

ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

---

# Βαθιά Μάθηση και Ανάλυση Συναισθήματος για Οικονομικά Δεδομένα

---

## Τσορμπατζούδης Δημοσθένης



Σχολή Θετικών Επιστημών

Τμήμα Πληροφορικής

Ιούνιος 2021

*Αυτή η πτυχιακή κατατέθηκε για  
την απόκτηση Πτυχίου Πληροφορικής*

**Copyright © 2021 Τσορμπατζούδης Δημοσθένης**

Typeset by the author with the X<sub>E</sub>L<sup>A</sup>T<sub>E</sub>X documentation system.

Ημερομηνία κατάθεσης, Ιούνιος 2021.



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ  
Σχολή Θετικών Επιστημών  
Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

---

Βαθιά Μάθηση και Ανάλυση Συναισθήματος  
για Οικονομικά Δεδομένα

---

## Τσορμπατζούδης Δημοσθένης

Επιβλέπων Καθηγητής:  
Τέφας Αναστάσιος

Σχολή Θετικών Επιστημών  
Τμήμα Πληροφορικής

Ιούνιος 2021



# Περίληψη

---

Στη παρούσα εργασία θα εξεταστεί η χρήση βαθιάς μάθησης για την πρόβλεψη χρονοσειρών οικονομικών δεδομένων. Πιο συγκεκριμένα θα χρησιμοποιήσουμε στα πειράματα μας δεδομένα ιστοιμίων νομισμάτων και χρυπτονομισμάτων. Επιπλέον θα κάνουμε χρήση τεχνικών ανάλυσης συναισθήματος τις οποίες θα συνδυάσουμε με νευρωνικά δίκτυα στις πειραματικές μας δοκιμές. Ως αποτελέσματα θα εξετάσουμε τις μετρήσεις κάθε πειράματος και το πως επηρεάζονται από τις διαφορετικές τεχνικές πρόβλεψης καθώς και τις διαφορετικές παραμέτρους. Τέλος θα προσπαθήσουμε να καταλήξουμε σε συμπεράσματα σχετικά με κατά πόσο η πρόβλεψη χρονοσειρών είναι δυνατή και με το κέρδος το οποίο επιφέρουν οι προβλέψεις από τα καλύτερα μοντέλα. Θα αναφερθούμε στις τεχνολογίες νευρωνικών δικτύων MLP, LSTM και GRU και στις τεχνολογίες ανάλυσης συναισθήματος *VADER* και *FinBERT*.

**Λέξεις Κλειδία:** Βαθιά Μάθηση, Ανάλυση Συναίσθηματος, Νευρωνικά Δίκτυα, MLP, LSTM, FOREX, Twitter, Bitcoin.

# Abstract

---

The purpose of this thesis is to find ways to interpretate Financial Data and achieve the best possible results. To reach our desired aim we are going to use Deep Learning techniques and Sentiment Analysis. We are going to probe into currencies exchange data using both conventional currencies (FOREX) and cryptocurrencies.

We will use Neural Networks stuctures to analyze our data and predict their future movements. The categories of Neural Networks that we use in our models are MLP, LSTM and GRU. To achieve better results and compare different approaches we will use both Classification and Regression predictions for our models.

Movements of the FOREX indeces have an instant impact on people's sentiment. We are going to analyze data from the social media platfrom *Twitter* to better understand how people feel about the ongoing ups and downs. The tools that we are going to use and *VADER* and *FinBERT*. Both state-of-art technologies trained of enormous amount of data that are going to help us understand the posts related to currencies exchange better.

All of our results will be evaluated by calculating various metrics and the P&L of the transactions using real world data.

**Keywords:** Deep Learning, Sentiment Analysis, Neural Networks, MLP, LSTM, FOREX, Twitter, Bitcoin.

Σε όσους προς γνώσιν πορεύονται



# Ευχαριστίες

---

Για την εκπόνηση αυτής της εργασίας θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή Τέφα Αναστάσιο για την στήριξη και την συμβολή του καθόλη την διάρκεια συγγραφής της εργασίας. Επίσης, θα ήθελα να ευχαριστήσω τον διδακτορικό ερευνητή Πάσσαλη Νικόλαο για την βοηθεία του. Τέλος, ένα μεγάλο ευχαριστώ στην οικογένεια μου που με στήριξε στη συγγραφή της εργασίας και σε όλη την υπόλοιπη ζωή μου.



# Περιεχόμενα

---

<b>Περίληψη</b>	iii
<b>Abstract</b>	v
<b>Ευχαριστίες</b>	ix
<b>Περιεχόμενα</b>	xi
<b>Λίστα Σχημάτων</b>	xiii
<b>Λίστα Πινάκων</b>	xv
<b>0 Εισαγωγή</b>	1
<b>1 Προηγούμενο Έργο</b>	3
<b>2 Βαθιά Μάθηση</b>	5
2.1 Ορισμός . . . . .	5
2.2 Metrics . . . . .	6
<b>3 Νευρωνικά Δίκτυα</b>	9
3.1 Perceptron . . . . .	9
3.2 Εισαγωγή στα Νευρωνικά Δίκτυα. . . . .	10
3.3 Αρχιτεκτονική . . . . .	10
3.4 Συναρτήσεις Ενεργοποίησης . . . . .	11
3.5 Συνάρτηση Κόστους . . . . .	12
3.6 Gradient Descent . . . . .	13
3.7 Εκπαίδευση Νευρωνικών Δικτύων . . . . .	15
3.7.1 Τροφοδότηση (Feedforward-propagation) . . . . .	15
3.7.2 Οπισθοδρόμηση (Back-propagation) . . . . .	16
3.8 Αναδρομικά Νευρωνικά Δίκτυα - RNN . . . . .	18
3.9 Δικτύα Μακράς Βραχυπρόθεσμης Μνήμης - LSTM . . . . .	18
3.10 Δίκτυα Περιφραγμένης Επαναλαμβανόμενης Μονάδας - GRU . . . . .	19

<b>4 Χρηματοοικονομική Αγορά</b>	<b>21</b>
4.1 FOREX . . . . .	21
4.1.1 Candlesticks . . . . .	22
4.2 Profit and Loss - P&L . . . . .	22
4.3 Commision . . . . .	23
<b>5 Επεξεργασία Φυσικής Γλώσσας</b>	<b>25</b>
5.1 Προεπεξεργασία Κειμένου . . . . .	25
5.2 Bag-of-Words (BoW) . . . . .	26
5.3 Term Frequency-Inverse Document Frequency (TF-IDF) . . . . .	26
5.4 Word2Vec . . . . .	27
5.5 VADER . . . . .	27
5.6 Bert . . . . .	28
<b>6 Δεδομένα</b>	<b>29</b>
6.1 Προεπεξεργασία Δεδομένων . . . . .	29
6.2 Προεπεξεργασία Δεδομένων Ισοτιμιών . . . . .	30
6.3 Προεπεξεργασία Λεξικολογικών δεδομένων . . . . .	31
6.4 Ετικέτες Δεδομένων . . . . .	31
<b>7 Πειράματα και Αποτελέσματα Βαθιάς Μάθησης</b>	<b>33</b>
7.1 Τεχνικά χαρακτηριστικά . . . . .	33
7.2 Πρόβλεψη κατεύθυνσης - Classification . . . . .	34
7.2.1 Αποτελέσματα Μοντέλων . . . . .	34
7.2.2 Εκπαίδευση μόνο με EUR / USD . . . . .	38
7.2.3 Άλλαγη τρόπου δημιουργίας ετικετών. . . . .	41
<b>8 Πρόβλεψη τιμής - Regression</b>	<b>47</b>
8.1 Αποτελέσματα Μοντέλων EUR / USD . . . . .	49
8.2 Πειράματα Threshold . . . . .	50
8.3 Πειράματα διαφορετικών Ισοτιμιών . . . . .	52
<b>9 Πειράματα και Αποτελέσματα Συναισθηματικής Ανάλυσης</b>	<b>55</b>
9.1 Ανάλυση Συναισθήματος . . . . .	55
9.2 Πειράματα με χρήση Vader . . . . .	55
9.3 Πειράματα με χρήση FinBert . . . . .	61
<b>10 Συμπεράσματα</b>	<b>65</b>
<b>Βιβλιογραφία</b>	<b>67</b>

# Λίστα Σχημάτων

---

0.1 Λογότυπο . . . . .	2
3.1 Η αναπαράσταση ενός νευρώνα 2 εισόδων . . . . .	9
3.2 Αναπάσταση των Hyperplane για 2 και 3 διαστάσεις . . . . .	10
3.3 Αρχιτεκτονική ενός πλήρους συνδεδεμένου νευρωνικού δικτύου με 3 χρυφά επίπεδα . . . . .	11
3.4 Συναρτήσεις Ενοργοποίησης . . . . .	12
3.5 Η εκπαίδευση ενός νευρωνικού δικτύου . . . . .	14
3.6 Η αναπαράσταση της συνάρτησης $J$ προς $w$ για την μείωση του κόστους . . . . .	14
3.7 Αναπαράσταση της διαδικασίας Gradient Descent με διαφορετικές τιμές στο learning rate . . . . .	15
3.8 Η αρχιτεκτονική ένος Αναδρομικού Νευρωνικού Δικτύου . . . . .	18
3.9 Τα περιεχόμενα ένος νευρώνα LSTM . . . . .	19
3.10 Τα περιεχόμενα ένος νευρώνα GRU . . . . .	19
4.1 <i>EUR/AUD</i> Candlesticks . . . . .	22
5.1 Η αρχιτεκτονική των 2 μεθόδων του Word2Vec . . . . .	28
5.2 Οι φάσεις εκπαίδευσης του μοντέλου <i>BERT</i> . . . . .	28
6.1 Οι τιμές της ισοτημίας <i>EURUSD</i> . . . . .	29
6.2 Οι τιμές ανοίγματος νομισμάτων . . . . .	30
6.3 Οι τιμές κλεισίματος του <i>Bitcoin - USD</i> . . . . .	30
6.4 Η ποσοστιαία μεταβολή των τιμών κλεισίματος του <i>EUR - USD</i> και <i>EUR - JPY</i> αντίστοιχα . . . . .	31
6.5 Ένα <i>tweet</i> πριν και μετά την προεπεξεργασία . . . . .	31
6.6 Παράδειγμα χρήσης Triple Barrier Labels . . . . .	32
7.1 Ένα νευρωνικό δίκτυο Classification . . . . .	35
7.2 Γράφημα τιμών P&L για το 2000-2020 . . . . .	38
7.3 Γράφημα συνολικού P&L για το 2000-2019 . . . . .	38
7.4 Γράφημα συνολικού P&L για το 2019-2020 . . . . .	39
7.5 Σύγχριση συνολικού P&L για το 2019-2020 με διαφορετικά δεδομένα εκπαίδευσης . . . . .	40
7.6 Γράφημα συνολικού P&L για διάφορες τιμές average future window . . . . .	42

7.7 Γράφημα συνολικού P&L για διάφορες τιμές barrier length . . . . .	45
8.1 Ένα νευρωνικό δικτύo Regression . . . . .	48
8.2 Πραγματικές και προβλεπόμενες τιμές της ισοτιμίας <i>EUR/USD</i> . . . . .	51
8.3 Κόστος δεδομένων εκπαίδευσης και αξιολόγησης . . . . .	51
8.4 Αποτελέσματα συνολικού P&L με χρήση διάφορων τιμών threshold . . . . .	52
9.1 Συνολικό συναίσθημα και τιμή κλεισίματος για <i>BTC / USD</i> με την χρήση VADER . . . . .	56
9.2 Ένα νευρωνικό δικτύo Regression με Vader . . . . .	57
9.3 Μεταβολές P&L με χρήση και μη συναισθηματικής ανάλυσης . . . . .	60
9.4 Μεταβολές συνολικού P&L με χρήση και μη συναισθηματικής ανάλυσης . . . . .	60
9.5 Κόστος δεδομένων εκπαίδευσης και αξιολόγησης με χρήση VADER . . . . .	61
9.6 Συνολικό συναίσθημα και τιμή κλεισίματος για <i>BTC / USD</i> με την χρήση FinBERT . . . . .	61
9.7 Ένα νευρωνικό δικτύo Regression με FinBERT . . . . .	62
9.8 Κόστος δεδομένων εκπαίδευσης και αξιολόγησης με χρήση FinBERT . . . . .	63
9.9 Σύγχριση συνολικού P&L για τις μεθόδους συναισθηματικής ανάλυσης VADER και FinBERT . . . . .	64

# Λίστα Πινάκων

---

3.1	Πίνακας Συναρτήσεων Ενεργοποίησης . . . . .	12
4.1	Ισοτιμίες του Ευρώ . . . . .	21
5.1	Παράδειγμα Bag-of-Words . . . . .	26
7.1	Πραγματικές και προβλεπόμενες κατευθύνσεις για <i>EUR/USD</i> . . . . .	34
7.2	Classification αποτελέσματα P&L με MLP για <i>EUR/USD</i> . . . . .	35
7.3	Classification αποτελέσματα F1 με MLP για <i>EUR/USD</i> . . . . .	36
7.4	Classification αποτελέσματα P&L με LSTM για <i>EUR/USD</i> . . . . .	36
7.5	Classification αποτελέσματα F1 με LSTM για <i>EUR/USD</i> . . . . .	37
7.6	Classification αποτελέσματα P&L με GRU για <i>EUR/USD</i> . . . . .	37
7.7	Classification αποτελέσματα F1 με GRU για <i>EUR/USD</i> . . . . .	37
7.8	Classification αποτελέσματα P&L με εκπαίδευση μόνο σε <i>EUR/USD</i> . . . . .	39
7.9	Classification αποτελέσματα Metrics με εκπαίδευση μόνο σε <i>EUR/USD</i> . . . . .	40
7.10	Classification αποτελέσματα P&L με χρήση smooth oracle labels με average future window = 5 . . . . .	41
7.11	Classification αποτελέσματα Metrics με χρήση smooth oracle labels με smoothing window = 5 . . . . .	41
7.12	Classification αποτελέσματα P&L με χρήση smooth oracle labels με average future window 5 εώς 10 . . . . .	42
7.13	Classification αποτελέσματα Metrics με χρήση smooth oracle labels με average future window 5 εώς 10 . . . . .	43
7.14	Classification αποτελέσματα P&L με χρήση triple barrier labels με barrier length = 7 . . . . .	43
7.15	Classification αποτελέσματα Metrics με χρήση triple barrier labels με barrier length = 7 . . . . .	44
7.16	Classification αποτελέσματα P&L με χρήση triple barrier labels με barrier length 5 εώς 10 . . . . .	44
7.17	Classification αποτελέσματα Metrics με χρήση triple barrier labels με barrier length 5 εώς 10 . . . . .	45
8.1	Πραγματικές και προβλεπόμενες τιμές για <i>EUR/USD</i> . . . . .	48
8.2	Regression αποτελέσματα P&L για <i>EUR / USD</i> . . . . .	50
8.3	Regression αποτελέσματα MAPE για <i>EUR / USD</i> . . . . .	50

8.4	Regression αποτελέσματα P&L με χρήση διάφορων τιμών threshold . . . . .	52
8.5	Regression αποτελέσματα P&L με χρήση διάφορων ισοτιμιών . . . . .	53
8.6	Regression αποτελέσματα MAPE με χρήση διάφορων ισοτιμιών . . . . .	53
9.1	Αποτελέσματα P&L για Sentiment Analysis με Vader . . . . .	58
9.2	Αποτελέσματα MAPE για Sentiment Analysis με Vader . . . . .	58
9.3	Αποτελέσματα P&L χωρίς την χρήση Sentiment Analysis . . . . .	59
9.4	Αποτελέσματα MAPE χωρίς την χρήση Sentiment Analysis . . . . .	59
9.5	Αποτελέσματα P&L για Sentiment Analysis με <i>FinBERT</i> . . . . .	63
9.6	Αποτελέσματα MAPE για Sentiment Analysis με <i>FinBERT</i> . . . . .	63

# 0

## Εισαγωγή

---

Τα τελευταία χρόνια με τις ραγδαίες εξελίξεις στον τομέα της Βαθιάς Μάθησης και την αύξηση της υπολογιστικής ισχύος όλοι και περισσότεροι κλάδοι διαφορετικών επιστημών συνδυάζονται με τις νέες τεχνολογίες με αποτέλεσμα την δημιουργία νέων κλάδων μελέτης και πειραμάτων. Μια τέτοια περίπτωση είναι η θεματολογία που θα μας απασχολήσει στην παρακάτω εργασία, το πως δηλαδή μπορεί να συνδυαστεί η βαθιά μάθηση με την οικονομική επιστήμη και πιο συγκεκριμένα στο πεδίο της πρόβλεψης τιμών ισοτιμιών. Παρότι η ίδεα της πρόβλεψης των μελλοντικών τιμών ισοτιμιών και δεικτών χρηματιστηρίου είναι τόσο παλιά όσο η ίδια η ίδεα τέτοιου είδους συναλλαγών οι προσπάθειες αυτές δεν είχαν τα επιθυμητά αποτελέσματα. Μαθηματικοί τύποι και συνδυασμοί προηγούμενων τιμών ήταν κάποιες από τις μέθοδους που χρησιμοποιήθηκαν για να βρέθει ένα μοντέλο ικανό να προβλέπει τις τιμές μακροπρόθεσμα. Οι μέθοδοι αυτοί δεν είχαν τα επιθυμητά αποτελέσματα χυρίως λόγω της έλλειψης υπολογιστικής ισχύος για την επεξεργασία του μεγάλου όγκου δεδομένων. Αυτό έχει αλλάξει τα τελευταία χρόνια καθώς με τις νέες τεχνολογίες που αναφέρθηκαν υπάρχουν οι πόροι διαθέσιμοι για την πρόβλεψη κινήσεων των δεικτών. Σε αυτό το σημείο θα αναφερθούμε στο κατά πόσο η πρόβλεψη των μελλοντικών τιμών είναι εφικτή και δεν αποτελεί έναν άπιαστο στόχο. Οι απόψεις διίστανται με κάποιους να θεωρούν ότι οι τιμές δημιουργούνται από όπειρους απρόβλεπτους παράγοντες και είναι ουσιαστικά τυχαίες [Mal73]. Αντιθέτως αρχετοί επενδυτές θεωρούν ότι οι κινήσεις του χρηματιστηρίου μπορούν να προβλεφθούν σε κάποιον βαθμό, αυτή την ίδεα ασπάζονται και πολλοί ερευνητές στον χώρο της Βαθιάς Μάθησης που έχουν αναφέρει μεθόδους στις οποίες γίνεται ορθή πρόβλεψη τιμών με την χρήση νευρωνικών δικτύων [KAYT90]. Μια ακόμα τεχνολογία η οποία έχει δώσει σημαντική ώθηση στην εύρεση τρόπων για πρόβλεψη τιμών μετοχών είναι η Ανάλυση Συναίσθημάτος. Με την άνοδο της υπολογιστικής ισχύος και την εξάπλωση των μέσων κοινωνικής δικτυώσης υπάρχει τρόπος να κατανοηθούν και αναλυθούν τα συναισθήματα που προκαλούν οι κινήσεις των δεικτών. Μέσω των αναλύσεων αυτών δίνονται ακόμα περισσότερα δεδομένα τα οποία μπορούν σε συνδυασμό με Βαθιά Μάθηση να χρησιμοποιηθούν για την πρόβλεψη τιμών [SMAN<sup>+13</sup>].

Σχήμα 0.1: Λογότυπο



# 1

## Προηγούμενο Έργο

---

Σε αυτό το κεφάλαιο θα κάνουμε μια αναφορά σε άλλες εργασίες οι οποίες έδωσαν τις βάσεις, ιδέες και έμπνευση για την συγγραφή της παρακάτω εργασίας. Οι εργασίες αυτές έχουν ως κοινό θέμα την πρόβλεψη τιμών μετοχών και την ανάλυση συναισθήματος: δύο θεματολογίες για τις οποίες έχουν συγγραφεί πολλά έργα και αποτελούν συνεχώς αναπτυσσόμενους κλάδους στον τομέα της Πληροφορικής Επιστήμης.

Οι Matthew Dixona, Diego Klabjanb και Jin Hoon Bangc στην δημοσίευση τους [DKB16] έκαναν πειράματα χρήσης Deep Neural Networks για Classification δύο επιλογών, μειώση ή αύξηση για τιμές εμπορευμάτων (commodity) και συναλλάγματος (FOREX). Τα πειράματα τους επέφεραν υψηλές επιδόσεις ακρίβειας στα δεδομένα αξιολόγησης και κατέληξαν σε χρήσιμα συμπεράσματα.

Οι C.L. Giles, S. Lawrence, Ah Chung Tsoi [GLT97] δημιούργησαν ένα σύστημα πρόβλεψης τιμών με χρήση νευρωνικών δικτύων το οποίο μπορούσε να διαχειριστεί και διαφορετικές συνθήκες δεδομένων όπως θόρυβος ή μικρό μέγεθος δειγμάτων. Το σύστημα επίσης έκανε δημιουργία και ενός συνόλου κανόνων έτσι ώστε να γίνεται πιο κατανοητή η στρατηγική.

Οι Zexin Hu, Yiqi Zhao, and Matloob Khushi [HZK21] χρησιμοποίησαν διάφορες κατηγορίες νευρωνικών δικτύων για την πρόβλεψη των τιμών του FOREX. Χρησιμοποίησαν CNN, LSTM και DNN για την δημιουργία προβλέψεων και σύγχριναν την κάθε κατηγορία. Επιπλέον δοκίμασαν και άλλες τεχνικές βαθιάς μάθησης καταλήγοντας ότι δεν μπορούν να φτάσουν τις επιδόσεις των νευρωνικών δικτύων σε αυτού του είδους τα προβλήματα.

Οι Tsantekidis , Passalis , Toufa, Saitas-Zarkias, Chairistanidis, και Tefas [PTP<sup>+</sup>20] χρησιμοποιούν ένα reinforcement learning περιβάλλον για να δημιουργήσουν έναν trading agent στην αγορά FOREX. Μέσω των προβλέψεων για τις κινήσεις της αγοράς ο agent έχει καταφέρει υψηλές τιμές PL. Μέσω αυτής της εργασίας καταλαβαίνουμε επίσης την σύνδεση που υπάρχει μεταξύ των ισοτιμίων του FOREX και πως η εκπαίδευση σε όλες τις ισοτιμίες θα αποφέρει καλύτερα αποτελέσματα ακόμα και αν αξιολογείται μόνο σε μια.

Οι Passalis, Tefas, Anastasios, Kannainen, Gabbouj και Iosifidis [PTK<sup>+</sup>20] χρησιμοποιούν ένα διαφορετικό τρόπο κανονικοποίησης δεδομένων χρονοσειρών ο οποίος

χρησιμοποιεί βαθιά μάθηση για την επιλογή του είδους κανονικοποίησης για κάθε δείκτη ώστε να επιφέρει τα μέγιστα δυνατά αποτελέσματα.. Τα συμπεράσματα τους απέδειξαν ότι με την χρήση του μοντέλου οι προβλέψεις είναι πιο ακριβείς σε σχέση με τις κανονικές μεθόδους κανονικοποίησης.

Στη συνέχεια θα αναφερθούμε σε εργασίες οι οποίες συνδύασαν Νευρωνικά Δίκτυα με Ανάλυση Συναισθήματος.

Οι Arman Khadjeh, Nassiroussi Saeed, Aghabozorgi, Teh Ying, Wah David και Chek Ling Ngo στην εργασία τους [KAYN15] συνδύασαν την συναισθηματική ανάλυση των επικεφαλίδων από άρθρα ειδησεογραφικών εφημερίδων για την πρόβλεψη τιμών FOREX. Τα αποτελέσματα τους επέφεραν υψηλές αποδόσεις ακρίβειας. Επίσης κατέληξαν στο συμπέρασμα ότι υπάρχει άμεση αλληλεξάρτηση μεταξύ των επικεφαλίδων και των τιμών των ισοτιμιών οι οποίες μπορούν να βρεθούν με την χρήση συναισθηματικής ανάλυσης.

Οι Sushree Dasa, Ranjan Kumar Beheraa, Mukesh kumarb, Santanu Kumar Rath [DBKR18] χρησιμοποίησαν δεδομένα από το Twitter για να αναλύσουν το συναίσθημα των δημοσιεύσεων που αφορούν τον δείκτη μιας μετοχής. Στη συνέχεια με χρήση νευρωνικών δικτύων LSTM συνδυάζουν τις προηγούμενες τιμές ενός χρονικού διαστήματος μαζί με τον μέσο όρο συναισθήματος για να προβλέψουν την επόμενη τιμή της. Αυτή η τεχνική χρησιμοποιείται και στην συνέχεια στην εργασία για τα πειράματα που αφοράνε πρόβλεψη τιμής (regression) σε συνδυασμό με ανάλυση συναισθήματος.

# 2

## Βαθιά Μάθηση

---

### 2.1 Ορισμός

Η βαθιά μάθηση αποτελεί υποκατηγορία της Μηχανικής μάθησης. Ως μηχανική μάθηση ονομάζουμε την διαδικασία με την οποία ο άνθρωπος προσπαθεί να αποτυπώσει την συμπεριφορά του και το πως παρακολουθεί και αλληλεπιδρά με το περιβάλλον σε ένα μοντέλο ώστε να είναι ικανό να τις επαναλάβει. Βασικές έννοιες στην μηχανική μάθηση είναι η πρόσκτηση δεδομένων (data) στο σύστημα από κάποιο περιβάλλον καθώς και η επανάληψη με την οποία το σύστημα προσπαθεί να αποσπάσει τις μέγιστες πληροφορίες από τα εν λόγω δεδομένα. Σκοπός είναι η συνεχής βελτίωση της ικανότητας της υπολογιστικής μηχανής για να εκτελέσει διάφορες λειτουργίες (tasks).

Βασικές χρήσης της μηχανικής μάθησης είναι :

- Η πρόβλεψη δεδομένων (regression)
- Ο διαχωρισμός σε ομάδες (clustering)
- Η ανάπτυξη στρατηγικής (agent)

Για τις παραπάνω ενέργειες η μηχανική μάθηση χωρίζεται σε τρείς κατηγορίες :

1. **Μάθηση με επίβλεψη (supervised learning).** Στην μάθηση με επίβλεψη κάθε στοιχείο των δεδομένων αντιστοιχεί σε μια άλλη τιμή. Σκοπός του αλγορίθμου είναι η αντιστοίχιση νέων δεδομένων πάνω στα οποία δεν έχει εκπαιδεύτει με τις τιμές του βασιζόμενος στα προϋπάρχοντα δεδομένα. Η μάθηση με επίβλεψη μπορεί να αναφέρεται είτε σε κατηγοριοποίηση δεδομένων (classification) σε μια συγκεκριμένη τιμή (π.χ. έγκριση ή απόρριψη αιτήματος δανείου, αναγνώριση προσώπων) είτε σε πρόβλεψη τιμής (regression) (π.χ. τιμή μετοχής, μελλοντικά κέρδη επιχείρησης).

2. **Μάθηση χωρίς επίβλεψη (unsupervised learning).** Στην μάθηση χωρίς επίβλεψη σκοπός του αλγορίθμου είναι να διαχωρίσει τα δεδομένα που του δόθηκαν σε ομάδες ανάλογα με τα χαρακτηριστικά τους. Για να επιτευχθεί αυτό το υπολογιστικό σύστημα καλείται να αποκαλύψει διάφορες συσχετίσεις και ομοιότητες που υπάρχουν ανάμεσα στα δεδομένα. Κάποια παραδείγματα μάθησης χωρίς επίβλεψη είναι ο διαχωρισμός φυσικών προσώπων σε υποψήφιο αγοραστικό κοινό (target group) για κάποιο προϊόν και η προβολή παρόμοιων προτιμήσεων για κάποιο φυσικό πρόσωπο (π.χ. ιστοσελίδες για προβολή ταινιών, μουσικές πλατφόρμες).
3. **Ενισχυτική μάθηση (reinforcement learning).** Σε αυτό το είδος μηχανικής μάθησης το σύστημα προσπαθεί μέσω συνεχούς αλληλεπίδρασης με το περιβάλλον του προβλήματος να φέρει εις πέρας κάποιον σκόπο. Αυτό επιτυγχάνεται με ένα σύστημα επιβράβευσης για τις δράσεις (actions) που οδηγούν στον σκοπό αυτό και τιμωρίας για τις αντίθετες. Αυτού του είδους μάθησης βρίσκεται πάντοι γύρω μας στον σύγχρονο κόσμο. Από την επίτευξη της νίκης σε επιτραπέζια παιχνίδια μέχρι πιο σύνθετα προβλήματα όπως η αυτόματη οδήγηση και η αγοραπωλησία μετοχών στο χρηματιστήριο.

## 2.2 Metrics

Η επίδοση ενός υπολογιστικού προβλήματος σε ένα πρόβλημα μηχανικής μάθησης πρέπει να είναι δυνατή να αξιολογηθεί και να μετρηθεί έτσι ώστε να επιτευχθεί η μεγιστοποίηση της απόδοσης. Ο πιο συχνός δείκτης μέτρησης για να αξιολογηθούν οι αποφάσεις του συστήματος είναι η ακρίβεια (accuracy), δηλαδή το ποσοστό των σωστών προβλέψεων. Σε πολλές περιπτώσεις είναι απαραίτητο να υπάρχουν περισσότερες μεταβλητές μετρήσεις της απόδοσης (metrics) για τα αποτελέσματα του συστήματος. Οι μετρικές αυτές είναι η ακρίβεια (precision), η ανάκληση (recall) και το F1-score.

Αρχικά θα αναλύσουμε τις περιπτώσεις του αποτελέσματος της αξιολόγησης για ένα δείγμα.

- Αληθές Θετικό (**True Positive**) ορθή ταξινόμηση ενός δείγματος ως θετικό.
- Ψευδές Θετικό (**False Positive**) λανθασμένη ταξινόμηση ενός δείγματος ως θετικό.
- Αληθές Αρνητικό (**True Negative**) ορθή ταξινόμηση ενός δείγματος ως αρνητικό.
- Ψευδές Αρνητικό (**False Negative**) λανθασμένη ταξινόμηση ενός δείγματος ως αρνητικό.

Οι συναρτήσεις υπολογισμού των μετρικών της αποδόσης του συστήματος :

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

$$precision = \frac{TP}{TP + FP} \quad (2.2)$$

$$recall = \frac{TP}{TP + FN} \quad (2.3)$$

$$F_1 = \frac{2 * (recall * precision)}{recall + precision} \quad (2.4)$$

$$F_\beta = (1 + \beta^2) \frac{recall * precision}{(\beta^2 * recall) + precision} \quad (2.5)$$

Όπου TP, TN, FP και FN ο αριθμός των στοιχείων που είναι Αληθές Θετικά, Ψευδές Θετικά, Αληθές Αρνητικά και Ψευδές Αρνητικά αντίστοιχα.

Η ακρίβεια (2.1) εκφράζει το ποσοστό των ορθά αξιολογημένων δειγμάτων ως προς το πλήθος τους. Η ακρίβεια (2.2) εκφράζει το ποσοστό των ορθά αξιολογημένων θετικών δειγμάτων ως προς το πλήθος των δειγμάτων που αξιολογήθηκαν ως θετικά. Η ανάληση (2.3) εκφράζει το ποσοστό των ορθά αξιολογημένων θετικών δειγμάτων ως προς το πλήθος των δειγμάτων που είναι θετικά. Το F1-score (2.4) εκφράζει το αρμονικό μέσο (harmonic mean) μεταξύ της ακρίβειας και της ανάλησης. Επιπλέον μια παραλλαγή του F1-score είναι το F $\beta$ -score (2.5) το οποίο χρησιμοποιεί βάρη για να δώσει έμφαση είτε στην ακρίβεια είτε στην ανάληση.



# 3

## Νευρωνικά Δίκτυα

---

### 3.1 Perceptron

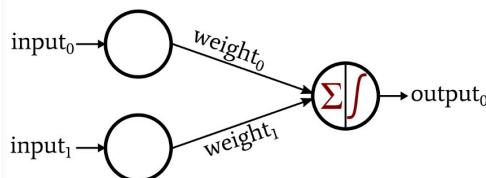
Το Perceptron αποτελεί την πιο απλή έκδοση Νευρωνικού Δικτύου και την βάση πάνω στην οποία εξελίχθηκαν τα επόμενα δίκτυα. Δημιουργήθηκε το 1957 από τον Frank Rosenblatt [Ros58]. Η χρήση του είναι ως δυαδικός ταξινομητής και προσπαθεί να βρεί το επίπεδο που χωρίζει τα σημεία κάθε κλάσης. Οι είσοδοι αποτελούν αριθμητικές τιμές και με την χρήση βαρών υπολογίζει το σταθμισμένο άθροισμα μεταξύ των εισόδων (3.1). Έπειτα με την χρήση βηματικής συνάρτησης (3.2) υπολογίζει το τελικό αποτέλεσμα από το άθροισμα (3.3).

$$z = w_1 * x_1 + w_2 * x_2 + \dots + w_n * x_n = w^T * x \quad (3.1)$$

$$step(z) = \begin{cases} 1, z \geq 0 \\ 0, z < 0 \end{cases} \quad (3.2)$$

$$c(x) = step(w^T * x) \quad (3.3)$$

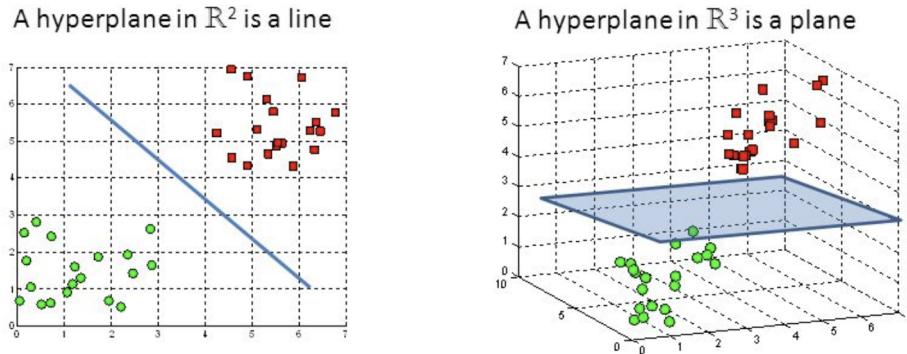
Σε πολλές περιπτώσεις του Perceptron χρησιμοποιείται και κατώφλι (bias) το οποίο είναι ένας νευρώνας με τιμή 1.



**Σχήμα 3.1:** Η αναπαράσταση ενός νευρώνα 2 εισόδων  
**Πηγή :** [Kei19]

Η έξοδος του Perceptron είναι  $[0,1]$  δηλαδή οι 2 τιμές που αντιπροσωπεύουν τις δύο κλάσεις που έχουμε στο πρόβλημα.

$$\sum w_i * w_i + b = 0 \quad (3.4)$$



**Σχήμα 3.2:** Αναπάσταση των Hyperplane για 2 και 3 διαστάσεις  
Πηγή : [Gan18]

Τυποθέτωντας ότι τα σημεία των δύο κλάσεων είναι γραμμικά διαχωρίσιμα τότε είναι βέβαιο ότι ο Perceptron θα καταφέρει να κάνει ορθή διαχώριση.

Η εκπαίδευση του Perceptron γίνεται με τον κανόνα του Hebb , σύμφωνα με τον οποίο η σύνδεση μεταξύ δύο νευρώνων επηρεάζεται από την ομοιότητα των εξόδων. Για τον Perceptron σε περίπτωση λάθος αποτελέσματος έχουμε ενίσχυση και μείωση των κατάλληλων βαρών ώστε να οδηγηθούμε στην σωστή πρόβλεψη.

$$w_{ij}(k+1) = w_{ij}(k) + \gamma(y_i - \hat{y}_i)x_i \quad (3.5)$$

Ως  $w_{ij}$  ορίζεται το βάρος της σύνδεσης μεταξύ του νευρώνα εισόδου  $j$  και του νευρώνα εξόδου  $i$ .  $x_i$  είναι η είσοδος του νευρώνα  $i$ ,  $\hat{y}_j$  είναι η έξοδος του νευρώνα  $j$ , ενώ το  $y_j$  αναπαριστά τον αντίστοιχο στόχο του νευρώνα  $j$ . Το  $\gamma$  αναφέρεται στο βήμα της εκπαίδευσης, ενώ το  $k$  αναπαριστά το βήμα στον χρόνο.

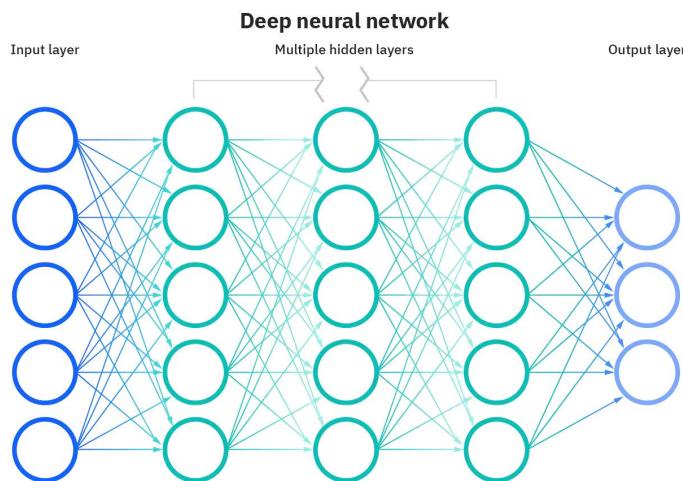
## 3.2 Εισαγωγή στα Νευρωνικά Δίκτυα.

Οπως αναφέραμε στο 3.1 ένας Perceptron είναι δυνατόν να διαχωρίσει μόνο μεταξύ δύο κλάσεων και υπό την προυπόθεση ότι τα δεδομένα τους είναι γραμμικά διαχωρίσιμα. Για τα υπόλοιπα προβλήματα έχουν δημιουργήσει Νευρωνικά Δίκτυα ώστε να παρέχουν λύση για επίλυση πιο περίπλοκων προβλημάτων. Τα νευρωνικά Δίκτυα έχουν επίσης την δυνατότητα να δημιουργήσουν γνώση , να εξορύξουν χρήσιμες πληροφορίες από τα δεδομένα και να λειτουργήσουν με μνήμη.

## 3.3 Αρχιτεκτονική

Την πάροχουν πολλά είδη αρχιτεκτονικών στα Νευρωνικά Δίκτυα. Μια δημοφιλής εκδοχή είναι τα νευρωνικά δίκτυα πολυεπίπεδων Perceptrons (MLP Neural Network) . Με τον όρο

Perceptron αναφερόμαστε σε κάθε νευρώνα του δικτύου. Το νευρωνικό δίκτυο χωρίζει τους νευρώνες σε επίπεδα που ο καθένας από αυτούς συνδέεται πλήρως με τους νευρώνες του επόμενου επιπέδου. Ως πρώτο επίπεδο είναι η είσοδος του δίκτυου ( $x$ ) , ως τελευταίου η έξοδος του ( $y$ ) και τα ενδιάμεσα επίπεδα ονομάζονται χρυφά επίπεδα. Σε διάφορες παραλλαγές που υπάρχουν τα χρυφά επίπεδα έχουν διαφορετικές συνδέσεις μεταξύ τους. Χαρακτηριστικό παράδειγμα μη-πλήρους συνδεδεμένου νευρωνικού δικτύου είναι τα Επαναλαμβανόμενα Νευρωνικά Δίκτυα (RNN) που ενδείκνυνται για προβλήματα όπως η ανάλυση χρονοσειρών.



**Σχήμα 3.3:** Αρχιτεκτονική ενός πλήρους συνδεδεμένου νευρωνικού δικτύου με 3 κρυφά επίπεδα  
Πηγή : [Edu20]

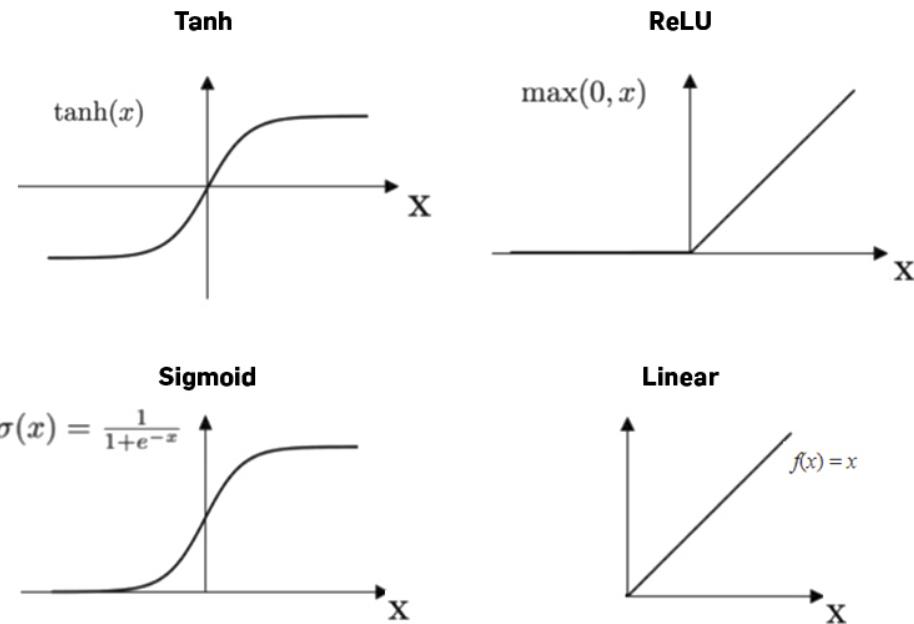
## 3.4 Συναρτήσεις Ενεργοποίησης

Τα Νευρωνικά Δίκτυα κάνουν χρήση διάφορων μη-γραμμικών συναρτήσεων ενεργοποίησης ώστε να αυξηθεί η ισχύ τους. Μερικές από αυτές είναι οι ReLU , PReLU , sigmoid , tanh και softmax.

Η συνάρτηση softmax  $\sigma : \mathbb{R}^K \rightarrow [0, 1]^K$

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \text{ for } i = 1, \dots, K \text{ and } z = (z_1, \dots, z_K) \in \mathbb{R}^K \quad (3.6)$$

Ως ζ έχουμε την λίστα με  $K$  πραγματικούς αριθμούς που είναι η είσοδος μας.



**Σχήμα 3.4:** Συναρτήσεις Ενεργοποίησης  
Πηγή : [FHT<sup>+</sup>19]

Συναρτήσεις Ενεργοποίησης		
Όνομα	Τύπος συναρτησης	Εύρος τιμών
tanh	$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	(-1,1)
ReLU	$\text{Relu}(x) = \max(0, x)$	[0,∞)
Σιγμοειδής	$\sigma(x) = \frac{1}{1+e^{-x}}$	(0,1)
Γραμμική	$f(x) = x$	(-∞, ∞)
PReLU	$\text{PReLU}(x) = \max(0, x) + \alpha * \min(0, x)$	(-∞, ∞)

**Πίνακας 3.1:** Πίνακας Συναρτήσεων Ενεργοποίησης.

### 3.5 Συνάρτηση Κόστους

Με τον όρο συνάρτηση κόστους (loss function) αναφερόμαστε στον τρόπο υπολογισμού του σφάλματος των προβλέψεων ενός νευρωνικού δικτύου σε σχέση με τα επιθυμητά αποτελέσματα. Το αποτέλεσμα τέτοιων συναρτήσεων είναι ένας ακέραιος αριθμός έτσι ώστε να μπορεί να γίνει μετρήσιμη σύγχριση αποτελεσμάτων μεταξύ διαφορετικών νευρωνικών δικτύων. Συνήθως ο συμβολισμός των συναρτήσεων κόστους είναι  $J$ .

Για κάθε είδος προβλήματος χρησιμοποιείται διαφορετική συνάρτηση κόστους.

Για προβλήματα πρόβλεψης τιμών χρησιμοποιείται το μέσο σφάλμα τετράγωνο (Mean squared error - MSE).

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 \quad (3.7)$$

Όπου  $N$  το πλήθος των δειγμάτων,  $Y_i$  η τιμή του δείγματος  $i$  και  $\hat{Y}_i$  η προβλεπόμενη τιμή του δείγματος  $i$ .

Για προβλήματα κατηγοροποίησης χρησιμοποιείται η cross entropy ή αλλίως αναφέρεται ως λογαριθμικό κόστος (logarithmic loss).

$$LogLoss = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij}) \quad (3.8)$$

Όπου  $N$  το πλήθος των δειγμάτων,  $M$  το πλήθος των κλάσεων,  $y_{ij}$  1 για το αν το δείγμα  $y_i$  ανήκει στην κλάση  $j$  και 0 αν δεν ανήκει και  $p_{ij}$  η πιθανότητα για να ανήκει στη κλάση αυτή.

## 3.6 Gradient Descent

Ως Gradient Descent ονομάζουμε την διαδικασία την οποία χρησιμοποιούμε στην εκπαίδευση των νευρώνων. Ως στόχο έχει την ελαχιστοποίηση της συνάρτησης κόστους με συμβολισμό  $J$ . Η επαναληπτική κίνηση του αλγορίθμου προς το ελάχιστο δηλαδή το σημείο με αρνητική παράγωγο. Στο σχήμα 3.5 βλέπουμε την συνάρτηση με κάθε βήμα να βρίσκεται όλο και πιο κοντά στο ελάχιστο.

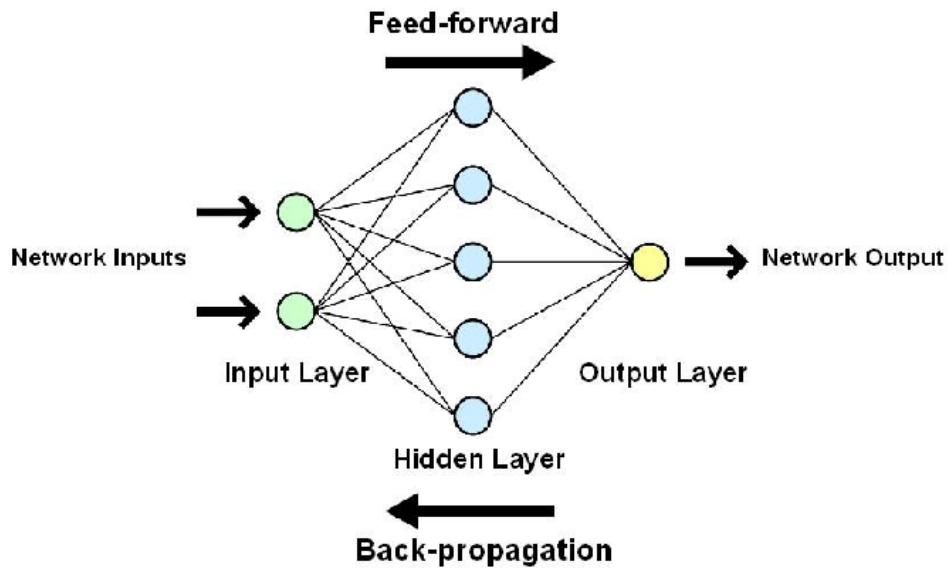
$$\Delta w_j = -\eta \frac{\partial J}{\partial w_j} \quad (3.9)$$

$$= -\eta \sum_i (target^{(i)} - output^{(i)}) (-x_j^{(i)}) \quad (3.10)$$

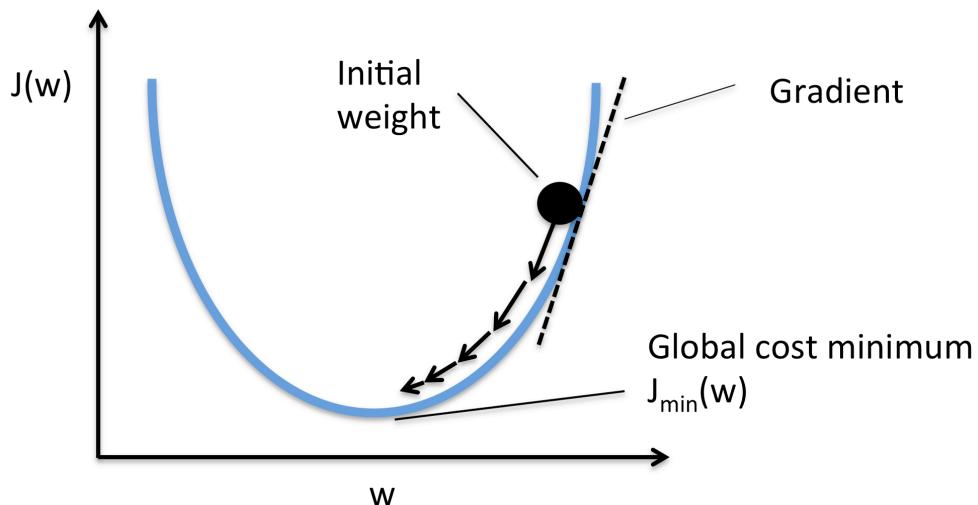
$$= \eta \sum_i (target^{(i)} - output^{(i)}) x_j^{(i)} \quad (3.11)$$

Όπου η το ρυθμός μάθησης (learning rate) και  $J(w)$  το κόστος (σφάλμα) για τα εν λόγω βάρη. Μπορούμε να αναπαραστήσουμε τη Gradient Descent ως μια καμπύλη στην οποία ξεκινάμε από ένα αρχικό σημείο και με κάθε βήμα βρισκόμαστε όλο και πιο κοντά στο ολικό ελάχιστο, όπως φαίνεται στο σχήμα 3.6.

Το βήμα εκπαίδευσης αντιπροσωπεύει τον ρυθμό με τον οποίο θέλουμε να κινείται ο αλγόριθμος για την εύρεση του ελάχιστου. Η χρήση μεγάλου βήματος εκπαίδευσης μπορεί να οδηγήσει στην μη σύγκλιση προς το επιθυμητό αποτέλεσμα ενώ η χρήση μικρού βήματος εκπαίδευσης να οδηγήσει σε πολύ αργή σύγκλιση που δεν είναι επιθυμητή.



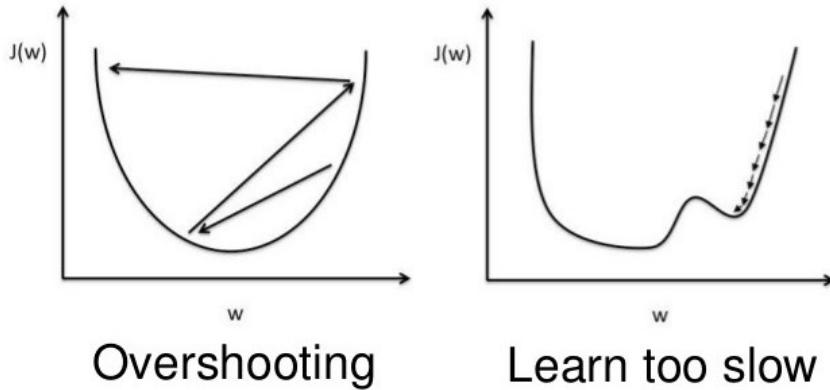
**Σχήμα 3.5:** Η εκπαίδευση ενός νευρωνικού δικτύου  
**Πηγή :** [Pai20]



**Σχήμα 3.6:** Η αναπαράσταση της συνάρτησης  $J$  προς  $w$  για την μείωση των κόστων  
**Πηγή :** [K19]

Όπως φαίνεται στο σχήμα 3.7 ένας μεγάλος ρυθμός μάθησης θα οδηγήσει στην μη σύγκλιση προς το ολικό ελάχιστο ενώ ένας μικρός ρυθμός μάθησης θα έχει πολύ αργή σύγκλιση και μπορεί να παγιδευτεί σε τοπικά ελάχιστα.

Εκτός από τον αλγόριθμο Gradient Descent υπάρχουν και άλλοι αλγόριθμοι βελτιστοποίησης.



**Σχήμα 3.7:** Αναπαράσταση της διαδικασίας Gradient Descent με διαφορετικές τιμές στο learning rate

Πηγή :[\[Kul18\]](#)

## 3.7 Εκπαίδευση Νευρωνικών Δικτύων

Ο σκοπός της δημιουργίας ενός Νευρωνικού Δικτύου είναι να είναι η ικανότητα δημιουργίας κάποιων αποτελέσματων. Τα αποτελέσματα αυτά θα παραχθούν από το Νευρωνικό Δίκτυο με βάση τις τιμές των βαρών των νευρώνων του. Η κατάλληλη παραμετροποίηση των βαρών γίνεται με την διαδικασία που ονομάζεται Εκπαίδευση Νευρωνικών Δικτύων (Neural Network Training). Στην διαδικασία αυτή το δίκτυο εκπαιδεύεται σε κάποια συγκεκριμένα δεδομένα (training set). Το πέρασμα των δεδομένων αυτών από το δίκτυο χωρίζεται σε δύο διαδικασίες, την τροφοδότηση και την οπισθοδόμηση.

### 3.7.1 Τροφοδότηση (Feedforward-propagation)

Στην διαδικασία της τροφοδότησης τα δεδομένα περνάνε από το δίκτυο και γίνεται ο υπολογισμός της εξόδου κάθε νευρώνα έως να υπολογισθεί το τελικό αποτέλεσμα. Τα αποτελέσματα κάθε επίπεδου περνάνε στο επόμενο.

Ο υπολογισμός αυτός γίνεται για κάθε επίπεδο με το άθροισμα των γινομένων από τα βάρη που ένωνουν το επίπεδο αυτό με το προηγούμενο με τις τιμές των νευρώνων του προηγούμενου επιπέδου.

$$a_i(l+1) = f\left(\sum_{j=1}^l (w_{ij} * a_j(l)) + w_{j0}\right) \quad (3.12)$$

Ως  $a_i(l)$  έχουμε την τιμή του νευρώνα  $i$  στο επίπεδο  $l$ . Ως  $l$  το πλήθος των νευρώνων στο επίπεδο  $l$  και ως  $w_{ij}$  το βάρος μεταξύ των νευρώνων  $i$  του επιπέδου  $l-1$  και  $j$  του επιπέδου  $l$ . Ως  $f$  συμβολίζουμε την επιλεγμένη συνάρτηση ενεργοποίησης (2.4).

Η συνολική διαδικασία της τροφοδότησης του Νευρωνικού Δικτύου δίνει την έξοδο  $y$

για μια είσοδο  $\vec{x}$  η οποία θα χρησιμοποιηθεί στην διαδικασία της οπισθοδρόμισης.

$$y = N(\vec{x}) \quad (3.13)$$

### 3.7.2 Οπισθοδρόμηση (Back-propagation)

Το νευρωνικό δίκτυο ανανέωνει τα βάρη του στην διαδικασία της εκπαίδευσης με την χρήση του Gradient Descent 3.6. Σε πολυεπίπεδα νευρωνικά δίκτυα υπολογίζουμε τις παραγώγους της συνάρτησης κόστους για κάθε βάρος αρχίζοντας από το τελευταίο επίπεδο όπως φαίνεται και στην εξίσωση 3.15. Η μέθοδος αυτή ονομάζεται Back-propagation. Η διαδικασία Back-propagation κάνει χρήση του κανόνα δέλτα (delta rule) που είναι η ανανέωση των βαρών σε ένα μονοεπίπεδο νευρωνικό δίκτυο.

Όπου  $J$  η συνάρτηση κόστους,  $z^l$  η είσοδος του επιπέδου  $l$  και  $a^l$  η έξοδος του.

$$\frac{dJ}{da^L} \cdot \frac{da^L}{dz^L} \cdot \frac{dz^L}{da^{L-1}} \cdot \frac{da^{L-1}}{dz^{L-1}} \cdot \frac{dz^{L-1}}{da^{L-2}} \cdots \frac{da^1}{dz^1} \cdot \frac{\partial z^1}{\partial x} \quad (3.14)$$

Ως  $\delta^l$  θέτουμε το σφάλμα στο επιπέδο  $l$ . Η τιμή  $\delta^l$  είναι μια λίστα με σφάλματα μεγέθους ίση με τον αριθμό των νευρώνων στο επίπεδο  $l$ .  $f^l$  είναι η συνάρτηση ενεργοποίησης στο επίπεδο  $l$ .

$$\delta^l := (f^l)' \cdot (W^{l+1})^T \cdots (W^{L-1})^T \cdot (f^{l-1})' \cdot (W^L)^T \cdot (f^L)' \cdot \nabla_{a^L} J \quad (3.15)$$

Η τιμή  $\delta^l$  μπορεί να υπολογισθεί και αναδρομικά με την σχέση 3.16.

$$\delta^{l-1} := (f^l)' \cdot (W^l)^T \cdot \delta^l \quad (3.16)$$

Άρα μπορούμε να υπολογίζουμε τις βαθμίδες των βαρών κάνοντας μερικούς υπολογισμούς πινάνω για κάθε βήμα. Αυτή είναι η διαδικασία Back-Propagation.

$$\begin{aligned} \delta^1 &= (f^1)' \cdot (W^2)^T \cdot (f^2)' \cdots (W^{L-1})^T \cdot (f^{L-1})' \cdot (W^L)^T \cdot (f^L)' \cdot \nabla_{a^L} J \\ \delta^2 &= (f^2)' \cdots (W^{L-1})^T \cdot (f^{L-1})' \cdot (W^L)^T \cdot (f^L)' \cdot \nabla_{a^L} J \\ &\vdots \\ \delta^{L-1} &= (f^{L-1})' \cdot (W^L)^T \cdot (f^L)' \cdot \nabla_{a^L} J \\ \delta^L &= (f^L)' \cdot \nabla_{a^L} J \end{aligned} \quad (3.17)$$

Η διαδικασία Back-propagation χωρίζεται σε τέσσερα βήματα υποθέτωντας ότι έχουμε τον ρυθμό μάθησης (learning rate)  $\alpha$  και έχουμε αρχικοποιήσει τυχαία τα βάρη  $w_{i,j}^k$

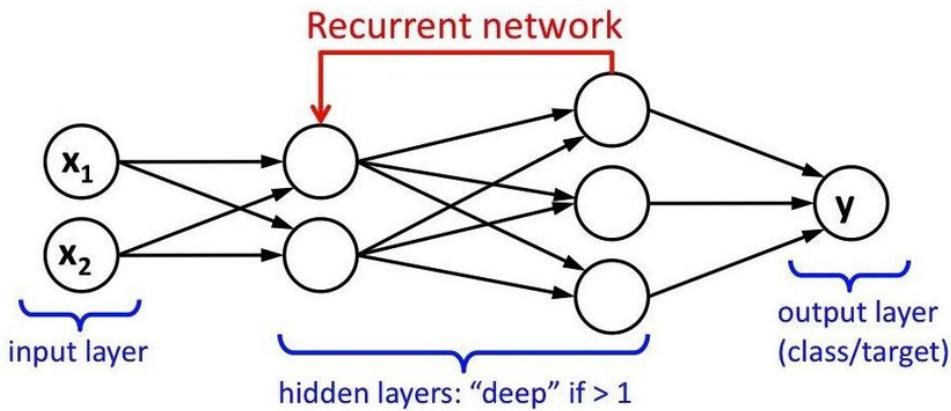
1. Υπολογισμός της φάσης του εμπρόσθιου περάσματος (forward phase) για κάθε ζευγάρι  $(\vec{x}_d, y_d)$  υπολογίζουμε και αποθηκεύουμε τις τιμές  $\hat{y}_d$ ,  $a_j^k$ ,  $o_j^k$  για κάθε νευρώνα

ξεκινώντας από το επίπεδο εισόδου μέχρι τον επιπέδο εξόδου.

2. Υπολογισμός της φάσης οπισθό περάσματος (backward phase) για κάθε ζευγάρι  $(\vec{x}_d, y_d)$  και υπολογίζοντας τις τιμές  $\frac{\partial E_d}{\partial w_{i,j}^k}$  για καθε βαρος  $w_{i,j}^k$  που ενώνει τον νευρώνα i στο επίπεδο k-1 με τον νευρώνα h στο επίπεδο k, ξεκινώντας από το επίπεδο εξόδου (output layer) προς το επίπεδο εισόδου (input layer).

- α) Υπολογισμός του σφάλματος για το τελευταίο επίπεδο  $\delta_i^m$ .
- β) Υπολογισμός των υπολοίπων σφαλμάτων για τα χρυφά επίπεδα  $\delta_j^k$  ξεκινώντας από το τελευταίο χρυφό επίπεδο  $k = m - 1$ .
- γ) Υπολογισμός του ατομικού σφάλματος  $E_d$  σε σχέση με το  $w_{i,j}$ .
- 3) Υπολογισμός των ατομικών βαθμίδων (individual gradient) για κάθε ζευγάρι είσοδος-εξόδος  $\frac{\partial E_d}{\partial w_{i,j}^k}$  για να βρούμε την συνολική βαθμίδα (total gradient)  $\frac{\partial E_d(X, \theta)}{\partial w_{i,j}^k}$  για όλο το σύνολο είσοδων-εξόδων  $X = (\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_n, y_n)$
- 4) Ενημέρωση των βαρών ανάλογα με τον βαθμό μάθησης α και την συνολική βαθμίδα (total gradient)  $\frac{\partial E_d(X, \theta)}{\partial w_{i,j}^k}$ .

Η υλοποίηση όλων αυτών των βημάτων μια φορά, δηλαδή το πέρασμα όλων των εισόδων στον αλγόριθμο ονομάζεται εποχή (epoch). Ο όρος αυτός αναφέρεται στις συνολικές φορές που θα περάσουν τα δεδομένα από τον αλγόριθμο και ορίζεται είτε από τον προγραμματιστή είτε έως το σφάλμα να γίνει μικρότερο από μια τιμή που έχει οριστεί.



**Σχήμα 3.8:** Η αρχιτεκτονική ένος Αναδρομικού Νευρωνικού Δικτύου  
**Πηγή :** [HS97]

### 3.8 Αναδρομικά Νευρωνικά Δίκτυα - RNN

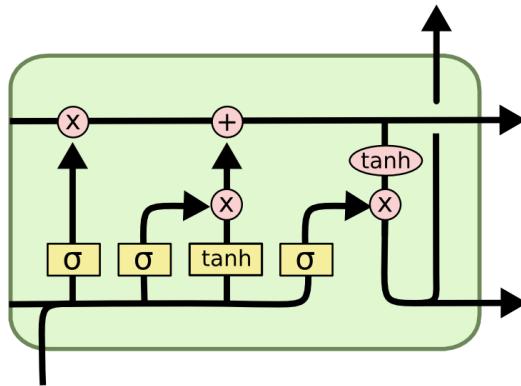
Η γενικότερη κατηγορία στην οποία ανήκουν τα Νευρωνικά Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης είναι τα Αναδρομικά Νευρωνικά Δίκτυα. Τα Αναδρομικά Νευρωνικά Δίκτυα χρησιμοποιούνται κυρίως σε προβλήματα που τα δεδομένα έχουν μια συγκεκριμένη αλληλουχία. Ως βασικό χαρακτηριστικό των RNN είναι η ανάδραση μεταξύ των νευρώνων διαφορετικών επιπέδων. Κάθε νευρώνας στέλνει την έξοδο του και σε νευρώνες του ίδιου ή του προηγούμενου επιπέδου αντί μόνο για τον επόμενο όπως γίνεται σε ένα κλασικό Multilayer Perceptron. Ουσιαστικά αυτό δημιουργεί ένα είδος βραχυπρόθεσμης μνήμης έτσι ώστε μελλοντικές αλληλουχίες που εξετάζονται από το δίκτυο να συσχετίζονται με προηγούμενες. Έτσι η είσοδος μια χρονικής στιγμής μπορεί να ανατροφοδοτήσει και το αποτέλεσμα της επόμενης χρονικής στιγμής. Με αυτή την τεχνική το νευρωνικό δίκτυο μπορεί να έχει μια έννοια του χρόνου και να επηρεάζεται από τα προηγούμενα αποτελέσματα.

Τα Νευρωνικά Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης χωρίζονται και αυτά με τη σειρά τους σε υποδεέστερες κατηγορίες με κύριες τα Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (Long Short-Term Memory - LSTM) και τα Δίκτυα Περιφραγμένης Επαναλαμβανόμενης Μονάδας (Gated Recurrent Unit - GRU).

### 3.9 Δικτύα Μακράς Βραχυπρόθεσμης Μνήμης - LSTM

Τα Νευρωνικά Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης είναι ένα είδος RNN το οποίο παρουσιάστηκε από τους Sepp Hochreiter και Jurgen Schmidhuber το 1997 [?]. Τα LSTM προσπαθούν να δημιουργήσουν συσχετίσεις μεταξύ απομακρυσμένων δειγμάτων (long term). Ένα LSTM ουσιαστικά αποτελείται από 4 επίπεδα, βασικό στοιχείο είναι το επίπεδο μνήμης (memory block). Τρείς πύλες αλληλοεπιδρούν με το επίπεδο μνήμης, η πύλη εισόδου

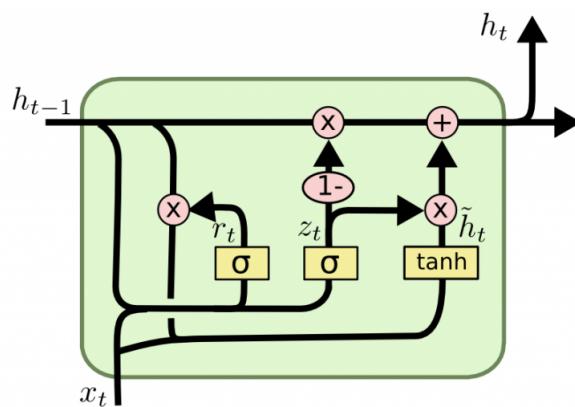
(input gate), η πύλη εξόδου (output gate) και η πύλη διαγραφής (forget gate). Μέσω της εκπαίδευσης τα LSTM μπορούν να κρατήσουν την ποιοτική πληροφορία και να την διατηρήσουν για μεγάλο χρονικό διάστημα. Χρησιμοποιούνται κυρίως σε προβλήματα πρόβλεψης χρονοσειρών (time series).



**Σχήμα 3.9:** Τα περιεχόμενα ένος νευρώνα LSTM  
Πηγή : [TPC<sup>+</sup>21]

## 3.10 Δίκτυα Περιφραγμένης Επαναλαμβανόμενης Μονάδας - GRU

Τα GRU είναι ένα ακόμα είδος RNN τα οποία έχουν λειτουργία παρόμοια με αυτή των LSTM. Η διαφορά τους είναι ότι έχουν λιγότερες παραμέτρους από ένα LSTM καθώς και η απουσία της πύλης εξόδου. Οι δύο πύλες που αλληλοεπιδρούν με το επόπεδο μνήμης είναι η πύλη επαναφοράς (reset gate) και η πύλη ανανέωσης (update gate).



**Σχήμα 3.10:** Τα περιεχόμενα ένος νευρώνα GRU  
Πηγή : [TPC<sup>+</sup>21]



# 4

## Χρηματοοικονομική Αγορά

---

### 4.1 FOREX

Με τον όρο Χρηματοοικονομική Αγορά ονομάζουμε το μέρος και την διαδικασία με την οποία πραγματοποιείται η ανταλλαγή / αγοραπωλησία (trading) διαφόρων περουσιακών στοιχείων. Το πιο δημοφιλές παράδειγμα Χρηματοοικονομική Αγορά είναι το Χρηματιστήριο (Stock Market) στο οποίο πραγματοποιείται trading μετοχών εταιριών. Η Χρηματοοικονομική Αγορά που πραγματοποιούνται περισσότερες συναλαγές είναι οι Συναλλαγές Ισοτιμιών Νομισμάτων (Foreign Exchange, FOREX) δηλαδή η αγορά που έχει δημιουργηθεί για την συναλλαγή των διαφορετικών νομισμάτων που υπάρχουν ανά τον κόσμο.

Μια ισοτιμία δύο νομισμάτων ονομάζεται η τιμή του ενός νομίσματος ως προς το άλλο την συγκεκριμένη χρονική στιγμή και συμβολίζεται ως X/Y, δηλαδή η ποσότητα του νομίσματος Y που μπορεί να αγοραστεί με μια μονάδα του νομίσματος X. Παράδειγμα η ισοτιμίας Ευρώ / Λίρα Στερλίνα Βρετανίας είναι 0.86 δηλαδή μπορεί να αγοραστεί 0.86 Λίρα Στερλίνα Βρετανίας με 1 Ευρώ.

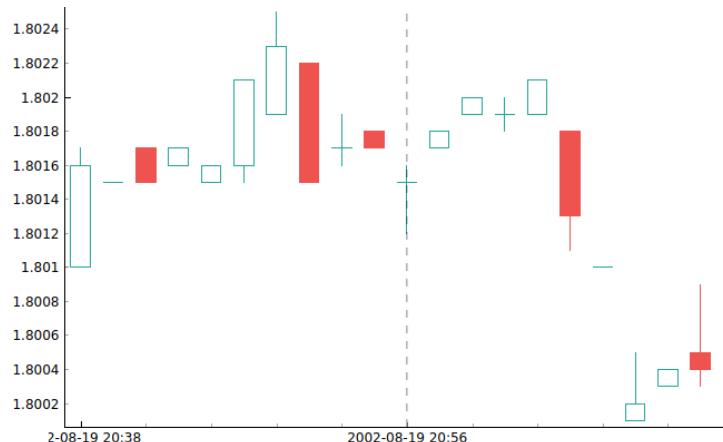
Ισοτιμία	Τιμή
EUR/USD	1.2125
EUR/JPY	132.88
EUR/BGN	1.9558
EUR/CZK	25.335
EUR/GBP	0.85710
EUR/DKK	7.4361
EUR/CHF	1.0883
EUR/TRY	10.1969
EUR/AUD	1.5687

Πίνακας 4.1: Ισοτιμίες των Ενρω

Οι συναλλαγές αυτές συμβαίνουν κατά κύριο μέρος μέσω του διαδικτύου που σημαίνει ότι μπορούν να αναληφθούν και να εξεταστούν για την δημιουργία χρήσιμων συμπερασμάτων. Μέσω των καθημερινών δεδομένων μπορεί να κατανοηθεί η κίνηση των αγορών. Επιπλέον με την άμεση αλληλεπίδραση των ατόμων που συμμετέχουν στις συναλλαγές αυτές με τα μέσα κοινωνικά δικτύωσης δημιουργούνται νέα δεδομένα για την κατανόηση των τάσεων των αγορών.

#### 4.1.1 Candlesticks

Τα δεδομένα των ισοτιμιών μπορούν να αναπαρασταθούν με την χρήση κεριών (candlestick). Με αυτό τον τρόπο μπορούν σε ένα γράφημα να αναπαρασταθούν οι τέσσερις τιμές της ισοτιμίας (τιμή ανοίγματος, κλεισίματος, μέγιστη και ελάχιστη) καθώς και το αν η τιμή ανοίγματος είναι μικρότερη από την τιμή κλεισίματος (αυξένση - μπλε ή πράσινο χρώμα) ή αν μεγαλύτερη (μειώση - κόκκινο χρώμα). Στα candlesticks η τιμή ανοίγματος και κλεισίματος περιέχονται στο παραλληλόγραμμο και η μέγιστη και ελάχιστη τιμή στις κάθετες γραμμές που διέρχονται από αυτό. Όπως βλέπουμε και στο σχήμα 4.1 όπου αναπαρίστανται είκοσι πέντε τιμές της ισοτιμίας Ευρώ - Αυστραλιανό Δολάριο με την χρήση της βιβλιοθήκης finplot.



Σχήμα 4.1: EUR/AUD Candlesticks

## 4.2 Profit and Loss - P&L

Με τον όρο Profit and Loss αναφερόμαστε στη δήλωση εισοδήματος για συγκεκριμένες κινήσεις. Ο όρος χρησιμοποιείται γενικότερα για πολλές οικονομικές δραστηριότητες όπως λειτουργία εταιρίας για συγκεκριμένο χρονικό διάστημα, επένδυση ποσού και συναλλαγές σε χρηματοοικονομικές αγορές. Στα πειράματα μας θα χρησιμοποιηθεί ως μέτρηση της αποτελεσματικότητας των μοντέλων μας με την χρήση της τεχνικής Backtesting, δηλαδή το να εφαρμόσουμε τις στρατηγικές διεκπεραίωσης συναλλαγών για ένα διάστημα του

παρελθόντος οπότε μπορούμε να γνωρίζουμε το αποτέλεσμα. Αυτή η τεχνική θα δώσει μια προσέγγιση για την λειτουργία των στρατηγικών αυτών σε πραγματικές συναλλαγές. Ένα υψηλό PL δίνει μεγαλύτερη εμπιστοσύνη προς τις στρατηγικές που το πετύχαν.

$$P\&L = \sum_i^N profit_i - loss_i \quad (4.1)$$

### 4.3 Commision

Με τον όρο Commision (προμήθεια) εννοούμε το ποσοστό του ποσού μιας συναλλαγής που κρατιέται από την υπηρεσία για τις υπηρεσίες της. Το πόσο αυτό συνήθως είναι λιγότερο από 0.1 % άλλα καθώς διενεργούνται μεγάλοι αριθμός συναλλαγών συμπεριλαμβάνεται στον υπολογισμό του P&L. Στα πειράματά μας η προμήθεια είναι 0.002 % η οποία θεωρείται ένα λογικό πόσο για συναλλαγές ισοτιμιών νομισμάτων.



# 5

## Επεξεργασία Φυσικής Γλώσσας

---

Η Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP) είναι η επιστήμη η οποία έχει ως σκοπό την ανάλυση και κατανοήση δεδομένων σε φυσική γλώσσα με την χρήση υπολογιστικών μηχανών. Με τον όρο φυσική γλώσσα αναφερόμαστε στο γεγονός ότι τα λεξικά δεδομένα δεν έχουν κάποια δομημένη μορφή και βρίσκονται στη μορφή στην οποία παράχθηκαν από τον ομιλητή. Για να μπορεί να υπολογιστής να είναι σε θέση να επεξεργαστεί τα γλωσσικά δεδομένα πρέπει τα αρχεία φυσικής γλώσσας να μετατραπούν σε διανύσματα (word embedding), για την διαδικασία αυτή υπάρχουν διάφορες τεχνικές. Η Επεξεργασία Φυσικής Γλώσσας έχει εξελιχθεί σε κύριο κλάδο της πληροφορικής επιστήμης με την άνοδο των μέσων κοινωνικής δικτυώσης καθώς μέσω αυτών υπάρχει πρόσβαση σε μεγάλο αριθμό δεδομένων γραπτού λόγου.

### 5.1 Προεπεξεργασία Κειμένου

Η Προεπεξεργασία Κειμένου είναι μια σημαντική διαδικασία καθώς βοηθάει στην απόσπαση της πληροφορίας από ένα αρχείο φυσικής γλώσσας.

- **Tokenization** Η διαδικασία διαχωρισμού των μερών ένος κειμένου σε μικρότερα κομμάτια (tokens). Τα κομμάτια αυτά μπορεί να είναι λέξεις, αριθμοί και σημεία στίξης.
- **Αφαίρεση stop-words** Η διαδικασία αφαίρεσης λέξεων που δεν περιέχουν πληροφορία και συμβάλλουν μόνο στη συντακτική δομή του κειμένου.
- **Stemming** Η διαδικασία αφαίρεσης της κατάληξης μιας λέξης, δηλαδή η λέξη *thinking* θα μετατραπεί σε *think*.
- **Lemmatization** Η διαδικασία επιστροφής μιας λέξης στην αρχική της μορφή. Έχει παρόμοιο αποτέλεσμα με το Stemming, όλα χρησιμοποιεί λεξικά με τις ρίζες των λέξεων και δεν αφαιρεί απλά την κατάληξη. Για παράδειγμα οι λέξεις *am*, *are* και *is* θα μετατραπούν σε *be* που είναι η αρχική τους λέξη, παρομοιώς η λέξη *though* θα μετατραπεί σε *think*.

## 5.2 Bag-of-Words (BoW)

Η διαδικασία Bag-of-Words έχει ως σκόπο την αποθήκευση και αναπαράσταση κειμένων φυσικής γλώσσας σε μορφή διανισμάτων (word embedding). Έχοντας έναν συγκεκριμένο αριθμό προτάσεων  $n$ , μετράμε τον αριθμό των ξεχωριστών λέξεων στις προτάσεις αυτές και δημιουργούμε η πίνακες μήκους ίσου με τον αριθμό των ξεχωριστών λέξεων. Σε κάθε στοιχείο των πινάκων που αντιστοιχεί σε μια λέξη έχουμε τον αριθμό που η λέξη αυτή εμφανίζεται μέσα στο κείμενο.

Όπως φαίνεται και στο παράδειγμα στον πίνακα 9.6 για 4 προτάσεις έχουμε 4 διανύσματα μεγέθους ίσου με τον αριθμό των ξεχωριστών λέξεων και σε κάθε θέση έχουμε τον αριθμό που η λέξη αυτή εμφανίζεται στη πρόταση.

Η αποθήκευση κειμένων με την μέθοδο BoW γίνεται ασύμφορη όταν ο αριθμός των ξεχωριστών λέξεων αυξάνεται καθώς το μεγαλύτερο μέρος των πινάκων καλύπτεται από μηδανικά.

Sentence	Bag-of-Words					
	the	red	dog	cat	eats	food
the red dog	1	1	1	0	0	0
cat eats dog	0	0	1	1	1	0
dog eats food	0	0	1	0	1	1
red cat eats	0	1	0	1	1	0

Πίνακας 5.1: Παράδειγμα Bag-of-Words

## 5.3 Term Frequency-Inverse Document Frequency (TF-IDF)

Μια πιο εξελιγμένη εκδοχή του μοντέλου BoW για αναπαράσταση κειμένου είναι η μέθοδος Term Frequency-Inverse Document Frequency (TF-IDF). Η μέθοδος αυτή υπολογίζει 2 παραμέτρους, την συχνότητα του όρου (TF) δηλαδή το πόσες φορές εμφανίζεται ο όρος στο κείμενο και στην συχνότητα αντίστροφων εγγράφων (IDF) δηλαδή το πλήθος των εγγράφων που περιέχουν τον όρο. Ο τελικός όρος είναι το γινόμενο των 2 μετρήσεων (TF-IDF).

$$TF_{t,d} = \frac{A_{t,d}}{N_d} \quad (5.1)$$

Όπου  $A_{t,d}$  οι εμφανίσεις του όρου  $t$  σε ένα έγγραφο  $d$  και  $N_d$  το πλήθος των όρων του εγγράφου  $d$ .

$$IDF_t = \log \frac{N}{N|t|} \quad (5.2)$$

Όπου  $N$  ο αριθμός των έγγραφων και  $|t|$  ο αριθμός των έγγραφων που περιέχουν τον όρο  $t$ .

$$TF - IDF_{t,d} = TF_{t,d} * IDF_t \quad (5.3)$$

## 5.4 Word2Vec

Βασικό μειονέκτημα των μεθόδων BoW και TF-IDF είναι πως δεν περιέχουν καμία σημασιολογική ερμηνεία της λέξης, δηλαδή 2 λέξεις συνώνυμες ή ακόμα και ταυτόσημες θα αντιμετωπιστούν ως διαφορετικοί όροι. Αυτό το μειονέκτημα καλύπτεται με τον αλγόριθμο Word2Vec.

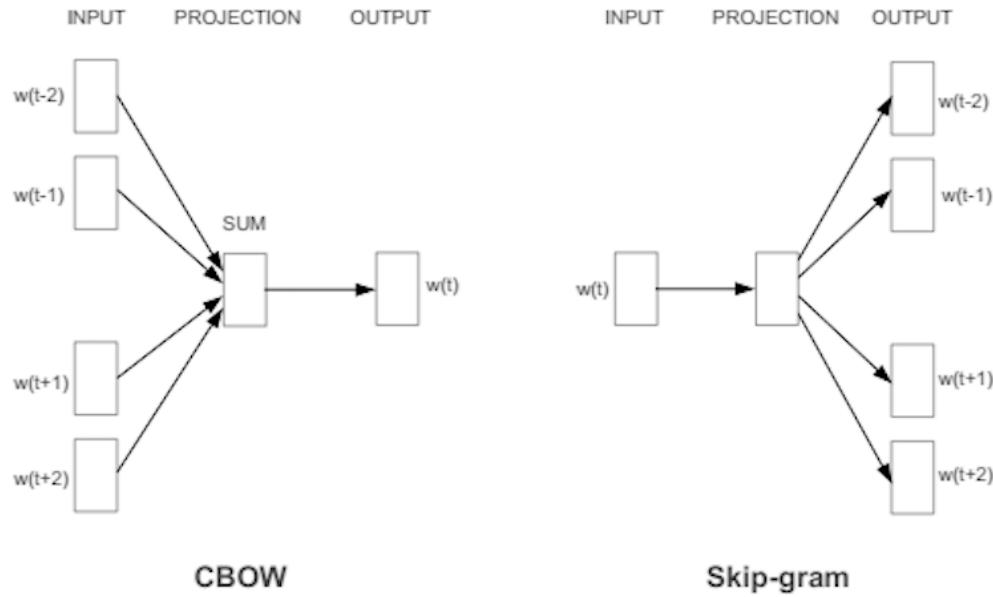
Το μοντέλο Word2Vec αναπτύχθηκε από τους Giuseppe Futia, Antonio Vetro, Alessio Melandri και Juan Carlos De Martin το 2018 [FVMM18]. Το μοντέλο περιλαμβάνει δυο λειτουργίες η Continuous Bag of Words (CBOW) και τη Skip Gram. Οι δυο αυτές λειτουργίες έχουν αντίθετους στόχους καθώς η μέθοδος CBOW προσπαθεί να προβλέψει μια λέξη έχοντας τα συμφραζόμενα ενώ η μέθοδος Skip Gram προσπαθεί από μια λέξη να προβλέψει διαφορετικές λέξεις που μπορεί να είναι συμφραζόμενες της. Για την κάθε μέθοδο απαιτείται εκπαίδευση με ένα νευρωνικό δίκτυο 2 στρωμάτων και ως είσοδο ένα σύνολο κειμένων. Σκοπός της εκπαίδευσης είναι η κωδικοποίηση των λέξεων σε διανυσμάτων λέξεων (word vector) μέσω των οποίων θα γίνει η κατάλληλη επιλογή των απαντήσεων για τις 2 μεθόδους.

Βασική λειτουργία του Word2Vec είναι η εύρεση ομοιότητας μεταξύ των λέξεων ώστε σημασιολογικά μοντέλα να αναπαρίστανται με την χρήση διανυσμάτων. Δημοφιλές παράδειγμα είναι η εξίσωση 5.4 η οποία αναπαριστά μαθηματικά την έννοια της λέξης βασίλισσας, δηλαδή είναι η έννοια της λέξης βασιλιάς αν αφαιρεθεί από αυτήν η έννοια της λέξης άντρας και προστεθεί η έννοια της λέξης γυναίκα.

$$\text{queen} = (\text{king} - \text{man}) + \text{woman} \quad (5.4)$$

## 5.5 VADER

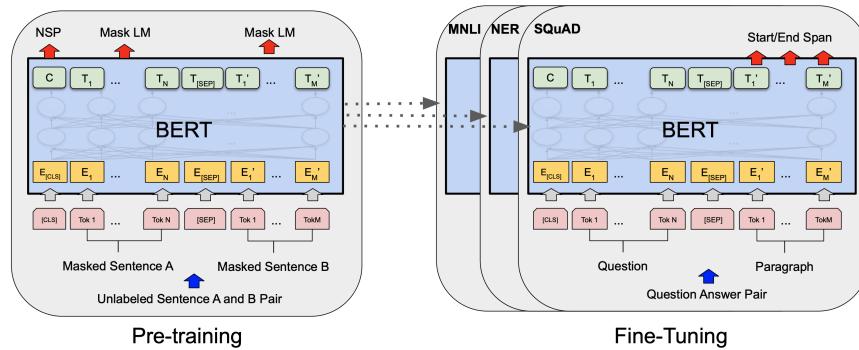
Η τεχνολογία VADER (Valence Aware Dictionary and sEntiment Reasoner) δημιουργήθηκε από τους Clayton J. Hutto και Eric Gilbert το 2014 [HG14]. Είναι ένα μοντελό εξειδικευμένο για την ανάλυση συναισθήματος σε μέσα κοινωνικής δικτύωσης. Βασίζεται σε λεξιλόγια και σε κανόνες για να καθορίσει τον βαθμό θετικού και αρνητικού συναισθήματος σε ένα κείμενο. Δεν απαιτείται εκπαίδευση.



**Σχήμα 5.1:** Η αρχιτεκτονική των 2 μεθόδων του Word2Vec  
**Πηγή :** [MSC<sup>+</sup>13]

## 5.6 Bert

Το εργαλείο Bert [DCLT18] (Bidirectional Encoder Representations from Transformers) είναι βασισμένο στην τεχνολογία *Transformers*, είναι προεκπαίδευμενο από την Google για διάφορα προβλήματα επεξεργασίας φυσικής γλώσσας. Για προβλήματα που αφορούν την επεξεργασία οικονομικολογικών κειμένων υπάρχει η έκδοση *FinBert* [Ara19]. Η εκπαίδευση του μοντέλου Bert γίνεται μέσω δυο φάσεων, του pretraining και του fine-tuning. Το pretraining είναι η προεκπαίδευση του μοντέλου σε προβλήματα εύρεσης κρυμμένης λέξης (Masked Language Modeling) και σε προβλήματα ακολουθίας προτάσεων (Next Sentence Prediction). Η διαδικασία του fine-tuning συμβαίνει από τον προγραμματιστή για να εξειδικεύσει το μοντέλο στο πρόβλημα που τον αφορά.



**Σχήμα 5.2:** Οι φάσεις εκπαίδευσης του μοντέλου BERT  
**Πηγή :** [DCLT18]

# 6

## Δεδομένα

---

Τα δεδομένα τα οποία χρησιμοποιούνται στα πειράματα είναι οι τιμές ισοτιμιών μεταξύ νομισμάτων. Οι ισοτιμίες που χρησιμοποιήσαμε είναι μεταξύ του Ευρώ και των : Δολάριο Η.Π.Α. ( EURUSD), Δολάριο Αυστραλίας (EURAUD), Δολάριο Καναδά (EURCAD) , Φράγκο Ελβετίας (EURCHF) , Λίρα Στερλίνας Βρετανίας (EURGBP) και Γιεν Ιαπωνίας (EURJPY). Τα δεδομένα περιέχουν την ημερομηνία και ώρα ως προς το λεπτό, την τιμή ανοίγματος (open), την τιμή κλεισμάτος (close) , την μέγιστη τιμή (high) και την ελάχιστη (low).

date	open	high	low	close
2019-12-31 16:55:00	1.12117	1.12118	1.12115	1.12118
2019-12-31 16:56:00	1.12118	1.12120	1.12115	1.12115
2019-12-31 16:57:00	1.12115	1.12115	1.12105	1.12105
2019-12-31 16:58:00	1.12105	1.12110	1.12099	1.12099
2019-12-31 16:59:00	1.12099	1.12115	1.12076	1.12076

**Σχήμα 6.1:** Οι τιμές της ισοτημίας EURUSD

Στο σχήμα 6.2 αναπαρίστανται οι τιμές ανοίγματος των ισοτιμιών EURUSD, EURCAD και EURGBP. Παρατηρούμαι ότι τα δεδομένα των ισοτιμιών δεν είναι διαθέσιμα όλα για τις ίδιες ημερομηνίες.

Επιπλέον χρησιμοποιήθηκαν τιμές του κρυπτονομίσματος *Bitcoin* καθώς και δημοσιεύσεις του μέσου κοινωνικής δικτύωσης Twitter (tweets). Τα tweets έχουν δημοσιευτεί μεταξύ τους διαστήματος από 2021-02-01 εως 2021-04-17 οπότε οι ανάλογες τιμές ισοτιμίας του *Bitcoin* θα εξεταστούν σε αυτό το διάστημα.

### 6.1 Προεπεξεργασία Δεδομένων

Τα Δεδομένα για να είναι σε θέση να γίνουν κατανοητά από το Υπολογιστικό Σύστημα πρέπει να εφαρμοστούν σε επεξεργασία. Τόσο τα δεδομένα τιμών των δεικτών που θα εξε-



**Σχήμα 6.2:** Οι τιμές ανοίγματος νομισμάτων



**Σχήμα 6.3:** Οι τιμές κλεισμάτος των Bitcoin - USD

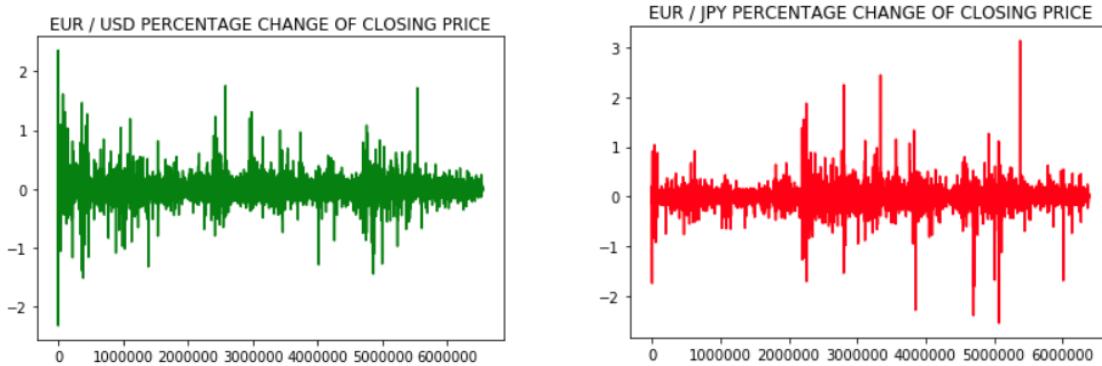
τάσουμε όσο και τα δεδομένα φυσικής γλώσσας.

