

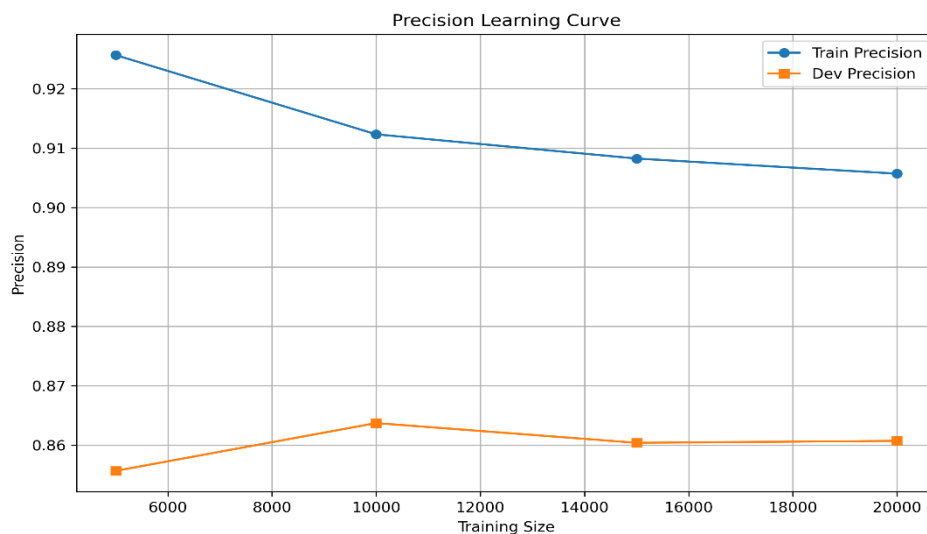
Μέρος Α'

---Αποτελέσματα Εκπαίδευσης/Αξιολόγησης---

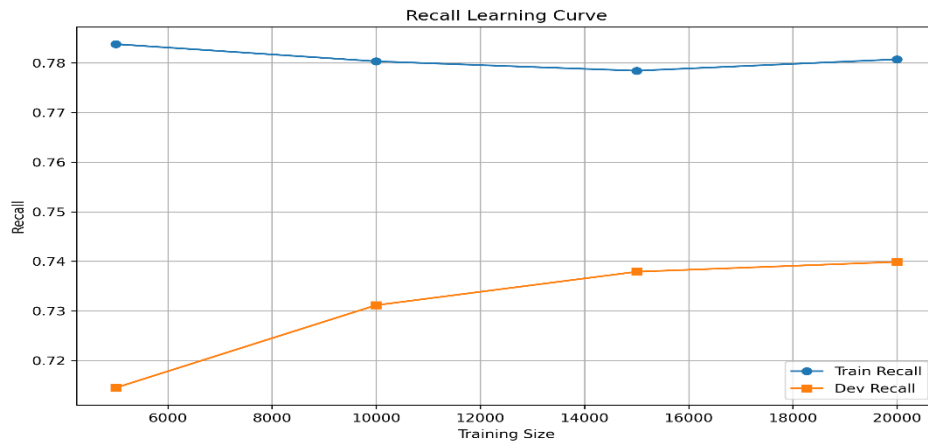
----Επιλογή Υπερπαραμέτρων----

Ανάλυση Αποτελεσμάτων NaiveBayesBernoulli

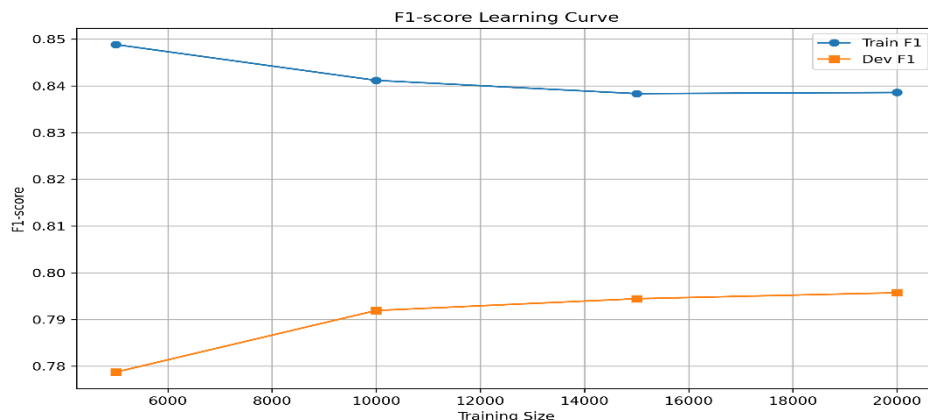
1. Καμπύλες Μάθησης συναρτήσει πλήθους παραδειγμάτων εκπαίδευσης (Θετικές κριτικές)



Παρατηρούμε ότι το **train precision** ξεκινά από υψηλή τιμή (περίπου **0.92**) και μειώνεται σταδιακά καθώς το μέγεθος των δεδομένων εκπαίδευσης αυξάνεται, φτάνοντας κοντά στο **0.90**. Αυτή η μείωση είναι αναμενόμενη, καθώς το μοντέλο γίνεται λιγότερο υπερπροσαρμοσμένο (overfitting) στα λίγα αρχικά δεδομένα. Αντίθετα, το **development precision** ξεκινά από χαμηλότερη τιμή (περίπου **0.86**) και βελτιώνεται ελαφρώς, σταθεροποιούμενο γύρω στο **0.86**. Η σταθεροποίηση αυτή δείχνει ότι το μοντέλο **διατηρεί καλή ισορροπία μεταξύ εκπαίδευσης και γενίκευσης**.



Παρατηρούμε ότι η **εκπαίδευση ξεκινά από περίπου 0.78** και παρουσιάζει μικρές διακυμάνσεις, καταλήγοντας κοντά στο **0.77**. Το **development recall** ξεκινά από **0.72**, αυξάνεται σταδιακά και σταθεροποιείται γύρω στο **0.74**. Αυτή η αύξηση του recall στο dev set δείχνει ότι το μοντέλο **βελτιώνεται στην ανίχνευση θετικών παραδειγμάτων καθώς μαθαίνει από περισσότερα δεδομένα**.



Το **F1-score**, το οποίο είναι ο σταθμισμένος μέσος όρος του Precision και του Recall, ακολουθεί παρόμοιο μοτίβο με τις δύο προηγούμενες καμπύλες. Το **train F1** ξεκινά κοντά στο **0.85** και μειώνεται ελαφρώς, φτάνοντας περίπου **0.83**. Αντίθετα, το **dev F1-score** ξεκινά από **0.78**, αυξάνεται σταθερά και σταθεροποιείται κοντά στο **0.80**. Αυτή η αύξηση είναι θετική, καθώς δείχνει ότι το μοντέλο **βελτιώνεται στην ισορροπία μεταξύ Precision και Recall** στο development set.

2. Πίνακας γενικής αξιολόγησης:

Μετρική	Τιμή
Test Class 0 Precision	0.7676
Test Class 0 Recall	0.8718
Test Class 0 F1-score	0.8164
Test Class 1 Precision	0.8517
Test Class 1 Recall	0.7360
Test Class 1 F1-score	0.7896
Test Macro Precision	0.8096
Test Macro Recall	0.8039
Test Macro F1-score	0.8030
Test Micro Precision / Recall / F1-score	0.8039

Για την Κλάση 0 (Αρνητικές Κριτικές):

- Το **Precision (76.76%)** είναι χαμηλότερο από το Recall (87.18%), πράγμα που σημαίνει ότι το μοντέλο εντοπίζει πολλές αρνητικές κριτικές, αλλά κάνει και κάποιες **λανθασμένες προβλέψεις**.
- Το **F1-score (81.64%)** δείχνει ότι το μοντέλο έχει **καλή συνολική απόδοση στην αρνητική κλάση**, αλλά υπάρχει κάποια ασυμμετρία λόγω του υψηλότερου Recall.

Για την Κλάση 1 (Θετικές Κριτικές):

- Το **Precision (85.17%)** είναι υψηλότερο από το Recall (73.60%), πράγμα που σημαίνει ότι το μοντέλο είναι πιο επιλεκτικό στις θετικές προβλέψεις και αποφεύγει πολλά false positives, αλλά χάνει κάποιες θετικές κριτικές.

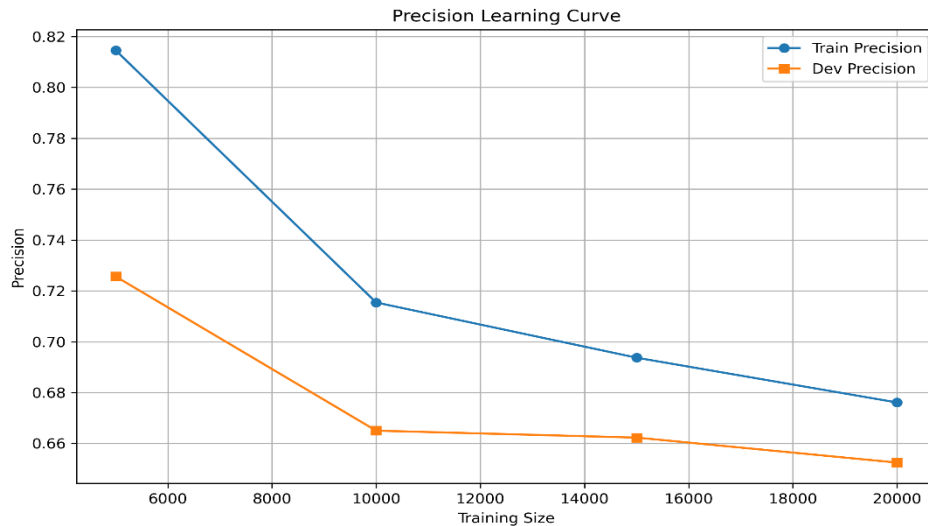
- Το **F1-score** (78.96%) δείχνει ότι υπάρχει μία σχετική ανισορροπία μεταξύ Precision και Recall.
- Το **Macro Precision (80.96%)** δείχνει ότι το μοντέλο έχει **καλή ακρίβεια συνολικά**, αλλά με μικρές διαφορές μεταξύ των δύο κλάσεων.
- Το **Macro Recall (80.39%)** είναι ελαφρώς χαμηλότερο, γεγονός που δείχνει ότι το μοντέλο δεν εντοπίζει όλες τις θετικές και αρνητικές κριτικές με την ίδια απόδοση.
- Το **Micro Precision, Recall και F1-score (80.39%)** είναι ίδια μεταξύ τους, κάτι που υποδεικνύει **ισορροπία μεταξύ των συνολικών σωστών προβλέψεων**.

