

Impact Determination of Social Media on Student's Academic Performance Using Neural Network

Dept. of Computer Science
& Engineering
Bangabandhu Sheikh Mujibur Rahman
Science and Technology University
Gopalganj, Bangladesh
Email: mnhshisir@gmail.com

Dept. of Computer Science
& Engineering
Bangabandhu Sheikh Mujibur Rahman
Science and Technology University
Gopalganj, Bangladesh
Email: ferdousahmed861@gmail.com

Abstract—Today, social media is an unbroken part of our lives. It plays a big role in education. But its outcome is positive for someone negative for someone. In this paper, we try to understand the effect of social media (facebook, twitter, linkedin, Instagram and so on) on students academic performance apply neural networks. Artificial Neural Network (ANN) is very useful for prediction such as weather prediction, stock market prediction and so on. This prediction can also be done using normal statistics, but inadequate when dealing with complex and highly nonlinear data. We will predict the impact of social media (good or bad) on learners academic performance. We use Artificial Neural Network (ANN) for predicting the effect of social media in the form of 1 or 0 (if good then 1, if bad then 0). Then we will match it with the actual data that will vary. In order to create the actual dataset, we are trying to collect data from almost all public universities of Bangladesh.

I. INTRODUCTION

Nowadays not only students but also every people are drastically attached to use social media. Obsessive use of social media is visible in our society. There are lot of students who have overpass much of their time beneath social media than study. Today social media plays a good role of communication. So it may be a means of rapid communication among students. We may learn many things using this media. After all it opens a vast area of communicating and learning. Students can use it both academic and non-academic purpose. Therefore, there may be an impact on social media on students life. This impact may be either positive or negative.

Institutions ranking is a vital part for every academic organization which primly depends on students perfection. The excellence of the students depends on their intellectual mindset, characteristics, creativity, academic performance etc. Almost all institutions require keep extra observation on students yearly performance for their future development. Based on the academic performance of students, a preliminary decision can be taken that how she might be. In this regard students quality is not at all but other related matters are also involved like social interaction is needed.

The use of social media plays a variety of roles in educations, which include providing a media to share ideas as well as allowing students to build their own communities to collaborate with each other, facilitating the art of learning and reaching out to the students in order to understand and

teach them. At the same time it has a negative or positive impact on students academic life. Although other reasons are responsible for negative or positive impact on students academic life like family condition, surrounding environment, integrity, knowledge and so on. Basically in this project we try to determine the impact of social media on students academic life. Prediction can be done using two different ways: Computational Intelligence (CI) and statistics. The first, The branch of mathematics that deals with the collection, organization, analysis, and interpretation of numerical data. Statistics is especially useful in drawing general conclusions about a set of data from a sample of the data. Statistics have solid and widely accepted mathematical foundations and can provide insights on the mechanisms creating the data. However, they frequently fail when there are nonlinearities (curse of dimensionality) involved. The second, Computational Intelligence (CI) is an embranchment of artificial intelligence. Evolutionary computation, neural networks and fuzzy systems can be evaluated through Computational Intelligence (CI). Artificial Neural Networks (ANN), an extremely popular class of CI models, has been widely applied to various predicting problems, because they are very generic, accurate and convenient mathematical models able to easily simulate numerical model components. The problem of nonlinearities, large dataset can be solved easily through Artificial Neural Networks (ANN).

The main contributions of this dissertation are described as follows: For this we use a real dataset from the Department of Computer Science and Engineering at Bangabandhu Sheikh Mujibur Rahman Science and Technology University (BSMRSTU). Here, we take the data of 142 students for results analysis. By using data mining technique we extract required information. We also use supervised neural network and decision tree algorithm. In this case, the Levenberg Marquardt back propagation algorithm is of use to train the neural network. Then, we divide the whole dataset into three parts, one is for training the network, another one is for validating the results and the last one is for testing the results. Therefore, the error percentage of the prediction will be lower. Decision tree is used to determine correctly/incorrectly classified instance, kappa statistics, true position, false position rate. Using our proposed neural network and decision tree one can find the effect of social media on his/her academic life with respect to parameters.

The dissertation is organized as follows:

In **Section II** This chapter first shows the state of the art methods of the prediction of nonlinear data. Then describe some of existing works for determining impact or effect of social media usages on students academic life. The limitations of these methods are clearly addressed, as these are the focuses of this dissertation, while **Section III** introduces our proposed solution for the prediction model. Description of data-set are described in **Section IV**. Implementation of our proposed technique has been described in **Section V**. **Section VI** includes experiments and evaluation to show the effectiveness of our proposed system. Concluding remark and some future directions of our work are described in **Section VII**.

II. RELATED WORK

A. Related works statistics methodology

Esam Alwagait et al. [1] shows on the impact of social media usage on students academic performance in Saudi Arabia. They used the statistics strategy in this regard.

Waleed Mugahed Al-Rahmi et al. [2] shows on The impact of social media use on academic performance among university students: A pilot study. They used quantitative approach and UTM library to draw the data for the study. A conceptual framework that identifies instruments by which using social media influences students academic performance through collaborative learning is supplied.

B. Related works non-statistics methodology

VO Oladokun et al. [3] shows the usefulness of Artificial Neural Network (ANN) in maintenance planning and management.

Fariborz Y Partovi et al. [4] focused on Predicting Students Yearly Performance using Neural Network: A Case Study of BSMRSTU. They used Artificial Neural Network (ANN) to predict and also proposed a system students yearly academic performance.

Md. Fahim Sikder et al. [5] focused on Predicting Students Yearly Performance using Neural Network: A Case Study of BSMRSTU. They used Artificial Neural Network (ANN) to predict and also proposed a system students yearly academic performance.

Martin T. Hagan et al. [6] focused on training feedforward networks with the Marquardt algorithm. The algorithm is tested on several function approximation problems, and is compared with a conjugate gradient algorithm and a variable learning rate algorithm. It is found that the Marquardt algorithm is much more efficient than either of the other techniques when the network contains no more than a few hundred weights.

V.O. Oladokun et al. [7] focused on Predicting students academic performance using artificial neural network: A case study of an engineering course. They used Artificial Neural Network (ANN) to predict and also proposed a system students academic performance.

Abbas Rohani et al. [8] focused on Prediction of tractor repair and maintenance costs using artificial neural network. They

try to evaluated as an alternative method for the prediction of machinery (specifically tractor) repair and maintenance costs.

Changsong chen et al. [9] focused on solar power forecasting using Artificial Neural Network(ANN). That's why they classified the weather type like as irradiance, relative humidity and temperature at the site of the photovoltaic power system.

Yakup Kara al. [10] They predict the direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. For this attempt ,they used two efficient models and compared their performances in predicting the direction of movement in the daily Istanbul Stock Exchange (ISE) National 100 Index. The models were based on two classification techniques, artificial neural networks (ANN) and support vector machines (SVM).

L. Ekonomou [11] try to predict Greek long-term energy consumption prediction using artificial neural networks. To select the best generalizing ability of several possible architectures the multilayer perceptron model (MLP) has been used. Author try to comparison of the predicted with the use of the ANN, Regression model and Support Vector Model Greek energy consumption with the actual one.

Holger R. Maier et al. [12] shows the prediction on current status and future directionson of water resource variables in river systems.

Holger R. Maier et al. [13] shows the prediction on current status and future directionson of water resource variables in river systems.

S Sebastianet al. [14] shows Evaluating students performance by artificial neural network using weka.

C. Limitations of exiting works

From the previous sections, it has been shown for predicting the impact of social media usages on students academic life use statistical method which may be complex for nonlinear and huge verity of data. It might be more complex for vast volume of data set.

III. PROPOSED SYSTEM

The overview of the system is shown in Figure: 1. It represents all the concepts of how to prediction will be done. First, we will collect information from educational institutions. After collecting the data we reprocess it using Microsoft Excel set. Using this tool we normalized the data. Then we construct a supervised neural network which trains according to data and extract information from the data. In our experimental case, we predict the influence of social media usages on their academic performance. We use Levenberg- Marquardt backpropagation algorithm to train the neural network. According to this system, we try to predict the influence of social media on students academic achievement in the form of 1(good) and 0(bad) and then we try to find out the accuracy percentage. Each part of our proposed system is described in the following subsections.

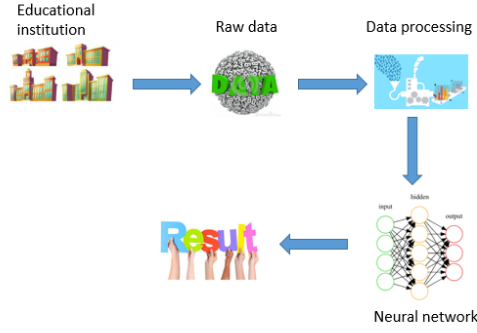


Fig. 1. System Overview.

