

Efficacy of Artificial Neural Network based Decision Support System for Career Counseling

Jasjit S. Sodhi^{1*}, Maitreyee Dutta² and Naveen Aggarwal³

¹Department of Biotechnology, Himachal Pradesh University, Shimla - 171005, Himachal Pradesh, India; jsodhi_2000@yahoo.com

²Department of Computer Science, N.I.T.T.R, Chandigarh - 160019, Punjab, India; d_maitreyee@yahoo.co.in

³Department of Computer Science, U.I.E.T, Punjab University, Chandigarh - 160014, Punjab, India; navagg@pu.ac.in

Abstract

Objectives: This paper presents the use of machine learning technique in order to eliminate the components affecting the human decision making process. To assist decision making in career counseling, an Artificial Intelligence model was implemented using Artificial Neural Network (ANN) in MATLAB for predicting vocational stream of pursuit based on the behavioral characteristic of the beneficiary. **Methods/Statistical Analysis:** The Differential Aptitude Test (DAT) battery; and Scientific Knowledge and Aptitude Test (SKAT) were used to assess an individual's specific abilities in different areas. The training data set for the ANN model was procured in the form of normalized scores based on occupational/vocational profiles as given by the authors of DAT battery and SKAT. The trained ANN was tested with normalized stanine scores of 100 tenth class students raised through cluster random sampling technique for predicting a vocational stream of pursuit. In order to evaluate its accuracy and efficiency, three techniques of classification were employed. The unclassified data was classified by using Discriminant function analysis, ANN and two classifications were obtained from the trained counselors. **Findings:** The classification result obtained from the above mentioned techniques were compared and was found that the ANN system and Discriminant function analysis agreed approximately by 91% over all the test cases. The results of the statistical method support the classification made by the ANN. The two counsellors were in agreement with ANN's classification output by approximately 81%. However, the counsellors disagreed with each other's prediction approximately by 27% over all the test cases. The experimental results support the hypothesis that the proposed machine learning technique performed better than the prediction made by counsellors. **Application/Improvements:** In the Indian scenario, the developed machine learning system may be used as a standalone system in places where there is the paucity of counsellors or can assist in the career decision making provided by human beings.

Keywords: Artificial Neural Network, Career Counseling, Decision Support System, Differential Aptitude Test Battery, Discriminant Function Analysis, Scientific Aptitude and Knowledge Test

1. Introduction

The role of decision making with the help of Artificial Intelligence in one form or the other has been successful in a number of cases for mainly unstructured and semi-structured problems. The methods used in solving problems using Artificial Intelligence are logical programming, application of fuzzy logic, neural networks, hybrid intelligent systems (like neuro-fuzzy expert system) and genetic algorithms. The choice of method

mainly depends on the type of the problem and whether the explicit data, information or knowledge is available in that domain. In the complex real world scenario the decision maker many times encounter problems for which no explicit data, information or knowledge is available. The decision maker then may take decisions based on heuristics or intuition that he had gained by experience in solving similar problems for which the exact replicas are not available. For these types of problems the machine's with a cognitive ability such as artificial neural networks

*Author for correspondence

are used. One of the traits of a human being is its ability to recognize patterns. Neural Networks possess the capability to mimic human capability of pattern recognition; once they learn the characteristics of the patterns to be found. The pattern recognition ability of the neural network is used in solving many complex business applications such as to predict the ratings of corporate bonds¹, bankruptcy prediction², credit card fraud detection³⁻⁵, predicting business failures⁶, classification of bone density⁷ and also in share/stock market or financial time series prediction⁸⁻¹⁰. In¹¹ described the design structure for an ANN based decision support system along with a fuzzy user interface for the selection of suitable course/curriculum. In¹² further improvised the ANN system developed by her using type-2 fuzzy sets. In the area of career counseling, In¹³ concluded that most expert systems for career selection that have been designed are specific to the curriculum of a particular University and are used for the selecting course/major of that university. There are several artificial intelligence techniques for creating such an expert system, but Artificial Neural Networks are superior as they are adaptive in nature. The Artificial Neural Networks are very flexible, greatly reliable and very accurate in making predictions. The flexibility of the design of ANN helps the user to add input parameters at a later stage and train the neural network again to respond to the new input parameters. The present study is organized as follows: Section 2 describes the Research Methodology used for the study. Section 3 describes results and their interpretation and Section 4 concludes the study with the scope of future work. Career guidance is a domain of trained counselors helping the student in choosing an appropriate vocational stream based on the beneficiary's aptitude, interest and the personality suitable for his/her work. The job of the counselor is on an individual basis and rather cannot be generalized for a group. The counselor based on the results of psychometric tests conducted for the beneficiary suggests the vocational streams he can pursue. However, the guidance and counseling process is flawed by several errors such as the lack of trained counselors in the institutions, counselors judgment might be fatigued by the high ratio between counselor and counselee's, lower productivity due to laborious process and biases due to external factors like interest of beneficiary, parent's expectations, lack of experience etc. In the context of India, the concept of guidance and counseling cell is still in the initial stages and the availability of trained manpower is also scarce. The most of the tools available in the market as analyzed¹³

are specific to the institute/university for which they have been developed and the manual process has an element of biases in human judgment. Hence, there is a strong motivation to develop an intelligent computer-based decision support system for helping the beneficiary in the selection of the stream of pursuit. The aim of the present study was to develop a decision support system using an artificial neural network for choosing the vocational stream of pursuit (science, administrative and secretarial, engineering and humanities) based on the beneficiary's aptitude measured by Differential Aptitude Test Battery (DAT) and Scientific Knowledge and Aptitude Test (SKAT). The results of developed model were further verified and evaluated against human counselors and a statistical tool i.e., discriminant analysis method.

2. Research Methodology

The ANN is a tool used for complex decision making, the ability of the neural network to learn from past performance and improve itself was the general idea behind its choice. The use of several statistical forecasting methods such as regression, discriminant function analysis etc., is also chosen for forecasting where the input factors are quantifiable. In¹⁴ asserted that the development and implementation of ANN for the selection of vocational stream consists of following phases:

- Phase 1: Collection of data to be used for the training of ANN.
- Phase 2: Transforming the data collected into neural network inputs and normalization of data.
- Phase 3: Construction of training set data.
- Phase 4: Select, Train, and Test Network.
- Phase 5: Deploy developed network application.
- Phase 6: Analysis and interpretation of results.

In this study, the target population was the students studying in 10th class in different schools in Chandigarh, India.

2.1 Sampling

A sample of 31 students was raised out of the target population mentioned above using cluster sampling technique. The data of 38 students was taken from a counseling cell. Further, to test the efficiency of the developed system 31 cases including boundary conditions were used as testing cases. In total 100 cases were used to test the developed system.

2.2 Psychological Tests

In the first phase the data collected were the scores based on DAT battery¹⁵ and SKAT. The author validate the¹⁶ as a measure of the scientific aptitude of the beneficiary. It consists of 72 multiple choice items. These items cover all branches of science and the test was of 60 minute duration. A translated and adapted version of DAT Form A used in this study was published by Manasyan, 32 Netaji Subash Marg, New Delhi. All of the DAT tests were multiple-choice tests with specified time limits. Each test can be administered individually or as a battery of tests. In the present study, the DAT and SKAT were used as criterion variables for Artificial Neural Network decision support system. The following were selected as the list of predict or variables for decision support in the selection of vocational stream of pursuit.

- Mechanical Reasoning Test Score (MR).
- Scholastic Aptitude Score (Verbal Reasoning (VR)+Numerical Ability (NR)).
- Abstract Reasoning Score (AR).
- Space Relations Score (Sp. R).
- Speed and Accuracy Score (S&A).
- Language Usage Score (LU).
- Scientific Knowledge and Aptitude Test Score (SKAT).

During the second phase of the study the data collected through Phase 1 above was considered as raw data and was transformed into normalized standard nine scores (stanine). The raw data of the candidate was converted into its percentile score in the second phase using the norm table provided by the author of DAT¹⁵ for 10th class students and then this percentile score was converted to stanine score using descriptive statistics such as mean and standard deviation. In the third phase of the study, the training set comprised of 3 subsets; the training dataset, the validation set, and testing set. The training set was used to improve the generalization of the ANN. The training dataset was constructed from the norms prepared by the author for 10th class students. In this study, the data set consists of 111 test cases having different values of input parameters and an expected target value namely: Engineering, Science, Clerical and Administration and Humanities from the available vocational streams of pursuit. About 68% of the total data (i.e., 75 cases) were used as the training set, 19% (i.e., 21 cases) as the validation set and 13% (i.e., 15 cases) were used as test set.

