



## **2<sup>η</sup> Εργασία Τεχνητής Νοημοσύνης**

**2021-22**

**Ιωάννης Μαστρογιάννης 3190114**

**Δημήτριος Παρασκευιώτης 3190164**

**Διδάσκων: κ. Ίων Ανδρουτσόπουλος**

Επιλέξαμε να υλοποιήσουμε σε Python τους αλγόριθμους μάθησης  
**Αφελής ταξινομητής Bayes και ID3.**

Η ανάπτυξη των 2 αυτών αλγορίθμων έγινε μέσω των προγραμμάτων  
PyCharm και Visual Studio Code

## **Αφελής ταξινομητής Bayes**

Για την ανάπτυξη του παραπάνω αλγορίθμου βασιστήκαμε στις διαφάνειες νούμερο 16 του μαθήματος, στα εργαστήρια και σε διάφορες πηγές που βρήκαμε στο διαδίκτυο.

Ο αλγόριθμος αυτός είναι σχεδιασμένος για Boolean και δυαδικές τιμές , οπότε ιδανικός για τα features μας (θετική(0)/αρνητική(1) κριτική).

Οι παρακάτω πίνακες και τα διαγράμματα έχουν δημιουργηθεί με σύνολο 500 λέξεων στο vocabulary.

Εδώ βλέπουμε τον πίνακα που μας δείχνει το ποσοστό **ορθότητας** (accuracy) στα **δεδομένα εκπαίδευσης** ανά 100 αρχεία

Results	Values		
		23800	0.72292
100	0.64132	23900	0.72276
200	0.67768	24000	0.72284
300	0.69292	24100	0.72264
400	0.6944	24200	0.72276
500	0.69756	24300	0.72268
600	0.70208	24400	0.72268
700	0.7026	24500	0.72248
800	0.70408	24600	0.72244
900	0.707	24700	0.7224
1000	0.70928	24800	0.7224
1100	0.7118	24900	0.72232
1200	0.71444	25000	0.72236

Προφανώς, έχουμε συμπεριλάβει μόνο την αρχή και το τέλος του πίνακα, λόγω του μεγέθους του

Στη συνέχεια βλέπουμε τον πίνακα που μας δείχνει το ποσοστό **ορθότητας** (accuracy) στα **δεδομένα ελέγχου** ανά 100 αρχεία

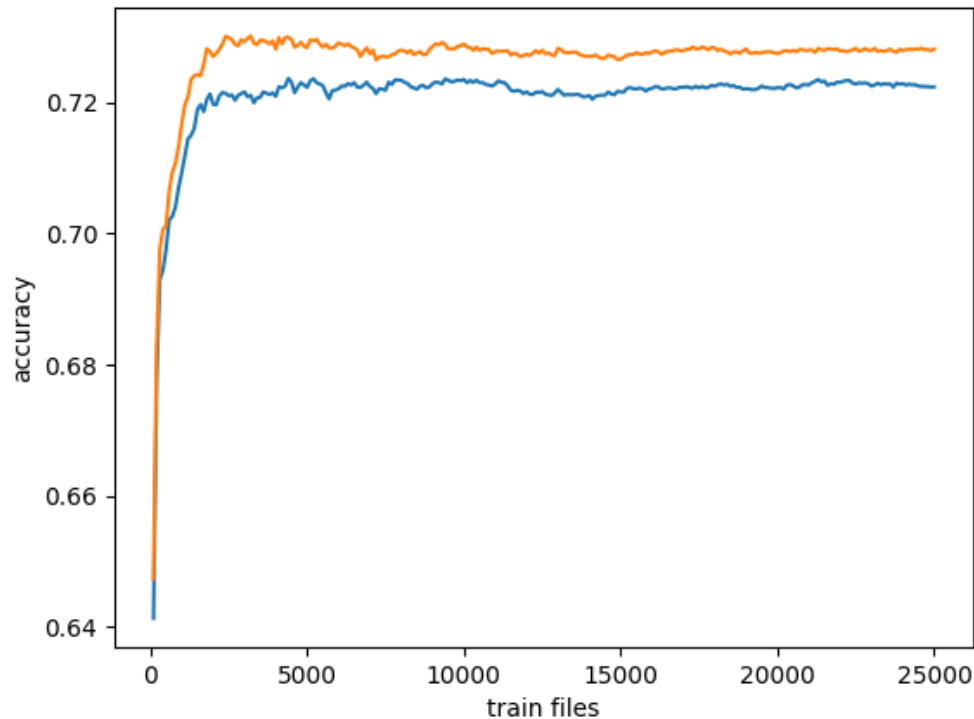
Results	Values
100	0.64732
200	0.68252
300	0.69764
400	0.70064
500	0.70108
600	0.70648
700	0.70932
800	0.71072
900	0.71324
1000	0.71672
1100	0.7196

24000	0.72792
24100	0.728
24200	0.72812
24300	0.728
24400	0.72816
24500	0.72808
24600	0.72828
24700	0.728
24800	0.72808
24900	0.72788
25000	0.72812

Control Run TODO P Process finished with exit code 0 (22 lines long)

Όπως και στον προηγούμενο πίνακα, έτσι και εδώ δεν συμπεριλάβαμε όλα τα δεδομένα του για εύλογους λόγους.

Η καμπύλη μάθησης φαίνεται παρακάτω και έχει ως άξονες το ποσοστό ορθότητας και τον αριθμό των δεδομένων εκπαίδευσης.

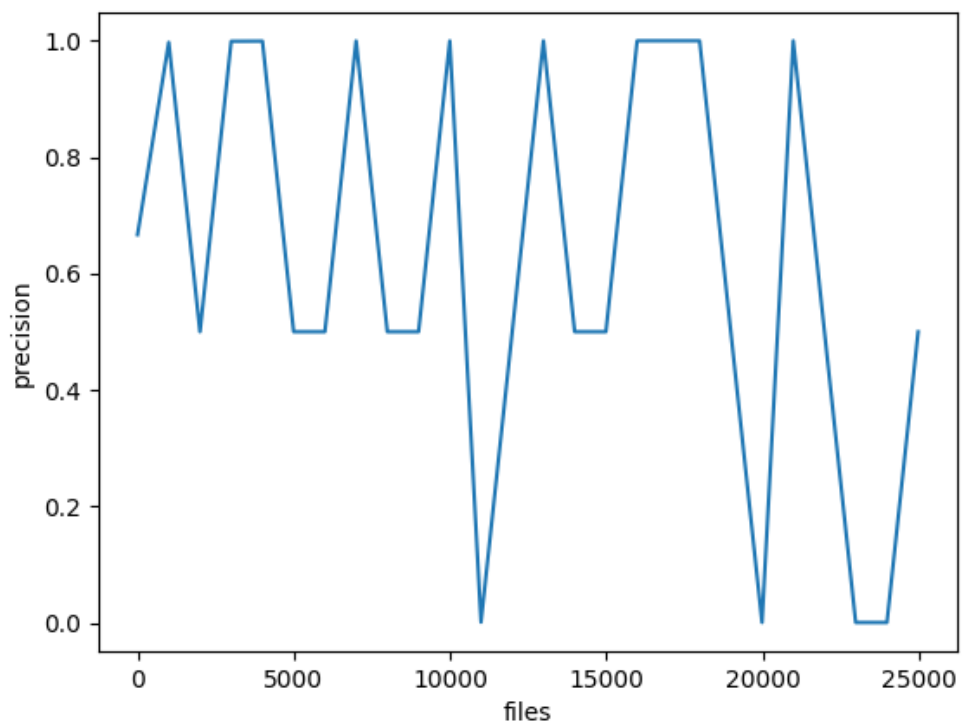


Ο πίνακας με τα αποτελέσματα **ακρίβειας** (precision), **ανάκλησης** (recall), **F1** για την αρνητική κατηγορία, συναρτήσει του πλήθους των παραδειγμάτων εκπαίδευσης είναι ο εξής:

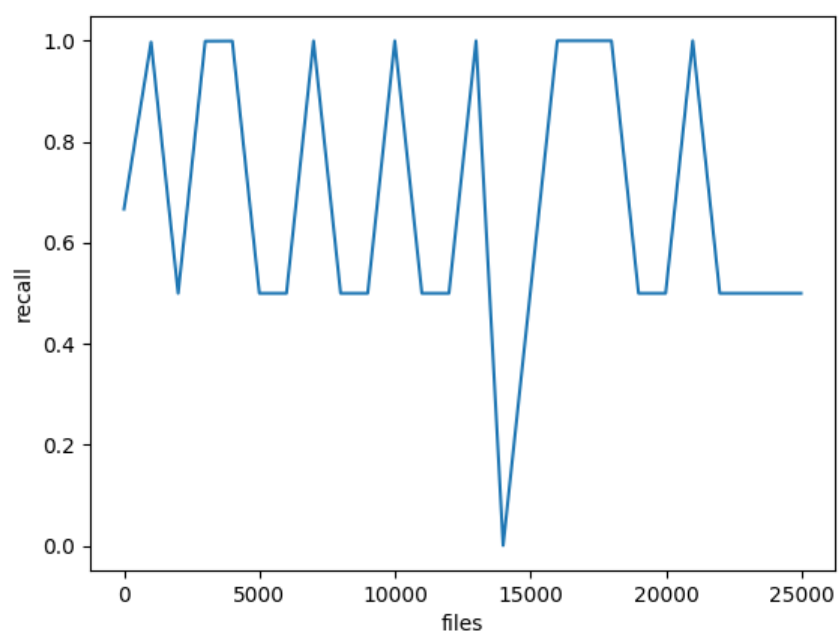
Results	Values
precision	0.5
recall	0.53
f1-score	0.51

Οι αντίστοιχες καμπύλες βρίσκονται παρακάτω.

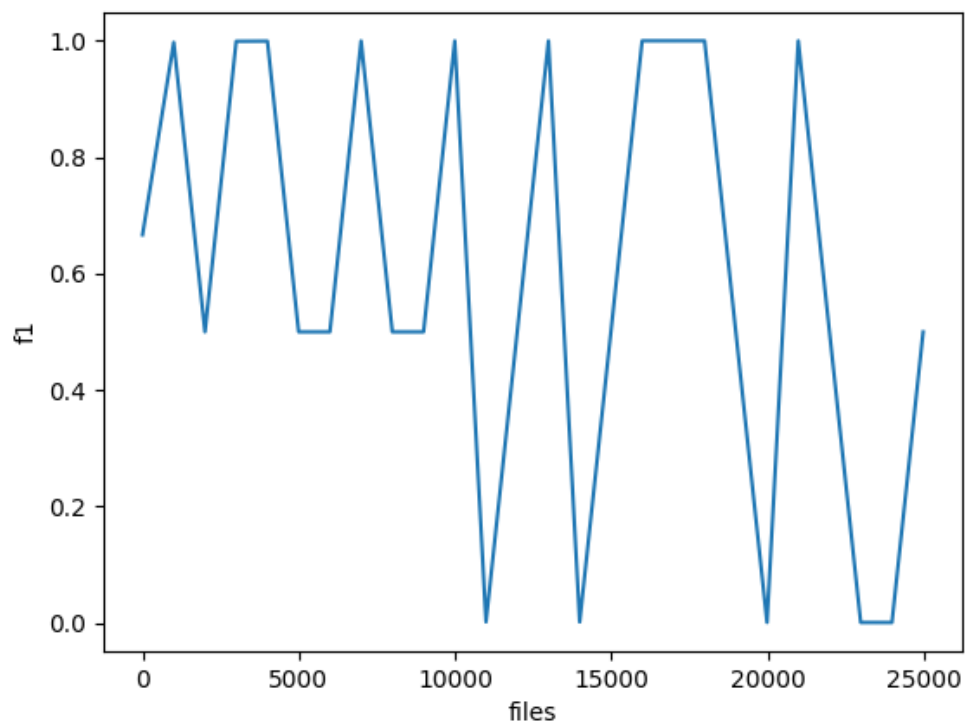
### Καμπύλη ακρίβειας (precision)



### Καμπύλη ανάκλησης (recall)



## Καμπύλη F1



## ID3 Αλγόριθμος

Στον αλγόριθμο που αναπτύξαμε, δεν χρησιμοποιήσαμε πριόνισμα ή πρόωρο τερματισμό της επέκτασης κάθε δέντρου.

Οι παρακάτω πίνακες και τα διαγράμματα έχουν δημιουργηθεί με σύνολο 100 λέξεων στο vocabulary.

