

שוק ההון-TIMING THE MARKETS

פרויקט בבינה מלאכותית, הפקולטה למדעי המחשב בטכניון, סמסטר אביב 2020



אבישי אביסרור 311378947
AVISHAYA@CAMPUS.TECHNION.AC.IL

יקיר יהודה 205710528
Y.YAKIR@CAMPUS.TECHNION.AC.IL

מבוא

שנת 2020 נפתחה עם מגפת הקורונה, שיצרה משבר עולמי בכל תחומי החיים, ובפרט בשוק ההון שהתרסק בצורה היסטורית וחסרת תקדים^[1].

מאז ומתמיד **משברים בשוק ההון נחשבו להזדמנות** – משקיעים נוהגים לנצל משברים על מנת לקנות ניירות ערך במחירים אטרקטיביים, ולהרוויח את העלייה המגולמת בהתאוששות מהמשבר.

הדרך לניצול הפוטנציאל הנ"ל רצופת מכשולים, ולהלן העיקריים שבהם:

1. קשה מאוד **לחזות** אילו חברות ישרדו את המשבר. כך למשל, לאחר משבר הסאב-פריים בשנת 2008 חברות רבות ירדו מנכסיהן ופשטו רגל^[2].

2. קשה מאוד **לתזמן את הכניסה לשוק** – זוהי בעצם נקודת המינימום של נייר הערך המועמד לקנייה, המאפשרת למשקיע את התזמון המושלם להיכנס לשוק במחיר אטרקטיבי.

שיטה נפוצה שהתפתחה במרוצת השנים להתמודדות עם הבעיה הראשונה היא **פיזור**. המשקיע הממוצע יחלק את הונו בין אפשרויות השקעה שונות, כדי להיות חסין מירידות ערך מאסיביות, ולייצר רשת ביטחון הגונה.

אחת הדרכים להשיג את הפיזור הנ"ל היא מכשיר פיננסי בשם **תעודת סל**. זוהי תעודה המונפקת על ידי גוף פיננסי (בנק או בית השקעות), ומאפשרת למשקיע לעקוב אחר נכס כגון **מדדי מניות** (מדדים המהווים ממוצע משוקלל של מספר רחב של מניות).

ההתמודדות עם הבעיה השנייה היא **שאלת מיליון הדולר של עולם ההשקעות**, ורבים מאמינים שבלתי אפשרי לחזות את התנהגות השוק. במרוצת השנים התפתחו שתי אסכולות מקבילות סביב השאלה הזו: Timing the markets VS Time in the markets^[3] (השקעה לטווח הארוך אל מול תזמון השוק).

בפרויקט זה ננסה לאתגר אלגוריתמים שונים של בינה מלאכותית ולמידה עמוקה עם היכולת לתזמן את השוק בזמן משבר, ולמקסם את הרווח של המשקיעים = "קנה בזול ומכור ביוקר".

תיאור הפתרון המוצע לבעיה

רקע:

ב-30 השנים האחרונות שוק ההון התפתח במספר מובנים:

- ידע **משברים** רבים, מהם ניתן ללמוד על אופן ההתאוששות של השוק.
- נוספו **אינדיקטורים** רבים שנועדו לצוד תופעות שונות בשוק במטרה להבין לאן פניו. האינדיקטורים הללו משמשים משקיעים רבים ככללי אצבע בתכנון ההשקעות שלהם.

ברצוננו, בהינתן נייר ערך מסוים, ויום מסוים בעיצומו של משבר, לתת חיזוי לגבי מידת הכדאיות של כניסה לשוק. יום שפוטנציאל הרווח בו הוא גדול יקבל ציון קרוב ל-1, ויום בעל פוטנציאל רווח מינימלי יקבל ציון קרוב ל-0. בהמשך נסביר כיצד להשתמש בציון הזה ברמה הפרקטית – ואף נציע מספר אסטרטגיות הנשענות על הציון.

לשם כך, נשתמש במשברים קודמים כדוגמאות אימון לאלגוריתם הלמידה שלנו. כל יום מסחר בעיצומו של משבר יהווה דוגמא.

- **התכונות** של יום המסחר הם אינדיקטורים ממקורות שונים בהם האלגוריתם יתחשב.
- **התיוג** של יום המסחר הוא פוטנציאל הרווח שלו (על פי חישוב שיפורט בהמשך).

תמצית הפתרון:

הפתרון יורכב משני שלבים עיקריים:

1. בניית Regressor: מטרתו, בהינתן עשרות תכונות אשר יפורטו, ומספר רחב של דוגמאות היסטוריות מהבורסה, לחזות את פוטנציאל הרווח של כל יום מסחר בזמן משבר (אותו ציון בין 0 ל-1 אותו הזכרנו, ונרחיב עליו בהמשך) בצורה המדויקת ביותר.
2. הפיכת המידע לידע: נרצה, בהינתן תוצאות ה-Regressor, לגזור משמעויות עסקיות. קרי, נבנה נוסחאות סגורות לאסטרטגיות השקעה המבוססות על ציוני ה-Regressor, ונבחן את ביצועיהן לאורך הזמן.

תיאור שלבי המערכת:

בשלב זה נתאר בפרוטרוט את החלקים השונים במערכת, וניגע באתגרים המרכזיים בכל אחד מהם.

1. Data Extractor

חלק זה אמון על הגדרת "משבר" על סמך patterns של ביצועים. הוא מקבל נתונים של ביצועים היסטוריים של נייר ערך, וקובע תקופות בהיסטוריה בהם נייר ערך זה היה במשבר.

ההגדרה למשבר היא תקופה בזמן שעונה על הקריטריונים הבאים :

- אורך מינימלי : לא כל יום של ירידות בשווקים ייחשב כמשבר. נרצה להגדיר סף מינימום (במונחים של ימי מסחר) שמצביע על משבר משמעותי מספיק.
- פגיעת מינימום : לא כל רצף של ימי ירידה מעיד על משבר. נרצה שהירידה תהיה חדה מספיק (או מתונה וארוכה), ולכן נרצה להגדיר סטנדרט למחיר בשיא המשבר (נקודת המינימום ביחס למחיר המקורי).

סיום המשבר ייקבע ליום הראשון בו נייר הערך חזר להיות קרוב למחיר שממנו התחיל המשבר.

הערות :

- האורך המינימלי, פגיעת המינימום, ורמת הקירבה למחיר המקורי (סוף המשבר) הם היפר-פרמטרים של המודל, וייקבעו בהתאם למדידות על סמך ניסויים שנבצע בהמשך.
- קיימים משברים בהם נצפו ימי עליות בודדים אשר לא מעידים על סוף המשבר (לדוגמה, כתוצאה מהזרמה ממשלתית של כספים). כדי להתמודד עם הבעיה, נגדיר היפר-פרמטר נוסף שמאפשר מספר מסוים של ימי תיקון.

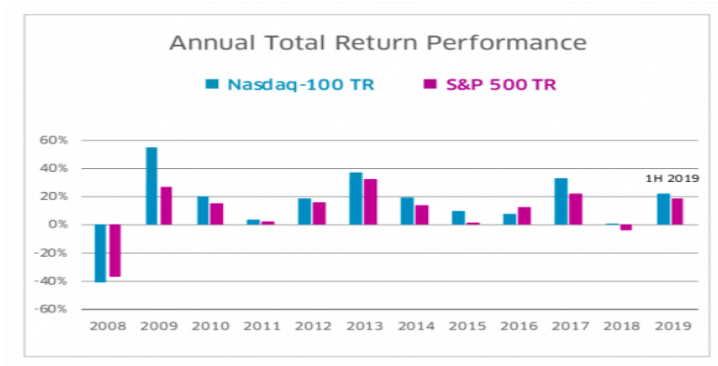
להלן ביצועי האלגוריתם על משברים בשלושת השנים האחרונות תחת פרמטרים של אורך מינימלי, פגיעת מינימום ומספר ימי תיקון שמצאנו כמדויקים ביותר (בכחול – ביצועי האמת של מדד S&P500 בבורסה האמריקאית, באדום – תקופות אותן האלגוריתם הגדיר כמשבר).



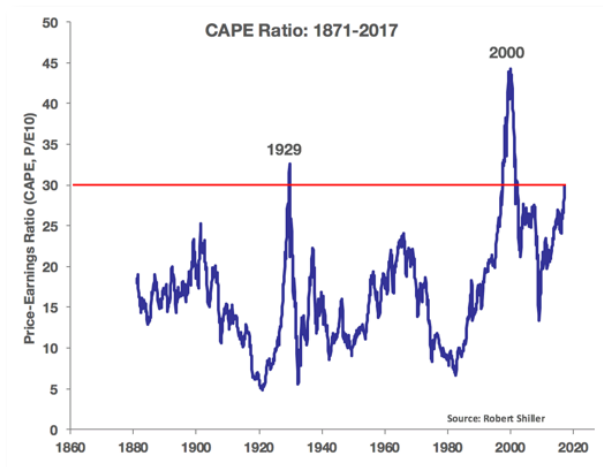
2. Data Combiner

חלק זה אמוך על איסוף אינדיקטורים ותכונות מימי המסחר שבתוך תקופות המשבר.
הגדרות :

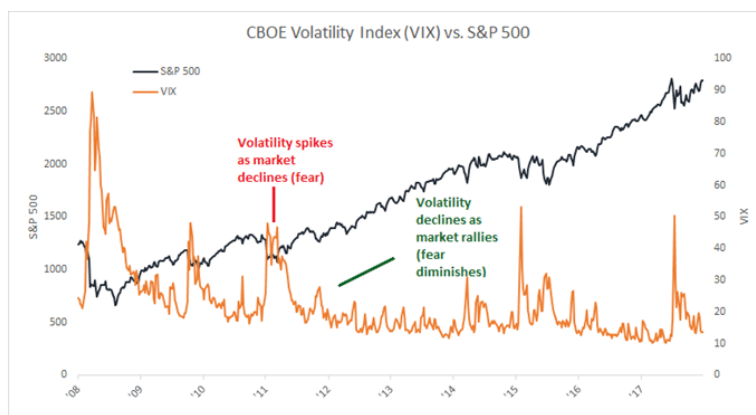
- Nasdaq, S&P500 : מדדים מובילים בעולם המניות האמריקאי. סחירים בבורסה באמצעות תעודות סל או קרנות מחקות (קרנות שקונות מניות בהתאמה להרכב המניות של המדד במטרה להתחקות אחר הביצועים שלו). ההיסטוריה מלמדת ששני המדדים מתנהגים בצורה דומה לאורך הזמן, ולכן כל אחד מהם מהווה אינדיקטור טוב לשני.



- אינפלציה : נתון המתאר את עליית המחירים במשק. נמדדת באמצעות מדד המחירים לצרכן, הבודק את השינוי במחירו של סל מוצרים קבוע (עלייה של מחירו גוררת עלייה באינפלציה). יש דרכים רבות להסביר את הקשר בין אינפלציה לבין ביצועי הבורסה. אחת מהן היא שירידה באינפלציה לרוב גוררת ירידה בריבית שנקבעת על ידי בנק ישראל, וזו מעודדת הוצאת כספים מפיקדונות והשקעה בשוק ההון. הגדלת מספר השחקנים בשוק מביאה לרוב להעלאת מחירים.
- Earning Ratio : היחס בין רווחי החברה לשווי המניה. זוהי דרך נפוצה לחשב רווח PER מניה, ולכן דרך מקובלת לנרמל את האטרקטיביות של החברות.
- Dividend Rate : מספר הדולרים שמשקיע מקבל מהחברה בה השקיע לכל מניה שקנה, כתוצאה מרווחי החברה. זוהי עוד דרך טובה לנרמל אטרקטיביות.
- Dividend Yield : היחס בין ה-Dividend Rate של החברה לבין מחיר המניה שלה.
- Price-Earnings Ratio : היחס בין שווי השוק של חברה לבין הרווח הנקי השנתי שלה. במדדים, מבצעים ממוצע משוקלל של היחס לפי החברות המשתתפות במדד.
- CAPE Ratio : מתאר את שווי השוק של מדד ה-S&P500 חלקי הרווח הממוצע, מתואם לאינפלציה, ב-10 השנים האחרונות. ערך גבוה של מדד זה לרוב יצביע על שוק שהוא overvalued (נסחר מעל לערכו), ולהפך. כך למשל בשנים 1929 ו-2000 (שנים ידועות במשברים, המדד עלה קיצונית למעלה).



- מדד הפחד (VIX) : מדד התנודתיות של בורסת האופציות בשיקאגו (CBOE), הנועד להתחקות אחר אלמנטים פסיכולוגיים של משקיעים בשוק. ירידה במדד נחשבת בעיני רבים למגמת התייצבות (כלומר ירידה במדד לרוב תבלום עליות או ירידות אגרסיביות). עלייה, לעומת זאת, נתפסת כפאניקה הנובעת מ"ראלי" (עליות חדות מאוד) או קריסה.



- APA : תאגיד נפט אמריקאי. משקיעים רבים רואים בו כמכשיר פיננסי שעוזר במעקב אחר מחיר הנפט בעולם. זה עשוי להיות אינדיקטור טוב לתקופות משבר כי משקיעים רבים עוברים להשקיע בסחורות בתקופות של אי וודאות בבורסה.
- GLD : קרן סל שמתיימרת לעקוב אחר מחיר הזהב בעולם. זה עשוי להיות אינדיקטור טוב לתקופות משבר כי משקיעים רבים עוברים להשקיע בסחורות בתקופות כנ"ל.
- TNX : תשואת אג"ח ממשלתי אמריקאי ל-10 שנים. ככלל, אג"חים ממשלתיים נחשבים לאפיק השקעה סולידי ובטוח, ואופן המסחר בהם קשור קשר הדוק לתקופות משבר. משקיעים רבים "בורחים" לאפיקים סולידים בעתות משבר.

תכונות בתדירות יומיות :

1. מחיר סגירה של S&P500 (Close) : מחיר המניה בתום יום מסחר.
2. מחיר גבוה של S&P500 (High) : העסקה היקרה ביותר (לפי מחיר מניה בודדת) שהתבצעה באותו יום.
3. מחיר נמוך של S&P500 (Low) : העסקה הזולה ביותר (לפי מחיר מניה בודדת) שהתבצעה באותו יום.
4. מחזור מסחר (Volume) : שווי כולל של העסקאות עבור מניה זו שבוצעו באותו יום.
5. מחיר סגירה של nasdaq (Nas-Close).
6. מחיר גבוה של nasdaq (Nas-High).
7. מחיר נמוך של nasdaq (Nas-Low).
8. מחזור מסחר של nasdaq (Nas-Volume).
9. ערך סגירה של מדד הפחד (Vix-Close).
10. ערך גבוה של מדד הפחד (Vix-High).
11. ערך נמוך של מדד הפחד (Vix-Low).
12. מחיר סגירה של APA (Apa-Close).
13. מחיר גבוה של APA (Apa-High).
14. מחיר נמוך של APA (Apa-Low).
15. מחזור מסחר של APA (Apa-Volume).
16. מחיר סגירה של GLD (Gld-Close).
17. מחיר גבוה של GLD (Gld-High).
18. מחיר נמוך של GLD (Gld-Low).
19. מחזור מסחר של GLD (Gld-Volume).
20. מחיר סגירה של TNX (Tnx-Close).
21. מחיר גבוה של TNX (Tnx-High).
22. מחיר נמוך של TNX (Tnx-Low).
23. מחזור מסחר של TNX (Tnx-Volume).

תכונות בתדירות חודשית :

1. אינפלציה.
2. Earning Ratio.
3. Dividend Rate.
4. Dividend Yield.
5. Price-Earning Ratio.
6. CAPE Ratio.

הערות :

1. **כדי שנוכל להשתמש במסווג בזמן אמת, נרצה שהתכונות יתבססו על העבר בלבד.**
לכן בעת סיווג של יום מסחר, המסווג לא ייקח בחשבון תכונות של יום המסחר הנוכחי, אלא של קודמיו.
2. מאותה סיבה, התכונות החודשיות הן במבט לאחור. כך למשל, ביום מסחר של חודש מרץ, נוכל להסתכל על הנתונים החודשיים של פברואר.
3. כל תכונה מפוצלת בפועל למספר תכונות, על מנת שנוכל לייצר זיכרון למודל על סמך תקופות זמן משמעותית יותר. כך למשל, נרצה לקחת בחשבון תכונות יומיות של חמישה ימים אחורה, ולא של יום בודד. האורך המדויק הוא היפר פרמטר של המודל (lookback).
4. עבור כל מדד שבחרנו לקחת, ניתן לראות שלקחנו מספר פרמטרים שלו (מחיר גבוה, מחיר נמוך, מחזור מסחר וכו'). הסיבה לכך היא שלכל פרמטר יש את החשיבות שלו. כך למשל, משקיעים רבים נוטים להתייחס לירידה דרסטית במחזור המסחר כסימן מקדים להיפוך מגמה (למשל, כאשר השוק מיצה תקופה של ירידות, מחזור המסחר יקטן והמגמה תתהפך).

3. Features Converter

כפי שראינו בחלק הקודם, חלק מהתכונות הן בתדירות יומית וחלק בתדירות חודשית. חלק זה אמון על "הפיכת" התכונות החודשיות ליומיות, כדי שנוכל להשתמש בהן במסווג (שכן הדוגמאות שהמסווג מקבל הן ברמה היומית).

חלק מהתכונות ברמה החודשית מתקבלות ביום הראשון של החודש, וחלק ביום האחרון של החודש. כדי לוודא שבכל יום, כל התכונות עליהן אנחנו מסתמכים זמינות לנו **מבעוד מועד** :

- עבור תכונות של סוף חודש – פיצלנו את ערך התכונה לכל יום בחודש הבא.
- עבור תכונות של תחילת חודש – פיצלנו את ערך התכונה לכל יום באותו החודש.

כך למשל, תכונה שהתקבלה ב31.3, תיוחס לכל דוגמא מימי חודש אפריל. תכונה שהתקבלה ב1.3, תיוחס לכל דוגמא מימי חודש מרץ.

4. Model

חלק זה אמון על תיוג הדוגמאות.

לכל משבר (כפי שהגדרנו, מאופיין בתאריך התחלה ותאריך סוף), נחשב את מחיר המינימום $price_{min}$ ואת מחיר המקסימום $price_{max}$.

נגדיר פוטנציאל רווח למשבר באופן הבא : $profit_{pot} = price_{max} - price_{min}$.

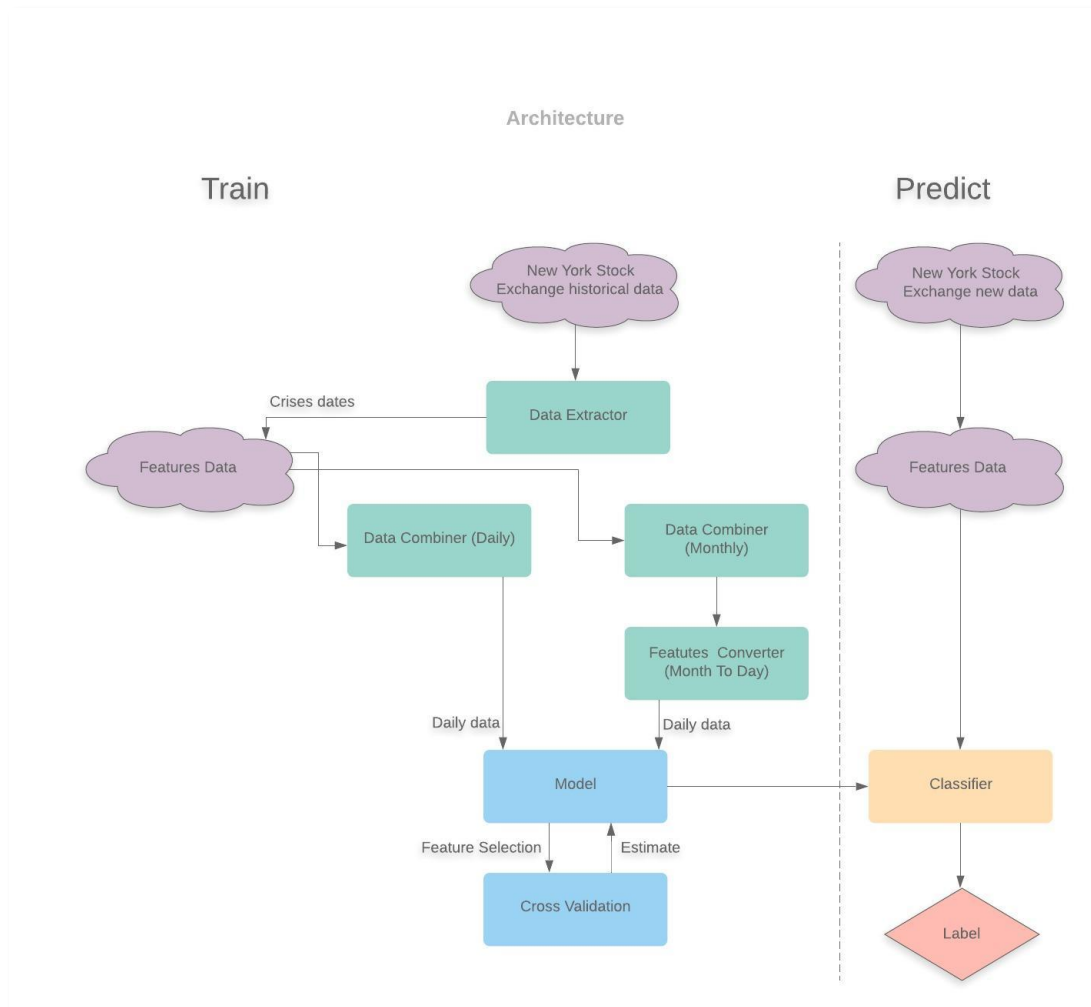
פוטנציאל הרווח של יום מסוים (דוגמה) בתוך המשבר יהיה : $Y = \frac{price_{max} - price_{day}}{profit_{pot}}$.

(הערה : הרעיון לחישוב הציון של יום למעשה מחקה את הקונספט של min-max normalization).

5. Cross Validation

חלק זה אמון על בחירת תתי-קבוצות שונות של פיצ'רים (בהתאם לאלגוריתמים שיפורטו בהמשך) ואלגוריתם למידה במטרה למצוא אלגוריתם ותת-קבוצה אידיאליים שייבחרו למסווג. הערכת הביצועים תיעשה על ידי K-fold Cross Validation. התוצאות יפורטו בפרק הבא.

מבט-על:



SVM, Random Forest, KNN אלגוריתמי

בשלב זה נתאר ניסויים שעשינו על אלגוריתמי בינה, בהם היינו עדים לקושי להתמודד עם בעיה המבוססת על Time Sequences. תפקידו של חלק זה להדגיש שאלגוריתמי בינה סטנדרטיים מתקשים לענות על הציפיות, וזוהי המוטיבציה למעבר לרשתות נוירונים עם זיכרון (יפורט בהמשך).

בחרנו מספר אלגוריתמים לבדיקה, ומספר דרכים לבצע Feature Selection.

לכל אלגוריתם, בדקנו את כל מנגנוני ה- Feature Selection, ועבור תת הקבוצה של תכונות שהתקבלה, בדקנו את רמת הדיוק שלה ע"י Cross-Validation.

חזרנו על הניסוי המתואר עבור מניפולציות שונות שביצענו על ה-Data.

: Data Manipulation

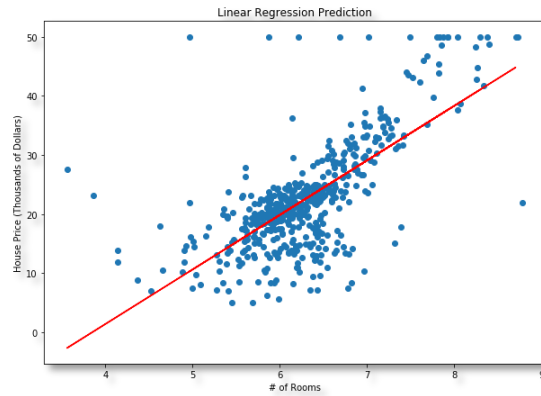
1. Original Data : שימוש ב-Data שתואר בחלק הקודם כפי שהוא.
2. Normalized Data : שימוש ב-Data מנורמל לפי אלגוריתם min-max normalization^[4].
3. Standardized Data : לכל פיצ'ר ולכל ערך, הערך המנורמל יהיה חיסור הממוצע מהערך המקורי, וחלוקה בשונות (שיטת z-score כפי שנלמדת בקורס מבוא למערכות לומדות).

: Feature Selection

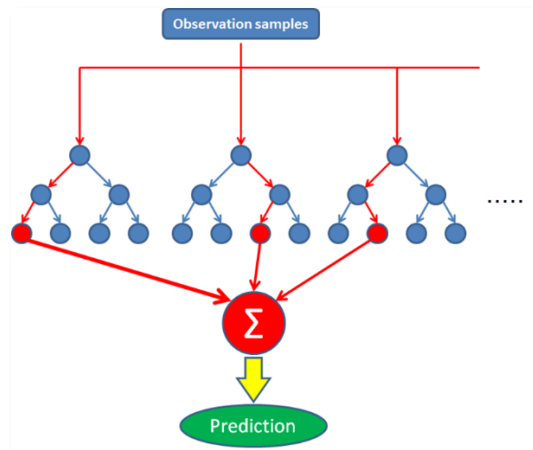
1. selectKBestFeatures : דירוג הפיצ'רים הטובים ביותר על פי פונקציית score, ובחירת k הטובים ביותר תחת הפונקצייה הזו.
פונקציית ה-score שהשתמשנו בה היא mutual_info_regression, המנקדת פיצ'רים בהתאם להערכת האנדרופיה שלהם^[5].
2. selectKExtraTree : מבוסס על ExtraTreeRegressor^[6]. אלגוריתם זה בונה יער של עצי החלטה על פי ה-Data, ולאחר מכן דוגם את היער עם הדוגמאות ומפחית את הסיכון ל-overfitting. בדרך זו, הוא נותן ציון בין 0 ל-1 לכלל הפיצ'רים שעומדים לרשותו. תחת ההנחה זו, בחרנו את k הטובים ביותר.

: אלגוריתמים

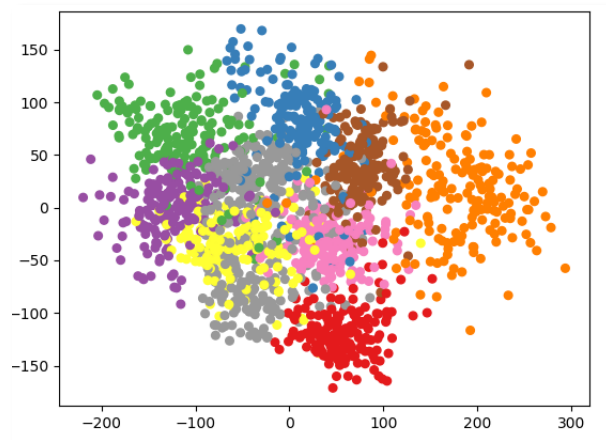
1. SVR : אלגוריתם Support Vector Regressor המבוסס על קירוב לינארי של דוגמאות. מטרתו למזער את סכום המרחקים הריבועיים (בעיית Least Squares) בין התיוגים האמיתיים לפרדיקציות של הדוגמאות במודל^[7].



2. Random Forest : יער של עצי החלטה. בקבלת דוגמה חדשה כלל העצים מנסים להעריך את ערך הרגרסיה עבור הדוגמה, וההחלטה מתקבלת על פי ממוצע משוקלל שלהם [8].



3. KNN : מתן ערך לדוגמה על פי ממוצע של K השכנים הקרובים ביותר על פי תכונות מוגדרות ומרחק מוגדר.



ניסוי – מטא אלגוריתם :

1. לכל data מקבוצת Data Manipulation :

a. לכל feature selector מקבוצת Feature Selection :

i. לכל algorithm מקבוצת האלגוריתמים :

1. לכל k (מספר הפיצ'רים) מ-1 עד כמות הפיצ'רים המקסימלית :

a. הפעל את K feature selector על ה-data וקבל תת קבוצה של פיצ'רים.

b. אמן את ה-algorithm על פי ה-data המצומצם לפי תת הקבוצה שנבחרה.

c. בדוק את הדיוק על פי 5-Fold Cross Validation.

בתוצאות הניסוי נציע לכל data, feature selector ו-algorithm את ה-k האידיאלי בלבד.

תוצאות הניסוי :

לכל הרצה נתאר את רמת הדיוק של האלגוריתם שהתקבל ואת מספר הפיצ'רים שהאלגוריתם בחר, כאשר יש סך הכל 122 פיצ'רים זמינים.

רמת הדיוק באה לידי ביטוי בפונקציית MAE (Mean Absolute Error) - ממוצע הפרשי הערכים בין החיזוי למציאות^[9]. כאמור, ככל שהערך שחוזר קטן יותר, כך הפרש הערכים קטן יותר, קרי החיזוי טוב יותר.

הפלט הגולמי של הניסוי מצורף בקובץ featureselectionRes.txt. בדו"ח זה נציג מידע סיכומי על התוצאות.

• Original Data :

		SVM	Random Forest	KNN
selectKBestFeatures	MAE	0.243	0.226	0.256
	Features	4	81	11
selectKExtraTree	MAE	0.234	0.225	0.23
	Features	8	20	6

• Normalized Data :

		SVM	Random Forest	KNN
selectKBestFeatures	MAE	0.243	0.226	0.252
	Features	39	81	12

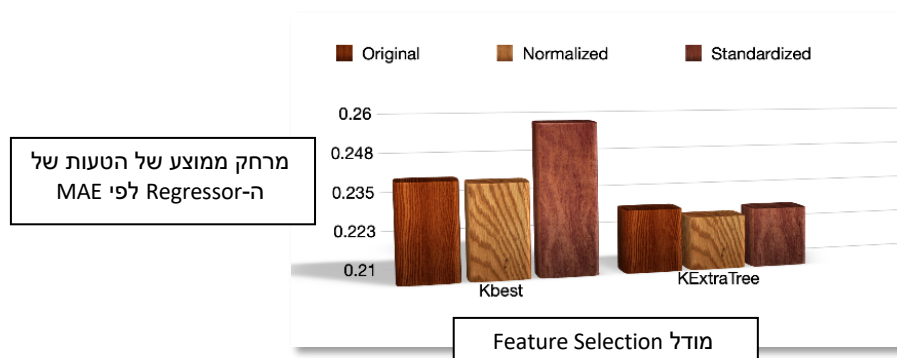
selectKExtraTree	MAE	0.228	0.223	0.231
	Features	13	22	5

• Standardized Data :

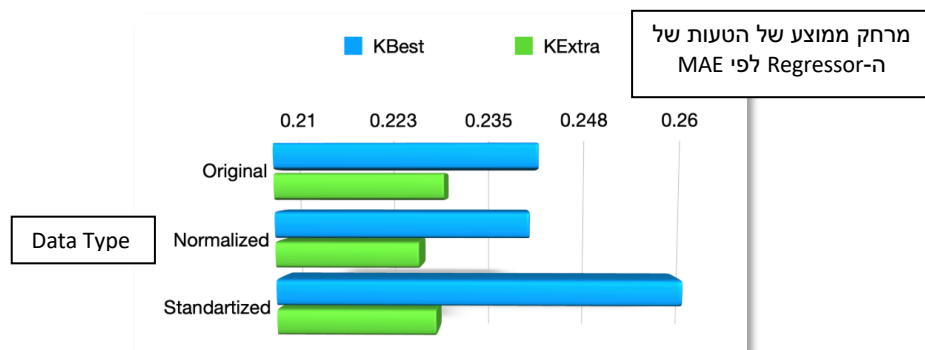
		SVM	Random Forest	KNN
selectKBestFeatures	MAE	0.292	0.225	0.259
	Features	102	82	14
selectKExtraTree	MAE	0.229	0.227	0.231
	Features	3	25	7

מסקנות :

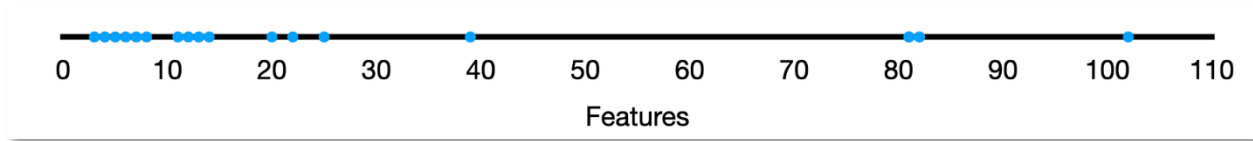
1. **selectKExtraTree** מביא לתוצאות טובות יותר (סטייה קטנה יותר) מ-selectKBestFeatures. בגרף : ממוצע תוצאות של selectKExtraTree לעומת ממוצע תוצאות של selectKBestFeatures בכלל האלגוריתמים שנבדקו.



2. **Normalized Data** מביא לתוצאות הטובות ביותר מבין כלל סוגי ה-Data. בגרף : ממוצע של Normalized data לעומת ממוצע של Standardized data, Original data בכלל האלגוריתמים שנבדקו.



3. האלגוריתמים לא מתאימים לבעיות הקשורות באופן הדוק ל-Time Sequences. ניתן לראות זאת הן ברמת הדיוק הנמוכה – מה שמעיד על יכולת דלה של הכללה, והן ברמת השונות של מספר הפיצ'רים שנלקחים בחשבון – מה שמעיד על אקראיות בסיווג. בגרף : פיזור של מספר הפיצ'רים שנבחרו (קרי, k הטוב ביותר) בניסויים השונים שתוארו.



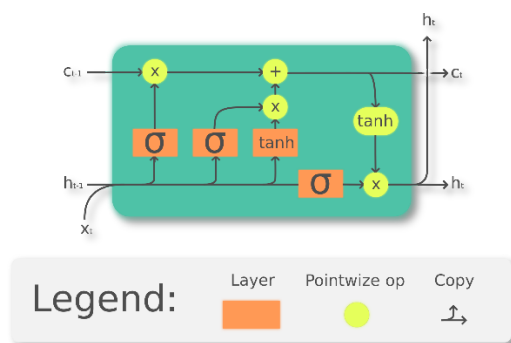
הערה : כזכור, במודל שלנו הציונים לכל דוגמה נעים בין 0 ל-1. לכן, אף על פי שערכי ה-MAE עשויים להשתמע כקטנים למראית עין, בראייתנו שגיאה של 0.2-0.3 מבטלת את הכדאיות להשתמש בחיזוי.

חקירת הבעיה באמצעות Deep Learning Algorithms ובפרט LSTM

רקע:

רשת LSTM^[10] הינה ארכיטקטורת RNN (Recurrent Neural Network). בניגוד ל- FeedForward Neural Network, רשת נזירונים שמתייחסת לכל דוגמה כ-Single data point, ל-LSTM יש ממשק המכיל Feedback connections, שמאפשרים לרשת להתייחס לאוסף הדוגמאות כאוסף סדרתי.

דוגמה טובה לכך היא ילד שרואה סרט ובו סצנה של שני אנשים רצים עם כדור. אם על סמך סצנות קודמות הוא יודע שמדובר בסרט על כדורסל, הוא יסיק שבסצנה ישחקו כדורסל בסבירות גבוהה. ההסקה הזו נובעת מההתייחסות לסצנה (הדוגמה) **כמידע עוקב לסצנות קודמות, ולא כ-Single data point**.
הייחוד העיקרי של LSTM הם האופרטורים בתוך תא ה-LSTM, שמאפשרים לו לזכור חלק מהמידע ולשכוח חלק אחר.



עקב כך, מספר מניפולציות קודמות שביצענו עשויות להשפיע, לטובה או לרעה, על רמת הדיוק של המסווג:

- **שימוש ב-data לא רציף**: כזכור, פיצלנו את ה-data על פי משברים. כתוצאה מכך LSTM מקבל דוגמאות רק מתקופות משבר והמידע איננו רציף.
- **שימוש ב"פוטנציאל הרווח"**: כזכור, הגדרנו "פוטנציאל רווח" כפונקציה של מחירי מינימום ומקסימום בתוך משבר מסוים. כלומר, פונקציית התיוג שלנו קשורה קשר הדוק להיות ה-data מחולק למשברים. ויתור על החלוקה למשברים – משמעותו ויתור על נוסחת "פוטנציאל הרווח". את משמעות ויתור זה על הארכיטקטורה נפרט בהמשך.
- **שימוש ב-lookback**: לרשת LSTM, כפי שהוסבר, יש מגננוני זיכרון מובנים, וניסיון "להעמיס" עליה זיכרון באופן ידני עשוי להביא למצב של כפילות מיותרת שתפגע בביצועים.
- **שימוש בנתונים חודשיים**: הנתונים החודשיים "נמרחים" בצורה זהה על פני ימים באותו חודש ועשויים להשפיע על התוצאות, כתכונות שאינן מפרידות היטב בין הדוגמאות.

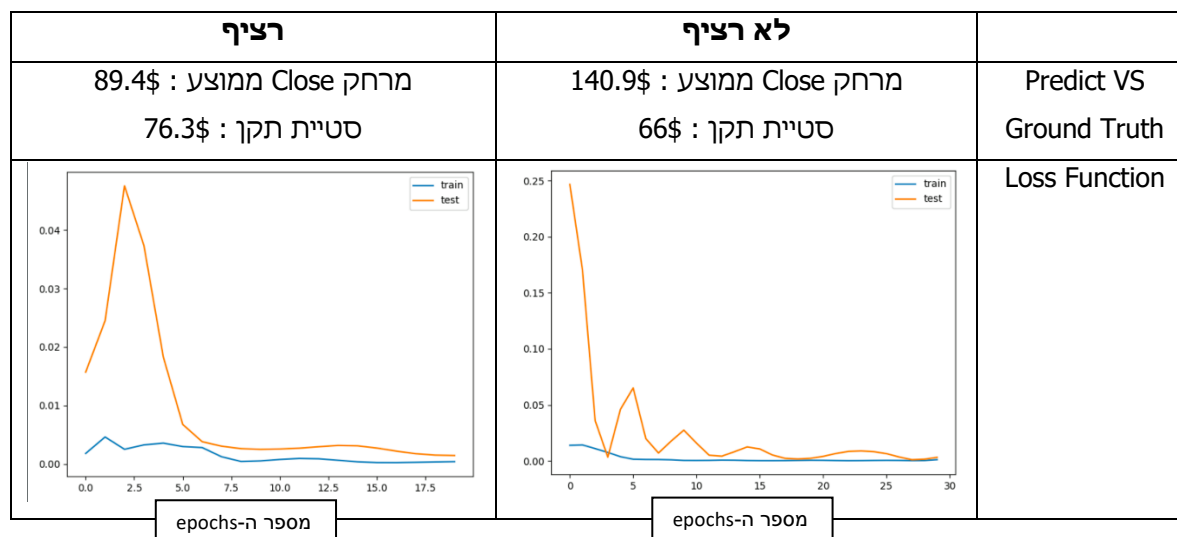
בניסויים שיתוארו בהמשך נבחן את הנקודות הנ"ל, ונקבל החלטות לגביהן.

ניסויים :

נסמן מעתה : Predict – תוצאות המסווג על קבוצת מבחן. Ground Truth – תוצאות אמיתיות על הקבוצה.
על מנת לבחון את ההשפעות של הנקודות שהועלו לעיל, ולבחור אסטרטגיה להמשך, נבצע מספר ניסויים ובהם נבדוק את טיב התוצאות על קבוצת מבחן :

1. **שימוש ב-data רציף מול שימוש ב-data תלוי-משברים.** על מנת להשוות בין השניים, ננסה בכל יום לחזות את מחיר הסגירה של יום המסחר הבא (שכן פונקציית התיוג המקורית איננה מוגדרת היטב עבור data רציף). את מדידת הדיוק נבצע על סמך המרחק הממוצע וסטיית התקן של החזוי של מחיר ה-Close מהתיוג האמיתי. כמו כן, נציג את התנהגות ה-Loss function^[11] (מרחק הפרדיקציה מהאמת כפונקציה של מספר ה-epochs) בכל אחד מן המקרים. על סמך ההשוואה נקבל החלטה באיזה data להשתמש.

תוצאה :

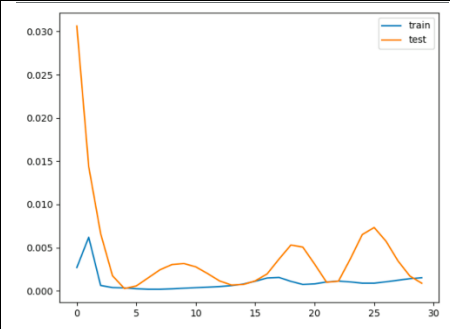
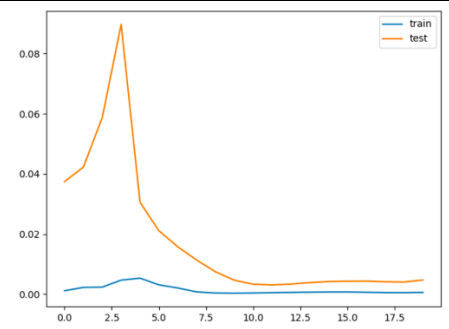


מסקנה :

- הרשת מתמודדת טוב יותר עם Data רציף. הסיבה לכך עשויות להיות רבות :
- הרציפות עצמה מסייעת למסווג בהסקת מסקנות מדויקות יותר.
 - העלייה הדרמטית במספר הדוגמאות כתוצאה מכך שלא עושים filter out לימים שאינם במשבר (בערך פי 3 דוגמאות).

2. **שימוש ב-lookback אל מול אי שימוש.** ההבדל ב-data בא לידי ביטוי בכך שבשימוש ב-lookback, לכל דוגמה יש תכונות של מספר ימים אחורה, וללא ה-lookback אנו מוותרים על כך. את מדידת הדיוק נבצע ע"י השוואת שני החזויים לתיוג האמיתי לאורך זמן.

תוצאה :

Without Lookback	With Lookback	
מרחק Close ממוצע : 71.5\$ סטיית תקן : 57.7\$	מרחק Close ממוצע : 161\$ סטיית תקן : 138.3\$	Predict VS Ground Truth
		Loss Function

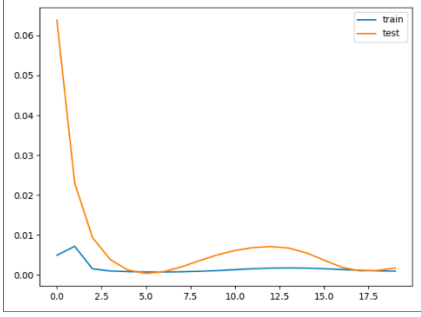
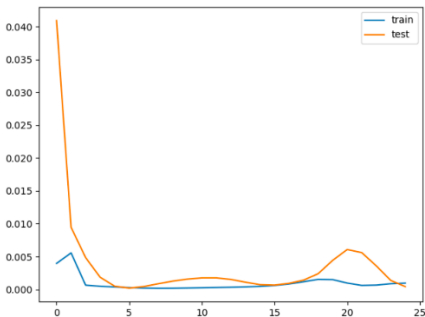


