Application of Rough Set Theory in Data Mining for Decision Support Systems (DSSs)

Mohammad Hossein Fazel Zarandi
Department of Industrial Engineering, Amir kabir university of Technology, Tehran, Iran,
Zarandi@Aut.ac.ir

Abolfazl Kazemi
Department of Industrial Engineering, Amir kabir university of Technology, Tehran, Iran,
Corresponding author, Aazemi@Aut.ac.ir

Abstract
Decision support systems (DSSs) are prevalent information systems for decision making in many competitive business environments. In a DSS, decision making process is intimately related to some factors which determine the quality of information systems and their related products. Traditional approaches to data analysis usually cannot be implemented in sophisticated Companies, where managers need some DSS tools for rapid decision making. In traditional approaches to decision making, usually scientific expertise together with statistical techniques are needed to support the managers. However, these approaches are not able to handle the huge amount of real data, and the processes are usually very slow. Recently, several innovative facilities have been presented for decision making process in enterprises. Presenting new techniques for development of huge databases, together with some heuristic models have enhanced the capabilities of DSSs to support managers in all levels of organizations. Today, data mining and knowledge discovery is considered as the main module of development of advanced DSSs. In this research, we use rough set theory for data mining for decision making process in a DSS. The proposed approach concentrates on individual objects rather than population of the objects. Finally, a rule extracted from a data set and the corresponding features (attributes) is considered in modeling data mining.

Keyword: Data Mining, Knowledge Discovery, Rough Set Theory.

1. Introduction

In modern decision support systems, data mining is the most prevalent tool used to search for and extract useful information from volumes of data and for combining databases, artificial intelligence, and machine learning. In other words, data mining is a powerful tool for nontrivial extraction of implicit, previously unknown, and potentially useful information from large data sets. Obviously, a fairly large classes of data mining tasks can be described as search for interesting and frequently occurring patterns or rules from databases. This technology uses machine learning, statistical and visualization techniques to discover and present knowledge in a form which is easily comprehensible to humans [1]. Data mining is used in a process called Knowledge discovery in databases (KDD). The discovered knowledge can be rules describing properties of data or relationships among data, frequently occurring patterns, clustering of the objects in the data set, etc. Most data mining systems in use have been designed using variants of traditional machine learning techniques [2].
Rough set theory is a powerful tool for data mining. This approach can be implemented for: a) Reduction of data sets; b) Finding hidden data patterns; and c) Generation of decision rules [3].
Rough set theory is a relatively new mathematical and artificial intelligence (AI) technique introduced in the early 1980's by Pawlak [13]. The technique is particularly suited to reasoning about imprecise or incomplete data, and discovering relationships among them. The main advantage of rough set theory is that it does not require any preliminary or additional information about data-like probability in statistics, or the value of possibility in fuzzy set theory. Recently, there has been a growing interest in rough set theory among researchers in modern intelligent information systems. The theory has found many real life applications in many areas. The primary applications of rough sets so far have been in data and decision analysis, databases, knowledge-based systems, and machine learning [4]. The paper expands the Kusiak's work [3] in this area. The concentration of this paper is on the application of rough set theory in data mining.
The rest of the paper is organized as follows: In the next section, data mining and knowledge discovery are reviewed. In section 3, the main aspects of rough set theory are explained. Section 4 uses rough set theory for rule extraction. Section five presents rule-structuring algorithms. Finally, in section six, conclusion and potential future works are presented.

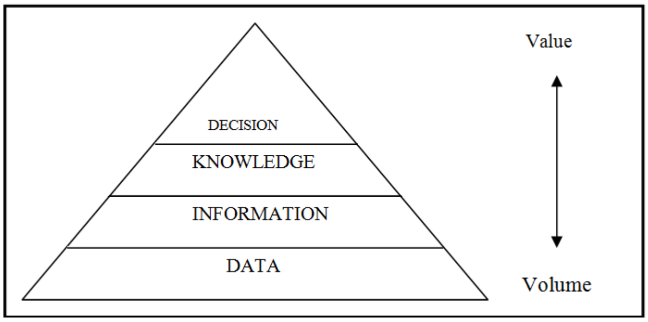
2. Data Mining

Data mining is the process of posing various queries and extracting useful information, patterns, and trends often previously unknown from large quantities of data possibly stored in a database [5]. Essentially, for many organizations, the goals of data mining include improving marketing capabilities, detecting abnormal patterns, and predicting the future based on past experiences and current trends. By increasing the size of a database, its supporting decision-making becomes more difficult. Moreover, the data might be from multiple sources and

multiple domains. Thus, the integrity of a data also should be considered in a database approach [5,6,7,8,9].
Some of the data mining techniques include those based on rough sets, inductive logic programming, machine learning, and neural networks, among others. The main tasks of data mining are classification (finding rules to partition data into groups), association (finding rule to make associations between data), and sequencing (finding rules to order data) [5]. With the increased use of computers, there is an ever increasing volume of data being generated and stored. The sheer volume held in corporate databases is already too large for manual analysis and, as they grow, the problem is compounded. Furthermore, in many companies data is held only as a record or archive. Data mining encompasses a range of techniques, which aim to extract value from volume and form the foundation of decision making. (see Figure 1) [10].

2.1. The Knowledge Discovery Process

The process of knowledge discovery via data mining can generally be divided into four basic activities: *Selection*, *Pre-Processing*, *Data mining*, and *Interpretation.*
Selection involves creating a target data set, i.e., the data set to about undergo analysis. It is a common misunderstanding to assume that the complete data set should be submitted to data mining software. This is not the case, partly due to the possibility that the data represents a number of different aspects of the domain, which are not directly related. Careful thought should therefore be given as to the purpose of the data mining exercise and a target data set created which reflects this propose.



- *Pre-Processing* involves preparing the data set for analysis by the data mining software to be used. This may involve activities such as resolving undesirable data characteristics such as missing data (non-complete fields), irrelevant fields, non-variant fields, skewed fields, and outlying data points. The pre-processing activities may result in the generation of a number of (potentially overlapping) subsets of the original target data set. Pre-Processing may also involve converting the data into a format acceptable to the data mining software being used. This initial collection and manipulation of data (during the selection and per-processing stages) in data mining process is sometimes referred to as collection and cleaning.
- *Data Mining* involves subjecting the cleaned data to be analyzed by the data mining software in an attempt to identify hidden trends or patterns, or to test specific hypotheses. It is recommended that any (apparently) significant obtained results be validated by using traditional statistical techniques.
- *Interpretation* involves analysis and interpretation of the generated results. This may return to previous stages to carry out additional activities in order to provide further necessary information.

Figure 2 shows an overview of the activities involved in Knowledge Discovery in Databases (KDD) [11].

The emerging of data mining and knowledge discovery in databases (KDD) is due to the fast development of information technologies. While database technology was successful for recording and managing data, data mining and KDD is aimed at developing methodologies and tools to automate the data analysis process and create useful information and knowledge from data to help decision-making [12]. Figure 3 shows the typical relative effort of a data mining project.

2.2. Steps of Data Mining

In general, the steps of data mining involves getting the data organized for mining, determining the desired outcomes to mining, selecting tools for mining, carrying out the mining, pruning the results so that only the useful ones are considered further, taking actions from the mining, and evaluating the actions to determine benefits [12]

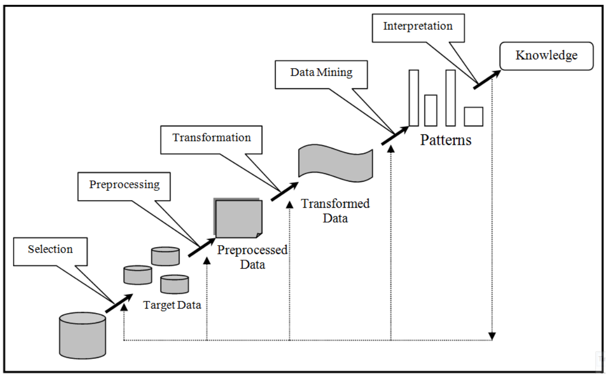


Figure 2. An overview of the steps comprising KDD process

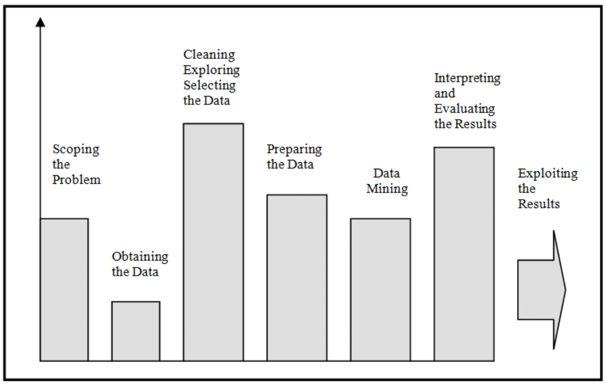


Figure 3. Relative effort of a data mining project

More specifically the steps of data mining are as follows:

• Identifying the data
• Massaging the data and Getting it ready
• Mining the data
• Getting useful results
• Identifying actions
• Implementing the action
• Evaluating the benefits
• Determining what to do next
• Carrying out the next cycle

2.3. Data mining technologies and techniques

As shown in Figure 4, data mining is an integration of multiple technologies. These include data management such as database management, data warehousing, statistics, machine learning, decision support, visualization, and parallel computing [5]. Data mining methods and tools can be categorized in different ways [12]. In application, data mining methods can be classified as clustering, classification, summarization, dependency modeling, link analysis and sequence analysis. Some methods are traditional and established and some are relatively new.

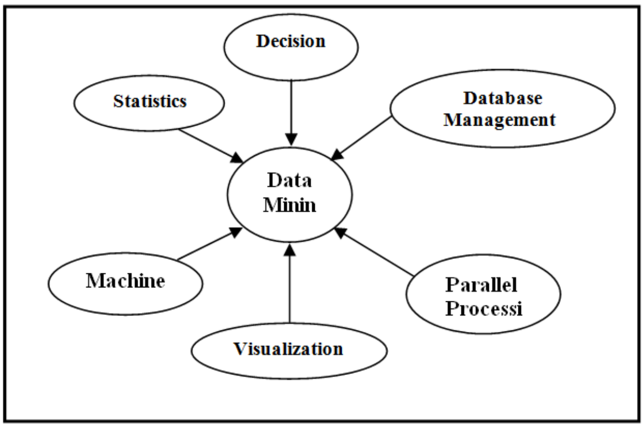


Figure 4. Data mining technologies

2.4. Method selection

Apart from general considerations such as cost and support, there are some technical dimensions to the data mining method selection. These include [12]:
• Uni-variate vs. multi-variate data. Most approaches assume independence of variable or simply consider a single variable at a time.
• Numerical vs. categorical or mixed data. Some methods are only suitable for numerical data, others only for categorical data. There are only a few cases, which allow mixed data.
• Explanation requirements or comprehensibility. Some tools give results, which are implicit to users (black box), while others can give causal and explicit representations.
• Fuzzy or precise patterns. There are methods such as decision trees, which only work with clear- cut definitions.
• Sample independence assumptions. Most methods assume independence of data patterns. If there are dependencies on the data patterns, it is necessary to remove or explore
• Availability of prior knowledge. Some tools require prior knowledge, which might be not available. On the other hand, some others do not allow input of prior knowledge causing a waste of prior knowledge.
It is important to be aware of the complexity of data, which tends to contain noise and erroneous components and has missing values.
Other challenges come from lack of understanding of the domain problem and assumptions associated with individual techniques. Therefore, data mining is rarely done in one step. It often requires to implement a number of approaches to use some tools to prepare data for other methods, or for validating purposes. As a result, multifunctional and integrated systems are required.

Table1
An attribute-value table of data

Store	E	Q	L	P
1	High	Good	No	Profit
2	Med.	Good	No	Loss
3	Med.	Good	No	Profit
4	Med.	Avg.	No	Loss
5	Med.	Avg.	Yes	Loss
6	High	Avg.	Yes	Profit
7	High	Avg.	No	Loss
8	Low	Good	No	Profit
9	Low	Avg.	No	Loss
10	Low	Avg.	Yes	Loss
11	Low	Good	Yes	Loss
12	High	Good	Yes	Profit

3. Rough set Theory

Rough set theory, proposed by Pawlak [13], is a new mathematical approach to data analysis. The basic idea behind this method is on classification of objects of interest into similar classes (clusters) containing objects which are indiscernible blocks of knowledge about reality and to find hidden patterns in data [14]. Of course, this theory has some overlaps with other methods of data analysis, e.g., statistics, cluster analysis, fuzzy sets, evidence theory,etc. However, it can be viewed its own right as an independent discipline. The rough set approach seems to be fundamental to AI and cognitive sciences, especially in the areas of machine learning, knowledge acquisition, decision analysis, knowledge discovery from data bases, expert systems, inductive reasoning, pattern recognition, decision support system, and data mining [3,14,15].
The rough set approach has many advantages. Among them, the following capabilities are more important:
• Providing efficient algorithms for finding hidden pattern in data.
• Identifying relationships that would not be found while using statistical methods.
• Allowing both qualitative and quantitative data.
• Finding minimal sets of data (data reduction).
• Evaluating significance of data.
• Generating sets of decision rules from data.

- Easy to understand.
- Offering straightforward interpretation of obtained results.

3-1 Approximation in rough set theory

Data are usually given in tables, called also attribute-value table, information table, or database. A database is a table, rows of which are labeled by objects, whereas columns are labeled by attributes. Entries of a table are attribute values. An example of a linguistic database is shown in Table 1. In this database, 12 store are characterized by four attributes:

E: empowerment of sales personnel.
Q: perceived quality of merchandises.
L: high traffic location.
P: store profit or loss.

Each subset of attributes in the database determines a partition of all objects into clusters having the same attribute values, i.e., displaying the same features expressed in terms of attribute values. In other words, all objects revealing the same features are indiscernible (similar) in view of the available information and they forms blocks, which can be understood as elementary granules of knowledge. These granules are called elementary sets or concepts, and can be considered as elementary building blocks (atoms) of our knowledge about the reality we are interested in. Elementary concepts can be combined into compound concept, i.e., concepts that are uniquely defined in terms of elementary concepts.

