

**עבודה מספר 1- הערכת שווי רכבים**

בעבודה זו נתמקד בבעיית הערכת שווי רכבי יד שניה. לצורך כך נעזר בסט הנתונים אשר פורסם באתר Kaggle-

<https://www.kaggle.com/c/1056lab-used-cars-price-prediction/overview>

**הנחיות:**

ניתוח הנתונים ומיצוי מאפיינים-

- א. הורד את סט הנתונים מהאתר (train set) וקרא אותם לתוך data frame.
- ב. בצע EDA- בדוק את התפלגות הערכים של המאפיינים השונים, אחוז ערכים חסרים בכל מאפיין, קורלציות בין המאפיינים, וכיוצא בזאת.
- ג. בצע עיבוד מקדים.

מדד שגיאה-

- ד. בחר מדד שגיאה מתאים עבור הבעיה. הסבר את הבחירה.

פיתוח המודלים-

- ה. בחר שניים מבין המודלים שנלמדו בהרצאות. ממש כל אחד מהמודלים והשתמש בו לצורך הערכת שווי רכבים. בחן והשווה את ביצועי המודלים. את הערכת הביצועים כאמור יש לבצע בשלוש שיטות שונות: hold-out, 10-fold cross validation ו-leave-one-out.

**הנחיות כלליות ודגשים:**

- ההגשה תתבצע במתכונת של מחברת Jupyter :
  - המחברת תכלול את כל הקוד שכתבתם.
  - המחברת תכלול תאי טקסט להסברים של מה שנעשה, ניתוח התוצאות והמסקנות.
  - גרפים יכללו כותרות ומקרא, ובנוסף הסבר לגרף.
  - יש להגיש את המחברת לתיבת ההגשה המתאימה.

- התרגיל יוגש ביחידים או בזוגות, באמצעות המודול בלבד. בהגשה בזוגות, מצופה מכל אחד מבני הזוג לשלוט בכל נדבכי העבודה והדו"ח. רק אחד מבני הזוג יגיש את העבודה במודל. יש לרשום שמות + מס' תעודות זהות בראש המחברת.
- עבודות דומות תיפסלנה ויינקטו צעדיים משמעותיים.
- לוח הזמנים להגשה- בהתאם למוגדר באתר.
- אין להשתמש ב- LLMs למיניהם.

• להלן קריטריונים לציון:

- ניתוח מעמיק של הבעיה והנתונים, הגדרת מטריקה מתאימה ו- target variable.
- בחירת מודלים מתאימים (לדוגמה – צריכה להיות סיבה טובה למה לא להשתמש במודלים שנחשבים "state of the art").
- אופן הצגת התוצאות והסקת מסקנות מתאימות מתוך התוצאות המוצגות.
- איכות התוצאות. בהקשר זה, יש לבחון אפשרות לשפר את התוצאות באמצעות ביצוע אופטימיזציה של ההיפר-פרמטרים.

בהצלחה!

צוות הקורס