



**ISM 4117 - 001**

**CRN# 12700**

**Data Mining & Predictive Analytics**

**Fall 2016**

**Tuesday 6:30pm – 9:20pm (FL 409)**

### **Professor Information**

Mary Schindlbeck, Ph.D.

Office Location: Fleming Hall - FL 317 – Boca Campus

Email: [mschind2@fau.edu](mailto:mschind2@fau.edu)

Office Phone: 561-297-3661

### **Office Hours**

Tuesday & Thursday 12pm- 1:45pm and 4pm-5pm

### **Required Text and Materials**

#### **Text:**

- Data Mining for Business Intelligence, Third Edition
- Authors: Galit Shmueli, Nitin R. Patel and Peter C. Bruce
- Publisher: John Wiley & Sons; 2016
- ISBN: 978-1118729274
  - Available on Amazon [https://www.amazon.com/Data-Mining-Business-Analytics-Applications/dp/1118729277/ref=asap\\_bc?ie=UTF8](https://www.amazon.com/Data-Mining-Business-Analytics-Applications/dp/1118729277/ref=asap_bc?ie=UTF8)
  - Each **new textbook** includes an Access Code, to obtain a six-month license to use XLMiner Educational edition.
  - To obtain the software, students must register on this page: [Software for Data Mining for Business Intelligence](#), enter your name, university name, email (key information will be sent to this address), Course Code (obtain from instructor), and Textbook code (enter **SDMB13**).
  - On the download page, change 32-bit to 64-bit only if you have confirmed that you have 64-bit Excel.
  - Complete instructions are listed in the first few pages of the text.

#### **Materials:**

- Microsoft Excel will be used to do calculations, data analysis and graphing
  - Microsoft Office 365 is available for free for all currently enrolled FAU students
  - For more information visit <http://www.fau.edu/oit/getoffice365>
- Data Mining Software: XLMiner® for Windows - a comprehensive data mining add-in for Excel, with neural nets, classification and regression trees, logistic regression, linear regression, Bayes classifier, K-means, association rules, clustering, and principal components analysis. Students enrolled in the class will have access to the XLMiner software in the instructional lab FL409.
- Additional materials will be assigned and distributed via Blackboard throughout the course.

### **Competencies:**

- Know how to use & navigate Blackboard 9 Platform
- Know how to take exams & quizzes on Blackboard 9 Platform
- If you do not know how to use Blackboard 9, please click on the Blackboard help link below & complete the tutorials in order to become comfortable using Blackboard:  
[https://help.blackboard.com/en-us/Learn/9.1\\_2014\\_04/Student](https://help.blackboard.com/en-us/Learn/9.1_2014_04/Student)
- Students should have a working knowledge of basic math (algebra) and Microsoft Excel.
- Students should have access to Excel spreadsheet software and students are assumed to be familiar at an intuitive level with general business practices of collecting, storing and using data.

### **Course Description**

- Introduces the core concepts of data mining (DM) and predictive analytics, the techniques, implementation, and benefits. Course also identifies industry branches that most benefit from DM and predictive analytics, such as retail, target marketing, fraud protection, health care and science, and web and e-commerce. Detailed case studies and using leading mining tools on real data are presented.

### **Course Prerequisites and Credit Hours**

- No course prerequisites  
3 Credit Hours

### **Course Learning Objectives**

The primary aim of this course is to acquaint students with the practical challenges that are encountered when solving real-world data mining problems. By allowing students to encounter the complete picture of going from business level problem formulation, through the analysis of quantity and quality of available data, as well as selection of suitable algorithms and evaluation of the obtained results. Students are expected to learn not only about the advantages of various data mining and predictive analytics methods, but also about their limitations. The main focus in this course will be on applying, in practice, knowledge and concepts from business models and on understanding the applicability of data mining and predictive analytics.

Students will reinforce the learning of data mining concepts by means of data analysis techniques to make better business decisions through proper data preparation, data exploration and simple tools for solving data mining problems. Students will be introduced to advanced concepts such as data mining applications, predictive analytics, web mining, text mining, and ethical aspects of data mining and predictive analytics. Students will learn to mine heterogeneous data and demonstrate proficiency in classification and prediction applications such as neural networks, linear regression, cluster analysis, market basket analysis and decision trees.

Working as a team, students will demonstrate proficiency in applying data mining analytical techniques on an advanced real world business problem that examines a large amount of data to discover new information in addition to analyzing and evaluating technique effectiveness with a less

than perfect constantly evolving technology by presenting a self-designed semester project. Commencing with several singular technique projects and concluding with the comprehensive semester project, students will reinforce their oral skills by way of presentations as well as written and critical thinking skills by the use of executive memos requiring quantitative analysis and evaluation.

**Grading Scale**

A	93.00-100%	C	73-76.99%
A-	90-92.99%	C-	70-72.99%
B+	87-89.99%	D+	67-69.99%
B	83-86.99%	D	63-66.99%
B-	80-82.99%	D-	60-62.99%
C+	77-79.99%	F	< 60 %

In fairness to all students, under no circumstances will grades be inflated at the end of the term. It would unethical to change an individual’s grade or offer an additional opportunity to improve a grade. An Excel spreadsheet has been provided on Blackboard to assist you in determining the grade you have earned for the course.

**Course Evaluation Method**

<b>6 Team Projects (5% each)</b>	<b>30%</b>
<b>Team Final Project</b>	<b>20%</b>
<b>Team Final Presentation</b>	<b>10%</b>
<b>Midterm Exam</b>	<b>20%</b>
<b>Final Exam</b>	<b>20%</b>
<b>Total</b>	<b>100%</b>

- **Exams:** The two exams (midterm and final) will include multiple choice questions and data mining application problems, administered on Blackboard during class and will **cover the content from the text, data mining techniques using XLMiner, material presented in the lectures and from the team assignments.**
  - Usually, students will be asked to interpret results from applying a specific data mining method, such as confusion matrices and classification false positive/negative rates.
  - Therefore, assignments, discussions, class attendance and good note taking are essential elements for success. Each exam has a time limit of 90 minutes.

- **Projects:** Team projects will emphasize a specific data mining method or principle.
  - The team should be of exactly 2 students. Finding a team partner is solely the students' responsibility. Choose your partner carefully, identify if your goals in this course are common and if the level of commitment is the same. If there are differences on these two basic criteria, chances are you will not collaborate effectively and there will be problems down the road. It is to your best advantage to document (email) your communications to avoid complications; if you feel more comfortable, use Google Docs to complete your projects and/or feel free to cc: your emails to me. If there is an odd number of the students in class, the instructor will have the discretion to place the remaining student in a team whose team members should do everything possible to work together as a team of three.
  - Your team will use the same data set for each of the first four projects, the last 2 assignments will contain a dissimilar set of data. For each assignment you will post all of the files you created in the Assignment Section of Black Board before the due date and time; penalty of 10% for each day exists for late submissions. Some teams will present their findings and other teams will participate in a discussion about the findings; our class will be similar to a project team. Individual assignments will not be accepted.
  - While you will collaborate with your peers on the team projects, each team member must contribute an equal portion of effort or work towards each phase of the project. Based on peer evaluations and observations of group work by the professor, team members contributing more than their fair share may receive higher grades than other group members. It is the group's responsibility to advise the professor of any issues or challenges that may hinder your ability to complete any phase of the project.

### Project Submissions:

- The project submission must include the following:
  - The actual Excel spreadsheet file(s) where the method/tool was applied.
  - The necessary additions such as confusion matrices, classification rates, etc., that reinforce the appropriate conclusions (can be added as worksheets to the original Excel file).
  - Memorandum if required that concisely presents, summarizes, and analyzes the results (draw meaningful conclusions and comparisons).
  - For each project submitted the file names will contain the assignment number, dataset name and the last names of the team members. **For example, for assignment 1 for students Jane Smith and Joe Cole using the insurance dataset, the file name should be ASG1\_Insurance\_Smith\_Cole.xxx** (.docx or .xlsx depending on the type of file).
  - **Worksheets in an Excel workbook must be appropriately named and unnecessary sheets must be removed.**
  - Detailed information about memos and project examples can be found on Blackboard.

**Participation & Discussion:**

- The team assignments, after submission, will be discussed in a class session. Far from everything will be clear and exact in these sessions – we will need a lot of input and brainstorming – a normal process when engaged in highly analytical work such as data mining and cleaning the data. Students are expected to actively participate and generate discussions on the techniques used and the results.
- Open discussions and participation are important elements of the course. Whether your techniques, methods and conclusions are correct or incorrect, the discussion grade will not be affected. The goal is to reach the best method and solution through sharing each group's ideas. Participation also includes bringing relevant current topics in the news into the classroom.

**Final Project & Presentation:**

- Final Project: The following is an overview of the final project.
  - A detailed document is provided on Blackboard regarding all requirements of the final data mining and predictive analytics project. The same rules and suggestions apply as stated above for the five team projects. Individual projects are not accepted.
  - A project proposal including the data source and data description must be pre-approved by the instructor by the proposal due date.
  - The project will require locating a large data set (between 3,000 and 10,000 records) with at least 8 variables of differing data types, preparing and understanding the data and addressing a business question suitable to the data chosen.
  - After exploring the dataset, the prepared data will be applied to the data mining and predictive analytic techniques (linear regression, classification trees, neural networks) previously used in class.
  - This project will demonstrate a comprehensive understanding of the course; additional details about the final project can be found on Blackboard.
- Presentation:
  - The final project will be presented to a peer audience and not exceed 20 minutes.
  - The presentation will include the analysis of each technique as well as a comparison/contrast of the techniques applied.
  - Discussion of the output can be included in PowerPoint slides and should be able to be demonstrated. Complete final project rubrics are available on Blackboard.
  - The team partners may or may not receive identical grades on the final project; it is expected that partners have contributed equally to the project. However, if this is not apparent during the presentation the grades will differ.

## **Additional Course Policies**

### **Missed Exams**

It is important that each exam be taken at the scheduled time and date. Any excusable absence (official athletic event, religious holiday, etc.) must be documented by a verifiable source and I must be notified at least one week prior to the exam. If you are absent from an exam due to illness or emergency, you must notify me by e-mail within 24 hours of the missed exam and provide verifiable documentation within one week of the exam date; the make-up policy is not applicable if you fail to report an absence as stated above. There will be two semester exams, each covering approximately one-half of the course material. A mid-term exam missed with prior documented approval as stated above may be made up by the Final exam. The score earned on the Final exam will be used for both the final and for the missed exam. An exam missed without prior approval and verifiable documentation that the unapproved absence was unavoidable as stated above cannot be made up.

### **Late Assignments**

Grade penalty equal to 10 percent of the project grade per day late will be applied after the project's due date.

### **Attendance Policy**

Learning is an interactive process and success in this course depends on the experiences the students bring to the classroom (our learning community). Therefore attendance is an important aspect of this course. Attendance will not be taken. However, you are responsible for everything that takes place in class. Additional homework assignments, their due dates, and changes to the tentative schedule will be announced in class. Occasionally, unannounced in-class exercises (or quizzes) will be given; if missed, these cannot be made up. Due to the cumulative nature of the material it is imperative that students keep up with the course materials on a daily basis. Attendance is strongly suggested and is a prerequisite for successful completion of this course. Missing classes will adversely affect your performance. The probability of successfully passing the tests in the course is directly dependent on regular attendance, studying the assigned materials and completing projects and lab exercises in a timely manner.

### **Etiquette and/or Netiquette Policy**

Each student is responsible for keeping up with the class schedule, checking your FAU email account, and checking the course Blackboard site on a regular basis. If you use a non FAU email address as your primary address, arrange for FAU email to be forwarded.

### **Anti-plagiarism Software**

Written components of any assignment or project may be submitted to anti-plagiarism software to evaluate the originality of the work. Any students found to be submitting work that is not their own will be deemed in violation of the University's honor code discussed below.



**Academic Irregularities:**

Cheating, plagiarism, copying, unauthorized collaboration, and hiring another person to do your assignments and exams are unacceptable, and are subject to disciplinary actions, including, but not limited to, an "F" in the course, a letter of fact on your student record, and a notation on your transcript in accordance with the policies of FAU and the College of Business. In cases where this has occurred, both the person who cheats/plagiarizes/copies/collaborates/hires another person AND the originator of the work will be punished. For information about the University's Honor Code, please refer to the policy statement under the section titled Selected College and University policies. For more information, please see FAU Regulation 4.001 at: [FAU Regulation 4.001](#)

**Selected University and College Policies**

**Code of Academic Integrity Policy Statement**

Students at Florida Atlantic University are expected to maintain the highest ethical standards. Academic dishonesty is considered a serious breach of these ethical standards, because it interferes with the university mission to provide a high quality education in which no student enjoys an unfair advantage over any other. Academic dishonesty is also destructive of the university community, which is grounded in a system of mutual trust and places high value on personal integrity and individual responsibility. Harsh penalties are associated with academic dishonesty. For more information, please see FAU Regulation 4.001 at: [FAU Regulation 4.001](#).

**Disability Policy Statement**

In compliance with the Americans with Disabilities Act Amendments Act (ADAAA), students who require reasonable accommodations due to a disability to properly execute coursework must register with Student Accessibility Services (SAS) in Boca Raton, SU 133 (561-297-3880); in Davie, LA 131 (954-236-1222); or in Jupiter, SR 110 (561-799-8585) - and follow all SAS procedures. The web site is: <https://fau.edu/sas>

**Religious Accommodation Policy Statement**

In accordance with rules of the Florida Board of Education and Florida law, students have the right to reasonable accommodations from the University in order to observe religious practices and beliefs with regard to admissions, registration, class attendance and the scheduling of examinations and work assignments. For further information, please see FAU Regulation 2.007 at: [FAU Regulation 2.007](#).

**University Approved Absence Policy Statement**

In accordance with rules of the Florida Atlantic University, students have the right to reasonable accommodations to participate in University approved activities, including athletic or scholastics teams, musical and theatrical performances and debate activities. It is the student's responsibility to notify the course instructor at least one week prior to missing any course assignment. For further information, please see [Academic Policies and Regulations](#).

**Incomplete Grade Policy Statement**

A student who is passing a course, but has not completed all work due to exceptional circumstances, may, with consent of the instructor, temporarily receive a grade of incomplete ("I"). The assignment of the "I" grade is at the discretion of the instructor, but is allowed only if the student is passing the course.

The specific time required to make up an incomplete grade is at the discretion of the instructor. However, the College of Business policy on the resolution of incomplete grades requires that all work required to satisfy an incomplete ("I") grade must be completed within a period of time not exceeding one calendar year from the assignment of the incomplete grade. After one calendar year, the incomplete grade automatically becomes a failing ("F") grade.

**Withdrawals**

- Any student who decides to drop is responsible for completing the proper process required to withdraw from the course.

**Grade Appeal Process**

- A student may request a review of the final course grade when s/he believes that one of the following conditions apply:
  - There was a computational or recording error in the grading.
  - Non-academic criteria were applied in the grading process.
  - There was a gross violation of the instructor's own grading system.

The procedures for a grade appeal may be found in [FAU Regulation 4.002](#).

**Disruptive Behavior Policy Statement**

- Disruptive behavior is defined in the FAU Student Code of Conduct as "... activities which interfere with the educational mission within classroom." Students who behave in the classroom such that the educational experiences of other students and/or the instructor's course objectives are disrupted are subject to disciplinary action. Such behavior impedes students' ability to learn or an instructor's ability to teach. Disruptive behavior may include, but is not limited to: non-approved use of electronic devices (including cellular telephones); cursing or shouting at others in such a way as to be disruptive; or, other violations of an instructor's expectations for classroom conduct.

**Faculty Rights and Responsibilities**

- Florida Atlantic University respects the right of instructors to teach and students to learn. Maintenance of these rights requires classroom conditions which do not impede their exercise. To ensure these rights, faculty members have the prerogative:
  - To establish and implement academic standards
  - To establish and enforce reasonable behavior standards in each class
  - To refer disciplinary action to those students whose behavior may be judged to be disruptive under the Student Code of Conduct





Tentative Course Schedule: **Note: all times apply to Eastern Time Zone, USA.**

Date	Lecture	Readings*	Assignments & Due Dates
8-23	Introduction to Data Mining & Data Mining Techniques Data Preprocessing	Shmueli-Chapter 1 Shmueli-Chapter 2	
8-30	Data Exploration: Data Visualization & Dimension Reduction	Shmueli -Chapters 3&4 BB-XLMiner Notes	Install XLMiner
9-6	Evaluation of Model Performance PRESENT & DISCUSS – Data Exploration	Shmueli -Chapter 5	Assignment 1 Data Exploration
9-13	Linear Regression & Logistic Regression Regression Lab: XLMiner and Excel	Shmueli –Chapters 6 & 10 BB-Regression Notes	
9-20	PRESENT & DISCUSS – Regression		Assignment 2 Regression
9-27	Classification & Regression Trees Decision Tree Lab: XLMiner	Shmueli -Chapter 9 BB-Decision Tree Notes	
10-4	PRESENT & DISCUSS - Decision Trees		Assignment 3 Decision Tree
10-10&11	<b>Fall Break</b>	<b>No Classes</b>	
10-18	<b>MIDTERM EXAM 6:30pm</b>		
10-25	Neural Networks in Data Mining Neural Networks Lab: XLMiner	Shmueli -Chapter 11 BB-NN Notes	<b>Final Project Proposal Due</b>
11-1	PRESENT & DISCUSS - Neural Networks		Assignment 4 Neural Network
11-8	Combining Methods: Ensembles &Uplift Association Rules-Market Basket Analysis	Shmueli -Chapter 13 Shmueli -Chapter 14 BB-MBA Notes	
11-15	Cluster Analysis K-Means/Clusters Lab: XLMiner	Shmueli -Chapter 15 BB-K-Means Notes	Assignment 5 Ensemble & MBA
11-18	Last day to drop or withdraw without receiving an “F” in the course.		
11-22	<b>FINAL PROJECT Presentations</b>	<b>Group 1</b>	Assignment 6 K-Means Cluster
11-29	<b>FINAL PROJECT Presentations</b>	<b>Group 2</b>	<b>Final Report Due Group 1</b>
12-6	<b>FINAL PROJECT Presentations</b>	<b>Group 3</b>	<b>Final Report Due Groups 2 &amp; 3</b>
12-7	<b>Reading Days</b>	<b>No classes</b>	
12-13	<b>FINAL EXAM 7:00pm*</b>		

\*Shmueli - Course Textbook 3e BB – Blackboard

**Online Resources and Reference Material List**

- Journals
  - [SIGKDD Explorations](#), a magazine of the [SIGKDD](#), the data miners professional group
  - [Data Mining and Knowledge Discovery](#), journal published by Springer
  - [Journal of Big Data](#), a Springer Open Journal.
  - [Journal of Data Science](#), International journal devoted to applications of statistical methods
  - [International Journal of Data Mining & Knowledge Management Process](#)
  - [Journal of Machine Learning Research](#)
- Books
  - D. Abbott, *Applied Predictive Analytics*, ISBN: 978-1-118-72796-6, John Wiley & Sons, 2014
  - C. Carlberg, *Predictive Analytics: Microsoft Excel*, Pearson Education, Inc., 2013
  - D. Delen, *Real-World Data Mining*, Pearson Education, Inc., 2015
  - J.W. Foreman, *Data Smart*, John Wiley & Sons, 2014
  - J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2011
  - C. N. Knaflitz, *Storytelling with Data*, John Wiley & Sons, 2015
  - F. Provost and T. Fawcett, *Data Science for Business: What you need to know about data mining and data-analytic thinking*, O'Reilly Media, 2013