Course Name Syllabus

Course Name: Data Preparation for Analytics

Instructor Information

Name: Tamara B. Sipes, Ph.D.

Email: tbsipes@gmail.com

Communication Policy:

You may contact me either by email or via remote meeting or in person during my office hours, TBD. **IMPORTANT:** Outside of office hours, I will try to answer your emails within 48 hours of receiving them.

Welcome

Welcome to the Data Preparation for Analytics class. This class will teach you how to get the most out of your data, big or small, and how to prepare it so that the information content is best represented to the learning schema. My teaching philosophy is to offer structured and clear explanation of the subject matter and ground it in a variety of hands-on examples, real-life case studies and hands on applications of the acquired new concepts, methods and technologies. This approach fosters deep understanding of the topics covered in the class and advances the readiness of students to apply the newly learned techniques in their own projects.

Course Purpose and Prerequisites

The course is intended for anyone learning about predictive analytics and data science as well as for data science practitioners. The purpose is to learn how to best prepare the data for analytics, to get familiar with the basic data preparation techniques as well as the advanced techniques.

Course Goal and Objectives

This class offers in-depth coverage of data preparation techniques and a step-by-step approach through a variety of methods while providing practical illustrations using real data sets. The hands-on exercises will anchor the learned concepts and offer valuable first-hand experience in cleaning, filtering, and preparing the data for mining and predictive or descriptive modeling. The goal is to transform the datasets so that their information content is best exposed to the mining tool, so that the gist of the data is more readily grasped than when the data is not cleaned and prepared for analytics.

Learning Objectives

By the end of this course, the student will be able to:

- Apply basic data preparation techniques to any dataset.

- Apply advanced data preparation techniques.

- Know how to prepare data for different modeling methods.

- Know how to deal with difficult data.

- Prepare the data so that its information content is best exposed to the learning method.

Course Materials/Textbooks

The required textbook (available online on amaon.com) is:

Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (Morgan Kaufmann Series in Data Management Systems) 3rd Edition

by Ian H. Witten (Author), Eibe Frank (Author), Mark A. Hall (Author)

Course Overview

This course has 6 lessons and 4 assignments, including the final exam, due by the

end of the course. The lessons are as follows:

- Lesson 1: Introduction to Data Preparation for Analytics
  - Background
  - The Process
  - Motivation and Importance
  - State of Affairs
  - Definition and Goal
  - Prerequisites
  - Examples
- Lesson 2: Input and Output
  - Inputs, Outputs, Models and Decisions
  - Concepts, Instances and Attributes
  - Obtaining the Data
  - Preparing the Input
  - Types of Measurement
  - Examples
  - Modelling Tools and Data Preparation
  - Stages of Data Preparation
- Lesson 3: Basic Data Preparation
  - Variables as objects
  - Removing variables
  - Sparcity
  - Monotonicity
  - Remapping
  - Outliers
  - Anachronisms
- Lesson 4: Advanced Data Preparation
  - Data cleaning
  - Data transformation
  - Data reduction
  - Discretization and concept hierarchies
- Lesson 5: Dealing with Difficult Datasets
  - Dealing with Inedaquate Attribute List
  - Dealing with Wide Datasets
  - Dealing with Imbalanced Data
  - Dealing with Big Data
- Lesson 6: Bringing it All Together
  - Preparing the Dataset

- Using the Prepared Data
- Hands-on exercises
- Summary

Online Course Structure

The course is organized using the course menu (left side of your screen):

Announcements	This is the first page you see upon entering your course. Your instructor will post weekly announcements and reminders here.
Introduction	Contains an introduction to the course and instructor biography.
Syllabus	Contains the course outline, learning objectives, weekly assignments and course details.
Lessons	If it's a fully online course, this section will have the instructor's weekly audio/ image lectures. The lectures are self-paced and can be replayed like a video movie (start, pause, rewind, etc.).
Discussion Board	Optional communication tool for the class
Assignments	Assignments, quizzes, Course Evaluation, and the Final Exam are available here.
Resources	Additional readings and handouts, web site links, and PowerPoint presentations are here.
Contacts	Instructor, student services and online learning support contact information is listed here.

Tools	Check your grades (My Grades), or access the Blackboard User Manual (User Manual) here.

Assignments Schedule and Grades

Assignment	Due Date	Grade %
Assignment #1	end of 4th week of class	20%
Assignment #2	end of 5th week of class	20%
Assignment #3	end of 6th week of class	25%
Assignment #4	end of 7th week of class	35%

Requirements

In order to satisfy course requirements, class participants must go through all the Lessons, complete all the required reading and complete all the assignments.

IMPORTANT! Late assignments (anything posted or sent after the due date) will be graded -1 point for each day late unless due to a verifiable medical or family emergency. Late assignments will be accepted at the discretion of the instructor and cannot be accepted more than 1 week late.

Grades

No late assignments or quizzes are accepted.

Grades are based on points and the letter grades are given as follows:

A+   97-100

A   94-96

A-   90-93

B+   87-89

B   84-86

B-   80-83

C+   77-79

C   74-76

C-   70-73

D+   67-69

D   65-66

F   0-64

You may check your grade anytime by clicking **Course Tools** and then **My Grades**. This will show you the points you have earned so far in this course.

Student Resources

On any Blackboard screen, there are tabs across the top and one is called the Student Tab. There is information on how to get started as a student and who to contact if you encounter any problems. There are also videos and written instructions on how to do some of the most common things in Blackboard.

Another one of these tabs is called FAQ (Frequently Asked Questions). If you click on the Students Category (on the left), you can find step-by-step directions for everything from sending email to uploading your assignments to posting a reply on the discussion board.

Campus Emergencies

In the event of an emergency, information will be posted at UC San Diego Extension (http://extension.ucsd.edu/). Extension students must access the website to find out the status of the emergency situation. Email and or phone lines may not be accessible. Information will be updated online as the situation progresses and an ALL CLEAR will be posted once the situation is resolved.

Code of Conduct

All participants in a course at UC San Diego Extension are bound by the University of California Code of Conduct, found at http://www.ucop.edu/ucophome/coordrev/ucpolicies/aos/uc100.html.

Academic Honesty Policy

The University is an institution of learning, research, and scholarship predicated on the existence of an environment of honesty and integrity. As members of the academic community, faculty, students, and administrative officials share responsibility for maintaining this environment. It is essential that all members of the academic community subscribe to the ideal of academic honesty and integrity and accept individual responsibility for their work. Academic dishonesty is unacceptable and will not be tolerated at the University of California. Cheating, forgery, dishonest conduct, plagiarism, and collusion in dishonest activities erode the University's educational, research, and social roles.

If students who knowingly or intentionally conduct or help another student perform dishonest conduct, acts of cheating, or plagiarism will be subject to disciplinary action at the discretion of UC San Diego Extension.