A. Data Collection

We try to collect qualitative data through direct interaction with individuals. It is a tough task to collect qualitative data from real world. Nevertheless, we are trying our best to collect data. At first we collect relative data from students of different universities of Bangladesh to determine the impact of social media on academy performance. To do this, first we create a Google form that forms allow to collect data from most university students of Bangladesh. We also try to collect data with a hard copy of the form from nearby university. We collect these data from the students via a form where was some questionnaires. Here we use 142 students data for the research which is taken from the different universities of Bangladesh. In this survey we consider nine factors. In these factors, we regard as not only academic data but also others personal information such as number of accounts in social media, highest time spent in social media (in hours), time spent for academic purpose in social media (in hours), time spent for entertainment in social media (in hours) etc.

B. Data Processing

Data processing is the way to produce a meaningful information from collected data. Working from raw data is a bit difficult as the data are nonlinear. Hence, we need to filter out some unnecessary data so that we have a right date for Neural Network training, testing and validation. For this, we used Microsoft excel tool for filtering these unnecessary data. Then we divide the nine factors in two parts. One part is the input part which uses eight factors among these nine factors and remaining one factor is called target value. This target value is the most essential part of the data because the neural network will find the pattern of the data and predict the result based on the target value [5]. Figure: 2 shows the steps of data preprocessing.

C. Neural Network

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information¹. The key element of this paradigm is the novel structure of the information processing system². Artificial neural networks are

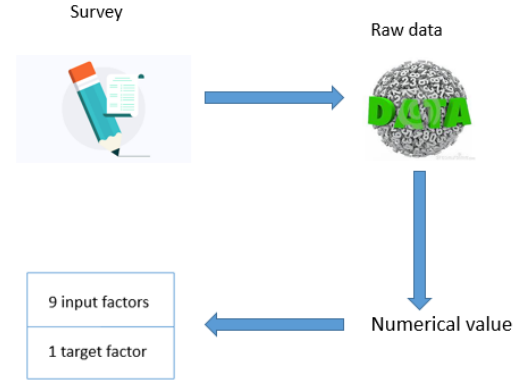


Fig. 2. Data processing.

relatively crude electronic networks of neurons based on the neural structure of the brain³. They process records one at a time, and learn by comparing their prediction of the record (largely arbitrary) with the known actual record⁴. The errors from the initial classification of the first record is fed back into the network, and used to modify the networks algorithm for further iterations⁵. To get the prediction first we make a neural network. Then we train the neural network with the preprocessed data. We used supervised neural network and train the network using feed-forward backpropagation with traingdm learning function and Levenberg-Marquardt back propagation algorithm. This method is the part of the neural network toolbox integrated in the MATLAB and neuroph studio. Supervised learning is the machine learning system of inferring a function from labeled training data. In supervised learning system two kinds of data are provided which called inputs and outputs. Then the system processes the inputs and compares its verdict outputs against the required outputs. Sometimes systems find some errors. These errors are back propagating to the system. Then the system adjusts the weights which control the network. This process happened recursively as the weights are frequently.

IV. DESCRIPTION OF DATA SET

A. Average CGPA

The students who get good result will inspire to continue his/her performance for the next time. On the other hand, the Students do not achieve expected results will be disappointed. That is why, future results depend on current result. Similarly current results depend on previous results. Hence, we consider last semester result for our experimental analysis.

B. Number of social media accounts

With so many social media, there are so many social media accounts, so we try to know the number of accounts they have (like facebook, twitter, instagram, skype etc).

¹<https://www.doc.ic.ac.uk>

²<https://www.doc.ic.ac.uk>

³<https://www.solver.com>

⁴<https://www.solver.com>

⁵<https://www.solver.com>

TABLE I. INPUT DATA TRANSFORMATION

S/N	Factors	Domain
1	Average CGPA	CGPA
2	Number of social media accounts	Number
3	Highest accessed account per day	Number
4	How much time spent on social media	1. less than 2 hours 2. 2 to 4 hours 3. 4 to 6 hours 4. More than 6 hours
5	How much time spent on social media for academic purpose	1. less than 1 hours 2. 1 to 2.30 hours 3. 2.30 to 4 hours 4. More than 4 hours
6	How much time spent on social media for entertainment	1. less than 1 hours 2. 1 to 2.30 hours 3. 2.30 to 4 hours 4. More than 4 hours
7	How much time spent for study without social media	1. less than 1 hours 2. 1 to 2.30 hours 3. 2.30 to 4 hours 4. More than 4 hours
8	Do you have any blog	1. Yes 2. No

C. Highest accessed account per day

Today, social media is an integral part of our lives. So we enter social media every day. But each of us has one or more social media accounts. Most people have a lot of social media accounts, they have access to the maximum number of accounts per day, and we have tried to find out through this parameter.

D. How much time spent on social media

Through this parameter, we have tried to find out that they spend the highest work hours in social media

E. How much time spent on social media for academic purpose

Here we have tried to know how much of the time the social media spend in that period, how much time they spend on more than one time

F. How much time spent on social media for entertainment

Through this, we have tried to find out how much time they spend on social media

G. How much time spent for study without social media

They have been trying to know how much time social media person spend for the day to study

H. Do you have any blog

Whether he has a blog or whether a blog regularly navigates or is connected to a blog

I. Effect

If he thinks the impact of social media on his academic life is positive, then he will fill one and if he thinks that the impact of social media is only on his academic life, then he will fulfill the last 0.

TABLE II. OUTPUT DATA TRANSFORMATION

S/N	Factors	Domain
1	What is the effect of social media on your academic performance	1. Good 2. Bad

V. IMPLEMENTATION DETAILS

A. Network selection

After the data has been processed and transformed, method of training has been chosen, it is necessary to then determine the topology of the neural network. The network topology describes the arrangement of the neural network [7]. Choosing the topology of the neural network is a difficult decision [3]. There are different types of network. We can select as our demand. Each network provides some advantages and disadvantages. For example, some networks are good for speed accuracy, while some of them are good for handling static variables and not continuous ones. Hence, from various network topologies such as Multilayer Perceptron, recurrent network, and time-lagged recurrent network were considered. Due to the nature of our work, the Multilayer Perceptron was selected. An MLP is a network of simple neurons called perceptrons. The perceptron computes a single output from multiple real-valued inputs by forming a linear combination according to its input weights and then possibly putting the output through some nonlinear activation function.

$$y = \varphi\left(\sum_{i=0}^n w_i x_i + b\right) = \varphi(W^T X + b) \quad (1)$$

where \mathbf{w} denotes the vector of weights, \mathbf{x} is the vector of inputs, b is the bias and φ is the activation function.

$$X = f(s) = B\varphi(As + a) + b \quad (2)$$

where \mathbf{s} is a vector of inputs and \mathbf{x} a vector of outputs. \mathbf{A} is the matrix of weights of the first layer, \mathbf{a} is the bias vector of the first layer. \mathbf{B} and \mathbf{b} are, respectively, the weight matrix and the bias vector of the second layer. The function φ denotes an element wise non linearity. A typical multilayer perceptron (MLP) network consists of a set of source nodes forming the input layer, one or more hidden layers of computation nodes, and an output layer of nodes. The input signal propagates through the network layer-by-layer⁶. It is so tough building a neural network model with number of nodes and hidden layers. Because small number of hidden layer lower the processing capability. Comparatively the system will slow down if a large number of hidden layer. We come to a conclusion from this paradigm, choose a network with a hidden layer and eight input processing elements, one output. Train this network with different learning rate and different hidden layer. For which learning rate and hidden layer our prediction came closer we took this network.

B. Data set grouping

In supervised training, the data is divided into 3 categories; the training set, verification set and the testing set. The training set enables the system to observe relationships between input data and resulting outputs, so that it can develop relationship between the input and the expected output. A total of 142 students records were used in the analysis. About 70% of the

⁶<http://www.helsinki.fi>

total data (i.e. 96candidates) were used as the training set, 15% (i.e. 21 candidates) as the testing set, and 15% (i.e. 21 candidates) used for cross validation.

C. Network training and validation process

In the implementation of the system first we take the preprocessed dataset and use them in the MATLAB. Then we make the neural network with ten hidden layers. To train the network we add the training algorithm, in this case we use Levenberg-Marquardt backpropagation algorithm. For applying this in MATLAB we use trainlm function. The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix in the following Newton-like update in the equation 1.

$$X_{k+1} = X_k [J^T J + I]^{-1} J^T e \quad (3)$$

Where, X is bias variable, J is Jacobian matrix, is scalar, I is identity matrix, e is vector of network errors. The network was trained with the number of runs set to three and the Epoch set to terminate at 14. The training performance is then evaluated using the following performance measures:

$$MSE = \frac{\sum_{i=0}^{PN} (d_{ij} - Y_{ji})^2}{NP} \quad (4)$$

where:

p = number of output of processing element.

N= no of exemplars in the data set.

