Available online at www.rajournals.in

RA JOURNAL OF APPLIED RESEARCH

ISSN: 2394-6709

DOI:10.47191/rajar/v7i10.13

Volume: 07 Issue: 10 October-2021

International Open Access

Impact Factor- 7.036

Page no.- 2596-2602

Academic Failure at University and Data Processing Methods Based on Decision Trees and Neural Networks: Research Methodology

Smail ADMEUR[1], **Hicham ATTARIUAS**[2]

[1]Computer Science and Systems Engineering Laboratory, Faculty of Sciences Tetouan, University Abdelmalek Essaadi Tetouan, Morocco

[2]Computer Science and Systems Engineering Laboratory, Faculty of Sciences Tetouan, University Abdelmalek Essaadi Tetouan, Morocco

ARTICLE INFO

Published Online: 27 October 2021

Corresponding Author: **Smail ADMEUR**

ABSTRACT

For a long time, academic failure among university students sparked heated controversy. Many educational psychologists try to figure it out and then explain it. Statisticians have tried to predict it. Our research (article) aims to classify students into several categories, as well as to use the decision tree and artificial neural networks to classify first-year students and identify variables that may explain the problem.

KEYWORDS: College Failure, Work Methodology, Data Processing, Decision Tree, Artificial Neural Networks.

I. INTRODUCTION

The enormous failure among the students of the university is a real problem on which the Moroccan Ministry of National Education, Vocational Training, Higher Education and Scientific Research has stumbled for years without succeeding in finding a solution. really satisfactory solutions.

The failure is mainly related to the first year of college. According to ministry figures, 25% of new students in open access faculties who do not have a selection system (law, arts and science faculties) do not make it past the first semester. Even more, 43% of students leave the university benches without having obtained a university degree (including the DEUG and the license). A large part of the students repeat a year and the rest drop out of university. The ministry also indicated that only 13.3% of students in open-access universities obtain their bachelor's degree in 3 years, etc.

Given the importance of success and failure in the field of higher education, and which is reflected mainly in the level of the first academic year, and because of the need for specialization orientation, the problem to be solved is to know which are, among the set of characteristics presented, those which explain the success or failure of the student in the first year.

The objective of our work is to propose a working methodology for our research by defining the instruments chosen for the processing and analysis of the data of our research. Through this article, we offer in a first part, an introduction to university failure, in a second part, we present the models designed through the decision tree and neural networks as models proposed for analysis and data processing. The third part concerns the presentation of our work methodology.

II. DECISION TREE

A. *Definition:*

A decision tree is a decision support tool that represents a set of choices in the form of a tree. The various possible decisions are located at the end of the branch (the "leaf" of the tree) and are reached according to the decisions taken at each stage. Decision trees are tools used in various fields such as security, data mining, and medicine. It has the advantages of readability and speed of execution. It is also a representation which can be calculated automatically by a supervised learning algorithm [1]. Decision trees are used in decision support areas in business intelligence or data mining. A major advantage of decision trees is that they can be calculated automatically from databases by supervised learning algorithms. These algorithms automatically select discriminated variables from unstructured and potentially large data. They can thus make it possible to extract logical cause-and-effect rules that did not initially appear in the raw data.

B. *Decision tree learning:*

Decision tree learning refers to a method based on the use of decision trees as predictive models. It is particularly used for data mining and machine learning [2]. In these trees, the leaves represent the values of the target variables, and the branches correspond to the combinations of input variables that lead to these values. In decision analysis, a decision tree can be used to clearly indicate the decision made and the process that led to the decision. In learning and data mining, the decision tree describes the data rather than the decision itself, and the tree will be used as a starting point for the decision-making process. This is a supervised learning technique: we use a data set for which we know the value of the target variable to build the tree, then we extrapolate the result to the test data set. . Decision trees are among the most popular algorithms in machine learning [3, 4].

Decision tree learning is a classic method in machine learning [5]. Its purpose is to create a model that predicts the value of a target variable from the value of several input variables. Each leaf (or terminal node of the tree) represents either a value of the target variable, or a probability distribution of the various possible values of the target variable. The combination of the values of the input variables is represented by the path from the root to the leaf.

