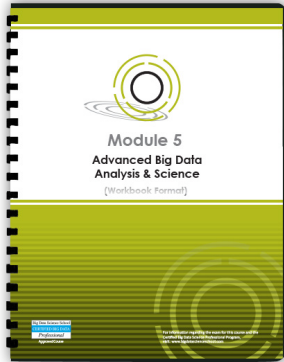


BDSCP Module 5: Advanced Big Data Analysis & Science

This course delves into a range of advanced data analysis practices and analysis techniques that are explored within the context of Big Data. The course content focuses on topics that enable participants to develop a thorough understanding of statistical, modeling, and analysis techniques for data patterns, clusters, and text analytics, as well as the identification of outliers and errors that affect the significance and accuracy of predictions made on Big Data datasets.

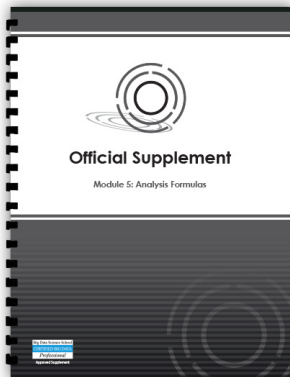
The following primary topics are covered:

- Statistical Models, Model Evaluation Measures (including cross-validation, bias-variance, confusion matrix & f-score)
- Machine Learning Algorithms, Pattern Identification (including association rules & apriori algorithm)
- Advanced Statistical Techniques (including parametric vs. non-parametric, clustering vs. non-clustering distance-based, supervised vs. semi-supervised)
- Linear Regression & Logistic Regression for Big Data
- Decision Trees for Big Data
- Classification Rules for Big Data
- K Nearest Neighbor (kNN) for Big Data
- Naïve Bayes for Big Data
- Association Rules for Big Data
- K-means for Big Data
- Text Analytics for Big Data
- Outlier Detection for Big Data



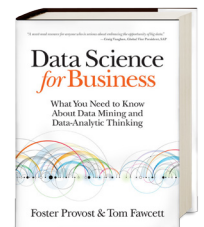
Duration: 1 Day

For more information, visit www.bigdatascienceschool.com.



Text Book

This BDSCP course module covers a range of in-depth topics that are described in the course booklet(s) and further elaborated by detailed technical coverage and case study examples in the accompanying *Data Science for Business* text book.



Self-Study Kit

The materials for this course module can be purchased separately as part of the Module 5 Self-Study Kit, which includes additional materials and study aids. These materials are designed to prepare you for Exam B90.05 and are also suitable for general remote, self-paced study purposes.

For ordering information, visit www.bigdatascienceschool.com/store.

