RESEARCH ARTICLE   OPEN ACCESS

Student Performance Prediction in Higher Education Using Lion – Wolf Optimization Algorithm

Mrs. V. Ananthi[1], Mrs. K. Mythili[2]

[1]M Phil Research Scholar, Department. of Computer Science, [2]Head & Associate Professor, Department of Computer Technology, Hindusthan College of Arts and Science, Coimbatore.
Corresponding Author : Mrs. V. Ananthi

ABSTRACT
Students are the future of the country and their education qualification becomes highly important for the growth of the country. Predicting the performance of a students with the help of prediction models, is the one of the recent trends in Data Mining area. Analyzing the factors affecting student's performance is the prerequisite that must be made before designing the performance improvement program. One of the greatest challenges each higher education institution facing currently is predicting the performance of the student. The main purpose of this research is to analyze and identify the factors such as schooling, family background that influence the students to select the area in higher education. There were several attempts made to predict the performance of the students, to achieve highly qualitative education, but the prediction accuracy is not acceptable. On the various researches done on Neural Network based methods, it is expected that this method can be adopted to predict the student's performance with more accuracy. We have proposed a prediction models based on Lion-Wolf training algorithm to build a model for predicting the performance of the higher education students by using their past education and general record.
Key words: Neural Network, Neuron and Lion-Wolf.

I. INTRODUCTION

Education is one of the key features for the improvement of a nation. It has gone through tremendous evolution based on the requirement of time and the emerging changes in the situation of the world. It gives a great chance for the people to understand the social, economic, good and profound issues of the humankind. A country needs more effective and knowledgeable individuals to drive its economy forward. There are numerous Indians around the world who are well known for their capacities and aptitudes. To bring India as one of the leading instruction center in the world and to bring the Indian Economy upward in the worldwide, we would need to identify some of the innovative ideas in the area of Education to make it more effective.

One of the key area is Analyzing Student's Performance where the classification task is employed to gauge student's performance and as there are several approach that is used for information classification. The choice tree methodology is employed here in information's like attendance, class test, seminar and assignment marks were composed from student's management system, to predict performance at the tip of the semester.

II. CLASSIFICATION TECHNIQUES

It is one of the key data mining techniques which are used for classifying each item in a data set into a predefined set of groups or classes. For example, a classification model can be used to classify whether a loan applicant is low, medium or high credit tasks based on machine learning. Classification divides data samples into target classes. The classification technique is used to predict the target class for each objects. In classification technique, the final data set is partitioned into two categories, training dataset and testing dataset. In this technique the items in a collection are assigned to target categories or classes. The main purpose of classification is to accurately predict the target class for each case in the data. Following mathematical techniques are used as part of Classification method,

- Back propagation
- Genetic algorithm Classification
- Association based Classification
- Decision tree
- Linear programming
- K-Nearest Neighbors
- Naive Bayesian Classification
- Artificial Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Associations.

III. EDUCATIONAL DATA MINING OVERVIEW

Educational Data Mining (EDM) is used to develop some methods that can help to explore the educational database, and using those methods to better understand student's performance, and the environment which they learn. The main important features of education data are context, sequence and time. In which context the techniques are proposed, and also whether the proposed methods correctly work based upon target results. Sequence used to identify the ordering of context used for get the target results. The main important feature is time, how the data is gathered based on session or period of interval. This isshown in Fig 3.1 Educational Data Mining System.
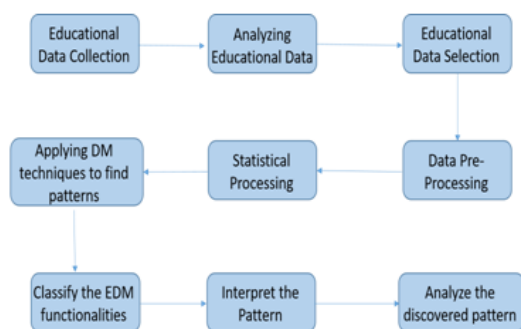

Fig 3.1 Educational Data Mining System

IV. NEED FOR PREDICTION MODEL

The knowledge obtained through application of Educational Data Mining can be used for various things such as offer suggestions to academic planners in higher education institutes to enhance their decision making process, take actions for underperforming students given in articles [1] (Ognjanovic et al., 2016). Many of factors are considered for EDM, some of them are environmental factors, society factors, school factors, college factors, individual factor, family factors etc., however the factors considered of prediction are not consistent [2] (Malvandi and Farahi, 2015). For effective prediction of student performance model, all factors related to student must be included based on the work results of [3] (Shetgaonkar, 2015).

