

## סילבוס קורס

כלים מתקדמים במדעי הנתונים  
4077999

### פרטי הקורס

<b>שנה אקדמית:</b> תשפב	<b>קמפוס:</b> באר שבע
<b>סוג הקורס:</b> בחירה	<b>מחלקה:</b> הנדסת תוכנה
<b>רמת הקורס:</b> תואר ראשון	<b>תחום:</b> כריה ואחזור מידע
<b>סמסטר:</b> ב	<b>שנת לימוד:</b> ד'
<b>נקודות זכות:</b> 3	<b>דרישות קדם:</b>
<b>נקודות ECTS:</b> 4.5	<b>דרישות במקביל:</b>
<b>צורת העברה:</b> פנים אל פנים, Project oriented.	<b>שפת הוראה:</b> עברית
<b>מתרגל/ים:</b>	<b>סביבת עבודה:</b> מעבדה
	<b>מרצה/ים:</b> מר יניב הדר
	<a href="mailto:yanivha6@ac.sce.ac.il">yanivha6@ac.sce.ac.il</a>

### מטרה

הרחבת הידע הנלמד בקורסים השונים במסלול מדעי הנתונים. לימוד גישות ואלגוריתמים העונים על האתגרים בעבודה עם נתוני ענק.

## תפוקות למידה

- עם סיום מוצלח של הקורס, הסטודנטים יהיו מסוגלים:
1. לאסוף, לאחסן ולייצג נתוני ענק לפני תחילת עיבוד מקדים.
  2. להציג וויזואליזציה של נתונים ומודלים.
  3. להתאים ולממש שיטה משולבת לסיווג נתונים מורכבים ואטומיים בקנה מידה ענק
  4. להתאים ולממש שיטה משולבת לקיבוץ נתונים מורכבים ואטומיים בקנה מידה ענק
  5. להכין נתונים ומודלים מנתונים מבוססי גרפים.
  6. להכין נתונים ומודלים מנתונים מבוססי זמן וסדרות עיתיות.
  7. להכין נתונים ומודלים המבוססים על עיבוד מרחבי.
  8. לבצע תהליך הערכה למודל והאתגרים הכרוכים בהערכת מודלים מנתוני ענק

## תוכן הקורס

שבוע	נושא	מקורות רלוונטים
1	איסוף המידע מהרשת והכנתו, CAP Theoram ו MongoDB	[1] פרקים 1-8
2	וויזואליזציה של נתונים – מעבדה בטבלאו וכלי וויזואליזציה נוספים	[1] פרקים 1-8
3	הנחיה אישית של הפרויקט	[3] פרק 2
4	מערכות קבצים עצומות ו Map-Reduce	[2] פרק 2
5	שיפורים לאלגוריתמי סיווג ואישיכול להתמודדות עם כמויות ענק של נתונים	[2] פרק 4
6	הנחיה אישית של הפרויקט	
7	כריית מידע בגרפים	[2] פרק 10
8	נתונים מבוססי זמן וסדרות עיתיות	[2] פרק 6
9	עיבוד מרחבי ב R	[7] פרקים 1-8
10	הנחיה אישית של הפרויקט	
11	האקתון	
12	הצגת פרויקטים	

## מקורות ספרות נדרשים ומומלצים

- ספר הקורס:
1. Charu C. Aggarwal, Social Network Data Analytics, Springer Science & Business Media, 18 Mar 2011
- מקורות נוספים:
2. [Jure Leskovec](#) , [Anand Rajaraman](#) , [Jeffrey D. Ullman](#), *Mining of Massive Datasets*
  3. Charu C. Aggarwal, ChengXiang Zhai, Mining Text Data, Springer Science & Business Media, 2012
  4. [Christopher D. Manning](#), [Prabhakar Raghavan](#), [Hinrich Schutze](#), *An Introduction to Information Retrieval*, Cambridge University Press, 2009
  5. [Han Jiawei](#) and [Micheline Kamber](#). *Data Mining: Concepts and Techniques, Second Edition*, Elsevier, 2006
  6. G?ron, Aur?lien. Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. O'Reilly Media, Inc., 2017
  7. Brett Lantz , Machine Learning with R, Packt Publishing 2013

8. Michael Dorman, Learning R for Geospatial Analysis, Packt Publishing 2014

פעילויות למידה מתוכננות ושיטות הוראה

שעות מעבדה שבועיות: 3. מספר מפגשים פרונטליים: 8-10.  
תכנית התנסות מעשית מתקיימת בקורס.  
הקורס יילמד במסגרת המעבדה. במהלך הקורס יינתנו מעבדות תכנותיות. הסטודנטים יכינו את הפרויקט בזוגות, יגישו דוחות ויציגו אותו בכיתה.  
ההוראה בקורס הינה פרונטאלית.

שיטות הערכה וקריטריונים

הערות	אחוז	קריטריון
חובת מעבר בציון 56 ומעלה בפרויקט.	40%	פרויקט:
	20%	האקתון:
יינתנו כ- 5 דוחות מעבדה להגשה. חובת מעבר בציון 56 ומעלה בממוצע הדוחות.	10%	דוחות:
	10%	הערכת מרצה:
חובת נוכחות בכל המפגשים בקורס. בהיעדרות הציון יחושב באופן יחסי. במפגשים בזום נדרשת מצלמה פתוחה.	20%	נוכחות:
במידה וציון הפרויקט ו/או ציון ממוצע הדוחות נמוך מ- 56, הציון הסופי בקורס הינו הנמוך מבין השניים.		הערות: