



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

Προγραμματιστικά Εργαλεία και Τεχνολογίες για Επιστήμη Δεδομένων

## Time Travel Project

Ορφανουδάκης Φίλιππος Σκόβελεφ

phil.orfa@gmail.com

AM: 03400107

M.Sc. , ΕΔΕΜΜ ΔΠΜΣ

Αθήνα, Φεβρουάριος 2021

# 1 Εισαγωγή

Στο πλαίσιο της αναφοράς θα παρουσιάσουμε με συνοπτικό τρόπο την στρατηγική που ακολουθήσαμε για την παρασκευή της μικρής και μεγάλης ακολουθίας, παραθέτοντας και το αντίστοιχο διάγραμμα αποτίμησης. Θα αναφέρουμε σύντομα τα στάδια προεπεξεργασίας και επιλογής των δεδομένων μας και τέλος θα γίνει μια μικρή αναφορά στον τρόπο σχεδίασης των διαγραμμάτων.

Συνεπώς η δομή της αναφοράς θα ακολουθήσει την εξής ροή:

- Ανάλυση Προεπεξεργασίας
- Μικρή Ακολουθία
- Μεγάλη Ακολουθία
- Σχεδίαση Διαγραμμάτων

Δεν θα προσπαθήσουμε να εμβαθύνουμε στην ανάλυση του κώδικα αλλά να δώσουμε μια διαισθητική εικόνα της κάθε στρατηγικής που ακολουθήσαμε.

## 2 Ανάλυση Προεπεξεργασίας

Πρωταρχικός μας στόχος ήταν να επιλέξουμε ένα υποσύνολο μετοχών για να χρησιμοποιήσουμε στην παρασκευή των 2 ακολουθιών και φυσικά να κάνουμε τις κατάλληλες τροποποιήσεις στα δεδομένα, έτσι ώστε να ικανοποιήσουμε τους περιορισμούς της εκφώνησης. Βασικές τροποποιήσεις που πραγματοποιήσαμε :

- Θέτουμε το  $\text{Volume} = 0.1 * \text{Volume}$ , καθώς σε κάθε συναλλαγή δεν μπορούμε να υπερέβουμε το 10%.
- Αφαιρούμε όποια γραμμή έχει μηδενικές ή αρνητικές τιμές σε ένα εκ των Open, Low, High, Close, καθώς θεωρούμε ότι είναι θόρυβος.
- Αφαιρούμε την στήλη "Open Int", καθώς δεν προσφέρει πληροφορία.
- Ταξινομούμε τα δεδομένα με βάση την ημερομηνία σε αύξουσα σειρά.
- Ορίζουμε την στήλη OP\_HI(Buy Open → Sell High) και LO\_CL(Buy Low → Sell Close) και θέτουμε σαν στήλη profit το μέγιστο από τα 2.

Στο σημείο αυτό, να αναφέρουμε πως η επιλογή της προσθήκης των δύο στηλών OP\_HI και LO\_CL θα μας βοηθήσει ύστερα με τα intra day transactions. Στη συνέχεια, πραγματοποιήσαμε έναν καθαρισμό των δεδομένων μας και με ορισμένα βήματα οδηγηθήκαμε στην τελική επιλογή των πακέτων μετοχών. Τα βήματα αυτά είναι τα εξής :

- Αφαιρούμε τα outliers ( $\text{mean} \pm 3 * \text{std}$ ) για τις στήλες High και Low, το σκεπτικό πίσω από αυτή την κίνηση είναι ότι όταν προχωρήσουμε στον υπολογισμό μέσω των όρων δεν θέλουμε μια τιμή να επηρεάσει την συνολική εικόνα.
- Υπολογισμός μέσου όρου για την στήλη High και Low.
- Προσθήκη νέας στήλης  $\text{diff} = \text{mean}(\text{High}) - \text{mean}(\text{Low})$  και ταξινόμηση με βάση αυτή τη στήλη σε φθίνουσα σειρά.
- Για την ακολουθία small επιλέγουμε τις πρώτες 100 μετοχές με βάση την παραπάνω ταξινόμηση ενώ για την ακολουθία large τις πρώτες 500.
- Τέλος, προσθέτουμε (αν δεν υπάρχουν ήδη) τις 30 μετοχές που αναφέρονται σε αυτό το [link](#), δηλαδή τις 30 μεγαλύτερες εταιρίες στο χρηματιστήριο.

Ολοκληρώσαμε την προεπεξεργασία και συνεχίζουμε στην ανάλυση της στρατηγικής για κάθε μια ακολουθία.

