

תרגיל למידת מכונה

חלק א: אנליזה וניבוי מחירי דירות

קבלן דירות תוהה כיצד לתמחר את מחירי הדירות אותן הוא מציע למכירה. לשם כך, הוא ממנה אתכן לעזור לו במשימה. לרשותו, נתונים של דירות אותן מכר בעבר עם המאפיינים הבאים:

- **Price:** Price of the house.
- **Bedrooms:** Number of bedrooms.
- **Bathrooms:** Number of bathrooms.
- **Sqft_living:** Square footage of the apartment's interior living space.
- **Sqft_lot:** Square footage of the land space.
- **waterfront:** A dummy variable for whether the apartment was overlooking the waterfront or not.
- **Floors:** Number of floors.
- **View:** An index from 0 to 4 of how good the view of the property was.
- **Condition:** An index from 1 to 5 on the condition of the apartment.
- **sqft_above:** The square footage of the interior housing space that is above ground level.
- **sqft_basement:** The square footage of the interior housing space that is below ground level.
- **yr_built:** The year the house was initially built.

Note: the indexes are from worst for best.

בניית מודל:

- בנו מודל **רגרסיה לינארית** לחיזוי מחירי הדירות (שימו לב שיש משתנים שצריך לקודד)

הערכת ביצועי המודל:

- השתמשו במדדים מתאימים להערכת איכות המודל (בדיקת טיב המודל נעשית על נתוני המבחן)

פרדיקציה:

- בחרו תצפית חדשה (המציאו דירה עם מאפיינים ריאליסטיים).
- בצעו חיזוי בעזרת המודל ובחנו האם התוצאה נראית סבירה. הסבירו.

חלק ב: אנליזה וניבוי של התקפי לב

פרופ' שלמה כהן, קרדיולוג מומחה בבית חולים ניו-ישראל, מבקש את עזרתכם לבנות מודל לחיזוי התקפי לב ובכך להציל חיי אדם. לרשותו המאפיינים הבאים אודות מטופליו:

- **Age** : Age of the patient
- **Sex** : Sex of the patient (1 = male; 0 = female)
- **exang**: exercise induced angina (1 = yes; 0 = no)
- **ca**: number of major vessels (0-3)
- **cp** : Chest Pain type chest pain type
 - Value 1: typical angina
 - Value 2: atypical angina
 - Value 3: non-anginal pain
 - Value 0: asymptomatic
- **trtbps** : resting blood pressure (in mm Hg)
- **chol** : cholestoral in mg/dl fetched via BMI sensor
- **fbs** : (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
- **rest_ecg** : resting electrocardiographic results
 - Value 0: normal
 - Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
 - Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
- **thalach** : maximum heart rate achieved
- **target** : 0= less chance of heart attack 1= more chance of heart attack

מפאת קוצר הזמן, הוא נאלץ להפקיד בידכם את המשימה במלואה ובפרט להבין מה כל משתנה מסמל בכוחות עצמכם. הוא סומך עליכם שתוכלו לעמוד במשימה. עליכם לענות על הסעיפים הבאים:

בניית המודל:

- בצעו קידוד מתאים למשתנים קטגוריאליים.
- בנו מודל **רגרסיה לוגיסטית** לחיזוי הסיכון להתקף לב.

הערכת ביצועים:

- חשבו את המטריקות Accuracy, Precision, Recall, F1
- הסבירו איזו מטריקה חשובה יותר בהקשר הרפואי ולמה

פרדיקציה:

- המציאו תצפית חדשה של מטופל עם פרטים ריאליסטיים.
- בצעו חיזוי עם המודל ופרשו את המשמעות הקלינית של התוצאה.

בהצלחה!