Εδώ βλέπουμε τον πίνακα που μας δείχνει το ποσοστό **ορθότητας** (accuracy) στα **δεδομένα εκπαίδευσης** ανά 1000 αρχεία

17000	0.8708	Results	Values
18000	0.8856	0	0.5
19000	0.90356	1000	0.5936
20000	0.9186	2000	0.61132
21000	0.93732	3000	0.63428
22000	0.95008	4000	0.65372
23000	0.96816	5000	0.66948
24000	0.9836	6000	0.68548
		7000	0.70408
Results	Values	8000	0.72096

Control ▶ Run ⋮ TODO ⓘ Control ▶ Run ⋮ TODO ⓘ

pre-built shared indexes; Reduce the pre-built shared indexes; Reduce the



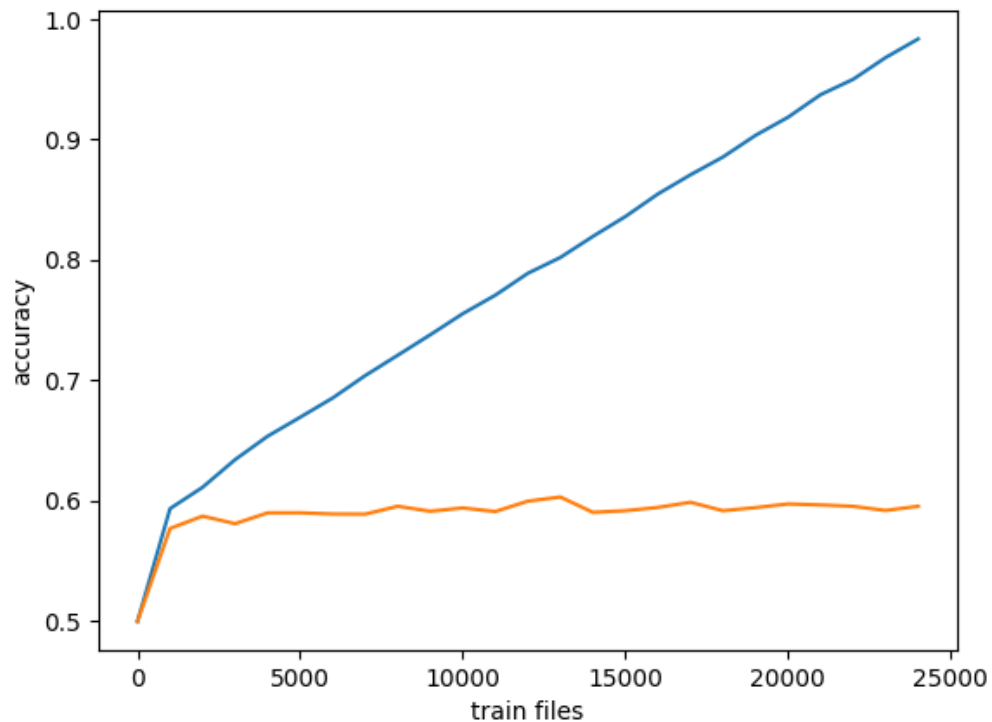
Στη συνέχεια βλέπουμε τον πίνακα που μας δείχνει το ποσοστό **ορθότητας** (accuracy) στα **δεδομένα ελέγχου** ανά 1000 αρχεία

Results	Values
0	0.5
1000	0.57712
2000	0.58736
3000	0.58112
4000	0.59008
5000	0.59012
6000	0.58916
7000	0.58904
8000	0.59556

17000	0.59888
18000	0.59188
19000	0.59444
20000	0.59748
21000	0.59668
22000	0.59552
23000	0.59208
24000	0.59556

Control ▶ Run ⋮ TODO ⓘ Process finished with exit code 0

Η καμπύλη μάθησης φαίνεται παρακάτω και έχει ως άξονες το ποσοστό ορθότητας και τον αριθμό των δεδομένων εκπαίδευσης.

