ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ – ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Πρόγραμμα Μεταπτυχιακών Σπουδών

«ΠΜΣ Πληροφορική»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	**Αλγοριθμικές Συναλλαγές με Χρήση Μεθόδων Βαθιάς Ενισχυτικής Μάθησης** **Algorithmic Trading Using Deep Reinforcement Learning**
Ονοματεπώνυμο Φοιτητή	**Νικόλαος Κουτσούκος**
Πατρώνυμο	**Θεόδωρος**
Αριθμός Μητρώου	**ΜΠΠΛ20039**
Επιβλέπων	**Σωτηρόπουλος Διονύσιος, Επίκουρος Καθηγητής**

Ημερομηνία Παράδοσης **Απρίλιος 2024**

Τριμελής Εξεταστική Επιτροπή

Σωτηρόπουλος Διονύσιος Επίκουρος Καθηγητής	Τσιχριντζής Γεώργιος Καθηγητής	Σακκόπουλος Ευάγγελος Αναπληρωτής Καθηγητής

ΠΕΡΙΛΗΨΗ

Η παρούσα εργασία διερευνά την εφαρμογή αλγορίθμων Βαθιάς Ενισχυτικής Μάθησης, συγκεκριμένα των Deep Q-Networks (DQN) και Double Deep Q-Networks (DDQN), στη δημιουργία αυτοματοποιημένων στρατηγικών συναλλαγών για την αγορά κρυπτονομισμάτων, με επίκεντρο το Bitcoin. Η μελέτη στοχεύει να αξιολογήσει εάν η εφαρμογή τεχνικών Βαθιάς Ενισχυτικής Μάθησης (Deep Reinforcement Learning) μπορεί να επιτύχει μεγαλύτερες αποδόσεις έναντι παθητικών επενδυτικών στρατηγικών. Χρησιμοποιώντας ιστορικά δεδομένα τιμών του Bitcoin, αναπτύξαμε μια προσαρμοσμένη προσομοίωση συναλλαγών για να εκπαιδεύσουμε και να αξιολογήσουμε τα αλγοριθμικά μοντέλα DQN και DDQN. Η αξιολόγηση των αποδόσεων βασίστηκε σε χρηματοοικονομικούς δείκτες απόδοσης και ρίσκου, κατά τη διάρκεια μιας περιόδου δοκιμών δύο ετών. Η μελέτη αυτή συμβάλλει στην επέκταση της συζήτησης σχετικά με τον ρόλο της Βαθιάς Ενισχυτικής Μάθησης στις χρηματοοικονομικές αγορές, καταδεικνύοντας την ικανότητα της να βελτιώνει τις στρατηγικές συναλλαγών μέσω της βέλτιστης επιλογής των βασικών συστατικών της. Σκοπός της συγκεκριμένης εργασίας είναι όχι μόνο να αποδείξει την δυναμική της Βαθιάς Ενισχυτικής Μάθησης στη βελτίωση της απόδοσης αλγοριθμικών χρηματοοικονομικών συναλλαγών, αλλά επίσης σκιαγραφεί κρίσιμες σκέψεις για την ανάπτυξη μοντέλων συναλλαγών, στον έντονα μεταβαλλόμενο κόσμο των κρυπτονομισμάτων.

ABSTRACT

This master thesis explores the application of Deep Reinforcement Learning algorithms, specifically Deep Q-Networks (DQN) and Double Deep Q-Networks (DDQN), in crafting automated trading strategies for the cryptocurrency market, focusing on Bitcoin. The study aims to assess whether the Deep Reinforcement Learning implementation can outperform traditional buy-and-hold strategies in terms of risk-adjusted returns. Utilizing historical Bitcoin price data, we developed a custom trading simulation to train and evaluate our DQN and DDQN models. Performance assessment was grounded on key financial indicators, across a testing period of two years. This study contributes to the expanding discourse on Deep Reinforcement Learning role in financial markets, demonstrating its capability to refine trading strategies through optimal selection of its key parameters. The current master thesis not only showcases Deep Reinforcement Learning potential in enhancing algorithmic trading performance but also delineates critical considerations for deploying Deep Reinforcement Learning models in the volatile realm of cryptocurrency.

Περιεχόμενα

1. ΕΙΣΑΓΩΓΗ 6
2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ 7
2.1. Συναλλαγές επί μετοχών 7
2.2. Συναλλαγές σε Κρυπτονομίσματα 8
2.3. Συναλλαγές σε Συναλλαγματικές Ισοτιμίες 9
3. ΤΟ BITCOIN ΩΣ ΚΡΥΠΤΟΝΟΜΙΣΜΑ 10
3.1. Η εμφάνιση του Bitcoin 10
3.2. Η αγορά των Κρυπτονομισμάτων 10
3.3. Η μεταβλητότητα της αγοράς του Bitcoin και οι ευκαιρίες της 12
4. ΜΕΘΟΔΟΛΟΓΙΑ ΠΡΟΒΛΗΜΑΤΟΣ 14
4.1. Περιγραφή Προβλήματος 14
4.2. Περιορισμοί Προβλήματος 15
4.3. Στόχος της Συναλλακτικής Δραστηριότητας 16
5. ΔΗΜΙΟΥΡΓΙΑ ΠΕΡΙΒΑΛΛΟΝΤΟΣ ΓΙΑ ΣΥΝΑΛΛΑΓΕΣ ΣΕ BITCOIN 18
5.1. Χώρος Καταστάσεων 18
5.2. Χώρος Ενεργειών 19
5.3. Συνάρτηση Ανταμοιβής 20
6. ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ – ΠΡΟΕΤΟΙΜΑΣΙΑ – ΚΑΝΟΝΙΚΟΠΟΙΗΣΗ 21
6.1. Συλλογή δεδομένων 21
6.2. Προετοιμασία Δεδομένων – Κανονικοποίηση 22
6.2.1. Χειρισμός τιμών που λείπουν και ταξινόμηση δεδομένων 22
6.2.2. Διαχωρισμός δεδομένων 22
6.2.3. Υπολογισμός Δεικτών Τεχνικής Ανάλυσης 23
6.2.4. Κλιμάκωση (Scaling) των Δεδομένων 24
6.2.5. Έλεγχος για Τιμές που λείπουν και Επαναφορά Ευρετηρίου 24
7. ΑΛΓΟΡΙΘΜΟΙ ΒΕΜ ΓΙΑ ΤΟΥΣ ΠΡΑΚΤΟΡΕΣ ΕΚΤΕΛΕΣΗΣ ΣΥΝΑΛΛΑΓΩΝ 25
7.1. Deep Q-Learning Μεθοδολογία και ο Αλγόριθμος Deep Q-Network 26
7.2. Double Q-Learning Μεθοδολογία και ο Αλγόριθμος Double Deep Q-Network 27
8. ΠΕΙΡΑΜΑΤΑ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ 29
8.1. Υπερ-παράμετροι 29
8.2. Παράμετροι Συναλλαγών 31
8.3. Βασική Παθητική Στρατηγική 32
8.4. Αξιολόγηση απόδοσης: DQN, DDQN, και στρατηγικής «Αγοράς και Διακράτησης» 32

9. ΣΥΜΠΕΡΑΣΜΑΤΑ 37
10. ΜΕΛΛΟΝΤΙΚΕΣ ΠΡΟΕΚΤΑΣΕΙΣ 38
11. ΒΙΒΛΙΟΓΡΑΦΙΑ 40

1. ΕΙΣΑΓΩΓΗ

Η έλευση της τεχνητής νοημοσύνης (Artificial Intelligent - AI) και της μηχανικής μάθησης (Machine Learning - ML) στον χρηματοοικονομικό τομέα έχει εισαγάγει μια νέα εποχή αυτοματοποιημένων συστημάτων συναλλαγών, μεταξύ των οποίων είναι οι αλγόριθμοι βασισμένοι στην καινοτόμο μεθοδολογία της Βαθιάς Ενισχυτικής Μάθησης (ΒΕΜ). Οι αλγόριθμοι ΒΕΜ, διαφαίνονται πολλά υποσχόμενοι στην προσαρμογή τους στην ασταθή και απρόβλεπτη φύση των χρηματοπιστωτικών αγορών. Η παρούσα έρευνα επικεντρώνεται στην εφαρμογή δύο συγκεκριμένων αλγοριθμικών στρατηγικών ικανών για εκμάθηση και βελτιστοποίηση σε πραγματικό χρόνο. Ειδικότερα, υλοποιούνται οι αλγόριθμοι Deep Q-Networks (DQN) και Double Deep Q-Networks (DDQN), στις συναλλαγές Bitcoin, με στόχο την συγκέντρωση μεγαλύτερων αποδόσεων προσαρμοσμένων στο κίνδυνο σε σχέση με παθητικές στρατηγικές «Αγοράς και Διακράτησης» επενδυτικών προϊόντων.

Η εξέχουσα θέση στην αγορά κρυπτονομισμάτων αλλά και η έντονη μεταβλητότητα του Bitcoin παρουσιάζουν τόσο ευκαιρίες όσο και προκλήσεις για τους επενδυτές, προκαλώντας την ανάγκη για πιο εξελιγμένες στρατηγικές συναλλαγών. Οι παθητικές μέθοδοι συχνά καθυστερούν να ανταποκριθούν σε γρήγορες αλλαγές της αγοράς, από την άλλη η υιοθέτηση της ΒΕΜ μπορεί να παρέχει μια περισσότερο προσαρμόσιμη εναλλακτική λύση. Η έρευνά μας, αξιοποιεί τους αλγόριθμους DQN και DDQN για την ανάπτυξη ενός «πράκτορα» (agent) συναλλαγών που ενημερώνεται από ιστορικά δεδομένα τιμών, χρησιμοποιώντας δείκτες τεχνικής ανάλυσης για τον εμπλουτισμό της λήψης αποφάσεων και την προσομοίωση των πραγματικών συνθηκών συναλλαγών χρησιμοποιώντας ένα νέο περιβάλλον βασισμένο στο OpenAI's Gym interface.

Οι συνεισφορές αυτής της μελέτης είναι διττές: Πρώτον, ενισχύουμε το πλαίσιο λήψης αποφάσεων των μοντέλων ΒΕΜ με μια στρατηγική επιλογή δεικτών τεχνικής ανάλυσης, προσφέροντας διαφοροποιημένες πληροφορίες για τις τάσεις της αγοράς. Δεύτερον, εισάγουμε μια μοναδική προσομοίωση συναλλαγών για την εφαρμογή backtesting, επιτρέποντας μια άμεση σύγκριση απόδοσης μεταξύ των DQN και DDQN πρακτόρων μας και των συμβατικών στρατηγικών.

Δομημένη για να παρέχει μια διεξοδική διερεύνηση των τεχνικών της ΒΕΜ στις συναλλαγές σε Bitcoin, η εργασία μας περιλαμβάνει μια περιεκτική βιβλιογραφική ανασκόπηση, την τυποποίηση του προβλήματος του αλγοριθμικού προβλήματος συναλλαγών, την λεπτομερή ανάπτυξη στρατηγικής με χρήση DQN και DDQN, την μεθοδολογία για την αξιολόγηση στρατηγικών, και μια ανάλυση των ευρημάτων μας, καταλήγοντας σε μια συζήτηση σχετικά με τις δυνατότητες της ΒΕΜ στις χρηματοοικονομικές αγορές και κατευθύνσεις για μελλοντική έρευνα.

Η συγκεκριμένη έρευνά εντάσσεται στον πολλά υποσχόμενο τομέα γνώσεων και ερευνών που σχετίζεται με την εφαρμογή της ΒΕΜ στις χρηματοοικονομικές αγορές, τονίζοντας τις δυνατότητες των αλγορίθμων ΒΕΜ να ενισχύσουν τις στρατηγικές συναλλαγών στον τομέα των κρυπτονομισμάτων. Αναλύοντας την απόδοση και τις διαδικασίες λήψης αποφάσεων των πρακτόρων DQN και DDQN, στοχεύουμε να φωτίσουμε τα πλεονεκτήματα και τους περιορισμούς της χρήσης της ΒΕΜ για συναλλαγές κρυπτονομισμάτων, παρέχοντας πληροφορίες που θα μπορούσαν να βοηθήσουν στην ανάπτυξη πιο εξελιγμένων και συστημάτων συναλλαγών στο μέλλον.

2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

Ως μια προηγμένη τεχνική τεχνητής νοημοσύνης, η ΒΕΜ έχει βρει σημαντικές εφαρμογές σε διάφορους τομείς και από αυτούς δεν θα μπορούσαν να λείπουν οι αυτοματοποιημένες συναλλαγές χρηματοοικονομικών προϊόντων. Η ασταθής και απρόβλεπτη φύση της αγοράς κρυπτονομισμάτων, ιδιαίτερα του Bitcoin, παρουσιάζει μοναδικές προκλήσεις και ευκαιρίες για την εφαρμογή της ΒΕΜ. Η ακόλουθη βιβλιογραφική ανασκόπηση εξετάζει τις θεμελιώδεις εξελίξεις, τις βασικές εφαρμογές στο επενδυτικό κομμάτι μετοχών, συναλλάγματος και κρυπτονομισμάτων, και τις προκλήσεις που αντιμετωπίζει η ενσωμάτωση της ΒΕΜ σε αυτά τα περίπλοκα χρηματοοικονομικά περιβάλλοντα.

Από τη θεμελιώδη εργασία του Nakamoto, S. (2008) [7] σχετικά με το Bitcoin και την τεχνολογία blockchain, η οποία έθεσε τις βάσεις για μεταγενέστερες εργασίες και καινοτομίες στους αλγορίθμους συναλλαγών κρυπτονομισμάτων, πολλές έρευνες έχουν πραγματοποιηθεί στον τομέα των αλγοριθμικών συναλλαγών.

Η μέθοδος ΒΕΜ έχει αναδειχθεί ως μια καίρια προσέγγιση στον τομέα των επενδυτικών συναλλαγών, συνδυάζοντας την ικανότητα λήψης αποφάσεων της τεχνητής νοημοσύνης με τη δυναμική των χρηματοοικονομικών αγορών. Οι θεμελιώδεις εργασίες των Mnih, V., et al. (2013, 2015) [17],[18] στον αλγόριθμο DQN έθεσαν τα θεμέλια για τις επακόλουθες προόδους στη ΒΕΜ οι οποίες και εφαρμόστηκαν στις χρηματοοικονομικές συναλλαγές. Ειδικότερα, η εργασία τους του 2013, "Playing Atari with Deep Reinforcement Learning," [17] παρουσίασε την καινοτόμο ιδέα του συνδυασμού «Βαθιών Νευρωνικών Δικτύων» «Deep Neural Networks - DNN» με Q-learning. Η συγκεκριμένη προσέγγιση βελτιώθηκε περαιτέρω στη δημοσίευσή τους του 2015 στο Nature, "Human-level control through deep reinforcement learning," [18] καταδεικνύοντας την εξαιρετική απόδοση του “Deep Q-network” αλγόριθμου στην εκμάθηση πολιτικών μέσω υψηλής διάστασης αισθητηριακές εισόδους (high-dimensional sensory inputs).

Όπως αναφέρεται στη μελέτη των Théate, T., & Ernst, D. (2021) [16] «Ο αλγόριθμος DQN υποφέρει από σημαντικές υπερεκτιμήσεις, αυτή η υπεραισιοδοξία βλάπτει την απόδοση του αλγορίθμου.». Αυτή η πρόκληση εξετάστηκε και ένας νέος αλγόριθμος προτάθηκε από τους Van Hasselt, H., Guez, A., & Silver, D. (2016) [15] οι οποίοι εισήγαγαν την έννοια του “Double Q-Learning” στη ΒΕΜ, αντιμετωπίζοντας την «μεροληψία/σφάλμα υπερεκτίμησης» (overestimation bias) που είναι εγγενής στην Q-learning. Μέσα από την έρευνά τους κατάφεραν να δείξουν ότι ο αλγόριθμος “Double DQN” (DDQN) ενισχύει τον αρχικό DQN αντιμετωπίζοντας και μειώνοντας τις υπερεκτιμήσεις στις λεγόμενες «συναρτήσεις τιμής» (value functions), οδηγώντας σε πιο σταθερά και αξιόπιστα αποτελέσματα μάθησης. Αυτή η βελτίωση επιτυγχάνεται χωρίς να απαιτούνται πρόσθετα δίκτυα ή παράμετροι, χρησιμοποιώντας το υπάρχον πλαίσιο του DQN. Η εφαρμογή του DDQN έχει δείξει σημαντικές βελτιώσεις στην απόδοση σε πολλά παιχνίδια (για παράδειγμα Atari 2600) αποδεικνύοντας την αποτελεσματικότητά του στην εύρεση καλύτερων πολιτικών και στην προώθηση των δυνατοτήτων της ΒΕΜ.

2.1. Συναλλαγές επί μετοχών

Οι καινοτόμες προσεγγίσεις της ΒΕΜ έχουν επηρεάσει σημαντικά τις συναλλαγές επί μετοχών. Οι Yang, H., Liu, X. Y., Zhong, S., & Walid, A. (2020) [1] τόνισαν τη σημαντικότητα των «στρατηγικών συνόλου» (ensemble strategies) στη ΒΕΜ, ενισχύοντας τη λήψη αποφάσεων και την ακρίβεια πρόβλεψης στη χρηματιστηριακή αγορά. Η μελέτη τους κατέδειξε την αποτελεσματικότητα του «συνόλου» (ensemble), συγκεκριμένων ΒΕΜ μοντέλων (PPO, A2C, DDPG) στις χρηματιστηριακές συναλλαγές, δείχνοντας ότι μια τέτοια συνδυαστική προσέγγιση οδηγεί σε πιο αξιόπιστες προβλέψεις και λήψεις αποφάσεων. Σε παρόμοιο πνεύμα, οι Zhang, Z., Zohren, S., & Stephen, R. (2020) [2] παρουσίασαν πώς η ΒΕΜ θα μπορούσε να επεξεργαστεί τεράστιες ποσότητες χρηματοοικονομικών δεδομένων για να βελτιστοποιήσει τις αποφάσεις συναλλαγών, ξεπερνώντας τις παθητικές ποσοτικές μεθόδους. Οι ερευνητές διαπίστωσαν ότι οι εξεταζόμενοι αλγόριθμοι ΒΕΜ - DQN, A2C και Policy Gradients - ήταν ικανοί να αναλύσουν αποτελεσματικά τα δεδομένα της αγοράς προκειμένου να λάβουν κερδοφόρες αποφάσεις συναλλαγών, επιδεικνύοντας ανώτερη απόδοση σε σύγκριση με τις

παθητικές ποσοτικές μεθόδους. Η έννοια της αντιμετώπισης των χρηματοοικονομικών συναλλαγών ως παιχνιδιού, όπως προτάθηκε από τον Huang, C. Y. (2018) [4], εισήγαγε τις αρχές της θεωρίας παιγνίων στις χρηματοοικονομικές αγορές χρησιμοποιώντας τη ΒΕΜ. Η εφαρμογή της ΒΕΜ από τον Huang, αντιμετωπίζοντας τις χρηματοοικονομικές συναλλαγές ως παίγνιο, έδειξε ότι αυτή η προσέγγιση θα μπορούσε να αποτυπώσει αποτελεσματικά τη σύνθετη δυναμική της αγοράς.

