Business Analytics I: Data Mining & Predictive Analytics

ECB353 Block 5 2019

Cornell College

General

Location:	College 118
Time:	MWF: 9:30 a.m. – 12:30 p.m.
	TTh: 12:30 – 3:00 p.m.
Instructor:	Cindy Bradley, Lecturer Business Analytics
Office Hours:	MTWTh: 3:00 – 4:00 p.m. Other times by appointment
Office:	College Hall 216
Office Phone:	319-895-4283
Email:	cbradley@cornellcollege.edu

Course Overview

This course introduces the data mining process as well as a variety of algorithms used in the area of predictive analytics. In the first week of the course we will cover the basics of the data mining process including data preprocessing, partitioning and model performance evaluation. During weeks two and three, we will introduce several common methods used to predict numerical and classification outcomes.

Course Learning Objectives

The content of this course supports the Educational Priorities and Outcomes of Cornell College and places emphasis on the following educational priorities:

- *Knowledge* – You will develop knowledge by learning how methods drawn from disciplines within statistics and computer science can be applied to solve a variety of real world problems.
- *Reasoning* – You will develop your analytical reasoning skills by learning how to evaluate and interpret results from various sources of data.
- *Inquiry* – You will learn how predictive analytics can be applied to solve many complex problems across a variety of disciplines.

Expectations of Students

Attendance Policy

Daily attendance is expected and required to succeed in this class. This is an upper level analytics course and will move at a rapid pace. Missing class WILL result in you failing significantly behind your peers. Excused absences for health, family-emergencies and college sponsored events are approved only with written documentation. If you need to miss class for any reason, you should notify me ahead of time. Each unexcused absence will result in a 25 point deduction in your final point grade.

Electronic Device Policy

We will spend considerable time on your laptops during this course. Come to class prepared with your laptop fully charged. Random web browsing or any unrelated use of electronic devices during class time is not acceptable. Other electronic devices (especially cell phones) may not be used in our classroom while class is in session. Abuse of the electronic device policy will impact your final grade to the same degree as an unexcused absence.

Late Work Policy

No credit will be assigned to daily homework assignments that are turned in late. Weekly projects turned in late will receive 25% deduction if received within 24 hours of the deadline, 50% deduction within 48 hours and zero credit thereafter. **All assignments/projects are to be submitted through Moodle and a strict submission cutoff will be enforced (note: late projects will need to be submitted directly to me via email).** Keep in mind computers fail… at some point your computer will freeze/crash and you will lose something. This is **not** an excuse for late work. Get started early and save often!

Course Components

Daily assignments & quizzes

Each session I will assign a lab assignment from the textbook to be completed outside of class and submitted prior to the next class session. Assignments will be graded based on completion and solutions to homework will be posted on Moodle after the submission deadline. On selected class days, there will be in class quizzes consisting of a few questions from the reading/previous lecture.

Weekly analytics projects

These projects are intended to develop your knowledge of the content discussed each week. For each project I will provide you with a business scenario, associated dataset and ask you to produce a number of different outputs, including a written management report of the results of your analysis to a potential decision maker.

Mid-term & Final Exam

A mid term exam will be given on the second Friday of the block. A final (comprehensive) exam will be given in class during the final day of the block. Both exams will contain a written short answer section and a programming section to be completed in XL Miner. The programming section will be open book and notes

Grading Policy

Daily assignments & quizzes	150 points
Weekly case projects (3 @ 150 points each)	450 points
Mid-term Exam	200 points
Final exam	200 points
Total Points	1,000 points

Course grades will be based upon the following grading scale:

A	93-100%	B+	87-89%	C+	77-79%	D+	67-69%
A-	90-92%	B	83-86%	C	73-76%	D	63-66%
		B-	80-82%	C-	70-72%	D-	60-62%

Academic Integrity

Cornell College expects all members of the Cornell community to act with academic integrity. An important aspect of academic integrity is respecting the work of others. A student is expected to explicitly acknowledge ideas, claims, observations, or data of others, unless generally known. When a piece of work is submitted for credit, a student is asserting that the submission is her or his work unless there is a citation of a specific source. If there is no appropriate acknowledgement of sources, whether intended or not, this may constitute a violation of the College's requirement for honesty in academic work and may be treated as a case of academic dishonesty. The procedures regarding how the College deals with cases of academic dishonesty appear in The Catalogue, under the heading "Academic Honesty."

Equality of Opportunity

Cornell College makes reasonable accommodations for persons with disabilities. Students should notify the Coordinator of Academic Support and Advising and their course instructor of any disability related accommodations within the first three days of the term for which the accommodations are required, due to the fast pace of the block format. For more information on the documentation required to establish the need for accommodations and the process of requesting the accommodations, see http://www.cornellcollege.edu/academic-support-and-advising/disabilities/index.shtml.

Resources

Textbook
Required
Data Mining for Business Analytics: Concepts, Techniques and Applications with XLMiner – Shmueli, Bruce, and Patel (ISBN: 978-1118729274)
Predictive Analytics, The Power to Predict Who Will Click, Buy, Lie or Die – Eric Siegel (ISBN:978-1119145677)

Software
You will be required to use a computer throughout this course. If you do not have a personal laptop, you should check one out through the library loaner laptop program. If you choose to use a personal computer for this class, the following software should be installed:

- **Microsoft Excel**: If you do not have Microsoft Excel, it is available for download via Cornell's Campus Agreement with Microsoft at https://products.office.com/en-us/student/office-in-education . Students will need to utilize their Cornell email address to obtain the software for free.
    - The Data Analysis Toolpak plug in should be enabled within Excel for this course
- **Analytical Solver –** Installation and license acquisition will be discussed the first day of class.
    - Course code is **ECBCB19SP** and book code is **SDMBI3**

Schedule Outline

Week 1: Preliminaries

Date	Topics	Read Prior to Class	Deadlines
Day 1 Monday, January 14	Course Introduction Overview of the Data Mining Process Installation of XL Miner	Chapter 1 & 2 (thru 2.3) – Shmueli Introduction & Chapter 1 - Siegel	
Day 2 Tuesday, January 15	Data Preprocessing & Exploratory Data Analysis	Chapter 5 – Powell, Baker (See Moodle) Chapter 3 – Siegel	
Day 3 Wednesday, January 16	Dimension Reduction Partitioning Introduction to XL Miner	Chapter 4 – Shmueli	Assign 1 (8 am)
Day 4 Thursday, January 17	Model Performance Evaluation	Chapter 5 – Shmueli	Assign 2 (8 am)
Day 5 Friday, January 18	Review Linear Regression	Chapter 7 – Camm (See Moodle)	Assign 3 (8 am)
Sunday, January 20			Project 1 (9 pm)

Week 2: Supervised Learning Models for Prediction & Classification

Date	Topics	Read Prior to Class	Deadlines
Day 6 Monday, January 21	Review Project 1 Advanced Topics in Linear Regression Building a Model in XL Miner	Chapter 6 - Shmueli	
Day 7 Tuesday, January 22	K Nearest Neighbors (k-NN)	Chapter 7 - Shmueli	Assign 4 (8 am)
Day 8 Wednesday, January 23	Classification and Regression Trees	Chapter 9 Shmueli Chapter 4 - Siegel	Assign 5 (8 am)
Day 9 Thursday, January 24	Naïve Bayes Review for Midterm Exam	Chapter 8 – Shmueli	Assign 6 (8 am)
Day 10 Friday, January 25	Midterm Exam		
Sunday, January 27			Project 2 (9 pm)

Week 3 – Supervised Learning Models for Prediction & Classification

Date	Topics	Read Prior to Class	Deadlines
Day 11 Monday, January 28	Review Project 2 Review Midterm Exam Logistic Regression	Chapter 10 - Shmueli	
Day 12 Tuesday, January 29	Neural Nets	Chapter 11 – Shmueli	Assign 7 (8 am)
Day 13 Wednesday, January 30	Discriminant Analysis	Chapter 12 - Shmueli	Assign 8 (8 am)
Day 14 Thursday, January 31	Enhancing Model Performance – Ensemble Methods	Chapter 13 – Shmueli Chapter 5 - Siegel	Assign 9 (8 am)
Day 15 Friday, February 1	Text Mining	Chapter 20 – Shmueli Chapter 7 - Siegel	Assign 10 (8 am)
Sunday, February 3			Project 3 (9 pm)

Week 4 – Wrap Up

Date	Topics	Read Prior to Class	Deadlines
Day 16 Monday, February 4	Review Project 3 Data and Ethics Discussion	Chapter 2 – Siegel	
Day 17 Tuesday, February 5	Final Exam Review	N/A	
Day 18 Wednesday, February 6	Final Exam	N/A	