

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.

1. TUTORIAL OVERVIEW

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. TUTORIAL FORMAT

This tutorial will be divided into four sessions, each covering a specific topic:

Session 1 – SAS University Edition

Session 2 – SAS Studio

Session 3 – SAS Enterprise Miner

Session 4 – Participant requests and questions