- **Επιλογή Υπερ-Παραμέτρων:**

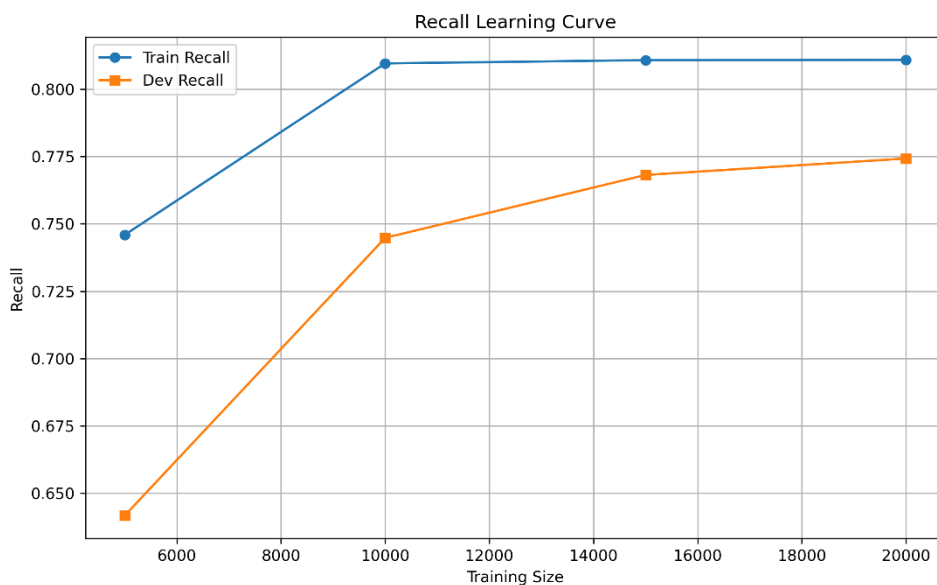
Δεδομένου ότι η συγκεκριμένος αλγόριθμος είναι ο πιο απλοϊκός στην εκπαίδευση και στην πρόβλεψη, πραγματοποιήσαμε πολλές δοκιμές σχετικά με τις τιμές των n πιο συχνών και των k πιο σπάνιων λέξεων καθώς και των m λέξεων με το υψηλότερο πληροφοριακό κέρδος. Αρχικά, λόγω του πολύ μεγάλου όγκου δεδομένων, ένα πολύ μεγάλο τελικό λεξιλόγιο θα έκανε χρονικά απαγορευτική την υλοποίηση του αλγορίθμου (και πόσο μάλλον των επόμενων) συνεπώς πειραματιστήκαμε με τιμές κοντά στο 5000, μικρές αυξομειώσεις (± 1000) δεν είχαν μεγάλη επιρροή στην τελική απόδοση, οπότε επιλέξαμε τις 5000 και για την εξοικονόμηση χώρου (αντί 6000 ή 7000) και για λίγο περισσότερη ακρίβεια (αντί 4000). Επίσης, και οι άλλες τιμές των δύο παραμέτρων, αρχικά δοκιμάστηκαν με πολύ μεγάλα νούμερα (15000 και 10000) οι οποίες μείωναν σημαντικά το πλήθος των χρήσιμων λέξεων και έδιναν τελική απόδοση περίπου στο 50%, πράγμα που καθιστούσε τον ταξινομητή σα τυχαίο. Συνεπώς μετά από αρκετές δοκιμές καταλήξαμε στην αφαίρεση των 5000 πιο συχνών και των 1000 πιο σπάνιων λέξεων.

Ανάλυση Αποτελεσμάτων Random Forest

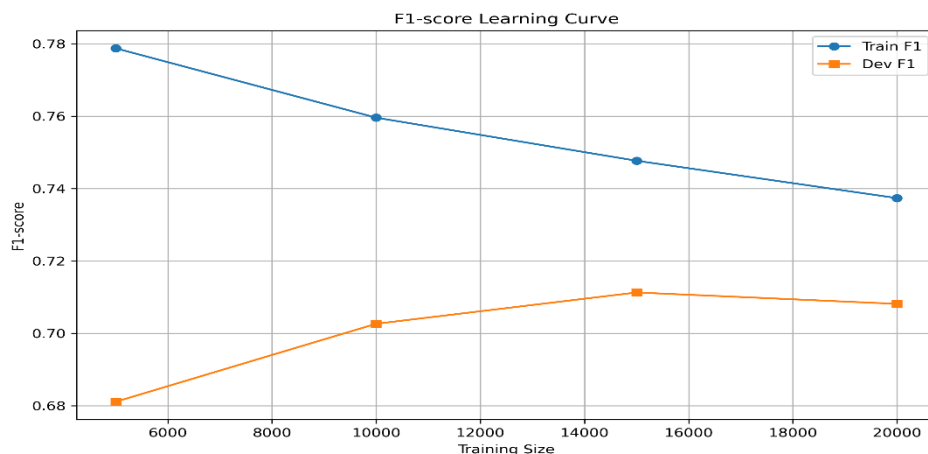
1. Καμπύλες Μάθησης συναρτήσεως πλήθους παραδειγμάτων εκπαίδευσης (Θετικές κριτικές)



Παρατηρούμε ότι η ακρίβεια του μοντέλου στην εκπαίδευση (Train Precision) ξεκινά από περίπου 0.82 αλλά μειώνεται σταδιακά καθώς αυξάνεται το μέγεθος των δεδομένων εκπαίδευσης, φτάνοντας κάτω από 0.74. Αντίστοιχα, η ακρίβεια στο Development Set ξεκινά από 0.72 και μειώνεται σταδιακά, καταλήγοντας γύρω στο 0.66.



