The Dominance-Based Rough Set Approach as a Business Analytical Tool

Ayrton Benedito Gaia do Couto
System Analyst
BNDES
Rio de Janeiro, RJ, Brazil
ayrtoncouto@gmail.com

Luiz Flavio Autran Monteiro Gomes
Professor
Ibmec/RJ
Rio de Janeiro, RJ, Brazil
autran@ibmecrj.br

***Abstract*— This study demonstrates how to extract essential information from a set of financial and non-financial business indicators, leading to a better understanding of company performance. The use of Rough Set Theory and the Dominance principle associated with the probabilistic relationship between conditions and decisions in decision algorithms, is justified by the possibility of there being uncertain data to yield an essential set of effectively consistent information. The analysis was based on the Brazilian publication Exame Melhores e Maiores 2013 which lists the 500 largest companies in various economic sectors ordered by net sales. The study reveals the importance of broadening the analysis of enterprise indicators, and shows a method of describing conclusions from data without referring to prior and posterior probabilities.**

Index Terms - Business Performance; Dominance principle; Rough Set Theory; Multi-Criteria Analysis.

I. Introduction

Analysis of financial and non-financial business indicators allows one to better understand company performance. The present study was based on the following research question: "How can patterns be inferred from the analysis of business indicators using a Multi-Criteria approach?". The study is based on a publication listing the 500 largest companies from several sectors of the Brazilian economy [1] and uses Multi-Criteria decision-supporting tools to collectively analyse the various attributes (financial and non-financial indicators) as condition and decision criteria. The choice of Rough Set Theory (RST) and Dominance-based Rough Set Approach (DRSA) associated with the probabilistic relationship between conditions and decisions in decision algorithms, as tools to support Multi-Criteria decision is justified by the possibility of there being uncertain (inconsistent) data. RST was chosen because it does not require any preliminary information about the data at hand, such as its probability distribution. Other theories might be used – e.g., the Fuzzy Set Theory, proposed by Lotfi Asker Zadeh in 1965 [2] as an extension of conventional (boolean) logic that introduces the concept of non-absolute truth and serves as a tool for dealing with uncertainties in natural language [3]. RST and Fuzzy Set Theory are independent approaches to the treatment of imperfect (incomplete) and uncertain (vague, indeterminate) knowledge [4]. RST has shown to be of fundamental importance to artificial intelligence (AI) and cognitive sciences, especially in the areas of machine learning, knowledge acquisition, decision analysis, knowledge discovery from databases, expert systems, decision support systems, inductive reasoning, and pattern recognition [4]. To support Multi-Criteria analysis, we used the jMAF (Dominance-Based Rough Set Data Analysis Framework) software tool [5], which is available for research purposes by the Institute of Computer Science of the Poznan University of Technology in Poland. The present paper comprises a short introduction to Rough Set Theory (RST), Decision Rule and Decision Algorithms, and to the Dominance-based Rough Set Approach (DRSA) in sections II, III and IV, respectively, a study conducted of the 20 (twenty) largest companies based on their net sales and business indicators in section V, and conclusions and remarks on future studies in section VI.

II. Rough Set Theory

RST had its origin with Zdzislaw Pawlak [6]: it proposes the treatment of data uncertainty using "lower and upper approximations" for a data set [7]. One of its concepts, the "indiscernibility relation," identifies objects that have the same properties, i.e., "indiscernible" objects, to be treated as similar or identical. An information system can be defined as a tuple S = (U, Q, V, f), where U is a finite set of objects, Q is a finite set of attributes, $V= \bigcup_{q\in Q} V_q$, where V_q is the domain of attribute q, and f: $U \chi Q \rightarrow V$ is a total function such that $f(x, q) \in Vq$ for every $q \in Q$, $x \in U$, known as an "information function" [8]. Given an information system S = (U, Q, V, f), $P \subseteq Q$, and $x,y \in U$, we say x and y are "indiscernible" through the set of attributes P in S if f(x,q) = f(y,q) for all $q\in P$. Therefore, all $P \subseteq Q$ generate a binary relation in U, known as an "indiscernibility relation", denoted by IND(P). Given that $P \subseteq Q$ and $Y \subseteq U$, the lower ($\underline{P}Y$) and upper approximations ($\overline{P}Y$) are defined as:

$$\underline{P}Y= \cup\{X\in U/P:X \subseteq Y\};\ \overline{P}Y= \cup\{X\in U/P: X \cap Y \neq \emptyset\} \quad (1)$$

The difference between $\underline{P}Y$ and $\overline{P}Y$ is called the "boundary region" of Y:

$$BNP(Y) = \underline{P}Y - \overline{P}Y \quad (2)$$

There is also the concept of accuracy:

$$\alpha P(Y) = \text{card}\ \underline{P}\ /\text{card}\ \overline{P} \quad (3)$$

which captures the degree to which the knowledge of set Y is complete.

There are two more fundamental concepts in RST: an information system's "reduct" and "core". The reduct is its essential part, i.e., the subset of attributes that provides the same quality of classification as the original set of attributes (it allows one to make the same decisions as if all condition attributes were there). The core is the most important subset of this knowledge [7], [8]. If R is a family of relations and R ∈ R [7], then R is termed "dispensable" in R if IND(R) = IND(R – {R}); otherwise, R is "indispensable" in R. A family R is "independent" if every R ∈ R is indispensable in R; otherwise, R is "dependent". From these, the following propositions were defined:

a) If R is independent and P ⊆ R, then P is also independent;

b) CORE(P) = ∩ RED(P), where RED(P) is the family of all "reducts" of P.

The family of reducts and the core, should there be any, can also be identified by building a Discernibility Matrix [10], [11], [12]. As a demonstration of RST's application, refer to the cited example of decision-making in determining the number of executives based on replicated and inconsistent data (Human Resources universe) [13].

III. Decision Rule and Decision Algorithms

Reference [14], [15]: let S = (U, A) be an information system. With every B ⊆ A we associate a formal language, i.e., a set of formulas For(B). Formulas of For(B) are built up from attribute-value pairs (a, v) where $a \in B$ and $v \in V_a$ by logical connectives ∧ (*and*), v (*or*), ~ (*not*) in the standard way. For any $\Phi \in For(B)$ by $\|\Phi\|_s$ we denote the set of all of objects $x \in U$ satisfying Φ in *S* and refer to as the meaning of Φ in *S*. The meaning of $\|\Phi\|_s$ of Φ in *S* is defined inductively as follows:

$$\|(a, v)\|_s = \{ x \in U : a(v) = x\} \quad \forall a \in B, \; v \in V_a \tag{4}$$

$$\| \Phi \vee \Psi \|_s = \| \Phi\|_s \cup \| \Psi \|_s , \tag{5}$$

$$\| \Phi \wedge \Psi \|_s = \| \Phi\|_s \cap \| \Psi \|_s , \tag{6}$$

$$\|\sim \Phi\|_s = U - \| \Phi\|_s . \tag{7}$$

A decision rule in S is an expression Φ → Ψ, read *if* Φ *then* Ψ, where Φ ∈ For(C), Ψ ∈ For(D) and C and D are condition and decision attributes, respectively; Φ and Ψ are referred to as *conditions* and *decisions* of the rule, respectively.

If a decision rule Φ → Ψ is true in S, we will say that the decision rule is *consistent* is S; otherwise, the decision rule is *inconsistent* in S. If Φ → Ψ is a decision rule and Φ and Ψ are P-basic and Q-basic formulas respectively, then the decision rule Φ → Ψ will be called a PQ-basic decision rule, (in short PQ-rule), or basic rule when PQ is known. A PQ-rule Φ → Ψ is *admissible* in S if Φ ∧ Ψ is satisfiable in S. Any finite set of basic decision rules will be called a *basic decision algorithm*. If all decision rules in a basic decision algorithm are PQ-decision rules, then the algorithm is said to be PQ-*decision algorithm*, or in short PQ-*algorithm*, and will be denoted by (P,Q). A PQ—algorithm is *admissible* in S, if the algorithm is the set of all PQ-rules admissible in S. The PQ-algorithm is *consistent* in S, if and only if all its decision rules are consistent (true) in S; otherwise the algorithm is *inconsistent* in S [7].

Some basic concepts are as follows [14], [15]: the number $supp_s(\Phi, \Psi) = card(\| \Phi \wedge \Psi\|_s)$ will be called the *support* of the rule Φ → Ψ in S. We consider a probability distribution $p_u(x) = 1/card(U)$ for $x \in U$ where U is the (non-empty) universe of objects of S; we have $p_u(X) = card(X)/card(U)$ for $X \subseteq U$. For any formula Φ we associate its probability in S defined by $\pi_s = p_u(\|\Phi\|_s)$.With every decision rule Φ → Ψ, we associate a conditional probability $\pi_s(\Psi \mid \Phi) = p_u(\|\Psi\|_s \mid \|\Phi\|_s)$ called *certainty factor* of the decision rule, denoted $cer_s(\Phi, \Psi)$. So it follows:

$$\begin{aligned} cer_s(\Phi, \Psi) &= \pi_s(\Psi \mid \Phi) \\ &= card(\| \Phi \wedge \Psi \|_s) / card(\| \Phi\|_s), \end{aligned} \tag{8}$$

where $\| \Phi\|_s \neq 0$. If $\pi_s(\Psi \mid \Phi) = 1$, then Φ → Ψ will be called a *certain decision* rule; if $0 < \pi_s(\Psi \mid \Phi) < 1$ the decision rule will be referred to as an *uncertain decision* rule.

A *coverage factor* of the decision rule, denoted $cov_s(\Phi \mid \Psi)$, defined by $\pi_s(\Phi \mid \Psi) = p_u(\| \Phi\|_s \mid \| \Psi\|_s)$. So it follows:

$$\begin{aligned} cov_s(\Phi \mid \Psi) &= \pi_s(\Phi \mid \Psi) \\ &= card(\| \Phi \wedge \Psi \|_s) / card(\|\Psi\|_s). \end{aligned} \tag{9}$$

Under a statistical interpretation, i.e., the certainty factor is the estimate of conditional probability that Ψ is true in S given Φ is true in S with the probability $\pi_s(\Phi)$. This coefficient is now widely used in data mining and is called *confidence coefficient*. The coverage factor is the estimate of conditional probability that Φ is true in S given Ψ is true in S with the probability $\pi_s(\Psi)$. The certainty and the coverage factors of decision rules express how exact is the knowledge (data) about the considered reality; they are not assumed to be arbitrary but are computed from the data, thus there are in a certain sense objective.

The number

$$\begin{aligned} \sigma_S(\Phi, \Psi) &= supp_s(\Phi, \Psi) / card(U) \\ &= \pi_s(\Psi \mid \Phi)\, \pi_s(\Phi) \end{aligned} \tag{10}$$

will be called the *strength* of the decision rule Φ → Ψ in S; this number express the ratio of all facts which can be classified by the decision rule to all facts in the data table.

Examples of decision rules admissible in Table I are given below:

1- (E, average) and (Q, good) → (P, loss)
2- (E, none) → (P, loss)
3- (E, average) and (Q, average) → (P, loss)

TABLE I. EXAMPLE WITH STORES AND INITIAL ATTRIBUTES

Store	E	Q	L	P
1	High	Good	No	Profit
2	Average	Good	No	Loss
3	Average	Good	No	Profit
4	None	Average	No	Loss
5	Average	Average	Yes	Loss
6	High	Average	Yes	Profit

For previous set at decision rules, the partial certainty and coverage factors, respectively:

$$\pi_s(\Psi \mid \Phi_1) = 1/2,\ \pi_s(\Psi \mid \Phi_2) = 1,\ \pi_s(\Psi \mid \Phi_3) = 1, \text{ and} \quad (11)$$

$$\pi_s(\Phi_1 \mid \Psi) = 1/3,\ \pi_s(\Phi_2 \mid \Psi) = 1/3,\ \pi_s(\Phi_3 \mid \Psi) = 1/3, \quad (12)$$

where Φ_i is the condition of rule "i", and Ψ = (P, loss). Let $\{\Phi_i \rightarrow \Psi\}_n$ be a set of decision rules such that all conditions Φ_i are pairwise mutually exclusive, i.e., $\| \Phi_i \wedge \Psi_j \|_s = \emptyset$, for any $1 \leq i, j \leq n$, $i \neq j$, and

$$\sum_{i=1}^{n} \pi_s(\Phi_i \mid \Psi) = 1 \quad (13)$$

The following property holds:

$$\pi_s(\Psi) = \sum_{i=1}^{n} \pi_s(\Psi \mid \Phi_i) \,.\, \pi_s(\Phi_i) \quad (14)$$

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

$$\pi_s(\Phi \mid \Psi) = \pi_s(\Psi \mid \Phi) \,.\, \pi_s(\Phi) \,/ \sum_{i=1}^{n} \pi_s(\Psi \mid \Phi_i) \,.\, \pi_s(\Phi_i) \quad (15)$$

It is easy to check that the set of decision rules statistic properties (13), (14) and (15), i.e.,

$$\pi(\Psi) = 1/2 \,.\, 1/3 + 1 \,.\, 1/6 + 1 \,.\, 1/6 = 1/2, \quad (16)$$

$$\pi(\Phi_1 \mid \Psi) = 1/3,\ \pi(\Phi_2 \mid \Psi) = 1/3,\ \pi(\Phi_3 \mid \Psi) = 1/3. \quad (17)$$

It can be easily seen that the relationship between the certainty factor and the coverage factor, expressed by the formula "15" is the Bayes' Theorem. It is interesting to observe that the total probability theorem is closely related with *modus ponens* (MP) and *modus tollens* (MT) inference rules. MP has the following form:

if	$\Phi \rightarrow \Psi$	is true
and	Φ	is true
then	Ψ	is true

If we replace truth values by corresponding probabilities we can generalize the inference rules as rough modus ponens (RMP), which has the form

if	$\Phi \rightarrow \Psi$	is true with probability $cer_s(\Phi, \Psi)$
and	Φ	is true with probability $\pi_s(\Phi)$
then	Ψ	is true with probability

$$\begin{aligned}\pi_s(\Psi) &= \textstyle\sum_{\Phi' \in C(\varphi)} cer_s(\Phi', \Psi)\, \pi_s(\Phi') \\ &= \textstyle\sum_{\Phi' \in C(\varphi)} \sigma_s(\Phi', \Psi)\end{aligned} \quad (18)$$

RMP enables us to calculate the probability of conclusion Ψ of a decision rule $\Phi \rightarrow \Psi$ in terms of strengths of all decision rules in the form $\Phi' \rightarrow \Psi$, $\Phi' \in C(\Psi)$. Similarly, if we replace truth values by probabilities we get the following rough modus tollens (RMT) inference rule

if	$\Phi \rightarrow \Psi$	is true with probability $cov_s(\Phi, \Psi)$
and	Ψ	is true with probability $\pi_s(\Psi)$
then	Ψ	is true with probability

$$\begin{aligned}\pi_s(\Phi) &= \textstyle\sum_{\Psi' \in D(\Phi)} cov_s(\Phi, \Psi')\, \pi_s(\Psi') \\ &= \textstyle\sum_{\Psi' \in D(\Phi)} \sigma_s(\Phi, \Psi')\end{aligned} \quad (19)$$

RMT enables us to compute the probability of condition Φ of the decision rule $\Phi \rightarrow \Psi$ in terms of strengths of all decision rules in the form $\Phi \rightarrow \Psi'$, $\Psi' \in D(\Phi)$. MP and MT are used to draw conclusions from logical axioms, whereas RMP and RMT are used to compute probabilities of decisions (conditions) in decision tables (decision algorithms).

IV. DOMINANCE PRINCIPLE

The key aspect of a Multi-Criteria decision is considering objects that are described by multiple criteria and that represent conflicting points of view. Criteria are attributes in domains with an ordering preference; e.g., in choosing a car, one may consider the price and fuel consumption to be characteristics that should serve as criteria in its acquisition, as one usually considers a low price to be better than a high price and moderate fuel consumption to be more desirable than high consumption. In general, other attributes such as colour and country of origin, the domains of which have no ordering preference, are not considered to be decision criteria – they are regular attributes. Therefore, the RST approach does not allow one to analyse multi-criteria decision problems because the analysis uses only regular attributes. Moreover, one cannot identify inconsistencies that violate the following dominance principle: "objects with a better evaluation or having at least the same evaluation (decision class) cannot be associated to a worse decision class, all decision criteria being considered". RST ignores not only the preference ordering in the set of attributes' values but also the "monotonic" relation of objects' evaluations regarding the condition attributes' values and decision attributes' values' order of preference (classification or degree of preference) [16], [17]. This problem is treated in an extension of RST called Dominance-based Rough Set Approach or DRSA [16], in which indiscernibility relations are replaced with dominance relations in the approximations of decision classes. Furthermore, due to the preferential ordering between decision classes, sets become approximations known as unions of "upward" and "downward" decision classes. Thus, for a tuple S = (U, Q, V, f), set Q is generally divided into condition attributes (set C) and decision attributes (set D). Assuming all condition attributes ($q \in C$) are decision criteria, S_q represents a non-classifiable relation in U with respect to criterion q such that xS_qy denotes "x is at least as good as y in regards to criterion q". Assuming the set of decision attributes D defines a partition of U into a finite number of classes, $Cl = \{Cl_t, t \in T\}$, $T = \{1, ..., n\}$ is a set of these classes such that each $x \in U$ belongs to one and only one $Cl_t \in Cl$. These classes are assumed to be ordered, i.e., for every $r,s \in T$ such that $r > s$, objects of Cl_r are preferable to objects of Cl_s. Therefore, objects can be approximated by unions of "upward" and "downward" decision classes, respectively: $Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s$, $Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s$, t=1, ...,n. The indiscernibility relation is thus substituted with a dominance relation. One says that x dominates y regarding $P \subseteq C$, denoted xD_Py, if xS_qy for all $q \in P$. The dominance relation is reflexive and transitive. Given that $P \subseteq C$ and $x \in \cup$, the "granules of knowledge" used in the DRSA approximations are:

- a set of dominating objects x, called the P-dominating set: $D_P^+(x) = \{y \in \cup : yD_Px\}$,

- a set of objects dominated by x, called the P-dominated set: $D_P^-(x) = \{x \in \cup : xD_Py\}$.

Using the $D_P^+(x)$ sets, the P-lower and P-upper approximations of $Cl_t^{\geq}$ are: $\underline{P}(Cl_t^{\geq}) = \{x \in \cup : D_P^+(x) \subseteq Cl_t^{\geq}\}$, $\overline{P}(Cl_t^{\geq}) = \bigcup_{x \in Cl_t^{\geq}} D_P^+(x)$, for *t*=1,...,n. Analogously, the P-lower and P-upper approximations of ($Cl_t^{\leq}$) are: $\underline{P}(Cl_t^{\leq}) = \{x \in \cup : D_P^-(x) \subseteq Cl_t^{\leq}\}$, $\overline{P}(Cl_t^{\leq}) = \bigcup_{x \in Cl_t^{\leq}} D_P^-(x)$, for *t*=1,...,n. The P-boundary sets of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ are: $Bn_P(Cl_t^{\geq}) = \overline{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq})$, $Bn_P(Cl_t^{\leq}) = \overline{P}(Cl_t^{\leq}) - \underline{P}(Cl_t^{\leq})$, for *t*=1,...,n. These approximations to the unions of "upward" and "downward" decision classes can be used to infer decision rules of the form "*if ... then ...*". For a given union of "upward" or "downward" of decision classes $Cl_t^{\geq}$ or $Cl_t^{\leq}$, *s,t* $\in T$, the rules induced under the hypothesis that objects pertaining to lower approximations $\underline{P}(Cl_t^{\geq})$ or $\underline{P}(Cl_t^{\leq})$ are positive and all others are negative suggest that an object be attributed to "at least one class Cl_t" or to "at most one class Cl_s", respectively. These rules are known as "certain decision rules" ($D_{\leq}$ or $D_{\geq}$) because they attribute objects to unions of decision classes without any ambiguity. Alternatively, if objects pertain to upper approximations, the rules are known as "possible decision rules"; thus, objects could pertain to "at least one class Cl_t" or "at most one class Cl_s". Finally, if objects pertain to the intersection $\overline{P}(Cl_s^{\leq}) \cap \overline{P}(Cl_t^{\geq})$ ($s<t$), the rules induced are known as "approximate rules", i.e., objects are between classes Cl_s and Cl_t. Therefore, if for each criterion $q \in C$, $V_q \subseteq \mathbf{R}$ (V_q is quantitative) and for each $x,y \in U$, $f(x,q) \geq f(y,q)$ implies xS_qy (V_q has a preferential ordering), decision rules can be considered to be of five types:

1- certain $D_{\geq}$-decision rules:

if $f(x,q_1) \geq r_{q1}$ *and* $f(x,q_2) \geq r_{q2}$ *and ...* $f(x,q_p) \geq r_{qp}$, *then* $x \in Cl_t^{\geq}$

2- possible $D_{\geq}$-decision rules:

if $f(x,q_1) \geq r_{q1}$ *and* $f(x,q_2) \geq r_{q2}$ *and ...* $f(x,q_p) \geq r_{qp}$, *then x possibly belongs to* $Cl_t^{\geq}$

3- certain $D_{\leq}$-decision rules:

if $f(x,q_1) \leq r_{q1}$ *and* $f(x,q_2) \leq r_{q2}$ *and ...* $f(x,q_p) \leq r_{qp}$, *then* $x \in Cl_t^{\leq}$

4- possible $D_{\leq}$-decision rules:

if $f(x,q_1) \leq r_{q1}$ *and* $f(x,q_2) \leq r_{q2}$ *and ...* $f(x,q_p) \leq r_{qp}$, *then x possibly belongs to* $Cl_t^{\leq}$, where $P = \{q_1, ..., q_p\} \subseteq C$, $(r_{q1}, ..., r_{qp}) \in V_{q1} \times V_{q2} \times ... \times V_{qp}$ and $t \in T$;

5- approximate $D_{\leq\geq}$-rules:

if $f(x,q_1) \geq r_{q1}$ *and* $f(x,q_2) \geq r_{q2}$ *and ...* $f(x,q_k) \geq r_{qk}$ *and* $f(x,q_{k+1}) \leq r_{qk+1}$ *and* $f(x,q_p) \leq r_{qp}$, *then* $x \in Cl_s \cup Cl_{s+1} \cup ... \cup Cl_t$.

Rules of types "1" and "3" represent "certain knowledge" extracted from a data table (or information system), rules of types "2" and "4" represent "possible knowledge", and the rule of type "5" represent "ambiguous knowledge".

V. Multi-Criteria analysis of business indicators

Reference [18], intangible assets as a decisive factor in obtaining competitive advantage characterises became increasingly important at the end of the 20th century. Organisations look for ways to better measure and present these intangible assets to managers and investors. Aggregate measures, such as ROI, ROE and operating profit, are no longer able to capture the complexity and values in the business environment. This environment is now process oriented, and the predominant aspects are related to identifying opportunities, speed of learning, innovation, quality, flexibility, reliability and response capacity, all of which must be measured. Analysts who use this type of non-financial information are capable of producing the most accurate projections; if they can communicate these attributes, they will more easily obtain funding from third parties. Studies show that investors want to understand the business model in greater depth, with a view into the main performance indicators, and they pay attention to non-financial aspects, which they use in guiding their decisions to invest. Table II shows these measures organised into nine categories: Financial (A), Product Quality (B), Client Satisfaction (C), Process Efficiency (D), Product and Process Innovation (E), Competitive Environment (F), Management Quality and Independence (G), Human Resources Administration (H), Social Responsibility (I).

TABLE II. MEASURES UNDER STUDY [18]

A - Financial
1- Net profit and profit/share
2- Cash flow
3- ROE (Return on Equity)
4- ROA (Return on Assets)
5- Sales
6- Return on Sales
7- Sales/Total Assets
8- Net Equity / Total Assets
9- Quality of accounting practices

B - Product Quality
10- % repeated sales
11- Clients that improve company image
12- Complaints within warranty
13- Client complaints

C - Client Satisfaction
14- Market research
15- Punctual delivery
16- Service response time
17- % returning customers
18- NAV (net asset value)
19- % contacted clients who effectively buy
20- Litigation with clients

D - Process Efficiency
21- Rate of defects
22- Product development time
23- Manufacturing cycle time
24- Time between ordering and delivery
25- Capacity for customization
26- Operational costs /employee
27- Sales/employee
28- COGS (Cost of goods sold)/ stock
29- Accounts receivable/sales
30- Capital investment
31- Plant/equipment age
32- Usage of installed capacity

E - Product/Process Innovation
33- R&D expenditures
34- % of patented products
35- Number of new patents
36- Number of new products
37- % of new product sales

F - Competitive Environment
38- Market share
39- Brand recognition
40- Potential competition
41- Tax/quota protection
42- % sales of patented products
43- Strategic alliances
44- Litigation w/ anti-trust legislation
45- Geographic diversification
46- Client diversification
47- Product diversification

G - Management Quality/Independence
48- Management continuity
49- Managers' experience/reputation
50- Managing council's involvement
51- Managing council's independence
52- Litigation with shareholders
53- Weakening of control
54- Managers' ethical behaviour
55- Value offered to investors

H - Human Resources Administration
56- Equal opportunity employment
57- Employee involvement
58- Profit sharing
59- Stock options plan
60- % candidates to positions in competing companies successfully recruited
61- Job/employee development
62- % new employees
63- Benefit policy

I - Social Responsibility
64- Minority protection
65- Performance in environmental activities
66- Involvement with communities
67- Litigation

Of these measures, those were sought that would generate the most aggregated value, as shown in Table III.

TABLE III. CAPACITY TO PREDICT VALUE (MEASURE, CATEGORY, MEAN SCORE) [18]

Measure	Ct	M	Measure	Ct	M
Clients that improve company image	B	4.00	Litigation w/ anti-trust legislation	F	2.64
Weakening of control	G	4.00	Profit sharing	H	2.62
Managing council's independence	G	4.00	Managing council's involvement	G	2.62
Sales	A	3.90	Time between ordering and delivery	D	2.49
Cash flow	A	3.70	R&D expenditures	E	2.45
Return on Sales	A	3.40	% sales of patented products	F	2.44
Usage of installed capacity	D	3.35	% returning customers	C	2.36
Market share	F	3.31	% of new product sales	E	2.31
Net profit and profit/share	A	3.30	Stock options plan	H	2.31
Client diversification	F	3.29	Performance in environmental activities	I	2.31
Accounts receivable/sales	D	3.29	% repeated sales	B	2.31
Capital investment	D	3.27	Number of new products	E	2.24
Potential competition	F	3.27	Market research	C	2.20
Geographic diversification	F	3.24	% contacted clients who effectively buy	C	2.16
Plant/equipment age	D	3.22	Capacity for customization	D	2.16
ROE (Return on Equity)	A	3.20	Client complaints	B	2.11
Product diversification	F	3.18	Service response time	C	2.10
Sales/ total assets	A	3.10	Rate of defects' amount	D	2.07
Net Equity / total Assets	A	3.10	Product development time	D	2.07
Quality of accounting practices	A	3.10	Litigation	I	2.07
Sales/employee	D	3.09	Minority protection	I	2.04
Operational/employee costs	D	3.05	Job/employee development	H	2.00
Tax/quota protection	F	3.04	Punctual delivery	C	2.00
Litigation with shareholders	G	3.00	% of patented products	E	1.98
ROA (Return on Assets)	A	3.00	Number of new patents	E	1.98
Management continuity	G	2.98	Litigation with clients	C	1.96
COGS (Cost of goods sold)/ stock	D	2.93	Involvement with communities	I	1.96
Managers' experience/reputation	G	2.89	Benefit policy	H	1.96
Value offered to investors	G	2.82	Complaints within warranty	B	1.95
Brand recognition	F	2.80	Employee involvement	H	1.93
Strategic alliances	F	2.76	Equal opportunity employment	H	1.90
Managers' ethical behaviour	G	2.75	% candidates to positions in competing companies successfully recruited	H	1.85
Manufacturing cycle time	D	2.73	% new employees	H	1.75
NAV (net asset value)	C	2.71			

The top 20 companies (for which all information needed for this study was available) were selected from among the 500 largest companies in various economy sectors from the net sales ranking published in Exame magazine [1]. They are shown in Table IV, where their names have been replaced with letters in the same order as their net sales ranking. The productivity metric "revenue created per employee" was unavailable for some companies; in these cases, the "net sales per employee" metric was adopted, considering "productivity" to be the ratio of output(s) to input(s) [19].

Based on Table III, we sought to select the information (measures) that aggregates the most value to a company. For example, "Market share" (mean 3.31) is preferable to "R&D expenditures" (mean 2.45). Market share can be considered to be an indicator of "efficacy" as the ratio of actual to expected outputs – in this case, (company's net sales / total net sales of the economy sector) x 100 [19]. This group of the 20 largest companies has representatives from the following sectors of the economy: wholesale, automotive industry, consumer goods, energy, mining, chemical and petrochemical, services, steelworks and metallurgy, telecommunications and retail. The data in Table IV were then analysed using the jMAF software application [5].

TABLE IV. THE 20 LARGEST COMPANIES BY NET SALES [1]

Company	Ranking	Net Sales (US$ millions)(1)	Legal Net Profit (US$ millions)(2)	Legal Profitability (%) (3)	Net working capital (US$ millions) (4)	General liquidity (index number) (5)	General debt (%) (6)	EBITDA (US$ millions) (7)	Net sales/employee (US$ millions) (8)	Market share (%) (9)
A	1	109,713.30	10,225.10	5.9	13,801.30	0.57	38.9	22,774.70	1.8	46.5
B	2	39,024.50	925.40	17.1	1,217.50	1.77	39.8	1,588.90	8.7	30.7
C	3	28,989.40	4,763.40	6.0	5,096.20	0.45	36.5	16,348.40	0.6	70.4
D	4	23,596.60	379.70	27.0	852.20	0.74	71.9	685.60	10.6	18.6
E	5	11,914.90	197.80	17.8	1,353.80	0.89	71.8	761.70	1.7	10.2
F	6	11,708.80	590.10	31.8	270.80	0.76	85.8	1,265.90	0.6	12.6
G	7	11,484.40	2,042.60	32.4	814.60	0.92	55.9	3,876.50	0.9	16.8
H	8	11,099.40	8.10	0.2	3,439.50	1.60	48.9	724.20	1.7	9.5
I	9	10,416.00	-357.80	-8.0	-780.20	0.43	74.6	1,138.30	2.1	17.2
J	10	9,617.20	514.40	12.0	-640.10	0.62	61.6	756.10	0.2	9.3
K	11	7,193.80	398.00	5.5	388.70	0.78	52.2	306.80	0.1	6.2
L	12	7,052.10	510.90	21.1	241.80	1.42	60.2	-33.70	0.1	8.1
M	13	6,584.70	5,142.20	29.7	-2,237.70	0.46	41.4	1,972.60	0.3	5.7
N	14	6,503.40	2,179.40	9.3	296.10	0.77	22.6	1,962.80	1.2	9.5
O	15	6,448.60	-468.60	-6.9	-325.30	0.46	50.0	511.80	0.8	12.8
P	16	5,761.80	-313.00	-3.8	557.40	0.58	43.6	37.60	0.4	11.4
Q	17	5,420.20	935.60	15.3	-209.20	0.28	54.5	1,767.30	0.4	6.3
R	18	5,371.20	-205.60	-4.6	1,314.30	0.31	80.9	1,403.70	0.3	10.6
S	19	5,164.10	341.50	10.2	1,684.90	1.06	60.2	820.00	0.3	5.6
T	20	5,027.30	52.80	2.2	250.10	0.67	59.8	1,242.40	0.9	2.1

(1) value of gross sales, from which returns, discounts and sales taxes were deducted
(2) nominal result for the period (not considering inflation), deducting income taxes and social contribution and adjusting interest on net equity
(3) main indicator of business excellence (return on investment): = profit (net profit, legal, adjusted/net equity, legal, adjusted) x 100
(4) short-term resources available for financing company's activities
(5) = (working assets + long-term receivables) / total enforceable; less than 1 implies solvency will depend on future profits, debt negotiation or sale of assets.
(6) business risk; = [(working liabilities + non-working liabilities)/adjusted total assets] x 100
(7) cash flow generated by business activity (profit before deducting interest, taxes on profits, depreciation and amortisation)
(8) net sales/mean number of employees; productivity indicator
(9) (net sales/total net sales in the economy sector) x 100

The attributes were considered according to their nature, which we denote by *gain* when increasing values mean greater advantage, *cost* when increasing values mean smaller advantage, and *none* otherwise. In this case, following a suggestion given by jMAF, the following "reducts" were chosen: {profitability, equity, ebitda, sales per employee, market share}. There was again an indication of reclassification of companies F and G as to the ordering by net sales, as shown in Table V.

```
**ATTRIBUTES
+ company: (nominal), none, description
- ranking: (integer), none
+ net_sale: (continuous), gain, decision
- net_profit: (continuous), gain
+ profitability: (continuous), gain
- net_ working _capital: (continuous), gain
+ equity: (continuous), gain
- debt: (continuous), cost
+ ebitda: (continuous), gain
+ sale_per_employee: (continuous), gain
+ market_share: (continuous), gain
decision: net_sale
```

Because the criterion chosen to order the largest companies was “net sales”, it was used as a “decision” criterion in this analysis. This kept a “1:1” relationship between the order and net sales. For classification, we used DRSA with the VC-DRSA method, which permitted reclassifying objects that violated the Dominance principle (companies F and G), as illustrated in Figure 1.

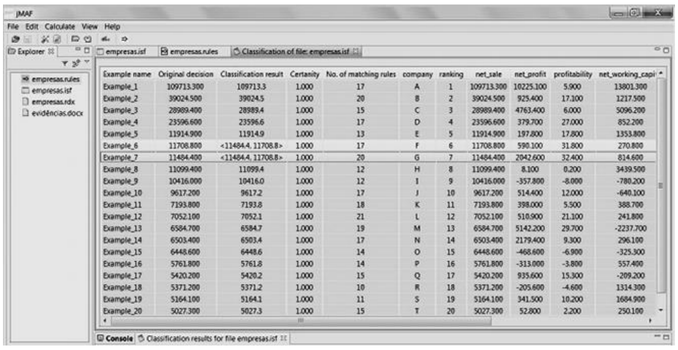

Fig. 1. Analysis with jMAF: reclassification results.

This analysis is shown in Table V.

TABLE V. COMPANIES F AND G VIOLATE THE DOMINANCE PRINCIPLE REGARDING THE CLASS “NET SALES”

Company	Ranking	Net Sales (US$ million)	Legal Net Profit (US$ million)	Legal Profitability (%)	Net working capital (US$ million)	General liquidity (index number)	General debt (%)	EBITDA (US$ million)	Net sales/employee (US$ million)	Market share (%)
A	1	109,713.30	10,225.10	5.9	13,801.30	0.57	38.9	22,774.70	1.8	46.5
B	2	39,024.50	925.40	17.1	1,217.50	1.77	39.8	1,588.90	8.7	30.7
C	3	28,989.40	4,763.40	6.0	5,096.20	0.45	36.5	16,348.40	0.6	70.4
D	4	23,596.60	379.70	27.0	852.20	0.74	71.9	685.60	10.6	18.6
E	5	11,914.90	197.80	17.8	1,353.80	0.89	71.8	761.70	1.7	10.2
F	6	11,708.80	590.10	31.8	270.80	0.76	85.8	1,265.90	0.6	12.6
G	7	11,484.40	2,042.60	32.4	814.60	0.92	55.9	3,876.50	0.9	16.8
H	8	11,099.40	8.10	0.2	3,439.50	1.60	48.9	724.20	1.7	9.5
I	9	10,416.00	-357.80	-8.0	-780.20	0.43	74.6	1,138.30	2.1	17.2
J	10	9,617.20	514.40	12.0	-640.10	0.62	61.6	756.10	0.2	9.3
K	11	7,193.80	398.00	5.5	388.70	0.78	52.2	306.80	0.1	6.2
L	12	7,052.10	510.90	21.1	241.80	1.42	60.2	-33.70	0.1	8.1
M	13	6,584.70	5,142.20	29.7	-2,237.70	0.46	41.4	1,972.60	0.3	5.7
N	14	6,503.40	2,179.40	9.3	296.10	0.77	22.6	1,962.80	1.2	9.5
O	15	6,448.60	-468.60	-6.9	-325.30	0.46	50.0	511.80	0.8	12.8
P	16	5,761.80	-313.00	-3.8	557.40	0.58	43.6	37.60	0.4	11.4
Q	17	5,420.20	935.60	15.3	-209.20	0.28	54.5	1,767.30	0.4	6.3
R	18	5,371.20	-205.60	-4.6	1,314.30	0.31	80.9	1,403.70	0.3	10.6
S	19	5,164.10	341.50	10.2	1,684.90	1.06	60.2	820.00	0.3	5.6
T	20	5,027.30	52.80	2.2	250.10	0.67	59.8	1,242.40	0.9	2.1

Considering that companies are ordered by net sales and that company G is superior to company F in all indicators, then by the Dominance principle, company G should be placed in the same class of net sales as company F, or in some class above. The jMAF software application inferred the following rules:

