האוניברסיטה העברית בירושלים

סילבוס

מבוא למדע הנתונים - 51127

תאריך עדכון אחרון 02-08-2023

נקודות זכות באוניברסיטה העברית: 5

תואר: בוגר

היחידה האקדמית שאחראית על הקורס: פסיכולוגיה

השנה הראשונה בתואר בה ניתן ללמוד את הקורס: 0

סמסטר: סמסטר ב'

שפת ההוראה: עברית

קמפוס: הר הצופים

מורה אחראי על הקורס (רכז): Yuval Hart

דוא"ל של המורה האחראי על הקורס: yuval.hart@mail.huji.ac.il

שעות קבלה של רכז הקורס: בתיאום עם המרצה

מורי הקורס:

ד"ר יובל הרט

תאור כללי של הקורס:
Data science combines data, analytics, and algorithms to gain insights and make useful predictions. This course will take a holistic approach to helping students understand the key factors involved, from data collection and exploratory data analysis to modeling, evaluation, and communication of results. Working on case studies in teams will provide students with hands-on experience with the data science process using state-of-the-art analysis tools. Emphasis will be given to the strengths, trade-offs, and limitations of each method to highlight the importance of merging analytical skills with critical quantitative thinking.

מטרות הקורס:
Provide students with concrete and tangible tools towards data science acumen. Students will learn the strengths, trade-offs, and limitations of each method to highlight the importance of merging analytical skills with critical quantitative thinking.

תוצרי למידה
בסיומו של קורס זה, סטודנטים יהיו מסוגלים:
After successful completion of this course you will be able to:
Explain and apply the data science process to analytics problems
Recognize ethical issues involved in applying the data science process
Learn to use basic Python and Colab notebooks for data analysis
Articulate core objectives and quantitative goals of the data science process
Understand and prepare different types of data for analysis
Use summary statistics and visualization to quickly explore data
Apply descriptive and predictive models from statistics and machine learning to data
Conduct detailed assessments of modeling approaches using quantitative metrics
Effectively communicate the outcome of data analysis with visual data stories
Work effectively in heterogeneous teams on a data science project of their choosing

דרישות נוכחות (%):
100

שיטת ההוראה בקורס: The course will employ methods of modern pedagogy - most of the classwork will be based on hands-on teamwork analysis of case studies and problem-sets. A combination of instructor guidance and peer instruction will facilitate students' ability to analyze different datasets, outline the limitations of

their analysis, and suggest paths for future analysis.

Case Method: Following the HBS Case Method practice aligns with the learning goals for this course. The nature of cases fits well for a course that aims at engaging students with a variety of data science problems from different disciplines. Each week will present a different real-world case study. Through case studies, students are developing and learning ideas in discussion with the instructor, exchanging perspectives, countering and defending points, and building on each other's ideas. Working in teams encourages transfer of knowledge and tools between students, and probing students viewpoints through polling questions allows to scale up the learning in the course.

Active Learning: Students will gain hands-on experience with computational data science methods using active learning. A dominant part of this course will be spent in hands-on computational activities. Due to the large size of the class these activities will be structured as small group work in a lab format, where students work together to solve data science problems that can be tackled with multiple viable approaches. Since this course has no prerequisites these activities will be structured carefully through tutorials and guided exploration using specially developed Jupyter notebooks.

Retrieval and Spacing: Core concepts of the course, such as the importance of the chosen data and its features, biases, metrics for evaluation, and analysis trade-offs -repeat throughout the course in different formats and instances. Moreover, the lab like format of the course allows to introduce many low-stakes testing events that strengthen students' learning and increase retention of knowledge for the long run.

Team based learning: Data science is a team effort, often involving experts from computer science, statistics, and various domain disciplines (e.g., business, science, or medicine). To achieve learning goal (10), students will work in small, diverse teams. Specific tools for team building, project management, and peer evaluation (see below) will support the work in teams. Experience from prior courses with similar focus on team projects suggests the need for designing appropriate grading rubrics and using online tools for peer evaluation.

Peer Assessment: Assessment of student participation in case studies, computational activities, and projects for a large class is challenging. Specific peer evaluation strategies and tools will be used to allow students to assess other members of their team as well as themselves. Peer assessment data will be used in assigning individual grades for the course.

<u>רשימת נושאים / תכנית הלימודים בקורס:</u>
Data collection and Exploratory data analysis
Data cleanup and sampling, missing data, data biases, data quality
Probability and measurement

Geometrical intuitions for data analysis
Clustering methods - supervised vs. unsupervised
Dimensional reduction
Predictive models (Introduction to machine learning)
Models evaluation and selection

חומר חובה לקריאה:
לפי מטלות הקורס

חומר לקריאה נוספת:

מרכיבי הציון הסופי :
הגשת עבודה מסכמת / פרויקט גמר / מטלת סיכום / מבחן בית / רפרט 70 %
מטלות הגשה במהלך הסמסטר: תרגילים / עבודות / מבדקים / דוחות / פורום / סימולציה ואחרות 30 %

מידע נוסף / הערות: