

DAY 1

Data Mining Description

- What is Data Mining?
- What Can Data Mining Do?
- How is Data Mining Used?
- Data Mining Process ([CRISP-DM](#))

Designing the Data Mining Project

- Business Questions
- Data Mining Objectives
- Identifying Data for Data Mining

Data Characterization

- Univariate and Multivariate Descriptions of Data
- Cleaning and Conditioning Data
 - Missing Data, Outliers, and Variable Formatting
- Effective Use of Data Visualization
- **Demonstration #1**

Feature Creation and Selection

- Binning Real-valued and Categorical Variables
- Correcting Problems with Variable Distributions
- Removing Redundant Variables
- **Demonstration #2**

Creating Data Samples

- Ways to Sample Data
 - Random and Stratified Sampling
 - Cross-validation
 - Bootstrapping
- How Much Data is Needed?

Model Scoring and Deployment

- How algorithms score models
 - Estimation
 - Classification
- How users (should) score models

DAY 2

Overview and Comparison of Data Mining Algorithms

- Predictive vs. Descriptive Models
- Data Mining Algorithm Taxonomy
- Data Mining vs. OLAP

Supervised Learning Algorithm Descriptions and Tips

- Decision Trees
- **Demonstration #3**
- Linear and Logistic Regression

- Neural Networks
- **Demonstration #4**
- Other Methods

Unsupervised Learning Algorithm Descriptions and Tips

- Association Rules
- Clustering
- Kohonen Self-Organizing Maps
- **Demonstration #5**

Model Ensembles and Deployment

- What are model ensembles?
 - Bagging, Boosting and Other Killer B's
- How to Create Model Ensembles
- Model Deployment

Data Mining Software Tools

- Types of Tools
 - General Purpose vs. Industry Specific
 - Free Tools vs. Commercial Tools
- Feature Comparisons

Data Mining Resources