Analytical Modeling, Evaluation, and Deployment Best Practices

by Deanne Larson

Module 0. About the Course (2 min)

Module 1. The Modeling Process (30 min)

- *The Data Science Process*
- *Types of Data Science Projects*
- *Modeling*
- *Project Type and Maturity*
    - Descriptive Analytics
    - Diagnostic Analytics
    - Predictive Analytics
    - Prescriptive Analytics
- Data Science Starting Point
    - Modeling Process: Define the Scope
    - Modeling Process: Collect the Data
    - Modeling Process: Explore the Data
    - Modeling Process: Data Reduction and Cleansing
    - Modeling Process: Build the Model
    - Modeling Process: Evaluation and Interpretation
    - Modeling Process: Model Deployment
- *Other Modeling Considerations*

Module 2. Overview of Common Algorithms and Uses (68 min)

- *Data Science Framework*
- *Approaches*
- *Techniques*
    - *Classification*
    - *Association*
    - *Sequencing*
    - *Forecasting*
- Algorithms
    - Decision Trees
    - Decision Tree Uses
    - K Nearest Neighbor: Part 1 & 2
    - K Nearest Neighbor Uses
    - Probability – Bayes Classification: Part 1-4
    - Neural Networks: Part 1-2
    - Neural Network Uses
    - Support Vector Machine: Part 1-2
    - Support Vector Machine Uses
    - Support Vector Machine vs. Neural Network
    - Statistical Learning
    - Linear Regression: Part 1 & 2
    - Linear Regression – Error Difference
    - Linear Regression Uses
    - Logistic Regression
    - Logistic Regression Uses
    - Other Regression Options
    - Ensemble Methods
    - Example of Ensemble Bagging
    - Example of Ensemble Boosting

  - Bagging and Boosting
  - Ensemble Method Uses
  - Clustering
  - Clustering Uses
  - Association – Market Basket
  - Association Rules
  - Association Uses
- *Anomaly Detection*
  - Model Selection
  - Many Models & Algorithms
  - Application of Analytics Models

Module 3. Tools for Model Evaluation (24 min)

- *Evaluation*
- *Bias/Variance Tradeoff*
- *Train and Test Sets*
- *Assessment of Results*
- *Hold-out Cross Validation*
- *K-fold Cross Validation Method*
- *Regression – Mean Squared Error*
- *Linear Regression Confidence and Prediction Intervals*
- *Logistic Regression – Significance Test*
- *Classification Accuracy*
- *Classification Accuracy – Other Measures*
- *Prediction Error Methods*
- *ROC Curve*
- *Evaluation – Customer Acceptance*

Module 4. Preparing for Deployment (11 min)

- *Deployment*
- *Deployment – Working Software*
- *Data Pipelines*
- *Data Pipelines – Part of Deployment: Part 1 – 2*
- *Deployment Operationalization*
- *Monitoring Models – Dashboard*
- *Life of the Model*

Module 5. Model Operations (18 min)

- *Model Operations: Part 1 – 2*
- *Data Ingestion*
- *Data Storage*
- *Data Integration and Synthesis*
- *Data Visualization*
- *Model Accuracy*
- *Model Retraining*
- *Model Retiring*
- *Machine Learning in Action*

Module 6. Model Metrics (29 min)

- *Machine Learning Metrics*
- *Metrics for Supervised Learning*
- *Classification Model Metrics*
    - Classification Model Metrics – Accuracy
    - Classification Model Metrics – Specificity
    - Classification Model Metrics – Log Loss
    - Classification Model Metrics – ROC
    - Classification Model Metrics – Area Under Curve
    - Classification Model Metrics – Ranking
    - Classification Model Metrics – Precision & Recall
    - Classification Model Metrics – Precision & Recall Curve
    - Reading the PRC
- *Normalized Discount Cumulative Gain (NDCG)*
- *Discount Cumulative Gain (DCG)*
- *Normalized Discount Cumulative Gain (NDCG) – Example*
- *Root Mean Squared Error*
- *Quantities of Error*