

**המחלקה: הנדסת תוכנה וחשמל**

**שם הפרויקט: מערכת משולבת למסחר בשוק ההון**

**Project Name:** Integrated Trading System

**הגדרת הפרויקט – Statement Of Work**

|  |  |
| --- | --- |
| **שם הסטודנט:** | **אבירם זוזות, חגי ז'ביליק, יריב מזרחי** |
| **מספר תעודת זהות:** | 313576654, 036828069 , 304965890 |
| **שם המנחה:** | **שטקל עמית** |
| **חתימת המנחה:** | **(חובה להחתים את המנחה לפני שהמסמך מוגש. החתימה מציינת שהתוצר אושר על ידו. יש לסרוק את הדף החתום ולהוסיף כעמוד הראשון של התוצר. ניתן לחילופין לצרף כעמוד שני למסמך עותק של מייל מהמנחה לפיו הוא מאשר אותו)** |
| **ייועץ מומחה:** | **(אדם אשר עוזר לסטודנט בתחום אשר דורש מומחיות מיוחדת, עזרה של לפחות 5 שעות, אם אין כזה יש למחוק את השורה)** |
| **תאריך ההגשה:** | **26/772020 \_** |

**תוכן עניינים:**

[**4. מבוא**](#_heading=h.kkso12wpidi3) **3**

[**5. מטרות הפרויקט**](#_heading=h.yxacif3mwl1v) **5**

[**6. סקירת ספרות ראשונית**](#_heading=h.rcv41j5vso46) **7**

[סקר שוק ראשוני](#_heading=h.ib53ad8lv3p0) 8

[**7. דרישות ראשוניות מהמערכת**](#_heading=h.dqseo9v43267) **10**

[דרישות פונקציונליות](#_heading=h.5648ryuex13f) 10

[דרישות לא פונקציונאליות](#_heading=h.kmc4tbyvr6sn) 10

[**8. תרשים בלוקים/ מודולים כללי של המערכת**](#_heading=h.aqk8zsnee13x) **12**

[**9. ניתוח פונקציונאלי ראשוני**](#_heading=h.svyomf80tn9w) **13**

[**10. הצגה ראשונית של חלופות מערכתיות**](#_heading=h.4283oqnx2u86) **14**

[**11. האמצעים/הכלים הנדרשים**](#_heading=h.mzo5s5gtuoqg) **17**

[**12. תוצרי הפרויקט**](#_heading=h.2gsfsbxfzdcl) **18**

[**13.תכנית עבודה ראשונית לביצוע הפרויקט**](#_heading=h.m491hqqqhbrq) **19**

[**14. פערים**](#_heading=h.tto4j72vzcbj) **21**

[**15. סיכונים עיקריים ודרכי התמודדות (בהיבטים טכניים ומנהלתיים בלבד)**](#_heading=h.s26febnn3jry) **22**

[**16.רשימת מקורות (ראשונית)**](#_heading=h.nup0werfgowj) **23**

# 4. מבוא

המסחר בשוק ההון האמריקאי היומי הינו בעל נפח מסחר גבוה ומגיע עד לחמישים מיליארד דולר*(1)*. המסחר מתבצע ע"י 3 סוגי סוחרים (המתחיל, המנוסה והמתקדם) אשר מנסים להרוויח ממחזור כסף מתגלגל בשוק. מניית אמזון עשתה ב- 5 שנים האחרונות (2015-2020) 600% תשואה, סוחר אשר ידע להשקיע את כספו ב- 2015 יכול היה להכפיל את ההשקעה שלו פי 6.

המסחר מתחלק בין השוק הפרטי למסחר הקבוצתי (בתי השקעות ובנקים), כאשר אופן המסחר מתבצע לרוב באופן אוטומטי(80%) . כיום חלק מהסוחרים מסתמכים על שיטות שונות, כאשר חלק מסתמכים על הניסיון שלהם והשאר מתבססים על אלגוריתם אוטומטי. סוחרים פרטיים מתקשים לעקוב אחר מספר רב של מניות, לזהות מגמות, לתזמן את המכירה והקניה, לשמור על קור רוח אל מול תנודות בשער מניה ועוד. גורמים אלו ואחרים מורידים סטטיסטית את מידת ההצלחה של הסוחר הממוצע אשר חשוף לבעיות רבות כגון: חוסר ידע, אי וודאות וכו'.

מכאן אנו שואבים את המוטיבציה לבנות מערכת מסחר משולבת לשוק ההון אשר תדע להעלות את אחוזי ההצלחה באופן סטטיסטי, על ידי הערכת מספר רב של הזדמנויות לרווח מקסימלי במקביל בכמה שפחות עבודה ידנית במינימום זמן עיבוד, ראשית למשקיעים הפרטיים ולכל גוף מוסדי אשר ירצה חוות דעת נוספת על הפתרונות הקיימים.

כיום חברות גדולות משתמשות בפתרונות כגון: אנשי מקצוע מנוסים או בתוכנות חיזוי אשר מסתמכות לרוב על נתונים יבשים משוק ההון. בפתרון שאנו נציע נשלב את שני הפתרונות הללו ונוסיף אלמנטים נוספים כגון: הרחבת מקורות המידע וחיבור בין עולמות תוכן שונים. האתגר הוא הערכת מקורות מידע רלוונטיים ומידת השפעתם באופן אוטומטי, התוכנה תדע לשנות באופן דינמי את בחירתם ומידת השפעתם של מקורות שונים. ביטוי של הצלחת המערכת הוא השגת תשואה של לפחות 10% ומעלה באופן סטטיסטי, כאשר עם הזמן נשאף גבוה יותר.

מטרת העל של המערכת היא לתת מענה לאנשים המעוניינים להניב תשואה על כספם בצורה יעילה ואוטומטית ככל הניתן, תוך מינימום סיכון.

חלוקת הפרויקט בין חברי הצוות תעשה בצורה מיניסטריאלית בלבד, כלומר לכל חבר צוות יש תחום שהוא אחראי על התקדמותו, הן בהיבט מנהלתי והן בהיבט פיתוח, אך על מנת שכל חברי הצוות יכירו את כל חלקי הפרויקט ושכל חבר צוות ירגיש שהוא תורם בכל תחום - כל חבר צוות יקבל משימות במס' תחומים.

חלוקת האחריות תתבצע באופן הבא:

* כריית מידע טקסטואלי ממספר רב של מקורות מידע באופן אוטומטי - חגי
* השלמת מידע חסר, תיקון מידע, הסרת מידע שאינו רלוונטי, והכנת מטריצה מתאימה לאימון - יריב
* אימון מודלים טקסטואלים, חילוץ יישויות, הבנת סמנטיקה וניפוק תובנה למע' - אבירם
* בניית מודלים המסתמכים על המודלים הטקסטואלים ועל הניתוח הטכני בפרט ומקורות מידע נוספים ככלל - יריב
* בניית מע' הכוללת צד שרת וממשק משתמש - אבירם

# 5. מטרות הפרויקט

1. מטרות:

יצירת מערכת אוטומטית המבוססת על למידת מכונה, אשר תבצע שימוש וסחר בניירות ערך בשוק ההון. זאת תוך ניתוח מודלים פיננסיים, ניהול סיכונים וניתוח הודעות מתפרצות רלוונטיות ממקורות מידע ברשת ומדיה חברתית, כאשר שאיפת המערכת לייצר תשואה חיובית בשוק ההון, תוך מתן דגש על אוטומציה וחיסכון בזמן.

1. יעדים:
   1. איסוף נתונים - בקשת מידע וריכוזו, מאתרי מדיה ומשוק המניות האמריקאי לאורך 20 השנים האחרונות.
   2. עיבוד מידע ראשוני -
   3. לימוד מודל באמצעות למידת מכונה - ייצור מודל על בסיס המידע שנאסף היודע לתאר את עבר התנהגות המניה וידע לחזות את העתיד.
   4. דירוג מניות - הערכת ניירות ערך בזמן אמת לפי פרמטרים סטטסטיים (נתוני מסחר בורסה) ודינמיים (ניתוח טקסטים במדיה חברתית).
   5. ניהול מסחר - פיתוח תהליך פקודות מסחר יומיות בבורסה.
   6. ניהול סיכוני תיק - בניית מודל מתאים ליישום תהליך ניהול סיכונים.
   7. הצגה - בניית ממשק משתמש להגדרת והפעלת המערכת.
2. מדדים:
   1. איסוף נתונים - איסוף נתונים טקסטואליים מ2 מקורות מידע לפחות, ועיבוד של 20 טקסטים כלכליים חדשים(או ידיעות חדשותיות) לפחות בכל יום.
   2. עיבוד מידע ראשוני - משתנה.
   3. לימוד מודל באמצעות למידת מכונה או למידה עמוקה- צפי לאחוז הצלחה גבוה בשלב ה-validation.
   4. דירוג מניות - ניהול רשימה של 10 המניות המומלצת למשתמש ברמה יומית.
   5. ניהול מסחר - בחירה מתוך רשימה של לפחות 100 ניירות( מניות) והשקעה בו זמנית ב- 10 מניות לפחות.
   6. ניהול סיכוני תיק - השגת יחס רווח לסיכון (מדד שארפ) של לפחות 0.5.

# 6. סקירת ספרות ראשונית

מסחר בשוק ההון מתבצע דרך הבורסה שהיא המוסד הרשמי בו מתקיים בפועל סחר החליפין וזאת באמצעות פעולות של קניית ומכירת ניירות ערך. כל הפעולות משותפות ב"ספר" פעולות אחד, העונה לכללים מוגדרים ומפוקחים לניהול משא ומתן על המחיר, כאשר מטרת המערכת לספק כלי מסחר לגופים פרטיים וציבוריים בשוק ההון(12).

מרבית המסחר מתבצע כיום נעשה באמצעות אלגוריתמים אוטומטיים, הניזונים ניתוחים כלכליים וטכניים של השוק, אשר אמורים לחזות את כיוון השוק. הניתוח מתבסס על חוקים פשוטים, מודלים מתמטיים ועל ידי למידת מכונה המקובלת כיום(13). החל מהקמתן של מערכות מסחר מקוונות וכלה בניתוח והמלצה לביצוע פעולות. המסחר האלגוריתמי תופס נפח גדל ועולה בסחר הכלכלי בעשורים האחרונים, כאשר השימוש בשיטת למידה עמוקה הפך לעיקר העשייה(14).

למרות שכיום עיקר העשייה מתמקדת בלמידה עמוקה, כריית טקסטים כלכליים מקבל את מירב תשומת הלב. בכך אתרים כלכליים, מדיות חברתיות, בלוגרים ואנשים בעלי דעה, אשר מפרסמים את משנתם או מפיצים שמועות ברחבי הרשת, גורמים לתנודות קיצוניות בשוק, לעיתים ללא הצדקה כלל. כאשר מוסיפים מידע זה אל מודלי הלמידה הקיימים כיום, ניתן לדייק אף יותר את תוצאות המודלים, להגיב בזמן אמת לתנודות בשוק ולשפר את יחס רווח-הפסד. הגישה המקובלת כיום היא שילוב בין כריית מידע טקסטואלי והבנה סמנטית-כלכלית של טקסטים אלו. כיום, מחקרים רבים בעולם מתעסקים בנושא זה(15), והדעה הרווחת היא כי שילוב של כריית טקסט ומודלים סטטיסטיים, הם הדור הבא של מודלי מסחר אלגוריתמי אוטומטי.

## סקר שוק ראשוני

קיימים בשוק מספר מוצרים המנתחים טקסטים מרשתות חברתיות, ניתוח מידע רלוונטי מהמדיה החברתית תופס תאוצה ואיתה ההבנה שאדם רלוונטי יכול להשפיע על מחיר מניה עם פרסום של משפט בודד. ניתן להסתכל על כמה מהמתחרים על מנת להבין את תמונת המצב:

1. SMA & NYSE**-** נוסדה 2011, רותמת מידע נרחב פיננסי באמצעות למידת מכונה. והחל מ2019 נוצר שיתוף פעולה עם NYSE. מטרתה לספק מידע ברור וזמין לגבי ניתוח שוק כדי להעריך מוצרים פיננסיים כדי להעלות תשואה ולהוריד את הסיכון. משתמשת בניתוח טקסטים, מידול ודירוג מקורות. החברה משלבת את ניתוח המידע במודלים מסחריים קיימים לניתוח שוק ההון. הזמן תגובה של הapi הוא 0.3 שניות.  
   יתרון בולט במוצרי החברה הוא מגוון האפשרויות אשר באמצעותן הלקוח יכול להתאים אישית את הניהול המידע או להכניס מידע מסוים לבחירתו. לעומת זאת, אחד החסרונות הבולטים במוצרי החברה הם החוסר בחיבור למערכת מסחר וכמו כן החוסר בהמלצות מדויקות מטעם החברה לביצוע פעולות עקב ניתוח המידע.
2. Bloomberg- מובילת דעה עולמית בתחום התקשורת וכלכלה אשר מספקת מידע פיננסי לחברות הסוחרות בוול סטריט. החל מ2013 משתפת פעולה עם חברת twitter ומבצעת ניתוח יומי של 500 מיליון טוויטים, מבליטה מידע פיננסי מוכוון על ידי ניתוח חברות, אנשים מובילי דעה, הצהרות חברה וכדומה.   
   יתרונה הבולט של החברה הוא צוות המומחים המנוסה שלה והאלגוריתמים בעלי עשור של שנות פיתוח בתחום של ניתוח מידע בזמן אמת מחברת טוויטר, אשר משתפת עימה פעולה בזירה הזו. החסרון מתאפיין בהספקת המידע אשר דורג גבוהה לפי דרך הניתוח שלה, אך אין בו קריאה לפעולה מסוימת מעבר לכך אין אפשרות לבצע פעולות מסחר אוטומטיות או ידניות כלל באתר החברה.

