Predictive Analytics & Data Mining Strategic Implementation

Course Summary

Description

Data mining is essentially a discovery process a process riddled with common yet elusive strategic pitfalls. Project failure is rarely due to poor model development. Rather, data mining projects often fall short of their potential due to flawed or overlooked assessment, business understanding, project definition and strategic planning specifically for information discovery.

If you are looking for an intensive vendor-neutral and non-promotional introduction to data mining best practices and an approach to predictive analytics, which is critical to modeling success, then this course is designed for you. There are no prerequisites for this course. However, participants will benefit by reviewing the CRISP-DM guide ahead of the training.

"Predictive Analytics & Data Mining: Strategic Implementation" offers a concentrated presentation of capabilities, limitations, risks, rewards, use cases, best practices, strategy and lifecycle management. Those in attendance will actively step through the industry standard process for data mining and realize why an advanced degree in statistics, mathematics or computer science is no longer needed to succeed in predictive analytics. Live working sessions reveal real-world obstacles and breakthroughs from which to interpret, learn and apply.

Objectives

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

- Make better business decisions based on information hidden within your data
- Develop a strong vocabulary and understanding of data mining terminology
- Communicate with confidence among your developers and consultants
- Plan and manage your data mining projects effectively from the start
- Experience firsthand that actual model-building is not as complicated as it might have seemed through the lecture segments
- Leave with resources, contacts and actionable plans to substantially reduce your project preparation time, costs and risks

Topics

- Introduction
- Use Case Workshop #1
- Use Case Workshop #2
- Extended Modeling Topics
- Wrap-Up and Next Steps

Audience

This course is designed for IT/IS Executives and Managers, Line-Of-Business Executives and Functional Managers, Technology Planners and Consultants.

Duration

Three days

Due to the nature of this material, this document refers to numerous hardware and software products by their trade names. References to other companies and their products are for informational purposes only, and all trademarks are the properties of their respective companies. It is not the intent of ProTech Professional Technical Services, Inc. to use any of these names generically

Predictive Analytics & Data Mining Strategic Implementation

Course Outline

I. Introduction

A. What is predictive analytics?
   1. Goal driven analysis of large data sets
   2. To identify an approach for allocating organizational resources
   3. That enhances performance on the organization's self-defined performance metrics
   4. To better achieve the organization's business objectives
   5. Using a repeatable, consistent strategy
   6. Beyond traditional statistics

B. Shift your thinking

C. The goal of modeling
   1. Physical systems
   2. Human behavior

D. Behaviors of interest

E. Setting up the game
   1. Project team
   2. Phased development cycle
   3. Definitions
   4. Data sandbox

F. Formulas vs. Model development

G. The conflict between algorithm objectives and business objectives

H. Why use predictive analytics?

I. Definition of data mining
   1. What data mining is not
   2. Why mine data?
   3. The key question is "so what?"

J. Successful data mining is goal-directed analysis

K. Traditional statistics are insufficient in today's world

L. What can data mining do?
   1. Data mining opportunities
   2. Data mining business goals
   3. Data mining analytic goals

M. Why the majority of data mining projects fail

N. How much data is needed to develop a model?
   1. How many variables?
   2. Rules of thumb
   3. Types of sampling

O. Experimental design
   1. Data sets used
   2. Types of data distribution
   3. Types of decision

P. Predictive analytics key technologies overview*

Q. Methods and techniques are detailed in the Model Development course

R. Who needs brains when you've got software?

S. Low-Risk / High-ROI project design

T. The business justification for predictive analytics: Goal driven analytics

U. Organizational predictive analytics opportunity identification

V. Incremental project design
   1. Single-tailed model development: Identify positive impacts
   2. Single-tailed model development: Identify negative impacts
   3. Two-tailed model development: Conflict resolution
   4. Ranking across the continuum: Adding resolution
   5. Subdividing dimensions: Adding detail
   6. Forecasting model development

W. A 'real world' standardized development process: The CRoss-Industry Standard Process for Data Mining (CRISP-DM)

II. USE CASE WORKSHOP #1 Implement CRISP-DM for a single-tailed model

III. Business Understanding (CRISP 1)

A. Determine business objective
   1. Background and business objectives
   2. Identify decision process
   3. Business success criteria

B. Identify performance metrics

C. Calculate current baseline levels of performance

D. Determine modeling objectives
   1. Requirements
   2. Assumptions
   3. Constraints
   4. Risks and contingencies
   5. Terminology
   6. Costs and benefits

Due to the nature of this material, this document refers to numerous hardware and software products by their trade names. References to other companies and their products are for informational purposes only, and all trademarks are the properties of their respective companies. It is not the intent of ProTech Professional Technical Services, Inc. to use any of these names generically

Predictive Analytics & Data Mining Strategic Implementation

Course Outline (cont'd)

7. Modeling goals
8. Modeling success criteria

E. Assess resource availability
1. Hardware resources
2. Sources of data and knowledge
3. Personnel sources

F. Produce project plan

G. Prepare Business Understanding Deliverables

IV. Data Understanding (CRISP 2)

A. Review data availability

B. Collect initial data
1. Data requirements planning
2. Selection criteria
3. Insertion of data
4. Construction of Output variable

C. Describe data
1. Volumetric analysis of data
2. Attribute types and values
3. Keys

D. Review assumptions and goals

E. Explore data
1. Statistical analysis
2. Data exploration
3. Suppositions for future analysis

F. Verify data quality

G. Data Understanding Deliverables
1. Initial data collection report
2. Data description report
3. Data exploration report
4. Data quality report

V. Data Preparation (CRISP-DM 3)

VI. Modeling (CRISP-DM 4)

VII. Evaluation (CRISP-DM 5)

VIII. Deployment (CRISP 6)

A. Plan deployment

B. Develop monitoring and maintenance plan

C. Produce final report

D. Project review

E. Deliverables
1. Deployment plan
2. Monitoring and maintenance plan
3. Final report

IX. USE CASE WORKSHOP #2 Second CRISP-DM pass for a two-tailed model implementation

A. Business Understanding (CRISP-DM 1)

B. Data Understanding (CRISP-DM 2)

C. Data Preparation (CRISP-DM 3)

D. Modeling (CRISP-DM 4)

E. Evaluation (CRISP-DM 5)

F. Deployment (CRISP 6)

X. Extended Modeling Topics

XI. Wrap-Up and Next Steps

A. PA&DM: "Model Development" Course

B. Certification Exam (for those who complete the series)

C. Product training courses

D. Keep learning!

E. Supplementary materials and resources

F. Conferences and communities

G. Get started on a project!

Due to the nature of this material, this document refers to numerous hardware and software products by their trade names. References to other companies and their products are for informational purposes only, and all trademarks are the properties of their respective companies. It is not the intent of ProTech Professional Technical Services, Inc. to use any of these names generically