

DATA ANALYSIS, FUNDAMENTAL FACTOR IN THE ELABORATION OF THE TOP ORGANIZATIONAL MANAGERIAL DECISION

Marian Sorin IONESCU^{a*}, *Olivia NEGOIȚĂ*^b

a, b. University Politehnica of Bucharest, Romania

DOI: 10.24818/IMC/2020/03.20

ABSTRACT

The processes for the elaboration of the top modern organizational managerial decision are characterized by a high degree of complexity and uncertainty. The high financial sustainability and their implicit strategic consequences determined on long-terms are only two relevant parameters identified in this scientific communication. Also the following components are monitored as theoretical, transposable entities, subsequently operational criteria of choosing decisional alternatives: the use of the probability calculus in the decision elaboration processes, influence of the primary decisions on future evolutions, as the decision quantifies the information value, as the attitude towards the risk arising from the business ecosystem affects data analysis. In this paper we comment about the features and the importance of these components in the general framework of Data Analysis.¹We consider and discuss some relevant examples of top decision such as: sending offers of the economic organizations' contracts, launching policies for new products on the marked – affected by a high uncertainty degree, extension of the production capacities in the financial – banking sector, decision to grant credits at the organizational or individual level, economic organizations providing utilities, strategies appropriate for the environment and economic consequences.

KEYWORDS: *data analysis, data mining, data science, deep learning, machine learning*

1. INTRODUCTION

In the study of the modern economic organizations, positioned in an dynamic and hyper- competitive ecosystem, the analysis imposed by Data Science (Albright & Winston, 2018) by the approach Data Mining is a process with very well-structured steps, some involving facilities provided by Information Technology (I.T.) or by the processes specific to Machine Learning and Deep Learning, involving an analytical approach with creative and innovative elements, a very good knowledge of the chosen business models that are operationally transposed and their strategic perspectives. The Data Mining approach (Berger et al., 2018) is a systemic paradigm, which must be understood both in its theoretical deepness and also in the complexity of the sampled data volumes, interpreted and subsequently stored. An important quest for use of the collected data has always been to make as accurate as possible business forecasts (Armstrong, 1986). It is worth noting that data mining from various sources such as by using distributed sensors from Internet of Things (IoT) may need specific technique for efficiency such as compressed sensing (Donoho, 2006). Regarding the issue of the chosen business models and subsequently transposed in an operational – strategic manner, the main aimed target is to find correlations, interactions between particular variables, specific to the main object of activity of the studied economic entity and the other variables involved in the approached systemic processes.

*Corresponding author. E-mail address: marian.ionescu@man.ase.r

We identify specific classifications and tasks of mathematical regressions, as follows:

- (i) The *Probability Estimations and Classifications* are predictive processes that from a set of elements evaluates whether a certain economic event belongs or not to the classification. For a classification task, an algorithm of the data mining type determines whether the appurtenance to this type of events is specific to a set or not, the monitored task is designated as *Scoring task* for that set or *Probability estimation*;
- (ii) The *Estimated value or Regression* allows the prediction of numerical values for certain individualized variables, with applicability in marketing strategies in the analysis of the choice and intensity of use by a final consumer of certain services and products, informally, classification processes predict whether a certain event will take place while regression predicts the amplitude of that event;
- (iii) Similar economic events are structured by sets of similar data, the focus of the economic organization's sale force is on portfolios of potential customers coming from competing economic entities but which are using a similar business model;
- (iv) Clustering focuses the individualities in sets, groups, having as decisional argument data similarities within the operationalized business model, aiming products that should be developed and offered, providing a superior added-value;
- (v) Coherent grouping, approach by *Frequent Item set Mining*, which identifies relations and connections between economic entities in case of systemic operationalities, their improvement possibility until optimization;
- (vi) *Profiling* represents the behavioral description characterizing the attitude of an individual, group or population for a certain market segment, this is a highly useful argument in the development of consolidation strategies and organizational growth;
- (vii) *Link prediction* is a facility identifying connections between items specific to data, initially identifying their existence and subsequently their deepness and interconnectivity degree;
- (viii) *Data Reduction* is the method by which large sets of data, even overwhelming, are restructured in smaller entities, thus being easier to understand and operate, containing significant, representative information for the economic organization;
- (ix) The *Casual Modeling* approach represents the way in which it is possible to understand economic events or actions of this type, with the capacity of influencing other systemic processes.

2. DEBATE ON THE SUBJECT

The learning processes, with applicability in the activities of the modern economic organizations are of two types: Supervised Learning and Unsupervised Learning" the terminology is subject to the concepts and paradigms specific to the approach Machine Learning.

These two approaches have many similarities but also profound operational and strategic systemic differences: Supervised Learning aims the prediction of the final targets for grouped sets of data. Unsupervised learning, like the punctual case of Clustering, refers to sets of data constituted on similarities but a relatively high degree of uncertainty may occur when it is transposed into operation for the entity studied in the business ecosystem. An essential condition used in the Supervised Data Mining is that the significant data for the approach of the main target of the economic organization shall be selected and evaluated.

We shall exemplify by considering that in the marketing studies evaluating the behavior of the end customers it is important to evaluate their behavior in certain time periods and their information history should be noted even if the information is incomplete, thus fixing the target data is the foundation for the development of investment strategies at the level of the economic organization, positioned in the business ecosystem.

The processes specific to data mining are those determining the patterns used for the elaboration of the new business models, the results of these processes are the key factors when the top organizational management makes decision and justifies the search for an optimum systemic operability. We suggest the structure of Table 1, after evaluating the operability itself, for data extraction.

Table 1. Proposed structure for data extraction
Historical Economic Data, Training Data, Total Specified Values

x	y	z	Set
10	True	Green	Accepted
8	True	Blue	Rejected
...
99.1	False	Red	Accepted

→ Data Mining → *Adopted business model*"

New Data" Sets of unknown values

x	y	z	Set
10	False	Blue	?

→ Model → *Accepted set, Probability: 0.9*

This enables the estimation both for sets value as well as for the probability of the variables values within the sets.

Important scientific and technological resources are allocated for the operationalization of the processes of Data Mining Process.

We consider that for modern economic organization, the most efficient, flexible and performing approach of processes specific to Data Mining is the structure of a Cross Industry Standard Process For Data Mining-CRISP-DM, described as follows:

- (i) Understanding the adopted business model represents the sequence in which the creativity of the data analyst has the most important role. They are at the disposal of those in charge with the processes specific to the scientific-technological tool data approaching the challenge from business problems to data mining tasks, a general evaluation of the problem is necessary, the team of the economic organization in charge with the solution's elaboration is the one deciding the type and mode of operation within the necessary scenario.
- (ii) Understanding data, in case a problem specific to business represents the end target, the raw material is found within these, condition that underlies the solution sought. The databases regarding the end consumers, real and potential customers, appropriate for studies and responses generated by marketing studies, contain disjoint sets of information but which are covering certain fields of interest aiming structures of population with different reliability degrees. The paths (directions) for the settlement of the occurred challenges by using processes specific to data analysis change the response operational directions and the efforts of the economic organization's team managing this problem are increased. An example for this purpose is the combat and monitoring of legislation fraud specific to health insurances is the medic are fraud, a very complex challenge, with financial implications of billion Euros. By approaching the business model chosen by the paradigms specific to data processes, we find that in this particular case there is not just a fraud detection, the fraud generators, medical providers with false claims, sometimes even final beneficiaries – patients operationally legitimize the services of the medical providers, use complex invoicing systems. Frauds are committed by subsets of identifiable customers (legitimated), all the parties involved in this process try to avoid fair public taxes.

- (iii) Data preparation is an essential process in the operationalization of the sampling technologies, data interpretation and use, this process aims data conversion from different initial forms in a form that is easier to understand and interpret. At the economic organization's level, data analysis specialists allocate a significant amount of time in the process of defining the variables used in systemic processes. It is one of the major points in which human creativity defines the business model and its role, the quality of the selected data and *mining solution* totally depends on the human decision and variables selection.
- (iv) Modelling represents the development of a modelling structure (Rissanen, 1978) or capturing pattern for data regularity, it is the main location where Data Mining algorithms are applied.
- (v) The evaluation is the step in which the results of the *Data Mining* processes are quantified and analysed and then validated, thus the organizational strategic development patterns are identifiable in terms of data. Evaluation is the systemic process both from a quantitative and qualitative point of view, categories of stakeholders with interests in adopting a certain business model for decision making as well as for understanding and assimilating the results of these types of models. Consequently, the framework's evaluation is important because it provides informational details of the generated economic performance, the vulnerabilities and limits of the adopted business models are evaluated from this point of view.
- (vi) Implementation is the step in which the precise results of *Data Mining* are operationally – strategically transposed at the level of the economic organization, at the same time the result of the investment yield is financially evaluated. For certain systemic intermediate targets, implementation is a process taking place automatically, among the arguments for this operability, we indicate the following, the business ecosystem is rapidly changed, the prevailing team, quantifies and evaluates the data which should have high degrees of adaptability and flexibility, the structure of the business models has multiple modelling tasks for the individuals in charge of understanding and processing data, this is a task with a high complexity degree, the relevant information is subsequently provided to the decision making organizational top management.
- (vii) The management by the team in charge with *Data Science* at the organizational level of the *Data Mining* processes such as a development cycle for software products. For different business models adopted and subsequently operationally – strategically transposed, the projects of *Data Mining* are cancelled and managed as engineering projects for software development dedicated for reaching certain targets indicated by the organizational top management. By using the perspective provided by the previously presented approach C.R.I.S.P., the possibility of an exploratory, iterative process occurs, models and strategies developing based on a design specific to software projects. Thus, for economic organizations using teams specialized in "Data Mining" type processes, both abilities regarding the development of software projects with variable complexity degrees as well as analytical abilities of the involved human operators are essential.

The Business Analytics perspective involves the systemic operability of a varied range of paradigms and analysis technologies, interpretation, the use and storage of data from the ecosystem in which the economic organization is located.

The differences between them occur when focusing on *Data Mining*, by automatic search for identification of knowledge, strategic patterns, data regularity. At the level of business analysis the recognition and selection process of the type of adopted analytical technologies is very important and include:

- (i) Statistics, referring to all the terms included in the calculation processes for numerical values (Hastie et al., 2003) that include amounts, average values, rates, yields, their totality is known as Summary Statistics. The other referral to Statistics (Gins et al., 2015) represents the referral to the

scientific apparatus necessary to all economic histories, understanding the positioning of the economic organization in the operational ecosystem, generation of strategic hubs, in this point a huge effort of understanding the selected and operationally transposed business models.

- (ii) The interrogative approach of data bases is the process by which a set of specific questions is defined for the statistical interrogation of the data sets and all located to the entire data bases system, the totality of information solutions are valid in order to provide answers to one or all of the interrogations generated by the organizational team which is responsible for the "DataMining" processes, their structuring is compliant to the paradigms Structured Query Language-S.Q.L. or facilities provided by graphical interfaces, of the "G. U. I." type which are a real help in generating interrogations of the type Query-By-Example(Q.B.E.).

Interrogations at the databases level are a mandatory systemic process when targeting subsets of data that confirm or disprove strategic consolidation and organizational development hypotheses, an example in this regard are software structures targeting marketing departments and the Sale Force.

The online analytical processes, On-Line Analytical Processing - O.L.A.P., provide the facile operationalization possibility of graphical interfaces, G.U.I. for very large sets of interrogations related to databases, their exploration, sampling and interpreting representative data, their storage inappropriate entities and making them available after wards to the deciding top organizational management.

O.L.A.P. is structured so as to facilitate a direct approach or visual exploration of data by the departmental responsible persons, it is practically a complementary utility in the understanding and collection of results specific to "Data Mining", applicable in case of implemented business models.

- (iii) "Data Warehousing" represents the location for the collection and storage of data from the entire economic organization occurring in the totality of the systemic transactions.
- (iv) The Regression Analysis is a fundamental method in in the set of analytical and statistical methods used in the operational strategic economic applications, estimative and predictive values are involved in the analysis of the data sets used.

We identify the possibility of developing exploring models versus predictive modelling that facilitate reaching the targets decided by the top organizational management.

We also note the systemic possibility of overlaying the different information technologies used, their identification and disjunction is the attribution of the organizational task force that manages the specific processes of Data Mining.

- (v) Data Mining addressing this approach by the concepts and paradigms specific to the competence field "Machine Learning" also two highly important components that must be studied and analysed in their deepness are those of Applied Statistics and Pattern Recognition.

The focus in these types of approaches is on the improvement of an intelligent agent's knowledge performance during the entire systemic processes.

The study Data Mining or K.D.D. – Knowledge Discovery and Data Mining begins by implementing concepts and paradigms specific to Machine Learning. This approach may include different models and fields of management performance improvement at the economic organization's level, such as robotics and computer vision that are not related to the competence field K. D.D.

The generated results are those of a Cognition Agent, organizational learning processes for the modern business ecosystem.

The K.D.D. processes are deeply linked to the operationalities specific to "Machine Learning" regardless of the business model adopted by the studied economic organization.

3. CASE STUDY AND PREDICTIVE MODELS

We identify a few questions specific to the business models adopted and addressed with these algorithms and technologies, to which a flexible and efficient answer must be provided, as follows:

- (i) The identification of the most profitable customer.
- (ii) The identification among the profiles of customers with high profitability and those with average profitability.
- (iii) The identification of the customers' features, particularities, behavior.
- (iv) Is new potential customers becoming a profitable customer? What is the potential revenue and investment yield generated by these customers?

A model represents a more simplified representation of the business reality, generated in order to be used in certain purposes decided by the deciding organizational top management. The simplification process is based on certain hypotheses and constrains regarding the information used and their traceability, on two levels, intrinsically and extrinsically.

Predictive processes, the elaboration of the management decision at the economic organization's level are forecasts for strategic events that will be faced by the studied entity, during data analysis, prediction is actually an estimation of an unknown value. Predictive models for credit-scoring functions are identifiable in their area of competence, regarding the customers' credit worthiness in the financial – banking industry, predictive models for spam filtering operationalities, estimating when an email is useful or spam, predictive models for detection of potential frauds, highly flexible in IT security.

For example, the structure of Table 2 below presents an approach for a supervised application of the classification problems.

Table 2. Example of classification

NAME	ANNUAL NET REVENUE	AGE	EMPLOYED	GOOD PAYER
Tudor	200,000 Ron	52	no	yes
Doina	55,000 Ron	38	yes	no
Claudiu	180,000 Ron	42	yes	no
Dora	72,000 Ron	28	no	yes

Descriptive modelling initially (primarily) aims not the estimation of a precise numerical value but the identification of a strategic perspective for the involved economic phenomena and system processes.

The creation of models specific to processes regarding data is known in the specialized literature as Model Induction, induction is a philosophical concept referring to extrapolation, generalization of paradigms, rules, fundamental operation laws, in this case being related to the analysed economic entities.

In the modern thinking, induction is the opponent of deduction, the latter starting from general rules or specific cases, creates other analogous particular cases, the use of business models adapted to the processes of sampling, processing, interpretation and storage of data, are considered to the procedures of probability deduction.

The Input Data for the induction algorithms specific to a certain type of business model are also known as Training Data, also being designated as Labeled Data because the value of the variable targets is assumed as being known.

The fundamental question that appears is regarding the selection process of one or more attributes, variables that successfully share sampling, by observing the condition of the variable target of economic performances ought at the level of top organizational decision maker.

The approach continues by focusing on the issue Supervised Segmentation, predictive models elaborated by being focused on certain particular, variable targets of economic performance that should be reached within the strategic programs, subsequently elaborated, which are operationally transposed. An extraction of patterns followed by data in a supervised manner, segmentation of the evaluated populations in two subgroups with different values for the monitored target variables represents the purpose of this type of approach. The purpose is to identify the best segmentation, possibly optimal, which will represent the base of the development of the subsequent strategic predictions.

Technologies and algorithms specific to Data Mining are implemented in the fundamental developed analysis that help us identify whether a variable contains information with a certain degree of relevance for the monitored target variable.

Predictive values may be estimated starting with data in different approaching modes, thus the working hypotheses is induced in which by a mathematical approximation Training Data are positioned into a certain segment, generated by the developed process. One very interesting example of application here has been reported by Li and Zhang (2014) concerning the random forest techniques applied to semiconductor manufacturing process fault detection.

