Washington University in St. Louis
SCHOOL OF ENGINEERING & APPLIED SCIENCE

T81-574: Foundations of Analytics

Fall 2019

INSTRUCTOR: Dihui Lai, PhD
Senior Data Scientist @ RGA
Contact info (dlai@wustl.edu; dlai@rgare.com)

Master of Engineering Management
The Henry Edwin Sever Institute | School of Engineering & Applied Science
Washington University in St. Louis

COURSE OVERVIEW

The steeply decreasing costs to gather, store, and process data has created a strong motivation for organizations to move toward "data driven" approaches to problem solving. As such, data analytics continues to grow rapidly in importance across industry, government, and non-profit organizations. This course seeks to equip students with a wide range of data analytics techniques that serve as the foundation for a broad range of applications including descriptive, inferential, predictive, and prescriptive analytics. Students will learn the process of building a data model as well as a variety of analytics techniques and under what situations they are best employed. Through lectures and practical exercises, students will become familiar with the computational mathematics that underpin analytics; the elements of statistical model and machine learning; model interpretation and assessment; structured and unstructured data analysis. Students will also undertake a project to build an analytical model using a "real-world" data set.

TEXTBOOKS

Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd Edition) ISBN: 0387848576
Daniel Jurafsky, James H. Martin. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition (2nd Edition) ISBN: 9780131873216

GROUPS

Please form a team of 4-5 members for the final project. Some homework will be assigned as group practice too. Please consider the following roles while you are forming a team: speaker, programmer, and project manager.

COURSE SCHEDULE

Date	Theme / Topics	Homework Due
08/27/2019	Introduction to data science/analytics • Linear algebra review • Python environment setup	Start to form teams
09/03/2019	Statistical description of structured data	Submit team proposals

	• Introduction to statistics: random variables, random distribution, histogram • Statistic distributions: Gaussian, Poisson etc.	
09/10/2019	Linear model and logistic regression • Correlation • Linear regression; logistic regression • Optimization method I: Newton method and gradient descent	Homework 1 Finalize teams
09/17/2019	Generalized Linear Model • Exponential family • Generalized linear model	
09/24/2019	Statistical Modeling Framework • Empirical modeling practices • Feature engineer; variable selection • Model evaluations	Homework 2
10/01/2019	Machine Learning I • Naïve Bayesian • KNN	
10/08/2019	Machine Learning II: Tree Algorithms • CART model • Random forest and GBM • Optimization method II	
10/15/2019	Mid-Term Presentation	Homework 3
10/22/2019	Machine Learning III: • Random process, random walk • Hidden Markov model • Neural network basics; stochastic gradient descent	
10/29/2019	Nature Language Process I • Text mining basics: regular expression, word count, bag of words • Word embedding; language model	
11/05/2019	Nature Language Process II • Part-of-speech tagging • Named entity recognition • Document classification	Homework 4
11/12/2019	Unsupervised Learning • PCA • K-mean clustering and hierarchical clustering	
11/19/2019	Image Analysis • Computer vision • Optical character recognition (OCR)	Homework 5
11/28/2019	Thanksgiving Break – no classes	
12/03/2019	Last official day of class Final presentation	Final paper

COURSE GRADINGS
Grades will be based on the assignment sets (50%), mid-term exam (20%) and the group term project (30%).

COURSE ASSIGNMENTS
There will be approximately 5 course assignments that will be assigned approximately weekly and due bi-weekly. Students will have two weeks to complete and submit the assignment. Assignments are due in class on the due date. Late assignments will not be accepted. The homework can be submitted electronically.

GRADING SCALE:

Letter Grade	%	Points Toward GPA		Letter Grade	%	Points Toward GPA
A+	≥ 97%	4.0		C+	77% - 79%	2.3
A	93% - 97%	4.0		C	73% - 76%	2.0
A-	90% - 92%	3.7		C-	70% - 72%	1.7
B+	87% - 89%	3.3		D+	67% - 69%	1.3
B	83% - 86%	3.0		D	65% - 66%	1.0
B-	80% - 82%	2.7		F	< 65%	0.0

Privacy and Security
Recording of class sessions either audio or video is prohibited without permission from the instructor and the other class members.

Collaboration:
With the exception of your team projects, all assignments are to be completed on your own. You are encouraged to discuss ideas and techniques broadly with other class members, but all written or presentation work, whether in preliminary or final form, is to be generated by you working alone. If in doubt - *ask*.

Language Sensitivity
When in the classroom, all students should speak English at all times. While meeting with classmates on a classroom project, speak a language that every student present (in your group) understands, without exception.

Professionalism
You are part of a professional, graduate program. Consequently, it is expected that your fellow students conduct yourselves in a *professional* manner. This includes being on time for classes and meetings, being prepared, and participating in class discussions, group activities, projects,

etc. The level of professionalism you exhibit throughout the course will impact your final grade. It directly affects the participation portion of the grade but is also taken into consideration in all other aspects of the course as it reflects the overall quality of professional performance.

Mental Health Service

Mental Health Services' professional staff members work with students to resolve personal and interpersonal difficulties, many of which can affect the academic experience. These include conflicts with or worry about friends or family, concerns about eating or drinking patterns, and feelings of anxiety and depression. See: shs.wustl.edu/MentalHealth

Sexual Harassment

Sexual harassment is a form of discrimination that violates university policy and will not be tolerated. It is also illegal under state and federal law. Title IX of the Education Amendments of 1972 prohibits discrimination based on sex (including sexual harassment and sexual violence) in the university's educational programs and activities. Title IX also prohibits retaliation for asserting claims of sex discrimination. The university has designated the Title IX Coordinator identified below to coordinate its compliance with and response to inquiries concerning Title IX. For more information or to report a violation under the Policy on Discrimination and Harassment, please contact:

Discrimination and Harassment Response Coordinators

Apryle Cotton, Asst. Vice Chancellor for Human Resources
Section 504 Coordinator
Phone: 314-362-6774
Email: apryle.cotton@wustl.edu
Leanne Stewart, Employee Relations Manager
Phone: 314-362-8278
Email: leannerstewart@wustl.edu

Title IX Coordinator

Jessica Kennedy, Director of Title IX Office
Title IX Coordinator
Phone: 314-935-3118
Email: jwkennedy@wustl.edu

You may also submit inquiries or a complaint regarding civil rights to the United States Department of Education's Office of Civil Rights at 400 Maryland Avenue, SW, Washington, DC 20202-1100 or by visiting the U.S. Department of Education website or calling 800-421-3481.