

Predictive Analytics Modeler



(Classroom)

Career path description

The Predictive Analytics Modeler career path prepares students to learn the essential analytics models to collect and analyze data efficiently. This will require skills in predictive analytics models, such as data mining, data collection and integration, nodes, and statistical analysis. The Predictive Analytics Modeler will use tools for market research and data mining in order to predict problems and improve outcomes.

ibm.com/training

General information

Delivery method

20% web-based and 80% instructor led

Version

2017

Product

IBM SPSS Modeler

Audience

Undergraduate senior students from IT related academic programs i.e. computer science, software engineering, information systems and similar others



Learning objectives

After completing this course, you should be able to:

- The importance of analytics and how its transforming the world today
- Understand how analytics provided a solution to industries using real case studies
- Explain what is analytics, the various types of analytics, and how to apply it
- Improve efficiency, sample records, and work with sequence data
- Explain data transformations, and functions
- Understand modeling, relationships, derive and reclassify fields
- Integrate and collect data
- Understand the principles of data mining
- Use the user interface of modeler to create basic program streams
- Read a statistics data file into modeler and define data characteristics
- Review and explore data to look at data distributions and to identify data problems, including missing values
- Use the automated data prep node to further prepare data for modeling
- User a partition node to create training and testing data subsets

Prerequisites Skills

- English Proficiency
- Basic Internet and web browser usage experience
- Basic analytics experience
- Exposure to the IBM Skills Academy Portal learning environment
- Exposure to the IBM Skills Academy Cloud hands-on labs platform

Duration

57.75 hours

Skill level

Basic – Intermediate

Hardware requirements

Classroom (ILT) setup requirements

Processor	Intel Core i7 CPU @ 2.7 GHz
GB RAM	8 GB
GB free disk space	60 GB
Network requirements	No
Other requirements	IBM ID

Notes

The following unit and exercise durations are estimates, and might not reflect every class experience. If the course is customized or abbreviated, the duration of unchanged units will probably increase.

Course Agenda

MODULE I – ANALYTICS OVERVIEW

Course I – Business Analytics Overview

Duration: 1 hour

Course introduction
Duration: 5 minutes

Unit 1. Analytics overview **Duration: 10 minutes**

Overview	This unit provides an understanding of the importance of business analytics in our world, society, and life.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Understand how analytics is transforming the world• Understand the profound impact of analytics in business decisions• Understand what is analytics and how it works• Understand why business analytics has become important in various industries

Unit 2. Analytics trends: Past, present & future **Duration: 15 minutes**

Overview	This unit explains how analytics has evolved over time.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Understand the history of analytics and how it has changed today• Understand how to analyze unstructured data• Understand how analytics is making the world smarter• Understand where the future of analytics lies

Unit 3. Towards a predictive enterprise **Duration: 5 minutes**

Overview	This unit explains the effects of business analytics in the corporate world that has led to its global adoption across geographies and industries.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Explain why successful enterprises need business analytics• Understand how business analytics can help turn data into insight

Unit 4. Analytics: Industry domains
Duration: 5 minutes

Overview	This unit highlights the application of analytics across major industries.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Understand how predictive analytics is transforming all types of organizations• Explain how analytics supports retail companies• Understand how analytics can reduce crime rates and accidents• Explain the use of analytics in law enforcement and insurance companies• Understand how analytics can affect the future of education

Unit 5. Case studies and solutions
Duration: 15 minutes

Overview	This unit covers real case studies and solutions of the adoption of business analytics across the world.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Understand the importance of business analytics• Comprehend how big data and analytics can help in understanding consumer/customer behavior• Explain how analytics can help manage assets• Understand how analytics can help combat fraud• Explain how analytics can help us to understand social sentiments

MODULE II – Business Analytics Foundations

Course I – Business Intelligence and Analytics 101

Duration: 1 hour

Course introduction
Duration: 5 minutes

Business Intelligence and Analytics 101
Duration: 1 hour

Overview

This course provides a collection of resources designed for participants to become familiar with business intelligence (BI) and analytics concepts. Participants will review materials to introduce themselves to terminology and practical business use cases for a high level understanding of BI and analytics. The course includes a pre-assessment for participants to measure their understanding of the content before taking the course, and a post-assessment for participants to gauge their learning after reviewing the materials.

