KNOWLEDGE DISCOVERY IN ARTIFICIAL NEURAL NETWORKS AND REGRESSION MODELS

Ojo, Adebola K. & Adeyemo, Barnabas A.
Department of Computer Science
University of Ibadan
Ibadan, Nigeria
adebola_ojo@yahoo.co.uk, esanadeyemo@gmail.com
*+2347032736013, +2348034747690

ABSTRACT
In this paper, Artificial Neural Networks (ANN) and Regression Analysis models were considered to determine which of them performs better. Prediction was done using one hidden layer and three processing elements in the ANN model. Furthermore, prediction was done using regression analysis. The parameters of regression model were estimated using Least Square method. To determine the better prediction, mean square errors (MSE) attached to ANN and regression models were used. Seven real series were fitted and predicted with in both models. It was found out that the mean square error attached to ANN model was smaller than regression model which made ANN a better model in prediction.

Keywords: Artificial Neural Networks, Regression, Least Square, Processing Element, Hidden Layer, Mean Square Error..

1. INTRODUCTION

Neural Networks (NN), also commonly referred to as Artificial Neural Networks, are information-processing models inspired by the way the densely interconnected, parallel structure of the brain processes information. In other words, neural networks are simplified mathematical models of biological neural networks. The key element of the NN is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements that are analogous to neurons, and tied together with weighted connections that are analogous to synapses. (Kosta Metaxiotis, 2010).

NN are capable of finding internal representations of interrelations within raw data. NN are considered to be intuitive because they learn by example rather than by following programmed rules. The ability to learn is one of the key aspects of NN. This typical characteristic, together with the simplicity of building and training NN, has encouraged their application to the task of prediction. Because of their inherent non-linearity, NN are able to identify the complex interactions between independent variables without the need for complex functional models to describe the relationships between dependent and independent variables.

Recently, the NN approach has been proposed as a substitute for statistical approaches for classification and prediction problems. The advantages of NN over statistical methods include the ability to classify in the presence of nonlinear relationships and the ability to perform reasonably well using incomplete databases. The comparison of the results from NN and statistical approaches indicated that neural networks offer an accurate alternative to classical methods such as multiple regression or autoregressive models (Feuston & Thurtell, 1994; AlFuhaid *et al.*, 1997).

Although the NN concept was first introduced in 1943 (McCulloh & Pitts, 1943), it was not used extensively until the mid-1980's owing to the lack of sophisticated algorithms for general applications, and its need for fast computing resources with large storage capacity. Since the 1980's, various NN architectures and algorithms were developed (*e.g.* the multi-layer perceptron (MLP) which is generally trained with the error backpropagation algorithm, Hopfield Network, Kohonen Network, *etc.*). Neural networks techniques are becoming useful as alternate approaches to conventional techniques or as components of integrated systems. They are used to solve complicated practical problems in various areas which include telecommunications. They are widely accepted as a technology offering an alternative way to tackle complex and ill-defined problems. They are being used in diverse applications in telecommunications.

ANNs are members of a family of statistical techniques, and flexible nonlinear regression models, discriminate models, data reduction models, and nonlinear dynamic systems (Sarle, 1994; Cheng and Tetterington, 1994). They are trainable analytic tools that attempt to mimic information processing patterns in the brain. Because they do not necessarily require assumptions about population distribution, economists, mathematicians and statisticians are increasingly using ANNs for data analysis. Not only do they not require assumptions about the underlying population but are also powerful forecasting tools that draw on the most recent developments in artificial intelligence research. Neural networks are used in a widening range of applications, including airline security control, investment management and risk control (Brockett, Cooper, Golden, and Pitaktong, 1994), bank insolvency Prediction (Al-Shayea, El-Refae, El-Itter and Al-Zaytoonah, 2010), industrial management and production (Satake, Morikawa, and Nakamura, 1994; Eberts and Habibi, 1995), as well as in forecasting stock price indexes and derivative securities (Hutchinson, Poggio, and Lo, 1994; Li, 1994; Fish, Barnes, and Milam, 1995; Shachmurove and Witkowska, 2001), and predicting exchange rates (Kuan and Liu, 1995) and thrift failures.

