Predicting Students' Graduation Outcomes through Support Vector Machines

Yulei Pang[a], Nicolas Judd[b], Joseph O'Brien[b], Michael Ben-Avie[b]

[a] Department of Mathematics, Southern Connecticut State University, USA, pangy1@southernct.edu

[b] Office of Assessment and Planning, Southern Connecticut State University, USA, {benaviem1, juddn1, obrienj14}@southernct.edu

***Abstract*—Low graduation rate is a significant and growing problem in U.S. higher education systems. Although previous studies have demonstrated the usefulness of building statistical models for predicting students' graduation outcomes, advanced machine learning models promise to improve the effectiveness of these models, and hone in on the "difference that makes a difference" not only on the group level, but also on the level of the individual student.**

In this paper we propose an ensemble support vector machines based model for predicting students' graduation. Up to about 100 features, including a set of psychological-educational factors, were employed to construct the predicting model. We evaluated the proposed model using data taken from a state university's longitudinal, cohort data sets from the incoming classes of students from 2011-2012 (n=350). The experimental results demonstrated the effectiveness of the model, with considerable accuracy, precision, and recall. This paper presents the results of analysis that were conducted in order to gauge the predictive capability of a machine learning algorithm to predict on-time graduation that took into consideration students' learning and development.

***Index terms*—** graduation outcome, machine learning, support vector machine, higher education

I. Introduction

Predicting graduation outcomes has been an area of interest for higher education institutions for decades because it allows them to develop strategic programs that will help to improve student performance and retention rates [1][2][3]. According to the Department of Education's Graduation Rate Survey, the national graduation rate (within 6 years) from the first institution attended for first-time, full-time bachelors degree-seeking students at 4-year postsecondary institutions was 60% in 2014 [4].

Methods of predicting graduation and other performance metrics have been discussed in the education literature for some time. These publications have discussed metrics at all levels of education. Typical methods in institutional research involve analyzing data with foundational statistical techniques such as t-tests and ANOVAs. More advanced techniques germane to the field include Cluster Analysis, Path Analysis and various regression techniques. For instance, Koker et al. have used logistic regression analysis to identify significant predictor variables of graduation status [5]. Cook et al. discussed path analysis as a method for investigating students progress through graduate programs [6]. Alzahrani et al. used multiple logistic regression to analyze predictor variables in a Dental Hygiene Baccalaureate program [7].

Machine Learning techniques have also seen some applications in education literature in recent years. In 2003, Kotsiantis et al. demostrated the effectiveness of a Naive Bayes algorithm in predicting dropout for students enrolled in a distance learning course on 'informatics' at the Hellenic Open University [8]. Related work also covered the topic of predicting drop-out in online courses offered at universities and in massive open online courses (MOOCs) [9]. Another such study conducted by Er et al. employs three techniques: instance-based learning Classifier, Decision Tree, and Naive Bayes [10].

While there has been a deluge of papers focusing on online coursework applying machine learning methods and a longstanding interest in predicting performance metrics at the university level, to these researchers' best knowledge nothing has been published applying advanced machine learning techniques to a robust cohort of students in order to predict their graduation outcomes. In light of this, we propose a machine learning ensemble, in particular, a voting population of Support Vector Machines (SVM) parameterized by a Simulated Annealing (SA) algorithm run on a cross-validated dataset predicting their graduation within four years of a subset of a longitudinal cohort dataset consisting of survey response items, demographic information and university performance metrics (e.g. GPA, Earned Credit).

Included in the dataset were students who had completed a continuing student survey when they were sophomores and juniors from the incoming classes of 2011 and 2012. As shown in Figure 1, of the students who entered the university in 2011, 51.7% of them graduated on-time. For the incoming class of 2012, this fell to 34.4% of students who completed the continuing student survey. The aim is to develop a model that can be used in the construction of an "at-risk student identification system" which seeks to identify the conditions that lead to student withdrawal and strengthen the conditions that promote students academic success and young adult development. The results presented is this paper serve as a first step towards the construction of such a system.

