| **מדעים דיגיטליים להיי-טק** | | |
| --- | --- | --- |
| פרויקט מס' \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  ***תכנית עבודה (מוקדמת)*** | | |
| שם הפרויקט: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | | |
| מבצעים: | | |
|  | שם: \_גלעד שפיר \_\_\_\_\_ | ת.ז. \_\_319092235\_\_\_\_ |
|  | שם: \_רועי ניר \_\_\_\_\_ | ת.ז. \_\_208847459\_\_\_\_ |
|  | | |
|  |  |  |
| מקום ביצוע הפרויקט: \_\_אוני' תל אביב, עבודה מרחוק\_\_\_\_  ***לשימוש המנחה:***  *הנני מאשר את תכנית העבודה המצורפת*  שם: \_\_\_דרור יעקובי\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  חתימה:\_\_\_ | | |
|  | | |

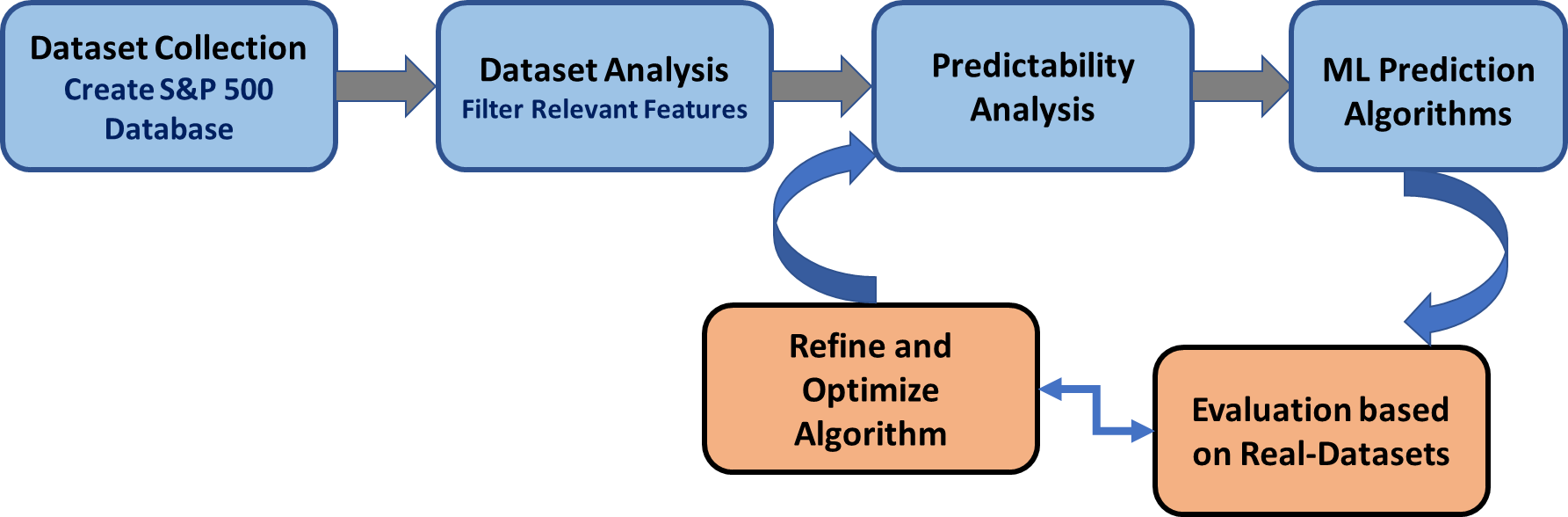
# תקציר

**מטרה:** מטרתו העיקרית של הפרויקט שלנו היא חיזוי מניות בשווקים פיננסיים. בניגוד לניתוח מניות רגיל העוסק בעיקר בניתוח גרפים, קריאת דוחות כספיים או עדכונים חדשותיים, אנו מתמקדים בבסיסי הנתונים המתארים את התנהגות המניות בנושאים כגון מחיר פתיחה, ווליום מסחר וכו'.

