Using Data Mining to Model Student Success

by

Becky Geltz

Submitted in Partial Fulfillment of the Requirements

for the Degree of

Master in Computing and Information Systems

in the

Computing and Information Systems

Program

Youngstown State University

December 2009

Using Data Mining to Model Student Success

Becky Geltz

I hereby release this thesis to the public. I understand that this thesis will be made available from the OhioLINK ETD Center and the Maag Library Circulation Desk for public access. I also authorize the University or other individuals to make copies of this thesis as needed for scholarly research.

Signature:

Becky Geltz, Student Date

Approvals:

John R. Sullins, Thesis Advisor Date

Thomas A. Bodnovich, Committee Member Date

Robert A. Hogue, Committee Member Date

Peter J. Kasvinsky, Dean of School of Graduate Studies and Research Date

TABLE OF CONTENTS

Weka

Analysis

Tables

Abstract

As funding for higher education through federal and state sources continues to decline, and a stronger call for accountability is placed upon higher education institutions to graduate students within the expected amount of time, colleges and universities are looking for ways to best leverage their resources to attract college-ready students who will enroll in their institutions, remain enrolled consistently, and earn their undergraduate degrees in a timely manner. Federal research conducted by the U.S. Department of Education's National Center for Education Statistics through the Integrated Postsecondary Education Data System (IPEDS) examines aggregate student enrollment, degree completions, and graduation rates. But to be truly helpful to the institutional researcher, unit record data is required. Only by examining the many attributes of each individual student can an institution determine the unique characteristics which will lead to student academic success – degree attainment. Because of the overall readability and the strong level of accuracy they can produce, decision trees are a good method for identifying the relationships between attributes in large datasets. Therefore, this study explores the use of data mining on higher education unit record data to develop a decision tree classification model of student success.

Acknowledgements

I would like to acknowledge the invaluable instruction, inspiration and guidance provided by Dr. Alina Lazar, Youngstown State University, in the field of data mining and the development and refinement of this project.

In addition, I would like to thank Dr. John Sullins for taking over mid-project, the responsibilities of thesis advisor and Mr. Thomas Bodnovich and Mr. Robert Hogue for their guidance, support, and inspiration and for serving on my thesis committee.

Finally, I wish to express my sincere thanks to the University of Waikato, Hamilton, New Zealand for making the Weka data mining software available online and free to the public; and to Ian H. Witten & Eibe Frank for their work developing instructions on using the Weka tool.

1. Motivation and Goals

1.1 Motivation

While politicians in both the state and federal governments are in agreement that a college education is necessary for economic recovery and technological superiority in the world, parents feel that a college education is necessary to obtain worthwhile employment. In Ohio, higher education institutions are continually hearing that business and industry are looking for an educated workforce. The belief is that if the state educates and retains those graduates, the jobs will come and the economy will flourish.

Moreover, given the current state of our nation's economy, higher education is being asked to justify the expense of a college degree. Parents are concerned that it is taking longer than the anticipated four years for their son/daughter to earn a four-year degree. The rising cost of tuition puts an added strain on household budgets and has many draining their savings. Skyrocketing student loan debt has student's pondering the continual dilemma of going further into debt, getting a job or working more hours and reducing their academic load, or dropping out of school all together.

In July of 2009, the President's Council of Economic Advisers (CEA) published a report forecasting employment opportunities for the next decade and outlining the groundwork required to prepare the labor force for the new millennium. The report posits that "well-trained and highly-skilled workers will be best positioned to secure high-wage jobs" and that "occupations requiring higher educational attainment are projected to grow much faster than those with lower education requirements." In support of that prediction the report goes on to state that the most current job growth has been experienced in those professions demanding higher education credentials and job loss in professions with lesser demands (Executive Office of the President, Council of Economic Advisors, 2009).

Given the downturn in the U.S. economy and subsequent 10%+ unemployment rate currently being experienced throughout the nation, this is the time when higher education can have its greatest positive impact. It is common knowledge that during times of economic recession, higher education experiences enrollment growth. In addition to current high school graduates pursuing their academic dreams for a better future, the recently unemployed or under-employed seek opportunities to acquire or refine skills necessary to be competitive in the diminished job market. “Therefore we need a comprehensive strategy to ensure that our education and training systems are strong and effective” (Executive Office of the President, Council of Economic Advisors, 2009).

In order to guarantee that students have worthwhile educational opportunities available, the federal and state governments are making strides in funding reformation. In particular the state of Ohio is preparing to rollout sweeping changes to how it allocates state subsidization of higher education. “Instead of funding institutions based on the number of students they enroll, the new formula would appropriate dollars based on colleges’ ability to retain and graduate students” (Moltz, 2009). This is a step, like many made in the past decade, toward greater accountability for colleges and universities and may present major challenges for institutions with open-enrollment policies. But as Dr. Watson Scott Swail, President and CEO of the Educational Policy Institute, opined in his August 2008 article *The Bell Curve Under a Different Cover*, “if our system is such that we let in a broad cross-section of students, then we have a moral and legal obligation to do what we can to help those students succeed."

Not only are the federal and state governments seeking to hold higher education more accountable, students and their parents are as well. People take notice when stories circulate about college graduates landing barely-above-minimum-wage jobs that not only do not provide a living wage but also do not provide enough income for graduates to meet the scheduled payments of their student loans (Perry, 2008). It was reported in October 2008 that "the latest generation of adults in the United States may be the first since World War II, …not to attain higher levels of education than the previous generations." The biggest declines are being experienced among the minority populations where it is believed that "the current generation is, on average, heading toward being less educated than its predecessor" (Jaschik, 2008). Perhaps conscious choices are being made to forego higher education because it is no longer perceived to have a good return on the investment. After all, the time required to earn a four-year degree seems to extend each year – the current average is near 5.5 years. Additionally, tuition rates for the most part rise every year. Combine the increased time-to-degree with predatory lending practices which target college students with everything from high interest educational loans to even higher interest credit cards, and it is no wonder that students are often split between spending more time in college and exiting further in debt or opting out of college altogether.

During the past few years across the nation, we have all as a society experienced the rapid ascent of utility costs, the push toward becoming "greener," and the increased costs associated with medical coverage. Higher education has not been immune. In fact while tackling those challenges, higher education has also been expected to provide "highly-skilled – and often expensive talent,… top-notch academic support, counseling, health services, and campus security,… a nicely maintained campus,… up-to-date libraries, labs, and other scientific resources" while satisfying government mandated responsibilities (Jacobs and Hyman, 2009). Higher education also finds itself confronted by a public that is crying out for if not a tuition freeze then a reduction in tuition costs.

1.2 Goals

In an effort to identify a solution to the problems presented, I chose to employ data mining - specifically the construction of a decision tree. Different from traditional statistics, which calculate the probability of a specific hypothesis, the end result of data mining is to identify the hidden pattern of connections within the data. Decision trees are just one several techniques used to display those patterns. Beginning with the attribute found by the data mining software to be the most significant and branching out from there, a decision tree provides a visual tree-like representation of the underlying connections leading to the final outcome.

By developing a decision tree model of student success, defined here as earning a baccalaureate/bachelor degree within six years, this thesis attempts to provide information necessary for determining how to increase the graduation rate at a public, four-year, open-enrollment institution. Additionally this increase would benefit the institution by satisfying the calls for accountability, optimizing institution's share of state and federal financial assistance, and hopefully make timely degree completion the norm.

In order to address the issue of how to increase the percentage of students who complete their studies and earn a baccalaureate (four-year) degree within six years, I attempted to identify the qualities of those students who earn their degrees in a timely manner. Next I used those qualities or criteria to examine a cohort of incoming students to predict which students should be successful and followed their progress. For those students who graduated, this information needs to be shared as much as possible with high school guidance counselors, so that high schools will know what preparation is necessary for their students to succeed in higher education. In order to best leverage limited recruiting dollars and guarantee critical completion-based state subsidy, focus should be placed on prospective students that exhibit these identified qualities. In addition, if students were predicted to graduate but did not, then these salient qualities or factors should also be identified. Examining student data may illuminate the key indicators that provide vital information on when, and in what manner, intervention should occur to facilitate the student's goal of obtaining a four-year degree. Students who are predicted not to graduate need to be further studied to identify how best to help them achieve success – perhaps fueling an argument for selective admissions or increased funding of student services.

To examine the multitude of student data and identify the relationships between the attributes, a decision tree will be constructed using the open-source data mining software package - Weka. It is the intent of this study to utilize the robust computing power of data mining software to quickly and accurately predict student success.

The organization of this paper follows:

1. Identify where applicable data are collected and at what frequencies.
2. Explain why a particular subset of data was chosen.
3. Tell how that subset was augmented to create the student cohort dataset.

4. Detail how and where the data were retrieved.
5. Disclose how the data were processed for analysis.
6. Display the frequency percentage distributions of the student cohort attributes.
7. Provide a synopsis of the data mining package selected.
8. Show the results of the application of the decision tree algorithm.
9. Breakdown the comprehensive analyses of the predicted versus actual outcomes by attribute.
10. Discuss the accuracy of the predictions and possible explanation for the unexpected results in some sub-categories in comparison to the research of others.
11. Share conclusions on why this method may be used to help meet the challenges facing higher education today.
12. Proffer recommendations for further research and analysis.

2. Data Collection and Processing

2.1 Data Colletion

In 1998, the Ohio Board of Regents, the state's higher education governing body, replaced its antiquated higher education data collection process, the Uniform Information System, with an updated and expanded data collection system, the Higher Education Information System (HEI). "The Higher Education Information (HEI) system contains data supplied by Ohio's colleges and universities. It is a comprehensive relational database that includes data on students, courses, faculty, facilities, and finances" (Ohio Board of Regents, 2009). The HEI System consists of several data modules, or primary data areas: Academic Programs; Enrollment; Facilities; Faculty-Staff; Financial; State

Grants and Scholarships (SGS) Financial Aid; and recently implemented Unit Record Tuition and Financial Aid.

Of interest in this study are the data submitted through the Enrollment data area. Data in this region include student demographic data (e.g., birth date, gender, race/ethnicity), state and county of residency, student enrollment data (e.g., course title, catalog number, credit hour for every course a student is enrolled in a given academic term), and student degree/certificate information. While some data elements are collected on an academic term basis (student and course enrollments) others are collected annually (degrees/certificates awarded). In addition to collecting a plethora of data values, the HEI System provides web access to an institution's submitted data to authorized users. This large volume of data can be used for verification purposes, benchmarking, longitudinal/trend analyses and in this case, data mining.

As Bailey (2006) stated, "with any kind of databases that contain multidimensional subjects and span multiple years, data mining is an ideal approach to identify hidden patterns and discover future trends of behaviors." Therefore, the HEI System serves as an excellent repository for and source of valuable and extensive data critical for this purpose. Because developing an accurate student success model using data mining requires a set of consistently captured data accumulated during the tracking of a specific entering cohort of students over a six-year period of time, a compilation of data related to first-time undergraduate students entering a Northeastern Ohio, public, four-year institution in 2001 was utilized in this research. Using the menu of predictors identified in the decision tree-related research by Herzog (2006), a subset of readily available HEI data elements was compiled. This subset was later supplemented with data from the institution's legacy data system in order to address college-readiness, student support (tutoring, supplemental instruction, etc.), and student financial need and financial aid awards.

2.2 Data Processing

The data files were initially downloaded from the HEI website or were obtained by querying the institution's legacy database. A university-issued student identifier served as a primary key between the files and facilitated the population of a cohort table stored in MS Excel spreadsheet format for processing. Each field was reviewed for its possible contribution to the classification model. If it became apparent that some fields were redundant, those fields were subsequently removed from the cohort table. In at least one case the range of different numerical data values for a specific field reached over 100. Because data mining algorithms typically examine each value looking for patterns within the data, fields with a large range of possibilities (e.g., federal financial aid awards, state financial aid awards, federal Work Study aid awards, student majors) were pared down through a discretization process to reduce the amount of differentiations and increase the accuracy of the classification model. Once the final table was populated, the table was converted to a comma separated value file required by Weka, the data mining package used in this study.

2.3 The Cohort

Cohort attributes were divided into five subcategories: General demographics; College readiness; Socio-economic/Financial data; Academic Ability; and Retention.

- General demographics included: gender; racial background; age; state residency; and housing status as of the first term of enrollment.
- College readiness included: ACT composite scores; high school graduating grade point averages; advanced placement credits; and academic intention (first term major could also be included to the extent that it is an indicator of an incoming student's academic focus).
- Socio-economic/Financial data included: student marital status (which one could argue is also a general demographic); financial dependence upon parents; calculated cost of attendance; 9-month estimated expected family contribution; financial aid determined need level; federal financial aid (excluding student loans); state aid; federal Work Study aid; institutional aid; other third party aid; and student loans.
- Academic ability was comprised of: first term academic load; attempted credit hours; credit hours earned; total quality points; and end of term grade point average; engagement in remedial English; remedial mathematics or Reading & Study Skills developmental coursework; and number of visits paid to the Center for Students Progress for assistance with peer mentoring, tutoring, and the like.
- Retention, included: returning the following spring term; continuous enrollment from entry fall term, to subsequent fall term including spring (but not accounting for summer term); and returning the following fall term.

The distributions of the values for these data are shown on the following charts.

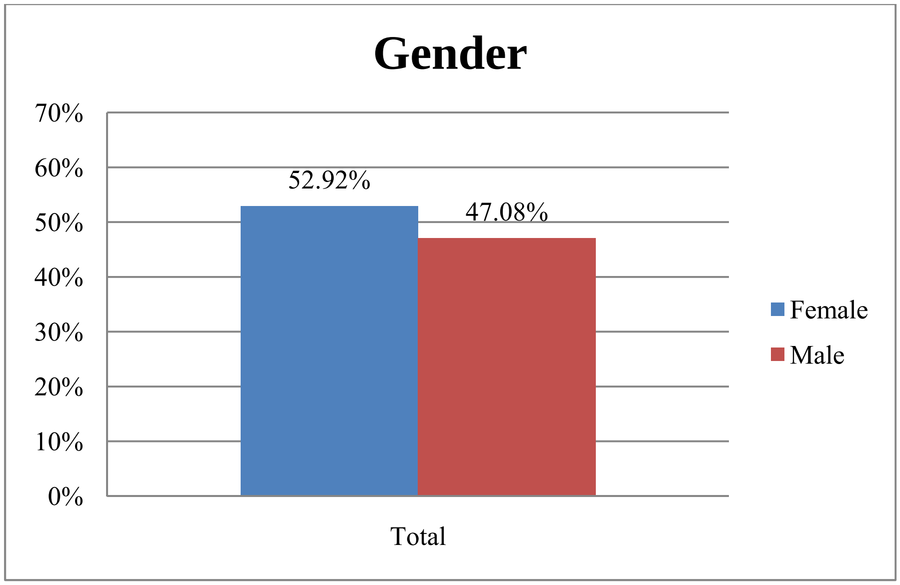

Figure 1

As has been the trend at this institution, just slightly more females than males (53% vs. 47%) were found in the 2001 cohort.

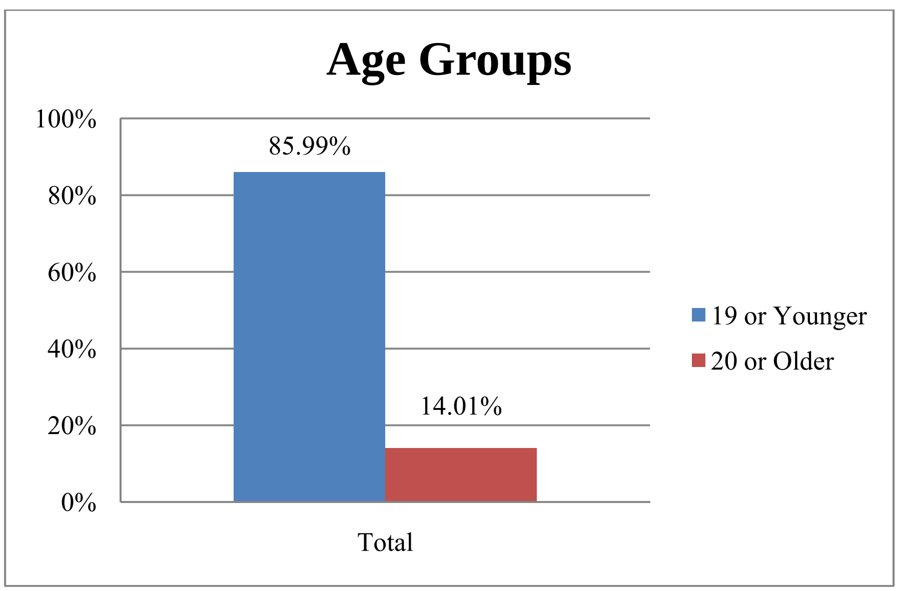

Figure 2

The clustering of years of birth around 1982 and 1983 indicates that most of the students in the cohort were in the traditional age group for college/university students. Nearly 86% of the students in the cohort were 19 years of age or younger.

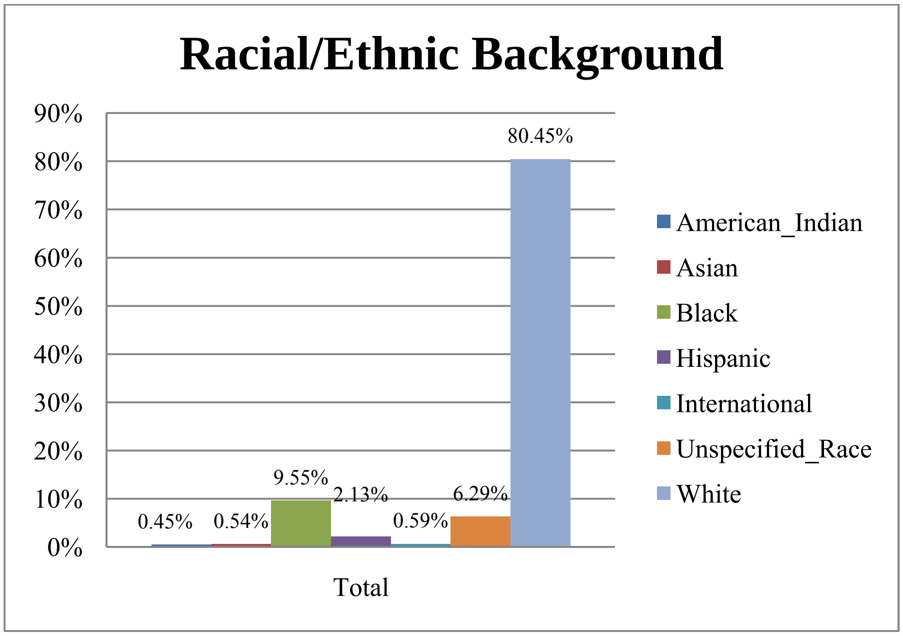

Figure 3

Slightly more than 80% of the students in the cohort were White. The next frequency in the category was Black with slightly less than 10%.

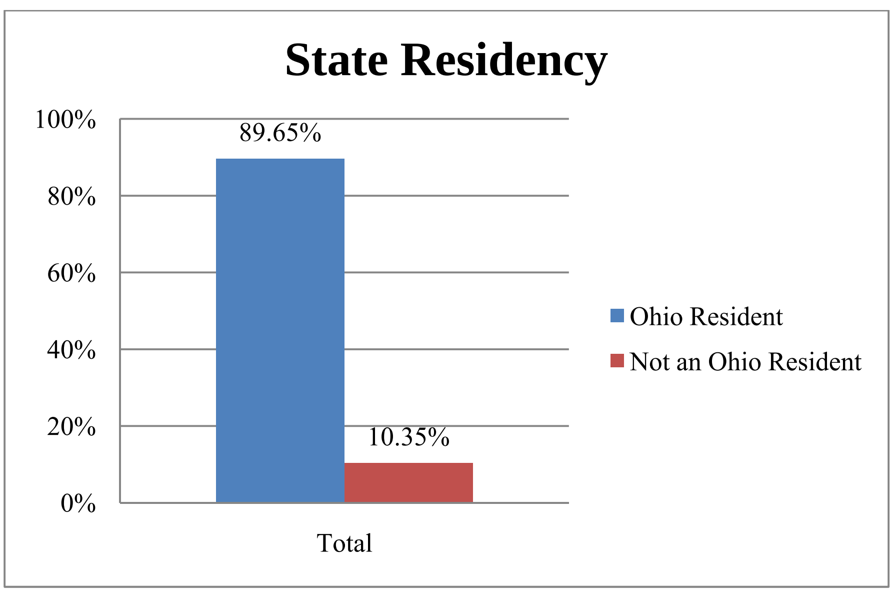

Figure 4

Close to 90% of the students were state residents at the time of application.

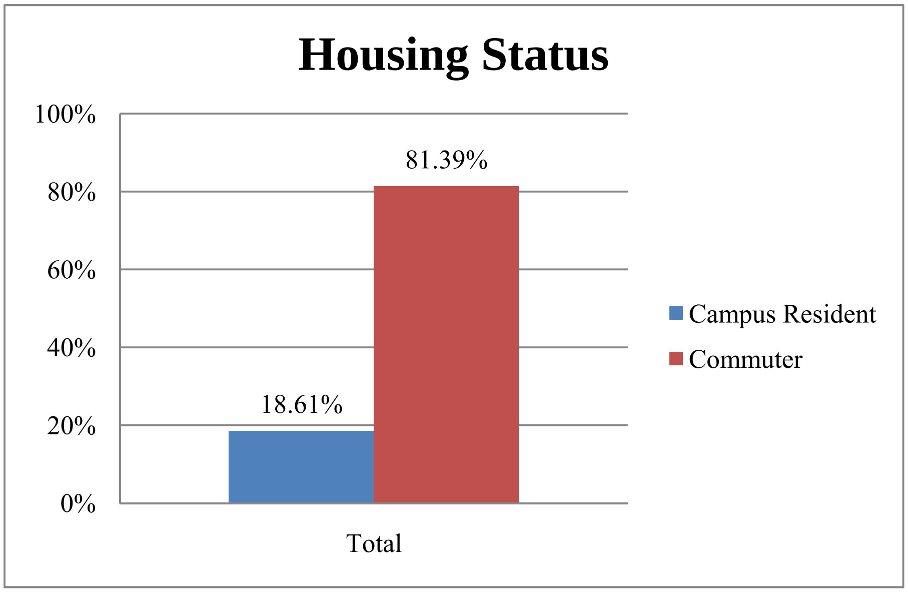

Figure 5

A minority of 18.6% of the cohort population fell into the campus resident category for the first term of enrollment.

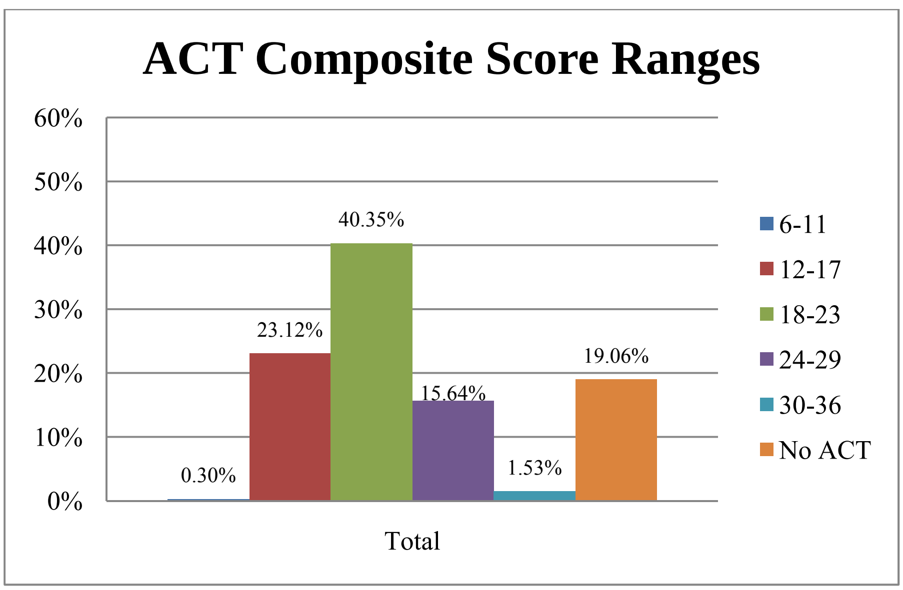

Figure 6

A majority of the cohort (over 80%) submitted an ACT composite score. Of that group 71% had a composite score of 18 or higher.

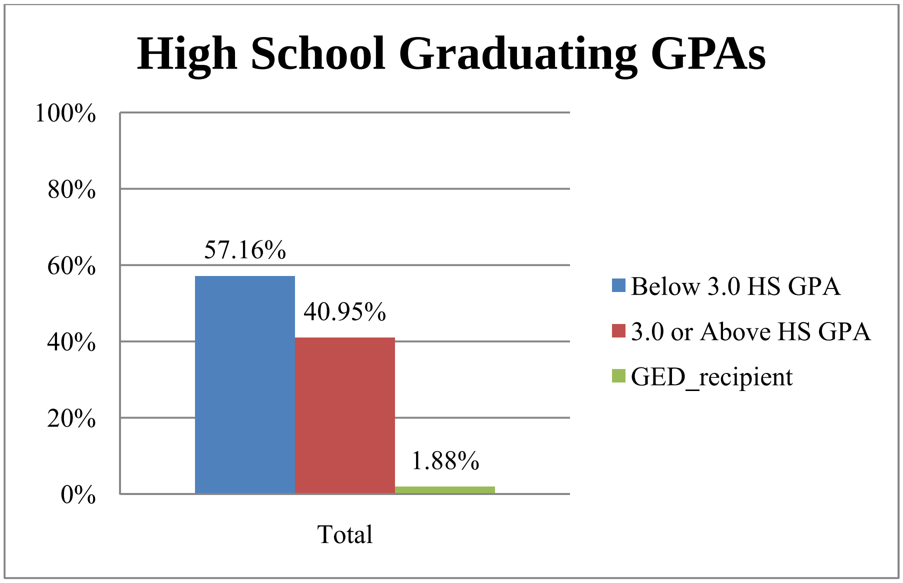

Figure 7

Most students entering in fall of 2001, submitted high school graduating overall grade point averages of below a 3.0.

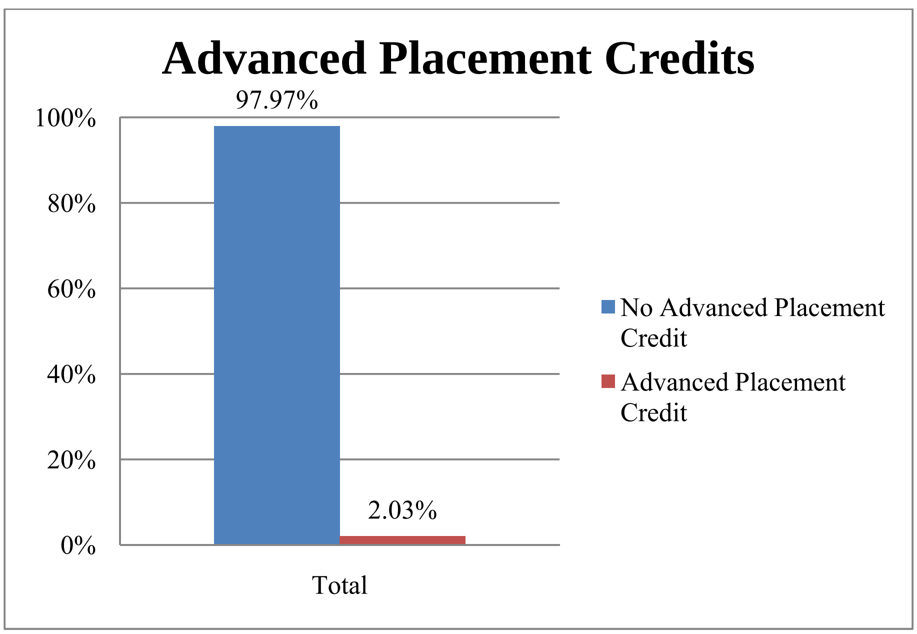

Figure 8

Nearly the entire student cohort group did not have any advance placement credits at the time of enrollment.

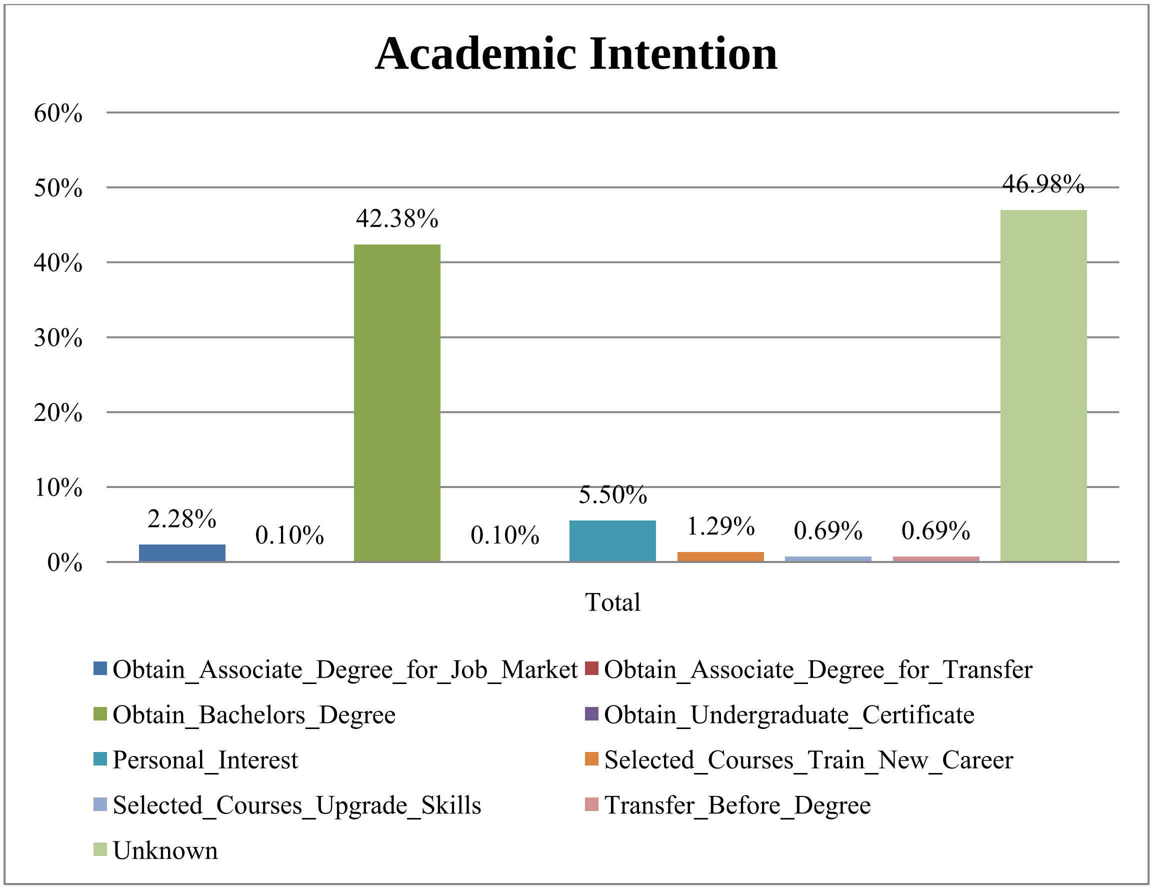

Figure 9

At the time of initial application to the institution, the majority of students in the cohort indicated that they intended to either obtain a bachelor's degree or that they did not know what their academic intention was (42.4% and 47% respectively).

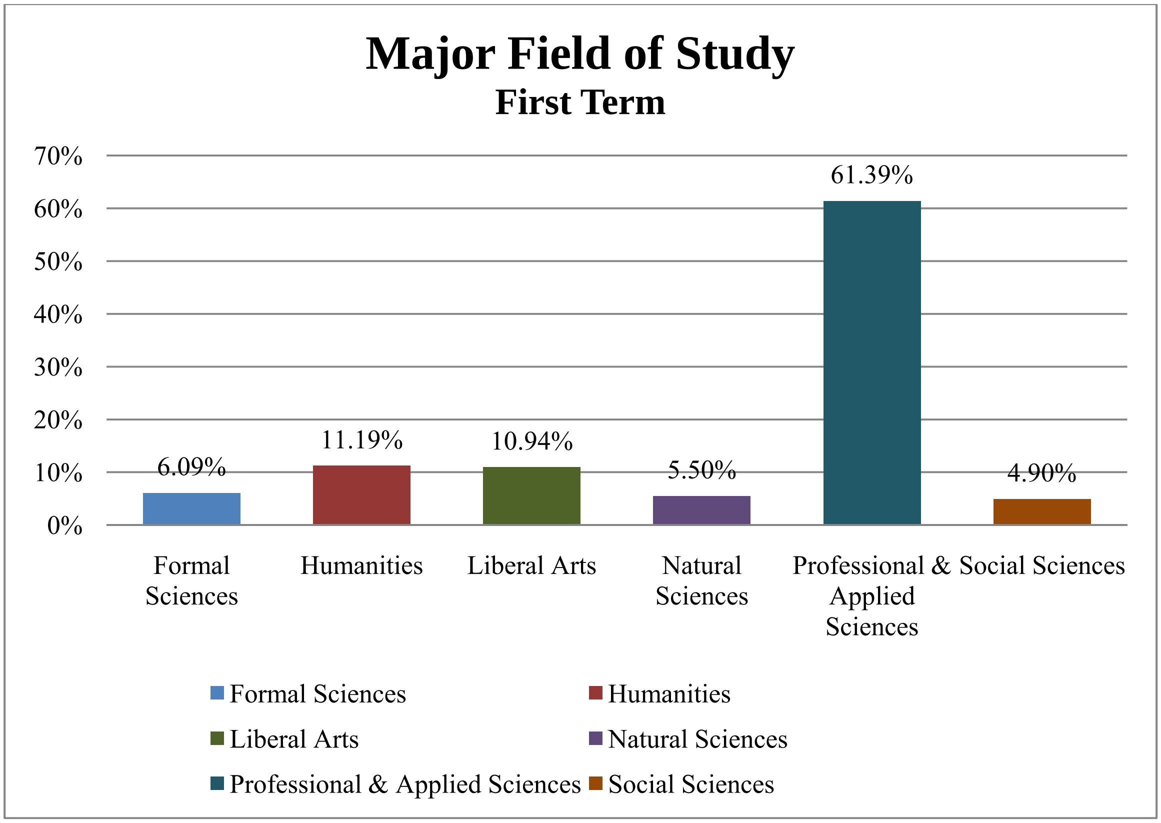

Figure 10

Though first term fields of study were widely dispersed, the Professional & Applied Sciences group held the most popular majors.

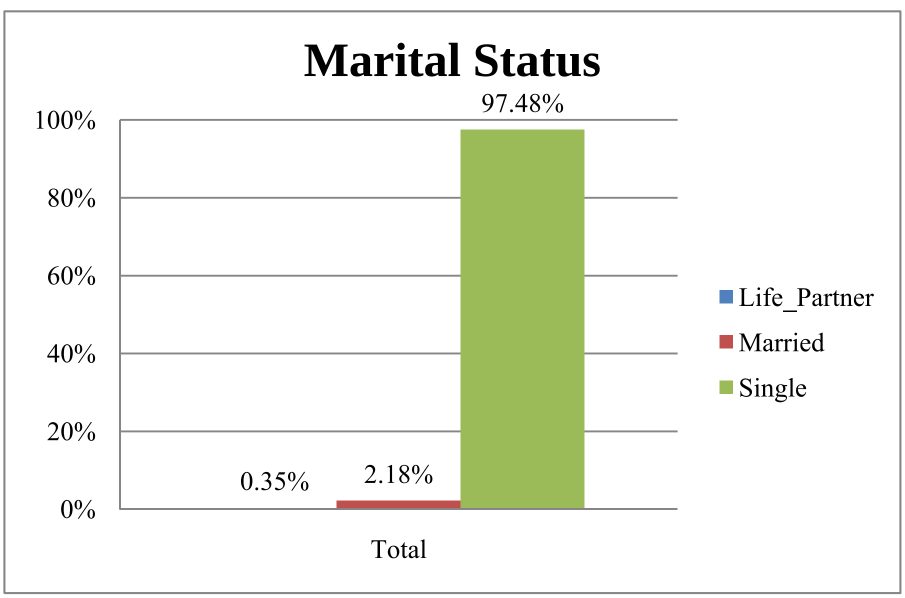

Figure 11

Of the students in this study (see Figure 11), approximately 97% were single, slightly more than 2% were married and less than 1% reported living with a life partner.

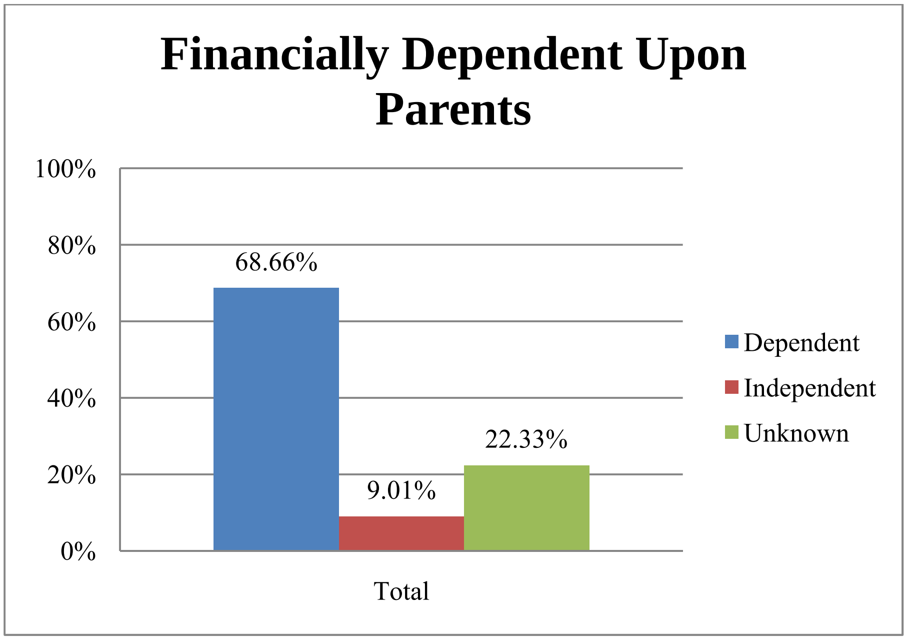

Figure 12

More than 2/3 of the cohort population was financial dependent upon their parents when they first entered the university.

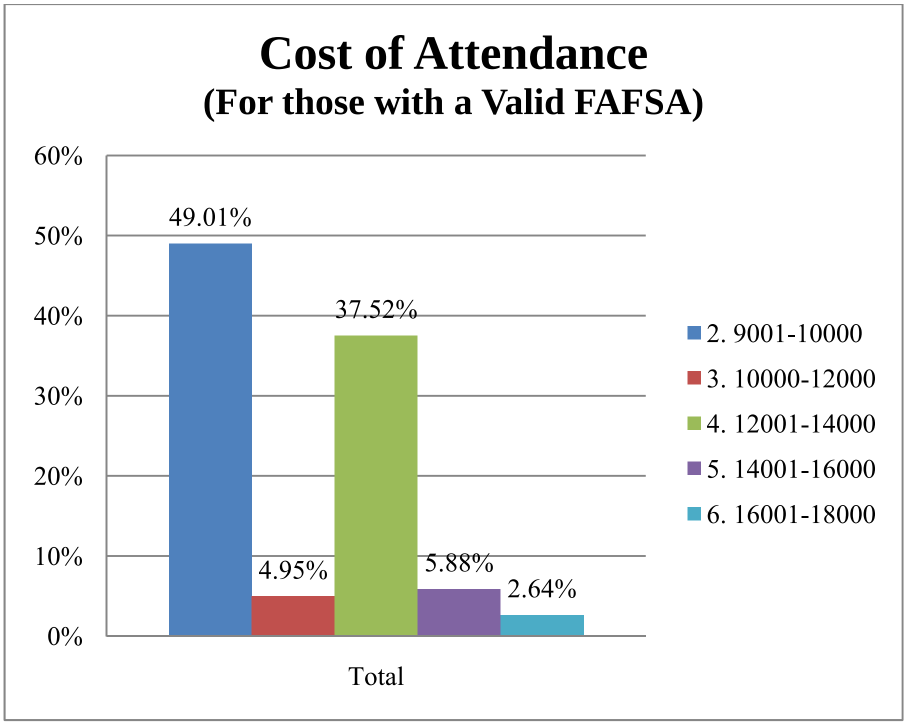

Figure 13

For those students submitting a valid FAFSA (Free Application for Federal Student Aid), the majority (49%) had an annual estimated cost of attendance equal to that of a full-time, in-state student.

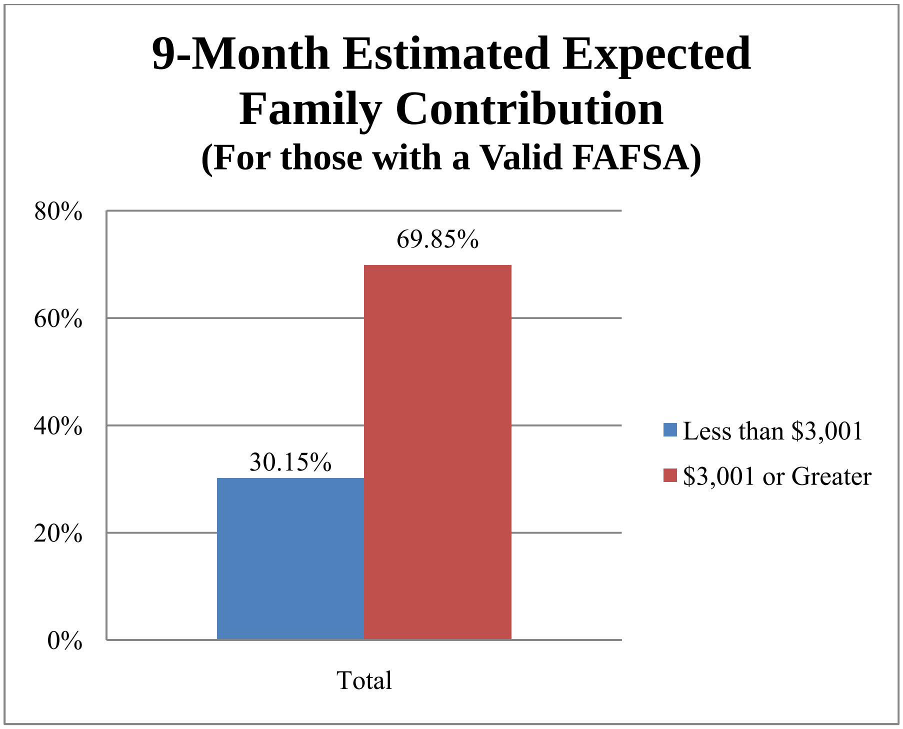

Figure 14

The 9-Month Expected Family Contribution (EFC) is the value the federal government believes a student's family is capable of paying as a result of calculations based upon the student's responses on the FAFSA. Most students had a 9-month expected family contribution of $3,000 or greater.

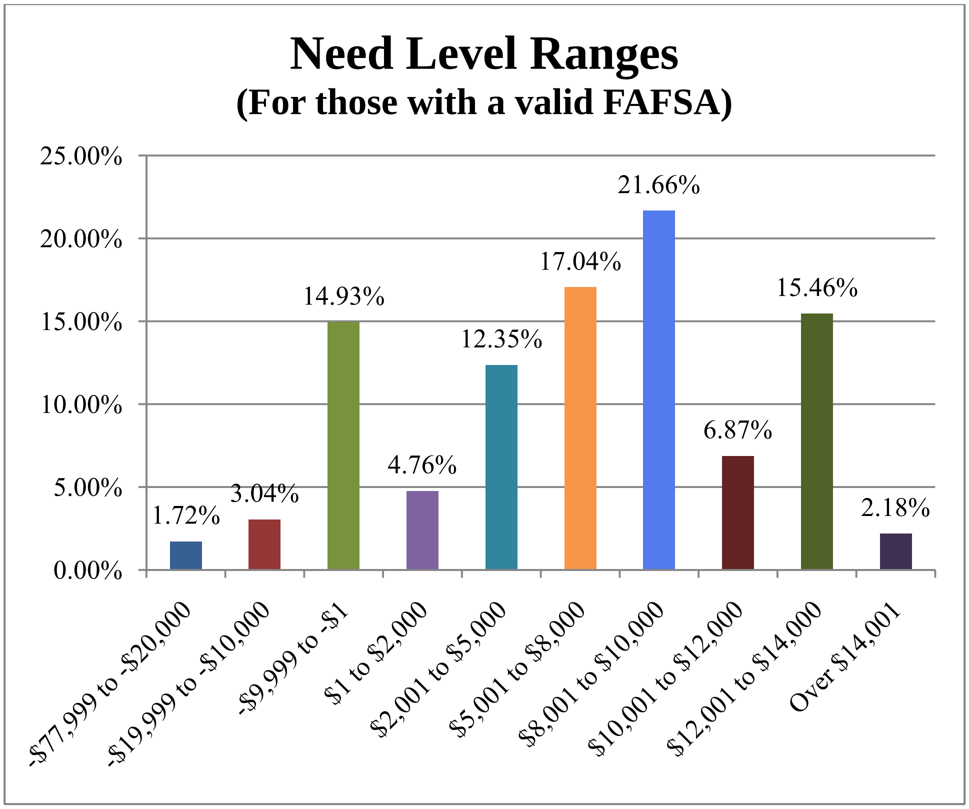

Figure 15

Need Level values are determined by examining the Cost of Attendance with respect to the 9-Month Estimated Family Contribution. The range of student need levels spanned from –$77,999 to over $14,001 with a majority of the students (over 80%) showing a financial need over$1 and up to over $14,001. It is noteworthy to point out that over 46% of the students showed a need level of over $8,000.

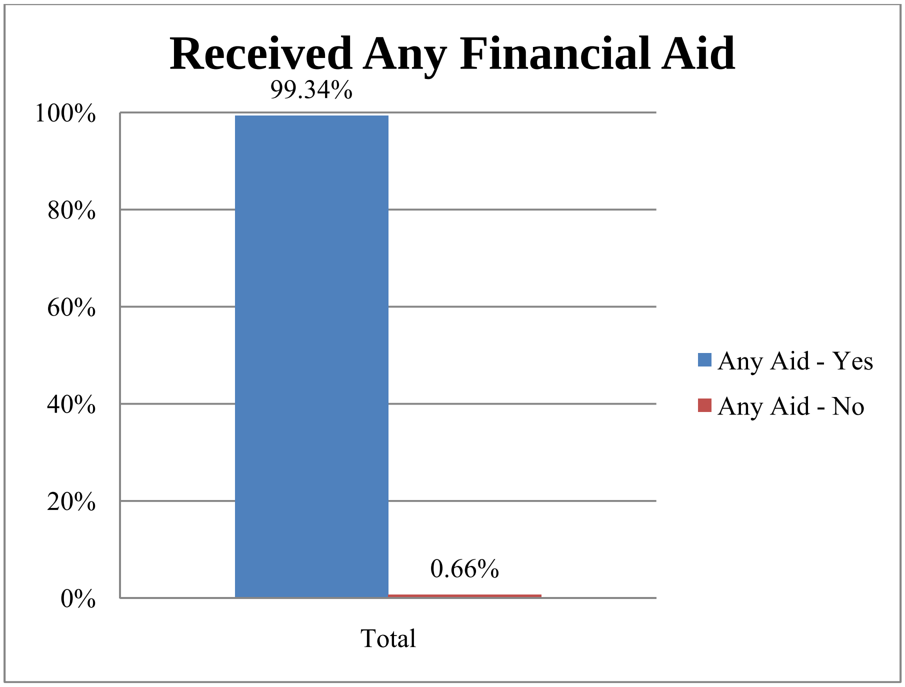

Figure 16

Nearly all students received some amount of financial aid.

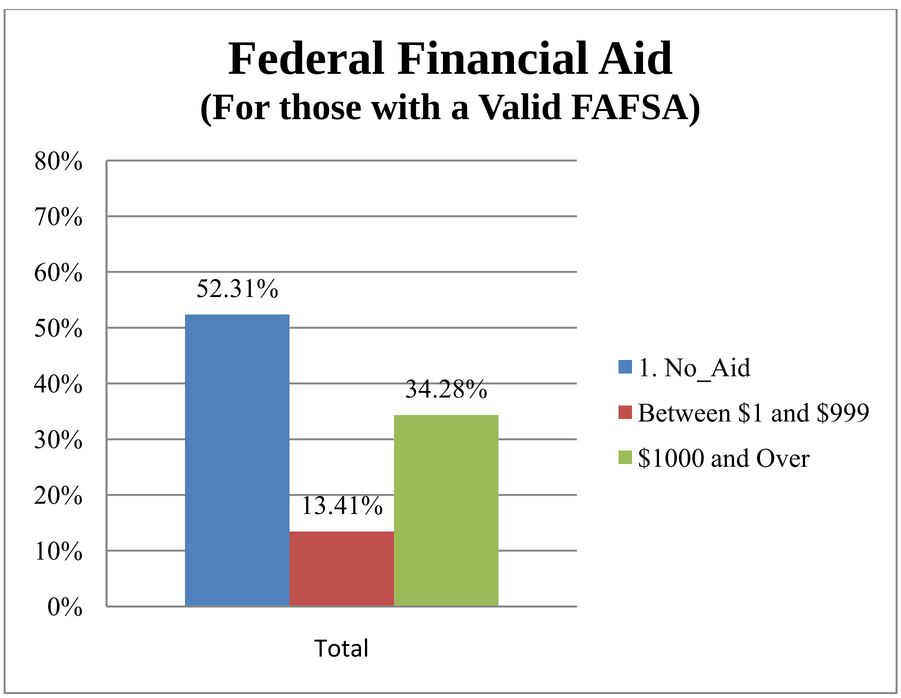

Figure 17

More than half of the cohort did not receive any federal financial aid. Of those receiving federal financial aid, most received $1,000 or more.

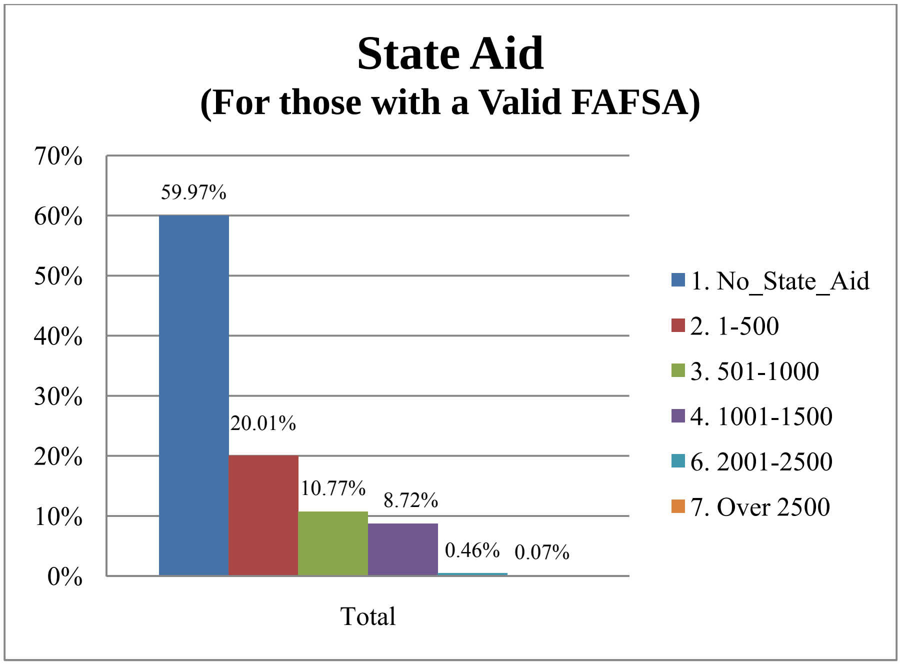

Figure 18

Just about 40% of the student population in this study received state aid. Most of those receiving state aid were awarded $500 or less.

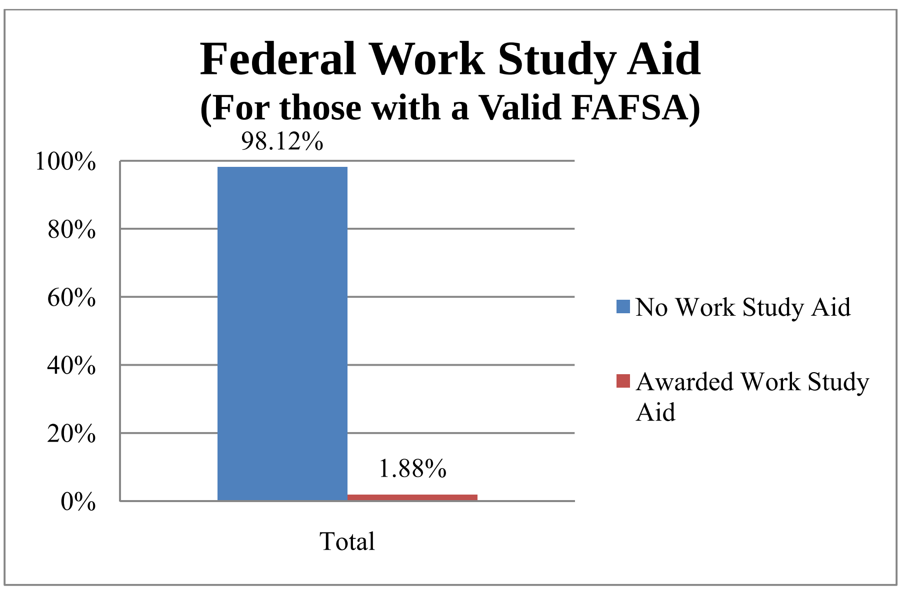

Figure 19

Less than 2% of the cohort participated in the federal work study program, which sponsors part-time jobs for students to earn money for college.

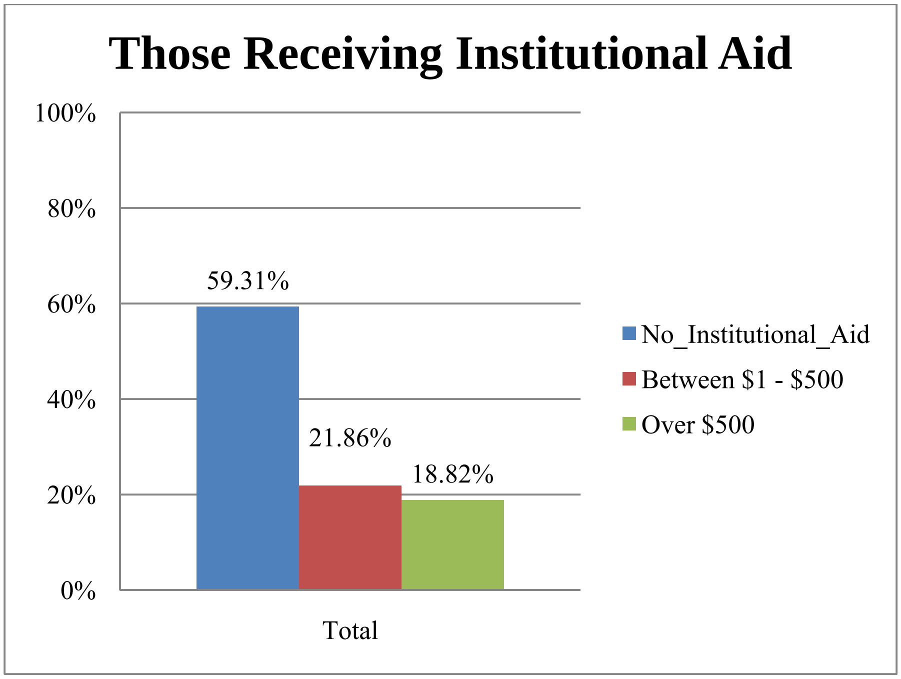

Figure 20

Institutional Aid includes fee remission for current and retired employees, their spouses and dependents, and internal and external scholarships administered by the institution. Just about 40% of the cohort student group received some amount of institutional aid.

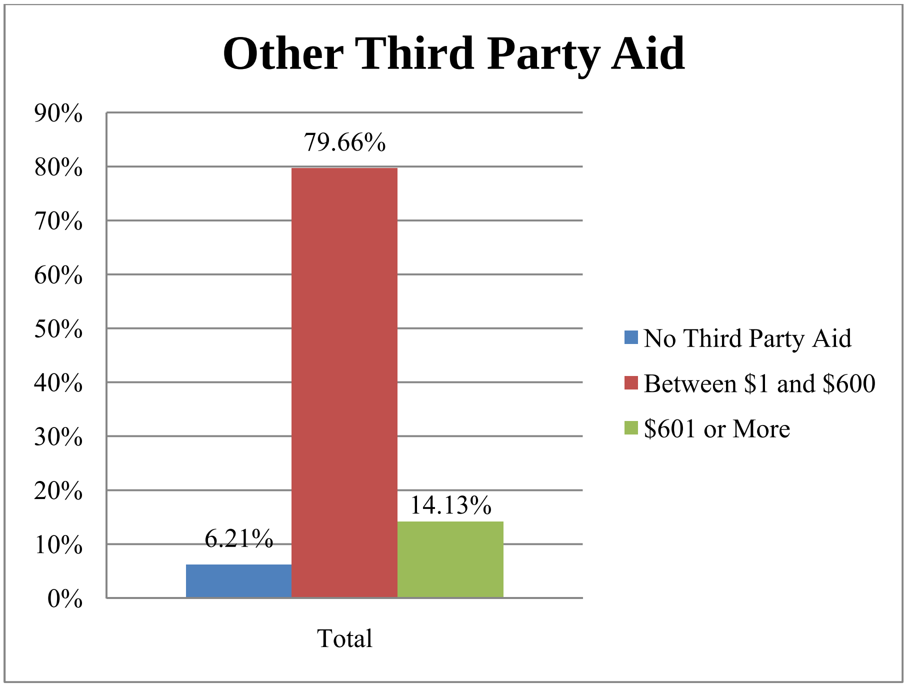

Figure 21

Agency funded money for retraining and local independent awards categorized as Other Third Party Aid was received by almost all (94%) of the students in the cohort.

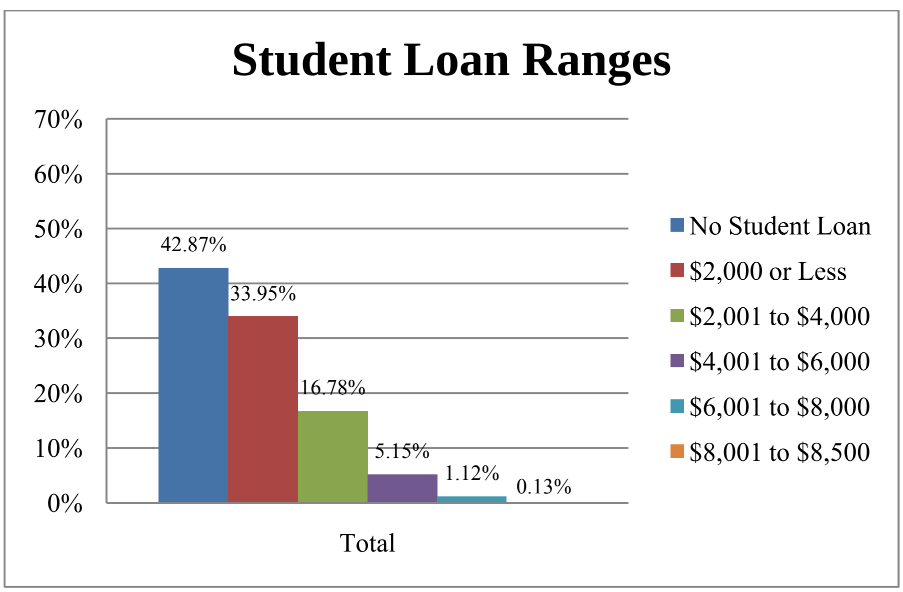

Figure 22

More than half (57%) of the cohort population took out a student loan during their 1st academic year of study – most in the amount of $2,000 or less.

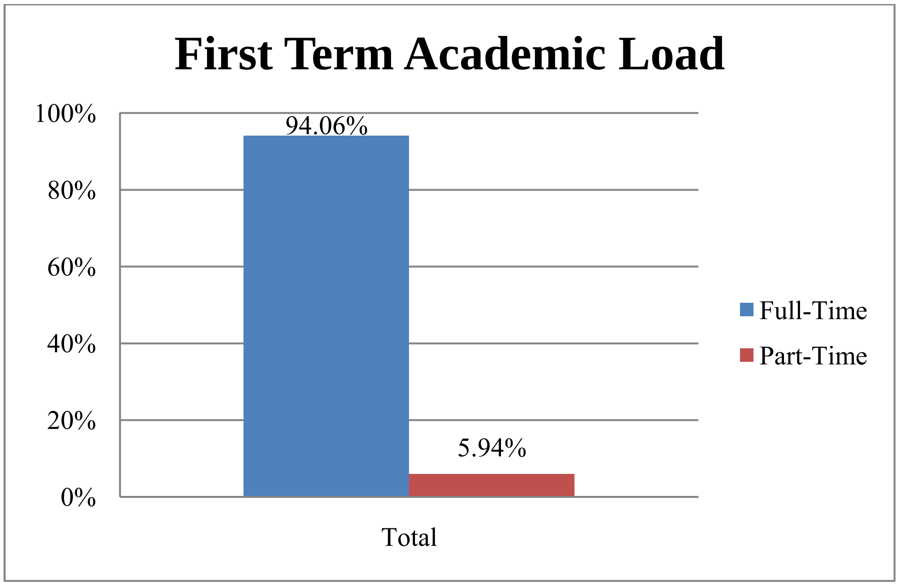

Figure 23

Almost all (94%) of this special population engaged in a full-time academic load of 12 or more credit hours of course work.

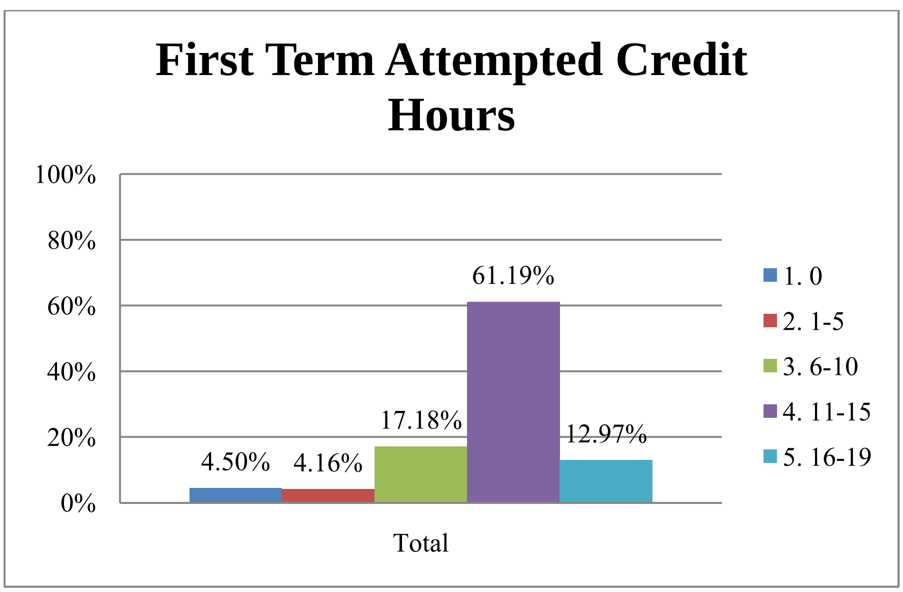

Figure 24

The distribution of first term attempted credit hours indicates that most students were enrolled between 11 – 15 credit hours in for credit courses. As this study attempts to predict students entering at a specified point and earning a bachelor's degree within six years, it is important to point out that in order to earn the required average of 126 credit hours for a bachelor degree within the specified timeframe, students need to complete roughly 16 credit hours per term to graduate within four academic years, 13 credit hours per term to graduate within five years, and 11 credit hours per term to graduate within six years. Additionally, in order to receive the maximum amount of financial aid a student is eligible to be awarded, a student must be enrolled in at least 12 credit hours of course work.

Note that 4.50% of the cohort population was engaged in coursework, in all likelihood auditing courses, which upon completion would earn them no academic credit. Further examination of this group may reveal that the academic intention of this group of students indicated that they were seeking something other than an academic certificate or degree.

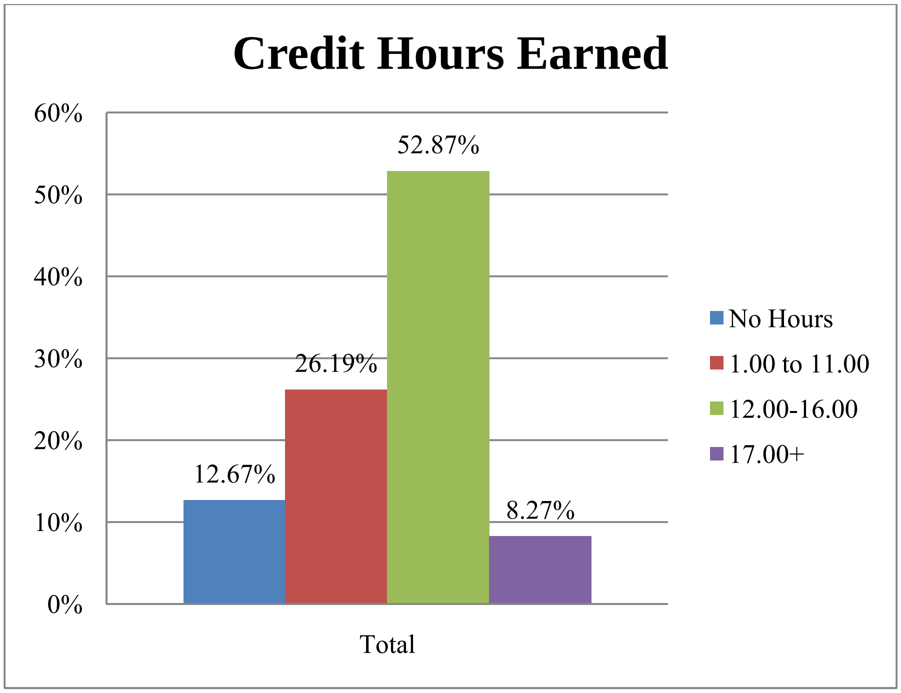

Figure 25

Where only 4.5% of the cohort knowingly enrolled in course work applicable for no academic credit, nearly 8% more (or 12.7%) actually earned no academic credit. This additional 8% can be attributed to students completely withdrawing from all their coursework after the enrollment census point, students failing all of their coursework for the term, or students who failed to officially withdraw from the institution earning non-attendance failing grades.

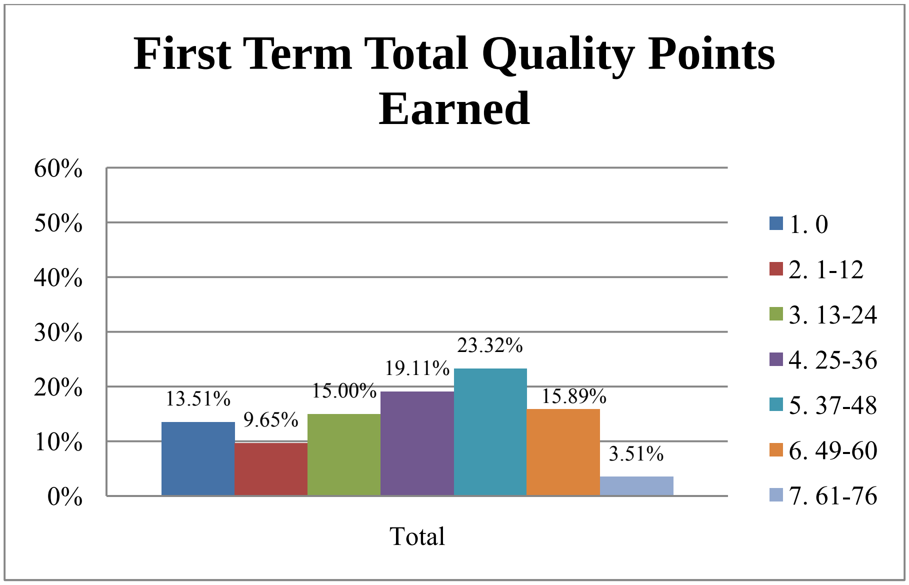

Figure 26

During the first term of enrollment for this cohort the institution had no policy in place for distinguishing those students earning non-attendance failing grades from those students actually earning failing grades. Embracing the belief that at least half of the students in the 0 quality points range either officially withdrew after the census point or earned non-attendance failing grades, the quality points (the values 4 through 0 assigned in accordance with letter grades earned in course work) earned resemble the normal distribution.

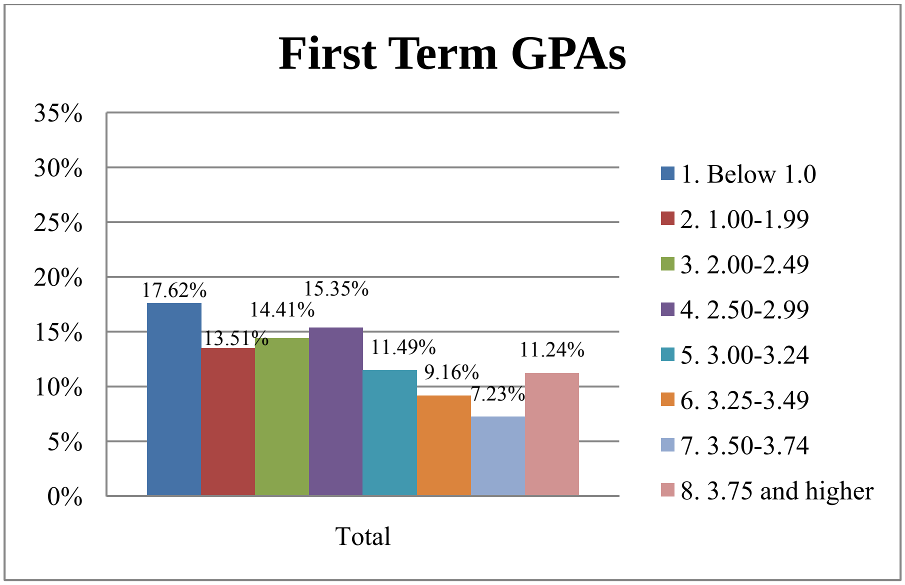

Figure 27

At the conclusion of the first academic term of study, more than half of the students in the cohort had grade point averages below 3.0 (61% vs. 39%).

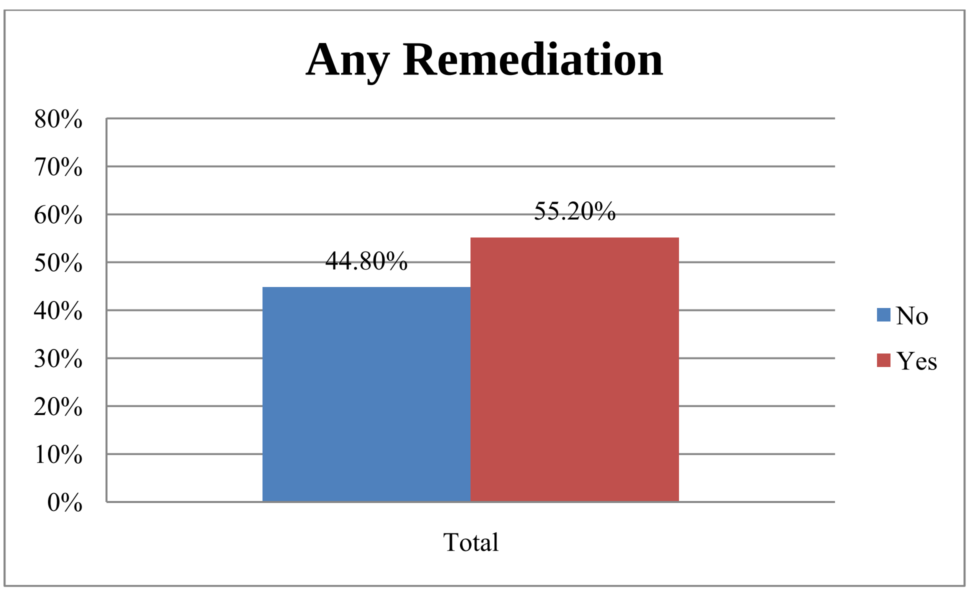

Figure 28

Interestingly, irrespective of placement testing recommendations, more than half (55%) of this population at sometime during their academic careers engaged in remedial coursework.

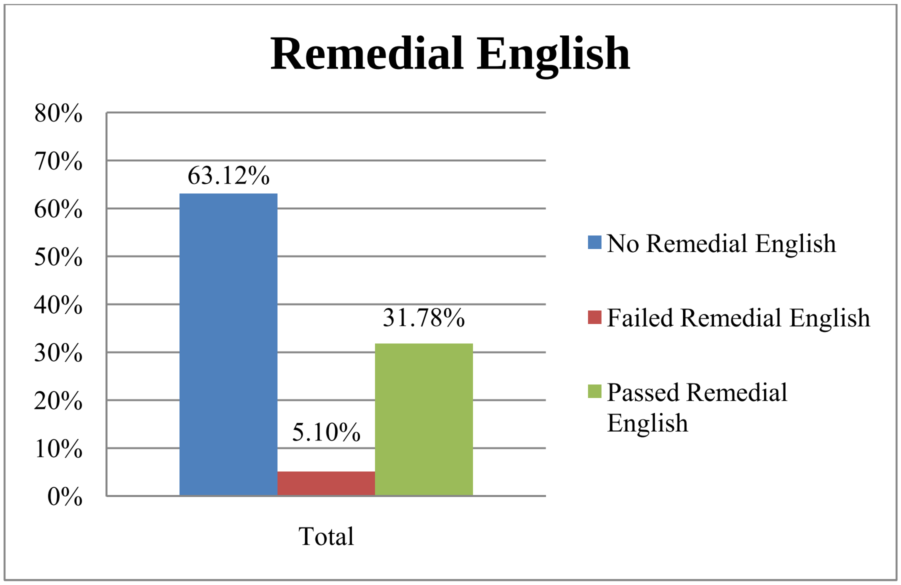

Figure 29

Just about 37% of the population engaged in remedial English.

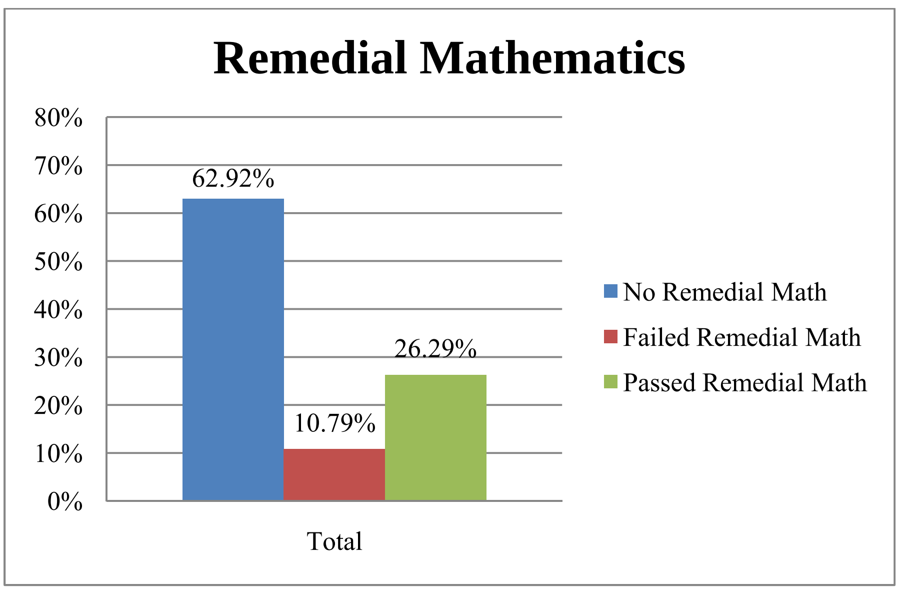

Figure 30

Likewise nearly 37% of the population engaged in remedial mathematics.

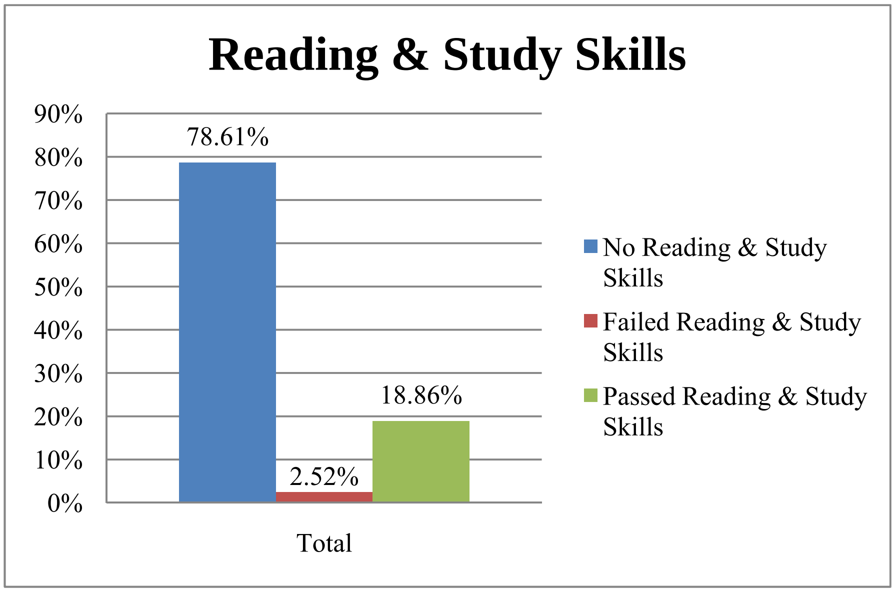

Figure 31

And a little more than 21% of the population engaged in Reading & Study Skills course work specifically "designed to develop students' skills essential for college studying" and believed to assist underprepared students in achieving a state of college readiness (YSU's Undergraduate Catalog, 2009).

Of the three categories of remedial coursework, remedial mathematics had the largest amount of students engaged with 749 out of the 2,020 in the cohort. This figure is just four more students than the number engaged in remedial English.

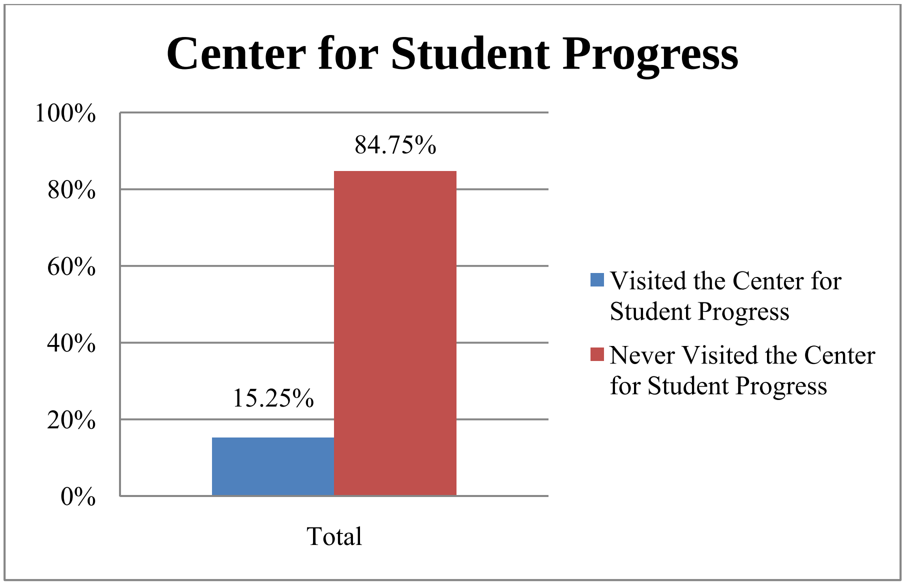

Figure 32

Only 15% of the cohort population visited the Center for Student Progress during their first term, thus taking advantage of any of the numerous the services they provide (e.g. tutorial, individual intervention, or supplemental instruction services) for helping students "acquire the skills and knowledge needed to become successful learners" (Center for Student Progress, 2009).

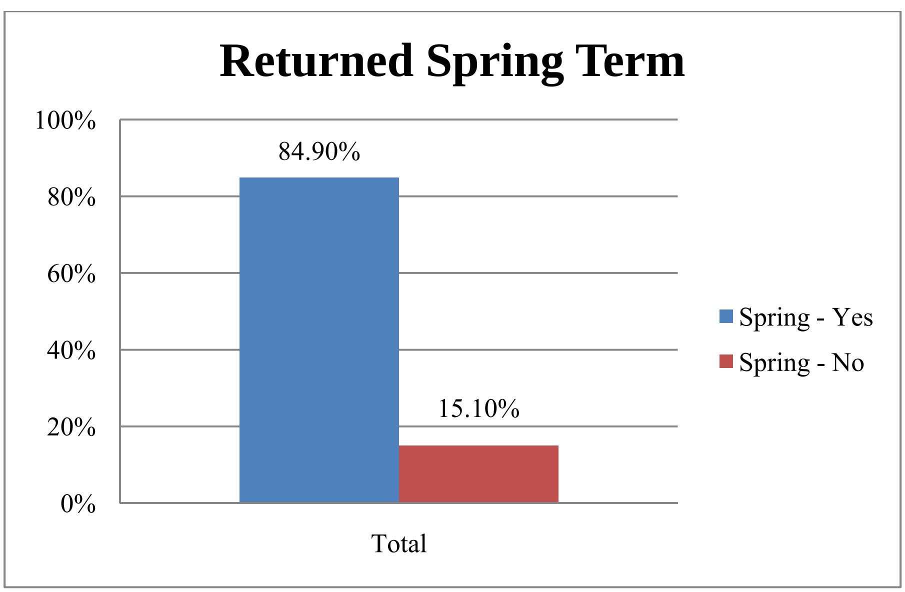

Figure 33

A majority (85%) of the entering student population continued to be enrolled the following spring term.

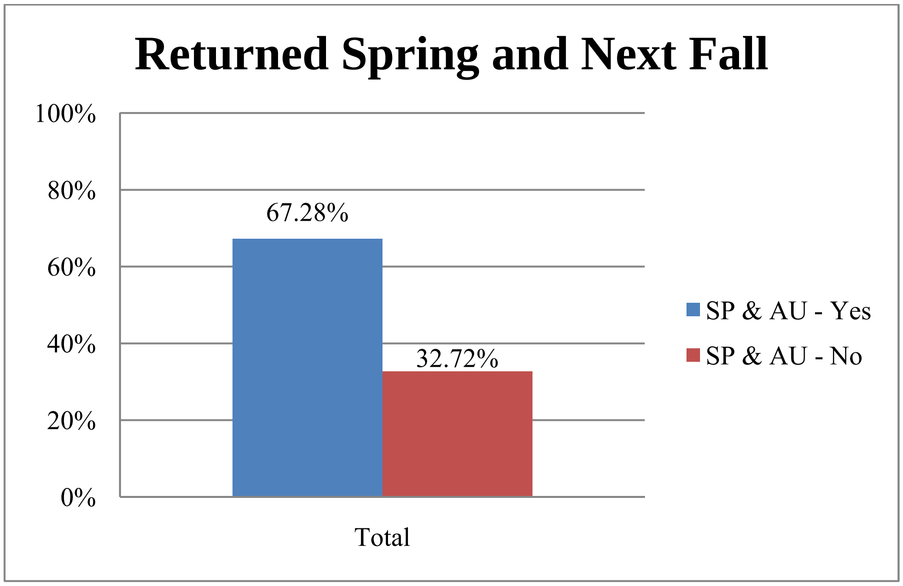

Figure 34

And slightly more than 67% of the entering cohort was enrolled the first three consecutive terms (fall 2001, spring 2002 and fall 2002).

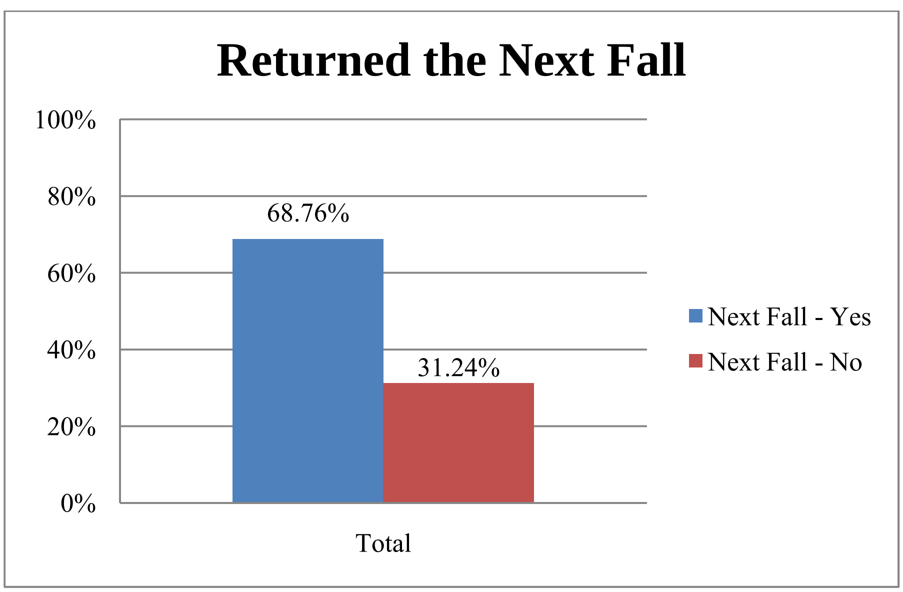

Figure 35

Additionally almost 69% returned to the institution the following fall term. This figure is down roughly 16% from those who continued through the first spring but up slightly from those students enrolled consecutively fall, spring and fall of the subsequent year – indicating that some students not enrolled the immediately following spring term do in fact still return the following fall.

2.4 Data Mining Tool – Weka

Herzog (2006) found that “when working with large data sets to estimate outcomes with many predictor variables, data-mining methods often yield greater prediction accuracy, classification accuracy, or both [than that of traditional statistics]”. Therefore rather than perform typical statistical analyses, data mining, in particular free-to-the-public, open source data mining software, was employed for this endeavor.

The open source data mining software, Weka, which stands for Waikato Environment for Knowledge Analysis, is a machine learning project undertaken by The University of Waikato. The primary goal of the “project is to build a state-of-the-art facility for developing machine learning (ML) techniques and to apply them to real-world data mining problems.” The software is actually a “workbench” of commonly known algorithms accessible through four different interfaces (The University of Waikato, 2009). The four interfaces are accessible via the Weka GUI Chooser.

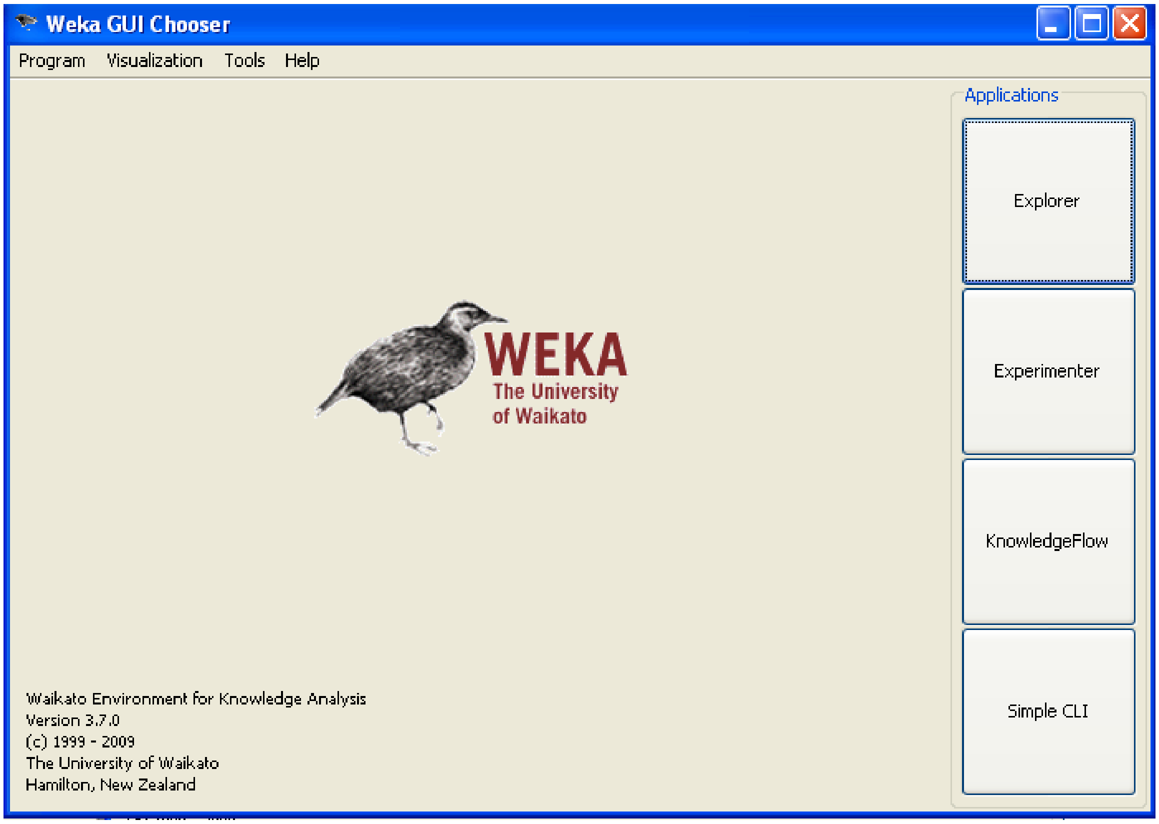

Figure 36

The Simple command line interface, or Simple CLI, is a text based interface to the workbench which requires the user to already be familiar with the software's facilities. The second option, the Knowledge Flow interface, requires an extensive amount of main memory to operate is consequently useful in analysis of small- to medium-sized datasets. It allows you to drag and drop icons representing the different algorithms on to the screen and design your own custom configurations for streamed data processing, again requiring some strong working knowledge of data analysis. The Experimenter interface provides assistance in determining which parameter values and algorithms will produce the strongest result for the problem at hand. And finally the Explorer interface, allows a novice user to easily upload a dataset and employ any of the software's features via menu selections and dropdown lists.

The software has been developed with an easy-to-use, intuitive style. Interface forms are set up to guide the user through the necessary steps in an appropriate order and, like other commercially available software packages, grey-out the selection items that are not available under the present conditions (Witten & Frank, 2005). The Explorer interface was used exclusively to perform the analysis on this project.

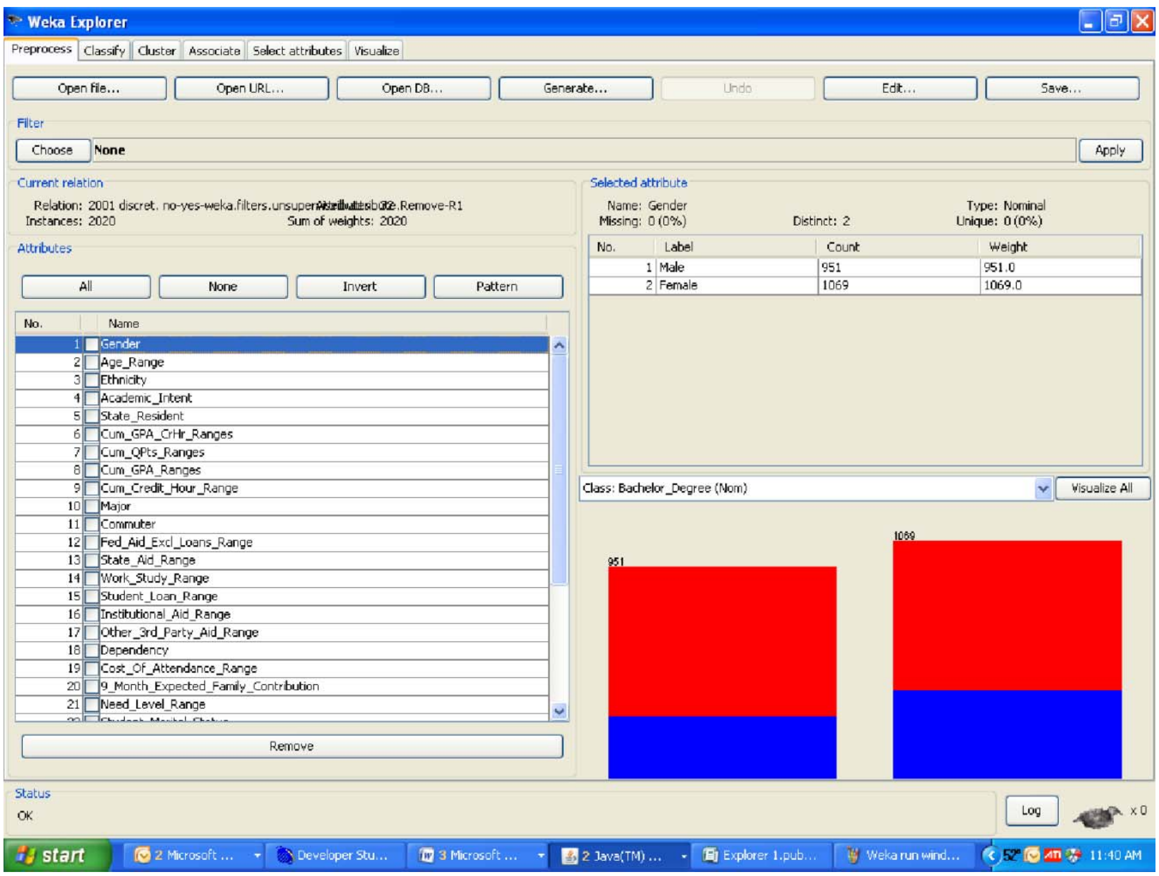

Figure 37

Upon entrance into the Explorer interface, the user must open an appropriately formatted data file. Weka accepts many types of data files, including comma-delimited

(.csv) files. In this study, .csv files were compiled because of the ease of formatting available with MS Excel 2007. After opening the data file the user is able to view and as needed remove attributes from the dataset via the Attributes window. This feature facilitates analysis by removing the attribute only from the Weka interface and not from the underlying dataset.

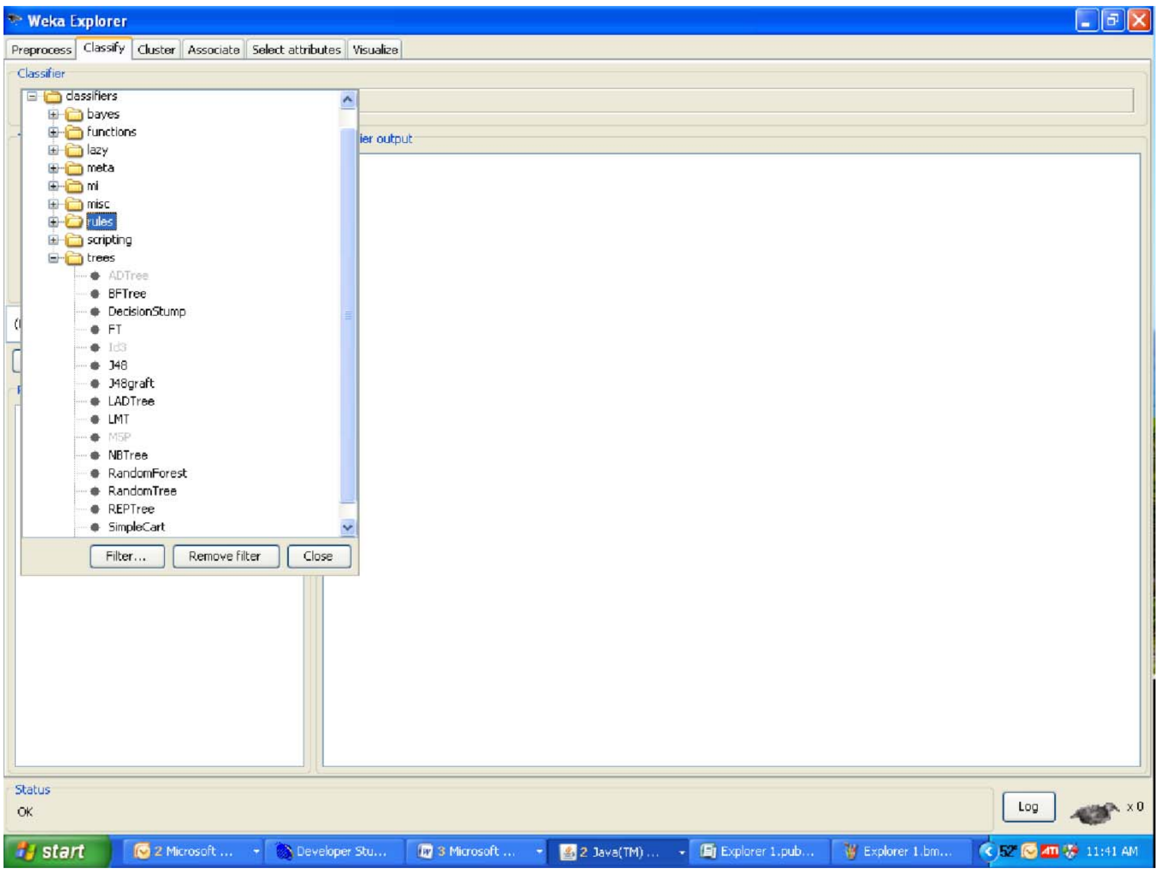

Figure 38

Once the desired attribute listing has been compiled, the user seeking to develop a decision tree then clicks on the Classify tab at the top of the screen to enter the next phase of processing. Here the user is able to access the many available classifier algorithms by clicking on the Choose button (not visible in this screen shot.)

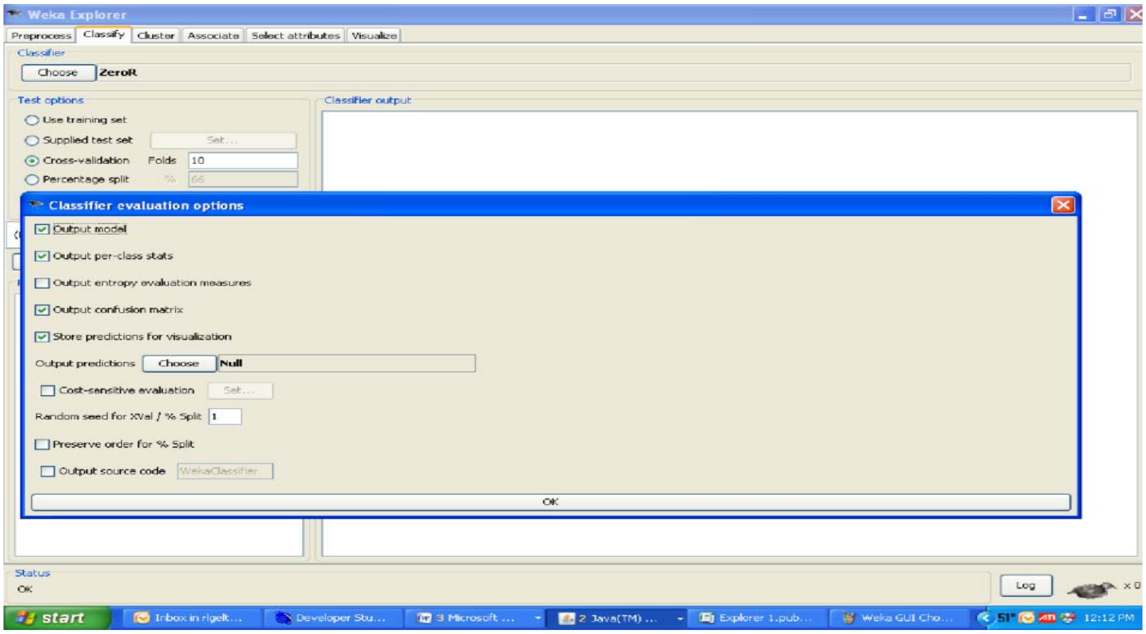

Figure 39

Next the user selects the desired evaluation options. It is important in this window to be sure to check the Output model, Output per-class stats, Output confusion matrix, and to Choose the file type for the Output predictions. These predictions are later appended to the original dataset in order to facilitate the development of MS Excel pivot tables and subsequent pivot charts for presenting the data analysis. Then from the Test Options dropdown box the user selects the target attribute, in this case Bachelor degree. Then the user clicks the Start button.