This paper makes the following contributions:

1) Proposes a predictive model of students' graduation outcomes based on ensembled support vector machines
2) Employed up to 100 students features, including demographic, functional, and psychological-educational features for building the prediction model

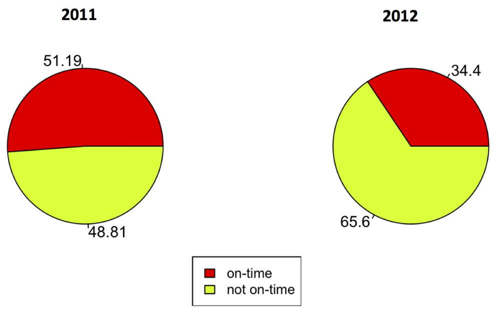

Fig. 1. Graduate outcome for students enrolled since 2011 v.s. 2012.

3) Evaluates the proposed technique based on an experimental study

The rest of this paper is organized as follows. Section II presents the background and the motivation of this study. Section III describes the technical details of the machine learning algorithms employed. Section IV provides a detailed explanation of the study including participants, data collection and analysis. Section V concludes the study and discusses possible future research directions.

II. Background

Southern Connecticut State University (SCSU) is a comprehensive metropolitan public university located in New Haven, Connecticut. SCSU promotes a data-driven process of educational change to provide opportunities for all students in the pursuit of knowledge, and helps those flourish who have not done so before. The Office of Assessment and Planning conducts longitudinal, cohort studies in order to identify patterns and anomalies in student persistence and graduation.

In 2007, SCSU initiated a comprehensive First-Year Experience (FYE) Program to promote student engagement, improve students' academic competencies, and boost retention rates. Only 50% of incoming students were enrolled in a first-year experience seminar in the program's first year; this provided an opportunity for the university to measure the impact of an FYE seminar on student success. The seminar participants demonstrated significantly higher rates of retention, higher GPAs, and more credits earned than students who did not participate in the seminar even three years later. This study identified a psychological-educational factor that is amenable to change–future orientation–for explaining the difference in outcomes between the FYE seminar and non-seminar students. The FYE self-assessment survey provides a measure of the nature and quality of students' college experiences [11].

The Southern Experience Survey (SES) was designed as a follow-up to the FYE self-assessment surveying students in their sophomore and junior years. It is a continuing student survey designed to elicit more information on the reasons why students stay or leave the university. This continuing student survey emerged during meetings of the university's Student Success Taskforce. The core of the survey was developed by students in a research methods course under the guidance of the same team that developed the FYE Self-Assessment. A taskforce, comprised of faculty and staff, then refined the survey. The SES continuing student survey was designed to complement the FYE Self-Assessment. Both measure the impact of college on students' learning and development.

During the 2014 pilot, an invitation to complete the online SES was sent to sophomores and juniors. The following year, both an online and paper version were administered. In 2016, the paper version was administered to juniors in their capstone courses and an online version was administered to sophomores.

All first-time, full-time undergraduate students are included in the longitudinal, cohort studies. The students are followed from New Student Orientation through graduation from the university, or subsequent enrollment in other colleges and universities. As each incoming class enters the university, a cohort dataset is established. A cohort dataset initially contains such demographic information as high school rank, high school GPA, SAT scores, gender, ethnicity, residential status, registration with Veterans Services and the Disability Resource Center, and English and Math placements. Each year, new data are added, including earned credits, cumulative GPA, registration status, and scores on surveys and direct performance-based assessments. Students' ID numbers are used to link their demographic characteristics with their scores from surveys and assessments to create comprehensive cohort datasets. There are currently 10 longitudinal, cohort datasets (n = 16,263).

III. Methodology

In this section, we will describe the algorithms that are referred to in this paper including support vector machine, ensemble learning.

A. Support Vector Machine

Support vector machine (SVM) are machine learning algorithms widely used for classification and prediction problems [12][13]. A SVM utilizes a kernel function to perform both linear and non-linear classifications. For a simple binary high-dimensional linear classification problem, the SVM algorithm builds a hyper-plane with the intention of maximizing the distance between the hyperplane and the nearest data points on each side, i.e. the margins. Figure 2 illustrates the mechanic of a linear classification for binary prediction [14]. SVMs are widely used in solving real problems posed in various application domains. They have been applied to text categorization as their application can significantly reduce the need for labeled training instances [15]. Their application to the classification of images have also been demonstrated [16]. Experimental results show that SVMs yield significantly higher search accuracy [10][8]. The medical sciences have used SVMs to classify gene proteins with high feasibility and accuracy predictions [17].