Αντιμετωπίζοντας τις πρακτικές προκλήσεις στην ανάπτυξη στρατηγικών ΒΕΜ, οι Xiong, Z., κ.ά. (2018) [5] εφάρμοσαν έναν αλγόριθμο Deep Deterministic Policy Gradient (DDPG) και διαπίστωσαν ότι ο «εκπαιδευμένος πράκτορας υπερέχει του Dow Jones Industrial Average και της μεθόδου mid-variance portfolio allocation σε όρους συσσωρευμένης απόδοσης (accumulated return)». Οι Chen, L., & Gao, Q. (2019) [13] συνέβαλαν στην έρευνα της ΒΕΜ στις χρηματοοικονομικές συναλλαγές, παρουσιάζοντας την αποτελεσματικότητα των αλγορίθμων DQN και Deep Recurrent Q-network (DRQN) σε σενάρια αγοράς μετοχών. Η μελέτη τους έδειξε ότι ένας πράκτορας που εκτελεί συναλλαγές βασισμένος στον DQN είχε καλύτερη απόδοση τόσο από τις στρατηγικές «Αγοράς και Διακράτησης» όσο και από τις τυχαίες στρατηγικές συναλλαγών με το “S&P500 ETF” ως περιουσιακό στοιχείο. Επιπλέον, κατέδειξε ότι ο πράκτορας που εκτελεί συναλλαγές βασισμένος στον DRQN, και ο οποίος ενσωματώνει ένα πλαίσιο επανάληψης, είχε καλύτερες επιδόσεις έναντι του πράκτορα DQN αναγνωρίζοντας και εκμεταλλευόμενος αποτελεσματικά μοτίβα που σχετίζονται με το χρόνο στα δεδομένα συναλλαγών. Ο Brim, A. (2020) [14] επικεντρώθηκε στις συναλλαγές ζευγών (pairs trading) χρησιμοποιώντας έναν αλγόριθμο “Double Deep Q-Network” (DDQN), τονίζοντας την προσαρμοστικότητα και τις δυνατότητες των προηγμένων μοντέλων ΒΕΜ στην εφαρμογή σύνθετων στρατηγικών συναλλαγών σε μετοχές. Ο αλγόριθμος DDQN του ερευνητή απέδειξε την ικανότητα να αναγνωρίζει και να εκμεταλλεύεται το μοτίβο “mean reversion” των συνολοκληρωμένων (cointegrated) ζευγών μετοχών, πραγματοποιώντας επιτυχημένες προβλέψεις και εκτελώντας μια στρατηγική διαπραγμάτευσης ζευγών στη χρηματιστηριακή αγορά. Μια βασική καινοτομία, της προσέγγισης του ερευνητή, στη διαδικασία εκπαίδευσής είναι η εισαγωγή ενός «πολλαπλασιαστή αρνητικών ανταμοιβών» (Negative Rewards Multiplier), ο οποίος ενθαρρύνει το σύστημα να επιλέγει πιο συντηρητικές ενέργειες.

Μια καινοτόμος προσέγγιση ΒΕΜ προτάθηκε από τους Pricope, T. V. (2021) και Théate, T., & Ernst, D. (2021) [16]. Παρουσίασαν μια λύση βασισμένη στη ΒΕΜ, που ονομάζεται αλγόριθμος «Trading Deep Q-Network» (TDQN). Ο αλγόριθμος TDQN, βασίζεται στον αλγόριθμο Deep Q-Network (DQN), στοχεύοντας στη μεγιστοποίηση του «Δείκτη Sharpe» κατά μήκος του επενδυτικού προϊόντος και ειδικότερα των μετοχών. Η έρευνα αντιπαραβάλλει την απόδοση του αλγορίθμου TDQN με παθητικές στρατηγικές συναλλαγών, συμπεριλαμβανομένων των «Αγορά και Διακράτηση», «Πώληση και Διακράτηση», «Παρακολούθησης Τάσης» με χρήση κινητών μέσων όρων και «Επιστροφής στο Μέσο» με χρήση κινητών μέσων όρων. Η αξιολόγηση δείχνει ότι ο TDQN όχι μόνο ξεπερνά κατά μέσο όρο αυτές τις παθητικές στρατηγικές αναφοράς, αλλά παρουσιάζει επίσης σημαντικά πλεονεκτήματα, όπως αυξημένη ευελιξία και ανθεκτικότητα απέναντι σε διαφορετικά κόστη συναλλαγών, σηματοδοτώντας μια πολλά υποσχόμενη πρόοδο στον τομέα των αλγοριθμικών συναλλαγών.

2.2. Συναλλαγές σε Κρυπτονομίσματα

Σημαντική πρόοδος στην ακαδημαϊκή έρευνα έχει σημειωθεί με τη χρήση της ΒΕΜ και στις συναλλαγές κρυπτονομισμάτων. Οι Wang, Z., & Fleiss, A. (2021) [21] διερεύνησαν τη χρήση ενός αλγορίθμου DQN για τις συναλλαγές σε Bitcoin, Ethereum, και Litecoin. Πρότειναν έναν πράκτορα DQN ο οποίος κατάφερε να επιτύχει μέση απόδοση 65.98%, αν και το αποτέλεσμα συνδυάστηκε με σημαντική μεταβλητότητα κατά τη διάρκεια της περιόδου διαπραγμάτευσης. Οι Mahayana, D., Shan, E., & Fadhl'Abbas, M. (2022) [22] ανέπτυξαν ένα μοντέλο ΒΕΜ για εφαρμογή σε αλγοριθμικές συναλλαγές στο Bitcoin (BTC/USD). Χρησιμοποιώντας δεδομένα ανά λεπτό και δείκτες τεχνικής ανάλυσης, το μοντέλο τους έχει ως στόχο να υπερβεί τις παθητικές μεθόδους συναλλαγών μέσα από σήματα/εντολές αγοράς, διακράτησης ή πώλησης του Bitcoin. Χρησιμοποίησαν τον αλγόριθμο Proximal Policy Optimization (PPO) και έδειξαν ότι οι πράκτορες/μοντέλα τους δεν μπόρεσαν να

έχουν καλύτερη απόδοση από τη βασική στρατηγική «Αγοράς και Διακράτησης» σε συγκριτικές δοκιμές.

Επιπλέον, μια σχετική εργασία εφαρμόστηκε από τους Ntourmas, I., & Sotiropoulos, D. (2022) [3]. Οι ερευνητές συνδύασαν δίκτυα Long Short-Term Memory (LSTM) για την πρόβλεψη τιμών του Bitcoin ανά λεπτό με ένα μοντέλο Recurrent Reinforcement Learning (RRL) για την διαπραγμάτευση σε Bitcoin. Στα αποτελέσματα τους αποδεικνύεται ότι η ενσωμάτωση των προβλέψεων LSTM στο μοντέλο RRL αυξάνει σημαντικά τα κέρδη, ξεπερνώντας τις επιδόσεις του μεμονωμένου μοντέλου RRL.

2.3. Συναλλαγές σε Συναλλαγματικές Ισοτιμίες

Ένα ακόμη επενδυτικό προϊόν που έχει επηρεαστεί σημαντικά από την εφαρμογή της μεθόδου της BEM είναι οι συναλλαγές σε συναλλαγματικές ισοτιμίες. Οι Grover, A. A., & Gabriel, R. S. (2021) [19] δημιούργησαν έναν αλγόριθμο BEM και τον δοκίμασαν στις ισοτιμίες EUR/USD, CHF/USD, GBP/USD και NZD/USD εστιάζοντας στη μεγιστοποίηση του κέρδους. Η έρευνα κατέδειξε ο προτεινόμενος αλγόριθμος συναλλαγών είναι κερδοφόρος, ειδικά όταν εφαρμόζεται στις ισοτιμίες EUR/USD και NZD/JPY. Έτσι, ο αλγόριθμος μπορεί να χρησιμοποιηθεί ως ένα αξιόπιστο εργαλείο backtesting για την επικύρωση των στρατηγικών διαπραγμάτευσης συναλλαγών. Τέτοιες στρατηγικές αποτελούν στρατηγικές που βασίζονται σε δείκτες τεχνικής ανάλυσης, όπως ο MACD ή ο RSI, και σε ειδήσεις, επιβεβαιώνοντας τη χρησιμότητά του στη βελτίωση των προσεγγίσεων της συναλλακτικής δραστηριότητας στη βάση ενός συνδυασμού τεχνικής και θεμελιώδους ανάλυσης της αγοράς.

Οι Carapuço, J., Neves, R., & Horta, N. (2018) [20] διερεύνησαν την μέθοδο της Reinforcement Learning (RL) εφαρμοσμένη σε βραχυπρόθεσμες συναλλαγές στην ισοτιμία EUR/USD, χρησιμοποιώντας νευρωνικά δίκτυα με Q-learning. Στη μελέτη δημιούργησαν ένα περιβάλλον προσομοίωσης αγοράς για να βελτιώσουν τις διαδικασίες εκπαίδευσης και δοκιμών, διασφαλίζοντας σταθερότητα και την ικανότητα γενίκευσης σε δεδομένα εκτός δείγματος. Στη συγκεκριμένη έρευνα οι καμπύλες εκμάθησης που χρησιμοποιούνται για την επικύρωση του δείγματος αποδεικνύουν τη δυνατότητα του μοντέλου να «αποκαλύπτει» σχέσεις χρηματοοικονομικών δεδομένων που είναι ωφέλιμες για τη λήψη αποφάσεων σε δεδομένα που δεν έχουν προς το παρόν εξεταστεί. Η προσέγγιση αποδεικνύεται κερδοφόρα, κατά την περίοδο 2010-2017, επιδεικνύοντας την αποτελεσματικότητά της για κερδοφόρες συναλλαγές σε δοκιμαστικά σύνολα δεδομένων.

3. ΤΟ BITCOIN ΩΣ ΚΡΥΠΤΟΝΟΜΙΣΜΑ

3.1. Η εμφάνιση του Bitcoin

Το 2009 σηματοδότησε ένα ορόσημο στην επίδραση του ψηφιακού κόσμου στις αγορές χρήματος και κεφαλαίου καθώς ήταν η πρώτη φορά που εφαρμόστηκε το δίκτυο blockchain του κρυπτονομίσματος Bitcoin. Η ιδέα ενός blockchain όπως το ξέρουμε σήμερα ξεκίνησε με την ανάπτυξη του Bitcoin το 2008 μέσω έργου του Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System" [7], το Bitcoin παρουσίασε στον κόσμο το πρώτο αποκεντρωμένο ψηφιακό νόμισμα, υποστηριζόμενο από την τεχνολογία blockchain—ένα κατανεμημένο καθολικό (distributed ledger) που εξασφαλίζει ασφαλείς, διαφανείς και αμετάβλητες συναλλαγές. Αυτή η καινοτομία όχι μόνο αμφισβήτησε τα συμβατικά χρηματοπιστωτικά συστήματα αλλά επίσης παρουσίασε μια νέα μέθοδο διευκόλυνσης των συναλλαγών χωρίς την ανάγκη για κεντρικές αρχές.

Το blockchain του Bitcoin λειτουργεί ως ένα ψηφιακό μητρώο/καθολικό που καταγράφει με ακρίβεια όλες τις συναλλαγές με το κρυπτονόμισμα, δημιουργώντας ένα ασφαλές, διαφανές και αδιάβλητο περιβάλλον ιδανικό για χρηματοοικονομικές συναλλαγές και όχι μόνο. Από την έναρξή του, ο χώρος της τεχνολογίας blockchain έχει διευρυνθεί, δημιουργώντας διάφορες κατηγορίες κρυπτονομισμάτων:

- **Bitcoin**: Το πρώτο και πιο γνωστό κρυπτονόμισμα. Λειτουργεί ως ένα αποκεντρωμένο ψηφιακό νόμισμα, διασφαλισμένο με κρυπτογραφία και ένα παγκόσμιο δίκτυο υπολογιστών, με τις συναλλαγές του καταγεγραμμένες δημόσια στο blockchain.
- **Altcoins**: Περιλαμβάνει κρυπτονομίσματα πέραν του Bitcoin, όπως το Ethereum, Litecoin, Ripple, και Bitcoin Cash, καθένα από τα οποία προσφέρει μοναδικά χαρακτηριστικά και πιθανές χρήσεις.
- **Stablecoins**: Σχεδιασμένα για να ελαχιστοποιούν την μεταβλητότητα, αυτά τα κρυπτονομίσματα είναι συνδεδεμένα σε σταθερά περιουσιακά στοιχεία όπως το δολάριο ΗΠΑ ή ο χρυσός.
- **Privacy Coins**: Κρυπτονομίσματα όπως τα Monero, Dash, and Zcash, τα οποία δίνουν έμφαση στη διασφάλιση του απορρήτου και της ανωνυμίας των χρηστών.
- **Utility Tokens**: Σχεδιασμένα για συγκεκριμένες χρήσεις εντός αποκεντρωμένων εφαρμογών (dApps) ή πλατφορμών blockchain, παραδείγματα αποτελούν το Binance Coin και Chainlink.
- **Security Tokens**: Αντιπροσωπεύουν πραγματικά επενδυτικά στοιχεία και υπόκεινται σε κανονιστική συμμόρφωση, με παραδείγματα να περιλαμβάνουν τα tZero (TZROP), Harbor (RMT), και Swarm Fund (SWM).

Η επιρροή της τεχνολογίας blockchain δεν αναλώνεται μόνο στα ψηφιακά νομίσματα. Το Blockchain έχει τη δυνατότητα να μεταμορφώσει ένα ευρύ φάσμα βιομηχανιών, από τη διαχείριση της εφοδιαστικής αλυσίδας έως την υγειονομική περίθαλψη και τα συστήματα ψηφοφορίας.

Καθώς η τεχνολογία έχει αναπτυχθεί και ωριμάσει, υπάρχει μια αυξανόμενη αναγνώριση των πιθανών πλεονεκτημάτων του blockchain. Μεγάλες εταιρείες και κυβερνήσεις σε όλο τον κόσμο διερευνούν τώρα τη χρήση της τεχνολογίας blockchain για τη βελτίωση της ασφάλειας, της διαφάνειας και της αποτελεσματικότητας σε μια ποικιλία εφαρμογών.

Παρά τα πολλά πιθανά οφέλη του blockchain, υπάρχουν επίσης προκλήσεις και ανησυχίες που πρέπει να αντιμετωπιστούν. Ζητήματα όπως η επεκτασιμότητα, η κατανάλωση ενέργειας και η ρυθμιστική αβεβαιότητα παραμένουν σημαντικά ζητήματα καθώς η τεχνολογία συνεχίζει να εξελίσσεται.

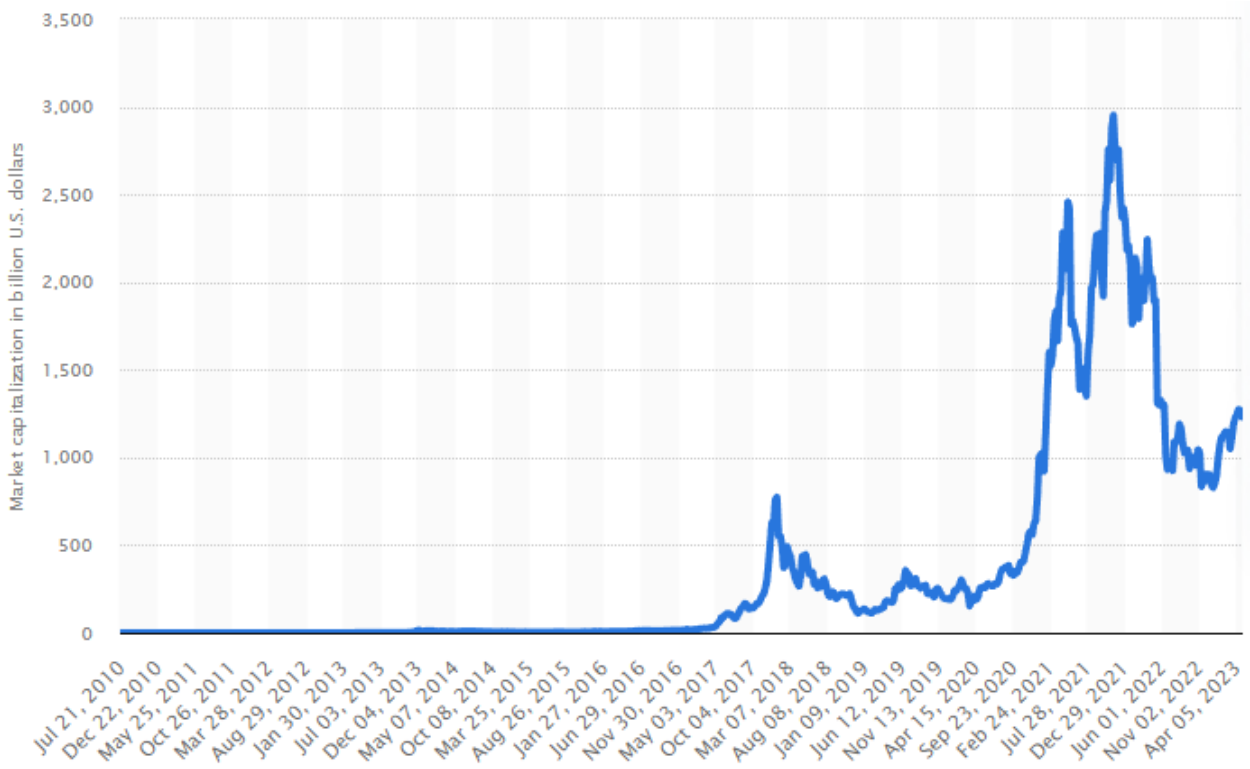
3.2. Η αγορά των Κρυπτονομισμάτων

Με τον όρο αγορά κρυπτονομισμάτων αναφερόμαστε στο συνολικό οικοσύστημα των κρυπτονομισμάτων και στις συναλλαγές τους. Το οικοσύστημα αυτό περιλαμβάνει όλες τις συνιστώσες της αγοράς, πώλησης και ανταλλαγής κρυπτονομισμάτων, καθώς και τους

συμμετέχοντες στην αγορά, όπως οι επενδυτές και τα ανταλλακτήρια κρυπτονομισμάτων, με όλα τα παραπάνω να λειτουργούν σε ένα παγκόσμιο αποκεντρωμένο πλαίσιο.

Σε αντίθεση με τις παθητικές χρηματοπιστωτικές αγορές, η αγορά κρυπτονομισμάτων είναι μια αποκεντρωμένη αγορά σε παγκόσμιο επίπεδο και ως εκ τούτου δεν υπάρχει κεντρική αρχή κεφαλαιαγοράς ή χρηματιστηριακός φορέας που να την ελέγχει. Αντίθετα, τα κρυπτονομίσματα διαπραγματεύονται σε μια ποικιλία διαφορετικών πλατφορμών και ανταλλακτηρίων, τόσο κεντρικών όσο και αποκεντρωμένων.

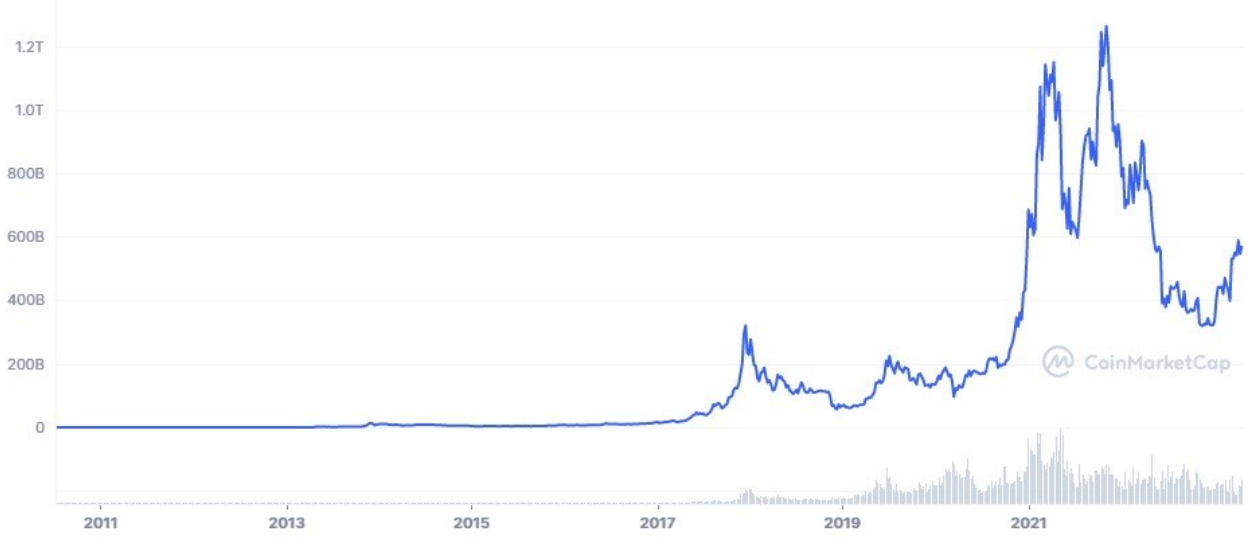
Η αγορά κρυπτονομισμάτων χαρακτηρίζεται από σημαντική μεταβλητότητα, με την κεφαλαιοποίηση της αγοράς αλλά και την αξία των μεμονωμένων κρυπτονομισμάτων να παρουσιάζουν έντονες διακυμάνσεις. Όπως απεικονίζεται παρακάτω στο Σχήμα 1, η συνολική κεφαλαιοποίηση της αγοράς όλων των κρυπτονομισμάτων έφτασε στο ανώτατο όριο των 3 τρισεκατομμυρίων δολαρίων ΗΠΑ τον Νοέμβριο του 2021, πριν μειωθεί σε περίπου 1.5 τρισεκατομμύρια δολάρια ΗΠΑ στις αρχές του 2023.



