Beyond Grade Point Average and Standardized Testing: Incorporating a Socio-Economic Factor in Admissions to Support Minority Success

Prof. Houshang Darabi, University of Illinois, Chicago

Dr. Houshang Darabi is an Associate Professor of Industrial and Systems Engineering in the Department of Mechanical and Industrial Engineering (MIE) at the University of Illinois at Chicago (UIC). Dr. Darabi has been the Director of Undergraduate Studies in the Department of MIE since 2007. He has also served on the College of Engineering (COE) Educational Policy Committee since 2007. Dr. Darabi is the recipient of multiple teaching and advising awards including the COE Excellence in Teaching Award (2008, 2014), UIC Teaching Recognitions Award (2011), and the COE Best Advisor Award (2009, 2010, 2013). Dr. Darabi has been the Technical Chair for the UIC Annual Engineering Expo for the past 5 years. The Annual Engineering Expo is a COE's flagship event where all senior students showcase their Design projects and products. More than 600 participants from public, industry and academia attend this event annually.

Dr. Darabi is an ABET IDEAL Scholar and has led the MIE Department ABET team in two successful accreditations (2008 and 2014) of Mechanical Engineering and Industrial Engineering programs. Dr. Darabi has been the lead developer of several educational software systems as well as the author of multiple educational reports and papers. Some of these products/reports have already been launched/completed and are now in use. Others are in their development stages. Dr. Darabi's research group uses Big Data, process mining, data mining, Operations Research, high performance computing, and visualization techniques to achieve its research and educational goals.

Ms. Elnaz Douzali, University of Illinois, Chicago

Elnaz Douzali is a graduate student and researcher at the University of Illinois at Chicago. She is a part of the Mechanical and Industrial Engineering Department and will receive her Masters of Science degree in Industrial Engineering in May 2018. Her research interests include Educational Data Mining, Process Mining, and Healthcare.

Samuel Thomas Harford, PROMINENT Labs UIC

Samuel Harford is a graduate research assistant at the University of Illinois at Chicago's Mechanical and Industrial Engineering Department. He received his BSIE in May 2016 from UIC and is currently pursuing his MS. Since 2015 he has done multiple projects in education data mining, some in collaboration with the Dean of Engineering. His research interests include healthcare and education data mining.

Dr. Peter C. Nelson, University of Illinois, Chicago

Peter Nelson was appointed Dean of the University of Illinois at Chicago's (UIC) College of Engineering in July of 2008. Prior to assuming his deanship, Professor Nelson was head of the UIC Department of Computer Science. In 1991, Professor Nelson founded UIC's Artificial Intelligence Laboratory, which specializes in applied intelligence systems projects in fields such as transportation, manufacturing, bioinformatics and e-mail spam countermeasures. Professor Nelson has published over 80 scientific peer reviewed papers and has been the principal investigator on over $30 million in research grants and contracts on issues of importance such as computer-enhanced transportation systems, manufacturing, design optimization and bioinformatics. These projects have been funded by organizations such as the National Institutes of Health, the National Science Foundation, the National Academy of Sciences, the U.S. Department of Transportation and Motorola. In 1994-95, his laboratory, sponsored by the Illinois Department of Transportation, developed the first real-time traffic congestion map on the World Wide Web, which now receives over 100 million hits per year. Professor Nelson is also currently serving as principal dean for the UIC Innovation Center, a collaborative effort between the UIC Colleges of Architecture, Design and the Arts; Business Administration; Medicine and Engineering.

Mr. Fazle Shahnawaz Muhibul Karim, University of Illinois, Chicago

Fazle Karim is an aspiring data scientist who is completing his PhD in the Mechanical and Industrial Engineering department at University of Illinois at Chicago. He received his BSIE in 2012 from the University of Illinois at Urbana Champaign. He is currently the lead data scientist at PROMINENT lab, the leading process mining research facility at the university. He has taught courses in Probability & Statistic in Engineering, Work Productivity Analysis, Quality Control & Reliability, and Safety Engineering. His research interest includes education data mining, health care data mining, and time series analysis.

Beyond Grade Point Average and Standardized Testing: Incorporating a Socio-Economic Factor in Admissions to Support Minority Success

Abstract

This paper proposes a revised approach to the admission process for freshman students entering the minority serving institute, the University of Illinois at Chicago (UIC). The purpose of the revised approach is to better evaluate an extremely diverse population of applicants. The details for the revised approach will be demonstrated through the use of data mining, statistical methods and association rule mining.

UIC is located in Chicago, Illinois and enrolls greater than 20,000 students from a wide spectrum of socio-economic neighborhoods. As a minority serving institute, it is of great concern to the University to better assess the capabilities of the diverse population of applicants.

Through this paper, the authors propose a process that integrates the socio-economic background of applicants into its admission process which allows for its applicants to be better evaluated. The proposed process increases the population of students with the potential to succeed, thereby increasing university retention rates. An extensive review on success and retention strategies that benefit not only minorities in Science, Technology, Engineering, and Mathematics but all students is provided. This process better assesses applicants from differing background and has the effect of increasing the population of minority students.

