

To Buy or to Create? Determining the Best Option for Student Success Models

Stephanie R. Thompson, Datamum / Rochester Institute of Technology

ABSTRACT

The number of vendors providing solutions for predicting student success keeps increasing. Each vendor claims to be able to use your data and to predict the likelihood of retention or graduation for you. This all comes with a financial cost and required time and effort to prepare and update data to feed their systems. Is creating an in-house system a better solution? This paper walks through one way to make that determination. If your institution has gone through changes impacting the continuity of data or has made other significant structural changes, finding an off-the-shelf solution can be more challenging. Should you include additional variables that the commercial products do not consider? A side benefit of creating the model in-house is validating multiple models on past students as well as working with those who will be using the model output to generate buy-in. Rolling out a commercial solution is of no benefit if those using it do not trust the results. SAS® has the tools to do the modeling whether you use Base SAS®, SAS® Enterprise Miner™, or SAS® Visual Analytics. Why not leverage what you have to support your students' success?

INTRODUCTION

Student success is becoming big business. There are now a wide variety of companies purporting to have a system to evaluate your students and identify those at risk of leaving your institution. With feeds of data from you, they will generate scores and profiles to be fed back to advisors to help them focus on the students in most need. They claim it is all seamless and easy on the part of the institution. However, this all comes at a significant cost. The cost is not only financial but also in the time required to generate the data feeds.

Most might agree that a student who is struggling academically is at risk of leaving. Low GPA, withdrawals from courses, and the failure of foundation courses required for a major are all signs of risk. Do we really need a model for those students? Probably not. However, there are other students who appear to be progressing but who do not stay until degree completion. The reasons they leave may be less intuitive and models may be helpful for these students. The students who have high GPAs and are on track for on time graduation that leave are the biggest mysteries. There seems to be no way to predict these students are at risk to leave. A successful model here would be a huge benefit to the institution.

Maybe the off the shelf solution can help your institution and are worth the expense. However, there are some situations when these may prove less effective. Colleges with a large diversity of incoming students with larger attrition rates seem to perform well with the standard models. There is more success in differentiating the students and patterns leading to departure are more common. How will the model perform if you have a relatively homogenous entering student cohort and a low attrition rate? Will the model still be as effective?

LOOKING AT YOUR INSTITUTION

As noted there are certain features of your student population that guide you in terms of deciding on the best modeling approach. A few of these are:

- Variability of student population
- Retention and graduation rates
- Changes in the academic calendar
- Changes in the academic organizational structure

Each of these factors can make modeling more of a challenge. For example, if you have a wide range of standardized test scores, high school GPAs, and demographics in your incoming freshman cohort, modeling may be able to find the factors that differentiate those who leave and those who stay. If, on the other hand, your incoming students have standardized scores in a narrow band, have similar high school performance, and are not particularly diverse, the model may have trouble predicting attrition. In this second case the students who leave look a lot like those who stayed.

Another issue to consider is your retention and graduation rates. Models may perform differently on a student cohort with an 82% rate versus a 97% rate. The lower rate cohort has more students that can be used to define the characteristics of risk. The 97% cohort has many fewer students who leave and many that leave may look just like ones who stayed. This is particularly true for models based on entry characteristics and without a lot of academic performance at your institution. Once students start receiving grades at your institution and other factors can be added, like athletics and student organizations, the more different they begin to look.

Maybe your institution has changed its academic calendar. A transition from semesters to trimesters or quarters can make historical data less effective in predicting future performance. A school on quarters has more data points in terms of course grades than a school on semesters. Does one translate to the other? It is hard to assume either way. Changes in calendar make it less likely that a model dependent on a long horizon of historical data is going to have a good fit looking forward. A model for one term retention may be successful but looking at building a 150% time to graduation model may be less so.

Changes in academic structure, similar to calendar changes, make it harder to translate historical data to the current structure. If majors move from one college to another, they may appear different to the model when in fact it should not. This might involve additional work on the data preparation to remove the appearance of the difference. It might be as simple as making sure the designation of the major remains consistent. It may also mean ignoring the variable of college in the model.

Changes in course requirements are much harder to capture and adjust for. Examples of this include requiring a grade of C versus D to move on to the next course in a sequence. Maybe an additional math requirement is added. There are also examples of where the total number of credit hours required to graduate increases or decreases. These factors are not easily adjusted for in model data preparation. It may be that they need to be recognized and understood within the process. Lowering total credit hours may have a one or two-year impact on improving graduation rate, but you should not consider it a long-term change or based on a factor other than the change in requirements.

A LOOK AT THE DATA

Now that the overall factors have been considered, it is time to look at the specific variables to be used in the model. Each of the commercial companies have their own list of variables but many do overlap. These include many of the common student characteristics and academic performance indicators. These should be variables that you use for your regular reporting and analyses. Even though the lists are similar and somewhat routine, there are still a few considerations that might impact your particular institution. These are reviewed below.

IS THE LIST COMPREHENSIVE?

Does the vendor list contain everything that you traditionally use to generate reports? It does seem that there is a common set of variables in the literature and commercial products. Review the list from the vendor to your reporting. Make sure you are comfortable with the list and that you understand any transformations that may be done.

Some institutions serve a large percentage of veterans and that may not be part of the model. Maybe there is some other aspect that you want to include such as specialized student housing, club memberships, athletics, and student interests. These factors may be important in predicting success for your school. It may be possible for the vendor include them or it may not. You will need to decide how important it is to include them.

On the other side, what will happen to the model if it includes a variable you do not collect? Will it still be able to generate a model? What happens to the scoring without the variable? This could become more of

a challenge as institutions become test optional or stop collecting some demographic information. Certain types of models can handle missing values but others cannot use a record with a missing value in generating the model. The more rows that are excluded the less robust the model may be.

WHAT ABOUT TRANSFORMATIONS?

High school GPA can be reported on a 100-, 4-, 5- or 6-point scale. Grades based on these different scales cannot be loaded as is into the variable for high school GPA in the model. They do not represent the same thing. Does the vendor handle this by adding a scale variable? If not, you may need to standardize the data before submission. If you, you also need to find out what scale the vendor requires. If you submit on a 4-point scale and a 100-point scale is anticipated, you will not get the right results.

Standardized test scores could be somewhat problematic. The ACT predominates in some areas of the country and the SAT in others. Some students take both exams. Does your institution super score the exams? If so, will that need to be done prior to submitting to the vendor. Are individual exam scores given to the vendor? Considering the changes in the SAT over recent years, will a concordance need to be generated? What about a concordance between ACT and SAT? These are important considerations.

WHEN ARE VARIABLES AVAILABLE?

The last consideration is data availability. Some models score students when they arrive on campus. Are all of those variables available for extraction from your system when needed by the vendor or your scoring program? Sometimes the data are collected but not entered until a later date.

Any variable that is not available for scoring needs to be evaluated to its importance. Maybe it is not critical for scoring at that time and only is used for models later on in a student's career. If it is an important variable, processes may need to be changed to have the data loaded sooner.

NUMBER AND TYPE OF MODELS

Once the variables have been reviewed, the next step is to look at the number of models that are run and what they are modeling. Will you run term to term or year to year models? Term to term may be more challenging since you have fewer data points available. Also, it is harder to score students in time to have the ability to make a meaningful intervention within their first term for example.

Will likelihood of retention or graduation be modeled? Maybe it is both. Will students be scored walking in the door, after the first term, throughout the term, or at some other point? Each model may have its own data requirements and frequency of pulling the data.

Increasing the number of models can increase the workload in data preparation. It may also be wise to determine if students will be included in multiple models at once or just one at a time. Evaluating a student for both retention and graduation at the same time could send conflicting scores to an advisor. Given that the significant factors in each model may be very different, a student could look at risk in one and not the other. Adding the layer of only modeling a student in one model at a time will add to the data pulling process but might be worth while in the end.

In looking at the vendor models, you can see that the models are not exceptionally complex and can be run easily in SAS. Depending on your SAS license, you can run the same type of models. There are many SAS products with modeling capabilities. Base SAS, SAS Enterprise Guide, SAS Enterprise Miner, and SAS Visual Analytics all have the ability to run models.

The best part about developing the models in house is that you can try and compare many types of models to find the best fit based in your criteria. Maybe it is logistic regression or a decision tree. You may find that neural networks or some other type of model works best on your data. Running them yourself allows you to review all of the model fit statistics. This can help you to fine tune your model to address the particular need you have. Vendors may not share your specific institution's model fit statistics with you. A general summary of fit for all of their clients may be provided but that does not confirm that their model is best for your institution.

All of these considerations are important in evaluating the workload you are willing and able to undertake in data preparation and modeling. This is true for both vendor products and in-house models.

VALIDATION

Validation is critical for any modeling project. It is a way to see how the model's predictions stack up to data where you know the outcome. It is where you will see how accurate the predictions are and where it is lacking or missing the mark altogether. When you run your own models, this is a transparent part of the work. It helps you see not only the overall accuracy of the models but also how it works for subpopulations. You can dig into the results in depth and see if there are groups that seem to be identified more or less accurately.

With the use of SAS, you are able to generate score code from your models and use it to see the predictions for your actual students. You can then integrate those results into reports or other summaries of the results of the models. Also, this can be compared to actual outcomes.

With access to the validation data you can dig into the confusion matrix and see, for example, are all of the false positives similar in any way. Do they share certain characteristics like GPA or college of major? Maybe there is something else that is similar. Running the validation yourself allows for this type of model evaluation.

Understanding how a vendor model is validated can be a little harder. They may state that there is a validation step and you may even need to provide the data. However, you more than likely will not be able to see the individual results of the validation. You may not even be given the scores for individual students from the validation step. This will make it a lot harder for you to determine if their model is performing to stated standards. Even seeing the confusion matrix does not give you enough information to truly evaluate performance.

Having access to the validation results will help you identify potential issues with the model and whether or not certain groups are more at risk of a score indicating attrition. Knowing these details can assist with building trust in the scores by those who will be using them. This will be discussed a little further on.

USE AND TIMING OF SCORING

Once the models are validated and put into production, the work of scoring begins. Scoring is running data through the results of a model to see what outcome it predicts or how likely it is that a student will leave the institution. Typically, scores are not at the extreme ends of the scale such as 100% likely to leave. Many scores are on a decimal scale that can be read as a percentage. The result of the scoring is dependent on the type of model is used.

Using the results of the validation, you can determine what the level of the score should be to alert an advisor that an intervention might be needed. If most of the students who left had scores over 80% then you may want to use that as a red flag zone for the system. The setting of this value will influence how many students are flagged for intervention. It is an art as well as a science.

Another consideration of scoring students is timing. The student data needs to be scored when the variables in the model are available. Scores will not be accurate when value are missing from the data. Results from partial data can be somewhat indicative of a result if it is on the high end of the scale. In those cases, the addition of variables may not lower the score significantly. However, the only way to know this is by looking at the score code to see the sign on the coefficient for the variables that are not available. It is possible that a missing value could swing the score from leaving to staying.

If the data needed to score students isn't all available until late in a term, that model will not be usable to identify students for intervention for term to term retention. A model looking at one-year retention might be a better option where you score students after one term and use the following term to work with the student. The ultimate goal of the model is to identify students for intervention. There needs to be enough time for the intervention to happen and have an impact on the student. This needs to be part of the considerations in the development of the model.

Some vendor models pull data on a daily basis. You need to consider how many of the input variables can actually change and how frequently they do. Standardized test scores and most demographics do not change once they are in the system. Financial aid need can change from year to year but is unlikely to change term to term. Variables like major and course withdrawals are examples of variables that could change within a term.

