SAS Machine Learning Specialist

1. **BASE SAS - 50 hrs**

   a) **SAS Programming 1: Essentials**

This course is for users who want to learn how to write SAS programs. It is the entry point to learning SAS programming and is a prerequisite to many other SAS courses. If you do not plan to write SAS programs and you prefer a point-and-click interface, you should attend the *SAS Enterprise Guide 1: Querying and Reporting* course.

Learn how to

- navigate the SAS windowing environment
- navigate the SAS Enterprise Guide programming environment
- read various types of data into SAS data sets
- create SAS variables and subset data
- combine SAS data sets
- create and enhance listing and summary reports
- validate SAS data sets.

Who should attend: Anyone starting to write SAS programs

Prerequisites

Before attending this course, you should have experience using computer software. Specifically, you should be able to

- understand file structures and system commands on your operating systems
- access data files on your operating systems.

   b) **SAS Programming 2: Data Manipulation Techniques**

This course is for those who need to learn data manipulation techniques using SAS DATA and procedure steps to access, transform, and summarize SAS data sets. The course builds on the concepts that are presented in the *SAS Programming 1: Essentials* course and is not recommended for beginning SAS software users.

Learn how to

- control SAS data set input and output
- combine SAS data sets
- summarize, read, and write different types of data
- perform DO loop and SAS array processing
- transform character, numeric, and date variables.

Who should attend: Business analysts and SAS programme

Prerequisites

Before attending this course, you should have at least six months of experience writing SAS programs or have completed the *SAS Programming 1: Essentials* course and used SAS for at least one month. Specifically, you should be able to

- submit a SAS program
- diagnose and correct syntax errors
- examine descriptor and data portions of a SAS data set
- access SAS data libraries
- read and create SAS data sets
- read Excel spreadsheets
- read delimited raw data files
- examine data errors when reading raw data files
- use SAS procedures to validate data
- clean invalid data
- create variables
- combine SAS data sets
- use global statements
- use labels and formats, including user-defined formats
- subset observations

produce summary reports using the FREQ and MEANS procedures

2. **Accelerating SAS® Code on the SAS® Viya® Platform (Programming for SAS Viya) – 12 hrs**

This course is for SAS programmers who want to learn how to run code on the SAS Viya platform. In the course, you first learn how you can easily run traditional SAS code in SAS Viya like you have in SAS 9. Then you learn how to modernize your SAS code to take advantage of SAS Cloud Analytic Services (CAS) on the platform to boost program performance. CAS is a high-performance in-memory analytic engine that performs massively parallel processing (MPP) on data of any size, working with multiple machines, CPUs, and threads for lightning-fast results. The course focuses on how to access, manage, analyze, and manipulate in-memory tables on the distributed CAS server using traditional SAS code as well as new coding techniques. This course is not intended for beginning SAS software users.
Learn how to

- Run traditional SAS code on the SAS Viya platform.
- Connect to SAS Cloud Analytic Services (CAS).
- Access and use caslibs on the CAS server.
- Load SAS data sets, Parquet files, CSV files, Microsoft Excel files, and DBMS tables into CAS.
- Save distributed CAS tables back to permanent storage in a variety of formats like SASHDAT, Parquet, and CSV.
- Modify DATA step code to execute in CAS.
- Modify SQL procedure code to execute in CAS using FedSQL.
- Use CAS-enabled procedures.
- Write CASL code to execute actions on the CAS server.

Who should attend: Anyone starting to write SAS programs

Prerequisites

Before attending this course, you should have nine months or more of SAS programming experience or have completed the SAS Programming 1: Essentials and SAS Programming 2: Data Manipulation Techniques courses.

Software Addressed

SAS Studio; SAS Viya

Course Contents

SAS Viya Platform Overview

- Introducing the SAS Viya platform.
- SAS Viya programming interfaces.
- SAS Viya servers and processing environments.

Running SAS Code on the SAS Compute Server

- Overview of running SAS code on the SAS Compute Server.
- SAS Viya Compute Server overview.
- Running SAS 9 Code on the Compute Server in SAS Viya.

SAS Cloud Analytic Services (CAS) Overview

- CAS fundamentals.
- Understanding caslibs.
-

Managing Data in SAS Cloud Analytic Services

- Loading data to in-memory tables.
- Accessing DBMS data.
- Saving and dropping in-memory tables.

Modifying DATA Step Code to Run in SAS Cloud Analytic Services (CAS)

- Modifying DATA step code to run in SAS Cloud Analytic Services (CAS).

Running SAS Procedures in SAS Cloud Analytic Services (CAS)

- Introduction to SAS procedures in SAS Viya.
- Running CAS-enabled SAS procedures.

Modifying SQL Code to Run in SAS Cloud Analytic Services (CAS)

- Modifying SQL code to run in CAS.
- Column data types in CAS.

Using the Native CAS Language (CASL)

- Introduction to CASL.
- Using CAS actions.

3.   Statistics You Need to Know for Machine Learning – 24 hrs

When it comes to using data, there are two main camps, traditional statistics and machine learning, and the two camps complement each other. Statistics remains highly relevant, irrespective of the "bigness" of data. Its role remains what it has always been, but it is even more important now. There is a need to transition from traditional statistical modeling to the machine learning world. This course introduces the statistical background necessary for machine learning using SAS Viya. Knowledge of statistics relevant to machine learning will prepare you to become a data scientist. The course prepares you for future instruction on doing machine learning (including its underlying methodology that has statistical foundations) and enables you to develop a deeper understanding of machine learning models.

This course is a prerequisite to many of the courses in the data science curriculum.

Learn How To

- Explain the relevance of statistics in big data and machine learning.
- Relate statistical and data science terminology.
- Generate descriptive statistics and explore data with graphs.
- Detect associations among variables.
- Perform linear regression for explanatory modeling.
- Compare explanatory modeling with predictive modeling.
- Describe the trade-off between bias and variance.
- Fit a logistic regression model for predictive modeling.
- Score new data.
- Explain the statistical foundations of machine learning.
- Discuss data difficulties and modeling issues and their statistical solutions.

Who Should Attend
- Anyone in the field of data science who does not yet have a deep understanding of statistical and machine learning concepts or wants to enhance their knowledge, which might include business analysts, data analysts, marketing analysts, marketing managers, data scientists, data engineers, financial analysts, data miners, statisticians, mathematicians, and others who work in allied areas

Prerequisites
- Before attending this course, you should have experience using computer software. It is beneficial if you have completed the equivalent of an undergraduate course in statistics covering distribution of data, p-values, hypothesis testing, and regression. No prior SAS experience is needed.

Course Outline

Statistics and Machine Learning
- Relevance of statistics in big data and machine learning.
- Terminology and vocabulary.
- Introduction to SAS Viya and SAS Studio.

Fundamental Statistical Concepts
- Introduction to statistical analysis.
- Descriptive statistics.
- Inferential statistics.

Explanatory Modeling Using Linear Regression
- Correlation and simple linear regression.
- Multiple regression and model selection.
- Model diagnostics.

Predictive Modeling Using Logistic Regression
- Introduction to predictive modeling.
- Categorical associations.
- Logistic regression model.
- Model deployment.

Statistical Foundations of Machine Learning
- Overview of machine learning.
- Data pre-processing for machine learning models.
- Model evaluation, estimation, and post-training tasks.

4.   **Machine Learning Using SAS® Viya® - 16 Hrs**

This course discusses the theoretical foundation for techniques associated with supervised machine learning models. A series of demonstrations and practices is used to reinforce all the concepts and the analytical approach to solving business problems. In addition, a business case study is defined to guide participants through all steps of the analytical life cycle, from problem understanding to model deployment, by illustrating data exploration, data preprocessing, feature selection, model training and validation, model assessment, and scoring. This course is the core of the SAS Viya Data Mining and Machine Learning curriculum. It uses Model Studio, the pipeline flow interface in SAS Viya that enables you to prepare, develop, compare, and deploy advanced analytics models. You learn to train supervised machine learning models to make better decisions on big data.

Learn How To

- Apply the analytical life cycle to a business need.
- Incorporate a business-problem-solving approach in daily activities.
- Prepare and explore data for analytical model development.
- Create and select features for predictive modeling.
- Develop a series of supervised learning models based on different techniques such as decision trees, ensembles of trees (forest and gradient boosting), neural networks, and support vector machines.
- Evaluate and select the best model based on business needs.
- Deploy and manage analytical models under production.

Who Should Attend

- Business analysts, data analysts, marketing analysts, marketing managers, data scientists, data engineers, financial analysts, data miners, statisticians, mathematicians, and others who work in correlated areas

Prerequisites

- Before attending this course, participants should have at least an introductory-level familiarity with statistics and machine learning concepts. You can gain this knowledge by first attending the Statistics You Need to Know for Machine Learning course. Previous SAS software experience is helpful but not required.

Course Outline

Introduction
- Machine learning in business decision making.
- Essentials of supervised prediction.
- Introduction to SAS

Viya.Data Preparation
- Data exploration.
- Feature extraction.
- Input transformations.
- Feature selection.
- Variable clustering (self-study).
- Best practices.
- Selecting your algorithm.

Decision Trees and Ensembles of Trees
- Introduction.
- Tree-structure models.
- Recursive partitioning.
- Pruning.
- Ensembles of trees.

Neural Networks
- Introduction.
- Network architecture.
- Network learning and optimization.

Support Vector Machines and Additional Topics
- Large-margin linear classifier.
- Methods of solution.
- Nonlinear classifier: Kernel Trick.
- Additional tools.

Model Assessment and Deployment
- Model assessment and comparison.
- Model deployment.

SAS Certification exams & Credentials

This course is mapped to One credentials as below:

1. SAS Certified Specialist: Base Programming Using SAS 9.4
2. SAS Certified Specialist: Machine Learning Using SAS Viya

Participants will be awarded SAS certified digital badges for these on clearing respective SAS exams

Besides participants will also be awarded participation certificate & Learn SAS digital badges for –

1. SAS Programming 1: Essentials
2. SAS Programming 2: Data Manipulation Techniques
3. Accelerating SAS® Code on the SAS® Viya® Platform
4. Statistics You Need to Know for Machine Learning
5. Machine Learning Using SAS® Viya®