Advanced Analysis of Data Involved in Organizational Business Processes

Marian-Sorin IONESCU
E.P.F. L.
Lausanne, Suisse

Politehnica University of Bucharest
Bucharest, Romania

Abstract. *The type of developed data analysis identified in the management decision study of the economic organization is the decisive factor for the success of modern corporations in the world in the modern globalization, the most extensive and complex factor that modern human society faces. New concepts and paradigms introduced by modern, fast, flexible, and in-game management decision-makers implicitly introduce software solutions developed in the broad concept of data exploitation, a concept known in the literature as "data mining". The process is one of analyzing significant amounts of data, extracting information considered to be relevant, based on mathematical algorithms and statistics. Data mining is, therefore, a non-trivial extraction of all previously unknown, potentially useful implicit information from significant volume data; it also conceptually refers to statistical analysis and logic within the operational and strategic information developments, for the elaboration of some decisional algorithms. By specific approach of "data mining", are developed the discovery of patterns, trends (trends), relationships, links and cleavages between different types and structures of data, especially unobvious, and paths difficult to identify intuitively. One of the key concepts for such studies is data warehouse, super-structured databases for understanding, studying and designing data-specific solutions, conceptually is an extrapolation of the operational organizational database that: (a) Combines various data from multiple sources with a multitude of inter-entity relationships; (b) Contains data with high accuracy and consistent individuality; (c) They are structured so as to allow quick access to waiting queues; (d) Allow follow-up of informational responses for relevant and meaningful questions at the level of top organizational management. The economic organizations that have or take the strategic decision to develop a theoretical-operational warehouse type, base their structures with efficient, flexible and performance algorithms and responses tailored to the specific business object.*

Keywords: *data analysis; management decision; information; data mining; data warehouse.*

Particularities of warehouses used in economic structures

We identify hubs to substantiate warehouses in the economic field as follows:

(a) The prediction, a process similar to the classification analysis, with the exception of the variables explaining the processes with a certain degree of continuity, here as the most popular algorithm usable in business, identifies the structures and mechanisms of mathematical regression;

(b) Cluster analysis, is the process of positioning concrete business observations in cluster areas, are particularly useful for economic organizations that focus their marketing and sales strategies on clusters with a high degree of individualization;

(c) It is the classification analysis that identifies variables that have operational connectivity with a quantifiable mathematical variable, usually binary;

(d) The Market Analysis Basket is the one that identifies all the products that are included by the final customer at the same acquisition

(e) The forecast predicts values of variables in temporary programming, the process is elaborated using mathematical extrapolation concepts and algorithms, from an economic past to a future of organizational consolidation and development;

The data extraction processes are of enormous theoretical and applicative complexity, the existing algorithms and processes are duplicated by software specialized in their operation. Advanced information algorithms make it possible to extrapolate data, but there is also mathematical formalism simplified by data exploration and highlighting.

Relative databases, extrapolation for online analytical processing

The individual databases in the operations used by the top organizational top management with the Excel facility are presented as rectangular individualized drawings with lines, columns and corresponding field records, while the complexity of applications is implicitly required for a new concept of "relational databases", a multitude of interconnected tables, related drawings with the same structure-lines, columns, records, links are explicitly presented.

For top decision management, the Excel application is a tool-generating feature in data mining, processing, and analysis, but does not show performance in storing, preserving significant data in the analyzed processes. For operational practice and development of strategic business model models, a software feature such as Access, S.Q.L. Server, Oracle, accessed in a modern, dynamic and competitive economic environment. The application-operational mix is achieved at the level of the economic organization when using algorithms and specialized procedures for importing data from databases within Excel for analysis and processing.

"OLAP" approach

The introduction of pivot-type cached tables is a relatively simple method of approaching the studied processes, allowing the data to be explored according to their category in Excel, the so elaborated thinking paradigm is extrapolated to the business models operating the online analytical processing, being a distinctive clear to online transaction processing, the processes known as OLAP algorithms.

Considering that in the question of economic models the queries are of the multidimensional analytical type, MDA, the OLAP analytical processing processes online, represent the most appropriate approach, including relational databases, written reports, and data mining; thus, a feature of flexible presentation and in concise time of managerial reports, marketing, financial analysis, sales, business models management, "B.P.M.".

OLAP-compliant databases are multidimensionally modeled, complex analytical queries and performance time allocations being identified. The navigation, hierarchical and relational databases are used in this approach. At the level of the economic organization, top decision-makers are asked questions that characterize the operational processes such as those related to customer relations, dispatching orders, paying them at predetermined deadlines, monitoring of financial-accounting fuses, pre- and post-sales services offered.

For proper resolution and generating economic performance for these types of queries, we believe that the OLAP processing approach is the right path. We propose a matrix-type operational structure called "Star Pattern", which lists the states of the economic organization, structured on many lines and few columns. These are not just standard entities for the Excel application, but in the OLAP methodology, they have particular features depending on the amount of data they conserve, so their size.

OLAP methodology has an individuality character, its development takes place after many studies and analyzes, its implementation is perfectly operable in a multitude of software features, OLAP algorithms and tools are positioned in the analysis services, the particularity of S.Q.L. from S.Q.L. Server within database software. In the OLAP approach, there is no possibility of Excel spreading any field as a surface within the pivot table, operations are preconceived in the value area only for rows, columns, and filths imposed by the data specification tracked and processed; the role of the pivotal array in the chosen thinking paradigm is the informational penetration of operational actions such as revenue, size, date, product type.

Within OLAP, the creation of a pivotal table enables the generation of a process of shifting operational hierarchies and its informational crossing, the final arrangement is a more flexible and flexible structure, easier to access, interpret and process data from the top organizational management. The databases used within the OLAP concepts and paradigms are all very large in size, and a sufficient time allocation is needed to predict the expected results; a "pre-processing" process is required to achieve the proposed goal more quickly, preprocessing takes place in the files presented in "OLAP Cubes (There is an analogy with the "Rubik Cube" where each smaller cubic dump contains the result of particular penetrations).

The S.Q.L. Server is required to generate cubes. Dedicated software services are available in Excel for all the activities of economic organizations, accessible to top decision-makers.

Future of data analysis in economic applications

We identify the PowerPivot option in the Microsoft Excel software suite, this appendix is not a constituent part of Excel but is accessible as an enhancement to the identification, sampling, processing of data on the Microsoft Web site.

The benefits of the PowerPivot facility are:

(i) The ability of top decision-makers to operate large data sets without upgrades at the hardware level for the enterprise's information architecture;

(ii) Access to tools processes discrete (separate) data tables within Excel.

By entering the "Data Model" for data storage, access and processing, PowerPoint is identifiable with an Excel subset, compatible with it. We mention that the Microsoft suite provides Excel software features, which in business operational practice is called "Excel's Self-Service Business Intelligence which includes smart tools, PowerPoint, PowerView, PowerQuery, PowerMap.

Existing subsections impugn at this level implementation and use of PowerPivot, the required data set is stored in three files in their raw form; operationally most of the data is identifiable in the Access file in ContosoSales.accdb, the data to be reported is located in the Excel Stores.xlsx file, the information delimited by commas in Geography.csv. Continuing the operational-functional analysis, we find in the ContosoSales database five remarkable tables, named Data, Sales, Products, Product Subcategories, ProductCategories (Table, Attach Charts).

General presentation of the economic model operation

Within the chosen economic model at the organizational level, several remarkable operational levels are required.

A first level is the introduction of data in the data processing model specific to Excel, which is functional in stages:

(i) Opening a new "Workbook" that is saved in the Power Pivot Tutorial.xlsx.

(ii) Opening a PowerPivot window that offers the opportunity to manage operational processes at the economic organization level

(iii) The Access From option in the PowerPivot Window panel is selected, followed by a list of From Database located in the file ContosoSales.accdb, the process operation is continued by selecting all five accessible files followed by the Finish command

(iv) The ExcelFile file in the From Other Source file is selected from the PowerPivot window, this is the decisive step in importing local tables into Stores.xlsx, they become part of Data Model without connections to other similar boards positioned in the technical- economics of the organization

(v) Repeat previous reasonability and operation for importing Geography paintings from Geography.csv commas that delimit the text file, eventually inducing an add-on of datasets in Data Model without connections to other panels

The second approach involves creating operationally usable connections between the Data Sources elements. The modeling of economic processes with the help of ExcelData offers the possibility of connecting the related tables in Excel without the inclusion of software applications. In specific operations, specific to the various business models adopted by each economic organization, depending on the profile of the activity carried out, with the help of the PowerPivot facility, solving the previous problems is possible with two types of approaches.

(i) The Store tab is indicated to relate to the Sales panel using the StoreKey, CreateRelationship fields in the PowerPivot Design location is an absolute required step.

(ii) Geography tab placed in the GeographyKey field with the picture in the Store, for this feature it is preferable to use the Create Relationships dialog box.

The third layer of the process approach is the one that allows for significant changes to sampling, understanding, and data processing. The identifiable data in the PowerPivot window, located with DataModel, cannot be manipulated with the usual worksheets in Excel, there are two useful features, the possibility of elaborating-creating measures, the possibility of creating-calculating the columns; both approaches require the use of PowerPivot along with Data Analysis Expressions, the DAX language. This is a flexible, powerful and efficient IT tool, quantifiable, measurable performance targets, field measurability, or possible data model modifications are solvable through the DAX approach, as detailed below.

We identify the ability to measure the summation of the locally located fields in the Values area of the pivot array if validating the sequence of instructions with the OK option and returning to the DataView within the PowerPivot window, it is possible to highlight TotalNet Revenue, later on in the Sales tab, the process of developing top management decisions, but also for substantiating organizational strategies.

Exemplifying for an economic organization specializing in the manufacture of industrial products - production management, so the columns involving calculus are added to the Products table, for two pictures used in the description of the economic process, ProductCategories and ProductSubcategories there is a feature of their "hiding".

Using the order succession, “=RELATED(ProductCategories[ProductCategoryName])”, offers the possibility of translating a field from an associated table into the current work table used by the decision-maker management of the studied organization.

By default, the command is used to generate the Product Subcategory column “=RELATED(ProductSubcategories[ProductSubcategoryName])”, thus the manufacturing process has an exhaustive description, allowing the decision-maker to intervene on it.

This operational feature is applicable to ProductCategories and ProductSubCategories, there is also the possibility of extrapolating it to other fields deemed unnecessary for pivot tables. Thus, starting from the Power Pivot window, a first item and Pivot Table are selected from the Pivot Table list, we identify the possibility of "pulling" the fields from any arrays into a pivotal array, usually in the framework of the business model.

There are two steps to be noted in the previous procedure. If the decision-maker returns to the PowerPivot window and make changes based on the targets, such as hiding multiple fields that are considered irrelevant or expanding the computational processes to multiple columns with essential data in the operational-strategic decision. It is offered the possibility of their subsequent creation, in the pivot table are highlighted the calculated and quantified results.

This computational analysis and business data models used in the decision-making management processes specific to the economic organizations demonstrate that PowerPivot is the most useful application in the "Power" suite that can be found in the Excel software feature.

Software for viewing

Viewing data processing with concrete representations and graphical diagrams provides the best understanding for the human decision-maker in imaginative mode. In Microsoft, Excel identifies tools that have a degree of excellence in management processes such as Power View and Power Map with easy results to transpose and graphically. In the specialty studies, there is the possibility of accessing Tableau Software's Tableau Public application where it is possible to see an overview of the general situation within the specialized market. The ultimate goal of all these IT applications is the graphical presentation of their data and processing processes, any more or less hidden tendency is identified and presented with much clarity.

Challenges in Data Mining Processes

Data Mining processes are operative to analyze enormous quantitative data structures, followed by processes of extracting information that is considered relevant to the work of the economic organization or in the process of managerial decision-making. Similarities but also major conceptual and paradigm differences are identifiable between Data Mining and KNOWLEDGE DISCOVERY, the latter identifies models, valid, innovative data procedures, usable in current organizational operations.

