CIS 8695 – BIG DATA ANALYTICS (CRN 93503)

Class Time: Wednesday 5:30-9:45pm
Class Location:
Instructor: Ling Xue Email: lxue5@gsu.edu
Office: RCB 907 Office Hours: By Appointment

Covid-19 Policy: Masks, social distancing, and vaccination are not mandatory for this class. However, for your own health and the safety of others, **I strongly urge you to mask up and practice social distancing all the time, and get vaccinated for taking this course**.

Catalog Description:
Data analytics is an interactive process of analyzing and exploring enterprise data to find valuable insights that can be exploited for competitive advantage. This course address data analytics models and methods, in the general business environment featured by the "big data" trends (increasing volume, variety, velocity, and veracity of data). The focus is on predictive analytics and related machine learning techniques.

Course Description:
With the widespread and economical availability of computing power and data management techniques, organizations have huge amounts of data (BIG DATA). Over the years, companies have come to realize that these data are of little use without systematic analysis to find trends, patterns, associations that help them make sense of their data and make better decisions. Data Analytics sits at the intersection of Information Systems, Statistics, Data Mining, Machine Learning and Data Management and helps organizations make better informed, data-driven decisions. Without the ability to transform the data into actionable intelligence, the volumes of organizational data add little value. Organizations willing to change the way they do business based on insights from analytics will prevail. Business analytics are becoming an even more critical capability for enterprises of all types and all sizes, and skilled business analytics professionals are in high demand.

In this course, you will learn to identify, evaluate, and capture business analytic opportunities that create value. It should make you alert to the ways that analytics can be used and abused in organizations. You will learn basic analytic methods and analyze data to identify trends and develop analytics models that inform business decisions. We focus on how to use data to develop insights and predictive capabilities using machine learning, data mining and forecasting techniques using R. The concepts learned in this class help you identify opportunities where business analytics can be used to improve performance and inform important decisions in organizations.

Course Credit: 3.0 Credit Hours

Course Objectives:
Upon completion of the course, students should be able to:
1. Demonstrate an understanding of data analytics and machine learning;

2. Identify, design and assess different data analytics and machine learning methodologies;
3. Prepare and formulate data collection, sampling, preprocessing;
4. Explore and develop descriptive and predictive analytic models;
5. Apply and assess different predictive modeling and machine learning techniques;
6. Evaluate efficacy of different analytics model implementations;
7. Demonstrate proficiency in the use of R.

Recommended Textbooks:
Galit Shmueli, Peter C. Bruce, Inbal Yahav, Nitin R. Patel and Kenneth C. Lichtendahl Jr. *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R*, Wiley, 2018. ISBN: 978-1118879368.

Additional articles that are required reading will be provided in class.

Required Software: R and R Studio

Homework Assignments:
A series of hands-on homework assignments will be provided to you to further explore the topic/technique covered in class.

Term Project:
A group-based term project is for students to integrate the analytics skillsets to solve business problems.

Typical class session:
Class sessions will comprise (1) lectures/discussions of relevant techniques, concepts and features, (2) instructor demonstrations, and (3) student lab sessions with in-class hand-on exercises. The purpose of this pedagogical approach is to introduce and reinforce ideas and skill sets so that you can master these on your own after class hours. To bring this knowledge to a highly proficient, professional level, you will have to spend time and effort outside of class reviewing and practicing the class material. To ensure that you have the basic knowledge that will allow you to function on your own after class, be sure to ask the instructor questions during class, either during the lecture/discussion, demo, or lab.

Class Attendance
All students are required to attend all classes and complete in-class exercises, except when precluded by emergencies, religious holidays or bona fide extenuating circumstances. If one or more class is missed, it is the student's responsibility to determine the specific material covered during their absence and make the necessary arrangements for making up what is missed.

In-class exercises are supposed to be completed in class and turned in at the end of each class session to earn class participation grades.

Course Grading

Grading Component	Percentage
Home Assignments	20%
Term Project	35%
Final Exam	35%
Class Participation (In-class Exercises)	10%
Total	**100%**

A+ = >=98	A = 93 – 97.9	A- = 90 – 92.9
B+ = 87 – 89.9	B = 82 – 86.9	B- = 80 – 81.9
C+ = 76 – 79.9	C = 71 – 75.9	C- = 68 – 70.9
D+ = 65 – 67.9	D = 60 – 64.9	F <= 59.9

Important Note

This syllabus provides a general guideline for the conduct of this course; however, deviations may be necessary. Updates will be given during the semester and posted online through iCollege. If the class cannot be held at the scheduled time or place, it may be held via an online forum.

Academic Honesty

Students may have general discussions about assignments with fellow classmates, but each student must develop his or her solution to each Mini-Project. It is each student's responsibility to keep his/her own work secure. DO NOT share computer files of Mini-Project Assignments with classmates. Failing to adequately protect one's work does not relieve the student from academic dishonesty charges.

University regulations will be enforced regarding dishonorable or unethical conduct (Cheating, Plagiarism, Falsification, Unauthorized Collaboration or Multiple Submissions). The penalties for incidents of academic dishonesty can lead to expulsion from the University (see General Catalogue p. 64, Student Handbook p. 130 or http://www2.gsu.edu/~wwwdos/codeofconduct_conpol.html). In this class, there will be zero tolerance for dishonorable or unethical conduct. Electronic or physical sharing of answers will be considered cheating and will not be tolerated.

Cheating on examinations involves giving or receiving unauthorized help before, during, or after an examination. Examples of unauthorized help include sharing information with another student during an examination, intentionally allowing another student to view one's own examination, and collaboration before or after an examination which is specifically forbidden by the instructor.

Submission for academic credit of a work product, or a part thereof, represented as its being one's own effort, which has been developed in substantial collaboration with assistance from another person or source, or computer based resource, is a violation of academic honesty. It is

also a violation of academic honesty to knowingly provide such assistance. Collaborative work specifically authorized by an instructor is allowed. (*Collaboration on all individual assignments is forbidden. If your instructor discovers that you have had unauthorized assistance or collaboration, the instructor is obligated to file a report with the Dean's Office.)*

If a student is charged with Academic Dishonesty, for each charge, a zero (0) will be given for the assignment, a minimum of point equivalent of one final grade (i.e. B- to a C-) will be deducted from the final course total points and a written Notice of Academic Dishonesty will be given to the Dean's office. The student will also receive a copy of the notice.

Unless specifically stated by the instructor, all exams and at-home assignments are to be completed by the student alone. Within-group collaboration is allowed on project work. Collaboration between project groups will be considered cheating unless specifically allowed by an instructor.

Copying work from the Internet without a proper reference will be considered plagiarism and subject to disciplinary action as delineated in the Student Handbook.

	Tentative Class Schedule			
	Date	**Topic**	**Due**	**Reading**
1	8/18	Introduction to data analytics and machine learning. **Regression-based Learning:** linear regression, logistic regression, ridge regression, lasso, etc.		Ch. 1, 2, 6, 10
2	8/25	**Time-series Forecasting:** trend and seasonality, regression-based forecasting, autocorrelation and ARIMA models, moving average and smoothing.		Ch. 16, 17, 18
3	9/1	**Tree-based Learning:** classification and regression trees (CART), random forest, boosting tree, etc.	HW1	Ch. 9
4	9/8	**More classifiers:** Naïve Bayes, Support Vector Machine (SVM), KNN, etc.	HW2	Ch. 7, 8
5	9/15	**Model Evaluation and Combination:** Model comparison, ensemble models, uplifting model, etc.	HW3	Ch. 5, 13
6	9/22	**Neural Network:** NN-based classification and prediction. **Dimension reduction**: PCA and SVD, clustering techniques, etc.	HW4	Ch. 11, 4, 15
7	10/1	**Recommendation techniques**: Association rules, collaborative filtering, etc. **Deeping Learning & AI:** applications in text classification, image recognition and classification, etc.	HW5	Ch. 14, 11
8	10/8	**Project Presentation.** **Final Exam**	TP	

**HW- Homework; TP- Term Project