In addition to the socio-economic based admission metric for such integration, this paper introduces a methodology and framework for a software to promote the success of students by recommending schedules based on their predicted performance. The types of information used in the development of this metric are available in almost every university or higher education institution. Therefore, the process in developing this metric can be implemented by other higher education institutions in the United States that have the potential to benefit from incorporating socio-economic factors into their admission procedure for applicants.

I. Introduction

Problem Description

The main objective of this paper is to develop a socio-economic indicator to more accurately predict the First Semester Grade Point Average (FS GPA) of applicants to the College of Engineering (COE) at the University of Illinois at Chicago (UIC). The predicted FS GPA is used to improve the admission of underrepresented students to the COE. Also, it is shown to utilize a threshold for the FS GPA to recommend support and retention strategies to improve the students' outcomes through a statistical and association rule mining software. The University has a metric, Metric 1, that uses an applicant's High School GPA and ACT Composite Score to predict their FS GPA. Metric 1 is a score that ranges from 0-40, representing the predicted FS GPA by a factor of 10. The university's current admissions policy is based on this scored Metric 1 of an applicant. It is hypothesized that a metric based solely on applicants' academic performance and standardized testing is not sufficient in predicting the potential of applicants. In this paper, it is argued that by including a socio-economic indicator into the current metric, the university can better predict how a student will perform and in return whether or not to accept that applicant. The result of this prediction is a new metric, Metric 2, that better predicts the FS GPA and has the added benefit of identifying may require additional support in order to succeed. Furthermore, the literature review provides a library of success support and retention system references to provide students of minorities their greatest chance of graduation.

Motivation

Students that enter UIC come from a wide variety of backgrounds. This study aims to benefit students that come from all different socio-economic areas and provides a support system for administration and academic advisors to promote student success. By promoting the success of students from poor socio-economic areas, this method will help promote the minority population at UIC. Based on initial analysis, there exists a high population of minority students that come from schools in poor socio-economic areas. As described by the National Academy of Sciences, the United States must research the role of diversity in the Science, Technology, Engineering, and Mathematics (STEM) workplace to ensure that the country is the global leader in science and technology[1].

As a minority serving institution, UIC is actively pursuing opportunities to better address its diverse pool of applicants. Traditionally, universities look at certain criteria of applications to determine students' candidacy into different programs. These criteria include measures such as ACT Composite score and High School GPA. These admission criteria are not sensitive to an applicant's race and socio-economic background. This opens an opportunity to expand on traditional admissions methods by incorporating a new metric for a student's socio-economic background and minority status.

As a nation, engineering occupations are projected to add over 130,000 jobs over the next decade in the United States[2]. While the need for engineers increases, it is known that the first to second year retention rate nationwide averages 66%[3] for students in engineering programs. When considering how to address the demand for qualified engineers, it is important to consider the possible places from which engineers could originate. This analysis is intended to more accurately predict the metric that is used to admit applicants into the COE. This process will increase the number of students in the COE that are more likely to succeed.

Increasing the number of engineers from disadvantaged socio-economic backgrounds has been a concern for several decades. For this reason, much research has been conducted to combat the problem at all levels of education. The focus of this paper is on the way underrepresented students are admitted to the COE. The use of varying admissions standards, indicators, and support systems are becoming important tools of several educational researchers. Some of these research topics will be described in the following text.

The remaining sections of this paper are as follows. Section II gives a Literature Review of key components related to this study. In Section III, the Data Collection and Methodology involved in developing a new metric for admission are discussed. Section IV goes into the evaluation of this new metric on the current population and introductions a system that can be used with the metric to support student success. Section V overviews the findings and discusses what is to come in terms of implementation.

II. Literature Review

There are many research projects exploring the factors that affect a student's performance in higher education. In this literature review, papers are explored that present ideas from socio-economic factors to addressing these difficulties in order to overcome them during post high school education.

Conditions Affecting Performance & Retention

One of the goals of UIC is to admit applicants that have the potential to be successful in a university environment. When analyzing such a diverse population of students, it is essential to know what affects student performance. Factors that contribute to academic success can either be internal (on campus) or external (off campus). Indicators of performance and retention in universities include high school academic performance, standardized testing, socio-economic factors and emotional health as described by the following authors.

As written by *Richardson et. al*, psychological and emotional health correlate with how a student performs at the university level and whether they complete their program. Richardson found that demographic and psychosocial factors, high school grade point average, SAT, ACT and self-efficacy were all correlated with a student's GPA in college[4]. *Conley et. al* also presents on how ACT scores and high school GPA predicted academic achievement best when combined with socioeconomic status, academic self-confidence and motivation[5]. For this study, the psychological and social predictors which are qualitative will not be explored. The focus will be towards quantitative predictors.

Geiser et. al found that although the traditional method of using ACT C and HS GPA were strong indicators of college performance, so were other factors such as family income and demographics[6]. Incorporating ACT C and HS GPA as predictors of FS GPA and college completion are present in a multitude of studies. A study performed by *DesJardins et. al* goes one step further and explores income, race and high school performance on college performance and retention. They also look at breaks from education and re-enrollment differences between income and race levels[7]. These papers demonstrate that although high school academic performance plays a key role in predicting college performance and retention, so do external factors such as family income, social and mental health[8,9].

Using statistical methods to model the behavior of individuals has been a trending topic of research[10]. Evidence of this claim lies in the numerous papers published in the field and are discussed in this paper. One study uses machine learning techniques to closely predict the GPA of a student early in their academic career[11]. In a recent paper, the academic performance of students was predicted using classification data mining techniques[12]. Predicting student performance was described by *Huang et. al* in his comparison of four different mathematical models[10]. The authors of this paper explore not only predictive models of student performance and retention but also the behavior of students in STEM.

In order to include a factor for students of minorities and underrepresentation in STEM, the following research studies must be explored. One study developed a criterion for high schools called Under Representation Score (URS) that combines a high school's Total Economically Disadvantaged Enrollment (TED) and College Readiness Index (CRI); that universities can use for their admissions process[13]. Similarly, *Walpole et. al* studies the differences between students who have low Socio Economic Scores (SES) and high SES[14]. They claim that the behaviors of these low SES students greatly differ from high SES students and that it's a main topic of policy makers.

Minority Student Support and Retention Strategies

To promote the success and graduation of all students in universities across the country, there have been many strategies and support programs developed. One research study suggests that programs designed to target first year students increases their probability of finishing their undergraduate degree[15]. A similar study shows that an intervention which provides supplemental instruction for freshman students run by senior students helps freshman focus, build more effective study skills and promotes their motivation and academic performance[16]. One of the instructions that improves the rate of retention in engineering students is whether a student comes into the university freshmen year ready to take calculus or pre-calculus. It is shown that a student who entered an engineering program ready for pre-calculus was more likely to graduate than a student who entered freshmen year ready for calculus[17]. Throughout this paper, pre-calculus will be considered as a helper course. These courses are designed to prepare students with the foundation required for calculus, when high school preparation was not sufficient. Although many papers have been published providing insight on promoting student performance and retention strategies[18], the area of this paper's focus is providing these programs specifically for minorities and students who are underrepresented in STEM in their programs.

May et. al claims that the success of minorities in STEM correlates back to a few factors: pre-college preparation, recruitment programs, financial assistance, intervention programs, admission policies and graduate school preparation[19]. They suggest that a national "will" towards providing these resources for minority underrepresentation in STEM will make drastic changes to the current STEM demographics in universities. To support these claims, *Tsui et. al* provides an intensive literature review on ten effective retention strategies designed to address students of underrepresented races/ethnics in STEM undergraduate programs. In a study of 20 programs, Tsui reviews the positive effect that pre-college summer bridge programs have for low income and minority students. Following up on that claim, mentoring is also a widely used intervention program for minorities that serves to increase their interest in STEM fields[20]. Studies have shown that minority students who actively participate in one-on-one mentorship positively correlate with the transition to college, higher grade point averages, lower attrition and clearly defined educational goals[21–24]. Tutoring, seminars, and research experience also positively supports the idea of increasing minority STEM student retention and success[25].

The importance of providing resources to help with college finances was illuminated by *Georges et. al.* In a national study of underrepresented students in 117 engineering programs, a significant relationship was found between minority retention rates and the average financial aid rewarded to students[26]. Providing financial resources and help to minorities in STEM has been the subject of many studies[27]. Other efforts to increase the achievement and retention of minority students in STEM are provided for review[28,29].

The awareness and application of these discussed support and retention strategies are significant in minority serving universities. UIC provides resources and programs discussed in the literature review to ensure STEM students who are underrepresented have the opportunity to succeed in today's world. A few of the programs designed to support minorities at the university in the study's engineering program are National Society of Black Engineers (NSBE), Society of Hispanic Profession Engineers (SHPE), Society of Women Engineers (SWE) and a unique program designed for this specific university to help minorities in engineering through recruitment and retention methods called MERP. These programs provide resources for tutoring, advising, access to financial assistance, scholarship opportunities as well as networking and internship opportunities.

Association Rule Mining in Education

Association rule mining is one of the most studied data mining method for supervised and unsupervised learning. The purpose of using association rule mining is to find rules that associate one attribute with another attribute. These rules create an *if* and *then* statement to associate different attributes[30]. *Agrawal, Imielinski and Swami et. al* proposed one of the earliest uses of association rule mining, in 1993, where they mine rules in a large database to find relationships between products people bought[31]. In a subsequent paper, *Agrawal and Srikant et. al* propose using sequential pattern mining to find patterns between "itemsets" that are based on time[32]. Most of the work that uses association rule mining in education is in web based education systems to detect patterns on what contents students are accessing together and what combination of tools they are using.

For web based education systems, sequential pattern mining can detect what content led students to access other online contents. Over the past 20 years, several researches have used association rules on web log files to detect incorrect student behavior[33], help teachers find interesting and unexpected learning patterns[34], develop frameworks to simply the process of uploading content for students based on their learning patterns[35], and to detect which learning materials should be recommended to each student[36]. Nonetheless, association rule mining has applications in content management systems[34,37,38] and adaptive and intelligent web-based educational systems[39–41]. In the cases of developing context based decision-support systems, more elaborate association rules like APriori can be preferred over other data mining algorithms[42]. Decision-support systems that use association rules are beneficial to multiple entities in a university. Researchers can study course patterns of students and interaction between courses. Deans can detect overlaps in curriculum within a college. Similarly, other entities can observe patterns beneficial for them[43]. In this paper, sequence pattern-mining is applied to detect patterns of courses students take each semester.

III. Data and Methodology

Data Collection

The dataset for this study consists of over 2,500 students at UIC. These students originate from 190 different High Schools. This study looks at students from the College of Engineering. The data for each student contains the following fields: High School Grade Point Average (HS GPA), ACT Composite Score (ACT C), High School Name, Major, the years and terms they registered for classes at the university, which courses they registered for, and grades received for each course.

Information about all the high schools, that were present amongst the students, was retrieved from USNews[44]. Data fields gathered from this sources included: College Readiness Index (CRI), Minority Percentage, Total Economically Disadvantaged Percentage (TED), City and Country. The U.S. Census Bureau[45] information provided FAFSA Application Percentage from each school (FAFSA), Country Per Capita Income (Country PCI) and City Per Capita Income (City PCI).

Methodology

UIC currently has an admissions process that strictly looks at ACT C and HS GPA to determine an applicant's acceptance into the school, which they calculate as Metric 1. The proposed metric, Metric 2, is used to improve college admissions by integrating a socio-economic indicator into the current admissions process. This indicator is known as the Under Represented Score (URS). URS was first mentioned in 2016 by *Darabi et. al* and is computed using CRI and TED, as shown in Equation 1[13]. Students that come from higher URS high schools are considered to be socio-economically disadvantaged.

$$URS = (100 - CRI + TED)/2 \tag{1}$$

The difficulty when using URS is the lack of information that USNews has on CRI and TED, which can deter using URS. The proposed model has two phases, as shown in Figure 1. The objective of Phase One is to calculate the URS score of high schools that have sent students to UIC, including high schools without CRI and TED. In Phase Two, the URS score is used with students' high school academic performance to predict their FS GPA.

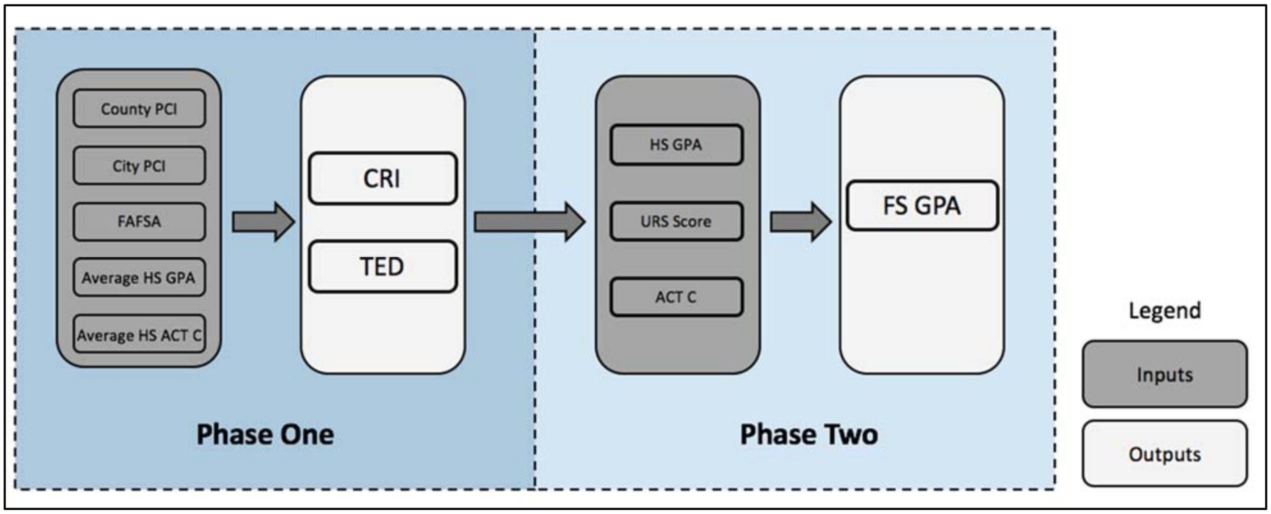

Figure 1. Phase One and Phase Two

In Phase One, the first step is to predict the CRI and TED of schools that do not publicly provided this information. CRI and TED are predicted using the dataset of high schools and known CRIs and TEDs. The following information is used to predict CRI and TED: FAFSA, City PCI, County PCI, Average HS ACT C and Average HS GPA.

To predict CRI and TED, data is split leaving 180 high schools for training and validation. The training set is then split, 60% of the schools were randomly selected for building a model and the remaining 40% are used to test the model. This process is repeated 200 times and the performance is averaged at the completion of all repetitions. The performances of these models were evaluated using Root Mean Square Error (RMSE), as shown in Equation 2. Where *n* is the number of observations in the testing set. Once the best models to predict CRI and TED were determined, it was applied on the high schools with unknown CRI and TED.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Actual_i - Predicted_i)^2} \qquad (2)$$

In Phase Two, a model is built to predict FS GPA, the basis of the proposed admission policy. The current process of the college is to use ACT C and HS GPA to predict FS GPA using a linear regression model. By incorporating the socio-economic indicator URS to predict FS GPA, the results will better admit applicants while also diversifying the college population. Similar to Phase One, this dataset is split such that 60% of students are randomly partitioned into the training set, and the remaining 40% are the testing set. Various modeling techniques are tested on the training set to measure its performance, RMSE. This process is repeated 200 times. At each iteration, the RMSE of the current Metric 1 is compared to the proposed Metric 2.

In the current admissions process, the university waitlists students with a predicted FS GPA of less than a certain threshold (we use CT to refer to this certain threshold). This paper proposes providing additional support to these students if admitted through a decision support system. The decision support system is a software developed using student trace data. A trace is defined as the historic set of courses a student takes each semester. The software uses a form of Sequential PAttern Discovery using Equivalence classes (SPADE) algorithm to study patterns of student grades and traces. Extended details of this algorithm and its application will be made available in future works. In addition to using the decision support system, the administration are proposed to provide the support and retention systems cited in the literature review to the underrepresented students in their engineering program.

IV. Results and Evaluation

The following section is divided into three parts: Phase One Prediction, Phase Two Prediction, and Providing Support to Students with a FS GPA of CT or below. Each section will show results indicating model performance.

Phase One Prediction

The objective of Phase One was to predict CRI and TED, so that it could be used to calculate URS. Several interpretable models with different parameters were trained on 60% of randomly chosen data and tested on the remaining 40% of the dataset. This process was repeated 200 times and the average RMSE was measured, as shown in Table 1.

Table 1. Phase One Model Testing

Model	Variables	RMSE (TED)	RMSE (CRI)
k-NN (k=5)	All Variables*	41.73120	23.08308
k-NN (k=1)	All Variables*	41.64231	23.10203
Decision Trees	All Variables*	22.96821	13.45639
Linear Regression	**All Variables***	**19.84266**	**12.75739**
Linear Regression	All Variables* exclude FAFSA	23.15377	15.47915

* All Variables include FAFSA, County PCI, City PCI, Average HS ACT C, and Average HS GPA

Based on the analysis, a linear regression model proved to be best when predicting CRI and TED, as it results in the lowest RMSE score. The linear equations are presented with their coefficients in Equation 3 and Equation 4. Using the predicted values of CRI and TED, URS is computed for students that did not previously have a CRI and TES. The RMSE score for computing URS is 11.68758, an RMSE score that is good for Phase 2, predicting FS GPA.

$$CRI = 213.30 - 3.707(FAFSA) - 0.00038(County\ PCI) - 0.0007(City\ PCI) - 6.30(Average\ HS\ ACT\ C) + 4.10(Average\ HS\ GPA) \quad (3)$$

$$TED = -85.33 + 92.80(FAFSA) + 0.00067(County\ PCI) + 0.00054(City\ PCI) + 1.294(Average\ HS\ ACT\ C) - 2.729(Average\ HS\ GPA) \quad (4)$$

Figure 2 illustrates that the predicted URS (using predicted TED and predicted CRI) performs well in estimating actual URS (using actual CRI and TED). Accurately predicting the URS of a high school is significant because it is used as a socio-economic indicator of the FS GPA prediction model.

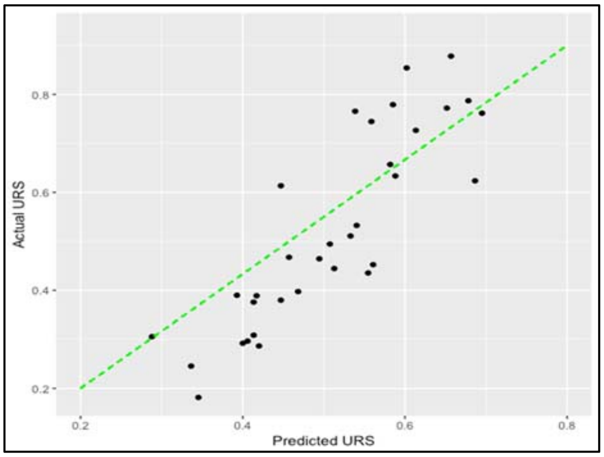

Figure 2. Predicted URS versus Actual URS

Phase Two Prediction

In Phase Two, FS GPA is predicted, an indicator to decide whether to admit a student or not. As mentioned, currently the admissions policy uses a linear regression with ACT C and HS GPA, to predict FS GPA. As shown in Table 2, adding a socio-economic indicator, URS, reduces the RMSE. Equation 5 shows the proposed linear regression model to predict FS GPA that includes URS. Alongside it is Equation 6, which is the original linear regression model used to predict FS GPA that does not include any socio-economic indicator.