DATA MINING is a significant step forward in discovering information through models positioned inside data using predictive techniques.

For business modeling, we identify two types of software entities:

(i) DATA MINING tools-algorithms and procedures generally usable in economic applications;

(ii) DATA MINING applications - techniques within a special application-developed a particular challenge, specific to a specific type of market and required by a certain type of users, final clients.

Predictive analysis at an economic organization level cannot be achieved in the modern, globalized economy without the concepts and paradigms of DATA MINING.
Their flexibility and precision increase the degree of economic efficiency, identifying the components of DATA MINING and KNOWLEDGE DISCOVERY & DATA MINING presents the following structure:

(i) The model-a function positioned in a one-dimensional or multi-dimensional parametric space, the representation is made by a linear function, of probability or fuzzy, we remark upon its elaboration the use of classification or clustering algorithms;

(ii) Preferential criteria-they have different nature, they induce the principle of order, interpolation, numerical approximation;

(iii) Selection algorithms-lead to selecting the model used, data used extracted from the database, preference criteria.

The most important problem that occurs in the application of DATA MINING concepts and issues is the classification, a challenge similar to the regression analysis, descriptive by the use of explanatory variables to make a prognosis for a variable with a certain degree of dependence. There is the possibility of two decisions, YES and NO, but it also appears opposed to a third way of INDEPENDENT, this frequently occurring challenges in the operational transposition of different business models is approachable with different types of algorithms, of which attention we grant a method using mathematical regressions. At the beginning of the developed study, the role of DATA PARTITIONING in the classification and concrete applications of DATA MINING, the analysis of this role goes from the fundamental mission of the concept, that of exploring data sets with hundreds of thousands or even millions of stored records.

There is a possibility, very useful, of sharing the data set in two or three sets, before the implementation and the operation of the chosen algorithm and procedures, corresponding to the organizational business model. From the total data volume, each subset has a certain percentage, which is a scientific analysis we choose randomly. We call the first subset, which has a majority of all data, "TRAINING DATA SET", the second subset is called "TESTING DATA SET", it contains the rest of the data, the cardinals of the two subsets gathered

algebraically generates a numerical value equal to the cardinal of the chosen set; each subset have known values of the dependent variable.

The algorithm used is trained to work with TRAINING DATA SET data, so a business model usable in the classification process is developed, a model that is later tested using the TESTING DATA SET subset. The top organizational decision-maker can be very high, but the fundamental question that arises is the level of flexibility of this model to be able to satisfy the classifications that occurred in the TESTING DATA SET.

We mention the specialized software such as StatTools and PalisadeNeuralTools, the latter with increased utility in Artificial Intelligence applications, operation and strategic development of business models there is also the possibility of specifying a third subset, a subset of data used in the predictions, called PREDICTION DATA SET, the chosen model is subsequently used to develop a process of classification of unknown values. The correctness of the predictions thus made is known only at the time when the set of real values of the dependent variable in the prediction data set, PREDICTION DATA SET.

Individual classification of economic processes

The most widespread method of individual classification used by the managerial decision-maker is the method called the "logistic regression", the process of estimating that a certain entity exists in a certain category - the set is probably quantified, the use of a nonlinear function is required. The logistic regression model or "Logit Model" is a statistically-probabilistic model used for the already existing sets in the study of the phenomena that are faced, in the present case, whether they are produced or not, "Pass-Fail", "Win-Lose".

The binary dependent variable is modeled with a logistic function, regression, thus estimating the set of values in the form of parameters of the logistic model searched; analyzing strictly mathematically, the binary logistic model, it has a dependent variable of two values, assimilated to the classical logical TRUE / FALSE, from which an indicator variable, "0/1".

