

### תקציר

דוח עסקי זה מציג את הפיתוח של כלי חדשני לקבלת החלטה והשקעה בשוק ההון מנתוני המדד S&P-500. כלי שנועד לעזור למשקיעים פרטיים ומנהלי השקעות בקבלת החלטות מושכלות, תוך שילוב של ניתוח טכני ונתוני חדשות כלכליות בזמן אמת. מטרת הדוח היא להציג את התועלות העסקיות של הכלי בתנאי שוק דינאמי של היום.



## דוח הבנה עסקית

השקעות בסיכון נמוך זה אצלנו-NexTrade

מחברים – אריאל קריב ומשי בר

## תוכן עניינים

1. רקע- ..... 1
- 1.1 קביעת יעדים עסקיים ..... 1
- 2.1 רקע עסקי ..... 2
2. יעדים עסקיים וקריטריונים להצלחה ..... 3
- 1.2 יעדים עסקיים ..... 3
- 2.2 קריטריוני הצלחה ..... 4
- קריטריונים אובייקטיביים וקריטריונים סובייקטיביים: ..... 4
3. הערכת מצב לפרויקט חיזוי סיכון השקעה ..... 5
4. מלאי משאבים ..... 6
5. דרישות, הנחות ואילוצים ..... 7
6. סיכונים ומקרי קצה ..... 7
7. מינוח ..... 8
8. עלויות ותועלות ..... 9
9. יעדי מדעי הנתונים וקריטריונים להצלחה ..... 10
- י. תוכנית פרויקט ..... 11
10. הערכה ראשונית של כלים וטכניקות ..... 11

## 1. רקע-

### 1.1 קביעת יעדים עסקיים

היעד העיקרי בפרויקט הוא להשתמש בנתוני המדד S&P-500 בשילוב עם ניתוח טכני של המדד וחדשות כלכליות, על מנת לסייע בקבלת החלטות השקעה מדויקות יותר. לדוגמא- מתי כדאי לקנות או למכור. הכוונה היא לנצל נתוני עבר ותובנות כדי לנסות לחזות את הסיכון בהשקעות במדד זה ולהציע כלים שיעזרו למשתמשים, כגון מנהלי השקעות, אנליסטים פיננסיים וגם אנשים פרטיים בכדי לשפר את ניהול הסיכונים שלהם.

#### ● מטרות:

- יצירת כלי ניתוח שייתן חיווי ברור לגבי כדאיות השקעה או מכירה של המדד, בהתבסס על נתונים כלכליים וניתוח טכני.
- פיתוח מודל שמנבא את רמת הסיכון של השקעות לפי תנאי שוק וחדשות עדכניות.

#### ● בעיות:

- ניהול סיכונים בתנאי שוק משתנים.
- צורך ביכולת ניתוח מתקדמת יותר שמבוססת על נתונים רחבי טווח וחדשות עדכניות כדי לשפר תחזיות השקעה.

#### ● משאבים:

- שימוש ב-Alpha Vantage API לנתוני העבר של מדד ה-S&P500 בתוספת של הניתוח הטכני של המדד ונתוני חדשות כלכליות.
- משאבים חומריים: מחשבים עם סביבה תומכת ב-Machine Learning.
- כוח אדם: צוות קטן (שני אנשים) המנהל את הפרויקט.

## 2.1 רקע עסקי

סוחרים פרטים וארגונים הזקוקים לכלי מתקדם שיאפשר ניהול סיכונים בשוק ההון באמצעות חיזוי תובנות השקעה בטווחים קצרים ובינוניים במדד ה-S&P 500.

### תרשימים ארגוניים ומשאבים:

- האנשים הפרטיים הם אלו שעתידיים להשתמש בפתרון זה בעקבות הרצון לפתרון המהיר (מבחינה רגולטורית הכלי יהיה לשימוש אישי בלבד).

### זיהוי בעיות ומטרות:

- הבעיה המרכזית היא קושי בקבלת החלטות השקעה בתנאי שוק תנודתיים.  
- מטרת הפרויקט היא לאפשר שימוש בנתונים טכניים וכלכליים כדי לשפר את ניהול הסיכון ואת יכולת החיזוי של תנודות המדד S&P-500.

### אזור הבעיה

- אזור הבעיה: הבעיה המרכזית היא בתחום ניהול השקעות וניהול סיכונים.  
- תיאור הבעיה: חוסר ודאות לגבי החלטות השקעה במדד ה-S&P 500 בשל תנודתיות בשווקים והשפעות של חדשות כלכליות ובנוסף חוסר יודע וניסיון של האדם שרוצה לסחור במדד.

### - תנאים מוקדמים:

- האדם הפרטי משתמש בנתונים פיננסיים אך לא בפתרון מבוסס Data Science מתקדם לחיזוי סיכונים.

### פתרון נוכחי

- תיאור הפתרון הנוכחי: פתרונות הניתוח מבוססים על ניתוחי נתונים ידניים ואנליזות בסיסיות של נתוני שוק עבר.

- יתרונות: - הגישה הנוכחית פשוטה לשימוש ואינה דורשת משאבי מחשוב כבדים.

### - חסרונות:

- חסרה בה יכולת לחזות מגמות בהתבסס על תנאים משתנים.  
- היכולת לנצל נתונים כלכליים עדכניים וניתוח טכני מתקדם בזמן אמת (יש עיכוב של חצי דקה).

## 2. יעדים עסקיים וקריטריונים להצלחה

### 1.2. יעדים עסקיים

מטרת הפרויקט היא לסייע למשקיעים פרטיים ולמנהלי השקעות בקבלת החלטות מושכלות יותר במדד ה-S&P 500 על ידי זיהוי הסיכון בהשקעה על פי תנודות השוק וחדשות כלכליות בזמן אמת. הפרויקט יהווה כלי המאפשר הבנה פשוטה יותר של השוק והפחתת ההסתברות לקבלת החלטות השקעה לא אופטימליות.

### משימות לפרויקט:

**- הגדרת הבעיה:** קבלת החלטות השקעה במדד ה-S&P 500 בתנאי שוק משתנים היא אתגר בפני משקיעים, ובמיוחד בפני משקיעים פרטיים אשר אינם מבינים בכלל את אופי השוק ואת הפרמטרים המשפיעים עליו כמו אירועים כלכליים וחדשות עדכניות על המדד. חוסר היכולת לחזות שינויים בזמן אמת מגביל את ניהול הסיכונים.

### - שאלות עסקיות:

1. כיצד אפשר להעריך את רמת הסיכון הנוכחית בהשקעה ב-S&P 500?
2. באילו מצבים כדאי למשקיע לשקול השקעה או מכירה בהתבסס על המדד הנוכחי ותחזיות עתידיות?
3. אילו חדשות כלכליות צפויות להשפיע באופן חזק על המדד, וכיצד נוכל לכלול אותן במודל החיזוי?

### - דרישות עסקיות:

- לפתח כלי ניתוח יעיל שייתן חיזוי על רמת הסיכון בהשקעה.
- להבטיח שהכלי מתאים לשימוש על ידי אנליסטים ומנהלי השקעות ונותן מענה בזמן אמת.
- לעקוב אחר נתונים חדשים ועדכניים כל העת, תוך שמירה על דיוק בניתוחים.

### **- יתרונות צפויים:**

- שיפור קבלת החלטות בשוק ה-S&P 500, תוך הפחתת הסיכון על ידי שימוש בנתונים עדכניים.

- כלי שיציע משוב על הזדמנויות השקעה והמלצות לסיכון, אשר יסייע בהקטנת השקעות מסוכנות ובהפחתת הפסדים.

### **2.2. קריטריוני הצלחה**

כדי להבטיח הצלחה של הפרויקט, נשתמש בקריטריונים להצלחה אובייקטיביים וסובייקטיביים:

#### **קריטריונים אובייקטיביים וקריטריונים סובייקטיביים:**

- **דיוק מודל:** המודל צריך להגיע לרמת דיוק של לפחות 85% בתחזיות ההשקעה והסיכון.

- **עדכון נתונים בזמן אמת:** הפתרון חייב לכלול חדשות כלכליות בזמן אמת עם עדכון שוטף לנתונים הפיננסיים.

- **זמינות המידע:** להבטיח שתוצאות הניתוח יהיו זמינות לשימוש המשתמשים תוך 30 שניות.

- **שיפור סיכונים:** הפחתה של לפחות 10% בהפסדים שנגרמים עקב החלטות שגויות על השקעות במדד ה-S&P 500.

- בכך, קריטריוני הצלחה יקבעו את מידת השגת היעדים העסקיים ויבטיחו שהפתרון ייתן ערך מוסף למנהלי ההשקעות ולמשתמשים בפרויקט.

### 3. הערכת מצב לפרויקט חיזוי סיכון השקעה

#### • סוגי נתונים זמינים לניתוח

- נתוני שוק: אנו נשתמש במפתח API של Alpha Vantage לקבלת נתונים על מדד ה-S&P 500. הנתונים כוללים נתונים היסטוריים, עדכונים בזמן אמת, ומדדים טכניים כגון- ממוצעים נעים, RSI, ו-MACD לצורך הניתוח.
- נתוני חדשות כלכליות: שירות ה-API מספק גם גישה לחדשות כלכליות היסטוריות וחדשות שוטפות לדוגמא: ממד מחירי הצרכן החודשי, מה שמסייע בניתוח השפעתן על המדד.
- פורמטים ואחסון: הנתונים זמינים בפורמט JSON וניתן לעבדם באמצעות פלטפורמות כמו – Python. האחסון נעשה בצורה מאובטחת בענן לצורך עיבוד וניתוח יעיל לאורך זמן.

#### • כוח האדם הדרוש כדי להשלים את הפרויקט

- יכולות צוות: צוות הפרויקט מורכב משני אנשים בעלי ידע בתחום מדעי הנתונים, ניתוח טכני, ידע בתחום שוק המניות ופיתוח ב-Python שיתוף הפעולה מתאפשר על ידי כלי ניהול גרסאות GitHub. - מה שמבטיח עבודה מסונכרנת ומעקב אחר שינויים.
- תמיכה נוספת: במידת הצורך, נשקול להתייעץ עם מומחים פיננסיים או אנשי דאטה מדעיים על מנת לשפר את ניתוח הנתונים או ליישם שיטות מתקדמות יותר.

#### • גורמי הסיכון הגדולים ביותר המעורבים

- מגבלת בקשות API: מאחר ו-Alpha Vantage מספקת API חינמי עם הגבלות על מספר הבקשות או הגבלת תקשורתית, יש סיכון של חוסר נתונים ברגעים קריטיים או עיכובים בגישה.
- תנודתיות שוק: תנודתיות גבוהה, חדשות כלל עולמיות מתפרצות עלולים לפגוע בדיוק החיזויים.
- זמני תגובה איטיים: שגיאות בתזמון או בתגובה בזמן אמת עלולות להשפיע על התובנות המופקות ועל תחזיות הסיכון.



## החוג למערכות מידע

- בעיות תקשורת: ברגע של בעיית אינטרנט כזו או אחרת הנתונים לא יעודכנו וכתוצאה מכך המשתמש לא יקבל אינפורמציה בזמן אמת.

### • תוכנית מגירה לכל סיכון

- ניהול הגבלות API: ננהל את הבקשות בקפידה ונצמצם בקשות מיותרות, וכן נבדוק אפשרות ל- Upgrading, לאפשרות בתשלום אם נדרשת גישה נוספת.
- מעקב אחר יעילות המודל: במהלך אימון המכונה נבצע בדיקות תכופות למודל כדי לאתר בעיות בנתונים או בשיטות הניתוח.
- עדכון המשתמש: ברגע של תקלה באינטרנט נודיע למשתמש שהנתונים כרגע אינם מעודכנים.