Table 2. Phase Two Model Testing

Model	RMSE
Original Model	0.7920418
Proposed Model	0.7121738

$$FS\ GPA = -0.20417 + 0.05097(ACT\ C) + 0.5007(HS\ GPA) - 0.25169(URS) \qquad (5)$$

$$FS\ GPA = -0.38891 + 0.05413(ACT\ C) + 0.49596(HS\ GPA) \qquad (6)$$

The proposed model is then applied to 200 repeated testing sets. In each iteration, the proposed model proved to perform better than the original. Since the difference between the predicted model and the original model is not normally distributed, as shown in Figure 3, a Wilcoxon Signed Rank Test is performed. The hypothesis of the Wilcoxon Signed Rank test is shown in Table 3. The p-value from the Wilcoxon Signed Rank Test of $2.2\text{x}10^{-16}$ indicates that proposed Model performs better than the original model so the null hypothesis is rejected.

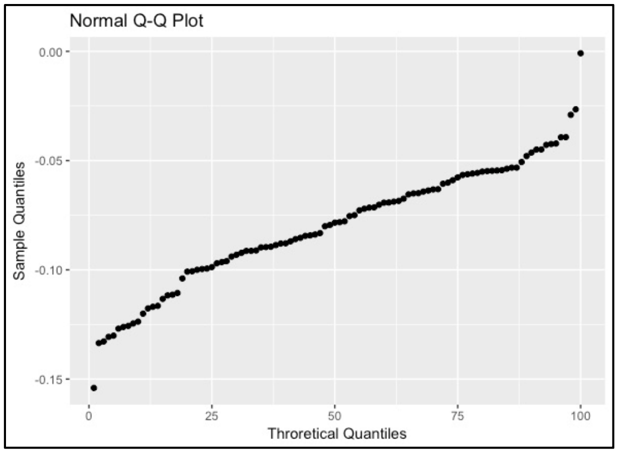

Figure 3. Phase Two Q-Q Plot

Table 3. Null and alternative hypothesis of the Wilcoxon Signed Rank Test

Null Hypothesis:	Difference between the pairs follows a symmetric distribution around zero
Alternative Hypothesis:	Difference between the pairs does not follow a symmetric distribution around zero

The following subsection outlines how predicted FS GPA is used to identify the at-risk students from high schools that are less prepared than others, and provide them support using the outlined decision support system. This may increase the enrollment of students of minorities and students who are underrepresented in STEM.

Providing Support to Students with a FS GPA of CT or below

Support to students who are predicted to have a FS GPA of CT or below is given using the decision support system software. The software uses sequence rule mining to detect a set of courses students should take if they decide not to take a helper course their first semester. The rules generated by the software were made on 70% of the student data and tested on the remaining 30%. The rules were applicable to a certain amount of the testing set and had an accuracy of 100%.

In addition to recommending a set of courses for students, the decision support system is able to detect an interesting pattern regarding a helper course, pre-calculus. The decision support system recommends students predicted to have a FS GPA of CT or below to have a B or higher in pre-calculus. This rule is further investigated by studying two cohorts of students with a FS GPA of CT or below, students that did not stay past four semesters, and students that did stay for at least four semesters. The authors studied the grades students received when taking pre-calculus as shown in Table 4. The provided analysis shows that students who receive a B or higher for pre-calculus tend to stay past for at least four semesters. This information is paramount, as it's propose that if a student takes pre-calculus, he or she is recommended to obtain a B or higher. Pre-calculus is the foundation of being successful in engineering. Performing well in this helper course will improve a student's likelihood of staying past 4 semesters.

Table 4: Percentage of students that obtained a particular grade in pre-calculus.

	A	B	C	D	F
< 4 Semesters	4.17%	8.33%	58.33%	8.33%	20.83%
≥ 4 Semesters	15.00%	30.00%	55.00%	0.00%	0.00%

As evident from the above table, students who took pre-calculus given that they had a FS GPA of CT or below and stayed less than four semesters received an A or B 12.5%. The students who took pre-calculus given that they had a FS GPA of CT or below and stayed at least four semesters received A or B 45% of the time. This shows that receiving at least a B in this helper course may affect the likelihood of remaining in the university for at least four semesters.

V. Conclusion & Future Work

Today's society aims to diversify the engineering workplace by increasing the pool of successful underrepresented students in STEM undergraduate programs. This paper proposes an alternative admission policy that incorporates a student's socio-economic status, URS, as an indicator.

The University of Illinois at Chicago currently has a metric, Metric 1, that is used in the admissions process. This paper recommends an alternative model for Metric 1 that is simple to incorporate into the existing system. The proposed new model, Metric 2, performs 10% better than the existing metric. A decision support system is also developed for students that have a predicted FS GPA of CT or below using sequence mining. The detailed expansion of this software for a decision support system applying a form of Sequential PAttern Discovery using Equivalence classes (SPADE) algorithm will be continued in future works. Furthermore, this paper advises the administration and academic advisors of UIC to provide the support and retention systems referenced in the literature review.

