Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia - Social and Behavioral Sciences 55 (2012) 409 – 418

Procedia
Social and Behavioral Sciences

INTERNATIONAL CONFERENCE ON NEW HORIZONS IN EDUCATION
INTE2012

A Genetic-Fuzzy Based Mathematical Model to Evaluate The Distance Education Students' Academic Performance

Osman Yıldız[a], Abdullah Bal[b], Sevinç Gülseçen[c], Fulya Damla Kentli[a]

[a] Informatics Department,Yildiz Technical University,Istanbul,Turkey
[b] Electronics and Communications Engineering Department, Yildiz Technical University, Istanbul,Turkey
[c] Informatics Department, Istanbul University, Istanbul,Turkey

Abstract

In distance education systems, it is very important to predict academic performance for both instructors and students during the course of the semester. If an instructor can properly assess and predict student performance early at the beginning of the semester, then the instructor can take action and arrange both the course content and the teaching style. This, in turn, contributes greatly to the success of students. In order to make such a prediction, constructing mathematical models is one of the most effective and efficient methods. Among many approaches, fuzzy logic-based models have the most appropriate topology. In this study, fuzzy logic model is used to model data of distance education and predict students' academic performances. In order to increase the success of fuzzy logic model, fuzzy membership functions are optimized by using genetic algorithms. As distance education data, when students enrolled in learning management system, how frequently they log on, and how long they stay online are used. By using this model and data of a 6 week-long study, students' success level at the end of the semester is predicted and the results are compared with the ground truth data.

Keywords: Fuzzy Logic, Genetic Algorithm, Distance Education, Academic Performance Evaluation

doi:10.1016/j.sbspro.2012.09.519

1. Introduction

Over the past several years, the use of distance education systems has grown rapidly, since they have more advantages than traditional education; an Internet connection is just enough to access the education system and there is no time restriction to attend the courses. In addition to these advantages, it is relatively cheap compared to traditional education.

Evaluation of student academic performance is one of the most important parts in educational systems. Instructors monitor students' learning processes and analyze their performance based on paper records and observation in traditional education. Distance education systems provide definite opportunities for instructors to observe students' academic performance. In order to establish this task, instructors encounter difficulties in observing academic performance during distance education. Therefore, weblogs stored in Learning Management Systems (LMS) could be helpful for observing and analyzing learners' academic performance, as well as predicting final marks. It is required to monitor and analyze students' activities in distance education systems by suitable scientific research methods. The ability to predict students' performance could be useful for instructors who can take a precaution failure or prevent students from dropping out.

In this study, we have incorporated three types of web logs: recency, frequency, and monetary. Recency is enrollment time of students that shows the number of days after the lectures have been uploaded to LMS. Frequency is defined as how frequently they logged on. How long they stayed online is regarded as Monetary. These data are incorporated into Fuzzy to obtain a prediction model. In order to increase the success of fuzzy logic model, fuzzy membership functions are optimized by genetic algorithms. By this mathematical model, we have tried to predict class grades of the students using only six weeks RFM data.

The rest of the paper is structured as follows: Section 2 Related Works presents the findings of related research literature. In Section 3, we briefly review theory of the Fuzzy Logic and Genetic Algorithm. In Section 4, we present classical fuzzy model and Genetic-Fuzzy Model to evaluate academic performance of students, and the experimental results are discussed. Finally, conclusion and further research are reported in Section 5.

2. Related Works

Several studies focus on the field of analyzing student performance in distance learning. Dimitris and Christos (2006) have employed genetic algorithms and decision trees to estimate academic performance of distance learning students'. Zafra and Ventura (2009) have incorporated multiple instance genetic algorithms to predict whether the students will fail or pass for a certain course. This prediction has been based on students' activities such as quizzes, assignments, and forums. Lykourentzou et al (2009) have performed a student achievement prediction method applied to 10-week introductory level e-learning. Vandamme, et al. (2007) have studied a model by means of neural networks procedure in SAS/Enterprise mining. They have categorized students into three groups as "the low-risk", "medium-risk" and "high-risk" who have a high probability of failing. Kotsiantis, et al. (2004) have conducted a supervised machine learning algorithm in which the training set was comprised of students' key demographic characteristics and their mark on a few written assignments. Ibrahim and Rusli (2007) have used neural network, decision tree and linear regression to estimate students' academic performance. In this work, they have employed demographic profile and students' first semester cumulative grade point averages (CGPA) to predict final CGPA.

In the literature, prediction studies involving Fuzzy Logic method are generally carried out associating with Artificial Neural Network (ANN). Taylan and Karagozoglu (2009) have presented a study that introduces a systematic approach to design of a fuzzy inference system based on a class of neural networks to assess the students' academic performance. The development method used a fuzzy system augmented by neural networks to enhance some of its characteristics like flexibility, speed, and adaptability, which is called the Adaptive Neuro-Fuzzy Inference System (ANFIS). Yusof, et al. (2009) have built an evaluation of student's performance and learning efficiency based on ANFIS, too. In that study, neural network and fuzzy have been used for predicting student's performance based on four criteria which are scores earned, time spent, number of attempts and help needed.

3. Fuzzy Logic and Genetic Algorithm

3.1 Fuzzy Logic

Fuzzy sets and fuzzy logic have been considered an effective tool to deal with uncertainties in terms of vagueness, ignorance, and imprecision (Cho, Cho, & Wang, 1997). While making an operation in a set, a 'b' element is a member of a set or not. On the other hand, when making an operation with fuzzy, 'b' element can be a member of two sets at the same time. Fuzzy Logic is a form of three main stages; fuzzification, rule evaluation and defuzzification. Fuzzification is the first step that transforms the crisp inputs into degrees of match with linguistic values. Rule evaluation is where knowledge derived from experts are formed which is then called fuzzy rules (Yusof, Zin, Yassin, & Samsuri,2009). Defuzzification transposes the fuzzy outputs to crisp values. Figure1 shows the stages of Fuzzy Logic.

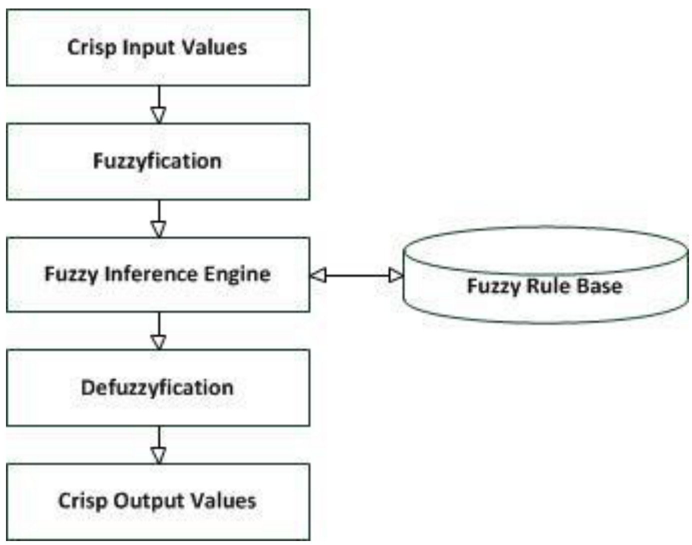


Fig. 1 Fuzzy Logic Stages

3.2 Genetic Algorithm

The Genetic Algorithm (GA) is an optimization and search technique based on the principles of genetics and natural selection (Haupt & Haupt, 2004). A GA works on a population of randomly generated input solutions symbolized by chromosomes that are often represented by binary strings. The population improves toward better solutions by applying genetic operators, such as crossover and mutation. In each generation, favorable solutions generate offspring that replace the inferior individuals. Crossover hybridizes the genes of two parent chromosomes in order to exploit the search space and constitutes the main genetic operator in GAs. The mutation is operated to provide the diversity of gene pool. An evaluation or fitness function plays the role to decide for the good or bad solutions (Cordon, Herrera, Hoffmann, & Magdalena, 2001). The major components and the principal structure of GA are shown in Figure 2.

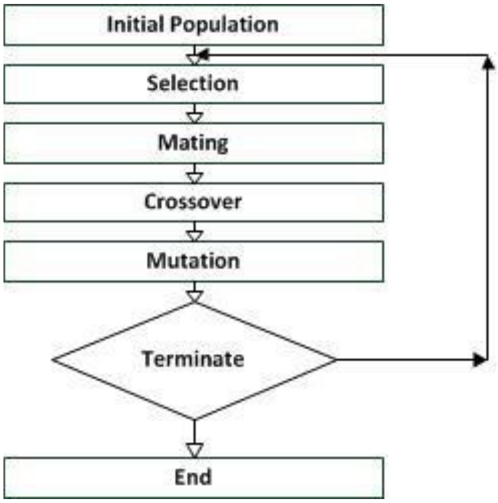


Fig. 2 Principal Structure of GA

4. Method and Results

In this study, the informatics course of the Yildiz Technical University (YTU), Matlab course is delivered in web-based education during 2010/2011 spring semester for Chemistry students.

Moodle (Modular object oriented developmental learning environment), a free learning LMS is utilized for this course. Moodle offers many activities that can be combined with learning material. It logs every activity report for each student that indicates the activities such as operation on course materials, quizzes, and messages for different times and days. This fact enables the authors to collect data only from the tutors involved in this learning process.

The Moodle records for Matlab course have supplied the data for the Fuzzy-RFM model. Matlab course is composed of 12 weeks and leads to an elective course. The total number of registered students is 62, 46 of whom, are selected and the model is formed by entering RFM values for the first six weeks. As the desired output, grade scores of 46 students are selected and introduced to model.

We have first used Classical Fuzzy Model to predict pass grades of the students. Then, we have improved the model by employing Genetic Optimization Algorithm to obtain higher prediction accuracy.

4.1 Classic Fuzzy Model

In this model, Fuzzy Memberships are organized for three input data sets; Recency, Frequency and Monetary. Academic Performance (AP) is used as output value of the model. Input values are determined as low (R1), medium (R2) and high (R3) for Recency, low (F1), medium (F2) and high (F3) for Frequency, low (M1), medium (M2) and high (M3) for Monetary. The output values are determined to be low (AP1), medium (AP2) and high (AP3). Figure 3 shows Classic Fuzzy Model.

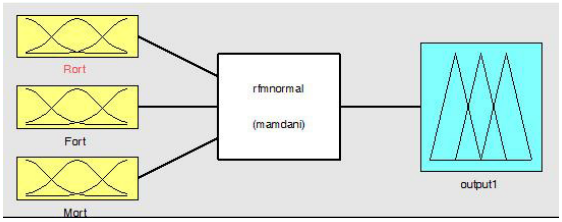


Fig. 3 Fuzzy RFM Model.

Input values R, F and M are shown in Figures 4, 5, and 6, respectively.

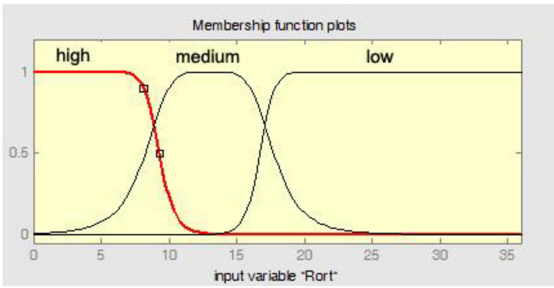


Fig. 4 Recency Membership Function Plots

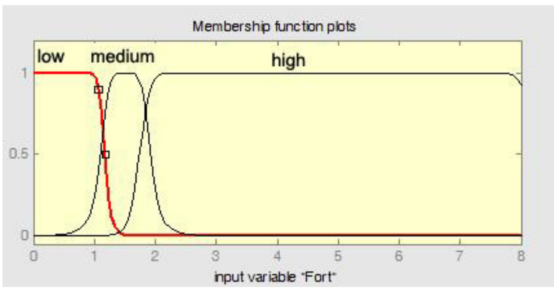


Fig. 5 Frequency Membership Function Plots

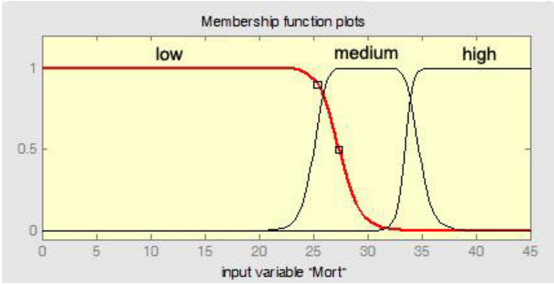


Fig. 6 Monetary Membership Function Plots

After determining Membership Functions, 27 rules are constructed according to expert opinion. For example, IF R (Recency) is low and F (Frequency) is low and M (Monetary) is low THEN AP (Academic Performance) is low.

The obtained results from classic fuzzy can be evaluated in two stages.

4.1.1 Determining Intervals

Expert opinion is the most vital factor in determining intervals and forming rules while designing the model with RFM data obtained from learning management system of students in distance education. Low, medium and high intervals for each value in RFM variables are formed. For instance, is Recency indicates to how long ago the student visit each course week after the course entered to the system. Firstly, intervals are equally installed. After this installation, the ratio of accuracy of students' academic performance has been found approximately %65. Then, we consult the expert to determine the intervals and the model is reformed. According to the expert, if the student enters the system in 1-7 days, it is high, 7-14 days, medium, and 14-36 days low. On the other hand, trial-error method is used in determining frequency that shows to how often each student stays in each course week. The ratio of accuracy between the alteration intervals according to the expert and prediction of students' academic performance has risen to %74.

4.1.2 Determining Membership Functions

After the determination of intervals in which membership function the model has got high value is determined by 4-type MF. Table 1 presents the results of the model formed before consulting expert opinion that show trimf function gives better result compared to other functions.

Table 1 Accuracy for different type of MF types.

Membership Function	**Accuracy (%)**
Trimf	65,25
Gauss2mf	64,81
Pimf	64,62
Gbellmf	64,28

Table 2 shows the results of models formed by varied MFs according to intervals formed by experts' opinions. It is observed that Gbellmf function gives better results than others.

Table 2 Accuracy for different type of MF types

Membership Function	**Accuracy (%)**
Gbellmf	74,40
Gauss2mf	73,40
Pimf	73,29
Trimf	72,90

4.2 Genetic-Fuzzy Model

Fuzzy Logic rules and membership functions are formed according to expert opinion, which is shown as one of the lack of the fuzzy logic. Therefore, fuzzy membership functions' intervals have been optimized by genetic algorithm in this section. Triangular-shaped (trimf) built-in membership function has been used. Figure 7 shows the membership function. There are 3 trimf shaped for all input variables that are R, F, M whose axis of vertex points are a,b,c, respectively. The other points of triangle are on the x-axis. a1 ve a2, b1 and b2, c1 and c2 have indicated the distance from a, b, c.

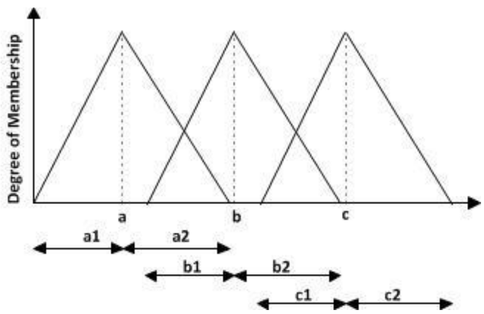


Fig. 7 Fuzzy Membership Intervals

The genetic algorithm works with chromosomes. The chromosome has 27 variables. The chromosome is written as row vector.

Chromosome=[Ra1, Ra, Ra2, Rb1, Rb, Rb2, Rc1, Rc, Rc2, Ra1, Fa, Fa2, Fb1, Fb, Fb2, Fc1, Fc, Fc2, Ma1, Ma, Ma2, Mb1, Mb, Mb2, Mc1, Mc, Mc2]

The genetic algorithm works with binary encodings. An example of binary encoded chromosome that has 27 variables, each encoded with 7 bits, is

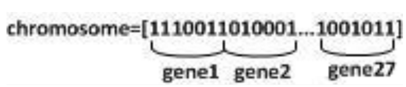


Figure 8 shows the shape of a member.

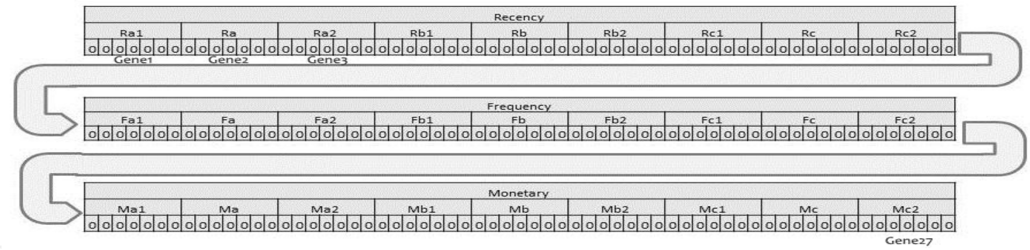
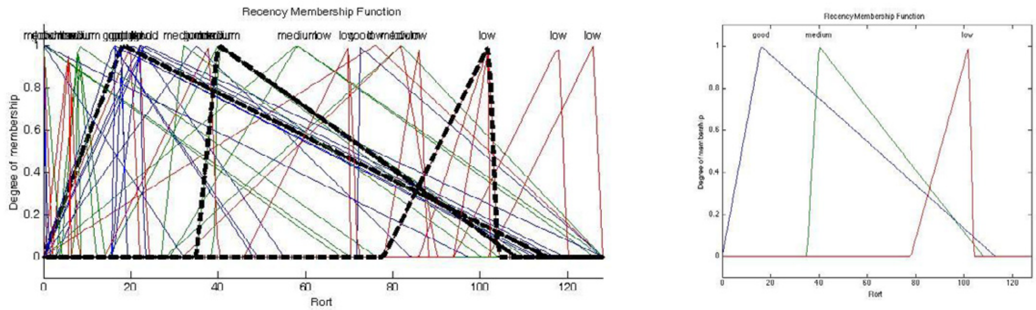


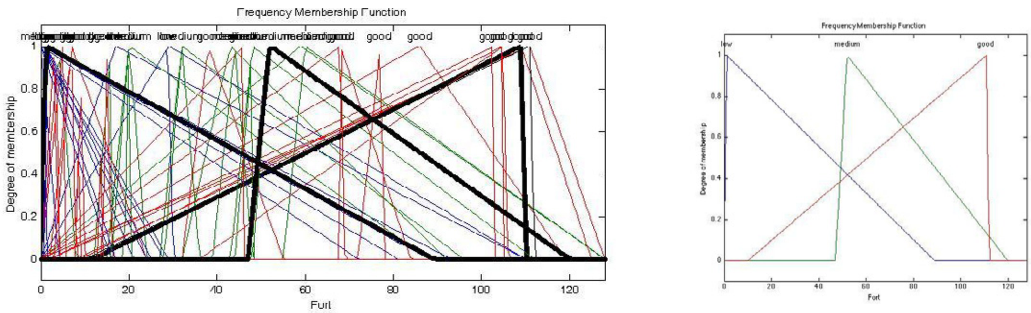
Fig. 8 Representation of Chromosome of Fuzzy Membership Function

The GA starts with 10 initial populations. After 85 iterations, the algorithm has stopped because costs are the same. Figure 9.(a),10.(a) and 11.(a) show optimization of Recency, Frequency, and Monetary Membership Function, respectively and Figure 9.(b),10.(b), and 11.(b) show optimized intervals of Recency, Frequency, and Monetary Membership Functions.



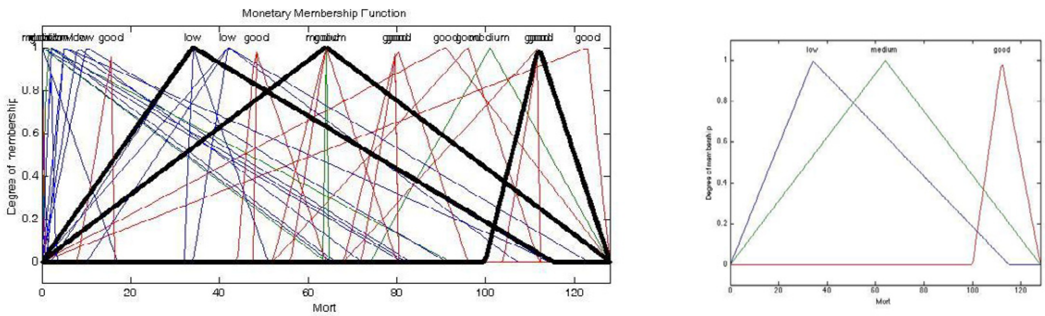
(a) (b)

Fig. 9. (a) Optimization of Recency Membership Function and (b) Optimized Membership Function



(a) (b)

Fig. 10. (a) Optimization of Frequency Membership Function and (b) Optimized Membership Function



(a) (b)

Fig. 11. (a) Optimization of Monetary Membership Function and (b) Optimized Membership Function.

After the optimization of intervals of membership functions, accuracy of prediction for distance education students' academic performance has risen to 84.52%.

5. Conclusion

To sum up, a new model, Genetic-Fuzzy Based Mathematical Model, has been formed to predict students' performances in distance education. Academic Performance have been attempted to predict just using 6-weeks data. Fuzzy membership intervals have been determined by genetic algorithm based optimization. Firstly, eight different models have been formed by using varied MF for the intervals formed in consultation with expert opinion. The best ratio of accuracy is 74.40 percent among formed models and it is an acceptable value. Then, using genetic algorithm, the optimal intervals for R, F, M were determined and prediction of students' academic performance has risen to 84.52%. In further studies, the optimization of rules used in the fuzzy logic system will be considered.

Acknowledgements

This research has been supported by Yıldız Technical University Scientific Research Projects Coordination Department. Project Number: 2010-16-03-KAP01

References

Yusof, N., Zin, N. A., Yassin, N. M., & Samsuri, P. (2009). Evaluation of Student's Performance and Learning Efficiency based on ANFIS. *2009 International Conference of Soft Computing and Pattern Recognition* (pp. 460-465). IEEE Computer Society.

Yusof, N., Zin, N. A., Yassin, N. M., & Samsuri, P. (2009). Evaluation of Student's Performance and Learning Efficiency Based on ANFIS. *International Conference of Soft Computing and Pattern Recognition* (pp. 460-465). Malacca: Soft Computing and Pattern Recognition.

Vandamme, J. -P., Meskens, N., & Superby, J. -F. (2007, 12). Predicting Academic Performance by Data Mining Methods . *Education Economics* , pp. 405-419.

Zafra, A., & Ventura, S. (2009). Predicting Student Grades in Learning Management Systems with Multiple Instance Genetic Programming. *Educational Data Mining 2009* , 307-314.

Chang, H.-C. (2010). Developing EL-RFM Model for Quantification Learner's Learning Behavior in Distance Education. *2nd International Conference on Educational Technology and Computer*, (pp. 452-454). Shanghai .

Cho, H.-J., Cho, K.-B., & Wang, B.-H. (1997). Fuzzy-PID Hybrid Control: Automatic Rule Generation Using Genetic Algorithm. *Elsevier Fuzzy Sets and Systems* , 305-316.

Cordon, O., Herrera, F., Hoffmann, F., & Magdalena, L. (2001). *Genetic Fuzzy Systems.* Danvers: World Scientific.

Geyik, A. K. (2007). Clustering- E-students in a Marketing Context: A Two-Stage Clustering Approach. *The 6th European Conference on e-Learning.* Copenhagen.

Hughes, A. M. (1994). *Strategic Database Marketing.* McGraw-Hill.

Haupt, R. L., & Haupt, S. (2004). *Practical Genetic Algorithms.* New Jersey: John Wiley & Sons.

Ibrahim, Z., & Rusli, D. (2007). Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision Tree And Linear Regressıon. *21st Annual SAS Malaysia Forum*, (pp. 1-6). Kuala Lumpur.

Kalles, D., & Pierrakeas, C. (2006). Analyzing Student Performance in Distance Learning with Genetic Algorithms and Decision Trees. *Applied Artificial Intelligence* , 655-674.

Kotsiantis, S. B., & Pintelas, P. E. (2005). Predicting Students' Marks in Hellenic Open University. *Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)* (pp. 664-668). Taiwan: IEEE Computer Society.

Kotsiantis, S., Pierrakeas, C., I.Zaharakis, & Pintelas, P. (2003). Efficiency of Machine Learning Techniques in Predicting Students' Performance in Distance Learning Systems. In *Recent advances in mechanics and the related fields* (pp. 297-306). University of patras.

Lykourentzou, I., Giannoukos, I., Mpardis, G., Nikolopoulos, V., & Loumos, V. (2009). Early and dynamic student achievement prediction in e-learning courses using neural networks. *JASIST* , 372-380.

Newell, F. (1997). *The New Rules of Marketing: How to use one-to-one relationship marketing to be the leader in your industry.* Newyork: McGraw-Hill Company.

Taylan, O., & Karagozoglu, B. (2009). An adaptive neuro-fuzzy model for prediction of student's academic performance. *Computers & Industrial Engineering* , 732-741.