Predicting Academically At-Risk Engineering Students: A Soft Computing Application

Necdet Güner, Abdulkadir Yaldır, Gürhan Gündüz, Emre Çomak, Sezai Tokat, Serdar İplikçi

Pamukkale University, Department of Computer Engineering, Denizli, Turkey,
nguner@pau.edu.tr; akyaldir@pau.edu.tr; ggunduz@pau.edu.tr;
.ecomak@pau.edu.tr; stokat@pau.edu.tr; iplikci@pau.edu.tr

Abstract: This paper presents a study on predicting academically at-risk engineering students at the early stage of their education. For this purpose, some soft computing tools namely support vectors machines and artificial neural networks have been employed. The study population included all students enrolled in Pamukkale University, Faculty of Engineering at 2008-2009 and 2009-2010 academic years as freshmen. The data are retrieved from various institutions and questionnaires conducted on the students. Each input data point is of 38-dimension, which includes demographic and academic information about the students, while the output based on the first-year GPA of the students falls into either at-risk or not. The results of the study have shown that either support vector machine or artificial neural network methods can be used to predict first-year performance of a student in a priori manner. Thus, a proper course load and graduation schedule can be transcribed for the student to manage their graduation in a way that potential dropout risks are reduced. Moreover, an input sensitivity analysis has been conducted to determine the importance of each input used in the study.

Keywords: at-risk students; least-square support vector classification; radial basis functions neural network; support vector classification

1 Introduction

There have been many new universities established in Turkey in recent years. As a result, the number of students studying at Turkish universities is increasing, which allows students with diverse backgrounds attend the same classes. Many students are failing in their studies, as a result of having different learning levels. Engineering students, especially those without a sufficient background in math and science, are more likely to fail in courses [1] [2].

Some of the students cannot manage to graduate within the expected period, which leads to economical losses for both the family and the public. These losses can be greatly reduced by taking necessary social and academic predictive measurements, if academically at-risk students can be identified in advance.

There are many studies on predicting the success of university students and the factors influencing their success. Some of this research has focused on the reasons for early withdrawal. For instance, Tinto [3] has observed that 73% of the withdrawals occur within the first two years. In addition, McGrath and Braunstein [4] have found that low grade point average (GPA) at the first year is the major factor causing the early withdrawal. Some scientific research revealed that one of the major factors assisting to predict the success of students is their first-year GPAs and that there is a direct correlation between the first-year GPAs and graduating successfully in time [4] [5].

Apart from these findings, it has been found that half of the engineering students in the United States withdraw within the first two years [6]. In Australia, it has been reported that only 20% of the students in Queensland University of Technology have managed to graduate within four years [7]. In addition, more than 25% of the students in Australia consider withdrawing seriously within the early years of their study [8]. Researchers have revealed that there is a strong relationship between the first year academic success and the continuation of a university education [5]. Therefore, it is of great importance to predict the first year success of students.

There have been numerous researchers investigating the factors that have influence the success of students. These studies can be divided into three groups, namely,

(i) Academic background of students [5] [9] [10] [11]

(ii) Social, economic, and educational levels of students' families [9] [12]

(iii) Physiological and individual properties of students [13] [14] [15] [16].

In the literature, there have been many research papers attempting to predict the GPAs of students by using data mining and Soft Computing (SC). For instance, in the study by Affendey et al. [17], the influencing factors contributing to the academic performance of the students have been ranked using the Bayesian Approach, Radial Basis Function Neural Networks (RBFNN). On the other hand, Vandamme et al. [18] have divided the students into three groups and then predicted the academic success of the students by using different methods such as discriminant analysis, neural networks, random forests, and decision trees. In another application, Oladokun et al. [19] have developed an artificial neural network model to predict the performance of the students who are entering universities through the National University Admission Examination in Nigeria. The model was able to correctly predict the performance of more than 70% of prospective students. Also, Huang [20] has used multiple regression and SC methods to obtain a validated set of mathematical models in order to predict academic performances of students in Engineering Dynamics Courses.

In this study, SC methodologies have been employed to predict the first-year engineering students who fall into an at-risk group. The at-risk is defined as the students who have a GPA less than 2.00 (out of 4.00). Therefore, it is important to predict first-year GPA's of the newly enrolled students. It has been known that academic performances of students can be improved through academic and other consultancy assistance by predicting their performances as early and accurate as possible [21] [22] [23].

Support Vector Classification (SVC) approaches are based on the Structural Risk Minimization and Statistical Learning Theory and handle the classification problem by converting it into either a quadratic programming problem in the conventional SVC case or a set of linear equations in the Least-Squares SVC case, respectively. The idea behind the use of SVC approaches in the prediction of the academic performances of the first-year university students is the fact that SVC models are simple to obtain and that they have higher generalization potential. The rest of this paper is organized as follows: In Section 2 the prediction problem is defined in detail, Section 3 describes the SC methods used herein, Section 4 outlines the Input-Sensitivity Analysis, Section 5 explains the obtained results and finally, the paper ends with the conclusions.

