

Πρόβλεψη Χρηματιστηριακών Τιμών

Ομάδα 3 : Καραγεώργης Στέργιος ΑΜ:2346 Πανουργιάς Χρήστος ΑΜ:2405

Ιούνιος 2022

Περιεχόμενα

- Ι. Εισαγωγή
- ΙΙ. Αλγόριθμος Λογισμικό
- ΙΙΙ. Μεθοδολογία Τεχνικές
- ΙV. Αποτελέσματα
- V. Συμπεράσματα
- VI. Βιβλιογραφία

Ι. Εισαγωγή

Παρουσίαση προβλήματος :

- ο Το χρηματιστήριο είναι ένας από τους σημαντικότερους τομείς επενδύσεων,
 - επομένως η πρόβλεψη της τάσης των τιμών του χρηματιστηρίου αποτελεί ένα θέμα υψίστης σημασίας για τους ερευνητές του οικονομικού τομέα.

Στόχος

Στόχος της παρούσας έρευνας, είναι η δημιουργία ενός μοντέλου πρόβλεψης για την βραχυπρόθεσμη πρόβλεψη της τάσης των τιμών

Ι. ΕισαγωγήΛόγοι επιλογής του θέματος

1

Η πρόβλεψη τιμών των μετοχών με τη βοήθεια αλγορίθμων μηχανικής μάθησης, βοηθά στον εντοπισμό της μελλοντικής αξίας των μετοχών της εταιρίας καθώς και των υπόλοιπων χρηματοοικονομικών περιουσιακών στοιχείων που συναλλάσσονται σε ένα χρηματιστήριο.

2

Ο κύριος σκοπός της πρόβλεψης των τιμών αφορά τη συνεχή αύξηση των κερδών.

3

Η πρόβλεψη της απόδοσης του χρηματιστηρίου παραμένει ένα δύσκολο έργο, καθώς η πρόβλεψη δύναται να επηρεαστεί από μια σειρά παραγόντων όπως σωματικοί, ψυχολογικοί, ορθολογική ή παρορμητική συμπεριφορά. Όλοι αυτοί οι παράγοντες συνδυάζονται για να κάνουν τις τιμές των μετοχών δυναμικές και ασταθείς, καθιστώντας έτσι δύσκολη την ακριβή πρόβλεψη τους.

ΙΙ. Αλγόριθμοι-ΛογισμικόΣύνολα Δεδομένων

Τα δεδομένα θα εξαχθούν από την παρακάτω πηγή:

https://www.kaggle.com/datasets/borismarjanovic/price-volume-data-for-all-usstocks-etfs

Η επίσημη γλώσσα υπολογιστών που θα χρησιμοποιηθεί είναι η Python 3

Θα αξιοποιηθούν βιβλιοθήκες για την καλύτερη κατανόηση του προβλήματος, όπως: το keras για το νευρωνικό δίκτυο, το numpy και scikit-learn για τις μεθόδους ταξινόμησης και παλινδρόμησης.

Θα χρησιμοποιηθεί η βιβλιοθήκη matplotlib για τη σύγκριση των αποτελεσμάτων και των σχεδιασμό τους σε γραφήματα.

ΙΙΙ. Μεθοδολογία - Τεχνικές

- ο Η Επιστήμη των Δεδομένων βασίζεται σε μεγάλο βαθμό στην πρόβλεψη μελλοντικών αποτελεσμάτων, όπως επίσης και στη μοντελοποίηση των αποτελεσμάτων
- ο Στο χρηματιστήριο συγκεκριμένα χρησιμοποιούνται μοντέλα χρονοσειρών για την πρόβλεψη των τιμών των μετοχών όπως επίσης και άλλοι αλγόριθμοι όπως νευρωνικά δίκτυα και στατιστικά μοντέλα.
- ο Στο παρών έργο θα υλοποιηθούν οι εξής αλγόριθμοι:
 - Ι. Ένα μοντέλο παλινδρόμησης
 - ΙΙ. Το στοχαστικό μοντέλο ΑΡΙΜΑ
 - ΙΙΙ. Ένα Νευρωνικό Δίκτυο LSTM

Περιγραφή Αλγορίθμων Μοντέλο ARIMA

- ARIMA (Autoregressive Integrated Moving Average)
 ονομάζεται ένα υπόδειγμα ARMA(p,q) που
 εφαρμόζεται σε μια ολοκληρωμένη σειρά d τάξης.
 - **p** που συμβολίζει τους όρους του αυτοπαλίνδρομου υποδείγματος
 - **q** που συμβολίζει τους όρους του υποδείγματος των κινητών μέσων
- Η τάξη του ARIMA (p,d,q) προσδιορίζεται με τη μεθοδολογία των Box – Jenkins ευρέως γνωστή στη στατιστική.

- Τα ολοκληρωμένα αυτοπαλινδρομικά μοντέλα κινητών μέσων όρων ARIMA είναι στοχαστικά μαθηματικά μοντέλα, τα οποία μας βοηθάνε να αναλύσουμε και να προβλέψουμε την εξέλιξη μεγεθών.
- Σε αντίθεση με τα ντετερμινιστικά μοντέλα (βλ. γραμμική παλινδρόμηση), η χρήση αυτού του μοντέλου βασίζεται στον υπολογισμό της πιθανότητας για την οποία η τιμή του μεγέθους βρίσκεται εντός κάποιου διαστήματος.

Περιγραφή Αλγορίθμων - Μοντέλο ΑΡΙΜΑ

- ο Στην γενική τους μορφή τα μοντέλα ARIMA περιέχουν τον τυχαίο παράγοντα (σφάλμα πρόβλεψης), τιμές του μεγέθους που εμφανίστηκαν σε προηγούμενες περιόδους και σχετικούς στοχαστικούς παράγοντες.
- ο Πιο συγκεκριμένα τα μοντέλα ΑRIMA γράφονται ως γραμμικός συνδυασμός των παραπάνω παραγόντων.

Μορφή Μοντέλου ΑΡΙΜΑ – Εξίσωση Μοντέλου

$$y_t' = c + \varphi_1 y_{t-1}' + \ldots + \varphi_p y_{t-p}' + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$
 intercept
$$\log_{(AR)}$$
 errors
$$\pmod{(MA)}$$

Εφαρμογή του Μοντέλου ARIMA στη μετοχή της TESLA

- ο Η εφαρμογή μοντέλων ARIMA προϋποθέτει την εξαγωγή βραχυπρόθεσμων προβλέψεων.
- ο Όπως αναφέρθηκε, τα συγκεκριμένα μοντέλα είναι γραμμικός συνδυασμός των παρελθοντικών τιμών της χρονοσειράς.
- ο Για αυτό το λόγο αν θέλουμε να προβλέψουμε την τιμή yt+1, απαιτείται αντίστοιχα γνώση των τιμών yt ,yt-1...yt-n-1. Οι οποίες θα γίνουν από το μοντέλο.
- ο Αντιλαμβανόμαστε λοιπόν ότι θα είναι λιγότερο αξιόπιστες αυτές οι προβλέψεις όσο αυξάνονται οι ημέρες που θέλουμε να προβλέψουμε.
- ο Ακολουθούν οι τρόποι επιλογής των παραμέτρων



Εύρεση των Παραμέτρων του μοντέλου

- ο Η παράμετρος d:
- Οι τιμές μίας μετοχής στο χρηματιστήριο δεν έχουν σταθερό μέσο όρο.
- Ομως οι επιστροφές (δηλαδή οι διαφορές) του χρηματιστηρίου τείνουν να έχουν σταθερό μέσο όρο.
- Εάν οι διαφορές δεν έχουν σταθερό μέσο όρο τότε θα πρέπει να συνεχίσουμε τον έλεγχο για τις διαφορές των διαφορών (d = 2) , συνεχίζοντας έτσι τους ελέγχους βρίσκουμε το ελάχιστο d το οποίο θα μας δημιουργήσει μία χρονοσειρά με σταθερό μέσο όρο.
- Αυτό επίσης θα είναι το βέλτιστο d για το πρόβλημα μας



Εύρεση των Παραμέτρων του μοντέλου

Η παράμετρος ρ:

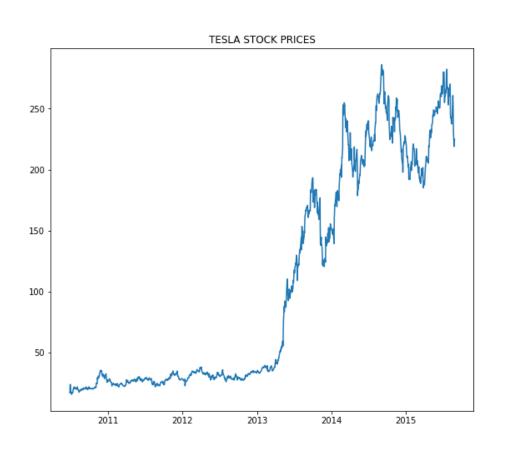
Εξετάζοντας την γραφική παράσταση της συνάρτησης μερικής αυτοσυσχέτισης (PACF) της διαφοροποιημένης σειράς, μπορούμε να προσδιορίσουμε προσωρινά την παράμετρο p.

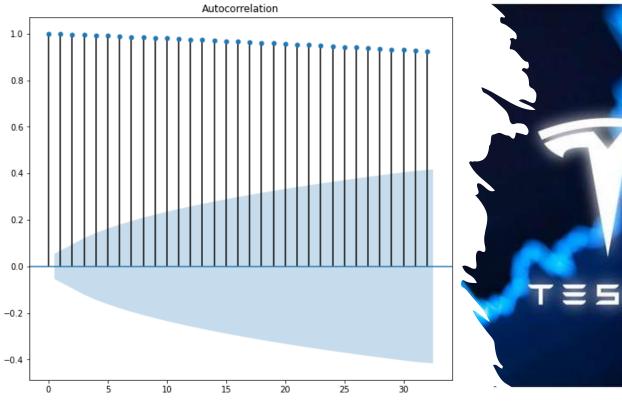
Η γραφική παράσταση PACF είναι μια γραφική παράσταση των συντελεστών μερικής συσχέτισης μεταξύτης σειράς και των καθυστερήσεων (lags) της ίδιας.



- ο Παρόμοια, εξετάζοντας την γραφική παράσταση της συνάρτησης αυτοσυσχέτισης (ACF) της διαφοροποιημένης σειράς, μπορούμε να προσδιορίσουμε προσωρινά την παράμετρο q.
- ο Τέλος έχοντας προσδιορίσει τις παραμέτρους p, d, q μπορούμε να υλοποιήσουμε το μοντέλο ARIMA.
- Έπειτα μπορούμε να το αξιολογήσουμε σύμφωνα με το πόσο καλά προσεγγίζει τις αληθινές τιμές και με το άν αυτές οι τιμές βρίσκονται εντός των τιμών των διαστημάτων εμπιστοσύνης.

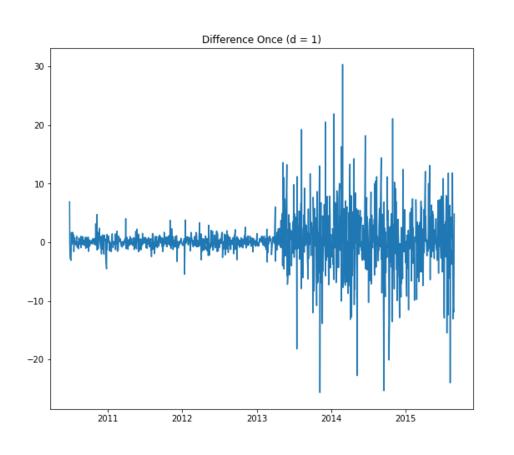
Χρονοσειρά / Αυτοσυσχέτιση

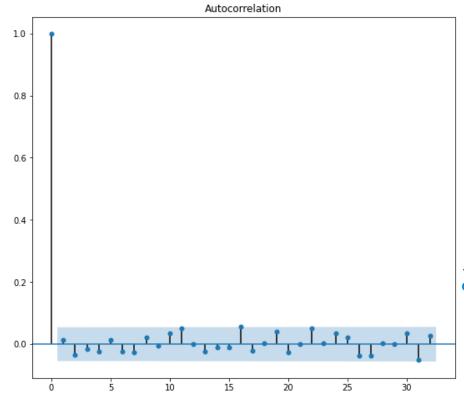






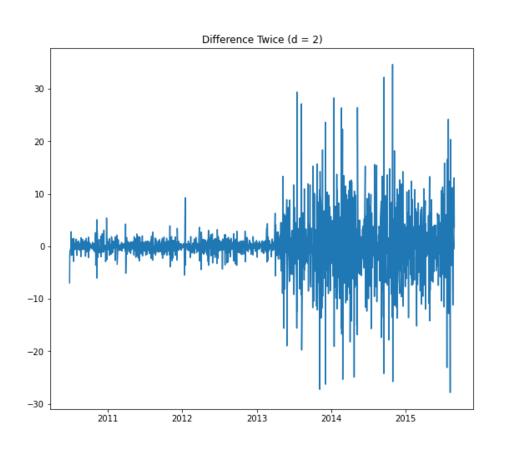
Χρονοσειρά / Αυτοσυσχέτιση (d=1)

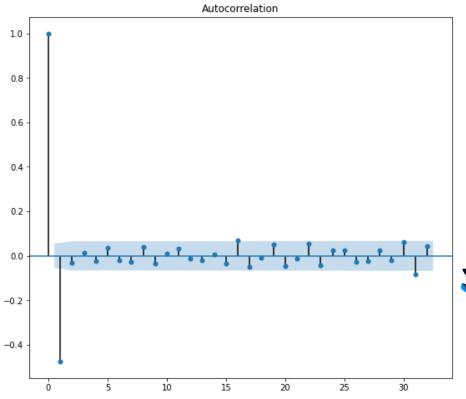






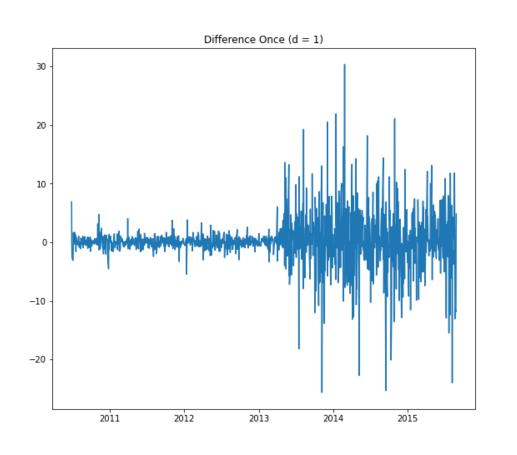
Χρονοσειρά / Αυτοσυσχέτιση (d=2))

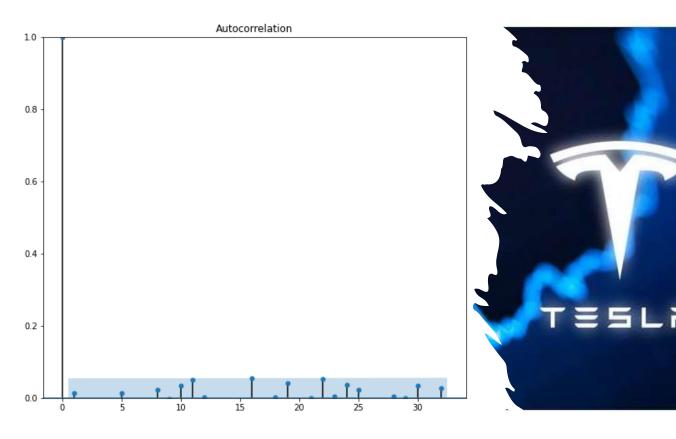




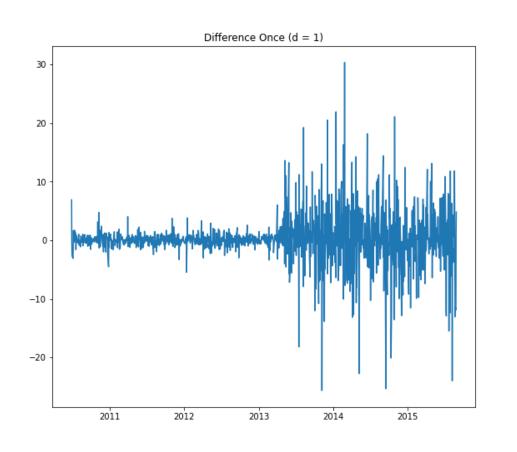


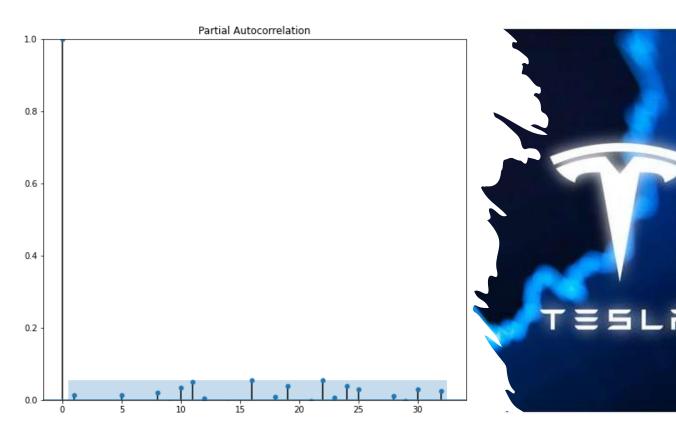
Λόγος επιλογής p=11



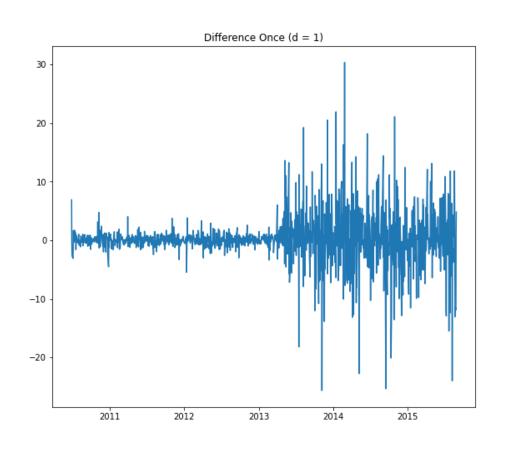


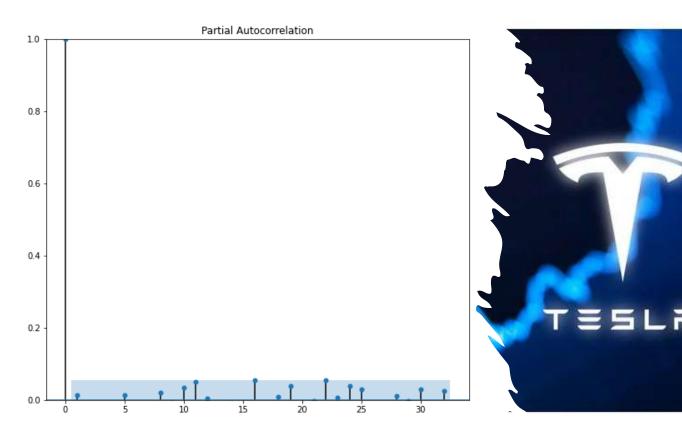
Λόγος επιλογής q=11



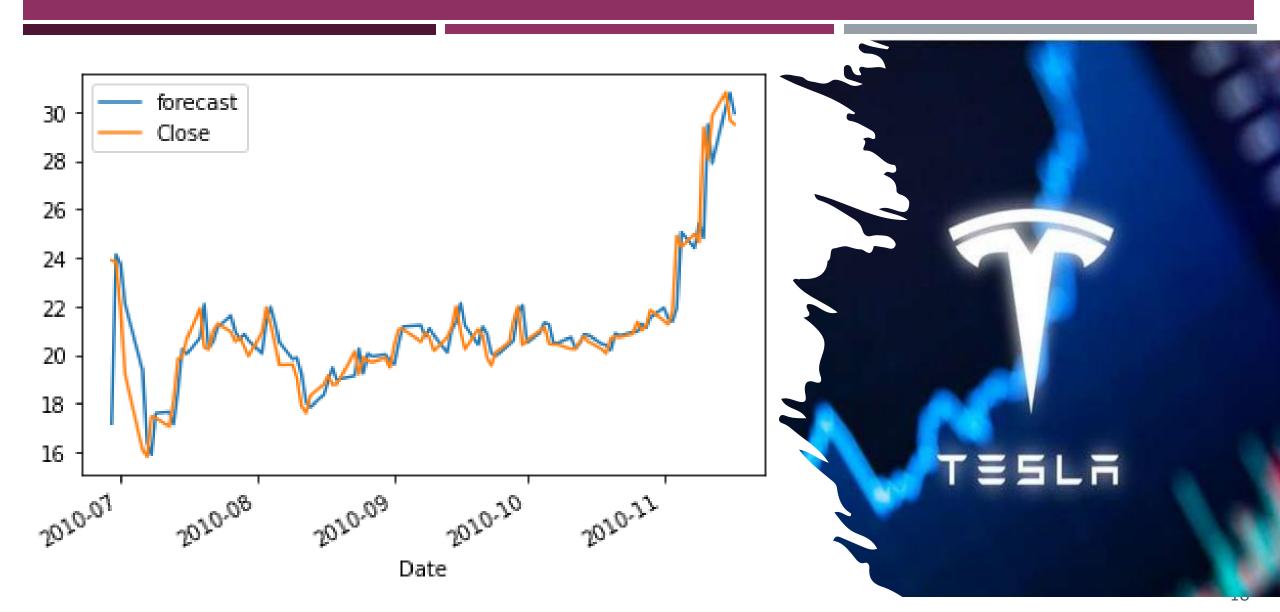


Λόγος επιλογής p=11

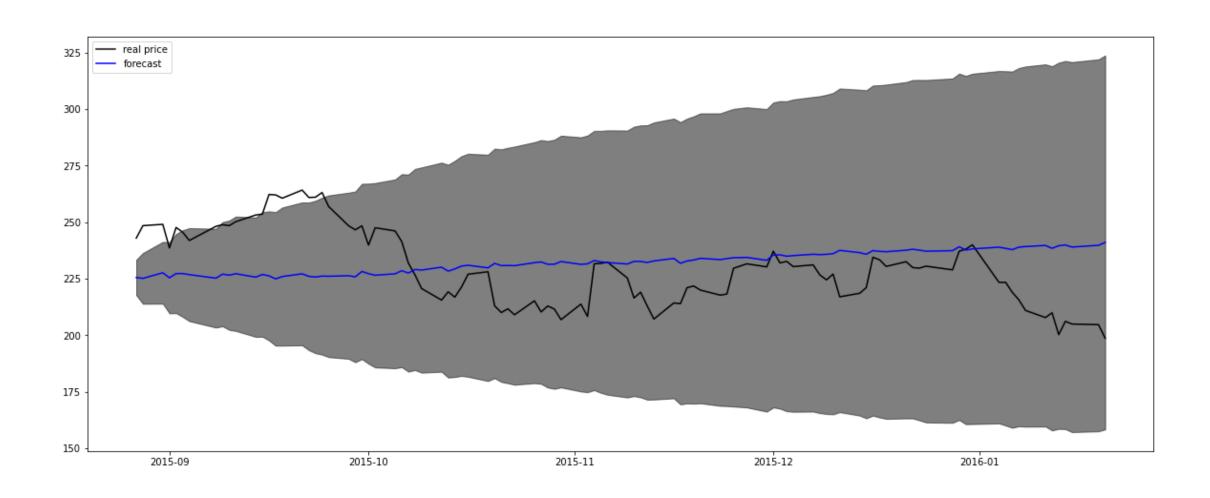




Οπτικοποίηση Αποτελέσματος ΑRIMA



Αξιολόγηση Μοντέλου

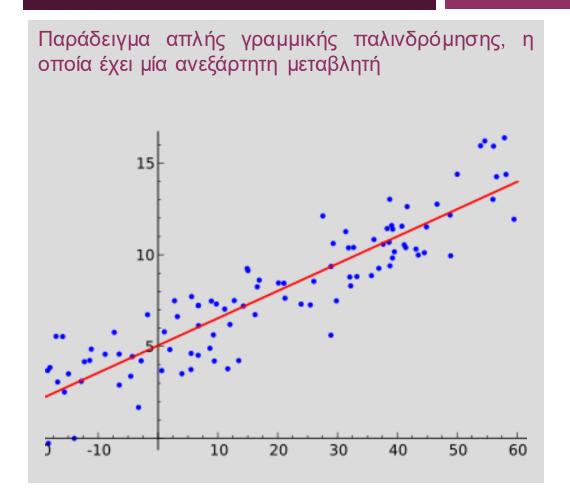


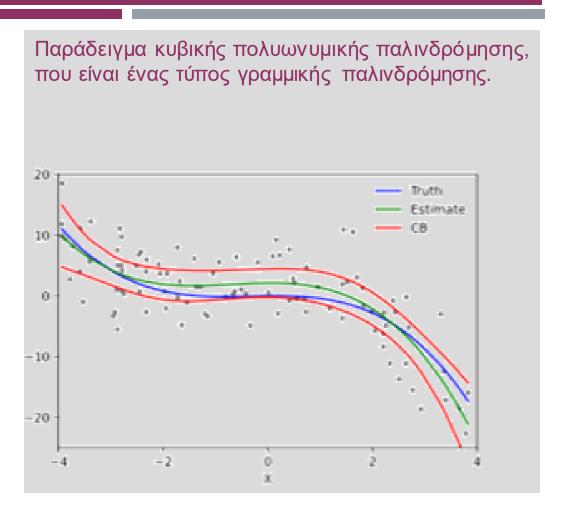


Περιγραφή Αλγορίθμων Γραμμική Παλινδρόμηση

- ο Γραμμική Μη γραμμική Παλινδρόμηση
 - Στη στατιστική, η γραμμική παλινδρόμηση είναι μια προσέγγιση για μοντελοποίηση σχέσεων μεταξύ μιας βαθμωτής εξαρτημένης μεταβλητής Υ και επεξηγηματικών μεταβλητών.
 - Στην περίπτωση μίας επεξηγηματικής μεταβλητής έχουμε απλή γραμμική παλινδρόμηση ενώ στην περίπτωση πολλών έχουμε πολλαπλή.
 - Στην γραμμική παλινδρόμηση,τα δεδομένα μοντελοποιούνται χρησιμοποιώντας γραμμικές λειτουργίες προγνωστικά, και οι άγνωστες παράμετροι μοντέλου υπολογίζονται από τα δεδομένα.

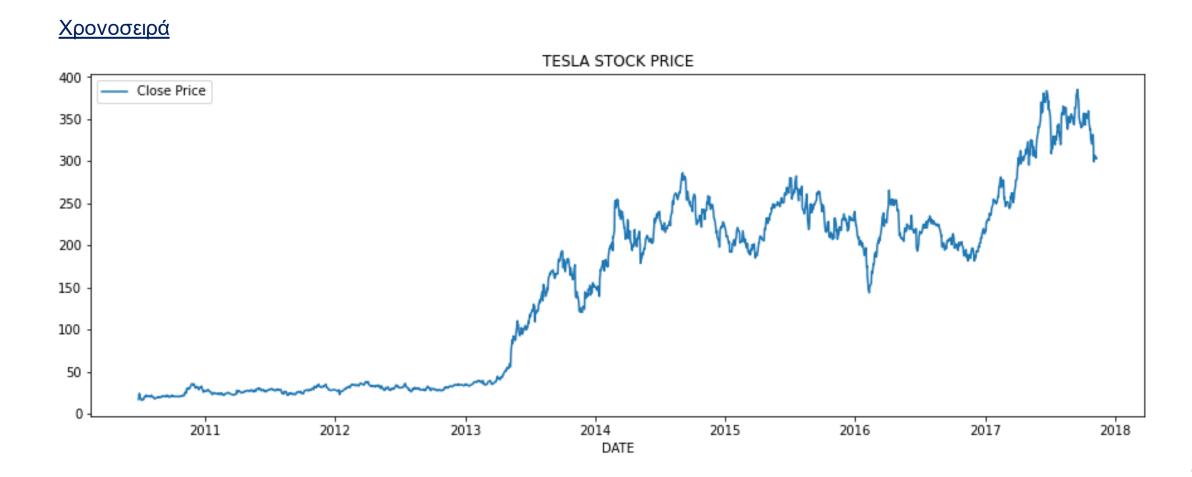
Μορφή Γραμμικής - Μη γραμμικής παλινδρόμισης





Εφαρμογή τουΜοντέλου Γραμμικής Παλινδρόμησης στη μετοχή της TESLA

- Ι. Αρχικά ανοίξαμε το αρχείο και προσθέσαμε τις αντίστοιχες βιβλιοθήκες που θα χρειαστούμε.
- ΙΙ. Στη συνέχεια διαβάσαμε το αρχείο και οπτικοποιήσαμε την χρονοσειρά.



