



Department of Computer Science and Engineering

Faculty of Engineering, University of Moratuwa

CS5122 Descriptive and Predictive Analytics

Semester 4, MBA in IT 2017 (Jan. - Mar. 2018)

Class LMS	https://online.mrt.ac.lk/course/view.php?id=8705
Schedule/ Hours	Sunday 1:00pm - 5:00pm at UOM (9 sessions) About 2-3 average hours/week self-study and on the Moodle expected
Instructor(s)	Dr. Dilum Bandara, dilumb@cse.mrt.ac.lk , 011 265 0152
Prerequisite(s)	None CS5618 Business Intelligence and MN5202 Analysis for Managerial Decision Making would be useful
Text(s)	Recommended Textbooks: <ul style="list-style-type: none">• Business Analytics – Methods, Models, and Decisions by J. Evans, Pearson Higher Education• Introduction to Applied Multivariate Analysis with R by B. Everitt and T. Hothorn, Springer• Time Series Analysis and Its Applications by R. H. Shumway and D. S. Stoffer, Springer• Related industry articles and research papers
Assessment	Distribution of marks are as follows: <ul style="list-style-type: none">• Homework (3) 30% (10 x 3)• Analytics Challenge (3) 30% (5 + 10 + 15)• Final Exam 40%
Course objectives	To provide a broader understanding of descriptive and predictive data analysis techniques and their applications. After completing this module, students will be able to: <ul style="list-style-type: none">• Interpret summarized data and graphs to understand the implications of data that they visualize• Explain the value of descriptive and predictive data analytics to strategic decision making in a modern business organization• Apply descriptive and predictive data analytics tools to analyze given datasets to explore emerging patterns in data and forecast future trends• Select and use appropriate statistical and machine learning approaches and data visualization techniques to support business decisions Required readings, case studies, homework, analytics challenges, and discussions are expected to enhance both the analytical and soft skills.
Syllabus	The goal for the class is to be broad rather than deep. Following is a tentative list of topics that might be covered in the class. We will select material adaptively based on the background, interests, and progress of the students. <ol style="list-style-type: none">1. Introduction to statistics and probability [1 class]<ul style="list-style-type: none">• Random variables, probability distributions, sampling2. Multi-variate data analysis [1 class]<ul style="list-style-type: none">• Correlation• Dimension reduction

3. Clustering [1 class]
 - K-mean, top-down, bottom-up
4. Regression analysis [1 class]
 - Linear and non-linear
5. Time series analysis [2 classes]
 - ARIMA model
 - Spectral analysis
6. Machine learning applications [2 classes]
 - k-NN, SVM, Random Forest, collaborative filtering
7. Big data analysis and visualization techniques [1 class]

Class policies

- Bring your own laptop for each class. Make sure necessary software are installed (Microsoft Excel or alternatives and R).
- Topics to be discussed in each class will be posted on Moodle, along with relevant readings for each topic. You are expected to keep up with the readings as we go, as they will help provide the foundation for the homework, project, and exam.
- All students are expected to actively participate in class, Moodle, and Yammer activities. Poor participation and/or poor performance in assigned course work can be grounds for failure in the course.
- Discussing and exchanging ideas through study groups are encouraged, as this usually leads to a better understanding. As part of the discussions, you may share ideas and thoughts, discuss the meaning of homework questions, or possible ways of approaching a solution. However, you must write homework solutions strictly. If one of your solutions is based on a key idea of someone else, you must acknowledge this in your homework, to avoid the perception of cheating. This form of collaboration is not an opportunity to copy answers from others.
- Group assignments are given to encourage teamwork and discussion/toleration of alternative ideas/views; hence, they need to be done as a group. A penalty will be enforced for doing group assignments individually.
- Plagiarism, copying another person's work, letting another person copy your work, giving or receiving aid during any test or examination is all strictly not allowed. Any student caught in any of these will receive a failing grade regardless of marks earned on other assessed work.
- Proper netiquette should be observed in using the Moodle and Yammer.
- Each assigned work will have either a deadline for submission or a specific date for performance. For each day delayed beyond a deadline, 10% of marks will be deducted. Not performing (e.g., not doing a presentation) on an assigned date will result in 0 marks unless there is a valid reason and another student/group is arranged as a replacement. Details of submission will be given with each assignment. All assignments must be submitted via the Moodle.
- The dictionary meaning of deadline is "the latest time or date by which something should be completed". Thus, as you may already experience during your internship, deadlines are supposed to be met.
- You may not use cell phones, mp3 players, etc., during the class. All laptops, smart phones, and tablets must be closed, unless you use it to take notes or search for additional contents relevant to the ongoing class discussion. The reason is to prevent distractions to other students, and to prevent the temptation to check email, Facebook, Twitter, etc.