Student Performance Prediction and Risk Factor Identification

COMP 5450/4220
May 2018

Aaron Raymond Buthelo
Lior Osherov
Jaclyn Solimine

Abstract:

Educational institutions are commonly faced with a shortage of important resources that are meant to enhance the academic experience and performance of their students. The overall utility of these resources can be optimized by ensuring that they are allocated to those students most in need or most deserving. This study investigated the use of Support Vector Machines to develop a classifier model capable of predicting a student's projected performance, with the aim of facilitating the allocation of academic resources. The performance of Linear, Polynomial, and Gaussian kernels in the creation of the SVM model was evaluated. In addition, PCA was used to observe the effect of reducing the dimensionality of the dataset on classification performance. Manual feature selection was conducted to determine the strongest indicators of likely academic success and failure. Overall, the models developed were successful in determining which students were likely to perform above or below average.

Keywords: Classification, Supervised Learning, Support Vector Machines, Principal Component Analysis

1. Introduction:

The focus of this study was to use machine learning techniques to create classifier system capable of determining a particular student's likelihood of academic success or failure based on their background information and past academic performance. As a preliminary finding of this study, those features or traits that were most indicative of success or failure in students were also able to be identified. Ultimately, following the successful development of the classifier, a framework could be developed to monitor, analyze, and track student performance in addition to identifying the specific factors that make the most significant impact on each individual's success.

Presently, many educational institutions are faced with a shortage of important resources that are meant to enhance the academic experience and performance of their students. These resources include forms of services and aid such as tutoring, scholarship aid, or specialized learning opportunities that often have limited availability due to restrictive budgets. These limitations on resources are also met with the challenge of ensuring that the proper aid is allocated to those students that would benefit from it the most, and that these students are identified as early as possible. For instance, students who are excelling academically or are expected to continue excelling may be targeted early on for opportunities to further their academic development, such as special programs, college application assistance, or scholarship money. Conversely, students who are struggling can be more easily identified so that the necessary intervention may occur before they slip through the cracks, and work can be done to ensure that specific risk factors they may possess can be adjusted.

For this study, the most significant features with respect to each student's academic success were selected through the use of manual data mining techniques to determine which features contributed the most to the variance present in the dataset. In addition, principal component analysis (PCA) was used as a feature reduction method with the aim of enhancing the performance of the classification process through the removal of redundant or highly correlated features. The performance of Support Vector Machine (SVM) models was evaluated based on their ability to accurately differentiate between binary and multi-class labels pertaining to student academic success. SVMs were chosen due to their ability to manage high-dimensional features spaces and for their resistance to overfitting. The performance of the classifiers were evaluated through the use of ROC curves, confusion matrices, and by observing the overall percentage of correctly classified instances.

The dataset used in this application is entitled "Student Performance Data Set" and is available through the University of California-Irvine Machine Learning Repository. The data set captures the academic performance of 1044 student "samples" in three parts: 1st Term Grades, 2nd Term Grades, and Final Grades. These grades were collected from both math and Portuguese classes over the course of an academic year. A feature set consisting of a variety of 30 socio-economic, demographic, and habitual traits was collected for each student in addition to their grades.

2. Background

The welfare of our society is maintained by providing high-quality education. The potential to predict a student's performance is very important in educational environment. There are various factors responsible for a student's performance. Factors can be personal, social or psychological. Data Mining techniques were used to attain this objective. The main function of data mining is to apply various methods and algorithms to discover and extract new pattern from the stored data. Educational Data Mining is applied on data related to education using techniques such as Decision trees, Naive Bayes, k-means and many others. [1] In earlier studies, data mining methodologies were used to classify students into clusters in order to enhance the teaching-learning process and thus the attempt to improve overall academic performance. In one research study, the k-means algorithm was used as a clustering technique to classify students according to their academic performance. Based on the clusters produced, the teachers were able to segregate students and aim his/her attention towards those predicted to under-perform in order to improve their performance [1].

In another study, a Multilayer Perceptron (MLP) based prediction application was proposed to predict the grade point average (GPA) of undergraduate students using a feature set containing students' previous academic history, regularity, no. of backlogs, degree of intelligence, working nature, discipline, social activities and grades. With this application it was possible to use the students' data to predict who was at risk of failure, and some proactive measures like extra classes & supporting material were offered to improve the academic progress of those students. To evaluate the performance of the proposed application, data was recorded from 134 third-year Computer Science Engineering Students at Vignan University. The model achieved 95.52% and 97.37% prediction accuracy with RBF (Radical basic function) and MLP, respectively. [2]

Another research study aimed to develop a model for analysis of student behavior through e-Learning based on data mining techniques using data collected from Suan Sunandha Rajabhat University. The student data set was composed of 5392 personal records and, to compare the accuracy of different algorithms, the model was created using both decision tree and Bayesian networks techniques. The results found showed that Bayesian networks demonstrated superior performance with a percentage accuracy of 91.32%. [3]

3. Approach

This section details the procedure and applicable theory of this study.

3.1 Dataset Preparation

For the purpose of this study, dataset preprocessing was required in order to ensure that the selected machine learning algorithms would be able to operate effectively. In its raw form, the Student Performance Dataset contained a mixture of categorical, binary, and continuous feature set variables. Each of these features were normalized on a [0,1] scale by (1) converting all binary variables to [0,1] (2) converting all categorical/multi-label features to dummy variables, where each label becomes its own binary class (i.e. a four-category feature becomes four separate binary features) and (3) converting all continuous features to represent values on the 0-1 range. After this step of the preprocessing, the original 30-feature input variable set was expanded to 67 features. A complete list of these features can be found in Appendix A.