## 6.2 Προεπεξεργασία Δεδομένων Ισοτιμιών

Τα δεδομένα τιμών για να είναι μπορούν να συγχρατήσουν τις πληροφορίες που περιέχουν με μικρότερο κόστος μετατρέπονται στην ποσοστιαία μεταβολή της τιμής σε σχέση με την τιμή της προηγούμενης μέρας.

Ο τύπος υπολογισμού της ποσοστιαίας μεταβολής είναι :

$$\text{percentage}_t = \frac{\text{price}_{t+1} - \text{price}_t}{\text{price}_t} * 100 \quad (6.1)$$



**Σχήμα 6.4:** Η ποσοστιαία μεταβολή των τιμών κλεισίματος του EUR - USD και EUR - JPY αντίστοιχα

## 6.3 Προεπεξεργασία Λεξικολογικών δεδομένων

Οι δημοσιεύσεις που λαμβάνονται από το μέσο κοινωνικής δικτύωσης *Twitter* πριν περαστούν από το σύστημα πρέπει να προεπεξεργαστούν για να είναι πιο εύκολη η κατανόηση τους. Αρχικά αφαιρούμε από το αρχείο όλες τις μη απαραίτητες πληροφορίες όπως το όνομα χρήστη και την τοποθεσία καθώς δεν θα εξεταστούν στα μοντέλα μας. Στη συνέχεια έχοντας μόνο την ημερομηνία και το περιεχόμενο του tweet θα το καθαρίσουμε από ειδικούς χαρακτήρες και links μετατρέποντας πρώτα όλους τους χαρακτήρες σε μικρά γράμματα.

**Σχήμα 6.5:** Ένα tweet πριν και μετά την προεπεξεργασία

## 6.4 Ετικέτες Δεδομένων

Με τον όρο ετικέτες δεδομένων αναφερόμαστε στις σταθερές τιμές με τις οποίες αναπαριστούμε την κατεύθυνση μιας ισοτιμίας. Οι τιμές που χρησιμοποιούμε είναι -1, 0 και 1 για να δείξουμε μείωση, αδράνεια και αύξηση της τιμής. Για την δημιουργία των ετικετών υπάρχουν διαφορετικοί τρόποι που ο καθένας έχει διαφορετικά αποτελέσματα και χρησιμοποιούνται ανάλογα το πρόβλημα και το επιθυμητό αποτέλεσμα.

Οι τρόποι που θα εξετάσουμε είναι :

- **Oracle Labels.** Με αυτή την μέθοδο συγχρίνουμε την τιμή κλεισίματος με την προηγούμενη της και αναπαριστούμε την κατεύθυνση με τον αντίστοιχο αριθμό.
- **Smooth Oracle Labels.** Με αυτή την μέθοδο συγχρίνουμε τα κινούμενα μέσα δυο τιμών ανοίγματος ανάλογα την παράμετρο που έχουμε ως μήκος. Ως κινούμενο μέσο αναφερόμαστε στον μέσο όρο τιμών για αριθμό δεδομένων ίσο με την παράμετρο average future window.

- **Triple Barrier Labels.** Στην μέθοδο αυτή έχουμε τρεία όρια με τα οποία υπολογίζουμε τις ετικέτες των δεδομένων. Τα δύο από αυτά είναι στην τιμή της ισοτιμίας, αν η τιμή ξεπεράσει το πάνω όριο θέτουμε ως 1 και αν ξεπεράσει το κάτω ως -1. Το τρίτο όριο είναι προς τον χρόνο. Αν δηλαδή περαστεί το όριο αυτό χωρίς να έχει περαστεί κάποιο άλλο όριο προηγουμένως τότε θεωρούμε ότι η αλλαγή της τιμής ήταν ελάχιστη και την θέτουμε ως 0. Όπως φαίνεται και στο σχήμα 6.6 έχουμε δύο όρια στον άξονα Y που αφορά την τιμή και ένα στον άξονα X που αφορά τον χρόνο. Αν η αλλαγή της τιμής είναι μεγάλου βαθμού τότε ένα από τα δύο οριζόντια όρια θα ξεπεραστούν, αν είναι μικρής κλίμακας όπως στο παρόντοιμο μας θα θεωρηθεί αδρανής.



**Σχήμα 6.6:** Παράδειγμα χρήσης *Triple Barrier Labels*  
Πηγή :

[https://mlfinlab.readthedocs.io/en/latest/labeling/tb\\_meta\\_labeling.html](https://mlfinlab.readthedocs.io/en/latest/labeling/tb_meta_labeling.html)

# 7

## Πειράματα και Αποτελέσματα Βαθιάς Μάθησης

---

### 7.1 Τεχνικά χαρακτηριστικά

Τα πειράματα έτρεξαν σε Jupyter Notebook στη γλώσσα προγραμματισμού Python έκδοση 3.7.6.

Οι βιβλιοθήκες που χρησιμοποιήθηκαν είναι :

- **numpy**, για πράξεις γραμμικής άλγεβρας.
- **pandas**, για την ανάγνωση και την επεξεργασία των DataFrames.
- **tensorflow**, για διάφορες λειτουργίες που αφορούν τα νευρωνικά δίκτυα όπως η δημιουργία , η εκπαίδευση και η αξιολόγηση τους.
- **datetime**, για τις λειτουργίες που αφορούν επεξεργασία ημερομηνίας ή ώρας.
- **os**, για την αλληλεπίδραση με το λειτουργικό σύστημα.
- **nltk**, για την επεξεργασία φυσικής γλώσσας.
- **matplotlib, finplot**, για την δημιουργία διαγραμμάτων.
- **re**, για τις πράξεις σε μεταβλητές τύπου String.
- **gensim**, για την πρόσβαση στο μοντέλο **Word2Vec**.
- **transformers**, για την πρόσβαση στο μοντέλο **Bert**.
- **vaderSentiment**, για την πρόσβαση στο μοντέλο **VADER**.

## 7.2 Πρόβλεψη κατεύθυνσης - Classification

Σε αυτά τα πειράματα προσπαθούμε να προβλέψουμε την κατευθύνση της μετοχής, δηλαδή να ταξινομήσουμε σε μια από τις κλάσεις άνοδος - στασιμότητα - κάθοδος οπότε το πρόβλημα υπάγεται στην κατηγορία προβλημάτων ταξινόμησης που γίνεται χρήσης μάθησης με επίβλεψη (Supervised Learning). Οι τρεις κατευθύνσεις αντιστοιχούν στις τιμές  $-1, 0, 1$ . Γίνεται τροφοδότηση των δεδομένων για τις ισοτιμίες συναλλάγματος όλων των νομισμάτων από το 2000 εώς το 2019. Στη συνέχεια θα πραγματοποιηθεί αξιολόγηση του μοντέλου πρόβλεψης με τον υπολογισμό του δείκτη PL (Profit and Loss) [4.2](#) και με τον υπολογισμό accuracy, precision, recall και F1-score [2.2](#) για τα δεδομένα εκπαίδευσης καθώς και για τα δεδομένα αξιολόγησης που έχουμε θέσει το διάστημα 2019 με 2020. Η επικέτα για την ταξινόμηση των δεδομένων την οποία θα χρησιμοποιήσουμε είναι η κατεύθυνση της τιμής σε σχέση με την αμέσως προηγούμενη τιμή (oracle labels). Το διάστημα χρόνου που χρησιμοποιείται είναι 1 ώρα.

Τα αποτελέσματα του συστήματος έχουν την μορφή :

Ημερομηνία	Πραγματικό Label	Προβλεπόμενο Label
2004-06-09 10:00:00	1	1
2004-06-09 11:00:00	1	1
2004-06-09 12:00:00	0	1
2004-06-09 13:00:00	-1	0
2004-06-09 14:00:00	-1	-1

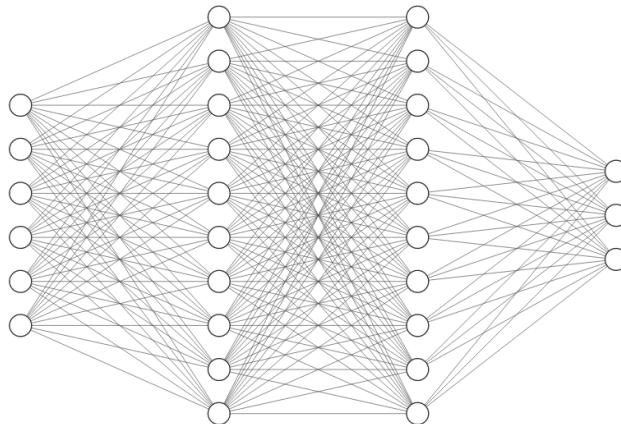
**Πίνακας 7.1:** Πραγματικές και προβλεπόμενες κατευθύνσεις για EUR/USD

Η δομή ένος νευρωνικού δικτύου με σκόπο το Classification της ισοτιμίας με αρχιτεκτονική 10 νευρώνες, 10 νευρώνες και window size = 7. Ως είσοδο δέχεται τις έξι τιμές των προηγούμενων χρονικών στιγμών και έπειτα από το πέρασμα των δυο επιπέδων νευρώνων έχουμε το επίπεδο έξοδο με τους τρεις νευρώνες που αναπαραστούν τις καταστάσεις long, stay , short (1, 0, -1). Ο αριθμός των επίπεδων , των νευρώνων ανά επίπεδο, του window size καθώς και το είδος του νευρώνα αλλάζουν ανά δοκιμή

### 7.2.1 Αποτελέσματα Μοντέλων

Το μοντέλο που θα ελέγξουμε χρησιμοποιεί την τεχνολογία MultiLayer Perceptron (MLP).

Ος βαθμός μάθησης έχουμε θέσει 0.001 και ως στόχος κέρδους (target profit) την τιμή 0.00004 (4e-5). Ος window size δηλαδή το πόσες θέσεις πίσω χρησιμοποιείται από τα



**Σχήμα 7.1:** Ένα νευρωνικό δίκτυο Classification  
**Πηγή :** Created with <http://alexlenail.me/NN-SVG/>

δεδομένα για τον υπολογισμό των προβλέψεων θέτουμε ως 7 και αλλάζει σε διαφορετικά πειράματα. Ως χαρακτηριστικά των κεριών (candle features) των συναλλαγών χρησιμοποιούμε int bar changes και int bar changes10 δηλαδή την διαφορά της συναλλαγής και την διαφορά της συναλλαγής με smoothing για 10 θέσεις. Ως συνάρτηση ενεργοποίησης χρησιμοποιείται η Rectified Linear Unit (ReLU).

Τα μοντέλα εκπαιδεύτηκαν σε όλες τις ισοτιμίες του Ευρώ για τις χρονιές 2000 εως 2019 και αξιολογήθηκαν στην ισοτιμία Ευρώ - Δολλάριο Ηνωμένων Πολιτειών με backtesting για την χρονική περίοδο 2019 - 2020 είναι :

Οι αρχιτεκτονικές που χρησιμοποιήθηκαν στα πειράματα είναι :

- **MLP model 1 :** 30 MLP, 15 MLP, windows size = 7
- **MLP model 2 :** 20 MLP, 20 MLP, windows size = 7
- **MLP model 3 :** 100 MLP, 100 MLP, 50 MLP, windows size = 20

Τα αποτελέσματα των μοντέλων με χρήση MLP για υπολογισμό P&L φαίνονται στον πίνακα 7.2 και για τα Metrics στον πίνακα 7.3 :

Αρχιτεκτονική	Train P&L	Test P&L
MLP model 1	-0.21	0.02
MLP model 2	-0.26	0.08
MLP model 3	-0.16	0.03

**Πίνακας 7.2:** Classification αποτελέσματα P&L με MLP για EUR/USD

Αρχιτεκτονική	Accuracy	Precision	Recall	F1 - score
MLP model 1	0.492	0.334	0.323	0.328
MLP model 2	0.475	0.252	0.330	0.286
MLP model 3	0.483	0.187	0.326	0.237

**Πίνακας 7.3:** Classification αποτελέσματα F1 με MLP για EUR/USD

Στη συνέχεια πραγματοποιούμε πειράματα με την χρήση της τεχνολογίας **Δικτύα Μακράς Βραχυπρόθεσμης Μνήμης - LSTM**. Τα νευρωνικά δίκτυα LSTM είναι μια υποκατηγορία των νευρωνικών δικτύων RNN [3.8] και ενδεικνύονται για την πρόβλεψη χρονοσειρών. Ως συνάρτηση ενεργοποίησης χρησιμοποιείται η Rectified Linear Unit (ReLU). Οι αρχιτεκτονικές που χρησιμοποιήθηκαν στα πειράματα είναι :

- **LSTM model 1** : 10 LSTM, 10 LSTM, 10 LSTM, windows size = 10, step size = 5
- **LSTM model 2** : 50 LSTM, 50 LSTM, windows size = 10, step size = 2
- **LSTM model 3** : 10 LSTM, 10 LSTM, windows size = 7, step size = 2
- **LSTM model 4** : 7 LSTM, 7 LSTM, windows size = 7 step size = 5
- **LSTM model 5** : 100 LSTM, 100 LSTM, windows size = 10, step size = 2

Τα αποτελέσματα των μοντέλων με χρήση LSTM για υπολογισμό P&L φαίνονται στον πίνακα 7.4 και για τα Metrics στον πίνακα 7.5 :

Αρχιτεκτονική	Train P&L	Test P&L
LSTM model 1	-0.19	0.03
LSTM model 2	0.01	0.13
LSTM model 3	-0.27	0.06
LSTM model 4	0.18	0.02
LSTM model 5	0.06	0.02

**Πίνακας 7.4:** Classification αποτελέσματα P&L με LSTM για EUR/USD

Έπειτα στα πειράματα μας χρησιμοποιούμε την τεχνολογία **Δίκτυα Περιφραγμένης Επαναλαμβανόμενης Μονάδας - GRU**. Με παρόμοια αρχιτεκτονική με τα δίκτυα LSTM τα νευρωνικά δικτύα GRU είναι επίσης υποκατηγορία των νευρωνικών δικτύων RNN. Ως συνάρτηση ενεργοποίησης χρησιμοποιείται η Rectified Linear Unit (ReLU).

Οι αρχιτεκτονικές που χρησιμοποιήθηκαν στα πειράματα είναι :

Αρχιτεκτονική	Accuracy	Precision	Recall	F1 - score
LSTM model 1	0.283	0.231	0.189	0.208
LSTM model 2	0.488	0.255	0.321	0.284
LSTM model 3	0.144	0.137	0.088	0.107
LSTM model 4	0.469	0.159	0.316	0.211
LSTM model 5	0.513	0.171	0.322	0.223

Πίνακας 7.5: Classification αποτελέσματα F1 με LSTM για EUR/USD

- **GRU model 1** : 7 GRU, 7 GRU, windows size = 20, step size = 5
- **GRU model 2** : 10 GRU, 10 GRU, windows size = 20, step size = 5
- **GRU model 3** : 25 GRU, 25 GRU, 25 GRU , windows size = 7, step size = 10
- **GRU model 4** : 50 GRU, 50 GRU, window size = 7, step size = 2

Τα αποτελέσματα των μοντέλων για υπολογισμό P&L φαίνονται στον πίνακα 7.6 και για τα Metrics στον πίνακα 7.7 :

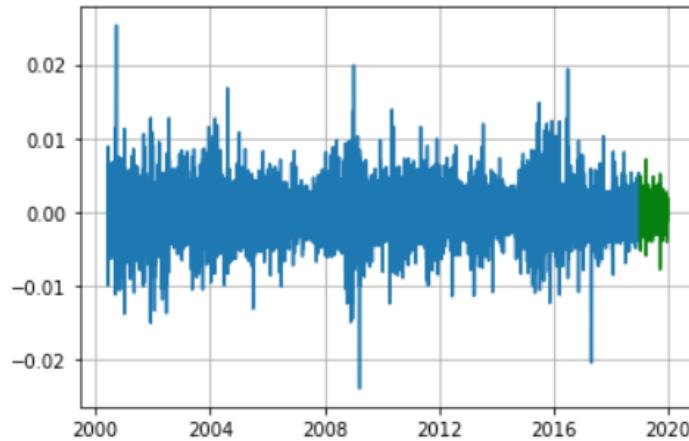
Αρχιτεκτονική	Train P&L	Test P&L
GRU model 1	0.02	0.04
GRU model 2	-0.19	0.06
GRU model 3	-0.41	0.09
GRU model 4	-0.20	0.02

Πίνακας 7.6: Classification αποτελέσματα P&L με GRU για EUR/USD

Αρχιτεκτονική	Accuracy	Precision	Recall	F1 - score
GRU model 1	0.510	0.341	0.345	0.343
GRU model 2	0.503	0.266	0.338	0.298
GRU model 3	0.518	0.174	0.316	0.224
GRU model 4	0.497	0.185	0.323	0.235

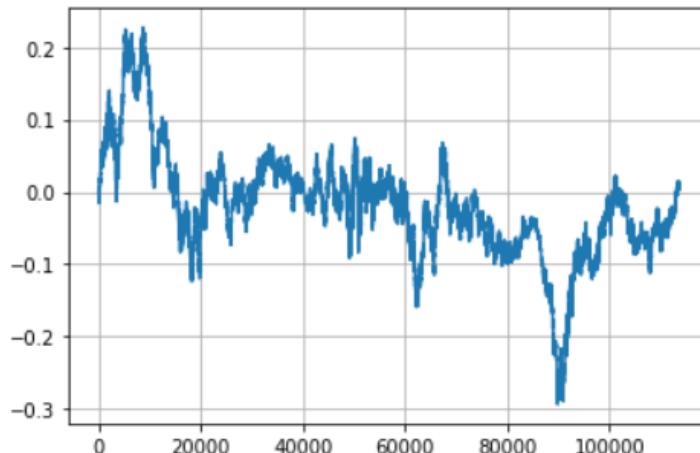
Πίνακας 7.7: Classification αποτελέσματα F1 με GRU για EUR/USD

Οι μεταβολές του P&L για το μοντέλο **LSTM model 2** για την διάρκεια των ετών εκπαίδευσης 2019 - 2020 και αξιολόγησης 2019 - 2020 φαίνονται στο γράφημα 7.2. Με μπλε χρώμα αναπαριστώνται οι μεταβολές του P&L για τα έτη εκπαίδευσης ένω με πράσινο για τη χρονική περίοδο αξιολόγησης.



**Σχήμα 7.2:** Γράφημα τιμών P&L για το 2000-2020

Οι μεταβολές του συνολικού P&L στο διάστημα εκπαίδευσης 2000 εώς 2019 εμφανίζονται στο γράφημα 7.3. Παρατηρούμε ότι υπάρχουν συνεχείς αυξομειώσεις χωρίς κάποια ιδιαίτερη υψηλή τιμή.

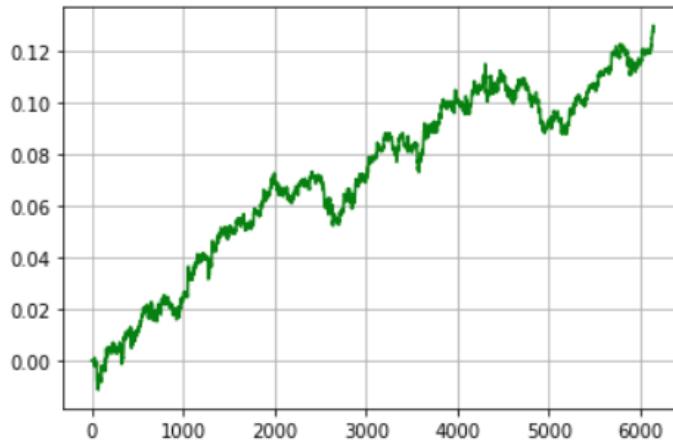


**Σχήμα 7.3:** Γράφημα συνολικού P&L για το 2000-2019

Οι μεταβολές του συνολικού P&L στο διάστημα αξιολόγησης 2019 εώς 2020 εμφανίζονται στο γράφημα 7.4. Παρατηρούμε ότι παρά μια αρχική μειώση στη συνέχεια υπάρχει συνεχόμενη άνοδος του συνολικού ποσού.

### 7.2.2 Εκπαίδευση μόνο με EUR / USD

Τα παραπάνω πειράματα εκπαιδεύτηκαν σε όλες τις ισοτιμίες του Ευρώ που αναφέρονται στη παράγραφο 6. Στη συνέχεια θα επαναλάβουμε τα πειράματα εκπαιδεύοντας τα μόνο στην



**Σχήμα 7.4:** Γράφημα συνολικού P&L για το 2019-2020

ισοτιμία EUR / USD και συγχρίνωντας τα αποτελέσματα με τις προηγούμενες μετρήσεις. Τα πειράματα χρησιμοποίησαν τις ετικέτες oracle labels. Επίσης το διάστημα μεταξύ των δεδομένων που χρησιμοποιήθηκε είναι μια ώρα.

