ARID ZONE JOURNAL OF ENGINEERING, TECHNOLOGY & ENVIRONMENT
AZOJETE June 2020. Vol. 16(2):243-254
Published by the Faculty of Engineering, University of Maiduguri, Maiduguri, Nigeria.
Print ISSN: 1596-2490, Electronic ISSN: 2545-5818
www.azojete.com.ng

ORIGINAL RESEARCH ARTICLE

COMPUTATIONAL MODEL OF ARTIFICIAL NEURAL NETWORKS AND ITS APPLICATIONS IN DATA MINING

J. Abah[1*], E. S. Alu[2], A. E. Chahari[3]
([1]Department of Computer Engineering, University of Maiduguri, Maiduguri, Nigeria
[2]Department of Agricultural Economics and Extension, Nasarawa State University, Keffi, Shabu-Lafia Campus, Lafia, Nigeria
[3]Department of General Studies Education, School of General Education, Federal College of Education, Yola, Nigeria)
**Corresponding author's email address: abah@unimaid.edu.ng*

ARTICLE INFORMATION

Submitted 11 July, 2017
Revised 15 October, 2019
Accepted 04 April, 2020

Keywords:
Artificial Neural Network
Databases
Data Mining
Information
Organizations

ABSTRACT

Data remain a very important ingredient required by any organization to make informed decision as it affects operations. Companies have been collecting data from various sources over the decades bringing about a very large volume of data warehouse. Unfortunately, most organizations build databases which are redundant and never used for any meaningful thing. While few companies use the data collected in their databases when taking strategic decisions others barely do same. However, for an organization to immensely derive benefits from the massive data warehouse, there is the need for an effective and efficient means of analysing the data with a view to extracting meaningful knowledge that is sufficient to achieve organizational goal. To achieve this, Artificial Neural Network (ANN) technique through the concept known as data mining is presented. The paper reviewed artificial neural network technique for data mining, examines the computational model behind this technique and analysed its use and application as a predicting or forecasting tool. Results shows that ANN' has capability in data management, analysis and able to provide desirable knowledge for management decision making processes. It is therefore recommended that data mining tools like ANN and others be applied to organization's databases which hitherto have not been minned in order to provide management with intelligence for decision making.

1.0 Introduction

As long as knowledge is another name of power, many organizations give much importance to knowledge. When reaching this knowledge, they make use of the data in their databases. Data is important since it enable them to learn from the past and to predict future trends and behaviours (Larose, 2005). Today, most of the organizations use the data collected in their databases when taking strategic decisions (Singh and Chauhan 2009). The process of using the data to reach this knowledge consists of two steps as to collecting the data and analysing the data. In the beginning, organizations are faced with difficulties when collecting the data so they have not enough data in order to make suitable analysis (Kartalopoulos and Stamatios, 1996). In the long run, with the rapid computerization, organizations are able to store huge amount of data easily. But at this time, they are faced with another problem when analysing and interpreting such large datasets. Traditional methods like statistical techniques or data management tools are not sufficient anymore (Berry and Linoff, 2000).

In order to manage this problem, the technique called Data Mining (DM) has been discovered to address this problem. According to Singh and Chauhan (2009) show that data mining is a useful and powerful technology that supports organizations to derive strategic information in their databases. It has been defined as: "The process of exploration and analysis by automatic or semi-automatic means of large quantities of data in order to discover meaningful patterns and rules" (Berry and Linoff, 1997; Berry and Linoff, 2000). The expression "meaningful patterns and rules" implies the followings; easily understood by humans, valid on new data, potentially useful and novel. Validating a hypothesis that the user wants to prove can also be accepted as a meaningful patterns and rules (Berry and Linoff, 2000). In sum, it is essential to derive patterns and rules that help us to reach strategic and unimagined information in data mining.

Few organizations use the data collected in their databases when taking strategic decisions. This has made most organization's databases redundant. The process of using accumulated data to reach knowledge consists of two steps; collecting the data and analyzing the data. The problem associated with the process of achieving these steps are first, most organizations are faced with difficulties when collecting the data so, they have not enough data in order to make suitable analysis and second, the technique of analysing the collected data in order to derive meaning or knowledge from it is often not sufficient and efficient to achieve the goal (Cilimkoviv, 2015).

In order to solve the stated problems, we explore ANN as a data mining technique to bring to bear, it's capability in data management and analysis and its ability to provide desirable knowledge for informed management decision making processes. Hence, we present ANN as a useful and powerful data mining technique that can support organizations to derive strategic information from their databases. The rest of this paper is organized as follows: Section two provides details about ANN; concept, models and topologies. Section three provides experimental procedure of ANN technique, section four gives the discussion of results and finally, section five gives the summary, conclusion and recommendation.

2. Artificial Neural Networks

An artificial neural network (ANN), just called a "neural network" (NN), is a mathematical model or computational model based on biological neural networks (Berry and Linoff, 1997). In other words, ANN is an emulation of biological neural system. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase (Singh and Chauhan, 2009; Zurada, 1992).

2.1 Concept of Neural Network

Neural networks are very sophisticated modeling techniques capable of modelling extremely complex functions. Neural networks learn by example. The neural network user gathers representative data, and then invokes training algorithms to automatically learn the structure of the data. It is a model for classification and prediction (continuous value) of an outcome variable (Beale and Jackson, 1990).

2.2 The Basic Artificial model

The artificial neuron was developed in an effort to model the human neuron (Kartalopoulos and Stamatios, 1996; Haykin and Simon, 1994). Inputs enter the neuron and are multiplied by their respective synaptic weights. They are then summed and processed by an activation function. Examples of activations functions includes the logistic or sigmoid function (1).

$$f(x) = 1/ (1+\exp(-x)) \quad (1)$$

while the second is the Gaussian function (2)

$$f(x) = \exp(-x^2). \quad (2)$$

Where the variable x represents the input vector in both equations.

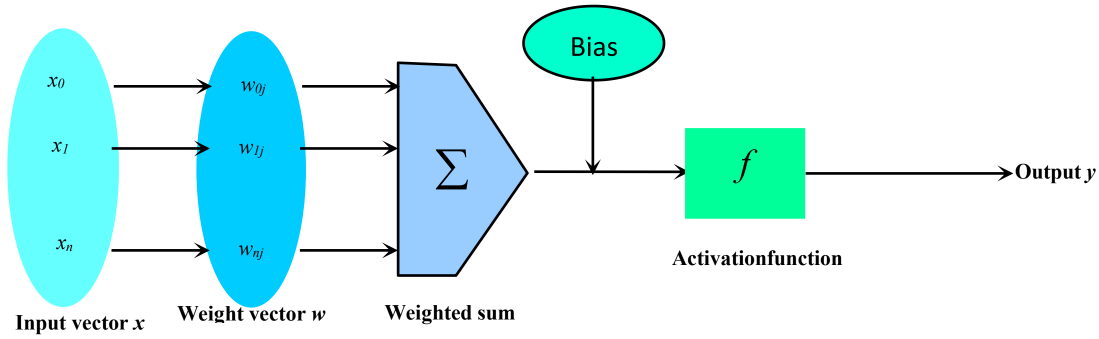

Figure 1: Artificial Neural Network Model

2.3 Neural Network Topologies

Feed Forward Neural Network: The feedforward neural network was the first and arguably simplest type of artificial neural network devised (Singh and Chauhan, 2009). In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network. The data processing can extend over multiple (layers of) units, but no feedback connections are present, that is, connections extending from outputs of units to inputs of units in the same layer or previous layers (Zurada, 1992).

Recurrent network: Recurrent neural networks that do contain feedback connections. Contrary to feedforward networks, recurrent neural networks (RNs) are models with bi-directional data flow. While a feedforward network propagates data linearly from input to output, RNs also propagate data from later processing stages to earlier stages (Singh and Chauhan, 2009).

2.4 Data Mining

Data mining is a set of techniques and procedures that can be developed from various data sources such as data warehouses or relational databases (Villanueva et al., 2018). Flat files without formats are made from this predictive analysis using statistical study techniques to predict or anticipate statistical measures of certainty based on existing facts. Data mining is seen as the evolution of information technology (Jiawei and Kamber, 2001), this is largely supported by the growth of the internet. Data mining is a multidisciplinary field that allows relevant information to be obtained from large amounts of data at the confluence among other disciplines namely: artificial intelligence, statistics, databases and information science. Figure 2 shows data mining taxonomy.

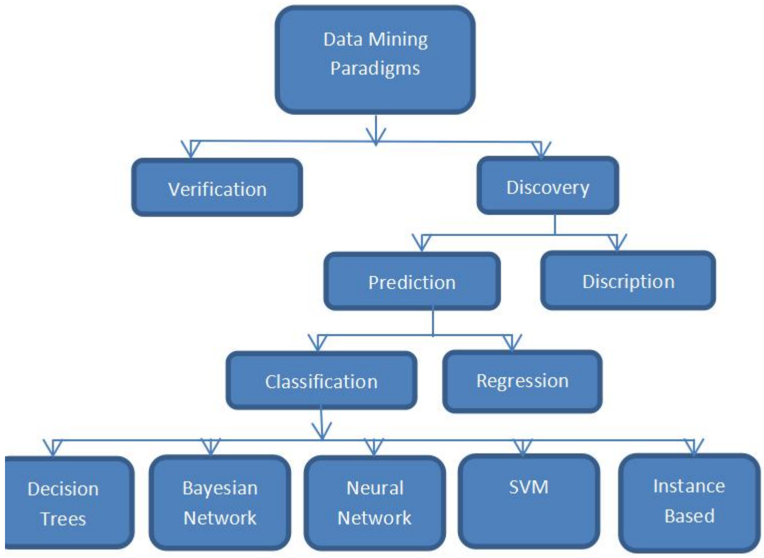

Figure 2: Data Mining Taxonomy (Maimon and Rokach, 2010)

2.4.1 Phases of Data Mining Task

The data mining tasks are implemented with the goal of discovering patterns of relevant and interesting information in large volumes. This is done with the development of four phases according to Villanueva et al. (2018), which are usually:

Filtering data;

Selection of variables;

Extracting knowledge and

Interpretation and evaluation.

In general, all techniques have been proved in educational settings according to Siemens and Baker (2012), and different case studies have been developed to evaluate the performance of different techniques and to meet the main goals of data mining.

