**תרגיל: ניתוח נתוני יין עם מודלים שונים**

**מטרת התרגיל**

1. להבין את תהליך ניתוח הנתונים (EDA) ולטפל בנתונים.
2. להתנסות בבנייה, כוונון והערכה של מודלים שונים.
3. לנתח את תוצאות המודלים ולהסיק מסקנות.

**שלב 1: טעינת הנתונים ובדיקה ראשונית**

1. טען את נתוני היין באמצעות הפונקציה load\_wine מספריית sklearn.
2. הצג את השורות הראשונות והאחרונות של הנתונים.
3. בדוק אם קיימים ערכים חסרים בנתונים.

**שלב 2: זיהוי סוגי עמודות**

1. בדוק אילו עמודות הן קטגוריאליות ואילו עמודות הן נומריות.
2. זהה אילו עמודות דורשות המרה באמצעות OneHotEncoder ואילו דורשות עיבוד באמצעות StandardScaler.

**שלב 3: ניתוח ראשוני עם העמודות הנומריות בלבד**

1. הגדר את העמודות הנומריות כמאפיינים (X) ואת עמודת target כעמודת המטרה (y).
2. חלק את הנתונים לסט אימון וסט בדיקה (80/20).
3. בנה מילון (dictionary) עם שלושת המודלים הבאים:
   * Logistic Regression
   * Decision Tree
   * Random Forest
4. בנה לולאה שמריצה את המודלים על הנתונים הנומריים בלבד, מפיקה דוח דיוק (accuracy report) ומציגה את תוצאות המודלים.

**שלב 4: הוספת StandardScaler**

1. בצע Scaling לעמודות הנומריות באמצעות StandardScaler.
   * השתמש ב-fit\_transform על סט האימון וב-transform על סט הבדיקה.
2. הרץ את הלולאה של המודלים שוב והשווה את התוצאות לתוצאות הקודמות.

**שלב 5: טיפול בערכים חריגים בעמודת 'Residual Sugar'**

1. זהה ערכים חריגים בעמודת 'Residual Sugar' באמצעות שיטת ה-IQR:
   * חשב את Q1, Q3 ו-IQR.
   * הגדר ספים תחתון ועליון והסר את הערכים הקיצוניים.
2. הרץ מחדש את הלולאה והשווה את התוצאות לתוצאות הקודמות.

**שלב 6: כוונון פרמטרים**

1. בחר את שיטת עיבוד הנתונים שהניבה את התוצאות הטובות ביותר.
2. נסה לכוונן פרמטרים עבור כל אחד מהמודלים:
   * **Logistic Regression:** כוונן את הפרמטר max\_iter.
   * **Decision Tree:** כוונן את הפרמטר max\_depth.
   * **Random Forest:** כוונן את הפרמטר n\_estimators.

**שלב 7: הצגת confusion matrix כ-Heatmap**

1. עבור המודל Logistic Regression עם הפרמטרים הכוונונים, הרץ את הנתונים והפק את confusion matrix והצג כ-Heatmap באמצעות ספריית seaborn.

**שלב 8: זיהוי הפיצ'רים החשובים בעץ ההחלטה**

1. שלוף את חשיבות הפיצ'רים מהמודל Decision Tree.
2. הצג את הפיצ'רים החשובים ביותר כגרף אופקי (Horizontal Bar Chart).

**שלב 10: הצגת עץ מספר 7 ב-Random Forest**

1. בחר את עץ מספר 7 מתוך ה-Random Forest.
2. הצג את מבנה העץ באמצעות הפונקציה plot\_tree מספריית sklearn.