SPECIFIC MACHINE LEARNING ALGORITHMS AS EFFICIENT SOLUTIONS FOR COMPLEX BUSINESS PROCESSES

Marian Sorin IONESCU, Olivia NEGOITA
FAIMA, *University POLITEHNICA Bucharest,*
Splaiul Independenței 313, București 060042, RO
marian.ionescu@man.ase.ro

Abstract
This paper presents and critically investigates some unsupervised and supervised algorithms specific to machine learning, which are particularly suited for business processes and their high dynamical evolution. We particularly highlight the general merits of machine learning and mathematical foundation in the big data era and its importance for applications to industrial and business processes, in the context of well- defined economic problems, with their variables and, especially, the predictive functions to be solved. Then, the paper discusses some existing algorithms and their mathematical formulations and merits, based on existing literature and recent usage in various economic applications. Such machine learning algorithms form today a calculation distribution framework, available in open-source software for storing data and running applications such as Handoop. We also reveal the importance of robustness and of data quality, as many of the business ecosystems are currently filled with dirty and useless data, Finally, we suggest that for precisely observing, quantifying and predicting the performance progress of organizational management, transparency for the users of machine learning is required, as opposed to traditional rather opaque machine learning processes. Finally, the proposed algorithms overview and discussion is helpful to provide first steps in learning how to apply machine learning to make your business more efficient, more effective and more profitable.

Keywords
Machine Learning, Business Process, Algorithms, Optimization, Transparency.

Introduction

Big data mining and analytics are playing an important role in knowledge discovery and decision making in the process industry and economics (Ge et al, 2019, p.20590). The modern business problems are today positioned in complex ecosystems characterized by high dynamics and challenges. Some of the emerging engineering scientific fields concern progress in computational engines applied to big data analytics in production and economics: statistical learning (Hastie, Tibshirani, & Friedman, 2001), compressed sensing (Donoho, 2006, p. 1289) machine learning (Shanga & You 2019, p.1010), deep learning (Anthony, M. & Barlet, P. 1999) and Artificial Intelligence (MacKay, 2003), are today applied to a large variety of domains, including the complex business processes.

In this late case, the decisional problems within different business models adopted by the top organizational management, based on the information and unique data, are confronted with an operational experience that may be practically unrepeatable. In this paper, we investigate the variables and equivalent economic performance factors that are defined in an economic organization to be optimized and/or predicted with various machine learning that is subsequently discussed.

The identification at the level of the final target for the subsequent transposition of the common tasks in operationality for the business problems is an improving decisional factor. There is also the possibility of segmenting the organizational activity by intermediate targets, which are mandatory to be reached until the final target. These should be broken down into intermediate sub-targets, which are easier to be reached, monitored, the global understanding of the problem being a simpler process.

In this work, the main research objective is to present some of the machine learning (ML) algorithms suitable for complex business processes and critically discuss their advantages and vulnerabilities, reflecting their status.

In Figure 1 we summarize the current state of the art of various existing ML algorithms, classified in *unsupervised* and *supervised* machine learning and their relation and impact on dimensionality reduction, data classification, process diagnosis, quality estimation and prediction of key process performance. For instance, unsupervised ML requires no human intervention and training and it would allow applications that get customized for specific uses and for individual people on the spot, *without pre-learned parameters*.

The methodology used in this paper is encompassing (i) first, an investigation of the variables needed to formulate economic problems, the importance of the used big data in economic and business, as well as their mathematical formulations, (ii) the presentation of a series of selected algorithms that are currently used in complex business processes and their mathematical background, easiness of use, advantages and vulnerabilities.