2.3 ANN Architecture

The fourth phase of the development of ANN comprised of choosing an effective neural network model and the learning algorithm. A multilayer feed forward network using backpropagation is almost a universally accepted standard paradigm for modeling, forecasting and classifying when the outputs are known. In the present study, ANN was implemented using multilayer feed-forward network with a variation backpropagation algorithm. Several key parameters of this ANN such as a number of hidden units, the transfer function, the learning rate, initial weights, and others were selected during this phase.

2.3.1 The Number of Input Nodes

The number of input nodes was problem dependent and in our case since there were 7 different input variables, the input layer had 7 nodes depicting each variable of the input vector.

2.3.2 The Number of Hidden Layers and Nodes

The number of hidden layers and the nodes in them were considered as critical parameters for the performance of the ANN. Several theoretical works in the literature^{17,18}. Suggested that to approximate any complex nonlinear function a single hidden layer was sufficient.

In¹⁹ suggested that using lesser number of neurons in the hidden layer would increase the training time of the neural network. Researcher's suggested that using more than one hidden layer may provide benefits in some complex problems. In²⁰⁻²² report ed using more than one hidden layer in their network design processes. This study also used two hidden layers in the ANN architecture. The numbers of hidden layer neurons are chosen based on heuristics using forward selection or backward selection technique. In this study, the forward selection approach was used and the best result was recorded with 7 nodes in the first hidden layer and 9 neurons/nodes in the second hidden layer Table 1.

2.3.3 Activation Function

The activation function is as important as the ANN architecture or the learning algorithm. It determine the output of a node in a network. In²³ suggested using logistic activation functions for classification problems and to use the hyperbolic tangent functions for forecasting problem.

Table 1. Optimal number of Neurons in the hidden layers

Number of neurons in the first hidden layer	Number of neurons in the second hidden layer	Mean Forecast error for 100 data points
5	5	14.70
6	5	10.66
7	5	10.97
8	5	11.07
9	5	12.75
6	6	11.93
6	7	12.34
6	9	12.52
7	7	11.43
7	8	9.18
7	9	5.13
7	10	7.23
7	11	7.53
7	12	10.08

The study was conducted using logistic transfer function in the hidden layer and the linear transfer function in the nodes of the output layer.

2.3.4 Training Algorithm

The aim of the training a neural network is to improve its ability to recognise similar inputs which were not used for its training. The back-propagation algorithm is a popular supervised learning algorithm used for solving classification problems.

The backpropagation algorithm suffers from slow convergence, inefficiency and lack of robustness and it is sensitive to the choice of learning rate and momentum parameter. A number of variations or modifications of backpropagation, such as the adaptive method quick-prop²⁴, and second-order methods²⁵⁻²⁷ etc., have been proposed by researchers to overcome the weakness of the backpropagation algorithm.

The study was conducted using a feed-forward four layer neural network. This net was trained using Levenberg-Marquardt algorithm, found to be taking the least training time (in seconds) among other known training algorithms Table 2.

2.3.5 Other Network Parameters

To select a best neural network model, other parameters having optimal value to achieve the desired goal were also considered Table 3.

2.3.6 Number of Output Nodes

The number of the neurons in the output layer is directly related to the problem under study. The study used four neurons in the output layer of the network. Each node in the output layer represented the option of a vocational stream of pursuit, namely engineering, science, clerical and administration and humanities for the stakeholder.

During the fourth phase of development, the network was trained using the training set chosen in Phase 3. The network used the current inputs to compute the output. Then the computed output was compared with the target output to determine the error, which was used to modify the weights to move closer to the target output. Some of the sample data was used for neural network validation. Among other performance measures such as training time and modeling time, the most effective measure is the classification accuracy of the net. Several accuracy measures such as root mean squared error, mean absolute error, the sum of squared error etc., have been reported by researchers. The performance of a trained network can be gauged by the errors on the training, validation, and test sets. The graph Figure 1 between mean square error v/s epoch (iterations) show similar characteristics for the validation and test set data depicting that no significant overfitting had occurred. The performance of the trained ANN was further investigate by performing a regression

Table 2. Selection of learning algorithm

Function	Technique	Time (seconds)
Trainrp	Resilient Backpropagation	1.59
Trainbfg	BFGS Algorithm	1.96
Trainlm	Levenberg-Marquardt	0.32

Table 3. Network parameters

Network parameter	Optimal value
Maximum epoch	100
Performance goal	0.0001
Minimum error gradient	1e ⁻¹⁰

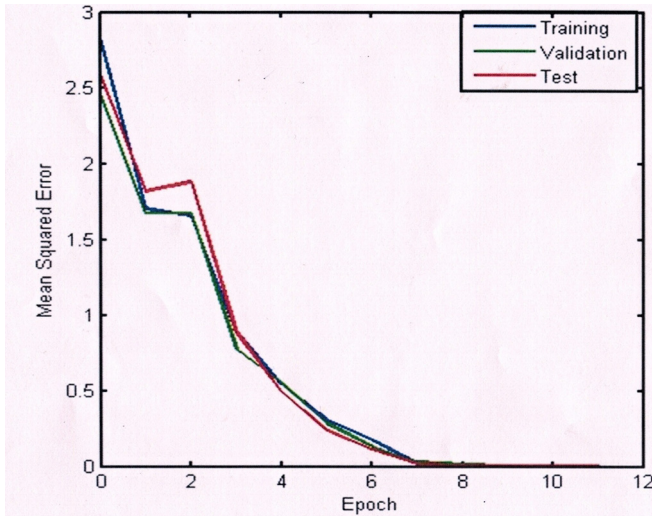


Figure 1. Graph of training, testing and validation error versus number of epochs for the ANN model.

analysis between the network output and the corresponding targets. The routine `postreg`²⁸ in MATLAB (neural network toolbox) was designed to perform this analysis. In the present study, it was observed that the R-value i.e., correlation coefficient between inputs and targets was close to 1. This indicated the best linear fit of weights for the trained ANN model.

Further, the evaluation of this decision support system was done on the basis of standard statistical measures like percentage error as

$$\text{Error}_i = \frac{(\text{stream}_{\text{actual}})_i - (\text{stream}_{\text{predicted}})_i}{(\text{stream}_{\text{actual}})_i} \times 100\% \quad (1)$$

For $I = 1, 2, \dots, N$ where N is a number of training data set consisting of training and validation data values. After calculating the classification error, classification accuracy was calculated as:

$$\text{Classification Accuracy (Training Set)} = 100 - \% \text{age of error in classification} \quad (2)$$

The average classification accuracy (Training Set) was calculated as 99.80% for the training set data. The trained neural network was tested with the sample data set to determine the accuracy of the prediction or classification. The trained ANN chooses the most suitable stream of pursuit for the candidate based on his stanine scores (converted raw score to normalized score) fed as input to the network. This involved the input variable data to the network without the output variable results.

2.4 Implementation of ANN using MATLAB

Once the network was trained to be sufficiently accurate, it was converted into a form that could be deployed with the application. During the fifth phase of the study, the ANN was implemented using MATLAB 7.8, R2009a. The GUI interface was prepared using GUIDE tool with MATLAB 7.8. The interface for the problem under study was developed using GUIDE tool Figure 2. The ANN was first trained (Figure 2.) by clicking the button “TRAIN THE NET”, while the other buttons were grayed (not selectable). After the training, the beneficiary enters the stanine scores obtained in the Aptitude test battery and click on button “SIMULATE THE NET” to obtain scores suggesting the suitable vocational stream of pursuit.

2.5 Performance Analysis of the Developed System

In this phase, the performance evaluation of the developed decision support system was conducted using a statistical measure such as percentage errors. The classification error for each pair of actual and predicted vocational stream of pursuit was calculated by Equation (1), for $i = 1, 2, \dots, N$, where N is the number of test cases. After calculating the classification error, classification accuracy was calculated as:

$$\begin{aligned} \text{Classification Accuracy (test data set)} \\ = 100 - \% \text{ error in classification} \end{aligned} \quad (3)$$

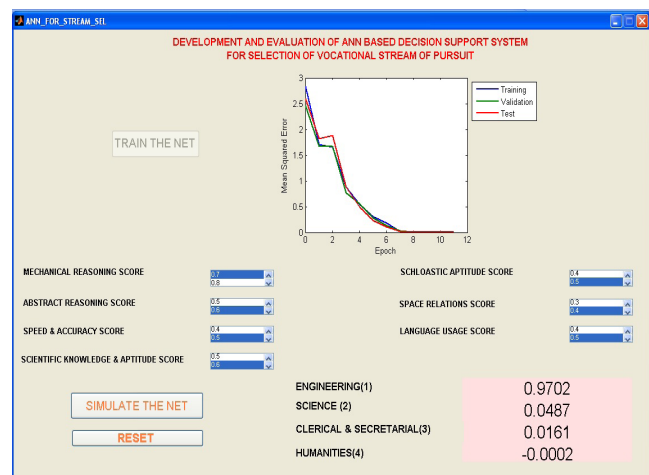


Figure 2. Screenshot of the graphical user interface for the developed ANN system

1 represented engineering stream, 2 represented science stream, 3 represented clerical and administration stream and 4 represented humanities stream. The average classification accuracy for the developed system was recorded to be 93.63% Table 4. This index of the accuracy of classification was quite high and validates the efficiency of ANN in predicting the stream of pursuit.

3. Results and Discussion

The two most common pattern recognition problems are regression analysis and classification. Regression analysis predicts the values of one or more continuous output variables for a given the set of input variables. The second problem is the classification and involves assigning input patterns to one set of discrete classes C_k where $k = 1, 2, \dots, c$. For prediction, two main statistical techniques are used 1) Regression analysis and multiple R2) Discriminant function analysis. Regression Analysis technique²⁹ is used when the data is parametric i.e., it fulfills the conditions of normality, linearity, and equal variance, whereas discriminant function analysis can be used with nonparametric or discrete data which is measured on nominal or ordinal scale. While training input variables were fed in ordinal scale (stanine scores), the discriminant function analysis was chosen to perform the classification of cases into four vocational streams of pursuit. Discriminant function analysis is a statistical technique used for classifying observations²⁹. It produces a classification Table 5, presenting categories and suggestive groups. Out of the 100 test cases, 14 cases were classified under stream 1, 18 cases under stream 2, 51 cases under stream 3 and 17 cases were found to group under stream 4. Results Table 6 revealed that four set of classifications were made using same test set data comprising of 100 cases. One classification was made using

Table 4. Classification results for ungrouped data

Stream	Classification Accuracy per stream (in %)	Average Classification accuracy over all streams (in %)
1	85.57318	93.63709
2	99.21819	
3	96.20912	
4	93.54787	

Table 5. Discriminant function analysis classification result

		Stream	Predicted Group Membership				Total
			1	2	3	4	
Original	Count	1	24	0	0	0	24
		2	0	29	0	0	29
		3	0	0	33	0	33
		4	0	0	0	25	25
	%	1	100.0	.0	.0	.0	100.0
		2	.0	100.0	.0	.0	100.0
		3	.0	.0	100.0	.0	100.0
		4	.0	.0	.0	100.0	100.0
		Ungrouped cases	14.0	18.0	51.0	17.0	100.0

a. 100.0% of original grouped cases correctly classified.

a statistical technique i.e., discriminant analysis through SPSS ver. 17.0, second classification by ANN technique and 3rd and 4th classification was made by experienced counselors who had around 10 to 15 years of experience in the field of guidance and counseling. On critical analysis of the results Table 6, it was found that the discrepancy in the classification for engineering stream with respect to ANN was 16.66% for Discriminant Function Analysis and Counselor 1. However, the prediction of Counselor 2 was found to be different from ANN by 33.33% for engineering stream. For the Science stream, a minimum discrepancy of 5.88% was found between ANN and Discriminant Function Analysis and maximum discrepancy of 41.17% was found between ANN and Counselor 2. For stream 3 (clerical and administration) percentage error in classification between ANN and Discriminant Function Analysis was 7.27% but the maximum percentage error in classification was 14.54% between ANN and Counselor 1. However, minimum difference in classification accuracy was obtained between ANN and Counselor 2 by 5.45% for stream 3. For stream 4 (Humanities), the output of ANN varied with the results obtained from Discriminant Function Analysis by 6.25%; with Counselor 1 by 25% and Counselor 2 completely agreed with the results of ANN.

4. Conclusion and Future Work

The classification made by ANN and Discriminant Function Analysis agreed by 91% over all streams. The

Table 6. Comparison of trained ANN outcome v/s discriminant analysis and human prediction

		Predicted outcome			% Error = $\frac{(\text{Ref. Value})_{\text{ANN}} - (\text{Pred. value})_{\text{DFA,C1,C2}}}{(\text{Ref. Value})_{\text{ANN}}} \times 100\%$		
		Discriminant Analysis (DFA)	Counselor 1	Counselor 2	DFA	Counselor 1	Counselor 2
Trained ANN outcome	Stream 1 (Engineering)=12 cases	14	14	8	16.66	16.66	33.33
	Stream 2 (Science)=17 cases	18	19	24	5.88	11.76	41.17
	Stream 3 (Clerical and Administration)=55 cases	51	47	52	7.27	14.54	5.45
	Stream 4 (Humanities)=16 cases	17	20	16	6.25	25	0
	Total Cases	100	100	100			

Table 7. Comparison of predicted output by counselor 1 v/s counselor 2 over all test cases

		Predicted outcome	% Diff. Error = $\frac{ (C1-C2) }{(C1+C2)/2} \times 100\%$
		Counselor 2 (C2)	
Classification by Counselor 1 (C1)	Stream 1(Engineering)=14 cases	8	54.54545
	Stream 2 (Science)=19 cases	24	23.25581
	Stream 3 (Clerical and Administration)=47 cases	52	10.10101
	Stream 4 (Humanities)=20 cases	16	22.22222
	Mean % DifferenceError		27.53113

Counselor 1 and ANN were found to agree on 83% over total cases. The Counselor 2 agreed with the ANN output on 80% cases over all the streams. The present study implied that the classification made by ANN was approximately same (91%) as obtained by a statistical method namely, discriminant function analysis. So, empirical data supports the classification made by ANN decision support system. However, before applying the Discriminant Function Analysis, the raw data and the predictor variables must fulfill conditions such normality, linearity, equal variance, etc., and require

laborious computations whereas no such conditions are required for the ANN. Further, it could be concluded that both counselors were at a variance of approximately 27% in classification over all test cases Table 7, which supports the hypothesis of this study that the decisions made by human beings might get affected by the numerous factors such as counselor biases, parent's aspirations, student's interest, stress, etc. The present study invariably showed that Artificial Neural Network performs successfully in complex decision making tasks where other methods fail to do so. The ANN system

developed out of the study could be used as a standalone guidance counseling system or help the guidance counselor in decision making for choosing the appropriate vocational stream for the student. Further, some additional input variables such as interest, social economic status, cognitive style, etc., may be included as input variables to improve the accuracy of prediction of the developed system. The study may be extended by taking a large sample of the target population by the developed system and an expert system (fuzzy system) may be integrated with to enable it to further choose occupational streams under the selected vocational stream for the beneficiary.