2.4.2 Neural Networks in Data Mining

In more practical terms neural networks are non-linear statistical data modelling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data (Singh and Chauhan, 2009). Using neural networks as a tool, data warehousing firms are harvesting information from datasets in the process known as data mining. The difference between these data warehouses and ordinary databases is that there is actual manipulation and cross-fertilization of the data helping users make more informed decisions. Neural networks essentially comprise three pieces (Zurada, 1992): The architecture or model; the learning algorithm; and the activation functions. Neural networks are programmed or "trained" to store, recognize, and associatively retrieve patterns or database entries; to solve combinatorial optimization problems; to filter noise from measurement data; to control ill-defined problems; in summary, to estimate sampled functions when we do not know the form of the functions." It is precisely these two abilities (pattern recognition and function estimation) which make artificial neural networks (ANN) so prevalent a utility in data mining.

According to Villanueva, Moreno and Salinas (2018), the first known work on Educational Data Mining titled "Explaining student grades Predicted by a neural network" (Gedeon and Turner, 1993). In this work, the researchers used neural networks were used to predict the final grades of students. Later, in 1994, Fausett and Elwasif published "Predicting performance from test

Corresponding author's e-mail address: abah@unimaid.edu.ng

scores using back propagation and counter propagation" in which neural networks are used to predict the student's performance (Fausett and Elwasif, 1994).

"Predicting and analyzing secondary education placement-test scores: A data mining approach" by Şen, Ucar and Delen (2012) relied on neural networks and decision trees to predict outcomes of high school students in Turkey. Bayesian networks were used to work educational situations, the first of them is "Predicting student's academic performance artificial using neural network: A case study of an engineering course" (Oladokun et al., 2008) where the researchers explain how to predict the performance a candidate might have if it is accepted in some university courses.

Additionally, Sundar (2013) in his work "A Comparative Study for Predicting Students Academic Performance using Bayesian Network Classifiers" made a comparison of obtained results using Bayesian networks in predicting student performance; Likewise, Bhise et al.,(2013) published "Importance of Data Mining in Higher Education System" their work is based on clustering to help instructors improving student performance.

Today, educational and business information systems store large volume of data and its origin can come from different sources, different formats and different granularity levels (Villanueva et al., 2018). The problems of business and educational data mining must be analyzed particularly due to their specific objective determines a singularity when it is solved by data mining techniques. For instance, in the domain of education, Data Mining for education is so important that in 2007 an international organization dedicated to researching this discipline was founded. Such is the importance has prompted this research area that it is estimated that by the year 2022 all research related to education will involve analysis and data mining (Baker and Yacef, 2009).

2.5 Feedforward Neural Network

One of the simplest Feedforward Neural Networks (FFNN), such as in figure 3, consists of three layers: an input layer, hidden layer and output layer. In each layer there are one or more processing elements (PEs) (Singh and Chauhan, 2009). PEs is meant to simulate the neurons in the brain and this is why they are often referred to as neurons or nodes.

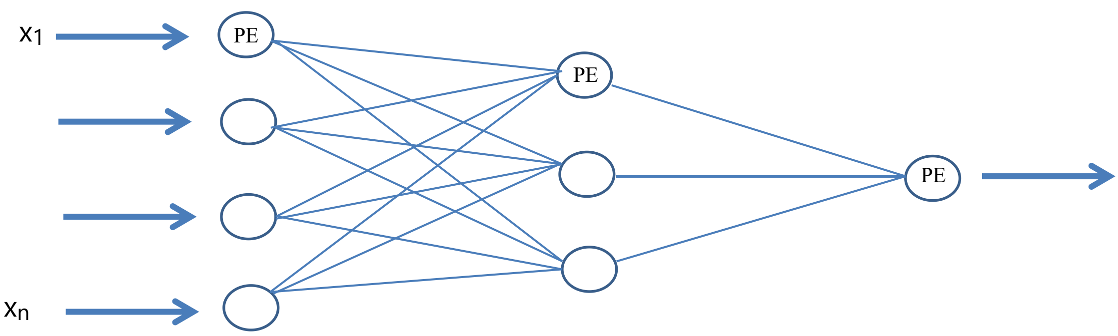

Figure 3: Feedforward Neural Network

A PE receives inputs from either the outside world. The simplified process for training a FFNN is as follows:

1. Input data is presented to the network and propagated through the network until it reaches the output layer. This forward process produces a predicted output.

2. The predicted output is subtracted from the actual output and an error value for the networks is calculated.

3. The neural network then uses supervised learning, which in most cases is back propagation, to train the network. Back propagation is a learning algorithm for adjusting the weights. It starts with the weights between the output layer PE's and the last hidden layer PE's and works backwards through the network.

4. Once back propagation has finished, the forward process starts again, and this cycle is continued until the error between predicted and actual outputs is minimized.

2.6 Artificial Neuron using Back Propagation Learning

The mathematical formulae used for Neural Network in Back propagation learning algorithm according to Dhar et al. (1996) presented in Figure 4, was used.

