

תואר ראשון

1221-7021- טיפול יישומי בנתונים

סמסטר א' תשפ"ג

קבוצה	יום בשבוע	שעה	תאריך בחינה	מרצה	דואר אלקטרוני
01	ה	15:30-18:00	כמפורט בלוח הבחינות	אליה ויגודה	elyawygoda@mail.tau.ac.il

שעת קבלה – בתיאום מראש

היקף הלימודים

היקף 3 ש"ס

ECTS 4 = 2 ש"ס – ECTS (European Credit Transfer and Accumulation System), ערך הניקוד של הקורס במוסדות להשכלה גבוהה בעולם שהינם חלק מ"תהליך בולוניה".

קורסי קדם נדרשים

סטטיסטיקה לניהול (בציון סף 78) (12215114) או סטטיסטיקה ב' (בציון סף 78) (10711117) או סטטיסטיקה לכלכלנים ב' (בציון סף 78) (10111109).

תיאור הקורס

תחומי ה-Data Science וה-Business Data Analytics כוללים טכנולוגיות ושיטות המאפשרות איסוף ועיבוד כמויות עצומות של נתונים ממגוון רחב של מקורות שאינם בהכרח מאורגנים. נתונים אלו, המכונים באופן פופולרי Big Data מאפשרים לגופים עסקיים קבלת תובנות עסקיות חשובות ושיפור משמעותי בביצועים העסקיים. שלב הכרחי בתהליכי ה-Data Science הינו שלב ה-Data Processing, שכולל השגת נתונים גולמיים ממקורות פנימיים וחיצוניים, עיבוד וארגון הנתונים, המרתם לפורמטים המתאימים לניתוח, וניהול הנתונים. נתונים גולמיים אלו יכולים להמצא במסגרת הארגון, באינטרנט, במאגרי מידע, או בידי גופים אחרים. הנתונים יכולים להופיע בצורה מאורגנת או בלתי מאורגנת (unstructured data), בפורמטים שונים, חלקם סטנדרטיים וחלקם יחודיים. קורס זה יתמקד בהבנת תהליך השגת הנתונים ממקורות שונים, לימוד כלים בסיסיים בתכנות (שפת Python) הכרות עם פורמטים ו-API מקובלים באינטרנט לצורך "שאיבת נתונים", היכרות עם בסיסי נתונים רלציוניים ושפת SQL לצורך ניהול נתונים.

תפוקות למידה

הכישורים הצפויים עם סיום הקורס בהצלחה:

1. הבנת תהליך השגת הנתונים ממקורות שונים.
2. תכנות בסיסי (בשפת Python).
3. "שאיבת נתונים" מהאינטרנט תוך שימוש בפורמטים API מקובלים.
4. הכרות עם בסיסי נתונים רלציוניים ושפת SQL לניהול נתונים.
5. הבאת נתונים לתצורה המתאימה לניתוח מתקדם בהמשך (כפי שנלמד בקורסים אחרים בפקולטה).

הערכת הסטודנט בקורס והרכב הציון

אחוז	מטלה	תאריך	גודל קבוצה/ הערות
30%	תרגילי בית		
70%	מבחן		

לקבלת ציון עובר בקורס יש לקבל ציון עובר בכל אחד ממרכיבי הציון הסופי - ובכל אחד מהתרגילים (כל התרגילים הם חובה)

- עפ"י תקנון האוניברסיטה תלמיד חייב להיות נוכח בכל השיעורים (סעיף 5).
- תלמיד, הנעדר משיעור המחייב השתתפות פעילה או שלא השתתף באורח פעיל, רשאי המורה להודיע למזכירות כי יש למחוק את שמו מרשימת המשתתפים. (התלמיד יחויב בתשלום בגין קורס זה)
- תנאי הכרחי לקבלת ציון עובר (60 לפחות) בקורס - הינו קבלת ציון 60 לפחות בבחינה, סטודנט שייכשל בבחינה - ציונו הסופי בקורס יהיה ציון הבחינה.

פירוט המטלות בקורס

בקורס ינתנו כחמישה תרגילי בית אותם יש להגיש במהלך הסמסטר לפי המפורט בהמשך.

מדיניות שמירה על טווח ציונים

בחוג לניהול מונהגת מדיניות שמירה על טווח ציונים. מדיניות זו מתייחסת לממוצע הציונים הסופיים בקורס מידע בנושא זה מתפרסם בהרחבה באתר החוג לניהול, בסעיף ציונים בתקנון.

הערכת הקורס ע"י הסטודנטים

בסיומו של הקורס הסטודנטים ישתתפו בסקר הוראה על מנת להסיק מסקנות לטובת צרכי הסטודנטים והאוניברסיטה.

אתר הקורס

אתר הקורס יהווה המקום המרכזי בו ימסרו הודעות לסטודנטים, לפיכך מומלץ להתעדכן בו מדי שבוע, לפני השיעור, ובכלל - גם בתום הסמסטר. (לצורך תיאום ענייני הבחינה למשל). שקפי הקורס יהיו באתר הקורס באתר. לתשומת לבכם - בכיתה ידונו גם נושאים (ובפרט דוגמאות) שאינם מופיעים בשקפים או מופיעים בכותרת בלבד. כל אלו הינם חלק בלתי נפרד מחומר הקורס.

נושא	תאריך	נושאים	קריאת עזר	הגשות
1		Data Processing as a Part of the Business Data Analytics Process Introduction to Data Science. The benefits of using Big Data for business insights. The data science process. Working methodologies in Data Science, CRISP-DM cycle, ETL – Extract, Transform, Load. Python – most commonly used recent technology tool for writing data related applications: Introduction, writing a first program and script.	- Data Science for business chapters 1,2. - Chapman et al. 2000 Learning Python chapters 1-3	
2		Python . Programing entities: variables, data types, dynamic typing, casting, expressions. Flow Control: conditional statements, iterations over data, loops.	Learning Python chapters 4, 6, 12-14	תרגיל 1
3		Python . Working methodologies in writing code. Memory model Functions.	Learning Python chapters 9, 16-18	
4		Python . Advanced data types and structures: tuples, lists, dictionaries. Strings fundamentals for text processing. Sorting.	Learning Python chapters 4, 7	תרגיל 2
5		Processing Textual-based Data Sources Using structured and unstructured textual sources for obtaining business data. Regular expressions as a tool for retrieval data from textual sources. Dictionaries and dictionary-based sentiment analysis example.	Learning Python chapter 7	תרגיל 3
6		Retrieving Big Data from External Sources Basic web scraper – for scraping online product information, reviews, and competitive information.	Relevant materials and pointers will be distributed in class	תרגיל 4
7		Managing and Storing Data Introduction to Relational Databases and Structured Query Language: Tables, Keys, Queries, Joins, Aggregation.	Database Systems chapters 2, 6	תרגיל 5

הערות:

- חלק משמעותי בקורס הינו תרגול ולימוד עצמי בנוסף לתרגולים השבועיים שינתנו בכיתה. משך לימוד רוב הנושאים יותר מהרצאה אחת. יתכן שחלק מההרצאות יוחלפו בהרצאות אורח או בנושאים עדכניים.

*התכנית הינה בסיס לשינויים. יתכנו כשלוש שעות מפגשי השלמה/תגבור נוספים במהלך הקורס במועדים שיקבעו ע"י המרצה במהלך הסמסטר.

קריאת רשות

- Chapman, Pete, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, and Rudiger Wirth. "CRISP-DM 1.0 Step-by-step data mining guide." (2000).
- Garcia-Molina, H., Ullman J. D., Widom J., "Database Systems: The Complete Book", Pearson Prentice Hall, 2009, 2002.
- Lutz, M., "Learning Python", 5th edition, O'Reilly Publisher.
- Provost, F., Fawcett, T., "Data Science for Business", O'Reilly Publisher, 2013