Σχήμα 1. Συνολική κεφαλαιοποίηση αγοράς κρυπτονομισμάτων. Ανάκτηση από Statista (n.d.). [27]

Από την εισαγωγή του Bitcoin το 2009, η αγορά κρυπτονομισμάτων έχει αυξηθεί σημαντικά, με το ίδιο το Bitcoin να παραμένει το μεγαλύτερο κρυπτονόμισμα με βάση την κεφαλαιοποίηση της αγοράς.

Στο Σχήμα 2 μπορούμε να δούμε ότι τον Νοέμβριο του 2021, η συνολική κεφαλαιοποίηση του Bitcoin ξεπέρασε τα 1.2 τρισεκατομμύρια δολάρια ΗΠΑ, καταδεικνύοντας τον σημαντικό αντίκτυπο και το ενδιαφέρον που έχει συγκεντρώσει. Από τις αρχές του 2023, η κεφαλαιοποίησή του συρρικνώθηκε σε περίπου 600 δισεκατομμύρια δολάρια ΗΠΑ, αντανακλώντας τη δυναμική και μεταβαλλόμενη φύση της αγοράς των κρυπτονομισμάτων.



Σχήμα 2. Κεφαλαιοποίηση του Bitcoin. Ανάκτηση από CoinMarketCap (n.d.). [28]

Η εξέλιξη του Bitcoin από μια καινοτόμο ιδέα σε μια κύρια οικονομική οντότητα υπογραμμίζει τη μοναδική του θέση στον χρηματοπιστωτικό τομέα. Η αποκεντρωμένη δομή του, η ανώτατη προσφορά που ισούται με 21 εκατομμύρια Bitcoin και η ανεξαρτησία του από τα παραδοσιακά χρηματοοικονομικά συστήματα και ρυθμιστικούς φορείς, παρουσιάζουν το Bitcoin τόσο ως μια νέα επενδυτική ευκαιρία όσο και ως ένα δυνητικό αντίβαρο στα συμβατικά χρηματοοικονομικά μέσα. Το αυξανόμενο ενδιαφέρον από θεσμικούς επενδυτές υπογραμμίζει περαιτέρω την αναδυόμενη νομιμοποίηση του Bitcoin και τις δυνατότητές του ώστε να καθιερωθεί ως μια νέα κατηγορία περιουσιακών στοιχείων εντός της ευρύτερης χρηματοοικονομικής αγοράς.

3.3. Η μεταβλητότητα της αγοράς του Bitcoin και οι ευκαιρίες της

Η μεταβλητότητα του Bitcoin είναι ένα από τα πιο αξιοσημείωτα χαρακτηριστικά του, συνδέοντας το έτσι με αυξημένους κινδύνους αλλά και υψηλές δυνητικές ανταμοιβές. Αυτή η έντονη μεταβλητότητα όχι μόνο τονίζει την κερδοσκοπική φύση του Bitcoin αλλά επίσης το τοποθετεί ως ένα ελκυστικό περιουσιακό στοιχείο τόσο για ιδιώτες όσο και για θεσμικούς επενδυτές που αναζητούν σημαντικές αποδόσεις. Η συνεχής λειτουργία της αγοράς κρυπτονομισμάτων ενισχύει περαιτέρω την καταλληλόλητα του Bitcoin για αυτοματοποιημένα συστήματα συναλλαγών, τα οποία είναι ειδικευμένα στην αξιοποίηση γρήγορων μεταβολών της αγοράς—κινήσεις που οι φυσικοί επενδυτές παραδοσιακά δυσκολεύονται να λειτουργήσουν αποτελεσματικά.

Το Bitcoin ως το κρυπτονόμισμα με το μεγαλύτερο μερίδιο αγοράς παρουσιάζει αυξημένη μεταβλητότητα και οι βασικοί παράγοντες που οδηγούν σε αυτή είναι:

- **Έλλειψη ρυθμιστικού πλαισίου**: Τα κρυπτονομίσματα εξακολουθούν να είναι σχετικά νέα και οι ρυθμιστικοί φορείς δεν έχουν ακόμη καλύψει τις ταχύρρυθμες εξελίξεις στον χώρο των κρυπτονομισμάτων. Αυτή η έλλειψη ρυθμιστικού πλαισίου μπορεί να δημιουργήσει αβεβαιότητα και αστάθεια, οδηγώντας σε διακυμάνσεις των τιμών. [8], [9]
- **Η έκθεση σε μέσα ενημέρωσης και εφαρμογές κοινωνικής δικτύωσης**: Οι ειδήσεις και η προβολή σε εφαρμογές κοινωνικής δικτύωσης μπορούν να επηρεάσουν σημαντικά τις τιμές των κρυπτονομισμάτων. Θετικά νέα, όπως μια νέα συνεργασία ή μια τεχνολογική ανακάλυψη, μπορεί να οδηγήσουν σε άνοδο της τιμής, ενώ τα αρνητικά νέα, όπως ένα περιστατικό hacking ή μια νέα κανονιστική συμμόρφωση, μπορεί να οδηγήσουν σε πτώση της τιμής. [10] [11]

- **Χειραγώγηση αγοράς**: Οι αγορές κρυπτονομισμάτων είναι σε μεγάλο βαθμό ανεξέλεγκτες, γεγονός που τις καθιστά πιο επιρρεπείς σε χειραγώγηση. Οι τιμές των κρυπτονομισμάτων μπορούν να επηρεαστούν εντόνως από αγοροπωλησίες μεγάλων ποσοτήτων ενός συγκεκριμένου κρυπτονομίσματος, κάτι που μπορεί να δημιουργήσει σημαντική αστάθεια στην αγορά. [12]
- **Περιορισμένη προσφορά**: Το Bitcoin, έχει περιορισμένη προσφορά με μόνο 21 εκατομμύρια να έχουν εξορυχθεί και να μην μπορούν να δημιουργηθούν άλλα. Καθώς η ζήτηση κυμαίνεται έναντι αυτής της σταθερής προσφοράς, η διακυμάνσεις της τιμής του είναι συχνά μεγάλες καθώς οι επενδυτές ανταγωνίζονται για την απόκτηση των περιορισμένων πόρων του. [10]

4. ΜΕΘΟΔΟΛΟΓΙΑ ΠΡΟΒΛΗΜΑΤΟΣ

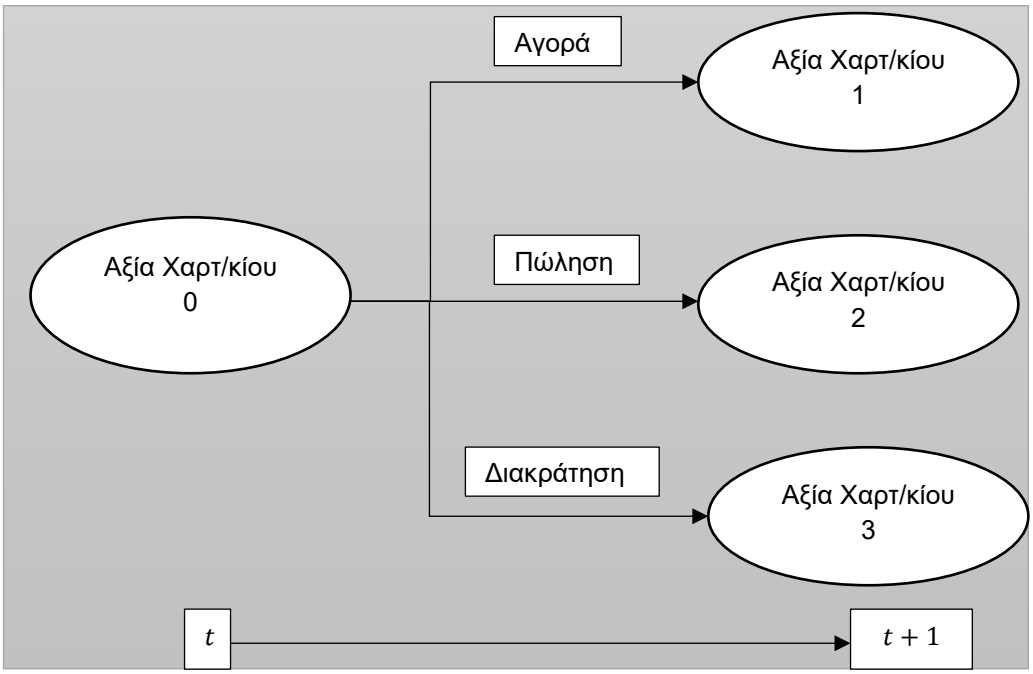
4.1. Περιγραφή Προβλήματος

Για την περιγραφή του προβλήματος ακολουθούμε την λογική παρόμοιων εργασιών όπως αναλύεται στα [1], [5] και [6]. Ειδικότερα, θα ακολουθήσουμε μια «Διαδικασία Απόφασης Markov» (Markov Decision Process - MDP) για τη διατύπωση του προβλήματος. Οι «Διαδικασίες Αποφάσεων Markov» είναι ένα μαθηματικό πλαίσιο που χρησιμοποιείται για τη μοντελοποίηση της λήψης αποφάσεων σε καταστάσεις όπου τα αποτελέσματα είναι εν μέρει τυχαία και εν μέρει υπό τον έλεγχο ενός υπεύθυνου λήψης αποφάσεων (του πράκτορα). Οι «Διαδικασίες Αποφάσεων Markov» παρέχουν μια διατύπωση του περιβάλλοντος λήψης αποφάσεων και είναι θεμελιώδεις για τη μελέτη της Ενισχυτικής Μάθησης.

Στην περίπτωσή μας υιοθετούμε την προσέγγιση ενός χαρτοφυλακίου με ένα μόνο περιουσιακό στοιχείο, δηλαδή το κρυπτονόμισμα Bitcoin. Μια «Διαδικασία Απόφασης Markov» χαρακτηρίζεται από το ακόλουθο σύνολο:

- Την **Κατάσταση** (State) $s = [p, h, b]$: ένα σύνολο που σχετίζεται με τις τιμές του $p \in \mathbb{R}_+^D$, την ποσότητα των Bitcoin που κατέχονται $h \in \mathbb{Z}_+^D$, και το υπόλοιπο του λογαριασμού $b \in \mathbb{R}_+$, όπου $D = 1$ υποδηλώνει το μοναδικό περιουσιακό στοιχείο δηλαδή το Bitcoin και $\mathbb{Z}_+$ υποδηλώνει τους μη αρνητικούς δεκαδικούς αριθμούς.
- Την **Ενέργεια** (Action) a: ένα σύνολο αποφάσεων που εφαρμόζεται σε όλο το D (Bitcoin), περιλαμβάνοντας τις ενέργειες πώλησης, "Αγοράς και Διακράτησης». Αυτές οι ενέργειες αντιστοιχούν σε μείωση, αύξηση ή καμία αλλαγή στα κατεχόμενα h, αντίστοιχα.
- Την **Ανταμοιβή** (Reward) $r(s, a, s')$: η άμεση ανταμοιβή που λαμβάνεται εκτελώντας την ενέργεια a στην κατάσταση s και μεταβαίνοντας στη νέα κατάσταση s'.
- Η **Πολιτική** (Policy) $\pi(s)$: η στρατηγική εκτέλεσης συναλλαγών στην κατάσταση s, ορισμένη ως η κατανομή πιθανότητας επί των ενεργειών a εντός της κατάσταση s.
- **Συνάρτηση Q-value** (Action-value function) - $Q_\pi(s, a)$: η αναμενόμενη ανταμοιβή που λαμβάνεται εκτελώντας την ενέργεια a στην κατάσταση s ενώ τηρείται η πολιτική π.

Η μετάβαση κατάστασης της διαδικασίας συναλλαγών του Bitcoin απεικονίζεται στο σχήμα 3:



Σχήμα 3. Οι προκύπτουσες αξίες του χαρτοφυλακίου μετά την επιλογή μιας επενδυτικής ενέργειας.

Σε κάθε κατάσταση, οι ακόλουθες ενέργειες είναι διαθέσιμες για να ληφθούν στο Bitcoin D,

- Πώληση: k ($k \in [1, h[D]]$, where $D = 1$) Bitcoin μπορούν να πωληθούν από τα τρέχοντα κατεχόμενα, όπου k είναι δεκαδικός αριθμός (64-bit ακέραιοι). Ως αποτέλεσμα, $h_{t+1} = h_t - k$.
- Διακράτηση: $k = 0$ και αυτό οδηγεί σε αλλαγή σε h_t.
- Αγορά: k Bitcoin μπορούν να αγοραστούν με αποτέλεσμα $h_{t+1} = h_t + k$.

Όσον αφορά την αξία του χαρτοφυλακίου, η οποία προσδιορίζεται από την σχέση $\boldsymbol{p^T h + b}$, όταν μια ενέργεια εκτελείται στον χρόνο t, και εν συνεχεία οι τιμές του Bitcoin ενημερώνονται στην $t+1$, οι αξίες του χαρτοφυλακίου μπορούν να προσαρμοστούν είτε στην "Αξία χαρτοφυλακίου 1", είτε στην "Αξία χαρτοφυλακίου 2" ή στην "Αξία χαρτοφυλακίου 3", αντιστοιχώντας στα αποτελέσματα που απεικονίζονται στο Σχήμα 3.

4.2. Περιορισμοί Προβλήματος

Ακολουθώντας τις μεθοδολογίες των ερευνών [1],[5] και επεκτείνοντας την ανάλυσή μας με την ενσωμάτωση επιπλέον παραμέτρων της αγοράς, έχουμε καθορίσει τους ακόλουθους περιορισμούς για το περιβάλλον συναλλαγών μας:

1. Περιορισμοί της Αγοράς των Κρυπτονομισμάτων:
    a. Οι εντολές Αγοράς και Πώλησης εκτελούνται στην τιμή κλεισίματος του τρέχοντος διαστήματος συναλλαγών για να διατηρηθεί η απλότητα και η προβλεψιμότητα.
    b. Υποθέτουμε ότι όλες οι συναλλαγές εκτελούνται στην αναμενόμενη τιμή, ανεπηρέαστες από το μέγεθος των εντολών ή τη ρευστότητα της αγοράς.
    c. Η εφαρμογή συναλλακτικής δραστηριότητας μέσω των αλγορίθμων Βαθιάς Ενισχυτικής Μάθησης δεν θα μπορούσε να επηρεάσει την αγορά κρυπτονομισμάτων.
2. Κόστη Συναλλαγών: Κάθε συναλλαγή, είτε αγορά είτε πώληση, συνεπάγεται ένα κόστος συναλλαγής ίσο με 0.1% της αξίας της συναλλαγής. Το συγκεκριμένο κόστος αποτυπώνεται

ως, $\boldsymbol{c_t = p^T k_t \times 0.1\%}$, εξασφαλίζοντας ότι η προσομοίωση αντικατοπτρίζει ρεαλιστικά σενάρια συναλλαγών.

3. Μόχλευση: Για την απλούστευση του μοντέλου και την εστίαση στα άμεσα αποτελέσματα των αποφάσεων συναλλαγών, η μόχλευση απαγορεύεται στο περιβάλλον συναλλαγών μας.
4. Περιορισμοί Μη-Αρνητικού Υπολοίπου:
    a. Αγορά: Ο πράκτορας είναι προγραμματισμένος να αγοράζει τη μέγιστη δυνατή ποσότητα Bitcoin, υπό την προϋπόθεση ότι αυτό δεν οδηγεί σε αρνητικό υπόλοιπο, βελτιστοποιώντας επιθετικές επενδυτικές στρατηγικές εντός των περιορισμών του προϋπολογισμού.
    b. Πώληση: Ο πράκτορας θα ρευστοποιήσει το σύνολο των διαθεσίμων Bitcoin όταν ληφθεί μια απόφαση πώλησης. Μια δικλείδα ασφαλείας εξασφαλίζει ότι το υπόλοιπο του πράκτορα παραμένει μη αρνητικό μετά τη συναλλαγή, λαμβάνοντας υπόψη τα πραγματοποιηθέντα κόστη.
5. Αρχικό υπόλοιπο: Κάθε προσομοίωση του μοντέλου μας ξεκινά με τον πράκτορα να διαθέτει ένα αρχικό υπόλοιπο 10,000 δολαρίων, τυποποιώντας το σημείο εκκίνησης για την αξιολόγηση της απόδοσης.

4.3. Στόχος της Συναλλακτικής Δραστηριότητας

Ο πρωταρχικός στόχος της ανάπτυξης πρακτόρων BEM στην προσομοίωση συναλλαγών Bitcoin είναι η ανάπτυξη και εφαρμογή αυτόνομων στρατηγικών συναλλαγών ικανών να πλοηγηθούν στην ασταθή και με έντονη μεταβλητότητα αγορά των κρυπτονομισμάτων. Αυτός ο στόχος επιδιώκεται με την πρόθεση μεγιστοποίησης της συνολικής αξίας των διαθεσίμων του πράκτορα, τα οποία περιλαμβάνουν τα μετρητά υπόλοιπα και την αξία οποιουδήποτε Bitcoin που κατέχει. Οι στρατηγικές αποφάσεις συναλλαγών λαμβάνονται με την προσδοκία κινήσεων της αγοράς, με στόχο την αγορά Bitcoin όταν οι τιμές αναμένεται να αυξηθούν και την πώληση όταν οι τιμές αναμένεται να μειωθούν, ενώ λαμβάνονται υπόψη με μεγάλη προσοχή τα κόστη συναλλαγών που επηρεάζουν άμεσα την καθαρή ανταμοιβή. Οι στρατηγικές των πρακτόρων σχεδιάζονται όχι μόνο για να ανταποκρίνονται στις άμεσες συνθήκες της αγοράς αλλά και για να βελτιστοποιούν τη μακροπρόθεσμη ανάπτυξη του χαρτοφυλακίου, ενσωματώνοντας παράγοντες για τα κόστη συναλλαγών και τη διαχείριση κινδύνου για να εξασφαλίσουν βιώσιμη κερδοφορία.

Η συνάρτηση ανταμοιβής έχει σχεδιαστεί γύρω από τις αλλαγές στην αξία του χαρτοφυλακίου λόγω των ενεργειών του πράκτορα (αγορά, διακράτηση, πώληση) και τις προκύπτουσες μεταβάσεις κατάστασης. Ορίζεται ως [1]:

$$r(s_t, a_t, s_{t+1}) = (b_{t+1} + p_{t+1}^T h_{t+1}) - (b_t + p_t^T h_t) - c_t$$

Όπου:

- $(b_{t+1} + p_{t+1}^T h_{t+1})$ αντιπροσωπεύει την αξία του χαρτοφυλακίου στο $t+1$,
- $(b_t + p_t^T h_t)$ υποδηλώνει την αξία του χαρτοφυλακίου στο t,
- και c_t είναι το κόστος συναλλαγής.

Οι φάσεις του μοντέλου μπορούν να περιγράφουν ως εξής:

1. **Αρχικοποίηση:** Τη χρονική στιγμή 0, ορίζονται η τιμή του Bitcoin p_0, το αρχικό υπόλοιπο λογαριασμού b_0, και τα κατεχόμενα Bitcoin $h = 0$. Η συνάρτηση Q-value για κάθε ζεύγος «κατάστασης-ενέργειας» $Q_\pi(s, a)$ ξεκινά από 0, με την πολιτική $\pi(s)$ να κατανέμει τις ενέργειες ομοιόμορφα σε όλες τις καταστάσεις.
2. **Βελτιστοποίηση:** Μέσω της αλληλεπίδρασης με την αγορά του Bitcoin, η Q-value για το ζεύγος «κατάστασης-ενέργειας», $Q_\pi(s, a)$, ενημερώνεται για να βελτιώσει τη στρατηγική συναλλαγών στο περιβάλλον του Bitcoin. Η βέλτιστη στρατηγική καθορίζεται με τη χρήση της εξίσωσης **Bellman**. Δηλαδή, η αναμενόμενη ανταμοιβή για τη λήψη της ενέργειας a_t στην κατάσταση s_t υπολογίζεται

ως το άθροισμα της άμεσης ανταμοιβής $r(s_t, a_t, s_{t+1})$ και της μελλοντικής ανταμοιβής από την επόμενη κατάσταση s_{t+1}, μειωμένη κατά έναν παράγοντα γ:

$$Q_\pi(s_t, a_t) = E_{s_{t+1}}[r(s_t, a_t, s_{t+1}) + \gamma E_{a_{t+1} \sim \pi(s_{t+1})}[Q_\pi(s_{t+1}, a_{t+1})]]$$

Ο στόχος είναι να αυξηθεί η καθαρή αξία του χαρτοφυλακίου μέσω ενημερωμένων αποφάσεων συναλλαγών που διευκολύνονται από τις τεχνικές BEM.

5. ΔΗΜΙΟΥΡΓΙΑ ΠΕΡΙΒΑΛΛΟΝΤΟΣ ΓΙΑ ΣΥΝΑΛΛΑΓΕΣ ΣΕ BITCOIN

Στην ανάπτυξη ενός μοντέλου Βαθιάς Ενισχυτικής Μάθησης (ΒΕΜ), προσαρμοσμένου για συναλλαγές σε Bitcoin, η δημιουργία ενός περιβάλλοντος συναλλαγών είναι πρωταρχικής σημασίας. Αυτή η ενότητα περιγράφει τον σχεδιασμό και την υλοποίηση ενός προσαρμοσμένου περιβάλλοντος προσομοίωσης συναλλαγών, το οποίο διαδραματίζει κρίσιμο ρόλο στη διευκόλυνση της αλληλεπίδρασης του πράκτορα με τα δεδομένα της αγοράς και την χάραξη στρατηγικών για συναλλαγές.

Αυτό το περιβάλλον, έχει δημιουργηθεί ως υποκλάση του gym.Env της OpenAI. Η κλάση TradingEnv βρίσκεται στον πυρήνα της αλληλεπίδρασης μεταξύ των πρακτόρων της ΒΕΜ και της αγοράς Bitcoin. Αυτό το περιβάλλον δεν προσομοιώνει μόνο τις πραγματικές συνθήκες συναλλαγών αλλά επιπλέον διαμορφώνει την πορεία «εκμάθησης» των πρακτόρων ορίζοντας τον χώρο καταστάσεων, τον χώρο ενέργειων και τον μηχανισμό ανταμοιβής.

Ο πυρήνας της διαδικασίας εκμάθησης των πρακτόρων της ΒΕΜ μέσα στο περιβάλλον συναλλαγών αποτελείται από τρεις βασικές λειτουργίες:

1. **Χώρος Καταστάσεων**: Ορίζει το σύνολο όλων των δυνατών καταστάσεων που μπορεί να συναντήσει ο πράκτορας στην αγορά του Bitcoin, περιλαμβάνοντας κινήσεις τιμών, δείκτες αγοράς και την κατάσταση του χαρτοφυλακίου.
2. **Χώρος Ενεργειών**: Προσδιορίζει τις δράσεις που είναι διαθέσιμες στον πράκτορα, συμπεριλαμβανομένων της αγοράς, διακράτησης και πώλησης Bitcoin, οι οποίες επηρεάζουν τη σύνθεση και την αξία του χαρτοφυλακίου.
3. **Συνάρτηση Ανταμοιβής**: Καθορίζει το ανατροφοδότηση που λαμβάνει ο πράκτορας για τις ενέργειές του, επηρεάζοντας άμεσα την ανάπτυξη της στρατηγικής, ποσοτικοποιώντας την επιτυχία των συναλλαγών σε όρους αλλαγών στην αξία του χαρτοφυλακίου.

Αυτά τα στοιχεία συλλογικά διασφαλίζουν ότι οι πράκτορες της ΒΕΜ εκπαιδεύονται σε ένα ολοκληρωμένο και ρεαλιστικό περιβάλλον, επιτρέποντας τη διαμόρφωση και βελτίωση αποτελεσματικών στρατηγικών συναλλαγών.

5.1. Χώρος Καταστάσεων

Στην “TradingEnv”, ο χώρος καταστάσεων αντιπροσωπεύεται από ένα 7-διάστατο διάνυσμα, εξοπλίζοντας τον πράκτορα με μια ολοκληρωμένη άποψη τόσο των συνθηκών της αγοράς όσο και της κατάστασης του χαρτοφυλακίου [1]. Κάθε συνιστώσα αυτού του διανύσματος προσφέρει ουσιαστικές πληροφορίες, καθοδηγώντας τη στρατηγική του πράκτορα για να βελτιστοποιήσει τις ανταμοιβές του:

1. $p_t \in \mathbb{R}_+^1$: Η τελευταία τιμή κλεισίματος του Bitcoin, η οποία και παρέχει άμεση εικόνα της τρέχουσας αγοραίας αξίας.
2. $b_t \in \mathbb{R}_+$: Το τρέχον ταμειακό υπόλοιπο (σε μετρητά) του πράκτορα, που δείχνει το διαθέσιμο κεφάλαιο του για συναλλαγές.
3. $h_t \in \mathbb{Z}_+^1$: Η ποσότητα των Bitcoin που κατέχονται, αντικατοπτρίζοντας την τρέχουσα επένδυση του πράκτορα στην αγορά.
4. $M_t \in \mathbb{R}^{30}$: Ο δείκτης “Moving Average Convergence Divergence” (MACD), ένας δείκτης τάσης (momentum) που ακολουθεί την τάση και αποτυπώνει τη σχέση μεταξύ δύο κινητών μέσων όρων της τιμής του Bitcoin, βοηθώντας στην ανάλυση τάσης.
5. $R_t \in \mathbb{R}_+^1$: Ο δείκτης “Relative Strength Index” (RSI), είναι ένας ταλαντωτής τάσης που μετράει την ταχύτητα αλλαγής των τιμών. Βοηθά στον εντοπισμό υπερ-αγορασμένων ή υπερ-πουλημένων συνθηκών μετρώντας την ταχύτητα και το μέγεθος των κινήσεων τιμών.
6. $C_t \in \mathbb{R}_+^1$: Ο δείκτης “Commodity Channel Index” (CCI), ένας ταλαντωτής που χρησιμοποιείται στην τεχνική ανάλυση για να προσδιορίσει πότε ένα επενδυτικό προϊόν μπορεί να βρίσκεται

σε κατάσταση υπερ-αγοράς ή υπερ-πώλησης. Είναι επίσης χρήσιμος για την αξιολόγηση της ισχύος και της κατεύθυνσης των τάσεων τιμών.

7. $X_t \in \mathbb{R}_+^1$: Ο δείκτης "Average Directional Index" (ADX), ένας δείκτης που αξιολογεί τη δύναμη μιας τάσης τιμής χωρίς να λαμβάνει υπόψη την κατεύθυνσή της, αυξάνεται σε αξία κατά τη διάρκεια ανοδικών και καθοδικών τάσεων.

Οι παράμετροι αυτοί ενημερώνουν συλλογικά τις αποφάσεις του πράκτορα, δίνοντάς του τη δυνατότητα να επιλέξει ενέργειες που στοχεύουν στη μεγιστοποίηση της αξίας του χαρτοφυλακίου του μέσω στρατηγικών αγοράς, διακράτησης ή πώλησης Bitcoin με βάση την εξελισσόμενη δυναμική της αγοράς.

5.2. Χώρος Ενεργειών

Ο χώρος ενεργειών οριοθετεί τις πιθανές ενέργειες που μπορεί να αναλάβει ο πράκτορας σε κάθε χρονικό βήμα, κατηγοριοποιημένες ως εξής:

- **Διακράτηση (Ενέργεια = 0)**: Ο πράκτορας επιλέγει να μην συναλλαχθεί, πιστεύοντας ότι τα τρέχοντα κατεχόμενα Bitcoin του είναι πιο επωφελή από οποιαδήποτε πιθανή συναλλαγή. Αυτή η απόφαση προκύπτει συνήθως από μια πρόβλεψη ότι καμία άμεση κίνηση της αγοράς δεν δικαιολογεί μια συναλλαγή.
- **Αγορά (Ενέργεια = 1)**: Ο πράκτορας αποφασίζει να αποκτήσει Bitcoin, υπολογίζοντας τη μέγιστη ποσότητα που μπορεί να αγοράσει με το διαθέσιμο υπόλοιπό του αφού ληφθεί υπόψη το κόστος συναλλαγής. Μια εντολή αγοράς προχωρά μόνο αν η αγορά δεν οδηγεί σε αρνητικό υπόλοιπο, υποδεικνύοντας τις αισιόδοξες προοπτικές της αγοράς του πράκτορα και την προσδοκία μελλοντικών αυξήσεων της τιμής για κέρδος.
- **Πώληση (Ενέργεια = 2)**: Ο πράκτορας επιλέγει να πουλήσει όλα τα κατεχόμενα Bitcoin, με τα συνολικά έσοδα να προκύπτουν αφού αφαιρεθούν τα κόστη συναλλαγής. Μια πώληση πραγματοποιείται μόνο αν διατηρείται ένα μη αρνητικό υπόλοιπο, αποτρέποντας καταστάσεις όπου το κόστος πώλησης υπερβαίνει τα διαθέσιμα ρευστά διαθέσιμα. Η ενέργεια της πώλησης πραγματοποιείται συνήθως για να κλειδωθούν τα κέρδη στα υψηλά της αγοράς ή για να μετριαστούν οι ζημίες εν αναμονή μιας πτωτικής τάσης.

Οι παραπάνω ενέργειες επηρεάζουν το χαρτοφυλάκιο του πράκτορα μιμούμενες τις αποφάσεις συναλλαγών όπως και σε ένα πραγματικό επενδυτικό περιβάλλον εκτέλεσης συναλλαγών, ενώ υπάρχει άμεση σύνδεση με τους περιορισμούς του προβλήματος που περιγράφονται στην Ενότητα 4.2. Ο χώρος δράσης διαμορφώνεται από διάφορους κρίσιμους παράγοντες:

- **Κόστη Συναλλαγών**: Αναπόσπαστο στοιχείο τόσο για την αγορά όσο και για την πώληση, τα κόστη συναλλαγών επηρεάζουν το καθαρό υπόλοιπο, ενσωματώνοντας μια ρεαλιστική προσομοίωση συναλλαγών.
- **Οικονομική Επάρκεια**: Οι ενέργειες αγοράς και πώλησης έχουν σχεδιαστεί έτσι ώστε να διατηρούν το υπόλοιπο του πράκτορα μη αρνητικό, τηρώντας τις θεμελιώδεις αρχές φερεγγυότητας στις χρηματοοικονομικές συναλλαγές.
- **Μεγιστοποίηση των Αγορών Bitcoin**: Ο πράκτορας στοχεύει να χρησιμοποιήσει το πλήρες υπόλοιπό του στις εντολές αγοράς Bitcoin, βελτιστοποιώντας τις μέγιστες πιθανές αποδόσεις.
- **Πλήρης Ρευστοποίηση κατά την Πώληση**: Η εντολή πώλησης συνεπάγεται πλήρη ρευστοποίηση των συμμετοχών Bitcoin, μια στρατηγική που απλοποιεί τη λήψη αποφάσεων και παράλληλα περιορίζει επενδυτικές στρατηγικές όπως οι μερικές πωλήσεις.
- **Διακριτές Ενέργειες**: Το μοντέλο απλοποιεί την εκμάθηση της πολιτικής του πράκτορα προσφέροντας διακριτά μεγέθη συναλλαγών – ο πράκτορας είτε πραγματοποιεί συναλλαγές με το μέγιστο δυνατό ποσό είτε δεν συναλλάσσεται καθόλου.

5.3. Συνάρτηση Ανταμοιβής

Η συνάρτηση ανταμοιβής διαδραματίζει έναν αναντικατάστατο ρόλο στην καθοδήγηση της διαδικασίας μάθησης του πράκτορα εντός της “TradingEnv” στον κώδικα που έχουμε υλοποιήσει. Αξιολογεί τις ενέργειες του πράκτορα με βάση την καθαρή αλλαγή στη συνολική αξία ενεργητικού του χαρτοφυλακίου μετά από κάθε ενέργεια, σε σχέση με την αρχή του κάθε «επεισοδίου» συναλλακτικής δραστηριότητας.

Η ανταμοιβή υπολογίζεται ως:

- *Το άθροισμα του τρέχοντος υπολοίπου μετρητών και της αξίας των διαθεσίμων Bitcoin μείον το αρχικό υπόλοιπο*.

Αυτός ο υπολογισμός αντικατοπτρίζει την καθαρή αλλαγή στην αξία του ενεργητικού ως αποτέλεσμα των ενεργειών του πράκτορα εντός ενός επεισοδίου.

- Μια **θετική ανταμοιβή** υποδηλώνει αύξηση στην αξία του χαρτοφυλακίου λόγω κερδοφόρων συναλλαγών, δίνοντας σήμα στον πράκτορα να εντοπίσει και να υιοθετήσει παρόμοιες στρατηγικές σε μελλοντικά σενάρια.
- Αντίθετα, μια **αρνητική ανταμοιβή** αντικατοπτρίζει μια μείωση στην αξία του χαρτοφυλακίου, υποδεικνύοντας ενέργειες που προκαλούν ζημίες και τις οποίες ο πράκτορας θα πρέπει να μάθει να αποφεύγει.

Το ιδιοχαρακτηριστικό, εντός του κώδικά μας, “self.current_reward” αποθηκεύει την υπολογισμένη ανταμοιβή μετά από κάθε βήμα, και αυτή η τιμή είναι αυτή που χρησιμοποιεί ο πράκτορας για να μάθει και να προσαρμόσει την πολιτική του. Ο στόχος του πράκτορα είναι να μεγιστοποιήσει τη σωρευτική ανταμοιβή κατά τη διάρκεια ενός επεισοδίου, το οποίο ευθυγραμμίζεται με τον στόχο μεγιστοποίησης της χρηματοοικονομικής απόδοσης από τις δραστηριότητες των συναλλαγών.

6. ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ – ΠΡΟΕΤΟΙΜΑΣΙΑ – ΚΑΝΟΝΙΚΟΠΟΙΗΣΗ

6.1. Συλλογή δεδομένων

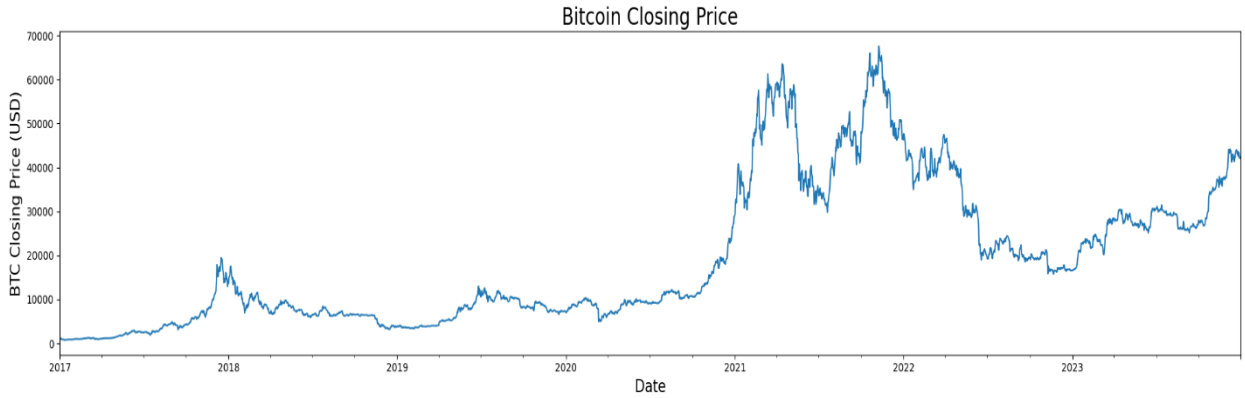
Τα δεδομένα για το Bitcoin αντλήθηκαν από το Yahoo Finance, χρησιμοποιώντας τη βιβλιοθήκη "yfinance" για πρόσβαση σε ιστορικά δεδομένα αγοράς μέσω του API του Yahoo Finance. Το σύνολο δεδομένων μας περιλαμβάνει συγκεκριμένα ιστορικές ημερήσιες τιμές Bitcoin (BTC-USD) για το διάστημα «01/01/2017 – 01/01/2024». Αυτή η επιλογή στοχεύει να καλύψει ένα ευρύ φάσμα συμπεριφορών της αγοράς του κρυπτονομίσματος, αποτυπώνοντας περιόδους ταχείας ανάπτυξης, σημαντικές διορθώσεις αγοράς και φάσεις σταθεροποίησης, παρέχοντας έτσι ένα πλούσιο σύνολο δεδομένων για ανάλυση.

Το σύνολο δεδομένων είναι δομημένο σε ένα πλαίσιο δεδομένων με 2.556 γραμμές, που αντιπροσωπεύουν τις ημέρες και 6 στήλες, που περιγράφουν λεπτομερώς τα ημερήσια στοιχεία συναλλαγών: Άνοιγμα, Υψηλό, Χαμηλό, Κλείσιμο, Προσαρμοσμένο Κλείσιμο και Όγκος. Ένα απόσπασμα αυτού του πλαισίου δεδομένων παρουσιάζεται στον Πίνακα 1.

Date	Open	High	Low	Close	Adj Close	Volume
2017-01-01	963.65802	1003.080017	958.698975	998.325012	998.325012	147775008
2017-01-02	998.617004	1031.390015	996.702026	1021.75	1021.75	222184992
2017-01-03	1021.599976	1044.079956	1021.599976	1043.839966	1043.839966	185168000
2017-01-04	1044.400024	1159.420044	1044.400024	1154.72998	1154.72998	344945984
2017-01-05	1156.72998	1191.099976	910.416992	1013.380005	1013.380005	510199008

Πίνακας 1: Δείγμα δεδομένων

Για τους σκοπούς της παρούσας εργασίας θα εστιάσουμε κυρίως στις ημερήσιες τιμές κλεισίματος του Bitcoin, οι οποίες λειτουργούν ως πρωταρχικός δείκτης για την ημερήσια αποτίμηση της αγοράς. Το Σχήμα 4 απεικονίζει γραφικά τις διακυμάνσεις του BTC- εντός του καθορισμένου συνόλου δεδομένων, απεικονίζοντας τη μεταβλητότητα και τις τάσεις της αγοράς κατά τη διάρκεια της περιόδου μελέτης.



Σχήμα 4: Τιμή Κλεισίματος BTC (USD)

6.2. Προετοιμασία Δεδομένων – Κανονικοποίηση

6.2.1. Χειρισμός τιμών που λείπουν και ταξινόμηση δεδομένων

Η αρχική φάση της προεπεξεργασίας δεδομένων επικεντρώθηκε στην αντιμετώπιση των τιμών που λείπουν, ένα βήμα ουσιαστικό για τη διατήρηση της ακεραιότητας και της αξιοπιστίας του συνόλου δεδομένων. Εξετάσαμε προσεκτικά το σύνολο δεδομένων για τιμές “NaN”, οι οποίες θα μπορούσαν να παραποιήσουν την ανάλυσή μας, και προετοιμαστήκαμε προκειμένου να είμαστε σε θέση να αφαιρέσουμε οποιεσδήποτε γραμμές περιείχαν τέτοιες ασυνέπειες. Ένας συστηματικός έλεγχος επιβεβαίωσε την απουσία τιμών “NaN” σε όλες τις στήλες, όπως απεικονίζεται στον Πίνακα 2. Αυτή η διαδικασία διασφαλίζει την «καθαρότητα» του συνόλου δεδομένων, με ένα πρωτόκολλο που έχει καθιερωθεί για την αφαίρεση οποιωνδήποτε μελλοντικών περιπτώσεων τιμών “NaN”:

Αριθμός τιμών NaN σε κάθε στήλη:	
Open	0
High	0
Low	0
Close	0
Adj Close	0
Volume	0

Πίνακας 2: Αριθμός τιμών NaN στο σύνολο δεδομένων

Στη συνέχεια, ορίσαμε τη στήλη “Date” ως ευρετήριο (index) των δεδομένων μας, παρέχοντας μια φυσική δομή χρονοσειρών στα δεδομένα. Αυτά τα ευρετηριασμένα δεδομένα ταξινομήθηκαν στη συνέχεια σε χρονολογική σειρά, ένα κρίσιμο βήμα για τη διατήρηση της ακολουθιακής συνέπειας των δεδομένων, ζωτικής σημασίας για την ανάλυση χρονοσειρών. Ο Πίνακας 3 παρέχει μια επισκόπηση του συνόλου δεδομένων μετά την ευρετηρίαση και την ταξινόμηση:

Date	Open	High	Low	Close	Adj Close	Volume
2017-01-01	963.65802	1003.080017	958.698975	998.325012	998.325012	147775008
2017-01-02	998.617004	1031.390015	996.702026	1021.75	1021.75	222184992
2017-01-03	1021.599976	1044.079956	1021.599976	1043.839966	1043.839966	185168000
2017-01-04	1044.400024	1159.420044	1044.400024	1154.72998	1154.72998	344945984
…	…	…	…	…	…	…
2023-12-28	43468.19922	43804.78125	42318.55078	42627.85547	42627.85547	22992093014
2023-12-29	42614.64453	43124.32422	41424.0625	42099.40234	42099.40234	26000021055
2023-12-30	42091.75391	42584.125	41556.22656	42156.90234	42156.90234	16013925945
2023-12-31	42152.09766	42860.9375	41998.25391	42265.1875	42265.1875	16397498810

Πίνακας 3: Δεδομένα μετά την ευρετηρίαση και την καθοδική ταξινόμηση

6.2.2. Διαχωρισμός δεδομένων

Το σύνολο δεδομένων διαχωρίστηκε σε δύο διακριτά τμήματα:

1. σετ εκπαίδευσης και
2. σετ δοκιμών.

Το συγκεκριμένο βήμα πραγματοποιήθηκε για την αποτελεσματική επικύρωση του μοντέλου και την αποφυγή του κινδύνου υπερμοντελοποίησης (overfitting).

- **Σετ Εκπαίδευσης**: Εκτείνεται από την 1η Ιανουαρίου 2017 έως τις 30 Δεκεμβρίου 2021, αυτό το τμήμα διευκόλυνε την αρχική φάση εκπαίδευσης του μοντέλου. Περιλαμβάνει 1.825 γραμμές δεδομένων, αντικατοπτρίζοντας μια ολοκληρωμένη περίοδο που προορίζεται να εκθέσει το μοντέλο σε διάφορες συνθήκες αγοράς.
- **Σετ Δοκιμών**: Σχεδιασμένο για την αξιολόγηση του μοντέλου, αυτό το σετ καλύπτει την περίοδο από τις 31 Δεκεμβρίου 2021 έως τις 31 Δεκεμβρίου 2023, περιλαμβάνοντας 781 γραμμές. Αυτή η μεταγενέστερη φάση επιτρέπει την αξιολόγηση των προβλεπτικών ικανοτήτων του μοντέλου σε δεδομένα που δεν έχουν προηγουμένως ελεγχθεί.

Η στρατηγική κατανομή των μη-επικαλυπτόμενων διαστημάτων για την εκπαίδευση και τις δοκιμές στηρίζει την ικανότητα του μοντέλου για γενίκευση σε νέα δεδομένα, ευθυγραμμιζόμενη με τις θεμελιώδεις πρακτικές της μηχανικής μάθησης. Το Σχήμα 5 αποτυπώνει οπτικά αυτές τις δύο περιόδους, απεικονίζοντας τη διαίρεση του συνόλου δεδομένων και το χρονικό εύρος που καλύπτει κάθε σετ.

Σχήμα 5. Σύγκριση των Σετ Δεδομένων Εκπαίδευσης και Δοκιμών που Χρησιμοποιήθηκαν στη Μεθοδολογία

6.2.3. Υπολογισμός Δεικτών Τεχνικής Ανάλυσης

Για να εμπλουτίσουμε το σύνολο δεδομένων μας με πολύτιμες πληροφορίες για τη στρατηγική ανάλυση, ενσωματώσαμε αρκετούς δείκτες τεχνικής ανάλυσης ευρέως αναγνωρισμένους για τη χρησιμότητά τους στις συναλλαγές:

- **MACD (Moving Average Convergence Divergence)**
- **RSI (Relative Strength Index)**
- **CCI (Commodity Channel Index)**
- **ADX (Average Directional Index)**

Χρησιμοποιώντας στον κώδικά μας τη βιβλιοθήκη “talib”, υπολογίσαμε αυτούς τους δείκτες για να αποτυπώσουμε διάφορες δυναμικές της αγοράς όπως την κατεύθυνση της τάσης και τη μεταβλητότητα. Αυτό το εμπλουτισμένο αναλυτικό υπόβαθρο υποστηρίζει τη διαμόρφωση διαφοροποιημένων στρατηγικών συναλλαγών. Σε ευθυγράμμιση με το μεθοδολογικό πλαίσιο που προτάθηκε από τους Yang, H., Liu, X. Y., Zhong, S., & Walid, A. (2020) [1], αυτοί οι δείκτες ενσωματώθηκαν απρόσκοπτα στο περιβάλλον συναλλαγών μας, με αποτέλεσμα την προσθήκη τεσσάρων στηλών στα σύνολα δεδομένων εκπαίδευσης και δοκιμής.

Αυτή η υπολογιστική διαδικασία πραγματοποιήθηκε στα πρωτογενή και ανεπεξέργαστα δεδομένα πριν από οποιαδήποτε κλιμάκωση (scaling) για να διατηρηθεί η αυθεντικότητα και η συνάφεια των δεικτών τεχνικής ανάλυσης σε διαφορετικές συνθήκες αγοράς. Κατά συνέπεια, αυτή η

προσέγγιση διασφαλίζει ότι οι δείκτες αντιπροσωπεύουν με ακρίβεια τις υποκείμενες δυναμικές της αγοράς κατά τη διάρκεια των φάσεων εκπαίδευσης και αξιολόγησης του μοντέλου.

6.2.4. Κλιμάκωση (Scaling) των Δεδομένων

Σε συνέχεια της ενσωμάτωσης των τεχνικών δεικτών στα ανεπεξέργαστα σετ δεδομένων, η διαδικασία της «κλιμάκωσης» (data scaling) αποτελεί το επόμενο κρίσιμο βήμα στην διαδικασία της προεπεξεργασίας. Αυτή η διαδικασία τυποποιεί το εύρος των ανεξάρτητων μεταβλητών ή χαρακτηριστικών, το οποίο είναι ζωτικής σημασίας για τα μοντέλα νευρωνικών δικτύων, διασφαλίζοντας ότι κάθε χαρακτηριστικό επηρεάζει εξίσου τα μαθησιακά αποτελέσματα του μοντέλου. Συγκεκριμένα, στην ανάλυσή μας, η κλιμάκωση διευκολύνει την ομοιόμορφη συνεισφορά όλων των χαρακτηριστικών στον αλγόριθμο μάθησης.

Στην υλοποίηση μας χρησιμοποιήσαμε την τεχνική “MinMaxScaler” από το module “sklearn.preprocessing”, ένα ευρέως αναγνωρισμένο εργαλείο για την προσαρμογή των κλιμάκων των χαρακτηριστικών σε ένα συγκεκριμένο εύρος, επιλεγμένο εδώ από το 0 έως 1. Αυτή η κλιμάκωση εφαρμόστηκε σχολαστικά τόσο στα σετ δεδομένων εκπαίδευσης όσο και στα σετ δεδομένων δοκιμών ανεξάρτητα, για να αποφευχθεί οποιαδήποτε πιθανή διαρροή δεδομένων μεταξύ τους. Επιπλέον, εφαρμόσαμε τις ίδιες παραμέτρους κλιμάκωσης και στα δύο σύνολα δεδομένων για να εγγυηθούμε τη συνέπεια στη μετατροπή των χαρακτηριστικών, διατηρώντας την ακεραιότητα της πειραματικής μας διάταξης.

Με την τήρηση αυτής της αυστηρής προσέγγισης, διασφαλίσαμε ότι τα δεδομένα μας κανονικοποιήθηκαν σωστά, βελτιστοποιώντας τα για τα επόμενα στάδια μοντελοποίησης και ανάλυσης νευρωνικών δικτύων.

6.2.5. Έλεγχος για Τιμές που λείπουν και Επαναφορά Ευρετηρίου

Η διασφάλιση της ακεραιότητας των δεδομένων απαιτεί προσεκτική διαχείριση των τιμών “NaN”. Η προσέγγισή μας για τη διατήρηση της ποιότητας και χρηστικότητας των συνόλων δεδομένων μας περιλαμβάνει:

1. **Αφαίρεση τιμών “NaN”**: Επιλέξαμε να αφαιρέσουμε τις γραμμές με τιμές “NaN” από τα σετ εκπαίδευσης και δοκιμών. Αυτή η στρατηγική υιοθετήθηκε λόγω της ελάχιστης παρουσίας τιμών “NaN” σε σχέση με το συνολικό μέγεθος των δεδομένων, διασφαλίζοντας ότι η αφαίρεσή τους δεν θα επηρέαζε αρνητικά την πληρότητα ή την ακεραιότητα του συνόλου δεδομένων.
2. **Επαναφορά του ευρετηρίου**: Μετά την αφαίρεση των τιμών “NaN”, επαναφέραμε τον δείκτη και των δύο συνόλων δεδομένων. Αυτή η προσαρμογή εξασφάλισε έναν ομαλό και συνεχή ακέραιο ευρετήριο σε όλα τα σετ δεδομένων, διευκολύνοντας την αποτελεσματική πρόσβαση στις γραμμές μέσω του ευρετηρίου στις επόμενες φάσεις μοντελοποίησης.

7. ΑΛΓΟΡΙΘΜΟΙ ΒΕΜ ΓΙΑ ΤΟΥΣ ΠΡΑΚΤΟΡΕΣ ΕΚΤΕΛΕΣΗΣ ΣΥΝΑΛΛΑΓΩΝ

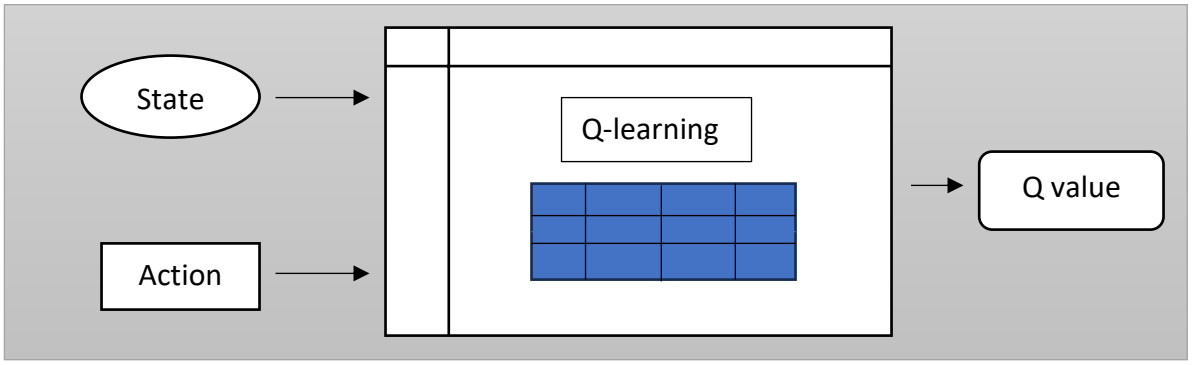
Σε αυτή την ενότητα, εμβαθύνουμε στους αλγορίθμους Βαθιάς Ενισχυτικής Μάθησης που υλοποιήθηκαν για την αντιμετώπιση του προβλήματος που περιεγράφηκε νωρίτερα. Όπως αναλύεται στην Ενότητα 4.3 και βασιζόμενοι στις αρχές που καθορίζονται στο [13], η πολιτική μάθησης ενός πράκτορα $\pi(s)$ σε ένα συγκεκριμένο περιβάλλον αφορά ουσιαστικά την εκμάθηση της κατανομής των ενεργειών για συγκεκριμένες καταστάσεις. Μόλις μια πολιτική $\pi(s)$ καθοριστεί, καθοδηγεί τη διαδικασία λήψης αποφάσεων του πράκτορα, επιτρέποντας την επιλογή της πλέον κατάλληλης ενέργειας για κάθε αντιμετωπιζόμενη κατάσταση.

Η ενισχυτική μάθηση (Reinforcement learning) διαιρείται σε δύο κύριες κατηγορίες με βάση τις μεθόδους βελτίωσης της πολιτικής: Την “value-based RL” και την “policy-based RL”. Η value-based μάθηση χρησιμοποιεί μια συνάρτηση - Q-function, $Q_\pi(s,a)$, για να προβλέψει τις αναμενόμενες αποδόσεις από μια κατάσταση s κατά την εκτέλεση μιας ενέργειας a υπό την πολιτική $\pi(s)$. Αυτή η προσέγγιση επιτρέπει σε έναν πράκτορα να βελτιστοποιήσει τις σωρευτικές ανταμοιβές επιλέγοντας τις ενέργειες με την υψηλότερη Q-value.

Η μέθοδος μάθησης “Q-learning” στην ουσία της, είναι μια μέθοδος για την κατανόηση των δυναμικών του περιβάλλοντος. Αντί να ενημερώνει την $Q(s_t,a_t)$ με βάση την αναμενόμενη τιμή της $Q(s_{t+1},a_{t+1})$, χρησιμοποιεί μια άπληστη (greedy) στρατηγική επιλέγοντας την ενέργεια a_{t+1} που μεγιστοποιεί την τιμή της $Q(s_{t+1},a_{t+1})$ για την επόμενη κατάσταση s_{t+1} [5]. Τα παραπάνω αποτυπώνονται στην ακόλουθη σχέση,

$$Q_\pi(s_t,a_t) = E_{s_{t+1}}[r(s_t,a_t,s_{t+1}) + \gamma \max_{a_{t+1}} Q_\pi(s_{t+1},a_{t+1})]$$

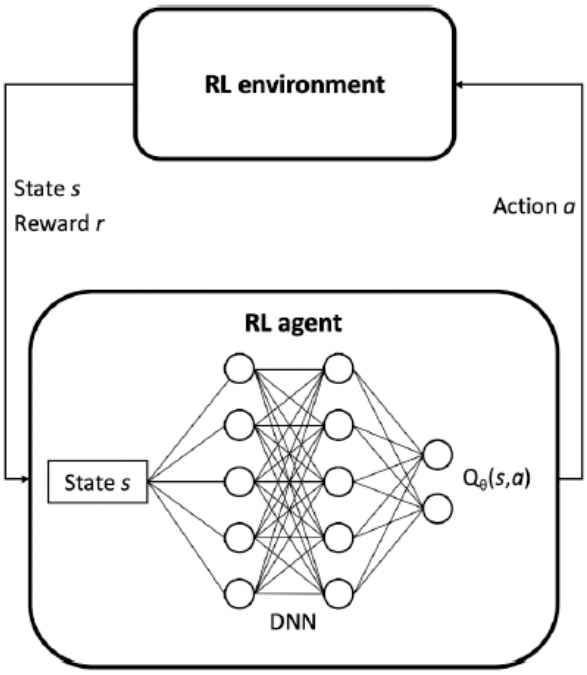
Η μέθοδος μάθησης “Q-learning” [23] είναι ένας αλγόριθμος εκτός πολιτικής (off-policy), ανεξάρτητος από μοντέλο (model-free), και βασισμένος στις τιμές (value-based) που επικεντρώνεται γύρω από έναν πίνακα “Q-table” που καταγράφει την Q-value για κάθε ζεύγος κατάστασης-ενέργειας. Το Σχήμα 6 απεικονίζει τη διαδικασία μάθησης “Q-learning”. Μέσω επαναληπτικών ενημερώσεων που καθοδηγούνται από την εξίσωση βελτιστοποίησης του Bellman, η συνάρτηση Q-function συγκλίνει σταδιακά προς την βέλτιστη $Q_\pi(s_t,a_t)$.



Σχήμα 6: Η Q Learning ανακτά τις τιμές του ζεύγους κατάστασης-ενέργειας από έναν πίνακα Q table. Ανάκτηση από 'Reinforcement Learning Explained Visually: Part 5 - Deep Q-Networks Step by Step' by Ketan Doshi, Towards Data Science, 2020 [26].

7.1. Deep Q-Learning Μεθοδολογία και ο Αλγόριθμος Deep Q-Network

Βασιζόμενοι στις μεθοδολογίες που αναλύονται στις πηγές [2], [13], [15] και [16], η “Deep Q-Learning” μεθοδολογία ενισχύει την “Q-learning” αντικαθιστώντας τον πίνακα “Q-table” με ένα «Βαθύ Νευρωνικό Δίκτυο» (Deep Neural Network - DNN). Αυτή η προσαρμογή αντιμετωπίζει τους περιορισμούς που σχετίζονται με τους συνεχείς ή απείρως διακριτούς χώρους καταστάσεων και ενεργειών, όπου η παραδοσιακή επανάληψη της Q-value μέσω ενός Q-table δεν είναι πρακτική. Με απλά λόγια, το DQN χρησιμοποιεί νευρωνικά δίκτυα για την προσέγγιση συναρτήσεων, όπου οι καταστάσεις αναπαριστώνται εντός της συνάρτησης τιμής [5]. Το Σχήμα 7 αναπαριστά την παραπάνω περιγραφή.



Σχήμα 7: Ο αλγόριθμος DQN. Ανάκτηση από Théate and Ernst (2021). [16]

Ο αλγόριθμος Deep Q-Network (DQN), μια προχωρημένη μορφή της Βαθιάς Ενισχυτικής Μάθησης, επεκτείνει τις δυνατότητες της κλασικής μεθοδολογίας Q-learning, για να διαχειριστεί αποτελεσματικά υψηλής διάστασης αισθητηριακές εισόδους. Ως ένας αλγόριθμος ανεξάρτητος από μοντέλο, ο DQN δεν απαιτεί ένα περιεκτικό μοντέλο του περιβάλλοντος, αλλά στηρίζεται αντ' αυτού σε «τροχιές» (trajectories) για την εκμάθηση των πολιτικών ελέγχου. Εμπίπτει στην οικογένεια των αλγορίθμων Q-learning, επικεντρώνοντας στην προσέγγιση της τιμής Q-value μέσω ενός DNN. Σε αυτό το πλαίσιο, η εκμάθηση της Q-value ισοδυναμεί με τον προσδιορισμό των βέλτιστων παραμέτρων θ του DNN. Αυτή η διαδικασία προσέγγισης περιλαμβάνει την ελαχιστοποίηση του μέσου τετραγώνου σφάλματος μεταξύ των τρεχουσών Q-values και των τιμών στόχων Q-values, βελτιστοποιώντας έτσι τη συνάρτηση, Q- function.

Για να γίνουμε πιο συγκεκριμένοι, όπως αναλύεται στο [15], ο αλγόριθμος Deep Q-Network (DQN) χρησιμοποιεί ένα εξελιγμένο πολυεπίπεδο νευρωνικό δίκτυο που, για κάθε δεδομένη κατάσταση s, παράγει ένα διάνυσμα τιμών ενέργειας $Q(s,\cdot;\theta)$ με το θ να αντιπροσωπεύει τις παραμέτρους του δικτύου. Αυτό το νευρωνικό δίκτυο απεικονίζει έναν «n-διάστατο» χώρο καταστάσεων σε έναν «m-διάστατο» χώρο ενεργειών, λειτουργώντας ουσιαστικά ως μετατροπή από $\mathbb{R}^n$ σε $\mathbb{R}^m$. Κλειδί για τον αλγόριθμο DQN είναι δύο θεμελιώδεις συνιστώσες: η ενσωμάτωση ενός Νευρωνικού Δικτύου με την ονομασία «δίκτυο στόχου» (target network) και της εφαρμογής της

μεθόδου «αποθήκη εμπειριών» (experience replay). Το «δίκτυο στόχου», που προσδιορίζεται από τις παραμέτρους του, θ^-, αντικατοπτρίζει το «βασικό δίκτυο» (main/online network) αλλά με μια κρίσιμη διάκριση – οι παράμετροί του συγχρονίζονται με το «βασικό δίκτυο» σε διαστήματα κάθε t βημάτων, με αποτέλεσμα $\theta_t^- = \theta_t$, και διατηρούνται σταθερές κατά τα ενδιάμεσα βήματα. Η «αποθήκη εμπειριών» περιλαμβάνει τη διατήρηση των παρατηρούμενων μεταβάσεων κατά τη διάρκεια μιας περιόδου, οι οποίες στη συνέχεια δειγματοληπτούνται ομοιόμορφα από αυτήν την ομάδα αρχειοθέτησης για να πραγματοποιηθούν οι ενημερώσεις του δικτύου.

Στον αλγόριθμο DQN, ο πράκτορας της ενισχυτικής μάθησης (reinforcement learning - RL) μαθαίνει επαναληπτικά τη σειρά των ενεργειών που βελτιστοποιούν μια «αντικειμενική συνάρτηση» (objective function), χρησιμοποιώντας την Εξίσωση Bellman για αναδρομή. Η Εξίσωση Bellman για τους DQN αλγόριθμους διατυπώνει το πώς η τιμή ενός ζεύγους κατάστασης-ενέργειας είναι η άμεση ανταμοιβή συν τις προεξοφλημένες μέγιστες μελλοντικές ανταμοιβές (discounted maximum future rewards). Η προσδοκία λαμβάνει υπόψη τη στοχαστική φύση του περιβάλλοντος. Στην πράξη, οι DQN προσεγγίζουν τη συνάρτηση Q-value χρησιμοποιώντας βαθιά νευρωνικά δίκτυα, όπου το θ αντιπροσωπεύει τα βάρη του δικτύου. Η Εξίσωση Bellman για τους DQN, ειδικά στο πλαίσιο της εκτίμησης της τιμής Q-value, εκφράζεται ως εξής:

$$Q(s, a; \theta) = \mathbb{E}[r + \gamma max_{a_{t+1}} Q(s', a'; \theta^-)|s, a]$$

Όπου:

- $Q(s, a; \theta)$ είναι η τιμή Q-value για την ανάληψη της ενέργειας a στην κατάσταση s, παραμετροποιημένη από το θ, που είναι τα βάρη του νευρωνικού δικτύου στο DQN.
- $\mathbb{E}$ συμβολίζει την προσδοκία για όλα τα δυνατά σενάρια.
- r είναι η άμεση ανταμοιβή που λαμβάνεται μετά την ανάληψη της ενέργειας a στην κατάσταση s.
- γ είναι ο «εκπτωτικός παράγοντας» (discount factor), που αντιπροσωπεύει τη διαφορά στη σημαντικότητα μεταξύ των μελλοντικών ανταμοιβών και των άμεσων ανταμοιβών. Είναι μια τιμή μεταξύ 0 και 1.
- s' είναι η νέα κατάσταση μετά την ανάληψη της ενέργειας a.
- $max_{a_{t+1}} Q(s', a'; \theta^-)$ αντιπροσωπεύει τη μέγιστη προβλεπόμενη ανταμοιβή που μπορεί να επιτευχθεί για την επόμενη κατάσταση s', που επιτυγχάνεται από οποιαδήποτε ενέργεια a' υπό τις παραμέτρους του «δικτύου στόχου» θ^- . Οι παράμετροι του «δικτύου στόχου» είναι ένα καθυστερημένο αντίγραφο των παραμέτρων του «βασικού δικτύου», παρέχοντας σταθερότητα στις ενημερώσεις εκμάθησης.

7.2. Double Q-Learning Μεθοδολογία και ο Αλγόριθμος Double Deep Q-Network

Ακολουθώντας την ίδια λογική με την ανάλυση των πηγών [6], [15] ενσωματώσαμε στην εργασία μας τον αλγόριθμο Double Deep Q-Network (DDQN). Ο DDQN είναι μια επέκταση του αλγορίθμου Deep Q-Network (DQN), που έχει σχεδιαστεί για να αντιμετωπίζει έναν συγκεκριμένο περιορισμό που εντοπίζεται στον αλγόριθμό DQN. Η κύρια καινοτομία του DDQN πηγάζει από την προσέγγισή του για μείωση του «σφάλματος υπερεκτίμησης» (overestimation bias) που υπάρχει στην προσέγγιση της τιμής Q-value του DQN.

Η βασική ιδέα του DDQN είναι εμπνευσμένη από τον αλγόριθμο Double Q-learning [24], ο οποίος αρχικά προτάθηκε για να μετριάσει την υπερεκτίμηση των τιμών ενεργειών στην τυπική Q-learning μεθοδολογία. Ο DQN, χρησιμοποιώντας ένα μοναδικό νευρωνικό δίκτυο (ή ένα ζεύγος σε περίπτωση που υπάρχει «δίκτυο στόχου») για την επιλογή και αξιολόγηση της ενέργειας, τείνει να υπερεκτιμά τις τιμές Q επειδή χρησιμοποιεί τη μέγιστη τιμή των ενεργειών ως προσέγγιση για τη μέγιστη αναμενόμενη τιμή. Ο DDQN διαχωρίζει αυτά τα δύο βήματα: ένα δίκτυο χρησιμοποιείται για την επιλογή της ενέργειας (το «βασικό δίκτυο»), και ένα άλλο για την αξιολόγηση της ενέργειας (το

«δικτύου στόχου»), μειώνοντας έτσι την υπερεκτίμηση, αποσυνδέοντας την επιλογή από την αξιολόγηση.

Στην Double Q-learning μεθοδολογία η διαδικασία μάθησης περιλαμβάνει δύο ξεχωριστές «συναρτήσεις τιμής», καθεμία από τις οποίες ενημερώνεται από εμπειρίες που ανατίθενται τυχαία. Αυτή η μέθοδος έχει ως αποτέλεσμα δύο διακριτά σύνολα βαρών, το θ και το θ', τα οποία χρησιμεύουν στον μετριασμό του σφάλματος υπερεκτίμησης που παρατηρείται στην Q-learning μεθοδολογία και, κατ' επέκταση, στον DQN αλγόριθμο.

Στο πλαίσιο του DDQN αλγορίθμου, η παραπάνω αρχή προσαρμόζεται στον τομέα της βαθιάς μάθησης, όπου διατηρούνται δύο σύνολα βαρών: τα βάρη του «βασικού δικτύου» θ και τα βάρη του «δικτύου στόχου» θ', για τη χωριστή διαχείριση των εργασιών της επιλογής ενέργειας και της αξιολόγησης της τιμής της ενέργειας. Ως εκ τούτου, η κύρια διαφορά μεταξύ του DQN και του DDQN έγκειται στην προσέγγισή τους για την εκτίμηση των Q-values για την επόμενη κατάσταση. Ενώ στον DQN χρησιμοποιείται το ίδιο δίκτυο τόσο για την πρόβλεψη όσο και για την επιλογή της ενέργειας με τη μέγιστη Q-value, στο DDQN χρησιμοποιεί μια διπλή στρατηγική όπου κατά τη διάρκεια κάθε ενημέρωσης [15]:

- Η ενέργεια επιλέγεται χρησιμοποιώντας τα τρέχοντα βάρη (θ) του «βασικού δικτύου» με βάση την άπληστη πολιτική (greedy policy).
- Η τιμή της επιλεγμένης ενέργειας αξιολογείται χρησιμοποιώντας τα βάρη του «δικτύου στόχου» (θ') παρέχοντας μια αμερόληπτη εκτίμηση των μελλοντικών ανταμοιβών.

Αυτός ο διαχωρισμός ευθυγραμμίζεται με την αρχή της “Double Q-learning” μεθοδολογίας, όπου οι ρόλοι των θ και θ' εναλλάσσονται περιοδικά για να διατηρηθεί η συμμετρία στις ενημερώσεις, διασφαλίζοντας μια ισορροπημένη και δίκαιη αξιολόγηση των πολιτικών.

Αντικατοπτρίζοντας τη στρατηγική των διπλών βαρών, η Εξίσωση Bellman για τον DDQN διατυπώνεται ως εξής:

$$Q(s,a;\theta) = \mathbb{E}[r + \gamma Q(s', \underset{a'}{arg\max}\, Q\left(s', a'; \theta\right); \theta' | s, a]$$

Αυτή η εξίσωση τονίζει τη μεθοδολογική απόκλιση από τον DQN, με τις παραμέτρους θ να χρησιμοποιούνται για την επιλογή ενέργειας και τις θ' για την αξιολόγηση της Q-value. Η λειτουργία του ορίσματος $argmax$ εκτελείται από το «βασικό δίκτυο» για να επιλεγεί η ενέργεια, αλλά η αξιολόγηση της Q-value αυτής της ενέργειας γίνεται χρησιμοποιώντας το «δικτύου στόχου». Μέσω της εφαρμογής αυτού του διπλού μηχανισμού, ο DDQN αλγόριθμός ενισχύει σημαντικά την ακρίβεια των εκτιμήσεων της Q-value, μειώνοντας τον κίνδυνο υπερεκτίμησης του αλγορίθμου DQN, οδηγώντας έτσι σε πιο ακριβείς και αξιόπιστες προσεγγίσεις πολιτικών σε περίπλοκα περιβάλλοντα λήψης αποφάσεων, όπως αυτό των κρυπτονομισμάτων.

8. ΠΕΙΡΑΜΑΤΑ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ

Στην παρούσα ενότητα παρουσιάζεται η εμπειρική αξιολόγηση των πρακτόρων των αλγορίθμων Deep Q-Networks (DQN) και Double Deep Q-Networks (DDQN), οι οποίοι και είναι υπεύθυνοι για την διαπραγμάτευση στην αγορά του Bitcoin. Η σύγκριση της επίδοσής τους θα γίνει με μια συμβατική παθητική στρατηγική «αγοράς-και-διακράτησης» μέσα στο πλαίσιο δοκιμών μας. Αυτή η συγκριτική ανάλυση είναι καθοριστική για την αξιολόγηση της ικανότητας λήψης αποφάσεων και της συνολικής επίδοσης των αλγορίθμων υπό ομοιόμορφες συνθήκες του περιβάλλοντος διαπραγμάτευσης, καθώς και για την εκτίμηση της αποτελεσματικότητάς τους σε σχέση με τις παθητικές επενδυτικές προσεγγίσεις. Αντιπαραβάλλοντας τις προσαρμοστικές στρατηγικές των DQN και DDQN που βασίζονται στους σχετικούς αλγόριθμους με τη στατική προσέγγιση της «αγοράς-και-διακράτησης», στοχεύουμε να τονίσουμε τα δυνητικά οφέλη και τις αδυναμίες της ενσωμάτωσης προηγμένων τεχνικών ενισχυτικής μάθησης σε περίπλοκα πλαίσια λήψης αποφάσεων. Τα ευρήματα που συζητούνται εδώ θα τονίσουν τα συγκριτικά πλεονεκτήματα κάθε στρατηγικής, ρίχνοντας φως στην πρακτικότητα και την αποτελεσματικότητά τους σε σενάρια συναλλαγών του πραγματικού κόσμου.

8.1. Υπερ-παράμετροι

Η διαδικασία εκπαίδευσης των πρακτόρων μας, που αποσκοπεί στη λήψη αποφάσεων εντός του ειδικά προσαρμοσμένου περιβάλλοντος, χρησιμοποιεί τις προηγμένες μεθοδολογίες των αλγορίθμων Deep Q-Networks (DQN) και Double Deep Q-Networks (DDQN).

Η αποτελεσματικότητα και οι αποδόσεις και των δύο μοντέλων DQN και DDQN επηρεάζονται σημαντικά από ένα σύνολο κρίσιμων παραμέτρων, γνωστών ως υπερ-παράμετροι. Αυτές οι υπερ-παράμετροι διέπουν βασικές πτυχές του καθεστώτος εκπαίδευσης, όπως η ισορροπία μεταξύ της εξερεύνησης νέων ενεργειών (exploration) και της εκμετάλλευσης/αξιοποίησης γνωστών στρατηγικών (exploitation), η αποτίμηση των μελλοντικών ανταμοιβών και ο ρυθμός με τον οποίο ο πράκτορας αφομοιώνει νέες πληροφορίες. Η προσαρμογή αυτών των υπερ-παραμέτρων είναι απαραίτητη για τη βελτίωση της διαδικασίας μάθησης του πράκτορα και για τη μεγιστοποίηση της απόδοσής του.

Παρακάτω, παρουσιάζουμε έναν πίνακα που περιγράφει τις κύριες υπερ-παραμέτρους που χρησιμοποιούνται στα μοντέλα μας. Αυτός ο πίνακας περιλαμβάνει τις τιμές τους και παρέχει μια συνοπτική επεξήγηση της συμβολής κάθε υπερ-παραμέτρου στη δυναμική εκπαίδευσης των πρακτόρων. Σκοπός της παράθεσης του παρακάτω πίνακα είναι μια διαφανή εικόνα στη διαμόρφωση των πρακτόρων ενισχυτικής μάθησης και των στρατηγικών παραμέτρων που διαμορφώνουν την πορεία τους προς τη μάθηση.

Υπερ-παράμετροι	Τιμή	Περιγραφή
Μέγεθος Μνήμης	200	Το μέγιστο μέγεθος της μνήμης buffer για την «αποθήκη εμπειριών». Καθορίζει πόσες παλαιότερες εμπειρίες αποθηκεύονται για τη μάθηση του πράκτορα.
Συχνότητα Ενημέρωσης Στόχου	1.000	Η συχνότητα (σε χρονικά βήματα) με την οποία τα βάρη του δικτύου στόχου ενημερώνονται για να ταιριάζουν με τα βάρη του βασικού δικτύου.
Gamma (γ)	0,95	Ο «εκπτωτικός παράγοντας» (discount factor) που χρησιμοποιείται για να δώσει προτεραιότητα στις άμεσες ανταμοιβές έναντι των μελλοντικών ανταμοιβών. Ένας παράγοντας ίσος με 0,95 σημαίνει ότι οι μελλοντικές ανταμοιβές αποτιμώνται στο 95% των άμεσων ανταμοιβών.
Epsilon (ε)	1,0	Το αρχικό ποσοστό εξερεύνησης. Ουσιαστικά είναι η πιθανότητα επιλογής μιας τυχαίας ενέργειας έναντι της καλύτερης ενέργειας σύμφωνα με το μοντέλο, για να ενθαρρυνθεί η εξερεύνηση του χώρου καταστάσεων-ενεργειών.
Epsilon Min (ε_min)	0,01	Το ελάχιστο ποσοστό εξερεύνησης. Αυτό διασφαλίζει ότι υπάρχει πάντα κάποιο επίπεδο εξερεύνησης, αποτρέποντας τον πράκτορα από το να εκμεταλλεύεται μόνο γνωστές στρατηγικές.
Epsilon Decay (ε_decay)	0,995	Ο ρυθμός με τον οποίο το ποσοστό εξερεύνησης μειώνεται μετά από κάθε επεισόδιο, επιτρέποντας στον πράκτορα να μεταβαίνει σταδιακά από την εξερεύνηση στην εκμετάλλευση καθώς μαθαίνει περισσότερα για το περιβάλλον.
Ρυθμός Μάθησης (Learning Rate)	0,001	Το μέγεθος βήματος με το οποίο ενημερώνονται τα βάρη του μοντέλου νευρωνικού δικτύου κατά τη διάρκεια της εκπαίδευσης. Ελέγχει την ταχύτητα και τη σταθερότητα της μάθησης.
Μονάδες Πρώτου Κρυφού Επιπέδου	24	Ο αριθμός των νευρώνων στο πρώτο κρυφό επίπεδο του νευρωνικού δικτύου.
Μονάδες Δεύτερου Κρυφού Επιπέδου	24	Ο αριθμός των νευρώνων στο δεύτερο κρυφό επίπεδο του νευρωνικού δικτύου.
Συνάρτηση ενεργοποίησης (Activation Function)	Συνάρτηση διορθωμένης γραμμικής μονάδας (ReLu)	Η συνάρτηση ενεργοποίησης που χρησιμοποιείται στα κρυφά επίπεδα.

Υπερ-παράμετροι	Τιμή	Περιγραφή
Ενεργοποίηση Επιπέδου Εξόδου	Γραμμική (Linear)	Η συνάρτηση ενεργοποίησης που χρησιμοποιείται στο εξωτερικό επίπεδο.
Συνάρτηση Απώλειας (Loss Function)	Μέσο Τετραγωνικό Σφάλμα (MSE)	Η συνάρτηση απώλειας που χρησιμοποιείται κατά τη διάρκεια της εκπαίδευσης του νευρωνικού δικτύου. Μετρά τον μέσο όρο των τετραγώνων των σφαλμάτων μεταξύ των προβλεπόμενων Q-values και των Q-values στόχου που υπολογίζονται στο «δίκτυο στόχου».
Βελτιστοποιητής (Optimizer)	Adam	Ο αλγόριθμος βελτιστοποίησης που χρησιμοποιείται για την ελαχιστοποίηση της συνάρτησης απώλειας.

8.2. Παράμετροι Συναλλαγών

Για να διασφαλίσουμε μια ακριβή αξιολόγηση της απόδοσης του μοντέλου συναλλαγών, έχουμε επιλέξαμε προσεκτικά μια σειρά από κρίσιμες παραμέτρους συναλλαγών που είναι απαραίτητες για προσεγγίσουν όσο το δυνατόν περισσότερο ένα πραγματικό περιβάλλον συναλλαγών. Αυτές οι παράμετροι είναι καθοριστικές για την ακριβή προσομοίωση των συνθηκών της αγοράς και επηρεάζουν άμεσα τον υπολογισμό βασικών χρηματοοικονομικών μετρήσεων όπως το Sharpe Ratio, η Σωρευτική Απόδοση (Cumulative Return), και η Μέγιστη Πτώση (Max Drawdown), που είναι ζωτικής σημασίας για την αξιολόγηση της αποτελεσματικότητας του μοντέλου μας.

Ειδικότερα, το «ακίνδυνο επιτόκιο» (risk-free rate) — μια βασική συνιστώσα για τον υπολογισμό του δείκτη Sharpe — προέκυψε από την μέση απόδοση του 10ετούς Κρατικού Ομολόγου των ΗΠΑ (10YTCMR) καθ' όλη τη διάρκεια της περιόδου δοκιμής μας. Αυτό το σημείο αναφοράς επιλέχθηκε για τη σταθερότητα και την αντιπροσωπευτικότητα μιας επένδυσης χωρίς κίνδυνο μακροπρόθεσμα. Σύμφωνα με την YCharts (n.d.) [25], τα ιστορικά δεδομένα αποκάλυψαν διακυμάνσεις από 1,49% στο τέλος του Δεκεμβρίου 2021 έως 3,88% στο τέλος του Δεκεμβρίου 2023. Υιοθετώντας μια προσέγγιση μέσης τιμής, καθορίσαμε το Ακίνδυνο Επιτόκιο στο 2%, με στόχο να παρέχει μια ισορροπημένη αντανάκλαση της απόδοσης χωρίς κίνδυνο κατά την περίοδο ανάλυσης.

Ο παρακάτω πίνακας περιγράφει με συνοπτικό τρόπο αυτές τις παραμέτρους συναλλαγών, παρέχοντας επεξήγηση τα κρίσιμα στοιχεία που αποτελούν τη βάση του πλαισίου αξιολόγησης του μοντέλου συναλλαγών μας.

Παράμετρος	Τιμή	Περιγραφή
Διάρκεια Περιόδου Δοκιμής	2 έτη	Η συνολική διάρκεια κατά την οποία δοκιμάστηκε το μοντέλο.
Ακίνδυνο Επιτόκιο	2%	Το θεωρητικό επιτόκιο απόδοσης μιας επένδυσης χωρίς κίνδυνο απωλειών.
Αρχική Επένδυση	$10.000	Το ποσό ($) του αρχικού κεφαλαίου που επενδύθηκε στην αρχή της περιόδου δοκιμής.

8.3. Βασική Παθητική Στρατηγική

Με παρόμοια λογική με τις μεθοδολογίες έρευνας που προωθούνται από τις πηγές [17] και [13], πραγματοποιούμε μια εκτενή σύγκριση μεταξύ της απόδοσης των πρακτόρων Deep Q-Network (DQN) και Double Deep Q-Network (DDQN) έναντι μιας παραδοσιακής επενδυτικής στρατηγικής. Αυτή η συγκριτική ανάλυση είναι ουσιώδης για αρκετούς λόγους. Πρώτον, μας επιτρέπει να αξιολογήσουμε την αποτελεσματικότητα των προηγμένων τεχνικών ενισχυτικής μάθησης εντός της πολύπλοκης δυναμικής της αγοράς κρυπτονομισμάτων. Δεύτερον, με την αξιολόγηση έναντι μιας συμβατικής στρατηγικής, μπορούμε να αξιολογήσουμε με μεγαλύτερη ακρίβεια την προστιθέμενη αξία που προσφέρουν αυτά τα εξελιγμένα μοντέλα στις αλγοριθμικές συναλλαγές.

Επιλέξαμε τη στρατηγική «Αγοράς και Διακράτησης» ως σημείο αναφοράς για την απλότητα και την ευρεία αποδοχή της ως βάση στις μελέτες χρηματοοικονομικής απόδοσης. Αυτή η στρατηγική, η οποία περιλαμβάνει την αγορά Bitcoin και τη διατήρησή τους για μεγάλο χρονικό διάστημα, ανεξάρτητα από τις μεταβολές της αγοράς, λειτουργεί ως μια ισχυρή μέθοδο σύγκρισης λόγω της παθητικής φύσης της. Είναι ιδιαίτερα χρήσιμο για την ανάδειξη της δυνατότητας των πρακτόρων μας να παράγουν ανώτερες αποδόσεις μέσω της ενεργού διαχείρισης, παρά την εκ των πραγμάτων απρόβλεπτη φύση της αγοράς.

Η στρατηγική της «Αγοράς και Διακράτησης» αναγνωρίζεται επίσης για την αποτελεσματικότητα της στο να παράγει σημαντικές αποδόσεις σε μεγάλες χρονικές περιόδους, καθιστώντας την έναν ισχυρό βασικό δείκτη που κάθε προηγμένος αλγόριθμος συναλλαγών θα πρέπει να επιδιώξει να ξεπεράσει. Δείχνοντας ότι οι πράκτορες DQN και DDQN μπορούν να ξεπεράσουν μια τέτοια παθητική στρατηγική, στοχεύουμε να τονίσουμε τα πρακτικά οφέλη της υλοποίησης τεχνικών Βαθιάς Ενισχυτικής Μάθησης στις χρηματοοικονομικές αγορές. Αυτή η σύγκριση όχι μόνο επιβεβαιώνει την αποτελεσματικότητα των μοντέλων μας αλλά συμβάλλει και στην ευρύτερη συζήτηση σχετικά με την εφαρμογή της τεχνητής νοημοσύνης στην ενίσχυση των επενδυτικών στρατηγικών.

8.4. Αξιολόγηση απόδοσης: DQN, DDQN, και στρατηγικής «Αγοράς και Διακράτησης»

Αυτή η ενότητα παρουσιάζει μια λεπτομερή σύγκριση της απόδοσης δύο αλγοριθμικών στρατηγικών συναλλαγών BEM, των Deep Q-Networks (DQN) και Double Deep Q-Networks (DDQN), έναντι της παθητικής στρατηγικής «Αγοράς και Διακράτησης». Η ανάλυσή μας εκτείνεται σε μια περίοδο δοκιμών 2 ετών από τις 31 Δεκεμβρίου 2021 έως τις 31 Δεκεμβρίου 2023, προσφέροντας πληροφορίες για την αποτελεσματικότητα των προηγμένων τεχνικών ενισχυτικής μάθησης στην ευμετάβλητη αγορά κρυπτονομισμάτων. Μέσω αυτής της συγκριτικής μελέτης, στοχεύουμε να αξιολογήσουμε τις δυνατότητες των στρατηγικών DQN και DDQN όχι μόνο να ξεπεράσουν τις παθητικές επενδυτικές στρατηγικές αλλά και να σταθούν αποτελεσματικά στις πολυπλοκότητες των δυναμικών της αγοράς. Για να αξιολογήσουμε ενδελεχώς τις δυνατότητες και την επίδοση αυτών των αλγορίθμων, τα πειράματά μας διεξάγονται με βάση δύο διακριτά σύνολα:

- το ένα χρησιμοποιώντας μια διαδικασία εκπαίδευσης για **«50 Επεισόδια/10 Μέγιστα Βήματα ανά Επεισόδιο»** και
- το δεύτερο για **«60 Επεισόδια/30 Μέγιστα Βήματα ανά Επεισόδιο»**.

Σημειώνουμε ότι κάθε επεισόδιο (episode) αντιπροσωπεύει μια πλήρη «δράση» του πράκτορα από μια αρχική κατάσταση σε μια τελική κατάσταση στο περιβάλλον. Με τον όρο «Μέγιστα Βήματα ανά Επεισόδιο» (max steps per episode) ορίζουμε τον μέγιστο αριθμό βημάτων (ή χρονικών βημάτων) που επιτρέπεται να κάνει ο πράκτορας σε ένα μόνο επεισόδιο. Στην περίπτωσή μας τα βήματα είναι οι ημέρες συναλλαγών.

Αυτές οι διαμορφώσεις μας επιτρέπουν να διερευνήσουμε τον αντίκτυπο της διαφορετικής διάρκειας επεισοδίων και του αριθμού των βημάτων στην αποτελεσματικότητα κάθε στρατηγικής, παρέχοντας μια λεπτομερή κατανόηση της προσαρμοστικότητας και των δυνατοτήτων βελτιστοποίησής τους σε διαφορετικά σενάρια συναλλαγών.

Η αξιολόγησή μας επικεντρώνεται σε πέντε βασικές μετρήσεις απόδοσης που παρέχουν συλλογικά μια ολιστική εικόνα του προφίλ «κινδύνου-απόδοσης», της λειτουργικής αποτελεσματικότητας και της ανθεκτικότητας κάθε στρατηγικής έναντι των πτωτικών (ή ανοδικών) τάσεων της αγοράς:

i. Σωρευτική Απόδοση (Cumulative Return),
ii. Ετησιοποιημένη Απόδοση (Annualized Return),
iii. Ετησιοποιημένη Μεταβλητότητα (Annualized Volatility),
iv. Δείκτης του Sharpe (Sharpe Ratio), και
v. Μέγιστη Πτώση (Max Drawdown).

Παρακάτω, ο Πίνακας 4 παρέχει μια λεπτομερή επισκόπηση των μετρικών απόδοσης για κάθε στρατηγική κατά τη διάρκεια της περιόδου δοκιμής. Αυτά τα δεδομένα αποτελούν τη βάση για τη συγκριτική μας ανάλυση και τα επακόλουθα συμπεράσματα σχετικά με τη σχετική αποτελεσματικότητα αυτών των στρατηγικών στη δυναμική αγορά των κρυπτονομισμάτων.

Μετρήσεις	Παθητική Στρατηγική	DQN (50 Επεισ./10 Βήμ.)	DDQN (50 Επεισ./10 Βήμ.)	DQN (60 Επεισ./30 Βήμ.)	DDQN (60 Επεισ./30 Βήμ.)
Σωρευτική Απόδοση (%)	14,68	-61,28	25,87	-40,11	12,7
Ετησιοποιημένη Απόδοση (%)	7,1	-37,8	12,2	-22,63	6,17
Ετησιοποιημένη Μεταβλητότητα (%)	47,06	33,9	46,63	33,42	47,1
Δείκτης του Sharpe	0,15	-1,12	0,26	-0,68	0,13
Μέγιστη Πτώση (%)	67,85	70,62	63,15	69,57	68,17

Πίνακας 4: Συγκριτικές μετρήσεις απόδοσης Παθητικής Στρατηγικής, DQN και DDQN κατά την περίοδο δοκιμής (2021-12-31 έως 2023-12-31)

Για να εμβαθύνουμε περισσότερο στα δυνατά και τα αδύνατα σημεία κάθε στρατηγικής συναλλαγών, παραθέτουμε παρακάτω την αξιολόγησή μας για τις μετρικές απόδοσης. Στόχος μας είναι να εντοπίσουμε τις πληροφορίες που θα μπορούσαν να αποκαλύψουν όχι μόνο τις δυνατότητες για υψηλότερες αποδόσεις αλλά και τους κινδύνους που συνδέονται με κάθε προσέγγιση:

- **Σωρευτικές και Ετησιοποιημένες Αποδόσεις**: Η στρατηγική DDQN επέδειξε σημαντικό πλεονέκτημα έναντι της παθητικής στρατηγικής στο αρχικό σετ πειραμάτων (50 επεισόδια/10 μέγιστα βήματα ανά επεισόδιο), με σημαντική βελτίωση στις σωρευτικές αποδόσεις. Αυτό

υποδηλώνει ότι οι προσαρμοστικές δυνατότητες λήψης αποφάσεων του DDQN μπορούν να αξιοποιήσουν τις ευκαιρίες για ανώτερες αποδόσεις πιο αποτελεσματικά από την παθητική προσέγγιση. Ωστόσο, καθώς η πολυπλοκότητα του σεναρίου αυξήθηκε (60 επεισόδια/30 μέγιστα βήματα ανά επεισόδιο), το χάσμα απόδοσης μεταξύ του DDQN και της παθητικής στρατηγικής μειώθηκε, υποδεικνύοντας διαφορετικά επίπεδα προσαρμογής υπό διαφορετικές συνθήκες αγοράς.

Η χαμηλότερη ετησιοποιημένη μεταβλητότητα που επέδειξαν και οι δύο στρατηγικές DQN και DDQN σε ορισμένες δοκιμές, σε σύγκριση με την παθητική στρατηγική, εγείρει ζητήματα σχετικά με τη διαχείριση κινδύνου. Υποδηλώνει ότι οι αλγοριθμικές στρατηγικές, μέσω της συνεχούς αξιολόγησης και προσαρμογής της αγοράς, μπορεί να διαθέτουν έναν έμφυτο μηχανισμό για τη μείωση της έκθεσης κατά τις περιόδους υψηλής μεταβλητότητας της αγοράς, ένα στοιχείο που αξίζει βαθύτερη διερεύνηση.

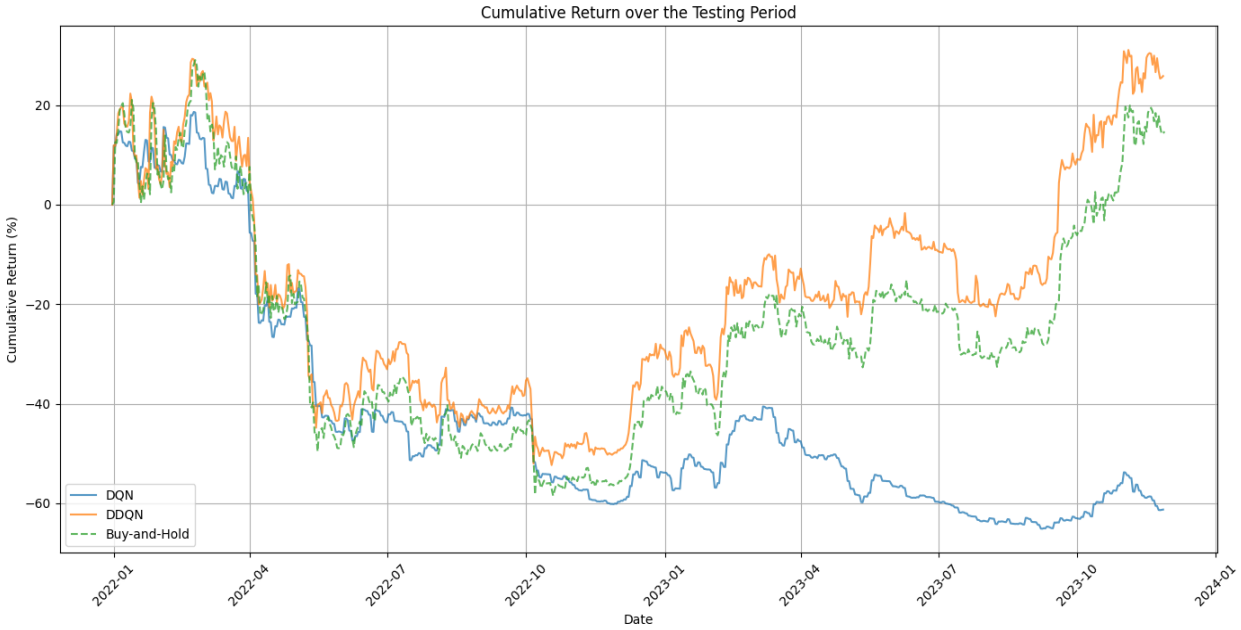
- **Μεταβλητότητα και Κίνδυνος**: Παρά την μεταβλητότητα της αγοράς, οι στρατηγικές DDQN και DQN επέδειξαν χαμηλότερη ετησιοποιημένη μεταβλητότητα σε σύγκριση με την παθητική στρατηγική σε ορισμένα σενάρια δοκιμών. Αυτό υποδηλώνει ότι η ενεργητική διαχείριση, μέσω αλγοριθμικών στρατηγικών συναλλαγών, μπορεί να προσφέρει ένα «μαξιλάρι» ασφαλείας απέναντι στην απρόβλεπτη φύση της αγοράς.

  Στο περιβάλλον των αγορών κρυπτονομισμάτων, που χαρακτηρίζεται από γρήγορες διακυμάνσεις τιμών και απρόβλεπτες τάσεις, η παρατηρούμενη χαμηλότερη ετησιοποιημένη μεταβλητότητα στις στρατηγικές DQN και DDQN αποκτά σημαντική σημασία. Η μείωση της μεταβλητότητας που επιδεικνύεται από τις στρατηγικές Βαθιάς Ενισχυτικής Μάθησης υποδηλώνει όχι μόνο ένα πιο σταθερό προφίλ απόδοσης αλλά συνεπάγεται επίσης έναν πιθανό μετριασμό του κινδύνου για τους επενδυτές. Ιδιαίτερα σε αγορές που είναι γνωστές για την υψηλή τους μεταβλητότητα, όπως τα κρυπτονομίσματα, το να επιτυγχάνεται χαμηλότερη μεταβλητότητα διατηρώντας παράλληλα ανταγωνιστικές αποδόσεις σηματοδοτεί μια αξιοσημείωτη πρόοδο στην αποτελεσματικότητα της στρατηγικής εκτέλεσης συναλλαγών.

- **Δείκτης Sharpe και Μέγιστη Πτώση**: Ο Δείκτης του Sharpe, ένα μέτρο απόδοσης προσαρμοσμένης στον κίνδυνο, ήταν πιο ευνοϊκός για τον DDQN στο πρώτο σετ (50 επεισόδια/10 μέγιστα βήματα ανά επεισόδιο), δείχνοντας έτσι τη δυνατότητά του να προσφέρει καλύτερες αποδόσεις ανά μονάδα κινδύνου. Ωστόσο, όλες οι στρατηγικές αντιμετώπισαν σημαντικές μειώσεις, αντανακλώντας το περιβάλλον υψηλού κινδύνου των αγορών κρυπτονομισμάτων.

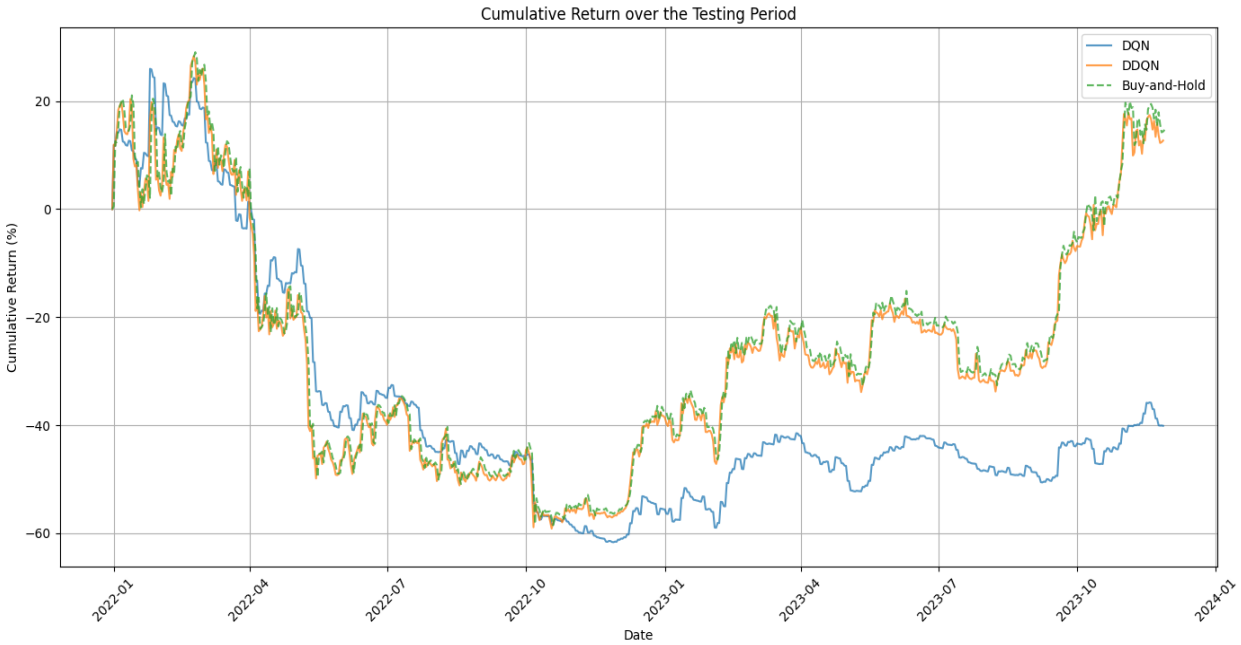
Ένα σημείο που πρέπει να αντιμετωπιστεί είναι οι παρατηρούμενες διακυμάνσεις στην απόδοση της στρατηγικής DDQN, ιδιαίτερα καθώς η πολυπλοκότητα των σεναρίων αυξάνεται. Αυτό μπορεί να υποδηλώνει την ευαισθησία του αλγορίθμου στις δυναμικές της αγοράς και στις ρυθμίσεις των υπερ-παραμέτρων. Ενώ το DDQN επέδειξε ανώτερη προσαρμοστικότητα σε απλούστερες ρυθμίσεις (50 επεισόδια/10 μέγιστα βήματα ανά επεισόδιο), η μειωμένη επίδοση σε πιο σύνθετα σενάρια (60 επεισόδια/30 μέγιστα βήματα ανά επεισόδιο) υποδηλώνει ότι οι δυνατότητες λήψης αποφάσεων του μοντέλου θα μπορούσαν να επωφεληθούν από περαιτέρω βελτίωση. Αυτή η βελτίωση θα μπορούσε να περιλαμβάνει την διερεύνηση της επίδρασης των διαφορετικών «ρυθμών μάθησης», την προσαρμογή για την εύρεση του βέλτιστου σημείου ισορροπίας μεταξύ των διαδικασιών εξερεύνησης και-εκμετάλλευσης των στρατηγικών και τη βελτιστοποίηση της αρχιτεκτονικής του δικτύου για να ενισχύσει την ικανότητά του να καλύπτει με μεγαλύτερο εύρος αντίληψης της διάφορες συνθήκες αγοράς.

