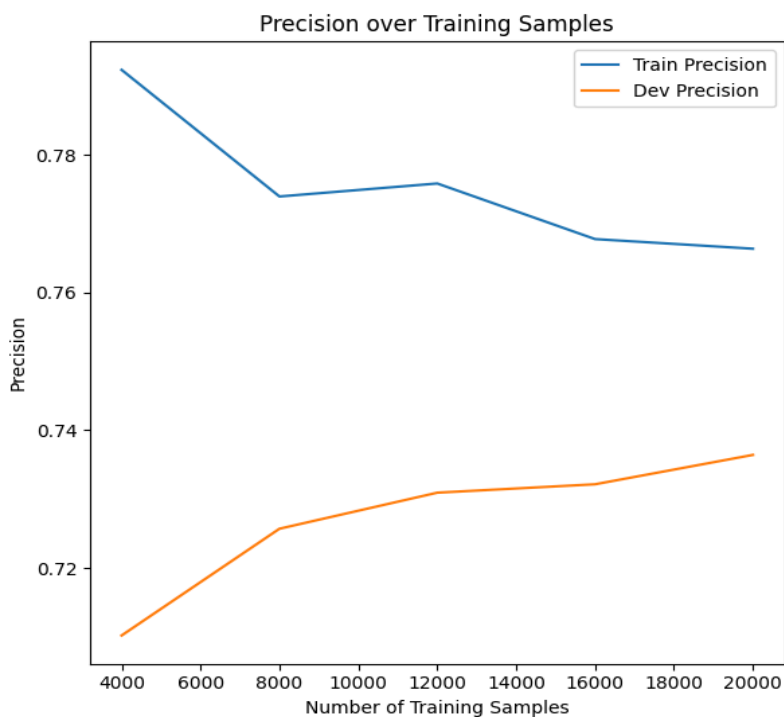
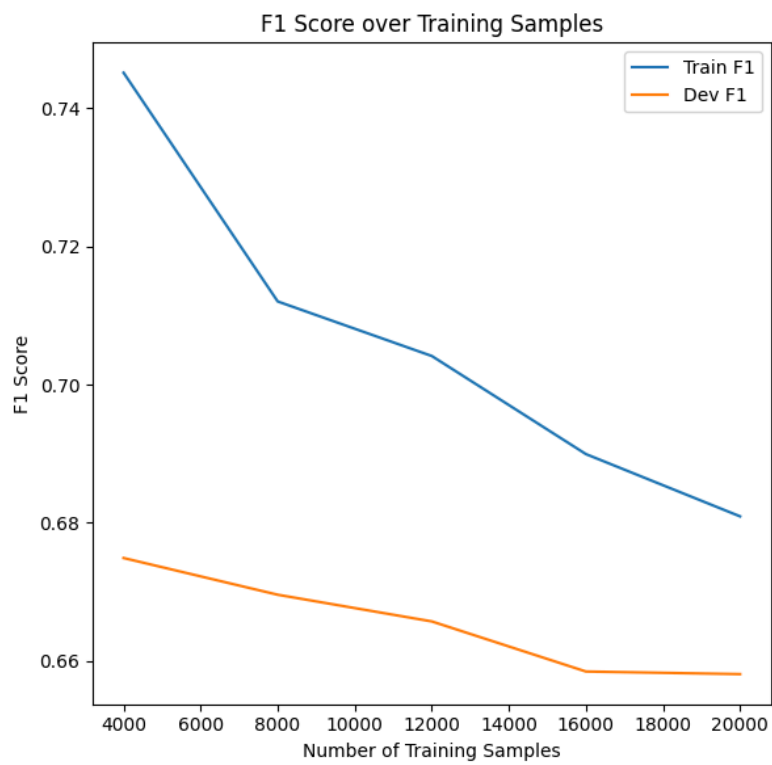
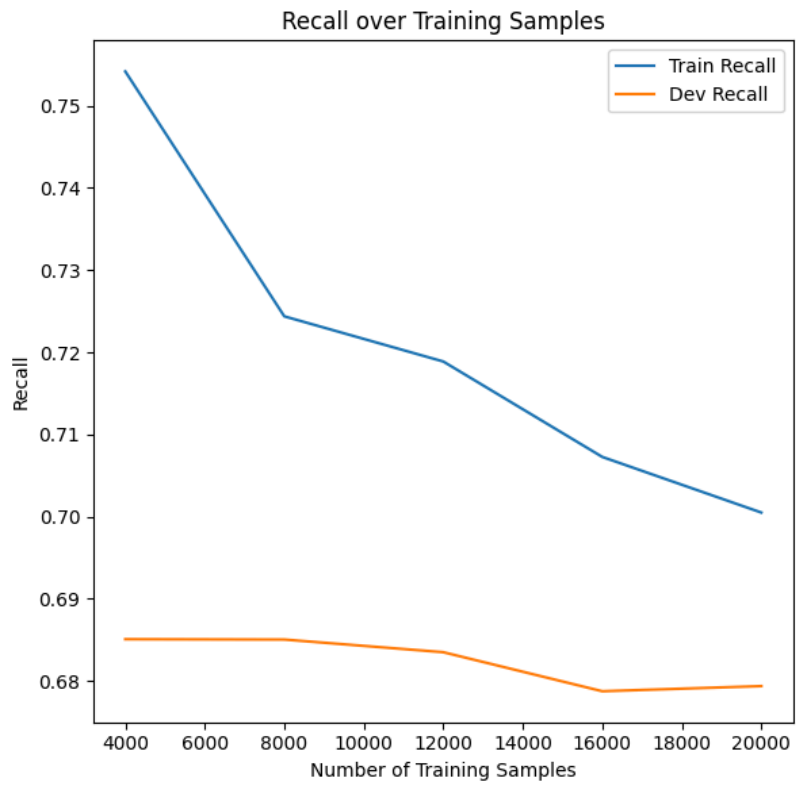