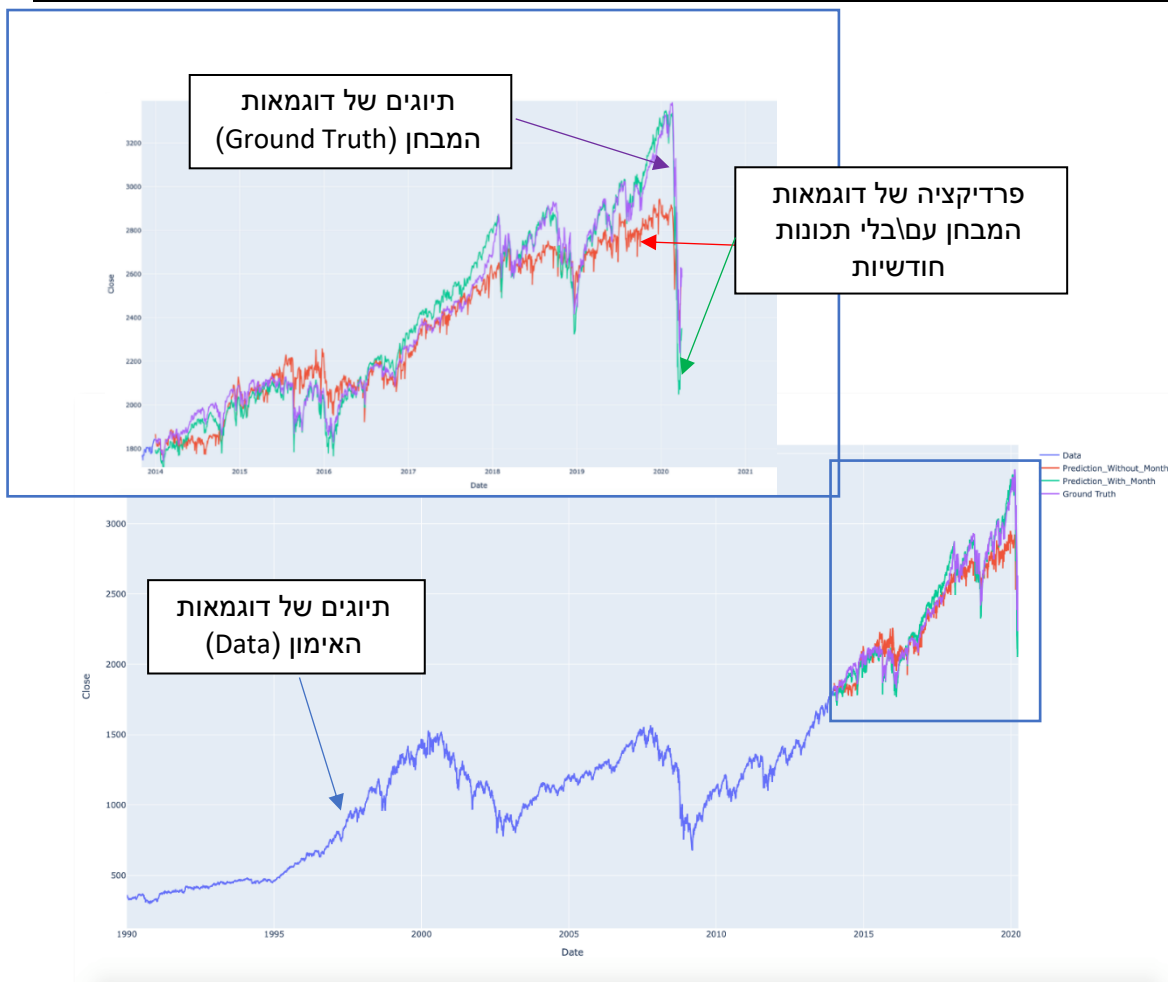
מסקנה :

החיזויים ללא Lookback טובים יותר, ואת ההשוואות הבאות נעשה על בסיס נתונים ללא Lookback. הערה – שימו לב ל-Scaling של פונקציות ה-Loss. אף על פי שלמראית עין With Lookback Loss Function נראית טובה יותר, בפועל הביצועים של Without Lookback Loss Function טובים משמעותית.

3. שימוש בתכונות יומיות בלבד אל מול שימוש בתכונות יומיות וחודשיות יחד. את מדידת הדיוק נבצע ע"י השוואת שני החזויים לתיוג האמיתי לאורך זמן.

תוצאה :

Without Month Attributes	With Month Attributes	
מרחק Close ממוצע : 92.9\$ סטיית תקן : 88.6\$	מרחק Close ממוצע : 48.9\$ סטיית תקן : 41.2\$	Predict VS Ground Truth
		Loss Function



מסקנה :

על אף פי שהתכונות החודשיות נמרחות על פני הימים השונים, הן מסייעות משמעותית בסיווג, וכדאי להשאיר אותן.

המסקנה הכוללת מהניסויים היא שאסטרטגיית הלמידה היא הבאה :

- נשתמש ב-Data רציף ולא תלוי משברים.
- **נחזה את מחיר ה-Close ולא את פוטנציאל הרווח.** היות שראינו בניסויים שחכם יותר להשתמש ב-Data רציף, נעבור לחזות את מחיר ה-Close, שכן פוטנציאל הרווח לא מוגדר היטב במקרה כזה. בהמשך נסביר איך נגדיר מחדש את מושג פוטנציאל הרווח במקרה כזה.
- נחזה על סמך התכונות ללא Lookback.
- ניקח בחשבון את התכונות החודשיות.

: Hyper Parameters Tuning

כעת, כשבחרנו אסטרטגיה, יש מספר היפר-פרמטרים באלגוריתמים שניתן לכוון על סמך ביצועים. על מנת להגיע לרמת הסיווג המיטבית, בדקנו את כל הפרמוטציות שבאות לידי ביטוי בטווח הערכים הבא :

Parameter	Possible Values	Description
Epochs	18,22,26,30	מספר מחזורי האימון של הרשת.
Batch Size	32,48,64,72,96	מספר דוגמאות האימון בכל סיבוב.
Units	30,40,50,60,70,80	מספר היחידות בתוך תא LSTM.

מדובר בבעיית אופטימיזציה ארבע-מימדית – רמת הדיוק של המסווג כפונקציה של שלושה פרמטרים. לכן, במקום לנתח את התוצאות על פני טבלה, נשתמש ב"מפת חום". ככל שהתוצאה מרשימה יותר – היא תהיה ירוקה יותר במפה. תוצאות גרועות במיוחד יסומנו באדום. המוטיבציה לשימוש בשיטה זו נעוצה בכך שהרשת מחזירה תוצאות באופן לא דטרמינסטי – ולכן נרצה לראות "אזורים" מוצלחים של תחומי ערכים לפרמטרים, ולא דווקא ריצה מושלמת אחת.

פונקציית הדיוק תלויה בשני פרמטרים :

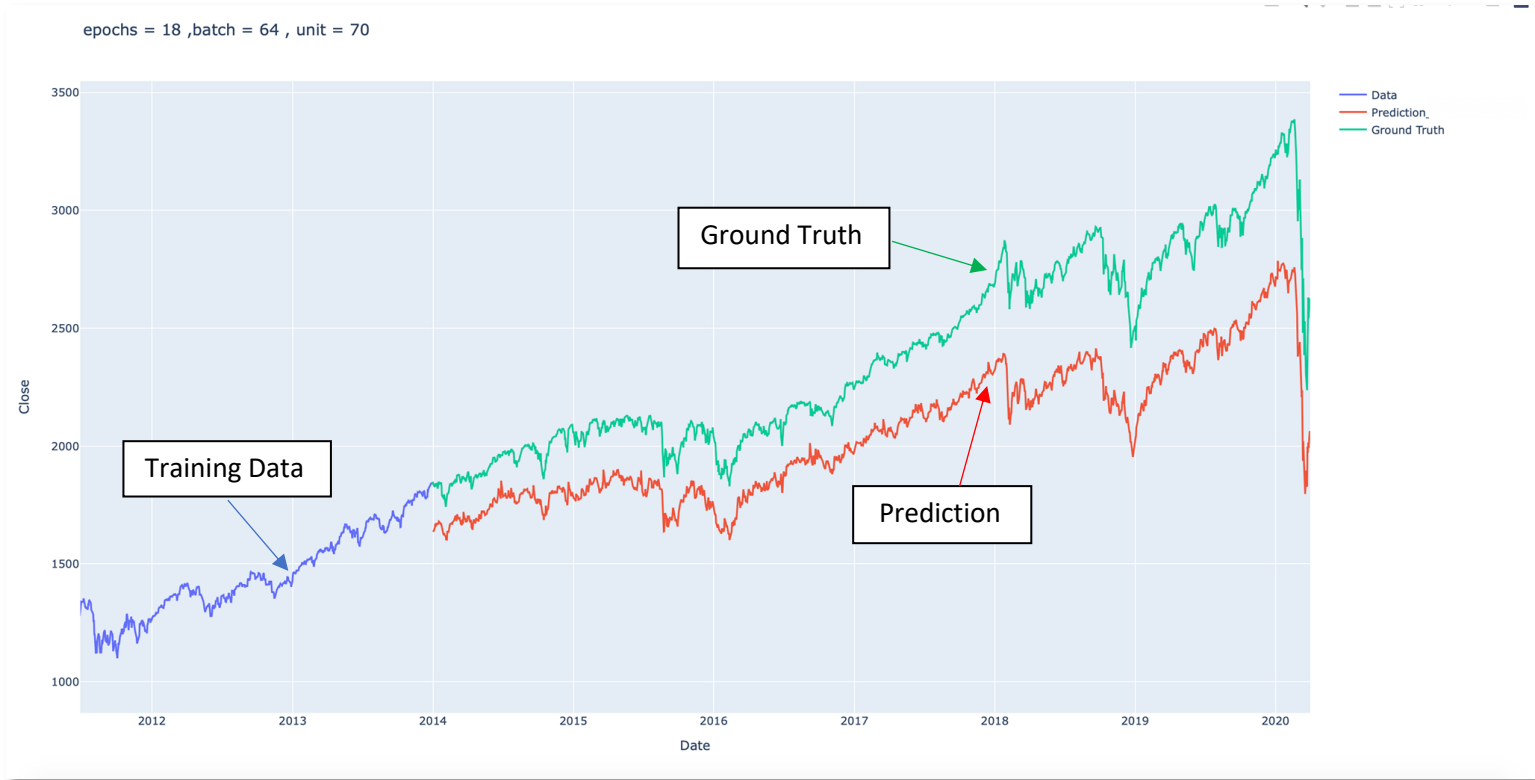
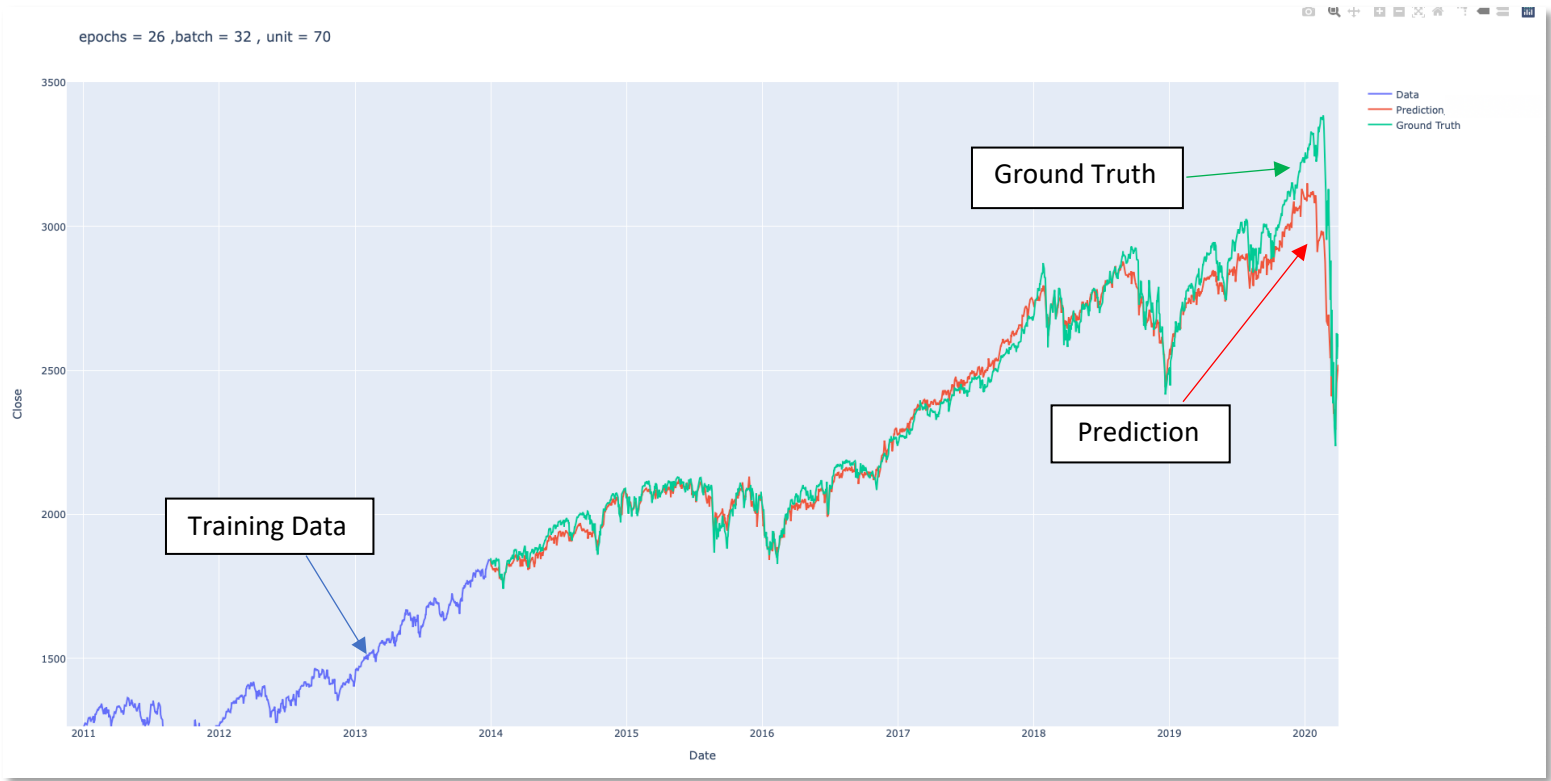
- Mean : המרחק הממוצע בין ה-label המקורי לבין prediction label.
- Std : סטיית התקן של המרחק ביניהם.