Για να διευκρινιστεί περαιτέρω η δυναμική απόδοσης των στρατηγικών συναλλαγών που εξετάζονται, παρουσιάζουμε μια σειρά από γραφικές απεικονίσεις. Τα γραφήματα 8 και 9 απεικονίζουν τις σωρευτικές αποδόσεις που επιτεύχθηκαν από το Deep Q-Network (DQN), το Double Deep Q-Network (DDQN) και την παθητική στρατηγική "Αγοράς και Διακράτησης" κατά τη διάρκεια της περιόδου δοκιμών. Αυτή η οπτική ανάλυση όχι μόνο συμπληρώνει τα αριθμητικά μας ευρήματα αλλά προσφέρει επίσης μια εικόνα για τις διακυμάνσεις της απόδοσης κατά τις διάφορες φάσεις της αγοράς, τονίζοντας την ανθεκτικότητα και την προσαρμοστικότητα κάθε στρατηγικής υπό διαφορετικές συνθήκες.



Σχήμα 8: Σωρευτικές Αποδόσεις Παθητικής Στρατηγικής, DQN και DDQN (50 Επεισόδια/10 Μέγιστα Βήματα ανά Επεισόδιο)

Όπως απεικονίζεται στο Σχήμα 8, η καμπύλη της σωρευτικής απόδοσης της στρατηγικής DDQN ξεπερνά αισθητά αυτή της παθητικής στρατηγικής κατά τη διάρκεια του πειράματος των «50 επεισοδίων/10 μέγιστα βήματα ανά επεισόδιο». Αυτή η οπτική σύγκριση όχι μόνο τονίζει τα ποσοτικά ευρήματα που παρουσιάζονται στον Πίνακα 4, αλλά προσφέρει επίσης μια διαισθητική κατανόηση των διαφορών απόδοσης μεταξύ των στρατηγικών κατά τις διάφορες φάσεις της αγοράς του Bitcoin.



Σχήμα 9: Σωρευτικές Αποδόσεις Παθητικής Στρατηγικής, DQN και DDQN (60 Επεισόδια /30 Μέγιστα Βήματα ανά Επεισόδιο)

Το Σχήμα 9 περικλείει οπτικά τις δυναμικές απόδοσης στο πλαίσιο των «60 επεισοδίων/30 μέγιστα βήματα ανά επεισόδιο», υποδεικνύοντας ότι η στρατηγική DDQN και η παθητική στρατηγική εμφανίζουν παρόμοια μοτίβα όσον αφορά τις σωρευτικές αποδόσεις.

Η ανάλυσή μας αποκαλύπτει ότι ενώ η στρατηγική DDQN μπορεί να ξεπεράσει την παραδοσιακή στρατηγική "Αγορά και Διακράτηση" όσον αφορά τις σωρευτικές και ετησιοποιημένες αποδόσεις υπό ορισμένες συνθήκες, το πλεονέκτημα απόδοσής της δεν είναι σταθερό σε όλα τα σενάρια δοκιμών. Τα διαφορετικά αποτελέσματα τονίζουν τη σημασία του πλαισίου και των συνθηκών της αγοράς στον καθορισμό της αποτελεσματικότητας των αλγοριθμικών στρατηγικών συναλλαγών στην αγορά του Bitcoin.

9. ΣΥΜΠΕΡΑΣΜΑΤΑ

Η παρούσα εργασία παρέχει μια ολοκληρωμένη διερεύνηση αλγοριθμικών συναλλαγών στην ασταθή αγορά κρυπτονομισμάτων, με στόχο την αξιολόγηση των δυνατοτήτων των προηγμένων τεχνικών Βαθιάς Ενισχυτικής Μάθησης να ξεπεράσουν τις παθητικές επενδυτικές στρατηγικές. Χρησιμοποιώντας τους αλγορίθμους Deep Q-Networks (DQN) και Double Deep Q-Networks (DDQN), συγκρίναμε αυτά τα μοντέλα με την παθητική στρατηγική "Αγοράς και Διακράτησης" σε μια περίοδο δοκιμών δύο ετών.

Η αξιολόγησή μας πραγματοποιήθηκε μέσω μιας σειράς πειραμάτων που σχεδιάστηκαν για να αξιολογήσουν αυστηρά την απόδοση των μοντέλων DQN και DDQN. Αυτά τα πειράματα, τα οποία διεξήχθησαν με διαφορετικά επεισόδια και βήματα, είχαν ως στόχο να αξιολογήσουν ενδελεχώς την απόδοση κάθε μοντέλου. Αναλύοντας βασικές μετρήσεις απόδοσης, επιδιώξαμε να κατανοήσουμε το προφίλ κινδύνου-απόδοσης, την λειτουργική αποτελεσματικότητα και την ανθεκτικότητα που έχει στην αγορά κάθε στρατηγική συναλλαγών.

Τα αποτελέσματα των πειραμάτων παρέχουν σημαντικές πληροφορίες για την αποτελεσματικότητα των αλγοριθμικών στρατηγικών. Ειδικότερα, το μοντέλο DDQN, επέδειξε σημαντικό πλεονέκτημα έναντι της παθητικής στρατηγικής στο αρχικό σετ των πειραμάτων, υποδεικνύοντας την ανώτερη προσαρμοστικότητα και τις ικανότητές του στη λήψη αποφάσεων. Ωστόσο, καθώς η πολυπλοκότητα των σεναρίων αυξήθηκε, το χάσμα απόδοσης μεταξύ του μοντέλου DDQN και της παθητικής στρατηγικής μειώθηκε, υποδηλώνοντας μια μεταβλητή προσαρμοστικότητα σε διαφορετικές συνθήκες αγοράς. Παρά την μεταβλητότητα της αγοράς κρυπτονομισμάτων, και οι δύο στρατηγικές DQN και DDQN εμφάνισαν χαμηλότερη ετησιοποιημένη μεταβλητότητα σε ορισμένα σενάρια δοκιμών, υπογραμμίζοντας έτσι τη δυνατότητα των αλγοριθμικών συναλλαγών να προσφέρουν ένα κάποιου είδους ανάχωμα έναντι της απρόβλεπτης αγοράς του Bitcoin.

Συμπερασματικά, η συγκεκριμένη μελέτη υπογραμμίζει τις υποσχόμενες δυνατότητες των μοντέλων Βαθιάς Ενισχυτικής Μάθησης, όπως τα DQN και DDQN, στις αλγοριθμικό συναλλαγές, παρέχοντας πολύτιμες πληροφορίες σχετικά με την απόδοσή τους σε σχέση με τις παθητικές επενδυτικές προσεγγίσεις. Παρά το γεγονός ότι η στρατηγική DDQN έδειξε δυνατότητες καλύτερης απόδοσης έναντι της παθητικής στρατηγικής "Αγορά και Διακράτηση" υπό ορισμένες συνθήκες, η μεταβλητότητα στην απόδοση σε διαφορετικά σενάρια δοκιμών τονίζει τη σημασία της περαιτέρω έρευνας και βελτιστοποίησης για την ενίσχυση της αποτελεσματικότητας αυτών των προηγμένων αλγορίθμων συναλλαγών.

10. ΜΕΛΛΟΝΤΙΚΕΣ ΠΡΟΕΚΤΑΣΕΙΣ

Παρόλο που αυτή η μελέτη παρέχει πολύτιμες πληροφορίες σχετικά με την εφαρμογή των Deep Q-Networks (DQN) και Double Deep Q-Networks (DDQN) στην αγορά κρυπτονομισμάτων, έχουν προκύψει αρκετοί παράμετροι για μελλοντική έρευνα που υπόσχονται να προωθήσουν περαιτέρω την κατανόηση και εφαρμογή αυτών των μοντέλων στις αλγοριθμικές συναλλαγές:

1. **Βελτίωση Αρχιτεκτονικής Μοντέλου**: Μελλοντικές εργασίες θα μπορούσαν να διερευνήσουν την ενσωμάτωση προηγμένων αρχιτεκτονικών νευρωνικών δικτύων, όπως τα Νευρωνικά Δίκτυα Συνέλιξης (Convolutional Neural Networks - CNNs) για την αναγνώριση μοτίβων, τα Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks - RNNs) για την καταγραφή χρονικών εξαρτήσεων, και μηχανισμούς προσοχής (attention mechanisms) για την ιεράρχηση σημαντικών πληροφοριών. Τέτοιες πρόοδοι θα μπορούσαν να ενισχύσουν σημαντικά την προβλεπτική ακρίβεια και την προσαρμοστικότητα των μοντέλων στις διακυμάνσεις της αγοράς.
2. **Βελτιστοποίηση Παραμέτρων**: Η περαιτέρω εμβάθυνση στη ρύθμιση βασικών υπερ-παραμέτρων, όπως ο Ρυθμός Μάθησης και της ισορροπίας μεταξύ εξερεύνησης και εκμετάλλευσης, είναι απαραίτητη. Οι προσαρμογές σε αυτούς τους τομείς θα μπορούσαν να επηρεάσουν εντόνως την αποδοτικότητα μάθησης και την ποιότητα της λήψης αποφάσεων των μοντέλων.
3. **Πολυπλοκότητα στο Πρόγραμμα Εκπαίδευσης**: Η ενίσχυση του πλαισίου εκπαίδευσης με την αύξηση του αριθμού των επεισοδίων και των βημάτων ανά επεισόδιο θα μπορούσε να παρέχει στα μοντέλα ένα πιο περιεκτικό και ολοκληρωμένο περιβάλλον μάθησης. Αυτή η προσαρμογή στοχεύει στη βελτίωση των δυνατοτήτων γενίκευσης των μοντέλων και την προσαρμοστικότητά τους σε απρόβλεπτες συνθήκες αγοράς.
4. **Προηγμένες Τεχνικές Διαχείρισης Κινδύνου**: Η ενσωμάτωση εξελιγμένων τεχνικών διαχείρισης κινδύνου στα πλαίσια των μοντέλων αποτελεί έναν άλλο κρίσιμο τομέα έρευνας. Τεχνικές όπως η βελτιστοποίηση της «Υπό Συνθήκη Αξία σε Κίνδυνο» (Conditional Value at Risk - CVaR) και δυναμικοί μηχανισμοί περιορισμού ζημιών όπως οι εντολές stop-loss μπορούν να προσφέρουν βελτιωμένη προστασία έναντι των θέσεων που οδηγούν σε ζημίες, διασφαλίζοντας μια ισχυρή άμυνα ενάντια σε σημαντικές πτώσεις της αγοράς.
5. **Εξερεύνηση Εναλλακτικών Δομών Ανταμοιβής**: Η διερεύνηση διαφορετικών δομών ανταμοιβής για την εκπαίδευση των μοντέλων είναι ουσιαστική για την ευθυγράμμιση της συμπεριφοράς τους με μακροπρόθεσμους επενδυτικούς στόχους. Οι συναρτήσεις ανταμοιβής που δίνουν έμφαση την ελαχιστοποίηση του «βάθος της μέγιστης πτώσης» (drawdown) και τις αποδόσεις προσαρμοσμένες στον κίνδυνο θα μπορούσαν να βελτιώσουν τις λειτουργίες των μοντέλων, προωθώντας πιο βιώσιμες στρατηγικές συναλλαγών.
6. **Σύγκριση με εναλλακτικές παθητικές στρατηγικές**: Επιπλέον, η αξιολόγηση της απόδοσης των μοντέλων έναντι μιας στρατηγικής «Πώληση και Διακράτηση» ("Sell and Hold") θα μπορούσε να προσφέρει πολύτιμες πληροφορίες, ιδιαίτερα στην πρόβλεψη και διαχείριση πτωτικών φάσεων των αγορών. Αυτή η σύγκριση θα μπορούσε να καταδείξει την ευελιξία και τη στρατηγική αξία των μοντέλων σε διάφορα σενάρια αγοράς.
7. **Βελτίωση και Διαφοροποίηση Δεικτών Τεχνικής Ανάλυσης**: Μελλοντικές μελέτες θα μπορούσαν να διεξαγάγουν μια λεπτομερή αξιολόγηση των χρησιμοποιούμενων δεικτών τεχνικής ανάλυσης, για τη βελτίωση της αποτελεσματικότητας του μοντέλου. Επιπλέον, η ενσωμάτωση άλλων δεικτών όπως ο «Στοχαστικός Ταλαντωτής» (Stochastic Oscillator) , τα επίπεδα Fibonacci (Fibonacci Retracement) και οι Ζώνες/Λωρίδες Bollinger (Bollinger Bands) θα μπορούσε να δώσει ώθηση σε νέες πληροφορίες για την πρόβλεψη των κινήσεων της αγοράς.

Εστιάζοντας και προωθώντας την έρευνα στα παραπάνω σημεία, η μελλοντική έρευνα δεν θα ενισχύσει μόνο τα στρατηγικά οφέλη των μοντέλων DQN και DDQN στις αλγοριθμικές συναλλαγές, αλλά θα συνεισφέρει επίσης πολύτιμες γνώσεις στον εξελισσόμενο τομέα της χρηματοοικονομικής τεχνολογίας. Αυτή η προσπάθεια θα προωθήσει τις καινοτομίες που διαχειρίζονται την κερδοφορία,

τον κίνδυνο και τις διαφοροποιημένες απαιτήσεις των δυναμικών περιβαλλόντων των αγορών χρήματος και κεφαλαίου.

11. ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] Yang, H., Liu, X. Y., Zhong, S., & Walid, A. (2020, October). Deep reinforcement learning for automated stock trading: An ensemble strategy. In Proceedings of the first ACM international conference on AI in finance (pp. 1-8).

[2] Zhang, Z., Zohren, S., & Stephen, R. (2020). Deep reinforcement learning for trading. The Journal of Financial Data Science.

[3] Ntourmas, I., & Sotiropoulos, D. (2022, July). Bitcoin Price Prediction and Automated Trading via LSTM Networks and Reinforcement Learning. In 2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA) (pp. 1-5). IEEE.

[4] Huang, C. Y. (2018). Financial trading as a game: A deep reinforcement learning approach. arXiv preprint arXiv:1807.02787.

[5] Xiong, Z., Liu, X. Y., Zhong, S., Yang, H., & Walid, A. (2018). Practical deep reinforcement learning approach for stock trading. arXiv preprint arXiv:1811.07522, 1-7.

[6] Lucarelli, G., & Borrotti, M. (2019). A deep reinforcement learning approach for automated cryptocurrency trading. In Artificial Intelligence Applications and Innovations: 15th IFIP WG 12.5 International Conference, AIAI 2019, Hersonissos, Crete, Greece, May 24–26, 2019, Proceedings 15 (pp. 247-258). Springer International Publishing.

[7] Bitcoin, Nakamoto S., 2008, “Bitcoin: A Peer-to-Peer Electronic Cash System”, https://bitcoin.org/bitcoin.pdf

[8] Badmus, G. (2019). A Global Guide to a Crypto Exchange Regulatory Framework. Journal of Law, Policy and Globalization.

[9] Lyócsa, Š., Molnár, P., Plíhal, T., & Širaňová, M. (2020). Impact of macroeconomic news, regulation and hacking exchange markets on the volatility of bitcoin. Journal of Economic Dynamics and Control, 119, 103980.

[10] Bakas, D., Magkonis, G., & Oh, E. Y. (2022). What drives volatility in Bitcoin market?. Finance Research Letters, 50, 103237.

[11] Aalborg, H. A., Molnár, P., & de Vries, J. E. (2019). What can explain the price, volatility and trading volume of Bitcoin?. Finance Research Letters, 29, 255-265.

[12] Gandal, N., Hamrick, J. T., Moore, T., & Oberman, T. (2018). Price manipulation in the Bitcoin ecosystem. Journal of Monetary Economics, 95, 86-96.

[13] Chen, L., & Gao, Q. (2019, October). Application of deep reinforcement learning on automated stock trading. In 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS) (pp. 29-33). IEEE.

[14] Brim, A. (2020, January). Deep reinforcement learning pairs trading with a double deep Q-network. In 2020 10th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0222-0227). IEEE.

[15] Van Hasselt, H., Guez, A., & Silver, D. (2016, March). Deep reinforcement learning with double q-learning. In Proceedings of the AAAI conference on artificial intelligence (Vol. 30, No. 1).

[16] Théate, T., & Ernst, D. (2021). An application of deep reinforcement learning to algorithmic trading. Expert Systems with Applications, 173, 114632.

[17] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602.

[18] Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. (2015). Human-level control through deep reinforcement learning. nature, 518(7540), 529-533.

[19] Grover, A. A., & Gabriel, R. S. (2021, March). Analysis of Algorithmic Trading with Q-learning in the forex market. In 2021 International Conference on Emerging Smart Computing and Informatics (ESCI) (pp. 73-77). IEEE.

[20] Carapuço, J., Neves, R., & Horta, N. (2018). Reinforcement learning applied to Forex trading. Applied Soft Computing, 73, 783-794.

[21] Wang, Z., & Fleiss, A. (2021). Deep Q-Learning for Trading Cryptocurrency. The Journal of Financial Data Science, 3(3), 121-127.

[22] Mahayana, D., Shan, E., & Fadhl'Abbas, M. (2022, October). Deep reinforcement learning to automate cryptocurrency trading. In 2022 12th International Conference on System Engineering and Technology (ICSET) (pp. 36-41). IEEE.

[23] Watkins, C. J., & Dayan, P. (1992). Q-learning. Machine learning, 8, 279-292.

[24] Hasselt, H. (2010). Double Q-learning. Advances in neural information processing systems, 23.

[25] YCharts. (n.d.). 10 Year Treasury Rate. Ανάκτηση από https://ycharts.com/indicators/10_year_treasury_rate

[26] Doshi, K. (2020). Reinforcement Learning Explained Visually (Part 5): Deep Q Networks, step-by-step. Towards Data Science. https://towardsdatascience.com/reinforcement-learning-explained-visually-part-5-deep-q-networks-step-by-step-5a5317197f4b

[27] Statista. (n.d.). Cryptocurrency market value. Ανάκτηση από https://www.statista.com/statistics/730876/cryptocurrency-maket-value/

[28] CoinMarketCap. (n.d.). Bitcoin market capitalization. Ανάκτηση από https://coinmarketcap.com/currencies/bitcoin/