The Hebrew University of Jerusalem

Syllabus

A Needle in a Data Haystack: Introduction to Data Science - 67978

Last update 05-11-2018

<u>HU Credits:</u>   3

<u>Degree/Cycle:</u> 2nd degree (Master)

<u>Responsible Department:</u> Computer Sciences

<u>Academic year:</u> 0

<u>Semester:</u> 1st Semester

<u>Teaching Languages:</u> English and Hebrew

<u>Campus:</u> E. Safra

<u>Course/Module Coordinator:</u> Dr. Dafna Shahaf

Coordinator Email: dshahaf@cs.huji.ac.il

<u>Coordinator Office Hours:</u> TBA

<u>Teaching Staff:</u>
  Prof Dafna Shahaf

<u>Course/Module description:</u>
  Data Science is an interdisciplinary field that deals with finding patterns in data. With the ever increasing amounts of digital data, the need for automated methods for data analysis is growing rapidly.
Data science employs techniques from many areas, including statistics, machine learning and databases. It has a wide range of applications, from science and technology through business and society.

<u>Course/Module aims:</u>

<u>Learning outcomes - On successful completion of this module, students should be able to:</u>

<u>Attendance requirements(%):</u>
  0

Teaching arrangement and method of instruction:

<u>Course/Module Content:</u>
  Tentative list of topics to be covered:

* Useful background: Statistical inference

* Similar Items, Distance Measures, Locality-Sensitive Hashing, Similarity-Preserving Summaries
- Text similarity measures and text processing

* Clustering, Hierarchical Clustering, Non-Euclidean Spaces

* Graph Analysis
- Social Networks, community detection, triangles, small world

* Recommendation Systems

Data Exploration, Visualization, and Feature Engineering

Experimental design
- Randomized trials vs. observational studies, causality

MapReduce, Hadoop

Dimensionality Reduction

Mining Data Streams

Required Reading:
 TBA

 Additional Reading Material:

 Course/Module evaluation:
 End of year written/oral examination 0 %
 Presentation 0 %
 Participation in Tutorials 5 %
 Project work 0 %
 Assignments 40 %
 Reports 0 %
 Research project 55 %
 Quizzes 0 %
 Other 0 %

Additional information:
 Required: Knowledge of at least one programming and/or scripting language.
Basic knowledge of algorithms and probability.