Understanding data availability and variability is important in understanding the scores regardless of the source of the models. Scoring daily or at set intervals will depend on these two factors. Also keep in mind how the scores will be shared and how often they are likely to be reviewed and acted upon. If the score changes daily but is only looked at once a term, is there a benefit to scoring that frequently?

OUTPUT FROM SCORING AND INTERFACE TO SYSTEMS

One of the last considerations with any model is how it will be shared and integrated into existing systems. The best result is to have the score, in terms of a risk scale, integrated into the existing system used by advisors. This is where a system from the same vendor can be appealing. They may have the mechanism to easily integrate this into a dashboard already used by advisors. These integrations may include the use of traffic light or dial style colored indicator that make interpretation of the scores very intuitive.

However, many systems allow for the addition of user specified information to be added to the dashboards. The data would need to reside in the same database as the other information. Many systems allow for the addition of user defined variables and the model scores can be fed to the database. In this case the addition of a dial or color-coded indicator may not be possible. Use of the score would rely more on training and understanding of the information.

Many institutions are creating their own dashboards and reports to supplement existing systems. These could be supported by an in-house data warehouse and analytics system. If you are unable to integrate the score into a supplied system, reports or dashboards designed for advisors can be created. This would add an additional step to their process of evaluating the needs of a student. Many institutions are designing visually appealing dashboards to convey information to information consumers. Designing the presentation of the scoring allows you to add other information deemed relevant as well as the ability to make it visually appealing. For example, you could add the factors that are relevant for each student in terms of risk for leaving based on your model.

BUILDING CONFIDENCE IN THE SCORE

Lastly, no modeling effort and score generation is of any value if the results are not trusted and acted on by those who should use them. This is where in-house modeling can have certain advantages. Developing the models yourself allows you to involve advisors and others in determining what variables should be included in the development of the models. They can be involved in the process and see the results of the validation. You can send them the scores for students they have worked with in the recent past and see if the model score corresponds to their impression of the student and their outcomes. You want to build confidence in the results so that they can be used to improve student outcomes. The true success in any modeling effort is the resulting action taken as a result of it and not in the creation of it.

Some may feel comfortable using a product from a vendor since it has some type of track record. Success at other institutions can build confidence in results and lead to broader adoption. However, it is important to make sure that the vendor's existing customers are similar to your institution. If most of the customers are community colleges or four-year mid-size public universities and you are a large, private, moderately selective school you need to determine if those results will translate. Are your student characteristics similar? How about retention and graduation rates? You need to review all of the factors mentioned earlier in this paper. Some models and systems are more effective on certain student populations than others. The vendor model will be more successful if your students and institution look similar to other clients.

Regardless of the choice of model, user buy in from the beginning is critical. Make sure that the users of the scores understand how they were generated and what the limitations to their accuracy are. No model is a crystal ball. Scores are just another tool that can be used to assist student achieve success and

graduation. It will not replace other tools nor the ability of the advisor to understand the issues impacting individual students that cannot be captured by any model.

CONCLUSION

The goal of this paper is to provide some considerations on developing student success models in-house using the SAS tools you likely already have. Many institutional research offices have the skill set to do this. A lot needs to be evaluated in terms of existing workload and institutional goals. In-house models allow you to best capture your students and build buy-in from the users of the model results. While the topics covered are not exhaustive, they provide a solid foundation to start you on this process.

Vendor models can be a good choice for institutions without in-house experience. They need to be evaluated in a comprehensive manner to make sure they will work for your student population. The cost could be offset by increases in student success.

Building your own models allows you to include things the vendors do not and to address concerns specific to your institution. You are in control of timing, data extraction, transformations, and what types of model to create. You are not confined to the options of a vendor.

Whichever path you choose, make sure you use the best available data and verify its accuracy. No model will be successful if the input is faulty. Remember, models are just one part of student success.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Stephanie R. Thompson
Datamum
901-326-0030
Stephanie@datamum.com
<http://www.datamum.com>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.