Τυχαίο δάσος: Στον αλγόριθμο αυτό καλούμαστε να υλοποιήσουμε τον αλγόριθμο του τυχαίου δάσος με την χρήση του ID3. Για την προεπεξεργασία των δεδομένων χρησιμοποιήθηκαν οι παράμετροι $n = 250$ παράλειψη πιο σπάνιων λέξεων, $k = 750$ παράλειψη πιο συχνών λέξεων και $m = 7500$ λέξεις με το μεγαλύτερο information gain που κρατούνται στο τελικό λεξιλόγιο. Τα χαρακτηριστικά του ID3 είναι $\text{max_features} = 1000$, $\text{min_samples_split} = 5$ και $\text{max_depth} = 11$. Τέλος οι παράμετροι του random forest που χρησιμοποιήθηκαν είναι $\text{num_of_trees} = 5$ και $\text{features_split_depth}$ ίδια με τον ID3. Το Random Forest αρχικά περιλαμβάνει πολλά ID3 δέντρα. Τα δέντρα χωρίζονται με τρόπο ώστε να μεγιστοποιείται το information gain μέχρις ότου να φτάσουν το μέγιστο βάθος. Στο τέλος γίνονται οι προβλέψεις των κατηγοριών με τη χρήση της μεθόδου του majority voting. Παρακάτω φαίνεται η ακρίβεια και τα υπόλοιπα metrics των αποτελεσμάτων που εμφάνισε η εκτέλεση του αλγορίθμου μας. Οι υπερπαραμέτροι επιλέχθηκαν όλες μέσω πειραμάτων με διαφορετες τιμές. Είναι εμφανές ότι υπήρξε περισσότερος χώρος για βελτίωση αποτελεσμάτων και περεταίρω tuning, αλλά προβλήματα όπως ο χρόνος εκτέλεσης από την υπολογιστική ισχύ που κατέχουμε δεδομένης της προθεσμίας καθώς και memory related προβλήματα όπως μεγάλα heap sizes που διέκοπταν συχνά το τρέξιμο του προγράμματος δεν μας διευκόλυναν την εξαγωγή καλύτερων αποτελεσμάτων.





Precision (Class 0, Class 1): [0.87349206 0.62582888]

Recall (Class 0, Class 1): [0.44024 0.93624]

F1 Score (Class 0, Class 1): [0.58542553 0.75019231]

Micro-Averaged Precision: 0.68824

Micro-Averaged Recall: 0.68824

Micro-Averaged F1: 0.68824

Macro-Averaged Precision: 0.7496604702487055

Macro-Averaged Recall: 0.68824

Macro-Averaged F1: 0.6678089198036007

Classification Report:

	precision	recall	f1-score	support
Class 0	0.87	0.44	0.59	12500
Class 1	0.63	0.94	0.75	12500
accuracy			0.69	25000
macro avg	0.75	0.69	0.67	25000
weighted avg	0.75	0.69	0.67	25000