

## **Predictive Analytics & Data Mining: Model Development**

### **Course Summary**

#### **Description**

The "Model Development" course dives into the data mining process at the tactical level. Participants will observe live demonstrations of machine learning methods and computer-aided pattern discovery techniques for extracting and interpreting complex patterns and relationships from large volumes of data. Attendees then participate in work-along labs that build upon an overall project.

This course is designed to be taken independently, yet is part of a larger course series that covers a 6-Phase Model Development Methodology for low-risk, high-impact projects. The scope of this course extends to the second and third phases: **Prepare** and **Build**.

You need not be an experienced statistician or mathematician to track well in this course – though traditional quantitative experts will benefit from the strategic referencing to the Plan phase and the pragmatic mind-shift required in this event. The machine learning algorithms are covered from a functional perspective. Modern software does a great job of handling the mathematical complexity. Those seeking a deep drill-down into the mathematical or theoretical underpinnings of predictive analytics algorithms should refer to other available academic offerings.

This vendor-neutral course utilizes popular commercial and open-source analytic tools in its demonstrations. The tools are used to illustrate the methods conveyed, but not to showcase the products. If you desire an intensive tactical orientation to predictive modeling methods, techniques and practice, then this event is designed for you.

#### **The Organizational Challenge For Analytics**

The rapid emergence of data processing and collection methods has propelled the IT industry into the age of big data. Organizations now contain so much data that it has become very difficult to understand just what all that data is telling us.

There is hardly a transaction that does not generate a computer record somewhere. All this data has meaning with respect to making better prospective organizational decisions and anticipating customer needs and preferences. But how do you discover those needs and preferences within data stores that contain endless amounts of seemingly incomprehensible numbers and facts? Predictive analytics does just that.

Only predictive analytics can transform the rapidly growing mass of inert data into actionable information assets. These assets can be leveraged for measurable gain through improved and automated decision-making. The practitioners who develop these transformative skills will be the leaders in their field.

The intent of this course is to present participants with a roadmap for data certification and preparation, model-building techniques, how various methods and tools apply to different kinds of data intensive problems, and how to overcome limitations that cause the majority of predictive models to under-perform.

#### **What Makes This Course Unique**

This course does not restrict or skew the presentation of data mining methods through a single product. Rather, the Model Development course gives broad consideration of the capabilities and limitations of all resources from a vendor-neutral perspective.

Live modeling demonstrations projected from the presenter's machine will precede the follow-along lab exercises. The exercises run from a design produced in the Planning phase. Participants will directly experience the natural messiness of data mining to discover what really works... as well as what doesn't and why. The instructor will show how to evaluate various features and available products based upon strengths, limitations, function, value, and general performance. The highly seasoned faculty possesses a wealth of pragmatic experience in applying predictive analytics across industries in current real-world applications. This course, like no other, insists upon making predictive analytics purposeful, measurable and actionable in a business or organizational setting.

## Predictive Analytics & Data Mining: Model Development

### Course Summary (cont'd)

#### Objectives

At the end of this course, students will be able to:

- Understand the purpose, function and impact of the 6-Phase Model Development Methodology
- Proceed through the general implementation of the two tactical phases: Prepare and Build
- Realize that model-building for actionable production need not be highly technical or complex
- Construct a valid data set and transform data for superior model performance
- Describe each of the five steps for preparing raw data for predictive analysis
- Select appropriate methods for each of the Four Core Analytic Project Types
- Assess the degree to which a model meets a predefined performance objective
- Leave with resources, contacts and actionable plans to substantially increase your analytic capabilities while minimizing dead ends

#### Topics

- Core Concepts
- Prepare Phase
- Build Phase
- Wrap-up and Next Steps

#### Audience

- **Data Scientists:** who desire to extend their analytical toolbox and underscore the scientist aspect of the role with formal process and hands-on methodological practice
- **Functional Analytic Practitioners:** Customer Relationship Managers, Risk Analysts, Business Forecasters, Statistical Analysts, Social Media and Web Data Analysts, Fraud Detection Analysts, Audit Selection Managers, Direct Marketing Analysts, Medical Diagnostic Analysts, Market Timers
- **Big Data Analysts:** who are under increasing pressure to transform their deluge of data from a liability to an asset
- **Project Leaders:** who desire to have a more detailed understanding of predictive modeling methods and techniques to better manage and interact with their practitioners
- **Business Analysts:** who must develop and interpret the models, communicate the results and make actionable recommendations
- **IT Professionals:** who wish to gain a better understanding of the data preparation, analytics and analytic sandbox development requirements to more fully support the growing demand for analytic IT support
- **Anyone Overwhelmed with Data and Starved for Actionable Insights**

#### Prerequisites

While this course is designed to be taken independently, it is important to understand its place and function within the overall Predictive Analytics & Data Mining Course Series. Registrants will be required to view a four-hour asynchronous "Core Concepts" orientation prior to attending this event. (ProTech PT0276).

#### Duration

Two days

## Predictive Analytics & Data Mining: Model Development

### Course Outline

#### I. Core Concepts

- A. Prerequisite four-hour preparatory orientation

#### II. Prepare Phase

- A. Process Objectives/Goals
- B. Sourcing and Combining Data
- C. Attribute Preparation
- D. Attribute Types (Scales)
- E. Missing Value Identification & Treatment
- F. Outlier Identification & Treatment
- G. Data Construction
- H. Data Representations
- I. Data Transformations
- J. Structuring Data for Modeling
- K. When and Why We Sample?
- L. Determining Sample Sizes
- M. Core Sampling Methods
- N. Train, Test, Validate
- O. Cross Validation
- P. Balanced Sampling
- Q. Condition Attribute Identification
- R. LAB: Data Sandbox Construction
- S. LAB: Clustering for Data Preparation

#### III. Build Phase

- A. Process Objectives/Goals
- B. Experimental Design: TRAIN Revisited
- C. Selecting Condition Attributes
- D. Analytic Model Assessment (General/Types)
- E. Statistics
- F. Tables

- G. Graphs
- H. Ensemble Modeling Conceptualization
- I. LAB: Classification Models
- J. Stage 1P – Positive Impact Models
- K. Stage 1N – Negative Impact Models
- L. Stage 2 – Conflict Resolution
- M. Stage 3 – Ranking Across the Continuum
- N. Stage 4R – Resolution Enhancement
- O. Stage 4P – Precision Enhancement
- P. Algorithm Selection for Classification Models
- Q. Logistic Regression
- R. Decision Trees
- S. K-Nearest Neighbor
- T. Neural Networks
- U. LAB: Forecasting Models
- V. Stage 3 – Ranking Across the Continuum
- W. Stage 4P – Precision Enhancement
- X. Algorithm Selection for Forecasting Models
- Y. Linear Regression
- Z. Bayesian Regression
- AA. Neural Networks
- BB. LAB: Clustering Models
- CC. Overview and Demonstration of Selected Software Tools

#### IV. Wrap-up and Next Steps

- A. Supplementary materials and resources
- B. Conferences and communities
- C. Get started on a project!
- D. Strategic Oversight and Collaborative Development