Paper 031-2012

# Using SAS® Enterprise BI and SAS® Enterprise Miner™ to Reduce Student Attrition

Matt Bogard, Western Kentucky University, Bowling Green, KY
Chris James, Western Kentucky University, Bowling Green, KY
Tuesdi Helbig, Ph.D., Western Kentucky University, Bowling Green, KY
Gina Huff, Western Kentucky University, Bowling Green, KY

## ABSTRACT

The true supremacy of the SAS® Enterprise Business Intelligence Server is the ability to utilize the power of SAS® Analytics to deliver real-time information to end users, who usually do not understand statistics, but have the ability to make a difference if they have easy access to the analyzed data. This paper describes the process of using SAS® Enterprise Miner to develop a model to score university students based on their risk of attrition and deliver easy-to-understand results to university personnel using SAS® EBI.

## INTRODUCTION

For over 17 years, the staff of the Office of Institutional Research (IR) at Western Kentucky University (WKU) has used SAS® for data analysis and reporting. Charged with developing a first-time, first-year student retention model, IR assembled an eclectic group of data points associated with previous academic endeavors, standardized test scores, financial aid, and preliminary student success initiatives. After a period of evaluating and testing IR's empirical models in SAS® Enterprise Miner, Base SAS®, and SAS/STAT®, the IR staff--along with key university administrators--finalized the predictive model. To distribute the results of the selected model dynamically to the WKU community, IR staff used SAS® 9.2 Enterprise Business Intelligence (EBI) Server to seamlessly integrate the statistical model results into dynamic dashboards and reports for use by university personnel.

## ISSUE TO ADDRESS

One long-standing issue in institutional research is assisting institutions in their efforts to improve the retention and graduation rates of their students (McLaughlin, Brozovsky & McLaughlin, 1998; Pascarella, 1982). Federal and state emphasis on graduating students in less time to improve the state and federal governments' return on investment has only increased the pressure for institutions to identify factors that may hinder graduation and intervene to address the problems. There are numerous issues with traditional retention and graduation studies. First, the models developed may explain little of the variance in retention and graduation rates. Second, the often-used logistic regression models are hard to explain to administrators and others who can actually intervene. Third, even if the models can be explained, researchers do not go on to identify the attrition risk for individual students. Fourth, getting the risk information to the proper individuals for intervention is often cumbersome. SAS® Enterprise Business Intelligence Server and SAS® Enterprise Miner provide an integrated platform for addressing these issues.

## MODELING STUDENT RETENTION: A SUMMARY OF EMPIRICAL METHODS FROM LITERATURE

The vast majority of the literature related to the empirical estimation of retention models includes a discussion of the theoretical retention framework established by authors such as Bean (1980), Braxton (2000), Braxton, Hirschy, & McClendon (2004), Chapman & Pascarella (1983), Pascarella &Terenzini (1978), Tinto (1975), Miller & Herreid (2008) and Dey & Astin (1993). Literature indicates that data mining or algorithmic approaches to prediction can provide superior results vis-à-vis traditional statistical modeling approaches (Delen, Walker, & Kadam, 2004; Delen, Sharda, & Kumar, 2007; Kiang, 2003; Li, Nsofor, & Song, 2009). However, little research in higher education has focused on the employment of data mining methods for predicting retention (Herzog, 2006). Similar to Herzog, this paper adds to the literature related to the application of data mining methods in predicting student retention, with an emphasis on model implementation and deployment using SAS® EBI and analytics tools.

## DATA AND VARIABLES

Three models were developed using data available upon the first day of the 1st term (pre-enrollment), the 5th week of enrollment, and full semester enrollment. The full semester model included all of the variables in the previous models, including some variables that were not available until the end of the first term. Similarly, the 5th week model included all variables in the pre-enrollment model, in addition to data not available until the 5th week.

## METHODS AND SAS® ENTERPRISE MINER SETTINGS

Three years of data for first-time, first-year degree-seeking students were partitioned into training (80%) and validation (20%) subsets and used to implement logistic regression, decision trees, neural networks, and ensemble models using SAS® Enterprise Miner. For logit and neural network models, missing data was imputed utilizing SAS® Enterprise Miner’s tree imputation. No imputation was implemented for the decision tree models. The EM diagram can be seen in Figure 1.

## LOGISTIC REGRESSION

Logistic regression is a popular tool used for retention modeling (Dey & Astin, 1993; Herzog, 2006; Miller & Herried, 2008) as it conveniently provides a formula for deriving predicted probabilities related to student behavior. Stepwise logistic regression optimized based on validation misclassification was used for pre-enrollment and 5th week variables, while the variable selection tool using default settings was used to select inputs for the full semester logistic model.

## NEURAL NETWORK

SAS® Enterprise Miner’s default settings for a multilayer perceptron architecture were utilized in all three time periods. SAS® Enterprise Miner also provides an automated model selection process through the ‘AutoNeural’ node, which conducts limited searches for optimizing network configurations based on some initial user settings. In this project, for all three time periods, we specified a single hidden layer architecture, which added hidden nodes one at a time using a variety of activation functions. Although some attempts have been made to quantify variable importance and selection in the context of neural networks (Gevrey, Dimopoulos & Lek, 2003), SAS Enterprise Miner does not directly accommodate variable selection for neural networks. Inputs selected via logistic regression using stepwise selection were utilized by the neural networks using pre-enrollment and 5th week data. When utilizing full semester data, inputs were selected by decision trees.

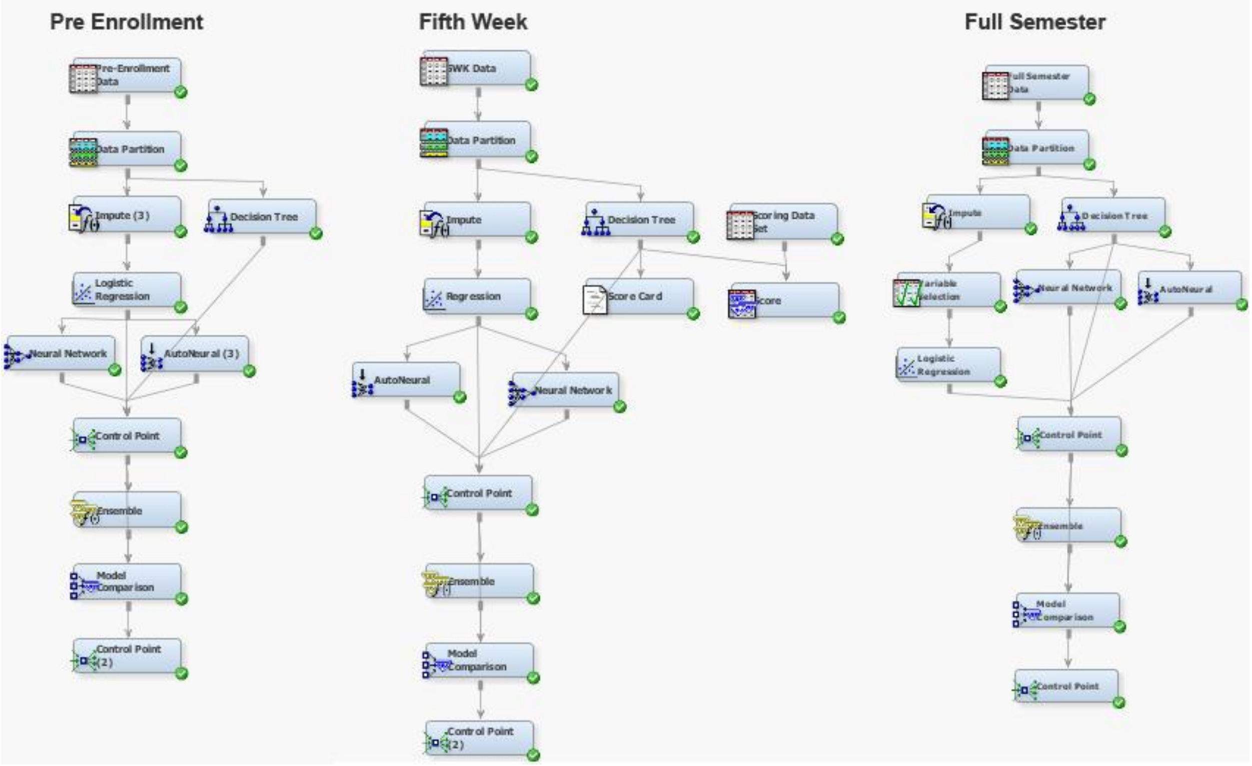
## DECISION TREE

One advantage of decision trees is that they have mechanisms for dealing with missing data as part of the split search algorithm utilized by SAS® Enterprise Miner. Decision trees were implemented autonomously in all cases using the default settings in SAS® Enterprise Miner, but optimized based on validation error.

## ENSEMBLE MODEL

An ensemble model can be thought of as a collection of a number of predictors (models) (Krogh & Sollich, 1997). Using SAS® Enterprise Miner, we implemented an ensemble model consisting of the logit, neural network, and decision tree models using the default average method.

**Figure 1 - SAS® Enterprise Miner Process Flow**



## RESULTS

With the primary goal of predicting attrition outcomes and minimizing generalization error, we produced metrics such as misclassification and percentage of correct predictions for each model. For each time period, the order from best to worst (in terms of percentage of correct predictions for the prediction of attrition) for each model was as follows:

| | |
|---|---|
| **Pre-Enrollment** | 1) autoneural 2) neural network 3) ensemble 4) logistic regression 5) decision tree |
| **5th Week** | 1) decision tree 2) autoneural 3) ensemble 4) logistic regression 5) neural network network |
| **Full Semester** | 1) ensemble 2) neural network 3) autoneural 4) logistic regression 5) decision tree |

As shown in Table 1, the practical differences between the performance of each model within each time period was small. Decision trees were preferred, based on reasoning similar to Delen (2010), who suggested that decision trees portray a more transparent model structure and explicitly illustrate the logical process associated with outcomes as opposed to neural networks or ensemble models. SAS® Enterprise Miner also provides English rules associated with decision trees, which are easier to explain to administrators and other nontechnical staff who may rely on the model results. For these reasons, we chose to move forward with decision trees as our champion model for implementation.

**Table 1 – Model Evaluation**

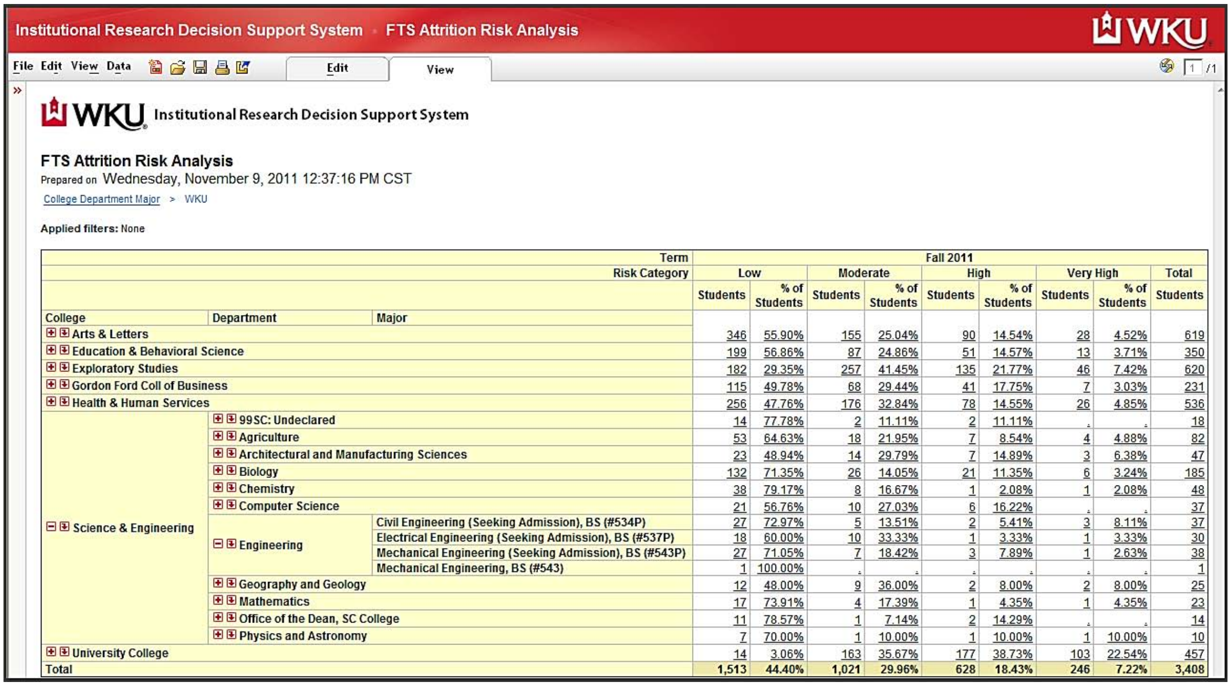
| Variables | Model | Overall Misclassification | Overall | ROC Index | Precision | Recall |
|---|---|---|---|---|---|---|
| | | | % Correct | | % Correct (Not Retained) | % Captured (Not Retained) |
| Pre Enrollment | Logit | 0.29 | 0.71 | 0.716 | 57.2954 | 26.3072 |
| | Neural Network | 0.29 | 0.71 | 0.716 | 58.9286 | 26.9608 |
| | AutoNeural | 0.28 | 0.72 | 0.715 | 59.8662 | 29.2484 |
| | Decision Tree | 0.30 | 0.70 | 0.681 | 53.4014 | 25.6536 |
| | Ensemble | 0.29 | 0.71 | 0.719 | 58.2090 | 25.4902 |
| 5th Week | Logit | 0.25 | 0.75 | 0.764 | 66.0274 | 39.3791 |
| | Neural Network | 0.26 | 0.74 | 0.759 | 64.9171 | 38.3987 |
| | Autoneural | 0.25 | 0.75 | 0.764 | 67.1348 | 39.0523 |
| | Decision Tree | 0.27 | 0.73 | 0.712 | 69.1244 | 24.5098 |
| | Ensemble | 0.25 | 0.75 | 0.764 | 66.1850 | 37.4183 |
| Full Semester | Logit | 0.21 | 0.79 | 0.813 | 75.2427 | 50.6536 |
| | Neural Network | 0.21 | 0.79 | 0.812 | 77.0950 | 45.098 |
| | AutoNeural | 0.21 | 0.79 | 0.808 | 76.7123 | 45.7516 |
| | Decision Tree | 0.21 | 0.79 | 0.801 | 75.1185 | 51.7974 |
| | Ensemble | 0.20 | 0.80 | 0.818 | 77.5000 | 50.6536 |

## MODEL IMPLEMENTATION, REPORTING, AND VISUALIZATION

Model selection and implementation into SAS® EBI is seasonal. The initial pre-enrollment attrition model did not explain the variance as well as the 5th Week model nor did it predict attrition as accurately. Because of the unavailability of some data and lower overall predictive accuracy, we decided not to utilize the pre-enrollment results for model deployment, but began our implementation with the 5th week decision tree model. After the fall semester concludes, the full semester ensemble model will be utilized.

Using the score code generated by SAS® Enterprise Miner, and defining formats using Base SAS®, each student's risk of attrition was classified into risk categories - i.e., low='less than 30%', moderate='between 30% and 36%', high='between 37% and 64%', very high='greater than 64%'. To dynamically implement the program generator and score code into the SAS® EBI production environment, an ETL was created in SAS® Data Integration Studio including both components. This ETL was submitted in batch every Monday morning before the open of business allowing the updated list of students to be scored for the week. The SAS® data set generated by the ETL was used to create an OLAP cube with student level information. This OLAP cube consisted of variables necessary for the statistical model and other demographic and academic variables useful to the WKU user community. Most importantly, variables related to the model were transformed into understandable, statistically simple variables that everyone involved in the retention effort could utilize. This allowed administrators, faculty, and professional staff at the institution to easily incorporate advanced analytics into their retention strategy on regular basis, without having to rely on manually generated lists of at-risk students, or less precise ad hoc reports generated solely on the basis of intuition. Figure 2 shows the default view of this report.

