

עבודה מס' 1 - הערכת שווי רכבים

בעבודה זו נתמקד בבעיית הערכת שווי רכבי יד שנייה. 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) (.data frame) וקרא אותו לתוך train set (train set) וקרא אותו לתוך .data frame.
- ב. בצע EDA - בזוק את התפלגות הערכים של המאפיינים השונים, אחו ערכים חסרים בכל מאפיין, קורלציות בין המאפיינים, וכיוצא בזאת.
- ג. בצע עיבוד מקדים.

מדד שגיאה -

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

פיתוח המודלים -

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

הנחיות כלליות ודוגשים:

- ההגשה תבוצע במתכונת של מחברת Jupyter :
- המחברת תכלול את כל הקוד שתכתבם.
- המחברת תכלול תא טקסט להסבירים של מה שנעשה,ניתוח התוצאות והמסקנות.
- גראפים יכלול כותרות ומקרא, ובנוסף הסבר לגרף.
- יש להציג את המחברת לתיבת ההגשה המתאימה.

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

• להלן קритריונים לציון:

- ণיתוח עמוק של הבעיה והנתונים, הגדרת מטריקה מתאימה ו-target variable.
- בחירה 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.
- aicoot התוצאות. בהקשר זה, יש לבחון אפשרות לשפר את התוצאות באמצעות ביצוע אופטימיזציה של ההיפר-פרמטרים.

בצלחה!

צוות הקורס