**STA 536 – Data Mining**

**Spring 20XX**

**Professor**: Andrew Crossett, PhD **Phone**: 610-436-2613

**Email**: acrossett@wcupa.edu **Office**: 25 University Ave. 111

**Office Hours**: Tuesday: 3:30-5:30

Wednesday: 1:00-3:00

Thursday: 3:00-4:00

**Prerequisites:** C or better in STA 512

**Required Materials**: *An Introduction to Statistical Learning with Applications in R:* James, Witten, Hastie, and Tibshirani

**Course Description:** The purpose of this course is to give you an introduction to many of the data mining techniques that are used to analyze a wide array of datasets. These include principal components analysis, clustering, classification, and ensemble approaches. We will be applying these methods using the statistical programming language R.

**Program Learning Outcomes:**

1) Demonstrate an understanding of probability and statistical inference, including the fundamental laws of classical probability, discrete and continuous random variables, expectation theory, maximum likelihood methods, hypothesis testing, power, and bivariate and multivariate distribution theory.

2) Demonstrate the ability to apply the elementary methods of statistical analysis, namely

those based on classical linear models, categorical methods, and non-parametric ideas to perform data analysis for the purposes of statistical inference.

3) Demonstrate proficiency in the effective use of computers for research data management and for analysis of data with standard statistical software packages, particularly SAS.

4) Learn to develop and critically assess design of experimental studies and the collection

of data.

5) Apply one or more methods of statistical inference to a particular area of interest,

particularly the program in the elective concentration.

6) Gain practical experience in statistical consulting and communicating with non-

statisticians, culminating with interaction with research workers at a local company as part

of the internship practicum.

**Student Learning Outcomes:**

1) To become familiar with the statistical programming language R [PLO 3]

2) To be introduced to many data mining techniques [PLOs 1,2,4,5]

3) To develop the tools necessary to analyze a wide variety of datasets [PLO 6]

**Attendance Policy:** I will not be taking attendance. However, I highly recommend that you attend class. Even though I will be posting some notes on the website, there will be a lot of supplemental material not in the notes. If you miss a class, your best bet to learn the material is to come to my office hours, meet with someone from class or, in extreme cases, contact me to meet at another time.

**Tentative Course Outline:**

**Week 1:** Introduction to R

**Weeks 2 and 3:** Visualizing Data

**Weeks 4 and 5:** Principal Components Analysis

**Week 7:** Exam

**Weeks 8 and 9:** Clustering

**Weeks 10 and 11:** Classification

**Weeks 12 and 13:** Regression Techniques

**Week 14:** Ensemble Approaches

**Evaluation & Grading:**

* You will have weekly homework assignments. These will be a combination of theory and application. The assignments are to be done using R.
* You will have one in-class exam. It will be administered on WEEK SEVEN
* Project: You will also be working on a project throughout the semester.
  + You will be responsible for finding a dataset (from Kaggle.com), coming up with an interesting research question, and then effectively carrying out the proposed analysis using the techniques we learned in class.
  + A 1-2 page proposal will be due by Tuesday, March 6th.
  + You will turn in a detailed report (4-8 pages) and then give a 15-20 minute presentation highlighting the report.
  + Presentations will be scheduled during finals week.
* Course Grade:
  + **Homework: 30% (SLO 1,2,3)**
  + **Exam: 30% (SLO 2,3)**
  + **Final Project: 40% (SLO 1,2,3)**

A letter grade will be assigned based on performance in the course, according to the following scale:

|  |  |  |  |
| --- | --- | --- | --- |
| **Grade** | **Quality Points** | **Percentage Equivalents** | **Interpretation** |
| A | 4.00 |  | Superior graduate attainment |
| A- | 3.67 |  |  |
| B+ | 3.33 |  | Satisfactory graduate attainment |
| B | 3.00 |  |  |
| B- | 2.67 |  |  |
| C+ | 2.33 |  | Attainment below graduate expectations |
| C | 2.00 |  |  |
| C- | 1.67 |  |  |
| F | 0 | < 70% | Failure |

D grades are not used. Refer to the Graduate Catalog for description of NG (No Grade), W, & other grades.

**Statements Common to All WCU Graduate Syllabi:**



**ACADEMIC & PERSONAL INTEGRITY**

It is the responsibility of each student to adhere to the university’s standards for academic integrity. Violations of academic integrity include any act that violates the rights of another student in academic work, that involves misrepresentation of your own work, or that disrupts the instruction of the course. Other violations include (but are not limited to): cheating on assignments or examinations; plagiarizing, which means copying any part of another’s work and/or using ideas of another and presenting them as one’s own without giving proper credit to the source; selling, purchasing, or exchanging of term papers; falsifying of information; and using your own work from one class to fulfill the assignment for another class without significant modification. Proof of academic misconduct can result in the automatic failure and removal from this course. For questions regarding Academic Integrity, the No-Grade Policy, Sexual Harassment, or the Student Code of Conduct, students are encouraged to refer to the Department Graduate Handbook, the Graduate Catalog, the *Ram’s Eye View*, and the University website at www.wcupa.edu.

**STUDENTS WITH DISABILITIES**

If you have a disability that requires accommodations under the Americans with Disabilities Act (ADA), please present your letter of accommodations and meet with me as soon as possible so that I can support your success in an informed manner. Accommodations cannot be granted retroactively. If you would like to know more about West Chester University’s Services for Students with Disabilities (OSSD), please visit them at 223 Lawrence Center. The OSSD hours of Operation are Monday – Friday, 8:30 a.m. – 4:30 p.m. Their phone number is 610-436-2564, their fax number is 610-436-2600, their email address is ossd@wcupa.edu, and their website is at www.wcupa.edu/ussss/ossd.

**REPORTING INCIDENTS OF SEXUAL VIOLENCE**

West Chester University and its faculty are committed to assuring a safe and productive educational environment for all students. In order to meet this commitment and to comply with Title IX of the Education Amendments of 1972 and guidance from the Office for Civil Rights, the University requires faculty members to report incidents of sexual violence shared by students to the University's Title IX Coordinator, Ms. Lynn Klingensmith. The only exceptions to the faculty member's reporting obligation are when incidents of sexual violence are communicated by a student during a classroom discussion, in a writing assignment for a class, or as part of a University-approved research project. Faculty members are obligated to report sexual violence or any other abuse of a student who was, or is, a child (a person under 18 years of age) when the abuse allegedly occurred to the person designated in the University protection of minors policy.  Information regarding the reporting of sexual violence and the resources that are available to victims of sexual violence is set forth at the webpage for the Office of Social Equity at <http://www.wcupa.edu/_admin/social.equity/>.

**EXCUSED ABSENCES POLICY**

Students are advised to carefully read and comply with the excused absences policy, including absences for university-sanctioned events, contained in the WCU Graduate Catalog. In particular, please note that the “responsibility for meeting academic requirements rests with the student,” that this policy does not excuse students from completing required academic work, and that professors can require a “fair alternative” to attendance on those days that students must be absent from class in order to participate in a University-Sanctioned Event.

**EMERGENCY PREPAREDNESS**

All students are encouraged to sign up for the University’s free WCU ALERT service, which delivers official WCU emergency text messages directly to your cell phone. For more information, visit www.wcupa.edu/wcualert. To report an emergency, call the Department of Public Safety at 610-436-3311.

**ELECTRONIC MAIL POLICY**

It is expected that faculty, staff, and students activate and maintain regular access to University provided e-mail accounts. Official university communications, including those from your instructor, will be sent through your university e-mail account. You are responsible for accessing that mail to be sure to obtain official University communications. Failure to access will not exempt individuals from the responsibilities associated with this course.