

סילבוס – תשפ"ד (2023-2024)

פרטי הקורס	
שם הקורס: אנליזה של ביג דאטה / Big Data Analysis	קוד הקורס: 110200303-241
שם החוג: מדעי המחשב - תואר ראשון	קמפוס: נביאים
שנה לתואר: ג	סמסטר: ב
סוג הקורס: רשות	מבנה הקורס: שיעור ותרגיל
שפת הוראה: עברית	מספר נ"ז: 3
ימים ושעות הלימוד: יום רביעי, 17:00 – 19:30	
צורת ההוראה בקורס: 2. שיעור פרונטלי, המשלב שימוש בכלים דיגיטליים בהיקף פחות מ-30%	
מידע נוסף/ הערות:	

פרטי המרצה
שם המרצה: חננאל פרל
פרטי התקשרות: chananelpe@edu.hac.ac.il

שעת קבלה: קבלה פרונטלית בכיתה – מייד אחרי השיעור ביום רביעי.
פגישה מקוונת בזום – לפי תאום מראש במייל.

מידע אקדמי

דרישות קדם:

תכנות בסיסי בפייטון

תיאור הקורס:

אנו חיים כיום בעידן בו מצטברת סביבנו כמות עצומה של מידע (Big Data) במגוון רחב של מקורות כגון: במערכות מידע של הארגון, מערכות טכנולוגיות עסקיות, מערכות ניטור ורשתות חברתיות. כמעט כל פעילות שלנו מייצרת פריט מידע שנשמר בצורה דיגיטלית. ניתוח נתונים אלו יכול להביא לתובנות חדשות והחלטות מיטביות בעולם העסקי, הציבורי או המחקרי.

בקורס זה, נלמד את העקרונות והטכניקות המעשיות של אנליזה וניתוח של ביג דאטה. נכיר את סוגי הנתונים השונים, ואת הכלים והפלטפורמות שאפשר להשתמש בהם כדי לאסוף ולאחסן, לעבד ולנתח את הנתונים.

נלמד את שפת SQL המהווה בסיס לניתוח ותשאול נתונים.

נלמד ונכיר את מערכת HADOOP ואת סכמת MapReduce, ונמשיך לטכנולוגיית Apache Spark, שהיא כיום אחת מהטכנולוגיות הנפוצות ביותר לצורך עיבוד נתונים בסביבת ביג דאטה.

נלמד את יסודות NoSQL ונתנסה עם MongoDB.

בנוסף נכיר חבילות פייטון נוספות של Visualization, Scraping, Pandas ועוד..

מטרות הקורס ופירוט תפוקות הלמידה:

עם סיום הקורס בהצלחה, הסטודנטיות/ים יוכלו:

1. להבין ולתאר את הכלים הנפוצים לאנליזה של ביג דאטה לדעת ולהבחין בין הסוגים והשימושים של כלים אלו.
2. לקרוא ולהבין, לתכנן ולכתוב שאילתות מורכבות בשפת SQL
3. לקרוא ולהבין, לתכנן ולכתוב שאילתות עבור NoSQL
4. לעצב ולבנות תוכנות קצרות ושאילתות במערכת Spark-SQL / Spark
5. למצוא דאטה עבור התרגיל, לנתח ולארגן אותו, ולבסוף לעצב ולהציג התובנות בכלי ויזואליזציה
6. להראות ולתאר את התרגיל מול הכיתה

דרישות הקורס:

השתתפות בשיעורים

חזרה על החומר

הגשת תרגילים

הצגת התרגיל המסכם בכיתה

דרישות נוכחות: מומלץ מאד להיות נוכח ופעיל בהרצאות הקורס, אך אין חובת נוכחות. יש תוספת ציון (בנוסף) למי שאכן יהיה נוכח פעיל בשיעור (עיון סעיף הבא - הערכת הקורס)

הערכת הקורס – הרכב הציון הסופי:

תרגילים: בקורס צפויים להינתן כ-8 תרגילי בית. הגשת התרגילים היא חובה. הגשת התרגילים היא בזוגות. בשקלול רכיב התרגילים בציון הסופי ילקחו ציוני כל התרגילים. משקל התרגילים בציון הסופי הינו 100%.

התרגילים בחציו השני של הסמסטר יהיו תרגיל מתגלגל שלמעשה יורחב ויתפתח לכדי תרגיל מסכם. בשיעור האחרון התרגיל המסכם יוצג בכיתה במשך כ-10 דקות לזוג (כל אחד מתוך הזוג יציג למשך 5 דקות). הצגה זו של התרגיל תשוקלל בתוך הציון הסופי של התרגילים.

