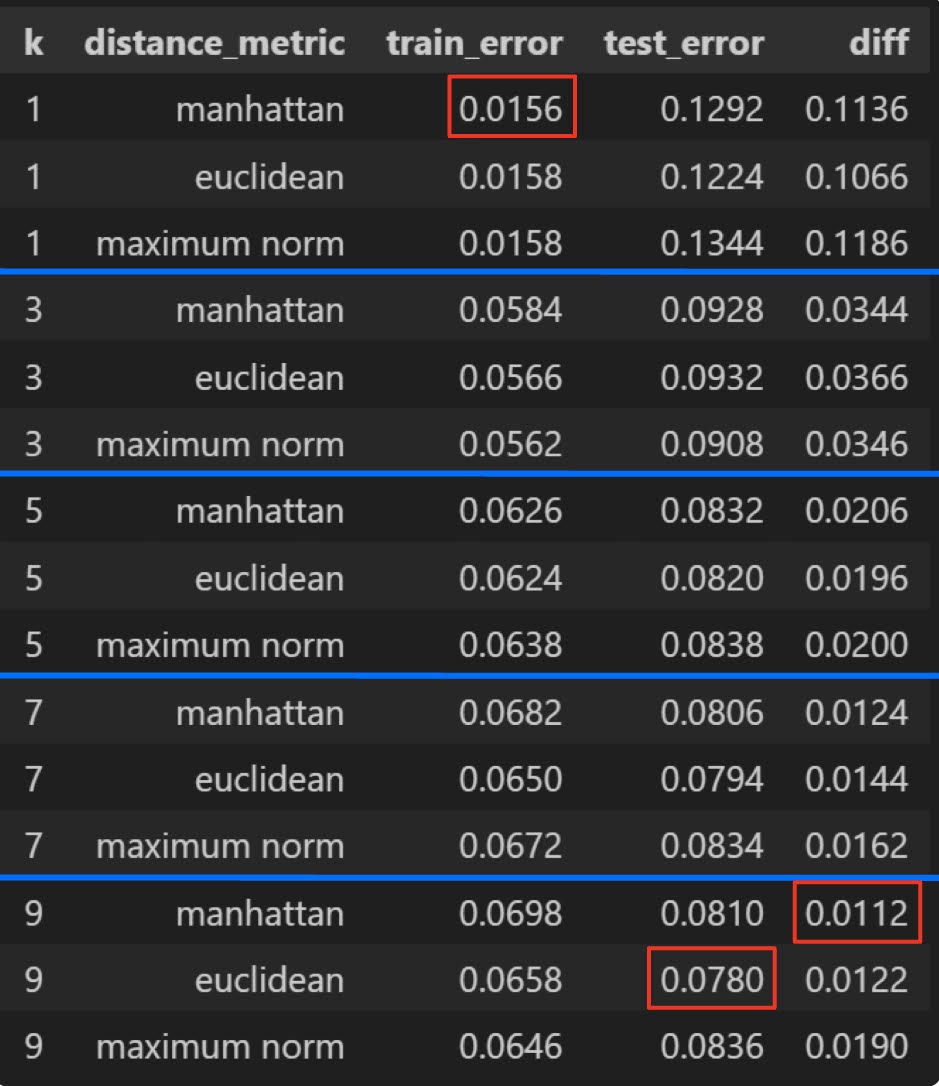
**מגישים:   
נועם סיידה 212631071  
אלון סויסה 211344015**

**חלוקת עבודה:**שאלה 1- נועם  
שאלה 2- אלון

**שאלה 1:**



סימנתי באדום את הערכים המינימליים בכל עמודה.

**ערכי הk וה-p הכי טובים לסיווג הם: p=2, k=9.**

זאת משום שהtest error הממוצע שלהם הוא הכי נמוך מבין כל שאר האפשרויות.

צריך למזער את ה-test error כי היא מציגה מה הטעות של המודל על העולם, בעוד שהtrain error היא הטעות של המודל על הdata שהוא כבר ראה והתאמן עליו, ולכן מסוגל לסווג אותו נכון גם אם הוא לא מכליל את הדפוס בדאטא ולא מזהה את החוק האמיתי שעל פיו הדאטא מחולק(overfitting).

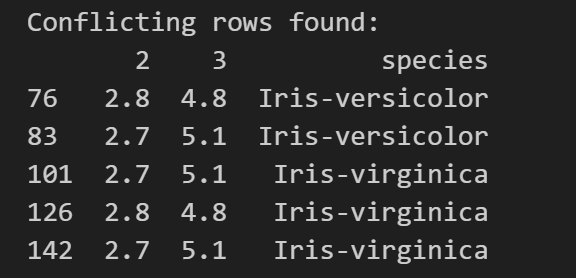
כֵּכְלל – ניתן לראות, בכל מטריקות המרחק, שככל שהגדלנו את הk:  
1. הtest error נהיה קטן יותר  
2. ה train error נהיה גדול יותר  
3. הdifference נהיה קטן יותר

כלומר, מודל עם ערכי k נמוכים יוצרים overfitting.

זאת משום שלמודלי knn עם ערכי k קטנים יש vc-dim גבוה. הם מסוגלים להתאים את עצמם לכל data שהוא, ולכן לא מכלילים את הדפוס בדאטא ומזהים את החוק האמיתי.

ניתן לראות שהtrain error הכי נמוך הוא כאשר k=1.   
זה בגלל הצורה שknn פועל: כאשר k=1, נקודה מסווגת כמו התיוג של השכן הכי קרוב שלה בבסיס של הknn.  
נשים לב, כל נקודה בtrain set היא השכן הכי קרוב לעצמה! ולכן המודל יסווג אותה כסיווג הנכון שלה.  
כלומר הtrain error צריכה להיות 0!

בפועל הtrain error היא קצת יותר מ0 כי יש כמה זוגות של דגימות עם תיוג שונה ופיצרים זהים, והמודל בטוח טועה על אחד מכל זוג, כי הוא מחזיר אותה תשובה לשניהם.

