

# Missouri University of Science & Technology

## Information Science and Technology 5420

BUSINESS ANALYTICS AND DATA SCIENCE  
Spring Semester 2016 – Syllabus

### Class Information

Class: Monday, Wednesday, and Friday 1:00-1:50  
Class Room: Butler-Carlton Hall (BCH) 213

### Contact Information

Instructor: Prof. Michael G. Hilgers Ph.D. CQF  
Email: hilgers@mst.edu  
Phone: 341-6484(office)  
Office: Fulton Hall 106D  
Office Hours: Monday, Wednesday & Friday (9:30-11:00) (Or by appointment: hilgers@mst.edu)

### Course Information:

---

**"By 2018, the United States alone will face a shortage of 140,000 to 190,000 people with deep analytical skills, as well as a shortfall of 1.5million data-savvy managers with the know how to analyze big data to make effective decisions."**

### Catalog Description

---

This course addresses the foundations of using statistics on "big data"<sup>1</sup> sets to impact business decision-making. Focus is applied examples using realistic data. Models implemented include regression (parametric/nonparametric), classification, decision trees, and clustering with analytical estimation accomplished using popular software.

### Extended Description

---

Data analytics facilitates realization of objectives by identifying trends, creating predictive models for forecasting, and optimizing business processes for enhanced performance. Three main categories of analytics are:

1. **Descriptive**—the use of data to find out what happened in the past;
2. **Predictive**—the use of data to find out what could happen in the future; and
3. **Prescriptive**—the use of data to prescribe the best course of action for the future

---

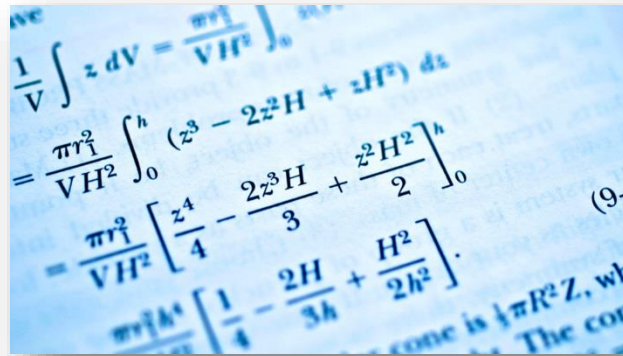
<sup>1</sup> "Big Data" is becoming a tired phrase. It simply means that the data being analyzed is too big for in-memory calculations or even residing on a single computer. This will be discussed extensively.

## Course Prerequisites:

**Calculus** – Ordinary and partial differentiation, optimization, norms, and some analytical geometry

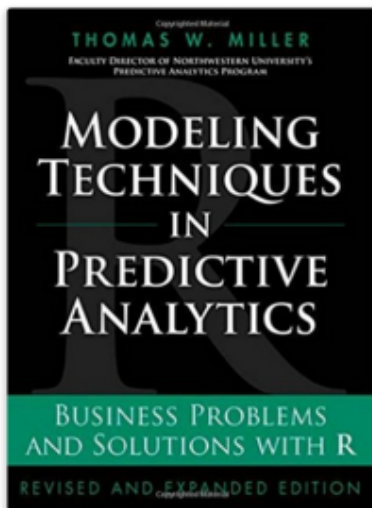
**Statistics** – Normal probability distributions, expected value, conditional expected value, mean, standard deviation, statistical inference

**Matrix Algebra** – Though not a required prerequisite, matrix notation and operations simplify many of the models. A basic introduction will be given.



**Foundations of Mathematical Models** -- The course will begin with an overview of mathematical models and will include topics relevant to data analytics. References to fundamental mathematical topics will be given as well as some homework problems to reinforce the skills needed to succeed in the course. So you will have a chance to come up to speed.

## Textbooks:



### By Thomas W. Miller Modeling Techniques in Predictive Analytics: Business Problems and Solutions with R, Revised and Exp (1st First Edition)

**[Hardcover]** Hardcover – October 26, 2014

[Be the first to review this item](#)

[See all formats and editions](#)

Hardcover  
from \$75.04

33 Used from \$75.04

26 New from \$97.51

It is very important to have this book and the R code that comes with it. Much of the semester will be spent unraveling the R solutions he provides. NOTE! There is another version of this book. You want the one with “Business Problems and solutions with R”) in the title.

## Software:

### R skills attract the highest salaries

Two recent salary surveys have shown that [R language](#) skills attract median salaries in excess of \$110,000 in the United States.

In the [2014 Dice Tech Salary Survey](#) of over 17,000 technology professionals, the highest-paid IT skill was R programming. While [big-data skills in general featured strongly](#) in the top tier, having R at the top of the list reflects the strong demand for skills to make sense of, and extract value from big data.

AVERAGE SALARY FOR High Paying Skills and Experience		
SKILL	2013	YR/YR CHANGE
R	\$ 115,531	n/a
NoSQL	\$ 114,796	1.6%
MapReduce	\$ 114,396	n/a
PMBok	\$ 112,382	1.3%
Cassandra	\$ 112,382	n/a
Omnigraffle	\$ 111,039	0.3%
Pig	\$ 109,561	n/a
SOA (Service Oriented Architecture)	\$ 108,997	-0.5%
Hadoop	\$ 108,669	-5.6%
Mongo DB	\$ 107,825	-0.4%

We will program in R. As can be seen from the comparison to the left, R is at the top of the pay scale. Be careful in the interpretation of this. Companies are hiring people who understand analytics and data science.

## Free sources of R



An excellent choice available at no cost is RStudio (<http://www.rstudio.com>). It is my preferred software ide.

## Course Elements

### Instructional Methods:

To achieve an understanding of the material, several techniques and methods will be used:

- I will lecture over foundational material. Typically, some form of notes will be provided, though I am not a strong supporter of PowerPoint in the classroom.
- Various sources of reading matter will be given including textbook, web, and personally developed notes. Please do the reading
- I like to work examples. Expect to spend some time studying data with software tools such as R.
- Analysis is not a spectator sport. Expect a homework problem or two most every night.

## Assignments:

---

Homework/programming assignments will be made frequently. Please observe the following:

- Work is to be done individually unless otherwise specified. If you submit the work of another person as your own, you will receive a zero for the assignment and your name given to the Vice Provost of Undergraduate/Graduate studies.
- Homework is to be completed on the specified date and time.
- Late homework will NOT be accepted and will result in a zero for that assignment unless prior arrangement is made.

## Course Content

---

*“Big Data is an emerging phenomenon. Computing systems today are generating 15 petabytes of new information every day— eight times more than the combined information in all the libraries in the U.S.; about 80% of the data generated every day is textual and unstructured data.”*

The following should give you an idea of how we will proceed in the course. It is not ridged in that I will slow down rather than lose everyone or speedup if I am boring everyone.

### Getting Started with Data Science

Basic data types, data representation, data types in R, data structures in R, basic descriptive, predictive, and prescriptive statistics with R applications, fitting a linear model, predicting with a linear model, visualization

### Advertising and Promotion Models

Numerous linear models with techniques for comparison; Non-trivial Visualizations

### Preference and Choice Models

Traditional Conjoint Analysis; Non-trivial Visualizations

### Market Basket Models and Analysis

Contingency Tables, Association Rules; Non-trivial Visualizations

### Text Analysis

Text Mining, Parsing, Corpus, Text Measures; Non-trivial Visualizations

### Sentiment Analysis

Text Mining, word strength, random graphs and forests

### Brand and Price Models

Hierarchical Bayes Models, Market Simulation

## Course Activity:

---

Given what is described about, the breakdown of activity is as follows:

Activity	Approximate Quantity	Course Percentage
Homework, R Programs, Projects	7 (One per section)	700 (100 pts each)
Midterm Examination or Project	1	150
Final Examination or Project	1	150

*R programs:* Practical applications of the concepts developed in class is extremely important. You will have a chance in this class to become fairly fluent in R by the end of the semester. An effort will be made to have all data sets to be on the order of a million records or more. Your programs will be done individually and strongly punished if copied.

*Test:* We will have two tests. The test will be drawn from lectures, examples, quizzes, and programs.

NOTE: It is very hard to build a course at this level. I may not give all 7 assignments and we must settle for 5 or 6. We will discuss and adjust the syllabus accordingly if this happens.

## Grading Breakdown:

---

Grades will be based on total points, as defined below. There may be bonus points from time to time, which would be added to whatever category the bonus applies to. Boundaries for grades may be adjusted downward slightly, if deemed needed.

**Grades:**

- A: 100% - 90%
- B: 89% - 80%
- C: 79% - 70%
- D: 69% - 60%
- F: Below 59%

## Learning Objectives

	Communication Skills	Critical Thinking	Information Technology	Teamwork and Leadership
Develop basic skills needed to understand and manipulate the mathematical models forming the foundations of data analytics		X		
Learn how to use R to visualize large multidimensional data sets and explain	X	X	X	
Identify proper mathematical model for a data set				
Perform Linear regression on a multi-dimension data set		X	X	
Be able to use nonparametric techniques		X	X	
Be able to use logistic models to analyze data sets		X	X	
Be able to use classification techniques to see patterns in data		X	X	
Be able to use clustering techniques to see patterns in data		X	X	
Be able to explain the role of business analytics in corporate environments	X			

## COURSE POLICIES

### Attendance:

Attendance is required, particularly as the assignments will be based on the important definitions and concepts presented in the lectures. You will likely want to ask questions. The class moves quickly and it is easy to fall behind and not get caught up. The more you miss class, the more material that will be foreign to you. If a student has missed an extended or excessive amount of classes or has failed to turn in multiple assignments, the instructor will send that student an Academic Alert. The alert will be emailed to the student and student's advisor. The student must meet with the instructor within three days or the instructor will send out another alert. If the student has not met with the instructor after the second alert, the instructor reserves the right to drop the student. If emergency circumstances arise, please contact the instructor soon to avoid penalties, and to try to catch up to the rest of the class.

## Academic Integrity Statement

---

(<http://registrar.mst.edu/academicregs/>):

Violations of the University's academic code include, but are not limited to, possession of or use of unauthorized materials during quizzes or tests; providing unauthorized information to another student; or copying the work of another person. Violations may result in academic penalties in addition to receiving an "F" on the assignment in question. (See page 30 of S&T's "Student Academic Regulations" handbook for further details about student standards of conduct relative to the system's Collected Rules and Regulations section 200.010.) The most common attempt at dishonesty is submitting the program of another person with only some changes to deceive me. These are easy to recognize and not be tolerated.

## Academic Alert System

---

(<http://academicalert.mst.edu/>):

S&T is committed to the success of its students by providing an environment conducive to teaching and learning. To ensure that every student takes full advantage of the educational opportunities and support programs on campus, the University has implemented an Academic Alert System, a web-based application. The purpose of the System is to improve the overall academic success of students by:

- Improving communication between students, instructors, and advisors;
- Reducing the time required for students to be informed of their academic status;
- Informing students of actions they need to perform in order to meet the academic requirements in the courses they are taking.

To assist you, I will initiate an academic alert for students who are not meeting academic course requirements through poor performance on assignments or poor attendance. When an alert is initiated, an email is immediately sent to the instructor, student, and advisor. You are encouraged to respond quickly to all academic alerts. If you fail to open the alert within one week, email notification is sent to your advisor.

## Disability Support Services

---

(<http://counsel.mst.edu/>):

If you have a documented disability and anticipate needing accommodations in this course, you are strongly encouraged to meet with me early in the semester. You will need to request that the Disability Services staff send a letter to me verifying your disability and specifying the accommodation you will need before I can arrange your accommodation. If you have a disability that might require academic accommodations, please visit Disability Support Services in 204 Norwood Hall (341-4211; [dss@mst.edu](mailto:dss@mst.edu)) very early in the semester.

## Classroom Egress Maps

---

(<http://registrar.mst.edu/links/egress/>):

Please familiarize yourself with the classroom egress maps posted on line so you will know where emergency exits are located.

## Cell Phones

---

Cell phones must be turned off will in the distance classroom.