

ΠΡΟΓΡΑΜΜΑΤΙΣΤΙΚΗ ΕΡΓΑΣΙΑ
ΗΜΕΡΟΜΗΝΙΑ ΠΑΡΑΔΟΣΗΣ: 31/05/2025

Περιγραφή

Καλείστε να αναπτύξετε ένα σύστημα ανάλυσης και πρόβλεψης τιμών μετοχών βασισμένο σε ιστορικά δεδομένα από το **Yahoo Finance API**. Η εργασία εστιάζει στην επεξεργασία των δεδομένων, στην αριθμητική ανάλυση με Πολυωνυμική Προσαρμογή Ελαχίστων Τετραγώνων, στην αριθμητική ολοκλήρωση για εκτίμηση μέσης τιμής, στην ανίχνευση ανωμαλιών με παραμετρικό δυναμικό κατώφλι και στην πρόβλεψη τιμών για την επόμενη ημέρα.

Είστε ελεύθεροι να επιλέξετε τη μετοχή της αρεσκείας σας.

1. Συλλογή Δεδομένων

- Χρησιμοποιώντας το **Yahoo Finance API**, κατεβάστε δεδομένα για τη μετοχή της επιλογής σας για 5 συνεχόμενες ημέρες, ανά 30 λεπτά. Για την εξαγωγή των ιστορικών δεδομένων, προτείνεται η χρήση της βιβλιοθήκης `yfinance` της Python, η οποία παρέχει εύκολη πρόσβαση στα δεδομένα μετοχών από το Yahoo Finance.
- Αποθηκεύστε τα δεδομένα σε αρχείο CSV και εμφανίστε τα πρώτα 5 αποτελέσματα.

2. Πολυωνυμική Προσαρμογή Ελαχίστων Τετραγώνων

- Εκτελέστε Πολυωνυμική Προσαρμογή Ελαχίστων Τετραγώνων με πολώνυμο βαθμού 1 (γραμμικό μοντέλο), βαθμού 2 (παραβολικό μοντέλο) και βαθμού 3 (κυβικό μοντέλο).
- Προσαρμόστε τα δεδομένα χρησιμοποιώντας τα πρώτα 80% ως “training” και τα τελευταία 20% ως “testing”.
- Υπολογίστε τα σφάλματα **Μέσο Απόλυτο Σφάλμα (MAE)** και **Μέσο Τετραγωνικό Σφάλμα (MSE)** για κάθε μοντέλο. Εμφανίστε σε πίνακα ποιο μοντέλο έχει την καλύτερη απόδοση.
- Εξηγήστε γιατί ένα πολώνυμο υψηλότερου βαθμού μπορεί να παρουσιάζει υπερπροσαρμογή (overfitting) στις προβλέψεις. Πώς μπορούμε να το εντοπίσουμε αριθμητικά; Ποια μέτρα μπορούν να περιορίσουν αυτό το φαινόμενο;

3. Πρόβλεψη Επόμενης Ημέρας

- Με βάση το καλύτερο μοντέλο (μικρότερο MAE), εκτελέστε πρόβλεψη τιμών για τα επόμενα **48 χρονικά σημεία** (30 λεπτά το καθένα, δηλαδή 24 ώρες). Παρουσιάστε τα αποτελέσματα σε γράφημα μαζί με την πολυωνυμική καμπύλη προσαρμογής.
- Πώς η επιλογή διαφορετικών βαθμών πολυωνύμων μπορεί να επηρεάσει την ποιότητα πρόβλεψης; Πώς μπορούμε να εντοπίσουμε αν η πρόβλεψη είναι ασταθής;

4. Ολοκλήρωση της Καμπύλης Τιμών για Μέση Τιμή

- Χρησιμοποιήστε τις αριθμητικές μεθόδους του σύνθετου **Κανόνα του Τραπεζίου** και του σύνθετου **Κανόνα του Simpson** για να υπολογίσετε το εμβαδόν κάτω από την καμπύλη των τιμών.
- Το εμβαδόν αυτό, διαιρεμένο με το διάστημα των 5 ημερών, θα δώσει την **εκτιμώμενη μέση τιμή της μετοχής** στο διάστημα αυτό.
- Υπολογίστε τη μέση τιμή για κάθε ένα από τα μοντέλα (1ου, 2ου, 3ου βαθμού) και συγκρίνετε τα αποτελέσματα σε πίνακα.
- Ποια είναι η επίδραση της επιλογής του βήματος (step size) στην ακρίβεια της αριθμητικής ολοκλήρωσης; Μπορείτε να εξηγήσετε γιατί μικρότερα βήματα βελτιώνουν την ακρίβεια;

5. Ανίχνευση Ανωμαλιών στα Ιστορικά Δεδομένα

- Χωρίστε τα ιστορικά δεδομένα σε διαστήματα των **5 χρονικών σημείων** (δηλαδή 2,5 ώρες).
- Για κάθε διάστημα, εφαρμόστε:
 - τον **Αλγόριθμο Aitken-Neville** και
 - τη **Μέθοδο Διηρημένων Διαφορών Newton**
- Προβλέψτε την **επόμενη τιμή** της μετοχής μετά το διάστημα των 5 σημείων.
- Συγκρίνετε την πρόβλεψη με την πραγματική τιμή από τα δεδομένα.
- Πώς επηρεάζει η μεταβλητότητα των τιμών τον καθορισμό του κατωφλίου ανωμαλίας; Μπορείτε να αιτιολογήσετε γιατί ένα πιο ευμετάβλητο σύνολο δεδομένων απαιτεί διαφορετική προσέγγιση στον ορισμό του κατωφλίου, σε σύγκριση με ένα πιο σταθερό σύνολο δεδομένων;

Δυναμικός Υπολογισμός Κατωφλίου Ανωμαλίας (δ):

$$\delta = 0,05 \cdot Range + 0,5 \cdot \sigma$$

όπου:

- **Range (Εύρος Τιμών):** Η διαφορά μεταξύ της μέγιστης και της ελάχιστης τιμής της μετοχής στο επιλεγμένο χρονικό διάστημα.
- **σ (Τυπική Απόκλιση):** Μετρική διασποράς των τιμών γύρω από τη μέση τιμή.

Παραδοτέο στην ιστοσελίδα του μαθήματος στο elearning (Εργασία 2025):

Αρχείο .ipynb που θα περιέχει τον κώδικα και τις απάντησεις στα ανωτέρω ερωτήματα. Εναλλακτικά, μπορείτε να παραδώσετε ένα αρχείο με τον κώδικά σας και ένα αρχείο σε μορφή .pdf με τις απαντήσεις σας.