## 4. מלאי משאבים

### משאבי חומרה

-חומרה נדרשת: מחשבים חזקים לעיבוד נתונים, עם חיבור רציף לאינטרנט ומעבדים מהירים כדי להאיץ את חישובי ה- Machine Learning.

### מקורות נתונים ומאגרי ידע

-מקורות נתונים זמינים: שימוש ב-API Alpha Vantage כמקור לנתוני S&P 500 ולחדשות כלכליות, נתוני העבר זמינים בפורמט JSON.

- אחסון הנתונים: הנתונים יאוחסנו במאגר SQL או ב-NoSQL, בהתאם לנפח, ויגובו בענן לצורך נגישות.

- מקורות נוספים: אין כרגע תכנון לרכוש נתונים נוספים, אך אפשר להוסיף בעתיד מאגרי נתונים דמוגרפיים או כלכליים משלימים לפי הצורך.

- בעיות אבטחה: אין מידע אישי ולכן אין צורך באבטחה מיוחדת.



## 5. דרישות, הנחות ואילוצים

### קביעת דרישות

- הגבלות אבטחה: יש לאחסן את הנתונים בצורה מאובטחת ולמנוע גישה בלתי מורשית לנתונים הרגישים.

- תזמון הפרויקט: כל שלב חייב להיות תואם ללוחות הזמנים שנקבעו מראש, כולל אבני דרך להערכת דיוק המודל והטמעתו.

### הבהרת הנחות

- גורמים כלכליים: ניתוח הפרויקט ידרוש הקצאת תקציב מוגבלת לייעוץ טכנולוגי והרחבת שירותי API (במידת הצורך).

- הנחות איכות נתונים: הנתונים מ-Alpha Vantage אמינים, אך נדרש לבצע ניקוי נתונים ותיקוף כדי להבטיח איכות גבוהה.

- ציפיות נותני החסות: הצפיות שלנו מהמודל הוא שהוא יצליח יותר מ-75%.

### אימות אילוצים

- גישה לנתונים: כל מפתחות ה-API והסיסמאות מאובטחים ונגישים.

- מגבלות משפטיות: אנו לא משתמשים בנתונים באופן מסחרי ולכן אין מגבלות.

- מגבלות תקציביות: כל העלויות הכרוכות בטכנולוגיה, אחסון ושירותי נתונים מתוקצבות ומאושרות.

## 6. סיכונים ומקרי קצה

### סיכונים אפשריים ותוכנית מגירה

#### 1. תזמון:

- סיכון: הפרויקט מתעכב עקב תקלות או מורכבויות בלתי צפויות.

- תוכנית מגירה: הגדרת לוחות זמנים גמישים והקצאת זמני רזרבה לניסויים עם המודל.

#### 2. פיננסי:

- סיכון: מגבלות תקציב שיכולות להגביל שימוש בנתונים או כלים נוספים.

- תוכנית מגירה: חיפוש חלופות חלופות חינמיות לנתונים ואופטימיזציה לקוד כדי להקטין צריכת משאבים.

### 3. נתונים:

- סיכון: בעיות באיכות הנתונים או כיסוי נתונים חלקי.
- תוכנית מגירה: ביצוע ניקוי נתונים ותיקוף, במקרה של בעיות קשות- חיפוש מקורות נתונים חלופיים.

### 4. תוצאות:

- סיכון: תוצאות המודל אינן מספקות את רמת הדיוק הנדרשת.
- תוכנית מגירה: התאמת המודל וביצוע כוונונים נוספים, שימוש במודלים מתקדמים יותר אם נדרש.

### 5. בניית המערכת:

- סיכון: הצוות לא מיומן ומכיר חלק מהסביבות עבודה וממבנה האלגוריתמים שהמערכת תבנה עליהם.
- תוכנית מגירה: שימוש בעזרת המנחה ולמידה עצמית של הצוות להבאת תוצאות מקסימליות.
- בכך נוכל להבטיח שהפרויקט מתנהל בהתאם למשאבים, לדרישות ולאילוצים הקיימים, ושכל הסיכונים הפוטנציאליים מקבלים מענה מתאים.

