

תואר ראשון

7021-1221- טיפול יישומי בנתונים

סמטר א' תשפ"א

קבוצה	יום בשבוע	שעה	תאריך בחינה	מרצה	דואר אלקטרוני
01	ג	12:00-14:30		דביר נתנאלי	dnetanely@tau.ac.il

שעת קבלה: בתיאום מראש

שיעורי תיגבור: נקבעו שני שיעורי תיגבור (ללא חובת נוכחות) לטובת חזרה על החומר ב-6.11 (12-10) וב-27.11 (13-11).

היקף הלימודים

היקף 3 ש"ס

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

תיאור הקורס

תחומי Data Science ו- 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 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 בלימוד כלים בסיסיים בתכנות (שפת Python). כמו כן נעסוק בהבנת תהליך השגת הנתונים ממקורות שונים, בהכרות עם פורמטים API- מקובלים באינטרנט לצורך "שאיבת נתונים", בהיכרות עם בסיסי נתונים רלציוניים ושפת SQL לצורך ניהול נתונים.

תפוקות למידה

עם סיום הקורס בהצלחה יוכל הסטודנט:

1. הבנת תהליך השגת הנתונים ממקורות שונים.
2. תכנות בסיסי (בשפת Python).
3. "לשאוב נתונים" מהאינטרנט תוך שימוש בפורמטים ו API מקובלים.
4. הכרות עם בסיסי נתונים רלציונים 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
70%	תרגילי בית		7 תרגילי בית
30%	מטלת סיום		

* 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 במטלת הסיום.

פירוט המטלות בקורס

בקורס ינתנו כשבעה 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 עינייני 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.

תכנית הקורס *

נושא	תאריך	נושאים	קריאת עזר	הגשות	הערות
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	תרגיל 2 תרגיל 3	

	תרגיל 4 תרגיל 5	Learning Python chapters 4, 7	Python. Advanced data types and structures: tuples, lists, dictionaries. Strings fundamentals for text processing. Sorting.		4
	תרגיל 6	Learning Python chapter 7	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.		5
	תרגיל 7	Relevant materials and pointers will be distributed in class	Retrieving Big Data from External Sources Basic web scraper – for scraping online product information, reviews, and competitive information.		6
	תרגיל 8	Database Systems chapters 2, 6	Managing and Storing Data Introduction to Relational Databases and Structured Query Language: Tables, Keys, Queries, Joins, Aggregation.		7

הערות :

חלק משמעותי בקורס הינו עבודה על משימות התיכנות שיינתנו כתרגילי בית 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.

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). <http://the-modeling-agency.com/crisp-dm.pdf>

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