UNIVERSITY of MASSACHUSETTS DARTMOUTH
Charlton College of Business
Decision and Information Sciences

COURSE: **Business Analytics & Data Mining, POM-681**
Prerequisite: POM 500 *(or equiv.)*

PROFESSOR: Dr. Bharatendra Rai
Office: Charlton Building, Room 326
Phone: (508) 910-6434
Email: brai@umassd.edu

1. COURSE DESCRIPTION

Introduction to business analytics and data mining. Topics covered include data mining, data visualization, methods for classification and prediction, decision trees, logistic regression, association rules, and text analytics. Applications of business analytics and data mining methodologies to a wide variety of real world business data are included. The course also makes use of leading software for analyzing big data.

2. COURSE OBJECTIVES

The emphasis of this course is on applications of business analytics and data mining, rather than on mastering the theoretical underpinnings of the techniques.
From this course you will:

- Be able to appreciate the enhanced data rich environment of today's global economy and get exposed to the related business intelligence service opportunities that exist
- Be able to provide a practical understanding of the key methods of classification, prediction, reduction and exploration that are at the heart of business analytics and data mining
- Be able to decide and use appropriate tools and techniques in different situations
- Be introduced to leading business analytics and data mining software
- Be able to gain the intellectual capital required to provide business analytics and data mining services
- Be able to write business reports and present them.

3. COMPETENCIES AND CONTACT HOURS

Competencies		Contact Hours
The student will be introduced to statistical thinking, definitions, techniques of organizing and describing information, estimating and testing. The student will understand:		
• **What is business analytics and data mining?**	3	3
• **Exploratory data analysis**	3	6
• **Classification and prediction**	3	9
• **Simple classification schemes**	3	12
• **Classification and regression Tree (CART)**	3	15
• **Logistic regression**	6	21
• **Cluster analysis**	3	24
• **Twitter data analysis**	6	30
• **Projects**	6	36

4. COURSE MATERIAL

Reading material: "R and Data Mining: Examples and Case Studies" available on the course website.

Software: **R.** R is a free software environment for statistical computing and graphics, and is widely used by both academia and industry. The advantage of the R software is that it can work on both Windows and Mac-OS. It is ranked no. 1 in the KDnuggets 2013 poll on top languages for analytics, data mining, and data science. RStudio is a user friendly environment for R that has become popular.

Steps for downloading & installing R/RStudio:
https://www.youtube.com/watch?v=PHi-6GWfMBQ&list=PL34t5iLfZddv8tJkZboegN6tmyh2-zr_T&index=1

Course website: Class website will be used for all the material related to this course.

5. UNIVERSITY CLASSROOM POLICIES

Academic Honesty: Students are expected to participate in the course within the guidelines of the Academic Ethical Standards published in the General Catalogue. Instances of academic dishonesty will be penalized to the greatest extent possible. Plagiarism is a serious offense.

To help you maintain academic integrity, the Carney Library offers many resources including http://www.lib.umassd.edu/get/refworks.html, the link to Refworks, a bibliography and database manager, as well as guides on using MLA, APA, and other citation standards. Before you write any paper, you should review the guides on Avoiding Plagiarism: http://www.lib.umassd.edu/find/plagiarism.pdf or http:/www.lib.umassd.edu/find/plagiarism.pdf.

Students with Disabilities: Disabled Student Services (DSS) provides support to both learning and physically disabled students. If you have a disability that requires accommodation you should contact DSS. If you have a documented disability and require accommodations to obtain equal access in this course, please meet with the instructor at the beginning of the semester and provide the appropriate paperwork from the Center for Access and Success. The necessary paperwork is obtained when you bring proper documentation to the Center for Access and Success, which is located in Group I (Liberal Arts Building), Room 016, and phone: 508-999-8711.