והרי התוצאות :

(Epochs, batch size, units)	Mean	std	(Epochs, batch size, units)	Mean	std	(Epochs, batch size, units)	Mean	std	(Epochs, batch size, units)	Mean	Std
(18,32,30)	94.3	78.9	(22,32,30)	230.8	194.5	(26,32,30)	70.1	81.4	(30,32,30)	213.2	195.4
(18,32,40)	73.1	57.4	(22,32,40)	64.7	75.6	(26,32,40)	50.8	61.4	(30,32,40)	170.5	158.6
(18,32,50)	58.7	65.3	(22,32,50)	64.2	64.4	(26,32,50)	70.8	87.9	(30,32,50)	235.7	187.3
(18,32,60)	176.2	114.8	(22,32,60)	115.4	73.1	(26,32,60)	60	57.7	(30,32,60)	91.6	105.2
(18,32,70)	57.1	38.5	(22,32,70)	106.8	71	(26,32,70)	56	66.9	(30,32,70)	116.1	83.2
(18,32,80)	114.5	124.7	(22,32,80)	175.5	114.5	(26,32,80)	121	115.4	(30,32,80)	242.3	203.1
(18,48,30)	288.6	158.7	(22,48,30)	302.7	154.8	(26,48,30)	383.8	180.8	(30,48,30)	73.1	69
(18,48,40)	287	202.6	(22,48,40)	147.3	135.3	(26,48,40)	167.3	112.1	(30,48,40)	112.2	91.5
(18,48,50)	88.3	51.5	(22,48,50)	278.8	186.4	(26,48,50)	214.9	132.6	(30,48,50)	296.1	142
(18,48,60)	285.9	140.2	(22,48,60)	240.7	150.6	(26,48,60)	152.1	123.9	(30,48,60)	140.1	99.5
(18,48,70)	164	81.5	(22,48,70)	107.6	73	(26,48,70)	86	62.5	(30,48,70)	343	174.3
(18,48,80)	75.3	41.4	(22,48,80)	54.7	57.2	(26,48,80)	245	156.2	(30,48,80)	171.3	139.5
(18,64,30)	435.3	178.4	(22,64,30)	65.8	58.7	(26,64,30)	94.9	68.1	(30,64,30)	108.8	67
(18,64,40)	181.3	84.1	(22,64,40)	248.2	103.2	(26,64,40)	111.5	75	(30,64,40)	111.2	56.1
(18,64,50)	396.4	152.8	(22,64,50)	120.4	67	(26,64,50)	121.5	75.5	(30,64,50)	125.7	64.8
(18,64,60)	190.9	145	(22,64,60)	50.7	45.9	(26,64,60)	184.3	106.5	(30,64,60)	84.5	53.2
(18,64,70)	343.8	130.9	(22,64,70)	107.6	71.4	(26,64,70)	99.5	62	(30,64,70)	112.1	54.6
(18,64,80)	320.6	140.7	(22,64,80)	52.6	42	(26,64,80)	181.1	98.3	(30,64,80)	185.3	87.9
(18,72,30)	149.2	83	(22,72,30)	54.5	42.1	(26,72,30)	182.9	96.5	(30,72,30)	91.2	97.6
(18,72,40)	79.1	75.4	(22,72,40)	64.3	42.9	(26,72,40)	88.1	53	(30,72,40)	177.8	106
(18,72,50)	160.6	52.7	(22,72,50)	182.7	110.6	(26,72,50)	233.6	119.2	(30,72,50)	95.8	108
(18,72,60)	176.1	74.4	(22,72,60)	74.1	56.1	(26,72,60)	135.2	73.6	(30,72,60)	258.4	130.6
(18,72,70)	229.7	83.5	(22,72,70)	122.1	60.8	(26,72,70)	132.2	69.5	(30,72,70)	59	49.1
(18,72,80)	251.1	112.2	(22,72,80)	52.5	44.5	(26,72,80)	178.8	96.2	(30,72,80)	130.6	75.4
(18,96,30)	87.5	58.1	(22,96,30)	148.2	155.2	(26,96,30)	343.5	222.5	(30,96,30)	327.6	190
(18,96,40)	84.8	52.2	(22,96,40)	75.2	46.8	(26,96,40)	108.2	118.5	(30,96,40)	356	223.5
(18,96,50)	46.1	41.4	(22,96,50)	126.2	74.5	(26,96,50)	147.2	69.7	(30,96,50)	429.2	250.9
(18,96,60)	143.6	85.4	(22,96,60)	183.9	116.5	(26,96,60)	185.3	68.5	(30,96,60)	163.6	173.5
(18,96,70)	57.5	49.4	(22,96,70)	142.6	90.5	(26,96,70)	118	83.6	(30,96,70)	131.1	65.8
(18,96,80)	183.7	104.3	(22,96,80)	178.5	114.1	(26,96,80)	258.4	218.3	(30,96,80)	122.5	92.8

מסקנות

- קבוצת הפרמטרים (Epochs=18, Batch Size=96, Units=50) מביאה לתוצאות האידיאליות ביותר בריצה הנ"ל.
- עבור תוצאות סיווג אידיאליות, קיים יחס הפוך בין Epochs ל-Batch Size. כך למשל, התוצאות הטובות ביותר התקבלו עבור (Epochs=22, Batch Size=72) ועבור (Epochs=26, Batch Size=32).
- נבחר את המסווג עם הפרמטרים (Epochs=26, Batch Size=32, Units=70).
- על מנת להמחיש את ההבדל בביצועים, נשווה בין זוג מסווגים – המסווג הנבחר, ומסווג שנתן תוצאות גרועות :



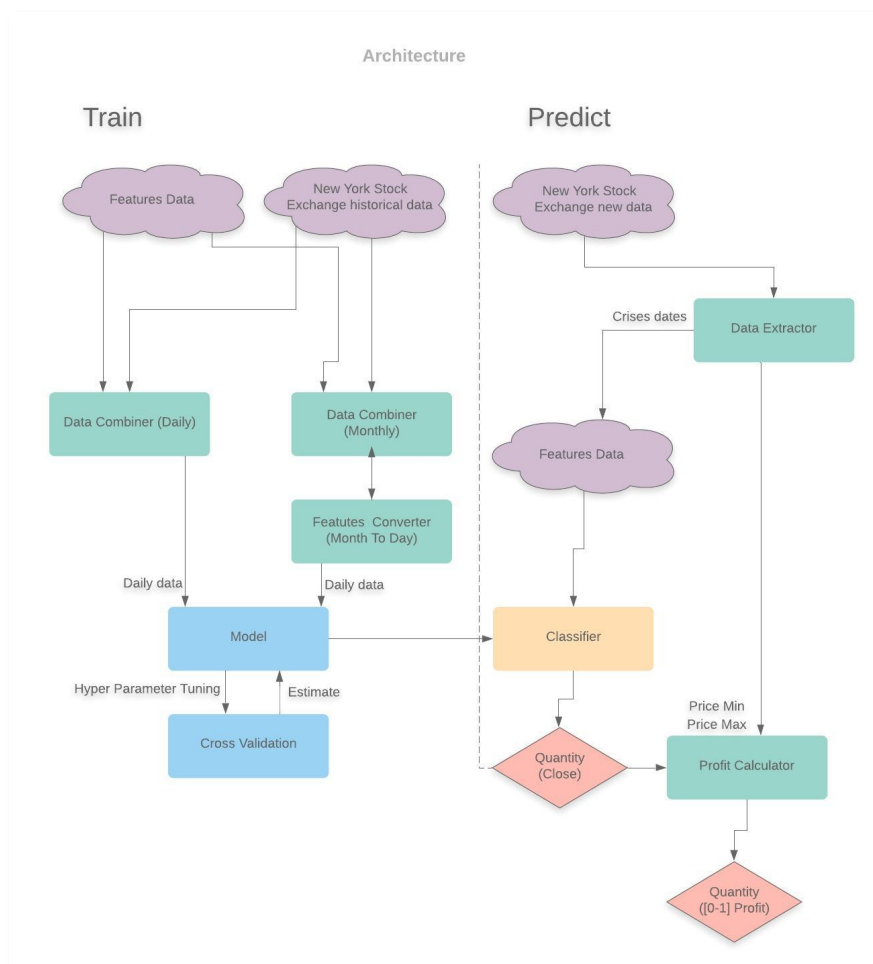
שינויים בארכיטקטורה כתוצאה משימוש ב-Data שאינו מחולק למשברים

כזכור, בפרק "תיאור הפתרון", בהסבר על פונקציית פוטנציאל הרווח, התבססנו על הפרמטרים $profit_{pot}$, $price_{max}$, $price_{min}$. הפרמטרים האלה מוגדרים היטב כל עוד ה-Data מחולק למשברים, אך בסעיף זה הראינו שמשתלם לאמן את המסווג על Data רציף. במקרה כזה, הפרמטרים הנ"ל לא מוגדרים היטב.

לכן, נשנה את הארכיטקטורה באופן הבא :

- **האימון (train)** אשר מתבצע על Data רציף של כל ימי המסחר, לא יקח בחשבון את הפרמטרים הללו, וכן כפי שהראינו, יחזה את מחיר הסגירה Close, במקום את פונקציית פוטנציאל הרווח.
- **הסיווג (predict)** יחזה את מחיר הסגירה Close לכל יום מסחר. ואז באופן ידני, לכל יום שנמצא בזמן תקופת משבר, נחשב על סמך החיזוי את פונקציית פוטנציאל הרווח המתאימה, כאשר :
 - $price_{max}$ – המחיר ממנו הבורסה "צנחה" טרם תקופת המשבר.
 - $price_{min}$ – מחיר המינימום אליו צנח המשבר **עד כה**.

וכך, נוכל עדיין לענות על השאלה המקורית ולהתמחות באסטרטגיית מסחר לימי משבר, ומאידך לא לוותר על Data שחינוי לאימון ולדיוק של המסווג. במבט על :



יישום אסטרטגיית מסחר על סמך תוצאות המסווג

רקע

בחלק זה נחזור להשתמש בפונקציית פוטנציאל הרווח שהגדרנו. המשמעות היא שהמסווג ייתן ציונים בין 0 ל-1 לימים במהלך משבר, וככל שהציון גבוה יותר, כך לכאורה פוטנציאל הרווח עולה.

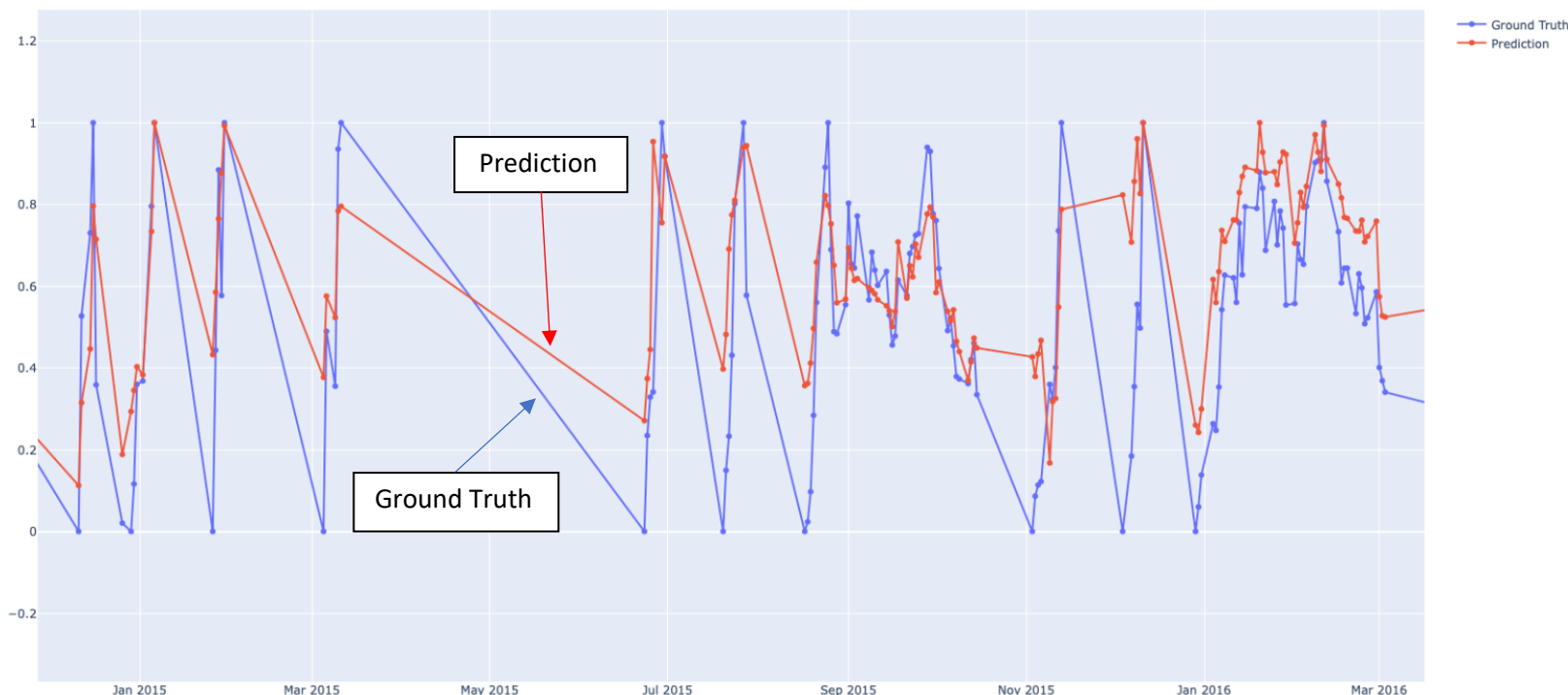
שאלה שעשויה להישאל היא איך מחלצים את פוטנציאל הרווח ממחיר הסגירה, שכן כפי שראינו בחלק הקודם, הרשת חזתה את מחיר הסגירה (Close) ולא את הפוטנציאל. התשובה לכך, אם כן, נעוצה בשינוי הארכיטקטורה שעשינו. חיזוי פוטנציאל הרווח יחושב מהחיזוי של Close, כשם שפוטנציאל הרווח האמיתי חושב מה-Close. כלומר :

$$y = \frac{\text{price_max} - \text{close_prediction}}{\text{profit_pot}}$$

הערה : להבדיל מפוטנציאל הרווח האמיתי, חיזוי פוטנציאל הרווח לאו דווקא יהיה בין 0 ל-1 (היות שהוא מבוסס על מחירי סגירה שחזינו ואינם מדויקים). עקב כך, תוצאות המסווג עבור פוטנציאל הרווח ינורמלו.

את התוצאות הנ"ל נרצה לתרגם לאסטרטגיית מסחר. בחלק זה נציע מספר אסטרטגיות שמהוות נוסחה סגורה, ומסתמכות על ניבויי המסווג בלבד. נבחן את ביצועי האסטרטגיות האלה בהינתן סכום כסף מסוים ותחת משברים מקבוצת ה-test שלנו. כלומר, נבדוק, בהינתן סכום התחלתי של 100K ואסטרטגיית מסחר, כמה כסף הצלחנו להרוויח במהלך משבר מסוים.

בגרף : snapshot של ביצועי המסווג עבור תקופות משבר מסוימות מתוך קבוצת ה-test.



הערה : בעמודים הבאים נציג מספר אסטרטגיות מסחר התלויות בפרמטרים k, k_1, k_2, k_3 . אלו הם היפר פרמטרים של האסטרטגיות, ובהמשך נבצע ניסויים והערכת ביצועים על מספר ערכים שונים שלהם, כדי להגיע לקבוצת ערכים עבורם האסטרטגיות אופטימליות.

אסטרטגיה #1 – All In Strategy

משקיעים יחכו ליום שהמסווג יציין כפוטנציאלי במיוחד, ויקנו מניות בכל כספם ביום זה. לאחר מכן, ימתינו ליום שהמסווג מחשיב כשיא (ציון נמוך במיוחד) וימכרו הכל.

יתרונות :

- התאמה לאנשים שלא רוצים להתעסק במסחר יום-יומי.
- השקעת הכסף בנקודה יחסית בטוחה ומימוש בנקודה משתלמת.

חסרונות :

- טעות במסווג תעלה ביוקר.
- משקיעים רבים מאמינים שיש להיכנס לשוק בפעילות ולא בבת אחת. ככל הנראה משקיעים מהסוג הזה פחות יתחברו לאסטרטגיית המסחר הזו.

אלגוריתם :

1. בכל יום, חשב את פוטנציאל הרווח $profit_day$.
2. קנה בכל כספך את נייר הערך ביום הראשון בו $profit_day$ גדול מ- k_1 .
3. ממש בכל כספך את נייר הערך ביום הראשון בו $profit_day$ קטן מ- k_2 .

אסטרטגיה #2 – Gradient Strategy

בכל יום נשווה את הציון לציון של היום הקודם. עלייה בציון תיחשב לעלייה בפוטנציאל (כמתבקש כפי שהגדרנו את פוטנציאל הרווח) ולכן נשקיע. ירידה בציון תיחשב לירידה בפוטנציאל ולכן נממש. סכום הקנייה/המימוש יהיה ביחס ישר להפרש בין הימים.

יתרונות :

- פיזור וכניסה לשוק בפעילות.
- מעט הסתמכות על ביצועי המסווג ביחס למציאות, ויותר הסתמכות על ביצועי המסווג ביחס לעצמו (ביום האתמול)

חסרונות :

- רצף ימי שפל עשויים להתפסס בעקבות פער נמוך בין כל יום לקודמו.



- עיסוק רציף ויומיומי (לא מתאים לכולם).

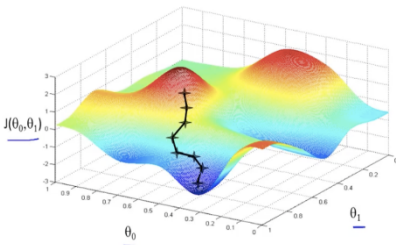
אלגוריתם :

1. בכל יום, חשב את פוטנציאל הרווח profit_day , ואת פוטנציאל הרווח של אתמול profit_yesterday .
2. נסמן : profit_gradient – הפרש הפוטנציאלים. invested_money – סכום הכסף המושקע ברגע נתון. uninvested_money – סכום הכסף שנותר לא מושקע ברגע נתון.
3. בכל יום בו profit_day גדול מ- profit_yesterday :

$$a. \text{abs}(\text{profit_gradient})^k * \text{uninvested_money} = \text{סכום הכסף להשקעה}$$

4. בכל יום בו profit_day קטן מ- profit_yesterday :

$$a. \text{abs}(\text{profit_gradient})^k * \text{invested_money} = \text{סכום הכסף למימוש}$$



אסטרטגיה #3 – Calculated Risk Strategy

נגדיר סף מינימום להשקעה וסף מקסימום למימוש. בכל יום בו פוטנציאל הרווח גדול מסף המינימום, נשקיע ביחס ישר לפוטנציאל היומי. בכל יום בו פוטנציאל הרווח קטן מסף המקסימום, נממש ביחס הפוך לפוטנציאל היומי (שכן ככל שהציון קטן יותר, כך השפל עמוק יותר מהגדרת פוטנציאל הרווח)

יתרונות :

- פיזור וכניסה לשוק בפעילות.
- מתמודד היטב עם רצפים (הן של ימי שיא והן של ימי שפל)

חסרונות :

- חוסר דיוק של המסווג עשוי להביא לכישלון חרוץ.
- עיסוק רציף ויומיומי (לא מתאים לכולם).

אלגוריתם :

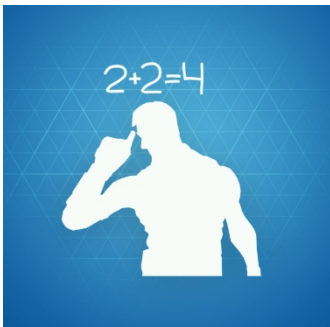
1. בכל יום, חשב את פוטנציאל הרווח profit_day .
2. נסמן : invested_money – סכום הכסף המושקע ברגע נתון. uninvested_money – סכום הכסף שנותר לא מושקע ברגע נתון.

3. בכל יום בו profit_day גדול מ- k_1 :

$$a. k_3 * \text{profit_day} * \text{uninvested_money} = \text{סכום הכסף להשקעה}$$

4. בכל יום בו profit_day קטן מ- k_2 :

$$a. k_3 * (1 - \text{profit_day}) * \text{invested_money} = \text{סכום הכסף למימוש}$$



תוצאות

נבחן את האסטרטגיות שהצגנו תחת ערכי k , k_1 , k_2 , k_3 שונים ועבור תקופת משבר שנבחרה אקראית.

אסטרטגיה #1 :

Date	Close	Label	#1 ($k_1=0.6$, $k_2=0.7$)	#2 ($k_1=0.6$, $k_2=0.6$)	#3 ($k_1=0.6$, $k_2=0.5$)
7/27/2015	2067.639893				
7/28/2015	2093.25				
8/17/2015	2102.439941	0.379858469			
8/18/2015	2096.919922	0.391843538			
8/19/2015	2079.610107	0.502045245			
8/20/2015	2035.72998	0.690282973	Buy 100k (+49 Stocks)	Buy 100k (+49 Stocks)	Buy 100k (+49 Stocks)
8/21/2015	1970.890015	1			
8/24/2015	1893.209961	1			
8/25/2015	1867.609985	1			
8/26/2015	1940.51001	1			
8/27/2015	1987.660034	1			
8/28/2015	1988.869995	0.829279827			
8/31/2015	1972.180054	0.848891757			
9/1/2015	1913.849976	1			
9/2/2015	1948.859985	1			
9/3/2015	1951.130005	0.950947818			
9/4/2015	1921.219971	0.960579123			
9/8/2015	1969.410034	0.911222847			
9/9/2015	1942.040039	0.895871396			
9/10/2015	1952.290039	0.877438427			
9/11/2015	1961.050049	0.845733305			
9/14/2015	1953.030029	0.813342015			
9/15/2015	1978.089966	0.786745725			
9/16/2015	1995.310059	0.700119088			
9/17/2015	1990.199951	0.782091219			
9/18/2015	1958.030029	1			
9/21/2015	1966.969971	0.854536009			
9/22/2015	1942.73999	1			
9/23/2015	1938.76001	0.969649017			
9/24/2015	1932.23999	1			
9/25/2015	1931.339966	1			
9/28/2015	1881.77002	1			
9/29/2015	1884.089966	1			
9/30/2015	1920.030029	1			
10/1/2015	1923.819946	0.88423357			
10/2/2015	1951.359985	0.94203803			
10/5/2015	1987.050049	0.78270773			
10/6/2015	1979.920044	0.730801197			
10/7/2015	1995.829956	0.79087313			
10/8/2015	2013.430054	0.61976777	Sell 49 Stocks (+98.6k)		
10/9/2015	2014.890015	0.564447031		Sell 49 Stocks (+98.7k)	
10/12/2015	2017.459961	0.408569404			Sell 49 Stocks (+98.9k)

בסיום המשבר, מימשנו את כל כספנו וישנו הפסד זעיר של כ-1%-2% כתלות ביום המדויק בו מימשנו (הפרמטר k2). ניתן לראות שהגדלה של k1 תסייע לשיפור. ננסה זאת :

Date	Close	Label	#4 (k1=1, k2=0.5)	#5 (k1=0.8, k2=0.6)
7/27/2015	2067.639893			
7/28/2015	2093.25			
8/17/2015	2102.439941	0.379858469		
8/18/2015	2096.919922	0.391843538		
8/19/2015	2079.610107	0.502045245		
8/20/2015	2035.72998	0.690282973		
8/21/2015	1970.890015	1	Buy 100k (+50.7 Stocks)	Buy 100k (+50.7 Stocks)
8/24/2015	1893.209961	1		
8/25/2015	1867.609985	1		
8/26/2015	1940.51001	1		
8/27/2015	1987.660034	1		
8/28/2015	1988.869995	0.829279827		
8/31/2015	1972.180054	0.848891757		
9/1/2015	1913.849976	1		
9/2/2015	1948.859985	1		
9/3/2015	1951.130005	0.950947818		
9/4/2015	1921.219971	0.960579123		
9/8/2015	1969.410034	0.911222847		
9/9/2015	1942.040039	0.895871396		
9/10/2015	1952.290039	0.877438427		
9/11/2015	1961.050049	0.845733305		
9/14/2015	1953.030029	0.813342015		
9/15/2015	1978.089966	0.786745725		
9/16/2015	1995.310059	0.700119088		
9/17/2015	1990.199951	0.782091219		
9/18/2015	1958.030029	1		
9/21/2015	1966.969971	0.854536009		
9/22/2015	1942.73999	1		
9/23/2015	1938.76001	0.969649017		
9/24/2015	1932.23999	1		
9/25/2015	1931.339966	1		
9/28/2015	1881.77002	1		
9/29/2015	1884.089966	1		
9/30/2015	1920.030029	1		
10/1/2015	1923.819946	0.88423357		
10/2/2015	1951.359985	0.94203803		
10/5/2015	1987.050049	0.78270773		
10/6/2015	1979.920044	0.730801197		
10/7/2015	1995.829956	0.79087313		
10/8/2015	2013.430054	0.61976777		
10/9/2015	2014.890015	0.564447031		Sell 50.7 Stocks (+102.1k)
10/12/2015	2017.459961	0.408569404	Sell 50.7 Stocks (+102.2k)	

בסיום המשבר, מימשנו את כל כספנו וישנו רווח זעיר של כ-2%, כתלות ביום המדויק בו מימשנו (הפרמטר k2).

אסטרטגיה #2 :

Date	Close	Label	Gradient	Actions (k=1)	uninvested	invested	Stocks
8/17/2015	2102.439941	0.379858469	0.011985069		100k	0k	
8/18/2015	2096.919922	0.391843538	0.110201707	Buy 11k (+5.2 Stocks)	89k	11k	5.2
8/19/2015	2079.610107	0.502045245	0.188237728	Buy 16k (+7.7 Stocks)	73k	27k	12.9
8/20/2015	2035.72998	0.690282973	0.309717027	Buy 22k (+10.8 Stocks)	51k	49k	23.7
8/21/2015	1970.890015	1	0				
8/24/2015	1893.209961	1	0				
8/25/2015	1867.609985	1	0				
8/26/2015	1940.51001	1	0				
8/27/2015	1987.660034	1	-0.170720173	Sell 4.2 Stocks (+8.3k)	59.3k	40.7k	19.5
8/28/2015	1988.869995	0.829279827	0.01961193	Buy 1k (+0.5 Stocks)	58.3k	41.7k	20
8/31/2015	1972.180054	0.848891757	0.151108243	Buy 9k (+4.4 Stocks)	49.3k	50.7k	24.4
9/1/2015	1913.849976	1	0				
9/2/2015	1948.859985	1	-0.049052182	Sell 1.3 Stocks (+2.5k)	51.8k	48.2k	21.9
9/3/2015	1951.130005	0.950947818	0.009631304	Buy 0k (+0 Stocks)	51.8k	48.2k	21.9
9/4/2015	1921.219971	0.960579123	-0.049356276	Sell 1.3 Stocks (+2.6k)	54.4k	45.6k	20.6
9/8/2015	1969.410034	0.911222847	-0.015351451	Sell 0 Stocks (+0k)	54.4k	45.6k	20.6
9/9/2015	1942.040039	0.895871396	-0.018432969	Sell 0 Stocks (+0k)	54.4k	45.6k	20.6
9/10/2015	1952.290039	0.877438427	-0.031705122	Sell 0.7 Stocks (+1.3k)	55.7k	44.3k	19.9
9/11/2015	1961.050049	0.845733305	-0.03239129	Sell 0.7 Stocks (+1.3k)	57k	43k	19.2
9/14/2015	1953.030029	0.813342015	-0.026596289	Sell 0.6 Stocks (+1.1k)	58.1k	41.9k	18.6
9/15/2015	1978.089966	0.786745725	-0.086626637	Sell 1.6 Stocks (+3.3k)	61.4k	38.6k	17
9/16/2015	1995.310059	0.700119088	0.08197213	Buy 4.9k (+2.4 Stocks)	56.5k	43.5k	19.4
9/17/2015	1990.199951	0.782091219	0.217908781	Buy 11.3k (+5.7 Stocks)	45.2k	54.8k	25.1
9/18/2015	1958.030029	1	-0.145463991	Sell 3.9 Stocks (+7.6k)	52.8k	47.2k	21.2
9/21/2015	1966.969971	0.854536009	0.145463991	Buy 6.6k (+3.4 Stocks)	46.2k	53.8k	24.6
9/22/2015	1942.73999	1	-0.030350983	Sell 0.8 Stocks (+1.6k)	47.8k	52.2k	23.8
9/23/2015	1938.76001	0.969649017	0.030350983	Buy 1.4k (+0.7 Stocks)	46.4k	53.6k	24.5
9/24/2015	1932.23999	1	0				
9/25/2015	1931.339966	1	0				
9/28/2015	1881.77002	1	0				
9/29/2015	1884.089966	1	0				
9/30/2015	1920.030029	1	-0.11576643	Sell 3 Stocks (+5.9k)	52.3k	47.7k	21.5
10/1/2015	1923.819946	0.88423357	0.05780446	Buy 2.6k (+1.3 Stocks)	49.7k	50.3k	22.8
10/2/2015	1951.359985	0.94203803	-0.1593303	Sell 4.1 Stocks (+8k)	57.7k	42.3k	18.7
10/5/2015	1987.050049	0.78270773	-0.051906534	Sell 1 Stocks (+2.1k)	59.8k	40.2k	17.7
10/6/2015	1979.920044	0.730801197	0.060071933	Buy 3.6k (+1.8 Stocks)	56.2k	43.8k	19.5
10/7/2015	1995.829956	0.79087313	-0.17110536	Sell 3.7 Stocks (+7.4k)	63.6k	36.4k	15.8
10/8/2015	2013.430054	0.61976777	-0.055320739	Sell 0.9 Stocks (+1.8k)	65.4k	34.6k	14.9
10/9/2015	2014.890015	0.564447031	-0.155877627	Sell 2.6 Stocks (+5.2k)	70.6k	29.4k	12.3
10/12/2015	2017.459961	0.408569404	0.100210685	Buy 7k (+3.4 Stocks)	63.6k	36.4k	15.7

בסיום המשבר, יש ברשותנו 63.6k כסף שאיננו מושקע, ו15.7 מניות ששווין 31.7k (לפי שווי של 2017 דולר למניה). כלומר, סה"כ ערך כספי של פחות מ-100k, ולכן אסטרטגיה זו נכשלה עבור א זה.

ננסה א נוסף :

Date	Close	Label	Gradient	Actions (k=0.5)	uninvested	invested	Stocks
8/17/2015	2102.439941	0.379858469	0.011985069		100k	0k	
8/18/2015	2096.919922	0.391843538	0.110201707	Buy 1.2k (+0.6 Stocks)	98.8k	1.2k	0.6
8/19/2015	2079.610107	0.502045245	0.188237728	Buy 3.2k (+1.5 Stocks)	95.6k	4.4k	2.1
8/20/2015	2035.72998	0.690282973	0.309717027	Buy 8.6k (+4.2 Stocks)	87k	13k	6.3
8/21/2015	1970.890015	1	0				6.3
8/24/2015	1893.209961	1	0				6.3
8/25/2015	1867.609985	1	0				6.3
8/26/2015	1940.51001	1	0				6.3
8/27/2015	1987.660034	1	-0.170720173	Sell 0.2 Stocks (+0.4k)	87.4k	12.6k	6.1
8/28/2015	1988.869995	0.829279827	0.01961193				6.1
8/31/2015	1972.180054	0.848891757	0.151108243	Buy 2k (+1 Stocks)	85.4k	14.6k	7.1
9/1/2015	1913.849976	1	0				7.1
9/2/2015	1948.859985	1	-0.049052182				7.1
9/3/2015	1951.130005	0.950947818	0.009631304				7.1
9/4/2015	1921.219971	0.960579123	-0.049356276				7.1
9/8/2015	1969.410034	0.911222847	-0.015351451				7.1
9/9/2015	1942.040039	0.895871396	-0.018432969				7.1
9/10/2015	1952.290039	0.877438427	-0.031705122				7.1
9/11/2015	1961.050049	0.845733305	-0.03239129				7.1
9/14/2015	1953.030029	0.813342015	-0.026596289				7.1
9/15/2015	1978.089966	0.786745725	-0.086626637				7.1
9/16/2015	1995.310059	0.700119088	0.08197213	Buy 0.5k (+0.25 Stocks)	84.9k	15.1k	7.35
9/17/2015	1990.199951	0.782091219	0.217908781	Buy 3.7k (+1.9 Stocks)	81.2k	18.8k	9.25
9/18/2015	1958.030029	1	-0.145463991	Sell 0.2 Stocks (+0.4k)	81.6k	18.4k	9.05
9/21/2015	1966.969971	0.854536009	0.145463991	Buy 1.8k (+0.9 Stocks)	79.8k	20.2k	9.95
9/22/2015	1942.73999	1	-0.030350983				9.95
9/23/2015	1938.76001	0.969649017	0.030350983				9.95
9/24/2015	1932.23999	1	0				9.95
9/25/2015	1931.339966	1	0				9.95
9/28/2015	1881.77002	1	0				9.95
9/29/2015	1884.089966	1	0				9.95
9/30/2015	1920.030029	1	-0.11576643	Sell 0.1 Stocks (+0.2k)	80k	20k	9.85
10/1/2015	1923.819946	0.88423357	0.05780446				9.85
10/2/2015	1951.359985	0.94203803	-0.1593303	Sell 0.25 Stocks (+0.5k)	80.5k	19.5k	9.6
10/5/2015	1987.050049	0.78270773	-0.051906534				9.6
10/6/2015	1979.920044	0.730801197	0.060071933	Buy 0.3k (+0.15 Stocks)	80.2k	19.8k	9.75
10/7/2015	1995.829956	0.79087313	-0.17110536	Sell 0.3 Stocks (+0.6k)	80.8k	19.2k	9.45
10/8/2015	2013.430054	0.61976777	-0.055320739				9.45
10/9/2015	2014.890015	0.564447031	-0.155877627	Sell 0.25 Stocks (+0.5k)	81.3k	18.7k	9.2
10/12/2015	2017.459961	0.408569404	0.100210685	Buy 0.8k (+0.4 Stocks)	80.5k	19.5k	9.6

בסיום המשבר, יש ברשותנו 80.5k כסף שאיננו מושקע, ו-9.6 מניות ששווין 19.4k (לפי שווי של 2017 דולר למניה). כלומר, סה"כ ערך כספי של 99.9, ולכן יש שיפור מסוים.

ניתן לראות שפוטנציאל המכירה לא ממומש במלואו לעומת פוטנציאל הקנייה, מכיוון שהוא מוכפל בסכום שהושקע, שהוא משמעותית נמוך יותר. לכן, ננסה לפצל את k ל- k_1, k_2 , כאשר k_1 יהיה הפרמטר עבור הקנייה, ו- k_2 עבור המימוש. נבחן מצב שבו k_2 קטן יותר, כדי שסכומי המימוש יעלו.

Date	Close	Label	Gradient	Actions (k1=2, k2=0.5)	uninvested	invested	Stocks
8/17/2015	2102.439941	0.379858469	0.011985069		100k	0k	
8/18/2015	2096.919922	0.391843538	0.110201707	Buy 1.2k (+0.6 Stocks)	98.8k	1.2k	0.6
8/19/2015	2079.610107	0.502045245	0.188237728	Buy 3.2k (+1.5 Stocks)	95.6k	4.4k	2.1
8/20/2015	2035.72998	0.690282973	0.309717027	Buy 8.6k (+4.2 Stocks)	87k	13k	6.3
8/21/2015	1970.890015	1	0				6.3
8/24/2015	1893.209961	1	0				6.3
8/25/2015	1867.609985	1	0				6.3
8/26/2015	1940.51001	1	0				6.3
8/27/2015	1987.660034	1	-0.170720173	Sell 2.7 Stocks (+5.4k)	92.4k	7.6k	3.6
8/28/2015	1988.869995	0.829279827	0.01961193				3.6
8/31/2015	1972.180054	0.848891757	0.151108243	Buy 2k (+1 Stocks)	90.4k	9.6k	4.6
9/1/2015	1913.849976	1	0				4.6
9/2/2015	1948.859985	1	-0.049052182	Sell 1.1 Stocks (+2.1k)	92.5k	7.5k	3.5
9/3/2015	1951.130005	0.950947818	0.009631304				3.5
9/4/2015	1921.219971	0.960579123	-0.049356276	Sell 0.9 Stocks (+1.7k)	94.2k	5.8k	2.6
9/8/2015	1969.410034	0.911222847	-0.015351451				2.6
9/9/2015	1942.040039	0.895871396	-0.018432969				2.6
9/10/2015	1952.290039	0.877438427	-0.031705122				2.6
9/11/2015	1961.050049	0.845733305	-0.03239129				2.6
9/14/2015	1953.030029	0.813342015	-0.026596289				2.6
9/15/2015	1978.089966	0.786745725	-0.086626637	Sell 0.8 Stocks (+1.6k)	95.8k	4.2k	1.8
9/16/2015	1995.310059	0.700119088	0.08197213	Buy 0.6k (+0.3 Stocks)	95.2k	4.8k	2.1
9/17/2015	1990.199951	0.782091219	0.217908781	Buy 4.2k (+2.1 Stocks)	91k	9k	4.2
9/18/2015	1958.030029	1	-0.145463991	Sell 1.75 Stocks (+3.4k)	94.4k	5.6k	2.45
9/21/2015	1966.969971	0.854536009	0.145463991	Buy 2k (+1 Stocks)	92.4k	7.6k	3.45
9/22/2015	1942.73999	1	-0.030350983				3.45
9/23/2015	1938.76001	0.969649017	0.030350983				3.45
9/24/2015	1932.23999	1	0				3.45
9/25/2015	1931.339966	1	0				3.45
9/28/2015	1881.77002	1	0				3.45
9/29/2015	1884.089966	1	0				3.45
9/30/2015	1920.030029	1	-0.11576643	Sell 1.3 Stocks (+2.5k)	94.9k	5.1k	2.15
10/1/2015	1923.819946	0.88423357	0.05780446				2.15
10/2/2015	1951.359985	0.94203803	-0.1593303	Sell 1 Stocks (+2k)	96.9k	3.1k	1.15
10/5/2015	1987.050049	0.78270773	-0.051906534				1.15
10/6/2015	1979.920044	0.730801197	0.060071933	Buy 0.35k (+0.18 Stocks)	97.25k	2.75k	1.33
10/7/2015	1995.829956	0.79087313	-0.17110536	Sell 0.6 Stocks (+1.1k)	98.35k	1.65k	0.73
10/8/2015	2013.430054	0.61976777	-0.055320739				0.73
10/9/2015	2014.890015	0.564447031	-0.155877627	Sell 0.3 Stocks (+0.65k)	99k	1k	0.43
10/12/2015	2017.459961	0.408569404	0.100210685	Buy 1k (+0.5 Stocks)	98k	2k	0.93

בסיום המשבר, יש ברשותנו 98k כסף שאיננו מושקע, 0.931 מניות ששווין 1.9k (לפי שווי של 2017 דולר למניה). כלומר, סה"כ ערך כספי של 99.9k.

המסקנה הסופית לגבי אסטרטגיה #2 היא שהיא נכשלה כישלון חרוץ. ניסוי של ערכי k רבים לא הצליח להביא אותה להצלחה.

אסטרטגיה #3 :

Date	Close	Label	k1=0.9, k2=0.65, k3=0.7	invested	uninvested	Stocks
8/17/2015	2102.439941	0.3798585				
8/18/2015	2096.919922	0.3918435				
8/19/2015	2079.610107	0.5020452				
8/20/2015	2035.72998	0.690283		100,000	0	
8/21/2015	1970.890015	1	Buy 70,000 (+35.52 Stocks)	30,000	70,000	35.52
8/24/2015	1893.209961	1	Buy 21,000 (+11.09 Stocks)	9,000	91,000	46.61
8/25/2015	1867.609985	1	Buy 6,300 (+3.37 Stocks)	2,700	97,300	49.98
8/26/2015	1940.51001	1	Buy 1,890 (+0.97 Stocks)	810	99,190	50.96
8/27/2015	1987.660034	1	Buy 567 (+0.29 Stocks)	243	99,757	51.24
8/28/2015	1988.869995	0.8292798				
8/31/2015	1972.180054	0.8488918				
9/1/2015	1913.849976	1				
9/2/2015	1948.859985	1				
9/3/2015	1951.130005	0.9509478				
9/4/2015	1921.219971	0.9605791				
9/8/2015	1969.410034	0.9112228				
9/9/2015	1942.040039	0.8958714				
9/10/2015	1952.290039	0.8774384				
9/11/2015	1961.050049	0.8457333				
9/14/2015	1953.030029	0.813342				
9/15/2015	1978.089966	0.7867457				
9/16/2015	1995.310059	0.7001191				
9/17/2015	1990.199951	0.7820912				
9/18/2015	1958.030029	1				
9/21/2015	1966.969971	0.854536				
9/22/2015	1942.73999	1				
9/23/2015	1938.76001	0.969649				
9/24/2015	1932.23999	1				
9/25/2015	1931.339966	1				
9/28/2015	1881.77002	1				
9/29/2015	1884.089966	1				
9/30/2015	1920.030029	1				
10/1/2015	1923.819946	0.8842336				
10/2/2015	1951.359985	0.942038				
10/5/2015	1987.050049	0.7827077				
10/6/2015	1979.920044	0.7308012				
10/7/2015	1995.829956	0.7908731				
10/8/2015	2013.430054	0.6197678	Sell 13.21 Stocks (+26,605)	26,848	73,152	38.03
10/9/2015	2014.890015	0.564447	Sell 11.08 Stocks (+22,325)	49,173	50,827	26.95
10/12/2015	2017.459961	0.4085694	Sell 10.58 Stocks (+21,347)	70,520	29,480	16.37

בסיום המשבר, יש ברשותנו 70.5k כסף שאיננו מושקע, ו16.37 מניות ששווין 33k (לפי שווי של 2017 דולר למניה). כלומר, סה"כ ערך כספי של 103.5k.

מדובר ברווח נקי של 3.5k.

ננסה היפר-פרמטרים אחרים :

Date	Close	Label	k1=0.85, k2=0.6, k3=0.5	invested	uninvested	Stocks
8/17/2015	2102.439941	0.3798585				
8/18/2015	2096.919922	0.3918435				
8/19/2015	2079.610107	0.5020452				
8/20/2015	2035.72998	0.690283		100,000	0	
8/21/2015	1970.890015	1	Buy 50,000 (+25.36 Stocks)	50,000	50,000	25.36
8/24/2015	1893.209961	1	Buy 25,000 (+13.21 Stocks)	25,000	75,000	38.57
8/25/2015	1867.609985	1	Buy 12,500 (+6.69 Stocks)	12,500	87,500	45.27
8/26/2015	1940.51001	1	Buy 6,250 (+3.22 Stocks)	6,250	93,750	48.49
8/27/2015	1987.660034	1	Buy 3,125 (+1.57 Stocks)	3,125	96,875	50.06
8/28/2015	1988.869995	0.8292798				
8/31/2015	1972.180054	0.8488918				
9/1/2015	1913.849976	1	Buy 1,562.5 (+0.82 Stocks)	1,562.50	98,437.50	50.88
9/2/2015	1948.859985	1	Buy 781.25 (+0.4 Stocks)	7,81.25	99,218.75	51.28
9/3/2015	1951.130005	0.9509478	Buy 371.46 (+0.19 Stocks)	409.78	99,590.21	51.47
9/4/2015	1921.219971	0.9605791	Buy 196.81 (+0.1 Stocks)	212.97	99,787.02	51.57
9/8/2015	1969.410034	0.9112228	Buy 97.03 (+0.05 Stocks)	115.93	99,884.06	51.62
9/9/2015	1942.040039	0.8958714				
9/10/2015	1952.290039	0.8774384				
9/11/2015	1961.050049	0.8457333				
9/14/2015	1953.030029	0.813342				
9/15/2015	1978.089966	0.7867457				
9/16/2015	1995.310059	0.7001191				
9/17/2015	1990.199951	0.7820912				
9/18/2015	1958.030029	1				
9/21/2015	1966.969971	0.854536				
9/22/2015	1942.73999	1				
9/23/2015	1938.76001	0.969649				
9/24/2015	1932.23999	1				
9/25/2015	1931.339966	1				
9/28/2015	1881.77002	1				
9/29/2015	1884.089966	1				
9/30/2015	1920.030029	1				
10/1/2015	1923.819946	0.8842336				
10/2/2015	1951.359985	0.942038				
10/5/2015	1987.050049	0.7827077				
10/6/2015	1979.920044	0.7308012				
10/7/2015	1995.829956	0.7908731				
10/8/2015	2013.430054	0.6197678				
10/9/2015	2014.890015	0.564447	Sell 10.91 Stocks (+21,974)	22089.94	77910.06	40.71397
10/12/2015	2017.459961	0.4085694	Sell 11.58 Stocks (+23,373)	45462.94	54537.06	29.12861

בסיום המשבר, יש ברשותנו 45.5k כסף שאיננו מושקע, ו29.12 מניות ששווין 58.7k (לפי שווי של 2017 דולר למניה). כלומר, סה"כ ערך כספי של 104.2k.

מדובר ברווח נקי של 4.2k.

הערות

1. לאורך הצגת תוצאות האסטרטגיות ולשם הפשטות, הוצגו מספרים לא מציאותיים של כמות מניות (כמובן שאין דבר כזה לקנות "חצי מניה"). ניתן להתגבר על הבעיה באמצעות –
 - a. השקעה של סכום כסף גדול יותר.
 - b. בחינה של האסטרטגיות ברמה המציאותית ע"י לקיחה של ערך שלם תחתון/עליון של כמות המניות המיועדות לקניה/מכירה.
2. את המניות שנותרו בסיומה של אסטרטגיה מסוימת המרנו לכסף על פי שער המניה ביום האחרון של המשבר, וזאת כדי לשערך את סכום הכסף שנותר, מה שעזר בהערכת טיב האסטרטגיה.

