Preview: ITEC5040 : Predictive Models and Classification Methods

Syllabus

Course Overview

This course focuses on predicting future outcomes based on historical data, as well as identifying patterns in current data that can be used to classify or group future observations. Learners complete their own analytics project through hands-on statistical techniques coupled with a broad understanding of contemporary predictive modeling and analytics classification methods.

Prerequisite(s): ANLT5030.

Technology Resources

Capella offers tutorials, labs, or a virtual desktop as part of this course. These resources offer software or guided practice in performing tasks related to achieving course competencies and completing assessments. If you require the use of assistive technology or alternative communication methods to participate in these activities, please contact DisabilityServices@Capella.edu to request accommodations.

Course Competencies **(Read Only)**

To successfully complete this course, you will be expected to:

1. Identify data that addresses a business problem or scenario.
2. Prepare data for use in predictive models and classification algorithms to address a defined need.
3. Select appropriate predictive modeling or classification methods for a problem or scenario.
4. Apply analytic tools and methods to develop predictive models or classification algorithms to address a defined need.
5.

Integrate relevant ethical, legal, global or cultural factors into predictive and classification models and algorithms.

6 Communicate the results of predictive models and classification algorithms to key stakeholders.

Course Prerequisites

Prerequisite(s): ANLT5030.

Syllabus >> Course Materials

Required

The materials listed below are required to complete the learning activities in this course.

Integrated Materials

Many of your required books are available via the VitalSource Bookshelf link in the courseroom, located in your Course Tools. Registered learners in a Resource Kit program can access these materials using the courseroom link on the Friday before the course start date. Some materials are available only in hard-copy format or by using an access code. For these materials, you will receive an email with further instructions for access. Visit the Course Materials page on Campus for more information.

Book

Miller, T. W. (2015). *Web and network data science: Modeling techniques in predictive analytics.* Upper Saddle River: Pearson ISBN: 9780133886443

Library

The following required readings are provided in the Capella University Library or linked directly in this course. To find specific readings by journal or book title, use Journal and Book Locator. Refer to the Journal and Book Locator library guide to learn how to use this tool.

- Acharya, S. & Chellappan, S. (2017). *Pro Tableau: A step-by-step guide.* Apress.
- Collica, R. S. (2017). *Customer segmentation and clustering using SAS Enterprise Miner* (3rd ed.). SAS Institute.
- de Ville, B., Neville, P. (2013). *Decision trees for analytics: Using SAS Enterprise Miner* (1st ed.). SAS Institute.
- Gupta, D. (2018). *Applied analytics through case studies using SAS and R: Implementing predictive models and machine learning techniques.* Apress.
- Heimann, R. (2015). *Big data legal perspective: Using data ethically* [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). *Data visualization essentials: Adding effective visuals and text* [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). *Data visualization essentials: Adding functionality and color* [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). Data visualization essentials: Choosing the type of visualization for the data [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). *Data visualization essentials: Data visualization versus infographics* [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). *Data visualization essentials: Defining the narrative* [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). *Data visualization essentials: Designing data visualizations* [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). *Data visualization essentials: Knowing your audience* [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). *Data visualization essentials: Traits and visual perception* [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). Data visualization essentials: Understanding the data being visualized [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). *Data visualization essentials: What is data visualization* [Video] Skillsoft Ireland Limited.
- Miner, G., Elder, J., Fast, A., Hill, T., Nisbet, R., & Delen, D. (2012). *Practical text mining and statistical analysis for non-structured text data applications.* Waltham, MA: Elsevier/Academic Press.

- O'Keefe, K. & O Brien, D. (2018). *Ethical data and information management: Concepts, tools and methods.* Kogan Page.
- Scott, S. (2017). *Data science essentials: Effective communication and visualization* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Data science essentials: Working with decision trees* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Assigning across document predictor variables* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Assigning within document predictor variables* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Autoregressive models* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Autoregressive Moving Average (ARMA) models* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). Predictive analytics: Decision tree characteristics [Video]. Skillsoft Ireland.
- Scott, S. (2017). *Predictive analytics: EDA and cluster segmentation modeling* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). Predictive analytics: Identify the regression technique [Video]. Skillsoft Ireland.
- Scott, S. (2017). Predictive analytics: Introduction to regression analysis [Video]. Skillsoft Ireland.
- Scott, S. (2017). *Predictive analytics: K-Means clustering termination procedures* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). Predictive analytics: Linear regression overview [Video]. Skillsoft Ireland.
- Scott, S. (2017). Predictive analytics: Logistic regression overview [Video]. Skillsoft Ireland.
- Scott, S. (2017). *Predictive analytics: Moving average models* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive Analytics: Overviewof K-Means Clustering* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Overviewof text mining* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Overviewof the k-NN algorithm* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Term frequency and inverse document frequency* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Text mining applications* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Text normalization* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Time series decomposition* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Time series overview*[Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Types of clustering techniques* [Video]. Skillsoft Ireland Limited.
- Scott. S. (2017). *Predictive analytics: Stationary and nonstationary data series* [Video]. Skillsoft Ireland Limited.
- Strickland, J. (2016). *Predictive analytics: Clustering* [Video]. Skillsoft Ireland Limited.
- Strickland, J. (2016). *Predictive analytics: Decision trees* [Video]. Skillsoft Ireland Limited.
- Strickland, J. (2016). *Predictive analytics: Time series analysis* [Video]. Skillsoft Ireland Limited.

- Watson, H. J. (2017). Data Visualization, Data Interpreters, and Storytelling. *Business Intelligence Journal, 22*(1), 5–10.

External Resource

Please note that URLs change frequently. While the URLs were current when this course was designed, some may no longer be valid. If you cannot access a specific link, contact your instructor for an alternative URL. Permissions for the following links have been either granted or deemed appropriate for educational use at the time of course publication.

- Gupta, D. (2018). Source Code for Applied Analytics Through Case Studies Using SAS and R. Retrieved from https://github.com/Apress/applied-analytics-case-studies-sas-r/blob/master/flight_delay.csv
- Kaggle. (2017). NFL Combine Data. Retrieved from https://www.kaggle.com/savvastj/nfl-combine-data
- Malekzadeh, M., Clegg, R. G., Cavallaro, A., and Haddadi, H. (2018). Protecting sensory data against sensitive inferences. In W-P2DS'18: 1st workshop on privacy by design in distributed systems, April 23–26, 2018, Porto, Portugal. New York, New York: ACM.
- Save Workbooks to Tableau Public. (n.d.). Retrieved from http://onlinehelp.tableau.com/current/pro/desktop/en-us/publish_workbooks_tableaupublic.html
- Schubert, S., & Lee, T. (2011). Time series data mining with SAS Enterprise Miner . Retrieved from https://support.sas.com/resources/papers/proceedings11/160-2011.pdf
- Tableau for Students. (n.d.). Tableau Desktop Retrieved from https://www.tableau.com/academic/students
- Tableau training and tutorials. (n.d.). Retrieved from https://www.tableau.com/learn/training
- Tableau. (n.d.). Tableau Starter Kits: Author a Viz in Tableau Desktop. Retrieved from https://www.tableau.com/learn/starter-kits/author/desktop
- Truxillo, C. (n.d.). SAS Enterprise Miner: Profiling Segments. | Transcript Retrieved from https://video.sas.com/detail/video/2436413682001/sas-enterprise-miner:-profiling-segments?autoStart=true&q=cluster
- UCI Machine Learning. (2016). SMS Spam Collection Dataset. Retrieved from https://www.kaggle.com/uciml/sms-spam-collection-dataset
- UCI Machine Learning. (2017). Internet Advertisement Data Set. Retrieved from https://www.kaggle.com/uciml/internet-advertisements-data-set
- Wujek, B. (n.d.). The ABCs of selecting clusters. Retrieved from https://video.sas.com/detail/video/4572850292001/the-abcs-of-selecting-clusters?autoStart=true&q=cluster

Suggested

The following materials are recommended to provide you with a better understanding of the topics in this course. These materials are not required to complete the course, but they are aligned to course activities and assessments and are highly recommended for your use.

