

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

Procedia Social and Behavioral Sciences 8 (2010) 299–306

---

---

**Procedia**  
Social and Behavioral Sciences

---

---

International Conference on Mathematics Education Research 2010 (ICMER 2010)

## Predicting Pre-university Students' Mathematics Achievement

Chun-Teck Lye<sup>a,\*</sup>, Lik-Neo Ng<sup>b</sup>, Mohd Daud Hassan<sup>c</sup>, Wei-Wei Goh<sup>d</sup>,  
Check-Yee Law<sup>e</sup>, Noradzilah Ismail<sup>f</sup><sup>a,b,c,d,e,f</sup>Centre for Foundation Studies & Extension Education, Multimedia University, Malaysia.

---

### Abstract

This study exploits three methods, namely the Back-propagation Neural Network (BPNN), Classification and Regression Tree (CART), and Generalized Regression Neural Network (GRNN) in predicting the students' mathematics achievement. The first part of this study utilizes enrolment data to predict the students' mid-semester evaluation result, whereas the latter part employs additional data to predict the students' final examination result. The predictive models' accuracy is evaluated using 10-fold cross-validation to identify the best model. The findings reveal that BPNN outperforms other models with an accuracy of 66.67% and 71.11% in predicting the mid-semester evaluation result and the final examination result respectively.

© 2010 Elsevier Ltd. Open access under [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/3.0/).**Keywords:** Decision Tree; Back-propagation Neural Network; Generalized Regression Neural Network; Mathematics; Pre-university

---

### 1. Introduction

A prediction model that is effective in identifying weak students and potential failures particularly in the initial phase, ahead of the final assessment provides opportunity for necessary remedial measures. The model can operate as an alert system in detecting at risk students and it is essential in improving and sustaining the students' achievement throughout their studies to reduce the dropout rates in universities. In addition, the model can work as an advisory reference for teachers/lecturers in preparing the course teaching and learning materials, it can be the alternative for placement examinations as well. All these applications not only support the education institutes in delivering a better quality education and producing excellence students, but also beneficial to the administrative tasks of the institutes. Motivated by these reasons, there have been numerous studies done in constructing predictive models in education for various purposes. Among them are Kotsiantis, Pierrakeas and Pintelas (2004) who employed six common machine learning techniques: Decision Trees, Artificial Neural Networks, Naïve Bayes classifier, Instance-Based Learning Algorithms, Logistic Regression and Support Vector Machines for the prediction of students' dropout at the Hellenic Open University. The models were trained using demographic data during the initial predictions with accuracy reaches 62%. The accuracy exceeds 82% after more data such as face-to-face meetings and written assignments are included for the subsequent predictions before the final examinations. Among all the machine learning techniques covered, the study also demonstrated that the Naïve Bayes algorithm is the most appropriate to be used for the construction of a software support tool.

---

\* Corresponding author.

E-mail address: [ctlye@mmu.edu.my](mailto:ctlye@mmu.edu.my)

Cripps (1996) developed and trained different Artificial Neural Networks to predict degree program completion, earned hours and GPA for college students in the developmental studies program at Middle Tennessee State University. The feed forward neural network with back-propagation learning algorithms and logistic activation function were used. The finding revealed that age, gender, race, ACT scores, and reading level are significant in predicting the degree program completion, earned hours, and GPA. In another related research, Nghe, Janecek and Haddawy (2007) used admissions information to predict GPA at the end of the first year to compare the accuracy of Decision Tree and Bayesian Network algorithms for predicting the academic performance of undergraduate and postgraduate students at two academic institutes: Can Tho University (CTU) and Asian Institute of Technology (AIT). The two models were able to achieve nearly the same levels of accuracy for predicting student performance: 73% for classification {fail, fair, good, very good} and 94% for nominal {fail, pass} at CTU, whereas the accuracy is 71% for classification and 93% for nominal at AIT. The finding of the study also showed that Decision Tree was consistently 3-12% more accurate than the Bayesian Network. Ibrahim and Rusli, D. (2007) used the students' demographic profile and the CGPA for the first semester of the undergraduate degree program to build three predictive models: Neural Network, Decision Tree and Linear Regression. The outcome of the study disclosed that all the three models generate more than 80% accuracy, in which the artificial neural network outperforms the other two models in predicting the final CGPA of the students upon graduation. Sheel, Vrooman, Renner and Dawsey (2001) applied Neural Networks and Classical Discriminant Analysis in their study to find an alternative for the mathematics placement examination at the Coastal Carolina University. By using the overall high school GPA, the SAT mathematics score, and the final grade in Algebra II as input, the prediction results showed that the Neural Networks (average of 81.4%) outperformed the classical discriminant analysis (average of 71.2%) in correctly predicting the recommended mathematics placement.

In another study, Vandamme, Superby and Meskens (2007) employed Decision Trees, Neural Networks and Linear Discriminant Analysis to identify the risk of three categories (low, medium and high) of students at the early stage of study based on the demographics and academic history of the first-year students in French-speaking Belgian universities. However, the predictive results produced by all the three models are not convincing, in which the best result was obtained with Discriminant Analysis with the overall correct classification rate of 57.35%. Oladokun, Adebajo and Charles-Obawa (2008) employed an artificial neural network for predicting the likely performance of a candidate being considered for admission into the Nigerian Universities. The neural network model which based on the multilayer perceptron topology was developed and trained using data such as ordinary level subjects' scores and subjects' combination, matriculation examination scores, age on admission, parental background, types and location of secondary school attended and gender. The test outcome shows that the model is capable to correctly predict the performance of more than 70% of the prospective students. In some recent researches, Affendy, Paris, Mustapha, Sulaiman and Muda (2010) conducted a study to rank influencing factors that contribute to the prediction of students' academic performance. They employed WEKA open source data mining tool such as Bayes, Trees and Functions to analyze attributes for predicting a higher learning institution's bachelor of computer science students' academic performance. The result disclosed that Naïve Bayes, AODE and RBFN performed best with an accuracy of 95.29%. Based on data mining techniques such as Feature Selection, CHAID Tree and Classification and Regression Tree, Kovacic (2010) used socio-demographic variables and study environment to explore the efficacy of enrolment data in pre-identifying successful and unsuccessful students at the Open Polytechnic of New Zealand. His finding showed that Classification and Regression Tree was the best model with an overall percentage of correct classification of 60.5%.

Three methods are exploited in this study, namely the Back-propagation Neural Network (BPNN), Classification and Regression Tree (CART), and Generalized Regression Neural Network (GRNN) in predicting the mathematics achievement of pre-university students. The focal point of this study is on the mathematics in view of the fact that mathematics has always been one of the core subjects in higher educational institutions due to its wide range of applications. This paper also focuses on the foundation study because the pre-university studies act as the foundation and bridge for the students to progress from the secondary/high school education to the higher learning education. More specifically, the first part of this study utilizes the enrolment data available during admission to build models for early prediction of study outcome preceding the mid-semester evaluation, whereas the second part employs some additional data attainable at certain stage during the study to build the predictive models prior to the final assessment. Moreover, the accuracy of these predictive models is evaluated using 10-fold cross-validation to identify the two models (one from each part) that are superior in predicting the mathematics achievement. This study

anticipates the best model to function as an alert tool in detecting at risk students or as an advisory reference for students (in study enhancement) and lecturers (in preparing teaching materials and setting question papers). The remaining of the paper is organized as follows. Section 2 briefly describes the data and methodology applied in this study. Section 3 discloses and discusses the findings. Finally, section 4 concludes the paper and suggests some potential enhancements intended for future research directions.

## 2. Data and Methodology

The importance of foundation studies and mathematics to the survival and success of students in higher education particularly in engineering related specialization is indisputable. Therefore, this study use the admission data and records of students' attendance, midterm and final evaluation result of Algebra and Trigonometry as the variables for this study. The data consist of information of 180 students who enrolled in the foundation in engineering, Multimedia University, during the first semester of their study. The variables and their description are presented in Table 1. The rationale behind the selection of these variables is based on the findings of some preceding studies. The study done by Kotsiantis, Pierrakeas and Pintelas (2004) has shown that prediction models that utilized demographic data alone is inadequate in predicting of students' academic performance. As for other related studies that make use of extra variables such as prior academic performance, proficiency in language and parental background, the prediction models yielded better prediction performance (Cripps, 1996; Oladokun, Adebajo and Charles-Obawa, 2008; Sheel, Vrooman, Renner and Dawsey, 2001; Vandamme, Superby and Meskens, 2007). Besides that, it is shown that the predictive ability of a model can be enhanced by including some additional attributes and recent data (Kotsiantis, Pierrakeas and Pintelas, 2004).

Table 1. The description of variables

<b>Data available during enrolment</b>		
<b>No.</b>	<b>Variable</b>	<b>Description</b>
1	Age	Students' age during enrolment
2	Gender	{Male, Female}
3	Ethnicity	{Malay, Chinese, Indian, Others}
4	Self-interest in Mathematics	{Yes, No}
5	Self-belief in Mathematics	{Very weak, Weak, Average, Good, Excellence}
6	Father's education level	{Tertiary, Secondary, Primary or below}
7	Mother's education level	{Tertiary, Secondary, Primary or below}
8	PMR English result	{A, B, C, D, E}
9	PMR Mathematics result	{A, B, C, D, E}
10	SPM English result	{A1, A2, B3, B4, C5, C6, D7, E8, G9}
11	SPM Modern Mathematics result	{A1, A2, B3, B4, C5, C6, D7, E8, G9}
12	SPM Additional Mathematics result	{A1, A2, B3, B4, C5, C6, D7, E8, G9}
<b>Data available at mid-semester and at the end of semester</b>		
<b>No.</b>	<b>Variable</b>	<b>Description</b>
13	Class attendance	Students' lecture class attendance (in percentage)
14	Lecturer's teaching experience	Lecturers' teaching experience (in year)
15	Midterm evaluation result	Students' mid-semester test results in Algebra and Trigonometry. Grade={A, B, C, D, F}
16	Final examination result	Students' final examination results in Algebra and Trigonometry. Grade={A, B, C, D, F}

This study is divided into two parts. The first part of the study constructs the three prediction models, i.e. BPNN, GRNN and CART by using the twelve variables (variable no.1 – no.12) to predict the midterm evaluation result (variable no.15) before the semester begins or at the beginning of the semester. These models can be employed as an

advisory reference for lecturers in preparing more suitable teaching materials with the appropriate level of proficiency based on the students’ predicted results and proficiency level. These models can also help the lecturers to identify the at risk students prior to the actual midterm evaluation, in which these students may need extra attention and assistance in their study. While for other students, the predicted outcomes can be used as a reference to motivate the students to improve and excel throughout the study. The second part of the study, the models are built by utilizing fifteen variables (variable no.1 – no.15) to predict the final examination result (variable no.16) after the mid-semester evaluation. At this stage, the prediction accuracy of the models at the first part is reviewed and improved by incorporating extra variables. Subsequently, these refined models can function as an alert system that identify the potential failures and provide opportunity for remedial actions to be carried out at the early stage. The predicted performance can further provide a clearer outlook for the students concerning their final examination result. As the accuracy of the predictive models is evaluated using 10-fold cross-validation to identify the best model in predicting pre-university students’ mathematics achievement, therefore a total of 30 models (each BPNN, CART and GRNN method has 10 models) are built in the first part and another 30 models are built in the second part. All data used in this study are normalized in order to perform a faster training, in which 90% (162 out of 180) of the normalized data are used for training, where as the remaining 10% of the data are used for out-of-sample tests.

Genetic algorithms (Goldberg, 1989; Holland, 1975; Mitchell, 1998) are renowned efficient global optimization tools because of their simplicity, derivative-free, and inherent parallel processing capability. Therefore, this study employs genetic algorithms to determine the best topology for the three-layer (input-hidden-output) feed-forward BPNN (Fausett, 1994; Patterson, 1998), as shown in Figure 1, by setting the initial population size equal to the number of variables, number of generations equal to 50, and root mean squared error (rmse) as the fitness function:

$$rmse = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \tag{1}$$

where:  $\hat{y}_i$  is the predicted value, and  $y_i$  is the actual value, and  $n$  is the total number of data under consideration.

The optimal weights and biases obtained from the genetic algorithms optimization are subsequently utilized as the initial weights of the BPNN. The gradient descent with momentum back-propagation (Chauvin and Rumelhart, 1995) with learning rate 0.01 and momentum constant 0.90 is applied to conduct supervised learning of the network. This study uses the hyperbolic tangent sigmoid transfer function in the hidden layer of the network, whereas the output layer uses the linear transfer function. In the out-of-sample test, the predicted outputs obtained are used in the comparative study between models.