### 3 Μικρή Ακολουθία

Η μικρή ακολουθία που κατασκευάσαμε έχει κέρδος **21677640348.653584 \$** μέσω **1000** κινήσεων.

Πρώτη μας κίνηση είναι να στρογγυλοποιήσουμε προς τα κάτω το Volume και να ορίσουμε τις βασικές μας παραμέτρους. Θέτουμε την αρχική και τελική ημερομηνία που προέκυψε με βάση τις μετοχές που επιλέξαμε καθώς και το  $\text{balance} = 1$ .

Η γενική ιδέα της στρατηγικής είναι ότι θέτουμε ένα χρονικό παράθυρο, μέσα στο οποίο θα πραγματοποιήσουμε την καλύτερη δυνατή συναλλαγή. Συνεπώς πρέπει να ορίσουμε το μέγεθος του παραθύρου το οποίο προκύπτει 35. Το νούμερο αυτό προέκυψε μέσα από αρκετές δοκιμές, αλλά η τάξη μεγέθους του είχε καθοριστεί από το πόσες διαθέσιμες μέρες έχουμε δια τον συνολικό αριθμό συναλλαγών\*2 (καθώς έχουμε 1 buy και 1 sell σε κάθε παράθυρο).

Έχοντας καθορίσει όλες τις απαραίτητες μεταβλητές θα εξηγήσουμε συνοπτικά τον αλγόριθμο κατασκευής της μικρής ακολουθίας.

- Ξεκινάμε με το αρχικό παράθυρο 35 ημερών και όσο δεν έχουμε ξεπεράσει την τελική ημερομηνία συνεχίζουμε
- Για αυτό το παράθυρο πραγματοποιούμε έναν έλεγχο για κάθε μετοχή ξεχωριστά ώστε να βρούμε το βέλτιστο κέρδος. Δεν πραγματοποιούμε κανένα έλεγχο για πιθανές intra day συναλλαγές και ψάχνουμε μόνο συναλλαγές της μορφής buy low -ι, sell high.
- Για κάθε μετοχή και για κάθε Low ελέγχουμε όλα τα High των επόμενων ημερών, και επιστρέφουμε το μέγιστο γινόμενο της διαφοράς (High-Low) επί το μέγιστο Volume που μπορούμε να αγοράσουμε.
- Ταξινομούμε τις παραπάνω μετοχές και επιστρέφουμε την μέγιστη.
- Την καταγράφουμε και πηγαίνουμε στο επόμενο χρονικό παράθυρο.
- Αφού ολοκληρωθεί ο παραπάνω αλγόριθμος έχουμε σε ένα αρχείο small.txt μια σειρά απο transactions. Το αρχείο αυτό το χρησιμοποιούμε πρώτα για την κατασκευή του διαγράμματος αποτίμησης και ύστερα το τροποποιούμε manually για να προσθέσουμε τον αριθμό των κινήσεων στην αρχή του.

