uCertify

Course Outline

Data Mining and Predictive Analysis

26 Apr 2024

1. Course Objective
2. Pre-Assessment
3. Exercises, Quizzes, Flashcards & Glossary

   Number of Questions
4. Expert Instructor-Led Training
5. ADA Compliant & JAWS Compatible Platform
6. State of the Art Educator Tools
7. Award Winning Learning Platform (LMS)
8. Chapter & Lessons

   Syllabus

   Chapter 1: Preface

   Chapter 2: An Introduction to Data Mining and Predictive Analytics

   Chapter 3: Data Preprocessing

   Chapter 4: Exploratory Data Analysis

   Chapter 5: Dimension-Reduction Methods

   Chapter 6: Univariate Statistical Analysis

   Chapter 7: Multivariate Statistics

   Chapter 8: Preparing to Model the Data

   Chapter 9: Simple Linear Regression

   Chapter 10: Multiple Regression and Model Building

   Chapter 11: k-Nearest Neighbor Algorithm

   Chapter 12: Decision Trees

   Chapter 13: Neural Networks

   Chapter 14: Logistic Regression

   Chapter 15: NaïVe Bayes and Bayesian Networks

   Chapter 16: Model Evaluation Techniques

   Chapter 17: Cost-Benefit Analysis Using Data-Driven Costs

   Chapter 18: Cost-Benefit Analysis for Trinary and -Nary Classification Models

Chapter 19: Graphical Evaluation of Classification Models
Chapter 20: Hierarchical and k-Means Clustering
Chapter 21: Kohonen Networks
Chapter 22: BIRCH Clustering
Chapter 23: Measuring Cluster Goodness
Chapter 24: Association Rules
Chapter 25: Segmentation Models
Chapter 26: Ensemble Methods: Bagging and Boosting
Chapter 27: Model Voting and Propensity Averaging
Chapter 28: Genetic Algorithms
Chapter 29: Imputation of Missing Data
Chapter 30: Case Study, Part 1: Business Understanding, Data Preparation, and EDA
Chapter 31: Case Study, Part 2: Clustering and Principal Components Analysis
Chapter 32: Case Study, Part 3: Modeling And Evaluation For Performance And Interpretability
Chapter 33: Case Study, Part 4: Modeling and Evaluation for High Performance Only
Chapter 34: Appendix A

Videos and How To

9. Practice Test

Here's what you get

Features

10. Live labs

Lab Tasks

Here's what you get

1. Course Objective

Data mining is the process of discovering useful patterns and trends in large data sets and predictive analytics is the process of extracting information from large data sets in order to make predictions and estimates about future outcomes. Data mining is becoming more widespread every day, because it empowers companies to uncover profitable patterns and trends from their existing databases. With uCertify's course Data mining and predictive analysis, you get a hands-on experience in data mining and you will learn what types of analysis will uncover the most profitable nuggets of knowledge from the data, while avoiding the potential pitfalls that may cost your company millions of dollars.

2. Pre-Assessment

Pre-Assessment lets you identify the areas for improvement before you start your prep. It determines what students know about a topic before it is taught and identifies areas for improvement with question assessment before beginning the course.

3. Exercises

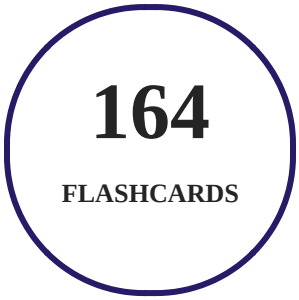
There is no limit to the number of times learners can attempt these. Exercises come with detailed remediation, which ensures that learners are confident on the topic before proceeding.

4. Quizzes

Quizzes test your knowledge on the topics of the exam when you go through the course material. There is no limit to the number of times you can attempt it.

5. flashcards

Flashcards are effective memory-aiding tools that help you learn complex topics easily. The flashcard will help you in memorizing definitions, terminologies, key concepts, and more. There is no limit to the number of times learners can attempt these. Flashcards help master the key concepts.



6. Glossary of terms

uCertify provides detailed explanations of concepts relevant to the course through Glossary. It contains a list of frequently used terminologies along with its detailed explanation. Glossary defines the key terms.