Εφαρμογή του Μοντέλου Γραμμικής Παλινδρόμησης στη μετοχή της TESLA

- ο <u>Ο στόχος (target) θα είναι :</u>
 - ο Close: Η τιμή της μετοχής στο κλείσιμο του χρηματιστηρίου
- ο Επειτα θα χρησιμοποιήσουμε αρκετά χαρακτηριστικά (features) τα οποία θα πρέπει να δημιουργήσουμε από τα στοιχεία που μας δίνονται στο σύνολο δεδομένων μας.
 - Τα ήδη υπάρχοντα χαρακτηριστικά που θα χρησιμοποιήσουμε είναι :
 - Open : Η τιμή της μετοχής στο άνοιγμα του χρηματιστηρίου.
 - Prev Close : Η τιμή της μετοχής την προηγούμενη μέρα.
 - Prev Volume : Το άθροισμα των ενεργών μετοχών την προηγούμενη ημέρα.
- Στην συνέχεια κατασκευάζουμε κάποια χρήσιμα χαρακτηριστικά χρησιμοποιώντας τα επόμενα.



Εφαρμογή του Μοντέλου Γραμμικής Παλινδρόμησης στη μετοχή της TESLA

Ως χαρακτηριστικά χρησιμοποιούμε:

- ο τους απλούς κινητούς μέσους (SMA : Simple Moving Average) των 5, 10, 20, 50, 100 και 200 ημερών τους οποίους τους υπολογίζουμε μέσω του χαρακτηριστικού Prev Close.
- Επειτα δημιουργούμε έναν βραχυπρόθεσμο και έναν μακροπρόθεσμο εκθετικό κινητό μέσο (EMA exponential moving average) από τους οποίους θα χρησιμοποιήσουμε ως χαρακτηριστικό την διαφορά τους.



Εφαρμογή του μοντέλου Γραμμικής Παλινδρόμησης στην Tesla

- Στην συνέχεια χρησιμοποιούμε ως χαρακτηριστικό την τάση του δείκτη MACD υπολογίζοντας τον εκθετικό του κινητό μέσο.
- Με σκοπό να εκπαιδεύσουμε το μοντέλο μας να υπολογίζει την τιμή σε μία απλοποιημένη μορφή υπολογίζουμε τον δείκτη σχετικής δύναμης.
- ο Με την ίδια λογική υπολογίζουμε το RSI για το Volume

$$RSI = 100 \times \frac{EMA[N] \ of \ U}{(EMA[N] \ of \ U) + (EMA[N] \ of \ D)}$$

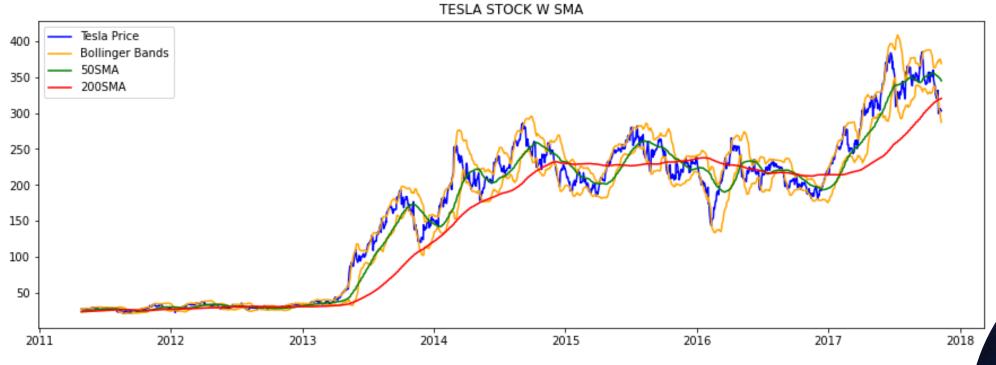


Εφαρμογή του μοντέλου Γραμμικής Παλινδρόμησης στην Tesla

- Θέλουμε να ξέρουμε πότε η μετοχή φτάνει σε ακραίες τιμές έτσι ώστε το μοντέλο μας να μπορεί να προβλέψει την πιθανότητα του ενδεχομένου επιστροφής της μετοχής στην προηγούμενη τιμή.
- Αυτό μπορούμε να το πετύχουμε προσθέτοντας Bollinger Bands, όπου αυτά στην ουσία είναι δύο δεσμίδες σε απόσταση διπλάσια της τυπικής απόκλισης από την μέση τιμή.
- Τέλος προσθέτουμε ως χαρακτηριστικά τις ποσοστιαίες μεταβόλες που έχουν γίνει σε ένα παράθυρο χρόνου x. Συγκεκριμένα επιλέγουμε τα x = 1, 2, 5,