**ביצוע:** הבנו שישנם גורמים רבים המשפיעים על המנייה מלבד הנתונים השטחיים שלה. נרצה לבדוק את המאפיינים והנתונים הרלוונטיים שיכולים לעזור לחיזוי המניה. לכן בשלב הראשון, נרצה לאגד מידע הרלוונטי לשווקים של מניות שונות, ולהבין מהו אותו מידע רלוונטי. כדי לבצע חיזוי כזה עלינו להסתכל על המדד שאליו המנייה שייכת. לכן, בשלב המקדים נרצה למצוא אלגוריתם שיסנן ויבחר את הדאטה והמאפיינים הרלוונטים של מניות. נתמקד במדדים וסקטורים אליהם שייכות המניות שבחרנו לחזות. **המדד** משקף את רמת המחירים הממוצעת של קבוצות מסוימות או ענפים מסוימים של מניות הנסחרות בבורסה לניירות ערך. המדד מחושב מתוך מחירי המניות, בדרך כלל על ידי חישוב ערכם המשוקלל. המדד משמש משקיעים הן להערכת רמת המחירים של ענף או קבוצה מסוימים, והן כבסיס להערכת ביצועים למדידת ביצועים של תיקי השקעות להשוואה של ביצועים אלה בטווחי זמן שונים. **הסקטור** הוא מגזר מניות מצומצם יותר המתייחס לקבוצת חברות הפועלות באותו ענף או עסק. לעתים קרובות משקיעים מקצים את השקעותיהם על פני מגזרים שונים כדי לגוון את תיקי ההשקעות שלהם ולהפחית סיכונים. בשלב הבא נרצה לבצע **חיזוי** לערך המניה, ונתרכז באי הוודאות של המודל ביחס למניה. בנוסף, נרצה למצוא את המדד ושיטה למדידת אי הוודאות. **שיטות:** נשתמש באלגורתמים של ML עדכניים של מודלים סטטיסטיים לניתוח סדרות בזמן.

**דאטה:** בפרויקט שלנו בחרנו להתמקד במדד s&p 500: זהו מדד ניירות ערך הכולל מניות של 500 תאגידים, רובם אמריקאיים. הסיבה שבחרנו במדד זה היא המבנה שלו. אלו הן מניות של חברות ציבוריות גדולות הנסחרות בבורסות אמריקאיות ממגוון רחב של סקטורים, דבר שבעינינו יכול לייצג בצורה טובה מאוד את המצב הכללי במשק ולהיות שימושי בעת חיזויים עתידיים על מדדים או מניות שונים.

**דיאגרמת בלוקים של הפרויקט**.:



**שלב ראשון:** איסוף וניתוח המידע.

1. **איסוף דאטה:** מציאת מקורות רלוונטים, איסוף דאטה היסטורי לאימון מודלים, לאגד הכל בדאטה בייס.
2. **ניתוח דאטה:** בחירת המניות והפיטצ'רים הרלוונטים לחיזוי כל מניה\מדד. שלב שזה יוכל להתבצע בדרך של שילוב ידע בתחום ואלגוריתמים לניתוח סדרות בזמן.

**שלב שני:** פיתוח אלגוריתמים.

1. **ניתוח אי-וודאות:** אי הוודאות בחיזוי יאפשר לנו להבין מהם ביצועי המודל האפשריים ויענה על שאלות בסיסיות הרלוונטיות לעולם המניות, למשל, מאיזה טווח כדאי לחזות את המניה.
2. **שימוש באלגוריתם:** חיזוי הכולל בחירת מודל מתאים וניתוח התוצאות.
3. **ניתוח תוצאות** על ידי שיטות סטטיסטיות.

# מוטיבציה

לפיתוח מודל לחיזוי מניות ב- S&P 500 העוסק בחוסר ודאות יכולה להיות מוטיבציה גבוהה מכיוון:

* השווקים הפיננסיים הם מטבעם בלתי צפויים, ופיתוח מודל שיוכל להסביר טוב יותר את אי הוודאות יכול לתת למשקיעים יתרון בקבלת החלטות השקעה מושכלות ורווחיות יותר.
* חיזוי של תנועות ה- S&P 500 יכול להיות בעל השלכות פיננסיות משמעותיות, הן עבור משקיעים בודדים והן עבור הכלכלה הרחבה יותר. על ידי פיתוח מודל חיזוי מדויק יותר, אתה יכול לעזור להפחית סיכונים ולמקסם תשואה.
* פיתוח מודל חיזוי המסביר את אי הוודאות בחישוב ערך המנייה בפועל דורש הבנה עמוקה של סטטיסטיקה, תורת ההסתברות וטכניקות למידת מכונה, מה שהופך אותו לבעיה מאתגרת ומגרה אינטלקטואלית לעבוד עליה. הסיפוק מהתמודדות מוצלחת עם בעיה מורכבת יכול להוות מניע רב עוצמה בפני עצמו.
* אי ודאות גבוהה בשווקים הפיננסיים עלולה ליצור סיכונים משמעותיים למשקיעים, להוביל לתנודתיות בשוק ולהפסדים גדולים. על ידי פיתוח מודל חיזוי המסביר את אי הוודאות, ניתן להפחית משמעותית סיכונים כאלו ולשפר את היציבות הכוללת של השווקים הפיננסיים.
* פיתוח מודל חיזוי שיכול לחזות אירועים בוודאות נמוכה\גבוהה ולספק מדד לכך יכול להיות בעל ערך מיוחד משום שלאירועים אלה יש לרוב השפעות משמעותיות על השווקים הפיננסיים. על ידי פיתוח מודל שיכול להסביר טוב יותר את האירועים הללו, אתה יכול לעזור למשקיעים לקבל החלטות מושכלות יותר והימנעות מהפסדים גדולים. פיתוח מודל חיזוי שיכול לחזות במדויק אירועים בוודאות גבוהה יכול להיות גם בעל ערך, מכיוון שהוא יכול לספק למשקיעים איתות ברור לקנות או למכור. על ידי זיהוי מוקדם של אירועים בוודאות גבוהה, משקיעים יכולים ליהנות ממגמות שוק ולהגדיל את התשואות שלהם.
* מוטיבציה נוספת יכולה להיות השגת שוויון "בתנאי המגרש", כלומר, אפשרות לתת כוח נוסף למשקיעים קטנים לעומת תאגידי ענק המעסיקים מקצוענים בתחום ובעלי השפעה רבה על המחירים עקב כוח השוק שלהם. אלגוריתם חזק שידע לנבא היטב את ערכי המניות יכול לאפשר גם למשקיעים פרטיים לגרוף חלק מהרווח וליצור שוויון הזדמנויות גדול יותר.