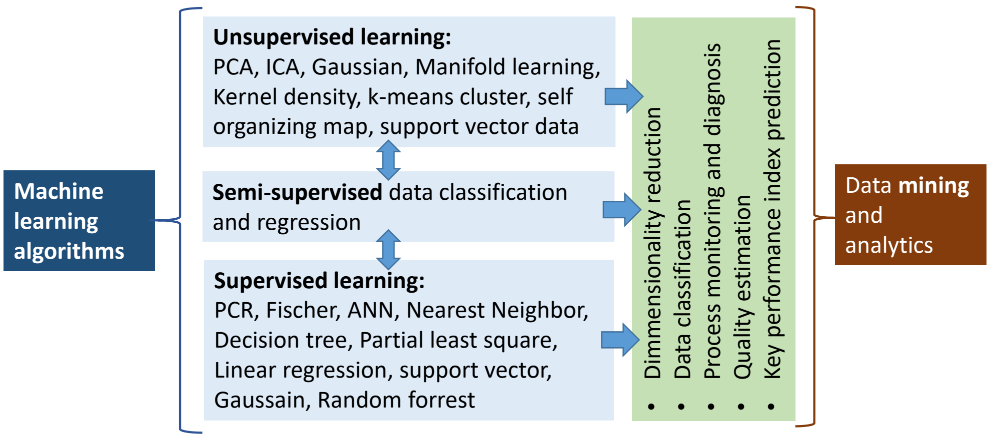

***Figure 1**. Machine learning algorithms in unsupervised and supervised learning and their relation to certain desired features of data mining and analytics.*

Variables and economic problems: fundamental mathematical formulations for unsupervised and supervised Machine Learning

By strategically approaching the problem and business model implemented at the economic organization's level, one should first identify the variables playing a significant role in the behavior of the final customers. Below, we list the main factors necessary to be considered for the economic performance of the organization.

(i) The estimation of the probability that a final customer may belong to a certain set or population for the targeted hierarchy.
(ii) The estimated value or “regression” makes an individual prediction of the numerical values and characteristic variables, classification predicts whether a certain event takes place - such regression shows the amplitude of the probability that this thing will happen;
(iii) Data providing identification similarities for people in a limited number of attempts, similar resemblances identify similar entities, features, and behavior;
(iv) Grouping events, coincidences, frequent analysis of items, associated operationality rules, analysis specific to the basket-market portfolio, correlation of the performed transactions;
(v) Behavioral description of the consumer, final customer, “profiling;
(vi) The prediction connection is the one forecasting the connections between data flows specific to the used items (entities), there is also the possibility of describing the “strength” of these connections;
(vii) The data flows restructuring by replacement of certain overwhelming sets of data with subsets of much smaller cardinality where information with a significant degree of relevance are conserved, by operationalizing them, the top organizational management has an increased speed of elaborating the decision and construction of the related strategies;
(viii) The “Casual Modelling” concept is involved in the decision making process in understanding which economic events or management actions are inter-influential.

It is worth noting that main aim here is a “predictive analysis” of the economic organization’s operational activity, which means *the anticipation of the evolution of the strategic vectors within the competitive ecosystem* for obtaining a competitive advantage. Today, analyzing big data using machine learning algorithms is a crucial method to help organizations to forecast future trends in the market, adapt and be more efficient. For manufacturing company such type of ML applied to big data may analyze the demands for the next year with the goal to predict future sales as profile and as amount and adapt the production. This requires collection of big valid and timely data and select data models incorporate predicting key parameters concerning the customers, the competition, and anticipating the demand for the products existing in the market. Such analysis of big data is expected to improve both the company strategy and decision-making process of the businesses in a very competitive global economy.

In this context the machine learning framework represents a proactive and appropriate science-based response for all the challenges faced by an economic organization. For instance, one can quote a few representative examples of relevant economic problems for machine learning algorithms:
(i) Detecting the behavior of the end consumer for a certain portfolio of products and services;
(ii) Detecting fraudulent financial operations in the virtual environment;
(iii) Anticipating costs for the transformation of an online commerce website according to the number of views;
(iv) Evaluating the insolvency risk for certain customers, scoring functions estimated in real-time;
(v) Predicting the interruption of a trade contract from the current operational activity.

Predictive analysis values all the opportunities of these technologies, having as innovative part is the vector leading to the accelerated development of the targeted features. In the world of business models and maximizing organizational profits, we identify two great challenges of this new type of algorithmic approach, as follows:

(i) Hardly predictable human behavior, which exceeds the totality of deterministic predictions;

(ii) In most of the real situations, the identification of a descriptive, explicative model should be relinquished, correlations between observations and data and information identification are identifiable, they may originate from the economic history of the studied organization;

In practice, the goal is to translate this process into mathematical specific variables, categorized as below:

(i) **Predictive variables**, attributes or parameters specific to economic organizations and their history, p variables that are predictive and associated with observation are components of a vector $x = (x_1, \dots, x_p)$ with p components, an assembly with N observations is consisting of the vectors $x^{(1)}, \dots, x^{(N)}$.

(ii) **Target variables** that should be reached, at this value events that take place in the business macro-environment are observable, generally noted with y with the same significance for indices as well as for the vector x.

