ACM Data Science Task Force Course Example

DA 350: Advanced Methods for Data Analytics

Denison University, Granville OH
Anthony Bonifonte
bonifontea@denison.edu

Bonifonte SP 19 DA 350-20200917T180426Z-001.zip

Knowledge Areas that contain competencies (knowledge, skills, and dispositions) covered in the course

Knowledge Area	**Total Number of Contact Hours**
Machine Learning (ML)	20
Programming, Data Structures, and Algorithms (PDA)	13
Data Acquisition, Management, and Governance (DG)	10
Professionalism (PR)	5
Big Data Systems (BDS)	4
Data Mining (DM)	4
Artificial Intelligence (AI)	3
Analysis and Presentation (AP)	3
Data Privacy, Security, Integrity, and Analysis (DPSIA)	1

Where does the course fit in your undergraduate Data Science curriculum?

This course is required for all DA majors in our program. It usually taken in the second semester of year 3 or the first semester of year 4 because of the advanced topics and math and cs prerequisites.

Is this course from or used in other curricula/majors?

The course was designed from scratch for our new DA program. Some math and CS majors take the course to broaden their skillset. While similar classes are offered in the math and CS department, this course is unique in being heavily application and implementation driven.

What is covered in the course?

This course is designed to develop students' understanding of the cutting-edge methods and algorithms of data analytics and how they can be used to answer questions about real-world problems. These methods can learn from existing data to make and evaluate predictions. The course will examine both supervised and unsupervised methods and will include topics such as dimensionality reduction, machine learning techniques, handling missing data, and prescriptive analytics.

What is the format of the course?

The course is usually face to face (pre-COVID) and is a mixture of in person lecture, project based labs, and discussion. It is usually taught on a 3 days/week basis in an 80min time slot. Total contact hours are roughly 63.

How are students assessed?

See syllabus for more details but graded work is a combination of multiple project based labs requiring creative mastery of concepts (45%), weekly short concept quizzes (20%), a presentation on student research into a cutting edge algorithm and its applications (15%), and a computation final challenging students to demonstrate and make creative use of what they've learned on a real world, "messy" problem (20%).

Course tools and materials

The course primarily uses R in conjunction with a variety of ML and other specific packages. Other instructors who teach this course have used Python in lieu of R. Datasets vary to take advantage of instructor and student interest and to avoid stagnation over repeated offerings. Care is taken to choose data from a variety of domains to emphasize interdisciplinary and diverse applications. The textbook is *An Introduction to Statistical Learning* by James et.al. Everything in the course is free, including software, data, and a publicly available electronic version of the textbook.

Why do you teach the course this way?

The course provides advanced majors experience and familiarity with many of the techniques and theories in modern Data Science. It was reviewed by our faculty governance process when originally proposed as a concept course for the major 5yrs ago. The course is generally considered the most challenging and technically valuable class in the major. Course objectives are available in detail on the syllabus, but include: Apply advanced data analytics methods to solve real world problems; generate predictions for classification and regression problems using cutting edge machine learning techniques; and describe structures in data using unsupervised algorithms. Students consider the course challenging but achievable. Students face particular challenges grappling with large data and run times, and they are challenged to think deeply about coding efficiency.

Body of Knowledge coverage

KA	Sub-domain	Competencies Covered	Hours

AP	Visualization	Visualizing ML outputs	3
AP	Foundational considerations	Presenting technically complex topics	2

AI	General	AI applications	3
BDS	Problems of Scale	Managing large data in practice	3
BDS	Techniques for Big Data applications	Applying methods with large data	1
DAMG	Data Acquisition	Acquiring data	2
DAMG	Data Reduction and Compression	Reducing data with PCA	2
DAMG	Data Cleaning	Foundational cleaning	3
DAMG	Information Extraction	Putting data into format for ML algs	3
DM	Outlier detection	Removing outliers	1
DM	Cluster Analysis	k-means and hierarchical	3
DPSI	Data privacy	Privacy concerns	1
ML	General	Interpretability, balancing complexity of models	2
ML	Supervised Learning	Knn, LDA, DT/RF, neural nets	14
ML	Unsupervised Learning	Clustering, PCA	4
PR	Ethical considerations	Ethics in data applications	2

PR	On automation	Applications in society	1
PR	Continuing professional development	Communicating value of student skillsets	2
PDA	Algorithmic thinking and problem solving	Learning new methods	5
PDA	Programming	Coding efficiently and using software	3
PDA	Algorithms	Analyzing algorithms	3
PDA	Basic complexity analysis	Predicting run times and coding efficiently	2

Additional topics
Prescriptive analytics – reinforcement learning
Prescriptive analytics – optimization (linear programming and extensions)

Other comments
None