Incomplete: According to the university catalogue, an incomplete may be given only in exceptional circumstances, at the instructor's discretion. The student must be passing at the time of the request, or sufficiently close to passing. If the work is not completed within one year of the recording of the incomplete grade, the grade will become an F (I). The incomplete policy for this course is that at least 80% of the course must be complete and an exceptional circumstance (for example, a medical issue) must exist. If you feel you require an incomplete for an exceptional reason, you need to email me and state your reasons for the incomplete in writing. I will then decide whether to consider granting the incomplete. If I agree to consider it, we will then have to meet to work out a specific course of action.

Link to all the student support services for UMassD:
http://www.umassd.edu/extension/studentresources/

6. COURSE POLICIES

Communication: When communicating via email (my email ID: brai@umassd.edu), make sure to include POM 681 in the subject area. Response time is expected to be within 24 hours.

Late assignments: This being an online class, it is extremely critical to complete you work as per the schedule. To encourage on-time submissions and to ensure you don't fall behind too much, assignments that are submitted after the due date will lose 50% points.

Proper grammatical writing: Make sure to use proper grammar in assignments that are submitted. Use of writing language that are common in instant text messaging will cause deduction of points.

Feedback: Assignments submitted on-time will be graded before the due date for next assignment and appropriate feedback will be provided.

Changes: The class schedule towards the end of this syllabus is tentative, and is subject to change at the instructor's discretion.

7. EVALUATION POLICY

Following is the break-up of the class grades:

Evaluation Type	Score
Assignments (8)	60%
Online quiz	10%
Project-1	15%
Project-2	15%
TOTAL	100%

Final grades would be determined based on all items discussed above with weightings as indicated. The course contents are designed to help you to be successful in your current or future profession. And therefore to pass this course, a student must develop and demonstrate basic understanding of the concepts, and good comfort level in interpretation and application. Given below are the grades based on the final course score:

Grade	*Final Score (%)*
A-, A, A+	90-100
B-, B, B+	80-89
C-, C, C+	70-79
D-, D, D+	60-69
F	00-59

Assignments:

Students will complete eight assignments during the course. Assignment related documents, instructions, and other details will be available on the course website.

Project-1: Details will be available on the course website.

Project-2: Details will be available on the course website.

8. CLASS SCHEDULE*

Topics and Readings	Upload to course website
Part-1: Articles from Flipboard - Introductory lecture video; Reading 'competing on analytics article by Davenport	Assignment-1 involving PowerPoint with narration using Flipboard articles.
Part-2: Data Visualization/Dashboadrs - Dashboard with Google Data Studio	Assignment-2 Dashboard with airline data.
Part-3: Installing R and RStudio **Getting started with R** - Exploring data, numerical summary and graphical summary using dim, names, str, attributes, head, tail, summary, quantile, sd, hist, plot, and pairs commands. Linear regression, and data partitioning	Assignment-3 involves exploring data and modeling with R & RStudio + QUIZ
Part-4: Visualization - Bar plot, Histogram, Box Plot, Scatter plot, 3D Scatter Plot, Multiple Plots, Pie Chart, Contour Plots, 3D Surface Plots	Assignment-4 involves exploratory visualization using 'US Airlines On-Time Performance' data + QUIZ
Project-1: Details will be available on the course website.	Project-1
Part-5: Text Mining - Text Mining basics, Text mining from pdf files, Word clouds in R, getting tweets from Twitter using R	Assignment-5 involves getting tweets from Twitter and carrying out text mining steps.
Part-6: Decision Trees & Logistic Regression - Supervised Segmentation, Theory: Decision trees and concepts of Logistic Regression (simple/ multinomial logistic)	Assignment-6 involves application of binary logistic regression and decision trees + QUIZ

Topics and Readings	Upload to course website
Part-7: Classification Models and Performance - Classification models, multinomial logistic regression, decision trees, Training & Validation, Confusion Matrix to assess model performance	Assignment-7 involves application of multinomial logistic regression and decision trees to medical data + QUIZ
Part-8: Cluster Analysis - K-means clustering, Hands-on examples	Assignment-8 involves application of cluster analysis for pharmaceutical industry
Project-2: Details will be available on the course website.	Project-2

*The class schedule is tentative, and is subject to change at the instructor's discretion.