פרויקט בלמידה חישובית

מגישים:

אביב זבולוני – ת.ז. 211313333  
אליה אטלן – ת.ז. 318757200

**מבוא:**

פרויקט זה עוסק בחיזוי מחלות. האלגוריתם שלנו מנסה לחזות 4 מחלות שונות:

1) הסיכוי של בן אדם לחלות בסכרת – [קישור לדאטא](https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset)  
2) הסיכוי בן אדם לקבל התקף לב – [קישור לדאטא](https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction)  
3) הסיכוי של בן אדם לחלות ב hypothyroid (תפקוד לא תקין של בלוטת התריס) – [קישור לדאטא](https://www.kaggle.com/datasets/yasserhessein/thyroid-disease-data-set)  
4) הסיכוי של בן אדם לקבל שבץ – [קישור לדאטא](https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset)  
  
באמצעות השילוב של ארבעה קבוצות נתונים הנ"ל, האלגוריתם מביא למשתמש 4 פלטים, כל פלט מחזיק את הסיכוי עבור כל מחלה.  
  
אבל, כל דאטא מתאים לקבוצת גילאים אחרת, ולכן האלגוריתם לא יחזה את הסיכוי של בן אדם לחלות בכל הארבעה, אלא כל בן אדם יוכל לקבל חיזוי עבור המחלה שמתאימה לתחום הגילאים שלו.

**הדאטא:**

1. **סכרת:**זה קובץ הדאטא הכי גדול שמצאנו (עם 250 אלף דוגמאות). בקובץ זה ישנם 3 מחלקות - 0 זה אומר שאין סכרת, 1 זה סיכון לחלות, ו 2 זה אומר שהבן אדם חולה בסכרת.  
   זה גם הקובץ הכי מסודר שמצאנו, מכיוון שאין שם נתונים בטקסט, הכל במספרים (F,T למשל).

הדאטא הזה מתאים לילדים בגילאים 1-13, ולכן המודל יכול לחזות את הסיכוי בלחלות בסכרת רק לקבוצת הגילאים הזאת.