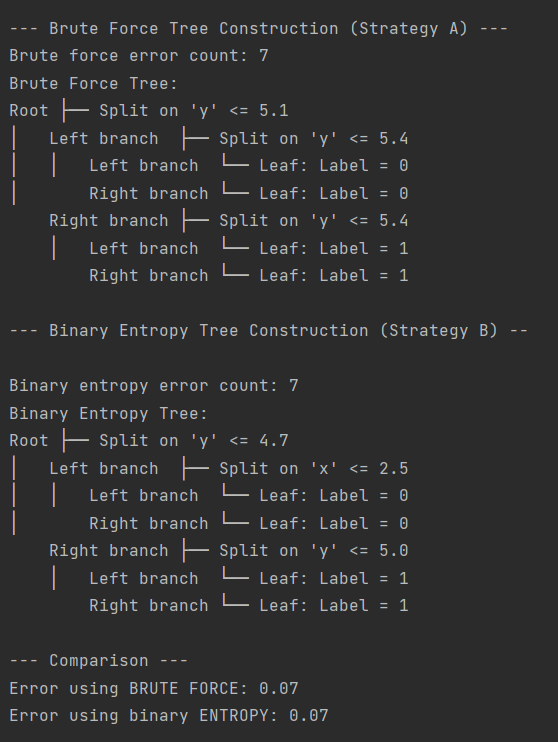


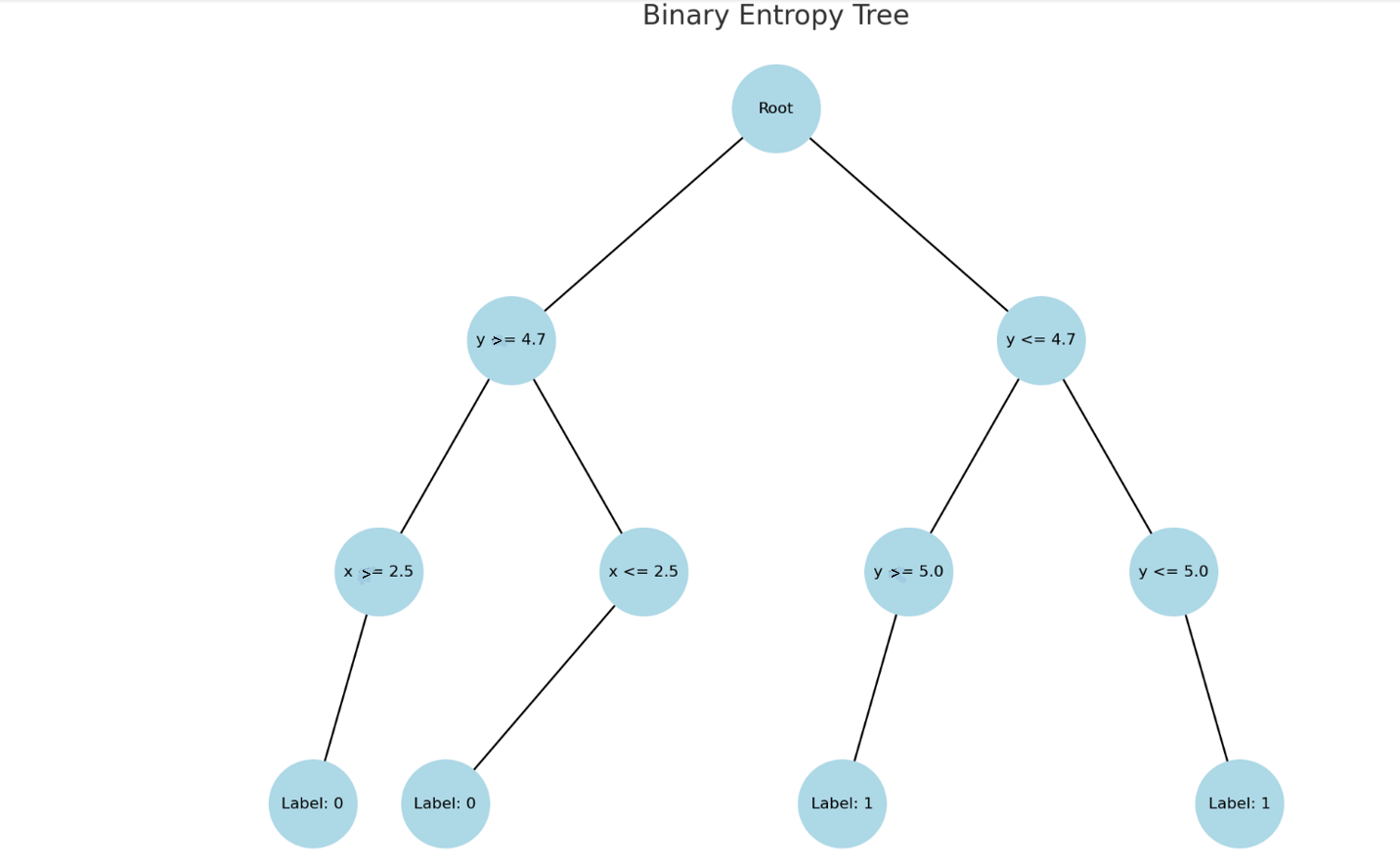
זאת בעוד הtest error ל k=1 הוא כ11%!.

לעומת זאת, בעוד שלk-ים גדולים הtest error הוא יותר גבוה וגם אין כל כך overfitting (סביבות האחוז וחצי הפרש).  
כלומר הk-ים הגדולים שניסינו הצליחו לזהות ברמה יותר טובה את החוק האמיתי.

נשים לב שלמרות הממצאים שלנו, צריך להיזהר לא לשים k-ים גדולים מדי.   
אם הk יהיה בערך בגודל של n – נסווג כמעט כל נקודה כתיוג של הרוב – וזה כמובן לא מוצא את החוק האמיתי בצורה טובה, ויתן תוצאות גרועות(50% accuracy במקרה שלנו).

**שאלה 2:**

****

הצגת העצים:  
תמונה שמכילה קו, עיגול, תרשים, צילום מסך

התיאור נוצר באופן אוטומטי

**ניתוח:**

**שוויון בביצועים:**

* שיעורי השגיאה הזהים מרמזים ששתי האסטרטגיות **יעילות באותה מידה** עבור מערך הנתונים הזה. ייתכן שזה נובע מכך ש:
  1. **מערך הנתונים פשוט מספיק**, כך שכל גישה מגיעה לאותן מסקנות.
  2. **מגבלת העומק ( K רמות)** כופה מבנה עץ דומה לשתי האסטרטגיות, מה שמגביל את הפוטנציאל לשוני.

**מגבלת עומק:**

* מגבלת העומק שהוטלה ככל הנראה מונעת מכל אסטרטגיה לחקור באופן מלא חלוקות ניואנסיות יותר, שייתכן והיו מובילות להבדלים.

**תפקיד האנטרופיה:**

* בעוד שבאופן תאורטי, האנטרופיה משפרת את איכות החלוקות, ייתכן שהיתרונות שלה **מעומעמים במערכי נתונים שבהם BRUTE FORCE כבר מספק חלוקות מיטביות** (למשל, במקרים פשוטים)

**OVER FITTING :**ניתן לראות שהפיצול האחרון לא השפיע כלל על התוצאות שכן לשתי העלים יש את אותו תיוק (בכל הפיצולים שלפני הסוף) לכן יכול להיות שעדיף להסתפק בשני פיצולים במקר הזה כדי להשיג תוצאות טובות יותר על סיווג של מידע חדש ולהימנע מ"OVER FITTING".