External Resource

Please note that URLs change frequently. While the URLs were current when this course was designed, some may no longer be valid. If you cannot access a specific link, contact your instructor for an alternative URL. Permissions for the following links have been either granted or deemed appropriate for educational use at the time of course publication.

- Teradata Corporation. (n.d.). Teradata university network. Retrieved from http://www.teradatauniversitynetwork.com/Register/

Optional

The following optional materials are offered to provide you with a better understanding of the topics in this course. These materials are not required to complete the course.

Library

The following optional readings may be available in the Capella University Library. To find specific readings by journal or book title, use Journal and Book Locator. Refer to the Journal and Book Locator library guide to learn how to use this tool. If the full text is not available, you may be able to request a copy through the Interlibrary Loan service.

- Chakraborty, G., Pagolu, M., & Garla, S. (2013). *Text mining and analysis: Practical methods, examples, and case studies using SAS.* SAS Institute.

External Resource

Please note that URLs change frequently. While the URLs were current when this course was designed, some may no longer be valid. If you cannot access a specific link, contact your instructor for an alternative URL. Permissions for the following links have been either granted or deemed appropriate for educational use at the time of course publication.

- Fernandez, G. (2011). *Statistical data mining using SAS applications* (2nd ed.). Boca Raton: CRC Press.
- Sarma, K. S. (2017). *Predictive modeling with SAS Enterprise Miner: Practical solutions for business applications* (3rd ed.). Cary, NC: SAS Institute. ISBN: 9781629602646.

Projects

Project >> Course Project

Project Overview

This course project has two deliverables:

- An Analytics Project Final Report.
- An Analytics Project Executive Summary Presentation.

In this course project you will identify a context to apply data analytic methods as appropriate for the topic you identify for your project. You will describe the scenario where your project takes place, the problem or opportunity you are trying to solve, or the question you are trying to answer, and the reasons you selected this topic. You will identify a data source to use for your course project, including an explanation of how this data can be used to address the problem or answer the question. If your topic or set of data is from your workplace, remember to follow confidentiality agreements and secure prior approvals as required by your organization. Keep in mind that you will submit information in your project deliverables about your topic, the business opportunity, and your data set. You will not include the actual data, personally identifiable information (PII), or other confidential information within any of the assignments. General descriptions of variables and procedures should be sufficient, but remember this project is not intended to be theoretical, so be sure that you are actually able to use and analyze the data you are describing.

For your project you will explain how the business question or opportunity and the data align or do not align, and if they do not align, discuss what you will do to rectify any misalignment. You will address any potential ethical, legal, global, or cultural factors that may present risk in using the data for the context of your project. You will audit and profile the data and identify data issues or potential data issues, and address the data issues, including any missing values or erroneous data.

Finally, in your final report you will you will summarize the work you did to address the business question or opportunity and the data set used for your analysis. In your report you will provide an explanation that justifies your choice of data analysis for the business question or opportunity you addresses in your project.

Your final report will:

1. Summarize your business question or opportunity, the source of data used, the analysis method used, and justification for your choice of that analytic method for the given data and business question or opportunity.
2. Address ethical, legal, global, or cultural factors for the analytic method applied in the context of your project.
3. Include the SAS code created during your analysis project as an appendix or attachment.

4. Justify your selection of the analytic method you used on your given data set to address your selected business question or opportunity.
5. Summarize and explain your analysis in business or industry language.

Your executive summary PowerPoint presentation will outline your analysis and is appropriate for an executive audience and:

1. Summarize the business question or opportunity being addressed.
2. Summarize the data you used to address the business question or opportunity.
3. Summarize the data cleansing that was completed and any changes that were made to the data before analyzing it.
4. Summarize the results of the analysis, including the methods used and how the results address the business question or opportunity.
5. Present the results in language, level of detail, and content appropriate for an executive audience.
6. Include a title page and references, not included in the slide limits indicated above.

Unit 1 >> Data Preparation and Cleansing

Introduction

This week you will:

- Study introductory information about data science methods.
- Prepare the software you will use in this course.
- Select a data analytics project context, topic, and data source.

In this course, you will learn about various modeling and forecasting techniques, as well as classification techniques. You will have one individual project on a topic and data set that you select that you will work on throughout the course. In this unit, you will:

- Select a problem or opportunity that you will try to address in your project
- Identify the data you will use to address it, under the assumption that you will be using one or more of the techniques you learn within the course to analyze your data and address the problem or opportunity.

Your data can be text or categorical data, or it could be numeric data, and your problem can be one that requires classification or prediction. The discussion within this unit provides you with an opportunity to gather feedback from your classmates and instructor on your problem or opportunity and on your data set, to make sure it is appropriate for this course. After you gather feedback during the discussion, you will submit the summary of your problem or opportunity and the data you plan to use to address it within this unit's assignment.

Learning Activities

u01s1 - Study

For an introduction to data science methods within the field of Information Technology, or, more specifically the web and networking, read:

- Miller, T. W. (2015). *Web and network data science: Modeling techniques in predictive analytics*. Upper Saddle River: Pearson.
    - Chapter A: Data Science Methods.

u01s1 - Learning Components

- Describe data science methods.

u01s2 - Software Preparation and Technology Access

In this course, you will be using software and technology that is needed to complete designated activities and assignments. There is no additional cost for this software and technology. Some software packages will be made available to you at no additional cost through Capella's subscription with Microsoft, while other software packages are available for free download through open-source licensing.

Capella University requires learners to meet certain minimum computer requirements. Please note that some software required for a course may exceed these minimum requirements. Check the requirements for the software you may need to download and install to make sure it will work on your device. Most

software will require a Windows PC. If you use a Mac, refer to Installing a Virtual Environment and Windows on a Mac.

The software and technologies below are strongly recommended to support you in completing the course objectives. If you have access to other tools that you believe may still meet the requirements of this course, please discuss your selected alternatives with your instructor.

If you use assistive technology or any alternative communication methods to access course content, please contact DisabilityServices@Capella.edu with any access-related questions or to request accommodations.

For this course, follow the instructions provided through the links below to download and install software or register for an account, as required.

SAS Statistical Software

Follow the instructions to set up SAS OnDemand for Academics (SODA).

- Check with your instructor for any additional instructions.
- Download or access the SAS data files for use in the assignments.
    - Locate the CourseID Data Files [ZIP], or links to data sets, in the assignment resources.
    - Open the file and take some time to explore the data sets to see how they are constructed.

Data Visualization Software

To prepare for the visualization portion of course project:

- Tableau for Students provides a one-year free license for Tableau. To get started preparing for the visualization portion of your project, download Tableau Desktop from https://www.tableau.com/academic/students. To obtain the license, follow the directions at the website https://www.tableau.com/academic/students.
- In addition to downloading Tableau and requesting a Tableau license directly from the company, students can also access Tableau, as well as SAS Visual Analytics for free through a partnership with Teradata. Students can register with Teradata University Network and gain access to Teradata, Tableau and SAS Visual Analytics here: https://www.teradatauniversitynetwork.com/Register

If you encounter any difficulties in the download and installation process, post a detailed question in the Ask Your Instructor section of the course. Your instructor should be able to help you or point you in the right direction for the answers you need.

Note: As a Capella learner, you have access to IT online resources through Capella's Skillsoft subscription, where you can find helpful materials.

u01s2 - Learning Components

- Prepare software as required for data analytics and data visualization activities.

u01d1 - Individual Project Opportunity and Data

This discussion provides you an opportunity to gather feedback on the problem or opportunity you are attempting to address within your project, and the data you would like to use to address it, before you submit your topic in to the Unit 1 Assignment 1 submission area for instructor feedback and topic approval. Please post a summary of the problem or opportunity you wish to address within your individual course project. Also summarize the data (including the variables and what each observation represents) that you plan to use to address the problem or opportunity.

Response Guidelines

Please respond to at least two of your classmates' initial responses. Focus your response on the clarity of their problem or opportunity statement, whether the data identified is appropriate for their problem or opportunity, and whether the data they have identified will allow them to apply classification or prediction techniques in order to address the problem or opportunity.

Course Resources
Graduate Discussion Participation Scoring Guide

u01d1 - Learning Components

- Select a data analytics project context, topic and data source.

u01a1 - Individual Project Topic Proposal

Assignment Overview

In this assignment you will select a topic for your course project, and describe the context for your topic selection. You will describe the scenario where your project takes place, the problem or opportunity you are trying to solve, or the question you are trying to answer, and the reasons you selected this topic. Your summary should also include the data source you plan to use for your course project, including an explanation of how this data can be used to address the problem or answer the question. If your topic or set of data is from your workplace, remember to follow confidentiality agreements and secure prior

approvals as required by your organization. Keep in mind that you will submit information in your project deliverables about your topic, the business opportunity, and your data set. You will not include the actual data, personally identifiable information (PII) data, or other confidential information within any of the assignments. General descriptions of variables and procedures should be sufficient, but remember that these assignments are not intended to be theoretical, so be sure that you are actually able to use and analyze the data you are describing.

Assignment Instructions

1. Write a 2–3 page double spaced paper summarizing your course project topic and the context of your topic selection.
2. Describe the scenario for your project, the problem or opportunity your project will try to solve, or a question you are trying to answer.
3. Provide the reasons you selected this topic.
4. Identify the data source for your project.
5. Identify any required permission or confidentiality agreements, if needed from your organization.
6. Provide an analysis of the alignment between the data set and the problem or opportunity, or question that is the focus of your project.

Submit your assignment documents by 11:59 on Sunday of this week.

Assignment Criteria

Your assignment will be scored on the following criteria:

- Identifies a topic and data set for a data analytics project that can be completed within a 10 week time frame.
- Articulates the business question, problem, or opportunity for a data analytics project.
- Explains the alignment of the business question, problem, or opportunity and the selected data set.
- Uses communication style and vocabulary appropriate for the target audience.

Assignment Requirements

- **Communication:** Communicate in a manner that is scholarly, professional, respectful, and consistent with expectations for professional practice in technology. Assignment submission must be original work that reflects critical thinking with clear organization of concepts and ideas.
- **Paper length:** Minimum of 2–3 pages, not including the title page or reference page.
- **References:** Cite your data and sources as appropriate.
- **Document format guidelines:** Use the Academic and Professional Document Guidelines linked in the resources as appropriate for your deliverable.
- **Font and font size:** Times New Roman, 11 pt.

Course Resources

Academic and Professional Document Guidelines [PDF]

Unit 2 >> Advanced Regression Techniques

Introduction

This week you will:

- Compare and contrast linear and logistic regression, in general, and multiple linear and multiple logistic regression methods.
- Describe regression analysis techniques.
- Evaluate the appropriateness of multiple regression for a given context and data set.
- Complete a linear regression lab.

In this unit, you will conduct linear and logistic regression analysis, learn about assumptions and considerations for both of these types of regression techniques, how to tell when you use each, and learn about multiple linear regression and multiple logistic regression techniques.

Learning Activities

u02s1 - Study

For a review of regression analysis, specifically linear and logistic regression analysis, watch the following short videos:

- Scott, S. (2017). *Predictive analytics: Introduction to regression analysis* [Video]. Skillsoft Ireland.
- Scott, S. (2017). *Predictive analytics: Identify the regression technique* [Video]. Skillsoft Ireland.
- Scott, S. (2017). *Predictive analytics: Linear regression overview*[Video]. Skillsoft Ireland.
- Scott, S. (2017). *Predictive analytics: Logistic regression overview*[Video]. Skillsoft Ireland.

u02s1 - Learning Components

- Describe data science methods.
- Compare and contrast linear and logistic regression, in general, and multiple linear and multiple logistic regression methods.

- Evaluate the appropriateness of multiple regression for a given context and data set.
- Describe regression analysis techniques.

u02d1 - Linear and Logistic Regression Applications

Evaluate the appropriateness of linear and logistic regression, in general, and multiple linear and multiple logistic regression, specifically, for your individual project. Detail the problem or issue you are trying to address and the data that you are using to do so. If you feel that either of these methods might be appropriate for your individual project, identify the response and explanatory variables you plan to use in your model(s) to address the problem or issue identified.

In addition to evaluating the appropriateness of multiple regression for your individual course project, think of a business question or problem from your current or future career that could be addressed using multiple regression. Explain the business question or problem, the type of data that might be used, and how you might use multiple regression to address this business problem using the type of data described.

Response Guidelines

Please respond to at least two of your classmates' initial responses. Do you agree or disagree with their assertion that linear or logistic regression is or is not appropriate for their data and the issue or problem they are attempting to address? Why or why not?

Course Resources
Graduate Discussion Participation Scoring Guide

u02d1 - Learning Components

- Describe data science methods.
- Select a data analytics project context, topic and data source.
- Compare and contrast linear and logistic regression, in general, and multiple linear and multiple logistic regression methods.
- Evaluate the appropriateness of multiple regression for a given context and data set.

u02a1 - Applied Advanced Regression Techniques Lab

Assignment Overview

Use the following resources as walk through tutorials to create a multiple linear regression model to predict delays in airline flights. You can use either SAS Studio or SAS Enterprise Guide to complete the tutorial for this lab.

- Gupta, D. (2018). *Applied analytics through case studies using SAS and R: Implementing predictive models and machine learning techniques.* Apress.
  - Chapter 6: Airline Case Study – Multiple Linear Regression Model Using SAS.

The data set can be downloaded here:

- Gupta, D. (2018). Source Code for Applied Analytics Through Case Studies Using SAS and R. Retrieved from https://github.com/Apress/applied-analytics-case-studies-sas-r/blob/master/flight_delay.csv

To learn about the background of the case study, read the following:

- Gupta, D. (2018). *Applied analytics through case studies using SAS and R: Implementing predictive models and machine learning techniques.* Apress.
  - Chapter 6: Airline Case Study – Predicting Flight Delays with Multiple Linear Regression Model.

Pay special attention to the splitting of the data set into train and test subsets, checking of assumptions about the individual variables and evaluations of relationships between the variables.

Assignment Instructions

Demonstrate your completion of the tutorial by including your code and output, along with a brief narrative of each step you took. At the end of your assignment, please evaluate your completion of this tutorial by discussing portions of the tutorial that were especially informative, surprising or challenging. A lab assignment template is provided in the resources for this assignment.

Assignment Criteria

Your assignment must meet this criteria, please refer to the scoring guide for this assignment.

- Identifies the strengths and challenges of creating a multiple linear regression model.
- Evaluates the steps that are needed to create a multiple linear regression model.
- Determines suitability of the application of a multiple linear regression model for a given context.
- Completes the tutorial to create multiple linear regression model.
- Describes the benefits of using a multiple linear regression model for a given context.
- Uses communication style and vocabulary appropriate for the target audience.

Assignment Requirements

- **Lab output:** Your code and output you generated for the tutorial.

- **Communication:** Communicate in a manner that is scholarly, professional, respectful, and consistent with expectations for professional practice in technology. Assignment submission must be original work that reflects critical thinking with clear organization of concepts and ideas.
- **Paper length:** Minimum of 2–3 pages, not including the title page or reference page.
- **References:** Cite your data and sources as appropriate.
- **Document format guidelines:** Use the Academic and Professional Document Guidelines linked in the resources as appropriate for your deliverable.
- **Font and font size:** Times New Roman, 11 pt.

Course Resources
Academic and Professional Document Guidelines [PDF]
Lab Assignment Template [DOCX]

u02s2 - Topic Resources

Optional

For a more in-depth discussion of variable selection and regression models, read:

- Sarma, K. S. (2017). *Predictive modeling with SAS Enterprise Miner: Practical solutions for business applications* (3rd ed.). Cary, NC: SAS Institute.
  - Chapter 3: Variable Selection and Transformation of Variables.
  - Chapter 6: Regression Models.

For a more in-depth discussion of linear and logistic regression, read:

- Fernandez, G. (2011). *Statistical data mining using SAS applications* (2nd ed.). Boca Raton: CRC Press.
  - Chapter 5: Supervised Learning Methods: Prediction.

Unit 3 >> Classification: Decision Trees

Introduction

This week you will:

- Compare and contrast decision tree methods.
- Describe decision tree techniques.
- Evaluate the appropriateness of decision trees for a given context and data set.
- Complete a decision tree lab.

In this unit, you will learn about decision trees used for classification and regression. You will also apply your new knowledge by using SAS Enterprise Miner to create a decision tree on internet advertisement data, and discuss the appropriateness of decision trees for your individual project's problem or issue and data.

Learning Activities

u03s1 - Study

For an overview of decision trees, watch the following short videos:

- Strickland, J. (2016). *Predictive analytics: Decision trees* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Decision tree characteristics* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Data science essentials: Working with decision trees* [Video]. Skillsoft Ireland Limited.

u03s1 - Learning Components

- Compare and contrast decision tree methods.
- Describe decision tree techniques.

u03d1 - Decision Tree Applications

Evaluate the appropriateness of decision tree methods for your individual project. Detail the problem or issue you are trying to address and the data that you are using to do so. If you feel that decision tree methods, either for classification or for regression, might be appropriate for your individual project, indicate why you believe this to be true, whether classification decision trees or regression decision trees are appropriate, and, if applicable, what response variable you plan to create your decision tree on.

In addition to evaluating the appropriateness of decision tree methods for your individual course project, think of a business question or problem from your current or future career that could be addressed using decision tree methods. Explain the business question or problem, the type of data that might be used, and how you might use decision tree methods to address this business problem using the type of data described.

Response Guidelines

Please respond to at least two of your classmates' initial responses. Do you agree or disagree with their assertion that decision tree methods are or are not appropriate for their data and the issue or problem they are attempting to address? Why or why not?

Course Resources
Graduate Discussion Participation Scoring Guide

u03d1 - Learning Components

- Select a data analytics project context, topic and data source.
- Compare and contrast decision tree methods.
- Describe decision tree techniques.
- Evaluate the appropriateness of decision trees for a given context and data set.

u03a1 - Decision Trees Lab

Assignment Overview

Use the following resources as tutorials for creating a classification decision tree on the internet advertisement data set using SAS Enterprise Miner:

- de Ville, B., Neville, P. (2013). *Decision trees for analytics: Using SAS Enterprise Miner* (1st ed.). SAS Institute.

The internet advertisement data set can be downloaded here:

- UCI Machine Learning. (2017). Internet Advertisement Data Set. Retrieved from https://www.kaggle.com/uciml/internet-advertisements-data-set

Assignment Instructions

Demonstrate your completion of the tutorial by including your code and output, along with a brief narrative of each step you took. At the end of your assignment, please evaluate your completion of this tutorial by discussing portions of the tutorial that were especially informative, surprising or challenging. A lab assignment template is provided in the resources for this assignment.

Assignment Criteria

Your assignment must meet this criteria, please refer to the scoring guide for this assignment.

- Identifies the strengths and challenges of creating a classification decision tree.
- Evaluates the steps that are needed to create a classification decision tree.
- Determines suitability of the application of a classification decision tree for a given context.
- Completes the tutorial to create a classification decision tree.
- Describes the benefits of using a a classification decision tree for a given context.
- Uses communication style and vocabulary appropriate for the target audience.

Assignment Requirements

- **Lab output**: Your code and output you generated for the tutorial.
- **Communication:** Communicate in a manner that is scholarly, professional, respectful, and consistent with expectations for professional practice in technology. Assignment submission must be original work that reflects critical thinking with clear organization of concepts and ideas.
- **Paper length:** Minimum of 2–3 pages, not including the title page or reference page.
- **References:** Cite your data and sources as appropriate.
- **Document format guidelines:** Use the Academic and Professional Document Guidelines linked in the resources as appropriate for your deliverable.
- **Font and font size:** Times New Roman, 11 pt.

Course Resources
Academic and Professional Document Guidelines [PDF]
Lab Assignment Template [DOCX]

u03s2 - Topic Resources

Optional

For a more in-depth overview of decision trees, read:

- Sarma, K. S. (2017). *Predictive modeling with SAS Enterprise Miner: Practical solutions for business applications* (3rd ed.). Cary, NC: SAS Institute.
    - Chapter 4: Building Decision Tree Models to Predict Response and Risk.

For a more in-depth discussion of decision trees, read:

- Fernandez, G. (2011). *Statistical data mining using SAS applications* (2nd ed.). Boca Raton: CRC Press.
    - Chapter 6: Supervised Learning Methods: Classification.

Unit 4 >> Classification: Cluster Analysis

Introduction

This week you will:

- Compare and contrast cluster analysis methods.
- Describe cluster analysis techniques.
- Evaluate the appropriateness of cluster analysis for a given context and data set.
- Complete a cluster analysis lab.

In this unit, you will learn about cluster analysis and customer segmentation methods. You will also apply your new knowledge by using SAS Enterprise Miner to cluster football players attending the NFL combine, and discuss the appropriateness of cluster analysis and segmentation for your individual project's problem or issue and data.

Learning Activities

u04s1 - Study

For an overview of cluster analysis and segmentation modeling, watch the following short videos:

- Strickland, J. (2016). *Predictive analytics: Clustering* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: EDA and cluster segmentation modeling* [Video]. Skillsoft Ireland Limited.

- Scott, S. (2017). *Predictive analytics: Types of clustering techniques* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Overview of the k-NN algorithm* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Overview of K-Means clustering* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: K-Means clustering termination procedures* [Video]. Skillsoft Ireland Limited.

For an overview of cluster analysis methods in SAS Enterprise Miner, read:

- Collica, R. S. (2017). *Customer segmentation and clustering using SAS Enterprise Miner* (3rd ed.). SAS Institute.
    - Part 1: The Basics.

u04s1 - Learning Components

- Compare and contrast cluster analysis methods.
- Describe cluster analysis techniques.

u04d1 - Cluster Analysis Applications

Evaluate the appropriateness of cluster analysis methods for your individual project. Detail the problem or issue you are trying to address and the data that you are using to do so. If you feel that cluster analysis might be appropriate for your individual project, indicate why you believe this to be true, what you will try to cluster using the methods (observations or variables), and the purpose of the use of clustering techniques for your project.

In addition to evaluating the appropriateness of cluster analysis methods for your individual course project, think of a business question or problem from your current or future career that could be addressed using cluster analysis methods. Explain the business question or problem, the type of data that might be used, and how you might use cluster analysis methods to address this business problem using the type of data described.

Response Guidelines

Please respond to at least two of your classmates' initial responses. Do you agree or disagree with their assertion that cluster analysis methods are or are not appropriate for their data and the issue or problem they are attempting to address? Why or why not?

Course Resources

Graduate Discussion Participation Scoring Guide

u04d1 - Learning Components

- Describe data science methods.
- Select a data analytics project context, topic and data source.
- Compare and contrast cluster analysis methods.
- Describe cluster analysis techniques.
- Evaluate the appropriateness of cluster analysis for a given context and data set.

u04a1 - Cluster Analysis Lab

Assignment Overview

Use the following resources as tutorials for using cluster analysis to segment football players in the football combine data set using SAS Enterprise Miner:

- Wujek, B. (n.d.). The ABCs of selecting clusters. Retrieved from https://video.sas.com/detail/video/4572850292001/the-abcs-of-selecting-clusters?autoStart=true&q=cluster
- Truxillo, C. (n.d.). AS Enterprise Miner: Profiling Segments | Transcript. Retrieved from https://video.sas.com/detail/video/2436413682001/sas-enterprise-miner:-profiling-segments?autoStart=true&q=cluster

The football combine data set can be downloaded here:

- Kaggle. (2017). NFL Combine Data. Retrieved from https://www.kaggle.com/savvastj/nfl-combine-data

Assignment Instructions

Demonstrate your creation of clustering of the football players by height, weight and forty-yard dash by including your code and output, along with a brief narrative of each step you took. At the end of your assignment, please evaluate your completion of this tutorial by discussing portions of the tutorial that were especially informative, surprising or challenging. A lab assignment template is provided in the resources for this assignment.

Assignment Criteria

Your assignment must meet this criteria, please refer to the scoring guide for this assignment.

- Identifies the strengths and challenges for using cluster analysis.
- Evaluates the steps that are needed for using cluster analysis.
- Determines suitability of the application for using cluster analysis for a given context.
- Completes the tutorial for using cluster analysis.
- Describes the benefits for using cluster analysis for a given context.
- Uses communication style and vocabulary appropriate for the target audience.

Assignment Requirements

- **Lab output**: Your code and output you generated for the tutorial.
- **Communication:** Communicate in a manner that is scholarly, professional, respectful, and consistent with expectations for professional practice in technology. Assignment submission must be original work that reflects critical thinking with clear organization of concepts and ideas.
- **Paper length:** Minimum of 2–3 pages, not including the title page or reference page.
- **References:** Cite your data and sources as appropriate.
- **Document format guidelines:** Use the Academic and Professional Document Guidelines linked in the resources as appropriate for your deliverable.
- **Font and font size:** Times New Roman, 11 pt.

Course Resources
Academic and Professional Document Guidelines [PDF]
Lab Assignment Template [DOCX]

u04s2 - Topic Resources

Optional

For an overview of cluster analysis, read:

- Sarma, K. S. (2017). *Predictive modeling with SAS Enterprise Miner: Practical solutions for business applications* (3rd ed.). Cary, NC: SAS Institute.
  - Chapter Two: Getting Started with Predictive Modeling.
    - Section 2.8.4 Variable Clustering Node.
    - Section 2.8.5 Cluster Node.

For a more in-depth discussion of cluster analysis and related techniques, read:

- Fernandez, G. (2011). *Statistical data mining using SAS applications* (2nd ed.). Boca Raton: CRC Press.
    - Chapter 4: Unsupervised Learning Methods.

Unit 5 >> Midpoint Check-In

Introduction

In this unit, you will continue work on your individual project, and will submit your midpoint review for your individual project. You will also have the opportunity within the discussion to get assistance from your classmates and instructor and share any resources or assistance you found helpful.

Learning Activities

u05s1 - Study

The following resources provide information about the ethical, legal, and cultural aspects for using data in predictive and classification models and algorithms.

- Heimann, R. (2015). *Big data legal perspective: Using data ethically* [Video]. Skillsoft Ireland Limited.
- O'Keefe, K. & O Brien, D. (2018). *Ethical data and information management: Concepts, tools and methods.* Kogan Page.
    - Chapter 3: Ethics, Privacy and Analytics.

To prepare for the visualization portion of individual project:

- Tableau for Students provides a one-year free license for Tableau. To get started preparing for the visualization portion of your project, download Tableau Desktop from https://www.tableau.com/academic/students To obtain the license, follow the directions at the website https://www.tableau.com/academic/students.
- In addition to downloading Tableau and requesting a Tableau license directly from the company, students can also access Tableau, as well as SAS Visual Analytics for free through a partnership with Teradata. Students can register with Teradata University Network and gain access to Teradata, Tableau and SAS Visual Analytics here: https://www.teradatauniversitynetwork.com/Register

u05s1 - Learning Components

- Prepare software as required for data analytics and data visualization activities.
- Describe ethical, legal, and cultural aspects for using data in predictive and classification models and algorithms for a given context and data set.

u05d1 - Key Elements of Analytics Project Reports

In this discussion, summarize what you believe to be the key elements or sections of analytic project reports. Remember to support your assertions with references. In addition, discuss the current status of your individual project, in preparation for your midpoint review due at the end of this unit. What barriers have you encountered? What questions do you have about your data, issue or problem you are attempting to address, technical tasks required to fit your model or on the interpretation or write-up of your analysis? Are there any additional resources you found outside of the course that have helped you get to your current status? If so, please feel free to share them with your classmates here.

Response Guidelines

Please respond to at least two of your classmates' initial responses. Discuss whether or not you agree with their list of key elements of analytics project reports. Identify any that you think should be added or removed and discuss why. Assist them in whatever ways you can by providing answers to their questions, technical assistance or additional resources, where appropriate.

Course Resources
Graduate Discussion Participation Scoring Guide

u05d1 - Learning Components

- Describe data science methods.
- Select a data analytics project context, topic and data source.
- List the key elements or sections of analytic project reports.
- Report on the status of a data analytics project.

u05a1 - Analytics Project Report Draft/Outline

Assignment Overview

For this assignment, write a paper to summarize the work you have completed on your individual course project to this point. This summary should include a definition of the business question or opportunity and the data source. Explain how the business question or opportunity and the data align or do not align, and if they do not align, discuss what you will do to rectify any misalignment. Address any potential ethical, legal, global, or cultural factors that may present risk in using the data for the context of your project. Audit and profile the data and identify data issues or potential data issues. Include output, graphs, and other evidence and explanations of your data profiling and audit activities. Address the data issues, including any missing values or erroneous data. Include a detailed specification, or explanation, and code for all data manipulation you completed to address the data issues.

Assignment Instructions

Write a 4-6 page paper that addresses each of the points above, including the code written and used within this assignment as an attachment or appendix (not included in the page limits above). Support your choices of methods with references, as appropriate.

Assignment Criteria

Your assignment must meet this criteria, please refer to the scoring guide for this assignment.

- Defines a business question, problem, or opportunity and the source of data.
- Explains how the data does or does not align with the business question, problem, or opportunity and how any misalignment will be rectified.
- Identifies ethical, legal, global, or cultural factors and potential risk in using the data for a given context.
- Audits the data to identify any data issues or potential data issues and include relevant output and graphs.
- Addresses missing values and erroneous data and include detailed explanation and code for all data cleansing and adjustments.
- Synthesizes multiple sources into key themes or findings.
- Uses communication style and vocabulary appropriate for the target audience.

Assignment Requirements

- **Output:** Your output, graphs, and other evidence and explanations of your data profiling and audit activities.
- **Communication:** Communicate in a manner that is scholarly, professional, respectful, and consistent with expectations for professional practice in technology. Assignment submission must be original work that reflects critical thinking with clear organization of concepts and ideas.
- **Paper length:** Minimum of 4–6 pages, not including the title page or reference page.

- **References:** Cite your data and sources as appropriate.
- **Document format guidelines:** Use the Academic and Professional Document Guidelines linked in the resources as appropriate for your deliverable.
- **Font and font size:** Times New Roman, 11 pt.

Course Resources

Academic and Professional Document Guidelines [PDF]

Unit 6 >> Text Analytics

Introduction

This week you will:

- Compare and contrast text analytics methods.
- Describe text analytics techniques.
- Evaluate the appropriateness of text analytics for a given context and data set.
- Complete a text analytics lab.

In this unit, you will learn about text analytics methods. You will also apply your new knowledge by using SAS Enterprise Miner to cluster and create a predictive model for the SMS spam collection data set and discuss the appropriateness of text mining and sentiment analysis for your individual project's problem or issue and data.

Learning Activities

u06s1 - Study

For an overview of text analytics and text mining, watch the following short videos:

- Scott, S. (2017). *Predictive analytics: Overview of text mining* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Text mining applications* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Text normalization* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Term frequency and inverse document frequency* [Video]. Skillsoft Ireland Limited.

- Scott, S. (2017). *Predictive analytics: Assigning within document predictor variables* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Assigning across document predictor variables* [Video]. Skillsoft Ireland Limited.

u06s1 - Learning Components

- Describe data science methods.
- Compare and contrast text analytics methods.
- Describe text analytics techniques.

u06d1 - Applications of Text Analytics

Evaluate the appropriateness of text analytics methods for your individual project. Detail the problem or issue you are trying to address and the data that you are using to do so. If you feel that text analytics might be appropriate for your individual project, indicate why you believe this to be true, what text you will be analyzing, whether you are attempting to classify the text or use the text to predict a response variable. If you are predicting, identify the response variable you will use the text to predict.

In addition to evaluating the appropriateness of text analytics methods for your individual course project, think of a business question or problem from your current or future career that could be addressed using text analytics methods. Explain the business question or problem, the type of data that might be used, and how you might use text analytics methods to address this business problem using the type of data described.

Response Guidelines

Please respond to at least two of your classmates' initial posts. Do you agree or disagree with their assertion that text analysis methods are or are not appropriate for their data and the issue or problem they are attempting to address? Why or why not?

Course Resources
Graduate Discussion Participation Scoring Guide

u06d1 - Learning Components

- Describe data science methods.

- Compare and contrast text analytics methods.
- Describe text analytics techniques.
- Evaluate the appropriateness of text analytics for a given context and data set.

u06a1 - Text Analytics Lab

Assignment Overview

Use the following resource as a tutorial for using text analysis to cluster the SMS spam data set using SAS Enterprise Miner:

- Miner, G., Elder, J., Fast, A., Hill, T., Nisbet, R., & Delen, D. (2012). Practical text mining and statistical analysis for nonstructured text data applications. Waltham, MA: Elsevier/Academic Press.
  - Read Tutorial Q: A Hands-On Tutorial on Text Mining in SAS - Analysis of Customer Comments for Clustering and Predictive Modeling.

Below is an overview of the steps to follow:

1. Create the project: Log on to SAS Enterprise Miner.
2. Create the diagram: Right click on Diagrams.
3. Create the library: Click on the dropdown menu under File.
4. Create the data source: Right click on the data source to choose the data set.
5. Text mining: Drag and drop the text mining node onto the project diagram.
6. Changing properties: Select parts of speech to ignore.
7. Results: Select Run.

Use class (ham and spam) as your response variable in your predictive model.

The SMS Spam Collection Dataset can be downloaded here:

- UCI Machine Learning. (2016). SMS Spam Collection Dataset. Retrieved from https://www.kaggle.com/uciml/sms-spam-collection-dataset

Assignment Instructions

Demonstrate your creation of the clusters and predictive model of the SMS spam collection data by including your code and output, along with a brief narrative of each step you took. At the end of your assignment, please evaluate your completion of this tutorial by discussing portions of the tutorial that were especially informative, surprising or challenging. A lab assignment template is provided in the resources for this assignment.

Assignment Criteria

Your assignment must meet this criteria, please refer to the scoring guide for this assignment.

- Identifies the strengths and challenges of using text analysis to cluster text data in data set.
- Evaluates the steps that are needed for using text analysis to cluster text data in data set.
- Determines suitability of the application of using text analysis to cluster text data in data set for a given context.
- Completes the tutorial to use text analysis to cluster text data in data set.
- Describes the benefits of using text analysis to cluster text data in data set for a given context.
- Uses communication style and vocabulary appropriate for the target audience.

Assignment Requirements

- **Lab output**: Your code and output you generated for the tutorial.
- **Communication:** Communicate in a manner that is scholarly, professional, respectful, and consistent with expectations for professional practice in technology. Assignment submission must be original work that reflects critical thinking with clear organization of concepts and ideas.
- **Paper length:** Minimum of 2–3 pages, not including the title page or reference page.
- **References:** Cite your data and sources as appropriate.
- **Document format guidelines:** Use the Academic and Professional Document Guidelines linked in the resources as appropriate for your deliverable.
- **Font and font size:** Times New Roman, 11 pt.

Course Resources
Academic and Professional Document Guidelines [PDF]
Lab Assignment Template [DOCX]

u06s2 - Topic Resources

Optional

For a more in-depth overview of text analytics, read:

- Chakraborty, G., Pagolu, M., & Garla, S. (2013). *Text mining and analysis: Practical methods, examples, and case studies using SAS.* SAS Institute.
  - Chapter 1: Introduction to Text Analytics.

- Sarma, K. S. (2017). Predictive modeling with SAS Enterprise Miner: Practical solutions for business applications (3rd ed.). Cary, NC: SAS Institute. ISBN:9781629602646
    - Chapter 9: Introduction to Predictive Modeling with Textual Data.

Unit 7 >> Time Series Methods

Introduction

This week you will:

- Compare and contrast time series methods.
- Describe time series method techniques.
- Evaluate the appropriateness of time series methods for a given context and data set.
- Complete a time series methods lab.

In this unit, you will learn about time series methods. You will also apply your new knowledge by using SAS Enterprise Miner to apply various time series methods to one small portion of the motion sensor data set and discuss the appropriateness of time series methods for your individual project's problem or issue and data.

Learning Activities

u07s1 - Study

For an overview of time series data and related methods, watch the following short videos:

- Scott, S. (2017). *Predictive analytics: Time series overview*[Video]. Skillsoft Ireland Limited.
- Strickland, J. (2016). *Predictive analytics: Time series analysis* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Time series decomposition* [Video]. Skillsoft Ireland Limited.
- Scott. S. (2017). *Predictive analytics: Stationary and nonstationary data series* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Autoregressive models* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics; Moving average models* [Video]. Skillsoft Ireland Limited.
- Scott, S. (2017). *Predictive analytics: Autoregressive Moving Average (ARMA) models* [Video]. Skillsoft Ireland Limited.

u07s1 - Learning Components

- Describe data science methods.
- Compare and contrast time series methods.
- Describe time series method techniques.

u07d1 - Differences Between Time Series Methods

Evaluate the appropriateness of time series methods for your individual project. Detail the problem or issue you are trying to address and the data that you are using to do so. If you feel that time series methods might be appropriate for your individual project, indicate why you believe this to be true, what time-related data you will be analyzing. Also indicate whether you believe there are any seasonal or other cyclical effects in your data and whether your time-related data might require any smoothing or averaging techniques.

In addition to evaluating the appropriateness of time series methods for your individual course project, think of a business question or problem from your current or future career that could be addressed using time series methods. Explain the business question or problem, the type of data that might be used, and how you might use time series methods to address this business problem using the type of data described.

Response Guidelines

Please respond to at least two of your classmates' initial responses. Do you agree or disagree with their assertion that time series methods are or are not appropriate for their data and the issue or problem they are attempting to address? Why or why not?