Here, we consider that the observed value y of the target variable is an implicit result of the superposition of the two contributions:

(i) A **function $\boldsymbol{F(x)}$ of the predictive variables**, a fully determined contribution of the variables x that belongs to the observation process (the equivalent of a signal that should be highlighted);

(ii) **A random noise $\boldsymbol{\varepsilon(x)}$**, which incorporates the combined effects of a large number of parameters that must be taken into account in the implementation of business models;

It should be noted that both F as well as ε remain unknown but the *objective specific to machine learning are to elaborate an approximation of the Function F starting from a set of observations*, this approximation, noted with f is designated "predictive function".
In practice, the elaboration of the economic organizational decision is very important, the evaluation of a value of a signal in relation to a noise, the probability that a certain effect might be randomly generated, statistical experiences show that a level of 5% indicates the fact that the observation is 20% generated by hazard.
Therefore, the Machine Learning (ML) approach mathematically represents a set of statistical and geometrical facilities correlated with procedures and information algorithms that allow the generation of an automation process ended with the definition of a predictive function f, the entire process is initiated by a set of observations called learning set.

The ML is finally a hybrid approach that takes over the advantages and facilities of the statistical analysis, Artificial Intelligence, Intelligent Business. In this context, **a Machine Learning Strategic Operational Model** is a specific algorithmic systemic structure that defines a predictive function f starting from a set of data specific to learning processes. The elaboration of this function f practically defines the learning process or training of the built model, the prediction itself is actually an evaluation of the function $f(x)$ of the function f with predictable variables of an observation x. It should be emphasized that ML allows the elaboration of systems that define and learn the own functioning rules, the identification of a program within which we find the predictive function f starting from data sampling and interpretation and finally, ML modelling allows data association and analysis.

In case of economic applications, we identify a particularity of the ML applications, namely the occurrence of the "over-learning" phenomena, which is operationally a situation in which the ML model overlaying the adopted business model reproduces with a high precision the data flow used in the learning process but does not allow subsequent extrapolations. We thus identify a challenge of this type of paradigm, the one of separating data into "training data (learning)" used in learning processes and "testing data", those evaluating the predictive models' quality. Operationally, in case of economic entities, processes of "crossed validation" are developed.

The foundation of this type of approach consists in the random separation of the initial data, in k parts of equal dimensions, for each of them a model is implemented for the data sets except for those that are used to predict the main business model agreed by the economic organization. At the end of the k ML approaches, the prediction process for each part is compared to the values obtained from the already-completed operational practice, a moderate value on a set of data is sampled from the estimated observations.

The careful analysis leads us to the primary observation that the volumes specific to "Big Data" may be symbolically characterized by "$V = Acronym\ Volume\ VVV$".

It is also identifiable the difficulty of the organizational entities of information technology to implement and systemically operate structures of storage and parallel treatment of the information. Another analysis suggests that "V" is the symbol for an overwhelming "$Variety$" (enormous dimensionality) is the set of parameters p characterizing each economic observation (status).

Statistically evaluating these situations, one can observe some important features of data, observations, and parameter dimensionality that should be considered:

(i) Dimensionality vulnerability is the one showing that each observation of a sampling used in the process of learning by machine, computer, sequencer, involves the use of a set of "p" parameters, statistically representative samples for an evaluated population, we assume it may have the dimensionality 10, implicitly in the elaboration of the top organizational decision, 10^p observations are necessary so it may have an important relevance degree.

(ii) Technologies used in the modern operationality are generically aiming at a dimensionality reduction, the information used is compressed and non-relevant or redundant parameters are eliminated from the systemic processes;

(iii) Fictional correlations in terms of images, views, set with high dimensionalities specific to Big Data show that is it highly possible that random phenomena may induce a higher degree of decisional influence;

(iv) The parallelization of algorithms is necessary because computer procedures face enormous quantities of observations and data.

The treatment of data representative for the economic organization involves the development of parallel IT architectures of the type provided by Handoop or MapReduce facilities, facilitating the selection, filtering, conversion, normalization and detection of a potential incoherence of the data flows. A significant principle of the paradigms specific to Big Data is the one showing that the optimum economic performance is generated due to the less sophisticated algorithms but with the capacity to operationalize large quantities of data. We do not consider that there is an objective choice for the interpretation of statistical data for the problems specific to business models, a subjectivity degree is always present and the simplest explanation is generally the most plausible in the economic phenomenology.

