Supervised Machine Learning Algorithms: Classification and Comparison

Dipak Ambhure
MIS - 111903143
Computer Engineering B.Tech
College of Engineering pune
ambhuredp19.comp@coep.ac.in

John Wang
EN849736
Computer Engineering M.Tech
California Institute of Technology
jOhn@gmai.com

Somnath Sone
RN9883
Computer Engineering B.Tech
College of Enginerring Pune
Somnat3@gmail.com

Fulan Devi
RN303
Computer Science BSc
Shahu College Latur
Fulan@gmail.com

***Abstract*—*Goal:* Supervised Machine Learning (SML) is the search for algorithms that reason from externally supplied instances to produce general hypotheses, which then make predictions about future instances. Supervised classification is one of the tasks most frequently carried out by the intelligent systems. This paper describes various Supervised Machine Learning (ML) classification techniques, compares various supervised learning algorithms as well as determines the most efficient classification algorithm based on the data set, the number of instances and variables (features).Seven different machine learning algorithms were considered:Decision Table, Random Forest (RF) , Naïve Bayes (NB) , Support Vector Machine (SVM), Neural Networks (Perceptron), JRip and Decision Tree (J48) using Waikato Environment for Knowledge Analysis (WEKA)machine learning tool.To implement the algorithms, Diabetes data set was used for the classification with 786 instances with eight attributes as independent variable and one as dependent variable for the analysis. The results show that SVMwas found to be the algorithm with most precision and accuracy. Naïve Bayes and Random Forest classification algorithms were found to be the next accurate after SVM accordingly.**

***Index Terms*—Machine Learning, Classifiers, Data Mining Techniques, Data Analysis, Learning Algorithms, Supervised Machine Learning.**

***Impact Statement-* Supervised learning provides you with a powerful tool to classify and process data using machine language. With supervised learning you use labeled data, which is a data set that has been classified, to infer a learning algorithm. The data set is used as the basis for predicting the classification of other unlabeled data through the use of machine learning algorithms.**

This paragraph of the first footnote will contain the date on which you submitted your paper for review. It will also contain support information, including sponsor and financial support acknowledgment. For example, â€œThis work was supported in part by the U.S. National Science Foundation under Grant BS123456â€.

The next few paragraphs should contain the authorsâ€™ current affiliations. Include detailed surface address and e-mail address only for the corresponding author. For example, F. A. Author is with the National Institute of Standards and Technology, Boulder, CO, USA. S. B. Author, Jr., was with Rice University, Houston, TX, USA. He is now with Colorado State University, Fort Collins, CO, USA. *T. C. Author is with the University of Colorado, Boulder, CO, USA (correspondence e-mail: author@uoc.edu).

I. INTRODUCTION

THIS Machine learning is one of the fastest growing areas of computer science, with far-reaching applications. It refers to the automated detection of meaningful patterns in data. Machine learning tools are concerned with endowing programs with the ability to learn and adapt [19]. Machine Learning has become one of the mainstays of Information Technology and with that, a rather central, albeit usually hidden, part of our life. With the ever increasing amounts of data becoming available there is a good reason to believe that smart data analysis will become even more pervasive as a necessary ingredient for technological progress.

There are several applications for Machine Learning (ML), the most significant of which is data mining. People are often prone to making mistakes during analyses or, possibly, when trying to establish relationships between multiple features [9]

Data Mining and Machine Learning are Siamese twins from which several insights can be derived through proper learning algorithms. There has been tremendous progress in data mining and machine learning as a result of evolution of smart and Nano technology which brought about curiosity in finding hidden patterns in data to derive value. The fusion of statistics, machine learning, information theory, and computing has created a solid science, with a firm mathematical base, and with very powerful tools.

One standard formulation of the supervised learning task is the classification problem: The learner is required to learn (to approximate the behavior of) a function which maps a vector into one of several classes by looking at several input-output

examples of the function. Inductive machine learning is the process of learning a set of rules from instances (examples in a training set), or more generally speaking, creating a classifier that can be used to generalize from new instances. The process of applying supervised ML to a real-world problem is described in Figure 1.

II. LITERATURE REVIEW

A. *Classification of Supervised Learning Algorithms*

*1) **Linear Classifiers**:* Linear models for classification separate input vectors into classes using linear (hyperplane) decision boundaries [6]. The goal of classification in linear classifiers in machine learning, is to group items that have similar feature values, into groups. [23] stated that a linear classifier achieves this goal by making a classification decision based on the value of the linear combination of the features. A linear classifier is often used in situations where the speed of classification is an issue, since it is rated the fastest classifier [21].Also, linear classifiers often work very well when the number of dimensions is large, as in document classification, where each element is typically the number of counts of a word in a document. The rate of convergence among data set variables however depends on the margin.

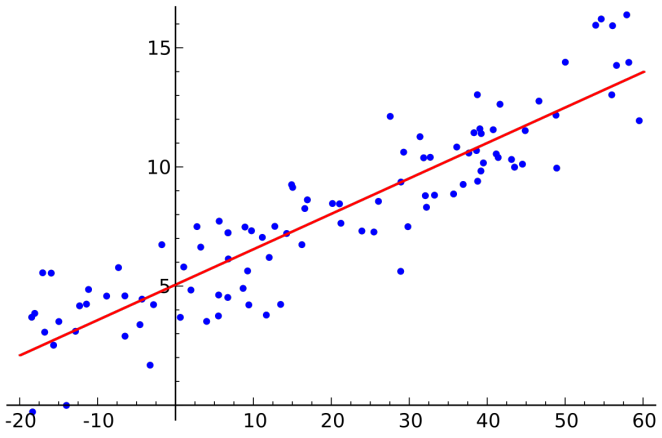

Fig. 1. **Linear regression**

*2) **Logistic regression**:* This is a classification function that uses class for building and uses a single multinomial logistic regression model with a single estimator. Logistic regression usually states where the boundary between the classes exists, also states the class probabilities depend on distance om the boundary, in a specific approach. This moves towards the extremes (0 and 1) more rapidly when data set is larger. These statements about probabilities which make logistic regression more than just a classifier. It makes stronger, more detailed predictions, and can be fit in a different way; but those strong predictions could be wrong.Logistic regression is one of the most commonly used tools for applied statistics and discrete data analysis. Logistic regression is linear interpolation[11].

*3) **Naive Bayesian (NB) Networks**:* These are very simple Bayesian networks which are composed of directed acyclic graphs with only one parent (representing the unobserved node)

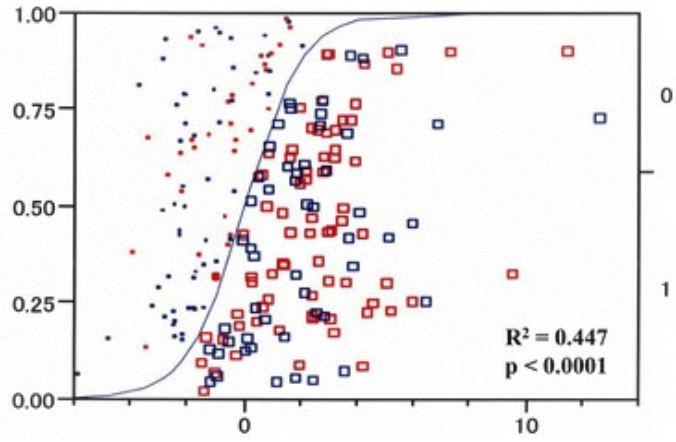

Fig. 2. **Logistic regression**

and several children (corresponding to observed nodes) with a strong assumption of independence among child nodes in the context of their parent [7].Thus, the independence model (Naive Bayes) is based on estimating [14]. Bayes classifiers are usually less accurate that other more sophisticated learning algorithms (such as ANNs).However, [5] performed a large-scale comparison of the naive Bayes classifier with state-of-the-art algorithms for decision tree induction, instance-based learning, and rule induction on standard benchmark datasets, and found it to be sometimes superior to the other learning schemes, even on datasets with substantial feature dependencies.

*4) **Multi-layer Perceptron**:* This is a classifier in which the weights of the network are found by solving a quadratic programming problem with linear constraints, rather than by solving a non-convex, unconstrained minimization problem as in standard neural network training [21].Other well-known algorithms are based on the notion of perceptron [17].Perceptron algorithm is used for learning from a batch of training instances by running the algorithm repeatedly through the training set until it finds a prediction vector which is correct on all of the training set.

*5) **Support Vector Machines (SVMs)**:* These are the most recent supervised machine learning technique [24].Support Vector Machine (SVM) models are closelyrelated to classical multilayer perceptron neural networks.SVMs revolve around the notion of a —margin—either side of a hyperplane that separates two data classes. Maximizing the margin and thereby creating the largest possible distance between the separating hyperplane and the instances on either side of it has been proven to reduce an upper bound on the expected generalisation error [9].

*6) **K-means**:* The simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priority.K-Means algorithm is be employed when labeled data is not available [1].General method of converting rough rules of thumb into highly accurate prediction rule.A boosting algorithm can provably construct single classifier with very high accuracy, say, 99

*7) **Decision Tree**:* Decision Trees (DT) are trees that classify instances by sorting them based on feature values. Each node in a decision tree represents a feature in an instance to be classified, and each branch represents a value that the node can assume. Instances are classified starting at the root node and sorted based on their feature values [9].Decision tree learning, used in data mining and machine learning, uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. More descriptive names for such tree models are classification trees or regression trees [20].Decision tree classifiers usually employ post-pruning techniques that evaluate the performance of decision trees, as they are pruned by using a validation set. Any node can be removed and assigned the most common class of the training instances that are sorted to it [9].

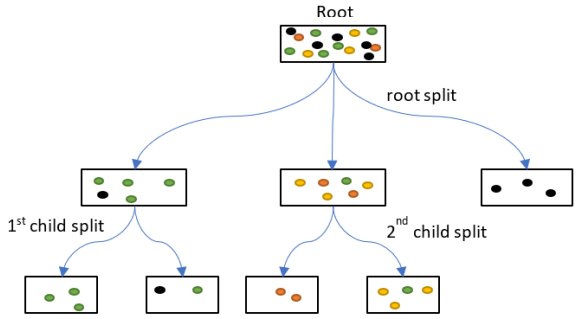

Fig. 3. **Decision Tree**

*8) **Neural Networks**:* Networks (NN) that can actually perform a number of regression and/or classification tasks at once,although commonly each network performs only one. In the vast majority of cases, therefore, the network will have a single output variable, although in the case of many-state classification problems, this may correspond to a number of output units (the post-processing stage takes care of the mapping from output units to output variables).Artificial Neural Network (ANN) depends upon three fundamental aspects, input and activation functions of the unit, network architecture and the weight of each input connection. Given that the first two aspects are fixed, the behavior of the ANN is defined by the current values of the weights. The weights of the net to be trained are initially set to random values, and then instances of the training set are repeatedly exposed to the net.Then, all the weights in the net are adjusted slightly in the direction that would bring the output values of the net closer to the values for the desired output.

III. FEATURES OF MACHINE LEARNING ALGORITHMS

Supervised machine learning techniques are applicable in numerous domains. A number of Machine Learning (ML) application oriented papers can be found in [18], [25]. Generally, SVMs and neural networks tend to perform much better when dealing with multi-dimensions and continuous features. On the other hand, logic-based systems tend to perform better when dealing with discrete/categorical features. For neural network

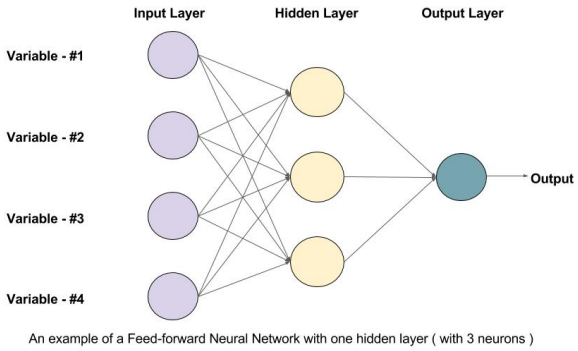

Fig. 4. **Neural Network**

models and SVMs, a large sample size is required in order to achieve its maximum prediction accuracy whereas NB may need a relatively small dataset.

There is general agreement that k-NN is very sensitive to irrelevant features: this characteristic can be explained by the way the algorithm works. Moreover, the presence of irrelevant features can make neural network training very inefficient, even impractical.Most decision tree algorithms cannot perform well with problems that require diagonal partitioning. The division of the instance space is orthogonal to the axis of one variable and parallel to all other axes. Therefore, the resulting regions after partitioning are all hyperrectangles. The ANNs and the SVMs perform well when multi-collinearity is present and a nonlinear relationship exists between the input and output features.

Naive Bayes (NB) requires little storage space during both the training and classification stages: the strict minimum is the memory needed to store the prior and conditional probabilities. The basic kNN algorithm uses a great deal of storage space for the training phase, and its execution space is at least as big as its training space. On the contrary, for all non-lazy learners, execution space is usually much smaller than training space, since the resulting classifier is usually a highly condensed summary of the data. Moreover, Naive Bayes and the kNN can be easily used as incremental learners whereas rule algorithms cannot. Naive Bayes is naturally robust to missing values since these are simply ignored in computing probabilities and hence have no impact on the final decision. On the contrary, kNN and neural networks require complete records to do their work.

Finally, Decision Trees and NB generally have different operational profiles, when one is very accurate the other is not and vice versa. On the contrary, decision trees and rule classifiers have a similar operational profile. SVM and ANN have also a similar operational profile. No single learning algorithm can uniformly outperform other algorithms over all datasets. Different data sets with different kind of variables and the number of instances determine the

IV. DISCUSSION

Figure 5 Describes the brief classification of Machine learning algorithms based on Data .

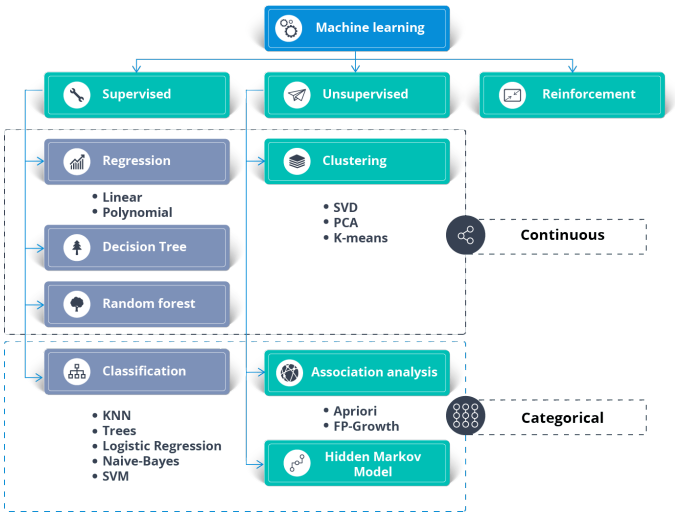

Fig. 5. Mchine Learning Broad Classification

UNITS FOR MAGNETIC PROPERTIES

Symbol	Quantity	Conversion from Gaussian and CGS EMU to SI[a]
Φ	magnetic flux	$1\ Mx \rightarrow 10^{-8}\ Wb = 10^{-8}\ V \cdot S$
B	magnetic flux density, magnetic induction	$1\ G \rightarrow 10^{-4}\ T = 10^{-4}\ Wb/m^2$
H	magnetic field strength	$1\ Oe \rightarrow 10^3/(4\pi)\ A/m$
m	magnetic moment	$1\ erg/G = 1\ emu \rightarrow 10^{-3}\ A \cdot m^2 = 10^{-3} A/m$
M	magnetization	$1\ erg/(G \cdot cm^3) = 1\ emu/cm^3 \rightarrow 10^3\ A/m$
4πM	magnetization	$1\ G \rightarrow 10^3/(4\pi)\ A/m$
σ	specific magnetization	$1\ erg/(G \cdot g) = 1\ emu/g \rightarrow 1\ A \cdot m^2/kg$
j	magnetic dipole moment	$1\ erg/G = 1\ emu \rightarrow 4\pi \times 10^{-10}\ Wb \cdot m$
J	magnetic polarization	$1\ erg/(G \cdot cm^3) = 1\ emu/cm^3 \rightarrow 4\pi \times 10^{-4}\ T$
χ, κ	susceptibility	$1 \rightarrow 4\pi$
$\chi\rho$	mass susceptibility	$1\ cm^3/g \rightarrow 4\pi \times 10{-}3\ m^3/kg$
μ	permeability	$1 \rightarrow 4\ \pi \times 10^{-7}\ H/m = 4\pi \times 10{-}7\ Wb/(A \cdot m)$
$\mu\rho$	relative permeability	$\mu \rightarrow \mu_r$
w,W	energy density	$1\ erg/cm^3 \rightarrow 10^{-1}\ J/m^3$
N,D	demagnetizing factor	$1 \rightarrow 1/(4\pi)$

TABLE I. **Some Basic Symbols** mainly used For Magnetic Properties of Machine

A. *Process to choose Algorithm*

1) Define adequately our problem (objective, desired outputs...).
2) Gather data.
3) Choose a measure of success.
4) Set an evaluation protocol and the different protocols available.
5) Prepare the data (dealing with missing values, with categorial values...).
6) Spilit correctly the data.
7) Differentiate between over and underfitting, defining what they are and explaining the best ways to avoid them.
8) An overview of how a model learns.
9) What is regularization and when is appropiate to use it.
10) Develop a benchmark model.
11) Choose an adequate model and tune it to get the best performance possible.

B. *Basic Terminology*

1) ***Algorithm****:* An Algorithm is a set of rules that a machine follows to achieve a particular goal. An algorithm can be considered as a recipe that defines the inputs, the output and all the steps needed to get from the inputs to the output. Cooking recipes are algorithms where the ingredients are the inputs, the cooked food is the output, and the preparation and cooking steps are the algorithm instructions.

2) ***Black Box Model****:* A Black Box Model is a system that does not reveal its internal mechanisms. In machine learning, "black box" describes models that cannot be understood by looking at their parameters (e.g. a neural network). The opposite of a black box is sometimes referred to as White Box, and is referred to in this book as interpretable model. Model-agnostic methods for interpretability treat machine learning models as black boxes, even if they are not.

C. *Essential Terminology*

A ***Dataset*** is a table with the data from which the machine learns. The dataset contains the features and the target to predict. When used to induce a model, the dataset is called training data.

An ***Instance*** is a row in the dataset. Other names for 'instance' are: (data) point, example, observation.

The ***Features*** are the inputs used for prediction or classification. A feature is a column in the dataset. Throughout the book, features are assumed to be interpretable, meaning it is easy to understand what they mean, like the temperature on a given day or the height of a person. The interpretability of the features is a big assumption. But if it is hard to understand the input features, it is even harder to understand what the model does.

The ***Target*** is the information the machine learns to predict.

A Machine Learning Task is the combination of a dataset with features and a target. Depending on the type of the target, the task can be for example classification, regression, survival analysis, clustering, or outlier detection.

The ***Prediction*** is what the machine learning model "guesses" what the target value should be based on the given features.