Course Resources
Graduate Discussion Participation Scoring Guide

u07d1 - Learning Components

- Describe data science methods.
- Compare and contrast time series methods.
- Describe time series method techniques.
- Evaluate the appropriateness of time series methods for a given context and data set.

u07a1 - Time Series Methods Lab

Assignment Overview

Use the following resources as tutorials for using time series methods to model the motion sensor data set using SAS Enterprise Miner and SAS Enterprise Guide:

- Schubert, S., & Lee, T. (2011). Time series data mining with SAS Enterprise Miner. Retrieved from https://support.sas.com/resources/papers/proceedings11/160-2011.pdf
- Additional Resource Document: Fitting ARIMA Models to Time Series Data in SAS Enterprise Guide.

Below is an overview of the steps to follow:

1. Fit a time series model using exponential smoothing in SAS Enterprise Miner.
2. Fit an autoregressive time series model (AR(1) or ARIMA(1,0,0)) using SAS Enterprise Guide.
3. Fit a moving average time series model (MA(1) or ARIMA(0,0,1)) using SAS Enterprise Guide.
4. Fit another autoregressive integrative moving average model (ARIMA) of your choice, using SAS Enterprise Guide.

The motion sensor data set can be downloaded here: https://www.kaggle.com/malekzadeh/motionsense-dataset

- Use sub_1.csv under the dws_1 folder, which represents subject 1's movement in all trials when moving down stairs. The variable you will use for the time series analysis is userAcceleration.x, which is the user acceleration in the x direction.
- The sub_1.csv file is also located in the resources for this assignment.

The reference for this data set is:

Malekzadeh, M., Clegg, R. G., Cavallaro, A., and Haddadi, H. (2018). Protecting sensory data against sensitive inferences. In W-P2DS'18: 1st workshop on privacy by design in distributed systems, April 23–26, 2018, Porto, Portugal. New York, New York: ACM.

Assignment Instructions

Demonstrate your creation of the time series models for the motion sensor data by including your code and output, along with a brief narrative of each step you took. At the end of your assignment, please evaluate your completion of this tutorial by discussing portions of the tutorial that were especially informative, surprising or challenging. A lab assignment template is provided in the resources for this assignment.

Assignment Criteria

Your assignment must meet this criteria, please refer to the scoring guide for this assignment.

- Identifies the strengths and challenges of creating a time series model.
- Evaluates the steps that are needed to create a time series model.
- Determines suitability of the application of a time series model for a given context.
- Completes the tutorial to create a time series model.
- Describes the benefits of using a time series model for a given context.
- Uses communication style and vocabulary appropriate for the target audience.

Assignment Requirements

- **Lab output**: Your code and output you generated for the tutorial.
- **Communication:** Communicate in a manner that is scholarly, professional, respectful, and consistent with expectations for professional practice in technology. Assignment submission must be original work that reflects critical thinking with clear organization of concepts and ideas.
- **Paper length:** Minimum of 2–3 pages, not including the title page or reference page.
- **References:** Cite your data and sources as appropriate.
- **Document format guidelines:** Use the Academic and Professional Document Guidelines linked in the resources as appropriate for your deliverable.
- **Font and font size:** Times New Roman, 11 pt.

Course Resources
Academic and Professional Document Guidelines [PDF]
Lab Assignment Template [DOCX]
Fitting ARIMA Models to Time Series Data in SAS Enterprise Guide [PDF]

u07s2 - Topic Resources

Optional

For a more in-depth overview of time series methods, read:

- Schubert, S., & Lee, T. (2011). Time series data mining with SAS Enterprise Miner. Retrieved from https://support.sas.com/resources/papers/proceedings11/160-2011.pdf

Unit 8 >> Reporting on Models and Analyses

Introduction

In this unit, you will complete the work on your individual project, and will submit your final technical report for your individual project. You will have an opportunity to discuss the technical report with your classmates and instructors to help identify any details or sections you might be missing or need to add to your technical report for your individual course project.

Learning Activities

u08s1 - Study

Review the following resources about the ethical, legal, and cultural aspects for using data in predictive and classification models and algorithms.

- Heimann, R. (2015). *Big data legal perspective: Using data ethically* [Video]. Skillsoft Ireland Limited.
- O'Keefe, K. & O Brien, D. (2018). *Ethical data and information management: Concepts, tools and methods.* Kogan Page.
    - Chapter 3: Ethics, Privacy and Analytics.

u08s1 - Learning Components

- Determine actions to mitigate risk due to ethical, legal, and cultural factor when using data in predictive and classification models and algorithms for a given context and data set.

u08d1 - Key Elements of Analytics Project Reports

In this discussion, summarize what you believe to be the best practices for creating detailed technical reports on models and analytic results. Remember to support your assertions with references. Answer the following questions:

1. What are the top 3-5 tips or best practices you have found for creating detailed technical reports on models and analyses? (Support each with a reference, to provide a source for further research and

information for your classmates.)
2. What is the purpose of the detailed technical report and what level of detail is required in a report of this type?
3. Who do you believe is the audience for the detailed technical report for analyses and models and what type of language should be used in a report of this type?
4. What is one section or detail that you didn't originally include in your technical report that you decided to add after further research on best practices for reports of this type?

Response Guidelines

Please respond to at least two of your classmates' initial responses. Discuss whether or not you agree with their list of best practices for reporting on analytics projects and results. Identify any that you think should be added or removed and discuss why.

Course Resources
Graduate Discussion Participation Scoring Guide

u08d1 - Learning Components

- Describe data science methods.
- Select a data analytics project context, topic and data source.
- Describe best practices for creating detailed technical reports on models and analytic results.

u08a1 - Analytics Project Final Report

Overview

This assignment is is final analysis report for your course project. You will summarize the work you did to address the business question or opportunity and the data set used for your analysis. In your report you will provide an explanation that justifies your choice of data analysis for the business question or opportunity you addresses in your project.

Instructions

1. Write a paper of 4-6 pages (not including appendices and references) that summarizes your business question or opportunity, the source of data used, the analysis method used, and justification for your choice of that analytic method for the given data and business question or opportunity.
2. Include a section to address ethical, legal, global, or cultural factors for the analytic method applied in the context of your project.

3. Include the SAS code created during your analysis project as an appendix or attachment.
4. Justify your selection of the analytic method you used on your given data set to address your selected business question or opportunity.
5. Summarize and explain your analysis in business or industry language.

Assignment Criteria

Your assignment must meet this criteria, please refer to the scoring guide for this assignment.

- Applies an appropriate analytic method to address a business question or opportunity.
- Identifies ethical, legal, global, or cultural factors for a given context and ways to mitigate risk in the applied analytic method.
- Identifies appropriate uses of the results and potential issues of the analytics project.
- Uses communication style and vocabulary appropriate for the target audience.
- Provides a logical argument in support of conclusions or recommendations.

Assignment Requirements

- **Output**: Your output, graphs, and other evidence and explanations of your project activities.
- **Communication:** Communicate in a manner that is scholarly, professional, respectful, and consistent with expectations for professional practice in technology. Assignment submission must be original work that reflects critical thinking with clear organization of concepts and ideas.
- **Paper length:** Minimum of 4–6 pages, not including the title page or reference page.
- **References:** Cite your data and sources as appropriate.
- **Document format guidelines:** Use the Academic and Professional Document Guidelines linked in the resources as appropriate for your deliverable.
- **Font and font size:** Times New Roman, 11 pt.

Course Resources

Academic and Professional Document Guidelines [PDF]

Unit 9 >> Data Visualization

Introduction

This week you will:

- Compare and contrast data visualization methods.
- Describe data visualization techniques.
- Evaluate the appropriateness of data visualizations for a given context and data set.
- Complete a data visualization lab.