The approach and settlement of these types of challenges lead to the appearance of actual technologies specific to Data Mining for economic organizations, easy to extrapolate for a multitude of business models.

Estimations are perfectly possible in case of marketing and sales departments regarding the population's transformation rate, dependence on a certain portfolio of products, occupational data, residency, age, gender, need of products and services, dimensionality of the allocated financial commissions.

In this scientific analysis, the Selection of Informative Attributes used in the partition processes, has a fundamental role.

The general framework considers two sets, the binary partition and attributes used for the presentation of these sets, we generally use the logics witch "Yes or No" so as to indicate whether one of the elements is related to one of the two subsets, the aimed "Attributes" and "Targets – Variables" are identified for this purpose.

We synthetically identify the following steps:

- (i) It is an ideal case when the attributes are shared in perfect groups, in the business operationality usually a group is purely idealized while the other is not.
- (ii) In case of accurate examples, the required conditions generate a single set (subset) of useful data, the existence of the second group of data is not justifiable.
- (iii) Not all the studied attributes are binary, there are many categories of attributes with three or more distinct values, sharing for analysis, use and storage are usually in two sets, but it is also possible in three or more according to the chosen business model and set strategic targets.
- (iv) Certain attributes have numeric values, whole or real, the segmentation process in this purpose is defining, the supervising segmentation is fully dependent on the use of numerical attributes.

The most used splitting criterion is known in the scientific literature as "Information Gain" being based on the measurement of the entropy purity, these were the bases of Shannon's information theory. The scientific term of entropy is perfectly applicable to the economic organizational entities, considering there is a set of properties specific to each member of the set:

$$Entropy = -p_1 \log(p_1) - p_2 \log(p_2) \quad (1)$$

Each p_i represents the probability that the property I belongs to one set, starting with $p_i = 1$, where all the set's members have the property i , and $p_i = 0$, when then on-members of the set have property i .

The evaluation of the informative quality brought by an attribute is important for the applications specific to business models, in the process of monitoring a certain indicated target, how much is the informational gain for the values assigned to the variables targets. If the entropy is the one defining

and characterizing the uncertainty and disorder of the studied economic processes, it is defined as *Information Gain (I.G.)* in the specialized scientific literature, quantifying how much an attribute improves the entropy during the entire developed segmentation process.

The I.G. is a function partitioning the quantity of information brought at the attributive level.

For example, a possible definition of *I. G.* is:

$$\begin{aligned} \text{Information Gain (I.G.)} &= I. G. (\text{Economic Organization, Customer}) \\ &= \text{Entropy}(\text{Economic Organization}) - [p(c_1) \times \text{entropy}(c_1) + p(c_2) \times \\ &\quad \text{entropy}(c_2)] \end{aligned} \quad (2)$$

The reference made is relatively related to the aimed purpose, sharing by using a single elements and observation if this is pure, does not represent a better option than dividing the set aiming the economic organization in two subsets of high dimensionality, relatively pure, even if none is totally pure in terms of information.

It remains to be seen what happens when attributes are numerical, thus numerical variables are discretized by selecting a disjunction point, sharing, a split point or several, the result is treated as for an attribute of the categorical type.

In certain economic case studies, the input data are separated in two or several categories, the information thus gained is operationalized in systemic processes for the evaluation of the segmentation created by discretizing the numerical attributes.

Conceptually, the approach of all the splitting points, the selection of a possibly optimal one provides the possibility of the best gained information performance.

Here in after, we evaluate the attributes selection with information gain, considered to be very useful in the problem of business models implemented operationally and strategically for modern economic organizations, looking for the performance improvement and growth of the investment yield rate.

We consider as straight forward and relevant for the operationalization of Information Gain, the example built with the help of a machine learning at University of California Irvine Machine Learning Repository (URL 1), in which a series of basic attributes are highlighted in the data sets, concerning here certain species of fungi possibly processed for commercialization, but also to be avoided, due to the lack of quality and toxicity (see Table 3).