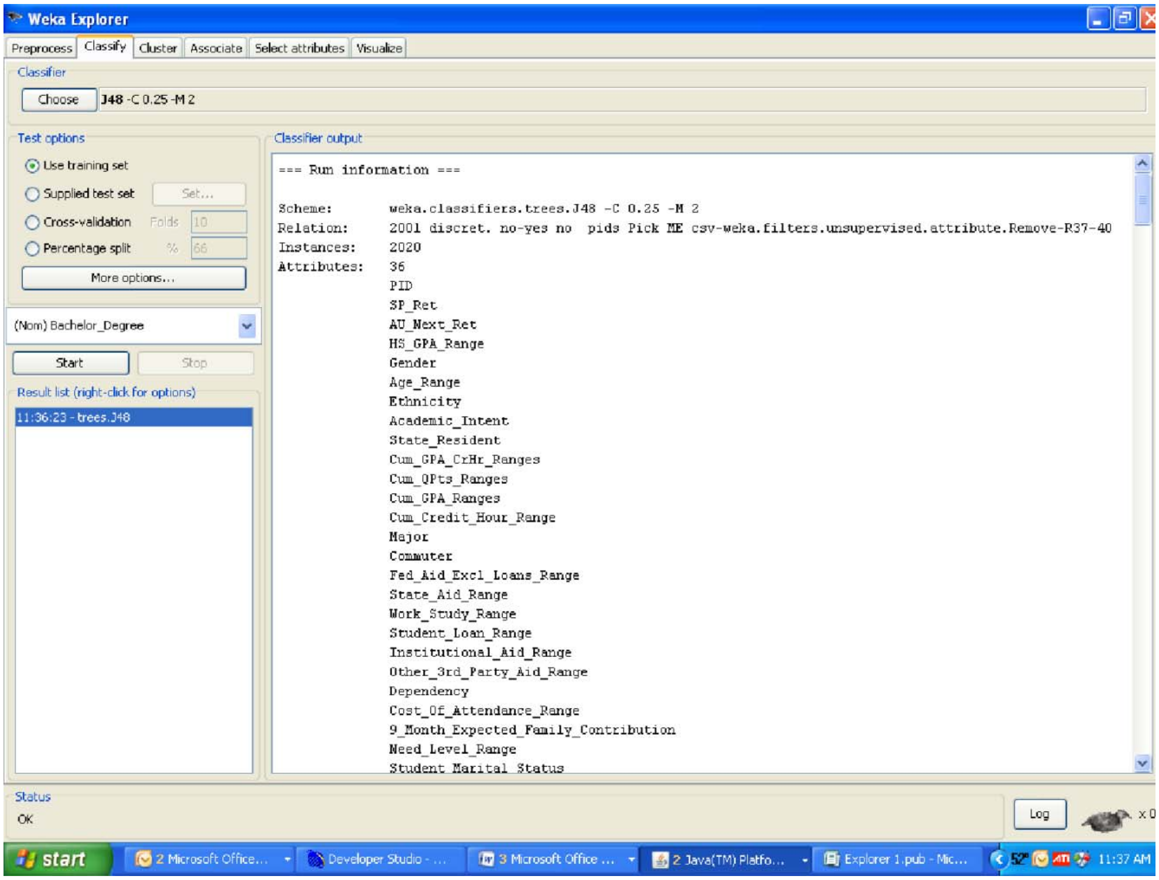

Figure 40

Within a few minutes, Weka produces the selected output and feeds it back to the user screen.

3. Analysis

The initial dataset contained attributes influenced by the published research of many in the field of institutional research supported by the recently published work of Bowen, Chingos and McPherson (2009). A majority of the data elements were selected based on their availability within the first academic year of study (e.g. ACT Composite test score, high school graduating grade point average, first term student credit hour load, returned spring term, etc.). For the data mining output, in this case a decision tree, to provide a meaningful predictor of future student success, it is important that the dataset be comprised of attributes significant to the accurate prediction of outcome as early as possible in a student's academic career - thus, affording the institution time to intervene.

Because "each technique employs a learning algorithm to identify a model that best fits the relationship between the attribute set and class label of the input data" (Tan, Steinbach & Kumar, 2006) , after determining the list of data elements desired for building the model, Weka was employed to process the data using the available decision tree classifiers. The J48 algorithm produced the strongest accuracy based on the initial dataset and was therefore chosen for developing the final decision tree model. Note that in order to increase the precision of the predictions, some original dataset attributes were removed or modified and others introduced.

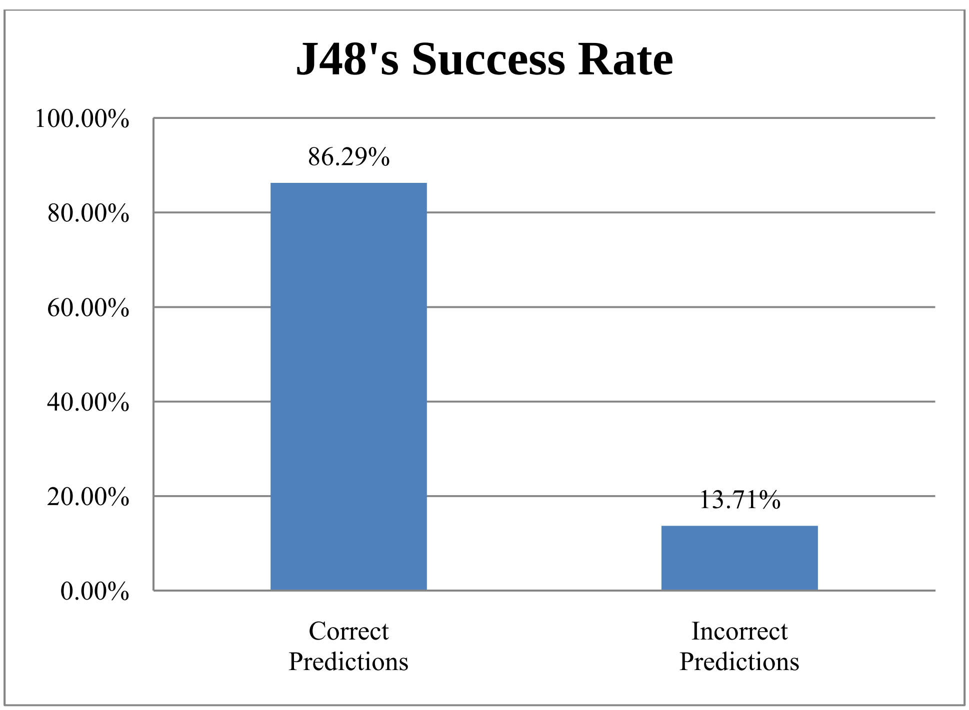

Figure 41

Once the beginning stages of analysis yielded less accurate results than expected, a re-evaluation of the dataset attributes took place.

The initial dataset consisted of the following data elements:

- Gender
- Year of birth
- Ethnicity/race
- Zip code
- Academic intent
- Student rank
- State residency
- First term attempted credit hours
- First term cumulative quality points
- First term cumulative grade point average
- First term cumulative total credit hours
- Major code
- Living arrangements
- Federal financial aid (excluding student loans) 2001-02
- State financial aid 2001-02
- Federal Work Study aid 2001-02
- Student loans 2001-02
- Institutional aid 2001-02
- Other third party aid 2001-02
- Dependency upon parents 2001-02
- Parent marital status 2001-02
- Student marital status (FAFSA) 2001-02
- Student marital status FAFSA code 2001-02
- Parental family size 2001-02
- Cost of attendance 2001-02
- 9-month estimated expected family contribution
- Need level
- Student marital status from legacy system
- Academic load
- High school CEEB code
- High school graduation year
- High school class standing
- Number of students in high school graduating class
- Advanced placement credit – Biology
- Advanced placement credit – Chemistry
- Advanced placement credit – English
- Advanced placement credit – Foreign Language
- Advanced placement credit – History
- Advanced placement credit – Math/Statistics

- Associate degree earned in 2 years, 3 years, 4 years, 5 years, 6 years, or 7 years
- Bachelor degree earned in 2 years, 3 years, 4 years, 5 years, 6 years, or 7 years
- Master degree earned in 5 years, 6 years, or 7 years
- Post Baccalaureate certificate earned in 6 years or 7 years
- Undergraduate certificate earned in 4 years, 5 years, or 6 years
- Visited the Center for Student Progress (yes/no), visited 1 - 5 times, visited 6 – 10 times, visited 11-15 times, visited 16-20 times, visited 21+ times
- Passed remedial English 1540T
- Failed remedial English 1540T
- Passed remedial English 1540
- Failed remedial English 1540
- Passed remedial math 1501
- Failed remedial math 1501
- Passed Reading & Study Skills 1510B
- Failed Reading & Study Skills 1510 B
- Passed Reading & Study Skills 1510A
- Failed Reading & Study Skills 1510A

After the initial data mining process was employed

the following data elements were removed:

- student zip code at time of application
- student rank, parental marital status
- student marital status (from the FAFSA form)
- parental family size
- high school graduation year
- high school class standing
- number of students in high school graduating class
- earned a master degree in 5 years, 6 years or 7 years
- earned a post baccalaureate certificate in 6 years or 7 years
- earned an undergraduate certificate in 4 years, 5 years or 6 years

the following data elements were discretized:

- year of birth – age ranges
- federal financial aid (excluding student loans)
- major field of study – first term

- state financial aid
- federal Work Study aid
- student loans
- institutional aid
- other third party aid
- cost of attendance
- 9-month expected family contribution
- need level
- advanced placement (AP) credits in biology, chemistry, English, foreign languages, history, or mathematics to a dichotomous (yes/no) field for any AP credits
- earned an associate degree in 2 years, 3 years, etc. to a dichotomous (yes/no) field for earned an associate degree ever
- earned a baccalaureate degree in 2 years, 3 years, up to 6 years to a dichotomous (yes/no) field for earned a baccalaureate within 6 years
- visited the Center for Student Progress
- failed remedial English and passed remedial English to a trichotomous field (did not take, failed, passed)
- failed remedial mathematics and passed remedial mathematics to a trichotomous field (did not take, failed, passed)
- failed Reading & Study Skills and passed Reading & Study Skills to a trichotomous field (did not take, failed, passed)

the following data elements were introduced:

- returned the immediately following spring term
- continued through spring and fall terms
- returned the subsequent fall term
- any financial aid*
- completed the FAFSA*
- any AP credits*
- any remediation*

*added during the final analysis stage to provide further context for appropriate interpretation

After incorporating changes in the dataset to increase the precision of the algorithm, the Weka software using the J48 decision tree classifier was able to achieve an 86.29% accuracy rate on student success predictions for the 2020 instances in the fall

2001 student cohort. The decision tree J48 produced utilizing the training data is very large - 272 branches with 227 leaves. As explained in the *Introduction to Data Mining* text book (Tan et al., 2006) by splitting the branches so many times, J48 may be overfitting the solution specifically to the training set data and may yield a lesser level of accuracy when applied to future datasets. Typically data mining software is invoked for processing extensive amounts of data with a large number of instances. The guiding principle behind data mining is that enormous amounts of data provided for analysis afford the data mining algorithm to learn which attributes are meaningless in predicting the outcome allowing the algorithm to prune those branches from the tree. The result is a smaller decision tree with greater prediction accuracy.

Therefore in this case it is believed that the enormity of this decision tree is due to the fact that the dataset itself was quite small – only 2020 instances; forcing J48 to split the tree into multiple branches in order to classify each instance. (See Appendix A for the results of the application of the J48 algorithm in Weka.) Methods for increasing the accuracy of decisions trees like boosting (Roe, Yang, Zhu, Liu, Stancu, & McGreagor 2004) or windowing (Long, Griffith, Selker, & D'Agosino, (1993) may provide avenues for future research.

A Simplified Version of the Resulting J48 Decision Tree

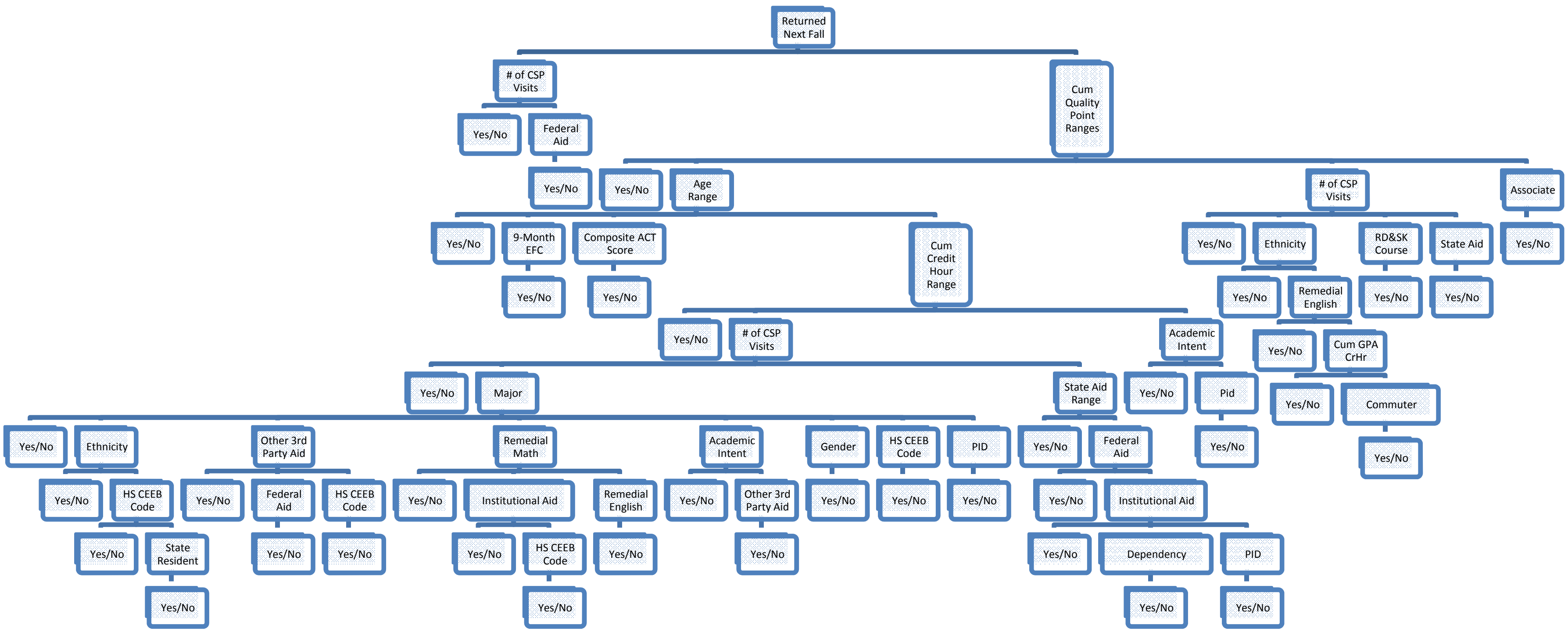

Figure 42

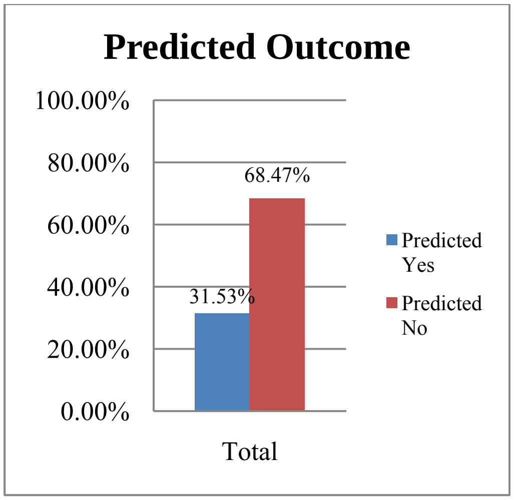

Figure 43

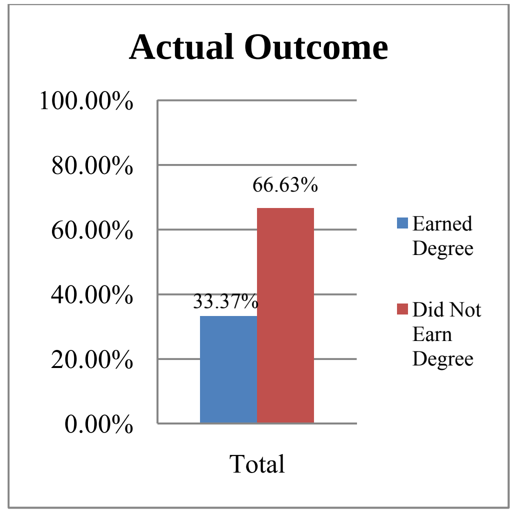

Figure 44

J48 predicted that a little more than 31% of the students in the cohort would earn a bachelor degree. In fact just over 33% actually successfully completed their degree requirements and earned a bachelor degree within six years of initial entrance.

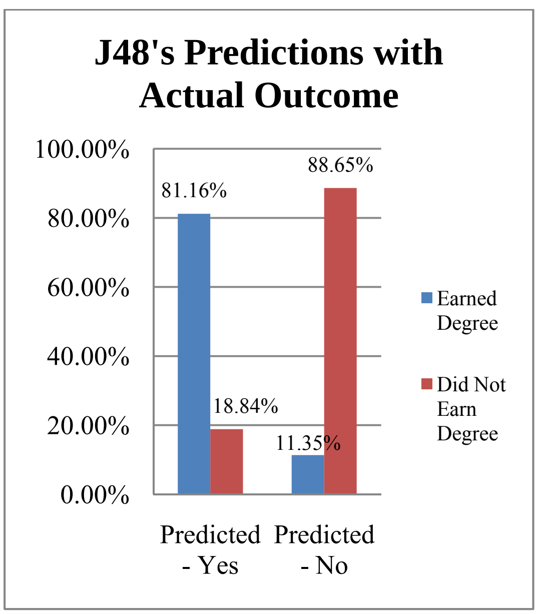

Figure 45

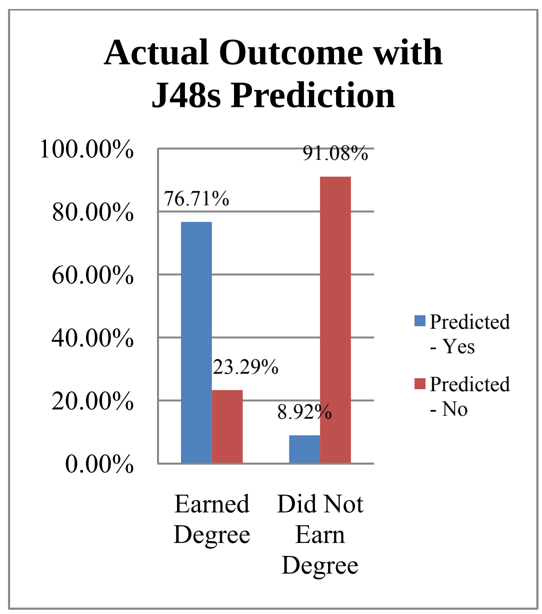

Figure 46

Of the 31.53% of the cohort predicted to earn a degree 81.16% did. Of the 68.47% predicted to not earn a degree 88.65% did not. There were 1,383 students predicted not to earn a degree in comparison to 637 students predicted to earn a degree.

More instances in the Predicted – No category provided J48 with enough examples to accurately predict over 89% of the final outcomes. This result in comparison to the 81% accuracy of the Predicted – Yes category provides an illustration of how the data mining algorithm obtains a higher level of accuracy with a greater number of instances.

Of the 674 students actually earning a degree, J48 only correctly predicted 76.71% of the outcomes. Of the 1,346 students not earning a degree, J48 correctly predicted 91.08% of the outcomes. These results further support the belief that the more instances available for analysis the greater the accuracy of the resulting decision tree. The chart of Students Earning a Bachelor Degree predicted versus actual outcomes follows as well as charts depicting predicted and actual percentages for each attribute.

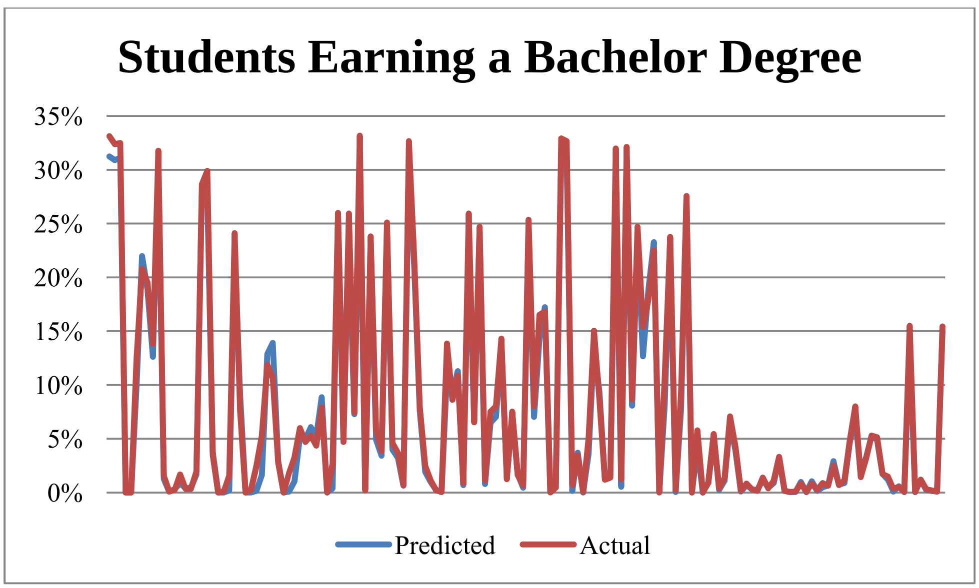

Figure 47

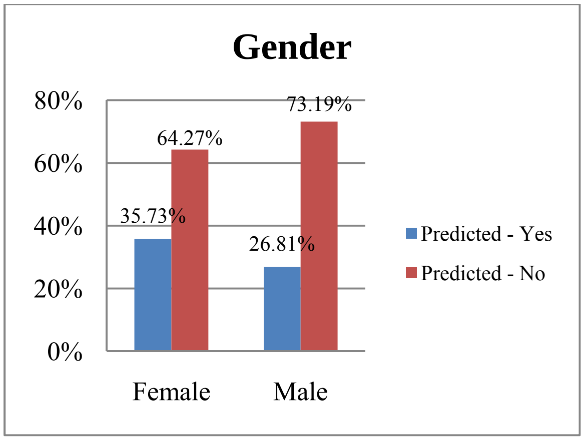

Figure 48

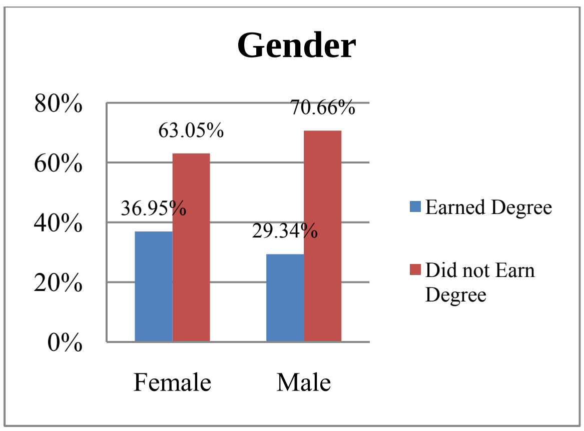

Figure 49

J48 performed slightly better predicting which female students would or would not graduate than it did for the male students.

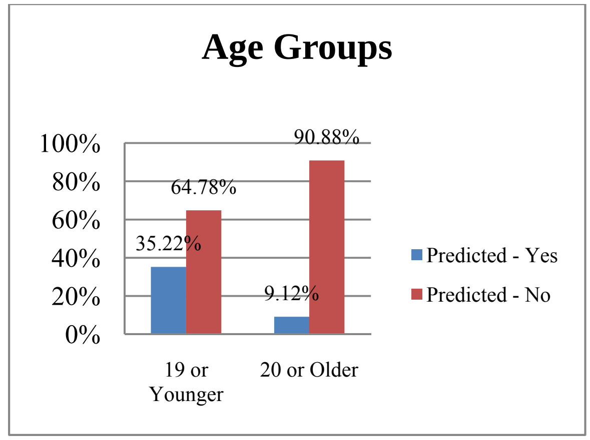

Figure 50

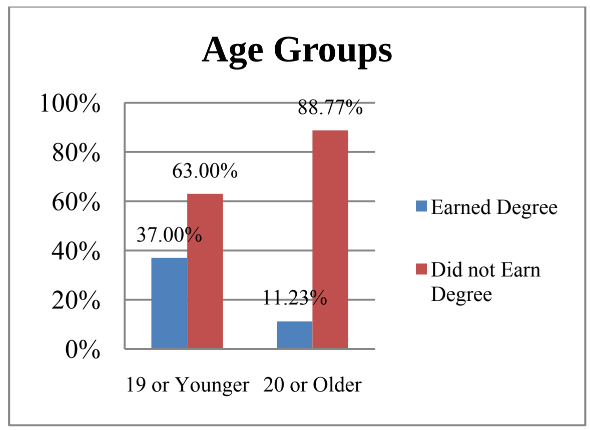

Figure 51

For both sub-categories, J48's predictions were off by about 2 percentage points.

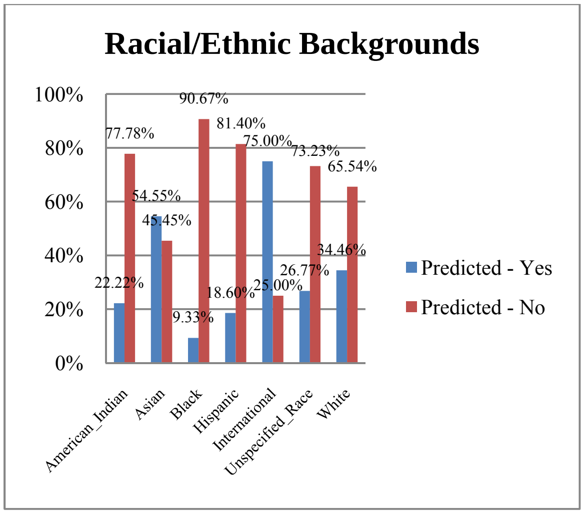

Figure 52

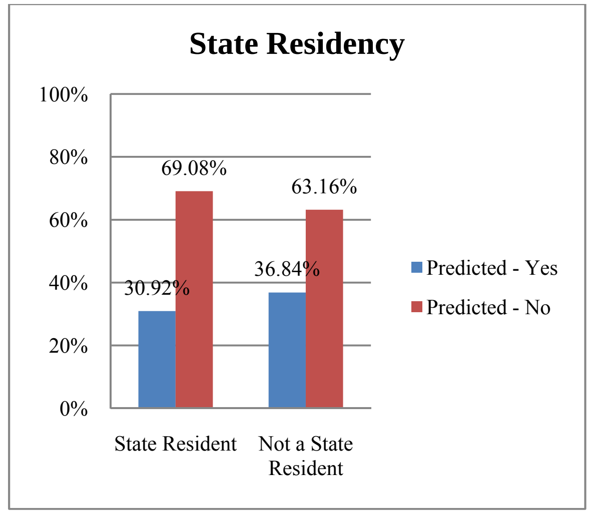

Figure 54

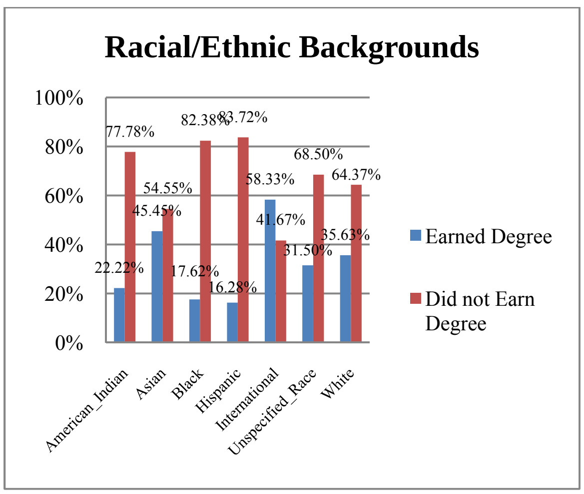

Figure 53

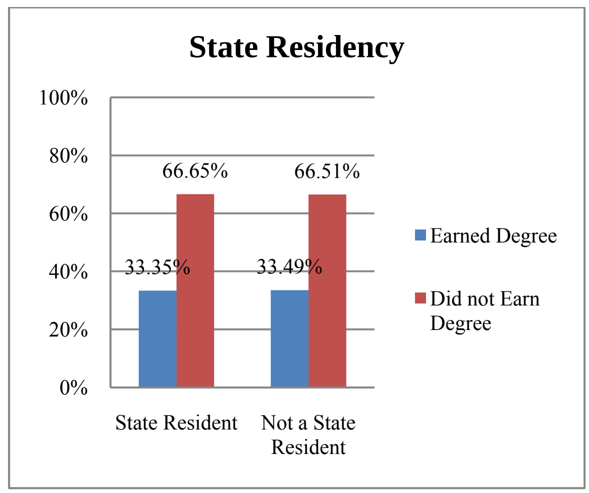

Figure 55

For American Indian, J48's prediction was exactly correct. For Black, International, and White, J48 under predicted the percentages earning degrees. For the remaining backgrounds, J48 over predicted the outcomes – most notably Asian, whose outcome was the exact opposite of the prediction.

Just about 3% more state residents and 3% less non-state residents earned a bachelor degree within six years of initial entrance.

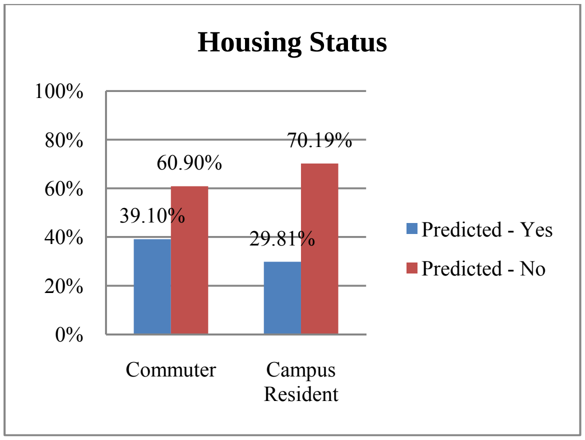

Figure 56

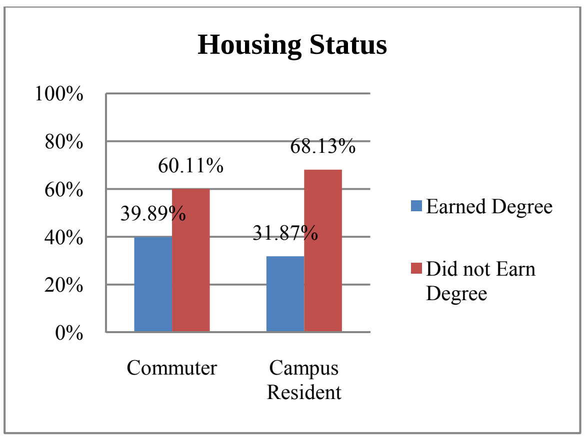

Figure 57

For commuters the J48 prediction was accurate within one percentage point. For campus residents it under predicted those earning a degree by approximately two percentage points.

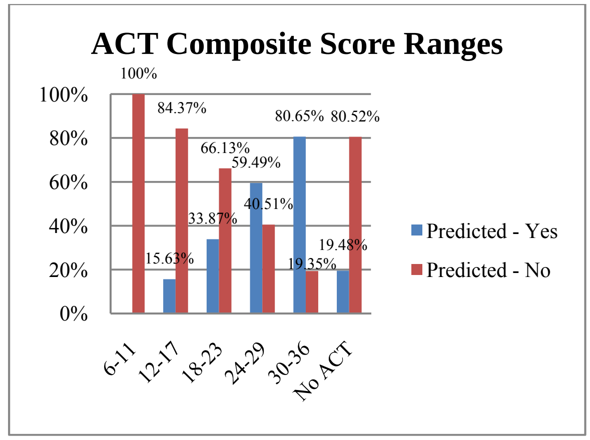

Figure 58

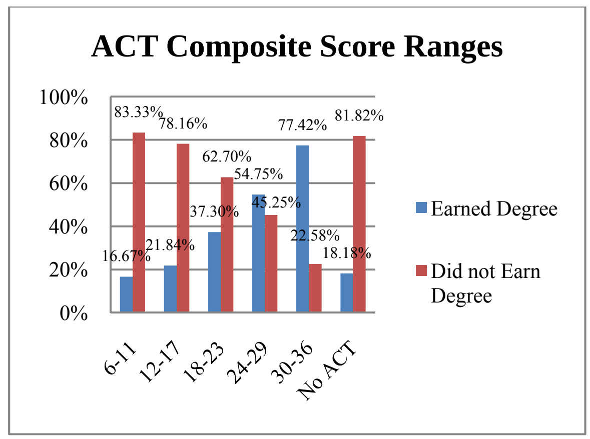

Figure 59

More students actually earned a degree than were predicted for scores ranging from 6 to 23. The unexpected lower graduation rate of those with a score of 24 or higher may due to students transferring out to another institution in pursuit of a program not offered at this institution. Unfortunately that information was not available at the time of this study.

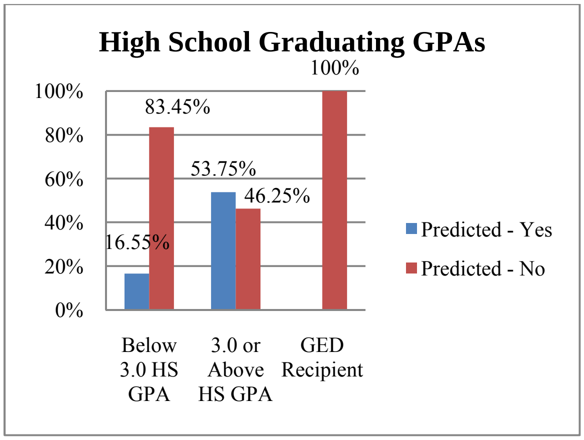

Figure 60

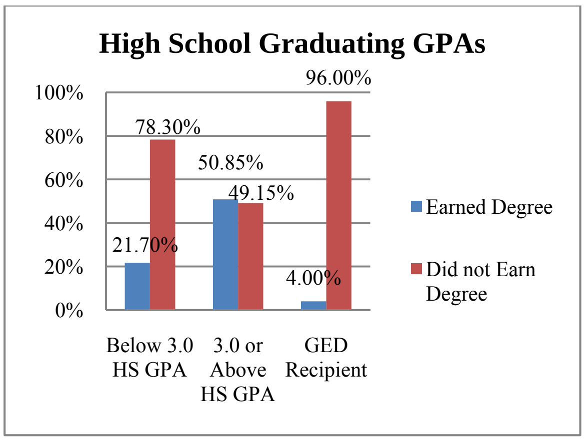

Figure 61

Slightly more than 5% of students with HS GPAs below 3.0 (≈ 60% of the cohort) earned a degree than were predicted. Where about 3% less of those at 3.0 or above earned a degree. This is the first sub-category identified where the absence of transfer out information surfaces as a potential critical factor for predictions.

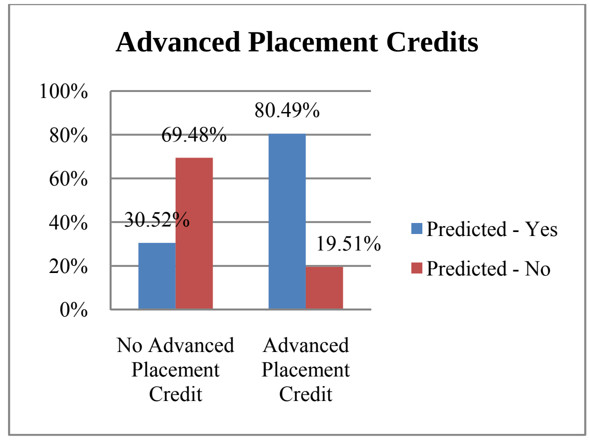

Figure 62

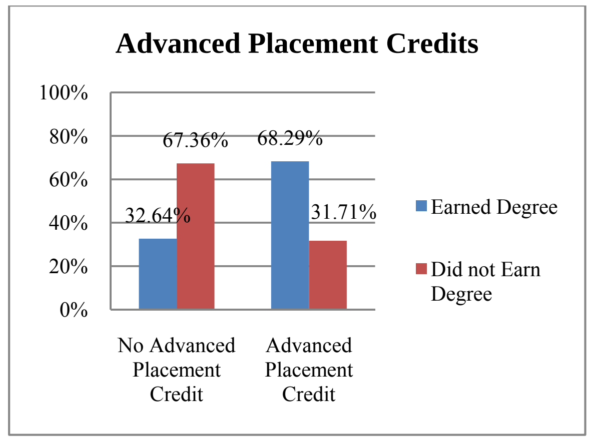

Figure 63

The predictions for those with advanced placement credits were very off - as 12% less of the students with advanced placement credits (≈2% of the cohort) earned a degree than what were predicted. Those with no advanced placement credits were within about 2% of the predicted percentage.

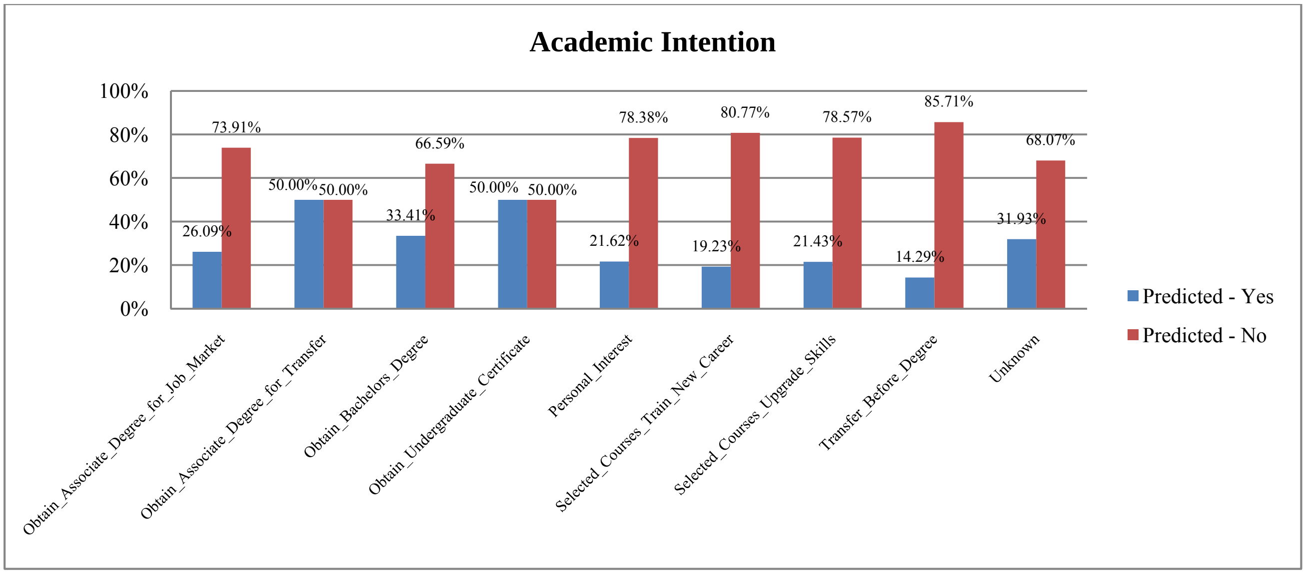

Figure 64

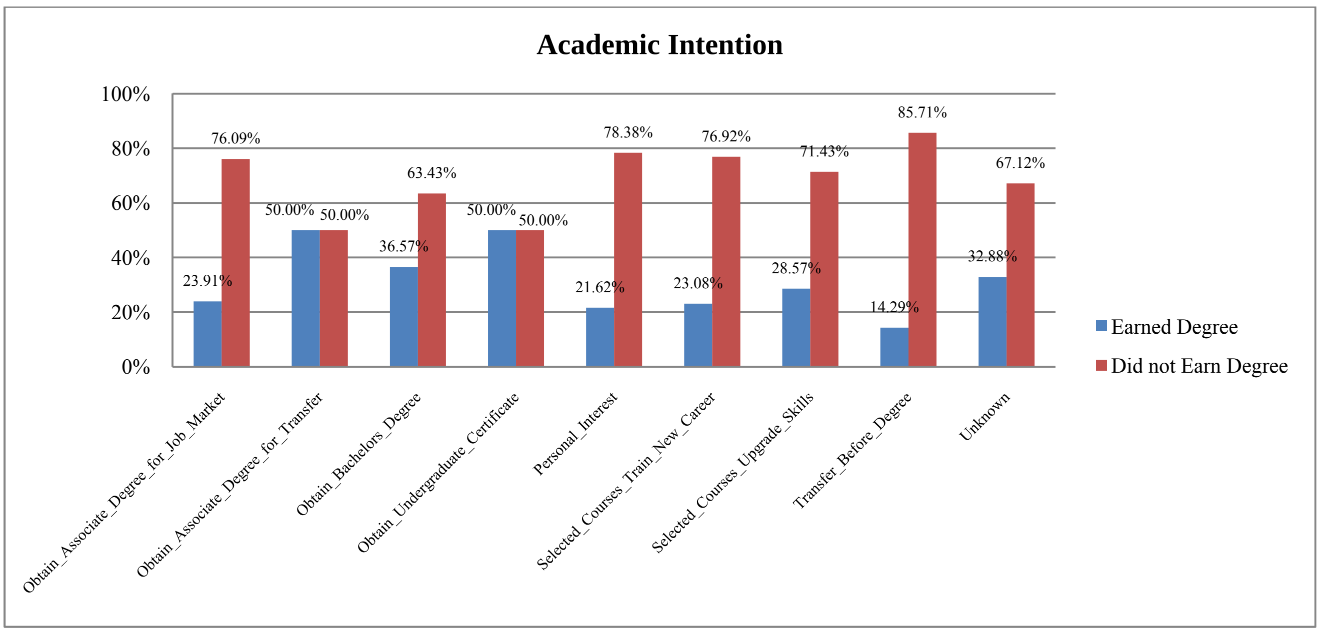

Figure 65

For the two largest subcategories (Unknown – n = 949 and Obtain_Bachelors_Degree – n = 856) J48's predictions were off by 2% and 3% respectively. The algorithm performed quite well for the subcategories of Obtain_Associate_Degree_for_Transfer, Obtain_Undergraduate_Certificate, Personal_Interest, and Transfer_Before_Degree.

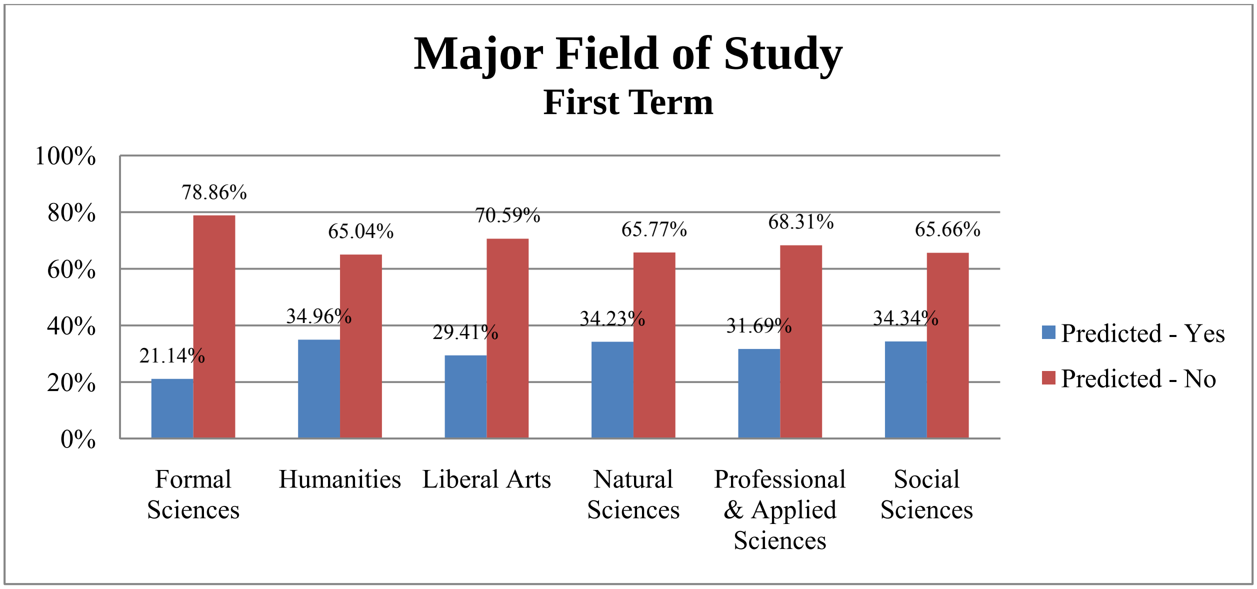

Figure 66

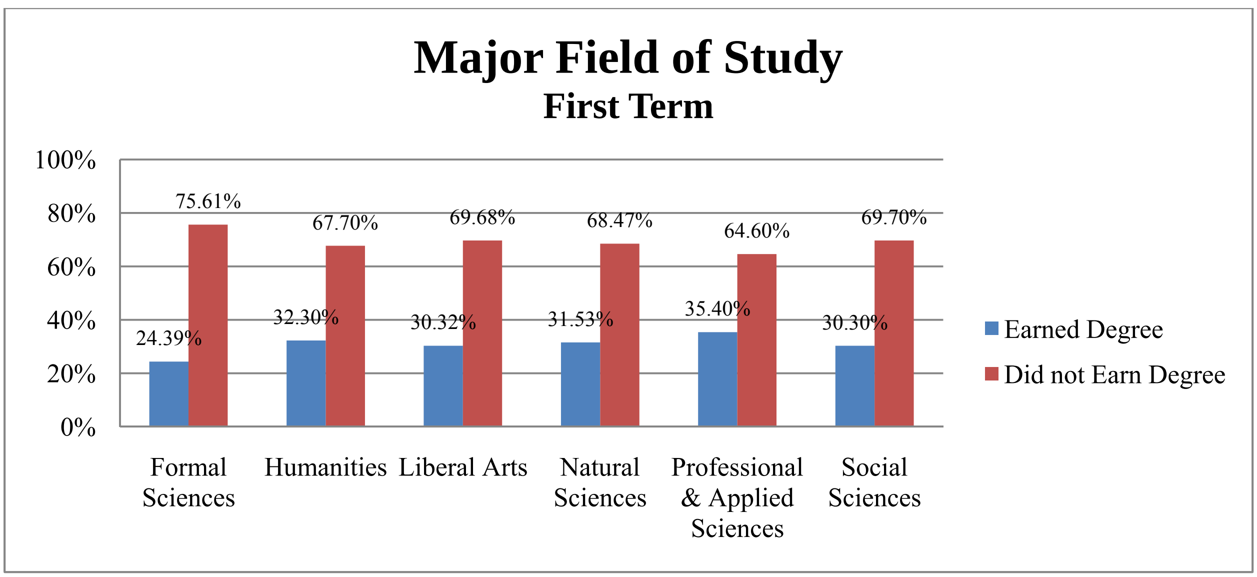

Figure 67

The original amount of student majors were downsized to the 6 major groups listed here in Figure 67 for ease in visual interpretation. J48s prediction for the largest sub-category, Professional & Applied Sciences, is lower by slightly less than 4 percentage points.

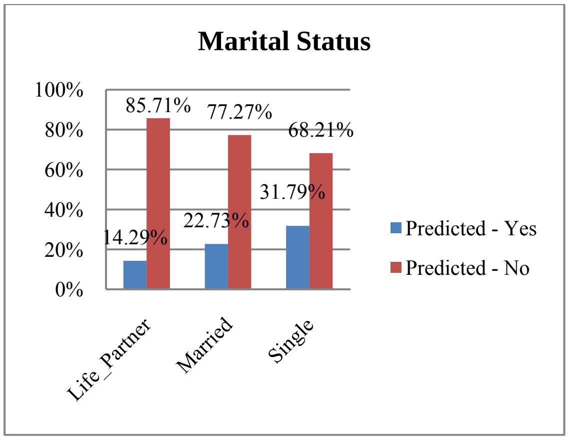

Figure 68

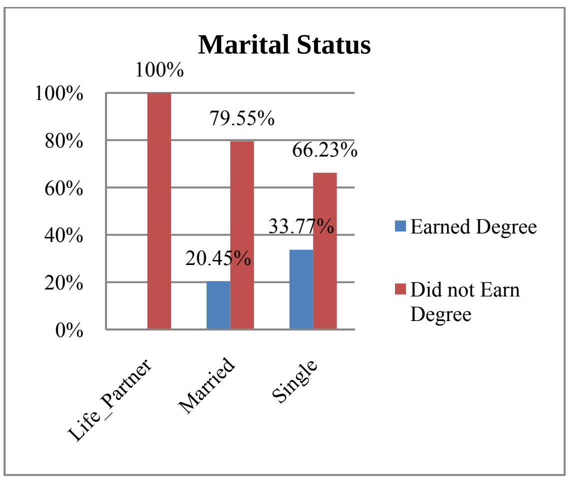

Figure 69

With the exception of the Life_Partner sub-category (n = 7) where no students earned a degree, the J48 predictions were off by about 2%.

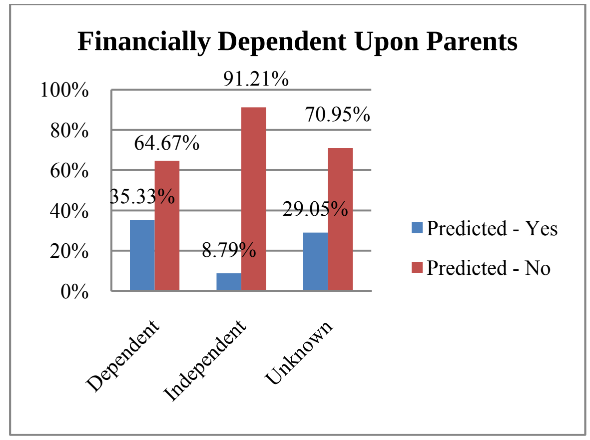

Figure 70

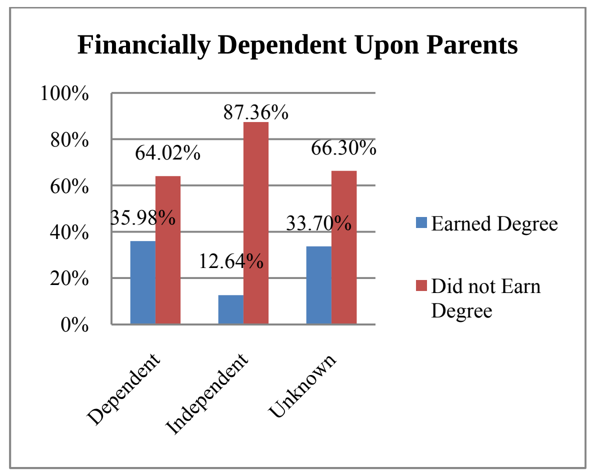

Figure 71

For the students that were financially dependent upon their parents (≈69% of the cohort), J48 came within 0.6% of the actual outcome. The Earned Degree predictions for the remaining subcategories, each with fewer instances, were lower than the actual outcomes by approximately 4%.

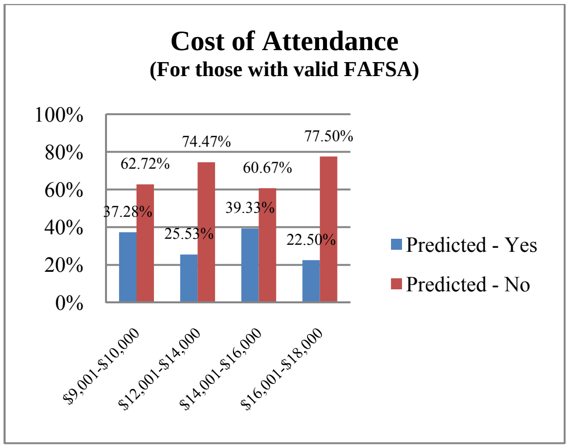

Figure 72

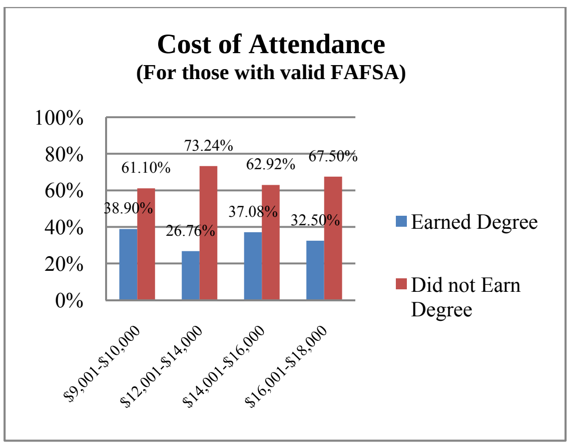

Figure 73

J48's predictions were between 1-2% different than the actual outcomes with the exceptions of $10,001-$12,000 and $16,001-$18,000, which were 5% and 10% different respectively.

Note: 505, or 25% of the cohort either did not complete or did not have a valid FAFSA required for determination of most financial awards.

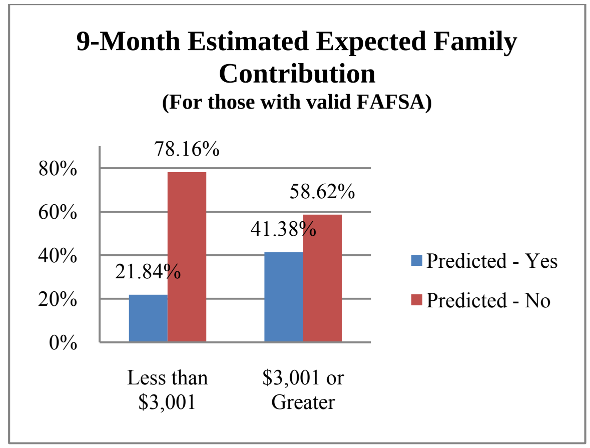

Figure 74

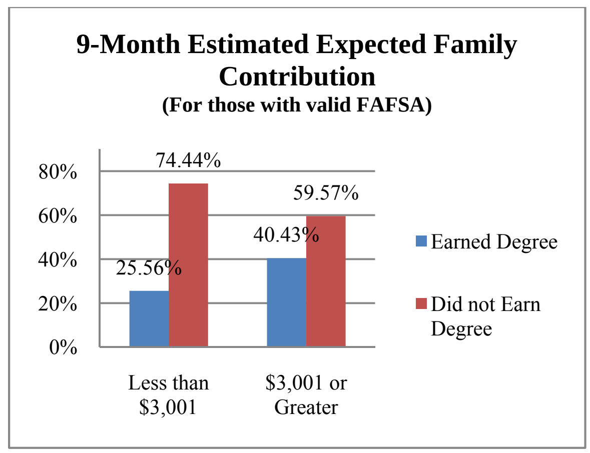

Figure 75

9-Month estimated expected family contribution predictions were fairly accurate for those students expected to pay $3,001 or more for their education than for those expected to pay less.

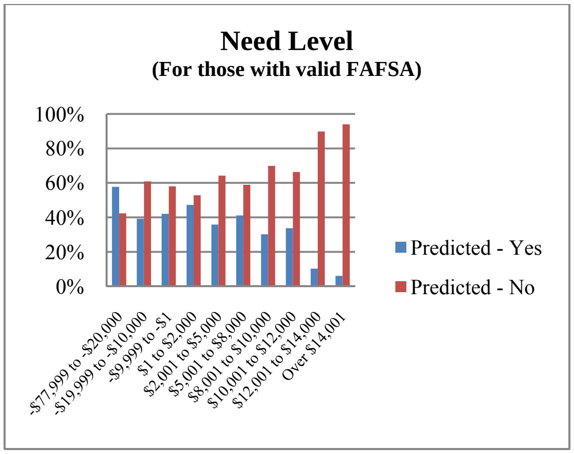

Figure 76

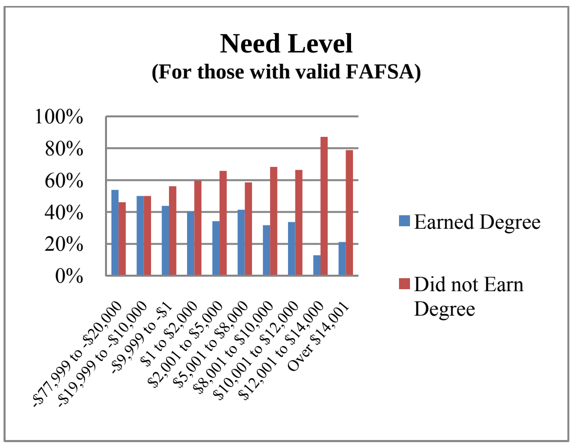

Figure 77

For the most part, the predicted and actual outcome charts are very similar. The findings for this category were consistent with common understanding that those who are more affluent have a greater tendency to graduate within the normal expected amount of time.

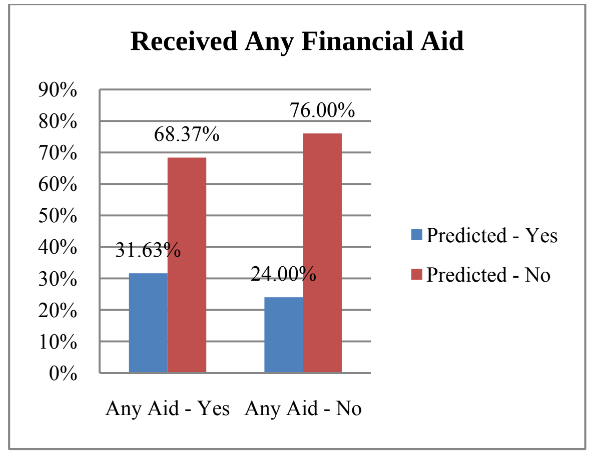

Figure 78

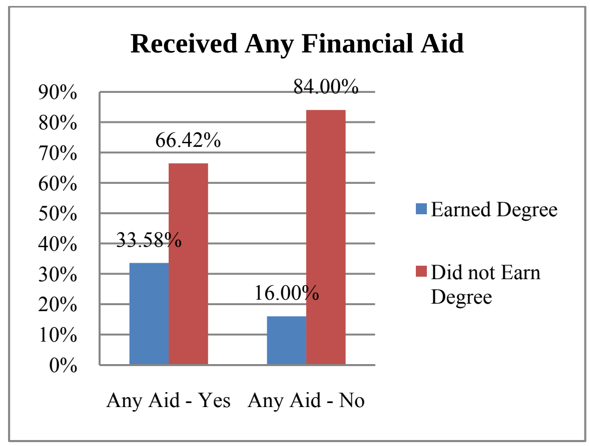

Figure 79

Consistent with sub-categories containing a large percentage of the cohort, the predictions for those students receiving any financial aid were closer to the mark than those receiving no financial aid.

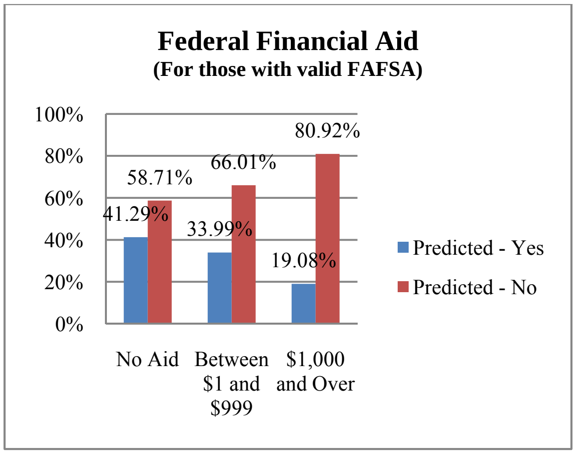

Figure 80

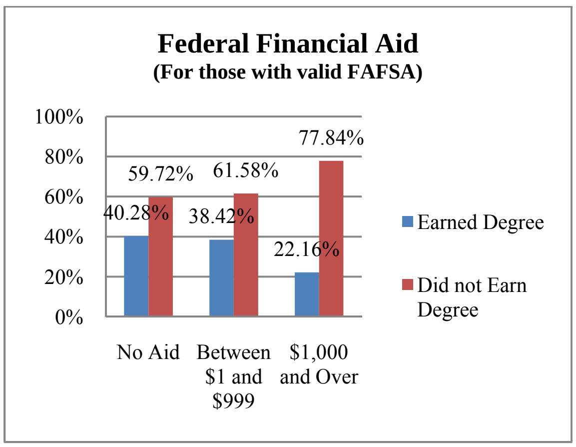

Figure 81

The accuracy of the predictions ranged between 1% and 4.5% different than the actual outcomes.

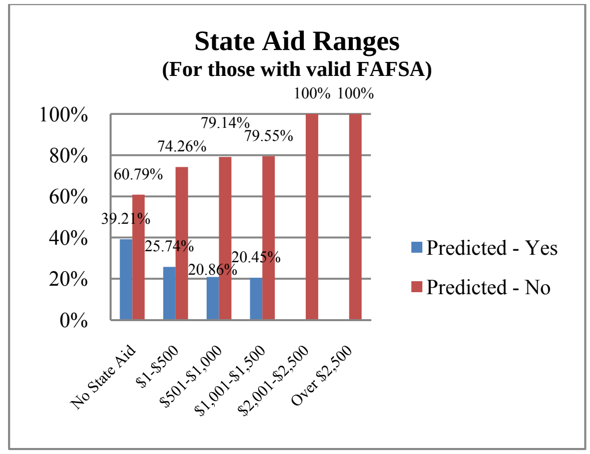

Figure 82

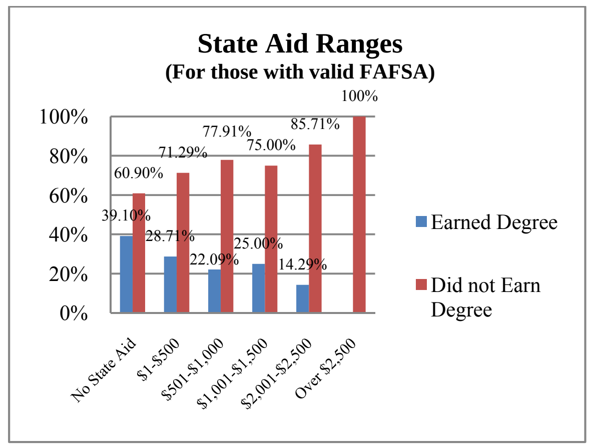

Figure 83

In all sub-categories a greater percentage of students earned degrees than was predicted with the exception of the one student that fell in the Over $2,500 range.

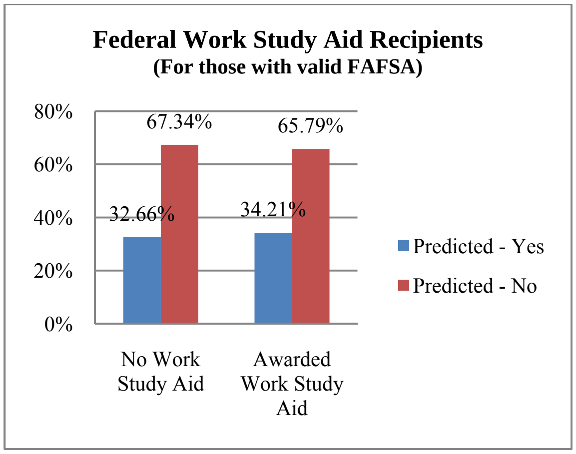

Figure 84

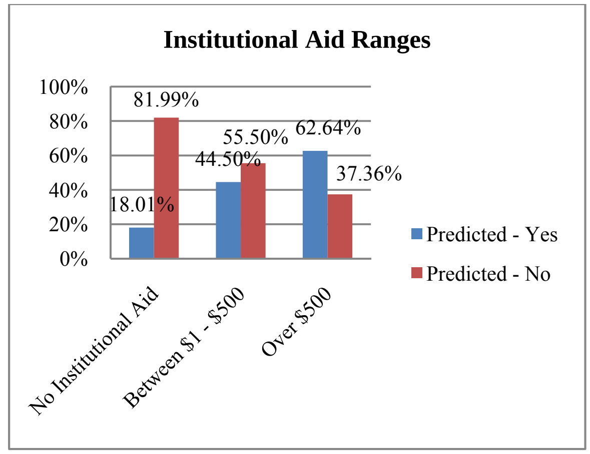

Figure 86

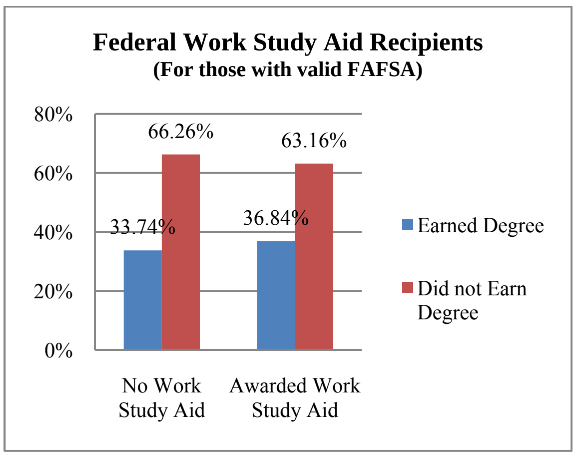

Figure 85

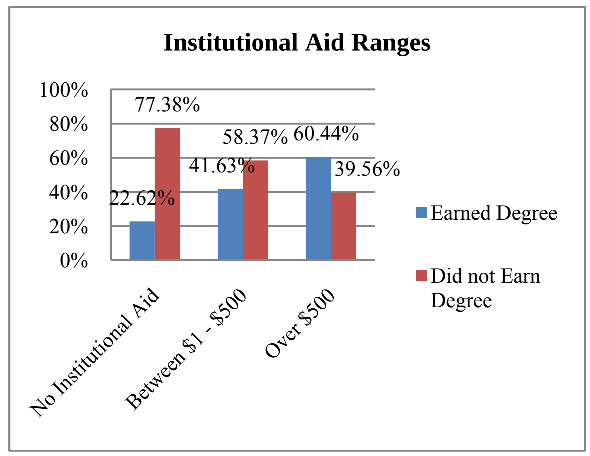

Figure 87

Slightly more students in both sub-categories earned degrees than were predicted.