### 7. מינוח

- כדי להבטיח תקשורת חלקה בין צוותי מדעי הנתונים והצוות העסקי, יש ליצור מילון מונחים הכולל:

#### - מונחים טכניים:

- רגרסיה: שיטה סטטיסטית שמשמשת לחיזוי ערך מספרי אחד (למשל- מחיר בית) על סמך ערך של משתנים אחרים (למשל- גודל הבית, מיקומו).
- תחזית: ניסיון לחזות את העתיד על סמך נתונים היסטוריים ומודלים סטטיסטיים או אלגוריתמים של למידה חישובית.
- סיווג: תהליך של הקצאת תצפיות לקטגוריות או כיתות שונות (למשל- סיווג תמונות לחיות או נופים).
- דשבורד: ממשק משתמש גרפי המציג נתונים ואינדיקטורים בצורה ויזואלית, בדרך כלל במטרה לספק תמונה כללית ומובנת של מצב מסוים.

API (Application Programming Interface): ממשק תכנות יישומים, המאפשר ליישומים שונים לתקשר ביניהם ולהחליף מידע.

Python: שפת תכנות פופולרית מאוד בתחומי מדעי הנתונים והלמידה החישובית, הידועה בקלות השימוש שלה ובספריות העשירות שלה.

### **- מונחים עסקיים:**

מדד ה-S&P 500: מדד מניות שמורכב מ-500 החברות הגדולות ביותר בשוק המניות האמריקאי, ומשמש כמדד רחב לשוק המניות האמריקאי כולו.

ניתוח טכני: שיטה לניתוח שוק ההון שמבוססת על ניתוח של נתונים היסטוריים של מחירים ונפחים, במטרה לחזות את התנהגות המחירים בעתיד.

תחזית סיכון: הערכה של הסיכון הכרוך בהשקעה מסוימת, כמו למשל הסיכון לאובדן הון.

מגמה כלכלית: כיוון כללי של שוק או כלכלה, כמו למשל מגמת עלייה או ירידה.

כיוון השוק: מציין אם השוק הכללי נוטה לעלות, לרדת או להישאר יציב.

דיוק תחזיתי: הוא מדד שמראה לנו כמה קרובה התחזית שלנו למציאות בפועל. במילים אחרות, זה כמו ניחוש משוכלל: אנחנו מנסים לחזות מה יקרה בעתיד, והדיוק התחזיתי אומר לנו כמה "טוב" הניחוש הזה שלנו.

דישדוש: מחירי המניות עולים ויורדים בצורה מתונה אך לא פורצים לרמות חדשות של עליות (שיאים) או ירידות (שפל).

### **8. עלויות ותועלות**

#### **עלויות משוערות:**

- איסוף נתונים: עלות רכישת מפתחות API מ-Alpha Vantage אם ידרשו יותר קריאות, וכן תוספות אפשריות למידע נוסף.

- פריסת תוצאות: הבחירה תהיה לפי צורך המשתמש שיוכל לבחור טווח של בין שבוע לשלושה חודשים, ושם נציג לו תחזית מגמה באחוזים.

- עלויות תפעול: תחזוקה של הנתונים והעדכונים השוטפים.

### תועלות צפויות:

- מטרה עיקרית: יצירת כלי לניהול סיכון והשקעות המבוסס על תחזיות שוק עדכניות על מדד S&P 500.
- תובנות נוספות: הבנה טובה יותר של השפעת חדשות כלכליות על השוק בזמן אמת.
- יתרונות אפשריים: הפחתת סיכונים והשגת יתרון תחרותי באמצעות תובנות מבוססות מדע נתונים.

### 9. יעדי מדעי הנתונים וקריטריונים להצלחה

#### יעדי מדעי הנתונים

- סוג הבעיה: גרסיה לתחזיות מחיר/תשואה עתידיות, סיווג רמות סיכון (למשל- גבוה, בינוני, נמוך).
- יעדים טכניים: היעד הטכני המרכזי של המערכת הוא פיתוח מודל מתקדם המסוגל לנבא את כיוון השוק (עלייה או ירידה) ברמת דיוק של לפחות 75%. בנוסף, המודל יוכל להעריך את רמת הדיוק התחזיתי שלו, אשר צפויה לנו בטווח של 80%-85%, תוך שמירה על עקביות ואמינות בניתוח.
- תוצאות רצויות: שיפור הדיוק של תחזיות המדד במונחי המגמה שיש לו (עליה, ירידה או דישדוש).

#### קריטריוני הצלחה במדעי הנתונים

- שיטות להערכת מודל: שימוש במדדי דיוק כמו RMSE (Root Mean Square Error) ו- MAE (Mean Absolute Error) במודלים הרגרסיביים ולאחר מכן נעשה ממוצע בין שניהם.
- אמות מידה להצלחה: היעד הטכני המרכזי של המערכת הוא פיתוח מודל מתקדם המסוגל לנבא את כיוון השוק (עלייה או ירידה) ברמת דיוק של לפחות 75%. בנוסף, המודל יוכל להעריך את רמת הדיוק התחזיתי שלו, אשר צפויה לנו בטווח של 80%-85%, תוך שמירה על עקביות ואמינות בניתוח.
- פריסת תוצאות: חלק מהצלחה נמדדת בהטמעת הדשבורד, שיעניק תובנות ברורות ומועילות למקבלי ההחלטות.

## 10. תוכנית פרויקט

כדי לייעל את תהליך הפרויקט, יש לוודא שתוכנית הפרויקט כוללת:

- תיאום עם כל המעורבים: כל בעלי העניין צריכים לאשר את התוכנית ואת כל שלבי הביצוע.

- הערכת זמן: לכל משימה יש לוח זמנים מותאם.

- נקודות החלטה ובדיקות: סימון שלבים כמו בניית מודל, תיקוף המודל, ובדיקות מערכת (בשלב אימון המודל).

## 11. הערכה ראשונית של כלים וטכניקות

### כלים:

- Python: בחירה ב-Python בשל התמיכה הרחבה שלה במדעי הנתונים ובספריות ל-ML כמו TensorFlow, scikit-learn, Pandas.

- ספריות ויישומים: power-bi או Streamlit לדשבורדים, NoSQL או-SQL לאחסון נתונים.

### טכניקות:

- מודלים: אנו נשתמש בשני סוגי מודלים ונאחד אותם בסוף בנייתם. המודל הראשון הוא LSTM (long Short Term memory), מודל מתאים במיוחד לנתונים רציפים כמו מחירי מניות, נפחי מסחר ושינויים באחוזים. הוא מסוגל לזהות דפוסים מתמשכים, כמו מגמות עולה או יורדת, ולחזות תנועות עתידיות בהתבסס על התנהגות העבר. בנוסף המודל יוכל לנתח את השינויים במדד S&P 500 לאורך זמן (למשל- תקופות זמן משתנות) ולחזות את כיוון השוק (אחוז עלייה/ירידה) לתקופה הבאה.

המודל השני הוא BERT (Bidirectional Encoder Representations from Transformers), BERT יכול לעבד מאמרים וכתובות חדשות ולזהות אם הטון הוא חיובי, שלילי או ניטרלי.

בנוסף המודל יכול להבין מושגים כלכליים מורכבים, לזהות גורמים משפיעים (כגון "רווחי חברות", "עלויות ריבית") ולהעריך את השפעתם האפשרית על המדד. לבסוף BERT ישמש לניתוח הכותרות והמאמרים הכלכליים (שמגיעים ממקורות Alpha Vantage) ולהקצות ציון רגשי לחדשות. הציון ישולב עם נתוני המדד שנאספו

## החוג למערכות מידע

כדי לשפר את התחזיות. התוצאות משני המודלים ישולבו במודל משוקלל (למשל- באמצעות רגרסיה או רשת נוירונים מתקדמת) כדי לתת תחזית כוללת.

לדוגמה : אם ה-LSTM מצביע על מגמה עולה, אבל ה-BERT מזהה חדשות שליליות משמעותיות, התחזית עשויה להיות ניטרלית במקום חיובית.