Deep Hallucination Classification

Proiectul se concentreaza are ca scop clasificarea imaginilor de tipul "Deep Hallucination" intr-una dintre cele 96 de clase. Cele doua metode de invatare automata utilizate in cadrul acestui proiect sunt Naive Bayes si KNN.

Modelul Naive Bayes

Detalii implemntare:

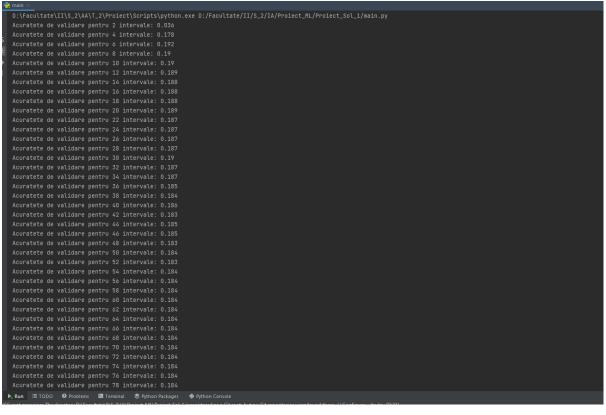
- Fiecare imagine este citita initial sub forma unui vector de pixeli, acest vector fiind de forma 64x64x3, intrucat o imagine este compusa din 64 de linii, fiecare continand 64 de coloane, astfel avand 64x64 de pixeli, iar pentru fiecare dintre acesti pixeli cunoscandu-se cele 3 valori RGB.
- -Aceste imagini sunt transformate intr-un vector 1D pentru a face prelucarea datelor mai usoara, iar apoi sunt adaugate intr-un vector de trasaturi care contine informatiile pentru fiecare imagine in parte, acest procedeu fiind identic pentru toate cele 3 seturi de date, cele de antrenament, de validare si de testare.
- -Valorile din cadrul acestor vectori de trasaturi sunt apoi discretizate, obtinandu-se astfel o forma prelucrata a celor trei vectori mentionati mai sus.
- -Se foloseste algoritmul Multinomial Naive Bayes (din cadrul bibliotecii sklearn) pentru a antrena modelul pe datele de antrenament (vectorul de trasaturi prelucrat si clasele corespunzatoare fiecarei imagini reprezentate in cadrul acestuia).
- -Dupa antrenare, modelul este folosit pentru a prezice clasele imaginilor de validare. Acuratetea modelului este evaluate utilizand datele de validare si este calculata ca raport intre numarul de predictii corecte si numarul total de imagini de validare.
- -Dupa afisarea ratei de acuratete a predictiei facuta pentru imaginile de validare este construita si afisata si matricea de confuzie, mai multe detalii despre aceasta fiind prezentate mai jos.
- -Acelasi procedeu se repeat si pentru imaginile de test, unde modelul nostru incearca in mod similar ca in cazul imaginilor de validare sa prezica clasele corespunzatoare acestora, aceste prezicerii fiind in final scrise intr-un csv.

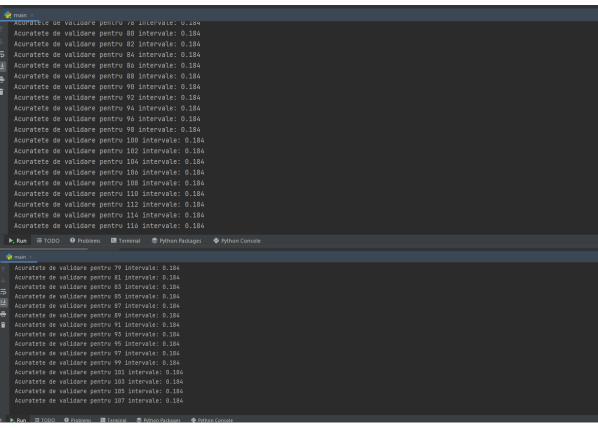
Hypermarameter tunning

De-a lungul numeroaselor rulari ale algoritmului au fost incercate multiple variante pentru parametrii, dar si pentru prelucrarea datelor.

- -In ceea ce priveste prelucrarea datelor, aspectul de care a trebuit tinut cont a fost numarul de interval in care dorim sa se faca discretizarea datelor. In final a fost aleasa valoarea 17, intrucat aceasta obtinea cel mai bun scor in cadrul competitiei, insa un lucru interesant de mentionat ar fi faptul ca valoarea 6 a fost cea alease initial, intrucat in cadrul datelor de validare obtinea acuratetea cea mai mare (ce e drept diferenta fiind una infima de numai 0.001), insa in cadrul competitiei submisia in care numarul de intervale era 6 fiind cu aproximativ 0.01 mai slaba decat cea in care numarul acestora era 17. Mai multe detalii despre acuratetea predictiilor pe testele de validare pot fi observate in cadrul tabelului si imaginilor de mai jos.
- -De asemenea s-a merita precizat si faptul ca s-a incercat si o a doua normalizare, predecesoare celei despre care am vorbit pana acum, in care scadeam din trasaturile unei imagini media acestora, rezultatul fiind mai apoi impartit la deviatia standard a imaginii. Am renuntat repede la aceasta abordare intrucat dupa numai cateva teste s-a observant faptul ca rata de acuratete obtinuta pe testele de validare este una mult mai scazuta.

Mai jos se pot observa date legate de valorile ratele de acuratete obtinute pentru diverse valori ale numarului de interval





```
D:\Facultate\II\s.2\Al\T.2\Sol.1_instanta_2\Scripts\python.exe D:\Facultate\II\s.2\IA\Proiect_NL/Sol_1_instanta_2\Scripts\python.exe D:\Facultate/II\s.2\IA\Proiect_NL/Sol_1_instanta_2\minin.py
Acoustate de validare pentro 3 intervale: 0.185
Acoustate de validare pentro 5 intervale: 0.183
Acoustate de validare pentro 7 intervale: 0.184
Acoustate de validare pentro 7 intervale: 0.186
Acoustate de validare pentro 13 intervale: 0.191
Acoustate de validare pentro 13 intervale: 0.186
Acoustate de validare pentro 13 intervale: 0.180
Acoustate de validare pentro 13 intervale: 0.180
Acoustate de validare pentro 23 intervale: 0.186
Acoustate de validare pentro 23 intervale: 0.188
Acoustate de validare pentro 23 intervale: 0.188
Acoustate de validare pentro 23 intervale: 0.188
Acoustate de validare pentro 23 intervale: 0.186
Acoustate de validare pentro 25 intervale: 0.184
Acoustate de validare pentro 35 intervale: 0
```