7. Expert Instructor-Led Training

uCertify uses the content from the finest publishers and only the IT industry's finest instructors. They have a minimum of 15 years real-world experience and are subject matter experts in their fields. Unlike a live class, you can study at your own pace. This creates a personal learning experience and gives you all the benefit of hands-on training with the flexibility of doing it around your schedule 24/7.

8. ADA Compliant & JAWS Compatible Platform

uCertify course and labs are ADA (Americans with Disability Act) compliant. It is now more accessible to students with features such as:

- Change the font, size, and color of the content of the course
- Text-to-speech, reads the text into spoken words
- Interactive videos, how-tos videos come with transcripts and voice-over
- Interactive transcripts, each word is clickable. Students can clip a specific part of the video by clicking on a word or a portion of the text.

JAWS (Job Access with Speech) is a computer screen reader program for Microsoft Windows that reads the screen either with a text-to-speech output or by a Refreshable Braille display. Student can easily navigate uCertify course using JAWS shortcut keys.

9. State of the Art Educator Tools

uCertify knows the importance of instructors and provide tools to help them do their job effectively. Instructors are able to clone and customize course. Do ability grouping. Create sections. Design grade scale and grade formula. Create and schedule assessments. Educators can also move a student from self-paced to mentor-guided to instructor-led mode in three clicks.

10. Award Winning Learning Platform (LMS)

uCertify has developed an award winning, highly interactive yet simple to use platform. The SIIA CODiE Awards is the only peer-reviewed program to showcase business and education technology's finest products and services. Since 1986, thousands of products, services and solutions have been

recognized for achieving excellence. uCertify has won CODiE awards consecutively for last 7 years:

- **2014**
    1. Best Postsecondary Learning Solution

- **2015**
    1. Best Education Solution
    2. Best Virtual Learning Solution
    3. Best Student Assessment Solution
    4. Best Postsecondary Learning Solution
    5. Best Career and Workforce Readiness Solution
    6. Best Instructional Solution in Other Curriculum Areas
    7. Best Corporate Learning/Workforce Development Solution

- **2016**
    1. Best Virtual Learning Solution
    2. Best Education Cloud-based Solution
    3. Best College and Career Readiness Solution
    4. Best Corporate / Workforce Learning Solution
    5. Best Postsecondary Learning Content Solution
    6. Best Postsecondary LMS or Learning Platform
    7. Best Learning Relationship Management Solution

- **2017**
    1. Best Overall Education Solution
    2. Best Student Assessment Solution
    3. Best Corporate/Workforce Learning Solution
    4. Best Higher Education LMS or Learning Platform

- **2018**
    1. Best Higher Education LMS or Learning Platform

2. Best Instructional Solution in Other Curriculum Areas
3. Best Learning Relationship Management Solution

- **2019**
  1. Best Virtual Learning Solution
  2. Best Content Authoring Development or Curation Solution
  3. Best Higher Education Learning Management Solution (LMS)

- **2020**
  1. Best College and Career Readiness Solution
  2. Best Cross-Curricular Solution
  3. Best Virtual Learning Solution

11. Chapter & Lessons

uCertify brings these textbooks to life. It is full of interactive activities that keeps the learner engaged. uCertify brings all available learning resources for a topic in one place so that the learner can efficiently learn without going to multiple places. Challenge questions are also embedded in the chapters so learners can attempt those while they are learning about that particular topic. This helps them grasp the concepts better because they can go over it again right away which improves learning.

Learners can do Flashcards, Exercises, Quizzes and Labs related to each chapter. At the end of every lesson, uCertify courses guide the learners on the path they should follow.

Syllabus

Chapter 1: Preface

- What is Data Mining? What is Predictive Analytics?
- Why is this Course Needed?
- Who Will Benefit from this Course?

