



מדעי הנתונים לניהול בשפת פייתון

ד"ר דרור כהן | המחלקה לניהול

מס הקורס 55066 | Introduction to Data Science for Management

הרצאות משולבות תרגילים במעבדת המחשבים

2

תשפ"ז

א

ב 18-20

סוג הקורס:

היקף נ"ז:

שנת לימודים:

סמסטר:

יום ושעה

שעת קבלה:

מייל מרצה:

קישור לאתר למדה:

Dror.Cohen@biu.ac.il

<https://lemida.biu.ac.il/course/view.php?id=105438>



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. הקורס מחולק להרצאה ותרגול, במהלך ההרצאה נלמדים העקרונות התיאורטיים ודרכי הפעולה של המודלים, יחד עם דוגמאות ליישום בקוד פייתון, ובמהלך התרגול תבוצע למידה פעילה של יישום קוד, והדרכה לקראת הגשת מטלות הבית.

מטרות/תוצרי הלמידה

ידע

1. הלומדים יגדירו את סוג האלגוריתם המתאים לשימוש, את משתני המטרה, ואת הצעדים

הנדרשים לניתוח נתונים ובנית מודלים של למידת מכונה, 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 פייתון 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 supervised learning: למידה בלתי מונחית (un-supervised learning) ודוגמאות לאלגוריתמים מתאימים.		
2	חקר נתונים ראשוני (Exploratory Data Analysis – EDA) חלק א' - וויזואליזציה של נתונים, בניית משתני דמה One-hot encoding, שיטות להתמודדות עם ערכים חסרים וערכים חריגים, קורלציה בין משתנים, שיטות לאיזון בין קבוצות במשתנה המטרה ואלגוריתם SMOTE		
3	חקר נתונים ראשוני (Exploratory Data Analysis – EDA) חלק ב' - הורדת מימד באמצעות PCA, t-SNE, דוגמאות ליישום חקר נתונים בקוד פייתון.	הגשת תרגיל בזוגות – ביצוע חקר נתונים ראשוני ועיבוד מקדים של סט נתונים.	

4	ניתוח אשכולות (clustering), אלגוריתם k-Gaussian mixture, DBSCAN, means models GMM Hierarchical clustering, יישום בקוד פייתון.	הגשת תרגיל בזוגות – יצירת אשכולות במערך נתונים
5	למידה בלתי מונחית ב' - זיהוי אנומליות (anomaly detection), אלגוריתמים Isolation Forest, LOF, DBSCAN. יישום בקוד פייתון.	
6	למידה מונחית א' - משימות סיווג (classification) בעזרת מודלי למידת מכונה קלאסיים ביניהם logistic regression, Naïve Bayes, עצי החלטה Random Forest, Extreme Gradient Boosting (XGboost), support vector machines (SVM).	
7	למידה מונחית ב' - משימות סיווג (classification) בחירת היפר-פרמטרים באמצעות אימות צולב (cross-validation), מדדים להערכת איכות המודל, זיהוי והתמודדות עם התאמת יתר, התמודדות עם מידע לא מאוזן, יישום בקוד פייתון.	הגשת תרגיל בזוגות – בניית מודל מסווג classifier
8	למידה מונחית ג' - משימות רגרסיה – רגרסיה לינארית, Random Forest regressor, support vector XGboost regressor, מדדים להערכת איכות המודל. יישום בקוד פייתון.	הגשת תרגיל בזוגות – בניית מודל רגרסיה
9	מבוא לרשתות נוירונים - מודלי רגרסיה וסיווג בעזרת רשת נוירונים בקישור מלא Fully connected יישום בקוד פייתון.	הגשת תרגיל בזוגות – בניית מודל רגרסיה באמצעות רשת נוירונים
10	חיזוי סדרות זמן. אלגוריתם SARIMA, ARIMA, PROPHET, יישום בקוד פייתון.	הגשת תרגיל בזוגות – בניית חיזוי סדרת זמן
11	חיזוי סדרת זמן באמצעות Recurrent Neural Network, GRU, LSTM, RNN Network יישום בקוד פייתון.	
12	שימוש ברשתות קונבולוציה CNN לניתוח תמונה, Transfer learning יישום בקוד פייתון.	הגשת תרגיל בזוגות – בניית מודל CNN לסיווג תמונות
13	מבוא לניתוח שפה טבעית, בניית מרחבי ייצוג embedding spaces, tf-idf, Word2Vec, מנגנון self attention, מבנה כללי של רשת טרנספורמר יישום בקוד פייתון.	

*ייתכנו שינויים בסילבוס בהתאם לקצב ההתקדמות



ציון סופי

תיאור התוצר	משקל בציון הסופי
ביצוע מטלות בית	50 % מהציון הסופי
מבחן מסכם	50 % מהציון הסופי



דרישות הקורס

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



דרישות קדם

שם הקורס	מס' הקורס
מבוא לתכנות בסביבת פייתון	55011



ביבליוגרפיה: תכנים לקריאה, צפייה והאזנה (רצוי עדכני)

ספרי הלימוד:

1. **Müller, A. C., & Guido, S. (2016).** *Introduction to machine learning with Python: A guide for data scientists*. O'Reilly Media.
2. **Géron, A. (2019).** *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (3rd ed.). O'Reilly Media.