Learning objectives

After completing this course, you should be able to:

- Explain what is analytics
 - Define various types of analytics
 - Demonstrate how to apply analytics
 - Describe business intelligence
 - Demonstrate how to apply business intelligence
-

MODULE III – PREDICTIVE ANALYTICS MODELER

Course I – Introduction to A Predictive Analytics Platform & Data Mining

Duration: 27.2 hours

Course introduction
Duration: 10 minutes

Unit 1. Introduction to data mining
Duration: 1 hour

Overview

In this unit, you will learn about data mining and its applications.

Learning objectives

After completing this unit, you should be able to:

- **List two applications of data mining**
 - **Explain the stages of the CRISP-DM process model**
 - **Describe successful data-mining projects and the reasons why projects fail**
 - **Describe the skills needed for data mining**
-

Exercise 1. Introduction to data mining (workshop)

Duration: 1 hour

Overview	In this exercise, you will learn how to apply data mining.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Understand data mining• Describe how to apply data mining in different scenarios

Unit 2. Working with modeler

Duration: 1 hour

Overview	In this exercise, you will learn about objects such as streams and nodes and you will acquire experience with the software.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Describe the MODELER user-interface• Work with nodes• Run a stream or a part of a stream• Open and save a stream• Use the online Help

Exercise 2. Working with modeler

Duration: 45 minutes

Overview	In this exercise, you will learn about MODELER's user-interface to create streams.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Create streams• Change streams• Generate a select node from the Table output

Exercise 3. Working with modeler (workshop)

Duration: 1 hour

Overview	In this exercise, you will learn how to build and run streams.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Create a stream that reads data and exports data to Microsoft Excel• Change and save a stream.• Create a new stream from an existing stream.• Make a stream neat using a SuperNode

Unit 3. A Data-mining tour
Duration: 1 hour

Overview	In this unit you will learn about building a model and then applying that model to future cases of a data-mining project.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Explain the basic framework of a data-mining project• Build a model• Deploy a model

Exercise 4. A Data-mining tour
Duration: 45 minutes

Overview	In this exercise, you are working as a data miner for a telecommunications firm and have to identify customers who are likely to churn.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Build a model using historical data• Deploy the model

Exercise 5. A Data-mining tour (Workshop)
Duration: 1 hour

Overview	In this exercise, you will learn to build a model using data of a test mailing and select groups with high response rates in the customer database.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Explore the data• Select modeling data• Build a CHAID model• Interpret of the fields added by model nugget• Explore the results

Unit 4. Collecting initial data
Duration: 1 hour

Overview	In this exercise, you will learn how to collect initial data. You will also learn how to describe data.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Explain the concepts of data structure, unit of analysis, field storage and field measurement level• Import Microsoft Excel files• Import text files• Import from databases• Export data to various formats

Exercise 6. Collecting initial data
Duration: 45 minutes

Overview	In this exercise you will learn how to import data from various data sources and report on the unit of analysis and fields' measurement levels .
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Import a Microsoft Excel file• Import a text file• Set fields' measurement levels

Exercise 7. Collecting initial data (workshop)
Duration: 1 hour

Overview	In this exercise you are working for a company selling sports products. You will import the company's data files and build a model to identify groups with high response rates .
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Import data• Determine the unit of analysis• Determine relationships between datasets• Set measurement levels

Unit 5. Understanding your data
Duration: 1 hour

Overview	In this exercise, you will learn how to explore data and assess it's quality .
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Audit the data• Explain how to check for invalid values• Take action for invalid values• Explain how to define blanks

Exercise 8. Understanding your data
Duration: 1 hour

Overview	In this exercise, you will handle a case study where you will import data and later assess its quality.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Audit the data• Define valid values and take action• Declare blank values

Exercise 9. Understanding your data (Workshop)

Duration: 1 hour

Overview	In this unit, you will learn how to examine a company's datasets, and take corrective action where needed.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Explore the data• Set ranges and take action• Declare blanks

Unit 6. Setting the unit of analysis

Duration: 1 hour

Overview	In this exercise, you will learn how to set unit of analysis in three different methods.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Set the unit of analysis by removing duplicate records• Set the unit of analysis by aggregating records• Set the unit of analysis by expanding a categorical field into a series of flag fields

Exercise 10. Setting the unit of analysis

Duration: 45 minutes

Overview	In this exercise, you will learn how to remove duplicate records from a customer dataset. You will also learn how to transform a transactional dataset into a dataset that has one record per customer.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Cleanse data by removing duplicate records• Expand a categorical field into a series of flag fields

Exercise 11. Setting the unit of analysis (Workshop)

Duration: 1 hour

Overview	In this exercise you will learn how to import data from several sources and create datasets with the required unit of analysis.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Remove duplicate records• Create a dataset where customers are unique in a company's purchases data• Create a dataset where customers are unique in a company's order lines data• Create a dataset where customers are unique in a company's mailing history data

Unit 7. Integrating data

Duration: 1 hour

Overview	In this exercise you will learn how to combine different datasets into a single dataset for analysis.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Integrate data by appending records from multiple datasets• Integrate data by merging fields from multiple datasets• Sample records

Exercise 12. Integrating data

Duration: 45 minutes

Overview	In this exercise, you will learn how to combine a number of datasets into a single dataset as a preparation for analysis and modeling.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Append records from two datasets• Merge fields from different datasets• Enrich a dataset with aggregated data• Sample records

Exercise 13. Integrating data (Workshop)

Duration: 1 hour

Overview	In this unit, you will learn how to combine a number of datasets into a single dataset to build models using the information from all these datasets.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Create single datasets• Enrich the data with zipcode information• Export a random sample

Unit 8. Deriving and reclassifying fields

Duration: 1 hour

Overview	In this exercise you will learn how to construct the final dataset for modeling by cleansing and enriching your data.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Use the Control Language for Expression Manipulation (CLEM)• Derive new fields• Reclassify field values

Exercise 14. Deriving and reclassifying fields

Duration: 45 minutes

Overview	In this exercise, you will learn how to cleanse and enrich a dataset to build models.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Cleanse data and derive fields for modeling• Cleanse data and reclassify fields for modeling

Exercise 15. Deriving and reclassifying fields (Workshop)

Duration: 1 hour

Overview	In this exercise, you will learn how to cleanse a company's data and enrich the data with a number of new fields so that better models can be built.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Compute the difference between amount spent and credit limit• Compute fields in a currency from a different currency• Create a segment field• Create a field returning the bonus

Unit 9. Looking for relationships

Duration: 1 hour

Overview	In this exercise you will learn methods used to examine the relationship between two fields.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Examine the relationship between two categorical fields• Examine the relationship between a categorical field and a continuous field• Examine the relationship between two continuous fields

Exercise 16. Looking for relationships

Duration: 45 minutes

Overview	In this exercise you will learn how to assess relationships and determine its strength by doing a demo.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Assess the relationship between churn and handset• Assess the relationship between churn and number of dropped calls• Assess the relationship between number of products and revenues

Exercise 17. Looking for relationships (Workshop)

Duration: 1 hour

Overview	In this unit, you will learn how to examine the relationships in datasets and find out which fields are related to response.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Examine the relationship between response and other factors in the dataset.

Unit 10. Introduction to modeling

Duration: 1 hour

Overview	In this exercise, you will learn about the modeling stage of the CRISP-DM process model.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• List three modeling objectives• Use a classification model• Use a segmentation model

Exercise 18. Introduction to modeling

Duration: 45 minutes

Overview	In this exercise you will learn about classification and segmentation using a synthetic dataset from a telecommunications firm.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Predict churn by running a CHAID model• Predict churn by running a Neural Net model• Compare the accuracy of these models• Find groups of similar customers, based on usage.