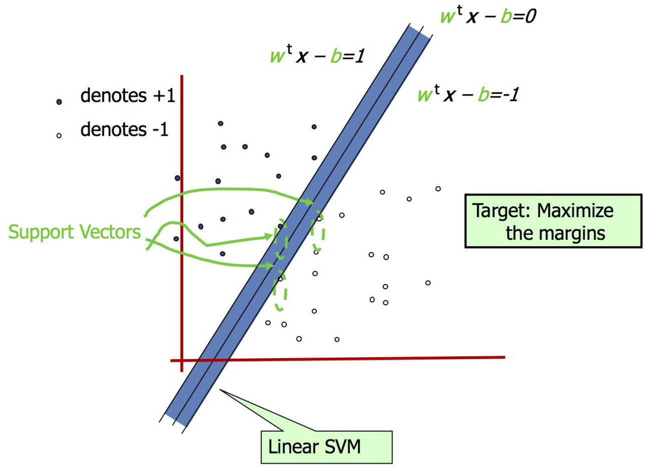

Fig. 2. A linear support vector machine.

Suppose that a training set S containing n data points is given as:

$$S = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_i, y_i), ..., (\mathbf{x}_n, y_n)\}$$

where $\mathbf{x}_i \in R^p$, i.e., p-dimensional space, and $y_i \in \{-1, +1\}$. The goal of SVM is to find a seperating hyperplane:

$$\mathbf{w}^t\mathbf{x} - b = 0 \tag{1}$$

and to solve the optimization problem:

$$min\left(\frac{1}{2}\mathbf{w}^t\mathbf{w} + C\sum \xi_j\right) \tag{2}$$

subject to the following constraints:

$$y_j\left(\mathbf{w}^t\mathbf{x}_j - b\right) \geq 1 - \xi_j \tag{3}$$

$$\xi_j \geq 0 \tag{4}$$

where $\mathbf{x}_j$ are the samples not on the correct side of the separating plane, C is the penalty parameter, and ξ_j is a slack variable. A slack variable is a variable that is added to an inequality constraint to transform it to an equality. SVM is a linear classifier, but in general, the feature vectors might not be linearly separable. To overcome this issue, the kernel trick is used. The original input space is mapped into a high-dimensional feature space using kernel functions where it becomes linearly separable. Some commonly used kernel functions are linear, polynomial, sigmoid and radial basis kernel function (RBF) [18]. The RBF is the kernel function most used with SVM and is used in our study.

$$k(x_i, x_j) = exp(-\gamma \parallel x_i - x_j \parallel^2) \tag{5}$$

Two parameters, C and γ , must be appropriately set in SVM. The accuracy will be very high in the training stage and very low in the testing stage, when the value of C is set too large [19]; while an extremely small value of γ may result in under-fitting, and excessively large value of γ may lead to over-fitting [18][20].

To obtain the better parameter values of SVM, simulated annealing (SA) approach will be performed to adjust the values for C and γ, which makes the optimal separating hyper-plane obtainable in both linear and non-linear classification problems [21].

B. *Ensemble Learning*

Ensemble learning is a paradigm where multiple learners, called base learners, are trained to solve the same problem [22][23]. Different from the single learning approaches which learn one hypothesis from training data, ensemble methods construct a set of hypotheses and combine them for better predictions. Figure 3 depicts a general picture on how to make use of ensemble learning.

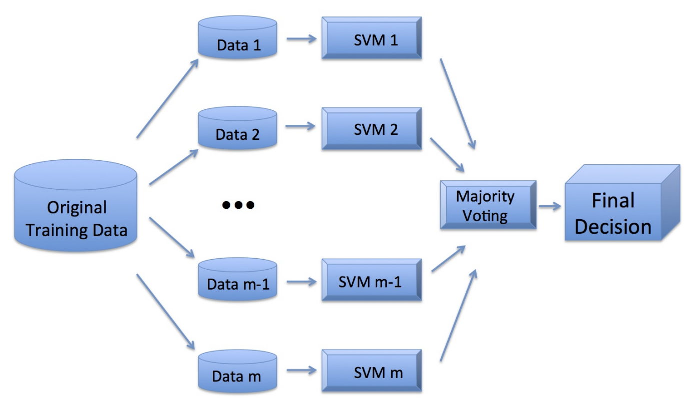

Fig. 3. Ensemble learning in linear support vector machine.