Yij=network output for exemplars i at processing element j

dij=desired output for exemplars i at processing element j

VI. EXPERIMENT AND EVALUATION

After running the neural network we have the predicted result. The result is best validated at 7 epochs out of 13 epochs. Figure: 3 shows the performance of the neural network.

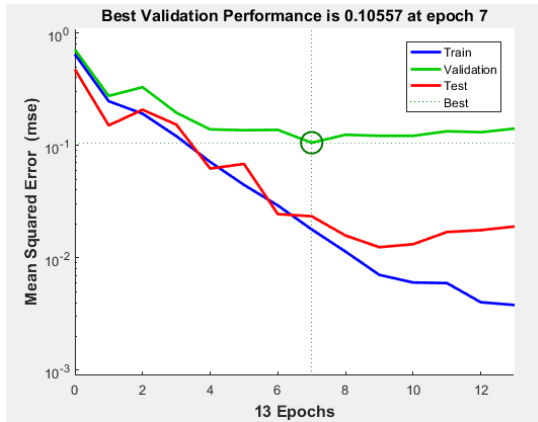


Fig. 3. Performance validation.

Error histogram shows maximum data are trained up with less error. We find the errors from the difference between targets results and output results. In the Figure: 4, shows the different errors histogram with 20 bins.

Regression analysis of Figure: 5 shows the consistency between input and output. Which tells us how well neural network are trained, validate and tested. It increases the reliability about network and dataset.

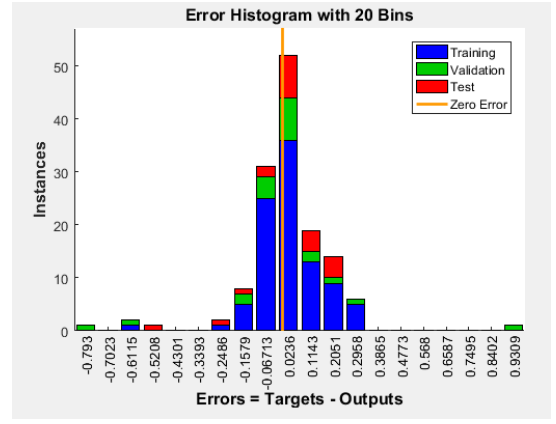


Fig. 4. Error histogram.

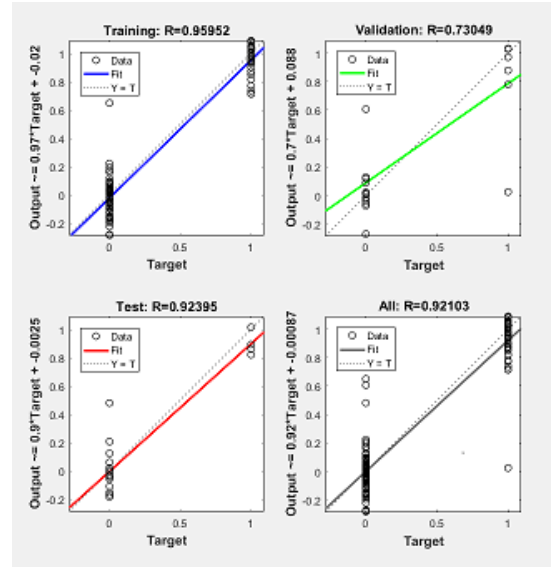


Fig. 5. Regression

Figure: 6 shows when the validation fail at the time of repetition. In this case at 13 epochs validation are failed.

In the table III we show the highest is 99.986% when students original result is 2.7 and we get lowest accuracy 91.547 % when original result is 3.83. In this results, we say that our system is better when students result is poor. Because good results depends many factors. That is why, its accuracy is small than the others.

The table III shows that our propose Neural Network predicted effect (1 means good and 0 means bad) and real dataset effect (1 means good and 0 means bad). It also shows the error percentage between real data value versus our proposed systems predicted value.

