

---

# Student's Academic Performance Prediction Using Factor Analysis Based Neural Network

Shamsuddeen Suleiman<sup>1</sup>, Ahmad Lawal<sup>1</sup>, Umar Usman<sup>1</sup>, Shehu Usman Gulumbe<sup>1</sup>, Aminu Bui Muhammad<sup>2</sup>

<sup>1</sup>Department of Mathematics, Statistics Unit, Usmanu Danfodiyo University, Sokoto, Nigeria

<sup>2</sup>Department of Mathematics, Computer Science Unit, Usmanu Danfodiyo University, Sokoto, Nigeria

## Email address:

macostamkd@gmail.com (S. Suleiman), suleman.shamsuddeen@udusok.edu.ng (S. Suleiman), madawaahmad@gmail.com (A. Lawal), uusman07@gmail.com (U. Usman), ugulumbe@gmail.com (S. U. Gulumbe), aminu.muhd@yahoo.com (A. B. Muhammad)

## To cite this article:

Shamsuddeen Suleiman, Ahmad Lawal, Umar Usman, Shehu Usman Gulumbe, Aminu Bui Muhammad. Student's Academic Performance Prediction Using Factor Analysis Based Neural Network. *International Journal of Data Science and Analysis*. Vol. 5, No. 4, 2019, pp. 61-66. doi: 10.11648/j.ijdsa.20190504.12

**Received:** July 6, 2019; **Accepted:** July 26, 2019; **Published:** August 26, 2019

---

**Abstract:** This study focused on the statistical technique using the neural network, hybrid models and factor analysis on constructing the new factors affecting students learning styles of the survey done among university students in predicting academic performance. The data were collected using survey questionnaires and students' academic records. The methodologies used were descriptive statistics, factor analysis, neural network and hybrid models technique using the following Learning algorithms; Levenberg-Marquardt (LM), Bayesian Regularization (BR), BFGS Quasi-Newton (BFG), Scaled Conjugate Gradient (SCG), Gradient Descent (GD) in artificial neural network model while for the second Hybrid model only the best two algorithms were used; Levenberg-Marquardt (LM), Bayesian Regularization (BR). The results showed ten new factors were successfully constructed using factor analysis and the proposed hybrid models show that though it took longer time and number of epochs to train the hybrid models by Bayesian Regularization Algorithms, and it gives more accurate predictions than both the Levenberg-Marquardt, Scaled Conjugate Gradient, Gradient Descent and BFGS Quasi-Newton (BFG) Algorithms. In a nutshell, the finding indicates that Bayesian Regularization is the best learning algorithms in both Neural Network and Hybrid models for predicting students' academic performance.

**Keywords:** Neural Network, Hybrid, Factor Analysis, Prediction, Learning Algorithms

---

## 1. Introduction

The applications of Machine Learning methods to predict students' performance based on student's background and term examination performances has turned to be helpful for foreseeing the different performance in various levels. Using such machine learning methods enables to timely predict the students who have a high chance of failing so that a remedy can be provided by a teacher to the student. It can even help to detect high caliber students of the institution and help them providing scholarship [1].

Universities and higher institutions of learning have been known as a breeding and grooming field for students' academic pursuit. That is why the path of the students' academic performances are necessary to be checked, in order

to help sustain the important roles students play in the society upon graduation, which may be either to continue their studies into the post-graduate program or become the main power for the industry, government and private sectors. Thus, the students' academic performances are critical in ensuring that those significant roles the students play in the society are maintained. This has motivated some higher institutions of learning to develop interest in predicting the paths of students, thus identifying which students will require assistance in order to graduate at the stipulated time or maintain their studies or even drop out of the school. This is brought about by the academic failure rate among students, which has led to a large number of debates [2].

The factor analysis describes the covariance relationship among many variables in terms of few underlying, but unobservable, random quantities called factors. Factors

analysis believes that the variables can be grouped by their correlations. It may be assumed that variables within a particular group are highly correlated among themselves, but they have relatively small correlations with variables in a different group. Then it can be said that each group of variables represents a single underlying construct (or factor) that is responsible for the observed correlations [3].

