

The Hebrew University of Jerusalem

Syllabus

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

Last update 27-09-2024

HU Credits: 3

Degree/Cycle: 1st degree (Bachelor)

Responsible Department: Computer Sciences

Academic year: 0

Semester: 2nd Semester

Teaching Languages: English

Campus: E. Safra

Course/Module Coordinator: Prof. Dafna Shahaf and Dr. Tom Hope

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

Coordinator Office Hours: TBA

Teaching Staff:

Prof. Dafna Shahaf

Course/Module description:

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.

Course/Module aims:

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

Understand what type of tools (and data) they need to approach a problem.

Attendance requirements(%):

0

Teaching arrangement and method of instruction:

Course/Module Content:

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, text embedding, contrastive learning
- * Language models
 - BERT, GPT, T5
- * Information extraction

-
- * *Clustering, Hierarchical Clustering, Non-Euclidean Spaces*
 - * *Graph Analysis*
 - *Social Networks, community detection, triangles, small world, graph embedding, link prediction*
 - * *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:

Grading Scheme:

Essay / Project / Final Assignment / Home Exam / Referat 75 %
Submission assignments during the semester: Exercises / Essays / Audits / Reports / Forum / Simulation / others 15 %
Mid-terms exams 10 %

Additional information:

Required: This is a grad class, or at least third year of BSc. Knowledge of at least one programming and/or scripting language.
Basic knowledge of algorithms and probability.
It is highly recommended to take an ML class before taking this one.