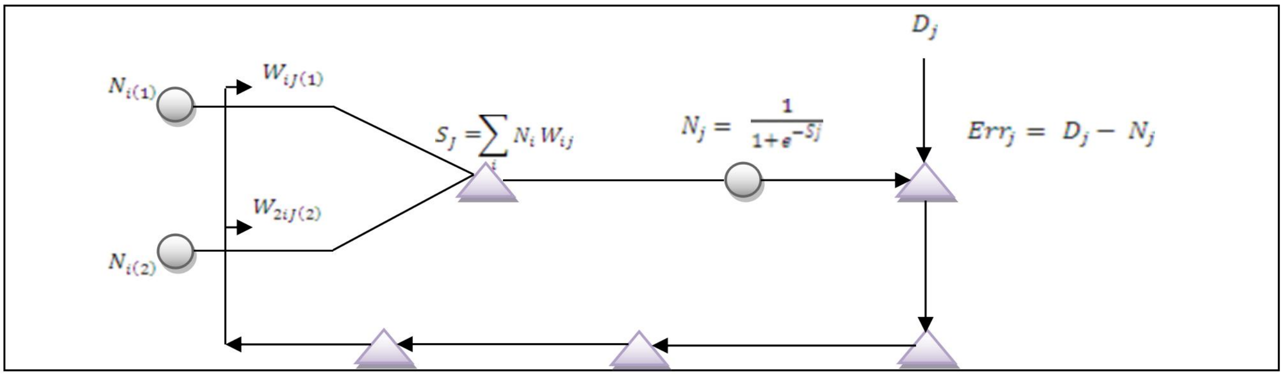

Figure 4: Artificial Neuron using Back Propagation Learning

Figure 4 depicts a single artificial neuron and the actual computation that occurs in neural network in a single epoch through the data.

2.7 Fitting a Network to Data

The equations in Figure 4 provides a good understanding on how neural network process data. However, for the purpose of clarity and understanding, we illustrate the actual computation that occurs in neural network in a single epoch through the data. Typically, neural network performs exceptionally well with large dataset but for the purpose of this illustration a small dataset is used as shown in Table 1. Refer to Figure 5 for the simulated neural network for the dataset provided in Table 1.

Table 1: Sample Dataset

Observation	Milk Score	Sugar Score	Price
1	0.2	0.9	0.65
2	0.1	0.1	0.34
3	0.2	0.4	0.42
4	0.2	0.5	0.49
5	0.4	0.5	0.72
6	0.3	0.8	0.83

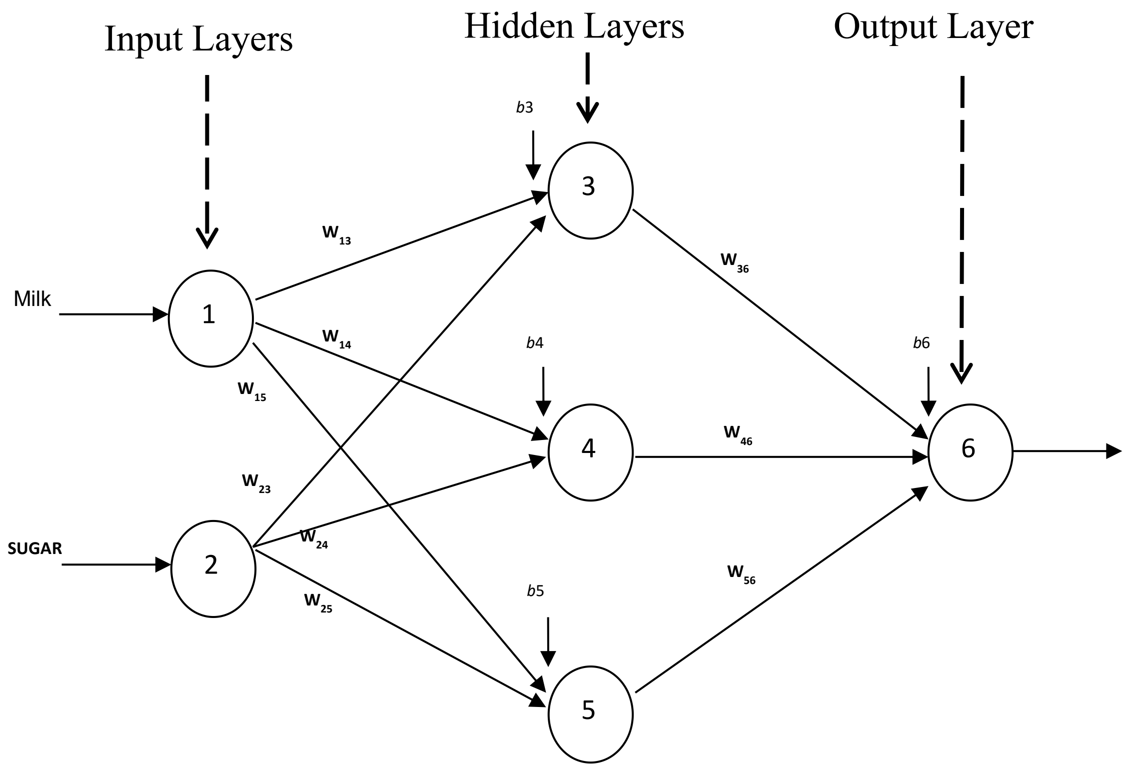

Figure 5: Simulated Neural Network

Considering the first observation: for the input at the input layer; milk = 0.2 and sugar = 0.9, the output of this layer is x1 = 0.2 and x2 = 0.9.

Hidden layer nodes take as input the output values from the input layer. The hidden layer in this example consists of 3 nodes. Each of the nodes receives input from all the input nodes (Figure 4). To compute the output of a hidden layer node, the weighted sum of the inputs is computed and then applies a certain function to it.

3. Experimental Procedure

The procedure describes the steps involved in the computational process. First we describe the artificial neural network as follows;

3.1 The Artificial Neural Network Model

Let x1, x2,..., xp be the set of input values to a node, the weighted sum is computed thus;

From the general algorithm $\lambda_j + \sum_{i=1}^{p} w_{ij}\, x_i$

We have;

$$y_{in} = b_j + \sum_{i=1}^{n} w_{ij} x_i \tag{3}$$

where: b, $w_{i,j}$,...,wn_j are weights that are initially set randomly then adjusted as the network "learns". the variable b, also known as the "bias" of node j, is a constant that controls the level of contribution of node j.

The next step is to apply a function called transfer function (that is, linear function, exponential function and a logistics/sigmoidal function).

In this paper, the sigmoid function in (1) is used for its practical value from the fact that it has a squashing effect on very small or very large value.

Using the logistic function then, the output of node j in the hidden layer is computed from the general formula:

$$output_j = g(b_j + \sum_{i=1}^{n} w_{ij}x_i) = \frac{1}{1+e^{-b_j + \sum_{i=1}^{n} w_{ij}x_i}} \tag{4}$$

as follows:

$$g(\lambda_j + \Sigma_{i=1}^{p} w_{ij}\, x_i) = \frac{1}{1+e^{-(\lambda_j + \Sigma_{i=1}^{p} w_{ij}\, x_i)}} \tag{5}$$

where:

W_{ij} is the weighted vector between i and j processing elements;

X_i is the input vector in i processing element;

p is the processing element

b_j and λ_j are the bias in (4) and (5) respectively.

3.2 Initializing the Weights

At this stage of the procedure, the values of bj and wij are typically initialize to small (but generally random) number in the range 0.00 ± 0.05. With this value range, we can then compute the outputs of Node3, Node4 and Node5. Suppose the following are the initial weights for;

Node 3 are $bnw_3=b_3=-0.3$, $w_{13}=0.05$ and $w_{23}=0.01$

Node 4 are $bnw_4=b_4= 0.2$, $w_{14}=-0.01$ and $w_{24}=0.03$

Node 5 are $bnw_5=b_5=0.05$, $w_{15}=0.02$ and $w_{25}=-0.01$

Node 6 are $bnw_6=b_6=-0.015$, $w_{36}=0.01$, $w_{46}=0.05$ and $w56=0.015$

where: the input x_1 and x_2 are 0.2 and 0.9 respectively as seen in observation 1 in table 1.

Using the following values in (2), the output of node3, Node4, Node5 are computed as follows;

$$Node_3 = \frac{1}{1+e^{-\{-0.3+(0.05)(0.2)+(0.01)(0.9)\}}} = 0.4302$$

$$Node_4 = \frac{1}{1+e^{-\{0.2+(-0.01)(0.2)+(0.03)(0.9)\}}} = 0.5560$$

$$Node_5 = \frac{1}{1+e^{-\{0.5+(0.02)(0.2)+(-0.01)(0.9)\}}} = 0.5112$$

The output from node3, node4 and node5 are fed into the output layer (node6) to compute the final output for the first observation in our sample data.

$$Node_6 = \frac{1}{1+e^{-\{-0.015+(0.01)(output3)+(0.05)(output4)+(0.015)(output5)\}}}$$

$$Node_6 = \frac{1}{1+e^{-\{-0.015+(0.01)(0.4302)+(0.05)(0.5560)+(0.015)(0.5112)\}}} = 0.5062$$

3.3 Training the Model

After initializing the weights, the next step is to train the model. Training the model means estimating the weights b_j (bias node) and w_{ij} (link between nodes) that lead to the best predictive results. The process illustrated above for computing the neural network output for an observation is repeated for all the observations in the training set. For each observation, the model produces a prediction which is then compared with the actual response value. Their difference is the error for the output node. This error obtained from the output node is use iteratively to update the estimated weights. In particular, the error for the output node is distributed across all the hidden nodes that led to it, so that each node is assigned "responsibility" for part of the error. Each of these node-specific errors is then used for updating the weights.

3.4 Back Propagation of Error

Back propagation is a process for using model errors to update weights ("learning"). As the name implies, errors are computed from the output layer back to the hidden layers. The error associated with output node is computed by the general formula;

$$Err_j = N_j\,(1-N_j)\,(D_j-N_j) \quad (6)$$

The above equation is similar to the ordinary definition of an error (Dj – Nj) multiply by a correction factor. The weights are then updated as follows:

$$ib_j^{new} = bj^{old} + lErr_j \quad (7)$$

$$w_{ij}^{new} = w_{ij}^{old} + lErr_j \quad (8)$$

where: l is a learning rate or weight decay parameter, a constant ranging typically between 0 and 1, which controls the amount of change in weights from one iteration to the other.

After the first iteration, the error associated with the output node is;

Err = 0.5062(1-0.5062) (0.65-0.5062) = 0.03594.

Hence the weight update for second iteration is computed as follows using l = 0.5

b_3: -0.3 + 0.5 (0.03594) = -0.282
b_4: 0.2 + 0.5 (0.03594) = 0.218
b_5: 0.05 + 0.5 (0.03594) = 0.068
b_6: -0.015 + 0.5 (0.03594) = 0.003
W_{13}: 0.08 + 0.5 (0.03594) = 0.068
W_{23}: 0.01 + 0.5 (0.03594) = 0.028
W_{14}: -0.01 + 0.5 (0.03594) = 0.008
W_{24}: 0.03 + 0.5 (0.03594) = 0.048
W_{15}: 0.02 + 0.5 (0.03594) = 0.038
W_{25}: -0.01 + 0.5 (0.03594) = 0.008
W_{36}: 0.01 + 0.5 (0.03594) = 0.028
W_{46}: 0.05 + 0.5 (0.03594) = 0.068
W_{56}: 0.015 + 0.5 (0.03594) = 0.033

The adjusted weights are used on the second observation (i.e. x_1=0.1 and x_2=0.1) to computer the second output of the model. This process is repeated for the entire training data (in this case 6). Hence one sweep through the data, is called an epoch, consisting of six (6) iterations.

For the purpose of this simulation, the network is trained using 5 epochs so, there will be a total of 30 iterations.

4. Results and Discussion

Table 2 reveal the computation result of the adjusted weights, hidden layers output and the final output for each epoch and the total sum of error associated with each epoch.

Table 2: Output Errors Generated for each of the Five (5) Epochs

EPOCH 1

s/n	x1	x2	price	bnw3	bnw4	bnw5	bnw6	w13	w23	w14	w24	w15	w25	w36	w46	w56	output3	output4	output5	output6	error
1	0.2	0.9	0.65	-0.3	0.2	0.05	-0.015	0.05	0.01	-0.01	0.03	0.02	-0.01	0.01	0.05	0.015	0.4302	0.556	0.511	0.5062	0.03595
2	0.1	0.1	0.34	-0.282	0.22	0.07	0.003	0.07	0.028	0.008	0.048	0.038	0.008	0.028	0.07	0.033	0.4323	0.5557	0.518	0.5175	-0.0443
3	0.2	0.4	0.42	-0.3042	0.2	0.05	-0.019	0.05	0.006	-0.01	0.026	0.016	-0.01	0.006	0.05	0.011	0.4273	0.5507	0.511	0.5035	-0.0209
4	0.2	0.5	0.49	-0.3146	0.19	0.04	-0.03	0.04	-0	-0.02	0.015	0.005	-0.02	-0.005	0.04	4E-04	0.4231	0.5469	0.506	0.497	-0.0017
5	0.4	0.5	0.72	-0.3155	0.18	0.03	-0.03	0.03	-0.01	-0.03	0.015	0.005	-0.03	-0.005	0.03	-5E-04	0.4245	0.5453	0.506	0.4964	0.05589
6	0.3	0.8	0.83	-0.2876	0.21	0.06	-0.003	0.06	0.022	0.002	0.042	0.032	0.002	0.022	0.06	0.027	0.4376	0.5615	0.519	0.5141	0.0789
																					0.1038

EPOCH 2

s/n	x1	x2	price	bnw3	bnw4	bnw5	bnw6	w13	w23	w14	w24	w15	w25	w36	w46	w56	output3	output4	output5	output6	error
1	0.2	0.9	0.65	-0.2481	0.25	0.1	0.037	0.1	0.062	0.042	0.082	0.072	0.042	0.062	0.1	0.067	0.4571	0.5827	0.538	0.5401	0.02731
2	0.1	0.1	0.34	-0.2344	0.27	0.12	0.051	0.12	0.076	0.056	0.096	0.086	0.056	0.076	0.12	0.081	0.4464	0.5697	0.532	0.5481	-0.0515
3	0.2	0.4	0.42	-0.2602	0.24	0.09	0.025	0.09	0.05	0.03	0.07	0.06	0.03	0.05	0.09	0.055	0.4446	0.568	0.528	0.5317	-0.0278
4	0.2	0.5	0.49	-0.2741	0.23	0.08	0.011	0.08	0.036	0.016	0.056	0.046	0.016	0.036	0.08	0.041	0.44	0.5639	0.523	0.5227	-0.0082
5	0.4	0.5	0.72	-0.2782	0.22	0.07	0.007	0.07	0.032	0.012	0.052	0.042	0.012	0.032	0.07	0.037	0.4419	0.5628	0.524	0.5201	0.04989
6	0.3	0.8	0.83	-0.2533	0.25	0.1	0.032	0.1	0.057	0.037	0.077	0.067	0.037	0.057	0.1	0.062	0.4554	0.5791	0.536	0.5366	0.07295
																					0.06264

EPOCH 3

s/n	x1	x2	price	bnw3	bnw4	bnw5	bnw6	w13	w23	w14	w24	w15	w25	w36	w46	w56	output3	output4	output5	output6	error
1	0.2	0.9	0.65	-0.2168	0.28	0.13	0.068	0.13	0.093	0.073	0.113	0.103	0.073	0.093	0.13	0.098	0.4735	0.5986	0.555	0.5613	0.02183
2	0.1	0.1	0.34	-0.2059	0.29	0.14	0.079	0.14	0.104	0.084	0.124	0.114	0.084	0.104	0.14	0.109	0.4549	0.5781	0.541	0.5668	-0.0557
3	0.2	0.4	0.42	-0.2337	0.27	0.12	0.051	0.12	0.076	0.056	0.096	0.086	0.056	0.076	0.12	0.081	0.4551	0.5784	0.539	0.5491	-0.032
4	0.2	0.5	0.49	-0.2497	0.25	0.1	0.035	0.1	0.06	0.04	0.08	0.07	0.04	0.06	0.1	0.065	0.4503	0.5741	0.534	0.5386	-0.0121
5	0.4	0.5	0.72	-0.2557	0.24	0.09	0.029	0.09	0.054	0.034	0.074	0.064	0.034	0.054	0.09	0.059	0.4524	0.5732	0.534	0.5348	0.04607
6	0.3	0.8	0.83	-0.2327	0.27	0.12	0.052	0.12	0.077	0.057	0.097	0.087	0.057	0.077	0.12	0.082	0.4661	0.5896	0.547	0.5505	0.06917
																					0.03733

EPOCH 4

s/n	x1	x2	price	bnw3	bnw4	bnw5	bnw6	w13	w23	w14	w24	w15	w25	w36	w46	w56	output3	output4	output5	output6	error
1	0.2	0.9	0.65	-0.1981	0.3	0.15	0.087	0.15	0.112	0.092	0.132	0.122	0.092	0.112	0.15	0.117	0.4832	0.608	0.564	0.5743	0.01852
2	0.1	0.1	0.34	-0.1889	0.31	0.16	0.096	0.16	0.121	0.101	0.141	0.131	0.101	0.121	0.16	0.126	0.4599	0.5831	0.546	0.578	-0.0581
3	0.2	0.4	0.42	-0.2179	0.28	0.13	0.067	0.13	0.092	0.072	0.112	0.102	0.072	0.092	0.13	0.097	0.4614	0.5845	0.545	0.5597	-0.0344
4	0.2	0.5	0.49	-0.2351	0.26	0.11	0.05	0.11	0.075	0.055	0.095	0.085	0.055	0.075	0.11	0.08	0.4564	0.5801	0.54	0.5483	-0.0144
5	0.4	0.5	0.72	-0.2423	0.26	0.11	0.043	0.11	0.068	0.048	0.088	0.078	0.048	0.068	0.11	0.073	0.4587	0.5795	0.541	0.5437	0.04373
6	0.3	0.8	0.83	-0.2205	0.28	0.13	0.065	0.13	0.09	0.07	0.11	0.1	0.07	0.09	0.13	0.095	0.4725	0.5958	0.554	0.5588	0.06686
																					0.02218

EPOCH 5

s/n	x1	x2	price	bnw3	bnw4	bnw5	bnw6	w13	w23	w14	w24	w15	w25	w36	w46	w56	output3	output4	output5	output6	error
1	0.2	0.9	0.65	-0.187	0.31	0.16	0.098	0.16	0.123	0.103	0.143	0.133	0.103	0.123	0.16	0.128	0.4891	0.6135	0.57	0.582	0.01654
2	0.1	0.1	0.34	-0.1788	0.32	0.17	0.106	0.17	0.131	0.111	0.151	0.141	0.111	0.131	0.17	0.136	0.4629	0.586	0.549	0.5847	-0.0594
3	0.2	0.4	0.42	-0.2085	0.29	0.14	0.077	0.14	0.102	0.082	0.122	0.112	0.082	0.102	0.14	0.107	0.4652	0.5882	0.549	0.566	-0.0359
4	0.2	0.5	0.49	-0.2264	0.27	0.12	0.059	0.12	0.084	0.064	0.104	0.094	0.064	0.084	0.12	0.089	0.4601	0.5837	0.543	0.5541	-0.0158
5	0.4	0.5	0.72	-0.2343	0.27	0.12	0.051	0.12	0.076	0.056	0.096	0.086	0.056	0.076	0.12	0.081	0.4625	0.5832	0.544	0.5491	0.04231
6	0.3	0.8	0.83	-0.2132	0.29	0.14	0.072	0.14	0.097	0.077	0.117	0.107	0.077	0.097	0.14	0.102	0.4764	0.5995	0.557	0.5638	0.06546
																					0.01318

From Table 2, we observed that the total sum of error decreases after every epoch but after a while it begins to increase (for a large dataset). The point of minimum validation error is a good indicator of the best number of epoch for training and the weights at that stage are most likely to provide the best error rate in new data. However, it should be noted that the number of observations in this simulation is too small for estimating the 13 weights, hence the reason for the large error margin between the actual price and the predicted price (Table 3). The negative margin signifies that the predicted value is more than the actual price. As the simulation runs from the sart epoch (epoch 1) to the final epoch (epoch 5), the predicted value seems to decrease hence, an increase in the margin. This is as a result of the learning process taking place from one epoch to the next.

Table 3 gives the predicted values on the sample dataset. The predicted values were obtained from the output values (output 6) of the last epoch (epoch 5).

Table 3: Predicted Values on a Sample Dataset

S/No.	x_1	x_2	Price	Predicted values	Margin (Price - Predicted value)
1	0.2	0.9	0.65	0.58	0.07
2	0.1	0.1	0.34	0.58	-0.24
3	0.2	0.4	0.42	0.57	-0.15
4	0.2	0.5	0.49	0.55	-0.06
5	0.4	0.5	0.72	0.55	0.17
6	0.3	0.8	0.83	0.56	0.27

5. Conclusion

In this paper, we try to bring to the fore an overview of Artificial Neural Networks (ANN) as a prediction tool. It is an attempt to build machine that will mimic brain activities and be able to learn. ANN usually learns by examples. Basic ANN is composed of three layers, input, output and hidden layer. Each layer can have number of nodes and nodes from input layer are connected to the nodes from hidden layer. Nodes from hidden layer are connected to the nodes from output layer. Those connections represent weights between nodes. It further describes one of the most popular ANN algorithms; The Back Propagation (BP) Algorithm. The idea behind BP algorithm is quite simple; output of ANN is evaluated against desired output. If results are not satisfactory, connection (weights) between layers are modified and the process is repeated again and again until error is small enough. Simple BP example is demonstrated in this paper with ANN architecture also covered.

In conclusion, ANN is able to perform classification and even discover new trends or patterns in data when supplied with enough samples. The dataset must be considerably large for ANN to perform well. From the sample simulation, we were able to discover patterns or trend in the price of the given commodities (milk and sugar in this case). As the simulation iterates from one epoch to the next, the predicted value decreases to a given value before it increases again. The minimum value attained before it starts to inrease again point to the best prediction. Similarly, the observed decrease in the margin between stated price and predicted value shows a pattern within the data which is not visible in the given dataset. This is an indication that in most cases neural networks perform better than the traditional statistical techniques. Thus, neural networks are becoming very popular with data mining practitioners, particularly in medical research, finance and marketing. This is because they have proven their predictive power through comparison with other statistical techniques using real data sets (Singh and Chauhan, 2009).

It is recommended therefore that organizations who want to make meaning from their huge data repository should apply ANN models. This is not to say that ANN are perfect but due to design problems neural systems need further research before they are widely accepted in industry. As software companies develop more sophisticated models with user-friendly interfaces, the attraction to artificial neural networks will continue to grow.

References

Baker, RSJ. and Yacef, K. 2009. The state of educational data mining in 2009: A Review and Future visions. Journal of Educational Data Mining, 1(1): 3–17.

Beale, R. and Jackson, T. 1990. Neural Computing: An Introduction. Adam Hilger, Bristol, England., pp.1-14.

Berry, MJA. and Linoff, GS. 1997. Data Mining Techniques. New York: Wiley and Sons., pp. 1-444.

Berry, MJA. and Linoff, GS. 2000. Mastering Data Mining: The art and Science of Customer Relationship Management. New York: Wiley., pp. 1-512.

Bhise, RB., Thorat, SS. and Supekar, AK. 2013. Importance of Data Mining in Higher Education System. IOSR Journal of Humanities and Social Science, 6(6): 18–21.

Dhar, V. and Stein R., 1996. Intelligent Decision Support Methods: The Science of Knowledge Work, Prentice Hall Business Publishing. Upper Saddle River, New Jersey., pp. 11-94.

Fausett, L. and Elwasif, W. 1994. Predicting performance from test scores using back propagation and counter propagation. Proceeding of 1994 IEEE International Conference on Neural Networks (ICNN'94), 5: 3398–3402.

Gedeon, TD. and Turner, HS. 1993. Explaining Student Grades Predicted by a Neural Network. In Proceeding of 1993 IEEE International Conference on Neural Networks (IJCNN-93-Nagoya, Japan), 1: 609–612.

Jiawei, H. and Kamber, M. 2001. Data mining: concepts and techniques. San Francisco, CA, Ltd: Morgan Kaufmann, pp. 377–385.

Larose, DT. 2005. Discovering knowledge in data: an introduction to data mining. Wiley-InterScience., pp. 40-241.

Maimon, O. and Rokach, L. 2010. Data Mining and Knowledge Discovery Handbook 2nd Ed. Springer New York Dordrecht Heidelberg London, 1-37.

Cilimkoviv, M., 2015. Neural Network and Back Propagation Algorithm, Institute of Technology Blanchardstown, Blanchardstown Road North, Dublin, Ireland. 1-13.

Oladokun, VO., Adebanjo, AT. and Charles, B. 2008. Predicting Student's Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course. The Pacific Journal of Science and Technology, 9(1): 72–79.

Şen, B., Uçar, E. and Delen, D. 2012. Predicting and analyzing secondary education placement-test scores: A data mining approach. Expert Systems with Applications, 39: 9468–9476.

Siemens, G. and Baker, RSJD. 2012. Learning analytics and educational data mining: Towards Communication and Collaboration. LAK'12: Proceedings of the ACM 2nd International Conference on Learning Analytics and Knowledge, Vancouver, BC, Canada. 252-254.

Sundar, PP. 2013. A Comparative Study for Predicting Students Academic Performance using Bayesian Network Classifiers. IOSR Journal of Engineering, 3(2): 37–42.

Villanueva, A., Moreno, LG. and Salinas, MJ. 2018. Data mining techniques applied in educational environments: Literature review. Digital Education Review, 33: 235-255.

Singh, Y. and Chauhan, A.S., 2009. Neural Network in Data Mining. Journal of Theoretical and Applied Information Technology, 5(6): 36-42.

Zurada, JM. 1992. Introduction to Artificial Neural Systems. West Publishing Co., St. Paul, New York, United States. pp. 1-764

Corresponding author's e-mail address: abah@unimaid.edu.ng