Over 4.5% more students in the no institutional aid range earned degree than were predicted. In comparison between 2.2 and 2.8% fewer students earned degrees in the remaining ranges.

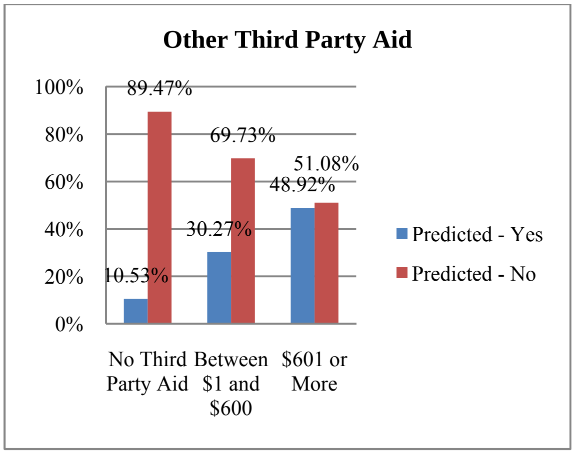

Figure 88

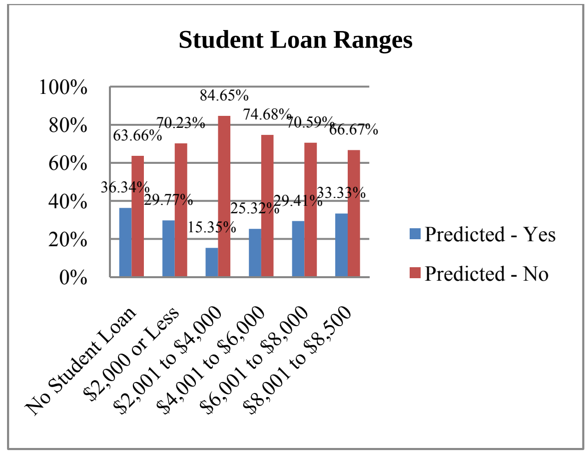

Figure 90

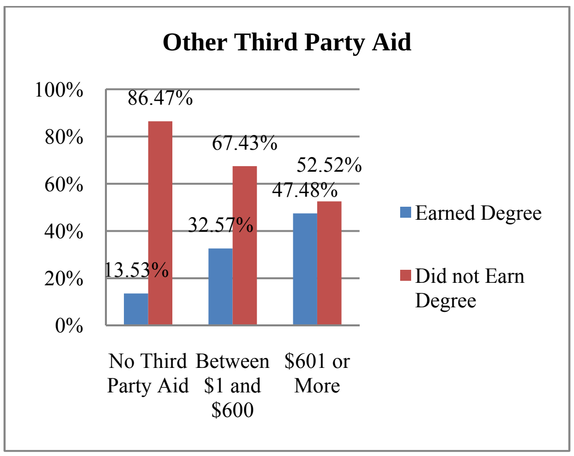

Figure 89

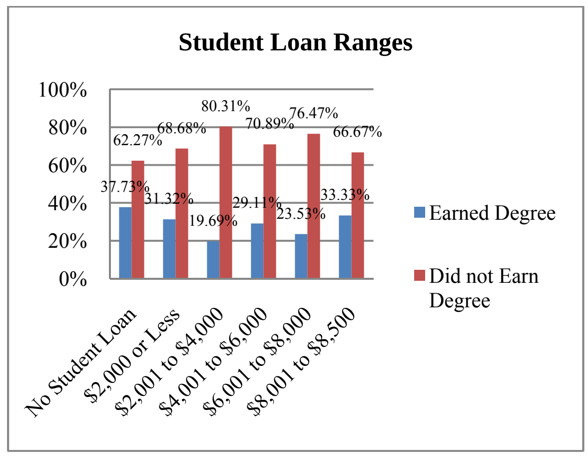

Figure 91

Other third party aid prediction percentages differed by 3%, 2.3%, and 1.44% respectively.

Student loan ranges prediction percentages differed between 1.39% and 5.88% with the exception of the $8,001 to $8,500 group which was exactly precise.

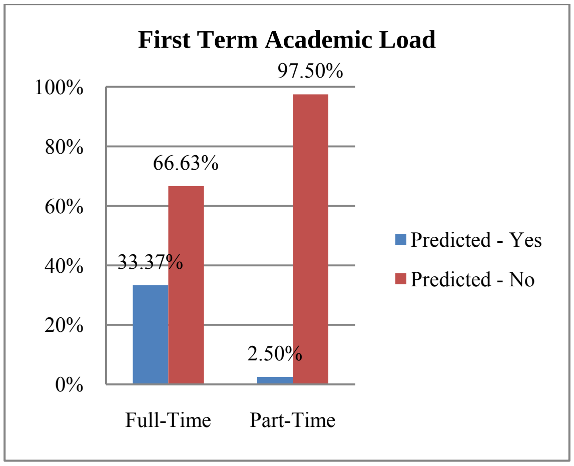

Figure 92

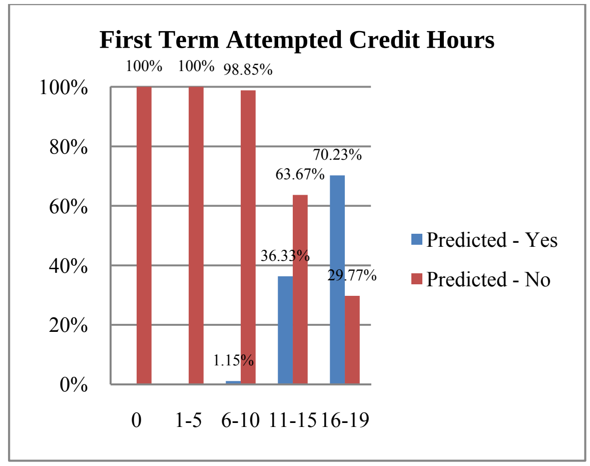

Figure 94

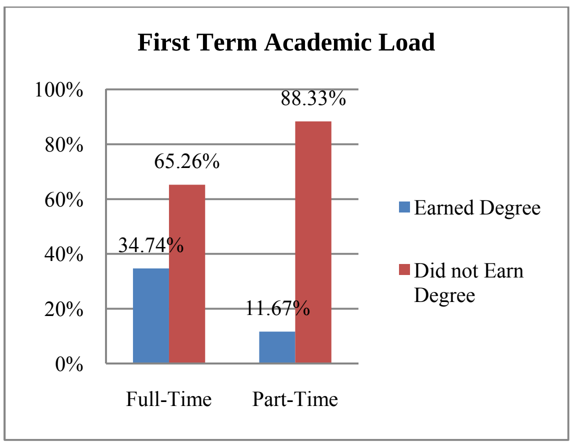

Figure 93

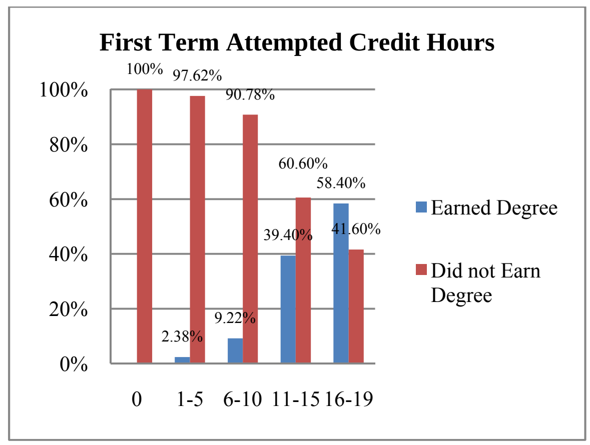

Figure 95

J48 performed well in predicting the percentage of full-time students who would later earn their degree but showed signs of difficulty with the part-time students' prediction. Again it may be worth stating that 1,900 of 2,020 students in the cohort attended full-time their first term.

In regard to the first term attempted credit hours sub-category, with the exception of the group of students attempting earn no credit hours their first term, J48 did not perform as well as expected. Once again this may be an affect of the missing transfer out information.

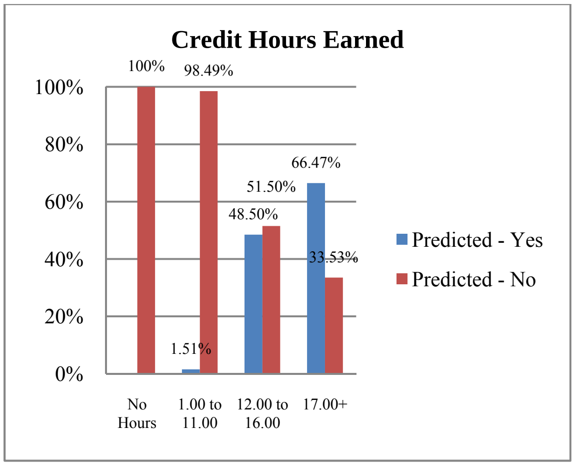

Figure 96

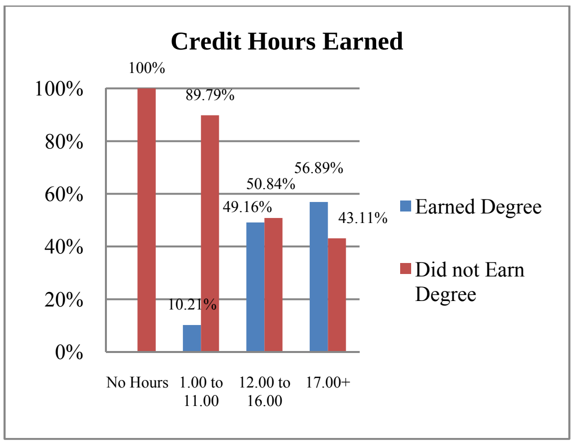

Figure 97

J48 performed well in predicting the percentage of students earning between 12 and 16 credit hours that would later earn their degree but showed a problem with the 17.00+ credit hour group's prediction. This may well be another attribute affected by the missing transfer out information.

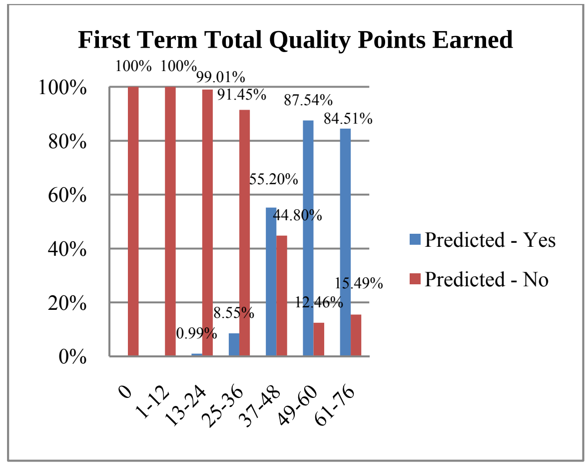

Figure 98

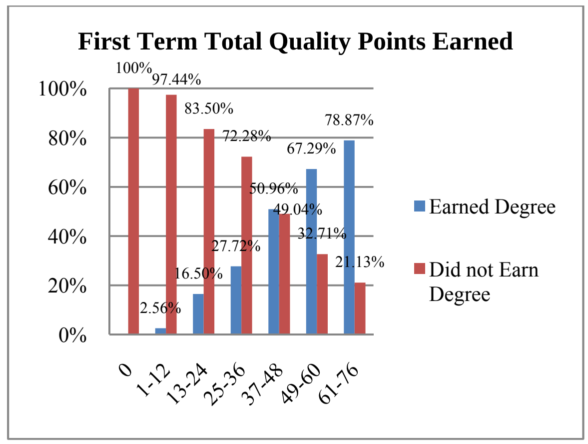

Figure 99

J48's predictions for students ending their first term with 13-24, 25-36, 49-60, and 61-76 quality points were significantly different than the actual values. This attribute also falls on the list of casualties with regard to transfer out information.

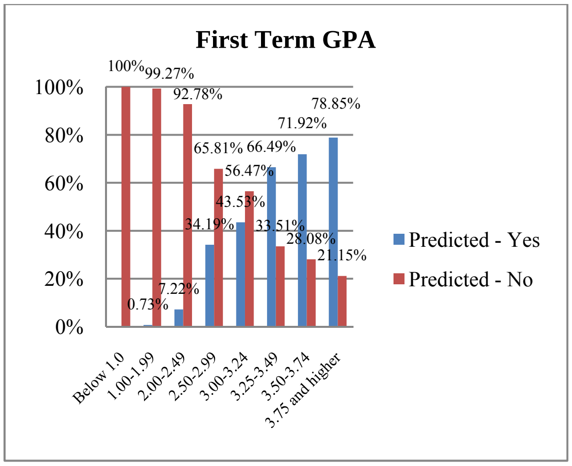

Figure 100

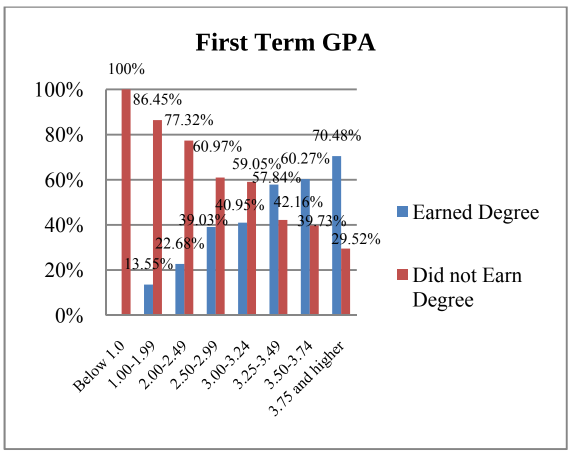

Figure 101

Once more the students intuitively expected to earn degrees within the six-year time period in actuality graduated in lesser percentages than predicted. Furthermore those students with lower first term GPAs graduated at a significantly higher rate. This latter issue is an area worthy of follow-up investigation.

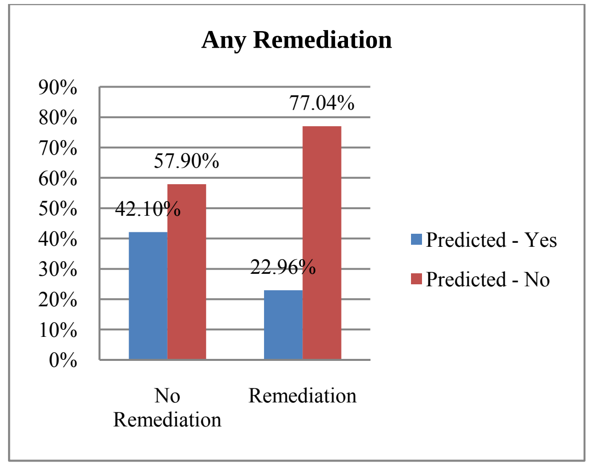

Figure 102

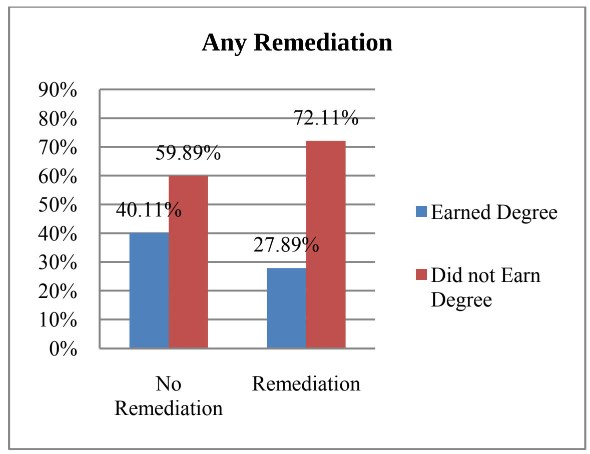

Figure 103

J48's prediction for those students with no remediation and for those with remediation was 2% and 5% different than the actual values respectively. The counter-intuitive result in regard to those engaging in remediation is also an item that should be further explored. Perhaps the introduction of an additional dataset would help increase the accuracy.

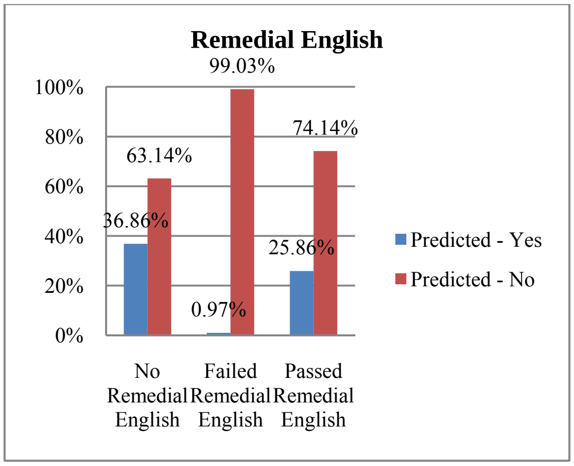

Figure 104

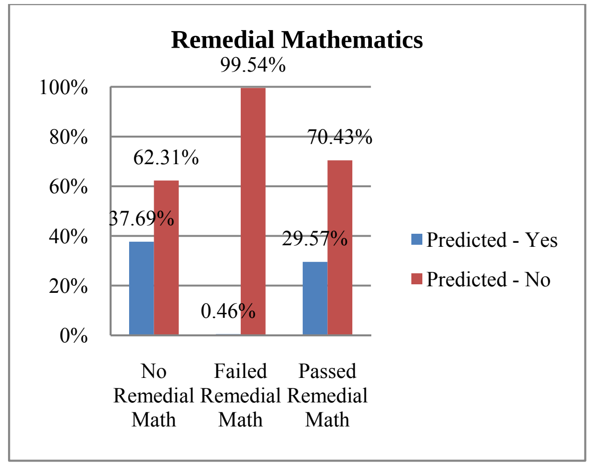

Figure 106

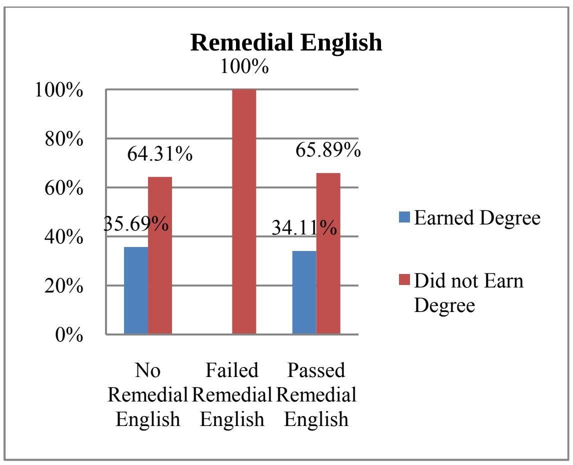

Figure 105

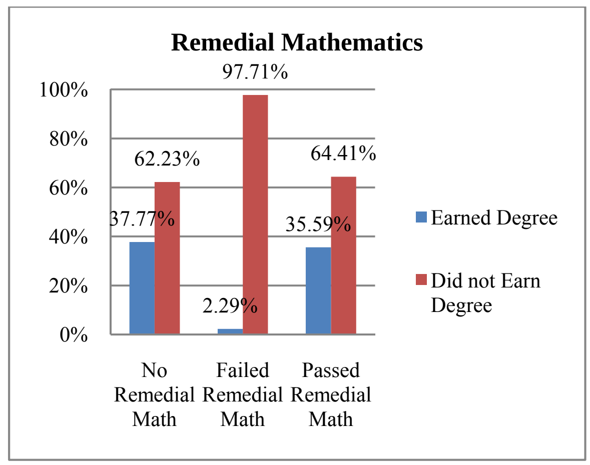

Figure 107

J48 did well in predicting the outcomes of those students engaging in no remedial English and those failing remedial English. However it did not do well predicting the outcomes of those passing remedial English.

Similarly J48's performed well in predicting the percent of the cohort that would earn a degree for those students with no remedial mathematics and for those failing remedial mathematics. However, once again its prediction for those passing the remedial coursework was off by 5%.

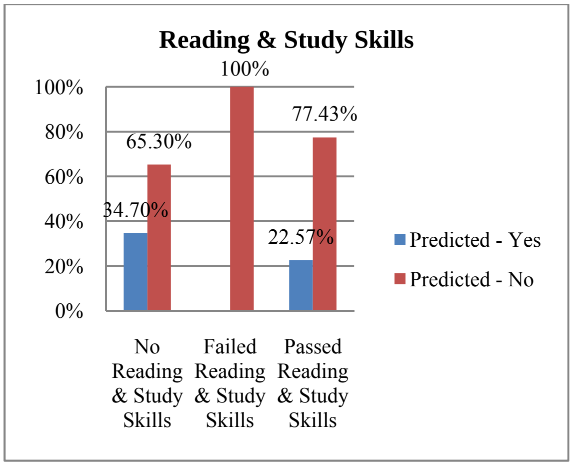

Figure 108

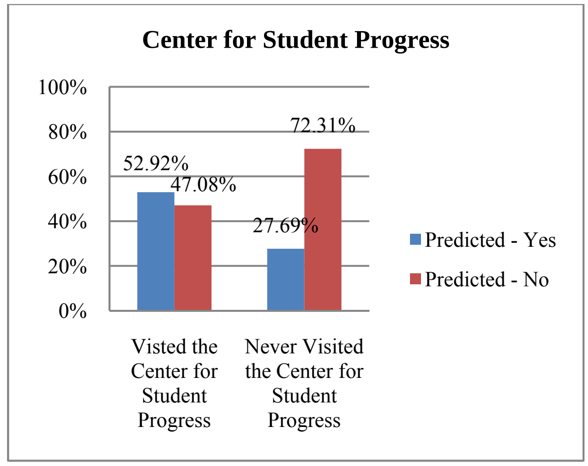

Figure 110

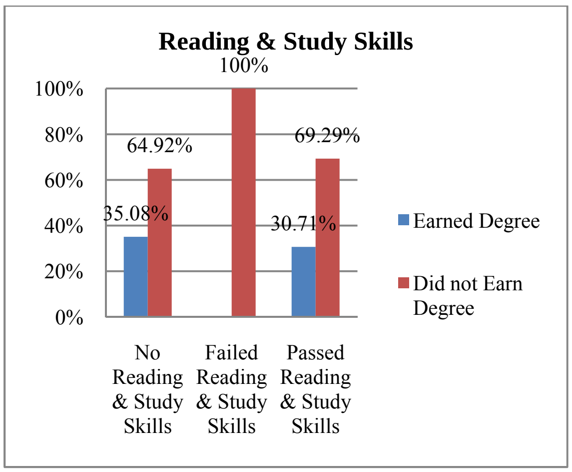

Figure 109

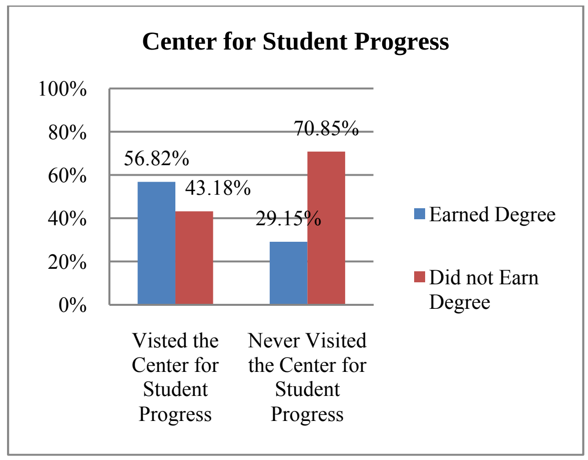

Figure 111

As was the case with the other remedial areas, J48 performed well on its predictions for those students not engaging in Reading & Study Skills coursework and those failing the coursework. Yet its prediction for those passing the coursework was significantly different. In this case J48 was off by over 8 percentage points.

The predictions for those students never visiting the Center for Student Progress during their first term were very close to the actual percentages yet the predicted percentages were off by 6 percentage points for those that did visit the Center.

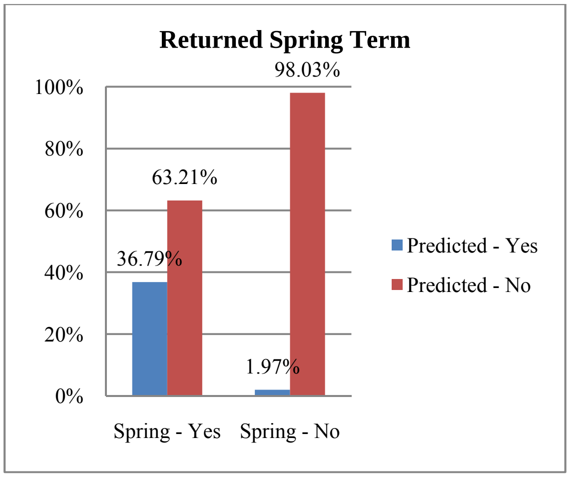

Figure 112

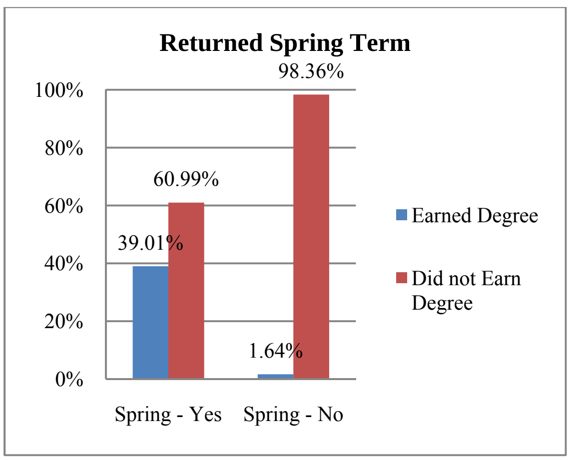

Figure 113

J48 did a stellar job predicting the percentage of students not enrolling the immediately following spring term that would go on to earn a degree and was off by a little more than 2% for your enrolling that spring term.

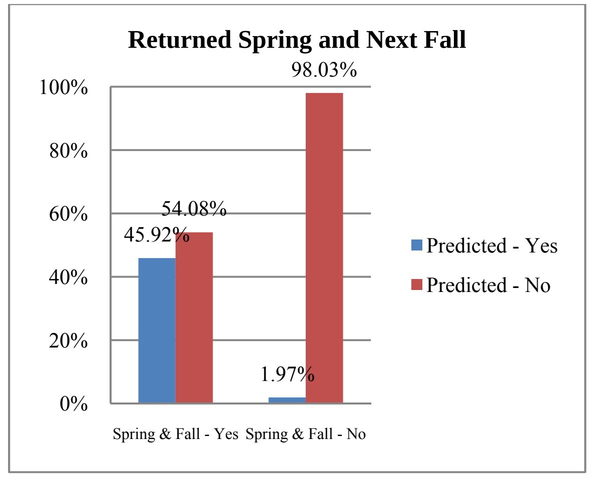

Figure 114

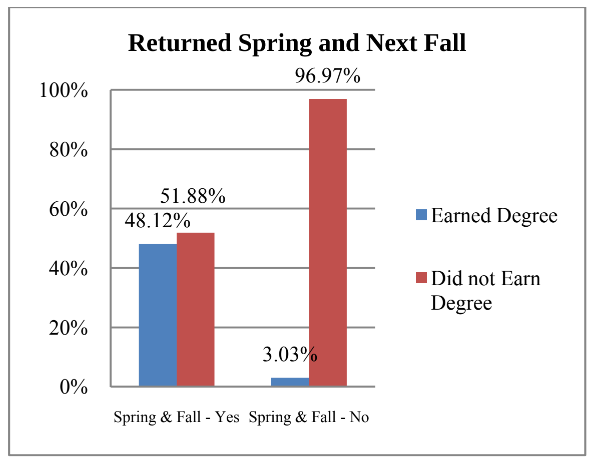

Figure 115

Likewise, J48 performed well in predicting the percentages of students earning a degree with six years for both those students consecutively enrolled fall, spring and the following fall terms and for those not consecutively enrolled.

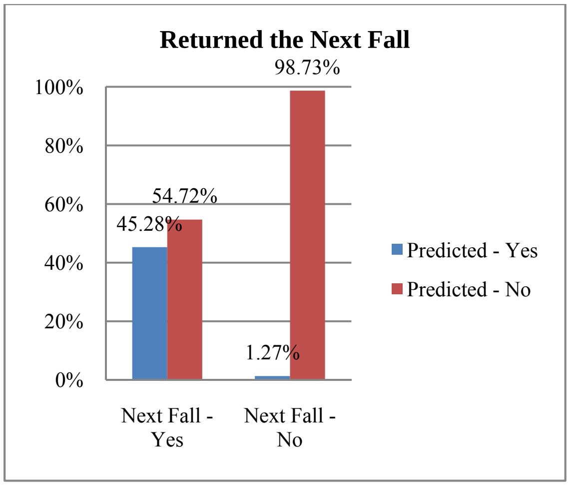

Figure 116

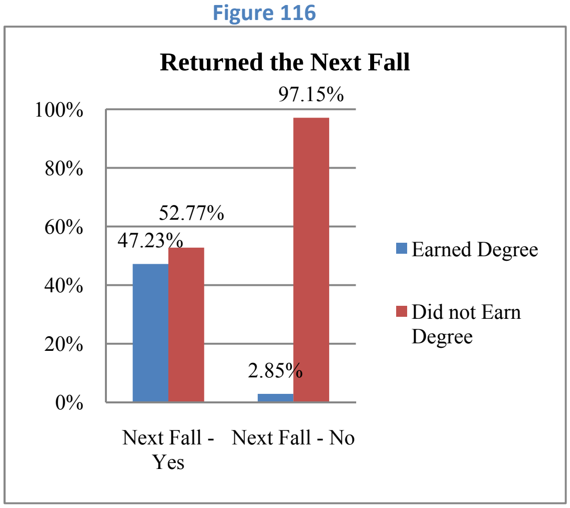

Figure 117

J48 also did well predicting the percentage of students enrolling and not enrolling the fall term for the second year who would go on to earn a degrees – with predictions within a little more than 2% of the actual amounts.

The table of statistics detailing the cohort distributions among the various data elements including the predicted and actual values can be found in Table 1.

4. Discussion

With few exceptions, the J48 predictions were very close to the actual outcomes experienced by the students in the cohort. Noticeable differences in the high school grade point average, first term grade point average, first term quality points earned, and first term credit hours may be explained by missing data – in particular data indicating whether or not students transferred out to another institution. Introduction of this missing data may support the statements of Bowen, Chingos & McPherson (2009) with regard to the predictive strength of high school grade point average of student degree attainment. Additionally the 86.29% accuracy rate provides a strong support for future utilization of data mining on student data for success prediction. In general the ease of use of the data mining software combined with the high rate of accuracy, make this method of prediction highly desirable. By allowing the software to perform the difficult computations, the researcher was able to focus on those elements of the process most familiar - selecting appropriate student data attributes and preparing the dataset for processing.

5. Conclusions

Data mining software provides a relatively easy way to quickly identify previously unknown relationships among the attributes within a student cohort dataset. These relationships may provide policy analysts with the necessary information for supporting operational changes in order to enhance a higher education institution's graduation rate. Thusly by increasing the number of college credentialed citizens within the state, a higher education institution will provide a desirable educated workforce to entice new business and industry to the region. As the economy has reached a significant low point and unemployment rates continue to climb though at a decreasing rate, the use of such predictions can have a dramatic affect on the institution's ability to provide outstanding service to the state as well as the students who enter its domain. Further the resulting changes may reduce the time to degree and subsequently the cost of higher education to the student while increasing the institution's subsidy allocation from the state.

5.1 Recommendations

Follow-up work is indicated for this study. The decision tree model produced with this dataset should be applied to future datasets to gauge and/or increase its accuracy and in all likelihood refine the decision tree model itself for subsequent use. Additionally, further analysis of the cohort dataset including the J48 predictions is necessary for developing profiles of each student category for communicating to high school guidance counselors, for effective institutional recruiting efforts, for academic advising and for identifying appropriate intervention. Moreover the introduction of transfer out data and possibly the expansion of the dataset to include teaching faculty attributes should be strongly considered in order to provide a stronger predictive result from the algorithm. Finally investigation of the counter-intuitive results with regard to remedial coursework should be explored.

REFERENCES

Bailey, B. (2006). Let the data talk: developing models to explain IPEDS graduation rates. New Directions for Institutional Research n 131 p 101-115 fall 2006.

Bowen, W., Chingos, M., & McPherson, M. (2009, October 16). Crossing the finish line completing at America's public universities. Presentation to the Ohio Association for Institutional Research and Planning fall conference 2009. Columbus, Ohio: The Ohio State University.

Center for Student Progress. Mission statement. Retrieved October 17, 2009, from http://www.ysu.edu/csp/historymission.shtml.

Executive Office of the President, Council of Economic Advisors. (2009, July). Preparing the workers of today for the jobs of tomorrow. Retrieved September 20, 2009, from http://www.whitehouse.gov/administration/eop/cea/Jobs-of-the-Future/.

Herzog, S. (2006, Fall). Estimating student retention and degree-completion time: decision trees and neural networks vis-à-vis regression. New Directions for Institutional Research, no 131.

Jacobs, L. & Hyman, J. (2009, July 1). 7 reasons why college is so expensive. U.S. News & World Report, Professors Guide. Retrieved July 9, 2009, from http://www.usnews.com/blogs/professors-guide.