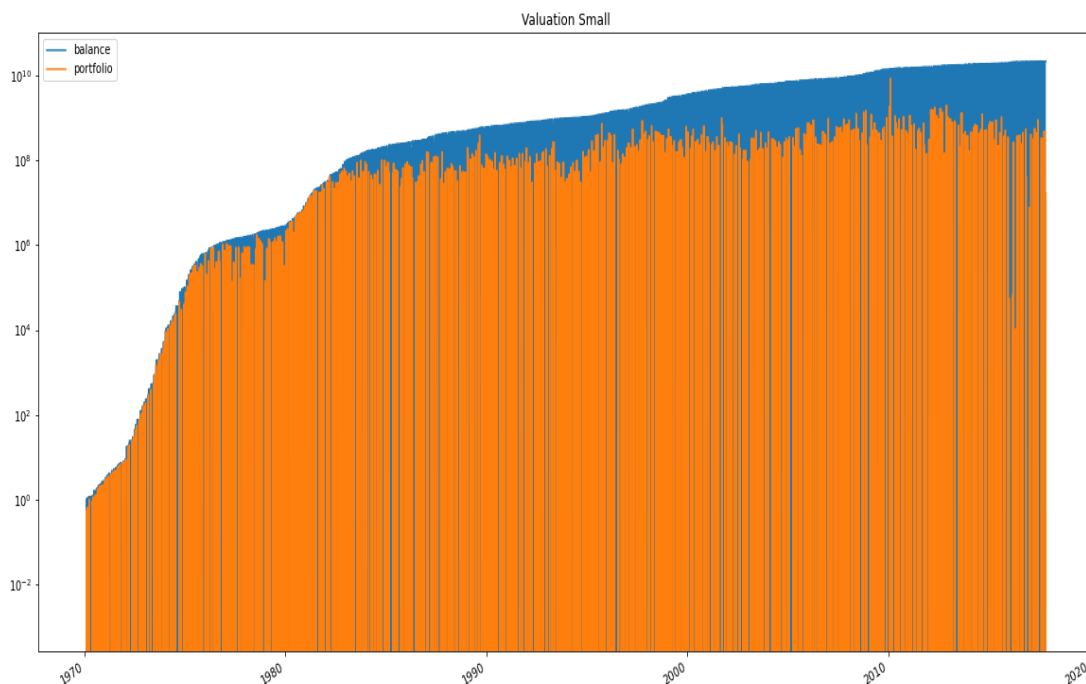


Figure 1: Διάγραμμα Αποτίμησης Small

## 4 Μεγάλη Ακολουθία

Η μεγάλη ακολουθία που κατασκευάσαμε έχει κέρδος **121776455955.61334 \$** μέσω **888442** κινήσεων.

Πρώτη μας κίνηση και εδώ είναι να στρογγυλοποιήσουμε προς τα κάτω το Volume, επί προσθέτως αφαιρούμε όποια γραμμή έχει profit (το είχαμε ορίσει στην προεπεξεργασία) μηδενικό ή αρνητικό. Στη συνέχεια ορίζουμε μια νέα στήλη vol\_prof που ισοδυναμεί με το  $\text{Volume} * \text{profit}$  (θα προτιμήσουμε να ασχοληθούμε με μια μετοχή που έχει μικρότερο profit αλλά πολλά μερίδια, συγκριτικά με μεγάλο profit αλλά ελάχιστα μερίδια). Στη συνέχεια ορίζουμε τις βασικές μας παραμέτρους. Θέτουμε την αρχική και τελική ημερομηνία που προέκυψε με βάση τις μετοχές που επιλέξαμε καθώς και το balance = 1.

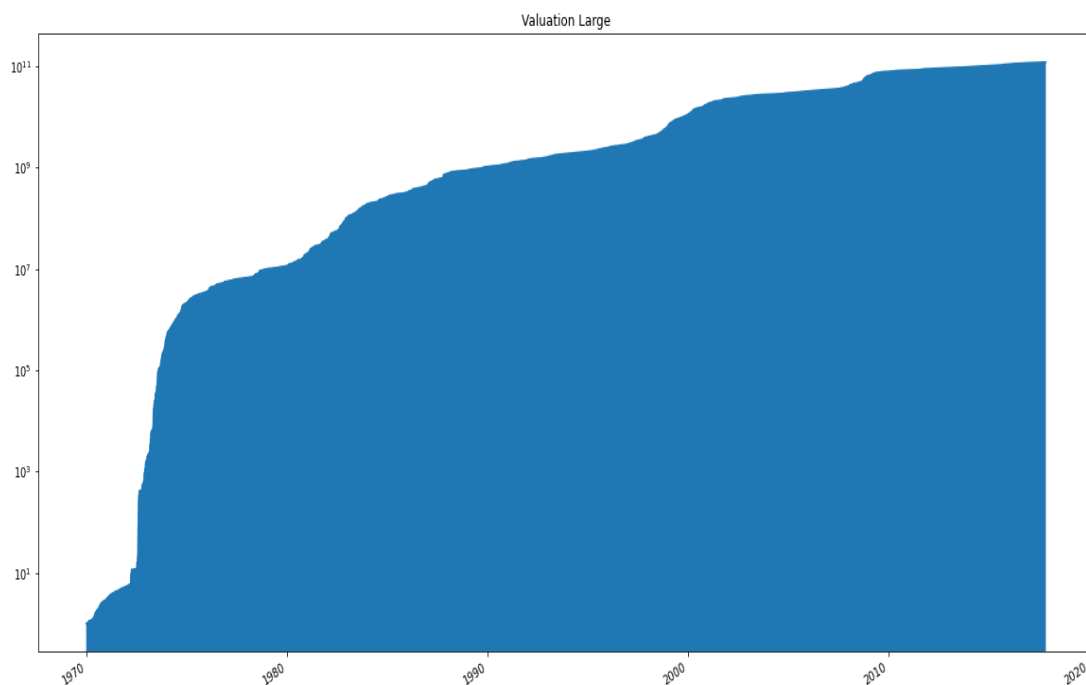
Η γενική ιδέα της στρατηγικής είναι ότι πραγματοποιούμε μόνο intra day transactions και με έναν τυχαίο αλλά ελεγχόμενο τρόπο προσπαθώ να επιλέξω τις βέλτιστες από όλες τις μετοχές. Συνεπώς σε αυτή την περίπτωση το sliding window = 1.

Μια καθοριστική παράμετρος για το πρόβλημα είναι πόσες συναλλαγές θα επιτρέψουμε ανά ημέρα. Εμπειρικά και μετά από δοκιμές καταλήξαμε στις 130. Η τάξη μεγέθους μπορεί να υπολογιστεί από το ότι έχω  $(1M \text{ transactions}/2)/(\text{διαθέσιμες μέρες})$ . Φυσικά αυτός ο αριθμός αρχικά προκύπτει περίπου 40 αλλά ειδικά στην δεκαετία του '70 και του '80 δεν έχουμε πολλές επιλογές με αποτέλεσμα να οδηγούμαστε σε πολύ λιγότερες.

Έχοντας καθορίσει όλες τις απαραίτητες μεταβλητές θα εξηγήσουμε συνοπτικά τον αλγόριθμο κατασκευής της μεγάλης ακολουθίας.

- Ξεκινάμε με την αρχική ημερομηνία και όσο δεν έχουμε ξεπεράσει την τελική, συνεχίζουμε
- Για την συγκεκριμένη ημερομηνία ταξινομούμε τις μετοχές σε φθίνουσα σειρά με βάση το vol\_prof και επιλέγουμε να ελέγξουμε τις 130 πρώτες για intra day transactions.
- Εισάγουμε μια μεταβλητή pseudobalance για να ελέγχουμε αν είναι υλοποιήσιμες όλες οι αγορές που ανήκουν στην ίδια ομάδα σύμφωνα με την εκφώνηση. Αν είναι υλοποιήσιμη την πραγματοποιούμε και αφαιρούμε το αντίστοιχο ποσό από το pseudobalance
- Στη συνέχεια ανάλογα με το ποιά από τις 2 intra day transactions έχει πραγματοποιηθεί την καταγράφουμε και συνεχίζουμε στην επόμενη ημέρα
- Αφού ολοκληρωθεί ο παραπάνω αλγόριθμος έχουμε σε ένα αρχείο large.txt μια σειρά απο transactions. Το αρχείο αυτό το χρησιμοποιούμε πρώτα για την κατασκευή του διαγράμματος αποτίμησης και ύστερα το τροποποιούμε manually για να προσθέσουμε τον αριθμό των κινήσεων στην αρχή του.

Όπως φαίνεται από το διάγραμμα δεν έχουμε μη μηδενικές τιμές για το portfolio καθώς κάθε μέρα πουλάμε ότι έχουμε αγοράσει.



**Figure 2:** Διάγραμμα Αποτίμησης Large

Πιθανές βελτιώσεις θα μπορούσαν να πραγματοποιηθούν αν εισάγουμε ολόκληρο το dataset με όλες τις μετοχές ή ακόμα και αν θέταμε μια σταδιακή αύξηση των επιτρεπόμενων transactions ανά τις δεκαετίες (πχ 20 ημερήσιες για τις δεκαετίες '70-'80, 70 για τη δεκαετία '90, 130 για τη δεκαετία '00 και 170 για τη δεκαετία '10). Παρόλα αυτά επιτυγχάνουμε ένα ικανοποιητικό ποσό και ένα αρκετά υψηλό πλήθος κινήσεων. Φυσικά βλέπουμε ότι δεν είναι γραμμική η εξάρτηση μεταξύ κινήσεις - κέρδους συγκριτικά με την Small ακολουθία καθώς στη περίπτωση της μεγάλης ακολουθίας δεν γνωρίζουμε αν παίρνουμε την βέλτιστη απόφαση.

## 5 Σχεδίαση Διαγραμμάτων

Για την σχεδίαση των διαγραμμάτων πραγματοποιήθηκε μια προσομοίωση που βασίστηκε στα .txt αρχεία με τις ακολουθίες.

Επομένως δεν κατασκευάζονται οι αντίστοιχες τιμές των μεταβλητών κατά τη διάρκεια του υπολογισμού των transactions καθώς θεωρήσαμε ότι θα επιβαρύνει τον χρόνο εκτέλεσης που ήδη κυμαίνεται στις 2 ώρες.