

DATA MINING OVERVIEW

SPSS

□ Platform: Windows □ Duration: 2 – 3 Hours

Target Audience

Business analysts and others who have heard of data mining and want to know what it entails. Those who feel data mining might be beneficial to their company, but don't know where to begin. Those who ask, "what do we need to consider if we start a data mining project."

Prerequisites

An interest in Data Mining.

Overview

This course is designed to provide a high-level view or roadmap to those considering or beginning a data mining project. Thus, while it will define and discuss each main step of the data mining process, it will not cover any step in exhaustive detail; it is simply a general overview of the concepts. For example, although it is organized around the Cross Industry Standard Process for Data Mining methodology (CRISP-DM), because all the different data mining methodologies, such as Berry and Linoff's Virtuous Circle, share most of the same core elements anyone interested in Data Mining should benefit from the course. In addition, although we will present examples of output from a variety of different SPSS software, including SPSS, AnswerTree, and Clementine, it is not necessary that you own all three of these packages. In many cases you can perform the same data mining task in two or even all three of them.

Course Content

Lesson Title	Lesson Description
Lesson 1 Introduction to Data Mining	This lesson describes what Data Mining is, summarizes the history of data mining and introduces CRISP methodology as a standard framework for performing a data mining project.
Lesson 2 Business Understanding	This lesson covers the Business Understanding phase of the CRISP methodology in which you define what business questions need to be answered and outlines the requirements for a successful answer.
Lesson 3 Framing a Research Question	This lesson discusses how your business questions must be phrased or converted into analysis questions that can be answered using available methods.
Lesson 4 Data Understanding and Preparation	This lesson discusses the main issues faced when thinking about data access in the context of data mining.
Lesson 5 Checking and Transforming Data	This lesson covers the issue of data cleaning and the resolution of data problems that must be accomplished before the analysis phase.
Lesson 6a Modeling: Statistical Data Mining Techniques	This lesson covers various statistical techniques that are commonly used in data mining, including Discriminant Analysis, Linear Regression, and Analysis of Variance. Additional topics covered are Logistic and Multinomial Regression, and also Time Series Analysis
Lesson 7a Modeling: Automated Data Mining Methods	This lesson covers automated modeling methods, including Cluster Analysis, Market Basket Analysis and Sequence Analysis. Additional topics covered are MARS Models, Decision Trees, Neural Networks, and Text Mining.
Lesson 8 Evaluation and Deployment	The objective of this lesson is to look at important issues that relate to the analysis you perform, your evaluation of it, and the deployment of your model