בחינה מסכמת: לא תתקיים בחינה מסכמת.

נוכחות: על כל נוכחות פיזית פרונטלית פעילה בשיעור, תנתן חצי (0.5) נקודה בונים לכל שיעור, ניתן לצבור מקסימום 6 נקודות בונים בכל הסמסטר.

תרגילים - מידע כללי:

הגשת התרגילים היא חובה. הגשת התרגילים היא בזוגות. בעבודה על פתרון התרגיל ניתן להתייעץ עם עמיתים אך העבודה חייבת להיות אישית. הגשה סדירה של התרגילים חשובה לשליטה טובה בחומר.

לא תתקבל הגשת תרגילים באיחור ללא אישור מראש.

מילואים ומחלה אינם מזכים בפטור אוטומטי מהגשת תרגיל. במקרים של אשפוז, מילואים, שמירת הריון, לידה, חתונה, שבעה ניתן להגיש בקשה לדחייה לכתובת המייל של המרצה ולהציג אישור מתאים (טופס 3010, או

תעודת מחלה וכד). לאחר הבקשה המרצה יבחן ויתאם עם התלמיד את התרגילים או חלקי התרגילים אותם יש להגיש, ואת אלה מהם התלמיד פטור.
איחור בהגשת תרגיל (ללא אישור המרצה), איחור ביום אחד – הפחתה של 4 נקודות. איחור ביומיים – הפחתה של 8 נקודות. לאחר מכן לא ניתן להגיש.
ערעור על ציון תרגיל יעשה בפנייה למייל של בודק התרגילים. מייל הבודק מפורסם באתר המודל. ערעור על ציון תרגיל יעשה לכל היותר עשרה ימים לאחר פרסום הציון. לא יתקבלו ערעורים לאחר מועד זה!

ציון עובר בקורס: 55

מידע נוסף/ הערות:

נושאי הלימוד ופירוט הלמידה הפעילה בהלימה לתפוקות הלמידה

נושא	נושאי הלימוד	פרקי הקריאה (חובה ורשות)	תפוקות למידה לשיעור	תיאור למידה פעילה
SQL	מבוא, שפה הצהרתית שאלות SQL בסיסי, בינוני ומתקדם	https://www.sqlite.org/lang.html https://dataschool.com/how-to-teach-people-sql/ https://mode.com/sql-tutorial/		למידה מבוססת מקרים בעיות ודוגמאות
JSON	מבוא	https://www.json.org/json-en.html		
NoSQL	שאלות Document DB JSON, BSON	https://www.mongodb.com/docs/manual/introduction/		למידה מבוססת מקרים בעיות ודוגמאות
HADOOP	HDFS, MapReduce	https://hadoop.apache.org/		הרצאה ושאלות
SPARK	RDD, DataFrame, DataSet, Spark-SQL	https://spark.apache.org/docs/latest/		הרצאה ושיתוף
ויזואליזציה	גרפים ודיאגרמות	לדוגמא: https://matplotlib.org/		

ביבליוגרפיה

קריאת חובה:

קריאת רשות:

- Modern Database Management – Jeff Hoffer, Ramesh Venkataraman, Heikki Topi ([link](#) | [toc](#) | [full PDF](#))
- Mining of Massive Datasets – Jure Leskovec, Anand Rajaraman, Jeff Ullman – (<http://www.mmds.org/> | [full PDF](#))
- Database System Concepts – Avi Silberschatz, Henry F. Korth, S. Sudarshan (<https://www.db-book.com/> | [toc](#))
- SQL and Its Applications – Raymond A. Lorie, Jean-Jacques Daudenarde ([link](#))
- SQL for Smarties: Advanced SQL Programming – Joe Celko ([link](#) | [full PDF](#))

חומר קריאה נוסף:

אתרי אינטרנט מומלצים:

<https://www.sqlite.org/quirks.html>

https://www.sqlite.org/lang_corefunc.html

https://www.sqlite.org/lang_aggfunc.html

<https://www.sqlite.org/windowfunctions.html>

<https://www.sqlite.org/json1.html>

<https://dataschool.com/learn-sql/>

<https://www.mongodb.com/docs/manual/reference/sql-comparison/>

<https://www.mongodb.com/docs/manual/reference/sql-aggregation-comparison/>

<https://spark.apache.org/docs/latest/sql-ref.html>

חומרי עזר:
לקריאה נוספת: