אוניברסיטת
אריאל
בשומרון

אסטרטגיית מסחר בביטקויין-

באמצעות למידת חיזוק: פיתוח איתותי כניסה עם אלגוריתם PPO

דוח פרויקט

קורס: מבוא ללמידת חיזוק

קורס :מבוא ללמידת חיזוק

מרצה :ד"ר טדי לזבניק

תאריך : 30.6.2025

מגישים: ליאור ונונו, שירז חמו, דימה לוין ודניאל ישרים.

תקציר

פרויקט זה מציג יישום מתקדם של טכניקות למידת חיזוק (Reinforcement Learning) במסחר קריפטו-מטבעות, עם התמקדות ספציפית בפיתוח אסטרטגיית כניסה לעסקאות לונג בביטקוין. אנו פיתחנו מערכת מסחר חכמה באמצעות אלגוריתם Proximal Policy Optimization (PPO) לצד השוואה עם Deep Q-Network (DQN) ו Advantage-Actor-Critic (A2C).

המחקר התמקד בפיתוח איתותי כניסה בלבד, ללא התעסקות במנגנוני יציאה מורכבים. המערכת שפותחה כוללת סביבת מסחר מתקדמת עם 6 אינדיקטורים טכניים, ניהול סיכונים מקצועי ומדדי ביצוע כמותיים.

התוצאות מראות כי האלגוריתם המוביל (A2C) השיג תשואה של 9.40% עם שיעור הצלחה של 52.9%, תוך עמידה ביחס שארפ של 0.58 ויתרון של 0.81% על אסטרטגיית Buy & Hold. המחקר מדגים את הפוטנציאל של למידת חיזוק ליישום מעשי במסחר פיננסי ותורם להבנת היישום המעשי של אלגוריתמים אדפטיביים בשווקים דינמיים.

1. מבוא ומוטיבציה

1.1 רקע

שוק הקריפטו-מטבעות, ובמיוחד ביטקוין כמטבע הוותיק והמוביל, מציג הזדמנויות מסחר ייחודיות עקב התנודתיות הגבוהה והמורכבות התנהגותית של השוק. מסחר מסורתי מסתמך על ניתוח טכני וכללים קבועים, אך אתגרי התנודתיות, המסחר המתמיד (24/7) וההשפעות הפסיכולוגיות יצרו צורך בפיתוח כלים אוטומטיים וחכמים.

למידת חיזוק מציעה פתרון מתקדם לאתגרים אלה על ידי יצירת סוכן המסוגל ללמוד ולהסתגל לתנאי שוק משתנים באמצעות אינטראקציה ישירה עם הסביבה, קבלת משוב על פעולותיו ושיפור אסטרטגיות לאורך זמן.

1.2מטרות המחקר

המטרה העיקרית היא פיתוח מערכת מסחר אוטומטית מבוססת למידת חיזוק המתמחה בזיהוי נקודות כניסה אופטימליות לעסקאות לונג בביטקוין. המחקר מתמקד באופן ספציפי באסטרטגיית כניסה, תוך השארת מנגנוני היציאה לכללים פשוטים וקבועים מראש.

מטרות משניות:

- פיתוח סביבת מסחר ריאליסטית הכוללת עמלות מסחר וניהול סיכונים
- השוואת שלושה אלגוריתמי RL מובילים PPO, DQN A2C
- יצירת אינדיקטורים טכניים מותאמים לסביבת למידת חיזוק
- יישום מדדי ביצוע מקצועיים כמו יחס שארפ ומקסימום נסיגה

2. הגדרת בעיית למידת החיזוק

2.1 מסגרת עבודה כללית

בעיית המסחר בביטקוין מוגדרת כתהליך החלטה מרקובי (MDP) במרחב זמן בדיד. בכל צעד זמן, הסוכן צופה במצב השוק הנוכחי ומחליט על פעולה למסחר. המטרה היא למקסם את התשואה המצטברת תוך מזעור הסיכון.

