual retirement

The analysis step of the RST process consists of determining the lower and upper approximations, obtaining the set of core attributes and attaining decision rules. For these purposes, ROSE2 (Rough Sets Data Explor) is widely used software that applies RST process and generates decision rules (Erden & Tüysüz 2014). Therefore, in this study, ROSE2 software was used for all the analyses and results were given in the next section.

3.3 Results

After forming the information and decision systems, the lower and upper approximations related to each level of decision attribute and the accuracies of prediction was determined. The results are given in Table 3.

Table 3. The lower and upper approximations related to the level of decision attribute and the accuracies of prediction

Classes	The number of units in lower approximation set	The number of units in upper approximation set	Prediction accuracy
1	174	194	0.8969
2	212	230	0.9217
3	94	106	0.8868
The quality of classification	0.9600		

As it can be seen in Table 3, the prediction accuracy is very close to 1,000 for all levels of decision attribute. These values show that the customers will most probably belong to one of the classes of insurance types that they have preferred depending on condition attributes. Also the quality of classification was obtained as 0.96. This value verifies the ability of rough set to deal with the uncertain knowledge of the data used in this study (Lazim *et al.* 2012).

After obtaining the lower and upper approximations, reductions were made by attaining indiscernibility matrix and function. As a result of the conducted reductions, the set of core attributes are acquired and given in Table 4.

Table 4. The set of core attributes

The set of core attributes	Age Gender Marital Status Level of Education Occupation Business Sector Level of Income Number of Children Owning a House Owning a Car Type of Insurance
Accuracy of classification	For all attributes: 0.9600
	For the core set: 0.9600

As it can be seen in Table 4, all condition attributes have taken place in the set of core attributes. According to this, it can be said that the insurance type preferences of customers are affected by all condition attributes which are handled in this study. Therefore decision rules will be formed according to the values of these attributes.
As the last stage, decision rules were obtained by taking the set of core attributes given in Table 4. In the study, a total of 38 decision rules whose rate of explanation changes between 6,67 % and 89,95 were obtained. The first five rules with highest rate of explanation were selected out of these rules and given in Table 5.

Table 5. Decision rules attained in the study

Rule no	Rules	Rate of explanation
1	(Occupation = Teacher) & (Business Sector = Public) & (Owning a House= No) => (Type of Insurance = Life insurance)	0.8995
2	(Gender = Female) & (Occupation = Teacher) & (Business Sector = Public) => (Type of Insurance = Life insurance)	0.7717
3	(Age = 26-33) & (Marital Status = Married) & (Business Sector = Private) & (Number of Children = 1) => (Type of Insurance = Individual retirement insurance)	0.4153
4	(Level of Education = University) & (Business Sector = Public) & (Number of Children = 2) => (Type of Insurance = Life insurance)	0.3836
5	(Level of Education = High School) & (Number of Children = 0) => (Type of Insurance = Health insurance)	0.2551

As it can be seen in Table 5, the decision rule with the highest rate of explanation (89,95 %) is Rule 1. Namely, teachers, who work in public sector and do not own a house, prefer the product of “Life Insurance” with the probability of 89.95 %. The other rules are interpreted similar to Rule 1. According to Rule 2, female teacher customers and working in public sector purchase the product of “Life Insurance” with the probability of 77.17 %. The result which is obtained from Rule 3 is that the customers who are between the age group 26-33, married, working in private sector and have one child opt for “Insurance of Individual Retirement” with the probability of 41.53 %. According to Rule 4, it can be seen that the customers with the educational level of university, working in public sector and have two children prefer the product of “Life Insurance” with the probability of 38.36 %. When Rule 5 which is the last rule given in the study is examined, it is seen that the customers who have graduated from high school and do not have any children prefer “Health Insurance” with the probability of 25.51 %.

4. Discussion and Conclusions

In this study, the problem of determining the insurance preferences of the customers of EPD was handled by taking the profile of customers into consideration. Producing data related to the profiles of the customers and determining the insurance product which will have the highest purchasing potential is a complex problem. In this study, this problem was solved by using Rough Set Theory (RST). With RST, the customer characteristics which are efficient on insurance type preferences of customers were determined and decision rules were formed by taking advantage of these data.
In the study, a total of ten condition attributes being age, gender, marital status, level of education, occupation, business sector, level of monthly income, number of children and owning a house and car were taken. Decision

attribute is the variable of insurance type that includes individual retirement, health and life insurance. As a result of the analysis, it was determined that all condition attributes that were handled have a significant role in determining the insurance type preferred by the customers.

When the decision rules related to the insurance type preferences of customers are examined, the customers who are teachers, working in public sector and do not have a house prefer the product of life insurance. The rate of explanation of this decision rule is 89.95 %. Namely, the possibility of preferring life insurance is 89.95 % for the customers with these characteristics. Moreover, it was determined that the customers who are females, teachers and working in public sector will prefer the product of life insurance with the probability of 77.17 %; the customers who are between the age group 26-33, married, working in private sector and have one child will prefer the product of individual retirement with the probability of 41.53 %; the customers who are the university graduates, working in public sector and have two children will prefer the product of life insurance with the probability of 38.36 %; and the customers who are graduates of high school and have no children will prefer the health insurance with the probability of 25.51 %.

Consequently, in the study, an efficient way of solution was developed for the executives who are aware of the fact that the key factor in increasing market share is hidden in the “recognition of the customer”. Detailed suggestions were presented to EPD executives who face this complex decision problem. The executives revised their marketing and target group strategies for each product in accordance with the decision rules obtained from this study. The strategies that were determined by the rules were exercised effectively in marketing phase by EPD. This study will shed light on both researchers who conduct scientific studies related to this topic and the executives facing similar kinds of administrative problems.

References

Ağırgün, B. (2009). *Bulanık Kaba Küme Yöntemi ile Nitelik İndirgemede Yeni Bir Algoritma*. Gazi University, Institute of Natural and Applied Sciences, PhD. Thesis, 149p.

Ahmady, A. (2010). *Rough Set Kansei Engineering: Multiple Users, Multiple Kanseis*. Department of Industrial and Manufacturing Engineering and the Faculty of Graduate School of Wichita State University, PhD. Thesis, 248p.

Ahn, B. S., Cho, S. S., & Kim, C. Y. (2000). The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Expert Systems with Applications*, 18, 65–74.

Aydoğan, E., & Gencer, C. (2007). *Kaba Küme Yaklaşımı Kullanılarak Veri Madenciliği Problemlerinde Sınıflandırma Amaçlı Yapılmış Olan Çalışmalar*. [Online] Available: http://w3.gazi.edu.tr/~ctemel/aydogan&gencer_2007.pdf (10.06.2014).

Aydoğan, E. K. (2008). *Veri Madenciliğinde Sınıflandırma Problemleri için Evrimsel Algoritma Tabanlı Yeni Bir Yaklaşım- Rough-Mep Algoritması*. Gazi University, Institute of Natural and Applied Sciences, PhD. Thesis, 149p.

Ayhan, S., & Erdoğmuş, Ş. (2010). Web Sayfası Tasarımında Kaba Küme Teorisi Tabanlı Kansei Mühendisliği Yaklaşımı. *Endüstri Mühendisliği Dergisi*, 23, 1, 16-27.

Ayhan, S. (2014). *Kaba küme ve destek vektör makineleri kullanılarak nitelik indirgeme ve sınıflandırma problemlerinin çözümü için bütünleşik bir yaklaşım*. Eskişehir Osmangazi University, Institute of Natural and Applied Sciences, PhD. Thesis, 199p.

Chen, Y., Miao, D. & Wang, R. (2010). A rough set approach to feature selection based on ant colony optimization. *Pattern Recognition Letters*, 31, 226-233.

Chen, J., & Zhang, C. (2011). Efficient Clustering Method Based on Rough Set and Genetic Algorithm. *Procedia Engineering*, 15, 1498 – 1503.

Çildağ, G. (2007). *Müşteri İlişkileri Yönetimi ve Sigortacılık Sektöründe Bir Uygulama*. Adnan Menderes University, Institute of Social Sciences, Master Thesis, 115p.

Erden, C., & Tüysüz, F. (2014). *An Application of Rough Sets Theory on Traffic Accidents*. OPT-i, An International Conference on Engineering and Applied Sciences Optimization, M. Papadrakakis, M.G. Karlaftis, N.D. Lagaros (eds.), Kos Island, Greece.

He, Y., Chen, D., & Zhao, W. (2008). Integrated method of compromise-based ant colony algorithm and rough set theory and its application in toxicity mechanism classification. *Chemometrics and Intelligent Laboratory Systems*, 92, 22–32.

Huang, C.C., Tseng, T.L., Fan, Y.N., & Hsu, C.H. (2013). Alternative rule induction methods based on incremental object using rough set theory. *Applied Soft Computing*, 13, 372-389.

Jaaman, S.H., Shamsuddin, S.M., Yusob, B., & Ismail, M. (2009). A Predictive Model Construction Applying Rough Set Methodology for Malaysian Stock Market Returns. *International Research Journal of Finance and Economics*, 30, 211-218.

Jagielska, I., Matthews, C., & Whitfort, T. (1999). An investigation into the application of neural networks, fuzzy logic, genetic algorithms, and rough sets to automated knowledge acquisition for classification problems.