הפרויקט מצריך פיתוח ויישום של טכנולוגיות חדשות ומתקדמות, כמו למשל אלגוריתמי למידה עמוקה וטכנולוגיות חיזוי חדשות, ובכך מעודד חדשנות ופיתוח טכנולוגי. חלופות למימוש הפרויקט עשויות לכלול שימוש בתוכנות חיזוי קיימות, או פיתוח מודלים חיזוי באמצעות שיטות פענוח נתונים סטטיסטיות. החידוש שאנחנו מביאים נכנס לתחום הסברה בלמידת מכונה (ML) שהיא אתגר חשוב ומתמשך, במיוחד כאשר יותר ויותר החלטות קריטיות מתקבלות על ידי מערכות אוטומטיות. כדי להבטיח שמערכות אלו יהיו הוגנות, חסרות פניות ואמינות, חיוני להיות מסוגל להסביר כיצד הן מגיעות להחלטותיהן.

הדבר מצריך פיתוח טכניקות ומסגרות לפירוש והסבר ההתנהגות של מודלים מורכבים של למידת מכונה, והבטחה שניתן להבין אותם ולבחון אותם על ידי בני אדם. חלופה אפשרית למימוש שלנו היא שימוש במודלים של Explainable AI. אנחנו נתמקד בשלב המקדים במודל בניתוח הגורמים שמביאים לקבלת החלטות במודלים שונים (למשל הגדרת הקלט והפלט).

# תכולת עבודה

במסגרת הפרויקט, הסטודנטים ילמדו על האלגוריתמים הסטטיסטיים העכשוויים ב ML המתרכזת בחיזוי\סיווג וניתוח סדרות בזמן. כדי לעשות זאת, הם יעיינו במקורות רלוונטיים ויבצעו ניתוח מעמיק של האלגוריתמים השונים וטכניקות הניתוח הסטטיסטיות השונות בעזרת כלים שונים. העבודה תבוצע ב-פייתון, Jupyter Notebook, ותאפשר לסטודנטים להכיר ספריות שימושיות לניתוח נתונים ופיתוח אלגוריתמי ML כמו - Pytorch/sklearn/Pandas.

בפרויקט נדרש להשתמש בתשתיות מדעיות, מתמטיות, הנדסיות ואלגוריתמיות הקיימות בשוק, על מנת לממש את הפרויקט ולהבין את המודלים הקיימים בשוק. יש לכך את התמיכה של מקורות רלוונטיים כמו קודים פתוחים ומאמרים שמטרתם להרחיב את הידע וההבנה של הסטודנטים בתחום. בסופו של דבר, הפרויקט מטרתו להעמיד את הסטודנטים ביכולת לפתח ולממש מודלים של למידת מכונה וניתוח סדרות בזמן, ולהבין את האלגוריתמים והטכניקות הסטטיסטיות הנדרשות לפיתוח פתרונות מתקדמים בתחום זה. הפרויקט יסייע לסטודנטים להבין את התהליך המכני של למידת מכונה וניתוח נתונים, ולחשוף אותם לכלים וטכניקות שונים שיסייעו להם להשתמש בצורה יעילה במודלים של למידת מכונה וניתוח סדרות בזמן. בנוסף, הפרויקט יחזק את יכולתם להתמודד עם הקושי שבפיתוח פתרונות לבעיות של אי ודאות עם נתונים מורכבים וסדרות זמן.