- Danger! Data Mining is Easy to do Badly
- “White-Box” Approach
- Algorithm Walk-Throughs
- Exciting New Topics
- The R Zone
- Appendix: Data Summarization and Visualization
- The Case Study: Bringing it all Together
- How the Course is Structured

Chapter 2: An Introduction to Data Mining and Predictive Analytics

- What is Data Mining? What Is Predictive Analytics?
- Wanted: Data Miners
- The Need For Human Direction of Data Mining
- The Cross-Industry Standard Process for Data Mining: CRISP-DM
- Fallacies of Data Mining
- What Tasks can Data Mining Accomplish
- The R Zone
- R References

- Exercises

Chapter 3: Data Preprocessing

- Why do We Need to Preprocess the Data?
- Data Cleaning
- Handling Missing Data
- Identifying Misclassifications
- Graphical Methods for Identifying Outliers
- Measures of Center and Spread
- Data Transformation
- Min–Max Normalization
- Z-Score Standardization
- Decimal Scaling
- Transformations to Achieve Normality
- Numerical Methods for Identifying Outliers
- Flag Variables
- Transforming Categorical Variables into Numerical Variables
- Binning Numerical Variables

- Reclassifying Categorical Variables
- Adding an Index Field
- Removing Variables that are not Useful
- Variables that Should Probably not be Removed
- Removal of Duplicate Records
- A Word About ID Fields
- The R Zone
- R Reference
- Exercises

Chapter 4: Exploratory Data Analysis

- Hypothesis Testing Versus Exploratory Data Analysis
- Getting to Know The Data Set
- Exploring Categorical Variables
- Exploring Numeric Variables
- Exploring Multivariate Relationships
- Selecting Interesting Subsets of the Data for Further Investigation
- Using EDA to Uncover Anomalous Fields

- Binning Based on Predictive Value
- Deriving New Variables: Flag Variables
- Deriving New Variables: Numerical Variables
- Using EDA to Investigate Correlated Predictor Variables
- Summary of Our EDA
- The R Zone
- R References
- Exercises

Chapter 5: Dimension-Reduction Methods

- Need for Dimension-Reduction in Data Mining
- Principal Components Analysis
- Applying PCA to the Houses Data Set
- How Many Components Should We Extract?
- Profiling the Principal Components
- Communalities
- Validation of the Principal Components
- Factor Analysis

- Applying Factor Analysis to the Adult Data Set
- Factor Rotation
- User-Defined Composites
- An Example of a User-Defined Composite
- The R Zone
- R References
- Exercises

Chapter 6: Univariate Statistical Analysis

- Data Mining Tasks in Discovering Knowledge in Data
- Statistical Approaches to Estimation and Prediction
- Statistical Inference
- How Confident are We in Our Estimates?
- Confidence Interval Estimation of the Mean
- How to Reduce the Margin of Error
- Confidence Interval Estimation of the Proportion
- Hypothesis Testing for the Mean
- Assessing The Strength of Evidence Against The Null Hypothesis

- Using Confidence Intervals to Perform Hypothesis Tests
- Hypothesis Testing for The Proportion
- Reference
- The R Zone
- R Reference
- Exercises

Chapter 7: Multivariate Statistics

- Two-Sample t-Test for Difference in Means
- Two-Sample Z-Test for Difference in Proportions
- Test for the Homogeneity of Proportions
- Chi-Square Test for Goodness of Fit of Multinomial Data
- Analysis of Variance
- Reference
- The R Zone
- R Reference
- Exercises

Chapter 8: Preparing to Model the Data

- Supervised Versus Unsupervised Methods
- Statistical Methodology and Data Mining Methodology
- Cross-Validation
- Overfitting
- Bias–Variance Trade-Off
- Balancing The Training Data Set
- Establishing Baseline Performance
- The R Zone
- R Reference
- Exercises

Chapter 9: Simple Linear Regression

- An Example of Simple Linear Regression
- Dangers of Extrapolation
- How Useful is the Regression? The Coefficient of Determination, r2
- Standard Error of the Estimate, s
- Correlation Coefficient r
- Anova Table for Simple Linear Regression

