

Elemeztem a Seaborn könyvtár egy adatkészletét. Készítettem diagramokat és alkalmaztam a gépi tanulási modellt. A seaborn könyvtár `load_dataset()` funkciójával egy kész adatkészletet használtam, amely az iris virágok adatait tartalmazza (virág szirmai, csészelevelei)

Készítettem egy pairplot ábrát és megvizsgáltam az adatpontok közötti kapcsolatokat. A `hue` paraméterrel a különböző virágfajokat színekre bontottam, így könnyebben észrevehetőek a különbségek a jellemzők között. A `sns.kdeplot` funkcióval sűrűségbecsléseket is készítettem az alsó háromszög mátrixon, hogy részletesebben is vizsgálhassam az egyes jellemzők eloszlását.

A gépi tanuláshoz `scikit-learn` könyvtárat használtam és egy klaszterezési algoritmust (KMeans). Ez a modell segít a virágfajok csoportosításában, megmutatja mennyire lehet őket szétválasztani a vizsgált adatok alapján.

A modell futtatása után megvizsgáltam az eredményeket és összehasonlítottam a gépi tanulási eredményeket az eredeti fajokkal.

1) Adatok betöltése és előfeldolgozás

A `sns.load_dataset("iris")` segítségével betöltöttem az Iris adathalmazt, és a `pairplot()` függvénnyel grafikonon ábrázoltam a különböző tulajdonságokat (`sepal_length`, `sepal_width`, `petal_length`, `petal_width`) és azok kapcsolatát a különböző `species` (fajok) között.

2) Adatok előkészítése

Az `iris['species']` oszlopot numerikus értékekre konvertáltam a `.astype('category').cat.codes` módszerrel, hogy a gépi tanulási modell kezelni tudja őket. Az `X` tartalmazza az összes jellemzőt (főként a mérési adatokat), míg a `y` a címkét (a virág fajtája).

3) Tanító és tesztadatok

A `train_test_split` segítségével véletlenszerűen felosztottam az adatokat tanító és tesztelő halmazokra.

4) Standardizálás

Az `X_train` és `X_test` adatokat a `StandardScaler` segítségével skáláztam, hogy a gépi tanulási modell jobban működjön.

5) Betanítás

A `RandomForestClassifier` modellt alkalmaztam a tanító adatokra (`X_train` és `y_train`).

6) Előrejelzés és értékelés

A modell előrejelzést készít a teszt adatokra (`y_pred`). Ezután kiértékeltem a modell teljesítményét az `Accuracy` (pontosság) és a `classification_report` segítségével.