The tree is generally built by separating the set of data into subsets based on the value of an input characteristic. This process is repeated on each subset obtained recursively, so it is a recursive partitioning. In data mining, decision trees can help in the description, categorization or generalization of a fixed data set. There are two main types of decision trees in data mining:

- Classification trees are used to predict to which class the target variable belongs; in this case the prediction is a class label,
- Regression trees are used to predict a real quantity (for example, the price of a house or the length of time a patient stays in a hospital), in this case the prediction is a numerical value.

The term Classification and Regression Tree Analysis (CART) is a generic term referring to the procedures described and introduced by Breiman [6]. The trees used in the case of regression and in the case of classification present similarities but also differences, in particular with regard to the procedure used to determine the separations of the branches.

III. Neural Networks

A. Introduction:

Many researchers from different fields are interested in Artificial Neural Networks (ARNs), including computer scientists, engineers, mathematicians, statisticians, physicists, cognitive scientists, neuroscientists, psychologists and linguists. The field of research is interdisciplinary, attracting researchers from different backgrounds with different motivations and goals. Motivated by the structure of the human brain, the models of RNA are many and diverse. Network research is a promising approach to artificial intelligence.

Neural networks were originally designed as very simplified mathematical models of our brains. For example, formal or artificial neural networks are learning systems that mimic the way we learn to perform complex tasks through trial and error and continuous correction. Specifically, these neural techniques are mathematical and algorithmic tools that have proven to be powerful and practical in solving nonlinear and complex problems.

By providing practical methods for solving large-scale problems, the field has developed rapidly and has been applied in most traditional engineering fields (diagnostics, control, character and image recognition, predictive time series, discrimination, etc.). Therefore, the artificial neural network is a mathematical model, which is expressed as a group of strongly connected computational units called formal neural networks or artificial neurons. In this way, they can be described by their architecture and their components.

Many terms in the literature today refer to the field of neural networks. The RNAs than the model manipulated, so it is better to say "artificial neural networks". Indeed, the biological neural networks (RNB) are much more complex than the mathematical models used in the artificial neural networks, consequently the RNAs are composed of the interconnection of basic unit of information processing, and its operating principle is based on that of biological neurons, a network has capacities for storing and processing information, attributable to the weight of the connections between each neuron. These weights are calculated through an adaptation process also called learning.

B. Historical overview:

The first steps in neural network theory were developed in 1943 by neurophysiologist Warren McCulloch and mathematician Walter Pitts. They introduced artificial neurons with the threshold, which could be arranged in networks. A few years later, in 1949 psychologist Donald Hebb devised the first rule of learning for artificial neural networks (Hebb's rule). His suggestion is based on increasing the strength of connection: if two neurons were active simultaneously, then the strength of the connection between them should be increased. We provide a brief history of early research carried out in chronological order.

After the work of James in 1890 who worked on associative memory, the work of Hermann von Helmholtz, Ernst Mach and Ivan Pavlov at the beginning of the 20th century consisted of multidisciplinary work in physics, psychology and neurophysiology at that time, he These were rather general theories without a precise mathematical model of a neuron [7, 8]. In 1943, the first steps in neural network theory were developed by neurophysiologist Warren McCulloch and mathematician Walter Pitts [9]. They introduced artificial neurons with the threshold, which could be arranged in networks. A few years later, in 1949 psychologist Donald Hebb devised the first learning rule for artificial neural networks called Hebb's rule [9]. His suggestion is based on increasing the strength of connection: if two neurons were active simultaneously, then the strength of the connection between them should be increased. In 1951 Minsky thus proposed the stochastic neural analog reinforcement calculator (SNARC), the first neural network simulator and the first operational implementation [10]. In 1957, Rosenblatt invented the perceptron which is a supervised learning algorithm of binary classifiers (i.e. separating two classes). It is a formal neuron equipped with a learning rule that automatically determines the synaptic weights so as to separate a supervised learning problem. If the problem is linearly separable, a theorem assures that the perceptron rule allows finding a separator between the two classes [11].

In 1960 Widrow developed ADALINE (Adaptive Linear Neuron or later Adaptive Linear Element) is a single-layer artificial neural network. ADALINE is based on the formal neuron of McCulloch and Pitts. It consists of a synaptic weight, a bias (a constant that is added to the input) and a summation function [12]. In 1969, in Perceptrons, Minsky, and Papert Criticizing Rosenblatt, they showed the limitations of perceptron-like neural networks, including the inability to deal with nonlinear or connectedness problems. The consequence of this was to drain most of the research funds towards symbolic artificial intelligence [13].

Based on the work of Grossberg, Kohonen and Anderson, the period from 1967 to 1982 is considered to be the stalling of ANR research. They continue under the cover of various domains [14, 15, 16]. In 1982, Hopfield proposed the spin glasses model [17]. While in 1985, Rumelhart and colleagues proposed gradient backpropagation and the multilayer perceptron [18]. In 1992, Wang and Mendal dealt with the problem of seismic deconvolution in a Hopefield network in order to reduce the computation time [19]. The same year, Murat and Rudman, detect the first arrivals of seismic waves [20]. In 1994, Roth and Tarantola classified the different signals by inversion [21]. In 1997, Zhang and, Paulson characterized the distribution of subsurface resistivity by the inversion of magnetotelluric data [22, 23]. Whereas in 2000, Drew and Monson (2000): solving problems of classification, prediction, pattern recognition, categorization, associative memory and optimization [24].

C. *Applications:*

Today, neural networks are used for all kinds of applications in various fields. for example, autopilots for airplanes, or even guidance systems for automobiles, automatic reading systems for bank checks and postal addresses, signal processing systems for various military applications, systems for speech synthesis, networks are also used to build computer vision systems, to forecast money markets, to assess financial or insurance risk, for various manufacturing processes, for medical diagnostics, for oil or gas exploration, in robotics, in telecommunications,… However, neural networks have a significant impact today, and it is certain that the importance of neural networks will increase in the future.

D. *Model of an artificial neuron:*

Artificial neuron networks are characterized by a neuron model, network architecture and associated objective functions and learning algorithms.

The neuron is essentially made up of an integrator, which performs a weighted sum of its inputs. Then, the result n of the sum is transformed by the transfer function f, which produces the output a of the neuron [25].

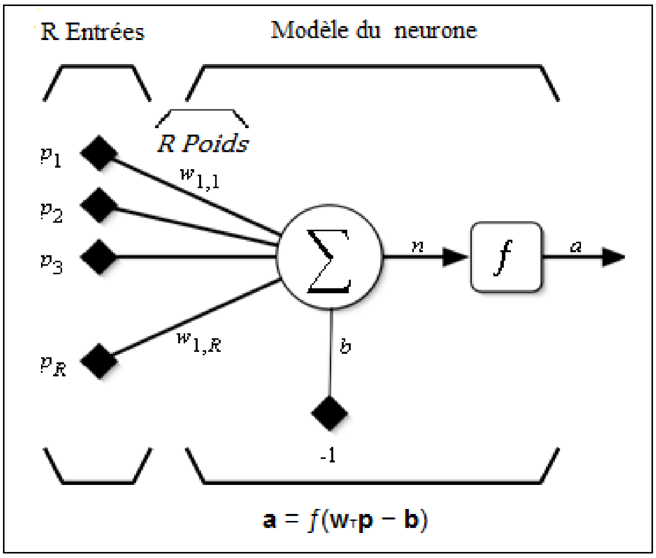

Figure 1: Model of an artificial neuron

According to the notation in the previous section, the R input of the neuron corresponds to the vector:

$$\mathbf{p = [p1p2 \bullet\bullet\bullet\ pR]\ T,\ et\ w = [w1{,}1w1{,}2 \bullet\bullet\bullet\ w1{,}R]}$$

T represents the weight vector of the neurons. The output n of the integrator is given by the following formula:

$$\mathbf{n = w_{1,1}\, p_1 + w_{1,2}\, p_2 + \cdot \cdot \cdot + w_{1,R}\, p_R - b}$$

We can also write in matrix form:

$$\mathbf{n = wTp - b}$$

This output is the weighted sum of the weight and the input minus the so-called neural bias b. The result of the weighted sum n is called the activation level of the neuron. Bias is also referred to as the activation threshold of neurons. When the activation level reaches or exceeds threshold b, the argument of f becomes a positive (or zero) number. Otherwise, it is a negative number.

IV. LEARNING MODES AND NEURAL NETWORKS

In this work part, we remember in review the different learning modes linked to artificial neural networks.

A. *Definition:*

Among the desirable properties of a neural network, the most fundamental is surely the ability to learn from its environment, to improve their performance through a learning process. But what then is learning? Learning is a dynamic and iterative process of changing the parameters of a network in response to the stimuli it receives from its environment. The type of learning is determined by how the parameter changes occur. Learning is a step in developing a neural network whose behaviour of the network is modified until the desired behaviour is obtained [26].

Learning results in a change in synaptic efficiency, that is, a change in the value of the weights that connect neurons from one layer to another. Let the weight wi, j connect the neuron i to its input j. At time t, a change **Δ**wi, j (t) in weight can be expressed simply as follows:

$$\mathbf{\Delta} \boldsymbol{wi,j}\mathbf{(t)} = \boldsymbol{wi,j}\mathbf{(t+1)} - \boldsymbol{wi,j}\mathbf{(t)}$$

During this step, the network structure is adapted to the weights of the connections, so that the output of the network approaches the desired output. Training requires certain designated examples (training samples) as well as a training algorithm. Therefore,

$$\boldsymbol{wi,j}\mathbf{(t+1)} = \mathbf{\Delta} \boldsymbol{wi,j}\mathbf{(t)} + \boldsymbol{wi,j}\mathbf{(t)}$$

With wi, j (t + 1) and wi, j (t) respectively representing the new and old values of the weight wi, j.

A well-defined set of rules for performing an adaptation process of this weight is called the network learning algorithm.

B. *Types of learning:*

There are two classes of learning algorithms: the first class of learning is supervised learning based on a set of samples consisting of input and output. The second unsupervised learning class, whose objective is to group samples into classes based on similarities or similarities between them [27, 28, 29].

C. *Supervised learning:*

In the case of supervised learning, there is a labelled data set or samples associated with a class by a professional (labelled training). This set of samples forms the basis for learning. The goal of a supervised learning algorithm is to build a classifier from a learning base or to classify a function. Such a function makes it possible to identify specific attributes of the class from the description of the object.

Supervised learning has an error signal, which not only calculates the satisfaction index, but also estimates that the local gradient indicates the direction in which the gradient adapts to the synaptic weight. Synaptic coefficients are evaluated on a training basis by minimizing the error between desired output and obtained output.

D. *Unsupervised learning:*

Unsupervised (self-organized) learning is about determining the classification based on a set of objects or a given situation (unlabeled samples) We have a large amount of data with an unknown structure, and we want find out if they have a group structure. This is to determine possible trends in the grouping of data into classes.

Unsupervised learning is generally based on a competitive process aimed at generating a model where the synaptic weights of neurons represent prototypes of stimuli. The quality of the resulting model should be assessed using a metric that measures the distance between stimuli and their prototypes. It is the competition process that makes it possible to select the prototype associated with each stimulus by searching for the neuron whose synaptic weight vector is closest to the stimulus in question.

E. *Learning rule:*

Several rules can guide the learning of neural networks [30].

Rule by Error Correction:

The first rule that can be used is based on correcting the errors observed at the output. Let ai (t) be the output that we get (output error) for neuron i at time t. This output results from a stimulus p (t) that is applied to the inputs of the network, one of the neurons of which corresponds to neuron i. Let di (t) be the output that we want to obtain (desired output) for this same neuron i at time t. So, ai (t) and di (t) will generally be different and it is natural to calculate the error ei (t) between what we get and what we want to get:

$$\mathbf{ei(t) = di(t) - ai(t)}$$

We must then look for a way to reduce this error as much as possible. In vector form, we obtain:

$$\mathbf{e(t) = d(t) - a(t)}$$

The weight of the network should be changed in a direction that minimizes the performance index E based on the error vectors ei (t), and then in the opposite direction to the gradient. We then speak of a direction of "descent" which is defined by:

$$\mathbf{\Delta} \boldsymbol{w}\mathbf{(t)} = -\boldsymbol{\eta} \nabla \boldsymbol{E}\mathbf{(t)}$$

Where η is a positive constant which determines the speed of learning which is called the learning rate. ∇E (t) indicate the gradient of E compared to the parameters weight w and time t. The rule is called "gradient descent".

Hebb's rule:

Hebb's rule was inspired by the work of neurophysiologist Donald Hebb, in this neurobiological context, Hebb sought to establish the form of associative learning at the cellular level [31]. Learning by Hebb's rule expresses the change in weight as a function of the correlation between the input p and the output a of a neuron. In the context of artificial networks, we can reformulate Hebb's statement in the form of a learning rule given by:

$$\boldsymbol{\Delta wj\ (t-1) = \eta p_j(t)\ a\ (t)}$$

Where η is the learning rate, pj (t) and a (t) correspond to presynaptic activity (input j of the neuron) and postsynaptic activity (output of the neuron) at the instant t respectively. This formula explicitly highlights the correlation between the input vector and the output vector.

Competitive learning rule:

As the name suggests, competitive learning involves competing neurons in the network to determine which neuron will be active at any given time. Competitive learning produces a "winner" as well as, sometimes, a set of neurons "neighbours" to the winner. Competitive learning produces a "winner" as well as, at times, a set of neurons "neighbours" to the winner and only that winner and, potentially, his neighbourhood benefit from an adjustment in their weight. We then say that learning is local because it is limited to a subset of the neurons in the network.

In their simplest form, neural networks that use competitive learning often consist of a single layer of output neurons, fully connected to the inputs. A winning neuron will change its synaptic weights by bringing them (geometrically) closer to an input stimulus p for which it beat all other neurons in the competition

In its simplest form, a neural network using competitive learning typically consists of a single layer of fully connected output neurons across the input. The winning neuron will make its synaptic weight (geometrically) close to the input stimulus p, so that it will defeat all other competing neurons, thereby altering its synaptic weight.

So, the competitive learning rule can be written as follows:

$$\begin{cases} \mathbf{\Delta w = \eta\ (y - w)} \text{ if the neuron is a winner} \\ \mathbf{0} \text{ otherwise} \end{cases}$$

Where η corresponds to a learning rate, y and w are the input vector and the weight vector respectively.

F. *Learning tasks:*

We have ended the notion of learning by listing different categories of tasks that can be performed with a neural network [32]:

G. *Association:*

There are two types: hetero-association and self-association. The problem of the latter consists in memorizing and storing a set of patterns (vectors), presenting them sequentially to the network. Then a part or a modified version of an original pattern is presented to the network, and the task is to exit the corresponding original mode. On the other hand, The problem of hetero-association is the association of two pairs of bosses: an entry boss and an exit boss, hetero-association involves supervised learning, while Self- rather, association requires unsupervised learning.

Approximation:

Let the function g be such that:

$$\mathbf{\mu = g(p),}$$

Where p is the argument of the function (vector), and µ the value (scalar) of this function evaluated at p. Now suppose that the function g () is unknown. Next, The approximation task includes designing a neural network that can combine the elements of the input and output pairs: {(p1, µ1), (p2, µ2),. . ., (pQ, µQ)}. This problem can be solved by performing supervised learning on the instances Q, where the pi representing the stimuli, and the µi representing the expected (desired) outputs of each of these stimuli, with i = 1, 2,. . ., Q. Where on the contrary, we can also say that supervised learning is a problem of function approximation.

Ranking:

Regarding this task, the network must learn to recognize a fixed number of categories of input stimuli (classes). In a first step, the network must begin with a supervised learning phase, in which stimuli are represented as inputs, and categories are used to form the desired outputs, generally using one output per category. Therefore, output 1 is associated with category 1, output 2 is associated with category 2, etc. For a problem involving Q categories, we can for example organize and fix the desired outputs d = [d1, d2,. . ., dQ] T Using the following expression:

$$\mathbf{d}\boldsymbol{i} = \begin{cases} \textbf{1 if the stimulus belongs to category i} \\ \textbf{0 otherwise, i = 1,. . ., Q.} \end{cases}$$

Later, in the recognition phase, it will suffice to be able to present any unknown stimulus to the network so that it can be classified in a category. A simple classification rule is, for example, to select the category associated with the largest output.

Ordered:

Command is another learning task that can be processed with a neural network. Consider a nonlinear dynamic system {u (t), y (t)} where u (t) denotes the input to the system and y (t) corresponds to the response of the system. In general, we want to control this system so that it behaves according to a reference model, usually a linear model, {r (t), d (t)}, where at any time t ≥ 0, we can produce a command u (t) such that:

$$\lim_{t\to\infty} |d(t) - y(t)| = 0$$

So that the output of the system closely follows that of the reference model. This can be achieved through certain types of supervised networks.

Prediction:

The concept of prediction is one of the most fundamental concepts in learning. This is a temporary signal processing issue. Suppose we have M past samples of a signal, x (t − 1), x (t − 2),. . . , x (t − M), Sampling is performed at a fixed time interval and the task is to predict the value of x at time t. This prediction problem can be solved by learning error correction, but with an unsupervised method,

Because the desired output values can be concluded directly from the time series. Specifically, the sample of x (t) can be used as the desired value and the error signal for the weight fit is calculated simply by the following equation:

$$\mathbf{e(t) = x(t) - \hat{x}\ (t \mid t - 1, t - 2, . . ., t - M),}$$

Where x (t) refers to the desired output and $\hat{x}$ (t | t − 1, t − 2,..., T − M) represents the observed output of the network given the previous M samples. Prediction is similar to building a physical model of the time series. Insofar as the network has neurons whose transfer function is non-linear.

V. METHODOLOGY OF WORK

University failure in the first year of college has long been a subject of debate. Many psychologists try to figure it out and then explain it. Many statisticians try to predict this. Our objective in this work is to focus on the explanatory factors of the first year of university success and their interactions based on neural networks, so that the students can be divided into three groups: the group of students having a high probability of success, the group of students who could possibly do moderately at university and the group of students who are likely to fail (or drop out) of college.

Our work methodology consists, based on subsequent work, of collecting as much relevant information as possible that allows us to analyze the intermediate and final academic performance of students and to define the factors that influence this performance. These factors which combine structural factors which are stable and processual factors which change. They can be classified into three sets of factors:

- The first of these factors concerns the student's personal information (his identity, his socio-family background, his school background, ...);
- The second of these factors concerns the expression of the student's involvement in their studies or their behavior in relation to them (participation in optional activities, meeting with their teachers to ask questions or obtain feedback from them. 'a partial examination, ...).
- The third of these factors brings together all the student's perceptions (the way in which he perceives the academic context, his teachers, the lessons, ...).

Once the data is collected, we do the data processing based on the methods of decision trees and neural networks. Then, we will present the different results obtained by comparing their performance with those of linear discriminated analysis.

VI. CONCLUSION

As a conclusion of this work, the objective already mentioned at the beginning of this article is to propose a working methodology of our research by defining the instruments chosen for the processing and analysis of the data of our research. Thus, we offered an overview on failure at the university level and especially in the first year, we defined the instruments chosen for the processing and analysis of the data of our research. Thus, we have presented models designed through decision tree and neural networks as proposed models for data analysis and processing. Finally, we proposed our work methodology for carrying out our research. We hope to propose the different stages of our research in future articles.

REFERENCES

1. https://fr.wikipedia.org/wiki/Arbre_de_d%C3%A9cision#cite_ref-2.
2. https://fr.wikipedia.org/wiki/Arbre_de_d%C3%A9cision_(apprentissage)#cite_note-Quinlan86-4.
3. S. Madeh Piryonesi et Tamer E. El-Diraby, « Data Analytics in Asset Management: Cost-Effective Prediction of the Pavement Condition Index », Journal of Infrastructure Systems, vol. 26, n° 1, mars 2020, p. 04019036 (ISSN 1076-0342 et 1943-555X,
DOI 10.1061/(ASCE)IS.1943-555X.0000512.
4. Xindong Wu, Vipin Kumar, J. Ross Quinlan et Joydeep Ghosh, « Top 10 algorithms in data mining », Knowledge and Information Systems, vol. 14, n° 1, janvier 2008, p. 1–37 (ISSN 0219-1377 et 0219-3116, DOI 10.1007/s10115-007-0114-2.
5. Lior Rokach, Data mining with decision trees : theory and applications, Hackensack (NJ), World Scientific Pub Co Inc, 2008, 244 p. (ISBN 978-981-27-7171-1, notice BnF n° FRBNF41351943).
6. Leo Breiman, Classification and regression trees, Monterey, CA, Wadsworth & Brooks/Cole Advanced Books & Software, 1984, 368 p. (ISBN 978-0-412-04841-8).
7. William James (1890). The Principles of Psychology. Henry Holt and Company, New York, USA.
8. W.S. McCulloh, et W.Pitts."A logical calculus of the ideas immanent in nervous activity".the bullettin of mathematical biophysics,,vol.,no 4, p. 115-133. 1943
9. D. 0. HEBB. (1949). The Organization of Behavior, A NEUROPSYCHOLOGICAL THEORY. McGill University. New York JOHN WILEY if SONS, Inc. London CHAPMAN i HALL, Limited.
10. In 1951 Marvin Minsky teamed with Dean Edmonds build the first artificial neural network that simulated a rat finding its way through a maze.
11. F. Rosenblatt (1958), The perceptron: a probabilistic model for information storage and organization in the brain, - repris dans J.A. Anderson & E. Rosenfeld (1988), Neurocomputing. Foundations of Research, MIT Press
12. B. Widrow, (1960). Adaptive sampled-data systems, IFAC Moscow Congress Record, Butterworth Publications, London.

13. Minsky, M., & Papert, S. (1969). Perceptrons : An Introduction to Computational Geometry, MIT Press, 268 p.
14. Grossberg, S.: Adaptive pattern classification and universal recoding. II. Feedback, expectation, olfaction, illusions. Biol. Cybern. 23, 187–202 (1976).
15. Kohonen, T.: Associative memory—a system-theoretical approach. Berlin, Heidelberg, New York: Springer 1977.
16. Anderson, J.A.: A simple neural network generating an interactive memory. Math. Biosci. 14, 197–220 (1972)
17. J J Hopfield, (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences Apr 1982, 79 (8) 2554-2558; DOI: 10.1073/pnas.79.8.2554
18. Rumelhart, D. E., Durbin, R., Golden, R., and Chauvin, Y. (1995). Ûackpropagation: The basic theory. In Y. Chauvin and D. E. Rumelhart (eds.),Backpropagation:Theory, Architectures, and Applications. Ðillsdale: NJ. Lawrence Erlbaum, 1-34. Chapter 1.
19. Wang, L.X. et Mendel, J.M. (1992). Adaptive m1mmum prediction error deconvolution and source wavelet estimation using Hopfield neural networks. Geophysics, 57, 670-679.
20. Murat, M.E. and Rudman, A.J. " (1992). Automated first arrival picking: a neural network approach. Geophysical Prospecting " 40,587-604, .
21. Röth, G., Tarantola, A. (1994). "Neural networks and inversion of seismic data". journal of geophysical research, VOL. 99, NO. B4, PAGES 6753-6768, APRIL 10,.
22. Y. Zhang, K. V. Paulson " (1997a). Magnetotelluric inversion using regularized Hopfield neural networks" Geophysical Prospecting, 45, 725–743,
23. Y. Zhang, K. V. Paulson " (1997b). Enhancement of Signal-to-noise Ratio in Natural-source Transient Magnetotelluric Data with Wavelet Transform" appl. geophys. 149 - 405–419,.
24. P. J. Drew, J. R. T. Monson, (2000). "Artificial neural networks".127(1),3–11.
25. http://informatique.coursgratuits.net/methodes-numeriques/reseaux-de-neurones-formels.php.
26. http://www.grappa.univ-lille3.fr/polys/apprentissage/sortie005.html
27. Antoine Cornuéjols, Laurent Miclet, Yves Kodratoff, Apprentissage Artificiel : Concepts et algorithmes, Eyrolles, 2002 (ISBN 2-212-11020-0)
28. Christopher M. Bishop, Pattern Recognition And Machine Learning, Springer, 2006
29. Guillaume Cleuziou, Une méthode de classification non-supervisée pour l’apprentissage de règles et la recherche d’information, 10 juillet 2006.
30. F. Rosenblatt (1958), The perceptron: a probabilistic model for information storage and organization in the brain, - repris dans J.A. Anderson & E. Rosenfeld (1988), Neurocomputing. Foundations of Research, MIT Press.
31. D. 0. HEBB. (1949). The Organization of Behavior, A neuropsychological theory. McGill University. New York John Wiley if SONS, Inc. London CHAPMAN i HALL, Limited.
32. Claude Touzet. Les réseaux de neurones artificiels, introduction au connexionnisme : cours, exercices et travaux pratiques. EC2, 1992, Collection de l’EERIE, N. Giambiasi. hal-01338010.