

Course content of Big Data with SAS Analytics:

The Advanced Analytics Certification program includes three learning modules, comprising a total of 9 courses.

Module 1: Predictive Modeling

Course 1: Applied Analytics Using SAS Enterprise Miner

This course covers the skills required to assemble analysis flow diagrams using SAS Enterprise Miner for both pattern discovery (segmentation, association and sequence analyses) and predictive modeling (decision trees, regression and neural network models).

Topics Covered

- Defining a SAS Enterprise Miner project and exploring data graphically.
- Modifying data for better analysis results.
- Building and understanding predictive models, including decision trees and regression models.
- Comparing and explaining complex models.
- Generating and using score code.
- Applying association and sequence discovery to transaction data.

Module 1 prepares you for the [Predictive Modeling certification exam](#).

Module 2: Advanced Predictive Modeling

Course 1: Neural Network Modeling

This course helps you understand and apply two popular artificial neural network algorithms – multilayer perceptrons and radial basis functions. Both the theoretical and practical issues of fitting neural networks are covered.

Topics Covered

- Constructing multilayer perceptron and radial basis function neural networks.
- Constructing custom neural networks using the NEURAL procedure.
- Choosing an appropriate network architecture and determining the relevant training method.
- Avoiding overfitting neural networks.
- Performing autoregressive time series analysis using neural networks.
- Interpreting neural network models.

Course 2: Predictive Modeling Using Logistic Regression

This course explores predictive modeling using SAS/STAT® software, with an emphasis on the LOGISTIC procedure.

Topics Covered

- Using logistic regression to model an individual's behavior as a function of known inputs.
- Selecting variables and interactions.
- Creating effect plots and odds ratio plots using ODS Statistical Graphics.
- Handling missing data values.
- Tackling multi collinearity in your predictors.
- Assessing model performance and comparing models.
- Recoding categorical variables based on the smooth weight of evidence.

- Using efficiency techniques for massive data sets.

Course 3: Data Mining Techniques: Predictive Analytics on Big Data

This course introduces applications and techniques for assaying and modeling large data. It presents basic and advanced modeling strategies, such as group-by processing for linear models, random forests, generalized linear models and mixture distribution models. You will perform hands-on exploration and analyses using tools such as SAS Enterprise Miner, SAS Visual Statistics and SAS In-Memory Statistics.

Topics Covered

- Using applications designed for big data analyses.
- Exploring data efficiently.
- Reducing data dimensionality.
- Building predictive models using decision trees, regressions, generalized linear models, random forests and support vector machines.
- Building models that handle multiple targets.
- Assessing model performance.
- Implementing models and scoring new predictions.

Course 4: Using SAS to Put Open Source Models Into Production

This course introduces the basics for integrating R programming and Python scripts into SAS and SAS Enterprise Miner. Topics are presented in the context of data mining, which includes data exploration, model prototyping, and supervised and unsupervised learning techniques.

Topics Covered

- Calling R packages in SAS.
- Using Python scripts in SAS.
- Integrating open source data exploration techniques in SAS.

- Integrating open source models in SAS Enterprise Miner.
- Creating production (score) code for R models.

Module 2 prepares you for the [Advanced Predictive Modeling certification exam](#).

Module 3: Text Analytics, Time Series, Experimentation and Optimization

Course 1: Text Analytics Using SAS Text Miner

In this course, you will learn to use SAS Text Miner to uncover underlying themes or concepts contained in large document collections, automatically group documents into topical clusters, classify documents into predefined categories, and integrate text data with structured data to enrich predictive modeling endeavors.

Topics Covered

- Converting documents stored in standard formats (Microsoft Word, Adobe PDF, etc.) into general-purpose HTML or TXT formats.
- Reading documents from a variety of sources (web pages, flat files, data elements in a relational database, spreadsheet cells, etc.) into SAS tables.
- Processing textual data for text mining (e.g., correcting misspellings or recoding acronyms and abbreviations).
- Converting unstructured text-based character data into structured numeric data.
- Exploring words and phrases in a document collection.
- Querying document collections using keywords (i.e., identifying documents that include specific words or phrases).
- Identifying topics or concepts that appear in a document collection.
- Creating user-influenced topic tables from scratch or by modifying machine-generated topics, or creating concepts using domain knowledge.

- Using derived topic tables or pre-existing user-influenced topic tables (or both) to enhance information retrieval and document classification.
- Clustering documents into homogeneous subgroups.
- Classifying documents into predefined categories.

Course 2: Time Series Modeling Essentials

In this course, you'll learn the fundamentals of modeling time series data, with a focus on the applied use of the three main model types for analyzing univariate time series: exponential smoothing, autoregressive integrated moving average with exogenous variables (ARIMAX), and unobserved components (UCM).

Topics Covered

- Creating time series data.
- Accommodating trend, as well as seasonal and event-related variation, in time series models.
- Diagnosing, fitting and interpreting exponential smoothing, ARIMAX and UCM models.
- Identifying relative strengths and weaknesses of the three model types.

Course 3: Experimentation in Data Science

This course explores the essentials of experimentation in data science, why experiments are central to any data science efforts, and how to design efficient and effective experiments.

Topics Covered

- Defining common terminology in designed experiments.
- Describing the benefits of multifactor experiments.
- Differentiating between the impact of a model and the impact of the action taken from that model.

- Fitting incremental response models to evaluate the unique contribution of a marketing message, action, intervention or process change on outcomes.

Course 4: Optimization Concepts for Data Science

This course focuses on linear, nonlinear and efficiency optimization concepts. Participants will learn how to formulate optimization problems and how to make their formulations efficient by using index sets and arrays. Course demonstrations include examples of data envelopment analysis and portfolio optimization. The OPTMODEL procedure is used to solve optimization problems that reinforce concepts introduced in the course.

Topics Covered

- Identifying and formulating appropriate approaches to solving various linear and nonlinear optimization problems.
- Creating optimization models commonly used in industry.
- Formulating and solving a data envelopment analysis.
- Solving optimization problems using the OPTMODEL procedure in SAS.

Module 3 prepares you for the [Text Analytics, Time Series, Experimentation and Optimization certification exam](#).