Several separate datasets can be derived from the original training data by different approaches. For each dataset, a model is trained and built. Each model produces a classification when applied to the test data. The final result, however, is an aggregation of all predictions, usually based on the majority votes of all the models built. It can be proved that ensemble learning can improve the prediction accuracy significantly. The intent of ensemble learning is to improve prediction accuracy and consistency. Suppose the accuracy of a basic learner is p, a probability value between [0, 1]. The likelihood that a basic learner may produce a false prediction is therefore $1-p$. When there are n learners, the possibility of making correct predictions P by majority voting is given in Formula 6:

$$P = \sum_{m=\lceil \frac{n}{2} \rceil}^{n} C_n^m (1-p)^{n-m} p^m \tag{6}$$

Where:

- n is the total number of learners;

- m is the number of learners that have predicted the labeling correctly;
- C_n^m is the combination of m out of n observations or learners;
- $C_n^m(1-p)^{n-m}p^m$ is the binomial distribution for each learner, i.e. selection m out of n values forms a binomial distribution.

It is easy to prove that the value of P is greater than p when $p > 1/2$. The larger the n, the number of learners, the more accurate the prediction. In recent years, besides the traditional classification techniques, the application of ensemble learning in detecting mislabeled data has been also introduced [24], [25]. The dataset can be divided into two training and test sets interactively. Each data point is tested several times. The final decision for each data point is performed by a majority voting procedure.

IV. A Cohort Study

In this section we evaluate the proposed technique by a cohort study. We discuss the data source, data curation, experimental procedures, algorithm, and report the results.

A. Data Source

For the purpose of this analysis, the cohort datasets of the incoming classes of 2011 and 2012 were considered. In the cohort datasets, the following types of information appear: demographic characteristics pre-college achievement, survey results, and college progress.

This research was conducted using data collected through the Southern Experience Survey (SES) combined with identified student data provided by the office of Institutional Research at Southern Connecticut State University, a medium-sized north-eastern public university, a part of the Connecticut State University system. Here are some sample questions from the survey:

"In what year did you enter SCSU?"

"I made the right decision in choosing SCSU."

" My professors' attitudes about SCSU influence my decision to remain here."

"My confidence in my academic skills and abilities has increased this semester."

"I am driven to succeed in college."

B. Data Curation

In order to target the sophomores and juniors who took the SES, students from outside of the incoming classes of 2011 and 2012 were removed from consideration. To get a proper measure of the time students took to graduate, transfer students were also excluded. 59.8% of the remaining students were from the incoming class of 2012 while 40.2% of the students were from the incoming class of 2011. Of the students who entered the university in 2011, 51.7% of them graduated on-time. For the incoming class of 2012, this fell to 34.4%.

Missing values were then removed from the data using case-wise deletion. Categorical variables were one-hot encoded, while the numeric variables were normalized. This resulted in fairly sparse data. While the original dataset had about 100 variables, the recoded data had approximately 150 variables. Some representative input variables are shown in Table I.

In the processed dataset, 44.0% of the students have graduated on-time and 56.0% have not. This is slightly different from the original dataset in which 41.3% had graduated on-time and 58.7% had not.

On the other hand, the within group percentages differ more noticeably. In 2011 54.5% of the students graduated on-time while in 2012 37.1% had graduated on-time. Cross-folds were created using proportionate allocation stratification. Folds were constructed such that the prevalence within each fold reflected the overall prevalence of the outcome variable.

C. Classifier

The research question, determining the predictability of on-time graduation, is a binary classification problem. As it is important to choose the best machine learning model for the question and data, we did some preliminary study in which different machine learning models including Random Forest, Linear SVM, and logistic regression were tried. As a result of this preliminary research, we found that Radial SVM performed best but it was clear that the parameters needed adjustment. Thus, the major results presented here were obtained through a Radial basis function kernel SVM the gamma and cost parameters of which were tuned by a Simulated Annealing algorithm (SA-SVM). A number of different approaches using this method were implemented. They are detailed below.

The first of these SVM models used 10-fold cross-validation (as shown in Fig 4) with a radial kernel support vector machine. In 10-fold cross-validation, we first divide the training set into 10 subsets of equal size. Sequentially one subset is tested using the classifier trained on the remaining 9 subsets. For our study the folds were constructed using a proportionate allocation stratification routine, meaning the prevalence of each fold reflects the overall prevalence of the dataset. Within the 9 folds of the training set, an oversampling procedure was performed so that the training set was comprised of 50% response cases and 50% non-response. Thus, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data that are correctly classified. The cross-validation procedure was implemented to address concerns about over-fitting. However, the kernel parameters γ and C were iteratively parametrized by a simulated annealing algorithm based on the performance of the models on the test set. Therefore, there was still some concern that models generated using this method would perform well on the training/test data but fail to generalize.

