

## מערכות המלצה מבוססות בינה מלאכותית

### מטלה 1 – מערכות המלצה שיתופיות – Collaborative Filtering

#### הוראות כלליות:

- יש להגיש את פתרון המטלה עד לתאריך **17.12.2025 בשעה 23:55** - כל איחור בהגשת הפתרון יגרור הורדת ציון.
- יש להגיש את העבודה בזוגות לאתר הקורס באופן אלקטרוני – רק מגיש אחד יעלה את הפתרון לאתר הקורס.
- יש להגיש לאתר הקורס קובץ זיפ (zip) המכיל שני קבצים:
  - Python Notebook (ipynb) מתועדת ומוכנה לבדיקה. מומלץ לבדוק כי המחברת עובדת בסביבת [Colab](#) טרם העלאה שלה.
  - דו"ח (קובץ PDF) המכיל מענה לכל הסעיפים במטלה זו. יש לציין במפורש את שמות המגישים ותעודות הזהות של המגישים בראש המסמך.

#### הוראות התרגיל:

בתרגיל זה תתבקשו לממש מערכת המלצה המבוססת על Collaborative Filtering בעבודה עם בסיס הנתונים [MovieLens](#). לצורך העבודה יש להוריד את הנתונים מאתר הקורס (ml-latest-small.zip) שנאספו וסופקו על ידי [grouplens](#). במטלה זו נתמקד בחיזוי הדירוג (rating) על בסיס הקובץ ratings.csv. ניתן ואף רצוי להשתמש בחבילות Python על מנת לפתור מטלה זו. לדוגמה שימוש בחבילה [Surprise](#).

#### 1. עיבוד מקדים וביצוע EDA:

- יש לבצע EDA ולהציג לפחות שלושה ניתוחים שונים ומשמעותיים אשר תרמו לכם להבנת הבעיה. יש לפרט כל ניתוח בעזרת תרשים ומלל שיסביר את משמעות הניתוח.
- במטלה זו ועבור שאר הסעיפים נתמקד אך ורק בעמודות: userId, movieId, rating. לכן, עליכם להסיר כל עמודה נוספת מהנתונים.

#### 2. יישום Collaborative Filtering – Memory Based:

- יש לממש את הגישות User-Based Collaborative Filtering ו-Item-Based Collaborative Filtering בעזרת KNN על בסיס פונקציית המרחק Cosine.
- עליכם להשתמש בטכניקת [5-fold cross-validation](#) לצורך דיווח התוצאות.
- עליכם להציג את המדדים הממוצעים עבור MSE ו-MAE.
- עליכם להציג את התוצאות עבור ערכי K שונים מסט הערכים [3,5,10] ולהסביר את המשמעות של כל התוצאות שהתקבלו.
- תוצר סופי עבור סעיף זה – עבור כל גישה (User-Based/Item-Based) יש לרכז את כל התוצאות שקיבלתם בשני איורים שונים – איור עבור כל מדד (MAE/MSE). ציר ה-x בכל איור יהווה ערך ה-K וציר ה-y יהווה ערך המדד שהתקבל.

### 3. יישום Collaborative Filtering – Model Based:

- a. יש לממש את המודלים SVD, SVD++, NMF
- b. עליכם להשתמש בטכניקת [cross-validation](#)-5 לצורך דיווח התוצאות.
- c. עליכם להציג את המדדים הממוצעים עבור [MSE](#) ו-[MAE](#).
- d. תוצר סופי עבור סעיף זה – יש להציג את התוצאות השונות בטבלה אחת כך שהשורות בטבלה יהוו המודלים השונים בעוד העמודות בטבלה יהוו המדדים השונים. יש להדגיש את הערך הטוב ביותר עבור כל מדד. יש לספק הסבר כללי ומסקנות על התוצאות שהתקבלו.

**בהצלחה!**