[redacted]

1. בתרגיל הבא יש לענות על השאלות באמצעות שימוש בקוד **פייתון ושימוש ב-Scikit-Learn**.

[redacted]

ענו על השאלות הבאות באמצעות הנתונים על מחירי יהלומים:

1. חלקו את המידע ל train וtest. באמצעות אלגוריתם KNN (K=3) בנו מודל שחוזה clarity של היהלום באמצעות נתוני carat, depth, price, table, y, x. לאחר מכן, חשבו את מדדי הaccuracy וf1_score עבור המודל שיצרתם.

2. חזרו על בניית המודל בסעיף 1 עבור ערכי k שונים. באמצעות seaborn ציירו גרפים של ביצועי המודלים עבור ערכי k שונים. כלומר, יש לצייר גרפים שבהם ציר ה-X הוא ערך ה-K של המודלים ואילו ערכי ה-Y הם מדדי הaccuracy וf1_score- שמתאימים לכל מודל.

3. חלקו את המידע ל train וtest. לאחר מכן, באמצעות אלגוריתמי KNN (K=1,3,5,7) וDecision Tree בנו מודלים שחוזים הcut של היהלום באמצעות עמודות carat, depth, price, table, and y, x. לפי מדד הaccuracy, מבין המודלים שיצרתם, איזה מודל חוזה את הcut טוב יותר?

4. חזרו על בניית המודלים בדומה לסעיף הקודם, רק הפעם בנו את המודל על ידי הוספת מידע מעמודות clarity וcolor (כלומר, בנו את המודלים בתוספת שתי העמודות הנוספות). האם הaccuracy של המודלים השתפרו?
רמז: יש להשתמש בLabelEncoder בסעיף זה

5. בדומה לסעיף 4, בנו מודלי (K=5) KNN החוזים את הcut של היהלום, רק שהפעם על מנת לבנות את המודלים, השתמשו בגדלים שונים של נתונים:
5%,10%,20%, 50%,75%,80%,90%,
כלומר, יש לבנות את המודל רק על ידי שימוש ב5% מהמידע, 10% מהמידע , וכו׳
צרו גרף שבו ציר ה-X הוא גודל הtrainset באחוזים ואילו ציר ה-Y הוא הaccuracy של המסווג.
הערה חשובה: חשוב להשתמש כ-test באותם נתונים בדיוק עבור כל המודלים