

## מבוא לאנליטיקה עסקית 1221.7020

### תרגיל מסכם: פרדיקציה

חברת ביטוח הרכב הידועה Sale'at Inc. מבקשת את שירותיך בחיזוי ערך לקוח – CLV (customer lifetime value). לשם כך, סיפקה החברה מסד נתונים ובו נתונים עבור 8637 לקוחות קיימים. הדאטה כולל מאפיינים דמוגרפיים, מאפייני רכישת פוליסות ותביעות, וערך לקוח לכל לקוח קיים (שמות המשתנים ברורים ומכילים הסבר מלא על כל משתנה).

מסד נתוני הלקוחות הקיימים נמצא בקובץ: CLV\_8637rows.csv

מטרתך היא לחזות את ה-CLV של כל אחד/ת מ-497 לקוחות חדשים, עבורם ידועים הערכים לכל המשתנים מלבד ה-CLV.

מסד נתוני הלקוחות החדשים נמצא בקובץ: CLV\_NoTargetVar\_PREDICT.csv

עליך לבחון מודלים מתאימים לחיזוי ה-CLV, ולבחור את המודל בעל הביצועים הכי טובים.

#### יש להגיש:

- א. קובץ CSV ובו 2 עמודות ושמותיהן: **CustomerID, CLV** עבור 497 הלקוחות החדשים.  
שם הקובץ יהיה **Predictions123456789.csv** כאשר את הספרות 1-9 יש להחליף במס' ת"ז שלך.
- ב. קובץ CSV שבו כתוב ה-MAE של המודל הסופי שבחרת לפרדיקציה, מחושב על כל מסד נתוני הלקוחות הקיימים (יש לחשב את ה-MAE ב-R ואז ליצור קובץ CSV עם עמודה אחת בשם **MAE** ובתוכה ערך ה-MAE שחושב)  
שם הקובץ יהיה **MAE123456789.csv** כאשר את הספרות 1-9 יש להחליף במס' ת"ז שלך.
- ג. קובץ R שבו השתמשת לפתרון. דגשים:
  - בתחילת הקובץ יש לקבוע `set.seed(123)` (לשם אחידות).
  - יש להגדיר את נתיב התיקיה בנפרד (`setwd`) ואת שם הקובץ בנפרד (`read.csv`)
  - שם הקובץ יהיה **Code123456789.R** כאשר את הספרות 1-9 יש להחליף במס' ת"ז שלכם.
  - יש להגיש קובץ R שעובד, מייצר את ה-CSV שהוגש (עם פקודת `write.csv`), ומחשב את ה-MAE שהוגש. יש לבדוק שכשאתם מריצים את הקוד ברצף מההתחלה עד הסוף מתקבל ה-CSV המוגש.
  - אין להגיש קוד שבהרצתו נוצרות הודעות שגיאה `error` (מותר להגיש קוד שמייצר הודעות אזהרה `warning`).
  - מותר להשתמש אך ורק במודלים שנלמדו במסגרת הקורס.
  - שימו לב – עליכם לכתוב את שמות הקבצים והעמודות בדיוק כפי שהוגדרו לעיל.

#### אופן הניקוד:

כל הגשה שתכלול קוד R שעובד ומייצר את התחזיות שבקובץ ה-CSV, ומחשב MAE בהתאם למה שהוגש, תיחשב כהגשה תקינה ותקבל ציון עובר.

ההגשות התקינות ינוקדו באופן יחסי לפי איכות התחזיות שתימדד עפ"י ה-MAE של הפרדיקציות. כלומר, ההגשות ידורגו עפ"י ה-MAE שלהן בנתוני האמת עבור 497 הלקוחות החדשים (ערך ה-CLV האמיתי עבור לקוחות אלו נמצא בידי צוות הקורס בלבד ולא יסופק לכם).

הגשות בעלות MAE זהה, יובילו לבדיקת הקוד. ניקוד ההגשה ייקבע בהתאם לאיכות היחסית של הפרדיקציה, ובאופן שיתקבל ממוצע 87 (להגשות התקינות).

בהצלחה!