|  |  |  |  |
| --- | --- | --- | --- |
|  | [bloomberg](https://www.bloomberg.com/company/press/bloomberg-launches-twitter-feed-optimized-trading/) | NYSE&SMA | **MediaTrade** |
| משיכת מידע ממדיה חברתית | V | V | V |
| ניתוח שער מניה | X | V | V |
| ניתוח דוחות חברה | X | V | V |
| המלצה לפעולה | X | X | V |
| מערכת מסחר אוטומטית | X | X | V |

ניתן לראות שחלק מהחברות בוררות ומציגות את המידע הרלוונטי למניה מסוימת בעוד חלק משלבות את המידע המנותח בהמלצותיהן, כאשר אין המלצה מספרית של פעולת מסחר (קנייה, מכירה).

מערכות מסחר שכן מבצעות פעולות אוטומטיות כרגע פועלות בעיקר לפי למידת מכונה ללא שילוב ניתוח מדיה חברתית. מטרת המערכת שלנו להציע מערכת מסחר כמותי (ולא להסתפק בניתוח או המלצות), בכך לחבר את העולמות, לנתח מידע רלוונטי אשר יכול להשפיע על מחיר המניה, להמליץ על פעולת מסחר ולבצע אותה בצורה אוטומטית בהתאמה אישית לרצון משתמש.

# 7. דרישות ראשוניות מהמערכת

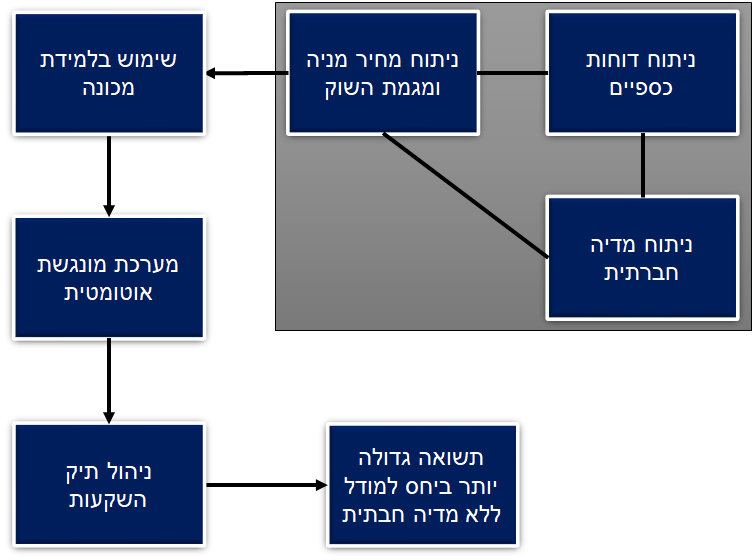
## דרישות פונקציונליות

* 1. המערכת תדע לאסוף את כלל המידע ברמה יומית על המניות בבורסה האמריקאית ותנתח את המידע במטרה להניב תובנות למשתמש.
  2. המערכת תאפשר בחירה של רמת הסיכון באחוזים.
  3. המערכת תאפשר בחירה של כמות העסקאות היומיות שהיא תבצע.
  4. המערכת תאפשר בחירה של אחוז כסף לסיכון.
  5. המערכת תבצע 2 אסטרטגיות לימוד מכונה במקביל על תיק מניות, האחת עם שילוב ניתוח מדיה חברתית והשניה ללא.
  6. המערכת תייצר דירוג מניות לפי המודל בשילוב מדיה חברתית.
  7. המערכת תדע לבצע פעולות קניה ומכירה בצורה אוטומטית לאור ההגדרות (סיכון, כמות עסקאות)
  8. המערכת תאפשר לחזור למצב התחלתי של בחירת פרמטרים להרצה (RESET)
  9. המערכת תציג בסוף היום:
     1. דירוג המניות.
     2. שווי התיק הנוכחי והשינוי ביחס ליום המסחר הקודם.
     3. כמות הפעולות שבוצעו ביום המסחר הקודם.
     4. התנהגות המניה שנבחרה ביום הקודם.