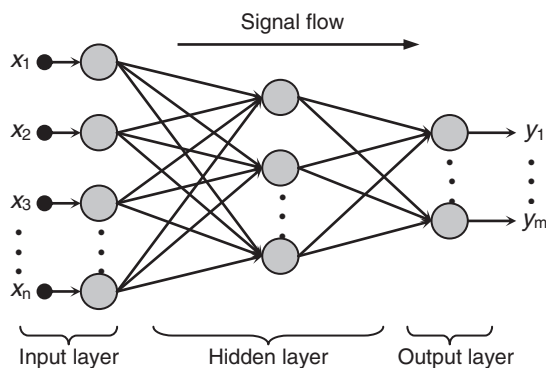


Figure 1: The standard architecture of a three-layer feed-forward network

The GRNN (Specht, 1991) has three layers: an input layer, a radial basis layer (Chen, Cowan and Grant, 1991), and a special linear output layer, and it is widely applied in function approximation studies. GRNN has a smoothing factor (or spread) where the larger the smoothing factor, the smoother the network function will be. In this study,

different smoothing factor are tested, in which the best smoothing factor is found incrementally [0, 10], with gradual step increment of 0.005. The best smoothing factor that yields the smallest rmse is then used in the out-of-sample test. As in the CART (Breiman, Friedman, Olshen and Stone, 1984), the algorithms search and split the possible variables and values into two most homogeneous parts (Figure 2). The process is repeated until the stopping criteria is fulfilled, where in this study, the aim of the CART algorithms is to achieve the best possible predictive accuracy. This study employed the re-substitution method to construct and search for the optimal tree. The best tree is subsequently use in the out-of-sample test.

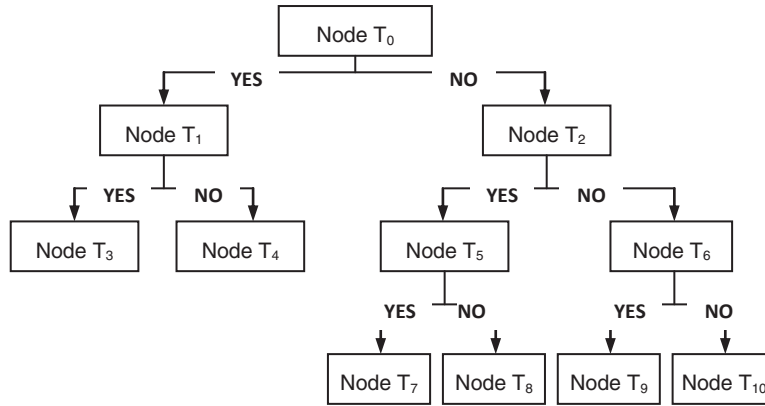


Figure 2: An example of a classification and regression tree

This study employs hit rate,  $h_k$  to evaluate the accuracy of the prediction models. If the absolute difference between the predicted value (rounded to the nearest integer) and the actual value of data  $i$  is defined as:

$$d_i = |\text{round}(\hat{y}_i) - y_i| \tag{2}$$

where:  $y_i = \{1, 2, 3, 4, 5\}$  represents the actual grade {F, D, C, B, A},  $\hat{y}_i$  (rounded to the nearest integer) is the predicted value. Then, the count of the predicted value that falls into respective category of  $d_i$  can be defined as:

$$\text{count, } c_{ki} = \begin{cases} c_{0i} = 1 & ; & d_i = 0 \\ c_{1i} = 1 & ; & d_i = 1 \\ c_{2i} = 1 & ; & d_i > 1 \end{cases} \tag{3}$$

Hence, the hit rate,  $h_k$  can be obtained by the following equation:

$$h_k = \frac{1}{n} \sum_{i=1}^n c_{ki} \times 100\% \quad ; \quad k = 0, 1, 2 \tag{4}$$

From Eq. (2), when  $d_i = 0$ , the predicted value (rounded to the nearest integer) is equal to the actual value, i.e. the predicted grade is the same as the actual grade; if  $d_i = 1$ , the predicted grade is at the difference of one-grade from the actual grade and if  $d_i > 1$ , the predicted grade is at the difference of two-grade or more from the actual grade. In another word, following the Eq. (3) and Eq. (4),  $h_0$  indicates the percentage of exact classification,  $h_1$  indicates the percentage of one-grade difference from the actual grade, and  $h_2$  indicates the percentage of two-grade or more differences from the actual grade. For instance, if the actual grade is C, the hit rate falls into  $h_0$  if the predicted grade is also C, or else if the predicted grade is B or D, which is one-grade different from the actual grade C, then the hit

rate will falls into  $h_1$ . Otherwise, the hit rate will falls into  $h_2$  if predicted grade is A or F, which is two-grade different or more from the actual grade (C). In addition, this paper also uses another universal performance measure in measuring prediction accuracy, which is the rmse or Eq. (1) to verify the performance of the models.

### 3. Findings

The outcomes of this study are partitioned into two tables: Table 2 and Table 3. As the performance of these predictive models is evaluated using 10-fold cross-validation, the table presents the average hit rate and root mean squared error of the BPNN, CART and GRNN models. In the first part of the study, the predictive results produce by all the three models (BPNN, CART and GRNN) shown in Table 2 are reasonable as preceding studies that employed practically the same attributes in their prediction models also yielded the same predicting accuracy, which is close to 66% (Kotsiantis, Pierrakeas and Pintelas, 2004; Kovacic, 2010). The outcomes shown in Table 2 also reveal that the BPNN outperforms other models in this study with the highest hit rate of 66.67%, as well as the lowest root mean squared error (0.7174). However, the cumulative hit rate ( $h_0+h_1$ ) of BPNN (91.67%) is lower than CART (93.89%) and GRNN (92.22%). This indicates that CART is better than other models in predicting the students' academic performance with the margin of error of one-grade difference. Moreover, CART also yielded the least hit rate of 6.11% in comparison to the other two models when considering the misclassification with two-grade or more differences, whereas BPNN produces the highest hit rate (8.33%) in such misclassification.

Table 2. The performance of the BPNN, CART and GRNN in predicting the midterm evaluation results

Model	$h_0$	$h_1$	$h_2$	rmse
BPNN	66.67	25.00	8.33	0.7174
CART	65.56	28.33	6.11	0.7195
GRNN	65.00	27.22	7.78	0.8051

The performance of the prediction models in the second part of study is summarized in Table 3. Additional variables (lecturers' teaching experience, students' class attendance and students' mid-semester evaluation results) are included into the models. The results disclose that the performance of BPNN (71.11%) is more superior in contrast to CART (66.11%) and GRNN (66.11%). The root mean squared error of the BPNN (0.4733) is the lowest among all the models under study as well. Furthermore, in the misclassification with one-grade difference, the BPNN has the least hit rate of 28.29% in comparison to other models which has a higher hit rate of 33.89%. As for the performance between CART and GRNN, although both models has the same hit rate, but the lower root mean squared error in Table 3 shows that the former is more superior. All models in the second part of the study did not have misclassification of two-grade or more throughout the prediction, i.e. the variance of the models in this part is smaller in comparison to the models constructed in the first part. Generally, the prediction accuracy of the models in the second part of the study (with extra variables), from the range of 66.11% to 71.11% is satisfactory as it is supported by some of the prior findings (Nghe, Janecek and Haddawy, 2007; Oladokun, Adebajo and Charles-Obawa, 2008; Sheel, Vrooman, Renner and Dawsey, 2001).

Table 3. The performance of the BPNN, CART and GRNN in predicting the final examination results

Model	$h_0$	$h_1$	$h_2$	rmse
BPNN	71.11	28.89	0	0.4733
CART	66.11	33.89	0	0.5022
GRNN	66.11	33.89	0	0.5477

All models show improvement in prediction accuracy in both performance measures when some latest data such as mid-semester evaluation and students' attendance were included. The accuracy of the models improve from the

range of 0.84% (CART) to 6.66% (BPNN) and 0.2173 (CART) to 0.2574 (GRNN) in the hit rate and rmse respectively. These improvements are consistent with the finding of Kotsiantis, Pierrakeas and Pintelas (2004). Furthermore, all predictive models show consistency in both parts of study in which the BPNN outperforms other models, whereas the CART is superior to GRNN.

#### 4. Conclusion

As the computing technologies and soft-computing skills develop, the data mining techniques and prediction accuracy of students' academic performance have improved extensively. Thus, this study uses the renowned BPNN, CART and GRNN to predict the mathematics achievement of pre-university students. By using the enrolment data, this study attempts to predict the students' mid-semester evaluation result at the early stage of their study. At a certain stage throughout the study, this study further collects some additional information and new prediction are made to predict the students' final examination result. The accuracy of the predictive models is evaluated via 10-fold cross-validation to discover the best model. The anticipation of this study is to build a model that is helpful in detecting at risk students, improving teaching and learning, and as a supporting tool for administrator. The findings reveal that the overall performance of BPNN outperforms the CART and GRNN models in predicting the mid-semester evaluation result with the highest hit rate of 66.67% and the lowest rmse of 0.7174, as well as in predicting the final examination result with the highest hit rate of 71.11% and the lowest rmse of 0.4733. Furthermore, the hit rates and rmse of all models also shown improvements from the range of 0.84% to 6.66%, and 0.2173 to 0.2574 respectively when recent variables are included in the predicting models. Although these results are satisfactory which the BPNN model is adequate to function as an advisory reference and alert tool in predicting the pre-university students' mathematics achievement, the model is anticipated to show improvement in its predictive accuracy with the inclusion of closely correlated attributes and a larger sample size. Consequently, the enhanced model can be implemented and it will definitely be beneficial to both lecturers and students.

#### References

- Affendy, L.S., Paris, I.H.M., Mustapha, N., Sulaiman, M.N. & Muda, Z. (2010). Ranking of influencing factors in predicting students' academic performance. *Information Technology Journal*, 9(4), 832-837.
- Breiman, L., Friedman, J.H., Olshen, R. & Stone, C.J. (1984). *Classification and regression trees*. Wadsworth.
- Chauvin. Y. & Rumelhart, D.E. (1995). *Back propagation: Theory, architectures, and applications*. Lawrence Erlbaum Associates. Hillsdale.
- Chen, S., Cowan, C.F.N. & Grant, P.M. (1991). Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on Neural Networks*, 2(2), 302-309.
- Cripps, A. (1996). Using artificial neural nets to predict academic performance. *Proceedings of the 1996 ACM symposium on Applied Computing*, 33-37.
- Fausett, L. (1994). *Fundamentals of neural networks: Architectures, algorithms, and applications*. Prentice Hall, New York.
- Goldberg, D.E. (1989). *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley.
- Holland, J.H. (1975). *Adaption in natural and artificial systems*. University of Michigan Press.
- Ibrahim, Z. & Rusli, D. (2007). Predicting students' academic performance: Comparing artificial neural network, decision tree and linear regression. Paper presented in the 21st Annual SAS Malaysia Forum, 5th September 2007, Shangri-La Hotel, Kuala Lumpur.
- Kotsiantis, S., Pierrakeas C. & Pintelas P. (2004). Predicting students' performance in distance learning using machine learning techniques. *Applied Artificial Intelligence*, 18(5), 411–426.
- Kovacic, Z.J. (2010). Early prediction of student success: Mining students enrolment data. *Proceedings of Informing Science & IT Education Conference (InSITE) 2010*.
- Mitchell, M. (1998). *An introduction to genetic algorithms*. The MIT Press.
- Nghe, N.T., Janecek, P. & Haddawy, P. (2007) A comparative analysis of techniques for predicting academic performance, *Proceeding of the 37<sup>th</sup> ASEE/IEEE Frontiers in Education Conference*, T2G, 7-12.

- Oladokun, V.O., Adebajo, A.T. and Charles-Obawa, O.E. (2008). Predicting students' academic performance using artificial neural network: A case study of an engineering course. *The Pacific Journal of Science and Technology*, 9(1), 72-79.
- Patterson, D.W. (1998). *Artificial neural networks: Theory and applications, 1<sup>st</sup> edition*. Prentice Hall.
- Sheel, S.J., Vrooman, D., Renner, R.S. & Dawsey, S.K. (2001). A comparison of neural networks and classical discriminant analysis in predicting students' mathematics placement examination scores. In *Computational Science-ICCS 2001*, Alexandrov, V.N., Dongarra, J., Juliano, B.A., Renner, R. & Kenneth Tan, C.J. (Eds.), Springer.
- Specht, D.F. (1991). A generalized regression neural network. *IEEE Transactions on Neural Networks*, 2, 568-576.
- Vandamme, J.P., Superby, J.F. & Meskens, N. (2007). Predicting academic performance by data mining methods. *Education Economics*, 15(4), 405-419.