

מסווג	פרמטרים	תוצאה
Decision Tree	{'criterion': 'entropy', 'max_depth': 9, 'min_samples_split': 12}	0.96
Random Forest	{'criterion': 'entropy', 'max_depth': 9, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 130}	0.96
AdaBoost	{'learning_rate': 1, 'n_estimators': 181}	0.97

פונקציית RandomizedSearchCV היא שיטה לחיפוש אקראי על פני מרחב של פרמטרים אפשריים לצורך אופטימיזציה של מודל למידת מכונה. היא חלק מהמודול sklearn.model_selection בספריית scikit-learn.

הסבר כללי על מה עושה RandomizedSearchCV:

היא מבצעת חיפוש אקראי (ולא חיפוש מקיף) על מרחב של פרמטרים כדי למצוא את הקומבינציה האופטימלית שתניב את הביצועים הטובים ביותר של המודל לפי קריטריון מסוים (לרוב "דיוק" או "שגיאה"). היתרון של חיפוש אקראי הוא שהוא יכול למצוא פרמטרים טובים בתוך מרחב חיפוש רחב, אבל תוך זמן קצר יותר מאשר חיפוש מלא של כל האפשרויות, כפי שעושה GridSearchCV.

איך זה עובד:

- הגדרת פרמטרים:** אנחנו מגדירים את טווח הערכים עבור כל פרמטר של המודל. לדוגמה, אם אנחנו רוצים לחפש את הפרמטרים עבור RandomForestClassifier, אנחנו יכולים להגדיר טווחים עבור (n_estimators מספר העצים), max_depth (עומק העץ), min_samples_split (מספר הדוגמאות המינימלי עבור פיצול עץ), וכדומה.
- בחירה אקראית של פרמטרים:** במקום לבדוק כל שילוב אפשרי של פרמטרים (כמו ב-GridSearchCV, RandomizedSearchCV) בוחר באופן אקראי מספר קומבינציות מתוך הטווחים שהוגדרו.
- הערכת המודל:** עבור כל קבוצה של פרמטרים שנבחרו, המודל מאומן ומוערך על פי קריטריון שנבחר (למשל דיוק). הערכה זו מבוצעת בדרך כלל על ידי שימוש ב-**cross-validation** - כלומר חלוקה של הנתונים לתת-קבוצות והערכה של המודל על כל תת-קבוצה בנפרד.
- בחירת המודל הטוב ביותר:** לאחר ביצוע החיפושים, הפונקציה מחזירה את המודל עם הפרמטרים שהניבו את הביצועים הטובים ביותר.