Exercise 19. Introduction to modeling (Workshop)

Duration: 1 hour

Overview	In this exercise, you will learn how to check a model's accuracy and use segmentation to cluster records.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Build a CHAID model to predict response• Assess the model's accuracy• Apply the model to other customers• Use the two step segmentation model to cluster records.

Course II – Advanced data preparation

Duration: 13.75 hours

Course introduction
Duration: 10 minutes

Unit 1. Using functions
Duration: 1 hour

Overview	In this unit, you will learn how to use various different kinds of functions.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Use date functions• Use conversion functions• Use string functions• Use statistical functions• Use missing value functions

Exercise 1. Using functions
Duration: 30 minutes

Overview	In this exercise, you will learn how to use functions to cleanse and enrich a dataset to build better models.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Use the date functions to derive fields• Use string functions to derive fields• Use statistical functions to derive fields• Use missing value functions to derive fields

Exercise 2. Using functions (Workshop)
Duration: 1 hour

Overview	In this exercise, you will work with data based on customers and their holiday destinations. You will then use this data to answer questions.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Import and instantiate the data• Compute an AGE field• Conditionally compute the sum over a series of fields• Derive a field taking blank values into account

Unit 2. Data transformations
Duration: 1 hour

Overview	In this exercise, you will learn how to apply various different nodes to modify fields and prepare your data for modeling.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Use the Filler node to replace values• Use the Binning node to recode continuous fields• Use the Transform node to change a field's distribution

Exercise 3. Data transformations
Duration: 1 hour

Overview	In this exercise, you will learn how to cleanse data using the Filler node and add new fields using the Binning node and the Transform node.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Use the Filler node to change storage• Use the Filler node to replace null values• Use the Filler node to replace strings• Do binning with equal counts• Do binning using a supervisor field

Exercise 4. Data transformations (Workshop)
Duration: 1 hour

Overview	In this unit, you will learn how to transform data by importing, replacing and recoding.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Import and instantiate the data• Correct spelling• Replace blanks with undefined values• Bin a field optimally with respect to target• Transform a field to change its distribution

Unit 3. Working with sequence data
Duration: 1 hour

Overview	In this exercise, you will learn how to work with sequenced data.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Use cross-record functions• Use the Count mode in the Derive node• Use the Restructure node to expand a continuous field into a series of continuous fields• Use the Space-Time-Boxes node to work with geospatial and time data

Exercise 5. Working with sequence data
Duration: 45 minutes

Overview	In this exercise, you will learn how to apply various transformations to sequence data.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Create a record identifier• Move an average• Restructure a transactional dataset• Use the Space-Time-Boxes node

Exercise 6. Working with sequence data (Workshop)
Duration: 1 hour

Overview	In this exercise, you will learn how to work on a dataset, and derive new fields.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Import the data• Derive a record identifier• Restructure the dataset• Analyze geospatial and time data

Unit 4. Sampling records
Duration: 1 hour

Overview	In this exercise, you will learn how to use the Sample node and various reasons for sampling records.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Use the Sample node to draw simple and complex samples• Partition the data into a training and a testing set• Reduce or boost the number of records

Exercise 7. Sampling records
Duration: 45 minutes

Overview	In this exercise, you will learn how sample data using various techniques, and use partitioning to select the best predictive model.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Draw a simple sample and a complex sample• Partition data into a training set and a testing set• Balance the data

Exercise 8. Sampling records

Duration: 1 hour

Overview	In this exercise, you will learn how to sample a data for a satisfaction survey and build a model to predict response to a campaign.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Import the data, instantiate the data and examine the response• Draw a random sample• Draw a stratified sample• Prepare for modeling by using a Type node• Run models on the training set and select the best model

Unit 5. Improving efficiency

Duration: 1 hour

Overview	In this exercise, you will learn how to work with SQL pushback, Set Globals node and parameters to optimize efficiency.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Use database scalability by SQL pushback• Use the Data Audit node to process outliers and missing values• Use the Set Globals node• Use parameters• Use looping and conditional execution

Exercise 9. Improving efficiency

Duration: 45 minutes

Overview	In this exercise, you will learn how to check data for outliers and extremes, compute standardized scores and use parameters and looping.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Use the Data Audit node to process outliers, extremes and missing values• Compute standardized scores using globals• Use parameters• Create a loop through values

Exercise 10. Improving efficiency (Workshop)
Duration: 1 hour

Overview	In this exercise, you will learn how to process outliers, extremes and missing values using the Data Audit node. You will use the Set Globals node to replace missing values, and will be introduced to automation by using parameters and looping.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Import and instantiate the data• Use globals to replace undefined values with the mean• Create a loop through the row fields in the Matrix node

Course III – Automated Data Mining

Duration: 14.8 hours

Course introduction
Duration: 15 minutes

Unit 1. Introduction to data mining
Duration: 30 minutes

Overview	In this unit you will learn about the principles of data mining.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Describe the featured included with modeler to automate data mining• Describe the phases of the CRISP-DM process model for data mining

Unit 2. The basics of using a modeler
Duration: 1 hour

Overview	In this unit you will review the basic features of the modeler user interface and learn how to perform common actions.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Use the modeler interface• Describe the components of the modeler user interface• Place nodes on the stream canvas• Connect and disconnect nodes• Edit and rename codes

Exercise 1. Adding nodes and creating streams in the modeler
Duration: 15 minutes

Overview	In this exercise, you will learn how to create adding nodes, and streams in modeler.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Create nodes• Create streams

Unit 3. Reading data files
Duration: 1 hour

Overview	In this unit you will learn how to read data files and define data characteristics.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Read a statistics data file into modeler• Use a statistics file node to read a statistics data file• Use the filter tab to filter and rename fields• Use the types tab to view measurement level and set field role• Save a modeler stream file

Exercise 2. Reading a data file and typing the data in the source node
Duration: 20 minutes

Overview	In this exercise, you will learn how to read a data file and type the data in the source node.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Read a data file• Type the data in the source node

Unit 4. Data exploration
Duration: 1 hour

Overview	In this unit you will learn about different issues concerned with data quality.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Review and explore data to look at data distributions• Identify data problems, including missing values• Describe the types of missing values for fields• Set missing values for fields• Use the data audit node to explore data distributions• Use the data audit node to impute missing data• Use the table node to view the data file

Exercise 3. Review missing values in modeler and use the data audit node on the charity data
Duration: 30 minutes

Overview	In this exercise, you will learn how to review missing values in the modeler, and to use the data audit node on the charity data.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Edit the source node• Identify what types of blank values are defined for fields• Add a data audit node to the stead• Review missing values

Unit 5. Automated data preparation
Duration: 1 hour

Overview	In this unit you will learn how to apply automated data preparation to the telecommunications customer data to continue the process of data preparation.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Use the automated data prep node to further prepare data modeling• Use the type node to set characteristics for fields• Describe the various features and capabilities of the automated data prep node• Use settings of the automated data prep node that are appropriate for the data and modeling objectives• Describe the types of output produced by the automated data prep node

Exercise 4. Practice using the ADP node to prepare data for modeling
Duration: 20 minutes

Overview	In this exercise, you will learn how to use the ADP node to prepare data for modeling.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Add an ADP node to the stream• Edit the ADP node• Run analysis on the ADP node

Unit 6. Data partitioning
Duration: 45 minutes

Overview	In this unit you will learn how to add a partition node to the stream.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Use a partition node to create training and testing data subsets• Describe rationale and use of a partition node to create data subsets• Set sizes of the training and testing partitions and other partition characteristics• Use a distribution node to view the distribution of a categorical field

Exercise 5. Use a partition node to split the charity data for modeling
Duration: 20 minutes

Overview	In this exercise, you will learn how to create training and testing partitions.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Use a partition node to create training and testing data subsets• Describe rationale and use of a partition node to create data subsets• Set sizes of the training and testing partitions• Use a distribution node to view the distribution of a categorical field