Jaschik, S. (2008, October 9). Falling behind. Inside Higher Ed. Retrieved October 11, 2008, from *http://insidehighered.com/news/2008/10/09/minority*.

Long, W., Griffith, J., Selker, H., & D'Agostino, R. (1993). A comparison of logistic regression to decision-tree induction in a medical domain" from Computers in

Biomedical Research, 26: 74-97, 1993.

Moltz, D. (2009, April 30). Adopting performance-based funding. Inside Higher Education. Retrieved May 1, 2009, from http://www.insidehighered.com/layout/set/print/news/2009/04/30/ohio 5/1/2009.

Ohio Board of Regents. (2009, November). Higher Education Information System, Mission and Purpose. Retrieved November 15, 2009, from http://regents.ohio.gov/hei/.

Perry, N. (2008, October 6, 2008). With no way out of trouble, more students likely to default. The Seattle Times. Retrieved October 11, 2008, from http://seattletimes.nwsource.com/html/localnews/2008231488_loandaytwo06m.html.

Roe, B., Yang, H., Zhu, J., Liu, Y., Stancu, I., & McGregor, G. (2005). Boosted decision trees as an alternative to artificial neural networks for particle identification. Nuclear Instruments and Methods in Physics Research A 543 (2005) 577-584.

Swail, W. (2008, August 22). The bell curve under a different cover. The Educational Policy Institute's – Week in Review. Retrieved August 30, 2009, from http://www.educationalpolicy.org/pub/wir/080822.html.

Tan, P., Steinbach, M., & Kumar, V. (2006). Introduction to Data Mining. Pearson Education, Inc. Upper Saddle River, New Jersey.

The University of Waikato. (2009). Weka machine learning project. Retrieved October 17, 2009. from http://www.cs.waikato.ac.nz/~ml/index.html.

Witten, I. & Frank, E. Data mining practical machine learning tools and techniques (2nd ed.). Elsevier, Inc. Maryland Heights, Missouri.

Youngstown State University. Undergraduate Student Bulletin. Courses, Reading & Study Skills. Retrieved October 19, 2009, from http://www.ysu.edu/catalog/files/Courses%20251-401.pdf , p.384.

Appendix A

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: 2001 discret. no-yes no pids Pick ME csv-weka.filters.unsupervised.attribute.Remove-R37-40

Instances: 2020

Attributes: 36

PID

SP_Ret

AU_Next_Ret

HS_GPA_Range

Gender

Age_Range

Ethnicity

Academic_Intent

State_Resident

Cum_GPA_CrHr_Ranges

Cum_QPts_Ranges

Cum_GPA_Ranges

Cum_Credit_Hour_Range

Major

Commuter

Fed_Aid_Excl_Loans_Range

State_Aid_Range

Work_Study_Range

Student_Loan_Range

Institutional_Aid_Range

Other_3rd_Party_Aid_Range

Dependency

Cost_Of_Attendance_Range

9_Month_Expected_Family_Contribution

Need_Level_Range

Student_Marital_Status

Load

Comp_ACT_Ranges

HS_CEEB_Code

Any_AP

Associate_Ever

Bachelor_Degree

#_of_CSP_Visits

Remedial_English

Remedial_Math

R&SK

Test mode: evaluate on training data

=== Classifier model (full training set) ===

J48 pruned tree

AU_Next_Ret = NO

| #_of_CSP_Visits = None: No (618.0/10.0)

| #_of_CSP_Visits = 1. 1-5

| | Fed_Aid_Excl_Loans_Range = 9. 1750-1999: No (2.0)

| | Fed_Aid_Excl_Loans_Range = 1. No_Aid: Yes (7.0)

| | Fed_Aid_Excl_Loans_Range = 5. 750-999: Yes (1.0)

| | Fed_Aid_Excl_Loans_Range = 7. 1250-1499: Yes (0.0)

| | Fed_Aid_Excl_Loans_Range = 3. 200-499: Yes (0.0)

| | Fed_Aid_Excl_Loans_Range = 4. 500-749: No (1.0)

| | Fed_Aid_Excl_Loans_Range = 11. 2250-2499: No (1.0)

| | Fed_Aid_Excl_Loans_Range = 6. 1000-1249: Yes (0.0)

| | Fed_Aid_Excl_Loans_Range = 8. 1500-1749: Yes (0.0)

| | Fed_Aid_Excl_Loans_Range = 2. 1-199: Yes (0.0)

| | Fed_Aid_Excl_Loans_Range = 10. 2000-2249: Yes (0.0)

| | Fed_Aid_Excl_Loans_Range = 12. 2500-2749: Yes (0.0)

| #_of_CSP_Visits = 5. 21+: No (1.0)

| #_of_CSP_Visits = 2. 6-10: No (0.0)

| #_of_CSP_Visits = 3. 11-15: No (0.0)

| #_of_CSP_Visits = 4. 16-20: No (0.0)

AU_Next_Ret = YES

| Cum_QPts_Ranges = 2. 1-12: No (72.0/5.0)

| Cum_QPts_Ranges = 3. 13-24: No (194.0/43.0)

| Cum_QPts_Ranges = 5. 37-48

| | Age_Range = 35-39: No (2.0)

| | Age_Range = 30-34: Yes (4.0/1.0)

| | Age_Range = 40-49: No (2.0)

| | Age_Range = 25-29: No (4.0/1.0)

| | Age_Range = 18-19

| | | Cum_Credit_Hour_Range = 7.00-11.00: No (4.0/1.0)

| | | Cum_Credit_Hour_Range = 1.00-6.00: Yes (0.0)

| | | Cum_Credit_Hour_Range = 12.00-16.00

| | | | #_of_CSP_Visits = None

| | | | | Major = Health Professions and Clinical Services

| | | | | | Other_3rd_Party_Aid_Range = 2. 1-100: Yes (6.0)

| | | | | | Other_3rd_Party_Aid_Range = 1. No_3rd_Party_Aid: No (3.0)

| | | | | | Other_3rd_Party_Aid_Range = 7. 1501-2500: Yes (1.0)

| | | | | | Other_3rd_Party_Aid_Range = 4. 501-600

| | | | | | | Fed_Aid_Excl_Loans_Range = 9. 1750-1999: Yes (2.0)

| | | | | | | Fed_Aid_Excl_Loans_Range = 1. No_Aid: No (2.0)

| | | | | | | Fed_Aid_Excl_Loans_Range = 5. 750-999: No (0.0)

| | | | | | | Fed_Aid_Excl_Loans_Range = 7. 1250-1499: No (0.0)

| | | | | | | Fed_Aid_Excl_Loans_Range = 3. 200-499: No (0.0)

| | | | | | | Fed_Aid_Excl_Loans_Range = 4. 500-749: No (0.0)

| | | | | | | Fed_Aid_Excl_Loans_Range = 11. 2250-2499: No (0.0)

| | | | | | | Fed_Aid_Excl_Loans_Range = 6. 1000-1249: No (0.0)

| | | | | | | Fed_Aid_Excl_Loans_Range = 8. 1500-1749: No (0.0)

| | | | | | | Fed_Aid_Excl_Loans_Range = 2. 1-199: No (0.0)

| | | | | | | Fed_Aid_Excl_Loans_Range = 10. 2000-2249: No (0.0)

| | | | | | | Fed_Aid_Excl_Loans_Range = 12. 2500-2749: No (0.0)

| | | | | | Other_3rd_Party_Aid_Range = 8. 2501-5000: Yes (0.0)

| | | | | | Other_3rd_Party_Aid_Range = 6. 1001-1500: Yes (0.0)
| | | | | | Other_3rd_Party_Aid_Range = 5. 601-1000: Yes (0.0)
| | | | | | Other_3rd_Party_Aid_Range = 3. 101-500
| | | | | | | HS_CEEB_Code <= 361870: Yes (2.0)
| | | | | | | HS_CEEB_Code > 361870: No (2.0)
| | | | | | Other_3rd_Party_Aid_Range = 9. 5001-8000: Yes (0.0)
| | | | | Major = Business Management and Marketing: Yes (38.0/9.0)
| | | | | Major = Public Administration and Social Service: Yes (3.0)
| | | | | Major = Computer and Information Sciences
| | | | | | Ethnicity = Black: No (0.0)
| | | | | | Ethnicity = White: No (4.0/1.0)
| | | | | | Ethnicity = Unspecified_Race: Yes (2.0)
| | | | | | Ethnicity = Hispanic: No (0.0)
| | | | | | Ethnicity = International: No (0.0)
| | | | | | Ethnicity = Asian: No (0.0)
| | | | | | Ethnicity = American_Indian: No (0.0)
| | | | | Major = Social Sciences: Yes (4.0)
| | | | | Major = Education
| | | | | | Remedial_Math = Failed: No (1.0)
| | | | | | Remedial_Math = Did_not_take: Yes (28.0/6.0)
| | | | | | Remedial_Math = Passed
| | | | | | | Institutional_Aid_Range = 1. No_Institutional_Aid: No (9.0/1.0)
| | | | | | | Institutional_Aid_Range = 2. 1-500
| | | | | | | | HS_CEEB_Code <= 365507: Yes (3.0)
| | | | | | | | HS_CEEB_Code > 365507: No (2.0)

| | | | | | | Institutional_Aid_Range = 6. 2001-3000: No (0.0)

| | | | | | | Institutional_Aid_Range = 7. 3001-4000: No (0.0)

| | | | | | | Institutional_Aid_Range = 3. 501-1000: Yes (1.0)

| | | | | | | Institutional_Aid_Range = 4. 1001-1500: No (1.0)

| | | | | | | Institutional_Aid_Range = 9. 5001-6000: No (0.0)

| | | | | | | Institutional_Aid_Range = 5. 1501-2000: No (0.0)

| | | | | | | Institutional_Aid_Range = 10. 6001-7000: No (0.0)

| | | | | | | Institutional_Aid_Range = 11. 7001-8000: No (0.0)

| | | | | | | Institutional_Aid_Range = 8. 4001-5000: No (0.0)

| | | | | Major = Biological and Biomedical Sciences

| | | | | | Academic_Intent = Obtain_Bachelors_Degree: Yes (2.0)

| | | | | | Academic_Intent = Obtain_Associate_Degree_for_Job_Market: Yes (0.0)

| | | | | | Academic_Intent = Personal_Interest: No (1.0)

| | | | | | Academic_Intent = Unknown

| | | | | | | Other_3rd_Party_Aid_Range = 2. 1-100: Yes (2.0)

| | | | | | | Other_3rd_Party_Aid_Range = 1. No_3rd_Party_Aid: No (0.0)

| | | | | | | Other_3rd_Party_Aid_Range = 7. 1501-2500: No (0.0)

| | | | | | | Other_3rd_Party_Aid_Range = 4. 501-600: No (2.0)

| | | | | | | Other_3rd_Party_Aid_Range = 8. 2501-5000: No (0.0)

| | | | | | | Other_3rd_Party_Aid_Range = 6. 1001-1500: No (0.0)

| | | | | | | Other_3rd_Party_Aid_Range = 5. 601-1000: No (0.0)

| | | | | | | Other_3rd_Party_Aid_Range = 3. 101-500: No (0.0)

| | | | | | | Other_3rd_Party_Aid_Range = 9. 5001-8000: No (0.0)

| | | | | | Academic_Intent = Selected_Courses_Train_New_Career: Yes (0.0)

| | | | | | Academic_Intent = Selected_Courses_Upgrade_Skills: Yes (0.0)

```
|   |   |   |   |   |   Academic_Intent = Transfer_Before_Degree: Yes (0.0)
|   |   |   |   |   |   Academic_Intent = Obtain_Associate_Degree_for_Transfer: Yes (0.0)
|   |   |   |   |   |   Academic_Intent = Obtain_Undergraduate_Certificate: Yes (0.0)
|   |   |   |   |   Major = Engineering
|   |   |   |   |   |   Ethnicity = Black: Yes (0.0)
|   |   |   |   |   |   Ethnicity = White: Yes (16.0/2.0)
|   |   |   |   |   |   Ethnicity = Unspecified_Race: No (3.0/1.0)
|   |   |   |   |   |   Ethnicity = Hispanic: Yes (0.0)
|   |   |   |   |   |   Ethnicity = International: Yes (0.0)
|   |   |   |   |   |   Ethnicity = Asian: Yes (0.0)
|   |   |   |   |   |   Ethnicity = American_Indian: No (1.0)
|   |   |   |   |   Major = Liberal Arts and General Studies
|   |   |   |   |   |   Ethnicity = Black: No (0.0)
|   |   |   |   |   |   Ethnicity = White
|   |   |   |   |   |   |   HS_CEEB_Code <= 363487: No (6.0)
|   |   |   |   |   |   |   HS_CEEB_Code > 363487
|   |   |   |   |   |   |   |   State_Resident = Yes: Yes (22.0/6.0)
|   |   |   |   |   |   |   |   State_Resident = No: No (2.0)
|   |   |   |   |   |   Ethnicity = Unspecified_Race: No (3.0)
|   |   |   |   |   |   Ethnicity = Hispanic: No (0.0)
|   |   |   |   |   |   Ethnicity = International: No (0.0)
|   |   |   |   |   |   Ethnicity = Asian: No (0.0)
|   |   |   |   |   |   Ethnicity = American_Indian: No (0.0)
|   |   |   |   |   Major = Security and Protective Services
|   |   |   |   |   |   Gender = Female: Yes (4.0)
```

```
|   |   |   |   |   |   Gender = Male: No (4.0/1.0)
|   |   |   |   |   Major = Natural Resources and Conservation: Yes (0.0)
|   |   |   |   |   Major = English Language and Literature: No (6.0/1.0)
|   |   |   |   |   Major = Area, Ethnic, Cultural, Gender Studies: Yes (0.0)
|   |   |   |   |   Major = Visual and Performing Arts
|   |   |   |   |   |   HS_CEEB_Code <= 365310: No (11.0)
|   |   |   |   |   |   HS_CEEB_Code > 365310: Yes (11.0/4.0)
|   |   |   |   |   Major = Legal Professions and Studies
|   |   |   |   |   |   PID <= 250530: No (2.0)
|   |   |   |   |   |   PID > 250530: Yes (3.0)
|   |   |   |   |   Major = Foreign Languages and Literature: Yes (0.0)
|   |   |   |   |   Major = Engineering Technology: No (1.0)
|   |   |   |   |   Major = Psychology
|   |   |   |   |   |   Gender = Female: No (9.0/2.0)
|   |   |   |   |   |   Gender = Male: Yes (3.0/1.0)
|   |   |   |   |   Major = Physical Sciences: Yes (7.0/1.0)
|   |   |   |   |   Major = Mathematics and Statistics: No (1.0)
|   |   |   |   |   Major = Communication and Journalism: No (3.0/1.0)
|   |   |   |   |   Major = Precision Production: Yes (1.0)
|   |   |   |   |   Major = Leisure and Fitness Studies
|   |   |   |   |   |   Remedial_Math = Failed: No (0.0)
|   |   |   |   |   |   Remedial_Math = Did_not_take
|   |   |   |   |   |   |   Remedial_English = Passed: No (2.0)
|   |   |   |   |   |   |   Remedial_English = Did_not_take: Yes (7.0)
|   |   |   |   |   |   |   Remedial_English = Failed: Yes (0.0)
```

```
|  |  |  |  |  |  Remedial_Math = Passed: No (5.0)
|  |  |  |  |  Major = Family and Consumer Sciences: Yes (3.0/1.0)
|  |  |  |  |  Major = Philosophy and Religious Studies: Yes (0.0)
|  |  |  |  #_of_CSP_Visits = 1. 1-5
|  |  |  |  |  State_Aid_Range = 3. 501-1000: Yes (5.0/2.0)
|  |  |  |  |  State_Aid_Range = 2. 1-500: Yes (11.0/1.0)
|  |  |  |  |  State_Aid_Range = 4. 1001-1500: No (4.0/1.0)
|  |  |  |  |  State_Aid_Range = 1. No_State_Aid
|  |  |  |  |  |  Fed_Aid_Excl_Loans_Range = 9. 1750-1999: Yes (0.0)
|  |  |  |  |  |  Fed_Aid_Excl_Loans_Range = 1. No_Aid
|  |  |  |  |  |  |  Institutional_Aid_Range = 1. No_Institutional_Aid
|  |  |  |  |  |  |  |  Dependency = I: Yes (0.0)
|  |  |  |  |  |  |  |  Dependency = D: Yes (16.0/3.0)
|  |  |  |  |  |  |  |  Dependency =  : No (11.0/3.0)
|  |  |  |  |  |  |  |  Dependency = X: Yes (0.0)
|  |  |  |  |  |  |  |  Dependency = Y: Yes (0.0)
|  |  |  |  |  |  |  Institutional_Aid_Range = 2. 1-500: Yes (17.0/4.0)
|  |  |  |  |  |  |  Institutional_Aid_Range = 6. 2001-3000: No (1.0)
|  |  |  |  |  |  |  Institutional_Aid_Range = 7. 3001-4000: No (1.0)
|  |  |  |  |  |  |  Institutional_Aid_Range = 3. 501-1000
|  |  |  |  |  |  |  |  PID <= 251467: Yes (2.0)
|  |  |  |  |  |  |  |  PID > 251467: No (2.0)
|  |  |  |  |  |  |  Institutional_Aid_Range = 4. 1001-1500: No (2.0/1.0)
|  |  |  |  |  |  |  Institutional_Aid_Range = 9. 5001-6000: Yes (1.0)
|  |  |  |  |  |  |  Institutional_Aid_Range = 5. 1501-2000: No (1.0)
```

```
|  |  |  |  |  |  |  Institutional_Aid_Range =  10. 6001-7000: Yes (0.0)
|  |  |  |  |  |  |  Institutional_Aid_Range =  11. 7001-8000: Yes (0.0)
|  |  |  |  |  |  |  Institutional_Aid_Range = 8. 4001-5000: Yes (0.0)
|  |  |  |  |  |  Fed_Aid_Excl_Loans_Range = 5. 750-999: Yes (0.0)
|  |  |  |  |  |  Fed_Aid_Excl_Loans_Range = 7. 1250-1499: Yes (0.0)
|  |  |  |  |  |  Fed_Aid_Excl_Loans_Range = 3. 200-499: No (2.0)
|  |  |  |  |  |  Fed_Aid_Excl_Loans_Range = 4. 500-749: Yes (0.0)
|  |  |  |  |  |  Fed_Aid_Excl_Loans_Range = 11. 2250-2499: Yes (0.0)
|  |  |  |  |  |  Fed_Aid_Excl_Loans_Range = 6. 1000-1249: Yes (1.0)
|  |  |  |  |  |  Fed_Aid_Excl_Loans_Range = 8. 1500-1749: Yes (0.0)
|  |  |  |  |  |  Fed_Aid_Excl_Loans_Range = 2. 1-199: Yes (0.0)
|  |  |  |  |  |  Fed_Aid_Excl_Loans_Range = 10. 2000-2249: Yes (0.0)
|  |  |  |  |  |  Fed_Aid_Excl_Loans_Range = 12. 2500-2749: Yes (0.0)
|  |  |  |  |  State_Aid_Range = 6. 2001-2500: Yes (0.0)
|  |  |  |  |  State_Aid_Range = 7. Over 2500: Yes (0.0)
|  |  |  |  #_of_CSP_Visits = 5. 21+: Yes (0.0)
|  |  |  |  #_of_CSP_Visits = 2. 6-10: Yes (9.0)
|  |  |  |  #_of_CSP_Visits = 3. 11-15: No (4.0/2.0)
|  |  |  |  #_of_CSP_Visits = 4. 16-20: Yes (2.0)
|  |  |  Cum_Credit_Hour_Range = 0.0: Yes (0.0)
|  |  |  Cum_Credit_Hour_Range = 22.00-26.00: Yes (3.0/1.0)
|  |  |  Cum_Credit_Hour_Range = 17.00-21.00
|  |  |  |  Academic_Intent = Obtain_Bachelors_Degree
|  |  |  |  |  PID <= 251102: Yes (11.0/2.0)
|  |  |  |  |  PID > 251102: No (4.0)
```

| | | | Academic_Intent = Obtain_Associate_Degree_for_Job_Market: No (0.0)

| | | | Academic_Intent = Personal_Interest: No (1.0)

| | | | Academic_Intent = Unknown: No (9.0/1.0)

| | | | Academic_Intent = Selected_Courses_Train_New_Career: No (0.0)

| | | | Academic_Intent = Selected_Courses_Upgrade_Skills: No (0.0)

| | | | Academic_Intent = Transfer_Before_Degree: No (1.0)

| | | | Academic_Intent = Obtain_Associate_Degree_for_Transfer: No (0.0)

| | | | Academic_Intent = Obtain_Undergraduate_Certificate: No (0.0)

| | | Cum_Credit_Hour_Range = 27.00-31.00: Yes (0.0)

| | | Cum_Credit_Hour_Range = 32.00-over: No (3.0/1.0)

| | Age_Range = 22-24

| | | Comp_ACT_Ranges = 2. 6-11: No (0.0)

| | | Comp_ACT_Ranges = No ACT: No (8.0/2.0)

| | | Comp_ACT_Ranges = 4. 18-23: No (0.0)

| | | Comp_ACT_Ranges = 5. 24-29: Yes (2.0)

| | | Comp_ACT_Ranges = 6. 30-36: No (0.0)

| | | Comp_ACT_Ranges = 3. 12-17: No (0.0)

| | Age_Range = 50-64: Yes (0.0)

| | Age_Range = 20-21

| | | 9_Month_Expected_Family_Contribution = 1. No_Family_Contribution: No (2.0)

| | | 9_Month_Expected_Family_Contribution = 5. 5001-7000: Yes (0.0)

| | | 9_Month_Expected_Family_Contribution = 9. 13001-20000: Yes (0.0)

| | | 9_Month_Expected_Family_Contribution = 2. 1-1000: Yes (0.0)

| | | 9_Month_Expected_Family_Contribution = 6. 7001-9000: Yes (0.0)

| | | 9_Month_Expected_Family_Contribution = 4. 3001-5000: Yes (2.0)

| | | 9_Month_Expected_Family_Contribution = 10. 20001-40000: Yes (0.0)

| | | 9_Month_Expected_Family_Contribution = 3. 1001-3000: Yes (1.0)

| | | 9_Month_Expected_Family_Contribution = 7. 9001-11000: Yes (0.0)

| | | 9_Month_Expected_Family_Contribution = 8. 11001-13000: Yes (0.0)

| | | 9_Month_Expected_Family_Contribution = 11. Over 40001: Yes (0.0)

| | Age_Range = Under_18: Yes (0.0)

| Cum_QPts_Ranges = 4. 25-36

| | #_of_CSP_Visits = None: No (238.0/68.0)

| | #_of_CSP_Visits = 1. 1-5

| | | Ethnicity = Black: No (2.0)

| | | Ethnicity = White

| | | | Remedial_English = Passed

| | | | | Cum_GPA_CrHr_Ranges = 3. 6-10: No (7.0/1.0)

| | | | | Cum_GPA_CrHr_Ranges = 4. 11-15

| | | | | | Commuter = Yes: No (14.0/6.0)

| | | | | | Commuter = No: Yes (3.0)

| | | | | Cum_GPA_CrHr_Ranges = 2. 1-5: No (0.0)

| | | | | Cum_GPA_CrHr_Ranges = 1. 0: No (0.0)

| | | | | Cum_GPA_CrHr_Ranges = 5. 16-19: Yes (1.0)

| | | | Remedial_English = Did_not_take: Yes (16.0/2.0)

| | | | Remedial_English = Failed: Yes (0.0)

| | | Ethnicity = Unspecified_Race: No (3.0/1.0)

| | | Ethnicity = Hispanic: No (2.0)

| | | Ethnicity = International: Yes (0.0)

| | | Ethnicity = Asian: No (1.0)

| | | Ethnicity = American_Indian: Yes (0.0)

| | #_of_CSP_Visits = 5. 21+: No (1.0)

| | #_of_CSP_Visits = 2. 6-10

| | | R&SK = Passed: Yes (3.0)

| | | R&SK = Did_not_take: No (4.0)

| | | R&SK = Failed: No (0.0)

| | #_of_CSP_Visits = 3. 11-15

| | | State_Aid_Range = 3. 501-1000: Yes (0.0)

| | | State_Aid_Range = 2. 1-500: Yes (0.0)

| | | State_Aid_Range = 4. 1001-1500: No (2.0)

| | | State_Aid_Range = 1. No_State_Aid: Yes (7.0/1.0)

| | | State_Aid_Range = 6. 2001-2500: Yes (0.0)

| | | State_Aid_Range = 7. Over 2500: Yes (0.0)

| | #_of_CSP_Visits = 4. 16-20: Yes (1.0)

| Cum_QPts_Ranges = 1. 0: No (56.0)

| Cum_QPts_Ranges = 6. 49-60

| | Associate_Ever = No: Yes (280.0/69.0)

| | Associate_Ever = Yes: No (8.0/2.0)

| Cum_QPts_Ranges = 7. 61-76

| | Associate_Ever = No: Yes (60.0/4.0)

| | Associate_Ever = Yes: No (2.0)

Number of Leaves : 227

Size of the tree : 272

Time taken to build model: 0.05 seconds

=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances	1743	86.2871 %
Incorrectly Classified Instances	277	13.7129 %
Kappa statistic	0.6873	
Mean absolute error	0.2074	
Root mean squared error	0.3221	
Relative absolute error	46.6442 %	
Root relative squared error	68.3008 %	
Total Number of Instances	2020	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.911	0.233	0.886	0.911	0.898	0.918	No
	0.767	0.089	0.812	0.767	0.789	0.918	Yes
Weighted Avg.	0.863	0.185	0.862	0.863	0.862	0.918	

=== Confusion Matrix ===

```
   a    b   <-- classified as
1226  120 |    a = No
 157  517 |    b = Yes
```

	2001 First-Time Undergraduate Cohort					J48 Predicted to Earn a Bachelor Degree						J48 Predicted NOT to Earned a Bachelor Degree					
	Actually Earned Bachelor Degree					Actually Earned Bachelor Degree						Actually Earned Bachelor Degree					
	Yes		No		Cohort	Yes		No		Part of Cohort		Yes		No		Part of Cohort	
	#	%	#	%	#	#	%	#	%	#	%	#	%	#	%	#	%
Gender																	
Female	395	58.61%	674	50.07%	1,069	310	59.96%	72	60.00%	382	35.73%	85	54.14%	602	49.10%	687	64.27%
Male	279	41.39%	672	49.93%	951	207	40.04%	48	40.00%	255	26.81%	72	45.86%	624	50.90%	696	73.19%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Age Ranges																	
Under_18	1	0.15%	1	0.07%	2		0.00%		0.00%		0.00%	1	0.64%	1	0.08%	2	100.00%
18-19	642	95.25%	1,093	81.20%	1,735	500	96.71%	111	92.50%	611	35.22%	142	90.45%	982	80.10%	1,124	64.78%
20-21	9	1.34%	85	6.32%	94	6	1.16%	2	1.67%	8	8.51%	3	1.91%	83	6.77%	86	91.49%
22-24	8	1.19%	70	5.20%	78	3	0.58%	2	1.67%	5	6.41%	5	3.18%	68	5.55%	73	93.59%
25-29	7	1.04%	38	2.82%	45	3	0.58%	1	0.83%	4	8.89%	4	2.55%	37	3.02%	41	91.11%
30-34	4	0.59%	16	1.19%	20	4	0.77%	2	1.67%	6	30.00%		0.00%	14	1.14%	14	70.00%
35-39	2	0.30%	14	1.04%	16	1	0.19%	1	0.83%	2	12.50%	1	0.64%	13	1.06%	14	87.50%
40-49	1	0.15%	25	1.86%	26		0.00%	1	0.83%	1	3.85%	1	0.64%	24	1.96%	25	96.15%
50-64		0.00%	4	0.30%	4		0.00%		0.00%		0.00%		0.00%	4	0.33%	4	100.00%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Race/Ethnicity																	
American_Indian	2	0.30%	7	0.52%	9	2	0.39%		0.00%	2	22.22%		0.00%	7	0.57%	7	77.78%
Asian	5	0.74%	6	0.45%	11	5	0.97%	1	0.83%	6	54.55%		0.00%	5	0.41%	5	45.45%
Black	34	5.04%	159	11.81%	193	15	2.90%	3	2.50%	18	9.33%	19	12.10%	156	12.72%	175	90.67%
Hispanic	7	1.04%	36	2.67%	43	6	1.16%	2	1.67%	8	18.60%	1	0.64%	34	2.77%	35	81.40%
International	7	1.04%	5	0.37%	12	7	1.35%	2	1.67%	9	75.00%		0.00%	3	0.24%	3	25.00%
Unspecified_Race	40	5.93%	87	6.46%	127	29	5.61%	5	4.17%	34	26.77%	11	7.01%	82	6.69%	93	73.23%
White	579	85.91%	1,046	77.71%	1,625	453	87.62%	107	89.17%	560	34.46%	126	80.25%	939	76.59%	1,065	65.54%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Resident of Ohio																	
Yes	604	89.61%	1,207	89.67%	1,811	458	88.59%	102	85.00%	560	30.92%	146	92.99%	1,105	90.13%	1,251	69.08%
No	70	10.39%	139	10.33%	209	59	11.41%	18	15.00%	77	36.84%	11	7.01%	121	9.87%	132	63.16%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Commuter																	
Yes	524	77.74%	1,120	83.21%	1,644	398	76.98%	92	76.67%	490	29.81%	126	80.25%	1,028	83.85%	1,154	70.19%
No	150	22.26%	226	16.79%	376	119	23.02%	28	23.33%	147	39.10%	31	19.75%	198	16.15%	229	60.90%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Composite ACT Score Range																	
No ACT	70	10.39%	315	23.40%	385	51	9.86%	24	20.00%	75	19.48%	19	12.10%	291	23.74%	310	80.52%
2. 6-11	1	0.15%	5	0.37%	6		0.00%		0.00%		0.00%	1	0.64%	5	0.41%	6	100.00%
3. 12-17	102	15.13%	365	27.12%	467	58	11.22%	15	12.50%	73	15.63%	44	28.03%	350	28.55%	394	84.37%
4. 18-23	304	45.10%	511	37.96%	815	224	43.33%	52	43.33%	276	33.87%	80	50.96%	459	37.44%	539	66.13%
5. 24-29	173	25.67%	143	10.62%	316	160	30.95%	28	23.33%	188	59.49%	13	8.28%	115	9.38%	128	40.51%
6. 30-36	24	3.56%	7	0.52%	31	24	4.64%	1	0.83%	25	80.65%		0.00%	6	0.49%	6	19.35%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%