Any union of elementary sets is called a crisp set, and any other set is referred to as rough (vague, imprecise).

With a set X, we can associate two crisp sets, called the lower and the upper approximation of X. The lower approximation of X is the union of all elementary sets, which are included in X, whereas the upper approximation of X is the union of all elementary sets, which have non-empty intersection with X. In other words, the lower approximation of set is of elements that surely belong to X, whereas the upper approximation of X is the set of all elements that possibly belong to X. The difference between the upper and the lower approximation of X is the boundary region. Obviously, a set is rough if it has non-empty boundary regions; otherwise the set is crisp. Elements of boundary region cannot be classified employing the available knowledge, either to the set or its complement. Approximations of sets are basic operations in rough set theory and are used as main tools to deal with vague and uncertain data. Let us illustrate the above ideas by means of data given in Table 1. Each store has a different description in terms of attributes E, Q, L and P; thus all stores are discernible when employing information provided by all attributes. However, stores 2 and 3 are indiscernible in terms of attributes E, Q and L since they have same attribute values. Similarly, stores 1, 2 and 3 are indiscernible with respect to attributes Q and L, etc.

Each subset of attributes determines a partition (classification of all objects into classes having the same description in terms of these attributes). For example, attributes Q and L aggregate all stores into the following classes: {1,2,3}, {4} and {5.6}. Thus, each database determines a family of classification patterns, which are used for further consideration. Let us consider the following problem: what are the characteristic features of stores making a profit (or having a loss) in view of information available in Table 1, i.e., we want to describe a set (concept) {1,3,6}, (or {2,4,5}, in terms of attributes E, Q and L. Of course, this question cannot be answered uniquely since stores 2 and 3 have the same values of attributes E, Q and L, but store 2 makes a profit, whereas store 3 has a loss. Hence, in view of information contained in Table 1, we can say for sure that stores 1 and 6 make a profit, and stores 4 and 5 have a loss, whereas stores 2 and 3 cannot be classified as making a profit or having a loss. That is, employing attributes E, Q and L, we can say that stores 1 and 6 surely make a profit, i.e., surely belong to the set {1,3,6}, whereas stores 1,2,3 and 6 possibly make profit, i. e., possibly belong to the set {1,3,6}. We say that the set {1,6} is the lower approximation of the set (concept) {1,3,6}, and the set {1,2,3,6} is the upper approximation of the set {1,3,6}. The set {2,3}, being the difference between the upper approximation and the lower approximation is referred to as boundary region of the set {1,3, 6}.

Now, let us give some formal notations and definitions. By a database, we understand a pair S=(U, A), where U and A are finite, non-empty sets, called the universe, and a set of attributes, respectively. With every attribute $a \in$ A, we associate a set V_a of its values to domain of a. Any subset B of A determines a binary relation I(B) on U, which is called an indiscernibly relation, defined as follows:

$$(x,y) \in I(B) \text{ if and only if } a(x)=a(y) \tag{1}$$

for every $a \in A$, where a(x) denotes the value of an attribute of element x. It can easily be seen that I(B) is an equivalence relation. The family of all equivalence classes of I(B), i.e., the partition determined by B, is denoted by U/I(B), or simply U/B; an equivalence class of I(B), i.e., the block of the partition U/B containing x is denoted by B(x). If (x,y) belongs to I(B), we say that x and y are B-indiscernible. Equivalence classes of the relation I(B) (or blocks of the partition U/B) are referred to as B-elementary sets or B-granules. Next, the indiscernibly relation is used to define two basic operations in rough set theory as follows:

$$B_*(X) = \underset{X \in U}{Y}\{B(X) : B(X) \subseteq X\} \tag{2}$$

$$B^*(X) = \underset{X \in U}{Y}\{B(X) : B(X) \cap X \neq \phi\} \tag{3}$$

These are called the B-lower and the B-upper approximation of X, respectively. The set $BN_B(x) = B^*(x) - B_*(x)$ is referred to as the B-boundary region of X. If the boundary region of X is the empty set, i.e., $BN_B(x) = \phi$, then, the set X is crisp (exact) with respect to B, i.e., if $BN_B(x) = \phi$, the set is referred to as rough (inexact) with respect to B.

3-2-Dependency attributes

Approximations of sets are strictly related to the concept of dependency (total or partial) of attributes. Suppose the set of attributes A is partitioned into two disjoint subsets C and D called condition and decision attributes, respectively. Databases with distinguished condition and decision attributes are referred to as decision tables. Intuitively, a set of decision attributes D depends totally on a set of condition attributes. In other words, D depends totally on C, if there exists a functional dependency between values of C and D.