2 Problem Definition

This research was conducted among the students who have enrolled in the Faculty of Engineering at Pamukkale University, a public university in Denizli, which is located in the southwest part of Turkey. To determine the academically at-risk students, we have used Machine Learning methods based on the data containing information about the students who enrolled in Pamukkale University Faculty of Engineering departments in academic years 2008-2009 and 2009-2010. The data are retrieved from Pamukkale University Students' Registry (PUSR) and Turkish Students Selection and Placements Centre (SSPC), which is responsible for the execution of University Entrance Exam (UEE).

Data about the academic background of students comprise the following: type of high school graduated, high school GPA, individual scores obtained from each or combined subject at the UEE, and numbers of correct and wrong answers given in each or combined subject at the UEE. Demographic data include gender, age, and the department of students, their parents' educational and socio-economic levels, their hometown distance to Pamukkale University, and their willingness of working part-time at the university. A total of 38 different types of data were considered for the 1050 Faculty of Engineering students, who enrolled in academic years 2008-2009 and 2009-2010 and are tabulated in Table 1 given here in the appendix.

Table 1
Data Retrieved from Pusr and SSPC

1. Gender
2. Year of birth
3. Department
4. Day/evening studies
5. Type of high school
6. High school graduation year
7. High school GPA
8. Distance of hometown to university
9. Mother alive/dead
10. Father alive/dead
11. Mother and father living together
12. Total number of siblings
13. Number of siblings studying at university
14. Father's education
15. Socio-economical level of the family*
16. Mother's education
17. Willing to work at the university
18. Attended to English preparatory school in university
19. High school graduation rank
20. Verbal score of the high school
21. Quantitative score of the high school
22. Equally weighted score of the high school
23. Number of correct answers in Math-1 test of the UEE
24. Number of correct answers in Science-1 test of the UEE
25. Number of correct answers in Math-2 test of the UEE
26. Number of correct answers in Science-2 test of the UEE
27. Number of false answers in Math-1 test of the UEE
28. Number of false answers in Science-1 test of the UEE
29. Number of false answers in Math-2 test of the UEE
30. Number of false answers in Science-2 test of the UEE
31. Quantitative-1 score of the UEE
32. Verbal-1 score of the UEE
33. Equally weighted-1 score of the UEE
34. Quantitative-2 score of the UEE
35. Equally weighted-2 score of the UEE
36. Physics test score of the UEE
37. Number of correct answers to complex numbers, logarithms, and trigonometry questions in the Math-2 test of the UEE
38. Number of correct answers to limit, derivatives, and integral questions in the Math-2 test of the UEE
39. University first year GPA

***** *Socio-economic levels of the families have been calculated as a combination of ten different data about students and their families collected by PUSR at the registration.*

It should be noted that some of the data are in binary form (e.g., gender), some of them are integers (e.g., total number of siblings), and the remaining are real numbers (e.g., high school GPA). No matter what the forms of the answers are, they all have been normalized into the interval [0, 1] in this study. Therefore, 1050 normalized data points of 39 dimensions have been used to obtain proper prediction models. The first 38 rows of Table 1 are taken as inputs for the prediction models, while the output falls into either at-risk or not, based on the first-year GPA of the students taken from row 39 of Table 1.

3 Soft Computing Methods (SC)

For all of the SC tools employed in this study, it is assumed that the data set $\mathcal{D}$ is collected for obtaining optimal model and has the form given below:

$$\mathcal{D} = \{\mathbf{x}_k ; y_k\}_{k=1}^{k=N} \quad (1)$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is n-dimensional k^{th} input vector, $y_k \in \{-1,+1\}$ is the corresponding binary output, and N is the total number of data, which is N = 1050 for this work. It is desired to find a model that represents the relationship between the input and output data points. Each SC tool used to obtain a proper model has its own modeling parameters, and different modeling parameters result in different models. Therefore, it is inevitable to search for the optimal modeling parameters in the parameter space. For this purpose, $\mathcal{D}$ is randomly divided into three parts: 600 for training, 200 for validation, and 250 for testing. Then, in order to find the best model for each SC tool, a grid search approach is adopted. In this approach, the modeling parameter space is divided by grids, and for each node (corresponding to specific parameter values) on the grid, a model is obtained using the training data set, and then, the model, which produces the least validation error based on the validation data set is chosen as the optimal model. Finally, optimal models for the SC tools are compared with each other by using the test data.

3.1 Support Vector Classification

The primal form of a SVC model is given by Equation (2), which is linear in a higher dimensional feature space F.

$$\hat{y}_i = \langle \mathbf{w}, \mathbf{\Phi}(\mathbf{x}_i) \rangle \quad (2)$$

where $\mathbf{w}$ is a vector in the feature space F, $\mathbf{\Phi}(.)$ is a mapping from the input space to the feature space, and $\langle . \rangle$ stands for the inner product operation in F. The SVC algorithms regard the classification problem as an optimization problem in dual space in which the model is given by Equation (3).

$$\hat{y}_i = \sum_{j=1}^{N_{Tr}} \alpha_j y_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{3}$$

where N_{Tr} is the number of training data, α_j is the coefficient corresponding to the training data $\mathbf{x}_j$, and $K(\mathbf{x}_i, \mathbf{x}_j)$ is a Gaussian kernel function given by,

$$K(\mathbf{x}_i, \mathbf{x}_j) = \left\langle \mathbf{\Phi}(\mathbf{x}_i), \mathbf{\Phi}(\mathbf{x}_j) \right\rangle = e^{-\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2}{2\sigma^2}} = K_{ij} \tag{4}$$

