

## **IS 665: Data Analysis for Information Systems**

### **Course Syllabus Spring 2017**

Instructor: Dr. Lin Lin  
*Office: 5600A Guttenberg Information Technology Center*  
*Phone: (973) 596-5212*  
*e-Mail: llin@njit.edu*

### **Description**

This graduate level introduction to data analysis, probability and statistics from an information systems perspective, including many of the techniques that are most relevant to the profession of Data Scientist for business, data and web analytics, as well as current research areas. The course emphasizes manipulation and analysis of relevant data sets. Course topics include the rudiments of probability and random variables, estimation, hypothesis testing, graphics and visualization, data warehousing and OLAP analysis, dashboard, scorecard, data mining algorithms, optimization techniques, DSS and knowledge systems.

At the end of this course, the student should be able to:

1. Build up a solid foundation of statistics and probability theories
2. Apply simple statistical analysis (such as descriptive statistics, regression analysis and ANOVA) to real world data sets
3. Design and construct data warehouse using SAP BI tools
4. Design and conduct data visualization using SAP Crystal Report
5. Design and construct dashboard using Tableau
6. Master commonly used data mining techniques such as neural networks, decision tree, association rules, clustering, genetic algorithm, SVM, Bayesian Networks, etc.
7. Apply data mining algorithms to real world data sets in the context of web mining, text mining, transaction mining, etc using RapidMiner and SPSS Modeler

### **Prerequisites**

Prior knowledge of statistics and basic knowledge of relational database is required.

## Required Texts:

*Business Intelligence and Analytics: Systems for Decision Support* (10<sup>th</sup> Edition)  
by Ramesh Sharda, Dursun Delen, and Efraim Turban

## Readings

The weekly schedule of readings, topics, and assignments will be in Moodle. Make sure you check Moodle every Monday – I post new materials on Sunday nights.

## Assignments (Individual and Team)

There will be several individual and team assignments over the semester. Details on each assignment will be posted in Moodle.

### READING ASSIGNMENTS:

Each week, each team will be assigned one paper to read. Teams are expected to develop a 5 – 7 page Powerpoint Slide set to summarize the assignment paper. The presentations should be posted to the weekly presentation forum. During class, teams will be randomly chosen to present their findings.

### TECHNICAL ASSIGNMENTS:

There will also be technical homework assignments in this course. Some of them are individual assignments and some will be team-based. More details will be posted on Moodle regarding these assignments.

### LABS:

There will be several labs IN CLASS. Your participation is expected for these labs, and absence during the lab results in 0 point for that lab.

## Projects (In teams)

**Objective:** To demonstrate the ability to apply Business Intelligence techniques to solve real world problems.

**Summary:** THREE projects will be assigned to teams throughout the semester.

### PROJECT ONE: Data Warehousing

Teams are expected to finish a series of tasks regarding data warehouse design, implementation and querying. This project is very structured, with the data created and maintained by the instructor. More details will be revealed later on Moodle.

### PROJECT TWO: Reporting

Teams are expected to find an interesting data set and visualize it using Tableau or Crystal report tools. Each team will then present the visualization model in a 10-minute presentation to the class. This happens in Mid-to-late March depending on our class progress.

### PROJECT THREE: Data Mining

Teams are expected to work with a real-world organization to gather data set, analyze it, and try to extract insightful information / knowledge using RapidMiner or SPSS Modeler

### **Late Assignments Policy**

Unexcused late submission of homework receives a 20% penalty. This means that you start with 8 out of 10 points as the maximum. Assignments submitted after graded assignments are returned or reviewed in class receive no credit.

## Grading

NJIT Academic Policy has grades for graduate courses assigned as follows:

GRADE	GPA	SIGNIFICANCE
A	4.0	Excellent
B+	3.5	Good
B	3.0	Acceptable
C+	2.5	Marginal Performance
C	2.0	Minimum Performance
F	0.0	Failure

Final grades for IS 684 will tentatively be assigned as follows. There may be slight modifications, depending on issues that arise during the semester.

- Labs - 10 %
  - Reading Assignments - 10 %
  - Technical Assignments - 30 %
  - Group Projects - 30 %
  - Final Exam - 20%
- Total: - 100%

Excellent participation demonstrated by preparation for discussion and thoughtful contributions (on-line and in class) will have the effect of **raising** a final letter grade by one value (e.g. B to B+, or B+ to A). Likewise, poor participation demonstrated by consistent lack of preparation for discussion and little or no thoughtful contributions (on-line and in class) will have the effect of **lowering** a final letter grade by one value (e.g. A to B+, B to C+).

## Honor Code

Any evidence of cheating in any form, including plagiarism, will be dealt with according to the honor code of NJIT (course failure and suspension or expulsion). Please note: There will be no warnings or chances with regard to cheating. Any discovered case of cheating will be immediately passed to the Dean of Students for further investigation. Cheating is not worth it.

You may not only fail this course but also be suspended from NJIT. The full text of the NJIT Honor Code is available for your review at <http://www.njit.edu/academics/honorcode.php>.

### Spring 2017 Outline/Weekly Schedule – Subject to Minor Modification

<b>Week</b>	<b>Theoretical Topics</b>	<b>Labs</b>
1	Introduction	N/A
2	Business Value of Data Analytics	Stats Lab I
3	Probability Theory & Statistics Basics (I)	Stats Lab II
4	Probability Theory & Statistics Basics (II)	Stats Lab III
5	Database & Data Warehouse (I)	SAP BI Lab I
6	Database & Data Warehouse (II)	SAP BI Lab II
7	Data Visualization I: Basics	Crystal Report Lab
8	Data Visualization II: Dashboard and Scorecard	Tableau Lab
9	Data Mining (I)	RapidMiner Lab I
10	Data Mining (II)	RapidMiner Lab II
11	Data Mining (III)	RapidMiner Lab III
12	Optimization	RapidMiner Lab IV
13	Project Presentations	