In order to address this concern about overfitting we constructed models using a stricter cross-validation methodology in which we divided the 10 folds into training (8 folds), test (1 fold) and validation (1 fold) sets. The construction of the

TABLE I
FEATURE LIST

Category	**Feature**	**Value Range**	**Description**
	Gender	$\{0, 1\}$	0 indicates female, and 1 indicates male;
	Age	$x \in N$ (Nature number)	the actual age of a student;
	New Haven Resident	$\{0, 1\}$	1 indicates Yes, and 0 indicates No;
Demographic	Nationality	$\{1, 2\}$	1 indicates US citizen, and 2 indicates non US citizen;
	Residency	$\{0, 1\}$	0 indicates In-State, and 1 indicates Out-of State,
	...	...	...
	GPA in Semester i	$0 \leq x \leq 4$	the actual GPA a student obtained in his/her ith semester;
Functional	Credits Earned in Semester i	$0 \leq x \leq 25$	the course credits a student obtained in his/her ith semester;
	Scholarship	$\{0, 1\}$	0 indicates No, and 1 indicates Yes,
	...	...	...
	Major	$\{1, 2, 3, 4, 5\}$	response to survey item: ”My major is a good fit for me”
Student-Environment Fit	Belonging	$\{1, 2, 3, 4, 5\}$	response to survey item: ”Southern is a big part of my life.”
	Diligence	$\{1, 2, 3, 4, 5\}$	response to survey item:”I settle for just passing courses.”
	...	...	...

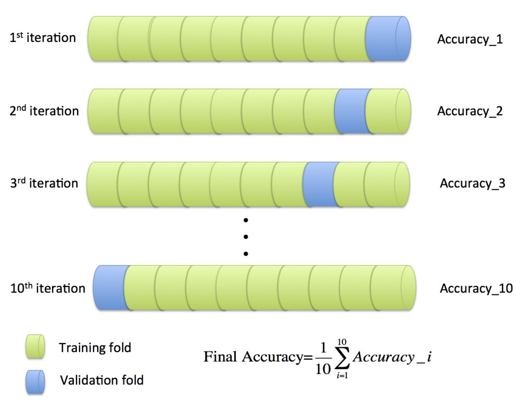

Fig. 4. 10-fold cross validation

folds for cross-validation were created using the same stratification procedure. Only the training sets were oversampled. The training and testing sets were used to parametrize the support vector machines, again via simulated annealing. This ultimately produces $n(n-1)$ models for n validation sets. After the models have been created, each is tested against its corresponding validation set to produce an overall accuracy. This score is averaged across all the models to produce a mean and standard deviation of accuracy. This method proved less effective at classification. We suspect that this is, at least in part, due to the significant decrease in training cases. As such, we expect the inclusion of more cases in future research to improve the accuracy of models generated in this way.

In an attempt to improve on the classification rates an ensemble of such models was constructed. Here, instead of testing individual models against their validation sets, models corresponding to the same validation set are placed in an ensemble voting scheme. Each model produces a label for each case in the validation set, referred to as a vote. The tally of votes for a case is considered to give the ultimate classification for the case.

D. Algorithm

Algorithm 1 provides psuedocode for SA-SVM and ensembled SA-SVM. The algorithm includes 3 stages. Line 1 to 3 demonstrates the steps related with variable initialization and data partition. Line 5 through line 7 involves building individual SA-SVMs. Line 8 to 16 is about the major voting mechanism.

Algorithm 1 Identifying the students who fail to graduate on-time through SVM ensembles.

Require: Inputs:
1) $TR = \{(X_1, Y_1), (X_2, Y_2), ..., (X_m, Y_m)\}$, which is a list of m students with their features information X and ground truth graduation outcome Y
2) S, which is a given student, with his feature information
3) Simulated Annealing SVM learning algorithm L
4) number of partitions I

Ensure: the graduation outcome of student S

```
1:  gra = 0 {initialization}
2:  non = 0 {initialization}
3:  Randomly divide TR into I partitions TR_1, ..., TR_I
4:  for n = 1 to I do
5:     Randomly pick up I − 1 partitions of TR as P_n
6:     SVM_n = L(P_n, TR − P_n) {Train an individual learner}
7:     label = SVM_n(S) {test S}
8:     if labels indicates graduation on time then
9:        gra++
10:    else
11:       non++
12:    end if
13: end for
14: if non ≥ gra then
15:    conclude S can not graduate on time {take the major voting}
16: end if
```

E. Evaluation Metrics

Four key terms are defined for assessing the performance of a classifier: true positives(tp), true negatives(tn), false positives(fp), and false negatives(fn). The terms positive and negative refer to the classifier's prediction, also known as the expectation, and the terms true and false refer to whether that prediction corresponds to the external judgment, also known as the observation [26]. These terms and their associations are illustrated in Table II for the prediction of students' on-time graduation:

TABLE II
A BINARY CLASSIFICATION.

	Truly graduated	**Truly not graduated**
Predicted graduate	tp	fp
Predicted not graduate	fn	tn

Three major measurement metrics are usually used to assess how well a binary classification performed: precision, recall and overall accuracy [26]. The accuracy is the degree of closeness of measurements of a quantity to the quantity's actual (true) value. The precision, also called reproducibility or repeatability, is the degree to which repeated measurements under unchanged conditions show similar results. Recall (in this context also referred to as the true positive rate or sensitivity) is the ratio of true positives over the sum of true positives and false negatives.

$$Accuracy = \frac{tp + tn}{tp + fp + fn + tn} \tag{7}$$

$$Precision = \frac{tp}{tp + fp} \tag{8}$$

$$Recall = \frac{tp}{tp + fn} \tag{9}$$

F. Results

TABLE III
PERFORMANCE OF SA-SVM

#year	**2011**		**2012**	
#measure	**mean**	**SD**	**mean**	**SD**
Accuracy	80.59	10.53	78.14	9.43
Precision	80.71	11.88	72.06	15.40
Recall	87.85	11.39	71.67	14.55

TABLE IV
PERFORMANCE OF NESTED-CV SA-SVM

#year	**2011**		**2012**	
#measure	**mean**	**SD**	**mean**	**SD**
Accuracy	69.95	8.04	67.81	7.17
Precision	72.51	8.37	57.86	11.16
Recall	76.97	10.29	57.19	11.75

TABLE V
PERFORMANCE OF ENSEMBLED-SVM

#year	**2011**		**2012**	
#measure	**mean**	**SD**	**mean**	**SD**
Accuracy	70.85	12.02	71.47	9.43
Precision	72.46	12.42	63.39	16.65
Recall	78.66	15.47	60.25	19.18

Table III, IV and V present the results in terms of means and standard derivation for accuracy, precision, and recall, respectively. Overall the results are good, and the average value is between 60% to 85%. The high accuracy demonstrates that the model can classify students who graduate on time or not correctly most of the time. The high recall ratio indicates that we can effectively identify the students who are most at risk of not graduating on-time so that an early warning could be created. The high precision shows that once a student is labeled as at-risk by the model, it is rarely wrong. In order to have a better insight of the data distribution, we provide the box plot of these data, as shown in Figure 5, 6, 7, 8 9, and 10.

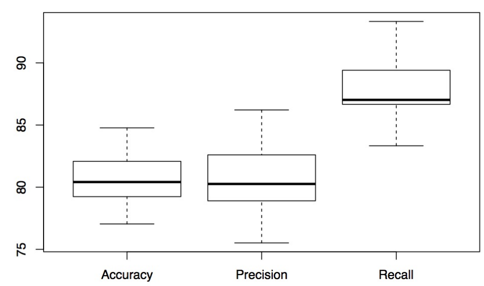

Fig. 5. The boxplot of result using SA SVM method for 2011 data

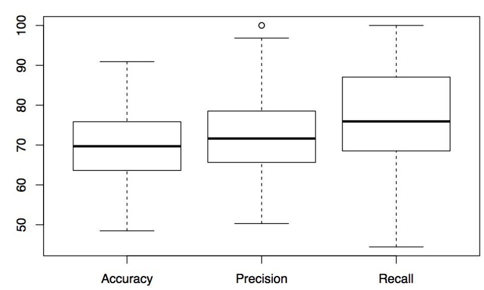

Fig. 6. The boxplot of result using nested-cv SA-SVM method for 2011 data

G. Threats to Validity

Similar to all other empirical studies, the experiments reported in this paper are quasi-experimental and thus prone to