UIC continues to be a thriving minority serving institute by implementing these strategies proposed in this paper. Through this paper, it is an aim to increase the minority enrollment and it is believed that an extensive understanding of the referenced strategies benefit both the emerging students and the university as a whole.

References

1. National Academy of Science. Expanding Underrepresented Minority Participation.
2. Emily Richards DT. *Occupational Employment Projectsion to 2022*.; 2013.
3. ACT. National Collegiate Retention and Persistence to Degree Rates. http://www.act.org/research/policymakers/pdf/retain_2012. Published 2012. Accessed November 1, 2016.
4. Richardson M, Abraham C, Bond R. Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychol Bull*. 2012. http://psycnet.apa.org/journals/bul/138/2/353/. Accessed February 1, 2017.
5. Conley D. Redefining College Readiness. *Educ Policy Improv Cent*. 2007. http://eric.ed.gov/?id=ED539251. Accessed February 1, 2017.
6. Geiser S, Santelices M. Validity of High-School Grades in Predicting Student Success beyond the Freshman

Year: High-School Record vs. Standardized Tests as Indicators of Four-Year. *Cent Stud High Educ*. 2007. http://eric.ed.gov/?id=ED502858. Accessed February 1, 2017.
7. DesJardins S, Ahlburg D, McCall B. The effects of interrupted enrollment on graduation from college: Racial, income, and ability differences. *Econ Educ Rev*. 2006. http://www.sciencedirect.com/science/article/pii/S0272775705000853. Accessed February 1, 2017.
8. Lotkowski VA |Robbins. SB |Noeth. RJ. The Role of Academic and Non-Academic Factors in Improving College Retention. ACT Policy Report. *Am Coll Test ACT Inc*. 2004.
9. Pritchard M, Wilson G. Using emotional and social factors to predict student success. *J Coll Stud Dev*. 2003. https://muse.jhu.edu/article/40374/summary. Accessed February 2, 2017.
10. Huang S, Fang N. Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Comput Educ*. 2013;61:133-145. doi:10.1016/j.compedu.2012.08.015.
11. Slim A, Heileman GL, Kozlick J, Abdallah CT. Predicting student success based on prior performance. In: *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*. IEEE; 2014:410-415. doi:10.1109/CIDM.2014.7008697.
12. Ahmed A, Elaraby I. Data Mining: A prediction for Student's Performance Using Classification Method. *World J Comput Appl Technol*. 2014. http://www.hrpub.org/journals/article_info.php?aid=1285. Accessed February 4, 2015.
13. Darabi H, Karim F, Harford S, Douzali E. A Comparative Analysis of Underrepresented Engineering Applicants Ad-mission Practices and their Academic Performance at the University of Illi-nois at. *123rd ASEE Annu*. 2016. https://www.asee.org/file_server/papers/attachment/file/0006/9121/cp_Final.pdf. Accessed February 1, 2017.
14. Walpole M, Allmendinger D, Astin AW, et al. Socioeconomic Status and College: How SES Affects College Experiences and Outcomes. *Rev High Educ*. 2003;27(1):45-73. doi:10.1353/rhe.2003.0044.
15. Lee D, Olson E, Locke B, Michelson S. The effects of college counseling services on academic performance and retention. *J Coll*. 2009. https://muse.jhu.edu/article/265820/summary. Accessed February 1, 2017.
16. Ning H, Downing K. The impact of supplemental instruction on learning competence and academic performance. *Stud High Educ*. 2010. http://srhe.tandfonline.com/doi/abs/10.1080/03075070903390786. Accessed February 2, 2017.
17. Richardson J, Dantzler J. Effect of a freshman engineering program on retention and academic performance. In: *32nd Annual Frontiers in Education*. IEEE; :S2C-16-S2C-22. doi:10.1109/FIE.2002.1158656.
18. Habley W, Bloom J, Robbins S. *Increasing Persistence: Research-Based Strategies for College Student Success*.; 2012. https://books.google.com/books?hl=en&lr=&id=PDmMpwbs_CsC&oi=fnd&pg=PR13&dq=Increasing+persistence:+Research-based+strategies+for+college+student+success&ots=N04OjJ1saP&sig=lZLAQhjMcSmlK1LMfa03aDpHM54. Accessed February 1, 2017.
19. May GS, Chubin DE. A Retrospective on Undergraduate Engineering Success for Underrepresented Minority Students. *J Eng Educ*. 2003;92(1):27-39. doi:10.1002/j.2168-9830.2003.tb00735.x.
20. Kendricks K, Nedunuri K. Minority student perceptions of the impact of mentoring to enhance academic performance in STEM disciplines. *J STEM*. 2013. http://search.proquest.com/openview/10816f7dc7be8af1566b42203505e6e6/1?pq-origsite=gscholar&cbl=27549. Accessed February 2, 2017.
21. Bordes V, Arredondo P. Mentoring and 1st-year Latina/o college students. *J Hisp High*. 2005. http://journals.sagepub.com/doi/abs/10.1177/1538192704273855. Accessed February 2, 2017.
22. Santos S, Reigadas E. Understanding the student-faculty mentoring process: Its effects on at-risk university students. *J Coll Student Retent*. 2004. http://csr.sagepub.com/content/6/3/337.short. Accessed February 2, 2017.
23. Santos S, Reigadas E. Latinos in higher education: An evaluation of a university faculty mentoring program. *J Hispanic High Educ*. 2002. http://jhh.sagepub.com/content/1/1/40.short. Accessed February 2, 2017.
24. Santos S, Reigadas E. Evaluation of a University Faculty Mentoring Program: Its Effect on Latino College Adjustment. 2000. http://eric.ed.gov/?id=ED454008. Accessed February 2, 2017.
25. Tsui L. Effective strategies to increase diversity in STEM fields: A review of the research literature. *J Negro Educ*. 2007. http://www.jstor.org/stable/40037228. Accessed February 2, 2017.
26. Georges A. Keeping What We've Got: The Impact of Financial Aid on Minority Retention in Engineering. *NACME Res Lett*. 1999. http://eric.ed.gov/?id=ED435534. Accessed February 2, 2017.