Παρατηρούμε ότι το Train Recall ξεκινά από 0.75 και αυξάνεται σταδιακά, φτάνοντας κοντά στο 0.81 καθώς αυξάνονται τα δεδομένα εκπαίδευσης. Αντίθετα, το Development Recall ξεκινά από 0.65, βελτιώνεται σταδιακά και φτάνει κοντά στο 0.75. Η αύξηση του Recall στο Development Set υποδεικνύει ότι το μοντέλο καταφέρνει να εντοπίσει περισσότερες θετικές κριτικές καθώς εκπαιδεύεται με περισσότερα δεδομένα.



Παρατηρούμε ότι η καμπύλη του F1-score, που ισορροπεί το Precision και το Recall, ακολουθεί μια ενδιάμεση συμπεριφορά. Το Train F1-score ξεκινά από 0.78 και μειώνεται ελαφρώς στο 0.74, ενώ το Development F1-score ξεκινά από 0.68, αυξάνεται ελαφρώς στο 0.70, αλλά στη συνέχεια μειώνεται. Η τάση αυτή υποδηλώνει ότι το μοντέλο γενικεύει καλύτερα όταν αυξάνεται το μέγεθος του dataset, αλλά η πτώση του Precision επηρεάζει αρνητικά το τελικό F1-score.

2. Πίνακας γενικής αξιολόγησης:

Μετρική	Τιμή
Test Class 0 Precision	0.7354
Test Class 0 Recall	0.5939
Test Class 0 F1-score	0.6571
Test Class 1 Precision	0.6449

Μετρική	Τιμή
Test Class 1 Recall	0.7689
Test Class 1 F1-score	0.7013
Test Macro Precision	0.6901
Test Macro Recall	0.6815
Test Macro F1-score	0.6793
Test Micro Precision / Recall / F1-score 0.6815	

Για την Κλάση 0 (Αρνητικές Κριτικές):

- Το **Precision (73.54%)** είναι υψηλό, που σημαίνει ότι όταν το μοντέλο προβλέπει μια αρνητική κριτική, συνήθως είναι σωστό.
- Το **Recall (59.39%)** είναι χαμηλό, που σημαίνει ότι το μοντέλο δεν εντοπίζει αρκετές αρνητικές κριτικές.
- Το **F1-score (65.71%)** δείχνει μια **ισορροπία**, αλλά το χαμηλό Recall επηρεάζει αρνητικά την επίδοση.

Για την Κλάση 1 (Θετικές Κριτικές):

- Το **Precision (64.49%)** είναι χαμηλότερο από την Κλάση 0, δείχνοντας ότι υπάρχουν αρκετές **λανθασμένες θετικές προβλέψεις**.
- Το **Recall (76.89%)** είναι υψηλό, που σημαίνει ότι το μοντέλο καταφέρνει να εντοπίσει τις περισσότερες θετικές κριτικές.
- Το **F1-score (70.13%)** είναι υψηλότερο από την Κλάση 0, που σημαίνει ότι η ισορροπία μεταξύ Precision και Recall είναι καλύτερη εδώ.
- Το **Macro Precision (69.01%)** είναι χαμηλότερο από το **Macro Recall (68.15%)**, δείχνοντας ότι το μοντέλο

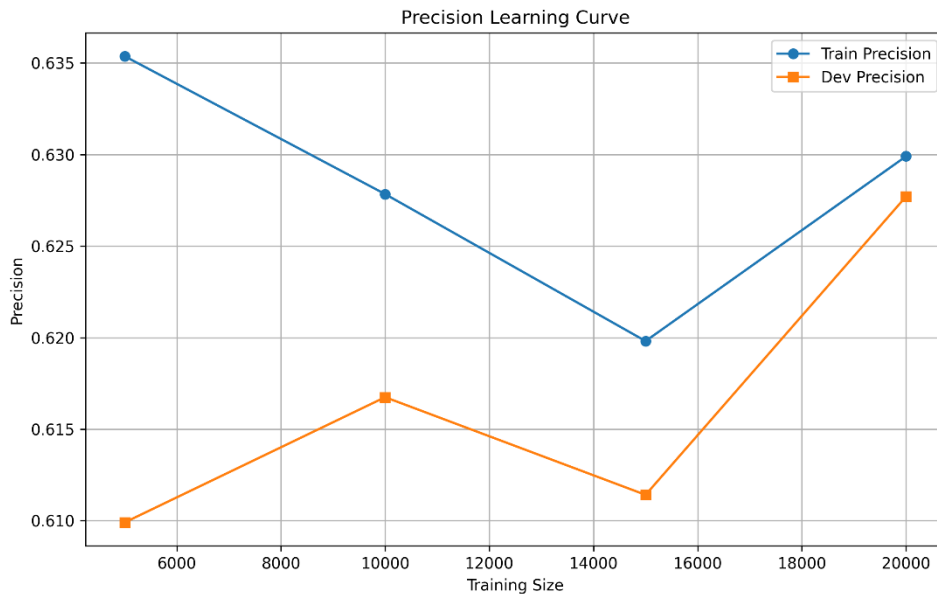
εντοπίζει περισσότερα παραδείγματα από όσα ταξινομεί σωστά.

- Το **Micro Precision / Recall / F1-score (68.15%)** είναι σχετικά χαμηλό, κάτι που δείχνει ότι το μοντέλο έχει **μέτρια συνολική απόδοση**.
- **Επιλογή Υπερ-Παραμέτρων:**

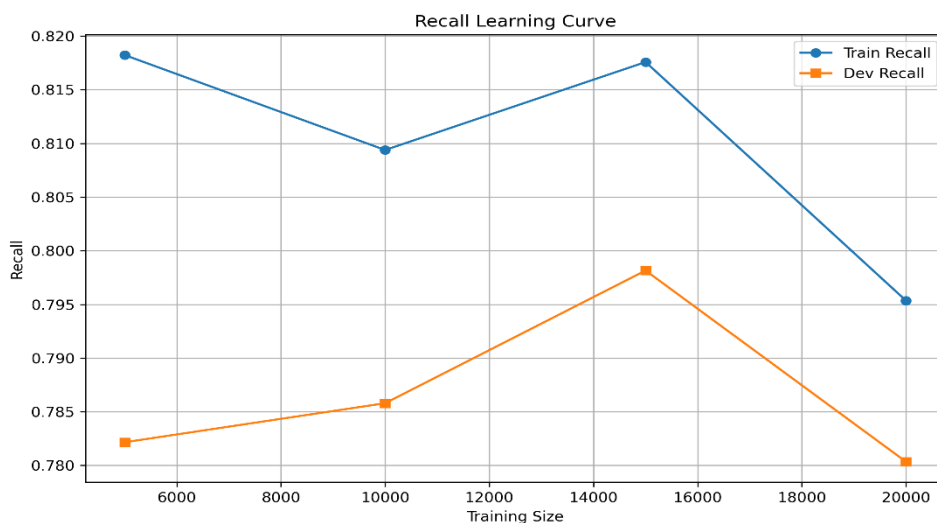
Εφόσον καταφέραμε να πετύχουμε αρκετά καλές επιδόσεις με τις τιμές των n πιο συχνών και των k πιο σπάνιων λέξεων καθώς και των m λέξεων με το υψηλότερο πληροφοριακό κέρδος, στον αλγόριθμο του naïve bayes, διατηρήσαμε τις ίδιες τιμές και σε αυτή την υλοποίηση. Για το πλήθος των δέντρων στο δάσος, κάναμε δοκιμές με διαφορετικές τιμές, οι χαμηλότερες τιμές από το 500 που δοκιμάσαμε (300, 150), μείωναν σημαντικά την απόδοση, ενώ τιμές μεγαλύτερες του 500 (700, 1000) έκαναν την υλοποίηση απαγορευτικά αργή. Επίσης, για το μέγιστο βάθος κάθε δέντρου ακολουθήσαμε την ίδια πρακτική, τιμές μικρότερες του 10 μείωναν πολύ την απόδοση και τιμές άνω του 10 πρόσθεταν επιπλέον χρόνο και δεν βελτίωναν την επίδοση. Το μοντέλο, σε μεγάλες τιμές των άνω δύο παραμέτρων παρουσίασε υπερπροσαρμογή. Τέλος, για την παράμετρο που καθορίζει τον αριθμό των χαρακτηριστικών που εξετάζονται σε κάθε διαχωρισμό του δέντρου (`max_features`), πραγματοποιώντας μια [αναζήτηση στο διαδίκτυο](#) βρήκαμε ότι για προβλήματα ταξινόμησης/κατηγοριοποίησης.

Ανάλυση Αποτελεσμάτων AdaBoost

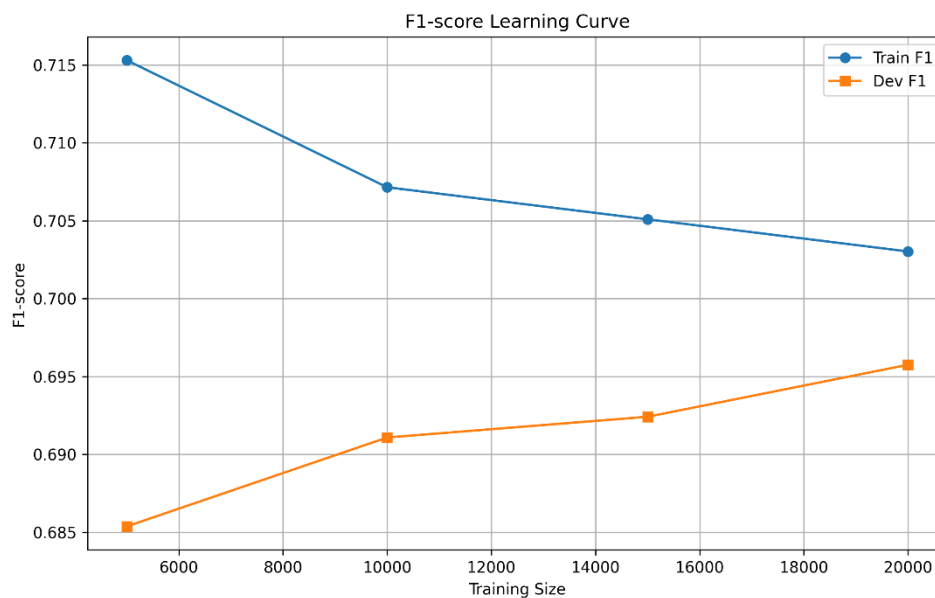
1. Καμπύλες Μάθησης συναρτήσεως πλήθους παραδειγμάτων εκπαίδευσης (Θετικές κριτικές)



Παρατηρούμε ότι στην αρχή, το **precision** του train set είναι υψηλό, αλλά μειώνεται καθώς αυξάνονται τα δεδομένα εκπαίδευσης, δείχνοντας ότι το μοντέλο γίνεται λιγότερο βέβαιο στις θετικές προβλέψεις του. Το dev precision αυξάνεται σταθερά, κάτι που δείχνει ότι το μοντέλο βελτιώνεται στη σωστή αναγνώριση των θετικών κριτικών.



Παρατηρείται πτωτική τάση στο train recall, αλλά το dev recall παρουσιάζει αύξηση μέχρι τα 15.000 δείγματα και στη συνέχεια μειώνεται, γεγονός που υποδηλώνει ότι το μοντέλο έχει μια σχετική σταθερότητα αλλά δεν βελτιώνεται σημαντικά μετά από ένα σημείο.



Παρατηρείται μια πτωτική τάση στο **train F1-score**, γεγονός που δείχνει ότι το μοντέλο χάνει κάποια ικανότητα να προσαρμόζεται στα δεδομένα εκπαίδευσης καθώς αυξάνεται το μέγεθος του dataset. Το dev F1-score αυξάνεται σταθερά, γεγονός που δείχνει βελτίωση της γενίκευσης.

