



## DATA MINING OVERVIEW

### HDT921

### Three Days

---

#### **Prerequisites**

Students should have at least some experience with coding SQL for any relational database management system plus at least a conceptual understanding of Data Warehousing.

#### **Who Should Attend**

This course is intended for users, power users, programmers, analysts, DBAs, Data Modelers, or anyone else who needs to do data mining.

#### **Course Description**

This course provides the students with the skills necessary to set up, execute, and interpret the output from data mining analysis tools. This course is based on the book: Mastering Data Mining by Michael J. A. Berry and Gordon S. Linoff published in 2000.

#### **Course Topics**

- Data Mining In Context
- Why Master The Art Of Data Mining?
- Data Mining Methodology: The Virtuous Cycle Revisited
- Customers And Their Life Cycles
- Data Mining Techniques And Algorithms
- What Data Should And Does Look Like
- Building Effective Predictive Models
- Taking Control: Setting Up A Data Mining Environment
- Who Needs Bag Balm And Pants Stretchers
- Who Gets What? Building A Best Next Offer Model For An Online Bank
- Please Don't Go! Churn Modeling In Wireless Communication
- Converging On The Customer: Understanding Customer Behavior In The Telecommunications Industry
- Who Is Buying What? Getting To Know Supermarket Shoppers
- Waste Not, Want Not: Improving Manufacturing Processes
- The Societal Context: Data Mining And Privacy



## INTRODUCTION TO DATA MINING HDT805

---

- I. DATA MINING IN CONTEXT**
  - A. What Is Data Mining?
  - B. What Can Data Mining Do?
  - C. The Business Context for Data Mining
  - D. The Technical Context for Data Mining
  - E. The Societal Context for Data Mining
  
- II. WHY MASTER THE ART OF DATA MINING?**
  - A. Four Approaches to Data Mining
  
- III. DATA MINING METHODOLOGY: THE VIRTUOUS CYCLE REVISITED**
  - A. Two Styles of Data Mining: Directed and Undirected
  - B. The Virtuous Cycle of Data Mining
  - C. Identifying the Right Business Problem
  - D. Transforming Data into Actionable Results
  - E. Acting on the Results
  - F. Measuring the Model's Effectiveness
  - G. What Makes Predictive Modeling Successful?
  
- IV. CUSTOMERS AND THEIR LIFE CYCLES**
  - A. Who Is the Customer?
  - B. The Customer Lifecycle
  - C. Targeting the Right Customers at the Right Time
  
- V. DATA MINING TECHNIQUES AND ALGORITHMS**
  - A. Different Goals Call for Different Techniques
  - B. Three Data Mining Techniques
  - C. Automatic Cluster Detection
  - D. Decision Trees
  - E. Neural Networks



## INTRODUCTION TO DATA MINING HDT805

---

- VI. WHAT DATA SHOULD AND DOES LOOKS LIKE**
  - A. What Should Data Look Like?
  - B. What Does Data Really Look Like?
  - C. How Much Data Is Enough?
  - D. Derived Variables
  - E. Case Study: Defining Customer Behavior
  - F. Dirty Data
  
- VII. BUILDING EFFECTIVE PREDICTIVE MODELS**
  - A. Building Good Predictive Models
  - B. Working with the Model Set
  - C. Using Multiple Models
  - D. Experiment!
  
- VIII. TAKING CONTROL: SETTING UP A DATA MINING ENVIRONMENT**
  - A. Getting Started
  - B. Case 1: Building Up a Core Competency Internally
  - C. Case 2: Building a New Line of Business
  - D. Case 3: Building Data Mining Skills on Data Warehouse Efforts
  - E. Case 4: Data Mining Using Tessera RME
  
- IX. WHO NEEDS BAG BALM AND PANTS STRETCHERS**
  - A. The Vermont Country Store
  - B. The Business Problem
  - C. The Data
  - D. The Technical Approach
  - E. The Future
  
- X. WHO GETS WHAT? BUILDING A BEST NEXT OFFER MODEL FOR AN ONLINE BANK**
  - A. Gaining Wallet Share
  - B. The Business Problem
  - C. The Data
  - D. Approach to the Problem
  - E. Building the Models
  - F. In a More Perfect World



## INTRODUCTION TO DATA MINING HDT805

---

- XI. PLEASE DON'T GO! CHURN MODELING IN WIRELESS COMMUNICATION**
  - A. The Wireless Telephone Industry
  - B. The Business Problem
  - C. Building a Churn Model: A Real-Life Application
  - D. The Data
  - E. Lessons about Building Churn Models
  
- XII. CONVERGING ON THE CUSTOMER: UNDERSTANDING CUSTOMER BEHAVIOR IN THE TELECOMMUNICATIONS INDUSTRY**
  - A. Dataflows
  - B. The Business Problem
  - C. The Data
  - D. A Voyage of Discovery
  
- XIII. WHO IS BUYING WHAT? GETTING TO KNOW SUPERMARKET SHOPPERS**
  - A. An Industry in Transition
  - B. Three Case Studies
  - C. Analyzing Ethnic Purchasing Patterns
  - D. Who Buys Yogurt at the Supermarket?
  - E. Who Buys Meat at the Health Food Store?
  
- XIV. WASTE NOT, WANT NOT: IMPROVING MANUFACTURING PROCESSES**
  - A. Data Mining to Reduce Costs at R. R. Donneley
  - B. Reducing Paper Wastage at Time Inc.
  
- XV. THE SOCIETAL CONTEXT: DATA MINING AND PRIVACY**
  - A. The Privacy Prism
  - B. Is Data Mining a Threat?
  - C. The Expectation of Privacy
  - D. Information in the Material World
  - E. Information in the Electronic World
  - F. The Promise of Data Mining