

Adatbányászat és információ keresés

Labor 8,9.

1. Elemezzük és tisztítsuk meg a dbalance.csv fájl adatait. Az attribútumok leírása, az értékek várt eloszlása a dbalance.txt fájlban található.
2. Hány adatunk van és mi az osztályelosztás?
3. Építsünk egy döntési fa és egy naiv Bayes-féle osztályozó modellt az adatokra.
4. Írassuk ki a konfúziós mátrixot, az átlag pontosságot, F1-mértéket a beépített metrika függvények meghívásával: <https://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics>
5. Az osztályozók paraméterezésével, ensemble tanulással, próbáljuk meg javítani a metrikákon.