



Predictive Analytics Modules and Learning Objectives

0 – Introduction to R (available with all other modules as background)

- Write and execute basic commands in R and RStudio
- Use best practices to write code in R

1 – Problem Definition and Project Management

- Determine whether a problem should be addressed with predictive analytics or traditional analysis techniques.
- Translate a vague question into one that can be analyzed with data, statistics, and predictive analytics to solve a business problem.
- Use case design and evaluation/prioritization based on available data and technology, significance of business impact, and implementation considerations to define the problem.
- Implement technology to efficiently utilize statistical and predictive analytics techniques, taking into account problem objectives and implementation constraints.
- Explain the key principles involved in creating and managing an effective predictive modeling team that can successfully manage your project from problem definition to implementation.

2 – Data Sources, Structures, and Manipulation

- Identify sources of data and the challenges created.
- Identify structured, unstructured, and semistructured data types.
- Read data from a variety of file formats and save data as a csv file.
- Identify the types of variables and terminology used in predictive modeling.
- Evaluate the quality of appropriate data sources for a problem.
- Employ common methods for cleaning data.
- Identify the regulations, standards, and ethics surrounding predictive modeling and data collection.
- Implement effective data design.
- Use common data blending techniques.

3 – Data Exploration and Visualization

- Describe and apply common data visualization techniques.
- Identify data anomalies and outliers using univariate exploration techniques.
- Bivariate data exploration. Use bivariate exploration to determine relationships, calculate correlation, and investigate conditional means.

4 – Feature Generation

- Describe and apply common data transformation techniques.
- Identify relationships among multiple variables using principal component analysis.
- Identify relationships and structure among multiple variables using clustering techniques.

- Explain the differences between features and variables and apply prior knowledge to create features.
- Understand various approaches to creating variables for modeling text data.

5 – Feature Selection

- Describe various filter-based selection techniques and then features.
- Apply algorithm-based selection and data mining techniques to select features.

6 – Model Development and Validation

- Differentiate types of business problems and understand their impact on model development and validation.
- Explain the limitations of traditional analytics techniques.
- Explain the difference between supervised and unsupervised learning.
- Explain the concepts of bias, variance, and model complexity, the bias-variance tradeoff, and its implications for building robust models.
- Explain cross-validation and the use of training, testing, and validation sets.
- Describe the different analytics techniques and the key dimensions of each.
- Construct a basic decision tree and a basic generalized linear model

7 – Further Topics in Model Selection and Assessment

- Apply various methods of assessing model performance and selection of the best model.
- Understand the advantages and disadvantages of cross validation and apply it during modeling.

8 – Further Topics in Decision Trees and Generalized Linear Models

- Apply decision trees in the contexts of classification and regression. Apply cost-complexity pruning.
- Apply generalized linear models to regression problems, including the use of weights and offsets, interactions, and lift charts.