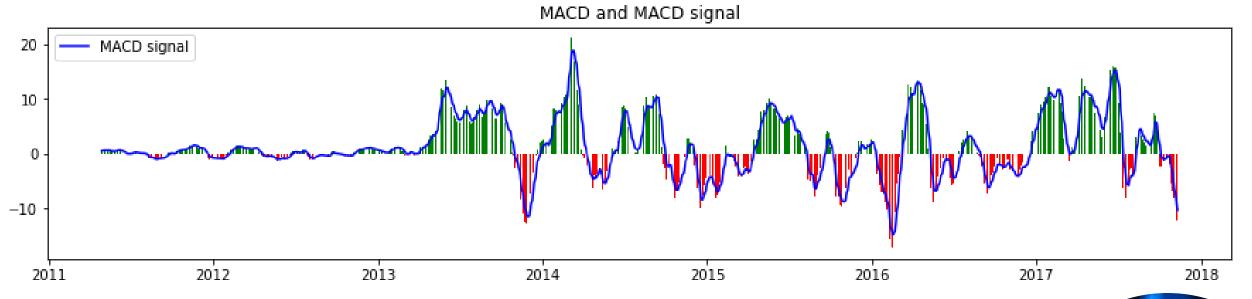


Αποτελέσματα Απλών Κινητών Μέσων



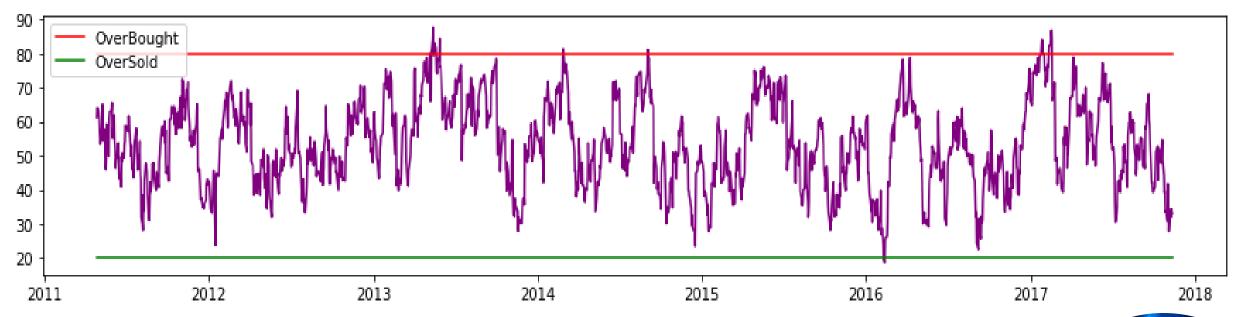


Δείκτης Κινητής Μέσης Σύγκλισης / Απόκλισης



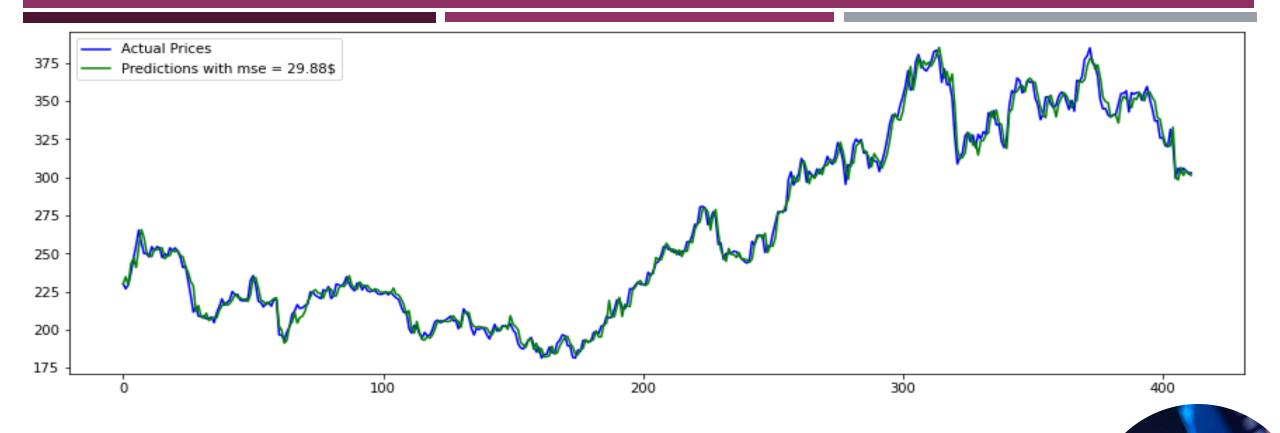


Υπεραγορασμένες / Υπερπωλημένες





Αποτέλεσμα Γραμμικής Παλινδρόμησης



Περιγραφή Αλγορίθμων Νευρωνικό Δίκτυο LSTM

- ο Στην διαδικασία πρόβλεψης των τιμών κάποιας μετοχής του χρηματιστηρίου ένας, ιδιαίτερα σημαντικός, παράγοντας που πρέπει να λάβουμε υπόψιν είναι τα παρελθοντικά δεδομένα της μετοχής αυτής.
- ο Ένα οποιοδήποτε άλλο νευρωνικό δίκτυο αποτυγχάνει στην διατήρηση, και έπειτα χρήση, αυτών των δεδομένων.
- Από την συλλογή των επαναλαμβανόμενων νευρωνικών δικτύων, το LSTM έχει την ικανότητα να διατηρεί, και έπειτα να χρησιμοποιεί, δεδομένα των οποίων το «χρονικό» κενό με το παρόν είναι «μεγάλο».
- ο Αυτή η ικανότητα του LSTM είναι διαισθητικά λογικό ότι θα μας φανεί χρήσιμη πρόβλεψη των τιμών μίας μετοχής.

Εφαρμογή του LSTM στη μετοχή της Tesla

- ο Στην Ορολογία του LSTM η μακροπρόθεσμη μνήμη καλείται cell state
- ο Έτσι από εδώ και πέρα θα χρησιμοποιείται αυτός ο όρος
- ο Η διαδικασία που ακολουθεί το LSTM που θα χρησιμοποιήσουμε χωρίζεται σε 4 μέρη

1ο μέρος :

- Αρχικά επιλέγουμε τι πληροφορία θα κρατήσουμε και τι πληροφορία θα αφαιρέσουμε από το cell state Ct-1.
- Αυτό θα το επιτύχουμε μέσω ενός layer με σιγμοειδή συνάρτηση ενεργοποίησης (sigmoid activation function) όπου τα ακραία σημεία ως έξοδο τα αντιλαμβανόμαστε ως εξής:
- \circ 1 \rightarrow το κρατάμε πλήρως
- \circ 0 \rightarrow το αφαιρούμε πλήρως



ο 2ο μέρος:

- Στο δεύτερο βήμα αποφασίζουμε εάν υπάρχει κάποια καινούργια πληροφορία που χρειάζεται να αποθηκεύσουμε στο cell state Ct
- Αυτό το επιτυχγάνουμε χρησιμοποίωντας 2 layers, το πρώτο layer έχει σιγμοειδή συνάρτηση ενεργοποίησης και μέσω αυτού επιλέγουμε ποιές τιμές θα ενημερώσουμε
- Το δεύτερο layer έχει υπερβολική εφαπτομένη ως συνάρτηση ενεργοποίησης (tanh) και από αυτό θα δημιουργηθεί ένα διάνυσμα, Ct, με νέες υποψήφιες τιμές που μπορούν να προστεθούν στο cell state Ct.



ο 3ο μέρος:

- Σε αυτό το σημείο έχουμε ήδη αποφασίσει τι πληροφορία θα κρατήσουμε και τι πληροφορία θα αφαιρέσουμε, και άρα μπορούμε πλέον να υπολογίσουμε ποίο θα είναι το καινούργιο cell state Ct.
- Έχουμε : Ct = ft * Ct-1 + it* C^{*}t όπου :
- o $ft = \sigma(Wf \cdot [ht-1, xt] + bf)$
- $\circ it = \sigma(Wi \cdot [ht-1, xt] + bi)$
- \circ C^{*}t = tanh(WC · [ht-1, xt] + bC)



ο 4ο μέρος:

- Τέλος, επιλέγουμε την πληροφορία που θα εξάγουμε (στην περίπτωση μας την τιμή της μετοχής).
- Πρώτα θα χρησιμοποιήσουμε ένα layer με σιγμοειδή συνάρτηση ενεργοποίησης το οποίο θα αποφασίσει ποία μέρη του cell state θα εξάγουμε.
- Έπειτα περνάμε το cell state από μία υπερβολική εφαπτομένη tanh η οποία κανονικοποιεί την πληροφορία του cell state μεταξύ των τιμών -1 και 1.
- Πολλαπλασιάζοντας αυτά τα δύο μεγέθη θα έχουμε την συνάρτηση που θα παράγει τις προβλέψεις μας.



Εφαρμογή του LSTM στη μετοχή της Tesla

Έτσι έχουμε : ot = $\sigma(\text{Wo} \cdot [\text{ht-1}, \text{ xt}] + \text{bo})$

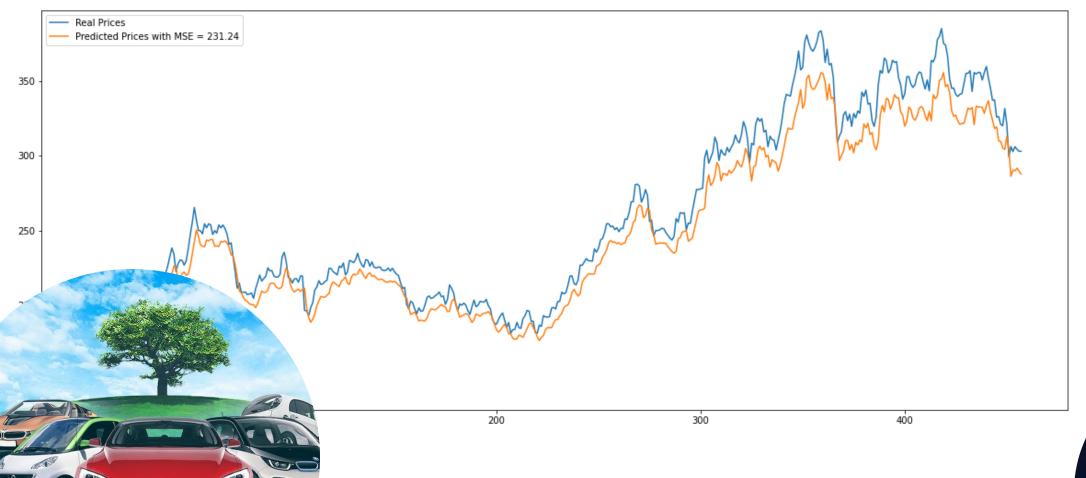
- \circ Άρα ht = ot * tanh(Ct)
- ο Για αυτό το νευρωνικό δίκτυο θα χρησιμοποιήσουμε τον βελτιστοποιητή Adam όπου:

$$w_{t} = w_{t-1} - \eta \frac{\hat{m}_{t}}{\sqrt{\hat{v}_{t}} + \epsilon}$$

ο Και ως συνάρτηση απώλειας την Mean Squared Error.



Αποτελέσμα LSTM





Δημιουργία Στρατηγικής

- ο Από ότι είδαμε, η γραμμική παλινδρόμηση και το νευρωνικό δίκτυο LSTM φαίνεται να εκπαιδεύονται αρκετά ικανοποιητικά. ΄
 - Αυτό που μας ενδιαφέρει είναι αν οι στρατηγικές που προκύπτουν είναι πιό κερδοφόρες από το εάν απλώς είχαμε επενδύσει χωρίς να ακολουθήσουμε κάποια στρατηγική. ΄
- ο Έτσι φτιάξαμε μία προσομοίωση στην οποία έχουμε έναν υποθετικό λογαριασμό αξίας 1000 δολάρια, έτσι ώστε να συγκρίνουμε αυτές τις στρατηγικές.

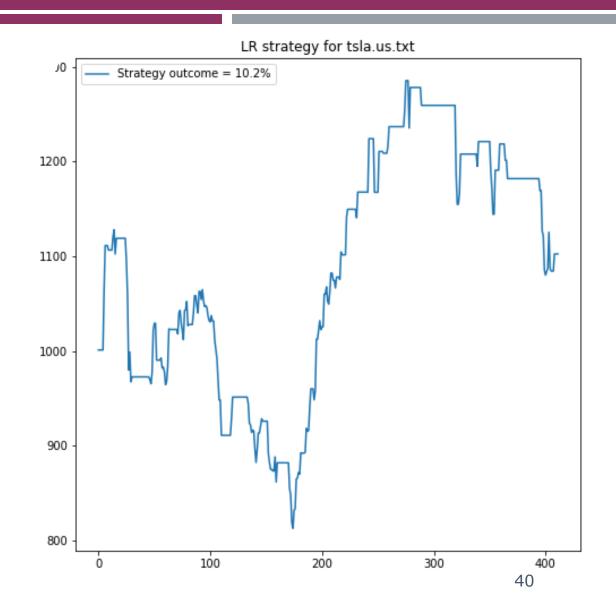


Στρατηγική με L-R στην Tesla

<u>1. Χωρίς στρατηγική : + 31.2 %</u>

2. Linear Regression: + 10.2 %

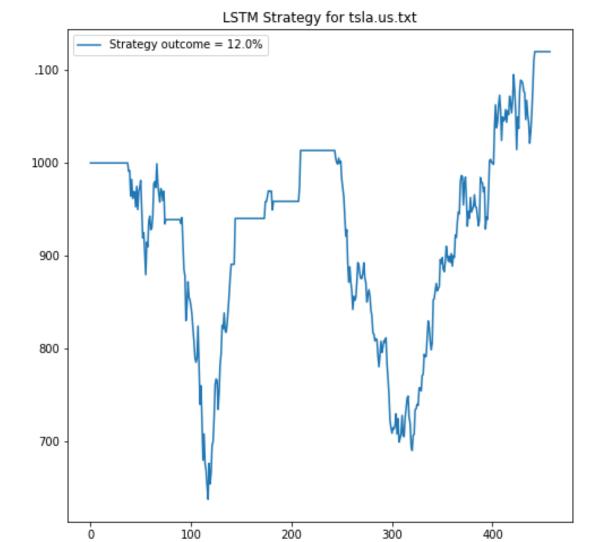




Στρατηγική με LSTM - Tesla

LSTM: + 12 %





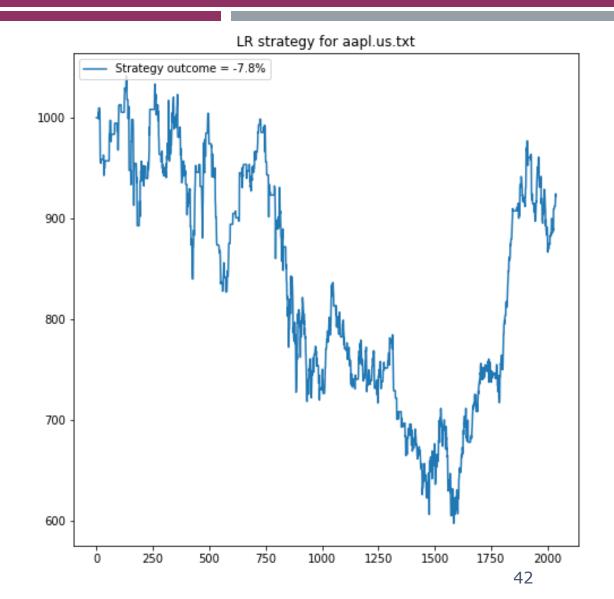
41

Στρατηγική με L-R στην Apple

1. Χωρίς στρατηγική : + 632.5 %

2. Linear Regression: -7.8%

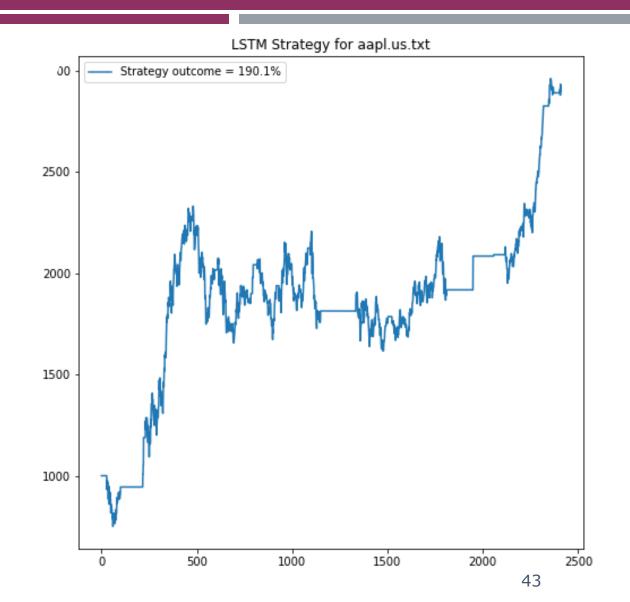




Στρατηγική με L-R στην Apple

LSTM: +190.1 %



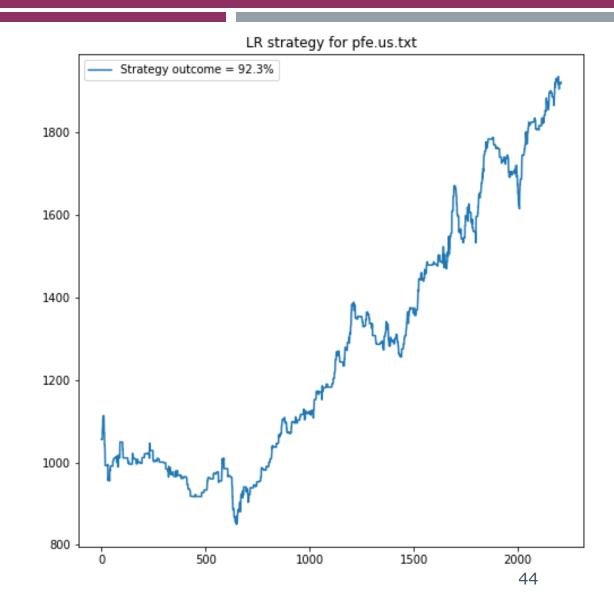


Στρατηγική με L-R στη Pfizer

1. Χωρίς στρατηγική: +212.2 %

2. Linear Regression: +92.3 %

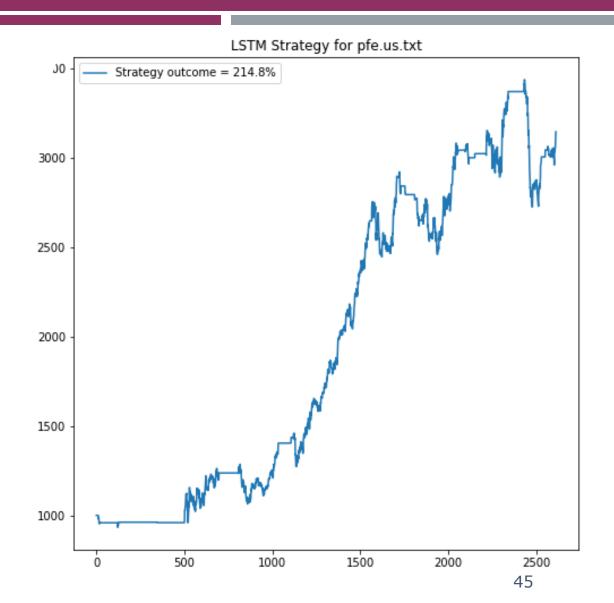




Στρατηγική με LSTM - Pfizer

LSTM: +214.8 %





Συμπεράσματα

- Σύμφωνα με την ανάλυση , την επεξεργασία αλλά και τη σύγκριση των τριών αυτών μοντέλων καταλήγουμε στα εξής:
 - 1. Και τα τρία μοντέλα προβλέπουν σε ικανοποιητικό βαθμό τις διακυμάνσεις των τιμών αν αναλογιστούμε τη δυσκολία του προβλήματος.
 - 2. Οι διαφορές μεταξύ τους είναι μικρές
 - 3. Και τα τρία μοντέλα αποκλίνουν όταν χρειάζεται να προβλέψουν περισσότερες μέρες.
 - 4. Άρα η διεξαγωγή αποτελεσμάτων για μακροχρόνιες προβλέψεις αποτελεί ένα θέμα που χρειάζεται παραπάνω έρευνα.

Συμπεράσματα

Σύμφωνα με την ανάλυση , την επεξεργασία αλλά και τη σύγκριση των τριών αυτών μοντέλων καταλήγουμε στα εξής:

- Και τα τρία μοντέλα προβλέπουν σε ικανοποιητικό βαθμό τις διακυμάνσεις των τιμών αν αναλογιστούμε τη δυσκολία του προβλήματος.
- ΙΙ. Οι διαφορές μεταξύ τους είναι μικρές
- ΙΙΙ. Και τα τρία μοντέλα αποκλίνουν όταν χρειάζεται να προβλέψουν περισσότερες μέρες.
- ΙV. Άρα η διεξαγωγή αποτελεσμάτων για μακροχρόνιες προβλέψεις αποτελεί ένα θέμα που χρειάζεται παραπάνω έρευνα.

Συμπεράσματα Στρατηγικής

Παρατηρούμε ότι γενικά η στρατηγική του LSTM έχει καλύτερη απόδοση από την στρατηγική της γραμμικής παλινδρόμης, και τα δύο όμως έχουν χειρότερη απόδοση από το να μην ακολουθούσαμε στρατηγική (εκτός της PFIZER)

Βιβλιογραφία

- o G'eron, A.Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow; O'Reilly Media, 2020
- https://iopscience.iop.org/article/10.1088/1742-6596/2161/1/012065/pdf
- https://kesmarag.gitlab.io/courses/mem205/
- H. L. Siew and M. J. Nordin, "Regression techniques for the prediction of stock price trend," 2012 International Conference on Statistics in Science,
- Business and Engineering (ICSSBE), Langkawi, 2012, pp. 1-5.
- https://colah.github.io/posts/2015-08-Understanding-LSTMs/?fbclid=IwAR3u8q2ccmPx2DjuO2wsUU9QXkLhzbuzUWDs4hdsorLxPWLxsGBiP6OXHQ
- https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average



Σας ευχαριστούμε πολύ για το χρόνο σας και θα χαρούμε να απαντήσουμε τις ερωτήσεις σας.

Ομάδα 3 : Καραγεώργης Στέργιος ΑΜ:2346 Πανουργιάς Χρήστος ΑΜ:2405

Ιούνιος 2022