מסקנות

1. אסטרטגיה #1 : ניתן לראות שהיא מוצלחת במונחי ROI (עלות מול תועלת) – רווחים נאים ביחס לתקופת ההשקעה (2% בכ-50 יום) אל מול זמן השקעה מינימלי.
2. אסטרטגיה #2 : כשלה, להערכתנו עקב חוסר יכולת של המסווג "להתאפק" עם השקעות בתחילת משבר, וזאת לאור מחיר מינימלי למשבר שהולך וקטן ככל שמעמיקים בו. הבעיה בכך היא שאת מרבית הכסף המסווג "מבזבז" על מחירים שלא מגלמים את פוטנציאל הרווח של התקופה.
3. אסטרטגיה #3 : המוצלחת ביותר (אך יש לציין כי דורשת התעסקות בתדירות גבוהה). הצליחה להגיע לרווחים מצוינים של 3.5%-4.2% בכ-50 יום. הסיבות להצלחה –
 - a. k_1 גבוה המבטא חוסר פשרה בימים פוטנציאליים לכניסה לשוק.
 - b. k_3 נמוך המבטא מעין סולידיות שמאפשרת להזניח אי דיוקים של המסווג לאורך זמן.
4. ככל שנמצאים בשלב עמוק יותר במשבר, כך המסווג מדייק יותר את הערכתו. לכן הערכתנו היא שהמסווג יכול לתת ביצועים מרשימים מאוד במשברים ארוכים.
5. חלק מאי הדיוקים נובעים, להערכתנו, מהקושי להגדיר מהו משבר ומהן נקודות הכניסה והיציאה שלו.

בחינת אסטרטגיות #1, #3 על קבוצת מבחן

על מנת לאשש את המסקנות שלעיל, נבדוק את הפרמטרים המוצלחים ביותר של אסטרטגיות #1, #3 על קבוצת מבחן (משבר אחר).

- אסטרטגיה #1 עם הפרמטרים $k_1=0.8$, $k_2=0.6$.
- אסטרטגיה #3 עם הפרמטרים $k_1=0.85$, $k_2=0.6$, $k_3=0.5$.

Date	Close	Label	Strategy #1	Strategy #3			
			actions	actions	uninvested	invested	stocks
10/3/2018	2925.51001	0			100k	0k	0
10/4/2018	2901.610107	0.084077613			100k	0k	0
10/5/2018	2885.570068	0.140504962			100k	0k	0
10/8/2018	2884.429932	0.144515854			100k	0k	0
10/9/2018	2880.340088	0.15890354			100k	0k	0
10/10/2018	2785.679932	0.491909073			100k	0k	0
10/11/2018	2728.370117	0.693519616			100k	0k	0
10/12/2018	2767.129883	0.557166402			100k	0k	0
10/15/2018	2750.790039	0.614648438			100k	0k	0
10/16/2018	2809.919922	0.406635066			100k	0k	0
10/17/2018	2809.209961	0.409132642			100k	0k	0
10/18/2018	2768.780029	0.551361343			100k	0k	0
10/19/2018	2767.780029	0.554879249			100k	0k	0
10/22/2018	2755.879883	0.596742845			100k	0k	0
10/23/2018	2740.689941	0.650179633			100k	0k	0
10/24/2018	2656.100098	0.947758751	Buy 100k (+37.65 Stocks)	Buy 47k (+17.7 Stocks)	53k	47k	17.7
10/25/2018	2705.570068	0.773728046			53k	47k	17.7
10/26/2018	2658.689941	0.938647927		Buy 24.9k (+9.35 Stocks)	28.1k	71.9k	27.05
10/29/2018	2641.25	1		Buy 14k (+5.3 Stocks)	14.1k	85.9k	32.35
10/30/2018	2682.629883	0.854429461		Buy 1.1k (+0.4 Stocks)	13k	87k	32.75
10/31/2018	2711.73999	0.75202284			13k	87k	32.75
11/1/2018	2740.370117	0.651304744			13k	87k	32.75
11/2/2018	2723.060059	0.712199901			13k	87k	32.75
11/5/2018	2738.310059	0.658551834			13k	87k	32.75
11/6/2018	2755.449951	0.598255305	Sell 37.65 Stocks (+103.75k)	Sell 6.35 Stocks (+17.5k)	30.5k	69.5k	26.4
11/7/2018	2813.889893	0		Sell 12.3 Stocks (+34.75k)	65.25k	34.75k	14.1
11/8/2018	2806.830078	0.062860178		Sell 5.8 Stocks (+16.33k)	81.58k	18.42k	8.3
11/9/2018	2781.01001	0.292760548		Sell 2.35 Stocks (+6.5k)	88.08k	11.92k	5.95
11/12/2018	2726.219971	0.780607839			88.08k	11.92k	5.95
11/13/2018	2722.179932	0.816580109	Buy 103.75k (+38.11 Stocks)		88.08k	11.92k	5.95
11/14/2018	2701.580078	1		Buy 44k (+16.3 Stocks)	44.08k	55.92k	22.25
11/15/2018	2730.199951	0.745170331			44.08k	55.92k	22.25
11/16/2018	2736.27002	0	Sell 38.11 Stocks (+104.3k)	Sell 10.2 Stocks (+28k)	72.08k	27.92k	12.05
11/19/2018	2690.72998	0.439109605		Sell 2.9 Stocks (+7.8k)	79.88k	20.12k	9.15
11/20/2018	2641.889893	0.910039172		Buy 36.3k (+13.8 Stocks)	43.58k	56.42k	22.95
11/21/2018	2649.929932	0.832514902					

- אסטרטגיה #1 : הביאה לרווחים של 4.3k (4.3% בתקופת זמן של חודש וחצי), וכזכור עם מינימום השקעה של זמן.
- אסטרטגיה #3 : בסיום המשבר, יש ברשותנו 43.58k כסף שאיננו מושקע, ו-22.95 מניות ששווין 60.8k (לפי שווי של 2649 דולר למניה). כלומר, סה"כ ערך כספי של 104.4k (4.4% רווח בתקופת זמן של חודש וחצי)

להערכתנו, התוצאות של אסטרטגיות אלו על קבוצת המבחן מהוות הוכחה נוספת למסקנות שהצגנו.

סיכום

התחבטנו רבות בשאלת החיזוי של הפרויקט. ידענו שאנחנו מעוניינים לעשות פרויקט בשוק ההון, אך הנחנו שמחקרים רבים ניסו לחזות התנהגות של מניה, ורצינו לבחון משהו חדש. עקב התפתחות נגיף הקורונה והמשבר שנוצר בעקבותיה, עלה הרעיון הבא - חשבנו שניתוח מגמות בשוק ההון בזמן משברים, מכיל בתוכו הן מאפיינים טכניים מעניינים, והן מאפיינים פסיכולוגיים שמשחקים תפקיד מכריע בהתנהלות של השוק בתקופות כאלה.

במהלך הפרויקט נתקלנו במספר נקודות מפתח :

- למדנו על הבורסה ואיך היא עובדת. קראנו על אסטרטגיות קיימות של משקיעים ווידאנו שהאינדיקטורים בהם הם משתמשים, יעמדו לרשות המסווג שלנו.
- הבנו שהאלגוריתמים שלמדנו בקורס "מבוא לבינה מלאכותית" לאו דווקא ישיגו את התוצאות אליהם ייחלנו, בגלל האופי הסדרתי של ה-Data. לאחר סקירה של מספר כתבות בנושא, החלטנו שכדאי לנסות ללמוד איך רשת LSTM עובדת.
- למדנו על רשתות נוירונים, רשתות LSTM - כיצד בונים אותן, מעריכים את הדיוק שלהן וכו'.

בסוף הפרויקט התחבטנו בשאלה – **איך הופכים Data לידע?**

אמנם תוצאות הפרויקט הם במונחים של עולם הפתרון (AI / Deep Learning), אך רצינו **שהמסקנות של הפרויקט יהיו במונחים של עולם הבעיה** (שוק ההון). הרציונל היה לספק תשובה ישירה לשאלת החיזוי, ולא רק להציג נתונים שקשה לגזור מהן מסקנה מדויקת או תוכנית עבודה. מכאן הגיע הרעיון של ניסוח אסטרטגיות מסחר בהתאם לתוצאות המסווג, וניתוחן.

כיוונים לפיתוח עתידי

1. **טיוב הגדרת משבר** : כפי שראינו לאורך המחקר, לפעולת התחימה של המשברים שביצענו בחלק הראשון יש חשיבות והשפעה רבה על כל תהליך הסיווג, ועל טיב התוצאות. כתוצאה מכך, בראייתנו, הדבר הנכון לעשות הוא לחשוב על פרמטרים נוספים שעשויים להגדיר משבר, ולבחון את דיוק ההגדרות על Data מהעבר.
2. **ניסויים על מניות ספציפיות** : לאורך המחקר נצמדנו לביצועי S&P500 שמגלם בתוכו מניות טכנולוגיה מובילות משוק המניות האמריקאי. יהיה מעניין לראות ביצועים של המסווג על חברה ספציפית. כאמור, חיזוי של מדדים על סמך תכונות עשוי להיות קשה יותר, היות שהוא לא מגלם ביצועים של חברה בודדת. במקרה של סיווג חברה בודדת ניתן לבחון הוספה של תכונות נוספות, שקשורות באופן ספציפי יותר לאותה החברה או למגזר שלה.

3. **בחינה של תכונות נוספות :** כידוע, S&P500 מורכב מביצועים של הרבה חברות, וחלקן דומיננטיות מאוד בהרכב של המדד. ניתן ורצוי לבחון תוספת של תכונות של חברות אלו, אשר מהוות חלק ניכר מהמדד, ולראות כיצד הן השפיעו על ביצועי המסווג.
4. **בחינה של אסטרטגיות המסחר לאורך עשרות משברים :** הניתוח שעשינו בשלב האחרון ומסקנותיו נובעות מניתוח של משברים בודדים. אנו מאמינים שניתוח על כמות גדולה יותר של משברים, ועל משברים מגוונים יותר באופיים (אורכם, "עומק" המשבר וכו') יכולים להביא לתמונת מצב טובה יותר של דיוק המסווג, וכן לכוונון טוב יותר של הפרמטרים באסטרטגיות.
5. **בחינה של הוספת שכבות לרשת :** כאמור, אחת המסקנות שהובילו אותנו לעבור לעסוק בלמידה עמוקה, היא שיש לבעיה הזו אופי "סדרתי", ולא ניתן להתייחס לדוגמאות כ-Single Data Points. כאמור, LSTM ידועה בתור רשת בעלת זיכרון. ברצוננו לבחון הוספה של שכבות נוספות לרשת שיסייעו בזיכרון של פרטים נוספים שעשויים לסייע בסיווג מדויק יותר. לאחר ההוספה יש לחזור על השלבים שהוצגו בדו"ח על מנת לבחור פיצ'רים והיפר-פרמטרים ולהעריך ביצועים.

הערות ונספחים

כיצד לשחזר את הניסויים?

בחלק זה נתאר כיצד ניתן לשחזר את הניסויים שתוארו בדו"ח זו.

ראשית, כדי לקבל הבנה טובה יותר, ניתן לעיין בקבצים README.txt (יסביר בקצרה על כל הקבצים שנמצאים תחת פרויקט זה) ו-libs.txt (יסייע בקונפיגורציות הדרושות לסביבת ה-Python).

הקובץ main.py מכיל את התוכנית הראשית של הפרויקט. התוכנית הראשית בנויה כך שיהיה קל לשחזר את הניסויים השונים והפעולות השונות שתיארנו :

- יצירת ה-CSV-ים – חלק זה אמון על יצירת ה-CSVים השונים ששימשו כקלטים לאלגוריתמים. על מנת לשחזר זאת ניתן להוריד את התיעוד מהקריאה לפונקציה csv_create().
- ניסוי SVR, Random Forest, KNN – חלק זה אמון על המדידות שביצענו באלגוריתמים אלה. על מנת לשחזר את הניסויים שכללו הרצה של האלגוריתמים, feature selection ו-cross validation, יש להסיר את התיעוד של part1. כדי לצפות בתוצאות הגולמיות של הניסויים מבלי הרצה מחדש, ניתן לעיין בקובץ featureselectionRes.txt.
- ניסוי LSTM לבחירת ת"ק של תכונות – על מנת לשחזר את הניסויים הללו שכללו השוואות בין פרמוטציות שונות של תתי קבוצות, תחת הרציונל שהוסבר בפרויקט זה, יש להסיר את התיעוד של part2. התוצאות של ניסויים אלה מובאות במלואן בדו"ח זה.
- ניסוי LSTM לכוונון היפר פרמטרים – על מנת לשחזר את הניסויים הללו, שכללו כוונון של היפר פרמטרים של המודל וחקר ביצועים (מהם יצרנו את מפת החום), תחת הרציונל שהוסבר בפרויקט זה, יש להסיר את התיעוד של part3. כדי לצפות בתוצאות הגולמיות של הניסויים מבלי הרצה מחדש, ניתן לעיין בקובץ param_lstm.txt.
- החלק האחרון הוא המסווג הסופי שנבחר כזה שנתן את התוצאות הטובות ביותר, על פי הרציונל שפורט בפרויקט. על מנת לבחון אותו יש להסיר את התיעוד של last part.
- חישוב התוצאות של אסטרטגיות המסחר בוצע באופן ידני, והוא מובא בתצורתו הגולמית בקובץ strategies.xls.

GitHub - <https://github.com/yakir-yehuda/AI-Project>

ביבליוגרפיה

[1]	https://www.maariv.co.il/business/world/Article-760391
[2]	https://he.wikipedia.org/wiki/%D7%94%D7%9E%D7%A9%D7%91%D7%A8_%D7%94%D7%9B%D7%9C%D7%9B%D7%9C%D7%99%D7%94%D7%A2%D7%95%D7%9C%D7%9E%D7%99_(2008)
[3]	https://zoefin.com/learn/market-timing-vs-time-in-the-market/
[4]	https://www.codecademy.com/articles/normalization
[5]	https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.mutual_info_regression.html
[6]	https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesRegressor.html
[7]	https://he.wikipedia.org/wiki/%D7%9E%D7%9B%D7%95%D7%A0%D7%AA_%D7%95%D7%A7%D7%98%D7%95%D7%A8%D7%99%D7%9D_%D7%AA%D7%95%D7%9E%D7%9B%D7%99%D7%9D
[8]	https://towardsdatascience.com/understanding-random-forest-58381e0602d2
[9]	https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_absolute_error.html#sklearn.metrics.mean_absolute_error
[10]	https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21
[11]	https://machinelearningmastery.com/loss-and-loss-functions-for-training-deep-learning-neural-networks/
[12]	https://commandwallstreet.com/%D7%9E%D7%A1%D7%97%D7%A8-%D7%90%D7%9C%D7%92%D7%95%D7%A8%D7%99%D7%AA%D7%9E%D7%99/%D7%90%D7%A1%D7%98%D7%A8%D7%98%D7%92%D7%99%D7%99%D7%AA-%D7%97%D7%96%D7%A8%D7%94-%D7%9C%D7%A9%D7%95%D7%A7/