Τα αποτελέσματα των πειραμάτων για τον υπολογισμό του P&L φαίνονται στον πίνακα 7.8

Αρχιτεκτονική	Train P&L	Test P&L
MLP model 1	-0.48	0.03
MLP model 2	-0.11	0.04
MLP model 3	-0.43	0.08
LSTM model 1	-0.20	0.03
LSTM model 2	-0.01	0.05
LSTM model 3	0.00	0.06
LSTM model 4	-0.12	0.02
LSTM model 5	0.02	0.02
GRU model 1	-0.11	-0.01
GRU model 2	0.02	0.04
GRU model 3	-0.21	0.04
GRU model 4	-0.40	0.01

**Πίνακας 7.8:** Classification αποτελέσματα P&L με εκπαίδευση μόνο σε EUR/USD

Τα αποτελέσματα των πειραμάτων για τον υπολογισμό των Metrics φαίνονται στον πίνακα 7.9

Αρχιτεκτονική	Accuracy	Precision	Recall	F1 - score
MLP model 1	0.492	0.289	0.328	0.307
MLP model 2	0.491	0.189	0.342	0.243
MLP model 3	0.442	0.246	0.356	0.291
LSTM model 1	0.487	0.258	0.312	0.282
LSTM model 2	0.432	0.192	0.305	0.236
LSTM model 3	0.178	0.235	0.208	0.221
LSTM model 4	0.413	0.168	0.156	0.162
LSTM model 5	0.443	0.179	0.242	0.206
GRU model 1	0.389	0.201	0.104	0.137
GRU model 2	0.402	0.256	0.199	0.224
GRU model 3	0.423	0.168	0.287	0.212
GRU model 4	0.389	0.182	0.250	0.211

Πίνακας 7.9: Classification αποτελέσματα Metrics με εκπαίδευση μόνο σε EUR/USD

Παρατηρούμε μια μείωση στις αποδόσεις των μοντέλων όταν αυτά εκπαιδεύονται μόνο με τα δεδομένα EUR/USD πάνω στα οποία εξετάζονται. Για το μοντέλο **GRU model 3** όπως φαίνεται και στο σχήμα 7.5 το συνολικό P&L του μοντέλου είναι υψηλότερο όταν εκπαιδεύεται σε όλες τις ισοτιμίες από ότι μόνο στην EUR / USD. Αυτό οφείλεται στο γεγονός ότι οι υπόλοιπες ισοτιμίες του Ευρώ που περιέχονται στα δεδομένα επηρεάζουν την ισοτιμία που εξετάζουμε και άρα τα δεδομένα τους μπορούν να χρησιμοποιηθούν για την εξαγωγή προβλέψεων. Στο συμπέρασμα που καταλήγουμε έχουν καταλήξει και οι Tsantekidis , Passalis , Toufa, Saitas-Zarkias, Chairistanidis, και Tefas [TPT<sup>+</sup>20] όταν εξέτασαν το κέρδος ένος trading agent που χρησιμοποιούσε δεδομένα από όλες τις ισοτιμίες για να μεγιστοποιήσει το κέρδος του.



Σχήμα 7.5: Σύγκριση συνολικού P&L για το 2019-2020 με διαφορετικά δεδομένα εκπαίδευσης

### 7.2.3 Αλλάγη τρόπου δημιουργίας ετικετών.

Τα παραπάνω πειράματα χρησιμοποίησαν ως ετικέτες oracle labels. Θα επαναλάβουμε τα πειράματα με την χρήση της μεθόδου δημιουργίας ετικετών **Smooth Oracle Labels**, όπως αυτή εξηγείται στο κεφάλαιο [Ετικέτες Δεδομένων](#), για διάφορες τιμές μήκους και θα συγχρίνουμε τα αποτελέσματα. Στα πειράματα χρησιμοποιήθηκαν τα μοντέλα με τις καλύτερες επιδόσεις από το κεφάλαιο [7.2.1](#) με την χρήση των oracle labels. Εκπαιδεύτηκαν σε όλα τα δεδομένα για την χρονική περίοδο 2000 εώς 2019 και αξιολογήθηκαν στη χρονική περίοδο 2019 εώς 2020.

Για τιμή average future window ίσο με 5.

Τα αποτελέσματα των πειραμάτων για τον υπολογισμό του PL φαίνονται στον πίνακα [7.10](#)

Αρχιτεκτονική	Train P&L	Test P&L
MLP model 1	-0.02	0.02
MLP model 3	-0.31	-0.02
LSTM model 2	0.20	-0.03
LSTM model 4	0.59	0.05
LSTM model 5	-0.32	0.03
GRU model 1	-0.20	0.03
GRU model 2	0.20	-0.03
GRU model 3	-0.54	0.08

**Πίνακας 7.10:** Classification αποτελέσματα P&L με χρήση smooth oracle labels με average future window = 5

Τα αποτελέσματα των πειραμάτων για τον υπολογισμό των Metrics φαίνονται στον πίνακα [7.11](#)

Αρχιτεκτονική	Accuracy	Precision	Recall	F1 - score
MLP model 1	0.521	0.306	0.343	0.324
MLP model 3	0.543	0.325	0.340	0.332
LSTM model 2	0.412	0.182	0.299	0.226
LSTM model 4	0.490	0.295	0.322	0.308
LSTM model 5	0.516	0.179	0.314	0.228
GRU model 1	0.408	0.239	0.300	0.266
GRU model 2	0.425	0.298	0.320	0.309
GRU model 3	0.503	0.301	0.288	0.294

**Πίνακας 7.11:** Classification αποτελέσματα Metrics με χρήση smooth oracle labels με smoothing window = 5

Δοκιμάζουμε το μοντέλο **LSTM model 4** με την χρήση smooth oracle labels με τιμές του average future window από το 5 έως το 10. Το συνολικό P&L των δοκιμών φαίνεται στο σχήμα 7.6



**Σχήμα 7.6:** Γράφημα συνολικού P&L για διάφορες τιμές average future window

Τα αποτελέσματα των πειραμάτων για τον υπολογισμό του PL με την χρήση του **LSTM model 4** για τις τιμές average future window 5 εώς 10 φαίνονται στον πίνακα 7.12

LSTM model 4		
Average Future Window	Train P&L	Test P&L
AFW = 5	0.59	0.05
AFW = 6	0.11	0.049
AFW = 7	0.20	0.033
AFW = 8	-0.19	0.033
AfW = 9	0.03	0.024
AFW = 10	-0.10	.001

**Πίνακας 7.12:** Classification αποτελέσματα P&L με χρήση smooth oracle labels με average future window 5 εώς 10

Τα αποτελέσματα των πειραμάτων για τον υπολογισμό των Metrics με την χρήση του **LSTM model 4** για τις τιμές avarege future window 5 εώς 10 φαίνονται στον πίνακα 7.13

<b>LSTM model 4</b>				
Average Future Window	Accuracy	Precision	Recall	F1 - score
AFW = 5	0.490	0.295	0.322	0.308
AFW = 6	0.479	0.177	0.298	0.222
AFW = 7	0.495	0.254	0.318	0.282
AFW = 8	0.480	0.193	0.303	0.236
AFW = 9	0.484	0.293	0.285	0.289
AFW = 10	0.464	0.197	0.306	0.240

**Πίνακας 7.13:** Classification αποτέλέσματα Metrics με χρήση smooth oracle labels με average future window 5 εώς 10

Στην συνέχεια θα γίνει χρήση της μεθόδου **Triple Barrier Labels** όπως εξηγείται στο κεφάλαιο **Ετικέτες Δεδομένων**. Ως μεταβλήτη μήκους ορίου, δηλαδή το όριο ως προς τον χρόνο θέτουμε την τιμή 7. Θα επαναλάβουμε τα πειράματα με τα μοντέλα με τις καλύτερες επιδόσεις από το κεφάλαιο 7.2.1 χρησιμοποιώντας τις νέες ετικέτες. Η εκπαίδευση θα γίνει στα δεδομένα από το 2000 έως το 2019 και η αξιολόγηση από το 2019 έως το 2020.

Τα αποτελέσματα των πειραμάτων για τον υπολογισμό του PL φαίνονται στον πίνακα 7.14

Αρχιτεκτονική	Train P&L	Test P&L
MLP model 1	-0.07	0.04
MLP model 3	-0.20	0.03
LSTM model 2	-0.04	0.10
LSTM model 4	0.10	0.03
LSTM model 5	0.05	0.02
GRU model 1	-0.01	0.01
GRU model 2	-0.05	0.02
GRU model 3	-0.21	0.02

**Πίνακας 7.14:** Classification αποτέλέσματα P&L με χρήση triple barrier labels με barrier length = 7

Τα αποτελέσματα των πειραμάτων για τον υπολογισμό των Metrics φαίνονται στον πίνακα 7.15

Αρχιτεκτονική	Accuracy	Precision	Recall	F1 - score
MLP model 1	0.502	0.326	0.357	0.341
MLP model 3	0.470	0.157	0.317	0.210
LSTM model 2	0.528	0.287	0.341	0.312
LSTM model 4	0.469	0.203	0.312	0.246
LSTM model 5	0.485	0.287	0.322	0.303
GRU model 1	0.520	0.302	0.301	0.301
GRU model 2	0.487	0.291	0.305	0.298
GRU model 3	0.515	0.300	0.348	0.322

**Πίνακας 7.15:** Classification αποτελέσματα Metrics με χρήση triple barrier labels με barrier length = 7

.  
Παρατηρώντας ότι το μοντέλο **LSTM model 2** είχε την καλύτερη απόδοση όσων αφορά το Precision και το Test P&L θα το χρησιμοποιήσουμε για τα επόμενα πειράματα μας δοκιμάζοντας τις τιμές barrier length από 5 εώς 10.

Τα αποτελέσματα των πειραμάτων για τον υπολογισμό του P&L φαίνονται στον πίνακα 7.16.

LSTM model 2		
Barrier Length	Train P&L	Test P&L
BL = 5	-0.18	0.06
BL = 6	0.04	0.02
BL = 7	0.04	0.10
BL = 8	-0.02	0.02
BL = 10	0.09	0.05

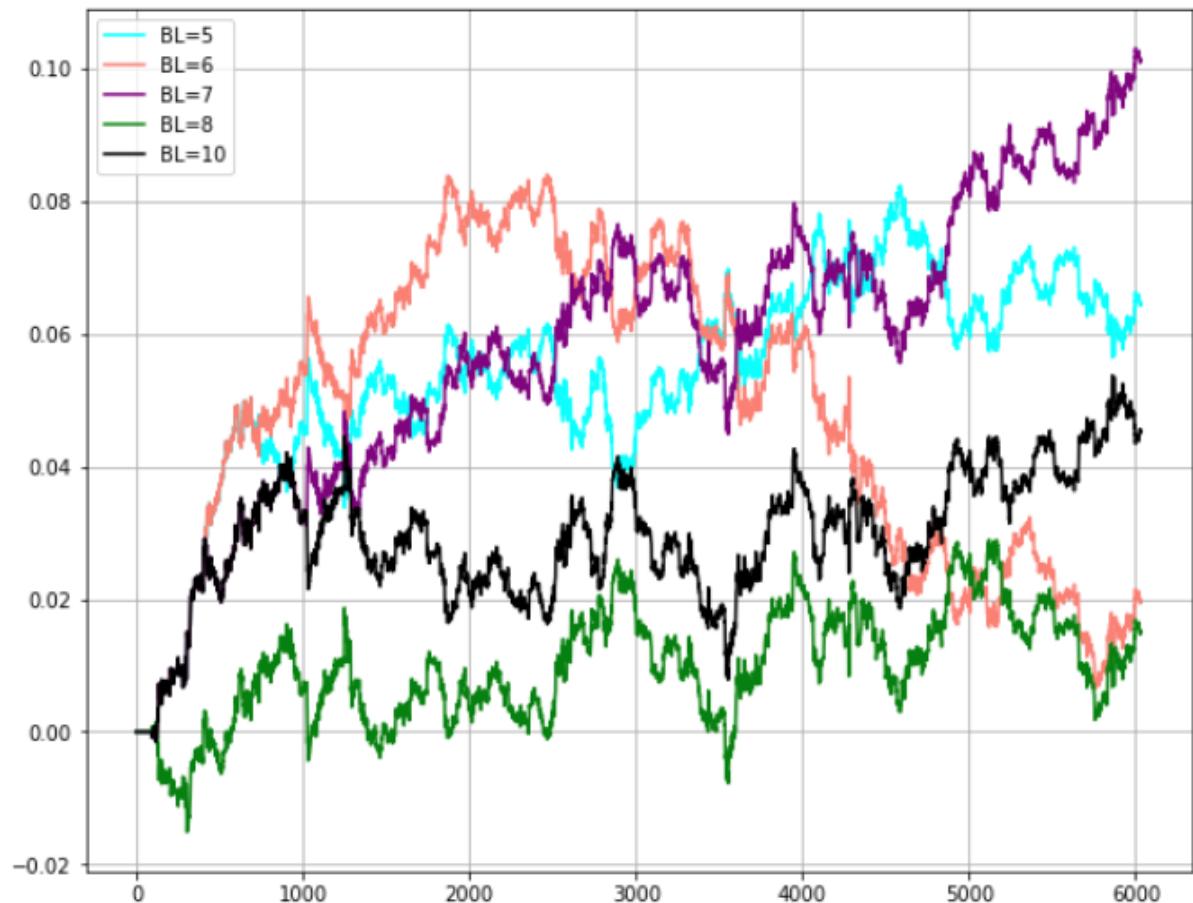
**Πίνακας 7.16:** Classification αποτελέσματα P&L με χρήση triple barrier labels με barrier length 5 εώς 10

.  
Τα αποτελέσματα των πειραμάτων για τον υπολογισμό των Metrics φαίνονται στον πίνακα 8.4.

Το συνολικό P&L των δοκιμών του δείγματος αξιολόγησης για τις τιμές του barrier length 5,6,7,8 και 10 φαίνεται στο σχήμα 7.7. Παρατηρούμε ότι το υψηλότερο P&L υφίσταται για την τιμή barrier length ίσο με 7 και το δεύτερο υψηλότερο για ίσο με 5.

LSTM model 2				
Barrier Length	Accuracy	Precision	Recall	F1 - score
BL = 5	0.477	0.184	0.321	0.234
BL = 6	0.507	0.274	0.341	0.304
BL = 7	0.528	0.287	0.341	0.312
BL = 8	0.528	0.271	0.335	0.299
BL = 10	0.513	0.247	0.328	0.282

**Πίνακας 7.17:** Classification αποτέλεσματα Metrics με χρήση triple barrier labels με barrier length 5 εώς 10



**Σχήμα 7.7:** Γράφημα συνολικού P&L για διάφορες τιμές barrier length



# 8

## Πρόβλεψη τιμής - Regression

---

Σε αυτού του είδους τα πειράματα προσπαθούμε να προβλέψουμε την τιμή της μετοχής. Δηλαδή ο στόχος είναι η πρόβλεψη μιας ακέραιας τιμής που να έχει την μικρότερη δυνατή απόσταση από την πραγματική τιμή για ένα συγκεκριμένο χρονικό διάστημα. Το είδος Βαθιάς Μάθησης που θα χρησιμοποιηθεί είναι μάθηση με επίβλεψη καθώς το υπολογιστικό σύστημα μέσα της εκπαίδευσης με ισοτιμίες κάποιου χρονικού διαστήματος θα δημιουργήσει τα μοντέλα που θα καταλήξουν για στις τιμές των ισοτιμιών για κάποιο άλλο χρονικό διάστημα που δεν έχει ξανα συναντήσει. Τα δεδομένα όπως αναφέρθηκε και στην παράγραφο 6.2 μετατρέπονται σε ποσοστιαίες μεταβολές για να επεξεργαστούν από το σύστημα. Η αξιολόγηση των αποτελεσμάτων θα πραγματοποιηθεί με τον υπολογισμό του μέσου σφάλματος απόλυτης τιμής ποσοστού (Mean absolute percentage error – MAPE) και του δείκτη PL (Profit and Loss) 4.2.

$$MAPE = \frac{100}{N} \sum_{t=1}^N \left| \frac{A_t - P_t}{A_t} \right| \quad (8.1)$$

Ο τύπος υπολογισμού της τιμής MAPE. Θεωρούμε ως  $t$  την τιμή της ισοτιμίας την χρονική στιγμή  $t$ ,  $P_t$  την πρόβλεψη της τιμής της ισοτιμίας την χρονική στιγμή  $t$  και  $N$  το πλήθος των χρονικών στιγμών για το οποίο υπολογίζουμε την τιμή αυτή.

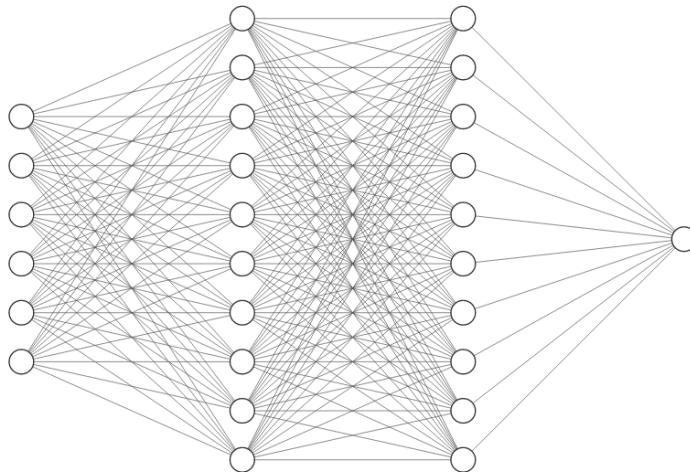
Τα αποτελέσματα των μοντέλων αφού μετατραπούν από ποσοστιαίες μεταβολές σε ακέραιες τιμές ισοτιμίας έχουν την μορφή όπως στον πίνακα 8.1.

Στο σχήμα 8.1 βλέπουμε την αρχιτεκτονική για ένα νευρωνικό δίκτυο πρόβλεψης τιμής. Η αρχιτεκτονική είναι : 10 νευρώνες, 10 νευρώνες και windows size ίσο με 7. Η είσοδος είναι οι 6 τιμές των προηγούμενων χρονικών περιόδων και ως έξοδος το προβλεπόμενο ποσοστό.

Τα μοντέλα προσέγγισης χρησιμοποιούν την τεχνολογία LSTM και GRU για την δημιουργία των νευρωνικών δικτύων. Ως συνάρτηση ενεργοποιήσης χρησιμοποιείται η *ReLU*

Ημερομηνία	Πραγματική τιμή	Προβλέπομενη τιμή
2000-06-09 10:10:00	0.94930	0.94940
2000-06-09 10:11:00	0.94950	0.94941
2000-06-09 10:12:00	0.94960	0.94961
2000-06-09 10:13:00	0.94970	0.94971
2000-06-09 10:14:00	0.95000	0.94982
2000-06-09 10:15:00	0.95010	0.95010

Πίνακας 8.1: Πραγματικές και προβλεπόμενες τιμές για EUR/USD



Σχήμα 8.1: Ένα νευρωνικό δικτύο Regression  
Πηγή : Created with <http://alexlenail.me/NN-SVG/>

και ως optimizer o *RMSprop*. Ως δεδομένα έχουμε τις τιμές κλεισίματος της ισοτιμίας *EUR/USD* για τις χρονιές 2016 - 2020. Χωρίζουμε τα δεδομένα σε δεδομένα εκπαίδευσης και σε δεδομένα αξιολόγησης με ποσοστό 75/25, άρα για την εκπαίδευση των μοντέλων χρησιμοποιούμε την χρονική περίοδο 2016 εώς 2019 και για την αξιολόγησή τους την περίοδο 2019 εώς 2020.

Για τον υπολογισμό του δείκτη P&L σε ένα πρόβλημα πρόβλεψης τιμής θα πρέπει να θεωρήσουμε ένα κατώφλι *a* (threshold) της ποσοστιαίας μεταβολής που προβλέψαμε. Με βάση αυτό το κατώφλι θα δημιουργήσουμε τις αποφάσεις short, stay και long με τις οποίες έπειτα θα υπολογίσουμε το P&L. Σε περίπτωση που η προβλεπόμενη ποσοστία μεταβολή που προβλέψαμε για την επόμενη τιμή είναι μικρότερη από - *a* τότε δίνεται η εντολή buy (-1) καθώς θεωρούμε ότι η τιμή θα πέσει. Στην περίπτωση που η προβλεπόμενη ποσοστία μεταβολή είναι μεγαλύτερη από - *a* και μικρότερη από *a* τότε δίνεται η εντολή stay (0) καθώς η τιμή αλλάζει ελάχιστα. Τέλος αν η προβλεπόμενη ποσοστία μεταβολή είναι μεγαλύτερη της τιμής *a* τότε δίνεται η εντολή long (1) καθώς η τιμή προβλέπουμε ότι θα αυξηθεί. Η τιμή του κατωφλιού αποφασίζεται με βάση την ισοτιμία που θέλουμε να προβλέψουμε, στην περίπτωση της ισοτιμίας *EUR / USD* καθώς οι μεταβολές των τιμών

είναι μικρής κλίμακας το κατώφλι που θέτουμε πρέπει να είναι επίσης μικρό. Παρομοίως για την πρόβλεψη τιμών άλλης ισοτιμίας με μεγαλύτερες ποσοστιαίες μεταβολές θα επιλεχθεί μεγαλύτερη τιμή κατωφλιού.

Οι αρχιτεκτονικές των μοντέλων που χρησιμοποιήθηκαν στα πειράματα είναι :

- **LSTM regression model 1** : 100 LSTM, 25 Dense, 1 Dense, windows size = 7, batch size = 100
- **LSTM regression model 2** : 100 LSTM, 50 LSTM, 25 Dense, 1 Dense, windows size = 7, batch size = 100
- **LSTM regression model 3** : 100 LSTM, 50 LSTM, 25 LSTM, 25 Dense, 1 Dense, windows size = 7, batch size = 100
- **LSTM regression model 4** : 200 LSTM, 100 LSTM, 25 LSTM, 25 Dense, 1 Dense, windows size = 10, batch size = 200
- **LSTM regression model 5** : 10 LSTM, 10 LSTM, 10 Dense, 1 Dense, windows size = 10, batch size = 200
- **GRU regression model 1** : 100 GRU, 50 GRU, 25 GRU, 25 Dense, 1 Dense, windows size = 7, batch size = 200
- **GRU regression model 2** : 100 GRU, 25 Dense, 1 Dense, windows size = 7, batch size = 200
- **GRU regression model 3** : 5 GRU, 5 Dense, 1 Dense, windows size = 7, batch size = 200
- **GRU regression model 4** : 10 GRU, 10 GRU, 10 Dense, 1 Dense, windows size = 7, batch size = 200
- **GRU regression model 5** : 20 GRU, 20 GRU, 20 GRU, 20 Dense, 1 Dense, windows size = 7, batch size = 200

## 8.1 Αποτελέσματα Μοντέλων *EUR / USD*

Για τα μοντέλα χρησιμοποιήθηκε threshold ίσο με 0.05.