In the field of ML applied to industrial processes and operational business, two fundamental types of learning processes called Supervised and Unsupervised Learning (see Fig. 1) are used together with their corresponding algorithms:

(i) **Supervised learning** is the form most used by the applications of Machine Learning type, it is made on the supposition that the deciding top management has a limited set of examples characterized by predictive variables $x^{(1)}, \dots, x^{(N)}$ for which target variables $x^{(1)}, \dots, x^{(N)}$ are known, the target in this learning step is to extrapolate the association between the explicative variable and target variable for the definition of a predictive function "f", the most used applications by this type for the business models are linear regressions and machines of support vectors;

(ii) **Unsupervised learning** (or "$clustering$") does not involve the existence of any previous labeling of learning data, the target is that the system is guided by itself and regrouping into categories of examples ($k - mean$ algorithm belongs to this category).

Key algorithms specific to Machine Learning: basics and discussions

In general, it is acknowledged that no ML algorithm has an exhaustive applicability. In practice, the decisional head-office is responsible for having in its staff experts in mathematical statistics that may develop an anticipative intuitive side, complementary to the Machine Learning mechanisms and procedures. In the following we describe and discuss some of the most used algorithms in ML.

Linear regression

This is one of the simplest approaches applicable and operational for certain business models. The mathematical formalism involves the use of a predictive function f bonding specific variables $x_1, \dots, x_p$ to the target variables, in the form of the convolution:

$$f(x) = \sum_{i=1}^{p} a_i p_i + b = a \times x + b,$$

In operational practice, b is replaced by $a_0 x_0$, where $x_0 = 1$.

Learning with this type of ML model in the algebraic calculation of all the coefficients a, b that are involved in minimizing errors and defining that the sum of the squares predictive values $f(x^i)$ and observable values y^i, Linear Regression is also known as the Method of the Smallest Squares. It is often used in the prediction of organizational models for marketing policies and also in the elaboration of econometric models.

The main advantage of the approach using Linear Regression is that the automatic learning process is practically the inversion of the matrix built on learning data, we thus identify an explicit mathematical formalism, easy to be numerically evaluated.
Identifiable vulnerabilities occur from the possibility that the linear dependency does not exist, the model may induce sometimes irrational values of learning data. Also, a big limitation is that the linear character of the model neglects all interactions between predictive variables.

The method of the nearest "k" neighbors

The method of the nearest "k" neighbors known as K Nearest Neighbours(K.N.N.) is a standard algorithm of supra-target classification, non-parametric, very well adapted to the problems of modern business models. The concepts and fundamental paradigm consist in the assumption that observation is similar to all the observations made near it, the assumption is that each observation made for a set of training is represented by a geometrical point in a space with several dimensions, "y" is symbolically the representation of predictive variables (Figure 2). The hypothesis is also made that there is a conceptual definition of distance in this space, the closest points "k" are searched in relation to the one for which the classification is required, the set of target variables is the majority among the closest points "k".The defined algorithm covers several variants, each of the "k" observations are interpreted depending on their distance to a new observation, the observations at the greatest distance are irrelevant operationally.

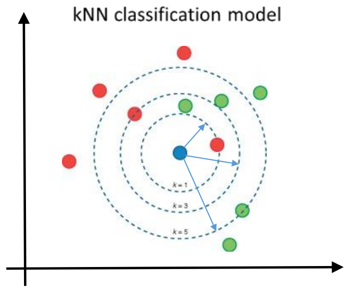

Figure 2. Illustration of a kNN classification model;
k has a strong effect on the performance

As an example of using this type of algorithm, we quote the case of the economic organization $NetFlix$ for the provision of the score granted by a film fan to a movie found in its portfolio, subsequently, there is the possibility of developing an appropriate portfolio-offer. Like other applications, Li and Zhang (2014) developed used k-neighbors rule technique for semiconductor manufacturing process fault detection and this was also used for the performance of fault diagnosis in batch processes, Ge at al., (2017).

Among the advantages provided by the method of the nearest "k" neighbors, we identify that at the data level assumptions should not be made regarding the distribution type used, a concept existing, the proximity (vicinity) one between similar entities that are grouped.

At the same time, we identify as vulnerabilities the sensitivity of the procedures and algorithms specific to noise, the fact that the operationally transposed learning processes have a strict local character, it could not be subject to strategic developments in wide ranges of time. In the case in which the deciding top management faces a high number of variables (for instance, >10 distance) the calculation becomes one with high allocated costs, so the use of the dimensionality reduction is needed.

Naïve Bayes Classification

This approach involves the elaboration of a specific supra-target algorithm of classification with a high-performance degree and a lot of simplicity in its transposition into operation. To understand its operation mode, we evaluate an anti-spam informatics filter, for each message, the filter performs a predictive activity for dual messages, one of "non-spam" type and the other of "spam" type, the frequency of the analyzed words is the one sending messages in one of the two classes. Having representative samples of messages, the estimation processes for these messages are relatively simple to be made, without inducing other additional hypotheses.

The analysis of this type of approach is focused on the occurrence frequency of the independent words, which is sending the messages arrived in the two sets, spam and non-spam. Positioning ourselves on the criteria of independence of such systemic operationalities, we find that they are false, which is why the adjective "naive" is used for this type of approach.

Naïve Bayes Classification is extrapolable in situations in which relations of conditioned dependence have a relevant complexity degree, their associated models are designated as Bayesian type of networks. The advantages of this type of approach are simplicity and flexibility of the algorithm in case of the independent hypotheses. Among vulnerabilities, we identify that the probabilistic predictive processes for different sets induce errors when the conditional independent hypotheses are not true.

Logistic regression

This is an algorithm of linear classification using the mathematical model of linear regression in cases when the target variable "y" has only two binary values "$0, 1$", we recommend the use of a function designated "$Function\ Score, S$" for predictive variables, expressed by the mathematical formalism:

$$S(x) = \sum_{i=1}^{p} a_i x_i$$

A searching and analysis process of the coefficients $a_1, \dots, a_p$ so the scoring function $S(x)$ is positive in case the appartenance probability at the first set is high or negative in case the appartenance probability to the second set is high, between the two an appartenance probability $P_1(x)$ to the first group is included, a function of logistic interpolation with the explicit form:

$$logit(S) = 1/[1 + \exp(-S)]$$

We consider it is the most appropriate mathematical formalism to be used in this situation in which the predictive variables $x_1, \dots, x_p$ are normally distributed, disjoint for each set.

The parametric probabilistic model is represented by the mathematical formalism:

$$P(y = 1|x) = lit(S(x)$$

The search for optimum parameters $a_1, \dots, a_p$is performed by numerical methods, totally different from the linear regression, without explicit calculation algorithms.

Among the method's advantages, we indicate a new observation to be quickly emphasized is the evaluation of a function of linear score S, the algorithm is eligible for processes of over-learning, the interpretation of a_i coefficients are very facile, the appurtenance probability to a certain set is also easy to predict in the business models chosen by economic organizations. Among vulnerabilities, *the use of the linearity hypothesis, blocking the interactions between variables*, thus leading to a process of recalculating the already-defined variables, the learning phase may take longer, along with the numerical operations for optimizing coefficients, the algorithm shows limitations of the binary targeted variables.

k-mean specific algorithm

k-means specific algorithm is the only one selected here from the set of **unsupervised algorithms**. Starting from an assembly of "N" observations $x^1, \dots, x^N$ each described by p variables, this algorithm creates a partitioning in k sets or homogeneous clusters, each observation having a correspondent point in a space with p dimensions. The algorithm finally aims to identify the partition of the N points in k clusters in which the sum of the squares of the distances between points in the group's center of gravity has minimal value (See Figure 3).

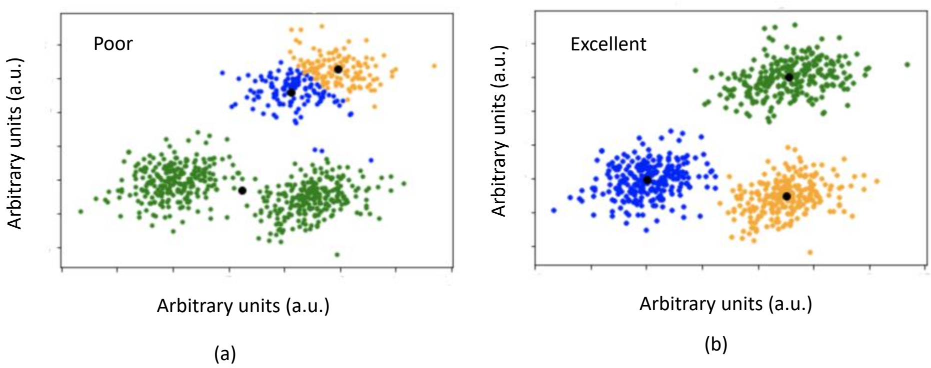

Figure 3. Illustration of K-means sensitivity to the initialization of the mean points. So, if a centroid is initialized to be a "far-off" point, it might end up with no points associated, and more than one cluster might end up linked with a single centroid

We notice the very high degree of complexity of this type of mathematical approach, the only possibility is a heuristic algorithm providing solutions with a certain approximation degree, after a prototype with k points is initially chosen, the algorithm is operationalized with a successive iterative structure.

We distinguish the following series of advantages for this type of algorithm: fast operability for economic organizations, the possibility of multiple launchings of the algorithm for the verification of the elaborated partition's correctness. Also, the algorithm shows the facility in processing large volumes of data, cases specific to the problems of the business ecosystem, correlated with other algorithms of the same type, facilitates the discovery of auxiliary data categories that may influence future observations involved in the elaboration of organizational strategies.

Among vulnerabilities we mention that the selection of the k sets represents the user's choice, an optimum way is not identifiable, there are negative influences in this regard, the obtained solution depends on the rule of the k points chosen at the beginning.

Decision trees

The decision trees are typical models of supra-target and non-parametric ML**,** used in the classification and regression processes specific to the systemic approach of the economic processes. No type of probabilistic model is used in this type of approach, the method being purely algorithmic. The fundamental idea consists in the classification of an observation based on a sequence of questions, operationality also designated "$Segmentation\ criterion$", which involves the predictable variables x_i, of the observation, each question represents a knot in the decisional tree.

The set of target variables is determined by a finality process represented by a final tree knot, where the final observation gets after the set of questions. The learning step consists of finding correct questions addressed for a set $x^1, \ldots, x^N$, of the observations N covered by labels $y^1, \ldots, y^N$. Finding an optimal tree in the meaning that minimizes the number of addressed questions is a difficult mission in the meaning of the highest complexity degree, the processes of dimensionality reduction must be therefore initiated, thus generating the occurrence of a heuristic method. The target is that the "leaves" of the tree are so homogeneous that they will ideally contain only x^i observations from a single set y^k.

The developed strategies consist of the association of the knots with well-defined criteria of segmentation, generating a cascade decision process with progressive growth.

Hereinafter we present a synthetic table of the business activity of an economic organization, the learning ML process of the decisional trees**,** the type is evaluated and structured during a calendar year, in which four predictive variables and one target variable, considered to be representative, are chosen.

Table 1**. **Example of a decision tree.

No.	Forecast	Sales	Advertising/Marketing	Ecosystem	Activity
1	Average performance	Average	Normal	Negative influences	Operational
2	Average performance	Average	Normal	Without influences	Operational
3	High performance	High	Aggressive	Without influences	Operational
4	High performance	High	Normal	Positive influences	Operational
5	Average performance	Average	Normal	Without influences	Operational

6	Average performance	Average	Normal	Without influences	Operational
7	Average performance	Average	Normal	Negative influences	Operational
8	Average performance	Average	Normal	Negative influences	Operational
9	Average performance	Average	Normal	Without influences	Operational
10	Average performance	Average	Normal	Without influences	Operational
11	High performance	High	Aggressive	Positive influences	Operational
12	High performance	High	Aggressive	Positive influences	Operational

The advantages induced by this approach for systemic economic processes are that explicative variables are perfectly possible both qualitative as well as quantitative, data preparation step has a reduced dimensionality if the interactions between variables are considered linearity hypotheses is no longer necessary.

Among the vulnerabilities identified for this method, we shall indicate the over-learning risk, in case of an incorrectly elaborated tree, the segmentation criterion negatively influences the entire predictive process, thus data are erroneously interpreted, generating false organizational strategic conclusions.

Random forest algorithm

This type of algorithm (Breiman, 2001) has in its structure most of the advantages generated by the approach using decision trees. The vulnerabilities induced by over-learning phenomena and high complexity are eliminated. This approach is built on three basic ideas (see also Figure 4), as follows:

(i) Starting from an initial sampling N of the observations $x^1, \dots, x^N$ in which each is described by the predictive p variables, a new B sampling is created, with the same N dimensionality, this is known in the operational practice of the organizational business models under the name of "$boootstrap$", we identify "B" different decision trees;

(ii) Among the p predictive variables in case of segmentation processes associated to the knot of a tree, in operation only a randomly chosen number $m < p$ is used, this selection aims the performance of the best segmentation possible;

(iii) The algorithm is practically a mix of weaker operational algorithms, B decision trees are generated, for the classification of a new observation x a process for running the B trees is initiated, the finality resulting from the selection of the majority set between the B predictions, we consider that is a typical example of the general method.

