

## מבוא להנדסת נתונים – מעבדה 4

לירון פז – 311280283

אדיר טולדנו - 313535379

### שאלה 1:

1-NN	
Accuracy	0.700709956

הרצנו את הנתונים על הקובץ המלא, קיבלנו שהמחלקה הצפויה (1) תואמת למחלקה המקורית (האמיתית) עם 70% דיוק.

### שאלה 2:

לאחר שהרצנו 30 פעמים את האלגוריתם מצאנו כי המסווג הטוב ביותר הינו: 19, ואחוז הדיוק שלו היה: 76.0416666%.

### שאלה 3:

לפי שאלה 2, כיוון שהמסווג הטוב ביותר היה 19, הרצנו עליו את כל הבדיקות.  
עבור  $\text{fold} = 2$  התקבלו תוצאות עקביות יחסית בעלי שונות נמוכה ביחס לשאר.  
בהינתן שאנו עובדים עם קובץ נתונים קטן, היינו ממליצים על  $k$  קטן כך שהשונות תצא נמוכה, למשל  $\text{Fold} = \{2,3\}$ .

### שאלה 4:

הרצנו את הבדיקה עבור  $K=5$ ,  $K=7$  בארבעת השיטות, המסקנה שלנו היא שנורמת Min-Max נותנת את אחוז הדיוק הגבוה ביותר (73.4375% עבור  $K=7$ ). לעומת זאת נורמת Sum הייתה בעל אחוז הדיוק הנמוך ביותר (69.53125% עבור  $K=7$ ).