## דרישות לא פונקציונאליות

* 1. ממשק המערכת - web.
  2. בניית צד שרת ייעודי למע' הכולל נתוני משתמש, פניה אל מס' מקורות מידע (מדיה חברתית, מידע רשמי מהבורסה האמריקאית וכו'), ביצוע מניפולציות על המידע, והכנת המידע אל ממשק המשתמש.
  3. ממשק המשתמש יעבוד אל מול השרת הייעודי למערכת, אשר יכיל בתוכו את ההגדרות (בחירה ועדכון). בנוסף הממשק יציג למשתמש את פלט המערכת.
  4. הפעלה: המערכת זקוקה למחשב בעל מערכת הפעלה מסוג 'windows 7' לפחות, בעלת גישה לאינטרנט ודפדפן פעילים.
  5. ביצועים: זמן הריצה של המערכת הוא פונקציה של זמן הרצת המודלים, כמות המניות ומספר הפעולות שהוגדרו לפעולה.
  6. אבטחה - כניסה למערכת תהיה דרך שם משתמש וסיסמא.

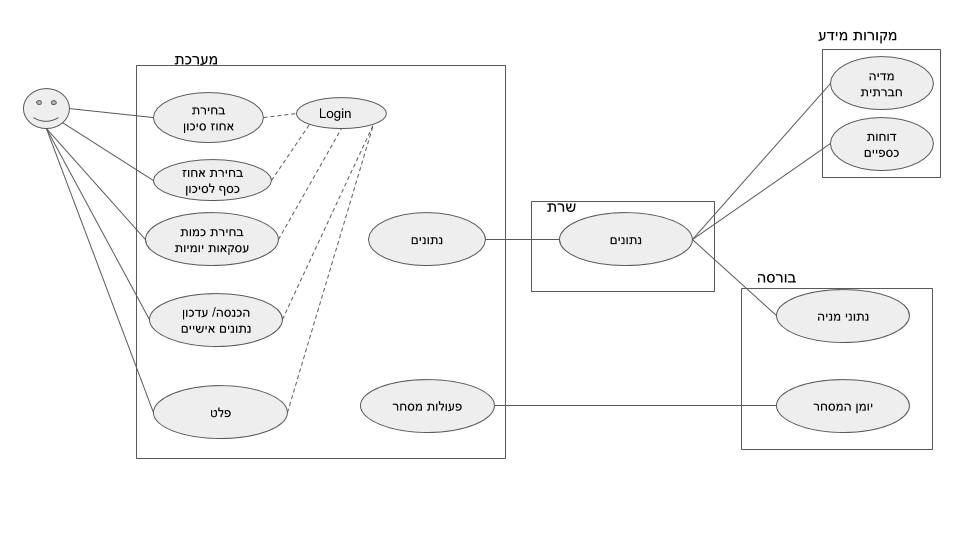
# 8. תרשים בלוקים/ מודולים כללי של המערכת

****

**הסבר תרשים בלוקים:**

השלב הראשוני הוא משיכת המידע משוק ההון האמריקאי, דוחות כספיים וניתוח מדיה חברתית. לאחר משיכת המידע המערכת תבצע מודל לניבוי מחיר המניה באמצעות למידת מכונה ותדרג את המניות לפי סדר המלצתה. לאור הגדרות המשתמש תבחר על איזה מניות לבצע פעולות מסחר כאשר בסוף התהליך נוכל להשוות את התשואה של התיק אל מול פעולות המסחר באלגוריתם ללא ניתוח מדיה חברתית.

# 9. ניתוח פונקציונאלי ראשוני

****

תיאור הUse case העיקרי במערכת:

שם התרחיש: מסחר בבורסה

שחקנים: לקוח(לקוחות פרטיים המעוניינים להשקיע), בורסה

1. לקוח מבצע התחברות למערכת.
2. לקוח בוחר הגדרות המותאמות אישית לצרכיו(נתונים אישיים, כמות עסקאות יומית, אחוז סיכון, אחוז כסף).
3. המערכת סורקת את נתוני המשתמש ומחפשת הזדמנויות עסקיות עבורו בעזרת נתונים מהבורסה וממדיות חברתיות.
4. המערכת מבצעת פקודות קנייה/מכירה באופן אוטומטי בהתאם להגדרות המשתמש.
5. המערכת מעדכנת במסך הבית את המשתמש על פעולות שהיא מבצעת, כמות הרווח/הפסד בכל רגע נתון.