By comparing to the algorithms specific to previously described ML algorithms, the random forest brings information with a high degree of accuracy and innovation, as follows:

(i) Estimation of the prediction error;

(ii) Estimation of the importance of the variables in a non-linear framework;

(iii) Defining a proximity formalism between observations;

We also identify the advantages provided by this approach, probably one of the most efficient available algorithms, provided by the current technical-scientific development for the high predictive degree of the operationalities and organizational strategies for business models. It should also be noted that this type of algorithm does not face the problem of over-learning, it is easily transposable into operation, with increased power in predictive strategies. One of the identified vulnerabilities is that the operational implementation of the algorithm is difficult due to its complexity, hardly intelligible character for the deciding decision trees.

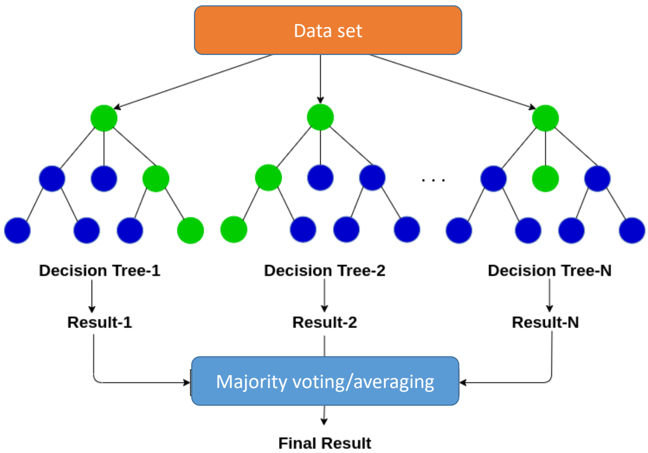

Figure 4. Illustration of a random forest algorithm

Support vector machines

Support vector machines (SVM) are also known as $Wide\ Edge\ Separators$. These are considered very powerful non-linear binary classification algorithms, in certain circumstances they are the perfect replacement for neuronal learning networks, inducing relatively higher costs (Tian, Shi & Liu, 2013). $The\ SVM$ principle consists in the elaboration of a separating non-linear band with a maximum width that separates two observation groups and uses them for issuing predictions.

A non-linear φ transformation is used, that sends the points $x^1, \dots, x^N$ from the original space with p dimensions, the number of predictive variables, to the new points, $\varphi(x^1), \dots, \varphi(x^N)$in a space with a dimensionality larger than p, where the separation is simpler, the identification of a "$Separating\ Linear\ Band$" is aimed, that with the help of this type of approach is a target easier to reach so the problem that initially seemed relatively complex is replaced with a simpler problem with reduced costs.

The advantages generated by such an approach consist in the efficient treatment of problems with a high number of dimensions, exactly as the problems faced by modern economic organizations. Support vector machines represent an alternative to neuronal systems more are cheaper and easier to train. One of the vulnerabilities is the difficulty to define the Function K($kernel\ trick$), the increased complexity of the algorithm, most of the time, the final generated performance is below the one of the approaches using the Random Forests.

Algorithms for dimensionality reduction

In the operational activity of the deciding organizational top management, the reduction of the number of explicative variables in several observations is a priority task, being justified by:

(i) The reduction of the execution time of a learning algorithm which is directly dependent on the size N of the learning algorithm and number of p variables;

(ii) Dimension N of a significant sample, containing the number of variables;

(iii) The development of a predictive model with an optimum number of variables that are easy to interpret, flexible and with increased visibility.

The reduction of the variables number is a permanent aim (Sarveziani, 2014), depending on the chosen organizational business, but retaining the final predictive target of the original variables, we identify three strategic directions for actions:

(i) "$Factorial\ Analysis$", explains the p variables observed $x_1, \ldots, x_p$, for $k < p$ for the factors $f_1, \ldots f_p$ designated "$Latent\ Variables$", operational practice determines the expression of each x_i as a linear combination of f_i;

(ii) The analysis of the main components is the inverse principle of the $Factorial\ Analysis$, it consists of searching a restrained number of linear combinations between the original variables $x_1, \ldots, x_p$ to create non-correlated components that explain the variables' essential compared to the original variables, the results obtained are very close to the one obtained following the factorial analysis;

(iii) "$Forward\ Selection$" modeling consists in the introduction one by one of the predictive variables in the chosen model, each step uses the non-used variables, the most correlated with the target variables, the process is stopped when the correlation level decreases below a previously set level according to the purpose and aimed accuracy degree.

Conclusions

Business modeling and optimization in economic organizations by paradigms specific to ML provides a very strong facility, flexible and adapted for the elaboration of optimal decisions by the top management. Without claiming to be a panacea for the top managerial decision, we consider that the procedures and algorithms specific to ML are an efficient alternative in this new field of human knowledge and research.

We have presented a discussion of algorithms for machine learning applied to business models of economic organizations and their challenges. Our analysis suggests that applying such ML to concrete industrial and business cases in existing complex ecosystems, results in some concrete and specific features that certainly contribute to a more efficient, more effective, and more profitable business:

(i) *Capability to achieve transposition into operationality*: intelligent solutions with a high complexity degree cannot reach levels of high efficiency if they are not transposable to a distribution framework;

(ii) *Capability to detect and use only the very useful data* for the economic organization and, at the same time, detect and eliminate dirty data with a high incoherence risk;

(iii) *Achieve transparency* as a characteristic to functionalities specific to Machine Learning, a feature for the improvement of the organizational performance, the learning process having a progressive growth;

(iv) *Leverage on appropriate competences*, specific to the implementation and optimization of ML paradigms;

(v) *Improved Proportionality* reflecting the quantity of energy and resources invested in the improvement and optimization of algorithms specific to the economic models in relation to the total revenue obtained at the economic organization's level.

Among the future trends and perspectives in the field, on top of the discussed algorithms, we would like to mention the following two ones: (a) confronted with the problem of huge amounts of Big Data, many not useful, ML can be deployed and used for filtering the data as per its significance and providing only the useful and functional data to gain time and efficiency, (b) use of ML for to address cybersecurity issues, which will require new classes of algorithms capable to serve as multiple layers of protection by automating complex tasks and detecting cyber-attacks on its own, automating responses to cyber-attacks without any human intervention.

References

Abernethy, J., Bartlett, P.L., Rakhlin, A., & Tewari, A. (2008). Optimal strategies and minimax lower bounds for online convex games. In *Proceedings of the 19th annual conference of computational learning theory*.

Anthony, M. & Barlet, P. (1999). *Neural Network Learning: Theoretical Foundations,* Cambridge University Press.

Boser, B.E., Guyon, I.M., & Vapnik, V.N. (1992) A Training Algorithm for Optimal Margin Classifiers. In *Proceedings of the 5th Annual Workshop on Computational Learning Theory (COLT'92), Pittsburgh, 27-29 July 1992* (pp.144-152).

Bousquet, O., & Elisseeff, A. (2002). Stability and generalization. *Journal of Machine Learning Research* 2, 499-526.

Breiman, L. (2001). Random forests, *Machine Learning* 45(1), 5-32.

Donoho, D.L. (2006), Compressed sensing, *IEEE Transaction on Information Theory* 52(4), 1289-1306. https://doi.org/10.1109/TIT.2006.871582

Ge, Z. Song, Ding, S.X., &Huang, B. (2017). Data mining and analytics in the process industry: the role of machine learning, *IEEE Access* 5, 20590-20616. https://doi.org/10.1109/ACCESS.2017.2756872

Gins, G., Van den Kerkhof, P. Vanlaer, J., & Van Impe, I.F.M. (2015). Improving classication-based diagnosis of batch processes through data selection and appropriate pretreatment. Journal of Process Contro 26, 90 -101.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*, Springer.

MacKay, D.J. (2003). *Information theory, inference and learning algorithms*. Cambridge Univesity Press.

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 Intell. Lab. Syst.* 136, 4757. https://doi.org/10.1016/j.chemolab.2014.05.003

Sarveniazi, A. (2014). An Actual Survey of Dimensionality Reduction. *American Journal of Computational Mathematics* 4, 55-72. http://dx.doi.org/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. https://doi.org/10.1016/j.eng.2019.01.019

Tian, Y., Shi, Y., & Liu, X. (2012), Recent advances on support vector machines research. *Technological and Economic Development of Economy* 18(1), 5-33. https://doi.org/10.3846/20294913.2012.661205