VII. CONCLUSION

Social media is may be a part of education today. Through social media, it is possible to reach the other end of the world on one side of the education system of the world. Today education system is largely dependent on social media. The use of social media among the students is increasing day

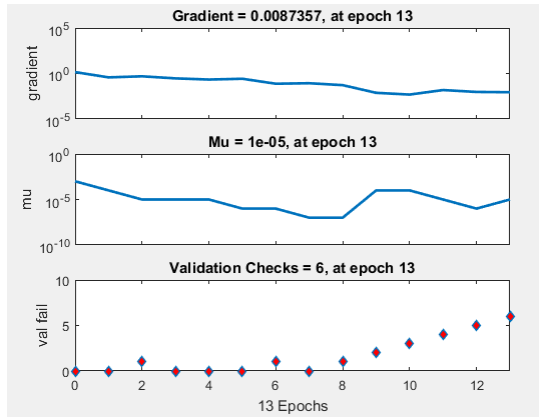


Fig. 6. State

TABLE III. RESULT

Average CGPA	Effect(0 means bad and 1 means good)	Predicted effect(0/1)	Error%
2.37	0	0	0%
3.73	1	0.999	1.00%
3.41	1	0.961	3.9%
2.23	0	0.035	3.5%
3.87	1	1	0%
3.5	1	0.786	21.4%
2.93	0	0	0%
3.27	0	0.042	4.2%
2.9	0	0	0%
2.23	0	0.001	1%
2.5	0	0.019	1.9%

by day. But excessive use of social media does not bring success in academic life. Basically, we wanted to show you that paper, social media has an impact on students academic performance, whether it is negative or positive. Using Artificial Neural Networks(ANN) algorithm so far, we have found that social media has an effect on the academic performance of students, this effect is congenial. It turns out that those who are very addicted to social media, their results are not very good, that is, their academic performance is at a lower level. It also shows that the performance of the students are good for those who do not spend too much time in social media. It is a remarkable that those who spend a lot of time in social media their performances are poor than those who spends a lot of time for academic study. With the capabilities of the proposed method, we plan to investigate and explore the following related problems, extensions:

We would like to investigate whether other efficient algorithms would lead to better discovery of predicting the

impact of the usages of social media on students academic performance.

We try to increase data volume and apply data mining for analyzing collected data.

REFERENCES

- [1] E. Alwagait, B. Shahzad, and S. Alim, "Impact of social media usage on students academic performance in saudi arabia," *Computers in Human Behavior*, vol. 51, pp. 1092–1097, 2015.
- [2] W. Al-Rahmi and M. Othman, "The impact of social media use on academic performance among university students: A pilot study," *Journal of information systems research and innovation*, vol. 4, no. 12, pp. 1–10, 2013.
- [3] V. Oladokun, O. Charles-Owaba, and C. Nwaozuru, "An application of artificial neural network to maintenance management," *Journal of Industrial Engineering International*, vol. 2, no. 3, pp. 19–26, 2006.
- [4] F. Y. Partovi and M. Anandarajan, "Classifying inventory using an artificial neural network approach," *Computers & Industrial Engineering*, vol. 41, no. 4, pp. 389–404, 2002.
- [5] M. F. Sikder, M. J. Uddin, and S. Halder, "Predicting students yearly performance using neural network: A case study of bsmrstu," in *Informatics, Electronics and Vision (ICIEV), 2016 5th International Conference on*. IEEE, 2016, pp. 524–529.
- [6] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the marquardt algorithm," *IEEE transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.
- [7] V. Oladokun, A. Adebajo, and O. Charles-Owaba, "Predicting students academic performance using artificial neural network: A case study of an engineering course," *The Pacific Journal of Science and Technology*, vol. 9, no. 1, pp. 72–79, 2008.
- [8] A. Rohani, M. H. Abbaspour-Fard, and S. Abdolhahpour, "Prediction of tractor repair and maintenance costs using artificial neural network," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8999–9007, 2011.
- [9] C. Chen, S. Duan, T. Cai, and B. Liu, "Online 24-h solar power forecasting based on weather type classification using artificial neural network," *Solar Energy*, vol. 85, no. 11, pp. 2856–2870, 2011.
- [10] Y. Kara, M. A. Boyacioglu, and Ö. K. Baykan, "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange," *Expert systems with Applications*, vol. 38, no. 5, pp. 5311–5319, 2011.
- [11] L. Ekonomou, "Greek long-term energy consumption prediction using artificial neural networks," *Energy*, vol. 35, no. 2, pp. 512–517, 2010.
- [12] B. Cetin, "Approaches to learning and age in predicting college students' academic achievement," *Journal of College Teaching & Learning (Online)*, vol. 13, no. 1, p. 21, 2016.
- [13] H. R. Maier, A. Jain, G. C. Dandy, and K. P. Sudheer, "Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions," *Environmental Modelling & Software*, vol. 25, no. 8, pp. 891–909, 2010.
- [14] S. Sebastian and J. J. Puthiyidam, "Evaluating students performance by artificial neural network using weka," *International Journal of Computer Applications*, vol. 119, no. 23, 2015.