4.1 Challenges in Existing Prediction Models

The important challenge is to predict the student performance based on their activities and performance in school levels. The attributes or variables considered for performance prediction is another challenge in student performance prediction model. The complexity in modelling is additional challenge of performance prediction model addressed by [4] (Wolff et al., 2014). The

model should be done in a way with reduced time consumption and with effective performance.

In NN based prediction model technique, optimal weight updation in training of neural network must be done properly without affecting actual output [5] (Guo et al., 2015). Neuron weight updation is vital, since the prediction accuracy depends on it. Even if the data selected for prediction decreases, the model must perform well without degradation in prediction accuracy [6] (Guar´ın et al., 2015).

V. PROPOSED PREDICTION MODEL

The success of students in education during their study is the main objective of the higher education system in India. There were several methodologies adopted by these higher educational institutions to improve the performance of the students and quality of education. The main aim is to create a model that classifies the instances correctly to predict the performance of students using these two prediction models Lion –Wolf Prediction Algorithm and Fuzzy Decision Tree Prediction Model. These methodology consists of several attribute of a student modelled that is collected based on several level testing done on a particular set of students. This involves student's complete academic details, his/her creative and other interpersonal skills and finally the level of interest towards the present educational approach.

5.1 Collecting Student's Record to Predict their Performance

Higher educational institute can use educational database which holds hidden information. This hidden information is helpful in improving students' performance in their academics. Successful prediction of students' performance leads to assure and achieve better quality in higher education system. There were set of predefined attributes which will be used to determine the performance of student. The hierarchy of performance prediction feature factors considered in this work is shown in Fig. 5.1
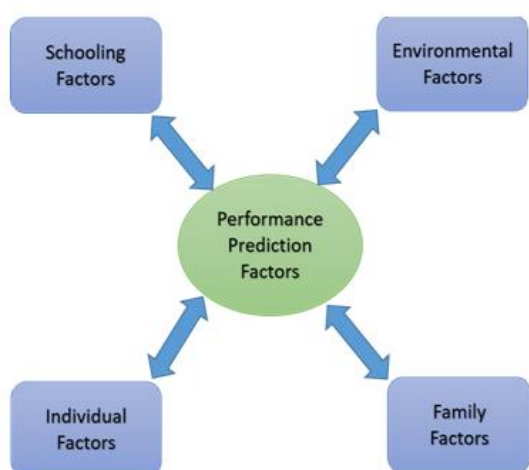
Figure 5.1 Performance Prediction Factors

5.1.2 Data Selection

Data selection is second step in proposed student performance prediction model. This is the fundamental block of prediction model, since the complexity in prediction is avoided by selection of relevant features from bulk data sets. Entropy function is used to select best feature for prediction in data selection process. Unique feature for semester performance prediction signifies the best in our approach. The details of entropy measures are discussed below

5.1.2.1 Entropy Function

Entropy measure is used to select certain feature or factors of respective students in this prediction model. Entropy value of each factors associated with students is calculated with respective to actual class values. The Shannon entropy function is given in Eqn. 5.1 by (Kosko, 1986b) as

$$H = \sum P \log p$$
(5.1)
Where, p is probability value.

5.2 Prediction Model Based on Lion-Wolf Neural Network

The prediction of student performance is one of the recent trending area and the problem construction for student performance prediction model is reflected below. The utmost intention of proposed student performance prediction model is to predict semester marks of separate students based on data collected from various influencing factors related to students.

Based on family, schooling, environmental and individuality factors X_i is the data collected from students S. The collected data is represented by $X_i = \{X_1, X_2\ldots, X_N\}$. The output of proposed NN based prediction model is

representing the marks of students in each semester given by SEM1, SEM2, SEM3, SEM4, SEM5, SEM6, SEM7 and SEM8.

Factors	Attributes	Values
Individual Factors	Gender	M/F
	Mobile	Yes, No
	Sports Interest	Yes, No
	Computer	Yes, No
Schooling Factors	Prev_Year %	12th %
	CGPA	Internal %
	SSLC/HSC Score	Subject scores
	Attendance %	Attend %
	Recent Result	Pass, Fail
	Type of School	Government, private
	Medium of Instruction	English, Regional
Environmental Factors	Location	Village, Town
	Neighbour	Good, Bad
Family Factors	Caste	SC/OBC/ GEN
	Parental Status	Yes, No
	Fathers/Mothers Education	Degree, Uneducated
	Fathers/Mothers Occupation	Govern/ Self
	Parents guidance	Yes, No

Table 5.1 List of Influencing Factors and its possible values

5.2.1 Implementation Factors

The input variables are selected based on four varying factors connected to student as listed in Table 5.1. The output variables represent the performance of a student is listed in Table 5.2.

Prediction O/P Variables	Expected % of Marks
SEM 1	
SEM 2	
SEM 3	
SEM 4	
SEM 5	
SEM 6	
SEM 7	
SEM 8	

Table 5.2 List of Output Variables

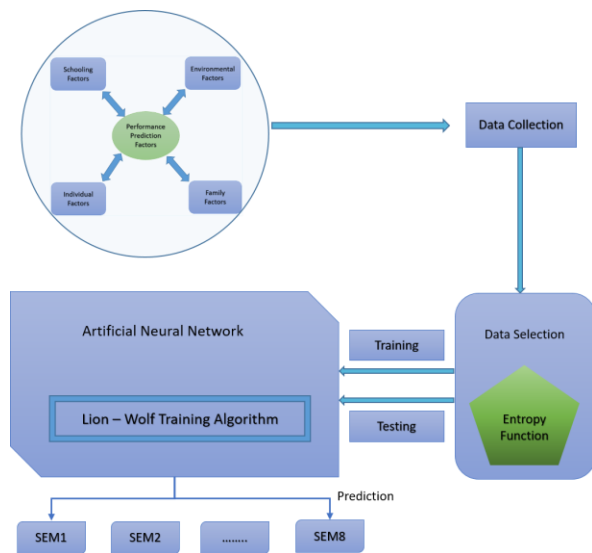The proposed student performance prediction model architecture is given in Fig. 5.2.

Figure 5.2. Lion – Wolf Prediction Model

5.2.2 Lion-Wolf Optimization Algorithm

The proposed Lion-Wolf optimization algorithm for neural network training is described in this section. Lion-Wolf optimization algorithm is developed by integrating Grey Wolf optimizer [7] (Mirjalili et al., 2014) and Lion optimization algorithm [8] (Rajakumar, 2014). The proposed optimization algorithm is majorly based on grey wolf optimizer, but the position updating of grey wolf optimizer is hybridized with lion optimization algorithm increasing convergence rate and so avoiding the local optimal problem. Grey wolf optimizer is a meta heuristic optimization algorithm which is based on leader ship behaviors of grey wolf. The hunting behavior of grey wolf is used for optimization.

VI. PERFORMANCE EVALUATION

The performance measures considered for performance evaluation are Mean squared Error and Root Mean Squared Error. The proposed student performance prediction model based on Neural Network is validated over comparative method by using MSE and RMSE. There are three following factors based on which performance evaluation is done for this student performance prediction model Varying Percentage of Training Data and Varying Number of Features

6.1 Study Based On Varying Percentage of Training Data

From the analysis curve shown in Fig 6.1, it is clear that proposed LION-WOLF perform well for different percentage of training data, thereby helping in enhanced performance prediction of students.
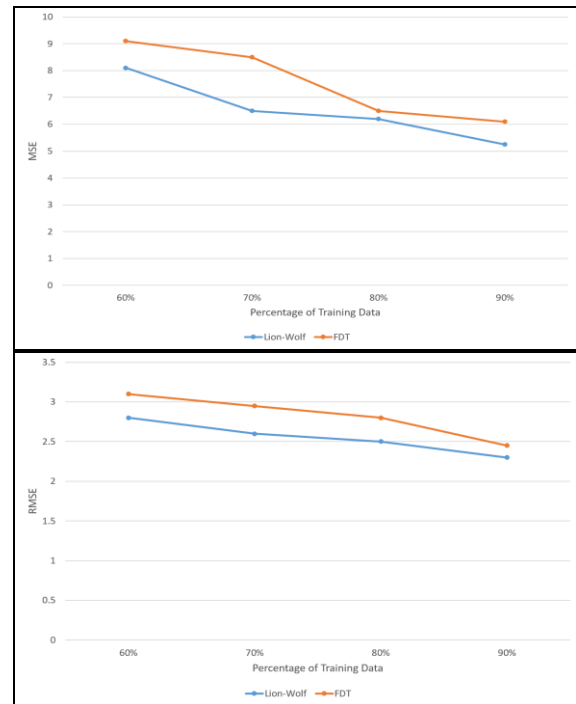

Fig. 6.1 Analysis Curve for Varying Percentage of Training Data.

6.2 Study Based On Varying Number of Features

The experimentation is performed with features size of 15, 20, 25, and 30 and corresponding MSE and RMSE values are validated for performance comparison over existing system. Fig. 6.2 shows the analysis curve for varying number of features.
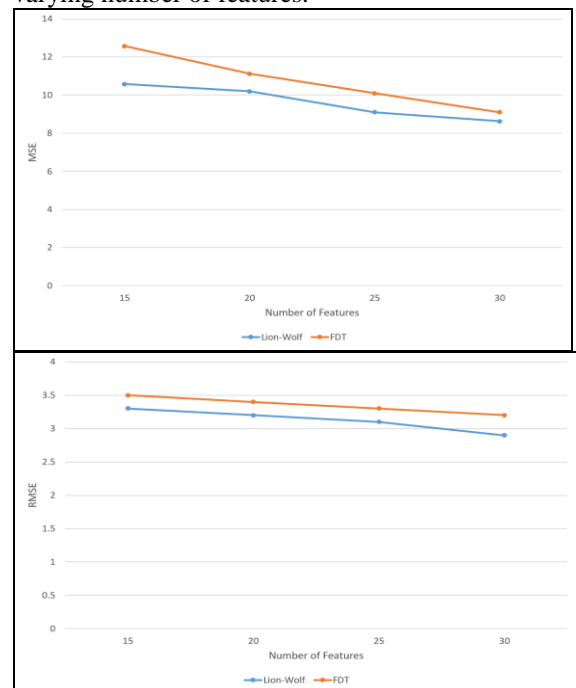

Fig. 6.2 Analysis Curve for Varying Number of Features.

VII.   CONCLUSION

The main objective of the proposed student performance prediction model is to predict semester marks of separate students based on data collected from various influencing factors related to students. The feature sets are given as input to neural network to predict the semester performance. In NN, optimal weight for prediction is selected using proposed Lion-Wolf training algorithm the selected features are given to the NN for prediction. From the collected data set, the best features for prediction are selected based on entropy function. Entropy measure is utilized to select relevant factors of student with good separability for prediction tasks. The selected features are given to the NN for prediction. In Neural Network Lion-Wolf optimization algorithm is used to select the optimal weight of neuron. The experimentation of proposed student performance prediction model was validated means of performance measures; MSE and RMSE. The proposed performance prediction model helps to identify the students with low academic performance initially, thereby providing guidance for underperforming students. The prediction results show that potential efficacy of proposed Lion-Wolf NN as a prediction tool was good compared to existing ANN based student performance prediction model with minimal MSE of 5.25 and minimal RMSE of 2.3.

REFERENCES

[1]. Ognjanovic, I., Gasevic, D. and Dawson, S. (2016), 'Using institutional data to predict student course selections in higher education', The Internet and Higher Education 29(1), 49–62.

[2]. Malvandi, S. and Farahi, A. (2015), 'Provide a method for increasing the efficiency of learning management systems using educational data mining', Indian Journal of Science and Technology 8(28), 1–8.

[3]. Shetgaonkar, P. R. (2015), 'Predicting the impact of different variables on students' academic performance using artificial intelligence', International Journal of Computer Science and Information Technologies 6(2), 1367–1370.

[4]. Wolff, A., Zdrahal, Z., Herrmannova, D. and Knoth, P. (2014), Predicting student performance from combined data sources, in 'Educational Data Mining', pp. 175–202.

[5]. Guo, B., Zhang, R., Xu, G., Shi, C. and Yang, L. (2015), Predicting students' performance in educational data mining, in 'Proceedings of IEEE International Symposium on Education technology', pp. 125–128.

[6]. Guar´ın, C. E. L., Guzm´an, E. L. and Gonz´alez, F. A., 'A model to predict low academic performance at a specific enrollment using data mining', IEEE Revista Iberoamericana de Tecnologias del Aprendizaje 10(3), 119–125, (2015).

[7]. Mirjalili, S., Mirjalili, S. M. and Lewis, A. (2014), 'Grey wolf optimizer', Advances in Engineering Software 69(1), 46–61.

[8]. Rajakumar, B. (2014), Lion algorithm for standard and large scale bilinear system identification: A global optimization based on lion's social behavior, in 'IEEE Congress on Evolutionary Computation', pp. 2116–2123