Table 3. Example of attributes in the classification of species of fungi for commercialization

Attribute Name	Possible Values
SHAPE	bell, conical, convex, flat, knobbed, sunken
SURFACE	fibrous, grooves, scaly, smooth
COLOR	brown, buff, cinnamon, grey, green, pink, purple, red, white, yellow
ODOR	almond, anise, creosote, fishy, foul, musty, none, pungent, spicy
GILL ATTACHMENT	attached, descending, free, notched
GILL SPACING	close, crowded, distant

Source: adapted and reproduced from (Albright & Winston, 2018)

All the sampled, operationalized and stored data within the organizational business model scan be found in structures of the type Data Ware House, huge assemblies of databases facilitating the elaboration of

data patterns, once Data Ware House are constituted, the deciding top organizational management shall monitor the following areas of interest:

- (i) Classification analysis seeks to identify variables that are connected to a well-defined variable, many times this is a binary process.
- (ii) Prediction is a process similar to Classification analysis, except the fact that it seeks to explain continuous variables, such as the credit balance, before approaching the well-defined variables, the regression should be quoted.
- (iii) Cluster Analysis is the one dividing the observation groups in clusters if these are or not included in a certain type of cluster.
- (iv) Market Basket Analysis is the operation searching for the accessed products and bought by the end users-customers.
- (v) Forecasting is the systemic operationality generating predictions for the values achieved in time of certain sets of variables with the help of patterns extrapolation processes, by using the economic history of data and correlating to the future evolution of the economic organization.

Data Mining can also be interpreted as a set, collection of complex algorithms which are operationalized for the identification of useful information (Shanga & You, 2019) within the adopted business models and patterns characterizing large sets of data.

It should be emphasized that the extrapolation of the previous concept is based on key ideas, such as tables, pivot tables, specialized maps (graphical representations), the facilities provided by the modern software products provide the analysis and studio possibility for certain data sets of a large dimensionality (Sarveniazi, 2014), made rapidly and easily intelligible for all members of the organizational deciding task force.

It should be mentioned that from the Microsoft Suite, Excel being an excellent support for data interpretation and analysis, other options from the software applications world, dedicated to these type of operationalities are Access, S.Q.L. Sever, Oracle, that are highly efficient for data storage and packing with a certain economic relevance.

This approach has certain particularities for data analysis, facility, flexibility and the highest reliability especially in terms of the business ecosystem's challenges is provided by Excel.

The methodology of using tables is not limited only to the level of the software tools, Excel or other applications of the Microsoft suite, is generally designated as On Line Analytical Processing, (O.L.A.P.), initially this name was used to make the distinction between the analysed and transactionally processed data types, also called On Line Transactional Processing (O.L.T.P.).

The most efficient structure for data analysis for economic processes has a star layout in which we can find the Table of Actions with several lines and only a few columns, as depicted in Table 4.

Table 4. Example of a simple table of to build a Beverage data base

1	A	B
2	Name Product label	Generated revenue
3	Beverage	142.578 Ron
4	Alcoholic beverages	41.137 Ron
5	Beverage	80.152 Ron
6	Carbonated beverages	17.754 Ron
7	Plain beverages	17.028 Ron
8	Hot beverages	26.227 Ron
10	Hot chocolate	4.085 Ron
11	Coffee Assortments	22.141 Ron

1	A	B
12	Natural juices	19.141 Ron
13	Dairy beverages	21.289 Ron
14	Food	1.187.171 Ron
15	Non-consumable products	314. 635.84 Ron
16	Final Total	1.644.385.60 Ron

Technology of the type O.L.A.P.is not the intellectual property of a certain specialized software corporation, it provides many versions already developed by experts in the field and it is accessible in many versions, adapted for a multitude of business models.

The facilities provided by this technology, within the applications MicroSoftare located in Analysis Services, which is an integral part of the module S. Q. L. Server – Data Bases Software.

The approach O.L.A.P.is different from the one provided by the specialized software in the "MicroSoft Suite", pivot tables may not be subject to the pulling process in any of the access areas, only actions are allowed in the area of values and only dimensions are accessed (allowed) in the area of lines, columns, specialized area filters. Certain dimensionalities have natural hierarchies, resulted from realities generated by specialized markets.

The databases found within this approaching paradigm have large dimensions, prevailing times, data analysis, processing and storage have increased dimensionalities, this is why we consider that the development of a pre-processing process is absolutely necessary.

Data pre-processing has a particular structure designated O. L. A. P. Cubes starting from an analogy with the Rubik cube, where each sub-cube contains the result of a process of whole fragmentation. As a facility provided to users, Excel, we identify as possible the construction of their own O.L.A.P. cubes. Finally, we would like to mention the concept of Data Model for the transmission and storage to the deciding factors of large tables, even huge as dimensionality and stored information. Power Pivot, Power View, Power Query, Power Map are subsequent software developments which are highly efficient for the business models of economic organizations, succeeding Excel's Self Service Business Intelligence (SELF SERVICE B.I).

A typical structure operationalized with the facility Power Pivot has the set of data stored in three different files, most of the relevant useful data are located in Access file. For example, ContosoSalesCom is structured in five entities (tables) of: Dates, Sales, Products, Product Subcategories, Product Categories.

Operationalization tables that are subsequently processed could come from different sources, combinations of software facilities such as Access, Excel if columns close as information for these tables, primary similar data and control key as usable are identifiable, Power Point is the operability with the highest degree of efficiency, close to optimization, thus being the most recommended.

4. CONCLUSIONS

The economic organizations, actors on the software products market, with applications for business models, use as strategic development vector the approach of the processes of the Data Analysis type. We conclude that the Excel and Power Point tools of MicroSoft tools have two essential advantages: (a) Systemic processes can operationalized, with enormous sets of data, millions of lines, avoiding the computer's blockage of the sequencer involved in this activity, and, (b) Power Point in corporates flexible, fast and reliable software solutions adapted to data tables, different to the classic Excel solutions.

An overview of the entire systemic process performed determines us to suggest the following operational-strategic pragmatic algorithmic structure for business models specific to economic organizations, as it follows:

- i. The data import process is initiated from three different sources in the software facility Excel Data Model, not in the physical tables Excel Tables;
- ii. Connections are created between tables in the software facility Data Model;
- iii. The applications from Data Model are modified so as to be used in pivot tables;
- iv. One or several pivot tables from the facility Data Model shall be elaborated.

The applications of these algorithms form just a part of the challenges of Data Analytics and Mining applied to business organisations and there are certainly many other challenges related to the decisional processes based on the applications of such algorithms and their outcomes.

REFERENCES

- Albright, S. C. & Winston, W. L. (2018). *Business Analytics: Data Analysis and Decision Making*. Stanford: Cengage Learning.
- Armstrong, S. (1986). Research on forecasting: a quarter-century review 1960-1984. *Interfaces*, 16(1), 89-103.
- Berger, P. D., Maurer, R. E. & Celli, G. B. (2018). *Experimental Design with Applications in Management, Engineering, and the Sciences*. Heidelberg, Germany: Springer International Publishing.
- Donoho, D. L. (2006). Compressed sensing. *IEEE Transaction on Information Theory*, 52(4), 1289-1306. doi:10.1109/TIT.2006.871582
- Gins, G., Vanden Kerkhof, P., Vanlaer, J. & Van Impe, I. F. M. (2015). Improving classification-based diagnosis of batch processes through data selection and appropriate pretreatment. *Journal of Process Control*, 26, 90-101.
- Hastie, T., Tibshirani, R. & Friedman, J. (2003). *The elements of statistical learning*. New York, NY: Springer.
- Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14, 465-471.
- Li, Y. & Zhang, X. (2014). Diffusion maps based k-nearest-neighbor rule technique for semiconductor manufacturing process fault detection. *Chemometrics and Intelligent Laboratory Systems*, 136, 47-57. doi: 10.1016/j.chemolab.2014.05.003.
- Sarveniazi, A. (2014). An Actual Survey of Dimensionality Reduction. *American Journal of Computational Mathematics*, 4, 55-72. doi: 10.4236/jasmi.2014.42006.
- Shanga, C. & You, F. (2019). Data Analytics and Machine Learning for Smart Process Manufacturing: Recent Advances and Perspectives in the Big Data Era. *Engineering*, 5(6), 1010-1016. doi: 10.1016/j.eng.2019.01.019.
- URL 1 (n.d.). *University of California Irvine Machine Learning Repository*, Retrieved July 20, 2019, from <http://archive.ics.ucl.edu/ml/datasets/Mushroom>.