



מדעי הנתונים לניהול בשפת Python

ד"ר דרור כהן | המחלקה לניהול

מס הקורס 55066 | Introduction to Data Science for Management

הרצאות משולבות תרגילים במעבדת המחשבים

2

תשפ"ה

סוג הקורס:

היקף נ"ז:

שנת לימודים:

סמסטר:

יום ושעה

שעת קבלה:

מייל מרצה:

dror.cohen.sh@gmail.com



תיאור הקורס ומטרות למידה

תקציר הקורס

הקורס הינו קורס מבוא למדעי הנתונים עבור תלמידי תואר ראשון בניהול. במהלך הקורס יילמדו אלגוריתמים ושיטות עבודה לניתוח נתונים, וויזואליזציה ובנית מודלים סטטיסטיים בלמידת מכונה.

מטרות/תוצרי הלמידה

ידע

1. הלומדים יגדירו את סוג האלגוריתם המתאים לשימוש, את משתני המטרה, ואת הצעדים הנדרשים לניתוח נתונים ו/או בנית מודלים על מנת לפתור בעיות שונות במדעי הנתונים.
2. הלומדים יתארו את אופן פעולתם של אלגוריתמים נבחרים בלמידת מכונה.

מיומנויות

הלומדים יכתבו קוד בתרגילים שבועיים להגשה בשפת python ובעזרתו ינתחו מערכי נתונים, ויבנו מודלים של למידת מכונה במגוון משימות.



למידה פעילה - תכנון מהלך השיעורים:

מס' השיעור	נושא השיעור	למידה פעילה/ הערכה תהליכית	קריאה/ צפיה נדרשת
1	היכרות עם סוגי משימות במדעי הנתונים והגדרת משתני מטרה. ביצוע אבחנה בין משימות למידה מונחית: (סיווג, רגרסיה, וניתוח סדרות זמן מונחית: supervised learning: למידה בלתי מונחית (ניתוח אשכולות, זיהוי אנומליות un-supervised learning) למידת חיזוק (Reinforcement learning) ודוגמאות לאלגוריתמים מתאימים.	הגשת תרגיל בזוגות – הגדרה ובנייה של משתני מטרה.	
2	חקר נתונים ראשוני (Exploratory Data Analysis – EDA) חלק א' - וויזואליזציה של התפלגות משתנים, בניית משתני דמה, התמודדות עם ערכים חסרים.	הגשת תרגיל בזוגות – ביצוע חקר נתונים ראשוני בשפת python, חלק א'	
	חקר נתונים ראשוני (Exploratory Data Analysis – EDA) חלק ב' - קורלציה בין משתנים. הורדת מימד באמצעות PCA, t-SNE.	הגשת תרגיל בזוגות – ביצוע חקר נתונים ראשוני בשפת python, חלק ב'	
3	ניתוח אשכולות (clustering) חלק א, אלגוריתם k-means.	הגשת תרגיל בזוגות – יצירת אשכולות במערכת נתונים בשפת python, חלק א'	
4	ניתוח אשכולות (clustering) חלק ב', אלגוריתם dbSCAN.	הגשת תרגיל בזוגות – יצירת אשכולות במערכת נתונים בשפת python, חלק ב'	

5	זיהוי אנומליות (anomaly detection) , אלגוריתמים Isolation Forest, LOF, dbscan .	הגשת תרגיל בזוגות – זיהוי אנומליות במערך נתונים בשפת python
6	משימות סיווג (classification) עצי החלטה Random Forest, Extreme Gradient Boosting (XGboost). בחירת היפר-פרמטרים באמצעות אימות צולב (cross-validation)	הגשת תרגיל בזוגות – בניית מודל מסווג בשפת python
7	משימות סיווג חלק ב' - Naïve Bayes , הערכת איכות המודל, זיהוי והתמודדות עם התאמת יתר, התמודדות עם מידע לא מאוזן.	הגשת תרגיל בזוגות – הערכת איכות המודל בשפת python
8	משימות רגרסיה – רגרסיה לינארית, XGboost regressor	הגשת תרגיל בזוגות – בניית מודל רגרסיה בשפת python
9	מבוא לרשתות נוירונים חלק א' - מודל רגרסיה בעזרת רשת נוירונים	הגשת תרגיל בזוגות – בניית מודל רגרסיה באמצעות רשת נוירונים בשפת python
10	חיזוי סדרות זמן. אלגוריתם ARIMA.	הגשת תרגיל בזוגות – בניית מודל סדרת זמן בשפת python
12	מבוא לרשתות נוירונים חלק ב' – חיזוי סדרת זמן באמצעות RNN Recurrent Neural Network	הגשת תרגיל בזוגות – בניית מודל לניבוי סדרת זמן באמצעות RNN
13	הרצאת העשרה כתלות בקצב ההתקדמות – שימוש ברשתות קונבולוציה CNN לניתוח תמונה, Transfer learning	הגשת תרגיל בזוגות – בניית מודל CNN לסיווג תמונות
14	הרצאת העשרה כתלות בקצב ההתקדמות – רשתות GANs, Autoencoder	זיהוי אנומליות עם רשתות AE

*ייתכנו שינויים בסילבוס בהתאם לקצב ההתקדמות ואפקטיביות הלמידה



ציון סופי

מסקל בציון הסופי	תיאור התוצר
50 % מהציון הסופי	ביצוע מטלות שבועיות בשפת python
50 % מהציון הסופי	מבחן מסכם



דרישות הקורס

דרישות מעבר - הגשת כל המטלות השבועיות לפני תאריך המבחן המסכם. ציון 60 משוקלל בין ציוני המטלות וציון המבחן המסכם.



שם הקורס	מס' הקורס
מבוא לתכנות בסביבת פייתון	55011

ביבליוגרפיה: תכנים לקריאה, צפייה והאזנה (רצוי עדכני)



ספרי הלימוד:

- Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python: a guide for data scientists*. " O'Reilly Media, Inc."
- Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc."