# סילבוס: מבוא למדע הנתונים Introduction to Data Science

מרצה: פרופ' סהרון רוסט [saharon@tauex.tau.ac.il](mailto:saharon@tauex.tau.ac.il)  
שעות קבלה: לפי תיאום באימייל

מתרגל: מר דורון גוראל [dorongurel@gmail.com](mailto:dorongurel@gmail.com)  
שעות קבלה: לפי תיאום באימייל

מטרת הקורס לייצר היכרות עם האספקטים השונים של התחום, ולשלב בין הבנה תיאורטית של עקרונות בסיסיים ורעיונות לבין התנסות מעשית בעבודה עם נתונים, בכל שלבי התהליך.

**הנושאים התיאורטיים שיכוסו בקורס כוללים:**

* איסוף וארגון של נתונים: סקרים, טבלאות, web scraping
* ניקוי והבנה של נתונים: סיכומים, ויזואליזציה, principal component analysis
* חשיבה הסתברותית: מומנטים, התפלגויות מותנות, פרדוקס סימפסון
* הסקה סטטיסטית: אמידה, בדיקת השערות
* בניית מודלים על נתונים: רגרסיה לינארית ולוגיסטית
* כלים ורעיונות בסיסיים בשיטות מודרניות: עצי החלטה, gradient descent, רגולריזציה, פירוקי bias-variance
* שיטות מודרניות לבניית מודלים: Random Forest, Boosting, Deep learning

**הקורס כולל עבודה מעשית נרחבת עם נתונים:**

* תכנות בסביבת Python
* Case studies מעשיים שילוו את הקורס:
  + איסוף וארגון נתונים מאתר אינטרנט כגון Ebay.com
  + הבנת הנתונים וניתוחם
  + בניית מודלים עליהם
* הרצאות אורח על Case studies נוספים

**רקע נדרש:**

* ידע בסיסי באלגברה לינארית וחשבון אינטגרלי ברמה דומה לקורס 1ב של בי"ס למתמטיקה בשתי הסדרות
* ידע בסיסי בהסתברות ברמת קורס מבוא להסתברות
* הכרות בסיסית עם סביבת מחשב ועקרונות כלליים של תוכנה ותכנות, ברמת קורס מבוא למחשבים
  + לא נדרשת היכרות עם Python, אך הנושא לא יילמד במפורש אלא יסופקו חומרים לסגירת פערים עצמית

**מרכיבי הציון:**

שיעורי בית (30%)

* יינתנו מדי שבוע, להגשה עצמית
* יכללו תרגילים תיאורטיים ומעשיים (משימות במסגרת case study)
* חובת הגשה: לפחות 2/3 מתוך התרגילים (לפי מספרם הסופי, שיהיה בסביבות 13)

מבחן סיום (70%)

**הערה:** שפת ההוראה בכיתה היא עברית (אלא אם תהיה דרישה אחרת), אך כל החומרים בקורס, כולל מצגות ושיעורי הבית יהיו ב**אנגלית** (השפה הבינ"ל של התחום).