SAS Tools for Educational Data Mining

Jennifer Sabourin
SAS Institute
100 SAS Campus Dr.
Cary, NC 27513
1. 919.531.3313
Jennifer.Sabourin@sas.com

Scott McQuiggan
SAS Institute
100 SAS Campus Dr.
Cary, NC 27513
1. 919.531.1119
Scott.McQuiggan@sas.com

Andre de Waal
SAS Institute
100 SAS Campus Dr.
Cary, NC 27513
1. 919.531.6575
Andre.Dewaal@sas.com

ABSTRACT

Researchers in the EDM community have always relied on sophisticated tools to analyze data and build models. As the amount of data that can be collected and stored grows, the need for tools capable of handling "big data" becomes ever more prevalent. SAS® Analytics U is a new initiative for making SAS data analysis and mining tools available for free to educational researchers and instructors. These tools are designed for handling very large data sets and can be run in the cloud, saving researchers valuable time and resources. Furthermore, SAS Analytics U provides a community of SAS educators and learners to share resources and information about SAS tools and techniques.

This tutorial aims to introduce researchers to the tools available through SAS Analytics U and how they can be applied to the field of Educational Data Mining. We will provide an overview of the SAS architecture and provide instruction on the key features of each tool in the suite. We will guide participants through examples using relevant educational data sources to help researchers understand how the tools can be applied to their own work.

REQUIREMENTS: In order to participate in the hands on exercises, please bring a laptop on which you have installed SAS University Edition. The free download is available at http://www.sas.com/en_us/software/university-edition/download-software.html. The download and installation may take up to 1 hour so there will not be time to get set up during the tutorial.

1. TUTORIAL DESCRIPTION

This tutorial will focus on introducing SAS to participants and guiding them through the use of the suite of tools using relevant educational data sets. The tools that will be covered include:

SAS® Programming Language. SAS programming language is a powerful language designed specifically for intensive data analysis. This highly flexible and extensible fourth generation programming language has a clear syntax and hundreds of language elements and functions. It supports programming everything from data extraction, formatting and cleansing to data analysis, building sophisticated models, and generating reports. The SAS programming language is at the heart of the SAS University Edition tools.

SAS® Studio. SAS Studio is the development environment for SAS University Edition and runs through the web browser as well as in the cloud. It offers a powerful GUI interface that allows novice programmers to interact with data and perform analyses without writing any SAS code themselves. However, the SAS code is all generated behind the scenes and is visible to help users learn.

SAS® Enterprise Miner. SAS Enterprise Miner helps users streamline the data mining process to create highly accurate predictive and descriptive models based on analysis of vast amounts of data. It includes innovative algorithms in the areas of statistics and machine learning to enhance the stability and accuracy of predictions, which can be verified easily by visual model assessment and validation. Users build process flow diagrams that serve as self-documenting procedures. These diagrams can be updated easily or applied to new problems without starting over from scratch. In addition to process flow diagrams, Enterprise Miner provides a programming interface for advanced users. Enterprise Miner allows integration with open source software for data manipulation and model comparison, the open standard PMML, and databases for scoring models without data movement.

Additional SAS tools that may be covered if it is of interest to the participants include tools for time series analysis, forecasting, matrix manipulations, and advanced statistics.

2. JUSTIFICATION

Educational data miners rely on computational tools to understand and explore their data. These tools must be robust and flexible in order to allow for innovation. They must be able to handle ever increasing amounts of data. Ideally, they are easy to use by both programmers and non-programmers alike due to the interdisciplinary nature of this research area. Finally, most researchers rely upon tools that are freely available and do not require excessive resources.

SAS University Edition is a new option that addresses many of these needs. This suite of powerful SAS software was made available to all learners for free in May of 2014. SAS Enterprise Miner, Text Miner, and Forecast Server have been available through SAS OnDemand for Academics since late 2010. However, the biggest barrier to adopting new tools is learning how to use them. SAS Analytics U is a community centered around these free offerings and is designed to support SAS learners and educators. This tutorial seeks to introduce participants to these resources and suite of tools and demonstrate how they can be applied to EDM research. The goal is that participants will be able to add another set of tools to their every growing toolbox for conducting EDM research.