- Outliers, High Leverage Points, and Influential Observations
- Population Regression Equation
- Verifying The Regression Assumptions
- Inference in Regression
- t-Test for the Relationship Between x and y
- Confidence Interval for the Slope of the Regression Line
- Confidence Interval for the Correlation Coefficient ?
- Confidence Interval for the Mean Value of y Given x
- Prediction Interval for a Randomly Chosen Value of y Given x
- Transformations to Achieve Linearity
- Box–Cox Transformations
- The R Zone
- R References
- Exercises

Chapter 10: Multiple Regression and Model Building

- An Example of Multiple Regression
- The Population Multiple Regression Equation

- Inference in Multiple Regression
- Regression With Categorical Predictors, Using Indicator Variables
- Adjusting R2: Penalizing Models For Including Predictors That Are Not Useful
- Sequential Sums of Squares
- Multicollinearity
- Variable Selection Methods
- Gas Mileage Data Set
- An Application of Variable Selection Methods
- Using the Principal Components as Predictors in Multiple Regression
- The R Zone
- R References
- Exercises

Chapter 11: k-Nearest Neighbor Algorithm

- Classification Task
- k-Nearest Neighbor Algorithm
- Distance Function
- Combination Function

- Quantifying Attribute Relevance: Stretching the Axes
- Database Considerations
- k-Nearest Neighbor Algorithm for Estimation and Prediction
- Choosing k
- Application of k-Nearest Neighbor Algorithm Using IBM/SPSS Modeler
- The R Zone
- R References
- Exercises

Chapter 12: Decision Trees

- What is a Decision Tree?
- Requirements for Using Decision Trees
- Classification and Regression Trees
- C4.5 Algorithm
- Decision Rules
- Comparison of the C5.0 and CART Algorithms Applied to Real Data
- The R Zone
- R References

- Exercises

Chapter 13: Neural Networks

- Input and Output Encoding
- Neural Networks for Estimation and Prediction
- Simple Example of a Neural Network
- Sigmoid Activation Function
- Back-Propagation
- Gradient-Descent Method
- Back-Propagation Rules
- Example of Back-Propagation
- Termination Criteria
- Learning Rate
- Momentum Term
- Sensitivity Analysis
- Application of Neural Network Modeling
- The R Zone
- R References

- Exercises

Chapter 14: Logistic Regression

- Simple Example of Logistic Regression
- Maximum Likelihood Estimation
- Interpreting Logistic Regression Output
- Inference: Are the Predictors Significant?
- Odds Ratio and Relative Risk
- Interpreting Logistic Regression for a Dichotomous Predictor
- Interpreting Logistic Regression for a Polychotomous Predictor
- Interpreting Logistic Regression for a Continuous Predictor
- Assumption of Linearity
- Zero-Cell Problem
- Multiple Logistic Regression
- Introducing Higher Order Terms to Handle Nonlinearity
- Validating the Logistic Regression Model
- WEKA: Hands-On Analysis Using Logistic Regression
- The R Zone

- R References
- Exercises

Chapter 15: NaïVe Bayes and Bayesian Networks

- Bayesian Approach
- Maximum A Posteriori (MAP) Classification
- Posterior Odds Ratio
- Balancing The Data
- Naïve Bayes Classification
- Interpreting The Log Posterior Odds Ratio
- Zero-Cell Problem
- Numeric Predictors for Naïve Bayes Classification
- WEKA: Hands-on Analysis Using Naïve Bayes
- Bayesian Belief Networks
- Clothing Purchase Example
- Using The Bayesian Network to Find Probabilities
- The R Zone
- R References

- Exercises

Chapter 16: Model Evaluation Techniques

- Model Evaluation Techniques for the Description Task
- Model Evaluation Techniques for the Estimation and Prediction Tasks
- Model Evaluation Measures for the Classification Task
- Accuracy and Overall Error Rate
- Sensitivity and Specificity
- False-Positive Rate and False-Negative Rate
- Proportions of True Positives, True Negatives, False Positives, and False Negatives
- Misclassification Cost Adjustment to Reflect Real-World Concerns
- Decision Cost/Benefit Analysis
- Lift Charts and Gains Charts
- Interweaving Model Evaluation with Model Building
- Confluence of Results: Applying a Suite of Models
- The R Zone
- R References
- Exercises