Formalizing, fie $X_1, \dots, X_k$, the string of potential explanatory variables, we choose the linear functional structure $b_0 + b_1X_1 + \ \dots + b_kX_k$, but there is no certainty that this is quantifiable between 0 and 1, rather than a probability, a choice of value identification between "FALSE OR TRUE" is made with 100% probability for nonlinear formalism

$$1/(1 + e^{-(b_0+\cdots+b_kX_k)})$$

The resulting function, $f(x) = 1/(1 + e^{-x})$ is an "S-logistic profile", as in figure ..., for negative large values the approximate curve of the curve as having the value "0", for positive values of x as having "1", the analyzed model, the logistic regression using in its operational analysis, this type of probability estimation function, for any type of observation located in category 1, if p is the probability of starting in Category 1, the model $p = 1/(1 + e^{-(b_0+\cdots+b_kX_k)})$ is considered an estimated value.

The equation is algebraically processed until it is presented under formalism:

$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1X_1 + \ \dots + b_kX_k$, natural logarithm $\ln\left(\frac{p}{1-p}\right)$ is the linear function of the potential explanatory variables, the ratio $p/(1-p)$ is called bad ratio, “odds ratio”.

The odds ratio has a broad expansion, depending on the specialized markets and the main business object of the studied economic organizations, which may be a company with a risk of insolvency-bankruptcy p, the odds ratio is $\frac{p}{1-p}$.

The logarithm used to express the oddos ratio, the quantity on the left side of the previously presented formalism is found in specialized scientific literature and under the name of "logit" or "log odds", in which case the logistic regression model shows logit as the linear function with explanatory variables.

The analysis of the mathematical formalism used reveals as a fundamental goal the correct understanding and interpretation of the regression coefficients used, thus:
If the coefficient b is positive, then if its X coefficient increases, the logarithm corresponding to odds will increase, the probability of positioning in category 1 also increases;
For a negative b coefficient, we get the proposition that denies the expression in the point (i);

It should be observed in the developed study that the explanatory variables are positively correlated with the beginning category 1, so b is positive, and which are positively correlated with the beginning in the group 0, so b is negative. The intensity of the coefficients b determines the importance of the mathematical formalism developed for the different organizational business models of the numerical values X, as well as the choice between the two 0,1 groups.

The use of other mathematical concepts and paradigms for the same issue, for example the case of "regular regression," assumes in its operational transposition the same vulnerabilities and challenges for top decision maker.

Classification by logistic regression is expressed by the parameters bi of the model that minimizes errors for all sets presenting concrete operational examples. It becomes obligatory to find the parameters b, which are most likely D data, so it is desirable to maximize a formalism of type, $p\left(\frac{B}{D}\right) = p\left(\frac{D}{B}\right) * \frac{p(W)}{p(D)}$, with the Bayes rule.

We continue to consider a uniform probability for all of the chosen values of b, the term $p(b_0, \dots, b_k) = c$ is of constant algebraic value, in most of their business applications it is neglected.

The maximization process, therefore, makes reference to:

$p(D|B) = \prod_i p(y = y_i\,|B, X_i)$, where i is a variable for all business models chosen for organizations, elements in the macro-economic environment.

If we introduce a transformation of a y which takes two discrete 0.1 values, a Gauss-type error assumption occurs $p(y = 1|X) = (0|X) = \frac{1}{1+e^{b*X}}$, which means further,
$p = (y = 1|X) = \frac{e^{B*X}}{1+e^{B*X}} = \frac{1}{1+e^{-(B*X)}}$, so $(y = y_i|X) = \frac{1}{1+e^{(1-2y_i)(B*X_i)}}$, the process of maximizing probability is reduced to finding a b that maximizes mathematical formalism
$\sum_i -\ln(1 + e^{(1-2*y_i)(B*X_i)})$,

A convex process optimization procedure that provides the optimal solution of the stable type, flexibility and efficiency is therefore identifiable.

The logistic regression is usable with a higher degree of operational performance in the probabilistic data models, along with the characteristic characterization of the influence of the variables is especially important for the set of regression coefficients used in the practice of managerial decision making, b_i. The primary objective of logistic regression is to give a score to the items (members) studied by the values of Xs; this score is made up of the estimated values for each "p" found from the logistic regression equation, solved as an algebraic equation after p. The classification of members with a high score is placed in the "1" category, the ranking of the ones with a lower score is placed in the 2nd category.

