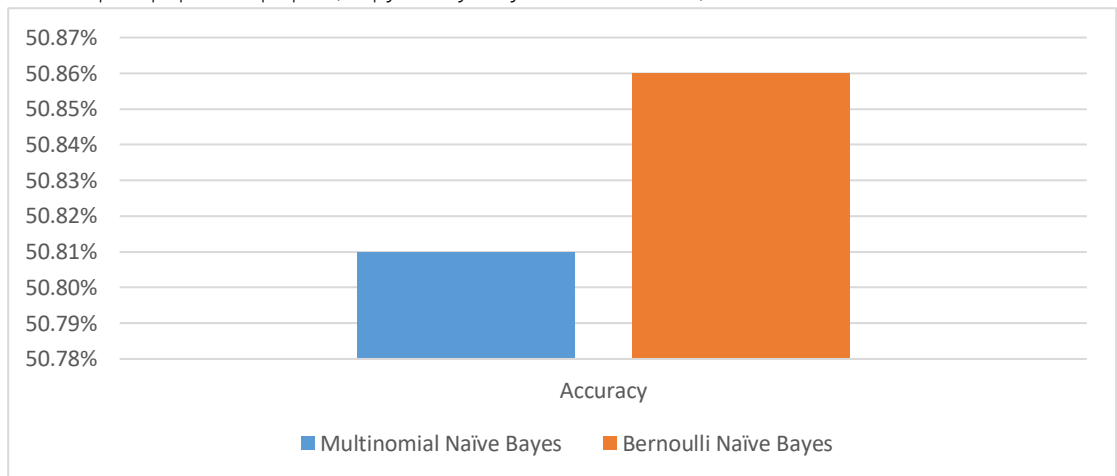
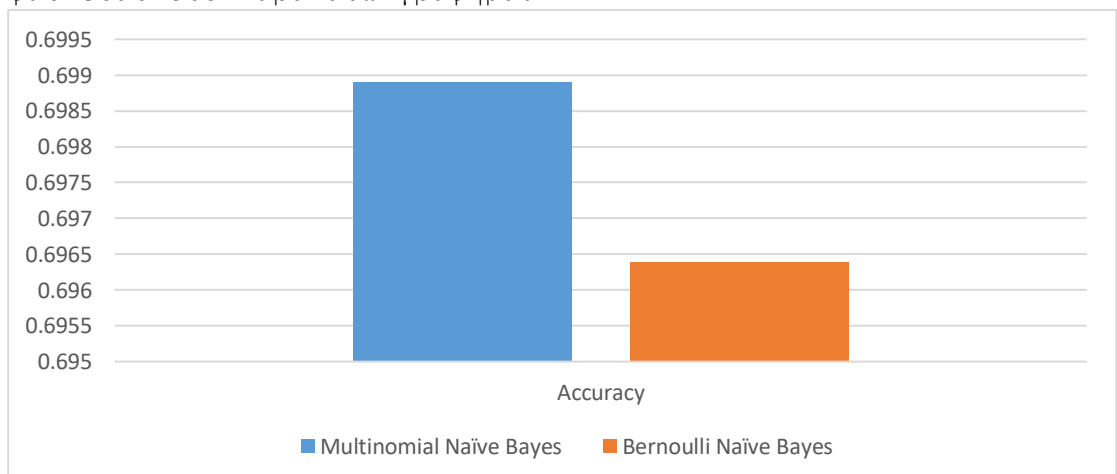


## Στέργιος Τσάνταλης (iis21125)

- a) Η προεπεξεργασία που έχω κάνει είναι όπου υπάρχει κείμενο όλο το κείμενο να είναι γραμμένο με πεζά γράμματα, δηλαδή όπου υπάρχει κεφαλαίο γράμμα να γίνεται πεζό. Ακόμη, σύμφωνα με την στήλη text και το κείμενο σε κάθε σειρά (για κάθε χρήστη) δημιουργήθηκε η στήλη tag που αντιπροσωπεύει το review για κάθε ταινία (pos για θετικό review και neg για αρνητικό review). Δημιουργώντας και εκτελώντας το process με το multinomial και Bernoulli naïve bayes classifier στο RapidMiner παρατηρούμε ότι καλύτερο accuracy έχει ο αλγόριθμος Bernoulli αλλά με πολύ μικρή διαφορά (της τάξεως του 0.04%).



- b) Ακολουθήθηκε ακριβώς η ίδια διαδικασία προεπεξεργασίας με του ερωτήματος a). Εκτελώντας τον κώδικα παρατηρούμε ότι το mean accuracy του αλγόριθμου Multinomial είναι καλύτερο με μικρή διαφορά σε σχέση με του Bernoulli όπως φαίνεται στο παρακάτω γράφημα.



Παρατηρούμε ότι με τη χρήση κώδικα έχουμε καλύτερο accuracy και γενικά καλύτερες μετρικές και για τους δύο αλγόριθμους σε σχέση με τη χρήση του λογισμικού

RapidMiner. Τέλος, με την χρήση κώδικα βλέπουμε ότι ο αλγόριθμος Multinomial είναι καλύτερος του Bernoulli, το ακριβώς αντίθετο ισχύει με τη χρήση του λογισμικού RapidMiner