V. CONCLUSION

Machine Learning is a technique of training machines to perform the activities a human brain can do, albeit bit faster and better than an average human-being. Today we have seen that the machines can beat human champions in games such as Chess, AlphaGO, which are considered very complex. You have seen that machines can be trained to perform human activities in several areas and can aid humans in living better lives.

Machine Learning can be a Supervised or Unsupervised. If you have lesser amount of data and clearly labelled data for training, opt for Supervised Learning. Unsupervised Learning would generally give better performance and results for large data sets. If you have a huge data set easily available, go for

deep learning techniques. You also have learned Reinforcement Learning and Deep Reinforcement Learning. You now know what Neural Networks are, their applications and limitations.

Finally, when it comes to the development of machine learning models of your own, you looked at the choices of various development languages, IDEs and Platforms. Next thing that you need to do is start learning and practicing each machine learning technique. The subject is vast, it means that there is width, but if you consider the depth, each topic can be learned in a few hours. Each topic is independent of each other. You need to take into consideration one topic at a time, learn it, practice it and implement the algorithm/s in it using a language choice of yours. This is the best way to start studying Machine Learning. Practicing one topic at a time, very soon you would acquire the width that is eventually required of a Machine Learning expert.

SUPPLEMENTARY MATERIALS

Machine learning is quickly becoming an important tool in modern materials design. Where many of its successes are rooted in huge datasets, the most common applications in academic and industrial materials design deal with datasets of at best a few tens of data points. Harnessing the power of machine learning in this context is, therefore, of considerable importance. In this work, we investigate the intricacies introduced by these small datasets. We show that individual data points introduce a significant chance factor in both model training and quality measurement. .

A. *Footnotes*

From the sci-fi classic "Bladerunner" to the recent films "Her" and "Ex Machina," pop culture is filled with stories demonstrating our simultaneous fascination with and fear of artificial intelligence (AI)[1]

VI. IMPACT OF MACHINE LEARNING OVER GLOBE

The past decade, and particularly the past few years, have been transformative for artificial intelligence (AI) not so much in terms of what we can do with this technology as what we are doing with it. Some place the advent of this era to 2007, with the introduction of smartphones. As I detail below, at its most essential, intelligence is just intelligence, whether artifact or animal. It is a form of computation, and as such a transformation of information. The cornucopia of deeply personal information that resulted from the willful tethering of a huge portion of society to the Internet has allowed us to pass immense explicit and implicit knowledge from human culture via human brains into digital form. Here we can not only use it to operate with human-like competence, but also produce further knowledge and behavior by means of machine-based computation.

For decades—even prior to the inception of the term—AI has aroused both fear and excitement as humanity contemplates creating machines in our image. This expectation that intelligent artifacts should by necessity be human-like artifacts blinded most of us to the important fact that we have been achieving AI for some time. While the breakthroughs in surpassing human ability at human pursuits, such as chess (Hsu, 2002), Go (Silver et al., 2016), and translation (Wu et al., 2016), make headlines, AI has been a standard part of the industrial repertoire since at least the 1980s. Then production-rule or "expert" systems became a standard technology for checking circuit boards and detecting credit card fraud (Liao, 2005).

Similarly, machine-learning (ML) strategies like genetic algorithms have long been used for intractable computational problems, such as scheduling, and neural networks not only to model and understand human learning, but also for basic industrial control and monitoring (Widrow et al., 1994). In the 1990s, probabilistic and Bayesian methods revolutionized ML and opened the door to some of the most pervasive AI technologies now available: search through massive troves of data (Bishop, 1995). This search capacity included the ability to do semantic analysis of raw text, astonishingly enabling Web users to find the documents they seek out of trillions of Web pages just by typing only a few words (Lowe, 2001; Bullinaria and Levy, 2007).

VII. SUMMARY

Machine Learning is rapidly changing the world, from diverse types of applications and research pursued in industry and academia. Machine learning is affecting every part of our daily lives. From voice assistants using NLP and machine learning to make appointments, check our calendar and play music, to programmatic advertisements — that are so accurate that they can predict what we will need before we even think of it. More often than not, the complexity of the scientific field of machine learning can be overwhelming, making keeping up with "what is important" a very challenging task. However, to make sure that we provide a learning path to those who seek to learn machine learning, but are new to these concepts. In this article, we look at the most critical basic algorithms that hopefully make your machine learning journey less challenging.

REFERENCES

[1] **Basics of Machine Learning:** *Alex S. Vishwanathan,S.V.N (2008).*Introduction to Machine Learning. Published by the press syndicate of the University of Cambridge, Cambridge, United Kingdom. Copyright © Cambridge University Press 2008.

[2] **ML Future:** *Bishop, C. M. (1995).* Neural Networks for Pattern Recognition. Clarendon Press, Oxford, England. 1995. Oxford University Press, Inc. New York, NY, USA © 1995.

[3] **ML Algorithms** *heng, J., Greiner, R., Kelly, J., Bell, D. Liu, W. (2002).* Learning Bayesian networks from data: An information-theory based approach. Artificial Intelligence Volume 137. Copyright © 2002. Published by Elsevier Science B.V.

[4] **Arificial intelligence Classification:** J. U. Duncombe,Infrared navigationâ"Part I: Process Towards AI, *IEEE Trans. Artificial Intelligence*, Setiono R. and Loew, W. K. (2000), FERNN: An algorithm for fast extraction of rules from neural networks, Applied Intelligence.

[5] **Theory beyond Machine :**Tapas Kanungo, D. M. (2002). A local search approximation algorithm for k-means clustering. Proceedings of the eighteenth annual symposium on Computational geometry. Barcelona, Spain: ACM Press.

[1]

[6] **Regression Techniques:** T. Hastie, R . Tibshirani, J. H. Friedman (2001) — The elements of statistical learning,‖ Data mining, inference, and prediction, 2001, New York: Springer Verlag.

[7] **Basic Format for Handbooks:** *Name of Manual/Handbook*, x ed., Abbrev. Name of Co., City of Co., Abbrev. State, year, pp. xxx-xxx.
[8] *Example Handbook 1: Transmission Systems for Communications*, 3rd ed., Western Electric Co., Winston-Salem, NC, 1985, pp. 44â€“60.
[9] **Linear Regression Techniques:** Lemnaru C. (2012). Strategies for dealing with Real World Classification Problems, (Unpublished PhD thesis) Faculty of Computer Science and Automation, Universitatea Technica, Din Cluj-Napoca.

[10] **Applications of AI:** Neocleous C. Schizas C. (2002). Artificial Neural Network Learning: A Comparative Review. In: Vlahavas I.P., Spyropoulos C.D. (eds)Methods and Applications of Artificial Intelligence. Hellenic Conference on Artificial IntelligenceSETN 2002.
[11] **Classification abd Regression:** *Rob Schapire (n.d)* Machine Learning Algorithms for Classifrication.
[12] **Computer Programs Configuration:** Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. Informatica 31 (2007). Pp. 249 – 268.
[13] **The Intranet Architecture:** S.L.Talleen. (1996, Apr.) Managing information in the new paradigm. Amdahl Corp., CA. [Online].
Basic Format for Computer Programs and Electronic Documents (when available online): ISO recommends that capitalization follow the accepted practice for the language or script in which the information is given.
[14] **Data Mining :** Timothy Jason Shepard, P. J. (1998). Decision Fusion Using a Multi-Linear Classifier. In Proceedings of the International Conference on Multisource-Multisensor Information Fusion.

[15] Witten, *I. H. Frank, E. (2005).* Data Mining: Practical machine learning tools and techniques (2nd ed.), ISBN: 0-12-088407-0, Morgan Kaufmann Publishers, San Francisco, CA, U.S.A. © 2005