By default, the total of the members who are in the "0" category are below the required limit, the rest are in the "1" category, in operational practice the value most commonly used in these applications is numerically equal to 0.5, generally any numerical value is accessible software, which we consider to be beneficial for organizational decision-maker top management to use it for different business models is StatTools; it is best to implement the logistic regression procedure.

Approaching the classification of probabilistic economic processes

A form used also in the classification of processes generated by economic models is the "Bayes Classification", it has a high degree of analogy to linear regression, but using a probabilistic inference methodology. From a causal structure with a single initial cause, Y, which is not subjected to a direct observation process, identifying k consequences and Xi observable variables the consequences as being conditionally independent, resulting in Y data, resulting

$P(X_i|Y, X_j, j <> i) = P(X_i|Y)$, with results easy to see.

In this approach, $P(Y)$ is proportional to the product of the conditional probabilities:

$p(Y|X_1, \dots, X_k) = \alpha p(Y) \prod_{i=1}^{k} p(X_i|Y)$

It is possible to estimate P (Y | X), the learning process within this approach requires a simple estimation of conditional probability $P(X_i|Y)$.

Business processes modeled by neural networks

The advantage that the neural network provides to predictions for economic phenomena is the degree of precision in relation to the rest of the approaches, especially on relationships with a high degree of non-linearity, the details of this process are transposed at the software level with the .xlsm files because the logistic function is interpreted as a macro command. The chosen neural network is likely to have any number of hidden layers and nodes, the process of choosing them is not a very obvious one, choosing within economic organizations is made with the help of specialized software.

Recent studies show that algorithms implemented with the help of neural networks are in a continuous process of adaptation, improvement, upgrading, unchanged, but the concepts and the paradigms of thinking in which they are positioned. It should be mentioned for this type of application that Palisade suite is accessible specialized software called NeuralTools, particularly useful and generates added value maximized and superior organizational economic performance.

Case Study: Protect spam mail

In operationalization of on-line communication within an economic organization, there is the possibility of predicting the analysis of an e-mail if it is an unwanted message, that is, of the type spam or not, (Y), we make the appreciation over the set of names and adjectives (Xi) of e-mail as a cause-by-cause approach. For a correct selection of these messages, a sampling process that determines spam and non-spam classification needs to be developed, so we have to estimate the probabilities $p(X_i|Y)$ and calculate the occurrence frequency of each word in each of the two classes.

Once the process is completed, it is estimated the probability that precise communication will be designated as spam or not, correlated with mathematical formalism, for an event Y, which has k consequences $X_1, \dots, X_k$), computable by

$p(Y|X_1, \dots, X_k) = \alpha p(Y) = \prod_{i=1}^{k} p(X_i|Y)$, (*)

where we obtain the value of α making the sum of the probabilities for all the possible values of Y that must be equal to 1, $\alpha[p(Y) \prod_{i=1}^{k} p(X_i|Y) + (1 - pY)) \prod_{i=1}^{k} p(X_i| - Y)] = 1$, the use of Bayes' rule determines that this network is the name of the Bayesian network.

For the application presented, spam filtering is estimated the probability that a message is 100% spam by applying mathematical formalism (*).

For a selection of messages based on the quantified probability value, a comparison process is required, so the value of the $p(Y)$ which may be the same for each message, this is an algorithmic structure with procedures commonly used in spam filtering operations targeting electronic mail between economic organizations. The method of induction by extrapolation, generalization is a creation by Pat LANGLEY, recalled in "Elements of Machine Learning". Morgan Kaufmann (1996), also in the paradigms used for this type of application, the concept of perceptron introduced by Rosenblat (1958).

The concept of "support vector machine" also has a significant significance in the issues addressed by Vanpik (1995). With the concepts and new paradigms introduced in these papers, the approach of concrete applications used in the strategic operational activity of modern economic organizations gains another image, both as an informational dimension and as a scientific depth.