27. Dowd A. Income and Financial Aid Effects on Persistence and Degree Attainment in Public Colleges. *Educ Policy Anal Arch*. 2004. http://eric.ed.gov/?id=EJ852304. Accessed February 1, 2017.
28. Arnold K, Pistilli M. Course signals at Purdue: Using learning analytics to increase student success. *Proc 2nd Int Conf*. 2012. http://dl.acm.org/citation.cfm?id=2330666. Accessed February 2, 2017.
29. Hurtado S, Newman C, Tran M. Improving the rate of success for underrepresented racial minorities in STEM fields: Insights from a national project. *New Dir*. 2010. http://onlinelibrary.wiley.com/doi/10.1002/ir.357/full. Accessed February 2, 2017.
30. Romero C, Ventura S. Educational data mining : A survey from 1995 to 2005. *Expert Syst Appl*. 2007;33:135-146. doi:10.1016/j.eswa.2006.04.005.
31. Agrawal R, Imielinski T, Swami A. Mining Association Rules between Sets of Items in Large Databases.
32. R. Agrawal RS. Mining Sequential Patterns. In: *Eleventh International Conference on Data Engineering, IEEE Computer Society Press*. Taipei, Taiwan; 1995:3-14.
33. Yu, P., Own, C., & Lin L. On learning behavior analysis of web based interactive environment. In: *In Proceedings of ICCEE*. Oslo/Bergen, Norway; 2001.
34. Wang W, Weng J. Learning portfolio analysis and mining in SCORM compliant environment. *Front Educ …*. 2004. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1408490. Accessed January 22, 2016.
35. Shen R, Han P, Yang F, Yang Q. Data mining and case-based reasoning for distance learning. *Int J*. 2003. http://www.igi-global.com/article/data-mining-case-based-reasoning/1614. Accessed February 8, 2017.
36. Markellou P, Mousourouli I, Spiros S. Using semantic web mining technologies for personalized e-learning experiences. *Proc web-*. 2005. http://www.hci.gr/files/461-826_Markellou.pdf. Accessed February 8, 2017.
37. Zaiane O. Web usage mining for a better web-based learning environment. 2001. https://era.library.ualberta.ca/files/jm214q67j. Accessed February 8, 2017.
38. Li J, Zaïane O. Combining usage, content, and structure data to improve web site recommendation. *Int Conf Electron Commer*. 2004. http://link.springer.com/chapter/10.1007/978-3-540-30077-9_31. Accessed February 8, 2017.
39. Lu J. Personalized e-learning material recommender system. *Int Conf Inf Technol*. 2004. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.200.4929&rep=rep1&type=pdf. Accessed February 8, 2017.
40. Freyberger J, Heffernan N. Using association rules to guide a search for best fitting transfer models of student learning. *Work Anal*. 2004. https://pdfs.semanticscholar.org/644c/da71ed69a42e1e819a753ffa5cb5c6ccbaec.pdf. Accessed February 8, 2017.
41. Merceron A, Yacef K. Tada-ed for educational data mining. In: *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*. ; 2005:267-287.
42. Agrawal R, Srikant R. Fast Algorithms for Mining Association Rules.
43. Smith W. Communications of the Association for Information Systems Applying Data Mining to Scheduling Courses at a University APPLYING DATA MINING TO SCHEDULING COURSES AT A UNIVERSITY. *Commun Assoc Inf Syst*. 2005;16(16):463-474. http://aisel.aisnet.org/cais. Accessed February 8, 2017.
44. Illinois High Schools. USNews. www.USNews.com/education/. Published 2016. Accessed November 1, 2016.
45. Bureau USC. Data Access and Dissemination Systems (DADS).