We also need a more general concept of dependency of attributes, called the partial dependency of attributes. Partial dependency means that only some values of D are determined by values of C. Formally, dependency can be defined in the following way:

Let C and D be subsets of A, such that $D \cap C \neq \phi$ and $D \cup C = A$. We say that D depends on C in a degree k ($0 \leq k \leq 1$), denoted $C \underset{k}{\rightarrow} D$, if

$$k = \gamma(C, D) = \sum_{X \in U/D} \frac{card(C, (X))}{card(U)} \tag{4}$$

where $\gamma(C, D)$ is an accuracy coefficient or accuracy of approximation of subsets C and D, and card(x) is the cardinality of X. If k =l we say that D depends totally on C, and if k<l, we say that D depends partially (in a degree k) on C. The coefficient k expresses the ratio of all elements of the universe, which can be properly classified to the block of the partition U/D employing attributes C and is called the degree of the dependency. For example, the attribute P depends on the set of attributes {E, Q, L} in the degree 2/3. It means that only four of six stores can be identified exactly by means of attributes E, Q and L as having a loss or making a profit.

3-3- Reduction of attributes

We often face the question of whether we can remove some data from a database preserving its basic properties, i.e., whether a table contains some surplus data. Let us express this idea more precisely:

Let C, $D \subseteq A$, be sets of condition and decision attributes, respectively. We say that $C' \subseteq C$ is a D-reduct (reduct with respects to D) of C, if C' is a minimal subset of C such that $\gamma(C, D) = \gamma(C', D)$. Hence, any reduct enables us to reduce condition attributes in such a way that the degree of dependency between condition and decision attributes is preserved. In other words, reduction of condition attributes gives the minimal number of conditions necessary to make specified decisions. In the database presented in Table l, {E,Q} and {E,L} are the only two reducts of condition attributes with respect to P, i.e., either the set {E,Q} or the set {E,L} can be used to classify stores instead of the whole set of condition attributes{E,Q,L}.

For large databases finding reducts on the basis of the definition given above is rather difficult because the definition leads to inefficient algorithm. Therefore, more efficient methods of reduct computation have been proposed.

3-4- Decision rules

Every dependency $C \underset{k}{\rightarrow} D$ can be described by a set of decision rules in the form "IF…THEN…", written $\Phi \rightarrow \Psi$, where Φ and Ψ are logical formulas describing conditions and decisions of the rule respectively, and are built up from elementary formulas (Attribute, value) combined together by means of prepositional connectives "AND", "OR" and "NOT" in the standard way. An example of a decision rule is given:

IF (E,Med.) AND (Q,Good) AND (L,No)
THEN (P,Loss) (5)

With every decision rule $\Phi \rightarrow \Psi$, we associate a conditional probability that Ψ is true in S, given Φ is true in S with the probability $\pi_s(\Phi)$ called a certainty factor and is defined as follows:

$$\pi_s(\Psi|\Phi) = \frac{card(|\phi \wedge \psi|_S)}{card(|\phi|_S)} \tag{6}$$

where $|\Phi|_S$ denotes the set of all objects satisfying Φ in S. As well, need a coverage factor

$$\pi_s(\Phi|\Psi) = \frac{card(|\phi \wedge \psi|_S)}{card(|\psi|_S)} \tag{7}$$

which is the conditional probability that Φ is true in S, given Ψ is true in S with the probability $\pi_S(\Psi)$. For the decision rule given above the certainty and coverage factors are ½ and 1/3, respectively. In other words, the probability that the correct decision be made by the decision rule is ½ and the rule covers one of the three decisions indicated by it.

Let $\{\Phi_i \rightarrow \Psi\}$ be a set of decision rules such that all conditions Φ_i are pairwise mutually exclusive, i.e., $\{\Phi_i \wedge \Phi_j\}_S = \Phi$, for any $1 \leq i, j \leq n, i \neq j$ and $\sum_{i=1}^{n} \pi_S(\Phi_i|\Psi) = 1$.

For any decision rule $\Phi \rightarrow \Psi$ the following formula is true:

$$\pi_S(\Phi|\Psi) = \frac{\pi_S(\psi|\phi).\pi_S(\phi)}{\sum_{i=1}^{n} \pi_S(\psi|\phi_i).\pi_S(\phi_i)} \tag{8}$$

The relationship between the certainty factor and the coverage factor, expressed by formula (8) is the Bayes theorem. The formula shows that any decision table satisfies this theorem. This property gives a new dimension to Baysian reasoning methods and enables us to discover relationships in data without referring to prior and posterior probabilities inherently associated with

Baysian philosophy. The above result is of special value for large databases.