In this unit, you will learn about visual analytics and data visualization methods. You will also apply your new knowledge by using Tableau to create a data visualization and discuss which data visualizations and visual analytic methods are appropriate for your individual project's problem or issue and data.

Learning Activities

u09s1 - Study

For an overview of data visualization, read the following:

- Watson, H. J. (2017). Data Visualization, Data Interpreters, and Storytelling. *Business Intelligence Journal, 22*(1), 5–10.

For an overview of visual analytics and aspects to consider when creating data visualizations, watch the following short videos:

- Scott, S. (2017). *Data science essentials: Effective communication and visualization* [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). *Data visualization essentials: What is data visualization* [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). *Data visualization essentials: Data visualization versus infographics* [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). *Data visualization essentials: Understanding the data being visualized* [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). *Data visualization essentials: Choosing the type of visualization for the data* [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). *Data visualization essentials: Defining the narrative* [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). *Data visualization essentials: Knowing your audience* [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). *Data visualization essentials: Designing data visualizations* [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). *Data visualization essentials: Traits and visual perception* [Video]. Skillsoft Ireland Limited.
- Khoury, J. (2015). *Data visualization essentials: Adding effective visuals and text* [Video]. Skillsoft Ireland Limited.

- Khoury, J. (2015). *Data visualization essentials: Adding functionality and color* [Video]. Skillsoft Ireland Limited.

u09s1 - Learning Components

- Prepare software as required for data analytics and data visualization activities.
- Compare and contrast data visualization methods.
- Describe data visualization techniques.
- Evaluate the appropriateness of data visualizations for a given context and data set.

u09d1 - Data Visualization Considerations and Best Practices

Discuss which data visualizations or visual analytics methods you will use for your individual project. Detail the problem or issue you are trying to address and the data that you are using to do so. Identify your audience for your visualization, the narrative you are trying to portray with the visualization, and the type of data visualization or visual analytics method that you feel is most appropriate for your individual project. Discuss any text, color or functionality choices you are making for your visualization and justify your choices using references.

In addition to evaluating the appropriateness of visual analytics methods for your individual course project, think of a business question or problem from your current or future career that could be addressed using visual analytics methods. Explain the business question or problem, the type of data that might be used, and how you might use visual analytics methods to address this business problem using the type of data described.

Response Guidelines

Please respond to at least two of your classmates' initial responses. Do you agree or disagree with their selection of the visualization method for their data and the issue or problem they are attempting to address? Why or why not?

Course Resources

Graduate Discussion Participation Scoring Guide

u09d1 - Learning Components

- Compare and contrast data visualization methods.

- Apply best practices for layout and design of an executive summary presentation.
- Describe data visualization techniques.
- Evaluate the appropriateness of data visualizations for a given context and data set.

u09a1 - Visual Analytics Lab

Assignment Overview

Use the following resources as tutorials for creating one or more data visualizations using your individual project data:

- Acharya, S. & Chellappan, S. (2017). *Pro Tableau: A step-by-step guide.* Apress.
- Tableau. (n.d.). Tableau Starter Kits: Author a Viz in Tableau Desktop. Retrieved from https://www.tableau.com/learn/starter-kits/author/desktop
- Tableau training and tutorials. (n.d.). Retrieved from https://www.tableau.com/learn/training
- Save Workbooks to Tableau Public. (n.d.). Retrieved from http://onlinehelp.tableau.com/current/pro/desktop/en-us/publish_workbooks_tableaupublic.html

Below is an overview of the steps to follow:

1. Import your data into Tableau.
2. Select and create at least one appropriate visualization using your individual project data in Tableau using the guides and starter kits above and the resources from this unit's study.
3. Save your Tableau workbook and upload it to Tableau Public.
4. Share the link to your Tableau workbook on Tableau Public within your assignment.

Assignment Instructions

Write a 1-3 page summary of your visualization, including the link to your visualization on Tableau Public. Explain the audience for your visualization and the narrative you are attempting to convey. Explain (and justify using references) your selection of visualization method or technique, text, color and functionality used within your visualization. Analyze the appropriateness of your visualization for your data and the issue or problem you are attempting to address within your individual project.

Assignment Criteria

Your assignment must meet this criteria, please refer to the scoring guide for this assignment.

- Identifies the strengths and challenges of creating data visualizations.
- Evaluates the steps that are needed to create data visualizations.
- Determines suitability of the application of data visualizations for a given context.
- Completes the tutorial to create data visualizations.
- Describes the benefits of using data visualizations for a given context.

- Uses communication style and vocabulary appropriate for the target audience.

Assignment Requirements

- **Lab output**: The link to your visualization on Tableau Public that you generated for the tutorial.
- **Communication:** Communicate in a manner that is scholarly, professional, respectful, and consistent with expectations for professional practice in technology. Assignment submission must be original work that reflects critical thinking with clear organization of concepts and ideas.
- **Paper length:** Minimum of 2–3 pages, not including the title page or reference page.
- **References:** Cite your data and sources as appropriate.
- **Document format guidelines:** Use the Academic and Professional Document Guidelines linked in the resources as appropriate for your deliverable.
- **Font and font size:** Times New Roman, 11 pt.

Course Resources

Academic and Professional Document Guidelines [PDF]

Unit 10 >> Presenting Analytics to Executives

Introduction

This week you will:

- Apply best practices for layout and design of an executive summary presentation.
- Create an executive summary with appropriate focus and detail for the target audience.

In this unit you will complete the work on translating the results your individual project into business language for an executive audience, and you will submit your analytics project executive presentation.

Learning Activities

u10s1 - Study

There are no assigned readings for this unit.

u10s1 - Learning Components

- Apply best practices for layout and design of an executive summary presentation.

u10a1 - Analytics Project Executive Presentation

Overview

For this assignment you will submit an executive presentation on the results of your analytics project.

Instructions

Create an 8-10 slide PowerPoint presentation that outlines your analysis and is appropriate for an executive audience. Within your presentation include the following:

1. Summarize the business question or opportunity being addressed.
2. Summarize the data you used to address the business question or opportunity.
3. Summarize the data cleansing that was completed and any changes that were made to the data before analyzing it.
4. Summarize the results of the analysis, including the methods used and how the results address the business question or opportunity.
5. Present the results in language, level of detail, and content appropriate for an executive audience.
6. Include a title page and references, not included in the slide limits indicated above.

Assignment Criteria

Your assignment must meet this criteria, please refer to the scoring guide for this assignment.

- Applies analytic methods to solve organizational issues.
- Creates an executive summary with appropriate focus and detail for the target audience.
- Organizes content so ideas flow logically with smooth transitions.
- Integrates visual elements that clarify or highlight key points.
- Writes clearly with few spelling, grammatical or mechanical errors.
- Uses required formats, templates and citation styles.

Assignment Requirements

- **Written communication:** Presentation is free of errors that detract from the overall message.
- **Length of presentation:** Approximately 8–10 presentation slides not including the title or references slides.
- **References:** Cite your data and sources as appropriate.

- **Document format guidelines:** Use the Academic and Professional Document Guidelines linked in the resources as appropriate for your deliverable.
- **Font and font size:** As appropriate for a PowerPoint presentation.

Course Resources

Academic and Professional Document Guidelines [PDF]

u10d1 - Course Reflection

In this discussion, reflect on your progress and learning throughout the course. Identify at least one topic that you feel you learned the most about throughout the course and identify at least one topic that you wish to learn more about. What were your favorite portions of the course and which were your least favorite? If you had to give incoming learners one piece of advice about the course, what would it be and why?

Response Guidelines

Please respond to at least two of your classmates' initial responses.

Course Resources

Graduate Discussion Participation Scoring Guide