

**2η 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**.

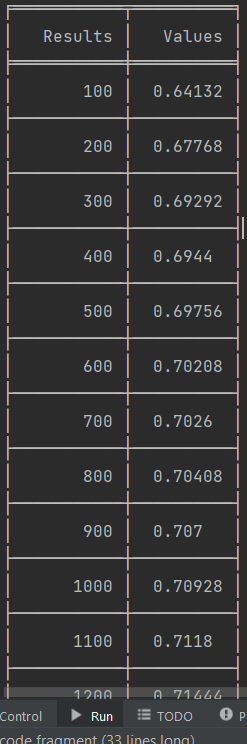
Η ανάπτυξη των 2 αυτών αλγορίθμων έγινε μέσω των προγραμμάτων PyCharm και Visual Studio Code

**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).

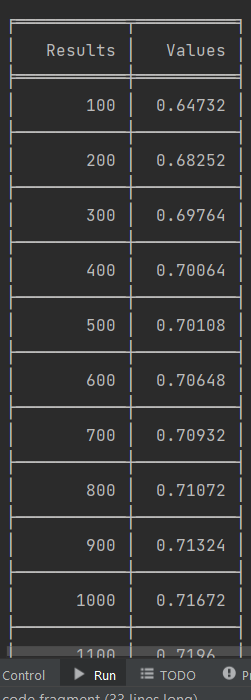
Οι παρακάτω πίνακες και τα διαγράμματα έχουν δημιουργηθεί με σύνολο 500 λέξεων στο vocabulary.

Εδώ βλέπουμε τον πίνακα που μας δείχνει το ποσοστό **ορθότητας** (accuracy) στα **δεδομένα εκπαίδευσης** ανά 100 αρχείαΕικόνα που περιέχει κείμενο, πίνακας αποτελεσμάτων

Περιγραφή που δημιουργήθηκε αυτόματα

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

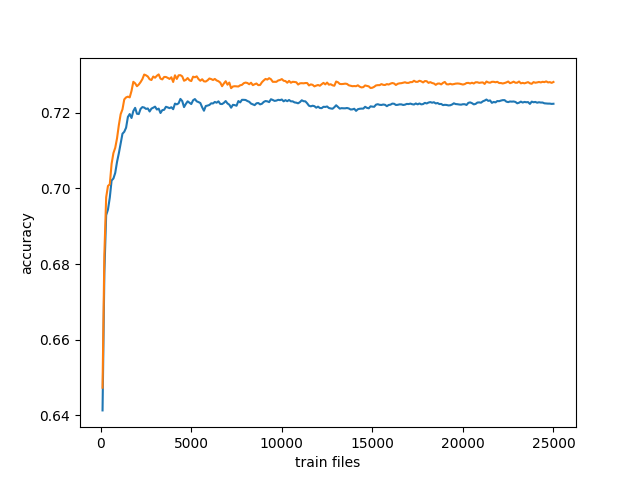
Στη συνέχεια βλέπουμε τον πίνακα που μας δείχνει το ποσοστό **ορθότητας** (accuracy) στα **δεδομένα ελέγχου** ανά 100 αρχεία

Εικόνα που περιέχει κείμενο, πίνακας αποτελεσμάτων

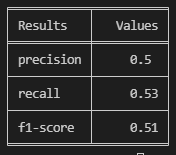
Περιγραφή που δημιουργήθηκε αυτόματα

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

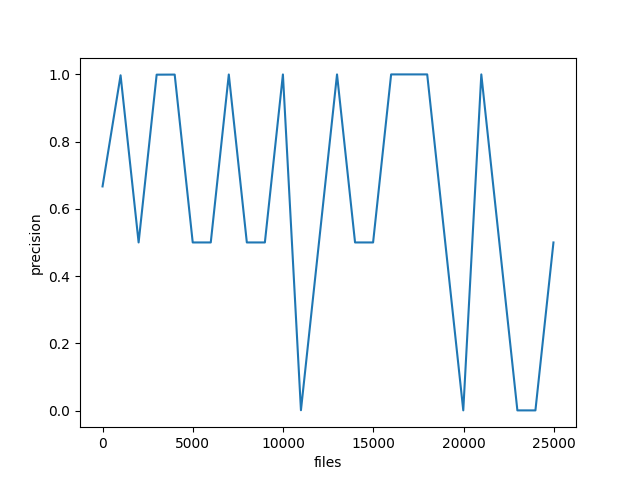
Η καμπύλη μάθησης φαίνεται παρακάτω και έχει ως άξονες το ποσοστό ορθότητας και τον αριθμό των δεδομένων εκπαίδευσης.