The student's academic performance was rated on a scale of 0-20, with the data provided in three parts: 1st Term Grades, 2nd Term Grades, and Final Grades. These output variables were processed to create new binary and multiclass output labels, which are outlined in Table 1. Figure 1 has also been provided as a visual representation of the class labels and distributions. These new classes were created in order to observe the success of the developed classifiers in differentiating between either two classes or multiple classes, while also ensuring that the classes were fairly balanced in terms of available sample data.

Table 1: Outline of new class output labels

Class Label Type	Range (Original Scale)	New Label
Binary	0-10	0
	11-20	1
Multiclass	0-5	1
	6-10	2
	11-15	3
	16-20	4

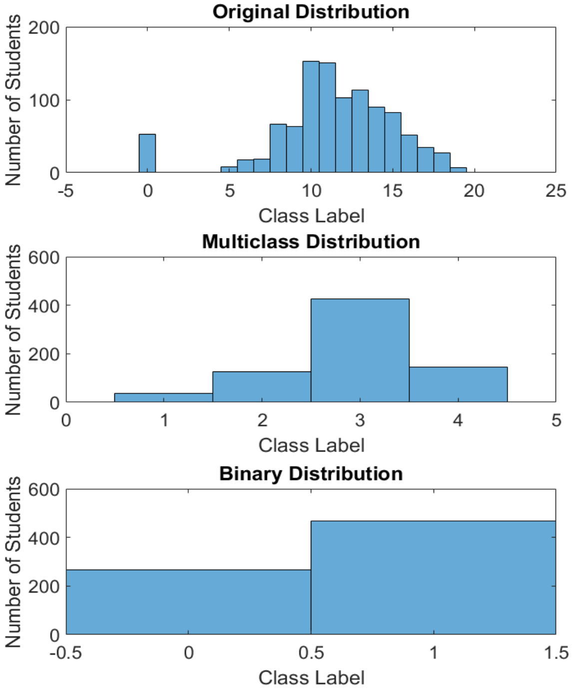


Figure 1: Class label distributions

In total, the performance of the each of the classification methods was evaluated over five different cases, as demonstrated in Table 2. For these cases, the effect of including the 1st and/or 2nd term grades on the accuracy of the models was observed. Ideally, these models will serve as an indication of just how soon in given academic year students can be classified into their respective academic performance classes. Specifically, the cases seek to determine if each of the models can accurately predict a student's projected performance based on only their background information, and how much this prediction improves when the recent grades of the student are included in the feature set.

Table 2: Overview of different Cases for evaluation

Case Number	Class Label Type	1st Term Grade Included in Feature Set?	2nd Term Grade Included in Feature Set?
1	Multiclass	No	No
2	Binary	Yes	No
3	Binary	Yes	Yes
4	Multiclass	Yes	No
5	Multiclass	Yes	Yes

The performance of each of these cases was considered with and without the application of PCA. In addition, the data samples were divided into two separate sets: a training set composed of 70% of the data, and a testing/validation set composed of 30% of the data.

3.2 Manual Feature Selection

Manual feature selection was conducted to observe (1) which features accounted for the greatest differences between the student performance classes (2) which features students within a single performance class had in common. In order to accomplish this, the means of the feature values for each individual class were determined. The percent differences between the means of each class were then calculated for each feature. In theory, a higher percent difference indicates that the members of one class varied more significantly from the members of another class with respect to that feature. A low percent difference indicates that members of the two classes in question did not vary with respect to that feature, therefore rendering that feature ineffective for differentiating between the two classes. This method was used to highlight the most salient features available in the collected data.

3.3 Feature Reduction using Principal Component Analysis

This section details the basic theory and application of PCA as it pertains to this study.

3.3.1 PCA Theory

Principal component analysis is a common feature reduction technique that is commonly used in the preprocessing stage of machine learning applications. The method uses orthogonal transformation to generate a principal component space from linear combinations of the original feature set. This process effectively converts highly correlated or redundant features into new, uncorrelated principal components. The original data points can then be projected onto this new principal component space, allowing them to be represented in a way that maximizes any variance that is present. The formula used for principal component analysis through singular value decomposition is given in Equation 1.

$$X = USV^T \tag{1}$$

Where:
X= *[m X n]* dataset with original feature space representation
U= *[m X m]* matrix of left singular vectors
S=diagonal matrix of singular values
V= *[n X n]* matrix of left singular vectors

The first principal component will represent the direction in which the data demonstrates the greatest spread, or variance. The second principal component will be an axis perpendicular to this first axis, and will capture the direction of the second greatest spread (and so on so on). By discarding the higher-ranking principal components (which are often regarded as “noise”), the dimensionality of the feature set can be greatly reduced with little loss of information. The machine learning algorithms will then be able to operate more efficiently on the reduced, uncorrelated "feature" set formed by the principal components.

3.3.2 PCA Application

For this study, PCA was applied to each of the cases listed in Table 2. The performance of these cases was then reevaluated in the new principal component space instead of the original feature space. In the dimensionality reduction stage of this process, the original data was projected onto the number of principal components required to explain 95% of the variance present in the data, and all other components were discarded.

3.4 Classification with Support Vector Machines

This section details the theory behind SVMs and their application to binary and multiclass classification problems.

3.4.1 SVM Theory

In machine learning, support vector machines (SVMs) are supervised learning models are used for classification and regression analysis. It is a discriminative classifier that attempts to find the optimal separating hyperplane which maximizes the margin present between the two classes of training data. New data examples can then be categorized according to which side of the hyperplane they lie on. In two-dimensional space, this hyperplane is a line dividing a plane in two parts where in each class lay in either side. Figure 2 demonstrates a binary support vector machine.

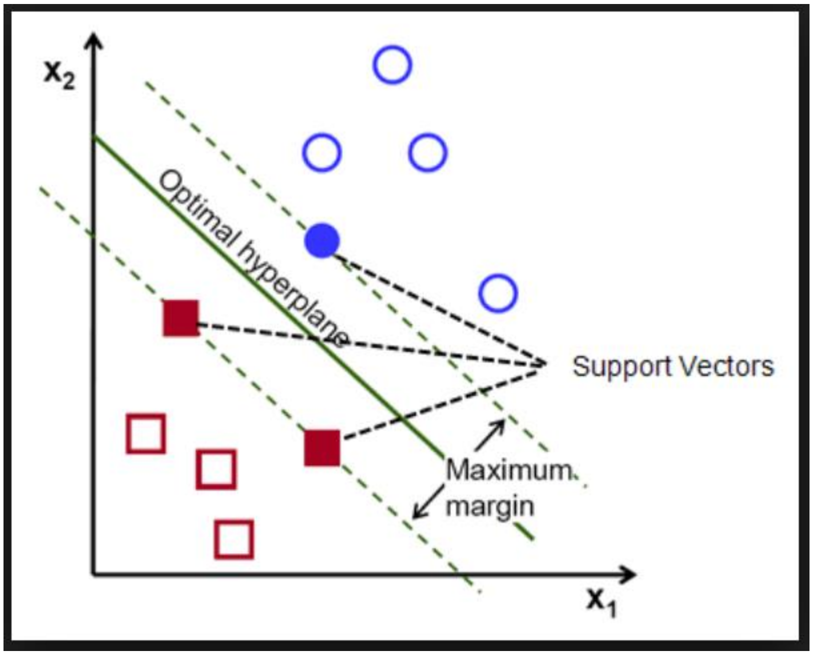


Figure 2: Demonstration of SVM model

Support vectors are the data points that lie closest to the decision surface (or hyperplane). They are the data points most difficult to classify. The hyperplane completely separates the vectors (cases) into two non-overlapping classes. The orientation of the hyperplane is adjusted until the margin, which is the distance between the hyperplane and the training instances on either side of the plane, is a large as possible. Typically, the training instances must be represented on a higher dimensional feature space to ensure that the different classes are linearly separable. The training instances are mapped to this feature space for classification using one of a variety of available kernel functions. While SVMs are typically used to perform binary classification, multiclass classification can be achieved by breaking the classification task into a set of multiple binary classification problems. The decision surface separating the classes is a hyperplane of the form is defined by Equation 2:

$$w^T x + b = 0 \tag{2}$$

Where: w = weight vector;
x = input vector;
b = the bias

3.4.2 Kernel Functions:

SVM algorithms use a set of mathematical functions that are defined as kernels. The function of a kernel is to map input data into a higher dimension so that it can be separated by a hyperplane, without the added computation expense of performing calculations in the high-dimensional space. SVM algorithms for different applications use different types of kernel functions in order to optimize classification performance. The different types of kernel functions investigated in this study are detailed below:

(1) Linear kernel: The simplest type of kernel, which performs well when working with high-dimensional feature spaces that are linearly separable.

(2) Polynomial kernel: (Eq 3) Used in applications where a non-linear decision surface is required.

$$k(\mathbf{x_i}, \mathbf{x_j}) = (\mathbf{x_i} \cdot \mathbf{x_j} + 1)^d \tag{3}$$

Where: d = the degree of the polynomial.

(3) **Gaussian kernel**: (Eq 4) Defined by a topology-type decision boundary, instead of a separating plane.

$$k(x, y) = \exp\left(-\frac{||x - y||^2}{2\sigma^2}\right) \tag{4}$$

3.4.3 SVM Application

In this study, support vector machines were used to classify students based on their academic performance for each of the five cases presented in Section 3.1, as well as the same five cases after undergoing dimensionality reduction through PCA. For the binary cases, a one-vs-one coding design was used to determine the optimal hyperplane orientation to separate the classes. The performance of linear, polynomial (3rd order), and Gaussian kernels was evaluated. Although SVMs are typically only applicable to binary classification tasks, they were also used to classify the multiclass cases in this study. For the K independent classes present in the multiclass cases, $K(K-1)/2$ binary SVM models were trained using the one-vs-all coding design, effectively transforming the multiclass classification task into a set of binary classification tasks. Receiver operating characteristic curves were used to determine how well the trained SVM models were able to classify the test data. The Matlab "fitecoc" and "predict" commands were used in the training and testing phases, respectively.

Observing the multiclass class label distribution in Figure 1, it is clear that available sample data is concentrated into Class 3, with very few sample instances available in the other class labels (Class 1 in particular). In order to compensate for this unbalanced dataset, the cost matrix associated with the SVM training procedure was adjusted to prevent the model achieving the highest accuracy by favoring the correct classification of Class 3 over the other Classes. This was done by increasing the cost of a false-positive classification of Class 3, as demonstrated in Tables 3a and 3b. By adjusting the cost matrix in this way, the model will be more negatively affected if it attempts to classify every sample as Class 3.

Table 3a: Original cost matrix

		Target Class			
		1	**2**	**3**	**4**
Output Class	**1**	0	1	1	1
	2	1	0	1	1
	3	1	1	0	1
	4	1	1	1	0

Table 3b: Modified cost matrix

		Target Class			
		1	**2**	**3**	**4**
Output Class	**1**	0	1	1	1
	2	1	0	1	1
	3	**2**	**2**	0	**2**
	4	1	1	1	0

4 Results

This section details the results and applicable discussion from the methods presented in Section 3.

4.4 Manual Feature Selection

After evaluating each of the 67 original features, it was determined that significant differences only existed between those students likely to pass and likely to fail in about 35 of the features. Some of the most significant findings are as follows:

- Goals for higher education: Having a goal for higher education was strongly correlated to better overall academic performance
- Being in a relationship: Relationships were highly correlated with a decrease in academic performance
- Past Academic Failures: Students with prior history of failures were more likely to demonstrate poor performance
- Education of Parents: Having more highly educated parents was correlated with improved performance. The mother's education level was of particular importance.
- Mother's Job: While the occupation of the father had little to do with academic success, those students with working mothers as opposed to unemployed mothers showed increased performance
- Study Time: An increase in the number of hours spent studying was significantly correlated with increase academic performance.
- Travel Time: Those students that lived closer to the school were at an increased risk of poor academic performance.

Many other features revealed slight differences between the different performance classes as well, but are not mentioned as a part of this report. By identifying these features, it is possible to determine which aspects of a student's background are contributing to their overall academic performance. Once students are classified into a particular performance group, these findings can be used to determine what corrective action can be taken in order to improve their academic standing.

4.5 Principal Component Analysis

Principal component analysis was conducted on the entire feature set for each of cases described in Section 3.1 (note that the feature set changes for each case, depending on which term grades were included and if they were included in a binary or multiclass format). The number of principal components required to explain 95% of the variance for each case are presented in Table 4. In addition, Figure 3 provides a visual representation of how the number of principal components for each case were chosen.

Table 4: Number of significant principal components required for each case study

Case Number	Description	# Principal Components
1	Multiclass (No term grades included)	36
2	Binary(1st term included)	37
3	Binary(Both terms included)	38
4	Multiclass(1st term included)	37
5	Multiclass(Both terms included)	37

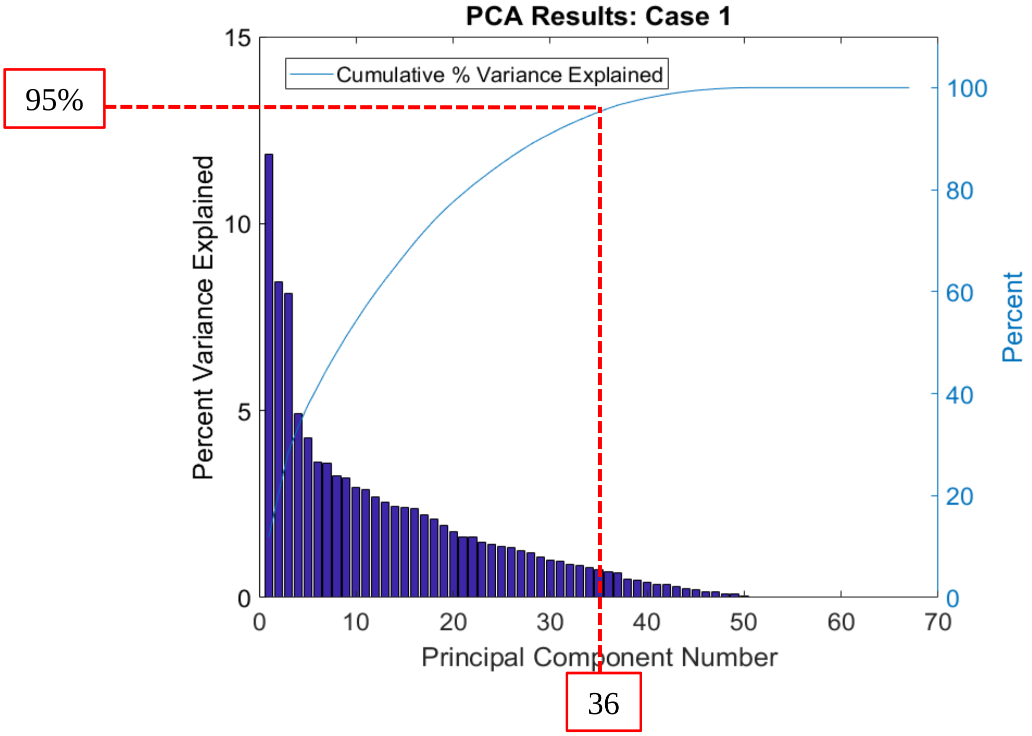


Figure 3: Representation of principal component selection process, Case study 1

Similar to the results observed in the manual feature selection, only 36-38 principal components are necessary to explain the majority of variance present in the data. Figure 4 demonstrates the improved data representation and decorrelation that is achieved by projecting the original data set onto the first three principal components. It is important to note here that more data points are now visible since the variance has been maximized.

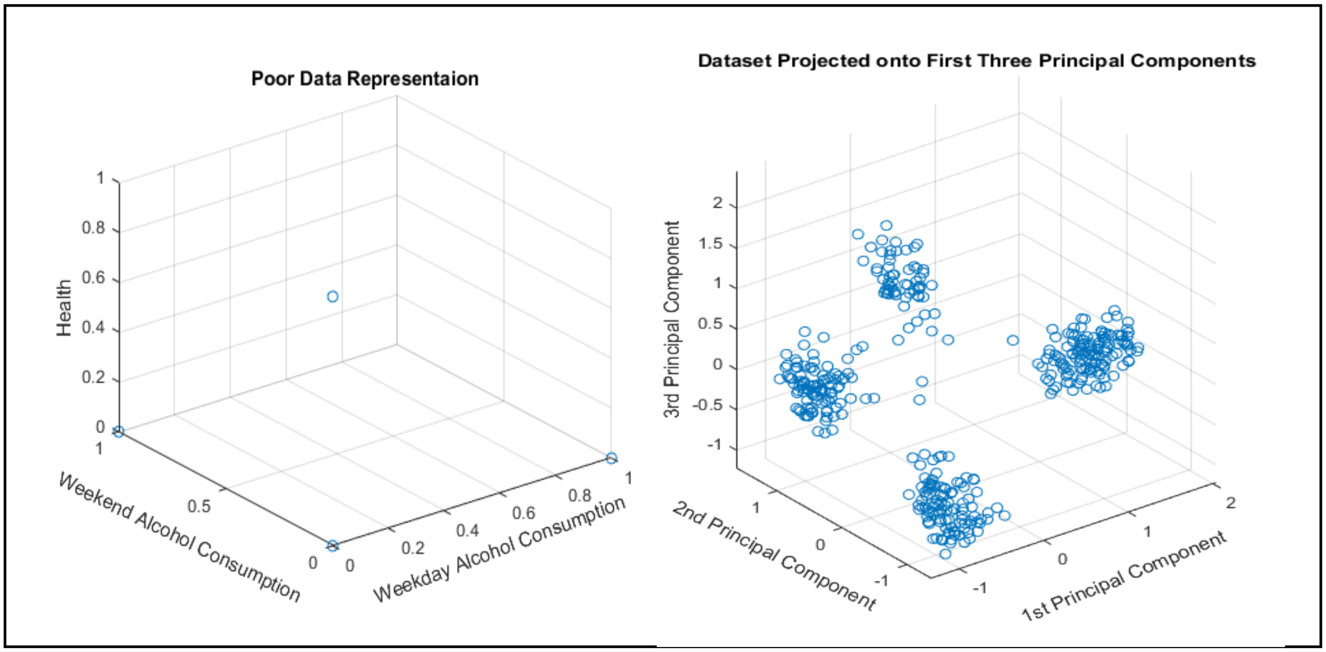


Figure 4: Improved Data representation after projection onto principal component space

4.3 SVM Results

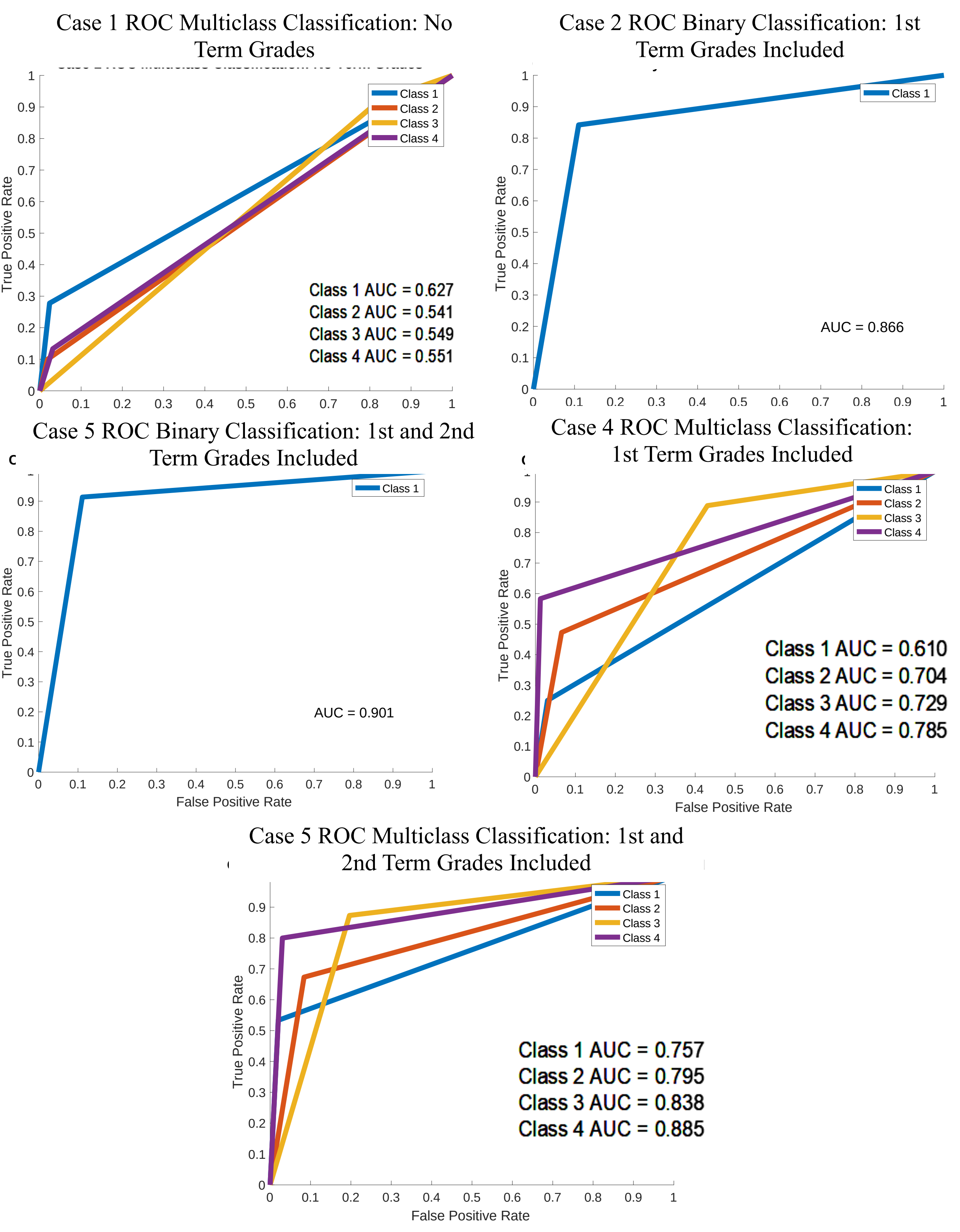
The performance of three different kernel types was tested on each of the five cases, with and without the application of PCA. The percent accuracy for each scenario is listed in Table 5.

Table 5 : SVM accuracy results

PCA?	Case Number	Kernel Used	Accuracy
No PCA	1	Linear	59.87
		Polynomial	58.28
		Gaussian	59.55
	2	Linear	85.98
		Polynomial	76.11
		Gaussian	64.01
	3	Linear	90.44
		Polynomial	85.03
		Gaussian	64.01
	4	Linear	70.02
		Polynomial	58.92
		Gaussian	56.05
	5	Linear	81.21
		Polynomial	74.84
		Gaussian	64.01
PCA	1	Linear	58.92
		Polynomial	52.23
		Gaussian	60.19
	2	Linear	84.39
		Polynomial	72.93
		Gaussian	62.42
	3	Linear	89.50
		Polynomial	74.84
		Gaussian	64.65
	4	Linear	73.24
		Polynomial	62.74
		Gaussian	58.92
	5	Linear	81.84
		Polynomial	69.43
		Gaussian	62.11

In nearly every case (with or without PCA), the classification accuracy of the SVM models trained using the linear kernel function outperformed the Gaussian and polynomial kernels. This indicates that the classes are linearly separable given the available features, and performance is not significantly improved by mapping the data to a higher feature space. In addition, the classification results found using the PCA reduced-dimensionality feature set were comparable to those found using the full feature set, despite one feature set being nearly half the size of the other. This suggests that the efficiency of the model can be improved by using PCA to reduce the dimensionality of the feature set, with little loss to the overall accuracy of the classification results.

The binary case proved to be sufficient in differentiating between the above and below-average students, with an accuracy of up to 90.44% when both the Term 1 and Term 2 grades were included. This information could be used to determine if a particular student is falling behind or likely to fall behind academically. Using the manual feature selection results outlined in 4.1, the appropriate interventions could be taken in order to remedy those behaviors that are able to be corrected, or to connect those students with the right resources to improve their performance. In order to understand which class labels were able to be most accurately predicted, the ROC curves were plotted for each case (for the linear kernel results without PCA) and are shown in Figure 5.



Case 1 ROC Multiclass Classification: No Term Grades

Case 2 ROC Binary Classification: 1st Term Grades Included

Case 5 ROC Binary Classification: 1st and 2nd Term Grades Included

Case 4 ROC Multiclass Classification: 1st Term Grades Included

Case 5 ROC Multiclass Classification: 1st and 2nd Term Grades Included

Figure 5: SVM results--Case study ROC curves

As expected, the classifier could more accurately predict a student's likely academic performance when information on the students' first and second term grades were included. Observing the ROC plot for Case 1, the area under each of the curves is close to 0.5. This indicates that the prediction accuracy of the classifier for the multiclass cases (without any term grade data) barely outperforms a random classifier. As more grades are included, the accuracy of the classifier increases for both the binary and multiclass cases. However, inclusion of the Second Term grades only provided a marginal improvement in the classifier performance for the binary case (an increase of~ 5% accuracy between Cases 2 and 3), and a slightly greater improvement for the multiclass case (an increase of~ 9% accuracy between Cases 4 and 5). This reveals that the classifier can accurately predict a student's likely academic performance following the availability of just their first term grades. In addition, the ROC curves for the multiclass cases reveal that Classes 3 and 4 were predicted with the most accuracy. This is significant, since it indicates that those students who are excelling the most academically could be easily identified (even after the inclusion of just the First Term grades in the feature set). This would facilitate the allocation of resources such as scholarship money or additional academic opportunities to the most deserving students. Conversely, the model was not able to differentiate Class 1 from the other classes as easily, so those students struggling the most (although able to be correctly characterized as "below average" by the binary classifier), could likely be misclassified in the multiclass classifier.

5 Conclusion

Overall, the SVM classifier models developed in this study are capable of accurately predicting a student's likely academic performance following the inclusion of data detailing their First Term grades. "Below Average" and "Above Average" binary classification was possible to within 85% accuracy after the First Term, and 90% accuracy following the Second Term. Although the overall classification capability of the multiclass classification models did not achieve the same accuracy as the binary models, the highest achieving students (Class 4) were able to be differentiated with an accuracy of 78.5% following the inclusion of their First Term grades and 88.5% following the inclusion of their Second Term grades. With the models developed, those students at risk of failure will be able to be identified and the proper intervention can take place. The appropriate action is determined with the help of those results obtained from the manual feature selection, which highlighted the most significant features impacting a student's academic performance. In addition, by successfully identifying those students who are projected to excel the most academically, the allocation of scholarship and educational opportunities can be facilitated.

A number of future steps can be taken in order to improve the performance of the SVM classifier model. The inclusion of additional training instances would greatly improve the ability of the classifier to differentiate between multiple classes, especially as it pertains to the Class 1 label in this study. Furthermore, improvements could be made in ensuring the unbalanced nature of the present data set does not lead the model to favor the accurate classification of one class over another. A hierarchical SVM scheme could be employed in order to first classify students as "above average" or "below average", followed by a second set of models capable of further classifying the new subsets students.

References

[1] A. Desai, N. Shah and M. Dhodi, "Student profiling to improve teaching and learning: A data mining approach," *2016 International Conference on Data Science and Engineering (ICDSE)*, Cochin, 2016, pp. 1-6.
doi: 10.1109/ICDSE.2016.7823947

[2] K. V. Krishna Kishore, Venkatramaphanikumar S and S. Alekhya, "Prediction of student academic progression: A case study on Vignan University," *2014 International Conference on Computer Communication and Informatics*, Coimbatore, 2014, pp. 1-6.
doi: 10.1109/ICCCI.2014.6921731

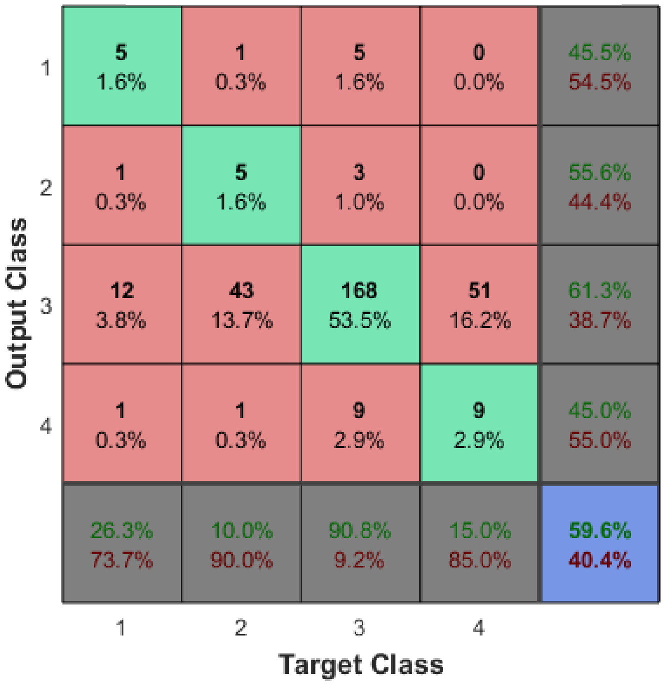
[3] Kularbphettong K. (2018) Analysis of Students' Behavior Based on Educational Data Mining. In: Silhavy R., Silhavy P., Prokopova Z. (eds) Applied Computational Intelligence and Mathematical Methods. CoMeSySo 2017. Advances in Intelligent Systems and Computing, vol 662. Springer, Cham.

Appendix A: Feature Set

	Feature Name
1	Activities
2	Address: Urban/Rural
3	Family Size (>3)
4	Family Help
5	Higher Education Goals
6	Internet Access
7	Nursery
8	Paid Tutoring
9	Parent Status
10	In a Relationship
11	School: GP
12	School help
13	Sex: M/F
14	Age 15
15	Age 16
16	Age 17
17	Age 18
18	Age 19
19	Age 20
20	Age 21
21	Age 22
22	Failures: 1
23	Failures: 2
24	Failures: 3
25	Failures: 4plus
26	Father: No Education
27	Father: Primary Education
28	Father: 5th thru 9th Grade
29	Father: Secondary Education
30	Father: Higher Education
31	Mother: No Education
32	Mother: Primary Education
33	Mother: 5th thru 9th Grade
34	Mother: Secondary Education
35	Mother: Higher Education
36	Mother Job: Teacher
37	Mother Job: Services
38	Mother Job: Health
39	Mother Job: At Home
40	Mother Job: Other
41	Father Job: Teacher
42	Father Job: Services
43	Father Job: Health
44	Father Job: At Home
45	Father Job: Other
46	Reason: Course
47	Reason: Reputation
48	Reason: Home
49	Reason : Other
50	Study Time: < 2hrs
51	Study Time: 2 to 5hrs
52	Study Time: 5 to 10hrs
53	Study Time: >10hrs
54	Travel Time: < 15min
55	Travel Time: 15 to 30min
56	Travel Time: 30min to 1hr
57	Travel Time: >1hr
58	Guardian: Father
59	Guardian: Mother
60	Guardian: Other
61	Family Relationship Quality
62	Free time
63	Go out
64	Weekday Alcohol Consumption
65	Weekend Alcohol Consumption
66	Health
67	Absences

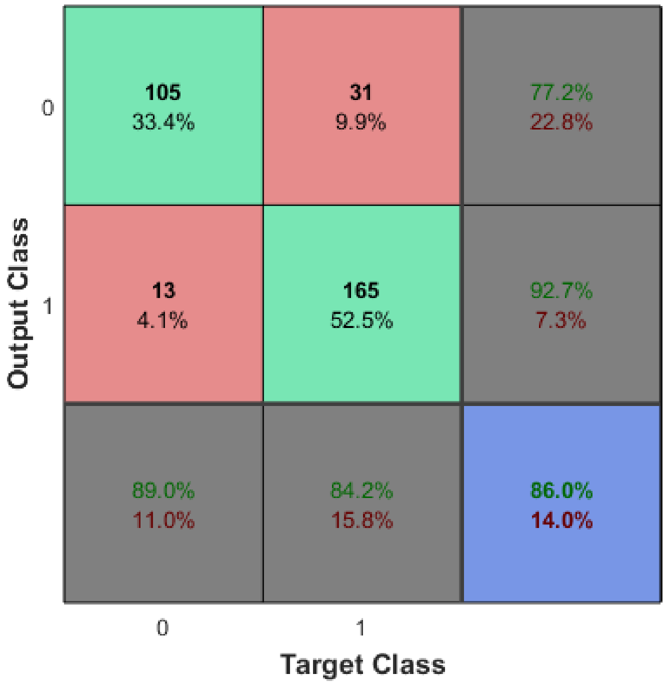
Appendix B: Confusion Matrices

Case 1 Confusion Matrix



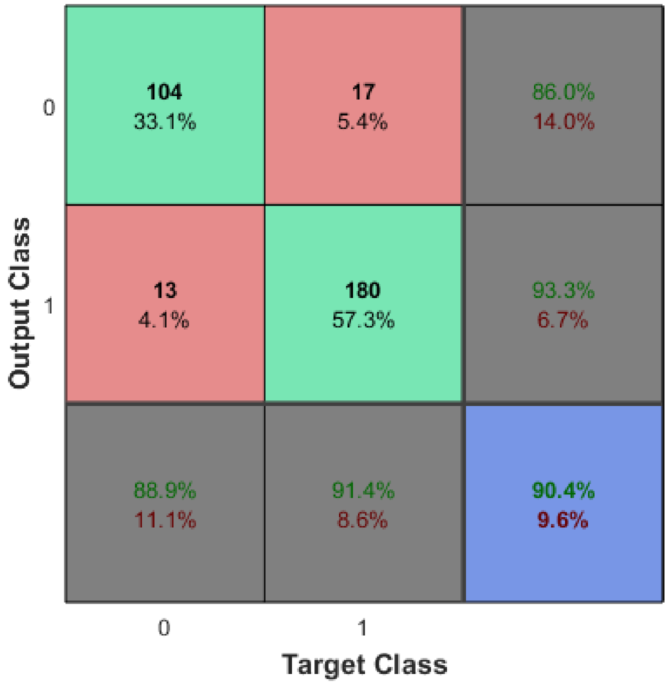
	1	2	3	4	
1	**5** 1.6%	**1** 0.3%	**5** 1.6%	**0** 0.0%	45.5% 54.5%
2	**1** 0.3%	**5** 1.6%	**3** 1.0%	**0** 0.0%	55.6% 44.4%
3	**12** 3.8%	**43** 13.7%	**168** 53.5%	**51** 16.2%	61.3% 38.7%
4	**1** 0.3%	**1** 0.3%	**9** 2.9%	**9** 2.9%	45.0% 55.0%
	26.3% 73.7%	10.0% 90.0%	90.8% 9.2%	15.0% 85.0%	**59.6%** **40.4%**

Case 2 Confusion Matrix



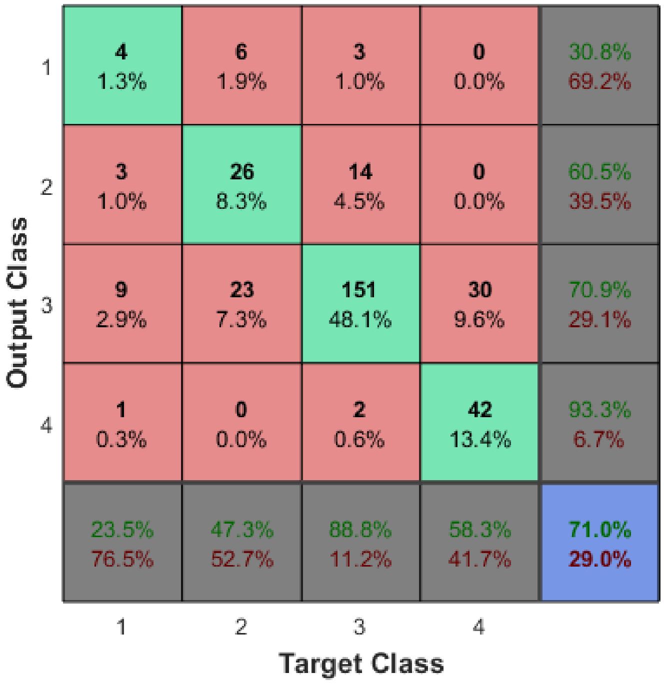
	0	1	
0	**105** 33.4%	**31** 9.9%	77.2% 22.8%
1	**13** 4.1%	**165** 52.5%	92.7% 7.3%
	89.0% 11.0%	84.2% 15.8%	**86.0%** **14.0%**

Case 3 Confusion Matrix



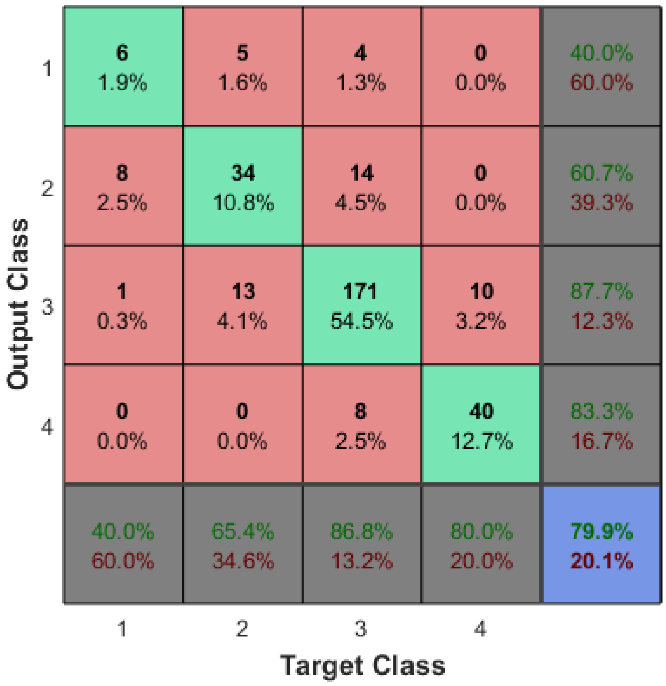
	0	1	
0	**104** 33.1%	**17** 5.4%	86.0% 14.0%
1	**13** 4.1%	**180** 57.3%	93.3% 6.7%
	88.9% 11.1%	91.4% 8.6%	**90.4%** **9.6%**

Case 4 Confusion Matrix



	1	2	3	4	
1	**4** 1.3%	**6** 1.9%	**3** 1.0%	**0** 0.0%	30.8% 69.2%
2	**3** 1.0%	**26** 8.3%	**14** 4.5%	**0** 0.0%	60.5% 39.5%
3	**9** 2.9%	**23** 7.3%	**151** 48.1%	**30** 9.6%	70.9% 29.1%
4	**1** 0.3%	**0** 0.0%	**2** 0.6%	**42** 13.4%	93.3% 6.7%
	23.5% 76.5%	47.3% 52.7%	88.8% 11.2%	58.3% 41.7%	**71.0%** **29.0%**

Case 5 Confusion Matrix



1	6 1.9%	5 1.6%	4 1.3%	0 0.0%	40.0% 60.0%
2	8 2.5%	34 10.8%	14 4.5%	0 0.0%	60.7% 39.3%
3	1 0.3%	13 4.1%	171 54.5%	10 3.2%	87.7% 12.3%
4	0 0.0%	0 0.0%	8 2.5%	40 12.7%	83.3% 16.7%
	40.0% 60.0%	65.4% 34.6%	86.8% 13.2%	80.0% 20.0%	79.9% 20.1%
	1	2	3	4	