**Figure 2 – Attrition Risk Categories by College, Department, and Major**


Institutional Research Decision Support System · FTS Attrition Risk Analysis
WKU
File Edit View Data
Edit
View
1 /1
WKU

Institutional Research Decision Support System

FTS Attrition Risk Analysis

Prepared on Wednesday, November 9, 2011 12:37:16 PM CST

College Department Major > WKU

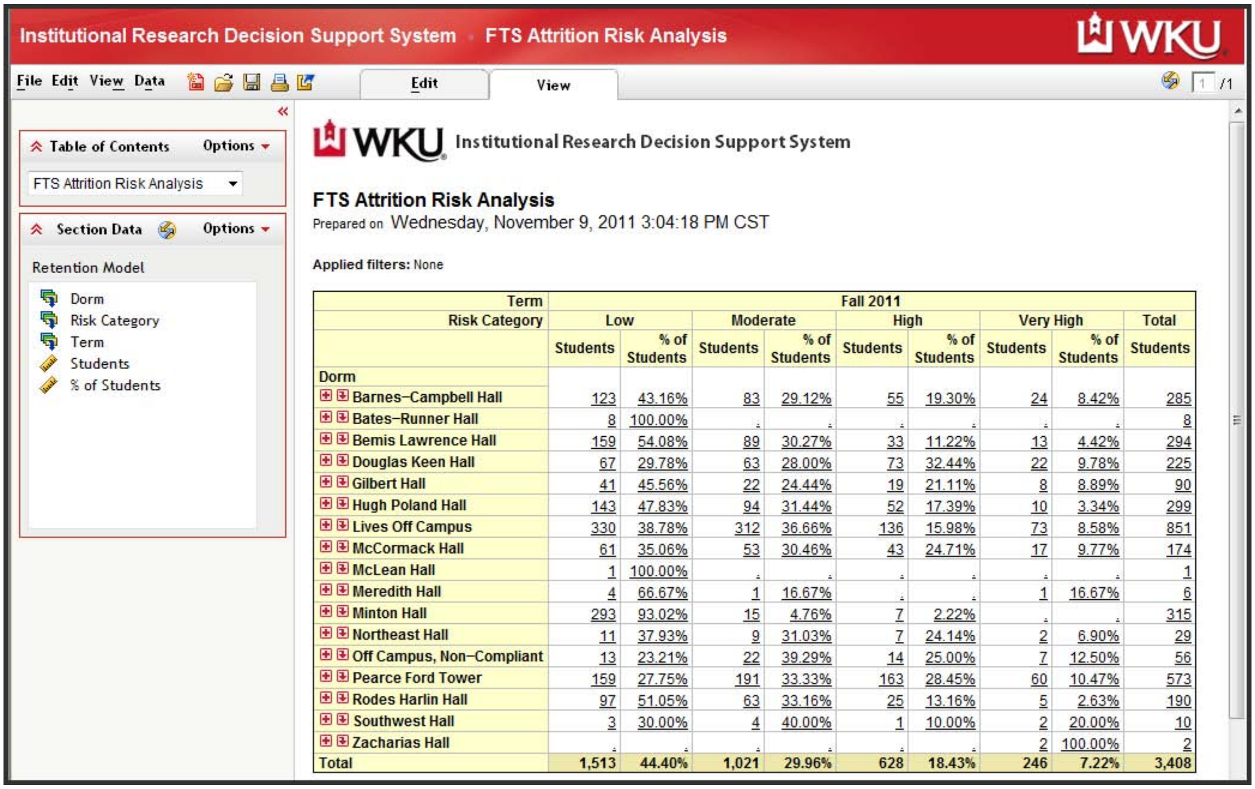
Applied filters: None

| | | Term | Fall 2011 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Risk Category | Low | | Moderate | | High | | Very High | | Total |
| | | | Students | % of Students | Students | % of Students | Students | % of Students | Students | % of Students | Students |
| College | Department | Major | | | | | | | | | |
| Arts & Letters | | | 346 | 55.90% | 155 | 25.04% | 90 | 14.54% | 28 | 4.52% | 619 |
| Education & Behavioral Science | | | 199 | 56.86% | 87 | 24.86% | 51 | 14.57% | 13 | 3.71% | 350 |
| Exploratory Studies | | | 182 | 29.35% | 257 | 41.45% | 135 | 21.77% | 46 | 7.42% | 620 |
| Gordon Ford Coll of Business | | | 115 | 49.78% | 68 | 29.44% | 41 | 17.75% | 7 | 3.03% | 231 |
| Health & Human Services | | | 256 | 47.76% | 176 | 32.84% | 78 | 14.55% | 26 | 4.85% | 536 |
| Science & Engineering | 99SC: Undeclared | | 14 | 77.78% | 2 | 11.11% | 2 | 11.11% | . | . | 18 |
| | Agriculture | | 53 | 64.63% | 18 | 21.95% | 7 | 8.54% | 4 | 4.88% | 82 |
| | Architectural and Manufacturing Sciences | | 23 | 48.94% | 14 | 29.79% | 7 | 14.89% | 3 | 6.38% | 47 |
| | Biology | | 132 | 71.35% | 26 | 14.05% | 21 | 11.35% | 6 | 3.24% | 185 |
| | Chemistry | | 38 | 79.17% | 8 | 16.67% | 1 | 2.08% | 1 | 2.08% | 48 |
| | Computer Science | | 21 | 56.76% | 10 | 27.03% | 6 | 16.22% | . | . | 37 |
| | Engineering | Civil Engineering (Seeking Admission), BS (#534P) | 27 | 72.97% | 5 | 13.51% | 2 | 5.41% | 3 | 8.11% | 37 |
| | | Electrical Engineering (Seeking Admission), BS (#537P) | 18 | 60.00% | 10 | 33.33% | 1 | 3.33% | 1 | 3.33% | 30 |
| | | Mechanical Engineering (Seeking Admission), BS (#543P) | 27 | 71.05% | 7 | 18.42% | 3 | 7.89% | 1 | 2.63% | 38 |
| | | Mechanical Engineering, BS (#543) | 1 | 100.00% | . | . | . | . | . | . | 1 |
| | Geography and Geology | | 12 | 48.00% | 9 | 36.00% | 2 | 8.00% | 2 | 8.00% | 25 |
| | Mathematics | | 17 | 73.91% | 4 | 17.39% | 1 | 4.35% | 1 | 4.35% | 23 |
| | Office of the Dean, SC College | | 11 | 78.57% | 1 | 7.14% | 2 | 14.29% | . | . | 14 |
| | Physics and Astronomy | | 7 | 70.00% | 1 | 10.00% | 1 | 10.00% | 1 | 10.00% | 10 |
| University College | | | 14 | 3.06% | 163 | 35.67% | 177 | 38.73% | 103 | 22.54% | 457 |
| Total | | | 1,513 | 44.40% | 1,021 | 29.96% | 628 | 18.43% | 246 | 7.22% | 3,408 |

## FLEXIBILITY OF OLAP

While the default view of this report is useful for academic units, other departments on campus can benefit by manipulating the report with a few clicks. Other hierarchies in the cube categorize the students based on other criteria, not related to academic unit. Because the cardinality of the underlying data is structured to represent each major the student is seeking, students with multiple majors would be counted multiple times in other views other than academic unit. The unique member count feature in SAS® OLAP Cube Studio circumvents this potential problem. The unique member count feature dynamically counts a selected hierarchy, comprised of a character variable; at whatever level the measure is totaled. We converted a unique student identification number into a character variable and developed a hierarchy named Student ID. This feature allowed users to take out the academic unit hierarchy and replace it with other hierarchies such as dorm, advisor, or origin, without duplication. Figure 3 shows the same report, but with a different view after a few changes to the report elements.

**Figure 3 – Attrition Risk Categories by Dorm, Floor, and Room Number**


Institutional Research Decision Support System
FTS Attrition Risk Analysis
WKU
File Edit View Data
Edit
View
1 /1
«
Table of Contents
Options
FTS Attrition Risk Analysis
Section Data
Options
Retention Model
Dorm
Risk Category
Term
Students
% of Students
WKU
Institutional Research Decision Support System

**FTS Attrition Risk Analysis**
Prepared on Wednesday, November 9, 2011 3:04:18 PM CST

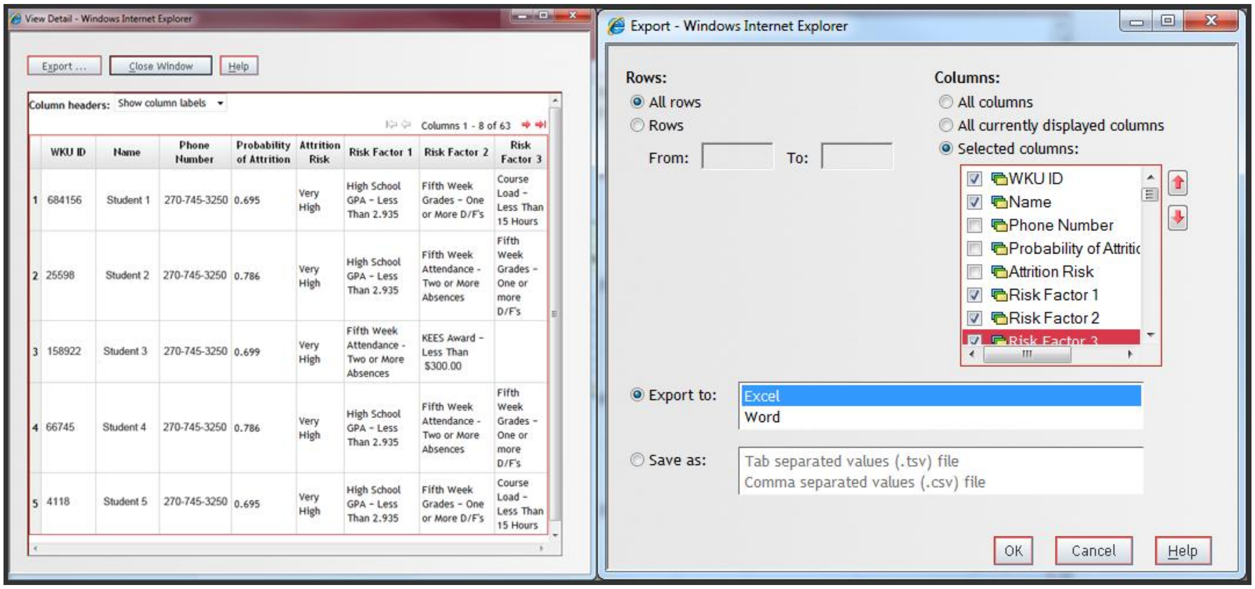
**Applied filters:** None

| Term | Fall 2011 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Risk Category | Low | | Moderate | | High | | Very High | | Total |
| | Students | % of Students | Students | % of Students | Students | % of Students | Students | % of Students | Students |
| Dorm | | | | | | | | | |
| Barnes–Campbell Hall | 123 | 43.16% | 83 | 29.12% | 55 | 19.30% | 24 | 8.42% | 285 |
| Bates–Runner Hall | 8 | 100.00% | . | . | . | . | . | . | 8 |
| Bemis Lawrence Hall | 159 | 54.08% | 89 | 30.27% | 33 | 11.22% | 13 | 4.42% | 294 |
| Douglas Keen Hall | 67 | 29.78% | 63 | 28.00% | 73 | 32.44% | 22 | 9.78% | 225 |
| Gilbert Hall | 41 | 45.56% | 22 | 24.44% | 19 | 21.11% | 8 | 8.89% | 90 |
| Hugh Poland Hall | 143 | 47.83% | 94 | 31.44% | 52 | 17.39% | 10 | 3.34% | 299 |
| Lives Off Campus | 330 | 38.78% | 312 | 36.66% | 136 | 15.98% | 73 | 8.58% | 851 |
| McCormack Hall | 61 | 35.06% | 53 | 30.46% | 43 | 24.71% | 17 | 9.77% | 174 |
| McLean Hall | 1 | 100.00% | . | . | . | . | . | . | 1 |
| Meredith Hall | 4 | 66.67% | 1 | 16.67% | . | . | 1 | 16.67% | 6 |
| Minton Hall | 293 | 93.02% | 15 | 4.76% | 7 | 2.22% | . | . | 315 |
| Northeast Hall | 11 | 37.93% | 9 | 31.03% | 7 | 24.14% | 2 | 6.90% | 29 |
| Off Campus, Non–Compliant | 13 | 23.21% | 22 | 39.29% | 14 | 25.00% | 7 | 12.50% | 56 |
| Pearce Ford Tower | 159 | 27.75% | 191 | 33.33% | 163 | 28.45% | 60 | 10.47% | 573 |
| Rodes Harlin Hall | 97 | 51.05% | 63 | 33.16% | 25 | 13.16% | 5 | 2.63% | 190 |
| Southwest Hall | 3 | 30.00% | 4 | 40.00% | 1 | 10.00% | 2 | 20.00% | 10 |
| Zacharias Hall | . | . | . | . | . | . | 2 | 100.00% | 2 |
| Total | 1,513 | 44.40% | 1,021 | 29.96% | 628 | 18.43% | 246 | 7.22% | 3,408 |

## DRILL-THROUGH TO DETAIL

Another powerful feature of any SAS® OLAP cube is the ability to drill-through to detail. This functionality allows these very reports to produce row level data with the click of a mouse. The detailed data of this report shows the risk indicators from the statistical model for the students selected. This detail data provides contact information for each student, as well as the student's probability of attrition, risk category, test scores, and other important data related to the student. Figure 4 shows the drill-through to detail table produced from clicking a number, or percent, on a report. Users can export this data to Microsoft® Excel, Word, or save the list as a comma or tab delimited file.

**Figure 4 – Drill-Through to Detail Table and Export Tool**



View Detail - Windows Internet Explorer

Column headers: Show column labels

Columns 1 - 8 of 63

| | WKU ID | Name | Phone Number | Probability of Attrition | Attrition Risk | Risk Factor 1 | Risk Factor 2 | Risk Factor 3 |
|---|---|---|---|---|---|---|---|---|
| 1 | 684156 | Student 1 | 270-745-3250 | 0.695 | Very High | High School GPA - Less Than 2.935 | Fifth Week Grades - One or More D/F's | Course Load - Less Than 15 Hours |
| 2 | 25598 | Student 2 | 270-745-3250 | 0.786 | Very High | High School GPA - Less Than 2.935 | Fifth Week Attendance - Two or More Absences | Fifth Week Grades - One or more D/F's |
| 3 | 158922 | Student 3 | 270-745-3250 | 0.699 | Very High | Fifth Week Attendance - Two or More Absences | KEES Award - Less Than $300.00 | |
| 4 | 66745 | Student 4 | 270-745-3250 | 0.786 | Very High | High School GPA - Less Than 2.935 | Fifth Week Attendance - Two or More Absences | Fifth Week Grades - One or more D/F's |
| 5 | 4118 | Student 5 | 270-745-3250 | 0.695 | Very High | High School GPA - Less Than 2.935 | Fifth Week Grades - One or More D/F's | Course Load - Less Than 15 Hours |

Export - Windows Internet Explorer

Rows:
● All rows
○ Rows
From: To:

Columns:
○ All columns
○ All currently displayed columns
● Selected columns:
☑ WKU ID
☑ Name
□ Phone Number
□ Probability of Attritic
□ Attrition Risk
☑ Risk Factor 1
☑ Risk Factor 2
☑ Risk Factor 3

● Export to: Excel
Word

○ Save as: Tab separated values (.tsv) file
Comma separated values (.csv) file

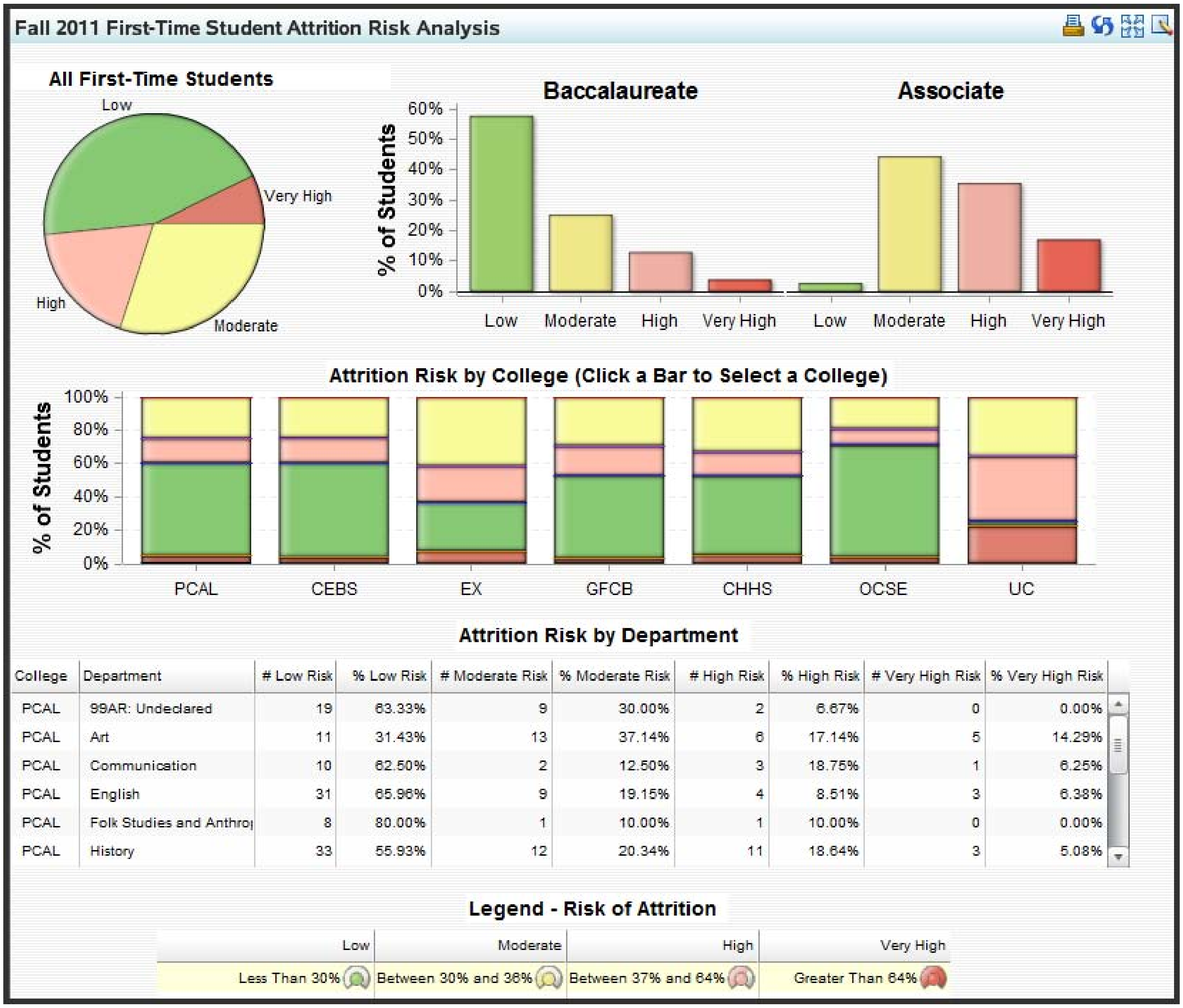
## MODEL VISUALIZATION

The final stage of this project involved dashboard creation and implementation. Displaying data in a tabular table isn't always the best way to investigate trends. Data visualization provides an at-a-glance view of the bigger picture. Dashboards provide an elegant platform for displaying aggregated data quickly and precisely. They excel in their capability to emphasize successful and/or problematic trends without having to churn through mounds of reports. Dashboards also allow critical indicators to supersede myopic details.

The SAS® BI Dashboard provided our staff the capability and flexibility to display the attrition model data in a user-friendly, visual environment. We initially hoped to display charts and graphs that would tell the story beginning with the high-level view and ending with a more precise depiction. We contemplated the number of indicators that would exist on the dashboard for fear that too many could be overwhelming for users. Interactions were set up between each indicator to dynamically show how one indicator related to another on the dashboard.

A particularly vibrant interaction was developed between the stacked bar chart - representing risk categories by college - and the departmental spark table. This interaction allows the stacked bar chart to feed the selected college to the department spark table, which consequently filters the table to departments in the selected college. This special feature allows users to quickly exam college and department information all on one interactive dashboard. As you can see in Figure 5, specific colors were selected to represent different levels of risk of attrition - i.e., low=green', moderate=yellow', high='pink', very high=red'. This coloring scheme quickly depicts the trends that exist with our students.

