

Introduction to Data Science for Business

Course Syllabus

Course Description	Students will learn what Data Science is, what it takes to become a data scientist, and how businesses can profit from these set of technologies. In the first part of the course, the Data Science process will be explained along with the structure of the platforms and the tools used. In the second part, the methods and algorithms used for Data Science will be illustrated. In the third part, we will discuss how data science techniques can be applied to sample business problems.	
Prerequisites	Basic (introductory) knowledge of probability and statistics to understand the material. Basic (introductory) knowledge of one (any) computer programming language is also useful to understand the examples shown in class. Python/R knowledge will be a plus.	
Material and Conduct	There is no particular book which will be followed in the course. Lecture slides will be provided and followed in each lecture. Practical (hands-on) examples will be shown (using Python) in the class and some hands-on work will also be required of students. Course projects may be theoretical or hands-on based on some sample data and using the tool of choice (preferably Python) by each student group (of two or three). Each group will present its work to the class in the final two weeks.	
Grading and Deliverables	Final Exam:	50%
	Project and Presentation:	40%
	Class Attendance and Participation:	10%
Classroom Conduct	In accordance with university rules, and out of courtesy to your fellow students, eating, or talking with fellow students are not permitted in the classroom. Also, please make sure to turn off your cell phone during the class period.	
Instructor Background	Dr. Kemal Ünaltuna received his Ph.D. Degree in Computer Science/Neural Networks from Syracuse University, Syracuse, New York. He worked for several years at AT&T Bell Labs as a data scientist working closely with various business departments. After creating Yapı Kredi Bank's Business Intelligence environment and participating in many Business Intelligence/Data Mining projects there, he co-founded Intellica and later EVAM, two companies which operate in the domain of Business Intelligence and Big Data Analytics. As the lead consultant at Intellica, he participated in more than 50 projects at corporate companies in telecommunications, banking, insurance, e-business and other sectors. As an adjunct professor, he has taught Business Intelligence, Expert Systems, Data Mining and MIS classes at Marmara University and Boğaziçi University.	
Office Hours	By appointment, or right after the class. Feel free to contact me via email or mobile.	
Contacts	Email: kunaltuna@gmail.com, kunaltuna@intellica.net, unal.tuna@boun.edu.tr Mobile: 533 358 33 51	

Tentative Schedule

Lecture 1	Introduction to course, course conduct, prerequisites, requirements, projects, grading, etc.
Lecture 2	What is Data Science? What is Machine Learning? What is Data Mining? What is Data Engineering? How can these technologies be used for Business? In this class, we will understand the data science process and each step in this process. CRISP methodology will be discussed: Business Understanding, Data Understanding, Data Preparation, Algorithm Selection, Modeling, etc.
Lecture 3	What are the platforms & tools to be used for Data Science? Can you do it on your laptop? Or else, what are the preferred architectures and platforms? Short discussion on Data Warehouse, Data Lake, Big Data, Relational Databases, Hadoop distributions, Wrangling tools, Mining tools, Python, R, Apache Spark, Presentation tools, , etc.
Lecture 4	Data Science Prerequisites: Basic Statistics required for Data Science. We will overview and remind ourselves basic probability, data types, distributions and histograms, basic statistics, Bayes Rule, sampling, etc.
Lecture 5	Data Preparation (data engineering): What does it mean and what platforms can be used? What are the data types and formats used for Data Science, particularly for segmentation, association discovery, classification/prediction, nearest neighbor, etc. Data reduction techniques, categorical to numerical coding and vice versa, normalization, substitution. What to do with empty fields, bad distributions, etc.?
Lecture 6	Supervised Learning: What is learning? Who is the supervisor? What type of business problems can be addressed using supervised learning? What are the data requirements for supervised learning? What is labeled data? What is data set partitioning? How do we measure the performance of our models? What is a confusion matrix? Sampling or no sampling and how? What is oversampling? Discussion of generalization versus overfitting. Should we and how can we prevent overfitting?
Lecture 7	Supervised Methods in Data Science I - Decision Trees: What is a decision tree? How does it help us in decision making? What are the data requirements to build decision trees? What is splitting? What type of splitting rules are there? Which one should we use? How big should the decision tree grow? What are random forests?
Lecture 8	Supervised Methods in Data Science II - Regression and Neural Networks: Discussion of linear regression and multinomial regression. How is logistic regression different from linear regression? What is odds ratio? What is the logistic function? What are the limitations of regression methods? What is a neural network? How is a single layer neural network equivalent to logistic regression? What are multilayer neural network? What are the advantages of using those?
Lecture 9	Supervised Methods in Data Science III – Time permitting, we will discuss SVM, kNN and other methods: We will briefly introduce alternative methods in supervised learning like support vector machines, nearest neighbor and others.
Lecture 10	Unsupervised Methods in Data Science I - Clustering: What is unsupervised learning? What type of business problems can be addressed using unsupervised learning? Clustering as a statistical segmentation method. K-means versus hierarchical clustering. How many clusters are good? How do we measure clustering outputs?
Lecture 11	Unsupervised Methods in Data Science II: - Association Discovery: What associations are we talking about? Associating itemsets, movies, products, etc. What is market basket analysis? What are the metrics in association discovery?
Lecture 12	Application of Data Science to the Customer Segmentation Problem. What is customer segmentation as a business problem? What does it serve? Demographic, behavioral and value segmentation. Rule based and statistical segmentation. Hybrid segmentation. Macro and micro segmentation.
Lecture 13	Application of Data Science to the Customer Attrition/Churn Problem. How can we use supervised methods for preventing customer churn/retention? What are the data requirements for churn prediction? How successful can we be expected to be and how do we measure success?
Lecture 14&15	Project Presentations