# 10. הצגה ראשונית של חלופות מערכתיות

**ניתוח החלופות נועד להוות כלי לפיתוח המערכת, כאשר אנו נדרג קריטריונים שונים:**

**פשוט הפיתוח (25%) – הכרות המפתח עם הסביבה, נוחות הפיתוח.**

**עלות (15%) – עלות השימוש בחלופה, הרשאות ורישיונות.**

**ביצועים (30%) – תמיכה בגורמים שונים (מערכות הפעלה, סביבות עבודה), תמיכה בפלטפורמות ועוד.**

**פונקציונאליות (30%) – האופציות והמטרות השונות הניתנות למימוש באמצעות החלופה.**

**הדירוג יתבצע מ 1 עד 10 כאשר המספר מסמל את מידת ההתאמה, 10 יהיה מתאים לחלוטין.**

**בסוף כל דירוג ירשם משקל סופי אשר יבטא את התאמת ההצעה לפרויקט ולצוות באותו אופן של הדירוג.**

תוצרי הביניים שירכיבו את הפרויקט לפי שלבים:

1. איסוף נתונים - בקשת מידע וריכוזו, מאתרי מדיה ומשוק המניות האמריקאי לאורך 20 השנים האחרונות.
2. עיבוד מידע ראשוני - ניקוי נתונים וסינון מידע ראשוני.
3. לימוד מודל באמצעות למידת מכונה - ייצור מודל על בסיס המידע שנאסף היודע לתאר את עבר התנהגות המניה וידע לחזות את העתיד.
4. דירוג מניות - הערכת ניירות ערך בזמן אמת לפי פרמטרים סטטיסטיים (נתוני מסחר בורסה) ודינמיים (ניתוח טקסטים במדיה חברתית).
5. ניהול מסחר - פיתוח תהליך פקודות מסחר יומיות בבורסה.
6. ניהול סיכוני תיק - בניית מודל מתאים ליישום תהליך ניהול סיכונים.
7. הצגה - בניית ממשק משתמש להגדרת והפעלת המערכת.

ממשק ויזואלי להגדרות המערכת (נפח מסחר, סיכון תיק, מספר פעולות)

פלט יומי של דירוג המניות.

פלט פקודות שהתבצעו.

שווי תיק ביחס ליום מסחר קודם.

|  |  |  |
| --- | --- | --- |
| **חלופה לסביבת העבודה** | **דירוג** | **הסבר קצר** |
| **שרת אינטרנטי** | פשוט הפיתוח: 2  עלות:4  ביצועים:8  פונקציונאליות:5  **משקל סופי: 5** | זו סביבת עבודה ייעודית למטרת פיתוח אלגוריתמים.  חסרונות: קשה מאד לשלב אותה עם מערכות נוספות.  ביצועים נמוכים ומגבלות אבטחה. |
| **Visual Studio.net** | פשוט הפיתוח: 5  עלות: 8  ביצועים: 8  פונקציונאליות: 9  **משקל סופי: 7.55** | סביבת עבודה זו נוחה ומוכרת לפיתוח.  יתרון: מתקשרת עם הרבה שפות.  חסרונות: ביצועים פחות טובים עקב דרך המימוש קרי העלאת הקוד לשרת. |

***טבלה 1 – חלופות לסביבת עבודה***

|  |  |  |
| --- | --- | --- |
| **חלופה לטכנולוגיית פיתוח** | **דירוג** | **הסבר קצר** |
| **#C** | פשוט הפיתוח: 8  עלות: 3  ביצועים: 5  פונקציונאליות: 8  **משקל סופי: 6.35** | יתרונות: ספריות עשירות, פונקציונליות, syntax ברור.  יתרון נוסף: ניסיון סביר.  חסרונות: ביצועים בינוניים. |
| **++C** | פשוט הפיתוח: 5  עלות: 3  ביצועים: 8  פונקציונאליות: 5  **משקל סופי: 5.6** | יתרונות: ביצועים טובים.  חסרונות: ממשק לא נוח ולא עשיר מספיק לעבודה. |
| **שרת אינטרנטי** | פשוט הפיתוח: 7  עלות: 2 (יקר)  ביצועים: 8  פונקציונאליות: 6  **משקל סופי: 6.25** | יתרונות: עלות אפסית, קהילה גדולה, ביצועים טובים.  חסרונות: ממשק דל. |
| **Python** | פשוט הפיתוח: 10  עלות: 10(חינם)  ביצועים: 6  פונקציונאליות: 9  **משקל סופי: 8.5** | יתרונות: יש ניסיון, קלה ונוחה מאד.  חסרונות: משמשת בעיקר להוכחת היתכנות ולאחר מכן ממשים באמצעות שפה יותר ייעודית. |

|  |  |  |
| --- | --- | --- |
| **Matlab** | פשוט הפיתוח: 10  עלות: 0(יקר מדי)  ביצועים: 9  פונקציונאליות: 9  **משקל סופי: 7.9** | יתרונות: יש ניסיון, קלה ונוחה מאד.  חסרונות: משמשת בעיקר להוכחת היתכנות ולאחר מכן ממשים באמצעות שפה יותר ייעודית. |

