

Beszámoló

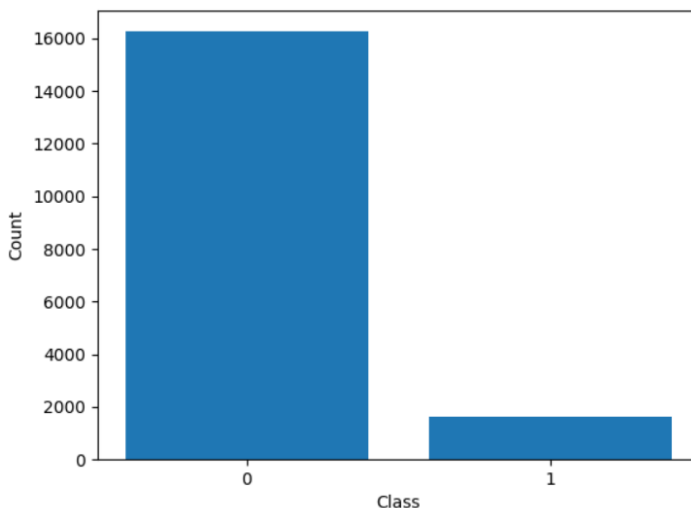
Az eddig végzett munkám a félévben:

A félév elején próbáltam jobban megismerkedni a pulzárokkal, hogyan is „működnek” illetve hogyan lehetne őket hatékonyan azonosítani.

Megcsináltam egy kezdetleges programot pythonban amiben feldolgozom az adatbázist, vizualizálom az adatokat és még egy gépi tanuló algoritmust is tanítottam, névszerint a Random forest osztályozót. Az első eredményeim viszont nem olyanok lettek mint amire én számítottam.

	precision	recall	f1-score	support
0	0.98	0.99	0.99	2439
1	0.94	0.84	0.89	246
accuracy			0.98	2685
macro avg	0.96	0.92	0.94	2685
weighted avg	0.98	0.98	0.98	2685

Az első eredmények meglepően jók lettek, ami gyanút ébresztett bennem. Mivel a kódot alaposan átnéztem, és nem találtam nyilvánvaló hibát (Ettől függetlenül természetesen előfordulhat hogy még is hibáztam valahol), arra a következtetésre jutottam, hogy az adatok eloszlása lehet torzító tényező - különösen a pozitív minták alacsony száma miatt.



Az alábbi ábra mutatja hogy össze hasonlítva mennyi negatív illetve pozitív mintám van. Jelenleg az adatbázist 70-15-15 arányban osztottam fel (tanító, validáló, teszt), de a pozitív minták alacsony száma miatt lehet, hogy ezen később módosítani fogok.

Ezen felül létrehoztam egy github repository-t amit majd folyamatosan frissítgetek a félév előrehaladtával., az alábbi linken érhető el: https://github.com/TheArtist/ThesisWork_PulsarDetectionWithMachineLearning . Egyenlőre csak a kezdetleges python kód és az adatbázis van feltöltve.

További tervek a félévre:

Ebben a félévben a programozás nagy részét meg szeretném csinálni, konkrétan:

- Kipróbálni a lehető legtöbb osztályozó algoritmust mint például Decision Tree, SVM, Logistic Regression stb.

- Ezeken felül szeretnék csinálni a félévben egy neurális hálót erre a feladatra, szerencsére ebben a félévben van pár mesterséges intelligenciás tantárgyam így talán majd azok is segítségemre lesznek.

- Ezen kívül még azon gondolkoztam hogy esetleg egy grafikus user interface-t csinálok a modelleimnek, ahova más userek feltudják tölteni az adatbázisaikat és lefuttatni az egyes modelleken - de ez már inkább ilyen fantázia szerű, ha nagyon jól haladnék esetleg tavasszal.

Emellett folyamatosan keresek új, releváns adatbázisokat, amelyeken tovább tudnék dolgozni, de eddig sajnos nem jártam sikerrel.