

Data Mining Techniques and the Decision Making Process in the Bulgarian Public Administration

Martin Gramatikov¹

Introduction to the topic

The current hype in the new public management literature and re-engineering movement has led to overuse of terms such as customer centric governance, public entrepreneurship electronic governance and so forth. Many of these concepts conceal large misconceptions and mechanic transfer of culture; rules and procedures from the private sector management to public governance domain. It is not our intention in that article to revive the long lasting theoretical argument between supporters and opponents of the New Public Management (NPM). What we are going to analyze here are the theoretical and practical implications of the use of data mining techniques and their applications in the Bulgarian public sector. In that article we will try to deliberate on the various aspects of the data mining and their importance for the development of effective, efficient and flexible public administration. As unit of analysis we determine the managerial decision making process as dependent variable and the data, information and knowledge gathering and analysis procedures and processes as independent variable which are deemed to strongly affect the qualitative and quantitative properties of the administrative decisions.

At micro level we will try to explain and predict the potential of the data mining techniques for leveraging the level of rationality in the decisions made in the public sector. From that particular perspective we will revise some parts of the theory of bounded rationality in line with the empirical evidences that the data mining techniques validate.

Overview of Data Mining

Data mining is branch of the applied informatics, which allows us to sift through large amounts of structured or unstructured data in attempt to find hidden patterns and/or rules. As a technique for data exploration data mining has its roots in the artificial intelligence research from the 80s and early 90s. Fayyad et al. (1996) defines the data mining methods as tools for searching databases with special algorithms to identify general patterns which can be used in the classification of the individual observations and making predictions thereof. According to Weiss and Indurkha, (1998) data mining is the search for valuable information in large volumes of data. Haskett (2000) gives narrower definition of data mining as “set of techniques used in an automated approach to exhaustively explore and surface complex relationships in very large datasets”. According to Hand (1998) “data mining is “... the process of secondary analysis of large databases aimed at finding unsuspected relationships which are of interest or value to the database owners.”

Knowledge discovery is often used as synonym of data mining because the exploration of raw data is intended to increase the substantial and procedural knowledge on the targeted processes. The transformation of data into knowledge means that the information is being derived from the collected facts and put into context. Haskett (2000) uses three-tier architecture to describe the facets of the data mining concept – applications, approaches and

¹ Assistant Professor, Department of Public Administration, University of Sofia, Bulgaria
martin@gramatikov.com

algorithms and models. Particularly data mining techniques extract from the raw data (information) potentially useful information such as: knowledge rules, patterns, regularities and other trends hidden in the data. In the process of transformation of information into knowledge several applications of the data mining are involved: data reduction, clustering, data classification, pattern discovering etc. In order to complete these tasks the contemporary data mining packages offer techniques such as: neural networks, inductive learning, decision trees, cluster analysis, link analysis, genetic algorithms, visualization etc.

At first glance data mining overlaps with the well known statistical methods and to a large extent that conceptual blurring is true. Most of the data mining techniques have their roots in the well developed statistical theories such as the Central Limit Theorem, General Linear Model etc. What is the most distinctive difference between the statistical methods and the data mining is the ultimate focus of the two tools. While in the statistics we are looking either for describing parameters of data or making inferences through statistics within confidence intervals in the data mining domain we are interested by trends, rules and patterns. In the latter we strive to find the general tendency hidden in the bulk of data while sacrificing the importance of the particular observation. Hand (1998) describes the difference between the statistical analysis and data mining as one between inductive and hypothetico-deductive approaches. While in the statistics we collect data with predefined set of questions in mind in the data mining the hypotheses emerge in the process of mining for hidden relations, rules and patterns.

Given the above mentioned difference we can say that the data mining is born out of the statistical theory but with its own goal: to provide a powerful tool for processing large amounts of data. At the beginning such techniques had been used in the business management – mainly in the marketing, sales and forecasting of the financial markets. Nowadays data mining is involved in many parts of the business process with most visible impact on the ‘customer relations’ function. Other areas are also opening up to the data mining: engineering, manufacturing, control and so forth. (Tam and Kiang, 1992; Chu & Widjaja, 1994). Functionally the data mining applications can perform any of the enlisted tasks or combinations thereof:

- Classify data into pre-existing categories
- Cluster data by mapping them into categories created during data analysis and determined by the data
- Provide a summary of the data in a sense that the raw data are not
- Describe dependencies between variables
- Find links between data fields
- Use regression to predict future values of data
- Model sequential patterns in the data that may indicate revealing trends

It is not hard to see the potential that the data mining offers to the public sector. Implementing public policies, delivering public services or governing public programs the public administrator has to make decisions on how, where and when to invest the public resources. As well as in the private sector the decision making in the public administration should be based on accurate and timely information. The data mining techniques offer to the decision maker from the public sector opportunities to optimize her decisions based on general trends extracted from historical data. With the knowledge that can be extracted from the data the public organization can level up its knowledge accumulation strategies and steps. In contrast with the procedural knowledge which is created and maintained at personal level the substantial knowledge which lies hidden in the data sets resides in the organizational domain. Hence the knowledge that can be derived with the data mining could serve first as a tool for better governance and second as a mean for sustaining the organizational knowledge. As Sauter (1999) asserts the decision support systems which embed data mining, virtual experience and other models can give the managers the background in processes and settings which they otherwise should acquire through longevity in the organizations. The thesis of Sauter is that in the contemporary organizations the turnover is much more dynamic than in the ‘traditional’ organizations. This causes inevitable pitfalls in the decision making process due to the lack of experience in particular problem-solving situations. Hence the decision making system and the data mining techniques in particular are viable solutions for bridging the gaps in the organizational knowledge acquisition and utilization processes.

Another argument for the case of using data mining in the public organizations is the knowledge intensiveness of the contemporary public management. The concepts of Information society, E-governance and the like are largely based on the assumption that the information is turning into the most valuable and therefore scarce resource in the contemporary society. It is not hard to see the implication of that assumption: those who know how to manage and use the available information will reap the benefits of the ICT and will enjoy the lighter side of the digital divide which is in its formation phases (Gramatikov, 2003). These trends are especially valid for the public administration organizations. The whole process of formulating, analyzing, implementing and evaluating public policies is about information processing. As the theory of bounded rationality puts it, the theory of the public administration can be oriented towards the process of decision making which is largely function of the information availability and processing capacity. According to Herbert Simon (1958, 1982, 1997) the individuals in organizational environment tend to act rationally, but are doomed to do so only to a certain extent. Two major limitations are to be blamed for the inevitability of the boundedness of the administrative decision with regard to its rationality: information inabilities and limitation of the processing powers of the administrators who have to assess the available information in a way which will allow them to set a comprehensive list of alternatives to the given problem and will identify the most rationale alternative. Newmann and Morgenstern (1964) add to these two factors a third one: the insecurity of the surrounding environment. To wrap up: the theory of bounded rationality defines the Administrative Man who, in contrast to the Economical Man from the traditional macro-economy theories, acts in situation that is far away from perfect. In that model of bounded rationality Simon describes the behavior of the administrator when he makes decisions as *satisficing*. This means that the Administrative Man takes into considerations only the information he regards as pertinent to the particular problem solving based on his past experience, skills and abilities, political factors and the like.

The model of bounded rationality predicts that large amounts of the relevant information will be left unused in the process of the administrative decision making and that plagues the rationality of the decision and hence the whole process of governance. Usually the perception is that this exemplifies the redundancy and inefficiency of the public administration. It is our thesis here that the information limits in the administrative decision making are not necessary a weak link in the process of implementation of the public policies. The main argument we get from the principle of parsimony which underlies many of the data mining techniques. Here we can quote Albert Einstein who said that “Everything should be made as simple as possible, but not simpler”. In the data mining as in the administrative decision making we need to process the available information in such a way as to find the proper balance between the existing information premises and the demand for outputs of the information processing process. In the former case the output is in the form of preferred alternative which is believed to bridge the gap that was identified and necessitated decision. In the latter case where knowledge discovery is required the output shapes as identified trends, rules and patterns, hidden in the mined data sets.

The principle of parsimony, that we deem to be common to both decision making and data mining, is in line with the theory of bounded rationality that professes that the administrator inevitably neglects part of the available information in order to expedite the decision making process and to actually make it possible. Same in the data mining: the large warehouses allow for running of precise statistical tools that will return a good deal of descriptive and inferential statistics. The essence here however is not in the robust pursue of the truth about the phenomena described with the variables recorded in the dataset. As in the administrative decision making the objective is to achieve the most adequate, effective, efficient and legitimate decision given the particular circumstances. Although in the public sector the decision has revealed much more attributes such as transparency, political expedience, compliance with the organizational vision etc. we think that the bottom line is the same: “getting more for less” if the New Public Management lingvo is to be used. Three of the most notable consequences of the application of the parsimony principle in the decision making and the data mining are:

- the stakeholders will be more prone to understanding the results
- less resources will be needed to monitor and evaluate the outcome in the future

- it is easier to replicate and cross-validate decisions or derived rules when only the most significant elements of the information environment is considered

In the data mining the quest for the simplest model with predictive potential is much more formalized than in the decision making. As it was already mentioned the administrator limits overcome the limits in the information premises through reliance to implicit knowledge embedded in experience, attitudes, sensing, prior knowledge, affection and other differing from the pure rational model means of deciding. In the data mining techniques we have to address the problem of deciding on which of the numerous variables stored in the warehouse we have to rely in order to gain the required knowledge at the lowest possible price. Four such models exist:

- forward entry method
- backward removal method
- forward stepwise method
- backward stepwise method

Each model provides unique features in its approach to enter the most suitable variables in the model that will be used for estimation and predictions. The techniques can be found in the various styles of decision making in the public organizations and further research on the artificial intelligence and administrative decision making can come up with important conclusions.

Another important comparison between the methods for assuring parsimony in the data mining techniques and the decision making as defined by the bounded rationality theory is the criteria according to which information should be considered as important for the particular situation. Let's take the case of a city planner who has to make decision on the possible alternatives of a brown field in the center of the city. Theoretically the administrator has enormous amount of information sources, types and forms to choose from in order to make the right decision. She can look back and analyze time series information in order to extrapolate and make predictions, survey the local residents' attitudes, analyze the public policies and the intentions of the key policy makers, take comparative approach and study the experience of comparable cities and countries. Having all that information will probably increase the depth and number of the thinkable alternatives and will allow for comprehensive analysis on which of the alternatives to choose. In most cases however the administrator would not have neither the required resources to provide all those resources nor the necessary time.

In the data mining techniques there are several frequently used criteria that solve the illustrated problem. The need for such criteria stems from the inability of the selection models to effectively address two problems of the selected models: usually the procedures tend to select too many variables and do not provide an unconditional "best" model. Unconditional "best model" is that model that best fits the concept that we want to investigate and collect knowledge about. The problem with the conditionality of the model comes from the fact that at each selection step the model contains "best" variables as predictors only on the condition that no other variables of interest exist. Alternative approach to the selection models is the use of thresholds for acceptance of variables in the model, only if they satisfy some criterion. Three such criteria popular in the data mining applications are the Akaike Information Criterion, Schwarz's Bayesian Criterion and Bayesian Information Criterion. These criteria provide sophisticated methods for inclusion of relevant information to the estimation and prediction model. Likewise in the decision making the selection of these portions of the information which prove to contain the highest contextual value here we also look for balance between the abundance of information and the need of reliable and efficient results.

Without the help of powerful hardware and software tools these methods for selection of the model which should be used as base for knowledge discovery would had been practically impossible. Today however it is not difficult for the public organizations to deploy knowledge discovery applications. The Achilles heel of the use of data warehousing and data mining applications in the public sector however is not in the potential of the technology but in the highly irrational process of decision making in the Bulgarian public administration. Apart of the discussed limits on the rationality the culture of many Bulgarian public organizations neglects the information processing as major component of the decision making process. It is difficult to say which factors

contribute most to that specific culture: historical background, political process, human capacity etc. The fact is that the decisions are predominantly based on information, derived from discrete data and very few strategic or operational decisions are based on information processed from continuous data.

In the recent years more and more focus is being diverted from the traditional public policy implementation towards the optimization of the public governance through management of the relations with the customers/citizens, streamlining of the organizational communications and processes, gathering of information on transactions etc. The classical bounded rationality theory tells us that the rationality of the decision is function of the availability of information and the capacity to process that information. In the case of even low scale services there are huge amount of factors, which influence the probability for success, or failure of her undertakings. Indeed the public administrator can gain information through qualitative or quantitative research methods but first these methods can hardly work in a complex and pressing environment and second it is very difficult to control for the various factors, which reflect on the process of governance. The complex structure, processes and environment oftentimes reveal existence of non-linear models which cannot be satisfactorily analyzed with the traditional statistical methods. Last but not least we have to count for the general lack of capacity to collect and analytically process data in the Bulgarian public sector. That trend can be attributed to several factors which we label: historical legacy, organizational inertia and human capacity.

The appeal of data mining in the public sector stems from the intensity of the digital data that is being collected in the public domain. Dating back to the 80s we can see policies and projects in the Bulgarian public administration which resulted in digital warehouses for various types of public information. After the proliferation of text storage systems different systems containing transactional, spatial, visual and mixed information had emerged. With the advancement of the technologies the public administration is catching up with the private sector and nowadays at both political and organizational level there are large-scale projects for building or upgrading of the IS of the public administration. Collecting large amounts of information and missing the analytical interpretation thereof puts the “data rich/information poor” problem ahead of the public administration.

The data mining techniques have the potential to bridge the gap between the two colliding expectations of the contemporary public administration. Effectiveness, efficiency, speed, flexibility on one hand are at continuous stride with fairness, equality, rule of law and legitimacy in the public administration. To give an example it will be very “managerial” to implement a public health program based on the findings that such program will yield positive monetary results. However the budget spending can be achieved at the cost of people left without accessible health care simply because they are not profitable. Clearly here the effectiveness is at odd with the social solidarity and fairness of the governance.

What data mining can do in the public administration is to provide more descriptive and predictive information on which the decision makers could rely and trust. It has been asserted by Herbert Simon that the administrative decision is boundedly rational and can never tally with the complete rationality. The data mining techniques however could mine the historical data and extract hidden predictive information which could direct the public administrator to the most effective and efficient solution in pursuing the fair, transparent, economic and just implementation of the public policies.

Of course data mining cannot be panacea for any problem of the public governance. For instance the public choice theory asserts that the behavior of the administrator could be irrational from perspective of the organizational domain but more rational from personal point of view. Moreover the human, technical and financial capacity development and maintenance are problems that the Bulgarian public sector has to face. We cannot test the descriptive and predictive power of the social choice literature now but want to bring in the context the existence of factors in the decision making which cannot be easily fitted in the perfect framework according to which the data mining techniques should leverage the level of rationality. For instance the decision making in the public administration is affected by factors and relationships which neither show linear patterns nor avail themselves to recording and subsequent exploration. The process of policy making, the

political control over the administration, the influence of international-level factors are all examples of phenomena that cannot be operationalized and meaningfully coded into tabular form.

However data mining can be a valuable tool in the hands of the decision makers aiming to predict the internal and external environment, adjust accordingly and hence to make more rational decisions. There are many domains in the public sector that avail themselves to quantification and measuring in one or another form. In the chapter of the practical applications of data mining we list some of those areas. Positively there are many other sectors of the governance where the appropriate collection and processing of data could provide invaluable tools in the decision-making quiver of the administrator. The first step to achieve this is to embrace the idea that the good decision is underlined by abundant and reliable data. Without the raw data no analyses and consequently decisions can be made and this is where the data mining stumbles now in the public administration systems.

The literature shows that the data mining is being widely used in the sector of law enforcement. Dube (1999) explores the utilization of data mining in the search of long-time missing children and the habits of the perpetrators. Data mining techniques are used in detection of network intrusions (Zhu et al., 2001). Recent cases of deface and denial of service attacks on various network points of the public administration in Bulgaria make the hypothesis real and comprehensible.

Use of Data Mining Techniques to support administrative decision making in Bulgaria

In line with the research interest we tried to find out to what extent the data mining is possible, affordable and legitimate in the Bulgarian Public sector? What is the downturn of the data mining – lack of data or lack of ability to generate useful information from it? The subjects of analysis were chosen on grounds of convenience and has nothing to do with inferential potential or ability to generalize. Even on the contrary: the difficulty to pinpoint public organizations employing data mining is eloquent evidence for the technological lag of the Bulgarian public administration. The problem can also be traced at political level where the public policies in Bulgaria relevant to the E-government development suffer from two deadly sins. The first is the inclination towards complete top-down approach to application of the ICT in the process of governance. Eliminating the role of the citizen, IT industry and public administrators the policies put the initiative in the hands of the political appointees. That supply-driven approach accounts for a good deal of the failed IT projects in the public administration. The second sin is the predominant emphasize of the public programs and projects on the citizen-administration scale and not on the dimension of the administrative decision making. For instance the recently published Strategy for development of E-Government in Bulgaria explicitly states that the main priority is the transfer of public services from paper-based mode to on-line Internet based services. We agree with the importance of the service delivery dimension and its significance for the citizens' trust in the public administration. However the public services are only one component of the public governance which is based on continuous decisions made by administrators at different levels of the hierarchy. These decisions can be significantly improved through deployment of ICT systems in the process of knowledge creation and management.

The empirical survey of the data mining applications in the Bulgarian public administration proved the initial assumption that it will face numerous problems. Most importantly it is the scarcity of technological, financial, organizational and human capacity that limits the proliferation of data mining applications in the public sector. Where we expected to see heavy involvement of the knowledge discovery methods we were either denied access or disappointed by a lack of such technologies. For instance the numerous attempts to get a glimpse on the possible applications of data mining techniques in the information systems of the Ministry of Interior were completely unsuccessful because of the “information security and sensitivity” issues. The implications that we can make out of that experience is that the methods of data processing in some of the public administration are far away from transparent. Of course for some institutions there is sound explanation for that trend but for others the “black box” approach is more a remnant from the administrative order and culture that is boldly denounced at political level but still predominates in some of the public organizations.

It was interesting to discover that the data mining techniques are popular in Bulgaria as “profiling”. We discovered two cases of organizations that claim to “profile” the data they collect in order to discover important patterns. The two organizations are kind of law-enforcement agencies: the customs administration and the hazard administration. Both authorities are developing the so called risk profiles in their respective areas. The risk profiles are nothing else than mining of historical data sets with variables that store different properties of persons, transactions, tangible or intangible goods etc. Target variable here is the dichotomous distribution of the probability of seeing adverse outcome - e.g. illegal export or trafficking in the case of the custom administration or tax evasion in the case of the hazard administration. Based on the records of previous cases the software for data mining creates profiles of cases in which the odds to encounter the unwanted outcomes are higher than the other cases. The most widespread approach for creating profiles is the clustering but other approaches are suitable as well: logistic regression, artificial neural networks etc.

The practical implications of the utilization of data mining techniques in these two cases are the improved predictive potentials of the administrative decisions. In the case of the custom administration this means that instead of checking randomly citizens and vehicles at the border the custom officers can concentrate on the subjects who pertain to the group with higher risk of conducting illegal activities. Thus the administrative decision making can guide the street-level administrators toward more effective and efficient behavior. Imagining the situation without the use of data mining applications we can reasonably expect that the non-IT methods for knowledge discovery such as experience accumulation, sensing, affection etc. would yield fewer results. On the other hand the hazard administration through knowledge discovery and target its efforts on the persons and legal entities that are more likely to circumvent the legal order. These two applications of the data mining techniques clearly show the potential of the knowledge discovery in the strive for improvement of the efficiency of the public administration and management.

It is not surprising that two of the three cases where we found data mining methods pertained to the law-enforcement sector. There are two factors that contribute to the suitability of the data mining in that policy area. First, there is well definable target variable in these sectors: inclination to felonies, crime rate, tax evasion and so forth. It is easy to specify conceptually clear variable and to measure it unambiguously. Second, there is long history of data collection in the law-enforcement sector. With regard to the public sector organizations it is undeniable that the law-enforcement (police, criminal investigation, intelligence, customs control, etc) are prominent with their data-collection and processing capabilities.

The third case is from completely different area of public governance - public health. Nevertheless the profound reform in the health cares sector the policy makers, somewhat influenced by external factors, managed to save the existing capacity in the field of public health promotion, awareness raising and disease prevention. The National Center of Public Health, Medical Ecology and Nurturing (NCPHMEN) is based within the organizational boundaries of the Hygiene-Epidemiology Service, which is authorized to implement the public health policies. NCPHMEN has an extensive background in data collection, processing and dissemination of analytical information, therefore it is not surprising to see that the service has fully embarked on the potential of the data mining applications to mine their data in order to classify, predict, cluster and describe the hidden connections between the variables. Some of the particular public health fields in which data mining technologies are being introduced are: cluster analysis of health related risks, analysis of geographic and spatial data for the purposes of risk management and health promotion campaigns as well as hybrid approaches between database management and data mining.

The challenge of data mining – privacy concerns

As Montana (2001) points out the data mining of various personal data sets such as “...medical, epidemiological, demographic, marketing and others – allow us to draw conclusions that are not otherwise possible or to focus marketing or other activities to groups we would otherwise know little about”. The issue of racial profiling at the airport security checks that emerged in the aftermath of September 11 is just another evidence of the powerful mechanisms that the public administration has in hands and the controversial nature of these methods.

The privacy concern is a major issue which must be accounted for before implementing data mining solutions in the public sector. Several domestic acts and the relevant EU legislation pose significant restrictions and checks and balances against misuse of the automated data processing of personal records. For instance the Privacy Law (2002) stipulate that personal data is the data which reveal the physical, psychological, cultural, family, economical, cultural and social identity of a natural person. The Law does not elaborate however on the applicability of the law over the hidden relationships between data which do not represent personal data per se but allow for inferences that can be labeled as personal. In that respect the data mining techniques open the door for general debate with profound legal, social and ethic implications. We want to point the attention towards the specific challenges of the data mining which have to be investigated further. Additional research is needed in order to find the balance between the potential of the knowledge discovery technologies in the pursue of the public interest and the privacy rights of the citizens. Most of the specifics stem from the fact that in the data mining we do not make specific case-based inferences but generalize about a groups of subjects, transactions or objects. Event though here there is no direct conclusion for particular person it can be fairly dangerous from a privacy point of view.

Conclusions:

The observation on the applications of the data mining techniques in Bulgarian public organizations gives us some valuable insights. First of all there we see that the organizational IT background matters when the knowledge management concept is at stake. We saw that only organizations with good capacity in IS management and culture of analytical knowledge discovery employ the data mining applications. One can argue that there are other factors that affect the chances of seeing application of data mining tools in the public sector and this probably is the case. However we think that these factors are mostly at managerial level and have to do with attitudes, management skills and styles and strategic thinking. Most of the public organizations assume the role of information processors and exist in information intensive environment. Facing the need for transformation of the sheer data into information and consequently the information into knowledge is ideology that must be developed bottom-up. Even to a greater extent than the rest of the ICT applications in the process of public governance we can expect that the knowledge management approach will fail if implemented top-down.

Knowledge discovery and management represents the direct appeal of the data mining techniques. A closer look at the organizational impact of the knowledge management shows us that data mining provides for organizational intelligence and memory preservation (Berry and Linoff, 1997). In the context of the public organizations where the political clout is important and inextricable factor we would expect to see more safeguards against organizational brain drains. Data mining seen as tool for knowledge discovery and management have the potential to significantly improve the consistency of the process of knowledge generation, preservation and utilization in the public organizations.

The brief analysis of the theoretical aspects of data mining confirms the fundamentals of the bounded rationality model. We can say that the artificial intelligence proved the empirical descriptive and predictive power of the bounded rationality theory, based on the assumption that the artificial intelligence and the human mind decision making have more similarities in the essential elements of the reasoning. Apparently more research is needed in order to deepen the validity and reliability of the statement, but we think that parsimony principle of the data mining is reference to the mechanisms of decision making as described by the bounded rationality model. A closer look at the information selection criteria used in the data mining applications is also a kind of evidence for the congruence between the two intellectual fields.

The brief presentation of the data mining techniques and some of their applications in the public sector prove the suitability of the knowledge discovery approach in the public sector. In the near future we can only expect the proliferation of data mining approaches and te pioneering approach of the private sector industries only strengthens this believe. Preparing the grounds of the knowledge management approach will require drastic changes in the public policies, legal frameworks, management attitudes and the concept of information systems the public sector.

References:

- Berry, Michael J.A., and Gordon Linoff (1997) *Data mining techniques : for marketing, sales, and customer support*. New York : Wiley
- Chu, C. H., & Widjaja, D. (1994). A neural network system for forecasting method selection. *Decision Support Systems*, 12, 13-24.
- Dan Zhu;G Premkumar;Xiaoning Zhang;Chao-Hsien Chu, 2001, Data mining for network intrusion detection: A comparison of alternative methods, , *Information Management Journal* v.4: 635-660
- Hand, David J, (1998), Data mining: Statistics and more?, *The American Statistician*, 2, 112-118
- Fayyad, Usama, Georges G. Grinstein and Andreas Wierse (eds.) (2002) *Information visualization in data mining and knowledge discovery*, San Francisco: Morgan Kaufmann
- Fulda, Joseph S. (2000), Data Mining and Privacy, *Albany Law Journal of Science and Technology*, 105
- Gramatikov, Martin (2003), Digital Divide: Major Patterns Found in the Web Sites of the Bulgarian Minicipalities in Wimmer, Maria (ed.) *Quo Vadis E-gov: 2003*, OCS: Vienna, pp. 72-94
- Haskett, Mitch, (2000) An introduction to data mining, part 1: Understanding the critical data relationship in the corporate data warehouse, *Enterprise Systems Journal*, v. 15: 32-34
- March, J. G., and Simon, H. A. (1958) *Organizations*, New York: Willey
- Montana John C, (2001) Data mining: A slippery slope, *Information Management Journal*, 35: 50-54
- Nancy A Dube, (1999) Globalsearch: The international search for missing and exploited children, *International Review of Law, Computers & Technology*, v. 13: 69-74
- Simon Herbert A. (1982) *Models of Bounded Rationality: Economic Analysis and Public Policy*, Cambridge Mass: MIT Press
- Simon, Herbert A. (1997) *Administrative Behavior*, New York: The Free Press
- Tam, K., & Kiang, M. (1992). Managerial applications of neural networks: The case of bank failure prediction. *Management Science*, 38(7), 926-947
- Sauter, Vicki L, (1999) Intuitive decision-making, Association for Computing Machinery. *Communications of the ACM*, v. 42: 109-115
- von Newmann, John, and Oskar Morgenstern (1964), *Theory of games and economic behavior*, New York, Wiley
- Weiss, S.H. and Indurkha, N (1998), *Predictive Data Mining: A Practical Guide*, Morgan Kaufmann Publishers, San Francisco, CA