

# 3η Εργασία Software Algorithms

ΣΑΜΙΟΣ ΓΡΗΓΟΡΗΣ 1115201500141

January 15, 2019

## 1. Αρχεία -Κλάσεις:

- **CryptoCurrencyRecommendation:** Κλάση για την διαχείριση User,Tweet κλάσεων και μέθοδοι για το recommendation
- **User:**Κλάση που αντιπροσωπεύει ένα χρήστη.Ο χρήστης διαθέτει ένα σύνολο από tweets που έχει ο ίδιος κάνει, το διάνυσμα για τα σκορ του στα κρυπτονομίσματα (cc),τα γνωστά και τα άγνωστα cc και το τελικό του αποτέλεσμα.
- **Tweet:**Κλάση που αντιπροσωπεύει ένα tweet.Διαθέτει το σκορ του,το id του ,την διανυσματική του αναπαράσταση και τις αναφορές στα cc.
- **ClusterManagement:**Βιβλιοθήκη και κλάσεις για την διαχείριση clustering.Χρήση Αλγορίθμου Kmeans PlusPlus-Assign-KmenasUpdate.
- **HashFunctions:**Κλάσεις για τις hashing συναρτήσεις.Cosine/Euclidean
- **InputOutput:**Κλάση για την διαχείριση παραμέτρων προγράμματος ,εισόδου και εξόδου.
- **HashManagementLSH:**Κλάση που διαχειρίζεται σύνολο από HashTable και βρίσκει κοντινούς γείτονες σημείων.
- **Distance:**Κλάση για την αποθήκευση ήδη υπολογισμένων αποστάσεων σημείων.

2. Εκτέλεση: recommendation -i "tweets" -v "tweets vectors" -b "coinsqueries.csv" -d "vaderlexicon.csv" -o outputfile -validation  
Μεταγλώττιση:make recommendation.
3. Συμβάσεις:Θεώρησα ότι όταν ένας χηστης έχει μηδενικό διάνυσμα σκορ στην περίπτωση του "cosine similarity" δεν ορίζεται το z τότε για αυτόν προτείνονται στην τύχη κρυπτονομίσματα.
4. Σχόλια:Παρατήρησα ότι το μεγαλύτερο MAE εμφανιζόταν στο CosineLSH B με μέσο όρο MAE 0.70 και το μικρότερο στο Clustering B με μο MAE 0.17.Γενικά το Clustering έπαιρνε περισσότερο χρόνο συγκριτικά με το CosineLSH,αλλά έβγαζε χαμηλότερο MAE.