6. References

1. Dutta S, Shekhar S. Bond rating: A non-conservative application of neural networks. Proceedings of the 2nd Annual IEEE International Conference on Neural Networks; San Diego, CA, USA. 1988 Jul. p. 443–50.
2. Odom MD, Sharda R. A neural network model for bankruptcy prediction. Proceedings of the IEEE International Joint Conference on Neural Networks; San Diego, CA. 1990 Jun. p. 163–8.
3. Brause R, Langsdorf T, Hepp M. Neural data mining for credit card fraud detection. Proceedings of 11th IEEE International Conference on Tools with Artificial Intelligence; Washington: IEEE press. 1999. p. 103.
4. Dorronsoro JR, Ginel F, Sanchez C, Cruz CS. Neural fraud detection in credit card operations. IEEE Transactions on Neural Networks. 1997 Jul; 8(4):827–34.
5. Ghosh S, Reilly DL. Credit card fraud detection with neural networks. Proceedings of the 27th Annual Hawaii International Conference on System Science; Wailea, HI, USA. 1994 Jan. p. 621–30.
6. Ibourk A, Aazzab A. The artificial neural networks for classification: Case of business failures. Journal of Research in Business and Management. 2016 Jul; 4(3):10–15.
7. Ozerdem MS, Akpolat V. Classification of bone density with using neural networks. Proceedings of the 15th Signal Processing and Communications Applications; Eskisehir: IEEE. 2007 Jun. p. 1–5.
8. Sathe SS, Purandare SM, Pujari PD, Sawant SD. Share Market Prediction using Artificial Neural Network. International Education and Research Journal. 2016 Mar; 2(3):74–5.
9. Zhang W, Li C, Ye Y, Li W, Ngai EWT. Dynamic business network analysis for correlated stock price movement prediction. IEEE Intelligent Systems. 2015 Mar-Apr; 30(2):26–33.
10. Tino P, Schittenkopf C, Dorffner G. Adaptive information systems and modelling in economics and management science, Report No. :46. Vienna (Austria): Vienna University of Economics and Business Administration; 2000 Mar.
11. Sajja PS. Fuzzy artificial neural network decision support system for course selection. Journal of Engineering and Technology. 2006; 19(1):99–102.
12. Sajja PS. Type-2 fuzzy user interface for artificial neural network based decision support system for course selection. International Journal of Computing and ICT Research. 2008 Dec; 2(2):96–102.
13. Waghmode ML, Jamsandekar. A study of expert system for career selection: Literature review. International Journal of Advanced Research in Computer Science and Software Engineering. 2015; 5(9):779–85.
14. Sodhi JS, Dutta M, Aggarwal N. Development and evaluation of ANN based decision support system for selection of vocational stream of pursuit. Proceedings of National Conference on Computing, Communication and Control; India. 2009. p. 5–8.
15. Bennett GK, Seashore HG, Wesman AG. Differential aptitude tests: Technical supplement. San Antonio. Tx: Harcourt Assessment; 1984.
16. Ganguly D, Ghosh D, Chatterji S, Mukherjee M. An investigation into the validity of a scientific knowledge and aptitude test. Psychometric Research and Service Unit. Calcutta: ISI; 1972.
17. Cybenko G. Approximation by superpositions of a sigmoidal function, Mathematical Control. Signals and Systems. 1989 Dec; 2(4):303–14.
18. Hornik K, Stinchcombe M, White H. Multilayer feed forward networks are universal approximators. Neural Networks. 1989; 2(5):359–66.
19. Zhang G, Patuwo BE, Hu MY. Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting. 1998 Mar; 14(1):35–62.
20. Srinivasan D. Evolving neural network for short-term load forecasting. Neuocomputing. 1998 Dec; 23(1-3):265–76.
21. Zhang X. Time series analysis and prediction by neural networks. Optimization Methods and Software. 1994; 4(2):151–70.
22. Chester DL. Why two hidden layers are better than one. Proceedings of the International Joint Conference on Neural Network; 1990 Jan. p. 265–8.
23. Klimasauskas CC. Applying neural networks. Part 3: Training a neural network. PC-AI; 1991 Jun. p. 20–4.
24. Falhman SE. Faster-learning variations of back-propagation: An empirical study. Proceedings of the 1988 Connectionist Models Summer School; 1989. p. 38–51.
25. Parker DB. Optimal algorithm for adaptive networks: Second order back propagation, second order direct propagation, and second order Hebbian learning. Proceedings of the IEEE International Conference on Neural Networks; 1987. p. 593–600.

26. Battiti R. First- and second-order methods for learning: Between steepest descent and Newton's method. *Neural Computation*. 1992 Mar; 4(2):141–66.
27. Cottrell M, Girard B, Girard Y, Mangeas M, Muller C. Neural modeling for time series: A statistical stepwise method for weight elimination. *IEEE Transactions on Neural Networks*. 1995 Nov; 6(6):1355–64.
28. Demuth H, Beale M. *Neural network toolbox user's guide*. The Math Works Inc; 1998 Jan.
29. Klecka WR. *Discriminant analysis. Quantitative Applications in the Social Sciences*. CA: Sage Publications; 1980.