Consequently, NN models have been used extensively as a tool for modeling, control, forecasting, and optimization in many fields of engineering and sciences such as process control, manufacturing, nuclear engineering, and pattern recognition. So also, regression analysis is used when two or more variables re thought to be systematically connected by a linear relationship. Regression analysis involves finding the best straight line relationship to explain how the variation in an outcome (or dependent) variable, Y, depends on the variation in a predictor (or independent or explanatory) variable, X. (Vinsnes et al 2001)

2. MATERIALS AND METHODS

2.1 Artificial Neural Networks

These are predictive models loosely based on the action of biological neurons. They provide models of data relationships through highly interconnected, simulated "neurons" that accept inputs, apply weighting coefficients and feed their output to other "neurons" which continue the process through the network to the eventual output. Some neurons may send feedback to earlier neurons in the network. Neural networks are "trained" to deliver the desired result by an iterative (and often lengthy) process where the weights applied to each input at each neuron are adjusted to optimize the desired output.

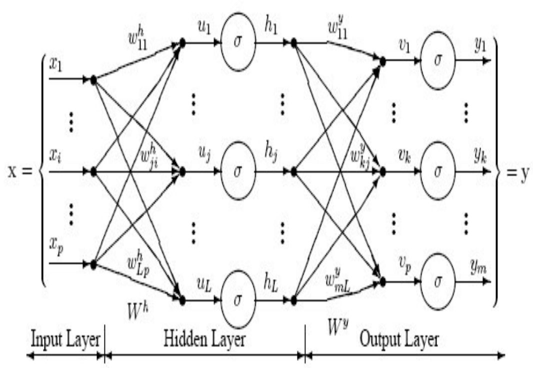

Figure 1: The Feedforward Perceptron Neural Network Model

Figure 1 above is a full-connected, three layer, feed-forward, perceptron neural network. The output from each input and hidden neuron is distributed to all of the neurons in the following layer. Also the values only move from input to hidden to output layers; no values are fed back to earlier layers (a Recurrent Network allows values to be fed backward). This network has an input layer (on the left) with three neurons, one hidden layer (in the middle) with three neurons and an output layer (on the right) with three neurons. There is one neuron in the input layer for each predictor variable. In the case of categorical variables, *N*-1 neurons are used to represent the *N* categories of the variable.

Input Layer — A vector of predictor variable values ($x_1 \ldots x_p$) is presented to the input layer. The input layer (or processing before the input layer) standardizes these values so that the range of each variable is -1 to 1. The input layer distributes the values to each of the neurons in the hidden layer. In addition to the predictor variables, there is a constant input of 1.0, called the *bias* that is fed to each of the hidden layers; the bias is multiplied by a weight and added to the sum going into the neuron.

Hidden Layer — Arriving at a neuron in the hidden layer, the value from each input neuron is multiplied by a weight (w_{ji}), and the resulting weighted values are added together producing a combined value u_j. The weighted sum (u_j) is fed into a transfer function, σ, which outputs a value h_j. The outputs from the hidden layer are distributed to the output layer.

Output Layer — Arriving at a neuron in the output layer, the value from each hidden layer neuron is multiplied by a weight (w_{kj}), and the resulting weighted values are added together producing a combined value v_j. The weighted sum (v_j) is fed into a transfer function, σ, which outputs a value y_k. The *y* values are the outputs of the network.

If a regression analysis is being performed with a continuous target variable, then there is a single neuron in the output layer, and it generates a single y value. For classification problems with categorical target variables, there are *N* neurons in the output layer producing *N* values, one for each of the *N* categories of the target variable.

2.2 The Neural Networks Topology

This neural network is called Time-Lagged Feedforward Network (TLFN). It is a Multi-Layer Perceptron (MLP) with memory components to store past values of the data in the network. The memory components allow the network to learn relationships over time. It is the most common temporal supervised neural network. It consists of multiple layers of processing elements (PEs) connected in a feedforward fashion. The PEs in NeuroSolutions are the orange circular icons and are called axons. The connections between the PEs are the icons with horizontal and diagonal lines between the axons and are called synapses. NeuroSolutions uses the backpropagation of errors to train the MLP. The smaller icons on top of the axons and synapses are called backpropagation components and pass the error backwards from the end of the network to the beginning. The green axons on top of the backpropagation components are called gradient search components and adjust the weights contained in the synapses and axons – this is how the network is trained.

Networks constructed using the "low complexity" setting has one hidden layer. Those with medium or high complexity have two hidden layers. The number of PEs in the first hidden layer is contained in the properties of the 2nd AXON from the left. The number of PEs in the second hidden layer is contained in the properties of the 3rd AXON from the left. The right most axon is called the criterion and reads the desired file from the attached file component and determines the error in the network. The axon 2nd from the right is the output axon and generates the actual network outputs. The axon on the far left is called the input axon and doesn't do anything but accept the input from the file component.

2.3 Linear Regression Model

Here we wish to determine the relationship between a single independent variable X and a dependent variable Y. The independent variable X is assumed to be a continuous mathematical variable controllable by the experiment suppose that the true relationship between Y and X is linear and that the observation Y at each level of X is a random variable. The expected value of Y for each value of X is E(Y/X) = a+bx where a and b are unknown constants. We also assume that each observation Y can be described by model:

$$Y = \beta_0 + \beta_1 X + \in_i$$

Where $\in_i$ is random variable with

$$E(\in_i) = 0$$
$$V(\in_i) = \sigma^2$$
$$E(\in_i \in_j) = 0$$

and $E(X_i \in_i) = 0$

Where Y is a dependent variable
X is an independent variable
E is an error or disturbance term

2.4 Least Square Estimators

To estimate β_0 and β_1 we consider n observations (X_1Y_1) (X_2Y_2) … (X_nY_n). If we write the line of best fit $\hat{Y} = \beta_0 + \beta_1 X$ where $\hat{Y}$ is used to distinguish between the observed value Y and the corresponding value $\hat{Y}$ on the line, then the least square criterion required that we minimized the sum of the square of the deviation between Y and $\hat{Y}$ i.e. Y = $\beta_0 + \beta_1 X + \in_i$

Then $\in_1 = Y - \beta_0 - \beta_1 X$

Squaring both sides,

$$e_i^2 = (Y - \beta_0 - \beta_1 X)^2$$

And summing over the n-observation we have,

$$\sum e_i^2 = (Y - \beta_0 - \beta_1 X)^2$$

Let $L = \sum e_i^2$ and minimizing L w.r.t. β_0 and β_1 and set each to zero.

$$\frac{\partial L}{\partial \beta_0} = \frac{\partial L}{\partial \beta_1} = 0$$
$$\frac{\partial L}{\partial \beta_1} = -2\sum (Y - \beta_0 - \beta_1 X)$$
$$= \sum (Y - \beta_0 - \beta_1 X) = 0$$
$$= \sum (Y - n\beta_0 - \sum_{i=1}^{n} \beta_1 X) = 0$$
$$\sum Y = n\beta_0 + \beta_1 \sum_{i=1}^{n} X) \qquad \ldots \qquad (1)$$

Also, differentially L w.r.t. β_1, we have,

$$\frac{\partial L}{\partial \beta_1} = -2\sum X_1 (Y - \beta_0 - \beta_1 X_1) = 0$$
$$= \sum X_1 (Y_1 - \beta_0 - \beta_1 X_1) = 0$$

Expanding,

$$\sum X_1 Y = \beta_0 \sum_{i=1}^{n} X_1 + \beta_1 \sum_{i=1}^{n} X_1^2$$

This implies that,

$$\sum X_1 Y = \beta_0 \sum_{i=1}^{n} X_1 + \beta_1 \sum_{i=1}^{n} X_1^2 \qquad \ldots \quad (2)$$

Equations (1) and (2) are called least square normal equations. Solving equations (1) and (2) simultaneously by substitution we have,

$$\beta_0 = \frac{\sum_{i=1}^{n} Y_i - \beta_i \sum_{i=1}^{n} X_i}{n} = \overline{Y} - \beta_i \overline{X}$$