**Figure 5 – Attrition Model Dashboard**



Attrition Risk by Department

| College | Department | # Low Risk | % Low Risk | # Moderate Risk | % Moderate Risk | # High Risk | % High Risk | # Very High Risk | % Very High Risk |
|---|---|---|---|---|---|---|---|---|---|
| PCAL | 99AR: Undeclared | 19 | 63.33% | 9 | 30.00% | 2 | 6.67% | 0 | 0.00% |
| PCAL | Art | 11 | 31.43% | 13 | 37.14% | 6 | 17.14% | 5 | 14.29% |
| PCAL | Communication | 10 | 62.50% | 2 | 12.50% | 3 | 18.75% | 1 | 6.25% |
| PCAL | English | 31 | 65.96% | 9 | 19.15% | 4 | 8.51% | 3 | 6.38% |
| PCAL | Folk Studies and Anthro| | 8 | 80.00% | 1 | 10.00% | 1 | 10.00% | 0 | 0.00% |
| PCAL | History | 33 | 55.93% | 12 | 20.34% | 11 | 18.64% | 3 | 5.08% |

Legend - Risk of Attrition

| Low | Moderate | High | Very High |
|---|---|---|---|
| Less Than 30% | Between 30% and 36% | Between 37% and 64% | Greater Than 64% |

## CONCLUSION

IR first launched SAS® EBI at WKU a year prior to the development and deployment of the attrition risk model. At that point we visited each college to show deans and department heads how to use the system. As expected, some administrators used the system daily, while others rarely logged on. Once the results of the attrition risk model were available via SAS® EBI, we took the time to meet with deans and department heads again to explain our model and how to interpret and use the dashboards and reports. With WKU's increased focus on retention, we have seen a more frequent use of data for decision making and policy analysis. While we cannot assess whether or not the model, deployment of results via EBI, or resulting interventions made a difference on student attrition until fall 2012, we have been able to realize one of our goals by turning analytics into action.

## REFERENCES

Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education*, *12*(2), 155–187.

Braxton, J. M. (Ed.) (2000). *Reworking the student departure puzzle*. Nashville, TN: Vanderbilt University Press.

Braxton, J. M., Hirschy, A. S., & McClendon, S. A. (2004). *Understanding and reducing college student departure*. (ASHE-ERIC Higher Education Report No. 30.3). San Francisco: Jossey-Bass.

Chapman, D. & Pascarella, E. (1983). Predictors of academic and social integration of college students. *Research in Higher Education*, *19*, 295-322.

Delen, D., Walker, G. & Kadam, A. (2004). Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial Intelligence in Medicine 34*(2) 113–127.

Delen, D., Sharda, R., & Kumar, P. (2007). Movie forecast guru: a web-based DSS for Hollywood managers, *Decision Support Systems 43*(4) 1151–1170.

Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Support Systems 49*, 498–506.

Dey, E. L. & Astin, A. W. (1993). Statistical alternatives for studying college student retention: A comparative analysis of logit, probit, and linear regression. *Research in Higher Education*, *34*(5).

Gevrey, M., Dimopoulos, I., & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Model., *160*, 249-264.

Herzog , S. (2006). Estimating student retention and degree-completion time: Decision trees and neural networks vis-à-vis regression. *New Directions for Institutional Research, 131*.

Kiang, M.Y. (2003). A comparative assessment of classification algorithms, *Decision Support Systems, 35*, pp. 441–454.

Krogh, A. & Sollich, P. (1997, January). Statistical mechanics of ensemble learning. *Physical Review E (Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics)*, *55* (1), 811-825.

Li, X., Nsofor, G.C., Song, L. (2009). A comparative analysis of predictive data mining techniques. *International Journal of Rapid Manufacturing 1* (2) 150–172.

McLaughlin, G. W., Brozovsky, P. V., & McLaughlin, J. S. (1998, February). Changing perspectives on student retention: A role for Institutional Research. *Research in Higher Education*, *39*(1), 1-17. Retrieved October 31, 2011, from http://www.jstor.org/stable/40196265.

Miller, T.E. & Herreid, C.H. (2008). Analysis of variables to predict first year persistence using logistic regression analysis at the University of South Florida. *College & University, 83*(3).

Pascarella, E.T. & Terenzini, P.T. (1978). The relation of students' precollege characteristics and freshman year experience to voluntary attrition. *Research in Higher Education, 9*, 347-366.

Pascarella, E. (1982). Studying student attrition. *New Directions for Institutional Research*. San Francisco: Jossey-Bass.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, *45*(1), 89–125.

## CONTACT INFORMATION

**Matt Bogard**
**Institutional Research, WKU**
**1906 College Heights Blvd. #11011**
**Bowling Green, KY 42101-1011**
**Work Phone: (270) 745-3250**
**Fax: (270) 745-5442**
**Email: Matt.Bogard@wku.edu**

sas Certified Predictive Modeler Using SAS® Enterprise Miner™ 6

**Chris James**
**Institutional Research, WKU**
**1906 College Heights Blvd. #11011**
**Bowling Green, KY 42101-1011**
**Work Phone: (270) 745-3250**
**Fax: (270) 745-5442**
**Email: Christopher.James@wku.edu**

sas Certified Advanced Programmer for SAS® 9

**Tuesdi Helbig, Ph.D.**
**Institutional Research, WKU**
**1906 College Heights Blvd. #11011**
**Bowling Green, KY 42101-1011**
**Work Phone: (270) 745-3250**
**Fax: (270) 745-5442**
**Email: Tuesdi.Helbig@wku.edu**

**Gina Huff**
**Institutional Research, WKU**
**1906 College Heights Blvd. #11011**
**Bowling Green, KY 42101-1011**
**Work Phone: (270) 745-3250**
**Fax: (270) 745-5442**
**Email: Gina.Huff@wku.edu**

sas Certified Advanced Programmer for SAS® 9