Table 1

	2001 First-Time Undergraduate Cohort					J48 Predicted to Earn a Bachelor Degree						J48 Predicted NOT to Earned a Bachelor Degree					
	Actually Earned Bachelor Degree					Actually Earned Bachelor Degree				Part of Cohort		Actually Earned Bachelor Degree				Part of Cohort	
	Yes		No		Cohort	Yes		No				Yes		No			
	#	%	#	%	#	#	%	#	%	#	%	#	%	#	%	#	%
High School Graduating GPA Ranges																	
Below_1.0	39	5.79%	122	9.06%	161	34	6.58%	6	5.00%	40	24.84%	5	3.18%	116	9.46%	121	75.16%
1.0-1.99	10	1.48%	175	13.00%	185	3	0.58%	3	2.50%	6	3.24%	7	4.46%	172	14.03%	179	96.76%
2.0-2.49	59	8.75%	275	20.43%	334	30	5.80%	8	6.67%	38	11.38%	29	18.47%	267	21.78%	296	88.62%
2.5-2.99	144	21.36%	329	24.44%	473	83	16.05%	25	20.83%	108	22.83%	61	38.85%	304	24.80%	365	77.17%
3.0-3.24	111	16.47%	189	14.04%	300	79	15.28%	25	20.83%	104	34.67%	32	20.38%	164	13.38%	196	65.33%
3.25-3.49	70	10.39%	89	6.61%	159	57	11.03%	17	14.17%	74	46.54%	13	8.28%	72	5.87%	85	53.46%
3.5-3.74	108	16.02%	78	5.79%	186	102	19.73%	21	17.50%	123	66.13%	6	3.82%	57	4.65%	63	33.87%
3.75_and_higher	131	19.44%	50	3.71%	181	128	24.76%	15	12.50%	143	79.01%	3	1.91%	35	2.85%	38	20.99%
GED_recipient	2	0.30%	36	2.67%	38	1	0.19%		0.00%	1	2.63%	1	0.64%	36	2.94%	37	97.37%
No GPA Information		0.00%	3	0.22%	3		0.00%		0.00%		0.00%		0.00%	3	0.24%	3	100.00%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Any Advanced Placement Credits																	
Yes	28	4.15%	13	0.97%	41	27	5.22%	6	5.00%	33	80.49%	1	0.64%	7	0.57%	8	19.51%
No	646	95.85%	1,333	99.03%	1,979	490	94.78%	114	95.00%	604	30.52%	156	99.36%	1,219	99.43%	1,375	69.48%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Academic Intention																	
Obtain_Associate_Degree_for_Job_Market	11	1.63%	35	2.60%	46	10	1.93%	2	1.67%	12	26.09%	1	0.64%	33	2.69%	34	73.91%
Obtain_Associate_Degree_for_Transfer	1	0.15%	1	0.07%	2	1	0.19%		0.00%	1	50.00%		0.00%	1	0.08%	1	50.00%
Obtain_Bachelors_Degree	313	46.44%	543	40.34%	856	235	45.45%	51	42.50%	286	33.41%	78	49.68%	492	40.13%	570	66.59%
Obtain_Undergraduate_Certificate	1	0.15%	1	0.07%	2	1	0.19%		0.00%	1	50.00%		0.00%	1	0.08%	1	50.00%
Personal_Interest	24	3.56%	87	6.46%	111	16	3.09%	8	6.67%	24	21.62%	8	5.10%	79	6.44%	87	78.38%
Selected_Courses_Train_New_Career	6	0.89%	20	1.49%	26	4	0.77%	1	0.83%	5	19.23%	2	1.27%	19	1.55%	21	80.77%
Selected_Courses_Upgrade_Skills	4	0.59%	10	0.74%	14	3	0.58%		0.00%	3	21.43%	1	0.64%	10	0.82%	11	78.57%
Transfer_Before_Degree	2	0.30%	12	0.89%	14	2	0.39%		0.00%	2	14.29%		0.00%	12	0.98%	12	85.71%
Unknown	312	46.29%	637	47.33%	949	245	47.39%	58	48.33%	303	31.93%	67	42.68%	579	47.23%	646	68.07%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Major Field of Study																	
Area, Ethnic, Cultural, Gender Studies		0.00%	1	0.07%	1		0.00%		0.00%		0.00%		0.00%	1	0.08%	1	100.00%
Biological and Biomedical Sciences	18	2.67%	58	4.31%	76	14	2.71%	4	3.33%	18	23.68%	4	2.55%	54	4.40%	58	76.32%
Business Management and Marketing	110	16.32%	204	15.16%	314	85	16.44%	22	18.33%	107	34.08%	25	15.92%	182	14.85%	207	65.92%
Communication and Journalism	8	1.19%	17	1.26%	25	3	0.58%	2	1.67%	5	20.00%	5	3.18%	15	1.22%	20	80.00%
Computer and Information Sciences	27	4.01%	90	6.69%	117	16	3.09%	6	5.00%	22	18.80%	11	7.01%	84	6.85%	95	81.20%
Education	143	21.22%	193	14.34%	336	111	21.47%	20	16.67%	131	38.99%	32	20.38%	173	14.11%	205	61.01%
Engineering	86	12.76%	88	6.54%	174	70	13.54%	9	7.50%	79	45.40%	16	10.19%	79	6.44%	95	54.60%
Engineering Technology	3	0.45%	24	1.78%	27	1	0.19%	1	0.83%	2	7.41%	2	1.27%	23	1.88%	25	92.59%
English Language and Literature	17	2.52%	37	2.75%	54	11	2.13%	3	2.50%	14	25.93%	6	3.82%	34	2.77%	40	74.07%
Family and Consumer Sciences	6	0.89%	18	1.34%	24	4	0.77%	2	1.67%	6	25.00%	2	1.27%	16	1.31%	18	75.00%
Foreign Languages and Literature	3	0.45%	3	0.22%	6	3	0.58%	2	1.67%	5	83.33%		0.00%	1	0.08%	1	16.67%
Health Professions and Clinical Services	28	4.15%	96	7.13%	124	21	4.06%	3	2.50%	24	19.35%	7	4.46%	93	7.59%	100	80.65%
Legal Professions and Studies	8	1.19%	18	1.34%	26	6	1.16%	3	2.50%	9	34.62%	2	1.27%	15	1.22%	17	65.38%
Leisure and Fitness Studies	22	3.26%	42	3.12%	64	15	2.90%	3	2.50%	18	28.13%	7	4.46%	39	3.18%	46	71.88%
Liberal Arts and General Studies	67	9.94%	154	11.44%	221	52	10.06%	13	10.83%	65	29.41%	15	9.55%	141	11.50%	156	70.59%
Mathematics and Statistics	3	0.45%	3	0.22%	6	3	0.58%		0.00%	3	50.00%		0.00%	3	0.24%	3	50.00%

	2001 First-Time Undergraduate Cohort					J48 Predicted to Earn a Bachelor Degree						J48 Predicted NOT to Earned a Bachelor Degree					
	Actually Earned Bachelor Degree					**Actually Earned Bachelor Degree**				**Part of Cohort**		**Actually Earned Bachelor Degree**				**Part of Cohort**	
	Yes		**No**		**Cohort**	**Yes**		**No**				**Yes**		**No**			
	#	%	#	%	#	#	%	#	%	#	%	#	%	#	%	#	%
Natural Resources and Conservation	1	0.15%	1	0.07%	2	1	0.19%		0.00%	1	50.00%		0.00%	1	0.08%	1	50.00%
Philosophy and Religious Studies	2	0.30%	3	0.22%	5	1	0.19%		0.00%	1	20.00%	1	0.64%	3	0.24%	4	80.00%
Physical Sciences	17	2.52%	18	1.34%	35	16	3.09%	4	3.33%	20	57.14%	1	0.64%	14	1.14%	15	42.86%
Precision Production	1	0.15%	3	0.22%	4	1	0.19%		0.00%	1	25.00%		0.00%	3	0.24%	3	75.00%
Psychology	17	2.52%	54	4.01%	71	11	2.13%	10	8.33%	21	29.58%	6	3.82%	44	3.59%	50	70.42%
Public Administration and Social Service	5	0.74%	23	1.71%	28	3	0.58%		0.00%	3	10.71%	2	1.27%	23	1.88%	25	89.29%
Security and Protective Services	18	2.67%	74	5.50%	92	10	1.93%		0.00%	10	10.87%	8	5.10%	74	6.04%	82	89.13%
Social Sciences	13	1.93%	14	1.04%	27	13	2.51%		0.00%	13	48.15%		0.00%	14	1.14%	14	51.85%
Visual and Performing Arts	51	7.57%	110	8.17%	161	46	8.90%	13	10.83%	59	36.65%	5	3.18%	97	7.91%	102	63.35%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Student Marital Status																	
Life_Partner		0.00%	7	0.52%	7		0.00%	1	0.83%	1	14.29%		0.00%	6	0.49%	6	85.71%
Married	9	1.34%	35	2.60%	44	7	1.35%	3	2.50%	10	22.73%	2	1.27%	32	2.61%	34	77.27%
Single	665	98.66%	1,304	96.88%	1,969	510	98.65%	116	96.67%	626	31.79%	155	98.73%	1,188	96.90%	1,343	68.21%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Student Dependency Upon Parents																	
Dependent	499	74.04%	888	65.97%	1,387	400	77.37%	90	75.00%	490	35.33%	99	63.06%	798	65.09%	897	64.67%
Independent	23	3.41%	159	11.81%	182	10	1.93%	6	5.00%	16	8.79%	13	8.28%	153	12.48%	166	91.21%
Unspecified	149	22.11%	291	21.62%	440	107	20.70%	24	20.00%	131	29.77%	42	26.75%	267	21.78%	309	70.23%
Unspecified - X	3	0.45%	6	0.45%	9		0.00%		0.00%		0.00%	3	1.91%	6	0.49%	9	100.00%
Unspecified - Y		0.00%	2	0.15%	2		0.00%		0.00%		0.00%		0.00%	2	0.16%	2	100.00%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Cost of Attendance																	
1. No FAFSA on file	162	24.04%	343	25.48%	505	114	22.05%	28	23.33%	142	28.12%	48	30.57%	315	25.69%	363	71.88%
2. 9001-10000	289	42.88%	454	33.73%	743	233	45.07%	44	36.67%	277	37.28%	56	35.67%	410	33.44%	466	62.72%
3. 10001-12000	25	3.71%	50	3.71%	75	20	3.87%	9	7.50%	29	38.67%	5	3.18%	41	3.34%	46	61.33%
4. 12001-14000	152	22.55%	416	30.91%	568	115	22.24%	30	25.00%	145	25.53%	37	23.57%	386	31.48%	423	74.47%
5. 14001-16000	33	4.90%	56	4.16%	89	26	5.03%	9	7.50%	35	39.33%	7	4.46%	47	3.83%	54	60.67%
6. 16001-18000	13	1.93%	27	2.01%	40	9	1.74%		0.00%	9	22.50%	4	2.55%	27	2.20%	31	77.50%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Need Level Ranges - For those with a valid/complete FAFSA																	
1. -77999 to - 20000	14	2.73%	12	1.20%	26	12	2.98%	3	3.26%	15	57.69%	2	1.83%	9	0.99%	11	42.31%
2. -19999 to -10000	23	4.49%	23	2.30%	46	16	3.97%	2	2.17%	18	39.13%	7	6.42%	21	2.31%	28	60.87%
3. -9999 to -1	99	19.34%	127	12.67%	226	83	20.60%	12	13.04%	95	42.04%	16	14.68%	115	12.64%	131	57.96%
5. 1-2000	29	5.66%	43	4.29%	72	24	5.96%	10	10.87%	34	47.22%		0.00%		0.00%		0.00%
6. 2001-5000	64	12.50%	123	12.28%	187	52	12.90%	15	16.30%	67	35.83%	5	4.59%	33	3.63%	38	20.32%
7. 5001-8000	107	20.90%	151	15.07%	258	92	22.83%	14	15.22%	106	41.09%	12	11.01%	108	11.87%	120	46.51%
8. 8001-10000	104	20.31%	224	22.36%	328	78	19.35%	21	22.83%	99	30.18%	15	13.76%	137	15.05%	152	46.34%
9. 10001-12000	35	6.84%	69	6.89%	104	27	6.70%	8	8.70%	35	33.65%	26	23.85%	203	22.31%	229	220.19%
10. 12001-14000	30	5.86%	204	20.36%	234	17	4.22%	7	7.61%	24	10.26%	8	7.34%	61	6.70%	69	29.49%
11. Over 14001	7	1.37%	26	2.59%	33	2	0.50%		0.00%	2	6.06%	13	11.93%	197	21.65%	210	636.36%
Grand Total	512	100.00%	1,002	100.00%	1,514	403	100.00%	92	100.00%	495	32.69%	5	4.59%	26	2.86%	31	2.05%

Table 1

	2001 First-Time Undergraduate Cohort					J48 Predicted to Earn a Bachelor Degree						J48 Predicted NOT to Earned a Bachelor Degree					
	Actually Earned Bachelor Degree					Actually Earned Bachelor Degree				Part of Cohort		Actually Earned Bachelor Degree				Part of Cohort	
	Yes		No		Cohort	Yes		No				Yes		No			
	#	%	#	%	#	#	%	#	%	#	%	#	%	#	%	#	%
9-Month Expected Family Contribution - For those with a valid/complete FAFSA																	
1. No_Family_Contribution	51	9.96%	259	25.85%	310	32	7.94%	11	11.96%	43	13.87%	19	17.43%	248	27.25%	267	86.13%
2. 1-1000	27	5.27%	100	9.98%	127	15	3.72%	5	5.43%	20	15.75%	12	11.01%	95	10.44%	107	84.25%
3. 1001-3000	94	18.36%	142	14.17%	236	71	17.62%	13	14.13%	84	35.59%	23	21.10%	129	14.18%	152	64.41%
4. 3001-5000	86	16.80%	131	13.07%	217	76	18.86%	14	15.22%	90	41.47%	10	9.17%	117	12.86%	127	58.53%
5. 5001-7000	53	10.35%	101	10.08%	154	45	11.17%	14	15.22%	59	38.31%	8	7.34%	87	9.56%	95	61.69%
6. 7001-9000	36	7.03%	58	5.79%	94	29	7.20%	10	10.87%	39	41.49%	7	6.42%	48	5.27%	55	58.51%
7. 9001-11000	23	4.49%	52	5.19%	75	21	5.21%	6	6.52%	27	36.00%	2	1.83%	46	5.05%	48	64.00%
8. 11001-13000	34	6.64%	35	3.49%	69	25	6.20%	2	2.17%	27	39.13%	9	8.26%	33	3.63%	42	60.87%
9. 13001-20000	69	13.48%	82	8.18%	151	59	14.64%	12	13.04%	71	47.02%	10	9.17%	70	7.69%	80	52.98%
10. 20001-40000	34	6.64%	39	3.89%	73	26	6.45%	3	3.26%	29	39.73%	8	7.34%	36	3.96%	44	60.27%
11. Over 40001	5	0.98%	3	0.30%	8	4	0.99%	2	2.17%	6	75.00%	1	0.92%	1	0.11%	2	25.00%
Grand Total	512	100.00%	1,002	100.00%	1,514	403	100.00%	92	100.00%	495	32.69%	109	100.00%	910	100.00%	1,019	67.31%
Any Aid																	
Yes	670	99.41%	1,325	98.44%	1,995	513	99.23%	118	98.33%	631	31.63%	157	100.00%	1,207	98.45%	1,364	68.37%
No	4	0.59%	21	1.56%	25	4	0.77%	2	1.67%	6	24.00%		0.00%	19	1.55%	19	76.00%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Federal Aid (Excluding Student Loans) Ranges - For those with a valid/complete FAFSA																	
1. No_Aid	319	62.30%	473	47.21%	792	264	65.51%	63	68.48%	327	41.29%	55	50.46%	410	45.05%	465	58.71%
2. 1-199		0.00%	3	0.30%	3		0.00%		0.00%		0.00%		0.00%	3	0.33%	3	100.00%
3. 200-499	28	5.47%	46	4.59%	74	26	6.45%	1	1.09%	27	36.49%	2	1.83%	45	4.95%	47	63.51%
4. 500-749	21	4.10%	34	3.39%	55	17	4.22%	1	1.09%	18	32.73%	4	3.67%	33	3.63%	37	67.27%
5. 750-999	29	5.66%	42	4.19%	71	21	5.21%	3	3.26%	24	33.80%	8	7.34%	39	4.29%	47	66.20%
6. 1000-1249	22	4.30%	46	4.59%	68	15	3.72%	5	5.43%	20	29.41%	7	6.42%	41	4.51%	48	70.59%
7. 1250-1499	24	4.69%	37	3.69%	61	17	4.22%	1	1.09%	18	29.51%	7	6.42%	36	3.96%	43	70.49%
8. 1500-1749	15	2.93%	38	3.79%	53	7	1.74%	4	4.35%	11	20.75%	8	7.34%	34	3.74%	42	79.25%
9. 1750-1999	45	8.79%	236	23.55%	281	31	7.69%	13	14.13%	44	15.66%	14	12.84%	223	24.51%	237	84.34%
10. 2000-2249	6	1.17%	20	2.00%	26	4	0.99%	1	1.09%	5	19.23%	2	1.83%	19	2.09%	21	80.77%
11. 2250-2499	3	0.59%	27	2.69%	30	1	0.25%		0.00%	1	3.33%	2	1.83%	27	2.97%	29	96.67%
12. 2500-2749		0.00%		0.00%			0.00%		0.00%		#DIV/0!		0.00%		0.00%		#DIV/0!
Grand Total	512	100.00%	1,002	100.00%	1,514	403	100.00%	92	100.00%	495	32.69%	109	100.00%	910	100.00%	1,019	67.31%
State Aid Ranges - For those with a valid/complete FAFSA																	
1. No_State_Aid	355	69.34%	553	55.19%	908	294	72.95%	62	67.39%	356	39.21%	61	55.96%	491	53.96%	552	60.79%
2. 1-500	87	16.99%	216	21.56%	303	60	14.89%	18	19.57%	78	25.74%	27	24.77%	198	21.76%	225	74.26%
3. 501-1000	36	7.03%	127	12.67%	163	25	6.20%	9	9.78%	34	20.86%	11	10.09%	118	12.97%	129	79.14%
4. 1001-1500	33	6.45%	99	9.88%	132	24	5.96%	3	3.26%	27	20.45%	9	8.26%	96	10.55%	105	79.55%
6. 2001-2500	1	0.20%	6	0.60%	7		0.00%		0.00%		0.00%	1	0.92%	6	0.66%	7	100.00%
7. Over 2500		0.00%	1	0.10%	1		0.00%		0.00%		0.00%		0.00%	1	0.11%	1	100.00%
Grand Total	512	100.00%	1,002	100.00%	1,514	403	100.00%	92	100.00%	495	32.69%	109	100.00%	910	100.00%	1,019	67.31%

	2001 First-Time Undergraduate Cohort					J48 Predicted to Earn a Bachelor Degree						J48 Predicted NOT to Earned a Bachelor Degree					
	Actually Earned Bachelor Degree					Actually Earned Bachelor Degree				Part of Cohort		Actually Earned Bachelor Degree				Part of Cohort	
	Yes		No		Cohort	Yes		No				Yes		No			
	#	%	#	%	#	#	%	#	%	#	%	#	%	#	%	#	%
Federal Work Study Aid Ranges - For those with a valid/complete FAFSA																	
1. No_Work_Study	498	97.27%	978	97.60%	1,476	393	97.52%	89	96.74%	482	32.66%	105	96.33%	889	97.69%	994	67.34%
2. 1-250	1	0.20%	7	0.70%	8	1	0.25%		0.00%	1	12.50%		0.00%	7	0.77%	7	87.50%
3. 251-500	3	0.59%	10	1.00%	13	1	0.25%	2	2.17%	3	23.08%	2	1.83%	8	0.88%	10	76.92%
4. 501-750	5	0.98%		0.00%	5	3	0.74%		0.00%	3	60.00%	2	1.83%		0.00%	2	40.00%
5. 751-1000	3	0.59%	3	0.30%	6	3	0.74%	1	1.09%	4	66.67%		0.00%	2	0.22%	2	33.33%
6. 1001-1250	1	0.20%	1	0.10%	2	1	0.25%		0.00%	1	50.00%		0.00%	1	0.11%	1	50.00%
7. 1251-1500		0.00%	2	0.20%	2		0.00%		0.00%		0.00%		0.00%	2	0.22%	2	100.00%
8. 1501-1750	1	0.20%	1	0.10%	2	1	0.25%		0.00%	1	50.00%		0.00%	1	0.11%	1	50.00%
Grand Total	512	100.00%	1,002	100.00%	1,514	403	100.00%	92	100.00%	495	32.69%	109	100.00%	910	100.00%	1,019	67.31%
Institutional Aid Ranges																	
1. No_Institutional_Aid	280	41.54%	958	71.17%	1,238	170	32.88%	53	44.17%	223	18.01%	110	70.06%	905	73.82%	1,015	81.99%
2. 1-500	174	25.82%	244	18.13%	418	148	28.63%	38	31.67%	186	44.50%	26	16.56%	206	16.80%	232	55.50%
3. 501-1000	80	11.87%	63	4.68%	143	72	13.93%	13	10.83%	85	59.44%	8	5.10%	50	4.08%	58	40.56%
4. 1001-1500	57	8.46%	35	2.60%	92	54	10.44%	4	3.33%	58	63.04%	3	1.91%	31	2.53%	34	36.96%
5. 1501-2000	17	2.52%	13	0.97%	30	16	3.09%	3	2.50%	19	63.33%	1	0.64%	10	0.82%	11	36.67%
6. 2001-3000	15	2.23%	14	1.04%	29	11	2.13%	4	3.33%	15	51.72%	4	2.55%	10	0.82%	14	48.28%
7. 3001-4000	7	1.04%	6	0.45%	13	5	0.97%	1	0.83%	6	46.15%	2	1.27%	5	0.41%	7	53.85%
8. 4001-5000	7	1.04%	2	0.15%	9	4	0.77%		0.00%	4	44.44%	3	1.91%	2	0.16%	5	55.56%
9. 5001-6000	26	3.86%	5	0.37%	31	26	5.03%	2	1.67%	28	90.32%		0.00%	3	0.24%	3	9.68%
10. 6001-7000	5	0.74%	3	0.22%	8	5	0.97%	1	0.83%	6	75.00%		0.00%	2	0.16%	2	25.00%
11. 7001-8000	6	0.89%	3	0.22%	9	6	1.16%	1	0.83%	7	77.78%		0.00%	2	0.16%	2	22.22%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Other Third Party Aid Ranges																	
1. No_3rd_Party_Aid	18	2.67%	115	8.54%	133	10	1.93%	4	3.33%	14	10.53%	8	5.10%	111	9.05%	119	89.47%
2. 1-100	262	38.87%	768	57.06%	1,030	178	34.43%	48	40.00%	226	21.94%	84	53.50%	720	58.73%	804	78.06%
3. 101-500	23	3.41%	44	3.27%	67	22	4.26%	7	5.83%	29	43.28%	1	0.64%	37	3.02%	38	56.72%
4. 501-600	239	35.46%	273	20.28%	512	196	37.91%	36	30.00%	232	45.31%	43	27.39%	237	19.33%	280	54.69%
5. 601-1000	31	4.60%	27	2.01%	58	28	5.42%	8	6.67%	36	62.07%	3	1.91%	19	1.55%	22	37.93%
6. 1001-1500	44	6.53%	53	3.94%	97	37	7.16%	9	7.50%	46	47.42%	7	4.46%	44	3.59%	51	52.58%
7. 1501-2500	43	6.38%	47	3.49%	90	36	6.96%	6	5.00%	42	46.67%	7	4.46%	41	3.34%	48	53.33%
8. 2501-5000	12	1.78%	18	1.34%	30	9	1.74%	2	1.67%	11	36.67%	3	1.91%	16	1.31%	19	63.33%
9. 5001-8000	2	0.30%	1	0.07%	3	1	0.19%		0.00%	1	33.33%	1	0.64%	1	0.08%	2	66.67%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%

	2001 First-Time Undergraduate Cohort					J48 Predicted to Earn a Bachelor Degree						J48 Predicted NOT to Earned a Bachelor Degree					
	Actually Earned Bachelor Degree					Actually Earned Bachelor Degree				Part of Cohort		Actually Earned Bachelor Degree				Part of Cohort	
	Yes		No		Cohort	Yes		No				Yes		No			
	#	%	#	%	#	#	%	#	%	#	%	#	%	#	%	#	%
Student Loan Ranges																	
1. No_Student_Loan	435	64.54%	718	53.34%	1,153	346	66.92%	73	60.83%	419	36.34%	89	56.69%	645	52.61%	734	63.66%
2. 1-1000	16	2.37%	45	3.34%	61	11	2.13%	3	2.50%	14	22.95%	5	3.18%	42	3.43%	47	77.05%
3. 1001-2000	145	21.51%	308	22.88%	453	112	21.66%	27	22.50%	139	30.68%	33	21.02%	281	22.92%	314	69.32%
4. 2001-3000	27	4.01%	82	6.09%	109	16	3.09%	4	3.33%	20	18.35%	11	7.01%	78	6.36%	89	81.65%
5. 3001-4000	23	3.41%	122	9.06%	145	13	2.51%	6	5.00%	19	13.10%	10	6.37%	116	9.46%	126	86.90%
6. 4001-5000	12	1.78%	25	1.86%	37	7	1.35%		0.00%	7	18.92%	5	3.18%	25	2.04%	30	81.08%
7. 5001-6000	11	1.63%	31	2.30%	42	8	1.55%	5	4.17%	13	30.95%	3	1.91%	26	2.12%	29	69.05%
8. 6001-7000	4	0.59%	10	0.74%	14	3	0.58%	2	1.67%	5	35.71%	1	0.64%	8	0.65%	9	64.29%
9. 7001-8000		0.00%	3	0.22%	3		0.00%		0.00%		0.00%		0.00%	3	0.24%	3	100.00%
10. 8001-8500	1	0.15%	2	0.15%	3	1	0.19%		0.00%	1	33.33%		0.00%	2	0.16%	2	66.67%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
First Term Credit Hour Load																	
Full-Time	660	97.92%	1,240	92.12%	1,900	514	99.42%	120	100.00%	634	33.37%	146	92.99%	1,120	91.35%	1,266	66.63%
Part-Time	14	2.08%	106	7.88%	120	3	0.58%		0.00%	3	2.50%	11	7.01%	106	8.65%	117	97.50%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
First Term Attempted Credit Hours																	
1. 0		0.00%	91	6.76%	91		0.00%		0.00%		0.00%		0.00%	91	7.42%	91	100.00%
2. 1-5	2	0.30%	82	6.09%	84		0.00%		0.00%		0.00%	2	1.27%	82	6.69%	84	100.00%
3. 6-10	32	4.75%	315	23.40%	347	4	0.77%		0.00%	4	1.15%	28	17.83%	315	25.69%	343	98.85%
4. 11-15	487	72.26%	749	55.65%	1,236	371	71.76%	78	65.00%	449	36.33%	116	73.89%	671	54.73%	787	63.67%
5. 16-19	153	22.70%	109	8.10%	262	142	27.47%	42	35.00%	184	70.23%	11	7.01%	67	5.46%	78	29.77%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
First Term Total Credit Hours Earned																	
0		0.00%	256	19.02%	256		0.00%		0.00%		0.00%		0.00%	256	20.88%	256	100.00%
1.00-6.00	10	1.48%	228	16.94%	238	2	0.39%		0.00%	2	0.84%	8	5.10%	228	18.60%	236	99.16%
7.00-11.00	44	6.53%	247	18.35%	291	4	0.77%	2	1.67%	6	2.06%	40	25.48%	245	19.98%	285	97.94%
12.00-16.00	525	77.89%	543	40.34%	1,068	427	82.59%	91	75.83%	518	48.50%	98	62.42%	452	36.87%	550	51.50%
17.00-21.00	72	10.68%	57	4.23%	129	65	12.57%	22	18.33%	87	67.44%	7	4.46%	35	2.85%	42	32.56%
22.00-26.00	8	1.19%	2	0.15%	10	7	1.35%	1	0.83%	8	80.00%	1	0.64%	1	0.08%	2	20.00%
27.00-31.00	8	1.19%	3	0.22%	11	7	1.35%	3	2.50%	10	90.91%	1	0.64%		0.00%	1	9.09%
32.00-over	7	1.04%	10	0.74%	17	5	0.97%	1	0.83%	6	35.29%	2	1.27%	9	0.73%	11	64.71%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
First Term Cumulative Quality Points																	
1. 0		0.00%	273	20.28%	273		0.00%		0.00%		0.00%		0.00%	273	22.27%	273	100.00%
2. 1-12	5	0.74%	190	14.12%	195		0.00%		0.00%		0.00%	5	3.18%	190	15.50%	195	100.00%
3. 13-24	50	7.42%	253	18.80%	303	3	0.58%		0.00%	3	0.99%	47	29.94%	253	20.64%	300	99.01%
4. 25-36	107	15.88%	279	20.73%	386	30	5.80%	3	2.50%	33	8.55%	77	49.04%	276	22.51%	353	91.45%
5. 37-48	240	35.61%	231	17.16%	471	216	41.78%	44	36.67%	260	55.20%	24	15.29%	187	15.25%	211	44.80%
6. 49-60	216	32.05%	105	7.80%	321	212	41.01%	69	57.50%	281	87.54%	4	2.55%	36	2.94%	40	12.46%
7. 61-76	56	8.31%	15	1.11%	71	56	10.83%	4	3.33%	60	84.51%		0.00%	11	0.90%	11	15.49%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%

Table 1

	2001 First-Time Undergraduate Cohort					J48 Predicted to Earn a Bachelor Degree						J48 Predicted NOT to Earned a Bachelor Degree					
	Actually Earned Bachelor Degree					Actually Earned Bachelor Degree				Part of Cohort		Actually Earned Bachelor Degree				Part of Cohort	
	Yes		No		Cohort	Yes		No				Yes		No			
	#	%	#	%	#	#	%	#	%	#	%	#	%	#	%	#	%
First Term GPA Ranges																	
1. Below 1.0		0.00%	356	26.45%	356		0.00%		0.00%		0.00%		0.00%	356	29.04%	356	100.00%
2. 1.00-1.99	37	5.49%	236	17.53%	273	2	0.39%		0.00%	2	0.73%	35	22.29%	236	19.25%	271	99.27%
3. 2.00-2.49	66	9.79%	225	16.72%	291	18	3.48%	3	2.50%	21	7.22%	48	30.57%	222	18.11%	270	92.78%
4. 2.50-2.99	121	17.95%	189	14.04%	310	90	17.41%	16	13.33%	106	34.19%	31	19.75%	173	14.11%	204	65.81%
5. 3.00-3.24	95	14.09%	137	10.18%	232	73	14.12%	28	23.33%	101	43.53%	22	14.01%	109	8.89%	131	56.47%
6. 3.25-3.49	107	15.88%	78	5.79%	185	96	18.57%	27	22.50%	123	66.49%	11	7.01%	51	4.16%	62	33.51%
7. 3.50-3.74	88	13.06%	58	4.31%	146	83	16.05%	22	18.33%	105	71.92%	5	3.18%	36	2.94%	41	28.08%
8. 3.75 and higher	160	23.74%	67	4.98%	227	155	29.98%	24	20.00%	179	78.85%	5	3.18%	43	3.51%	48	21.15%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Any Remediation																	
Yes	311	46.14%	804	59.73%	1,115	203	39.26%	53	44.17%	256	22.96%	108	68.79%	751	61.26%	859	77.04%
No	363	53.86%	542	40.27%	905	314	60.74%	67	55.83%	381	42.10%	49	31.21%	475	38.74%	524	57.90%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Remedial English																	
Did_not_take	455	67.51%	820	60.92%	1,275	382	73.89%	88	73.33%	470	36.86%	73	46.50%	732	59.71%	805	63.14%
Failed		0.00%	103	7.65%	103		0.00%	1	0.83%	1	0.97%		0.00%	102	8.32%	102	99.03%
Passed	219	32.49%	423	31.43%	642	135	26.11%	31	25.83%	166	25.86%	84	53.50%	392	31.97%	476	74.14%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Remedial Mathematics																	
Did_not_take	480	71.22%	791	58.77%	1,271	393	76.02%	86	71.67%	479	37.69%	87	55.41%	705	57.50%	792	62.31%
Failed	5	0.74%	213	15.82%	218		0.00%	1	0.83%	1	0.46%	5	3.18%	212	17.29%	217	99.54%
Passed	189	28.04%	342	25.41%	531	124	23.98%	33	27.50%	157	29.57%	65	41.40%	309	25.20%	374	70.43%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Reading & Study Skills Course Work																	
Did_not_take	557	82.64%	1,031	76.60%	1,588	449	86.85%	102	85.00%	551	34.70%	108	68.79%	929	75.77%	1,037	65.30%
Failed		0.00%	51	3.79%	51		0.00%		0.00%		0.00%		0.00%	51	4.16%	51	100.00%
Passed	117	17.36%	264	19.61%	381	68	13.15%	18	15.00%	86	22.57%	49	31.21%	246	20.07%	295	77.43%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Center for Student Profess # of Visits																	
None	499	74.04%	1,213	90.12%	1,712	380	73.50%	94	78.33%	474	27.69%	119	75.80%	1,119	91.27%	1,238	72.31%
1. 1-5	136	20.18%	110	8.17%	246	106	20.50%	22	18.33%	128	52.03%	30	19.11%	88	7.18%	118	47.97%
2. 6-10	21	3.12%	10	0.74%	31	17	3.29%	2	1.67%	19	61.29%	4	2.55%	8	0.65%	12	38.71%
3. 11-15	14	2.08%	8	0.59%	22	10	1.93%	2	1.67%	12	54.55%	4	2.55%	6	0.49%	10	45.45%
4. 16-20	4	0.59%	2	0.15%	6	4	0.77%		0.00%	4	66.67%		0.00%	2	0.16%	2	33.33%
5. 21+		0.00%	3	0.22%	3		0.00%		0.00%		0.00%		0.00%	3	0.24%	3	100.00%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Earned an Associate Degree																	
Yes	25	3.71%	48	3.57%	73	10	1.93%	1	0.83%	11	15.07%	15	9.55%	47	3.83%	62	84.93%
No	649	96.29%	1,298	96.43%	1,947	507	98.07%	119	99.17%	626	32.15%	142	90.45%	1,179	96.17%	1,321	67.85%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%

Table 1

	2001 First-Time Undergraduate Cohort					J48 Predicted to Earn a Bachelor Degree						J48 Predicted NOT to Earned a Bachelor Degree					
	Actually Earned Bachelor Degree					Actually Earned Bachelor Degree				Part of Cohort		Actually Earned Bachelor Degree				Part of Cohort	
	Yes		No		Cohort	Yes		No				Yes		No			
	#	%	#	%	#	#	%	#	%	#	%	#	%	#	%	#	%
Continued to Following Spring Term																	
Yes	669	99.26%	1,046	77.71%	1,715	514	99.42%	117	97.50%	631	36.79%	155	98.73%	929	75.77%	1,084	63.21%
No	5	0.74%	300	22.29%	305	3	0.58%	3	2.50%	6	1.97%	2	1.27%	297	24.23%	299	98.03%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Consecutively Enrolled (Fall, Spring, Fall)																	
Yes	654	97.03%	705	52.38%	1,359	507	98.07%	117	97.50%	624	45.92%	147	93.63%	588	47.96%	735	54.08%
No	20	2.97%	641	47.62%	661	10	1.93%	3	2.50%	13	1.97%	10	6.37%	638	52.04%	648	98.03%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%
Returned Next Fall																	
Yes	656	97.33%	733	54.46%	1,389	509	98.45%	120	100.00%	629	45.28%	147	93.63%	613	50.00%	760	54.72%
No	18	2.67%	613	45.54%	631	8	1.55%		0.00%	8	1.27%	10	6.37%	613	50.00%	623	98.73%
Grand Total	674	100.00%	1,346	100.00%	2,020	517	100.00%	120	100.00%	637	31.53%	157	100.00%	1,226	100.00%	1,383	68.47%

Table 1