

COURSE OUTLINE

BUSINESS KNOWLEDGE SERIES

Advanced Analytics in a Big Data World

OVERVIEW

Presented by Bart Baesens, Ph.D. or Christophe Mues, Ph.D., Professors at the School of Management, or Cristian Bravo, Ph.D., Assistant Professor, Business Analytics, University of Southampton (UK); or Wouter Verbeke, Ph.D., Assistant Professor, Business Informatics, University of Brussels (Belgium); or Stefan Lessmann, Ph.D., Professor, School of Business and Economics, Humboldt University (Germany)

In today's big data world, many companies have gathered huge amounts of customer data about marketing success, use of financial services, online website usage, and even fraud behavior. Given recent trends and needs such as mass customization, personalization, Web 2.0, one-to-one marketing, risk management, and fraud detection, it becomes increasingly important to extract, understand, and exploit analytical patterns of customer behavior and strategic intelligence. This course helps clarify how to successfully adopt recently proposed state-of-the-art analytical and data science techniques for advanced customer intelligence applications. This highly interactive course provides a sound mix of both theoretical and technical insights as well as practical implementation details and is illustrated by several real-life cases. The instructor will extensively report on both his research and consulting experience in the field. References to background material such as selected papers, tutorials, and guidelines are also provided.

Learn how to:

- apply a series of powerful, recently developed, cutting-edge analytical and data science techniques
- ensure the practical application of these techniques to optimize strategic business processes and decision making
- explore a futuristic vision of how emerging data science techniques might change your key business processes
- deploy, monitor, and optimally backtest analytical models.

Who should attend:

Those involved in estimating, monitoring, auditing, or maintaining models for various types of customer intelligence; those involved with using data mining techniques for various types of customer intelligence, job titles including business analysts in various settings (e.g. risk management, manufacturing, telco, retail, advertising, public, pharmaceutical, and so on), marketing/CRM managers, fraud managers, customer intelligence managers, risk analysts, CRM analysts, marketing analysts, senior data analysts, and data miners

Prerequisites:

Before attending this course, you should know how to

- preprocess data (such as sampling, missing values, outliers, categorization, and so on)
- develop predictive models using logistic regression
- develop predictive models using decision trees
- develop descriptive models using basic segmentation techniques
- quantify the performance of predictive models (such as lift curves, ROC curves, and so on).

Outline:

Refresher: The Analytics Process Model

- basic nomenclature (definition of customer, definition of target, and so on)
- data collection and preprocessing (sampling, missing values, outliers, categorization, weights of evidence, and so on)
- predictive versus descriptive analytics (data science)
- putting analytics to work
- analytic model requirements (performance, interpretability, operational efficiency, economical cost, compliance)
- key application areas (CRM, risk management, fraud, on-line analytics)

Refresher: Decision Trees

- splitting/stopping/assignment decision
- key algorithms: C4.5 (See5), CART, CHAID
- recommendations for using decision trees in a business context

Regression Trees

- splitting/stopping/assignment decision
- case study: using regression trees for loss forecasting

Ensemble Methods

- bootstrapping
- bagging
- boosting
- random forests

Alternative Rule Representation Formats

- rule types (propositional, oblique, M-of-N, fuzzy, and so on)
- decision tables (lexicographical ordering, contraction methods, and so on)
- decision diagrams
- case study: decision tables and diagrams for customer scoring
- case study: decision tables for textual knowledge verification

Neural Networks

- multilayer perceptrons (MLPs)
- MLP types (RBF, recurrent, and so on)
- weight learning (backpropagation, conjugate gradient, and so on)
- overfitting, early stopping, and weight regularization
- architecture selection (grid search, SNC, and so on)
- input selection (Hinton graphs, likelihood statistics, brute force, and so on)
- self-organizing maps (SOMs) for clustering
- case study: using SOMs for country corruption analysis

Support Vector Machines (SVMs)

- linear programming
- the kernel trick and Mercer theorem
- SVMs for classification and regression
- multiclass SVMs (one- versus-one, one-versus-all coding)
- hyperparameter tuning using cross-validation methods
- case study: benchmarking SVM classifiers

Opening up the Neural Network and SVM Black Box

- business applications of neural networks and SVMs
- rule extraction methods (pedagogical versus decompositional approaches such as Neurorule, Neurolinear, Trepan, and so on)
- two-stage models (combining white-box linear models with black-box high-performing neural networks)
- case studies

Bayesian Network Classifiers

- Naive Bayes
- Tree augmented Naive Bayes (TAN)
- unrestricted Bayesian network classifiers
- Bayesian inference
- case study: Bayesian networks for churn prediction

Survival Analysis

- censoring
- time-varying covariates
- survival probabilities versus hazard rates
- Kaplan Meier analysis
- parametric survival analysis (evaluating model fit, maximum likelihood estimation, and so on)
- proportional hazards regression (partial likelihood, estimating survivor functions, and so on)
- time varying covariates
- competing risks
- mixture cure modeling
- evaluating survival analysis models (ROC analytics, interpretability, and so on)

Social Network Learning and Inference

- social networks and applications (churn, email, fraud, and so on)
- social network components (nodes, edges, weights)
- representing social networks (sociograms, adjacency matrix, adjacency list)
- social network metrics (geodesic, closeness, betweenness, and so on)
- community mining (Girvan-Newman algorithm, graph partitioning approaches, min cut, ratio cut, min max cut, and so on)
- social network based inference
- relational classifiers (relational nearest neighbor classifier, probabilistic relational nearest neighbor classifier, relational logistic regression)

- featurization
- collective inference procedures (Gibbs sampling, Iterative Classification, Google PageRank)
- bipartite networks
- multipartite networks
- case study: Gotcha fraud detection tool

Evaluating and Monitoring Analytical Models

- quantitative evaluation
- data set split up (out of sample, out of time, out of universe)
- Traffic Light method
- backtesting classification models (SSI, ROC, Brier score, binomial test, Hosmer-Lemeshow test, and so on)
- backtesting regression models (SSI, correlation, MSE, CAP curve, REC curve, and so on)
- benchmarking (Champion-Challenger approach, Spearman rank order correlation, Kendall's tau, Goodman-Kruskal gamma)
- qualitative evaluation
- data quality (data accuracy, data completeness, data recency, and so on)
- model design
- documentation
- corporate governance and management oversight

Other Learning Algorithms and Applications (Short)

- semi-supervised learning
- fuzzy techniques
- evolutionary algorithms
- ant colony optimization
- online analytics applications
- social media analytics applications
- process analytics applications