- Hands-On Analysis

Chapter 17: Cost-Benefit Analysis Using Data-Driven Costs

- Decision Invariance Under Row Adjustment
- Positive Classification Criterion
- Demonstration Of The Positive Classification Criterion
- Constructing The Cost Matrix
- Decision Invariance Under Scaling
- Direct Costs and Opportunity Costs
- Case Study: Cost-Benefit Analysis Using Data-Driven Misclassification Costs
- Rebalancing as a Surrogate for Misclassification Costs
- The R Zone
- R References
- Exercises

Chapter 18: Cost-Benefit Analysis for Trinary and -Nary Classification Models

- Classification Evaluation Measures for a Generic Trinary Target
- Application of Evaluation Measures for Trinary Classification to the Loan Approval Problem

- Data-Driven Cost-Benefit Analysis for Trinary Loan Classification Problem
- Comparing Cart Models With and Without Data-Driven Misclassification Costs
- Classification Evaluation Measures for a Generic k-Nary Target
- Example of Evaluation Measures and Data-Driven Misclassification Costs for k-Nary Classification
- The R Zone
- R References
- Exercises

Chapter 19: Graphical Evaluation of Classification Models

- Review of Lift Charts and Gains Charts
- Lift Charts and Gains Charts Using Misclassification Costs
- Response Charts
- Profits Charts
- Return on Investment (ROI) Charts
- The R Zone
- R References
- Exercises
- Hands-On Exercises

Chapter 20: Hierarchical and k-Means Clustering

- The Clustering Task
- Hierarchical Clustering Methods
- Single-Linkage Clustering
- Complete-Linkage Clustering
- k-Means Clustering
- Example of k-Means Clustering at Work
- Behavior of MSB, MSE, and Pseudo-F as the k-Means Algorithm Proceeds
- Application of k-Means Clustering Using SAS Enterprise Miner
- Using Cluster Membership to Predict Churn
- The R Zone
- R References
- Exercises
- Hands-On Analysis

Chapter 21: Kohonen Networks

- Self-Organizing Maps

- Kohonen Networks
- Example of a Kohonen Network Study
- Cluster Validity
- Application of Clustering Using Kohonen Networks
- Interpreting The Clusters
- Using Cluster Membership as Input to Downstream Data Mining Models
- The R Zone
- R References
- Exercises

Chapter 22: BIRCH Clustering

- Rationale for BIRCH Clustering
- Cluster Features
- Cluster Feature TREE
- Phase 1: Building The CF Tree
- Phase 2: Clustering The Sub-Clusters
- Example of Birch Clustering, Phase 1: Building The CF Tree
- Example of BIRCH Clustering, Phase 2: Clustering The Sub-Clusters

- Evaluating The Candidate Cluster Solutions
- Case Study: Applying BIRCH Clustering to The Bank Loans Data Set
- The R Zone
- R References
- Exercises

Chapter 23: Measuring Cluster Goodness

- Rationale for Measuring Cluster Goodness
- The Silhouette Method
- Silhouette Example
- Silhouette Analysis of the IRIS Data Set
- The Pseudo-F Statistic
- Example of the Pseudo-F Statistic
- Pseudo-F Statistic Applied to the IRIS Data Set
- Cluster Validation
- Cluster Validation Applied to the Loans Data Set
- The R Zone
- R References

- Exercises

Chapter 24: Association Rules

- Affinity Analysis and Market Basket Analysis
- Support, Confidence, Frequent Itemsets, and the A Priori Property
- How Does The A Priori Algorithm Work (Part 1)? Generating Frequent Itemsets
- How Does The A Priori Algorithm Work (Part 2)? Generating Association Rules
- Extension From Flag Data to General Categorical Data
- Information-Theoretic Approach: Generalized Rule Induction Method
- Association Rules are Easy to do Badly
- How Can We Measure the Usefulness of Association Rules?
- Do Association Rules Represent Supervised or Unsupervised Learning?
- Local Patterns Versus Global Models
- The R Zone
- R References
- Exercises

