**Big Data & Economics, כלכלה בעולם הביג-דאטה 1011.3193.01; Wed., 18:00-21:00, Berglas 012**

Lecturer: Dr. Roy Sasson ([roy.sasson@gmail.com](mailto:roy.sasson@gmail.com) ; office hours coordinated by email)  
Teaching Assistant: Elad Guttman (eladg9@gmail.com ; office hours coordinated by email)

The Eitan Berglas School of Economics, Faculty of Social Sciences, Semester A 2018-2019

**Description and goals of the course:**

This course will teach you how to analyze data and apply automatic machine learning methodologies during your career. The methodologies you will learn are applied in most modern firms today - from small startups you read about in the news, to giant retailers such as Walmart and Amazon, tech giants such as Google, Facebook and Airbnb, and Banking mammoths such as Citi, JP Morgan and others.

The aim of this course is to make you feel comfortable and familiar with analytics, economic and business problems and provide you with basic scripting experience that brings out business-economic insights and predictions from data.

The course was constructed for students without prior background in programming. Unlike courses such as Econometrics, no statistical or mathematical theory is involved in this course, but it does require a basic understanding of what is linear regression, t-statistic, R^2, some prior experience of working with data in software such as Stata, and feeling comfortable with reading text in English.

This course takes the approach of “learning by doing”, with minimum theory involved. Learning is done by the following flow:

1. Describing an economic/business challenge and the data involved (ie, “You have an online store. A new user enters your website. What product should the system offer her? At what price?”).

2. Learning a data analysis technique that addresses this challenge, and evaluating the solution’s success (ie – “automatically find similar users to the new user, based on past behavior. Offer the new user a product she is most likely to buy, bundle it, estimate the level of competition in the industry”).

3. Generalization - writing a program (“hands on work”) in R that utilizes the technique using a new data set and a new business/economic problem, with similar characteristics.

By the end of this course, you will know how to make data-driven economic decisions, forecasts and modeling. You will obtain hands-on experience in data analysis and modeling using R (the most commonly used tool by Data Scientists today, which is also a free, open source platform). You will become familiar with how automatic Big-Data analysis work-flows are used by firms to optimize their performance, and get exposed to free online data resources.

**Units: 3 hours**

**Grading:**

1. Homework assignments (applying data analysis techniques & writing code in R): 30%
2. Concluding project (students will be graded based on the performance of their model): 70%

**Prerequisites: Passing grades in Statistics A,B & Econometrics.**

**Note: Although I will be speaking in Hebrew, all materials of this course (presentations, R code etc) are in English. You should register to this class only if you feel comfortable reading English texts, and have a laptop you can bring to all classes.**

**Course Layout (materials might change, depending on progress):**

|  |  |  |
| --- | --- | --- |
| **Unit** | **Data Analysis & ML tools we will learn** | **Business Application** |
| 1 | Exploring data and basic summary statistics using R.  Understanding the basic data operations: clean, merge, aggregate, group by and plot. | Making sense of data: understanding your customers and business environment using very basic summary statistics and pivot tables. This will account for about 40% of the course. |
| 2 | Linear Regression | Store Sales Forecasting |
| Wine Price Forecasting |
| Assessing abnormal Stock Returns: the Impact of special economic events on stock indices |
| 3 | Logistic Regression | Online marketing: predicting clicks on ads |
| Modeling credit risk (load default probability) |
| Moneyball: predicting NBA playoff participation |
| 4 | Classification Trees | Predicting travelers’ bookings. |
| Modeling a medical expert: Logistic Regression vs. Classification Trees. |
| 5 | Natural Language Processing | Sentiment Analysis of online user reviews |
|
| 6 | Clustering: the K-Means algorithm | Recommendations systems: Clustering customers into segments, based on implicit rankings. |
| 7 | Collaborative Filtering (“Users who bought this, also bought..) |
| 8 |  | Introduction to Big-Data technologies and how they work in practice |

Reading: Data Mining - Concepts and Techniques by Jiawei Han, Micheline Kamber, Jian Pei