The kernel function handles the inner product in the feature space, and thus, the explicit form of $\mathbf{\Phi}(\mathbf{x})$ does not need to be known. In the model given by Equation (3), a training point $\mathbf{x}_j$ corresponding to a non-zero α_j value is referred to as the support vector. The primal form of the classification problem is as follows:

$$\min_{\mathbf{w}, b, \xi, \xi^*} P = \frac{1}{2}\left\|\mathbf{w}\right\|^2 + C\sum_{i=1}^{N_{Tr}} \xi_i \tag{5}$$

subject to the constraints,

$$y_i \left\langle \mathbf{w}, \mathbf{\Phi}(\mathbf{x}_i) \right\rangle \leq 1 - \xi_i, \; i = 1, \ldots, N_{Tr} \tag{6a}$$

$$\xi_i \geq 0, \; i = 1, \ldots, N_{Tr} \tag{6b}$$

where ξ_i's are slack variables, $\|.\|$ is the Euclidean norm, and C is a regularization parameter. By adding the constraints to the primal form of the classification problem, the Lagrangian can be obtained as

$$L_P = \frac{1}{2}\left\|\mathbf{w}\right\|^2 + C\sum_{i=1}^{N_{Tr}} \xi_i - \sum_{i=1}^{N_{Tr}} \alpha_i \left(y_i \left\langle \mathbf{w}, \mathbf{\Phi}(\mathbf{x}_i) \right\rangle - 1 + \xi_i \right) - \sum_{i=1}^{N_{Tr}} \mu_i \xi_i \tag{7}$$

where α_i's and μ_i's are Lagrange multipliers. First-order conditions of the primal optimization problem are obtained by taking partial derivatives of L_P with respect to the design variables and then setting them to zero as follows:

$$\frac{\partial L_P}{\partial \mathbf{w}} = 0 \mapsto \mathbf{w} = \sum_{i=1}^{N_{Tr}} \alpha_i y_i \mathbf{\Phi}(\mathbf{x}_i) \tag{8}$$

$$\frac{\partial L_P}{\partial \xi_i} = 0 \mapsto C - \alpha_i - \mu_i = 0, \quad i = 1, \ldots, N_{Tr} \tag{9}$$

Now, the dual form of the optimization problem becomes a Quadratic Programming (QP) problem as:

$$\min_{\boldsymbol{\alpha}} D = \frac{1}{2}\sum_{i=1}^{N_{Tr}}\sum_{j=1}^{N_{Tr}}\alpha_i\alpha_j y_i y_j K_{ij} - \sum_{i=1}^{N_{Tr}}\alpha_i \tag{10}$$

subject to the constraints,

$$\sum_{i=1}^{N_{Tr}}\alpha_i y_i = 0 \text{ and } 0 \le \alpha_i \le C,\ i = 1,\ldots,N_{Tr} \tag{11}$$

Solution of the QP problem given by equations (10) and (11), yields the optimum values of α_i's [24]. Furthermore, when only the support vectors are considered, the model becomes as follows:

$$\hat{y}_i = \sum_{\substack{j=1 \\ j\in SV}}^{\#SV}\alpha_j y_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{12}$$

where #SV stands for the number of support vectors in the model. The SVC model given by Equation (12) is sparse in the sense that the whole training data are represented by only support vectors. The parameters of SVC are the regularization parameter C and the kernel parameter σ.

3.2 Least-Square Support Vector Classification

Least-squares support vector classification (LSSVC) is a variety of SVC, which has almost the same level of capability in classification and regression as SVC [25] [26]. LSSVC finds optimal value of the cost function given in Equation (13) subject to equality constraints instead of inequality ones in the SVC case. Therefore, it is desired to minimize the following:

$$\frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{2}\sum_{i=1}^{N_{Tr}}\xi_i^2 \tag{13}$$

subject to

$$y_i\left(\left\langle \mathbf{w}, \boldsymbol{\Phi}(\mathbf{x}_i)\right\rangle + b\right) = 1 - \xi_i,\ i = 1,\ldots,N_{Tr} \tag{14}$$

Because the optimization problem is built on linear equations, computational burden of LSSVC is less than that of SVC. On the other hand, SVC is sparser than LSSVC in the sense that the former contains less number of support vectors in the model than the latter. However, both approaches exhibit similar classification performances. Yet, we have employed both approaches in this study for the sake of comparison. Equation (15) is obtained when Eqs. (13-14) are presented in dual optimization form with Lagrange multipliers.

$$L(\mathbf{w},b,\boldsymbol{\alpha},\boldsymbol{\xi}) = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{2}\sum_{i=1}^{N_{Tr}}\xi_i^2 - \sum_{i=1}^{N_{Tr}}\alpha_i\left[y_i\left(\langle\mathbf{w},\boldsymbol{\Phi}(\mathbf{x}_i)\rangle + b\right) - 1 + \xi\right] \tag{15}$$

where $\alpha_i \in \mathbb{R}^n$ are the Lagrange multipliers. The first-order conditions for optimality are as follows:

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \rightarrow \mathbf{w} = \sum_{i=1}^{N_{Tr}}\alpha_i\boldsymbol{\Phi}(\mathbf{x}_i), \tag{16a}$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N_{Tr}}\alpha_i = 0, \tag{16b}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha_i = -C\xi_i,\ i = 1,\ldots,N_{Tr}, \tag{16c}$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow y_i = \langle\mathbf{w},\boldsymbol{\Phi}(\mathbf{x}_i)\rangle + b + \xi_i,\ \ i = 1,\ldots,N_{Tr} \tag{16d}$$

With the elimination of $\mathbf{w}$ and ξ_i, a set of linear equations are obtained as given by Equation (17), the solution of which contains Lagrange multipliers and the bias term.

$$\begin{bmatrix} 0 & \mathbf{Y}^T \\ \mathbf{Y} & \mathbf{ZZ}^T + C^{-1}I \end{bmatrix}\begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix} \tag{17}$$

where the matrix is a $(N_{Tr}+1)\times(N_{Tr}+1)$ square matrix,

$$\mathbf{Z}^T = \left[y_1\boldsymbol{\Phi}(x_1),\ldots,y_N\boldsymbol{\Phi}(x_{N_{Tr}}) \right] \tag{18a}$$

$$\mathbf{Y}^T = \left[y_1, y_2,\ldots,y_{N_{Tr}} \right] \tag{18b}$$

$$\boldsymbol{\alpha}^T = \left[\alpha_1,\alpha_2,\ldots,\alpha_{N_{Tr}} \right] \tag{18c}$$

and C is a scalar parameter. Similar to SVC, the output value of LSSVC is computed by Equation (12) after Lagrange multipliers and bias values are found. In contrast to SVC, Lagrange multipliers in LSSVC might be positive or negative. It should be noted that the number of support vectors in the model is the same as the number of training data. The inner product $\langle\boldsymbol{\Phi}(\mathbf{x}_i),\boldsymbol{\Phi}(\mathbf{x}_j)\rangle$ is handled by the Gaussian kernel function as in the SVC case.

3.3 Radial Basis Function Neural Networks

Radial basis function neural networks (RBFNN) are special artificial neural network structures in which the hidden units are activated with respect to the distance between the input vector and a predefined centre vector. RBFNN can provide a nonlinear model for the target dataset with its simple and yet fast learning network structure [27], and therefore, it is a sensible alternative to use complex polynomials for function approximation.

In a RBFNN, there is only one hidden layer that uses neurons with radial basis function (RBF) activation functions. RBFs implement localised representations of functions, and they are real valued functions whose outputs depend on the distance of the input from the stored centre vector of each hidden unit [28]. Thus, it has its peak value at the centre and decreases in each direction along the centre. Different functions, such as multi-quadratics, inverse multi-quadratics, and bi-harmonics, could be used as RBF. A typical selection is a Gaussian function for which the output of the i^{th} hidden unit is written as

$$y = \sum_{i=1}^{\#HU} w_i \exp(-\|\mathbf{x}_k - \mathbf{v}_i\|^2 \big/ 2\sigma_i^2)\,,\ i = 1,\ldots,\#HU \tag{19}$$

where $v_i \in \mathbb{R}^n$ is *n*-dimensional centre vector of the RBF of the i^{th} hidden neuron, σ_i is the width of RBF of the i^{th} hidden neuron, *#HU* is the number of hidden units, and w_i is the weight of the i^{th} hidden unit. An RBFNN is completely determined by choosing the dimension of input-output data; number of RBFs; and values of v_i, σ_i and w_i. The function approximation or classification performance of RBFNN is obtained by defining all these parameters. The dimension of input-output data is problem dependent and defined clearly at the beginning. Choice of the number of RBFs plays a critical role and depends on the problem under investigation. For simplicity in calculations, σ_i values are all taken equal to σ. In this study, *#HU* and σ are grid searched to choose the best values for validation data. In the training phase, hidden unit neurons are added using an orthogonal least squares algorithm to reduce the output error of network until the sum-squared error goal is reached [29].

4 Input-Sensitivity Analysis

By input-sensitivity analysis, it can be determined to what extent the output of the SVC model is sensitive to each input of the model. In this respect, the partial derivative of the output $\hat{y}(\mathbf{x})$ with respect to each input is needed. Let us remember that the input-output relationship of the SVC model is

$$\hat{y}(\mathbf{x}) = \sum_{\substack{j=1 \\ j\in SV}}^{\#SV} \alpha_j y_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{20}$$

where $\mathbf{x}_j$'s are the support vectors, $\mathbf{x} \in \mathbb{R}^n$ is n-dimensional input vector and $K(\mathbf{x}, \mathbf{x}_j)$ is a Gaussian kernel function given by

$$K(\mathbf{x}, \mathbf{x}_j) = e^{-\frac{\|\mathbf{x}-\mathbf{x}_j\|}{2\sigma^2}} = e^{-\frac{(x_1-x_{j1})^2+(x_2-x_{j2})^2+\ldots+(x_n-x_{jn})^2}{2\sigma^2}} \tag{21}$$

Then, the input-output relationship becomes

$$\hat{y}(\mathbf{x}) = \sum_{i\in S} \alpha_j y_j K(\mathbf{x}, \mathbf{x}_j) = \sum_{i\in S} \alpha_j y_j e^{-\frac{(x_1-x_{j1})^2+(x_2-x_{j2})^2+\ldots+(x_n-x_{jn})^2}{2\sigma^2}} \tag{22}$$

Now, the partial derivatives can be written by

$$\frac{\partial \hat{y}(\mathbf{x})}{\partial \mathbf{x}_k} = \frac{\partial \sum_{i\in S} \alpha_j y_j e^{-\frac{(x_1-x_{j1})^2+(x_2-x_{j2})^2+\ldots+(x_n-x_{jn})^2}{2\sigma^2}}}{\partial \mathbf{x}_k} \tag{23}$$

The derivative in Equation (23) can be calculated as

$$\begin{aligned}
\frac{\partial \hat{y}(\mathbf{x})}{\partial \mathbf{x}_k} &= \frac{\sum_{i\in S} \alpha_j y_j \partial e^{-\frac{(x_1-x_{j1})^2+(x_2-x_{j2})^2+\ldots+(x_n-x_{jn})^2}{2\sigma^2}}}{\partial \mathbf{x}_k} \\
&= \sum_{i\in S} \alpha_j y_j \frac{\partial e^{-\frac{(x_1-x_{j1})^2+(x_2-x_{j2})^2+\ldots+(x_n-x_{jn})^2}{2\sigma^2}}}{\partial \mathbf{x}_k} \\
&= \sum_{i\in S} \alpha_j y_j \frac{(x_{jk}-x_k)}{\sigma^2} e^{-\frac{(x_1-x_{j1})^2+(x_2-x_{j2})^2+\ldots+(x_n-x_{jn})^2}{2\sigma^2}} \\
&= \sum_{i\in S} \alpha_j y_j \frac{(x_{jk}-x_k)}{\sigma^2} K(\mathbf{x}, \mathbf{x}_j)
\end{aligned} \tag{24}$$

For a SVC model obtained by the data set $\{\mathbf{x}_i ; y_i\}_{i=1}^{i=N}$, it is possible to build a sensitivity vector for the k^{th} input as

$$\mathbf{s}_k = \left[\frac{\partial \hat{y}(\mathbf{x}_1)}{\partial \mathbf{x}_k} \quad \frac{\partial \hat{y}(\mathbf{x}_2)}{\partial \mathbf{x}_k} \quad \ldots \quad \frac{\partial \hat{y}(\mathbf{x}_N)}{\partial \mathbf{x}_k} \right] \tag{25}$$

Thus, the norm $\|\mathbf{s}_k\|$ of the sensitivity vector can be regarded as a numerical measure that indicates the sensitivity of the output to the k^{th} input for the SVC model obtained by the data set $\{\mathbf{x}_i; y_i\}_{i=1}^{i=N}$. For large sensitivity of the output to the k^{th} input, we obtain relatively large $\|\mathbf{s}_k\|$ values and vice versa. That being $\|\mathbf{s}_k\| = 0$ means no sensitivity to the k^{th} input, *e.g.* no matter how much the k^{th} input is changed the output is not affected. By comparing the sensitivity vectors regarding to all inputs, it is possible to determine the relative sensitivities of the inputs. Moreover, some inputs having very small sensitivities can be discarded from the data set and then the SVC model can be re-obtained with the new data set.

Similar to the case given for SVC case, using RBFNN input output equation given in (19), the partial derivative of output variable $\hat{y}(\mathbf{x})$ with respect to each input vector $\mathbf{x}_k$ can be obtained as

$$\frac{\partial \hat{y}(\mathbf{x})}{\partial x_k} = \sum_{i=1}^{\#HU} w_i \frac{\left(x_k - v_i(k)\right)}{\sigma^2} e^{-\frac{\|\mathbf{x}-\mathbf{v}_i\|}{\sigma^2}} \tag{26}$$

The sensitivity analysis of input variables is made by using Equation (26). The last four inputs 35, 9, 32 and 34 can be pruned as they have relatively lower sensitivity than other inputs. In this study, as also highlighted in the literature [30], the sensitivity analysis is initially examined at first hand prior to the design of the classifier structures. But, as the pruning of the last 4 inputs does not change the results significantly, the pruning of the network structure is not conducted in order to see the whole effect of the questionnaire.

5 Results and Discussions

Each SC method used in this study has its own parameter set to be optimized. To find the optimal parameter set, a grid search approach is adopted, where the parameter space is divided into grids. A node in the grid corresponds to a parameter set. In the grid search, validation performances of the models for each nodes (parameter sets) are calculated, and then, the parameter set having the least validation error is determined as the optimal parameter set. Table 2 tabulates the optimal parameter sets found by grid search for each method employed in the study. The optimal parameter sets are given in the second column. The optimal parameter sets, training, validation and test performance for each method, can be seen in columns 3-5 in Table 2.

Table 2
Optimal Parameters and Obtained Results

Method	Parameters	Train %	Validation %	Test %
SVC	$C = 0.1$, $\sigma = 0.6$	98.66	73.50	68.80
LSSVC	$C = 1.6$, $\sigma = 121$	91.50	78.50	75.60
RBFNN	$\#HU = 67$, $\sigma = 1.3$	77.17	78.00	77.60

As can be seen in the table, all methods exhibit satisfactory validation and test performances almost over 70%. However, the LSSVC and RBFNN yield better results than SVC. It can also be seen that the validation and test results for LSSVC and RBFNN methods are close to each other. The reason for the SVC approach to give relatively weak performance can be attributed to the fact that the SVC model may go into over-fitting. This can be observed if the performances of the methods in Table 2 are examined. The less training error the method produces, the more over-fitting and the worse generalization it does.

As a result of sensitivity analysis performed for the three methods, normalized $\mathbf{s}_k$ results have been presented with a bar graph in Figure 1. Also, the actual $\mathbf{s}_k$ values and inputs according to sensitivity ranks are presented in Table 3.

It is assumed that any input *k* which has a normalized $\mathbf{s}_k$ value less than 0.33 can be regarded as having a low impact on the student's academic success in the first semester. These inputs are year of birth, high school graduation year, mother alive/dead, number of siblings studying at university, number of correct answers in math-1 test of the UEE, number of false answers in science-1 test of the UEE, number of false answers in science-2 test of the UEE, quantitative-1 score of the UEE, verbal-1 score of the UEE, equally weighted-1 score of the UEE, quantitative-2 score of the UEE and equally weighted-2 score of the UEE, and indicated with '*' in Figure 1. Based on this sensitivity analysis, it is observed that some inputs have less impact on the output than others and these inputs can be discarded in further applications.

Table 3
Sensitivity Analysis Results

	SVC		**LSSVC**		**RBFNN**	
Sensitivity Rank	Input k	$\mathbf{s}_k$	Input k	$\mathbf{s}_k$	Input k	$\mathbf{s}_k$
1	38	61.1627	25	53.87	1	134.0328
2	20	49.1408	26	46.748	17	121.1353
3	26	47.2159	36	43.451	4	114.5789
4	22	43.6456	37	34.691	11	101.6467
5	25	40.8373	24	32.603	10	100.3562
6	21	38.4838	19	30.843	38	98.3192
7	7	37.5234	1	27.208	18	94.1968
8	19	36.6386	38	26.353	26	87.8470
9	37	35.7801	20	25.487	14	81.8787
10	36	29.7689	3	20.379	3	81.6844
11	8	25.3821	7	20.233	20	79.4239
12	24	23.4596	22	19.637	7	74.7644
13	29	23.0989	21	14.656	8	72.6266
14	1	21.672	28	14.495	19	71.6740
15	27	20.2662	15	14.104	16	68.9167
16	6	17.6121	23	12.798	36	66.8006
17	34	17.4067	5	12.228	12	66.4493
18	3	16.3039	4	11.784	5	65.0252
19	14	15.1906	2	11.573	22	62.9082
20	31	13.9646	33	11.39	25	62.2902
21	28	13.2361	8	10.419	15	59.9410
22	33	12.0748	29	10.278	21	56.4296
23	15	10.8849	14	9.7945	24	55.4738
24	2	10.279	31	7.0194	37	54.6353
25	4	10.1053	32	6.5261	13	49.9793
26	30	9.9314	6	6.0264	23	47.3972
27	12	7.3607	10	5.3529	27	43.3737
28	23	5.7258	27	5.2585	30	42.4624
29	35	5.7243	34	4.7037	28	37.8810
30	16	4.2286	11	4.4312	31	37.4403
31	11	4.2085	30	3.6643	33	36.6683
32	17	3.4682	35	3.4183	29	35.8899
33	5	2.8061	17	2.7849	9	34.9133
34	10	2.8017	12	2.7604	2	29.9388
35	18	2.5429	16	1.7147	6	27.5962
36	32	2.4576	18	1.3411	35	26.2895
37	9	1.5825	9	0.97898	32	22.6935
38	13	1.5804	13	0.95468	34	20.8956

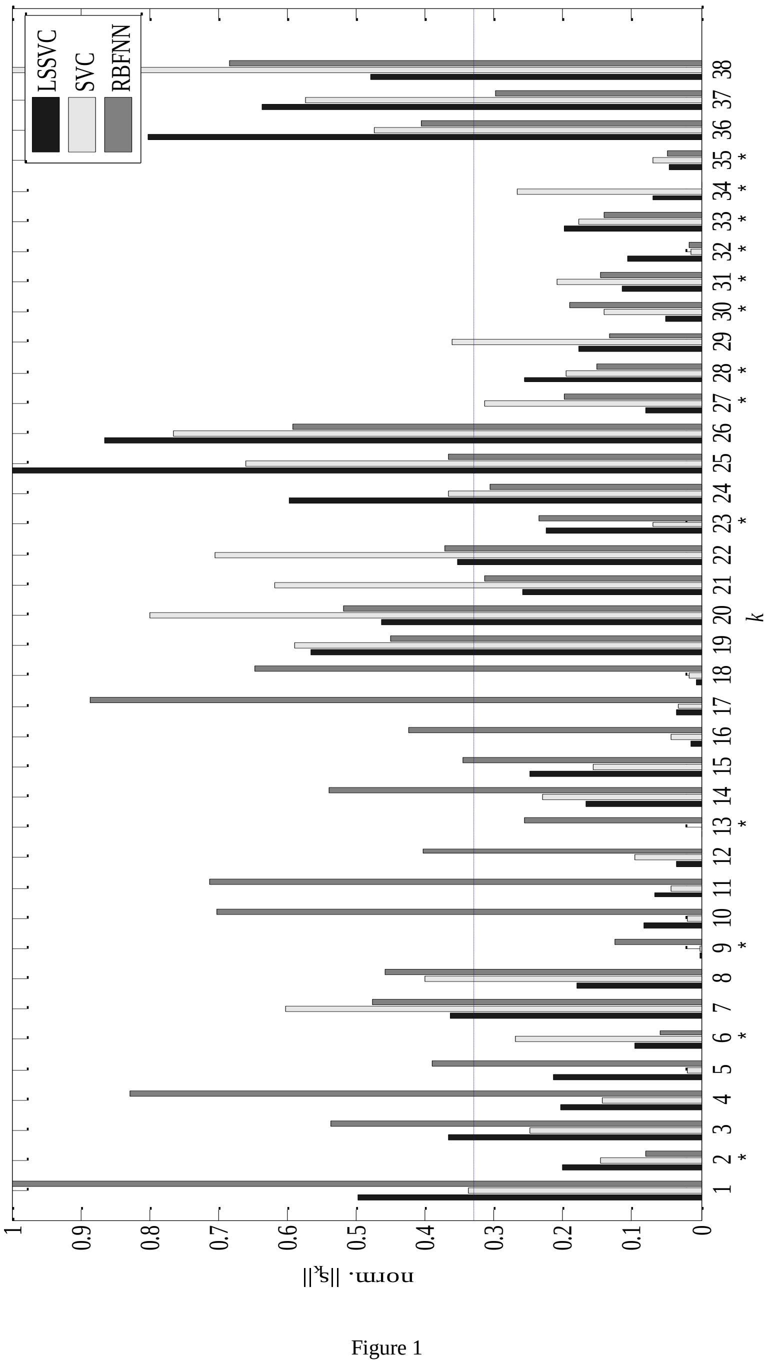

Figure 1
Input-Sensitivity Analysis Report

Conclusions

In this paper, a study on predicting academically at-risk engineering students newly enrolled to a university has been presented. For this purpose, some SC tools, namely, Support Vectors Machines and Artificial Neural Networks have been used, because of their high generalization capabilities. The data containing information 1050 students are retrieved from PUSR and SSPC, which are responsible for the execution of UEE. In the study, it has been assumed that the first-year success of an engineering student is mainly dependent on the performance in the centralized UEE, high school performance, and socio-economic and educational level of the family. Therefore, the data used in the study have been prepared accordingly. The results revealed that all the soft computing tools we have used yielded satisfactory prediction performances for both test and validation data. To be specific, both LSSVC and RBFNN provide more than 75% validation and test performance, whereas SVC provides 73.50% for validation and 68.80% for testing. The reason for the SVC method to give relatively weak performance can be explained by the fact that it makes more over-fitting than others as can be seen in Table 2.

Moreover, based on the obtained models a sensitivity analysis has been conducted, which has revealed that some inputs in the study can be ignored since the output is less sensitive to them than others. The results of this analysis can be used in similar applications in future.

Based on these SC approaches, a computer application may be developed to provide an academic counseling service for freshman engineering students, by means of which student advisors can predict the students' GPA scores at the end of the first semester by entering the required data into the application and can warn them when necessary. It is planned at the Engineering Faculty of Pamukkale University to apply such computer software to the freshman students who will enroll to the faculty in 2013-2014 academic year.

In conclusion, either support vector machine-based methods or RBFNN's can be used to predict first-year performance of a student based on a priori knowledge and data. Thus, a proper course load per semester and graduation schedule can be developed for a student to manage their graduation in a way that potential drop-off risks are reduced.

Acknowledgement

The authors gratefully acknowledge the help of the Pamukkale University Students' Registry and Turkish Students Selection and Placements Centre in Ankara for providing the necessary data.

References

[1] P. Broadbridge, S. Henderson, Mathematics Education for 21st Century Engineering Students - Final Report, Melbourne, Australian Mathematical Sciences Institute, 2008

[2] P. Kent, R. Noss, Mathematics in the University Education of Engineers, A Report to the Ove Arup Foundation, London, the Ove Arup Foundation, 2003

[3] V. Tinto, Leaving College; Rethinking the Causes and Cures of Student Attrition, Chicago, University of Chicago Press, 1994

[4] M. McGrath, A. Braunstein, "The Prediction of Freshmen Attrition", College Student Journal, Vol. 31, pp. 396-408, 1997

[5] S. M. DeBerard, D. J. Julka, G. I. Spielmans, "Predictors of Academic Achievement and Retention among College Freshmen: A Longitudinal Study", College Student Journal, Vol. 38, pp. 66-85, 2004

[6] M. Crawford, K. J. Schmidt, "Lessons Learned from a K-12 Project", Proceedings of the 2004. American Society for Engineering Education Annual Conference and Exposition, Washington, American Society for Engineering Education, pp. 1-13, 2004

[7] R. H.Cuthbert, H. L. MacGrillivray, "Investigation of Completion Rates of Engineering Students", In D'Arcy- A.L. Warmington, M. Victor, G. Oates, C. Varsavsky. (eds.) 6th Southern Hemisphere Conference on Mathematics and Statistics Teaching and learning (El Calafate DELTA' 07), 26-30 November 2007, El Calafate, Argentina, pp. 35-41, 2007

[8] K. L. Krause, "Serious Thoughts about Dropping Out in First Year: Trends, Patterns and Implications for Higher Education", Studies in Learning, Evaluation, Innovation and Development, Vol. 2, No. 3, pp. 55-68, 2005

[9] J. R. Betts, D. Morell, "The Determinants of Undergraduate Grade Point Average", The Journal of Human Resources, Vol. 34, No 2, pp. 268-293, 1999

[10] K. McKenzie, R. Schweitzer, "Who Succeeds at University? Factors Predicting Academic Performance in First Year Australian University Students", Higher Education Research and Development, Vol. 20, No. 1, pp. 21-33, 2001

[11] N. W. Burton, L. Ramist, Predicting Success in College: SAT Studies of Classes Graduating Since 1980, College Board Report No 2001-2. New York, College Entrance Examination Board, 2001

[12] S. M. R. Ting, "Predicting Academic Success of First-Year Engineering Students from Standardized Test Scores and Psychosocial Variables", International Journal of Engineering Education, Vol. 17, No. 1, pp. 75-80, 1998

[13] S. Museus, D. Hendel, "Test Scores, Self-Efficacy and the Educational Plans of First-Year College Students", Higher Education in Review, Vol. 2, pp. 63-88, 2005

[14] T. Farsides, R. Woodfield, "Individual Differences and Undergraduate Academic Success: the Roles of Personality, Intelligence and Application", Personality and Individual Differences, Vol. 34, pp.1225-1243, 2003

[15] J. D. A. Parker, L. J. Summerfeldt, J. Marjorie, M. J. Hogan, S. A. Majeski, "Emotional Intelligence and Academic Success: Examining the Transition from High School to University", Personality and Individual Differences, Vol. 36, pp. 163-172, 2004

[16] S. Trapmann, B. Hell, J .O. W. Hirn, H. Schuler, "Meta-Analysis of the Relationship between the Big Five and Academic Success at University", Zeithschrift für Psychologie, Vol. 215, No. 2, pp. 132-151, 2007

[17] L. S. Affendey, I. H. M. Paris, N. Mustapha, N. Sulaiman, Z. Muda, "Ranking of Influencing Factors in Predicting Students' Academic Performance," Information Technology Journal, Vol. 9, No. 4, pp. 832-837, 2010

[18] J. P. Vandamme, N. Meskens, J. F. Superby, "Predicting Academic Performance by Data Mining Methods", Education Economics, Vol. 15, No. 4, pp. 405-419, 2007

[19] V. O. Oladokun, A. T. Adebanjo, O. E. 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

[20] S. Huang, "Predictive Modeling and Analysis of Student Academic Performance in an Engineering Dynamics Course", Doctoral Dissertation, Utah State University, Logan Utah, 2011

[21] J. M. Braxton, A. S. Hirschy, S. A. McCladon, Understanding and Reducing College Student Departure: ASHE-ERIC Higher Education Report. San Francisco, John Wiley and Sons Inc., 2004

[22] G. D. Kuh, J. Kinzie, J. H. Schuh, E. J. Witt, Student Success in College Creating Conditions That Matter, San Francisco, John Wiley and Sons Inc., 2010

[23] M. L. Upcraft, J. N. Gardner, B. O. Barefoot, Challenging and Supporting the First-Year Student. A Handbook for Improving the First Year of College, San Francisco, John Wiley and Sons Inc., 2005

[24] E. Alpaydın, Introduction to Machine Learning, Cambridge, The MIT Press, 2010

[25] J. A. K. Suykens, J. Vandewalle, "Least Squares Support Vector Machine Classifiers", Neural Processing Letters, Vol. 9, No. 3, pp. 293-300, 1999

[26] D. Tsujinishi, S. Abe, "Fuzzy Least Squares Support Vector Machines for Multi-Class Problems", Neural Networks, Vol. 16, pp. 785-792, 2003

[27] N. B. Karayiannis, S. Behnke, “New Radial Basis Neural Networks and Their Application in a Large-Scale Handwritten Digit Recognition Problem”, In L. Jain, A. M. F. Fanelli (eds.) Recent Advances in Artificial Neural Networks: Design and Application. Florida, CRC Press, 2000

[28] R. J. Schilling, J. J. Carroll, A. F. Al-Ajlouni, “Approximation of Nonlinear Systems with Radial Basis Function Neural Networks”, IEEE Transactions on Neural Networks, Vol. 12, No. 1, pp. 1-15, 2001

[29] S. Chen, C. F. N. Cowan, P. M. Grant, “Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks”, IEEE Transactions on Neural Networks, Vol. 2, pp. 302-309, 1999

[30] D. S. Yeung, I. Cloete, D. Shi, W. W. Y. Ng, Sensitivity Analysis for Neural Networks, Natural Computing Series, Berlin Heidelberg, 2010