***טבלה 2 – חלופות לטכנולוגיית שפת פיתוח***

# 11. האמצעים/הכלים הנדרשים

כיוון שהמוצר הסופי שלנו הוא מערכת, נתייחס הן לצד הClient והן לצד הServer.

בצד **הClient** נשתמש בפלטפורמת ANGULAR מבית GOOGLE. פלטפורמה זו היא מהפופולריות בעולם מערכות הWEB, מכווינה את השוק בהיבטים ארכיטקטוניים וכן בהיבטים עיצוביים, תוך השמת דגש על מחקר מקיף וחווית משתמש מירבית, כיאה למוצרים המוכרים מבית Google. את הClient נפתח בעזרת Visual studio code מבית Microsoft בגרסת 1.30 ומעלה.

בצד **הServer** נקים API בסביבת NET. FRAMEWORK מבית Microsoft אשר תשמש לנו הן לצורך הבאת מידע למשתמש תוך ביצוע מניפולציות והתאמת המידע לתצוגה, והן לפניות לAPI'S חיצוניים כדוגמת הבורסה האמריקאית, אליה נרצה להתחבר על מנת לשלוח פקודות קנייה/מכירה, או על מנת לקבל את המידע שצברנו מהרשתות החברתיות ולבצע עליו שאילתות כדי לתת תובנה למשתמש. כלי הפיתוח שישמש אותנו כאן הוא Visual studio 2019, כאשר אנו נעבוד בגרסת .NET שהיא 4.5.1 ומעלה.

מהלכי הScraping, חילוץ טקסט, חילוץ יישויות וסמנטיקה ובניית המודלים יתבצעו בשפת Python בעזרת כלי הפיתוח PyCharm ובנוסף בעזרת ספריות כמו scikit-learn, keras, nltk המשמשות לחילוץ טקסט ולאימון מודלים. את התובנות שיוסקו מהמודל נשמור בDataBase על מנת שצד הServer יוכל לפנות לקבלת תובנות או לביצוע תובנות בעצמו למשתמש.

**כלל הכלים המוזכרים כאן ניתנים להורדה ברישיון ללא תשלום, מתוקף היותנו סטודנטים.**

פערי הידע החסרים לנו מפורטים בצורה מסודרת בסעיף 14 המוקצה לכך.

**Page Break**

# 12. תוצרי הפרויקט

* **פרסום ספר פרויקט.**
* **מערכת משולבת לסחר בשוק ההון.**
* **פיתוח אלגוריתם לאיסוף ועיבוד מידע, חילוץ ישויות והבנת סמנטיקה טקסטואלית ממקורות ממדיות חברתיות.**
* **פיתוח אלגוריתם המתחשב בתוצרי האלגוריתם הטקסטואלי, ובנוסף בניתוחים טכניים וסטטיים של שוק ההון.**

# 13.תכנית עבודה ראשונית לביצוע הפרויקט

# 14. פערים

פרויקט זה מכיל מס' עולמות תוכן, שכל מרכיב בו הוא עולם ומלואו בפני עצמו. הפרויקט מכיל בתוכו בניית מע' WEB, תוך שילוב טכנולוגיות חדשניות בעולם המע'. בתחום זה אבירם ידריך וילווה את חברי הצוות על מנת להשלים את פערי הידע עקב ניסיונו המקצועי בנושא זה.

בנוסף, פרויקט זה מכיל ביצוע SCRAPING על אתרים, ניתוח טקסט, הבנת סמנטיקה והקשרים, חילוץ ישויות, בניית ML ואימון מודלים. פערי ידע בנושאים אלו יושלמו בעזרת קורסים שאנו לומדים במכללה(בינה מלאכותית, למידת מכונה, התמחות AI שחגי מבצע וינחה אותנו על בסיסה) וכן בעזרת ניסיון מקצועי ורכישת קורסים מקוונים.

בהיבט עולם התוכן, שוק ההון, פערי הידע כיצד להתנהל בעולם זה, רכישת מושגים בסיסיים, הכוונה למקורות מידע רלוונטיים יתבצעו בעזרת עמית(המנחה) ויריב, שלהם ניסיון רב בעולם שוק ההון. בשלב זה איננו רואים פער בהיבט ציוד מבוקש החסר לנו לצורך התקדמות בפרויקט.

# 15. סיכונים עיקריים ודרכי התמודדות (בהיבטים טכניים ומנהלתיים בלבד)

סיכונים בפרויקט זה יכולים להתבטא במס' חזיתות, הן הטכנית והן המנהלתית.

בהיבט הטכני, קיים סיכון באינטגרציה וניגון כלל חלקי המע' והאלגוריתם בצורה מיטבית, שכן קיימים חלקים מרכזיים במע' הנשענים על תוצרים של חלקים אחרים. בנוסף קיים סיכון טכני בכך שבניית אלגוריתמים ושימוש בהם על מנת לחלץ מידע מטקסט ולהצליח לספק ממנו תובנות הוא מורכב במגבלת הזמן, שכן תחום זה כרגע הוא בפיתוח עולמי וחברות רבות לא הצליחו למטב מידע מטקסט לטובת הסקת תובנות לשוק ההון.

בהיבט המנהלתי, עבודה בצוות של 3 מפתחים טומנת בחובה קשיים בהסתנכרנות ועבודת הצוות עלולה להיתקל בקשיים.

# 16.רשימת מקורות (ראשונית)

1. <https://en.wikipedia.org/wiki/Stock_market_prediction>
2. <https://he.wikipedia.org/wiki/%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>
3. [https://www.bloomberg.com/company/press/bloomberg-launches-twitter-feed-optimized-trading](https://www.bloomberg.com/company/press/bloomberg-launches-twitter-feed-optimized-trading/)
4. <https://www.investopedia.com/articles/markets/031814/can-tweets-and-facebook-posts-predict-stock-behavior-and-rt-if-you-think-so.asp>
5. [**Stocktwits - The largest community for investors and traders**](https://stocktwits.com/) **-** See what actual investors and traders are saying in real time about the stocks, crypto, futures, and forex you care about for free.
6. [**https://www.bloomberg.com**](https://www.bloomberg.com/company/press/bloomberg-launches-twitter-feed-optimized-trading/)**-**  Bloomberg and Twitter expand relationship to provide clients with a feed of financially relevant Tweets, designed for algorithmic trading ([Bloomberg Launches a Twitter Feed Optimized for Trading | Press](https://www.bloomberg.com/company/press/bloomberg-launches-twitter-feed-optimized-trading/)**)**
7. [**https://www.socialmarketanalytics.com/#about-sma**](https://www.socialmarketanalytics.com/#about-sma)**-** A Leader in Unstructured Financial Data, SMA has built Intellectual Property in four Major FinTech Areas: Natural Language Processing, Textual Parsing, Topic Modeling, and Source Accurac
8. [NYSE: The New York Stock Exchange](https://www.nyse.com/index)-As the global leader in listings, New York Stock Exchange has been the venue of choice for innovators, visionaries and leaders for over 225 years. To help companies access capital and navigate global markets, NYSE offers a [unique market model](https://www.nyse.com/market-model), [unmatched network](https://www.nyse.com/network), [brand visibility](https://www.nyse.com/global-visibility) and [core services](https://www.nyse.com/nyse-services).
9. [Developing Trading Models Using Machine Learning on Financial News and Social Media Data](https://www.youtube.com/watch?v=Qe3Tx5Ct_B0) - סרטון טוב
10. [Machine Learning for Smarter Trading: 14 Companies You Should Know](https://builtin.com/artificial-intelligence/machine-learning-for-trading)
11. [**https://www.benzinga.com**](https://www.benzinga.com/)**-** Stock Market Quotes, Business News, Financial News, Trading Ideas, and Stock Research.
12. ***Stock Market Prediction based on Social Sentiments using Machine Learning*** *- By IEEE, published in* [*2018 International Conference on Smart City and Emerging Technology (ICSCET)*](https://ieeexplore.ieee.org/xpl/conhome/8521418/proceeding)*, on Jan 5, 2018.*
13. ***Semantic Aware Visual Abstraction of Large Scale Social Media Data With Geo-Tags****- By IEEE, published on August 15, 2019.*
14. ***Data Science in Economics -*** *By Amir Mosavi Prof, published on March 19, 2020.*
15. ***Deep Learning for Financial Applications A Survey-*** *By Department of Computer Engineering, TOBB University of Economics and Technology, Ankara, Turkey.**published on May 2, 2020.*