

## תכנות מתקדם - 3101803

מרצים: מר פרץ אור, מר גוטמן דוד

- משך הבחינה: שלוש שעות (180 דקות).
- חומר פתוח.
- מועד א'.
- הבחינה מכילה 2 חלקים אשר יש לענות על כל השאלות. יש לצרף את הפתרון לתיבת ההגשה בפורמט py או ipynb בלבד.
- **לאחר 3 שעות, תיבת ההגשה תיסגר באופן אוטומטי. יש לנהל את זמן הבחינה היטב ולהגיש בזמן הקצוב. בחינות אשר יאחרו את המועד לא ייבדקו.**

# בהצלחה!

### (35 נק') חלק א'

**הגדרה:** פלינדרום הוא מילה, מספר, משפט או כל רצף סמלים אחר, שקריאתו מימין לשמאל ומשמאל לימין היא זהה. למשל, המספר 12321 והמחרוזת "abba" הינם פלינדרומים, אך המחרוזת "abc" אינה פלינדרום, שכן קריאתה מימין לשמאל היא שונה מהמילה המקורית.

1. (15 נק') - כתבו את הפונקציה הרקורסיבית is\_palindrome אשר מקבלת ביטוי (מספר או מחרוזת) ומחזירה True אם הביטוי הינו פלינדרום, אחרת False.
2. (3 נק') - צרו רשימה המכילה 3 ערכים לבחירתכם, כאשר 2 מהם פלינדרומים והשלישי לא.
3. (5 נק') - הריצו את הפונקציה is\_palindrome עבור כל אחד מהערכים ברשימה מסעיף (2).

### ללא תלות בסעיפים הקודמים:

(12 נק') - נתונה הסדרה האינסופית הבאה:

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots$$

כתבו פונקציה רקורסיבית בשם harmonic אשר מקבלת מספר שלם, חיובי וגדול מ 1 ומחזירה את הסכום של הסדרה הנ"ל עד לאיבר ה-n.

למשל, עבור  $n = 3$  הפונקציה תחזיר  $1 + \frac{1}{2} + \frac{1}{3} = \frac{11}{6}$

## (65 נק') חלק ב'

לבחינה זו, מצורף קובץ בשם customers.csv אשר מכיל מידע לקוחות במרכז קניות מסחרי.

CustomerID: ת"ז

Gender: מגדר

Age: גיל

Income: הכנסה חודשית, באלפים

Spending Score: אחוז בזבז חודשי מהמשכורת

## (30 נק') - תחקור ראשוני, שאלות וגרפים

1. (2 נק') - קראו את קובץ ה-csv ל-DataFrame.
2. (2 נק') - מהו ממוצע גילאי הגברים אשר מכניסים מעל 20 אלף ש"ח?
3. (4 נק') - מהי כמות הגברים מול כמות הנשים?
4. (4 נק') - מי **מבזבז** יותר - נשים או גברים (ממוצע)?
5. (4 נק') - מי **מרוויח** יותר - נשים או גברים (סכום כולל)?
6. (4 נק') - מהי כמות הלקוחות אשר מכניסים מעל 15 אלף ש"ח ומבזבזים לפחות 40%?
7. (4 נק') - עבור נשים אשר גילן צעיר מ-40, מהו ממוצע ההכנסות שלהן?
8. הציגו 2 גרפים (מסוגים שונים) לבחירתכם:
  - a. (3 נק') - גרף 1 - יענה על התשובה לשאלה מספר 4.
  - b. (3 נק') - גרף 2 - יענה על התשובה לשאלה מספר 5.

## (20 נק') - עיבוד נתונים

1. (3 נק') - **במידה וישנם** ערכים חסרים בטבלה, מחקו את רשומות אלו.
2. (3 נק') - השתמשו בקידוד one-hot-vector והמירו את המשתנים קטגוריאליים.
3. (3 נק') - צרו עמודת Target אשר שווה 1 אם הלקוח מבזבז מעל ל-50% מהמשכורת שלו, אחרת 0.
4. (3 נק') - מחקו את עמודות CustomerID, Spending Score.
5. (5 נק') - עבור כל עמודה נומרית, נסמן  $max$  - האיבר המקסימלי של העמודה. בצעו נרמול לנתונים, כך שעבור כל ערך  $a$  בטבלה, הערך החדש יחושב:
 
$$\frac{a}{max}$$
6. (3 נק') - פצלו את אוסף הנתונים ל- $data$ ,  $labels$  כאשר  $data$  הינו אוסף הנתונים ללא עמודת Target, והמשתנה  $labels$  הינו עמודת Target. צרו קבוצות אימון ומבחן בהתאם לחלוקה של 25% לקבוצת המבחן.

## (15 נק') - מודל למידה

1. (5 נק') - הריצו עצי-החלטה עם העומקים  $depth = 2 to 10$ , כאשר בכל ריצה שמרו ברשימה את מדד הדיוק של האלגוריתם על קבוצת המבחן. הדפיסו את התוצאה ובחרו את העומק הרצוי.
2. (5 נק') - בצעו אימון על המודל אשר בחרתם בסעיף (1) והציגו את מדד הדיוק ומטריצת הבלבול (Confusion Matrix).
3. (5 נק') - רשמו מסקנה **אחת** עבור המודל.