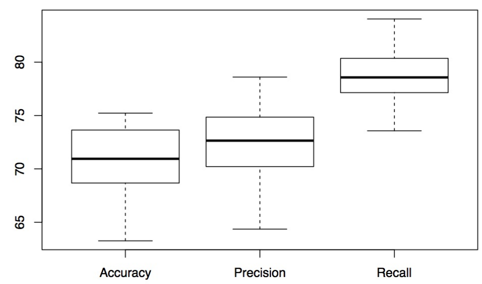

Fig. 7. The boxplot of result using ensemble SVM method for 2011 data

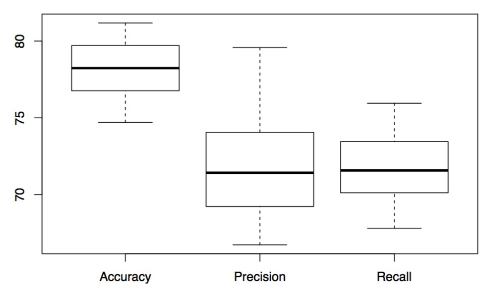

Fig. 8. The boxplot of result using SA SVM method for 2012 data

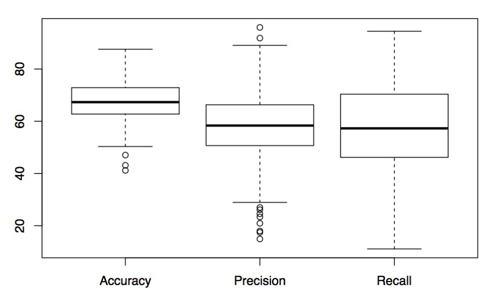

Fig. 9. The boxplot of result using nested-CV SA-SVM method for 2012 data

possible experimental threats.

Internal threats: The research was conducted based on the data collected from students at only one university of a seventeen-campus system. During the first year of the study, the survey was administered online and students self-selected to take the survey. In subsequent years, a paper survey was administered during capstone courses. While the survey was developed by the university's Student Success Taskforce, the survey did not evaluate the implementation of a particular intervention.

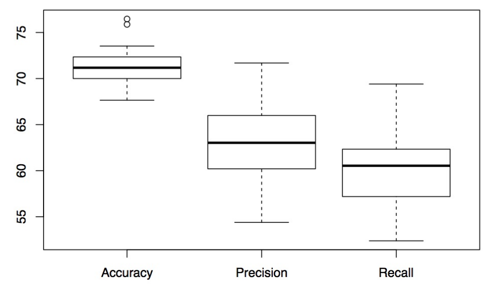

Fig. 10. The boxplot of result using ensemble SVM method for 2012 data

External threats: In this experimental study, for the purpose of performing the classification and prediction, we used some open source packages in the R and python programming languages, where external threats may be introduced [27].

Construct threats: The goodness of a classification model usually partly relies on the selection of features. In this study, we employed three categories of students' information for building the predictive model. The usage of other students information, as features, may affect the results. Other sources of students' information may influence results. Also, we measured the performance of the proposed technique based on metrics including accuracy, precision and recall. In practice, there are some other metrics to validate a classification model and the results may be different.

V. Conclusions and Future Work

This paper proposes a technique which employs a support vector machine ensemble to predict on-time graduation rates of students. Further experiments are needed to replicate and validate the results presented in this paper. In particular, experiments need to be conducted in other institutions. Also, the idea of the proposed technique is to use machine learning algorithms, which are grounded on features study. In practice, there are some other ways to construct features. Future research will explore further in this direction. We will use feature selection techniques to identify the most important factors. Understanding which factors are the most critical to student classification will inform the nature and quality of interventions targeted at improving graduation outcomes and student performance. There exists some other clustering or classification techniques, like the techniques we used in a previous study [28], and we will explore the possibility of applying them when developing an early warning system for withdrawal from higher education.

Acknowledgment

The authors are grateful for the excellent contributions of the following students who worked on this study in the Office of Assessment and Planning: Patrick Iben; Brain Darrow, Jr.; and Kevin Kuna. The authors are appreciative of the data and support provided by Dr. Richard Riccardi, Associate Vice President of Institutional Effectiveness.