Τα αποτελέσματα των μοντέλων για τον υπολογισμό του P&L είναι :

Τα αποτελέσματα των μοντέλων για την τιμή MAPE είναι :

Αρχιτεκτονική	Test P&L
LSTM regression model 1	0.0034
LSTM regression model 2	0.0123
LSTM regression model 3	0.0083
LSTM regression model 4	0.0156
LSTM regression model 5	0.0074
GRU regression model 1	-0.0047
GRU regression model 2	0.0035
GRU regression model 3	0.0004
GRU regression model 4	0.0055
GRU regression model 5	0.0078

Πίνακας 8.2: Regression αποτελέσματα P&L για EUR / USD

Αρχιτεκτονική	MAPE
LSTM regression model 1	0.1838
LSTM regression model 2	0.153
LSTM regression model 3	0.103
LSTM regression model 4	0.1655
LSTM regression model 5	0.2367
GRU regression model 1	0.3195
GRU regression model 2	0.3811
GRU regression model 3	0.2365
GRU regression model 4	0.3624
GRU regression model 5	0.212

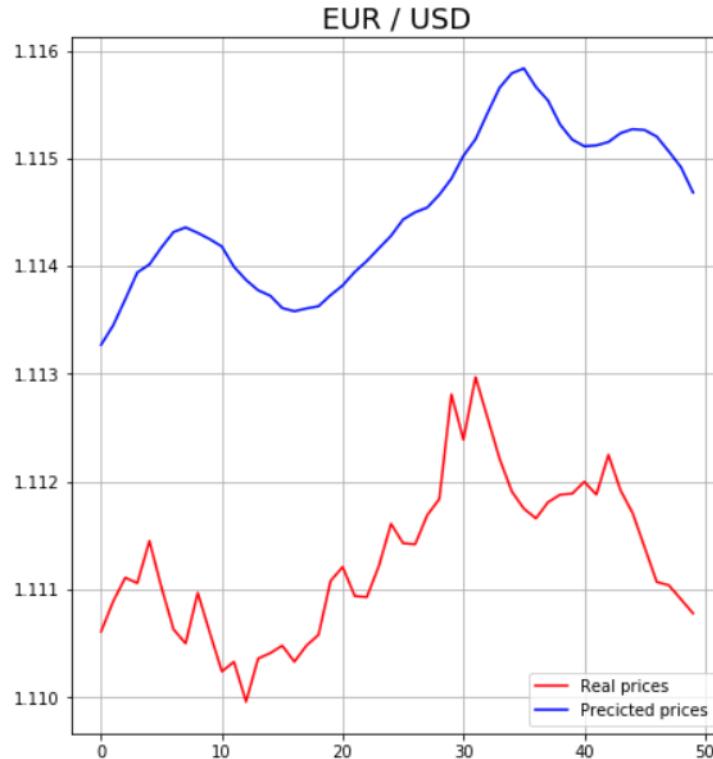
Πίνακας 8.3: Regression αποτελέσματα MAPE για EUR / USD

Τα αποτελέσματα του **LSTM regression model 4** αναπαρίστανται στο σχήμα 8.2 όπου βλέπουμε τις πραγματικές τιμές τις ισοτιμίας *EUR/USD* σε σχέση με τις προβλεπόμενες για 50 δείγματα.

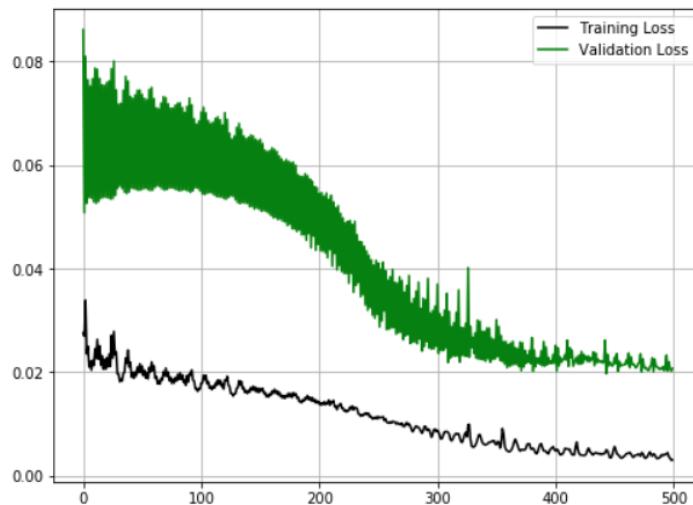
Για τον υπολογισμό της συνάρτησης κόστους κάνουμε χρήση της συνάρτησης MSE καθώς το πρόβλημα είναι πρόβλεψη τιμής. Στο σχήμα 8.3 φαίνεται το πως διαμορφώνεται το κόστος των δεδομένων εκπαίδευσης και αξιολόγησης για 500 εποχές στο μοντέλο **LSTM regression model 4**.

## 8.2 Πειράματα Threshold

Όπως αναφέραμε προηγούμενως για τον υπολογισμό του P&L σε ένα πρόβλημα Regression απαιτείται να υπάρχει η μεταβλήτη κατωφλιού (threshold). Η μεταβλήτη αυτή ορίζεται ανάλογα τα μέγεθη των ποσοστιαίων μεταβολών έτσι ώστε να μετατρέψει τις προβλεπόμενες τιμές σε κινήσεις long, short και stay. Στα προηγούμενα μοντέλα ορίσαμε το



**Σχήμα 8.2:** Πραγματικές και προβλεπόμενες τιμές της ισοτιμίας EUR/USD



**Σχήμα 8.3:** Κόστος δεδομένων εκπαίδευσης και αξιολόγησης

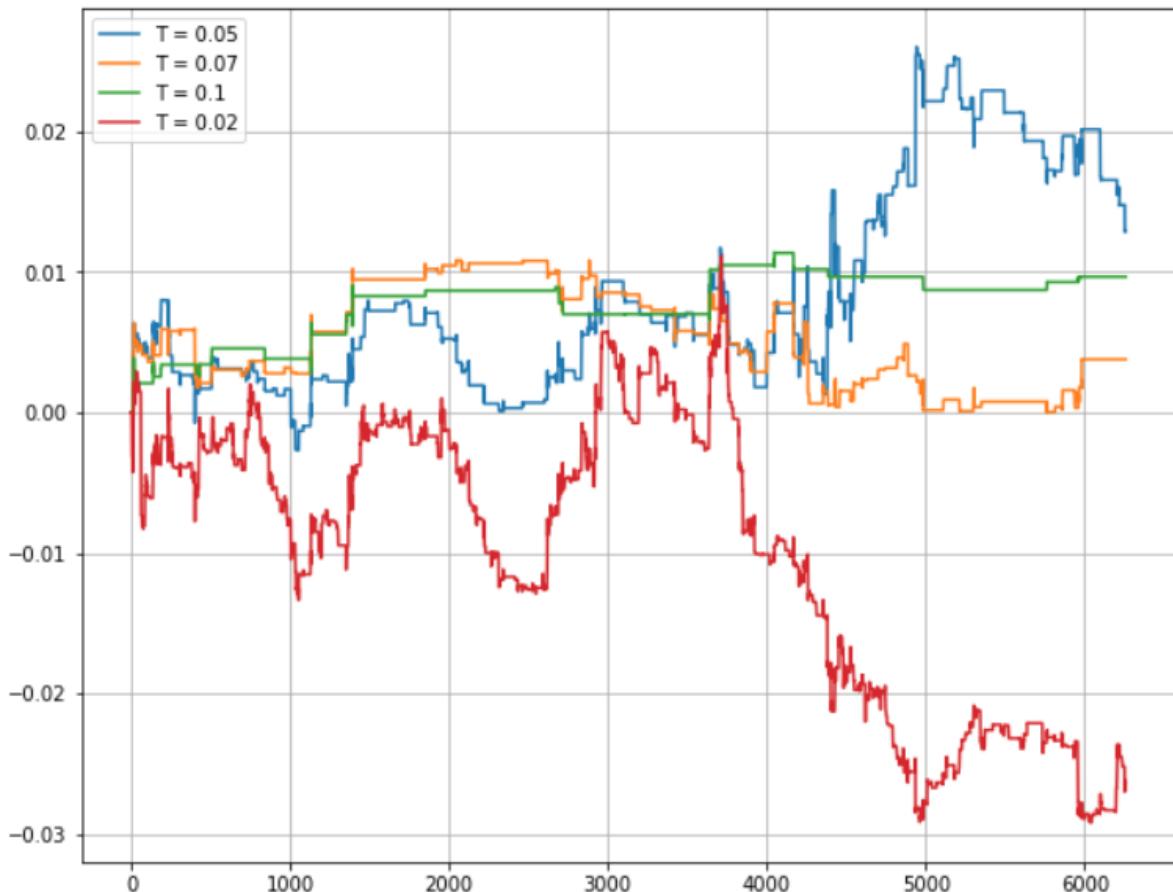
threshold ίσο με 0.05 . Στα επόμενα πειράματα θα δοκιμάσουμε την χρήση διαφορετικών τιμών threshold για το μοντέλο **LSTM regression model 4** το οποίο μας έδωσε το υψηλότερο P&L στις προηγούμενες δοκιμές.

Η συνολική πορεία του P&L για την κάθε τιμή κατωφλιού φαίνεται στο σχήμα 8.4. Παρατηρούμε ότι το υψηλότερο P&L αντιστοιχεί στην τιμή 0.05. Επιπλέον παρατηρούμε ότι όσο

LSTM regression model 4	
Threshold	Test P&L
T = 0.05	0.0156
T = 0.07	0.0289
T = 0.1	0.0114
T = 0.02	0.0265

Πίνακας 8.4: Regression αποτελέσματα P&L με χρήση διάφορων τιμών threshold

πιο μικρό το threshold τόσο πιο πολλές αγορές και πωλήσεις γίνονται με αποτέλεσμα το P&L να μεταβάλλεται περισσότερο, στο παράδειγμα όπου το threshold είναι ίσο με 0.002 η πορεία του P&L αλλάζει συνέχεια, αντίθετα στο παράδειγμα όπου threshold = 0.1 υπάρχουν ελάχιστες μεταβολές στο συνολικό P&L.



Σχήμα 8.4: Αποτελέσματα συνολικού P&L με χρήση διάφορων τιμών threshold

### 8.3 Πειράματα διαφορετικών Ισοτιμιών

Οι δοκιμές των προηγούμενων πειραμάτων έγιναν όλες στην ισοτιμία EUR / USD. Θα πραγματοποιήσουμε πειράματα με την χρήση δεδομένων από άλλες ισοτιμίες νομισμάτων

για να συγχρίνουμε τα τελικά αποτελέσματα.

Στις δοκιμές θα χρησιμοποιήσουμε το μοντέλο **LSTM regression model 4** το οποίο έδωσε την καλύτερη απόδοση στην ισοτιμία *EUR / USD*. Η χρονική περίοδο εκπαίδευσης είναι 2016 εως 2019 και της αξιολόγησης 2019 εως 2020.

<b>LSTM regression model 4</b>		
Iσοτιμία	Threshold	Test P&L
<i>EUR / USD</i>	0.05	0.0156
<i>EUR / JPY</i>	0.01	0.0078
<i>EUR / CAD</i>	0.05	0.0089
<i>EUR / AUD</i>	0.1	0.0019
<i>EUR / GBP</i>	0.05	0.0045
<i>EUR / NOK</i>	0.05	0.0071

**Πίνακας 8.5:** Regression αποτελέσματα *P&L* με χρήση διάφορων ισοτιμών

<b>LSTM regression model 4</b>	
Iσοτιμία	MAPE
<i>EUR / USD</i>	0.1655
<i>EUR / JPY</i>	0.1200
<i>EUR / CAD</i>	0.1721
<i>EUR / AUD</i>	0.2101
<i>EUR / GBP</i>	0.1988
<i>EUR / NOK</i>	0.3010

**Πίνακας 8.6:** Regression αποτελέσματα *MAPE* με χρήση διάφορων ισοτιμών



# 9

## Πειράματα και Αποτελέσματα Συναισθηματικής Ανάλυσης

---

### 9.1 Ανάλυση Συναισθήματος

Στα πειράματα αυτά θα γίνει συνδυασμός νευρωνικών δικτύων και συναισθηματικής ανάλυσης για την πρόβλεψη τιμής για συγκεκριμένη χρονική περίοδο. Οι μέθοδοι συναισθηματικής ανάλυσης που θα χρησιμοποιηθούν στα πειράματα αναλύθηκαν στο κεφάλαιο 5. Θα εξεταστεί η ισοτιμία αμερικάνικου δολλαρίου με το χρυπτονόμισμα *Bitcoin* καθώς είναι ισοτιμία που έχει συνεχώς νέα *tweets* επηρεασμένα από τις εξελίξεις καθώς και έναν μεγάλο όγκο διαθέσιμων δεδομένων. Τα λεξιλογικά δεδομένα χωρίζονται ανά ώρα καθώς μικρότερο χρονικό όριο δεν θα ήταν επαρκές για να περιέχει τον χρόνο που χρειάζεται για να γίνει εμφανής μέσω δημοσιεύσεων η αλλαγή της ισοτιμίας. Για κάθε χρονική περίοδο διαλέγουμε τυχαία 100 δημοσιεύσεις που θα χρησιμοποιήσουμε για την εξαγωγή του συμπεράσματος.

### 9.2 Πειράματα με χρήση Vader

Στα πρώτα πειράματα θα γίνει χρήση της μέθοδου *Vader* [HG14] η οποία για κάθε *tweet* θα υπολογίσει το θετικό, ουδέτερο και αρνητικό συναίσθημα. Στη συνέχεια ο μέρος όρος των συναίσθημάτων των *tweets* της κάθε ώρας, συγκεντρώνονται για να δημιουργηθεί η εικόνα για το πως αισθάνθηκαν οι χρήστες του μέσου κοινωνικής δικτύωσης για τις αλλαγές τις ισοτιμίας *BTC / USD* την συγκεκριμένη ώρα. Οι τιμές που θα εξετάσουμε είναι οι ποσοστιαίες μεταβολές της τιμής κλεισίματος. Για κάθε είδος συναίσθημας έχουμε το ποσοστό του για το συγκεκριμένο χρονικό διάστημα. Εύκολα παρατηρούμε από το σχήμα 9.1 ότι ενώ στην αρχή της εβδομάδας που εξετάζουμε το ποσοστό των δημοσιεύσεων με αρνητικό συναίσθημα είναι μεγαλύτερο από αυτών των θετικών, στο τέλος της εβδομάδας αυτό εχει αλλάξει με τις θετικές δημοσιεύσεις να είναι περισσότερες. Αυτή η αλλαγή στο συναίσθημα οφείλεται στην άνοδο της τιμής του *Bitcoin* κατά 20% μέσα στο διάστημα που

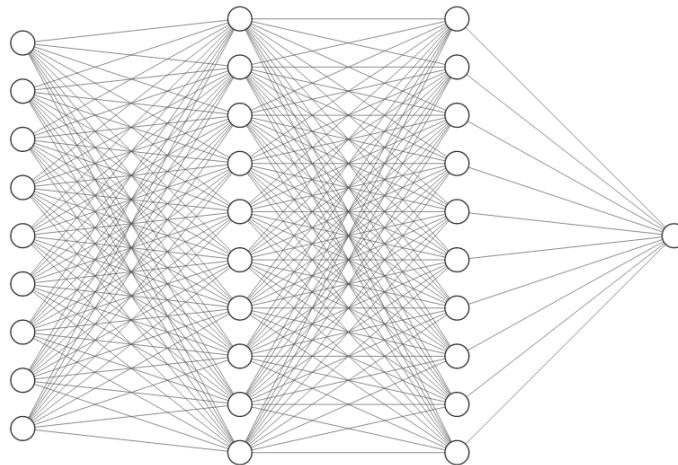
εξετάζουμε.

	<b>date</b>	<b>neg</b>	<b>neu</b>	<b>pos</b>	<b>close</b>
	2021-04-17 21	0.01481	0.88436	0.10083	50469.05
	2021-04-17 22	0.04000	0.82446	0.13554	50298.07
	2021-04-17 23	0.02889	0.86407	0.10705	50504.32
	2021-04-18 00	0.02353	0.85517	0.12129	50876.72
	2021-04-18 01	0.02439	0.89093	0.08466	50563.59
	...	...	...	...	...
	2021-04-23 17	0.04766	0.86388	0.08842	59386.59
	2021-04-23 18	0.03144	0.88844	0.08012	60006.66
	2021-04-23 19	0.02480	0.90967	0.06556	60633.78
	2021-04-23 20	0.03763	0.89142	0.07092	60713.98
	2021-04-23 21	0.04712	0.89354	0.05935	60802.38

**Σχήμα 9.1:** Συνολικό συναισθημα και τιμή κλεισίματος για BTC / USD με την χρήση VADER

Χρησιμοποιούμε την μέθοδο Regression για να προβλέψουμε την τιμή της ισοτιμίας την επόμενη χρονική θέση. Ως είσοδος στο νευρωνικό δίκτυο έχουμε τις ποσοστίαις μεταβολές των προηγούμενων θέσεων που ορίστηκαν από το windows size καθώς και τα ποσοστά του θετικού , αρνητικού και ουδέτερου συναισθήματος των δημοσιεύσεων της προηγούμενης ώρας. Παράδειγμα για ένα μοντέλο με αρχιτεκτονική 10 LSTM, 10 LSTM και windows size = 7 η γραφική του αναπάρασταση θα είναι όπως φαίνεται στο σχήμα 9.2. Με τις 6 προηγούμενες τιμές ποσοστιαίς μεταβολής και τα 3 ποσοστά συναισθήματος ως είσοδοι και την τιμή της ισοτιμίας ως έξοδος.

Τα μοντέλα προσέγγισης χρησιμοποιούν τις τεχνολογίες LSTM και GRU. Ως συνάρτηση ενεργοποίησης χρησιμοποιείται η ReLU και ως optimizer o RMSprop. Για λεξικολογικά δεδομένα έχουμε ένα dataset με δημοσιεύσεις στο μέσο κοινωνικής δικτύωσης Twitter για τις ημερομηνίες 2021 - 4 - 17 και ώρα 21:00 έως τις 2021 - 4 - 24 έως τις 21:00. Τα αποτελέσματα και τα συμπεράσματα που θα συγκεντρώθουν με την χρήση αυτών των δεδομένων μπορούν να χρησιμοποιήθουν και σε μεγαλύτερα datasets στην περίπτωση που είναι διαθέσιμα. Χωρίζουμε τα δεδομένα σε δεδομένα εκπαίδευσης και σε δεδομένα αξιολόγησης με ποσοστό 75/25. Άρα η εκπαίδευση θα γίνει περίπου τις πρώτες 5 μέρες και η αξιολόγηση τις επόμενες 2 μέρες της εβδομάδας που έχουμε τα διαθέσιμα δεδομένα. Η αξιολόγηση όπως και στα πειράματα Regression θα γίνει με την τιμή MAPE και τον υπολογισμό του P&L θέτωντας threshold.



**Σχήμα 9.2:** Ένα νευρωνικό δικτύο Regression με Vader  
**Πηγή :** Created with <http://alexlenail.me/NN-SVG/>

Οι αρχιτεκτονικές των μοντέλων που χρησιμοποιήθηκαν στα πειράματα είναι :

- **LSTM SA regression model 1** : 20 LSTM, 10 LSTM, 10 Dense, 1 Dense, windows size = 7, batch size = 1
- **LSTM SA regression model 2** : 50 LSTM, 10 LSTM, 10 Dense, 1 Dense, windows size = 5, batch size = 1
- **LSTM SA regression model 3** : 100 LSTM, 5 LSTM, 5 Dense, 1 Dense, windows size = 7, batch size = 1
- **LSTM SA regression model 4** : 20 LSTM, 7 LSTM, 7 Dense, 1 Dense, windows size = 5, batch size = 1
- **LSTM SA regression model 5** : 10 LSTM, 10 LSTM, 10 Dense, 1 Dense, windows size = 5, batch size = 1
- **GRU SA regression model 1** : 10 GRU, 10 Dense, 1 Dense, windows size = 7, batch size = 1
- **GRU SA regression model 2** : 10 GRU, 10 GRU, 10 Dense, 1 Dense, windows size = 5, batch size = 1
- **GRU SA regression model 3** : 5 GRU, 5 Dense, 1 Dense, windows size = 7, batch size = 1
- **GRU SA regression model 4** : 10 GRU, 5 GRU, 5 Dense, 1 Dense, windows size = 7, batch size = 1

- **GRU SA regression model 5 :** 7 GRU, 7 Dense, 1 Dense, windows size = 7, batch size = 1

Για τον υπολογισμό του P&L χρησιμοποιήθηκε threshold ίσο με 0.1 . Αυτό συμβαίνει διότι οι ποσοστιαίς μεταβολές της ισοτιμίας *BTC / USD* είναι μεγάλες οπότε ένα υψηλότερο threshold σε σχέση με τις προηγούμενες ισοτιμίες έχει υψηλότερη απόδοση.

Τα αποτελέσματα των μοντέλων είναι :

Αρχιτεκτονική	Test P&L
LSTM SA regression model 1	0.0463
LSTM SA regression model 2	0.1523
LSTM SA regression model 3	0.0591
LSTM SA regression model 4	0.0940
LSTM SA regression model 5	0.1201
GRU SA regression model 1	0.1114
GRU SA regression model 2	0.0952
GRU SA regression model 3	0.0564
GRU SA regression model 4	0.0989
GRU SA regression model 5	0.0861

**Πίνακας 9.1:** Αποτελέσματα P&L για Sentiment Analysis με Vader

Τα αποτελέσματα των μοντέλων για την τιμή MAPE.

Αρχιτεκτονική	MAPE
LSTM SA regression model 1	2.1384
LSTM SA regression model 2	2.2069
LSTM SA regression model 3	2.6964
LSTM SA regression model 4	1.9389
LSTM SA regression model 5	1.9389
GRU SA regression model 1	2.3689
GRU SA regression model 2	1.9683
GRU SA regression model 3	3.7037
GRU SA regression model 4	2.6849
GRU SA regression model 5	2.2788

**Πίνακας 9.2:** Αποτελέσματα MAPE για Sentiment Analysis με Vader

Αντίστοιχα τα ίδια αποτελέσματα χωρίς την προσθήκη των τιμών των ποσοστών των συναισθημάτων είναι :

Αρχιτεκτονική	Test P&L
LSTM SA regression model 1	0.0609
LSTM SA regression model 2	0.1072
LSTM SA regression model 3	0.0756
LSTM SA regression model 4	0.1072
LSTM SA regression model 5	0.0659
GRU SA regression model 1	0.0458
GRU SA regression model 2	0.0828
GRU SA regression model 3	0.0285
GRU SA regression model 4	0.0609
GRU SA regression model 5	0.0479

Πίνακας 9.3: Αποτελέσματα P&L χωρίς την χρήση Sentiment Analysis

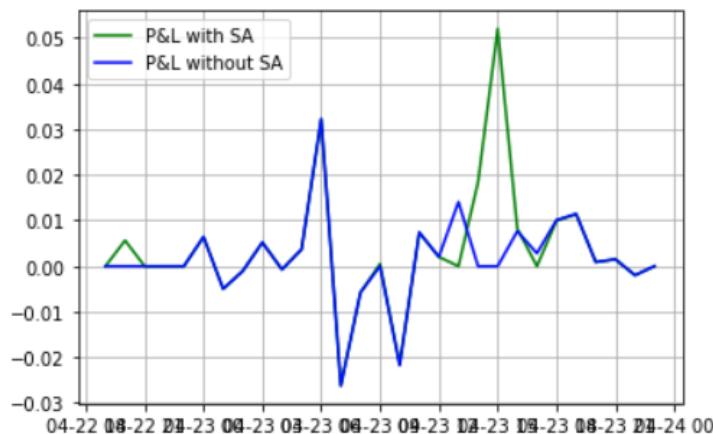
Αρχιτεκτονική	MAPE
LSTM SA regression model 1	1.4583
LSTM SA regression model 2	1.8261
LSTM SA regression model 3	1.9878
LSTM SA regression model 4	1.9643
LSTM SA regression model 5	1.4583
GRU SA regression model 1	1.3671
GRU SA regression model 2	1.9817
GRU SA regression model 3	1.5118
GRU SA regression model 4	1.4583
GRU SA regression model 5	1.5630

Πίνακας 9.4: Αποτελέσματα MAPE χωρίς την χρήση Sentiment Analysis

Μια σημαντική παρατήρηση είναι η ικανότητα επίτευξης μεγαλύτερης ακρίβειας στα μοντέλα πρόβλεψης με την χρήση των ποσοστών συναισθηματικής ανάλυσης των δημοσιεύσεων της προηγούμενης ώρας. Ήτοντας πρόβλεψη με την χρήση συναισθηματικής ανάλυσης τα μοντέλα πρόβλεψης αύξησαν τον μέσο δείκτη P&L κατά 39.12 % για τιμή threshold = 0.1 . Παρομοίως μέση τιμή MAPE αυξήθηκε κατά 44.32 %. Επίσης σημαντική σημειώση για τα αποτελέσματα P&L είναι ότι κατά την χρονική περίοδο που εξετάσαμε τις τιμές της ισοτιμίας *BTC / USD* η τιμή ανέβηκε κατά 20 % οπότε όλες οι μετρήσεις P&L είναι αυξημένες. Σε ισοτιμίες με μικρότερες διακυμάνσεις οι τιμές του P&L θα ήταν μικρότερου μεγέθους όπως συνέβη σε προηγούμενα πειράματα.

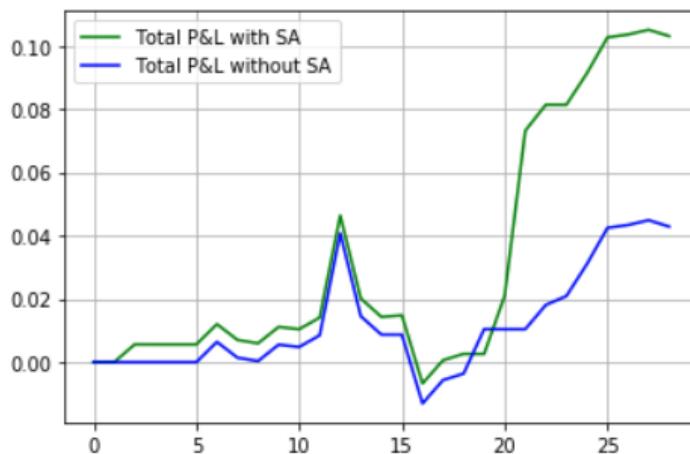
Οι τιμές του P&L κάτα την διάρκεια της χρονικής περίοδου αξιολόγησης για το μοντέλο **GRU SA regression model 1** για την χρήση ή μη συναισθηματικής ανάλυσης φαίνονται στο γράφημα 9.3. Παρατηρούμε ότι παρότι στην αρχή φαίνονται να παίρνουν τις ίδιες

απόφασεις πρός το δεύτερο μισό της χρονικής περίοδου το μοντέλο που δεν χρησιμοποιεί συναισθηματική ανάλυση δεν καταφέρνει να φτάσει τα υψηλά ποσοστά του μοντέλου που χρησιμοποιεί. Με μπλε χρώμα βλέπουμε την πορεία των μεταβολών του P&L για το μοντέλο που δεν λαμβάνει υπόψιν τις τιμές συναισθηματικής ανάλυσης και με πράσινο αυτού που το κάνει.



**Σχήμα 9.3:** Μεταβολές P&L με χρήση και μη συναισθηματικής ανάλυσης

Παρόμοια πορεία έχει και το συνολικό P&L των μοντέλων παρατηρώντας σημαντική αύξηση προς το τέλος της αξιολόγησης.



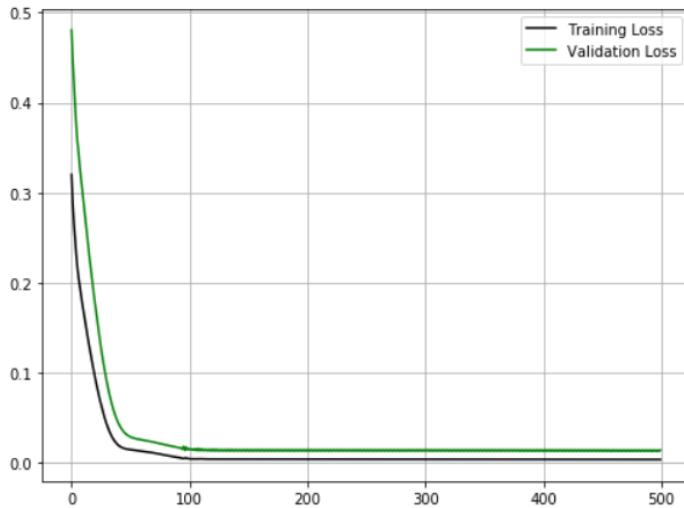
**Σχήμα 9.4:** Μεταβολές συνολικού P&L με χρήση και μη συναισθηματικής ανάλυσης

Η συνάρτηση κόστους που χρησιμοποιούμε είναι η mean squared error. Η πορεία του κόστους εκπαίδευσης και κόστους αξιολόγησης στο μοντέλο **GRU SA regression model 1** φαίνονται στο γράφημα 9.5.

Οι τελικές τιμές είναι :

loss : 0.0035

validation loss = 0.01411



**Σχήμα 9.5:** Κόστος δεδομένων εκπαίδευσης και αξιολόγησης με χρήση VADER

### 9.3 Πειράματα με χρήση FinBert

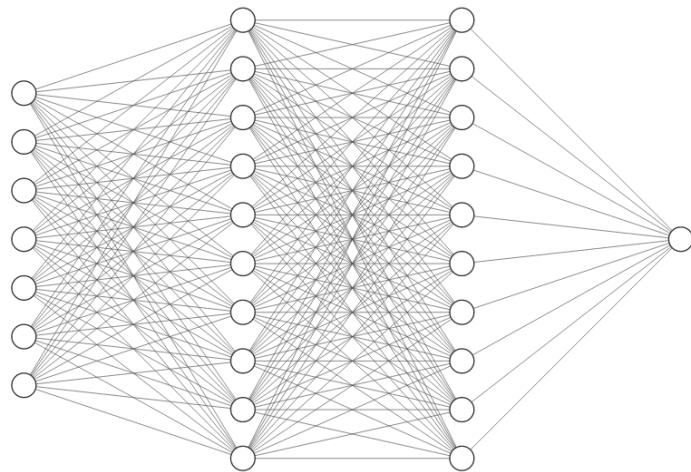
Στη συνέχεια των πειραμάτων μας γίνεται χρήση της μέθοδου **Bert** και πιο συγκεκριμένα στο pretrained μοντέλο FinBert [Ara19]. Το μοντέλο FinBert υπολογίζει για κάθε δημοσίευση τη μεταβλητή score που αναφέρει σε ακέραιο αριθμό την κλίμακα του συναισθήματος του κειμένου καθώς και τη μεταβλητή prediction όπου περιέχει σε αλφαριθμητική τιμή την κατάταξη του κειμένου σε positive, negative ή neutral. Όπως και στα προηγούμενα πειράματα θα εξεταστεί η τιμή κλεισμάτος ανά ώρα και το μέσο score συναισθήματος για το συγκεκριμένο χρονικό διάστημα.

	date	score	close
	2021-04-17 21	5.940530	50542.21
	2021-04-17 22	5.054524	50371.00
	2021-04-17 23	12.142456	50568.62
	2021-04-18 00	3.257516	50955.33
	2021-04-18 01	2.689329	50652.08
	2021-04-18 02	0.055414	49907.98
	2021-04-18 03	-1.016669	49911.42
	2021-04-18 04	-0.877423	49934.42
	2021-04-18 05	3.119814	49802.35
	2021-04-18 06	0.707514	49166.86

**Σχήμα 9.6:** Συνολικό συναισθήμα και τιμή κλεισμάτος για BTC / USD με την χρήση FinBERT

Χρησιμοποιούμε την μέθοδο Regression για να προβλέψουμε την τιμή της ισοτιμίας

την επόμενη χρονική θέση. Ως είσοδος στο νευρωνικό δίκτυο έχουμε τις ποσοστιαίες μεταβολές των προηγούμενων θέσεων που ορίστηκαν από το windows size καθώς και τα ποσοστά του θετικού, αρνητικού και ουδέτερου συναισθήματος των δημοσιεύσεων της προηγούμενης ώρας. Παράδειγμα για ένα μοντέλο με αρχιτεκτονική 10 LSTM, 10 LSTM και windows size = 7 η γραφική του αναπάρασταση θα είναι όπως φαίνεται στο σχήμα 9.7. Με τις έξι προηγούμενες τιμές ποσοστιαίας μεταβολής και το ποσοστό συναισθήματος ως είσοδος και την προβλεπόμενη τιμή της ισοτιμίας ως έξοδος.



**Σχήμα 9.7:** Era νευρωνικό δικτύο Regression με FinBERT  
**Πηγή :** Created with <http://alexlenail.me/NN-SVG/>

Χρησιμοποιούμε τις ίδιες αρχιτεκτονικές με τα μοντέλα που χρησιμοποιήθηκαν και στα πειράματα με την χρήση Vader και την ίδια τιμή για threshold = 0.1 για τον υπολογισμό του P&L.

Τα αποτελέσματα των πειραμάτων για τον υπολογισμό του PL φαίνονται στον πίνακα 9.5.

Τα αποτελέσματα των πειραμάτων για την τιμή MAPE φαίνονται στον πίνακα 9.6.

Η συνάρτηση κόστους που χρησιμοποιούμε είναι η mean squared error. Η πορεία του κόστους εκπαίδευσης και κόστους αξιολόγησης στο μοντέλο **LSTM SA regression model 5** φαίνονται στο γράφημα 9.8.

Οι τελικές τιμές είναι :

loss : 0.0014

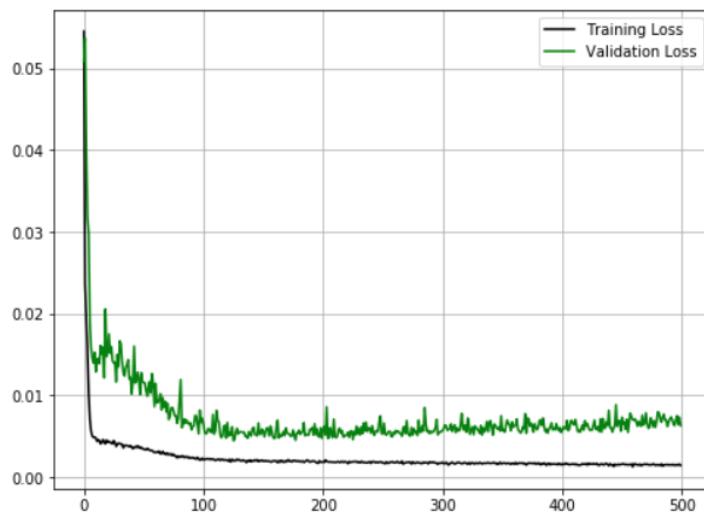
validation loss = 0.0063

Αρχιτεκτονική	Test P&L
LSTM SA regression model 1	0.0461
LSTM SA regression model 2	0.0364
LSTM SA regression model 3	0.0616
LSTM SA regression model 4	0.0136
LSTM SA regression model 5	0.0195
GRU SA regression model 1	0.0563
GRU SA regression model 2	0.1086
GRU SA regression model 3	0.0490
GRU SA regression model 4	0.0609
GRU SA regression model 5	0.1189

**Πίνακας 9.5:** Αποτέλεσμα P&L για Sentiment Analysis με FinBERT

Αρχιτεκτονική	MAPE
LSTM SA regression model 1	2.0135
LSTM SA regression model 2	2.3750
LSTM SA regression model 3	2.0600
LSTM SA regression model 4	2.3843
LSTM SA regression model 5	1.8462
GRU SA regression model 1	3.1034
GRU SA regression model 2	2.2604
GRU SA regression model 3	2.9810
GRU SA regression model 4	2.5745
GRU SA regression model 5	1.9729

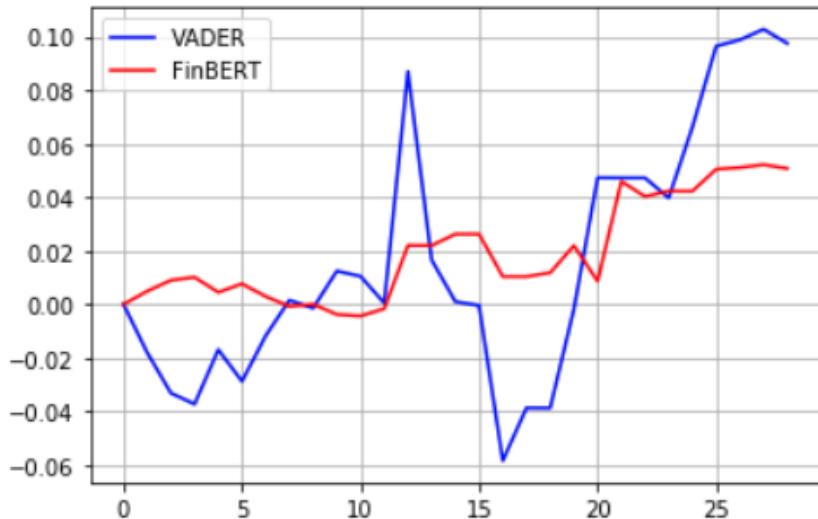
**Πίνακας 9.6:** Αποτέλεσμα MAPE για Sentiment Analysis με FinBERT



**Σχήμα 9.8:** Κόστος δεδομένων εκπαίδευσης και αξιολόγησης με χρήση FinBERT

Στη συνέχεια γίνεται σύγκριση μεταξύ των μεθόδων FinBERT και VADER χρησιμοποιώντας το μοντέλο **GRU SA regression model 5**. Παρατηρούμε ότι η απόδοση της μεθόδου

*FinBERT* είναι πιο αποτελεσματική από την μέθοδο *VADER*. Όπως φαίνεται στο σχήμα 9.9 όπου αναπαριστάται το συνολικό P&L του μοντέλου για την κάθε μέθοδο για την χρονική περίοδο αξιολόγησης.



**Σχήμα 9.9:** Σύγκριση συνολικού P&L για τις μεθόδους συνναϊσθηματικής ανάλυσης VADER και FinBERT

Για την τιμή MAPE στο συγκεκριμένο μοντέλο για την χρήση *VADER* έχουμε τίμη : 2.2788 ενώ με την χρήση *FinBERT* τίμη : 1.9729. Γενικότερα παρατηρείται μια αύξηση της τιμής MAPE κατά 9% για την χρήση του μοντέλου *VADER* σε σχέση με του *FinBERT*.

# 10

## Συμπεράσματα

---

Συνοψίζοντας, ο σκόπος της εργασίας η εξέταση του τρόπου της πρόβλεψης των τιμών ισοτιμιών με τη χρήση νευρωνικών δικτύων και συναισθηματικής ανάλυσης και του τρόπου λειτουργίας των μοντέλων σε διαστήματα χρόνου που δεν έχουν ξανασυναντήσει και το κέρδος το οποίο καταφέρνουν να βγάλουν. Συμπεραίνουμε ότι υπάρχει τρόπος να δημιουργηθούν κερδοφόρα μοντέλα νευρωνικών δικτύων υπό την προϋπόθεση πως υπάρχει επαρκής όγκος δεδομένων να εκπαιδεύτουν. Υπάρχουν πολλές διαφορετικές τεχνικές και διαφορετικοί παράμετροι οι οποίοι οδηγούν σε κερδοφόρα αποτελέσματα. Στα δικά μας πειράματα για την πρόβλεψη της ισοτιμίας *EUR / USD* επιτεύχθηκε 13 % αύξηση στον αρχικό κεφάλαιο με την μεθόδο πρόβλεψης κατεύθυνσης τιμής με την χρήση της τεχνολογίας LSTM. Η αύξηση αυτή παρατηρήθηκε για την χρονική περίοδο 2019 – 2020 με την χρήση δεδομένων όλων των ισοτιμιών. Στην ισοτιμία *Bitcoin / USD* επιτεύχθηκε 12 % αύξηση του αρχικού κεφαλαίου με τον συνδυασμό νευρωνικών δικτύων και συναισθηματικής ανάλυσης με *FinBERT*.



# Βιβλιογραφία

---

- [Ara19] Dogu Araci. Finbert: Financial sentiment analysis with pre-trained language models. *CoRR*, abs/1908.10063, 2019. [28](#), [61](#)
- [DBKR18] Sushree Das, Ranjan Behera, Mukesh Kumar, and Santanu Rath. Real-time sentiment analysis of twitter streaming data for stock prediction. *Procedia Computer Science*, 132:956–964, 01 2018. [4](#)
- [DCLT18] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2018. cite arxiv:1810.04805Comment: 13 pages. [28](#)
- [DKB16] Matthew Francis Dixon, Diego Klabjan, and Jin Hoon Bang. Classification-Based Financial Markets Prediction Using Deep Neural Networks. *SSRN Electronic Journal*, 2016. [3](#)
- [Edu20] IBM Cloud Education. Neural networks. Available : <https://www.ibm.com/cloud/learn/neural-networks>, 2020. [Online]. [11](#)
- [FHT<sup>+</sup>19] Junxi Feng, Xiaohai He, Qizhi Teng, Chao Ren, Honggang Chen, and Yang Li. Reconstruction of porous media from extremely limited information using conditional generative adversarial networks. *Physical Review E*, 100, 09 2019. [12](#)
- [FVMM18] Giuseppe Futia, Antonio Vetro, Alessio Melandri, and Juan Carlos De Martin. Training neural language models with sparql queries for semi-automatic semantic mapping. In Anna Fensel, Victor de Boer, Tassilo Pellegrini, Elmar Kiesling, Bernhard Haslhofer, Laura Hollink, and Alexander Schindler, editors, *SEMANTICS*, volume 137 of *Procedia Computer Science*, pages 187–198. Elsevier, 2018. [27](#)
- [Gan18] Rohith Gandhi. Support vector machine — introduction to machine learning algorithms svm model from scratch. Available : <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>, 2018. [Online]. [10](#)
- [GLT97] C.L. Giles, S. Lawrence, and Ah Chung Tsoi. Rule inference for financial prediction using recurrent neural networks. In *Proceedings of the IEEE/IAFE 1997 Computational Intelligence for Financial Engineering (CIFEr)*, pages 253–259, 1997. [3](#)
- [HG14] Clayton J. Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Eytan Adar, Paul Resnick, Munmun De

- Choudhury, Bernie Hogan, and Alice H. Oh, editors, *ICWSM*. The AAAI Press, 2014. 27, 55
- [HS97] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, November 1997. 18
- [HZK21] Zexin Hu, Yiqi Zhao, and Matloob Khushi. A survey of forex and stock price prediction using deep learning. *Applied System Innovation*, 4(1), 2021. 3
- [K19] Karthikeyan K. Gradient descent clearly explained. Available : <https://karthikeyanvk.medium.com/gradient-descent-clearly-explained-4be98793e894>, 2019. [Online]. 14
- [KAYN15] Arman Khadjeh Nassirtoussi, Saeed Aghabozorgi, Teh Ying Wah, and David Chek Ling Ngo. Text mining of news-headlines for forex market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment. *Expert Systems with Applications*, 42(1):306–324, 2015. 4
- [KAYT90] T. Kimoto, K. Asakawa, M. Yoda, and M. Takeoka. Stock market prediction system with modular neural networks. In *1990 IJCNN International Joint Conference on Neural Networks*, pages 1–6 vol.1, 1990. 1
- [Kei19] Robert Keim. How to use a simple perceptron neural network example to classify data. Available : <https://www.allaboutcircuits.com/technical-articles/how-to-perform-classification-using-a-neural-network-a-simple-perceptron-example>, 2019. [Online]. 9
- [Kul18] Chaitanya Kulkarni. Learning rate tuning and optimizing. Available : <https://medium.com/@ck2886/learning-rate-tuning-and-optimizing-d03e042d0500>, 2018. [Online]. 15
- [Mal73] Burton. G. Malkiel. *A Random Walk Down Wall Street*. Norton, New York, 1973. 1
- [MSC<sup>+</sup>13] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. 2013. 28
- [Pai20] Aravind Pai. Cnn vs. rnn vs. ann – analyzing 3 types of neural networks in deep learning. Available : <https://www.analyticsvidhya.com/blog/2020/02/cnn-vs-rnn-vs-mlp-analyzing-3-types-of-neural-networks-in-deep-learning/>, 2020. [Online]. 14
- [PTK<sup>+</sup>20] Nikolaos Passalis, Anastasios Tefas, Juho Kannainen, Moncef Gabbouj, and Alexandros Iosifidis. Deep adaptive input normalization for time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 31(9):3760–3765, 2020. 3

- [Ros58] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958. [9](#)
- [SMAN<sup>+</sup>13] Sadi Evren Seker, Cihan Mert, Khaled Al-Naami, Ugur Ayan, and Nuri Ozalp. Ensemble classification over stock market time series and economy news. In *2013 IEEE International Conference on Intelligence and Security Informatics*, pages 272–273, 2013. [1](#)
- [TPC<sup>+</sup>21] Toni Toharudin, Resa Pontoh, Rezzy Caraka, Solichatus Zahroh, Youngjo Lee, and Rung Chen. Employing long short-term memory and facebook prophet model in air temperature forecasting. *Communication in Statistics- Simulation and Computation*, 01 2021. [19](#)
- [TPT<sup>+</sup>20] Avraam Tsantekidis, Nikolaos Passalis, Anastasia-Sotiria Toufa, Konstantinos Saitas-Zarkias, Stergios Chairistanidis, and Anastasios Tefas. Price trailing for financial trading using deep reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–10, 2020. [3](#), [40](#)