4- Rule extraction

The content of large-scale data sets containing numerical and categorical information can not be easily interpreted unless the information is transformed into a form that can be understood by human users. The rule extraction algorithms are designed to identify patterns in such data sets and express them as decision rules. The rule extraction concept is illustrated in case 1.

4-1- Rule- structuring algorithm

The rule-structuring algorithm groups entries of the rule-attribute matrix. The attribute corresponding to the entries are called marked attributes. The steps of the rule-structuring algorithm are outlined below [13]:

Step1) Select from the rule- attribute matrix a marked attribute with the maximum equivalence classes. Break a tie arbitrarily.

Step2) Cluster the rules within each equivalence class.

Step3) Re-engineer the marked entries of the rule- attribute matrix according to the principles:

1. Merge clusters: cluster rules with equivalent attribute values.
2. Attribute value replacement: replace attribute values with a value range and merge the corresponding rule with the same decisions.
3. Column removal: remove entries that are not used by the rules.

Step 4) Stop, if a satisfactory matrix structure has been obtained

Case 1

Consider the data set in Table 2 with ten objects, four attributes (features) A1, A2, A3, A4, and the decision (outcome) D. The attributes denote process parameters (e.g., temperature, pressure) and the decision is the component performance D={high, medium, low}. A rule extraction algorithm transforms the data set of Table 2 into the decision rules of Figure 5. The two sets of numbers in square brackets behind each rule describe its properties and are: [Rule support, Rule coverage, Discrimination level] and [list of supporting objects].

Rule support is the number of all objects in the data set that have the property described by the condition of the rule. Rule coverage is the proportion of the objects in the training set that are identifiable by this rule. Discrimination level measures the level of precision which a rule represent the corresponding objects.

Table 3 indicates that the number of attributes used to describe all objects in the data set are A1, A2, and A4. The algorithm used to generate rules in Figure 5 minimized the number of decision rules. The rule extraction algorithms may consider other criteria, including the generation of all possible rules. An algorithm minimizes the number of attributes included in decision rules.

Table 2
Data set

NO.	Attributes				D
	A1	A2	A3	A4	
1	0	0	1	2	Low
2	2	1	0	1	High
3	1	0	1	0	Low
4	0	3	3	0	Medium
5	3	1	2	1	High
6	0	2	2	3	Medium
7	3	1	1	3	High
8	1	0	0	2	Low
9	0	2	0	1	Medium
10	1	3	1	0	Low

Rule 1: *IF* (A1=3) *THEN* (D= High); [2,66.67%, 100%] [5,7]

Rule 2: *IF* (A1=1) *THEN* (D= Low); [3,75%, 100%] [3,8,10]

Rule 3: *IF* (A3=3) *THEN* (D= Medium); [1,33.33%, 100%] [4]

Rule 4: *IF* (A1=2) *AND* (A2=1) *THEN* (D =High); [1,33.33%, 100%] [2]

Rule 5: *IF* (A2=2) *THEN* (D =Medium); [2,66.67%, 100%] [6,9]

Rule 6: *IF* (A3=1) *AND* (A4=2) *THEN* (D= Low); [1,25%, 100%] [1]

Figure 5. Decision rules extracted with a rough set algorithm

Table 3
Patterns corresponding to the rule of Figure 5

NO.	**Attributes**				**D**	
	A1	**A2**	**A3**	**A4**		
5	3	1	2	1	High	*Rule* 1
7	3	1	1	3	High	*Rule* 1
3	1	0	1	0	Low	*Rule* 2
8	1	0	0	2	Low	*Rule* 2
10	1	3	1	0	Low	*Rule 2*
4	0	3	3	0	Medium	*Rule 3*
2	2	1	0	1	High	*Rule 4*
6	0	2	2	3	Medium	*Rule 5*
9	0	2	0	1	Medium	*Rule 5*
1	0	0	1	2	Low	*Rule* 6

Case 2

Consider the data in Table 4 for eleven objects, eight attributes, and the decision D. When the data from each source is considered independently, the rough set algorithm generates the rule sets shown in Figures 6 and 7. The patterns resulting from the rule sets of Figures 6 and 7 are shown in Table 5. The steps of the rule-structuring algorithm are illustrated next with the matrix of Table 5.

Step 1) The two attributes A1 and A3 are the most frequently used by the rules. The attribute A3 is arbitrarily selected among the two.

Step 2) The rules 2, 6, and 7 included in the equivalence class (A3 in [3,3]) are clustered (see Table 6). The rules 1, 4, and 11 in the equivalence class (A3 in [2,2]) are clustered (see Table 6).

Step 3) Applying the data engineering principles, to the matrix of Table 6 results in the matrix in Table 7.

Table 4.
Data set generated from two different sources

NO.	Attributes								D
	A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	
1	High	3.7	2	True	0.95	1.83	Blue	1.12	Yes
2	Low	2.3	3	True	1.13	2.18	Blue	2.35	No
3	Medium	2.1	1	False	0.59	2.38	Red	2.42	Yes
4	Medium	3.4	2	True	0.76	1.62	Black	1.54	Yes
5	High	4.2	1	False	0.42	1.32	Black	1.83	Yes
6	Low	3.1	3	False	1.37	3.89	White	1.29	No
7	Low	4.9	3	True	1.41	2.24	Red	2.48	No
8	Medium	2.4	1	False	1.53	4.29	Red	2.25	No
9	High	4.6	1	False	1.74	4.09	White	1.98	Yes
10	Medium	3.3	1	True	1.62	4.36	Black	2.63	No
11	Medium	2.7	2	False	1.88	3.42	Orange	1.69	Yes

Rule S1-1: IF (A5 in [1.50,1.65]) THEN (D=No); [2,40%,100%] [8,10]

***Rule* S1-2:** *IF* (A3 in [3,3]) *THEN* (D=No); [3,60%,100%] [2,6,7]

***Rule* S1-3:** *IF* (A8 in [1.75,2.15]) *THEN* (D=Yes); [2,33.33%,100%] [5,9]

***Rule* S1-4:** *IF* (A5 in [0.53,0.85]) *THEN* (D=Yes); [2,33.33%,100%] [3,4]

***Rule* S1-5:** *IF* (A3 in [2,2]) *THEN* (D=Yes); [3,50%,100%] [1,4,11]

Figure 6. Decision rules derived from Source 1 data.

***Rule* S2-1:** *IF* (A1=High) *THEN* (D=Yes); [3,50%,100%] [1,5,9]

***Rule* S2-2:** *IF* (A6 in [2.35,3.50]) *THEN* (D=Yes); [2,33.33%,100%] [3,11]

***Rule* S2-3:** *IF* (A6 in [4.22,4.45]) *THEN* (D=No); [2,40%,100%] [8,10]

***Rule* S2-4:** *IF* (A1=Low) *THEN* (D=No); [3,60%,100%] [2,6,7]

***Rule* S2-5:** *IF* (A6 in [1.55,1.75]) *THEN* (D=Yes); [1,16.67%,100%] [4]

Figure 7. Decision rules derived from Source 2 data.

Table 5
Patterns of rules from Figures 6 and 7

No.	A1	A2	A3	A4	A5	A6	A7	A8	D	Rules
1	High	3.7	2	True	0.95	1.83	Blue	1.12	Yes	S1-5 , S2-1
2	Low	2.3	3	True	1.13	2.18	Blue	2.35	No	S1-2 , S2-4
3	Medium	2.1	1	False	0.59	2.38	Red	2.42	Yes	S1-4 , S2-2
4	Medium	3.4	2	True	0.76	1.62	Black	1.54	Yes	S1-4 , S1-5 , S2-5
5	High	4.2	1	False	0.42	1.32	Black	1.83	Yes	S1-3 , S2-1
6	Low	3.1	3	False	1.37	3.89	White	1.29	No	S1-2 , S2-4
7	Low	4.9	3	True	1.41	2.24	Red	2.48	No	S1-2 , S2-4
8	Medium	2.4	1	False	1.53	4.29	Red	2.25	No	S1-1 , S2-3
9	High	4.6	1	False	1.74	4.09	White	1.98	Yes	S1-3 , S2-1
10	Medium	3.3	1	True	1.62	4.36	Black	2.63	No	S1-1 , S2-3
11	Medium	2.7	2	False	1.88	3.42	Orange	1.69	Yes	S1-5 , S2-2

Table 6
Clustered matrix

No.	A1	A2	A3	A4	A5	A6	A7	A8	D	Rules
2	Low	2.3	3	True	1.13	2.18	Blue	2.35	No	S1-2 , S2-4
6	Low	3.1	3	False	1.37	3.89	White	1.29	No	S1-2 , S2-4
7	Low	4.9	3	True	1.41	2.24	Red	2.48	No	S1-2 , S2-4
1	High	3.7	2	True	0.95	1.83	Blue	1.12	Yes	S1-5 , S2-1
4	Medium	3.4	2	True	0.76	1.62	Black	1.54	Yes	S1-4 , S1-5 , S2-5
11	Medium	2.7	2	False	1.88	3.42	Orange	1.69	Yes	S1-5 , S2-2
5	High	4.2	1	False	0.42	1.32	Black	1.83	Yes	S1-3 , S2-1
9	High	4.6	1	False	1.74	4.09	White	1.98	Yes	S1-3 , S2-1
8	Medium	2.4	1	False	1.53	4.29	Red	2.25	No	S1-1 , S2-3
10	Medium	3.3	1	True	1.62	4.36	Black	2.63	No	S1-1 , S2-3
3	Medium	2.1	1	False	0.59	2.38	Red	2.42	Yes	S1-4 , S2-2

Table 7
Matrix transformed by the attribute value replacement and column removal principles.