References

[1] T. Wongkhamdi and P. Seresangtakul, "A comparison of classical discriminant analysis and artificial neural networks in predicting student graduation outcomes," in *Proceedings of the Second International Conference on Knowledge and Smart Technologies*, vol. 2010, 2010.

[2] G. Zhang, T. J. Anderson, M. W. Ohland, and B. R. Thorndyke, "Identifying factors influencing engineering student graduation: A longitudinal and cross-institutional study," *Journal of Engineering Education*, vol. 93, no. 4, pp. 313–320, 2004.

[3] R. D. Wessel, J. A. Jones, L. Markle, and C. Westfall, "Retention and graduation of students with disabilities: Facilitating student success." *Journal of Postsecondary Education and Disability*, vol. 21, no. 3, pp. 116–125, 2009.

[4] Undergraduate retention and graduation rates. [Online]. Available: https://nces.ed.gov/programs/coe/indicator_ctr.asp

[5] M. Koker and D. D. Hendel, "Predicting graduation rates for three groups of new advanced-standing cohorts," *Community College Journal of Research and Practice*, vol. 27, no. 2, pp. 131–146, 2003. [Online]. Available: http://dx.doi.org/10.1080/713838115

[6] M. M. Cook and A. Swanson, "The interaction of student and program variables for the purpose of developing a model for predicting graduation from graduate programs over a 10-year period," *Research in Higher Education*, vol. 8, no. 1, pp. 83–91, 1978.

[7] M. J. Alzahrani, E. M. Thomson, and D. B. Bauman, "Predictors of student success in an entry-level baccalaureate dental hygiene program," *American Dental Hygienists Association*, vol. 81, no. 2, pp. 51–51, 2007.

[8] S. B. Kotsiantis, C. Pierrakeas, and P. E. Pintelas, "Preventing student dropout in distance learning using machine learning techniques," in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*. Springer, 2003, pp. 267–274.

[9] D. Delen, "A comparative analysis of machine learning techniques for student retention management," *Decision Support Systems*, vol. 49, no. 4, pp. 498 – 506, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167923610001041

[10] E. Er, "Identifying at-risk students using machine learning techniques: A case study with is 100," *International Journal of Machine Learning and Computing*, vol. 2, no. 4, p. 476, 2012.

[11] "First-year experience: A comparison study," *Journal of Assessment and Institutional Effectiveness*, vol. 2, no. 2, p. 143170, 2012.

[12] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[13] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural computation*, vol. 12, no. 5, pp. 1207–1245, 2000.

[14] M. Tan, "Support vector machine & its applications," http://www.cs.cmu.edu/ awm/tutorials, Nov 2004.

[15] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *The Journal of Machine Learning Research*, vol. 2, pp. 45–66, 2002.

[16] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *Proceedings of the ninth ACM international conference on Multimedia*. ACM, 2001, pp. 107–118.

[17] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine learning*, vol. 46, no. 1-3, pp. 389–422, 2002.

[18] K.-R. M ºller, S. Mika, and R. §, "An introduction to kernel-based learning algorithms."

[19] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[20] B. Schölkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[21] S.-W. Lin, Z.-J. Lee, S.-C. Chen, and T.-Y. Tseng, "Parameter determination of support vector machine and feature selection using simulated annealing approach," *Applied soft computing*, vol. 8, no. 4, pp. 1505–1512, 2008.

[22] Z.-H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: many could be better than all," *Artificial intelligence*, vol. 137, no. 1, pp. 239–263, 2002.

[23] H.-C. Kim, S. Pang, H.-M. Je, D. Kim, and S.-Y. Bang, "Support vector machine ensemble with bagging," in *Pattern recognition with support vector machines*. Springer, 2002, pp. 397–408.

[24] H. Valizadegan and P.-N. Tan, "Kernel based detection of mislabeled training examples." in *SDM*, 2007.

[25] X. Zhu, X. Wu, and Q. Chen, "Eliminating class noise in large datasets," in *ICML*, vol. 3, 2003, pp. 920–927.

[26] D. Olson and D. Delen, *Advanced data mining techniques*. Springer Verlag, 2008.

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[28] Y. Pang, X. Xue, and A. S. Namin, "Identifying effective test cases through k-means clustering for enhancing regression testing," in *Machine Learning and Applications (ICMLA), 2013 12th International Conference on*, vol. 2. IEEE, 2013, pp. 78–83.