The problem of blocking e-mail messages that are not required at the level of a modern economic organization has a negative connotation, both from a technological and informational point of view, but also from the financial point of view of the induced costs. It is advisable that the IT department, in close contact with the organizational R&D department, implements, adapts specialized software within the organization such as "SpamAssassin", widely used in spam filtering, or modified versions of it to the peculiarities and constraints generated by the type of business model used and the economic activity unfolded.

We find integrated underlying software tools, algorithms, and facilities specific to the quasi-totality of the challenges coming from the modern business environment, for the management of electronic mail. We also note that the vast majority of SpamAssassin's identifiable anti-spam facilities allow the development of spam, non-spam classification processes, with end-user classified messages. Certain computer applications have purely scientific, statistical correlations, such as Bayesian methods. In upgraded versions of business environment dynamics, SpamAssassin achieves scientific, operational performance for business models chosen through the implementation and use of the perceptron concept, thus facilitating a weighting according to various criteria imposed by the end-user of this technology.

Using the support vector machine within these approaches offers increased performance, as a difficulty that may arise at this level is the high degree of complexity of this technology (http://spamassassin.apache.org/).

References

Albright, S.C., & Winston, W.L. (2018). *Business Analytics, Data Analysis and Decision Making*. CENGAGEbrain. Retrieved from www.cengagebrain.com.

Anthony, M., & Bartlett, P. L. (2009). *Neural network learning: Theoretical foundations*. Cambridge University Press.

Hammond, J. S., Keeney, R. L., & Raiffa, H. (2015). *Smart choices: A practical guide to making better decisions*. Harvard Business Review Press.

Nilsson, N. J., & Nilsson, N. J. (1998). *Artificial intelligence: a new synthesis*. Morgan Kaufmann.

Nilsson, N.J. (2010). *The Quest for Artificial Intelligence: A History of Ideas and Achievements.* Cambridge/New York, NY: Cambridge University Press.

OWL Web Ontology Language Reference. World Wide Web Consortium, 2004.

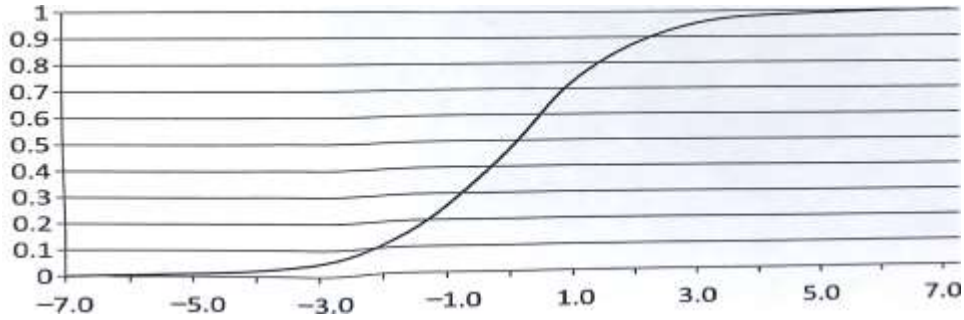
Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386.

Russell, S.J., & Norvig, P. (2003). *Artificial intelligence: a modern approach*. 2nd. Edition, Pearson Education.

Sowa, J. F. (Ed.). (1991*). Principles of semantic networks: Explorations in the representation of knowledge,* San Mateo, CA: Morgan Kaufmann.

Vapnik, V. (1995). *The nature of statistical learning theory*. Springer Science & Business Media.

Annexes



S-Shaped Logistic Curve

	A	B
1	Row Labels	Revenue
2	Drink	
3	Alcoholic Beverages	
4	Beverages	
5	Carbonated Beverages	
6	Drinks	
7	Hot Beverages	
8	Chocolate	
9	Coffee	
10	Pure Juice Beverages	
11	Dairy	
12	Food	
13	Non-Consumable	
14	Grand Total	

Hierarchy in the Foodmart Database
Source: Albright, S.C., & Winston, W.L. Business Analytics