[RULES]
#Certain at least rules
1: (ebitda >= 22774.7) => (net_sale >= 109713.3) |CERTAIN, AT_LEAST, 109713.3|
2: (equity >= 1.77) => (net_sale >= 39024.5) |CERTAIN, AT_LEAST, 39024.5|
3: (market_share >= 30.7) => (net_sale >= 28989.4) |CERTAIN, AT_LEAST, 28989.4|
4: (market_share >= 18.6) => (net_sale >= 23596.6) |CERTAIN, AT_LEAST, 23596.6|
5: (profitability >= 17.8) & (sale_per_employee >= 1.7) => (net_sale >= 11914.9) |CERTAIN, AT_LEAST, 11914.9|
6: (profitability >= 17.8) & (market_share >= 10.2) => (net_sale >= 11484.4) |CERTAIN, AT_LEAST, 11484.4|
7: (profitability >= 31.8) => (net_sale >= 11099.4) |CERTAIN, AT_LEAST, 11099.4|
8: (equity >= 0.89) & (sale_per_employee >= 1.7) => (net_sale >= 11099.4) |CERTAIN, AT_LEAST, 11099.4|
9: (sale_per_employee >= 1.7) => (net_sale >= 10416.0) |CERTAIN, AT_LEAST, 10416.0|
10: (market_share >= 16.8) => (net_sale >= 10416.0) |CERTAIN, AT_LEAST, 10416.0|
11: (profitability >= 12.0) & (market_share >= 9.3) => (net_sale >= 9617.2) |CERTAIN, AT_LEAST, 9617.2|
12: (equity >= 0.78) & (ebitda >= 306.8) & (market_share >= 6.2) => (net_sale >= 7193.8) |CERTAIN, AT_LEAST, 7193.8|
13: (equity >= 0.78) & (market_share >= 6.2) => (net_sale >= 7052.1) |CERTAIN, AT_LEAST, 7052.1|
14: (profitability >= 12.0) & (equity >= 0.62) => (net_sale >= 7052.1) |CERTAIN, AT_LEAST, 7052.1|
15: (profitability >= 17.1) => (net_sale >= 6584.7) |CERTAIN, AT_LEAST, 6584.7|
16: (ebitda >= 16348.4) => (net_sale >= 6584.7) |CERTAIN, AT_LEAST, 6584.7|
17: (sale_per_employee >= 1.2) => (net_sale >= 6503.4) |CERTAIN, AT_LEAST, 6503.4|
18: (market_share >= 12.6) => (net_sale >= 6448.6) |CERTAIN, AT_LEAST, 6448.6|
19: (market_share >= 11.4) => (net_sale >= 5761.8) |CERTAIN, AT_LEAST, 5761.8|
20: (equity >= 0.62) & (market_share >= 9.3) => (net_sale >= 5761.8) |CERTAIN, AT_LEAST, 5761.8|
21: (profitability >= 12.0) => (net_sale >= 5420.2) |CERTAIN, AT_LEAST, 5420.2|
22: (profitability >= 5.5) & (market_share >= 6.2) => (net_sale >= 5420.2) |CERTAIN, AT_LEAST, 5420.2|
23: (market_share >= 5.7) => (net_sale >= 5371.2) |CERTAIN, AT_LEAST, 5371.2|
24: (market_share >= 5.6) => (net_sale >= 5164.1) |CERTAIN, AT_LEAST, 5164.1|

The Table VI summarizes the previous rules (support, strength and coverage factor for each rule). Considering the strength number, for example, 95% of the 20 largest companies satisfy the rule “*if* market share equal to or greater than 5.6% *then* net sale equal to or greater US$ 5,164.10 million” (rule 24).

From the Coverage factor, we get the following characterization of companies (“inverse algorithm”):

• 66% with net sale equal to or greater US$ 10,416.00 million, have sale per employee equal to or greater US$ 1.7 million/employee (rule 9);

• 66% with net sale equal to or greater US$ 10,416.00 million, have market share equal to or greater 16.8% (rule 10);

• 70.5% with net sale equal to or greater US$ 5,420.20 million, have profitability equal to or greater 5.5% and market share equal to or greater 6.2% (rule 22);

• 100% with net sale equal to or greater US$ 5,371.20 million, have market share equal to or greater 5.7% (rule 23).

Hence we have about 20 largest companies:

• With net sale equal to or greater US$ 10,416.00 million, are most probably sale per employee equal to or greater US$ 1.7 million/employee and market share equal to or greater 16.8%;

• With net sale equal to or greater US$ 5,420.20 million, are most probably profitability equal to or greater 5.5% and market share equal to or greater 6.2%;

• With net sale equal to or greater US$ 5,371.20 million, are most probably market share equal to or greater 5.7%.

TABLE VI. RULE, SUPPORT, STRENGTH AND COVERAGE FACTOR

Rule	support	Strength	Coverage Factor
1	1	0.05	1.0
2	1	0.05	0.5
3	3	0.15	1.0
4	4	0.2	1.0
5	2	0.1	0.4
6	4	0.2	0.57
7	2	0.1	0.25
8	3	0.15	0.375
9	6	0.3	0.66
10	6	0.3	0.66
11	6	0.3	0.6
12	5	0.25	0.454
13	6	0.3	0.5
14	7	0.35	0.583
15	7	0.35	0.583
16	2	0.1	0.153
17	7	0.35	0.5
18	8	0.4	0.533
19	9	0.45	0.562
20	8	0.4	0.5
21	9	0.45	0.529
22	12	0.6	0.705
23	18	0.9	1.0
24	19	0.95	1.0

VI. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

It is hard to obtain companies' non-financial indicators because, traditionally, only a company's financial health is analysed. However, in the face of the need to establish long-term strategies, surveying non-financial indicators may signal where it is possible to aggregate value to the company's business relative to changes in the competition, markets and technology.

Furthermore, financial and non-financial metrics associated with the use of "balanced scorecards" allow one to connect current actions to future goals by providing a reference framework for the implementation of a long-term strategy [20]. The present study shows that the analysis can be extended by taking other financial and non-financial indicators into account to infer patterns that are often hidden in the data. Despite the relatively small set of 20 largest companies in a universe of the 500 largest by net sales, this study showed the importance of broadening the analysis of business indicators when the goals are, for example, to produce a more critical view for investment decisions, to identify the performance of probable competitors and to ascertain market trends. As the study was based on a Multi-Criteria analysis, Rough Set Theory and the Dominance principle were used to identify and treat data that, in some cases, proved to be inconsistent or to create paradoxes when objects' classifications did not fit into some reference standard ("decision class") – e.g., the case of companies F and G, see Table V. In this case, for an investment decision in a company, for example, company F would most likely not have so much of an advantage (in net sales) relative to company G. Thus, the analysis of these 20 largest companies using the jMAF software application [5] allowed us to infer hidden patterns when a given set of condition attributes was selected as the "reduct", i.e., equivalent to the set of all attributes (attribute "net sales" as the decision criterion).

In statistical data analysis based on Bayes' Theorem, we assume that prior probability about some parameters without knowledge about the data is given. The posterior probability is computed next, which tells us what can be said about prior probability in view of the data. In the Rough Set approach the meaning of Bayes' Theorem is unlike. It reveals some relationships in the database, without referring to prior and posterior probabilities, and it can be used to reason about data in terms of approximate (rough) implications. It identifies probabilistic relationship between conditions and decisions in decision algorithms and can be used to give explanation (reasons) for decisions. Let us also stress that Bayes' theorem in the rough set approach has a new mathematical form based on strength of decision rules, which simplifies essentially computations and gives a new look on the theorem [14], [15].

Regarding the ordering of companies by net sales as suggested in the cited study, it is possible to address this in future studies with the analysis results shown in Table V by processing these with the jRank application (*Ranking using Dominance-based Rough Set Approach*) [21], which also applies concepts of Rough Set Theory and the Dominance principle to establish a classification order. Furthermore, it is also possible to perform a broader analysis with the entire 500 largest companies to infer patterns of these companies' financial and non-financial indicators.

By the way, the attempt in unifying logic and probability to logical sentences is shown in [22]. And for data mining applications for example, the acquisition of probabilistic, rather than deterministic, predictive models is of primary importance [23].

ACKNOWLEDGMENT

The work leading up to this paper was partially supported by CNPq/Brazil, through Research Project n. 305732/2012-9. The authors thank the Institute of Computer Science of the Poznan University of Technology, Poland, for allowing them to use the jMAF software in their research.

REFERENCES

[1] Exame, "Exame melhores e maiores 2013" [Exame Best and Biggest 2013].São Paulo: Abril. 1044E. ed., 2013.

[2] L.A. Zadeh, "Fuzzy sets", Information and Control 8, 338–353, 1965.

[3] L.F.A.M. Gomes and C.F.S. Gomes, “Tomada de decisão gerencial: enfoque multicritério” [Management decision-making: multi-criterion approach], 5. ed. São Paulo: Atlas, 2014.

[4] Z. Pawlak, J. Grzymala-Busse, R. Slowinski and W. Ziarko, “Rough sets”, Communications of the ACM, 38, 11, 89-95, 1995.

[5] J. Blaszczynski, S. Greco, B. Matarazzo, R. Slowinski and M. Szeląg, “jMAF”, Institute of Computer Science. Poznan University of Technology, Poland. Available from: http://www.cs.put.poznan.pl/jblaszczynski/Site/jRS.html. 2012.

[6] Z. Pawlak, “Rough sets”, Int. J. Comput. Inf. Sci. 11, 341-356, 1982.

[7] Z. Pawlak, “Rough sets. Theoretical aspects of reasoning about data”, Kluwer Academic Publishers, Dordrecht, 1991.

[8] Z. Pawlak and R. Slowinski, “Rough set approach to multi-attribute decision analysis”, European Journal of Operational Research, Invited Review, 72, 443-459, 1994.

[9] Z. Pawlak, “Rough sets and decision analysis”, Information Systems & Operational Research. 38, 3, 132-144, 2000.

[10] A. Skowron, “Rough sets in KDD”, In: Z. Shi, B. Faltings and M. Musen, (eds.), Proc. 16th IFIP World computer congress, Beijing, China. Publishing House of Electronic Industry, 2000.

[11] Y. Yao and Y. Zhao, “Conflict analysis based on discernibility and indiscernibility”, Proceedings of IEEE Symposium on Foundations of Computational Intelligence, 302-307, 2007.

[12] D. Chen, S. Zhao, L. Zhang, Y. Yang and X. Zhang, “Sample pair selection for attribute reduction with rough set”, IEEE Transactions on Knowledge and Data Engineering, 24, 11, 2080-2093, 2012.

[13] A.B.G. Couto and L.F.A.M. Gomes, “A tomada de decisão em recursos humanos com dados replicados e inconsistentes: uma aplicação da teoria dos conjuntos aproximativos” [Decision making in human resources with replicated and inconsistent data: an application of rough set theory], Pesquisa Operacional, 30, 3, 657-686, 2010.

[14] Z. Pawlak, “Rough sets and decision analysis”, INFOR: Information Systems and Operational Research 38 (2), 132-144, 2000.

[15] Z. Pawlak, “Rough sets, decision algorithms and Bayes’ theorem”, European Journal of Operational Research, 136, 181-189, 2002.

[16] R. Slowinski, S. Greco and B. Matarazzo, “Rough set and rule based multicriteria decision aiding”, Pesquisa Operacional, 32, 2, 213-269, 2012.

[17] W. Kotlowski and R. Slowinski, “On nonparametric ordinal classification with monotonicity constraints”, IEEE Transactions on Knowledge and Data Engineering, 25, 11, 2576-2589, 2013.

[18] E.S.U. Pace, L.F.C. Basso and M.A. Silva, “Indicadores de desempenho como direcionadores de valor” [Performance indicators as value drivers], RAC, 7, 1, 37-65, 2003.

[19] A.A. Bandeira, “Indicadores de desempenho: instrumentos à produtividade organizacional” [Performance indicators: tools for organisational productivity], Rio de Janeiro: Qualitymark, 2009.

[20] R.S. Kaplan and D.P. Norton, “Using the balanced scorecard as a strategic management system”, Harvard Business Review, 150-161, 2007.

[21] M. Szelag, R. Slowinski, S. Greco, J. Blaszczynski and S. Wilk, “jRank”, Institute of Computer Science, Poznan University of Technology, Poland. Available from: http://www.cs.put.poznan.pl/mszelag/Software/jRank/jRank.html.2013.

[22] S. Russell, “Recent developments in unifying logic and probability”, Communications of the ACM, 58, 7, 88-97, July 2015.

[23] D. Slezak, W. Ziarko, “The investigation of the Bayesian rough set model”, International Journal of Approximate Reasoning, 40, 81-91, 2005.