Chapter 25: Segmentation Models

- The Segmentation Modeling Process

- Segmentation Modeling Using EDA to Identify the Segments
- Segmentation Modeling using Clustering to Identify the Segments
- The R Zone
- R References
- Exercises

Chapter 26: Ensemble Methods: Bagging and Boosting

- Rationale for Using an Ensemble of Classification Models
- Bias, Variance, and Noise
- When to Apply, and not to apply, Bagging
- Bagging
- Boosting
- Application of Bagging and Boosting Using IBM/SPSS Modeler
- References
- The R Zone
- R Reference
- Exercises

Chapter 27: Model Voting and Propensity Averaging

- Simple Model Voting
- Alternative Voting Methods
- Model Voting Process
- An Application of Model Voting
- What is Propensity Averaging?
- Propensity Averaging Process
- An Application of Propensity Averaging
- The R Zone
- R References
- Exercises
- Hands-On Analysis

Chapter 28: Genetic Algorithms

- Introduction To Genetic Algorithms
- Basic Framework of a Genetic Algorithm
- Simple Example of a Genetic Algorithm at Work
- Modifications and Enhancements: Selection
- Modifications and Enhancements: Crossover

- Genetic Algorithms for Real-Valued Variables
- Using Genetic Algorithms to Train a Neural Network
- WEKA: Hands-On Analysis Using Genetic Algorithms
- The R Zone
- R References
- Exercises

Chapter 29: Imputation of Missing Data

- Need for Imputation of Missing Data
- Imputation of Missing Data: Continuous Variables
- Standard Error of the Imputation
- Imputation of Missing Data: Categorical Variables
- Handling Patterns in Missingness
- Reference
- The R Zone
- R References

Chapter 30: Case Study, Part 1: Business Understanding, Data Preparation, and EDA

- Cross-Industry Standard Practice for Data Mining
- Business Understanding Phase
- Data Understanding Phase, Part 1: Getting a Feel for the Data Set
- Data Preparation Phase
- Data Understanding Phase, Part 2: Exploratory Data Analysis

Chapter 31: Case Study, Part 2: Clustering and Principal Components Analysis

- Partitioning the Data
- Developing the Principal Components
- Validating the Principal Components
- Profiling the Principal Components
- Choosing the Optimal Number of Clusters Using Birch Clustering
- Choosing the Optimal Number of Clusters Using k-Means Clustering
- Application of k-Means Clustering
- Validating the Clusters
- Profiling the Clusters

Chapter 32: Case Study, Part 3: Modeling And Evaluation For Performance And Interpretability

- Do You Prefer The Best Model Performance, Or A Combination Of Performance And Interpretability?
- Modeling And Evaluation Overview
- Cost-Benefit Analysis Using Data-Driven Costs
- Variables to be Input To The Models
- Establishing The Baseline Model Performance
- Models That Use Misclassification Costs
- Models That Need Rebalancing as a Surrogate for Misclassification Costs
- Combining Models Using Voting and Propensity Averaging
- Interpreting The Most Profitable Model

Chapter 33: Case Study, Part 4: Modeling and Evaluation for High Performance Only

- Variables to be Input to the Models
- Models that use Misclassification Costs
- Models that Need Rebalancing as a Surrogate for Misclassification Costs
- Combining Models using Voting and Propensity Averaging
- Lessons Learned
- Conclusions

Chapter 34: Appendix A

- Data Summarization and Visualization
- Part 1: Summarization 1: Building Blocks Of Data Analysis
- Part 2: Visualization: Graphs and Tables For Summarizing And Organizing Data
- Part 3: Summarization 2: Measures Of Center, Variability, and Position
- Part 4: Summarization And Visualization Of Bivariate Relationships

12. Practice Test

Here's what you get

Features

Each question comes with detailed remediation explaining not only why an answer option is correct but also why it is incorrect.

Unlimited Practice

Each test can be taken unlimited number of times until the learner feels they are prepared. Learner can review the test and read detailed remediation. Detailed test history is also available.

Each test set comes with learn, test and review modes. In learn mode, learners will attempt a question and will get immediate feedback and complete remediation as they move on to the next question. In test mode, learners can take a timed test simulating the actual exam conditions. In review mode, learners can read through one item at a time without attempting it.