A1	A3	A5	A6	A8	D
Low	3	1.13-1.41	2.18-3.89	1.29-2.48	No
High	2	0.95	1.83	1.12	Yes
Medium	2	0.76-1.88	1.62-3.42	1.54-1.69	Yes
High	1	0.42-1.74	1.32-4.09	1.83-1.98	Yes
Medium	1	1.53-1.62	4.29-4.36	2.25-2.63	No
Medium	1	0.59	2.38	2.42	Yes

5- Conclusion and Future works

In traditional approaches to decision making, usually scientific expertise together with statistical techniques have been needed to support the managers. However, these approaches are not able to handle the huge amount of real data, and the processes are usually very slow. This research has reviewed the basic concepts of data mining and rough set theory for decision making. Moreover, the potential applications of rough set theory in data mining were investigated. It is shown that this approach is a viable method for extraction of meaningful knowledge and making prediction for an individual data objects rather than a population of objects. Moreover, it was demonstrated that a rule extracted from a data set and corresponding features (attributes) can be considered as one of many models describing a data set.

6-References

[1] W.Frawley, G.Piatetsky, C.Matheus, Knowledge Discovery in Databases: An Overview, AI Magazine, Fall 1992, PP. 213-228.

[2] "State of the Art" in Byte Magazine October 1995, Contains three articles on Data Mining and Data Warehousing.

[3] A.Kusiak, Rough Set Theory: A Data Mining Tool for Semiconductor manufacturing, IEEE Transaction on Electronics Packing Manufacturing, Vol.24, No.1, January 2001.

[4] T.Munakata, Z.Pawlak, Rough Control Application of Rough Set Theory to Control, Fourth European Congress on Intelligent Techniques and Soft Computing, Aachen, Germany, September 2 – 5, 1996, Proceedings, Volume 1, PP. 209 – 218.

[5] B.Thuraisingham, Data Mining: Technologies, Techniques, Tools, and Trends, CRC Press, 1999.

[6] V.Ganti, J.Gebrke, R.Ramakrishnan, Mining Very Large Databases, Computers, August 1999, PP. 38-45.

[7] J.Han, L.Lakshmanan, R.Ng, Constraint-Based Multidimensional Data Mining, Computers, August 1999, PP. 45-50.

[8] J.Hellerstein, R.Avnr, A.Chou, C.Olston, T.Rath, P.Haas, Interactive Data Analysis; the Control Project, Computers, August 1999, PP. 51-58.

[9] N.Ramakrishnan, A.Grama, Data Mining: from Serendipity to Science, Computers, August 1999, PP.34.

[10] P.R.Limb, G.J.Meggs; Data Mining: Tools and Techniques, BT Technology Journal, 1994,Vol.12, No.4, PP.32-41.

[11] U.Fayyad, G.S.Piatetsky, The KDD Process for Extracting Useful Knowledge from Volumes of Data, Communications of the ACM, 1996, V.39, No.11, PP.27-34.

[12] X.Z.Wang, Data Mining and Knowledge Discovery for Process Monitoring and Control, Springer- Verlag, 1999.

[13] Z.Pawlak, Rough Sets, International Journal of Information and Computer Science, 1982, Vol.11, No.5, PP.344-356.

[14] D.E.Nelson, High Range Resolution Radar Target Classification: A Rough Set Approach, Ph.D. Dissertation, Ohio University College of Engineering and Technology, 2001, PP.11-30.

[15] Z.Pawlak, AI and Intelligent Industrial Applications: The Rough Set Perspective, Cybernetics and Systems, An International Journal, April 2000, Vol.31.

[16] A.Kusiak, A.Kern, H.Kernstine, T.Tseng, Autonomous Decision-Making: A Data Mining Approach, IEEE Transaction on Information Technology in Biomedicine, Vol.4, No.4, December 2001.

[17] A.Kusiak, T.Tseng, Data Mining in Engineering Design: A Case Study, Proceedings of the 2000 IEEE International Conference on Robotics & Automation, Sanfrancisco, CA, April 2000.

[18] L.Zhai, L.Khoo, S.Fok, Feature Extraction Using Rough Set Theory and Genetic Algorithms: An Application for The Simplification of Product Quality Evaluation, Computers & Industrial Engineering, 2002, No.43, PP.661-672.

[19] A.Kusiak, C.Kurasek, Data Mining of Printed- Board Defects, IEEE Transaction on Robotics & Automation, April 2001, Vol.17, No.2.