3. PRESENTERS

The presenters for this tutorial include both researchers who are active in the EDM community and trained SAS educators who are experienced in leading tutorials of SAS products.

Jennifer Sabourin. Sabourin has a dual role as a research scientist and software developer on the Curriculum Pathways team at SAS Institute. As a research scientist she works on identifying research questions and using machine learning and analytical techniques to improve the efficacy of Curriculum Pathways products. She also serves as a consultant aiding external researchers with using SAS

software to better understand and make decisions from their educational data. As a software developer she works on creating innovative applications for K-12 that are offered at no-cost.

Sabourin received her Ph.D. from North Carolina State University in 2013. Her graduate work focused on data mining and artificial intelligence in game-based learning environments. She has been an active member of the EDM community since beginning her graduate work.

Scott McQuiggan. McQuiggan leads SAS Curriculum Pathways, an interdisciplinary team focused on the development of no-cost educational software in the core disciplines at SAS Institute Inc. Curriculum Pathways includes more than 1,500 resources, tools, and apps for K-12 education used in all 50 states and more than 90 countries around the world. He regularly uses data mining and analytics to better understand the behaviors exhibited in Curriculum Pathways resources and improve the efficacy of the products themselves.

McQuiggan received his PhD in computer science from North Carolina State University, where his research focused on affective reasoning in intelligent game-based learning environments. He also holds an MS in computer science from North Carolina State University and a Bachelor of Science in computer science from Susquehanna University. Scott is co-author of the book, Mobile Learning: A Handbook for Developers, Educators, and Learners.

André de Waal. De Waal is an Analytical Consultant with SAS Institute and his work focuses on teaching users how they can use SAS to best meet their analytic needs. He received his Ph.D. in theoretical computer science from the University of Bristol during 1994. He spent the next year in Germany and Belgium continuing his research in Logic Programming and Automated Theorem Proving. During 1996 he returned to South Africa to take up his position as lecturer at the School of Computer Science and Information Systems at the then Potchefstroom University for Christian Higher Education (which later became the North-West University), where he was later promoted to Associated Professor. During 1999 he became one of the founder members of the Centre for Business Mathematics and Informatics at the same university. He became responsible for the Data Mining Program in the Centre and shifted his research focus to include Neural Networks and Predictive Modeling. He joined SAS Institute in Cary, NC during December 2010 to take up the position of Analytical Consultant in the Global Academic Program.

4. PROPOSED FORMAT

This tutorial will be presented as interactive instructions where users will be guided through the tools using relevant education data with a focus on techniques that are commonly required in the EDM community. The tutorial will also include an overview of SAS and its commitment to education research by a leading SAS executive. We also seek to gain feedback from participants prior to the event so that we can tailor the sessions to specific needs or questions. A tentative schedule (subject to conference timings) is below:

Session 1: Introduction and SAS Studio

9:00-9:15	Introduction – Introduction of presenters and participants and overview of SAS Analytics U
9:15-10:30	SAS Studio

Coffee Break

Session 2: SAS Studio

11:00-12:30	SAS Studio

Lunch Break

Session 3: Keynote and SAS Enterprise Miner

14:00-14:30	Keynote – A SAS executive (TBD based on final scheduling) will present an overview of SAS and its commitment to education by discussing tools made available to researchers and products made available to K-12 educators and students.
14:30-16:00	SAS Enterprise Miner

Coffee Break

Session 4: Participant Requested Instruction

16:30-17:30	Additional Instruction – based on the goals of the participants we will delve deeper into aspects of the tools already presented or introduce additional tools as listed in the tutorial description.
17:30-18:00	Conclusion

In addition to the tutorial, instructional materials will be made available to participants. We will also provide guidance on avenues for further learning through online instruction.