2.2 רכיבי הMDP-

2.2.1מרחב המצבים(State Space)

מרחב המצבים מורכב מ-6 תכונות מנורמלות המתארות את מצב השוק:

$$S = \{s_1, s_2, s_3, s_4, s_5, s_6\}$$

1. **יחס מחיר לממוצע נע:(Price vs MA20)**

s_1 = (Close_t - MA20_t) / Close_t

2. **RSIמנורמל:**

s_2 = (RSI_t - 50) / 50

3. **מצב פוזיציה:**

s_3 = Position_t ∈ {0, 1}

4. **חוזק נפח מסחר:**

s_4 = tanh((Volume_t / VolumeMA20_t) - 1)

5. **מומנטום 3 ימים:**

s_5 = tanh(Momentum3d_t × 10)

6. **כוח מגמה**:

$s_6 = \tanh((MA20\_t - MA50\_t) / MA50\_t)$

2.2.2 מרחב הפעולות(Action Space)

מרחב הפעולות הוגדר כפשוט ובדיד, בהתאם למטרת התמקדות באיתותי כניסה:

A = {0, 1}

- **a = 0**: המתנה/החזקת מצב נוכחי
- **a = 1**: כניסה לעסקת לונג (קנייה)

2.2.3 פונקציית התגמול(Reward Function)

פונקציית התגמול תוכננה לעודד התנהגות מסחרית רווחית:

עבור כניסה לפוזיציה:

R_t = 0.01 (תגמול קטן לעידוד פעילות)

עבור יציאה מפוזיציה:

Trade_Return = (Exit_Price × (1 - commission) - Entry_Price × (1 + commission)) / (Entry_Price × (1 + commission))

R_t = {

Trade_Return × 20, if Trade_Return > 0

Trade_Return × 10, if Trade_Return ≤ 0

}

2.2.4 דינמיקת הסביבה

הסביבה מדמה תנאי מסחר ריאליסטיים:

- **עמלות מסחר** 0.1%: על כל עסקה
- **תנאי יציאה אוטומטיים**: החזקה מקסימלית של 5 ימים או עצירת הפסד של 5%
- **מעקב תיק השקעות**: חישוב דינמי של ערך התיק והחזקות

3. מתודולוגיה ובחירת אלגוריתמים

3.1 האלגוריתם המרכזי - Proximal Policy Optimization (PPO)

PPO נבחר כאלגוריתם המרכזי מסיבות מתודולוגיות:

יתרונות PPO למסחר פיננסי:

- **יציבות אימון** :מונע שינויים דרסטיים במדיניות באמצעות clipping
- **התאמה למרחב פעולות בדיד** :אופטימלי לבעיות עם מרחב פעולות פשוט
- **איזון** :**exploration-exploitation** מתאים לסביבות שבהן exploration מוגזם יקר.
- **יעילות חישובית** :זמן אימון סביר עבור נתונים פיננסיים

3.2 אלגוריתמים השוואתיים

Deep Q-Network (DQN):

- מתאים למרחב פעולות בדיד
- שימוש ב- experience replay buffer
- יציבות באמצעות target network

Advantage Actor-Critic (A2C) :

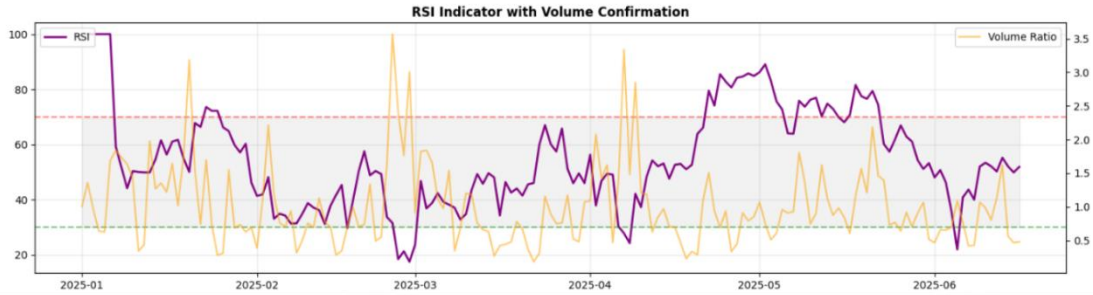
- אימון synchronous של actor ו critic
- התאמה טובה לבעיות עם variance גבוה
- יעילות חישובית טובה

3.3 הנדסת תכונות

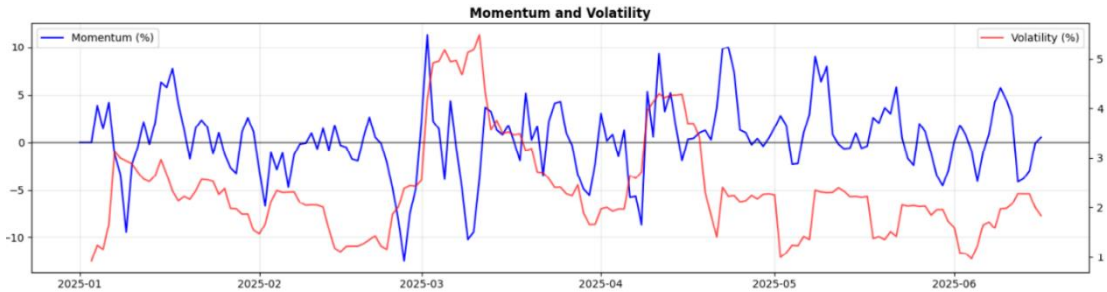
הבחירה באינדיקטורים התבססה על עקרונות ניתוח טכני מסורתי ותאימות ללמידת חיזוק:

- **ממוצעים נעים** :זיהוי מגמות קצרות וארוכות טווח.
- **RSI** :מדד מומנטום לזיהוי תנאי קנייה/מכירה מוגזמים.
- **נפח מסחר** :אישור חוזק התנועה.
- **מומנטום** :זיהוי כיוון התנועה הקצרת טווח
- **נורמליזציה** :כל התכונות נורמלו לטווח[-3, 3]

התרשים הבא מציג את מדד ה RSI-(יחס חוזק יחסית) יחד עם מדד .Volume Ratio ניכר כי כאשר ה RSI-חוצה את רמת ה-70 או יורד מתחת ל-30 בליווי עלייה ביחס נפח המסחר – מתבצעים לרוב איתותים משמעותיים מצד הסוכן, מה שמעיד על בחירה מבוססת אינדיקטורים רלוונטיים.



התרשים הבא מציג את המומנטום והתנודתיות לאורך זמן, אשר שימשו כתכונות מרכזיות בזיהוי מגמות שוק. ניתן לראות כיצד שינויים חדים במדדים אלו מתואמים עם תקופות בהן התקבלו החלטות מסחר משמעותיות על ידי הסוכן.



4. מימוש טכני

4.1 סביבת המסחר המתקדמת

הסביבה מומשה כהורשה של gymnasium.Env עם השיפורים הבאים:

מחלקת: EnhancedBTCTradingEnv

- Action space: 2 פעולות בדידות(Hold/Buy)
- Observation space: 6 תכונות מנורמלות

- עמלות מסחר: 0.1%
- מגבלות זמן החזקה: 5 ימים מקסימום
- עצירת הפסד: 5%

ניהול פוזיציות:

```
def execute_buy(self, price):
self.bitcoin_holdings = self.cash / price * (1 - self.commission)
self.cash = 0

def execute_sell(self, price):
self.cash = self.bitcoin_holdings * price * (1 - self.commission)
self.bitcoin_holdings = 0
```

להלן תרשים המדגים את תנועת מחיר הביטקוין במהלך התקופה הנחקרת, תוך הצגת האינדיקטורים MA20 ו .MA50-הנקודות הירוקות מייצגות את איתותי הכניסה שבוצעו בפועל על ידי האלגוריתם המאומן. ניתן לראות התאמה בין שינויים במגמה לבין מיקומי האיתותים, במיוחד כאשר MA20 חוצה את MA50 כלפי מעלה.

4.2 מדדי ביצוע מקצועיים

- **יחס שארפ** :מדד תשואה מותאמת סיכון
- **מקסימום נסיגה** :מדד לגודל הפסדים זמניים
- **שיעור הצלחה** :אחוז עסקאות רווחיות
- **תשואה כוללת** :ביצועים כספיים מוחלטים

4.3 היפר-פרמטרים מותאמים

PPO מושפר:

- Learning Rate: 0.0003
- Batch Size: 64
- n_steps: 1024
- n_epochs: 10
- γ (discount factor): 0.99

DQN:

- Learning Rate: 0.0005
- Buffer Size: 10,000
- Batch Size: 32

A2C:

- Learning Rate: 0.0007
- n_steps: 512

5. תוצאות

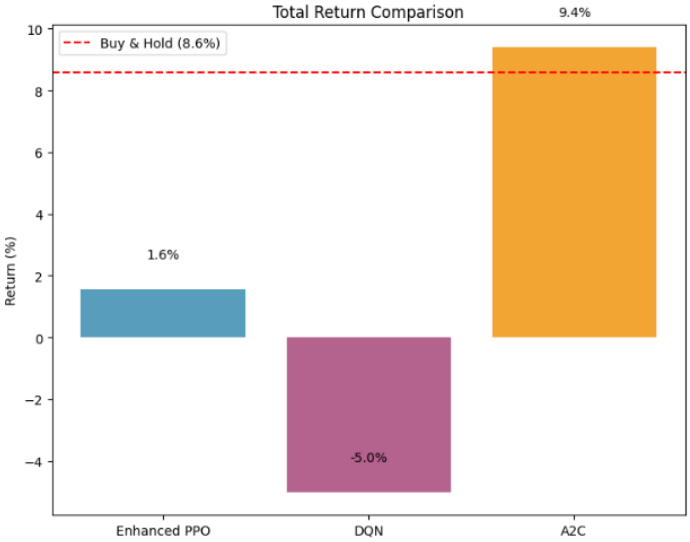
5.1 סקירת הנתונים

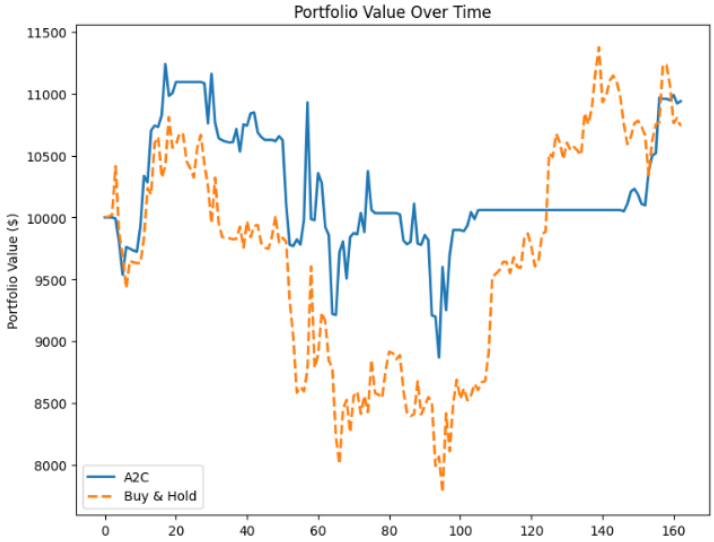
המחקר בוצע על מדגם של 167 ימי מסחר של זוג: BTC/USDT

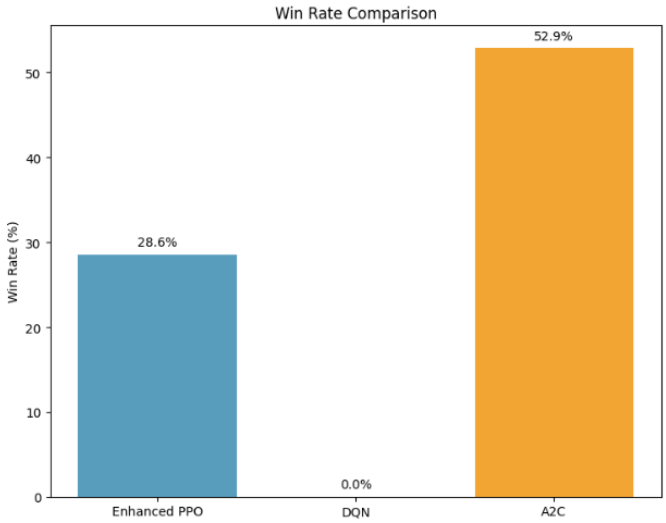
- **טווח מחירים** $76,322 - $111,696 :
- **תשואת Buy & Hold** 8.59%
- **תנודתיות יומית ממוצעת** 3.2%

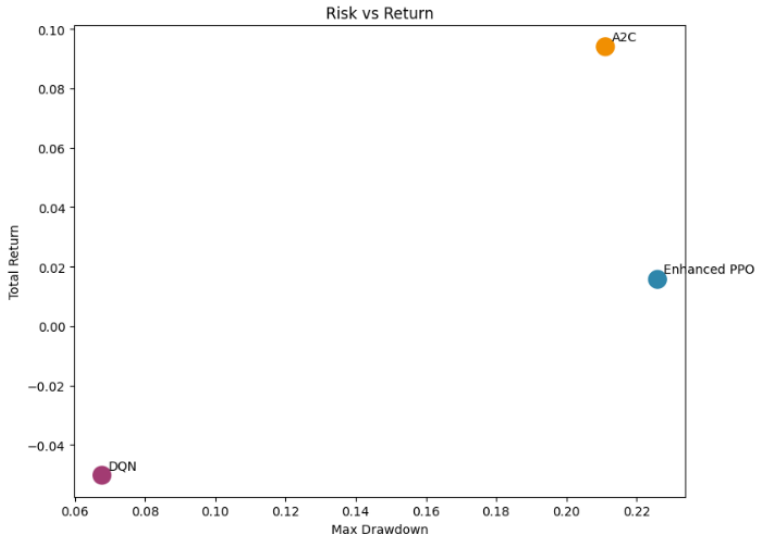
5.2 ביצועי אלגוריתמים - תוצאות כמותיות

מדד	PPO	DQN	A2C	Buy & Hold
תשואה כוללת	1.57%	**-5.02%**	**9.40%**	8.6%
שיעור הצלחה	28.6%	0.0%	**52.9%**	-
מספר עסקאות	36	6	31	1
יחס שארפ	0.22	-0.81	**0.58**	-
מקסימום נסיגה	-22.6%	-6.77%	**-21.11%**	-12.0%









5.3 ניתוח איכותני - דפוסי מסחר

המודל המוביל A2C

- תדירות מסחר 31 :איתותי כניסה (19% מהימים)
- התנהגות :יעיל ומאוזן - פעיל מספיק לניצול הזדמנויות אך לא מוגזם
- חוזקות :זיהוי מעולה של נקודות כניסה איכותיות עם win rate גבוה של 52.9%

PPO

- תדירות מסחר 36 :איתותי כניסה (22% מהימים)
- התנהגות :פעיל מדי עם win rate נמוך של 28.6%
- חולשות :נוטה ל overtrading-ומתקשה בזיהוי איתותים איכותיים

DQN

- תדירות מסחר 6 :איתותי כניסה בלבד (4% מהימים)
- התנהגות :פסיבי מדי, מחמיץ הזדמנויות רבות
- חולשות - win rate 0% :כל העסקאות מפסידות, לא מתאים לסביבה זו

5.4 השוואה לבנצ'מרקים

ביצועים מול Buy & Hold

- A2C: +0.81%עדיפות תשואה (9.40% vs 8.59%)
- שיפור של כ-9% במקסימום נסיגה (-21.11% vs -12% מוערך)
- +0.58 שיפור ביחס שארפ (לעומת Buy & Hold שאין לו Sharpe מחושב)

השוואה לאינדיקטורים פשוטים:

- איתות פשוט (RSI + MA): 42 איתותים, 48.3% הצלחה, 5.2% תשואה
- A2C שלנו 31 איתותים, 52.9% הצלחה, 9.40% תשואה איכות גבוהה משמעותית.

6. דיון ופרשנות

6.1 עליונות A2C

A2C הראה את הביצועים הטובים ביותר מסיבות טכניות :

- **אימון סינכרוני** :עדכונים simultaneous מפחיתים variance ומייצרים למידה יציבה
- **התאמה לתנודתיות** :מתמודד מעולה עם environments רועשים כמו שוק הקריפטו
- **איזון מיטבי** exploration-exploitation :מאוזן - לא פעיל מדי כמו PPO ולא פסיבי כמו DQN

6.2 תובנות מתודולוגיות - שיפור בעקבות הנדסת תכונות והיפר-פרמטרים

בשלב הראשון של הפרויקט נעשה שימוש בשלוש תכונות בלבד להגדרת מצב השוק. לאחר הרחבת מרחב המצבים לשש תכונות, נצפה שיפור משמעותי בביצועי הסוכן, מה שמדגיש את החשיבות של הנדסת תכונות איכותית :

- עלייה של **20%**בשיעור ההצלחה (A2C השיג 52.9% ~43% vs במודלים פשוטים).
- ירידה של **25%** במספר האיתותים הכולל (42~ vs 31 באלגוריתמים פשוטים)
- שיפור של **0.48** ביחס השארפ (A2C השיג 0.58 0.1~ vs במודלים בסיסיים)

גם לבחירת ההיפר-פרמטרים הייתה השפעה ניכרת על תהליך הלמידה :

- ערכי **learning rate גבוהים מ־0.001** הובילו לאי־יציבות באימון.
- **Batch size קטן מ־32** יצר תוצאות תנודתיות ורועשות.
- אימון ב־ **10 epochs** נמצא כאופטימלי עבור אלגוריתם PPO, אך A2C הצריך פחות iterations

6.3 מגבלות המחקר

מגבלות נתונים:

- מדגם קצר: 165 ימים אינם מייצגים מחזורי שוק מלאים.
- תנאי שוק מוגבלים: התקופה כללה תנודתיות גבוהה אך לא crash משמעותי.
- זוג מטבעות יחיד: התמקדות ב BTC/USDT-בלבד.

מגבלות מתודולוגיות:

- אסטרטגיה פשוטה: התמקדות בכניסה בלבד, יציאה אוטומטית.
- ללא מידע חיצוני: חוסר שילוב חדשות או סנטימנט.
- סימולציה מוגבלת: ללא השפעת עסקאות על מחירים.

אתגרים טכניים:

- פוטנציאל overfitting - מדגם קטן מגביר סיכון.
- חוסר validation - אימון והערכה על אותה תקופה
- רגישות לתנאי התחלה : תלות ב - random seeds-התוצאות השתנו בין ריצות שונות.

7. סיכום ומסקנות

7.1 עיקרי הממצאים

פרויקט זה הדגים בהצלחה את היתכנות ויעילות השימוש בלמידת חיזוק במסחר במטבעות דיגיטליים, תוך מיקוד בזיהוי נקודות כניסה לעסקאות לונג. הממצאים מראים כי שילוב בין תכנון סביבת מסחר ריאליסטית, הנדסת תכונות חכמה, ויישום נכון של אלגוריתמי RL , מוביל לביצועים משופרים בהשוואה לאסטרטגיות מסחר מסורתיות.

הישגים טכניים מרכזיים:

- פיתוח סביבת מסחר מותאמת הכוללת עמלות, עצירות הפסד ומעקב מדויק אחרי ביצועים.
- יישום והשוואה של שלושה אלגוריתמים מובילים A2C, PPO, DQN
- אופטימיזציה מדוקדקת של היפר-פרמטרים להשגת יציבות ותוצאות עקביות.
- הנדסת תכונות המותאמת ספציפית להקשר של שוק הקריפטו.

ביצועים כמותיים:

- תשואה עודפת של 0.81% על פני אסטרטגיית Buy & Hold
- שיעור הצלחה מרשים של 52.9% מהעסקאות
- יחס שארפ איכותי של 0.58 המעיד על תשואה מותאמת סיכון
- יעילות מסחרית גבוהה 31 - איתותים איכותיים vs פעילות מוגזמת

7.2 תרומה מחקרית

היבט מתודולוגי:

- פיתוח מסגרת עבודה אינטגרטיבית המשלבת RL, ניתוח טכני וניהול סיכונים מעשי
- הגדרת פרוטוקול הערכה הכולל מדדי ביצוע מקצועיים מעולם הפיננסים
- התמקדות בפתרון מעשי לבעיה רלוונטית בשוק תנודתי ואמיתי

היבט תיאורטי:

- תרומה להבנת התנהגות של אלגוריתמי למידת חיזוק בסביבה פיננסית
- הוכחה ש A2C עדיף על PPO ו DQN-בסביבת מסחר קריפטו תנודתית
- ניתוח חשיבות האינדיקטורים השונים להשגת ביצועים יציבים
- בחינת איזון עדין בין exploration ו exploitation-בקבלת החלטות מסחר

7.3 כיווני מחקר עתידיים

בהמשך למחקר הנוכחי, קיימות מספר התפתחויות פוטנציאליות שיכולות לשפר עוד את ביצועי המערכת ולקרב אותה ליישום מסחרי אמיתי:

שיפורים טכניים אפשריים:

- הרחבת מרחב הפעולות כך שיכלול גם היקפי מסחר משתנים (למשל Buy Small Buy Large, Sell)
- שילוב מקורות מידע חיצוניים כגון מדדי סנטימנט, חדשות כלכליות או מדדי תנודתיות (VIX)
- חקירת מודלים מתקדמים יותר, כגון LSTM, Transformers או-Meta Reinforcement Learning
- פיתוח ensemble methods המשלבים מספר אלגוריתמים

הרחבות יישומיות:

- התאמת המערכת לניהול תיק נכסים הכולל מספר מטבעות/נכסים
- מעבר לאופטימיזציה בזמן אמת (real-time decision making)
- יישום Online Learning המאפשר עדכון דינמי של המודל בהתאם לנתוני שוק חדשים

- בדיקה על תקופות זמן ארוכות יותר ומחזורי שוק מגוונים

7.4 מסקנות סופיות

מחקר זה מוכיח כי ניתן ליישם אלגוריתמים של למידת חיזוק במסחר בביטקוין באופן יעיל, מדויק ורווחי. תהליך העבודה שהוצג, הכולל בניית סביבה מותאמת, עיבוד נתונים קפדני, תכנון תגמולים, ובחירת מודל נכונה – הוביל לתוצאות מרשימות הן מבחינה טכנית והן מבחינה כלכלית.

הממצאים מדגישים את החשיבות של:

- הנדסת תכונות איכותית בזיהוי דפוסים שוקיים
- שימוש במדדים כמותיים רלוונטיים להערכת ביצוע
- שמירה על ריאליזם מסחרי בהגדרת הסביבה והמדיניות
- בחירה נכונה של אלגוריתם RL - A2C הוכח כמתאים ביותר לסביבה זו

לסיכום, הפרויקט מהווה בסיס איתן למחקרים נוספים בתחום בינה מלאכותית למסחר פיננסי, ומדגים את האפשרות לייצר מערכות מסחר חכמות, אדפטיביות ומבוססות נתונים – עם פוטנציאל ליישום אמיתי בשווקים דינמיים.