Neurocomputing, 24, 37-54.
Kaya, Y., Yeşilova, A. & Tekin, R. (2011). *İstenmeyen Elektronik Postaların(Spam) Filtrelenmesinde Kaba Küme Yaklaşımının Kullanılması*. Electric-Electronic and Computer Symposium, Fırat University, 148-153.
Lazim, Y. M., Rahman, M. N. A., & Mohamed, F. (2012). *Clustering Model of Multimedia Data By Using Rough Sets Theory*. International Conference on Computer & Information Science (ICCIS), 336-340.
Lee, S., & Vatchsevanos, G. (2002). An application of rough set theory to defect detection of automotive glass. *Mathematics and Computers in Simulation*, 60, 225–231.
Leung, Y., Fischer, M.M., Wu, W.Z., Mi, J.S. (2008). A rough set approach for the discovery of classification rules in interval-valued information systems. *International Journal of Approximate Reasoning*, 47, 233-246.
Liou, J. & Tzeng, G. (2010). A Dominance-based Rough Set Approach to customer behavior in the airline market. *Information Sciences*, 180, 2230-2238.
Namdeo, J., & Jayakumar, N. (2014). Predicting Students Performance Using Data Mining Technique with Rough Set Theory Concepts, *International Journal of Advance Research in Computer Science and Management Studies*, 2, 2, 367-373.
Narlı, S., & Sinan, O. (2011). Biyoteknoloji Tutum Ölçeğinin Değerlendirilmesinde Matematiksel Bir Yaklaşım: Kaba Küme Veri Analizi. *Kuram ve Uygulamada Eğitim Bilimleri, Educational Sciences: Theory & Practice*, 11, 2, 709-726.
Pawlak, Z. (1982). Rough sets. *International Journal of Computer and Information Sciences*, 11, 5, 341-356.
Pawlak, Z., Grzymala-Busse, J., Slowinski, R., & Ziarko, W. (1995). Rough Sets. *Communications of ACM-Emerging Technologies- AI*, 38, 11, 89-95.
Rahman, M. N. A., Lazim, Y.M., Mohamed, F., & Saany, S.I.A. (2013). *Implementation of Rough Sets Theory Based Multimedia Data Clustering under Web Services*. The 2nd International Conference on Information Science and Technology, 23, 256-264.
Stepaniuk, J., & Kierzkowska, K. (2003*).* Hybrid Classifier Based on Rough Sets and Neural Networks. *Electronic Notes in Theoretical Computer Science*, 82, 4, 228–238.
Stokic, E., Brtka, V., & Srdic, B. (2010). The synthesis of the rough set model for the better applicability of sagittal abdominal diameter in identifying high risk patients. *Computers in Biology and Medicine*, 40, 786-790.
Telçeken, S., & Doğan, M. (2004). Kaba Kümeler Teorisi Yardımı İle Büyük Veri Topluluklarının Analizi. *Eleco2004*, 1, 410-414.
Tsang, E.C.C., Yang, W.X., & Chen, D.G. (2008). *The Relationships Between The Inclusion Degree And Measures On Rough Set Data Analysis Based On Regular Probability Spaces*. Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, 2412-2416.
Uyar, M. & Kaya, Y. (2011). *Kaba Küme Yaklaşımıyla Güç Kalitesindeki Bozulma Türlerinin Sınıflandırılması*. 6th International Advanced Technologies Symposium (IATS'11), Elazığ, Turkey, 225-229.
Wan, S., Lei, T.C., Huang, P.C., & Chou, T.Y. (2008). The knowledge rules of debris flow event: A case study for investigation Chen Yu Lan River, Taiwan. *Engineering Geology*, 98, 102–114.
Wang, X., Yang, J., Jensen, R., & Liu, X. (2006). Rough set feature selection and rule induction for prediction of malignancy degree in brain glioma. *Computer Methods and Programs in Biomedicine*, 83, 147-156.
Wu, W.W. (2009). Exploring core competencies for R&D technical professionals. *Expert Systems with Applications*, 36, 9574-9579.
Yao, J.T., & Herbert, J.P. (2007). Web-Based Support Systems with Rough Set Analysis. Rough Sets and Intelligent Systems Paradigms, *Lecture Notes in Computer Science*, 4585, 360-370.
Zhao, Y., Yao, Y., & Luo, F. (2007). Data analysis based on discernibility and indiscernibility. *Information Sciences*, 177, 4959–4976.
Zhai, L.Y., Khoo, L.P., & Fok, S.C. (2002). Feature extraction using rough set theory and genetic algorithms—an application for the simplification of product quality evaluation. *Computers & Industrial Engineering*, 43, 661–676.

Acknowledgments: The author of this study would like to thank Mr. Emre Kaygısız, the Manager of EPD which is an agency of Allianz Insurance for his sincere contributions and supports.