2. Πίνακας γενικής αξιολόγησης:

Μετρική	Τιμή
Test Class 0 Precision	0.7286
Test Class 0 Recall	0.5418
Test Class 0 F1-score	0.6215
Test Class 1 Precision	0.6183
Test Class 1 Recall	0.7768

Μετρική	Τιμή
Test Class 1 F1-score	0.6886
Test Macro Precision	0.6735
Test Macro Recall	0.6593
Test Macro F1-score	0.6550
Test Micro Precision / Recall / F1-score 0.6593	

Για την Κλάση 0 (Αρνητικές Κριτικές):

- Η ακρίβεια (**precision 77.68%**) σημαίνει ότι όταν το μοντέλο προβλέπει μια αρνητική κριτική (Class 0) προβλέπει μια αρνητική κριτική, συνήθως είναι σωστό.
- Η ανάκληση (**recall 54.18%**) είναι χαμηλή, κάτι που σημαίνει ότι το μοντέλο χάνει αρκετές πραγματικές αρνητικές κριτικές (δηλαδή ταξινομεί αρκετές αρνητικές κριτικές λανθασμένα ως θετικές).
- Το F1-score (**62.15%**) είναι χαμηλότερο από το αντίστοιχο της Κλάσης 1, επιβεβαιώνοντας ότι η ταξινόμηση των αρνητικών κριτικών δεν είναι τόσο αποτελεσματική.

Για την Κλάση 1 (Θετικές Κριτικές):

- Η ακρίβεια (**precision 61.83%**) δείχνει ότι όταν το μοντέλο ταξινομεί μια κριτική ως θετική, έχει δεν έχει πολύ μεγάλη πιθανότητα να είναι σωστό. Αυτή η τιμή είναι χαμηλότερη από την ακρίβεια της Κλάσης 0, κάτι που σημαίνει ότι το μοντέλο κάνει περισσότερα λάθη στις θετικές προβλέψεις (false positives).
- Η ανάκληση (**recall 77.68%**) είναι υψηλή, που σημαίνει ότι το μοντέλο αναγνωρίζει το των πραγματικών θετικών κριτικών. Αυτό δείχνει ότι το μοντέλο δίνει προτεραιότητα στη μη απώλεια των θετικών δειγμάτων, έστω και αν κάνει κάποια λάθη.

- Το **F1-score (68.86%)** είναι καλύτερο από το F1 της Κλάσης 0, γεγονός που δείχνει ότι η ταξινόμηση των θετικών κριτικών είναι πιο ισορροπημένη.
- Το **macro precision (67.35%)** δείχνει ότι, κατά μέσο όρο, το μοντέλο κατά περίπου 70% να κάνει σωστές θετικές προβλέψεις.
- Το **macro recall** σημαίνει ότι το μοντέλο κατά μέσο όρο ανακτά το 65.93% των πραγματικών θετικών και αρνητικών περιπτώσεων.
- Το **macro F1-score (0.6550)** είναι σχετικά χαμηλό, δείχνοντας ότι η ισορροπία μεταξύ precision και recall δεν είναι τόσο καλή.

Τα **micro precision, recall και F1-score έχουν την ίδια τιμή (0.6593)**, κάτι που υποδηλώνει ότι η συνολική απόδοση του ταξινομητή είναι σταθερή. Ο ταξινομητής δεν έχει μεγάλη ασυμμετρία στις κατηγορίες.

• Επιλογή Υπερ-Παραμέτρων:

Όπως και πριν, για τους ίδιους λόγους, έχουμε διατηρήσει τις τιμές των n πιο συχνών και των k πιο σπάνιων λέξεων καθώς και των m λέξεων με το υψηλότερο πληροφοριακό κέρδος. Για το πλήθος των αδύναμων ταξινομητών, κάναμε δοκιμές με διαφορετικές τιμές, οι χαμηλότερες τιμές από το 800 που δοκιμάσαμε (500, 300, 100), μείωναν σημαντικά την απόδοση, ενώ τιμές μεγαλύτερες του 800 (1000) έκαναν την υλοποίηση απαγορευτικά αργή. Ο λόγος που επιλέξαμε τόσους πολλούς ταξινομητές είναι διότι το μέγιστο βάθος κάθε ένα περιοριζόταν από την εκφώνηση συνεπώς έπρεπε να κερδίσουμε σε κάτι άλλο. Δεδομένου ότι κάθε ταξινομητής με την σειρά του επιλέγει τα features με το μεγαλύτερο κέρδος πληροφορίας ακόμα και αυτή η τιμή των ταξινομητών έκανε υπερβολικά αργό όλη την διαδικασία με αποτέλεσμα να μην έχουμε πολλά περιθώρια ρυθμίσεως της τιμής αυτής με σκοπό καλύτερη απόδοση.