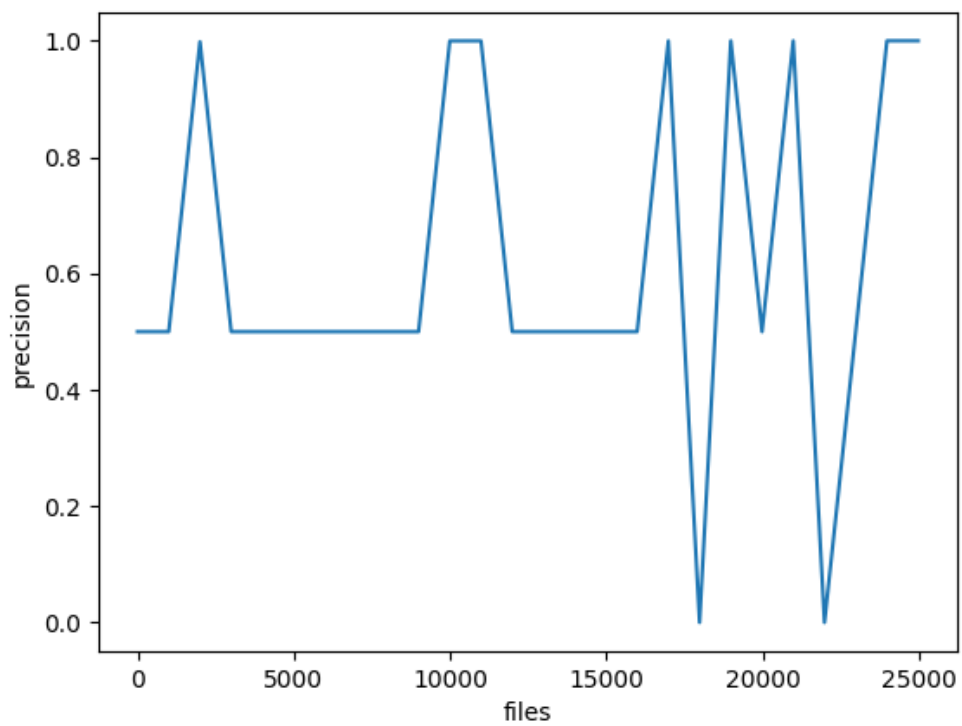


Ο πίνακας με τα αποτελέσματα **ακρίβειας** (precision), **ανάκλησης** (recall), **F1** για την θετική κατηγορία, συναρτήσει του πλήθους των παραδειγμάτων εκπαίδευσης είναι ο εξής:

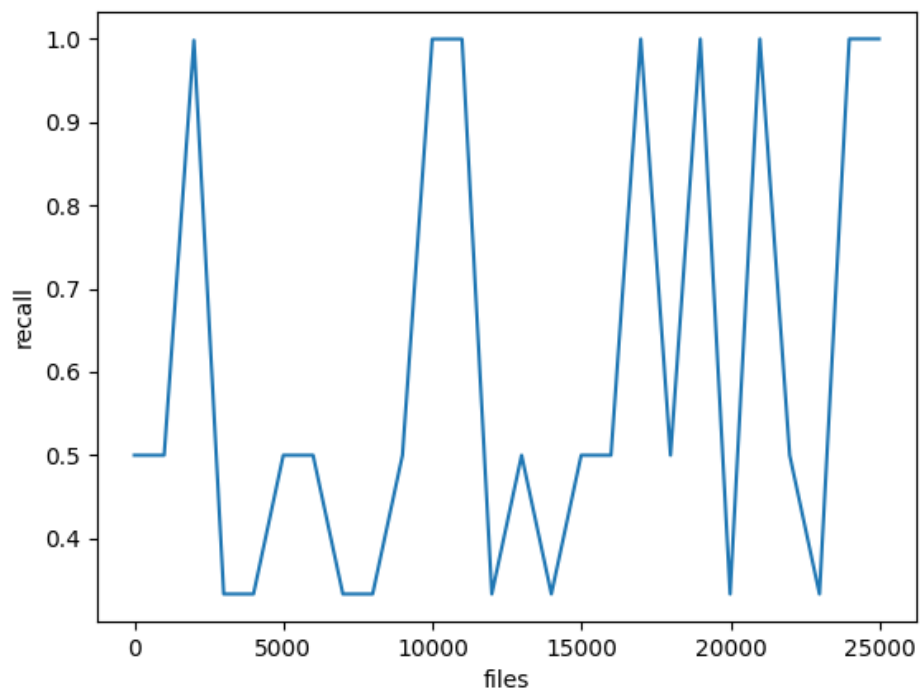
Results	Values
precision	0.999864
recall	0.999864
f1	0.999864

Process finished with exit

### Καμπύλη ακρίβειας (precision)



### Καμπύλη ανάκλησης (recall)



## Καμπύλη F1

