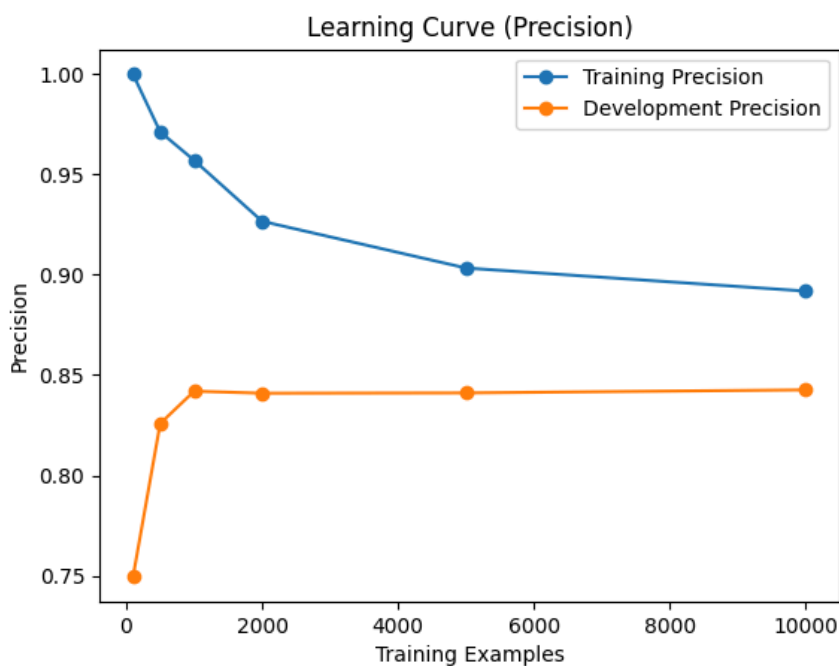
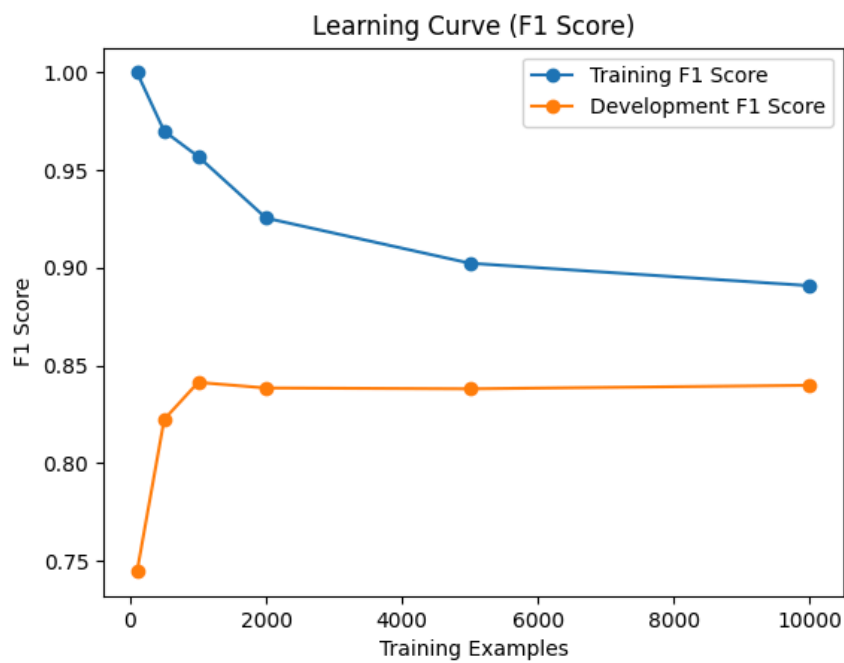
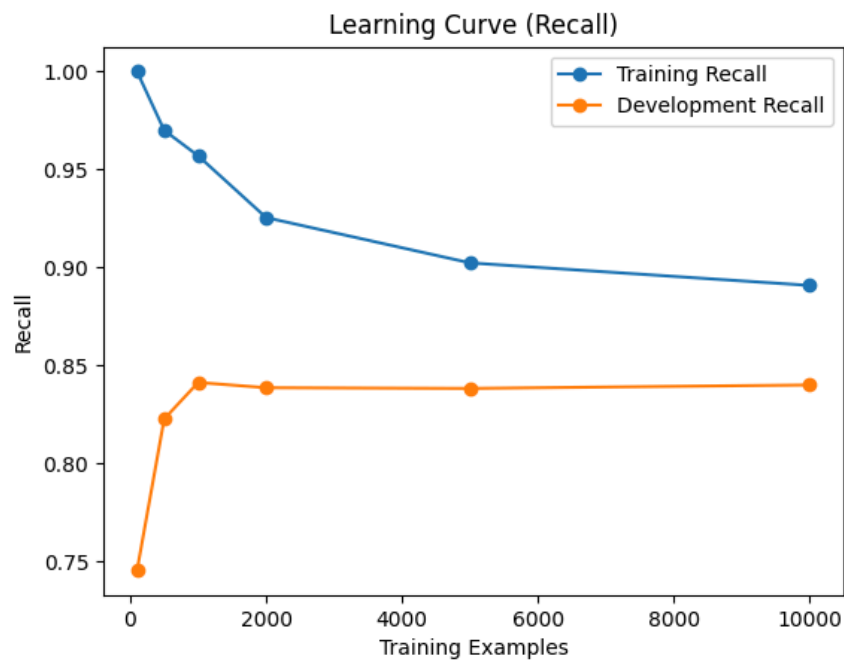


Αφελής Ταξινομητής Bayes (πολυμεταβλητή μορφή Bernoulli): Μετά την δημιουργία του vocabulary απο την μέθοδο DatabaseLoader που περιγράφηκε παραπάνω κάνουμε ένα initialise τον Bayes με το δομημένο vocabulary απο την μεθοδο. Η μέθοδος train υπολογίζει τα θετικά και αρνητικά δεδομένα και θέτει conditional probabilities. Η μέθοδος predict καθορίζει την τάξη του κάθε δεδομενου και επιστρέφει 1 αν τα θετικά > αρνητικά και 0 αν τα αρνητικά >= θετικά για κάθε λέξη στο vocabulary. Η μέθοδος evaluate υπολογίζει δεδομένα για τον ταξινομητή όπως το precision, recall and F1 ξεχωριστά για τα θετικά και τα αρνητικά όπως και τα micro-averaged και macro-averaged δεδομένα και με αυτά αξιολογεί το πως τρέχει ο αλγόριθμος. Προφανώς, η plot_learning_curve χρησιμοποιείται για να φτιάξει τα διαγράμματα για precision, recall, και F1. Τέλος, η main μέθοδος κάνει load τα reviews, δημιουργεί και οριστικοποιεί το vocabulary και ξεκινάει το training με 100, 200, 500, 1000, 2000, 5000 και 10000 δεδομένα. Όταν ολοκληρωθεί το training ξεκινάει το evaluation όπου υπολογίζει όλα τα προαναφερόμενα στατιστικά και εκτυπώνει τα plots.

Παρακάτω βλέπουμε τα διαγράμματα για το precision, recall και F1:





Classification Report: (k=100, n=100, m=5000)

	precision	recall	F1	support
Class 0	0.81	0.88	0.85	12500
Class 1	0.87	0.80	0.83	12500

accuracy			0.84	25000
macro-avg	0.84	0.84	0.84	25000
weighted-avg	0.84	0.84	0.84	25000

Accuracy: 0.8401

Macro-averaged Precision: 0.8426

Macro-averaged Recall: 0.8401

Macro-averaged F1 Score: 0.8398

Micro-averaged Precision: 0.8401

Micro-averaged Recall: 0.8401

Micro-averaged F1 Score: 0.8401