|  |  |
| --- | --- |
| BIA6303 | Dr. Myles P. Gartland |
| **COURSE NO AND SECTION** | **PROFESSOR** |
|  |  |
| Predictive Models | (816) 501-4563 office |
| **COURSE TITLE** | (816) 501-4650 fax |
| Fall 2016 | Myles.gartland@rockhurst.edu |
|  | **CONTACTS** |
|  |  |
| Office- Conway 201 (I am in most days after 10am and welcome to see anyone, but to make sure try email first!) | |

Required texts: **1) Practical Machine Learning**

**2) Python Machine Learning**

**3) Introduction to Machine Learning**

**Recommended Text (for more depth)**

**Introduction to Statistical Learning by James (good for R)**

Wiki Book- <https://en.wikipedia.org/wiki/Book:Machine_Learning_-_The_Complete_Guide>

**Course Description** The course will teach advanced statistical techniques to discover information and build predictive models from large sets of data. Emphasis is place on applications for marketing research and operations. Methods will include multiple and logistic regression, propensity models principal component analysis, market basket analysis, longitudinal data analysis and product launch models. Statistics and exposure to any programming language is required. The primary software tool for this class will be R. Python and Spark will also be incorporated. Prerequisites BIA6301, BIA6311, BIA6312 or consent of the Program Director.

This class goes into more depth of some of the areas introduced in the ADM class. It looks at different forms of models. It also allows students to explore different options.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. **Web Page:** A web page has been set up for this class on Blackboard at <http://courses.rockhurst.edu> . If you have any difficulty accessing this web site please email me at [myles.gartland@rockhurst.edu](mailto:myles.gartland@rockhurst.edu). Data sets, slides and other materials will be posted on this site.

2. **Attendance:** For a technical course, attendance is of great importance. Therefore, students who are absent more than (1) time, may have their final course grade reduced by at least one full letter. If a student needs to miss more than twice, it is recommended they enroll in this class at a later time.

3. **Lecture Material:** In any given class, I would expect each student to complete and have a grasp of the content from the textbook chapters before the class. In each class, I will only highlight some of the common themes or more difficult aspects of each chapter. However, I will not go over all the material you will be responsible for. During class I will also randomly call on students and ask them to give a brief explanation or discussion of a topic from the assigned chapter. If a student cannot do this (because they are unprepared), I reserve the right to lower some component of their final grade.

4. **Objectives**

1. To be able recognize business problems that can be solved with data mining applications.
2. To be able to effectively communicate results to a non-technical audience.
3. To advance your knowledge of predictive and classification models introduced in Applied Data Mining
4. To be able to prepare and evaluate basic and advanced predictive and classification models.
5. To allow the student to explore different models and parameters using various data.
6. To be effective using the Python AND R platforms for data mining and other data science applications.
7. To be able to distinguish ways of implementation of data mining models into production.

**5. Technology**

In each class session, students will need to bring a laptop capable of downloading and running open-source software and their associated packages. The laptop also needs the capability of accessing the university’s wireless internet. It is also important that the person has administrative privileges for their laptop in order to load and modify programs and packages easily. The laptop may be a Windows, Mac or Linux based system. It is highly recommended that the laptop be running a more recent operating system (i.e. Windows 7 and above or Mac 10.8 or above) and have at least 4GB of RAM. Higher amounts of RAM (8 GB or more) is also recommended.

Further, this class will do a couple of exercises using AWS and/or Azure. Cost for using this platform will be at the student’s expense.

**6. Assignment**

Assignments will be given to practice the concepts and techniques we learned in class. While the course will be taught using Python and R, you are welcome to use any program you wish to complete the assignment (unless otherwise specified). While I encourage you to communicate with each other, the assignments are individual. Meaning I should not see identical outputs or copies of code between students.

Assignments are due at 5:45p the night of class (so 5:45p Wed). 10% will be reduced for each day or partial day late (so penalized at 5:46p Wed).

The weekly “tune-up” is a small set of questions posted on BB that makes sure you have an understanding of the concepts of each model. Most questions start with “in your own words, explain…..”.

**7. Project**

Each student will create a data-mining project of their own choosing. Data Sets from your company are encouraged. However, the data set will have certain requirements. If a student does not have access to his or her own company data project, I will happily provide you one. I also encourage you to look at Kaggle competitions for good projects. More information on the requirement for this will be discussed in class. This project is expected to be of much higher complexity than the project in BUS6301.

**Course Grade:**

Weekly tune-up 10%

Assignments 65%

Project 25%

**Tentative Schedule**

Schedule can change slightly due to time it takes to get through a topic or other interests of the students.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **Topic(s)** | **Gollapudi** | **Raschka** | **Alpydin** | **Assignment (TENATIVE)** |
| 8/24 | Regularization and GLM for Data Mining | Ch 10 | Ch 4, 10 | Ch 1, 4.6 | Load Python and IPython Notebook. Use 2.7x distribution of Anaconda. Review Ch 3-6, 8 of Data Analysis with Python Book. |
| 8/31 | -Bayesian Models (and review classification from AMD) | Ch 9 | Pg 80-96 | Ch 3,16 |  |
| 9/7 | -Kernel methods/SVM | Ch 6 | Ch 3,6 | Ch 13,19 | Assignment 1 Due |
| 9/14 | -Ensembles/Gradient Descent | Ch 13 | Ch 7 | Ch 17 |  |
| 9/21 | -Ensembles/Gradient Descent pt 2 | Ch 4 | Ch 7 | Sec 10.6 |  |
| 9/28 | -Neural Nets and Deep Learning (intro) | Ch 11 | Ch 2, 12-13 | Ch 11 | Assignment 2 Due |
| 10/5 | -Recommender Systems | Ch 12 | Ch 8,9 | Ch 18 |  |
| 10/11 | -Machine Learning with Spark | Ch 3, 14 | Ch 8 |  | Assignment 3 Due (10/11, latest) |
| 10/16, 11:59pm |  |  |  |  | Latest I will accept the project paper/project without penalty |

**Scale**

93-100 A

87-92 B+

80-86 B

70-79 C

0-69 F

### Data Mining Websites

<http://www.kdnuggets.com/>

[www.r-bloggers.com/‎](http://www.r-bloggers.com/)

**Rigor:** As you can note: obtaining an “A”, will require superior performance, well above the quality for the other students. All “A” work should be of excellent quality, giving concise, insightful analysis. “B” work will consist of covering the major points quite well. “C” work consist of adequate coverage, but nothing insightful or inspirational.

**Rockhurst University and Helzberg School Policies** **and Statements**

**Online Wall Street Journal Subscription**

All Helzberg School of Management students are given an online subscription to the Wall Street Journal.  Using Google Chrome students can sign-up for the online journal at <https://accounts.rockhurst.edu/wsj/> (use your Rockhurst username and password).  Access to your online journal is at [www.wsj.com](http://www.wsj.com).  Should you have questions or problems contact the Helzberg School front desk (C201) at x4090 or email [lynn.ross@rockhurst.edu](mailto:lynn.ross@rockhurst.edu).  The customer service number for the online journal is 800-369-2834 or [onlinejournal@wsj.com](mailto:onlinejournal@wsj.com). Student guides can be found in C201.

**Academic Honesty Policy:**

Plagiarism and cheating will not be tolerated. The Rockhurst University Catalog provides examples of academic dishonesty and outlines the procedures, penalties, and due process accorded students involved in academic dishonesty. All infractions will be immediately referred to the Dean's office.  In your research paper, make sure you provide citations for all ideas and information that are not your own. This includes *copying, or the offering, requesting, receiving or using of unauthorized assistance or information in examinations, texts, reports, computer programs, term papers or other assignments*.

Undergraduate Policy:

<http://catalog.rockhurst.edu/content.php?catoid=8&navoid=380#Academic_Honesty_Policy>

Graduate Policy:

<http://catalog.rockhurst.edu/content.php?catoid=9&navoid=452#Academic_Honesty_Policy>

**Helzberg School Student Complaint Process**

Students who feel that they have a nontrivial complaint, either academic or nonacademic, should contact either Paul Nunez, Director of Undergraduate Advising, or Matt Honeycutt, Director of Graduate Advising, for guidance on submission.

**Course Assessments:**

Rockhurst University and The Helzberg School are committed to a comprehensive, ongoing assessment process. On occasion students will be expected to participate in aspects of this program so that the institution can document its strengths and identify opportunities for improvement. Student involvement is integral to the University’s success in this endeavor.

**ADA Statement**

RU official policy: “Rockhurst University is committed to providing reasonable accommodations for students with disabilities. Please contact Sandy Waddell, Director, Access Office (Massman Hall Room 7, (816) 501-4689, [sandy.waddell@rockhurst.edu](mailto:sandy.waddell@rockhurst.edu)) to provide documentation and request accommodations. If accommodations have already been approved by the Access Office, please communicate with the instructor of this course regarding these arrangements by the second week of class in order to coordinate receipt of services.”

**Student Resources**

In keeping with our commitment to *cura personalis*, “care of the whole person,” Rockhurst University provides a variety of programs and services beyond those of the Helzberg School of Management that are designed to support you as you pursue your education and navigate challenges, both expected and unexpected.  These include, but are not limited to: the Dean of Students Office, Dining Services, Bookstore, Counseling Center, Career Services, Access/Disability Services, Campus Ministry, Student Health Insurance and Clinic, Student Organizations & Greek Affairs (alumnus advisors), Athletics/Intramurals/Workout spaces, International and Multicultural Student Services, Security, Financial Aid, Student Accounts, Computer Services, The Registrar, The Library, Service Learning, and the Learning Center.  For information on any of these, please visit the University’s website at [www.rockhurst.edu](http://www.rockhurst.edu) – or call/visit Student Development at 816-501-4030 or Massman Hall, Room 1.

**Crisis Management**

The Crisis Management Team for Rockhurst University, out of a concern for the safety and welfare of all community members, urges you to familiarize yourself with Campus Emergency Procedures as well as emergency, evacuation, and shelter signage located within and outside of buildings across campus.  They also request that you sign up for Rock@lerts, the University’s emergency communications system.  All community members must comply with University staff instructions during regularly scheduled fire/tornado drills as well as actual emergencies.  Information regarding the above resources can be found at [www.rockhurst.edu/emergency](http://www.rockhurst.edu/emergency).

**University Communication with Students**

Clear and timely communication allows students to receive information related to policies, programs, events, and other practical matters (i.e. billing, grades, etc.) affiliated with their education at Rockhurst. The University has established several key routes for communication with students including:

* Rockhurst E-mail Account: All Rockhurst students receive a University e-mail account. All e-mail communication from the University is directed exclusively to the Rockhurst electronic mailbox system. Students are expected to access their e-mail account on a regular basis (daily is recommended) in order to stay abreast of important and time-sensitive information. University departments, faculty, and staff will routinely use e-mail to communicate important campus, academic, and extra-curricular/co-curricular information. It is the responsibility of each student to clean their e-mail boxes to allow capacity for incoming messages (i.e. empty deleted items, keep a limited number of sent items, etc.). For further information on your Rockhurst e-mail account, please see Computer Services (Conway Hall 413; x4357; [*www.rockhurst.edu*](http://www.rockhurst.edu/)).
* Addresses and Phone Numbers: Students are required to maintain accurate local, billing, permanent, and emergency contact information so that attempts to communicate by the U.S. Postal System as well as phone will not be impeded. To make changes to your addresses or phone numbers of record, please visit the Rockhurst website ([*www.rockhurst.edu*](http://www.rockhurst.edu/); Rockweb section under Registrar) or contact the University Registrar (Massman Hall 110).
* Campus Mailboxes: Undergrad Students who live in Rockhurst’s residence halls and Townhouse Village are assigned a campus mailbox located in Massman Hall where U.S. Postal Service mail and other University communications are delivered. For assignment inquiries please connect with Residence Life (Massman Hall 3) and for mailbox malfunctions please speak with the Mailroom staff (Massman Hall, Lower Level).

***Updated July 2016***