13. Live Labs

The benefits of live-labs are:

- Exam based practical tasks
- Real equipment, absolutely no simulations
- Access to the latest industry technologies
- Available anytime, anywhere on any device
- Break and Reset functionality
- No hardware costs

Lab Tasks

An Introduction to Data Mining and Predictive Analytics

- Analyzing a Dataset

Data Preprocessing

- Handling Missing Data
- Creating a Histogram
- Creating a Scatterplot
- Creating a Normal Q-Q Plot
- Creating Indicator Variables

Exploratory Data Analysis

- Analyzing the churn Dataset
- Exploring Categorical Variables
- Exploring Numeric Variables
- Exploring Multivariate Relationships
- Investigating Correlation Values and p-values in Matrix Form

Dimension-Reduction Methods

- Creating a Scree Plot
- Profiling the Principal Components
- Calculating Communalities
- Validating the Principal Components
- Applying Factor Analysis to a Dataset

Univariate Statistical Analysis

- Estimating the Confidence Interval for the Mean
- Estimating the Confidence Interval of the Population Proportion

Multivariate Statistics

- Performing a t-test for Finding the Difference in Means
- Performing a z-test for Finding the Difference in Proportions
- Performing a Chi-Square Test for Homogeneity of Proportions
- Performing a Chi-Square Test for Goodness of Fit of Multinomial Data
- Analyzing a Variance

Preparing to Model the Data

- Balancing the Training and Testing Datasets

Simple Linear Regression

- Plotting Data with a Regression Line
- Measuring the Goodness of Fit of the Regression
- Performing Regression with Other Hikers
- Verifying the Regression Assumptions
- Determining Prediction and Confidence Intervals
- Assessing Normality in Scrabble
- Applying Box-Cox Transformations

Multiple Regression and Model Building

- Approximating the Relationship between the Variables in a Scatterplot
- Identifying Confidence Intervals
- Creating a Dot Plot
- Determining the Sequential Sums of Squares
- Analyzing Multicollinearity
- Applying the Best Subsets Procedure in a Regression Model
- Applying the Stepwise Selection Procedure in a Regression Model
- Applying the Backward Elimination Procedure
- Applying Forward Selection Procedure
- Using the Principal Components as Predictors in Multiple Regression

k-Nearest Neighbor Algorithm

- Running KNN
- Calculating the Euclidean Distance

Decision Trees

- Plotting a Classification Tree

Neural Networks

- Running a Neural Network

Logistic Regression

- Creating a Plot for Logistic Regression
- Interpreting Logistic Regression and Odds Ratio for a Dichotomous Predictor

NaïVe Bayes and Bayesian Networks

- Calculating Posterior Odds Ratio
- Calculating the Log Posterior Odds Ratio
- Calculating the Numeric Predictors for Naive Bayes Classification

Model Evaluation Techniques

- Estimating Costs for Benefit Analysis

Cost-Benefit Analysis Using Data-Driven Costs

- Analyzing Cost-benefit Using Data-driven Misclassification Costs

Cost-Benefit Analysis for Trinary and -Nary Classification Models

- Analyzing the Cost-Benefit for the Trinary Loan Classification Problem

Hierarchical and k-Means Clustering

- Using Single-linkage Clustering
- Using Complete-linkage Clustering
- Finding Clusters in Data

Kohonen Networks

- Using a 3x2 Kohonen Network
- Interpreting Clusters

Measuring Cluster Goodness

- Plotting Silhouette Values of a Dataset
- Applying Cluster Validation to a Dataset

Association Rules

- Viewing the Output Sorted by Support

Segmentation Models

- Predicting Income Using Caps and No Caps Groups

Genetic Algorithms

- Using Genetic Algorithms to Train a Neural Network

Here's what you get

63	63	02:02
LIVE LABS	VIDEO TUTORIALS	HOURS

GET IN TOUCH:

3187 Independence Drive
Livermore, CA 94551,
United States

+1-415-763-6300

support@ucertify.com

www.ucertify.com