

# Northeastern University

## College of Professional Studies

### Course Syllabus

#### **6040 Data Mining Applications** **Summer 2017, 6-week term** **July 10 – September 02, 2017**

Instructor Name: Marcus Ellis  
E-mail: [m.ellis@northwestern.edu](mailto:m.ellis@northwestern.edu)  
Phone Number: 336-692-0279

The very best way to reach me is by text 336-692-0279. Other than that email is perfectly fine. I won't have specific office hours as I will let you tell me if you need to meet and from there we can set up a time that works for both of us.

#### **Required Text(s)/Software/Tools:**

Secrets of Analytical Leaders: Insights from Information Insiders ... ISBN 10: 1935504347 ISBN 13: 9781935504344.

R Graphics Cookbook (available free here):  
<http://ase.tufts.edu/bugs/guide/assets/R%20Graphics%20Cookbook.pdf>

Exploratory Data Analysis with R (available free here): <https://leanpub.com/exdata>

You will need to install R <https://cran.r-project.org/bin/windows/base/>  
and R Studio <https://www.rstudio.com/products/rstudio/download/>

#### **Course Prerequisites**

There are no formal prerequisites but you should feel comfortable with some statistical packages such as SAS or R or Python or even some basic SQL coding. We will be using R extensively in this course. You should also have taken some college level statistics or econometrics..

#### **Course Description**

Introduces the theories and tools for intensive data analysis methods and data mining techniques such as rule-based learning, decision trees, clustering, and association-rule mining. Also covers interpretation of the mined patterns using visualization techniques. Offers students an opportunity to gain the knowledge and experience to apply modern data-mining techniques for effective large-scale data pattern recognition and insight discovery. Data analysis software is introduced. Student teams evaluate, analyze, and report data for the methods used and insights discovered during case studies.

#### **Course Outcomes**

- Consider the application of data mining into the leadership of an organization to achieve stakeholder alignment and business objectives (understand why you need data in order to compete)
- Discern between noise and “sound data” with which to begin your modeling to answer business questions
- Develop your project that encompasses data mining, EDA, and storytelling in a way that conveys actionable insights to your peers and leadership
- Create an analysis driven culture within an organization through professional leadership methods
- Describe the various algorithms available to you for modeling and be able to explain why you choose one over another.

## Course Methodology

Each week, you will be expected to:

1. Review the week's learning objectives.
2. Complete all assigned readings.
3. Complete all lecture materials for the week.
4. Participate in the Discussion Board.
5. Complete and submit all assignments and tests by the due dates.

## Participation/Discussion Board

<i>Discussion Forums</i> Weeks 1 to 8 @ max. 10 pts/week	<b>10%</b>
<i>Quizzes:</i> 4 quizzes @ max. 10 pts/quiz	<b>20%</b>
<i>Assignments:</i> 5 @ max pts/100 each	<b>50%</b>
<i>Final Presentation</i> (Week 6) max pts/100	<b>20%</b>
<i>Total:</i>	<b>100 %</b>

## Communication/Submission of Work

In the Assignments folder, click on the View/Complete Assignment link to view and each assignment. Attach your completed assignments here and click Submit to turn them in to me. Once your assignment has been graded, you will be able to view the grade and feedback I have provided by clicking on My Grades in the Tools module from the Northeastern University Online Campus tab.

## Grading/Evaluation Standards

All written contributions should follow the APA writing style, in particular, the requirements how to lay out a paper, as well as how to cite and reference correctly. You can purchase the APA style guide from the bookstore, but the purchase of the guide is not essential. You can get an excellent guide to APA style on the Purdue Online Writing Lab page. These resources will guide you in creating references for your paper correctly.

## Discussion Participation and Grading Criteria

You are required to participate in the discussion sessions. Your initial posting for each discussion question is the end of Day 3 (11:59 PM, ET). You are to respond to at least 2 of your classmate's postings by the end of Day 7 (11:59 PM, ET).

<b>Score (grading points)</b>	<b>Description for Scoring Each Week's Discussions</b> <i>The length of the initial contribution to the discussion topic should not exceed 250 words. Actively responding to another student's initial submission means entering at least one response to that initial submission. Short responses such as "I agree," will not be counted as a valid response. You must provide the rationale as to why you agree or disagree with another student's submission.</i>
5.00	<b>Exceptional Participation – Met all of these conditions:</b> <b>High quality of comments</b> —Student explored others' comments and built on others' insights. The contributions are outstanding and represent a high-value added discussion which brings in new insights, material, and references. Author builds on discussions of others and has several High-level contributions during the week.
4.00	<b>High Level Participation – Met all these conditions:</b> <b>High quality of comments</b> —Student explored others' comments and built on others' insights. The contributions are outstanding and represent a high- value added discussion which brings in new insights, material, and references. Author builds on discussions of others and has several High-level contributions during the week.
3.00	<b>Average Participation – Met all of these conditions:</b> <b>Average quality of comments</b> —Author has been quite active in discussions during the week and made many valuable contributions, building on and enhancing others' comments by providing contributions which include references and citations to works of others on the topic.
2.00	<b>Low Level (Minimal) Participation – Met all these conditions:</b> <b>Low quality of comments</b> —Author has been active in discussions during week and made some valuable contributions, building on and enhancing others' comments by providing contributions which include references and citations to works of others on the topic.
1.00	<b>Minimal Participation</b> <b>Minimal quality of comments</b> —Author has been in discussion during the week but tends to repeat others or make opinion-related statements. Quite below average postings.
0.00	<b>Inadequate Participation</b> <b>Minimal quality of comments</b> —Author is not active and postings would be considered to be of a personal nature that do not contribute to the knowledge of

the course. Postings are well below average as they merely restate or provide personal opinions.

### Class Schedule / Topical Outline

Week	Dates	Topic	Assignments
1	7/10-7/16	Introduction to Data Mining	Introduction to the tm Package Text Mining in R
2	7/17-7/23	Finding Patterns in Data and EDA	EDA of data and Intro Regression and ggplot
3	7/24-7/30	Data Munging and Data Wrangling	Data Munging and written project proposal
4	7/31-8/06	Project Kick-Off	Find your data, explore your data, prepare your data.
5	8/7-8/13	Basic Data Mining Algorithms 1	Data Mining Algorithms 1
6	8/14-8/20	Basic Data Mining Algorithms 2	Data Mining Algorithms 2
7	8/21-8/27	Basic Data Mining Algorithms 3	Data Mining Algorithms 3
8	8/28-9/03	Show Time	Presentations

### Academic Integrity Policy

The University views academic dishonesty as one of the most serious offenses that a student can commit while in college and imposes appropriate punitive sanctions on violators. Here are some examples of academic dishonesty. While this is not an all-inclusive list, we hope this will help you to understand some of the things instructors look for. The following is excerpted from the University's policy on academic integrity; the complete policy is available on the [Office of Student Conduct and Conflict Resolution web page](#).

*Cheating* – intentionally using or attempting to use unauthorized materials, information or study aids in an academic exercise

*Fabrication* – intentional and unauthorized falsification, misrepresentation, or invention of any data, or citation in an academic exercise

*Plagiarism* – intentionally representing the words, ideas, or data of another as one's own in any academic exercise without providing proper citation

*Unauthorized collaboration* – instances when students submit individual academic works that are substantially similar to one another; while several students may have the same source material, the analysis, interpretation, and reporting of the data must be each individual's independent work.

*Participation in academically dishonest activities* – any action taken by a student with the intent of

gaining an unfair advantage

*Facilitating academic dishonesty* – intentionally or knowingly helping or attempting to violate any provision of this policy

For more information on Academic Integrity, including examples, please refer to the [Office of Student Conduct and Conflict Resolution web page](#).

## **College of Professional Studies Policies and Procedures**

For comprehensive information, please see the [Registrar University Catalogs page](#) as well as the [Student Resources](#) page of the [Northeastern University College of Professional Studies](#) website.

## **Student Accommodations**

Northeastern University and the Disability Resource Center (DRC) are committed to providing disability services that enable students who qualify under Section 504 of the Rehabilitation Act and the Americans with Disabilities Act Amendments Act (ADAAA) to participate fully in the activities of the university. To receive accommodations through the DRC, students must provide appropriate documentation that demonstrates a current substantially limiting disability.

For more information, visit [the Disability Resource Center Getting Started page](#).

## **End-of-Course Evaluation Surveys**

Your feedback regarding your educational experience in this class is very important to the College of Professional Studies. Your comments will make a difference in the future planning and presentation of our curriculum.

At the end of this class, please take the time to complete the evaluation survey at [the NEU EvaluationKit website](#). Your survey responses are **completely confidential**. Surveys will be open for the last two weeks of the class. An email will be sent to your HuskyMail account notifying you when surveys are available.

## **Online Proctoring**

In this class, some tests may be administered remotely by an online authentication and proctoring service called Examity®, which gives you the flexibility to schedule exams at your convenience and take them wherever you want.

To prepare for using Examity®, you will need to meet the following technical requirements:

- Working webcam and microphone which can be tested at [www.testmycam.net](http://www.testmycam.net)
- An Internet connection of at least 3Mbps ([www.speedtest.net](http://www.speedtest.net))
- Chrome/ Mozilla/ Safari/ Internet Explorer/ Microsoft Edge browser
- Up to date Operating system (Windows or Mac OS)

Please click on the link below to run an automated systems check:

[Examity Computer Readiness Check](#)

If you do not pass the systems check or have any questions or concerns, you can contact Examity's® technical support team 24/7 via email at [support@examity.com](mailto:support@examity.com) or phone at (855) 392-6489. Please tell your instructor immediately if your computer/equipment does not meet the standard to use online proctoring.

## **Northeastern University Online Copyright Statement**

Northeastern University Online is a registered trademark of Northeastern University.

All other brand and product names are trademarks or registered trademarks of their respective companies.

This course material is copyrighted and Northeastern University Online reserves all rights. No part of this publication may be reproduced, transmitted, transcribed, stored in a retrieval system, or translated into any language or computer language, in any form or by any means, electronic, mechanical, magnetic, optical, chemical, manual, or otherwise, without the express prior written permission of Northeastern University Online.

Copyright 2017 © by Northeastern University Online  
All Rights Reserved