ANNs are powerful tools for modeling. They can learn and identify correlated patterns between input datasets and corresponding target values. Once trained, ANNs can be used to predict the outcome of new independent input data [4]. This feature is employed in various areas, such as computing, medicine, engineering, and many others. Neural networks have a remarkable ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques [5]. A neural net resembles a brain in two respects: the learning process is used by the network to acquire knowledge, and the synaptic weights (inter-neuron connection strengths) are used to store the knowledge [6].

Dimensionality reduction is an important part of a pattern recognition system. It is a process in which we present a system having many degrees of freedom by a smaller number of degrees of freedom.

Factor analysis, principal component analysis (PCA) and linear discriminant analysis are some of the techniques, which can be used for dimension reduction. Factor analysis can be considered as an extension of PCA and its approximation is more elaborate than that of PCA [3]. PCA and Neural Networks were also combined in the studies [7-9] to predict students' performance related variables. However, none of these researches seeks to integrate factor analysis in to Artificial Neural Network to predict student performance which the current work is addressing.

The research [7] used two nonparametric methods to model entrepreneurial intentions: principal component analysis (PCA) and artificial neural networks (ANNs). PCA was used to perform feature extraction in the first stage of modelling, while artificial neural networks were used to classify students according to their entrepreneurial intentions in the second stage. Four modelling strategies were tested in order to find the most efficient model. Dataset was collected in an international survey on entrepreneurship self-efficacy and identity. Variables describe students' demographics, education, attitudes, social and cultural norms, self-efficacy and other characteristics.

The study [8] proposed a new classification method based on Principal Component Analysis (PCA) and Learning Vector Quantization (LVQ) neural network. PCA was first used to reduce the dimensionality of original data about students' learning strategies. 5 principal components were extracted to create a PCA-LVQ classification model. The classification result of the proposed model was compared with those produced by a simple LVQ network model and a standard BP network model. The simulation results show that compared with the other two networks, the PCA-LVQ model

has a better performance in training speed, classification accuracy and generalization ability.

The researcher [10] programmed a model based on the multilayer perception algorithm. The result from the test data evaluation showed that the programmed Artificial Neural Network model was able to correctly predict and classify the performance of students with Mean Correct Classification Rate (CCR) of 97.07%.

The researcher [11] used student data stored in a Moodle server and predicted student success in course, based on four learning activities - communication via emails, collaborative content creation with wiki, content interaction measured by files viewed and self-evaluation through online quizzes. Next, a model based on the Multi-Layer Perceptron Neural Network was trained to predict student performance on a blended learning course environment. The model predicted the performance of students with correct classification rate (CCR) of 98.3%.

The research [9] surveyed a large number of students across different districts in Bangladesh. Pre-processing was done thoroughly; the use of data balancing, dimensionality reduction, discretization and normalization in combinations has allowed them to derive the best model that could predict the academic performance based on different factors during the adolescence.

The paper [1] proposed a deep learning neural network (DNN) to predict students' performance category. The result of the proposed neural network was compared with other machine learning algorithms which show that the proposed DNN model outperforms the other machine learning algorithms and its accuracy stood at 84.3%.

The aim of this paper is to predict student's academic performance using Factor Analysis based Artificial Neural Network.

The data used in this paper are secondary data collected from the student's record in Department of Mathematics Statistics Unit of Usmanu Danfodiyo University, Sokoto and the questionnaire adopted is designed by [12].

## 2. Method

### 2.1. Factor Analysis

The factor analysis describes the covariance relationship among many variable in terms of few underlying, but unobservable, random quantities called factors. Factors analysis believes that the variables can be grouped by their correlations. It may be assumed that variables within a particular group are highly correlated among themselves, but they have relatively small correlations with variables in a different group. Then it can be said that each group of variables represents a single underlying construct (or factor) that is responsible for the observed correlations [3].

This factor model postulates that  $X$  is linearly dependent upon common factors and errors, and the factor analysis model can be described as [3].

$$(X - \mu)_{p \times 1} = L_{(p \times m)} \times F_{(m \times 1)} + \mathcal{E}_{(p \times 1)} \quad (1)$$

$X$  : Be an observation random vector with  $p$  components  
 $\mu$  : Mean of  $X$   
 $\Sigma$  : Covariance of  $X$   
 $\Sigma$  : Unobservable random variables called common factors

$\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_p$  : Addition source of variations called errors on specific factors

The coefficient  $l_{ij}$  is called the loading of  $i^{th}$  variable on  $j^{th}$  factor, so that the matrix  $L$  is the matrix of factor loadings. The  $p$  deviation  $X_1 - \mu_1, X_2 - \mu_2, X_3 - \mu_3, \dots, X_p - \mu_p$  are expressed in terms of  $p + m$  random variable  $F_1, F_2, F_3, \dots, F_m, \epsilon_1, \epsilon_2, \epsilon_3, \dots, \epsilon_p$  which are *unobservable* (John and Wichern, 1998).

Bartlett’s test of sphericity and the Kaiser-Meyer-Olkin measure of sampling adequacy are both tests that can be used to determine the factoriability of the matrix as a whole. The results value of Bartlett’s test of sphericity is significant

( $p < 0.001, p = 0.000$ ) [3].

**2.2. Neural Network Model**

Artificial Neural Network (ANN) model proposed by [13] was used in this research work. The model is given below with consideration of logistic transfer function.

$$y = f(X.W) + e_i = \alpha X + \sum_{h=1}^H \beta h g(\sum_{i=0}^I \gamma h_i x_i) + e_i \quad (2)$$

Where  $g(.) = \frac{1}{1+e^{-x}}$

$Y$  is the output variable

$X$  is the input variable

$\alpha$  is the weight of the input unit (s)

$\beta$  is the weight of the hidden unit (s)

$\gamma$  is the weight of the output unit (s)

$g(.)$  is the logistic transfer function

$e_i$  is the error term

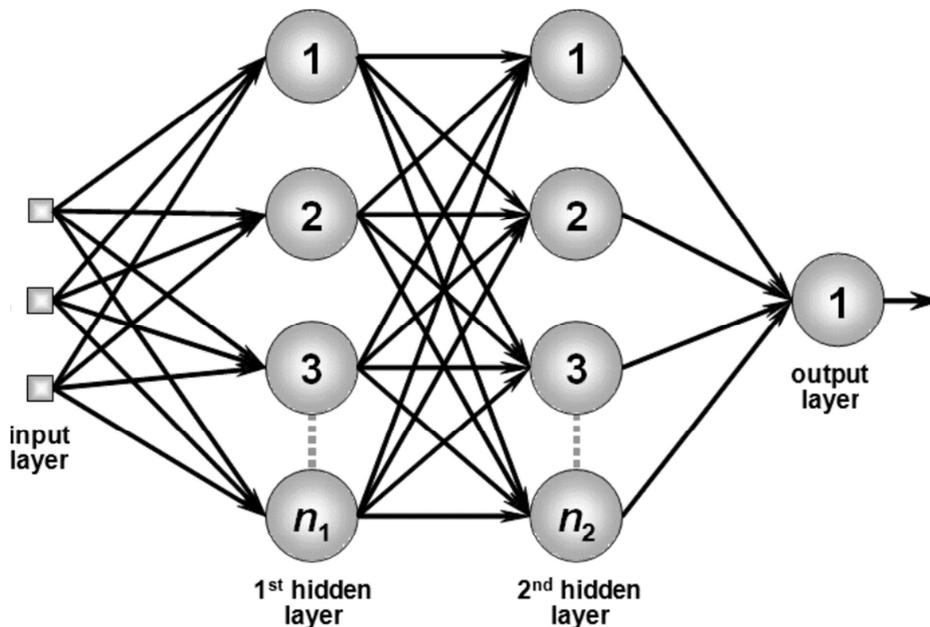


Figure 1. Architecture diagram.

**2.3. The Proposed Hybrid Model**

In this study, a hybrid models were proposed for Student’s Academic Performance Prediction Using Factor Analysis Based Neural Network, When we manipulate high dimensional data with neural network, many characteristic variables provide enough information, but too many network inputs go against designing of the hidden-layer of the network and take up plenty of storage space as well as computing time, and in the process interfere the convergence of the training network, even influence the accuracy of recognition finally. Factor Analysis (FA) concentrates the information that is carried by numerous original indexes which form the index system, and then stores it to the factor,

and can according to the precision that the actual problem needs, through controlling the number of the factors, to adjust the amount of the information. In this research we make full use of the advantages of FA and the properties of neural network structures to establish FA-ANN algorithm. That is Student’s Academic Performance and the learning styles: visual, tactile, auditory, kinesthetic, group, and individual. The new algorithm reduces dimensions by FA, and carry on network training and simulation with low dimensional data that we get, which obviously simplifies the network structure, and in the process, improves the training speed and generalization capacity of the neural network.

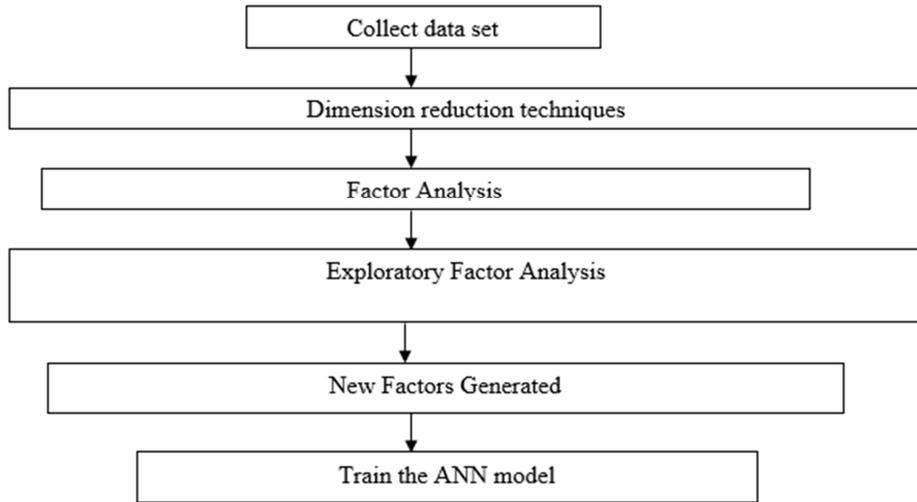


Figure 2. Flowchart of modelling process for hybrid model.

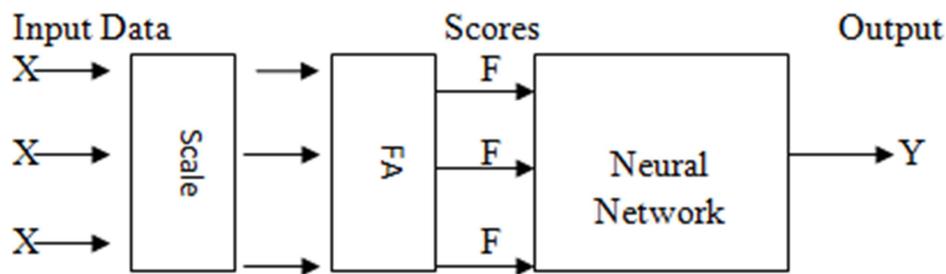


Figure 3. Schematic Diagram of the Hybrid model.

**2.3.1. Levenberg-Marquardt Algorithm (LM)**

The Levenberg-Marquardt algorithm, which is an approximation to the Newton’s method is said to be more efficient in comparison to other methods for convergence of the Backpropagation algorithm for training a moderate-sized feed forward neural network [14]. As the cost function is a sum of squares of nonlinear function, the Hessian matrix required for updating the weights and biases need not be calculated and can be approximated as

$$H = J^T(x)J(x) \tag{3}$$

The updated weights and biases are given by

$$x_{k+1} = x_k - [J^T(x)J(x) + \mu I]^{-1}J^T(x)e(x) \tag{4}$$

where  $\mu$  is a scalar and I is the identity matrix.

**2.3.2. Bayesian Regularization (BR)**

Regularization as a mean of improving network generalization is used within the Levenberg-Marquardt algorithm. Regularization involves modification in the performance function. The performance function for this is the sum of the squares of the errors and it is modified to include a term that consists of the sum of squares of the network weights and biases. The modified performance function is given by

$$F_{reg} = \beta SSE + \alpha SSW \tag{5}$$

where SSE and SSW are given by

$$SSE = \sum_{q=1}^N e_q^2(x) \tag{6}$$

$$SSW = \sum_{j=1}^n w_j^2 \tag{7}$$

Where  $n$  is the total number of weights and biases,  $w_j$  in the network. The performance index in (5) forces the weights and biases to be small, which produces a smoother network response and avoids over fitting. The values  $\alpha$  and  $\beta$  are determined using Bayesian regularization in an automated manner [15, 16].

**3. Results**

Table 1. Total Variance explained.

Factor	Total	% of Variance	Cumulative %
1	2.386	7.953	7.953
2	2.149	7.165	15.118
3	1.807	6.024	21.142
4	1.738	5.794	26.936
5	1.37	4.568	31.504
6	1.354	4.512	36.016
7	1.12	3.734	39.75
8	1.032	3.441	43.191
9	0.989	3.297	46.488
10	0.973	3.242	49.73

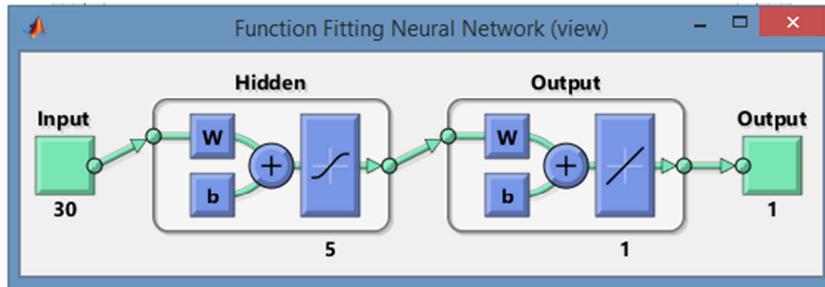


Figure 4. Predict Neural Network Model Architecture for Academic Performance.

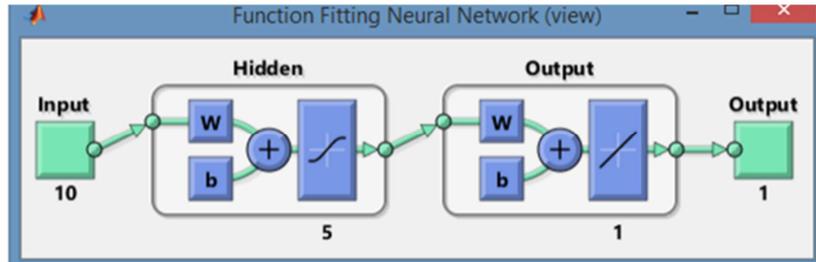


Figure 5. Predict Hybrid Model Architecture for Academic Performance.

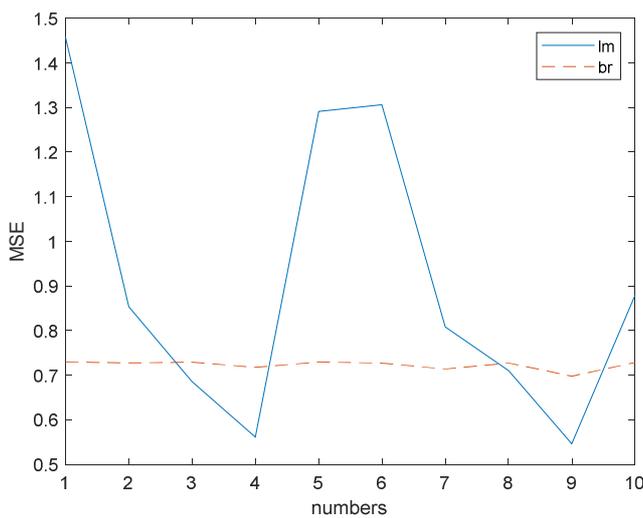


Figure 6. Neural Network Models performance to Academic Performance.

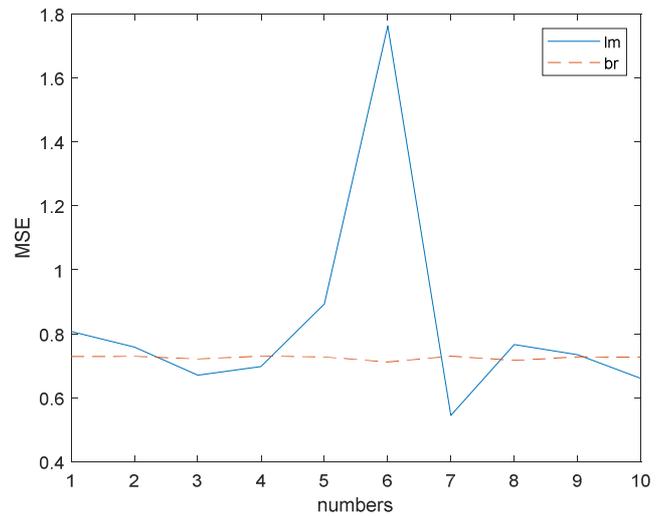


Figure 7. Hybrid Models performance to Academic Performance.

Table 2. Summary statistics of the Neural Network and Hybrid models’ performances.

Model	Training algorithm	Minimum	maximum	mean	Standard deviation
NN model	LM	0.5466	1.4583	0.9103	0.3268
	BR	0.6981	0.7302	0.7231	0.01035
Hybrid model	LM	0.5446	1.7629	0.8294	0.3410
	BR	0.7116	0.7302	0.7249	0.006321

### 4. Discussion

Table 1 displays the total variance explained at ten stages for factors that affecting students’ learning style among the undergraduate students in Mathematics Department. Ten factors were extracted because their eigenvalues greater are than 1. Using these factors can explain up to 50 percent of

the variability in the data.

The result of the training neural network, Mean Squared Error (MSE) represents the average squared difference between outputs and targets. Lower values of MSE are better and zero means no error.

In order to train, validate and test the neural networks developed using the LM, BFG, GD and the SCG algorithms, we have divided the data set in the following way: 70% of it

for the training process, 15% for the validation process and the remaining 15% for the testing process. In order to train and test the neural networks developed using the BR algorithm, we have divided the data set in the following way: 70% of it for the training process, 15% for the testing process and the remaining 15% was not allocated (in order to obtain a relevant comparison of the final results achieved by the algorithms). In all the cases, the samples have been randomly chosen as to cover the specified percentages. In order to train the neural networks, we have used the mean square error (MSE) as an objective function. When training a network with this function, if there are multiple outputs having different ranges of values, the accuracy is optimized for the output element that has a wider range of values and is less optimized relative to the output element with a smaller range of values.

Table 2 summarized the results of training the proposed networks using the two training algorithms discussed above. Each entry in the table represents 10 different trials, with random weights taken for each trial to rule out the weight sensitivity of the performance of the different training algorithms.

The average time required for training the hybrid models using the Levenberg-Marquardt algorithm was generally the least, whereas, maximum time is required for training the network using Scaled Conjugate Gradient algorithm for Neural Network Model. The training algorithm employing Bayesian Regularization continuously modifies its performance function and hence, takes more time on average to train compared to the Levenberg-Marquardt.

It can be seen from Figure 6 and Figure 7 that for the hybrid model, Levenberg-Marquardt algorithm has high average mean squared error for predicting Academic performance and Bayesian Regularization outperforms LM in consistency. Similarly, it took on average longest epoch (iterations) to train Neural Network model using BR. For the hybrid model, where the network size becomes smaller and thus more complex to train in terms of the network parameters to optimize, Bayesian Regularization turned the best results in all the two models but with the longest average number of iterations.

## 5. Conclusion

The results of the proposed hybrid models show that though it took longer time and number of epochs to train the hybrid model by Bayesian Regularization Algorithms, and it gives more accurate predictions than the Levenberg-Marquardt. In a nutshell, the finding indicates that Bayesian Regularization is the best learning algorithms in both Neural Network and Hybrid models for predicting students' academic performance. Based on the conclusion of this paper, it is recommended that the neural network algorithm can be used to predict student performance since it has the ability to model it very well.

## References

- [1] Bendangnuksung and Prabu P (2018) *Students Performance Prediction using Deep Neural Net* International Journal of Applied Engineering Research (13) 2 pp. 1171-1176.
- [2] Bresfelean, N. and Ghisoiu, N. (2005) Determining Students' Academic Failure Profile Founded on Data Mining Methods," *Proceedings of the ITI 2005 30th Int. Conf. on Information Interfaces*, 2005, Cavtat, Croatia, June 23rd - 26th, 2008, pp. 317-322.
- [3] Johnson R. A. and Wichern D. W., (1998), *Applied Multivariate Statistical Analysis. Fifth Edition*. Prentice-Hall, Inc, Upple Saddle River.
- [4] Zou, J., Han, Y., & So, S. S. (2009). *Overview of artificial neural networks*. In *Artificial Neural Networks* (pp. 14-22). Humana Press.
- [5] Shebany, M. et al. (2014). *Artificial neural network: a brief overview*. In *International Journal of Engineering Research and Applications*, Volume 4 (Issue 2), Version 1, pp. 07-12.
- [6] Hajek, Milan (2005). *Neural networks*, University of KwaZulu-Natal.
- [7] MarijanaZekić-Sušac, NatašaŠarlija and Sanja Pfeifer (2013) Combining PCA Analysis And Artificial Neural Networks in Modelling Entrepreneurial Intentions of Students *Croatian Operational Research Review (CRORR)*, Vol. 4, pp. 306-317.
- [8] Hu S, Yan G. and Jiang H (2015) Study of Classification Model for College Students' M-Learning Strategies Based on PCA-LVQ Neural Network *8th International Conference on BioMedical Engineering and Informatics (BMEI 2015)* pp. 742-746.
- [9] Ahamed A. T. M. S., Tanzeem N. M. and Rahman R. M (2017), An intelligent system to predict academic performance based on different factors during adolescence, *Journal of Information and Telecommunication*, 1: 2, 155-175.
- [10] Asogwa O. C. and Oladugba A. V., (2015) "Of Students Academic Performance Rates Using Artificial Neural Networks (ANNs)." *American Journal of Applied Mathematics and Statistics*, 3 (4), 151-155. doi: 10.12691/ajams-3-4-3.
- [11] Zacharis N. Z. (2016), predicting student academic performance in blended learning using artificial neural networks, *International Journal of Artificial Intelligence and Application (IJAIA)*, Vol. 7, No. 5, September 2016 pp 17-29.
- [12] Reid, J. (1984). *Perceptual Learning Style Preference Questionnaire*. Retrieved October 28, 2010 from <http://lookingahead.heinle.com/filing/l-styles.htm>.
- [13] Anders, U. (1996) Model selection in neural networks, *ZEW Discussion Papers 96-21*. Retrieved from <http://hdl.handle.net/10419/29449>.
- [14] Hagan, M. T., & Menhaj, M., (1994) "Training feed-forward networks with the Marquardt algorithm", *IEEE Trans. Neural Networks*, Vol. 5, No. 6, pp 989-993.
- [15] Foresee, F. D. & Hagan, M. T., (1997) "Gauss-Newton approximation to Bayesian regularization", *International Joint Conference on Neural Networks*.
- [16] Mackay, D. J. C., (1992) "Bayesian interpolation", *Neural Computation*, Vol. 4, No. 3, pp 415-447.