**שלבים בהכנת ה-Dataset**

1. הסרת שורות ריקות (null בעמודות של ביטוי הגנים).
2. "איחוד" של פרובים שונים השייכים לאותו גן כך שתהיה שורה אחת לכל גן בדאטאסט, רמות הביטוי יחושבו על-פי ממוצע.
3. הסרת גנים שלא מקושרים לאף go term.
4. הורדנו רשימה של go terms המקושרים ל-apoptosis: <http://amigo.geneontology.org/amigo/search/ontology?q=apoptosis>  
   ותייגנו כל גן כ-TRUE אם ברשימת ה-terms המקושרים אליו יש terms המקושרים ל-apoptosis, ו-FALSE אחרת.

סה"כ: 1728 גנים שמקושרים ל-terms של apoptosis, מתוך 16080 גנים מתויגים.

אחרי ניסיון הרצה ראשון של decision tree (ללא כוונון פרמטרים):

score: 0.7949834162520729

recall score: 0.11560693641618497

accuracy score: 0.1016949152542373

* כלומר יש דיוק יחסית גבוה אבל זה לא אומר שהביצועים באמת טובים (פשוט כי יש הרבה יותר FALSE בלייבלים)
* רעיונות לשיפור: לבדוק אם עדיף להשתמש בדרך שונה מממוצע כדי להגיע לערכים יחידים של רמות ביטוי לכל גן (למשל לבחור אחד...)