Substituting β_0 in equation (2) we have

$$\sum X_i Y_i = \left(\frac{\sum_{i=1}^{n} Y_i - \beta_i \sum_{i=1}^{n} X_i}{n} \right) \sum X_i + \beta_i \sum_{i=1}^{n} X_i^2$$
$$n\sum X_i Y_i = \left(\sum_{i=1}^{n} Y - \beta \sum_{i}^{n} X_i \right) \sum X_i + \beta_i \sum_{i}^{n} X_{i1}^2$$
$$n\sum X_i Y_i = \left(\sum Y_i\right)\left(\sum X_i\right) - \beta_i \left(\sum X_i\right)^2 + n\beta_i \left(\sum X_i\right)^2$$
$$n\sum X_i Y_i - \left(\sum Y_i\right)\left(\sum X_i\right) = \beta_i \left(n\sum X_i^2\right) - \left(\sum X_i\right)^2$$
$$\hat{\beta}_i = \frac{n\sum X_i Y_i - \left(\sum Y_i\right)\left(\sum X_i\right)}{\left(n\sum X_i^2 - \left(\sum X_i\right)^2\right)}$$

The regression equation or the line of best fit is now

$$\hat{Y} = \beta_0 + \beta_1 X$$

Note that $\beta_1 = \dfrac{\sum (X - \overline{X})(Y - \overline{Y})}{\sum (X - \overline{X})^2}$

2.5 Prediction

Prediction can be said to be expectation about a given aspect of social behaviour that may be verified by subsequent observations. Within this general conception, the term is used in two principal senses: for deductions from known to unknown events within a conceptually static system and for statement about future outcomes based on recurring sequences of events. Given the sequence X_t which is zero mean stationary stochastic process, we want to predict the value of X_{t+m}, that is, m step ahead predictor and denote by $X_t(m)$ (Shangodoyin and Ojo 2002).

3. RESULTS AND DISCUSSION

The real series used is the payment and consumption lists of a telecommunication industry in Nigeria. Regression and ANN models are fitted using the series.

3.1Artificial Neural Networks (ANN) Learning Curves

Payment list of telecommunication industry in Nigeria was considered as our input and using the model - one hidden layer and three PEs - the learning curve was generated which showed the plot of the MSE across iterations.

The learning curves generated show the training MSEs for 5 runs.

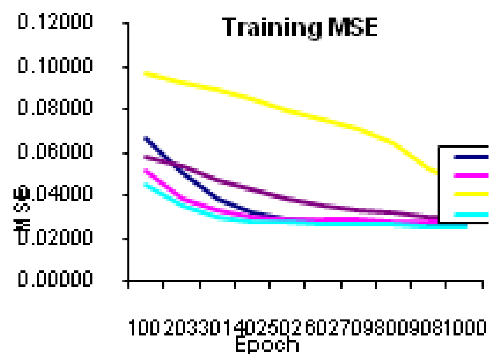

Figure 2: Training MSE for 5 Runs as a Function of Epoch

3.2 Prediction Equations for Regression Models: Series 1 to 7

$$\hat{Y} = 2.6E+08-1052975X \quad -0.395X_2 \tag{1}$$

$$\hat{Y} = 1.7E+07-252186X_1-0.109X_2 \tag{2}$$

$$\hat{Y} = 3.6E+08-442018X_1-0.205X_2 \tag{3}$$

$$\hat{Y} = 3.8E+07-671283X_1+3.489E-03X_2 \tag{4}$$

$$\hat{Y} = 2.7E+07-141358X_1-2.76E-02X_2 \tag{5}$$

$$\hat{Y} = 2.9E+07-60156.1X_1-2.14E.02X_2 \tag{6}$$

$$\hat{Y} = 2.4E+08+831100.0X_1-0.222X_2 \tag{7}$$

3.3 Prediction Performances of Regression and Artificial Neural Networks (ANN) Models

Table 1: Prediction Performances of Regression Models

S/N	Series 1	Series 2	Series 3	Series 4	Series 5	Series 6	Series 7
1	64000266.71	12820133.47	139338358.40	38241603.75	22806733.45	24070328.35	96216537.92
2	68261000.42	10762905.03	128425945.70	37639085.40	22619523.54	23458775.82	98415342.43
3	63803448.58	10618854.97	122052070.50	37003535.31	22025949.33	23768579.09	95358822.46
4	59345896.74	10474804.92	115678195.60	36367985.22	21432375.12	24078382.35	92302302.48
5	65493560.77	10773409.65	121591915.20	35682160.66	21163213.93	24029649.94	109945299.3
6	69623620.90	10994598.21	128041342.10	35046282.02	20916574.05	23971478.07	114699592.3
7	75410835.14	10866862.61	122364909.70	41894257.77	20885262.08	28261762.63	108909561.8
8	86343089.26	10396441.69	119169344.60	33717992.94	20707232.59	23830030.90	126465100.0
9	84924026.37	11393559.95	118288468.20	33058289.30	20536604.60	23727377.09	121009114.8
10	86412407.96	10912297.01	115360871.00	32404320.22	20664910.88	23776715.81	109877400.2
11	85624057.09	11334428.50	115262955.90	31717208.20	20427126.98	23596778.36	111907942.3
12	91683956.83	10433678.67	97146106.11	31170611.76	19597070.93	23070980.77	97797811.56
13	75194078.70	10221370.22	87536144.33	30549464.18	19248198.33	22781637.35	78395338.36
14	58830904.72	9479918.44	90916857.32	29911393.28	19273187.32	22687018.43	94558013.27
15	43446115.47	9673293.21	80845251.29	29268155.13	19120200.36	22665100.40	66193016.25
16	38232021.35	9758512.74	80713633.38	28573841.45	19182321.39	23342287.77	70731556.27
17	33017927.23	9843732.28	80582015.46	27879527.76	19244442.41	24019475.13	75270096.28
18	17777349.35	9783130.00	73396476.05	27260735.29	19160099.70	23930827.36	66564240.34
19	2922868.84	9136223.16	65752602.65	26630157.51	18698439.52	23616871.69	58052031.29
20	94082566.00	8869369.06	100389753.60	25780617.28	19205200.58	23942157.13	78845503.48
21	93770009.18	9336668.38	110691975.40	25073737.47	19254382.47	23921196.95	99402471.63
22	98706338.00	9141748.04	115489552.90	24434527.17	19131731.98	24043421.79	109899083.6
23	93228887.13	8909668.54	122004444.00	23778337.95	18918002.18	23953505.92	141690878.1
24	69495486.08	6361472.10	112097057.60	22977463.65	18445505.68	22867533.55	139762700.1
MSE	5130795	249447.1	4066876	1048972	256039.8	215749.4	4597734

Table 2: Prediction Performances of Neural Networks Models

S/N	Series 1	Series 2	Series 3	Series 4	Series 5	Series 6	Series 7
1	92627512.00	116920840.0	18413458.0	114416600.0	31271328.0	6625410.5	25266454.0
2	73450688.00	96658064.0	16162011.0	64530476.0	35329680.0	5841036.5	24572956.0
3	58597120.00	111162376.0	21662062.0	81173688.0	35408960.0	12975855.0	23989382.0
4	68956264.00	133909320.0	21875212.0	114570016.0	31802264.0	13929094.0	27512110.0
5	70684096.00	140757264.0	20919198.0	112765016.0	33236144.0	12663933.0	22407344.0
6	72926232.00	112658800.0	22023948.0	131352656.0	33289278.0	9729835.0	24404934.0
7	77139792.00	105567464.0	23430034.0	121069432.0	33304424.0	9264748.0	26216800.0
8	77474184.00	133486264.0	22292602.0	146081088.0	35080348.0	14480406.0	25504930.0
9	72911320.00	125376424.0	23565562.0	83724976.0	34427248.0	6110275.0	22861042.0
10	78238520.00	107528224.0	21936576.0	105088216.0	32893624.0	13882375.0	21530270.0
11	87508632.00	122894208.0	19928304.0	86913016.0	32292676.0	10109232.0	20923698.0
12	48929184.00	115671424.0	17787642.0	72983232.0	33252742.0	11424229.0	24779408.0
13	25982154.00	100756200.0	20652828.0	85049632.0	32524444.0	8294920.0	23284242.0
14	31899624.00	98615256.0	21553846.0	86878904.0	23488582.0	14381958.0	24440340.0
15	27892686.00	58078316.0	16589868.0	56840148.0	20648248.0	5704066.0	24103836.0
16	25165332.00	54128612.0	16327041.0	49198336.0	17999766.0	9559687.0	25520302.0
17	26220930.00	58337652.0	19886952.0	50099684.0	12676265.0	11877238.0	20554178.0
18	29080484.00	63351488.0	12785328.0	54715848.0	11254111.0	8077683.0	16379986.0
19	25376612.00	22930672.0	11421811.0	55648372.0	4875461.0	9213269.0	24873712.0
20	38098072.00	177545952.0	21393192.0	109916600.0	39816936.0	53871773.0	37294704.0
21	43723292.00	97976864.0	6931492.0	63615300.0	34594028.0	8285889.5	13733245.0
22	73142672.00	128396232.0	17314218.0	66541524.0	35123612.0	13333269.0	29777192.0
23	69052728.00	111271732.0	22238326.0	129284760.0	17948192.0	10236190.0	22661730.0
24	62405712.00	64994544.0	13936185.0	54053564.0	11478405.0	7733728.0	24103596.0
MSE	**0.016421**	**0.031374**	**0.028438**	**0.014321**	**0.025876**	**0.052776**	**0.034173**

From Tables 1 and 2 above, we could see the prediction of Artificial Neural Networks and regression models for series 1 to 7 resulting from regression equation model and training of runs as a function of epoch in artificial neural networks. Also, from the two tables, we could see the mean square error (mse) attached to each of the series. The mse attached to artificial neural networks was smaller when compared with the regression models. his implies that artificial neural networks model perform better than regression model when it comes to prediction.

4. CONCLUSION

The study shows the efficacy of artificial neural network in prediction. Having examined statistical and artificial neural networks models for prediction, the future values of payment were predicted using the appropriate models and the performance of the artificial neural networks compared with the regression. Artificial Neural Networks had the ability to analyze complex patterns quickly and with the higher degree of accuracy. Artificial Neural Networks had the lowest mean square error compared to statistical models.

REFERENCES

A.G. Vinsnes et al. (2001) Regression Analysis, Blackwell Science Ltd, Journal of Clinical Nursing, 10, 455-462, pg 462.

AlFuhaid, A. S., El-Sayed, M. A., & Mahmoud, M. S. (1997). Cascaded artificial neural networks for short-term load forecasting. *IEEE Transactions on Power Systems*, *12*(4), 1524–1529. doi:10.1109/59.627852

Cheng, B. and D.M. Titterington, (1994), "Neural Networks: A Review from a Statistical Perspective," Statistical Sciences, 9(1), 2-54

Feuston, B. P., & Thurtell, J. H. (1994). Generalized nonlinear regression with ensemble of neural nets: the great energy predictor shootout. *ASHRAE Transactions*, *100*(2), 1075–1080.

James M. Hutchinson; Andrew W. Lo; Tomaso Poggio (1994).A Nonparametric Approach to Pricing and Hedging Derivative Securities Via Learning Networks. The Journal of Finance, Vol. 49, No. 3, Papers and Proceedings Fifty-Fourth Annual Meeting of the American Finance Association, Boston, Massachusetts, January 3-5, 1994. (Jul., 1994), pp. 851-889.

Kelly E. Fish, James H. Barnes and Milam W. Aiken Assistant Artificial neural networks: A new methodology for industrial market segmentation Industrial Marketing Management Volume 24, Issue 5, October 1995, Pages 431-438

Kosta Metaxiotis (2010). Intelligent Information Systems and Knowledge Management for Energy: Applications for Decision Support, Usage, and Environmental Protection. *National Technical University of Athens, Greece. Information Science Reference, USA*

Kuan, Chung-Ming and Liu, T. (1995), Forecasting exchange rates using feedforward and Recurrent Neural Networks, Journal of Applied Econometrics, 10, pp 347 – 364.

Li,W.K. (1994). Time series models based on generalized linear models: some further results. *Biometrics*, **50**, 506-511.

McCulloch, W. S. and Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, 5:115-133.

Ostrom, C. W., Jr., (1990), Time Series Analysis, Regression Techniques, Second Edition: Quantitative Applications in the Social Sciences, v. 07-009: Newbury Park, Sage Publications.

Qeethara Kadhim Al-Shayea, Ghaleb A. El-Refae, Shurouq Fathi El-Itter and Al-Zaytoonah. (2010) Neural Networks in Bank Insolvency Prediction. IJCSNS, Vol.10 No.5, May 2010

Ray Eberts and Shidan Habibi (1995).Neural network-based agents for integrating information for production systems. International Journal of Production Economics.Volume 38, Issue 1, March 1995, Pages 73-84

P. L. Brockett, W. W. Cooper, L. L. Golden and U. Pitaktong (1994), A Neural Network Method for Obtaining an Early Warning of Insurer Insolvency, The Journal of Risk and Insurance, Vol. 61, No. 3, 1994, PP. 402-424.

Sarle, Warren S. (1994), "Neural Networks and Statistical Models," *Proceedings of the Nineteenth Annual SAS Users Group International Conference*, April, pp 1-13.

Shangodoyin, D. K and Ojo, J. F (2002). Elements of Time series Analysis: *Rashmed Publication*. (ISBN 978-8024-32-7)

Tsuyoshi Satake, Katsumi Morikawa and Nobuto Nakamura(1994). Neural network approach for minimizing the makespan of the general job-shop. International Journal of Production Economics.Volume 33, Issues 1-3, January 1994, Pages 67-74

Y. Shachmurove and D. Witkowska (2000) Utilizing Artificial Neural Network Model to Predict Stock Markets, University of Pennsylvania, technical University of Lodz.

BIBLIOGRAPHY

[1] Alexandre L. Gonçalves, Fabiano Beppler, Alessandro Bovo, Vinícius Kern, Roberto Pacheco (2006). A Text Mining Approach towards Knowledge Management Applications. ww.eurocris.org/Uploads/.../**2006**11%20...IR%20**2006**/Proceedings.pdf

[2] Allen, J. 1995. Natural Language Understanding

[3] Ananiadou, S., Chruszcz, L., Keane, J., Mcnaught, J., and Watry, P. 2005. The national centre for text mining: aims and objectives. In Ariadne 42, Jan. 2005.

[4] Rebecca, Hwa, (2002). An Overview of Text Mining www

[5] EBSCO 2009. EBSCO services overview, accessed from hhttp://web.ebscohost.com/ehost/ (27 October 2009)

[6] E. Riloff and R. Jones, (1999) "Learning Dictionaries for Information Extraction Using Multi-level Boot-strapping," in *the Proceedings of AAAI-99*.

[7] Frantzi, K., Ananiadou, S. and Mima, H. 2000 Automatic recognition for multi-word terms. International Journal of Digital Libraries 3(2), 117-132.

[8] Gartner Group, Text mining White Paper , June 2000 http://www.xanalys.com/intelligence_tools/products/text_mining_text.html

[9] Good, I. J. 1965. The Estimation of Probabilities: An Essay on Modern Bayesian Methods. MIT Press, Cambridge, MA, 1965.

[10] Hirschman, L., Park, J.C., Tsujii, J., Wong, L., and Wu, C. 2002. Accomplishments and challenges in literature data mining for biology. In Bioinformatics, 18(12), pp. 1553-1561.

[11] Khomra, T. 2005. Ai: Artificial intelligence. it4307 / 5307, 2005.

[12] M. Grobelnik, D. Mladenic, and N. Milic-Frayling, (2000) "Text Mining as Integration of Several Related Research Areas: Report on KDD'2000 Workshop on Text Mining," 2000.

[13] Manning, C.D. and Schütze, H. 2001. Foundations of Statistical Natural Language Processing. MIT Press, Cambridge, MA.

[14] McCallum, A. 2005. Information Extraction: Distilling Structured Data from Unstructured Text. ACM Queue, 3(9), November 2005

[15] Nadeau, D. 2007. Semi-Supervised Named Entity Recognition. A PhD Thesis Submitted to Ottawa-Carleton Institute for Computer Science, School of Information Technology and Engineering, University of Ottawa, Canada.

[16] Porter, M. 1980. An Algorithm for suffix stripping. *Program*, pages 130-137. Retrieved on September 8, 2011.

[17] Salton, G., Allan, J. and Buckley, C. 1994. Automatic Structuring and Retrieval of Large Text Files. *Communications of the ACM,* 37(2):97-108, Feb 1994.

[18] Text mining White Paper http://textmining.krdl.org.sg/whiteppr.html