Unit 7. Predictor selection for modeling
Duration: 1 hour

Overview	In this unit you will learn about the feature selection node and how it can help in data modeling.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Use the feature selection node to select inputs for modeling• Describe the features and settings of the feature selection node• Describe the model output from feature selection• Generate a filter node to use the selected fields

Exercise 6. Use the feature selection node to select fields and predict a response
Duration: 20 minutes

Overview	In this exercise, you will use the feature selection node to select fields and predict a response.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Use the feature selection node to select fields• Predict a response

Unit 8. Automated models for categorical targets
Duration: 1 hour

Overview	In this unit you will learn how to use the auto classifier node to create an ensemble model that predicts a categorical target.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Describe the features and settings of the auto classifier node• Describe and use the components of the model output from the auto classifier node

Exercise 7. Using the auto classifier node to construct a model in order to predict a response
Duration: 20 minutes

Overview	In this exercise, you will learn how to use the auto classifier node to construct a model in order to predict a response.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Use the auto classifier node to construct a model in order to predict a response

Unit 9. Model evaluation
Duration: 1 hour

Overview	In this unit, you will learn how to evaluate and understand the predictions of a model.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Use the analysis mode to get a summary of predictions• Use the select node to analyze the testing partition data• Use a matrix node to examine the percent accuracy of predictions• Use a distribution node to graphically display the relationship between a categorical prediction and the target• Use a histogram node to graphically display the relationship between a continuous predictor and the target

Exercise 8. Evaluate the model created to predict the field response
Duration: 30 minutes

Overview	In this exercise, you will learn how to evaluate the model created in order to predict the field response.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Use an analysis node to evaluate model predictions• Use a distribution node to evaluate model predictions• Use a histogram node to evaluate model predictions

Unit 10. Automated models for continuous targets
Duration: 1 hour

Overview	In this unit you will learn how to use the auto numeric node to create an ensemble model to predict a continuous target.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Describe and use the features of the auto numeric node• Describe and use the components of the model output from the auto numeric node• Use various nodes for model evaluation

Exercise 9. Develop a model to predict the total spending
Duration: 30 minutes

Overview	In this exercise, you will learn how to develop a model to predict total spending by the respondent.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Add an auto numeric node to the stream• Use an analysis node to evaluate the auto numeric model

Unit 11. Deploying models
Duration: 50 minutes

Overview	In this unit you will learn how to use a model to score new data.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Describe what needs to be modified to create a scoring stream for new data• Describe the deployment options in modeler• Export scored data to another file format

Exercise 10. Use a scoring stream to make predictions
Duration: 20 minutes

Overview	In this exercise, you will learn how to use a scoring stream to make predictions.
Learning objectives	After completing this exercise, you should be able to: <ul style="list-style-type: none">• Use a scoring stream to make predictions

Unit 12. Course summary
Duration: 45 minutes

Overview	In this unit you will learn how to use modeler to perform an automated data mining project.
Learning objectives	After completing this unit, you should be able to: <ul style="list-style-type: none">• Understand the principles of data mining• Use the user interface of modeler to create basic program streams• Read a statistics data file into modeler and define data characteristics• Review and explore data to look at data distributions and to identify data problems, including missing values• Use the automated data prep node to further prepare data for modeling• User a partition node to create training and testing data subsets• Use the feature selection node to select inputs for modeling• Use the auto classifier node to create an ensemble model to predict a continuous target• Use a model to score new data

IBM Official Badges and Associated Job Roles

IBM Official Badges	Predictive Analytics Modeler: Explorer I Mastery Award
Associated Job Roles	<ul style="list-style-type: none">• Business Intelligence Analyst• Big Data Developer• Data Warehouse Developer

For more information

To learn more about this career path and others, see ibm.biz/ibmskillsacademy

To learn more about validating your technical skills with IBM Open Badges, see www.youracclaim.com

To stay informed about the IBM Skills Academy, see the following sites:

Facebook: www.facebook.com/ibmskillsacademy