# תוצרי הפרויקט

בחינת הפרויקט תהיה כרוכה בדרך כלל בהצגה של הסטודנט בפני ועדת הבוחנים. המצגת צריכה לספק סקירה ברורה של הפרויקט, כולל המטרות, המתודולוגיה והתוצאות. התוצאות יפורטו דוח מפורט המתאר את המתודולוגיה ותוצאות הפרויקט.

בסיום הפרויקט, נכוון להצגה, שימוש באלגוריתמים ושימוש בקוד ותיעוד של הפרויקט שיכולים לשמש חוקרים או מפתחים אחרים. כלי ויזואליזציה או ממשקים אינטראקטיביים המקלים על חקר וניתוח נתונים. עבודת מחקר המציגה את ממצאי הפרויקט ומשמעותם בהקשר לתחום הרלוונטי. הניתוחים כוללים

* ניתוח נתונים חקרני: זה כולל שימוש בשיטות גרפיות ותיאוריות כדי לסכם את המאפיינים העיקריים של הנתונים וזיהוי דפוסים או חריגות כלשהן.
* בדיקת השערות: זה כולל ניסוח של השערה לגבי הנתונים ושימוש במבחנים סטטיסטיים כדי לקבוע אם ההשערה נתמכת בנתונים. זה כולל שימוש במודלים של רגרסיה כדי להעריך את הקשר בין משתנה תלוי למשתנה בלתי תלוי אחד או יותר.
* ניתוח סדרות זמן: זה כרוך בשימוש במודלים סטטיסטיים כדי לנתח ולחזות נתוני סדרות זמן, אשר נתקלים לעתים קרובות ביישומים כגון תחזיות פיננסיות, חיזוי מזג אוויר וחיזוי תנועה.

# לוח זמנים

בסעיף זה יש צורך לתאר את רשימת אבני הדרך של הפרויקט, עם תיאור מילולי קצר (2-3 שורות) של כל אבן דרך ותאריך יעד. יש לכלול בין 8-12 אבני דרך (הוסיפו שורות כנדרש), בתוכן:

* תאריך יעד להגשת דו"ח מעקב
* תאריך יעד להגשת הפוסטר
* תאריך יעד לסיום הפרויקט

| **אבן דרך** | **פירוט (2-3 שורות)** | **תאריך יעד לביצוע** | **הערות** |
| --- | --- | --- | --- |
| תוכנית עבודה |  | 20/4/2023 | פגישה הבאה עם המנחה |
|  |  |  |  |
| איסוף וניקוי הדאטה | איסוף וטעינה של כל הדאטה  משימות נוספות: פתיחת גיט לפרויקט + שמירת כל קטעי הקוד | 10/5/2023 |  |
| ניתוח הדאטה | פרה-פרוססינג על הדאטה, ניתוח ניקוי הדאטה והכנתו לעבודה עם מודלים | 20/5/2023 | שלב זה כולל למידה של כלים סטטיסטיים לניתוח סדרות בזמן |
| שלב מקדים: מחקר | מחקר על מודלים קיימים.  הגדרת הבעיה: אי-וודאות ומשימה של חיזוי בציר הזמן ואי וודאות של המודל (ערך סף) | 10/6/2023 | שלב זה כולל למידה ומחקר של מודלים של חיזוי |
| שלב ראשון | חיזוי מחירי מניות בודדות באמצעות מודלים של למידת מכונה | 20/6/2023 |  |
| הגשת דו"ח מעקב |  | 1/7/2023 | פגישה הבאה עם המנחה |
| ניתוח ובדיקת תוצאות | ניתוח ביצוע המודדים בהתאם למדדים השונים, השוואת תוצאות. הגדרות כלים להשוואה סטטיסטית. | 1/8/2023 | שלב זה כולל מחקר ולמידה של ניתוח סטטיסטי |
| שלב שני | חיזוי ערכים של סקטורים (מקבצי מניות) | 1/8/2023 | שלב זה כולל בתוכו את איסוף וניקוי הנתונים הרלוונטיים |
| שלב שלישי (אם יהיה זמן) | חיזוי "מדדים" | 1/9/2023 | בדיקת השפעות של סקטורים אחד על השני וחיזוי ערך המדד שבחרנו |
| הגשת קבצי הפרויקט + דו״ח מסכם (ספר הפרויקט) |  | 24/8/2023 | הגשה לבדיקה פנימית, שבוע לפני הצגת הפרויקט |
| הגשת פוסטר |  | 24/8/2023 | הגשה שבוע לפני הצגת הפרויקט |
| סיום הפרויקט – מפגש הצגת פרויקט | הצגת פרויקט לבדיקה חיצונית | 1/9/2023 | בחינה חיצונית |