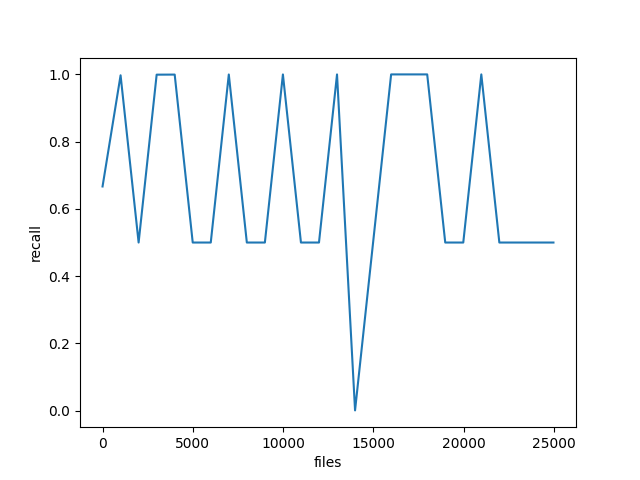
Ο πίνακας με τα αποτελέσματα **ακρίβειας** (precision), **ανάκλησης** (recall), **F1** για την αρνητική κατηγορία, συναρτήσει του πλήθους των παραδειγμάτων εκπαίδευσης είναι ο εξής:

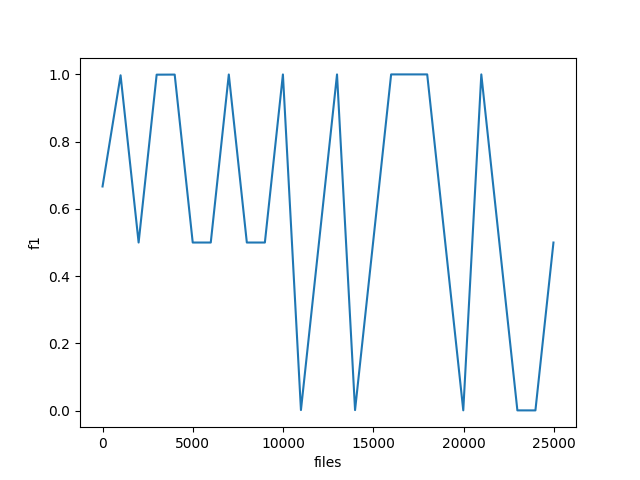


Οι αντίστοιχες καμπύλες βρίσκονται παρακάτω.

**Καμπύλη ακρίβειας (precision)**

**Καμπύλη ανάκλησης (recall)**

****

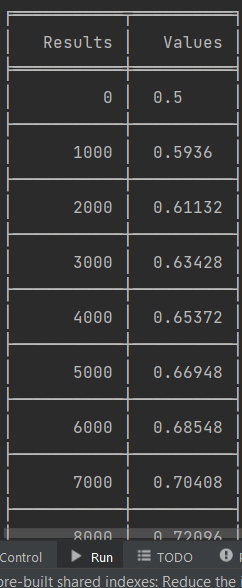
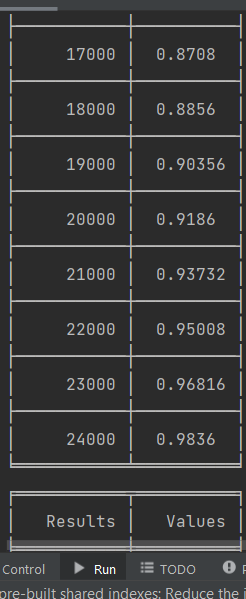
**Καμπύλη F1**

**ID3 Αλγόριθμος**

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

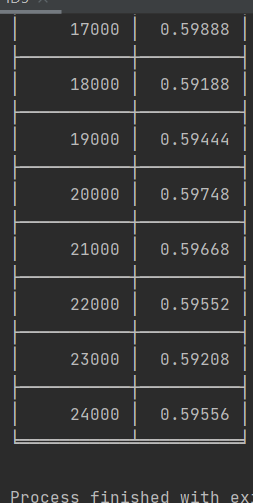
Οι παρακάτω πίνακες και τα διαγράμματα έχουν δημιουργηθεί με σύνολο 100 λέξεων στο vocabulary.

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

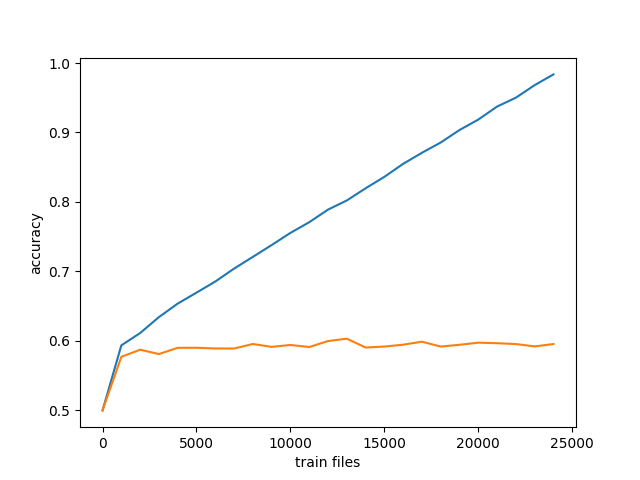


Στη συνέχεια βλέπουμε τον πίνακα που μας δείχνει το ποσοστό **ορθότητας** (accuracy) στα **δεδομένα ελέγχου** ανά 1000 αρχεία

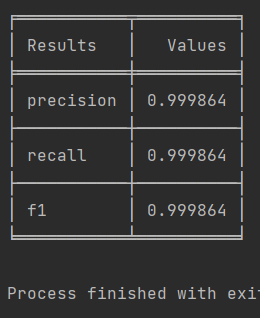
Εικόνα που περιέχει κείμενο, πίνακας αποτελεσμάτων

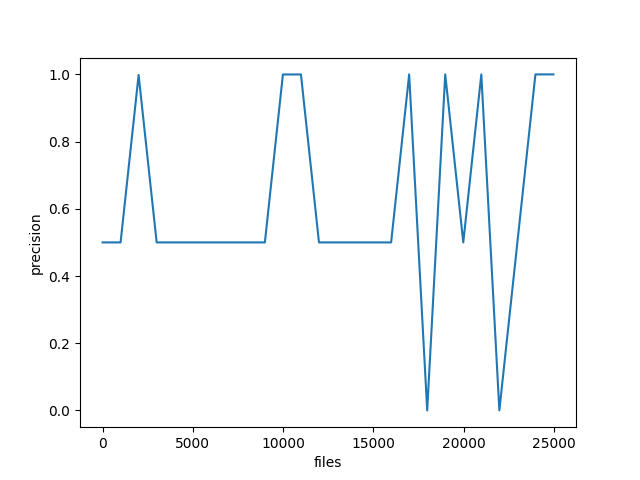
Περιγραφή που δημιουργήθηκε αυτόματα

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

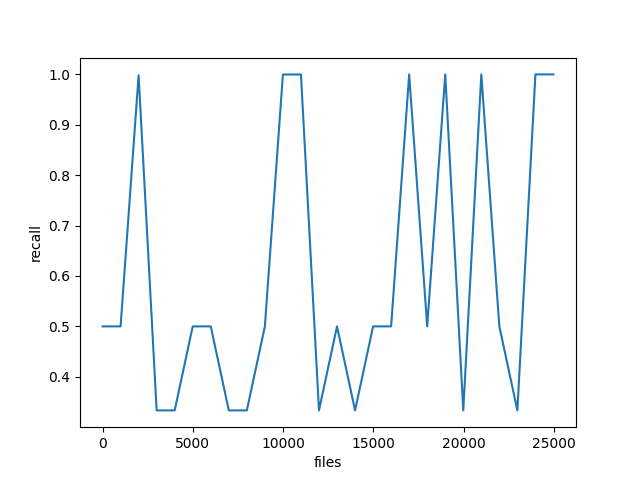


Ο πίνακας με τα αποτελέσματα **ακρίβειας** (precision), **ανάκλησης** (recall), **F1** για την θετική κατηγορία, συναρτήσει του πλήθους των παραδειγμάτων εκπαίδευσης είναι ο εξής:

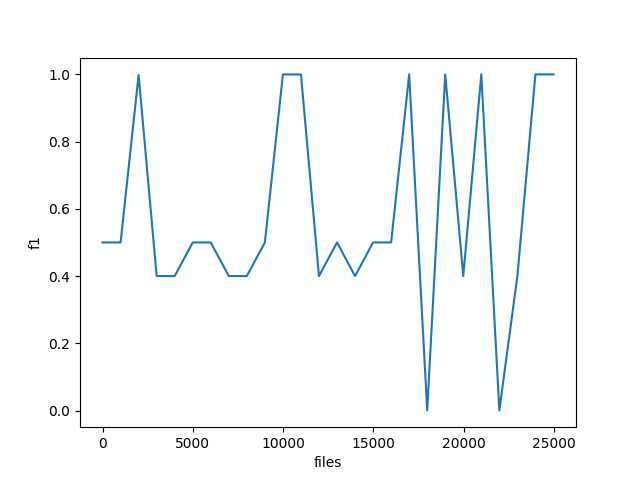


**Καμπύλη ακρίβειας (precision)**

**Καμπύλη ανάκλησης (recall)**

****

**Καμπύλη F1**

****