**פרויקט גמר במערכות מידע**

**חיזוי מניות**

**מנחה: מר יהונתן איתי סגמן**

**מוסד אקדמי: המרכז האקדמי רופין**

**חוג: כלכלה ומנהל עסקים בהתמחות מערכות מידע**

**תאריך: 10.7.2022**

**מגישות:**

**דיאנה מדבדובסקי 316080456**

**אלה בן דוד 206821332**

**איב נחום 313235020**

תוכן

[**תקציר** 3](#_Toc107234839)

[**מבוא** 4](#_Toc107234840)

[**שיטות** 11](#_Toc107234841)

[**הדאטה מהו וכיצד נאסף:** 11](#_Toc107234842)

[**האלגוריתמים והשיטות הסטטיסטיות:** 12](#_Toc107234843)

[**תוצאות** 15](#_Toc107234844)

[**סטטיסטיקה תיאורית והסקה סטטיסטית:** 15](#_Toc107234845)

[**תוצאות הרצת האלגוריתמים ו-evaluation:** 17](#_Toc107234846)

[**מסקנות ודיון** 20](#_Toc107234847)

[**מסקנות ותובנות:** 20](#_Toc107234848)

[**מה למדתי מהעבודה?** 21](#_Toc107234849)

[**יתרונות וחסרונות וכיצד היה ניתן להתגבר עליהם:** 22](#_Toc107234850)

[**מבט לעתיד:** 22](#_Toc107234851)

[**ביבליוגרפיה:** 23](#_Toc107234852)

[**נספחים:** 24](#_Toc107234855)

# **תקציר**

עבודה אקדמית זו מכילה את אחד הנושאים המרתקים והמעניינים בשוק הפיננסי בישראל ובעולם, שכן לאורך השנים משקיעים וחוקרים רבים מחפשים את התשובה לשאלה - כיצד ניתן לחזות ולהעריך את התשואה על נייר ערך מסוים. מתוך שאלה כללית זו בחרנו לחקור ולנתח את השפעת הדוחות הכספיים המפורסמים כמידע האמין והמדויק ביותר שקיים על כל חברה ציבורית, ובדרך זו לבחון את אופן השפעתם על תשואתה של מניית החברה. במחקרנו הסתמכנו על טפסי 10K של חברות ציבוריות המונפקות בבורסה האמריקאית במטרה לחזות באופן המדויק ביותר האם תהיה עלייה או ירידה בערך המניה וכן לחזות את ערך השינוי השנתי הצפוי בערכה בהסתמך על הנתונים שבידנו. במהלך העבודה נתקלנו בקשיים רבים בהכנה אופטימלית של הנתונים וניקויים עבור הרצת המודלים. בנתונים היו בין היתר ערכים חסרים, ערכי אפס מרובים, כמות גדולה של פיצ'רים, קורלציה בין פיצ'רים ומספר רב של ערכי קיצון אשר פוגעים בדיוק המודלים. במסגרת הפרויקט ביצענו הרצת אלגוריתמים מתקדמים של למידת מכונה. בין היתר רגרסיה לוגיסטית, KNN , עצי החלטה, יערות רנדומיים ורגרסיה ליניארית. כל זאת במטרה למצוא את המודל המדויק ביותר לחיזוי ביצועי מניות. זאת ועוד ביצענו אופטימיזציה לכל אחד מהמודלים בדרך הנכונה עבורו. כחלק מתהליך האופטימיזציה של מודל היערות הרנדומיים אשר נמצא בהמשך כיעיל ביותר מצאנו ערך סף אופטימלי (**נספח טו**), דירגנו את המשתנים לפי מידת תרומתם למודל (feature importance) וכן ביצענו אופטימיזציה במודל באמצעות מציאת השילוב האופטימלי של מס' העצים ועומק העץ המרבי אשר נותנים את הטעות הקטנה ביותר.   
על אופן התמודדותנו עם הקשיים המפורטים לעיל והשימוש הנרחב באלגוריתמים נפרט בהמשך העבודה בפרק השיטות. במחקרנו עבדנו עם שלוש מחברות colab בכל אחת מהמחברות יצאו תוצאות קצת שונות. התוצאות הטובות ביותר התקבלו במודל יערות רנדומיים לפני הורדת הפיצ'רים על בסיס Feature selection בו הצלחנו לנבא ב-62%. כמו כן, תוצאות הרגרסיה הלינארית טובות ביותר התקבלו במודל הפולינומי. בתחילת הפרויקט הגדרנו הצלחה אם נצליח לנבא ביצועי מניה ביותר מ-70%. פירוט התוצאות ומשמעותן בהמשך העבודה. המסקנה העיקרית אשר עלתה מעבודתנו היא שקשה מאוד לחזות ביצועי מניות בהסתמך על נתונים פיננסיים בלבד. נמליץ לעשות מחקרי המשך בתחום מרתק זה ע"י מודלים מתוחכמים יותר ואופטימיזציה גבוהה יותר.

# **מבוא**

הדוחות הכספיים הינם השתקפות לנעשה בארגון העסקי, הם מהווים את חומר הגלם העיקרי לניתוח ואבחון מצבה הכלכלי של החברה ולקביעת סיכויי התפתחותה או הישרדותה בעתיד - מכאן ניכרת השפעתם של הדוחות הכספיים על תשואת המניה. בעבודתנו נתייחס ל - 2 שאלות מחקר, כאשר המרכזית שבהן תיענה באמצעות קלסיפיקציה ואילו השנייה תענה ע"י שימוש ברגרסיה לינארית.  
1 .בהינתן נתוני טפסי 10K לשנה מסוימת של חברה האם ניתן לחזות אם ערך המניה שלה

יעלה או ירד בשנה העוקבת על בסיס אינדיקטורים פיננסיים בלבד.

2 .בהינתן נתוני טפסי 10K לשנה מסוימת של חברה האם ניתן לחזות את שיעור השינוי בערך

המניה בשנה העוקבת על בסיס אינדיקטורים פיננסיים בלבד.  
על בסיס הנתונים שבידנו ושאלות המחקר שהעלנו, ביצענו סקירה נרחבת על התחום הנחקר בה גילינו כי מחקרים רבים עסקו סביב השאלה האם ניתן לחזות את מחיר המניה ואם כן כיצד כדאי לסייע למשקיעים, ולמקבלי החלטות בביצוע נכון והחלטות השקעות מושכלות. במשך שנים ניסו חוקרים וכלכלנים להתחקות אחר המקור להתנהגותה של מניה נסחרת בשוק ולנסות למצוא מהו המפתח לחיזוי התנהגותה בעתיד של המניה. יש 2 גישות נפוצות ומקובלות לחיזוי מחיר המניות: האחת, גישה טכנית המבוססת על ניתוח טכני חוקרת את התנהגות השווקים בהתבסס על גרפים למטרת הבנת המגמה השולטת בשוק , השיטה מתבססת על חקר פסיכולוגיית המשקיעים וניסיון לצפות את התנהגותם על פי תקדימי העבר, ובכך לצפות גם את מחיר המניה. הגישה השנייה היא שיטת הניתוח הפונדמנטלי מסתמכת ככלל על לימוד מצבה הכלכלי של חברה מסוימת ,על ביצועיה הכספיים וחוזקה הפיננסי. סוחר המשתמש בניתוח פונדמנטלי בדרך כלל יהיה מעודכן תמיד לגבי זמני הפרסום של האינדיקאטורים הכלכליים, יעקוב בזמן אמת אחרי הודעות כלכליות ואחרי כל ידיעה כלכלית שמופצת אודות החברה שהמשקיע מעוניין לרכוש. ה-“תמונה הגדולה” של הניתוח הפונדמנטלי טובה לחיזוי מגמות ארוכות-טווח במניות, אך היא אינה נותנת מידע מספיק לגבי השקעה לטווח הקצר בינוני שם האווירה בשווקים והפסיכולוגיה משפיעה באופן ישיר על ביצועי המניות. על פי גישת הניתוח הפונדמנטלי, בטווח הקצר השווקים עשויים לתמחר נייר ערך באופן שגוי, אבל - בטווח הארוך, השוק יתמחר את נייר הערך במחיר ה"נכון". על פי שיטה זו ניתן ליצור רווח על ידי רכישת נייר ערך אשר מתומחר באופן "שגוי" ולחכות שהשוק יזהה את ה"טעות" ויבצע תמחור מחודש לנייר הערך. לסיכום, ההבדל העיקרי בין שתי גישות הינו כי בעוד האחת מסתמכת על דפוסי העבר החוזרים ונשנים לשיטתה, השנייה מתבססת על מידע עדכני מההווה המצוי כיום, הן במישור האישי של החברה ומצבה והן בהיבט של המעגלים החיצונים הסובבים את החברה, המגזר העסקי הספציפי, המשק המקומי כולו וכלכלת העולם (Fama, 1965). פאמה הינו כלכלן אמריקאי ופרופסור למימון אשר הגה את תאוריית השוק היעיל (Efficient Market Theory), על פי תיאוריה זו, שוק הון יעיל משקף בכל רגע נתון את מלוא האינפורמציה הידועה על נייר ערך, כולל שקלול מלא של תחזיות והערכות המומחים לגביו. כלומר, מחירו של נייר ערך משקף בכל רגע את ערכו הכלכלי המלא. כל הודעה חדשה של החברה, או הערכה אנליטית חדשה שמתפרסמת בציבור, תגרור תיקון מידי במחיר נייר הערך, כך שישקף גם את האינפורמציה החדשה כלומר לא ניתן להכות את השוק על ידי איתור מידע חדש או ניתוח מחדש של חברה או מניה מה שבעצם שומט את הקרקע מתחת לרגלי חסידי הניתוח הפונדמנטלי. אחת הנגזרות של תיאוריית השוק היעיל היא תאוריית ההליכה האקראית (Random Walk Theory). על פי תיאוריה זו התנהגות המניה בעבר אינה יכולה לסייע בחיזוי התנהגותה בעתיד. תאוריית ההליכה האקראית שומטת את הקרקע לשיטת הניתוח הטכני המנסה לנבא את כיוונן של מניות בהתאם למגמות התנועה שלהן מהעבר. הנושא שנוי במחלוקת ומעורר ויכוח תמידי בעיקר עם בתי השקעות שטוענים שכן ניתן להשיג תשואות עודפות בעזרת ניתוח אנליטי מקצועי. בעבודתנו אנו נתמקד בשיטת ניתוח פונדמנטלי מהסיבה שהנתונים מורכבים מאינדיקטורים פיננסים וננסה לגבש דעה לגבי הנושא השנוי במחלוקת על ידי העמקת הידע בתחום בעזרת מאמרים אקדמאיים ומחקרים קודמים שנעשו ושימוש באלגוריתמים עליהם נפרט בסקירה ובפרק השיטה. העקרון הבסיסי להערכת נכסים קובע שוויו של נכס הוא הערך הנוכחי של זרמי המזומן המוקנים לבעל הנכס בתוקף בעלותו. עקרון זה חל על נכסים פיננסים (מניות, אג"ח). בהקשר למניות זרמי המזומן הרלוונטיים הם הדיבידנדים העתידיים להם זכאי מחזיק המניה. אם ידועים הדיבידנדים בוודאות הרי שמתקבלת נוסחת הערכה פשוטה: V0=∑t=0 D(t)/(1+i)t כאשר V0 מייצג את שווי הנוכחי של המניות, D(t) דיבידנד למניה המשולם בזמן T , I שער הריבית חסרת סיכון בו יכול המשקיע להלוות וללוות. הקושי בהערכת מניות לעומת אגרות חוב או תוכניות השקעה כרוך בצורך לחזות את זרם הדיבידנד העתידי קושי זה הוא כפול כי ראשית קיימת אי וודאות לגבי מידת הצלחת הפרויקטים בהם פועלת החברה ואשר מהכנסותיהם ישולם הדיבידנד. שנית, האינפורמציה העומדת לרשות המשקיע פחותה מהאינפורמציה הקיימת בפירמה ותלויה בפרוצדורת דיווח בלתי סדירה ובלתי עדכנית. מקור האינפורמציה החשוב הוא הדוחות הכספיים (אמיר ברנע ,1980). במחקר יושמו מודלים שונים על מנת לחקור את הנושא, המאמר הראשון עסק במודל KNN (K-Nearest Neighbor) ומודלים לא ליניאריים שבאמצעותם ניסו לחזות את ביצועי המניות הירדניות. למטרה זו נלקח מדגם של 5 חברות גדולות הנסחרות בבורסה הירדנית שנבחרו באקראי, נתוני המדגם נלקחו מהתקופה שבין 4 ביוני 2009 ל-24 בדצמבר 2009,לכל אחת מהחברות היו כ200 רשומות (Khalid & Hassan & Ismail & Mohammed,2013). מחקרים עסקיים אחרונים התרכזו בתחומים של תחזיות עתידיות של תנועות מחירי המניות. חוקרים, וקהילות עסקיות מעוניינים להניח כי התרחשות עתידית תלויה בנתונים בהווה ובעבר(Kim,2003). הם להוטים לזהות את תחזית מחיר המניה של תנועות בשוקי המניות ,עם זאת כפי שהוזכר עוד במבוא, נתונים פיננסיים נחשבים כנתונים מורכבים לחיזוי, כפי שמוסבר בהשערת השוק היעיל (EMH) שהועלה על ידי (FAMA,1965). תיאוריה זו מגשרת על הפער בין המידע הפיננסי לבין שוק המניות, הוא טוען שהפער שנוצר הוא אך ורק בגלל כניסה של מידע חדש ושבעצם השוק תמיד נמצא בשיווי משקל כך שאי אפשר לנצל את השוק לצורך השגת תשואה מקניה ומכירה. בנוסף לכך מחקרים אישרו את התיאוריה הנגזרת מתיאורית EMH, שמחירי המניות אינם רודפים אחר הליכה אקראית (Random walk) ושתחום המניות צריך להמשיך להיחקר. מניה לא מאופיינת רק במחיר שלה, אלא גם בפיצ'רים נוספים כמו מחיר הסגירה המייצג את המשתנה החשוב ביותר לניבוי מחיר מניה ספציפית ע"פ המחקר הנוכחי. במחקר השתמשו בטכנולוגיית כריית נתונים שעוזרת בניתוח נפח גדול של נתונים עסקיים ופיננסיים, והיא מיושמת על מנת לקבוע תנועות של מניות. לשם כך לקחו נתוני מניות קודמים, ונתונים על פעילות החברות ואת מצב השוק. שילוב גישות סיווג כריית נתונים בחיזוי מניות מניב ערך עתידי עבור ערכי מניות של חברות על סמך נתונים היסטוריים. חיזוי זה משתמש בשיטות שונות של גישות סיווג כגון רשתות עצביות, רגרסיה, אלגוריתם גנטי, אינדוקציה(שיטת לימוד מן הפרט אל הכלל) של עץ החלטות ו-KNN .בנוסף לKNN השתמשו בחישובים מתמטיים ובמודלים ויזואליים. החישובים כללו הערכת שגיאות , סכום כולל של שגיאות בריבוע (ESS),שגיאה ממוצעת(MSE),וRoot Mean Square errors (RMSE) . על פי התוצאות אלגוריתם KNN היה מהימן עם יחס שגיאה קטן(RMSE) עבור כל המניות החברות שנבדקו כאשר השגיאה המקסימלית הייתה 0.037, כך שהתוצאות היו רציונליות גם כאשר התוצאות הוצגו בצורה ויזואלית ב Plot graph , כלומר בהתאם לנתוני מחירי המניות בפועל תוצאות התחזית היו קרובות. ניתן להבין מכך ששימוש בטכניקות כריית נתונים מציעות לעולם הפיננסים ניתוח חיזוי שימושי של תנועות שוק המניות (Khalid et. al ,2013). עם העמקת המחקר בנושא טכניקות למידת מכונה גילו שהם מציגים יכולות רחבות העוזרות לדמות בעיות מורכבות יותר בניגוד לתוכנית הרגרסיה הליניארית המסורתית המשמשת בדרך כלל לתיאור הקשר בין התשואה העתידית של המניה למאפייני החברה. תחום הפיננסים חווה התפתחות מהירה של אלגוריתמים מבוססי עצים ופרדיגמות רשת עצבית אשר מצליחים להמחיש את ההתנהגות המורכבת של המניות. שיטות לא ליניאריות אלה הוכיחו את עצמן כיעילות בחיזוי מחירי המניות ובבחירת מניות שיכולות לגבור על השוק הכללי (Zheng & Ziqin & Guangwei,2019). במחקר הנוכחי ניסו לחזות תשואה עודפת למניות בשוק הסיני בתקופת המסחר בין 8 בפברואר 2013 ל8 באוגוסט 2017 תוך שימוש במודל היער האקראי (RF), עבור כל מניה נלקחה תת-תקופת המסחר מכיל 252 ימי מסחר (כשנה בקירוב) שעליו התאמן סט האימון .במחקר נבדקו 2 דרכי הניתוח , הטכני(עבור טווח קצר) והפונדמנטלי (עבור טווח ארוך) על מנת לנסות לחזות את מגמת המחירים בטווח הארוך והקצר. הוכח ש2 השיטות הובילו לתשואות עודפות יוצאות דופן ב5 שנים האחרונות, כאשר יחסי שארפ מחושבים להיות 2.75 ו5 עבור הערך הנקי של תיק ההשקעות (מדד שארפ הוא מדד לביצועים של תיקי השקעות, הוא מודד את עודף התשואה ליחידת סיכון על נכסי השקעה. ככל שהתשואה גבוהה יותר ביחס לסיכון כך המדד גבוה יותר. המדד פותח על ידי הכלכלן האמריקאי וחתן פרס נובל לכלכלה ויליאם שארפ ב-1966). הוכח כי באמצעות מודלים כמו רשתות עצביות מלאכותיות ,עצי החלטה (DT) , יערות אקראיים (RF) וכולי הסיווג ויעילות החיזוי של מניות משופרים באופן משמעותי. יישום עצי החלטה במודל על ידי אימון סט שהורכב ממשתנים פיננסים וטכניים כאחד השיגו תשואה שנתית של 25.55% ומדד שארפ של 1.59 באסטרטגיית בחירת המניות השבועית שלהן , על ידי כך גברו על השוק הכללי. נמצא גם כי יעילות סיווג המניות הוכח כטוב יותר במודל עץ החלטה מאשר אלגוריתם לינארי בתיקים מגוונים בעלי סיכון ובהמחשת קשרים בסדר גבוה יותר בין תשואות מניות למשתנים פיננסים שכן גם הספרות מצביעה על כך שהראשון מבצע ביצועים טובים יותר מהאחרון בחיזוי נתוני השוק הפיננסי (Leung et al., 2000; Enke and Thawornwong, 2005). הממצאים האמפיריים היו מבטיחים, מה שמצביע על כך שהזדמנות רווח קיימת בטווח הקצר ניתנת לניצול באמצעות למידת מכונה, גם במקרה של שוק בוגר. פונדמנטליסטים סבורים כי מחיר המניה יתכנס בטווח הארוך בהתבסס על מאפייני החברה הבסיסיים שלה, בעוד אנליסטים טכניים נוטים לבצע תחזיות על מגמות המניה באמצעות אינדיקטורים מחיר ונפח, כפי שהם טוענים כי כל המידע כבר משתקף במחיר של נייר ערך. ידוע כי תשואות המניות קשורות למאפייני החברה במגוון דרכים. בספרות רבות פירשו תשואות מניה במונחים של מודל ליניארי פשוט או רב-משתני, שבו התשואה הצפויה ככל הנראה מושפעת על ידי יחס ספר למחיר(BP) (Brennan et a, 1997), שווי שוק (Market Cap) (Fama and French, 1993),רווחים ליחס מחיר (Basu, 1983), רווחיות, השקעה (Fama and French, 2015), לכן לניתוח פונדמנטלי נלקחו המשתנים הנ"ל ואילו לניתוח הטכני נלקחו משתנים כמו לקחו הממוצע הנע של מחיר המניות בתקופות מסוימות , מומנטום של המניה, משתנה שמתייחס לתנודתיות מניה במהלך M ימי המסחר האחרונים שחושבה לפי סטיית התקן של התשואות היומיות, ומשתנה שמתייחס לסטיית התקן של סדרות זמן של נפח מסחר במהלך M ימי המסחר האחרונים. מתוך המחקר הנוכחי של (Khalid et. al,2013) ניתן לראות כי ההשפעה של מספר העצים היא משמעותית. התיק חושף רווחיות יוצאת דופן בתקופת התנודות בשוק .יחס שארפ יכול להגיע ל- 2.75 כאשר מספר העץ מוגדר כ- 60. עוד מצוין כי מספרי עצים גדולים יותר מובילים לדיוק גבוה יותר עבור האימון בדגימה, אך בשום אופן אינם מרמזים על ביצועי אסטרטגיה טובים יותר מחוץ למדגם. נמצא גם כי סיווג פחות מדויק של מניות כאשר מספר המחלקות גדל. תיק ההשקעות שמקורו במרחב התכונות הפונדמנטלי/טכני מציג ביצועים יוצאי דופן לאחר עלויות העסקה, , נחשפת צמיחה מתמדת ברווח בחמש השנים האחרונות של המחקר. לפיכך, ניתן לסכם כי לפי המתודולוגיה בה השתמשו היא אמינה וכי לאלגוריתם למידת המכונה שנלקח עבור המודל יש יכולת טובה של הכללה בעת יצירת סיווגי מניות הגיוניים מבחינה כלכלית. מחקר אחר ניסה לבנות מודל מסוג רגרסיה לוגיסטית לצורך חיזוי ביצועי מניות הרשומות בבורסה באפגניסטן Syed & Muhammad & Irfan & Adnan,2018)). נלקחו יחסים פיננסיים וחשבונאיים שונים כמשתנים בלתי תלויים וביצועי מניות (הוגדרו כ"טובים" או "ירודים") כמשתנה תלוי. התוצאה מראה כי יחסים פיננסיים וחשבונאיים מנבאים באופן משמעותי את ביצועי המניה. המחקר מורכב מתקופת המדגם של נתונים שנתיים מ-2011-2015 וכולל 109 חברות לא פיננסיות רשומות בבורסה הנ"ל. המדגם נבחר על סמך נתונים זמינים של שווי שוק. המשתנים שנלקחו היו צמיחת מכירות, יחס חוב להון, יחס ספר למחיר, רווח למניה, תשואה להון ויחס נוכחי לצורך חיזוי ביצועי המניה. הממצאים מצביעים על כך שהתחזית שלנו הייתה מדויקת ב-89.77 אחוזים עבור חיזוי טוב וגם רע של ביצועים רעים של המניה. אמנם לא שקלו משתנים מאקרו-כלכליים כדי לחזות את ביצועי התשואות של המניה, אבל ששת היחסים החשבונאיים והפיננסיים הספציפיים של החברות היו טובים מספיק כדי לחזות את ביצועי המניות. הtresh hold שנקבע במחקרם הוא 0.5 . התוצאות סיפקו את הראיות לכך שרווח למניה (earning per share) ,יחס ספר מחירים (price book ratio),תשואה על ההון העצמי (return on equity),היחס הנוכחי (current ratio), צמיחת המכירות (sales growth) והחוב להון עצמי (debt to equity) משמשים כמשתנים המצליחים לחזות את ביצועי החברה כטובים או גרועים , התוצאה הסופית מנבאת ש89.7% חברות ביצועיהם ירדו ו87.2% חברות ביצועיהם יעלו Syed et. al,2018)). מחקר זה מוסיף לתרומה למחקר חיזוי מניות על ידי שימוש במודל הרגרסיה הלוגיסטית של Altman (1968) and Ohlson (1980) תוך שימוש בטכניקת ML, על מנת להגדיל את ביצועי המודל שולבו יחסים פיננסים וחשבונאיים שהומלצו על ידי המודלים של פאמה (Fama & French,1988 ,2017). רוב המחקרים הקודמים מראים ששיטות לחיזויים סטטיסטים בלתי מוצלחות. השיטה הרגילה של OLS הייתה נקודת האמצע לכל השיטה המסורתית. מחקרים קודמים הראו שרק לעיתים רחוקות הוכיחו הצלחה של חיזוי רק בשל קיומו של רעש ואי לינאריות בנתונים בעבר , כיום השיטות הלא לינאריות מציגות שיטות מתקדמות לחיזוי מניות. ממצא זה יכול לחזק את ההחלטה שלנו לא לבחור ברגרסיה לינארית או רגרסיה לוגיסטית ולבחור בעצם במודל מתקדם יותר לביצועים טובים יותר. חשוב להבין שההתעסקות הבלתי פוסקת בניסיון להרכיב מודל שיחזה באופן טוב את ביצועי המניות נובע מהתועלת הגדולה למשקיעים ובעלי מניות ,הוא יאפשר להם לקבל החלטות שקולות בעת השקעה בשוק ההון, בייחוד בתקופות זמן שהבורסה לא יציבה. אחת התקופות האלו הייתה בשנים 2006-2007 שמדינות שבהן הייתה כלכלה מפותחת חוו קריסות מקומיות ובינלאומיות שהובילו להתרסקויות בלתי צפויות ברחבי העולם (קריסת הבורסה) .מבחינה היסטורית ועד היום לא ניתן למצוא מודל שלם ,מדויק ומקיף לחיזוי ביצועי שוק המניות. כפי שכבר ניתן להבין ממאמרים קודמים מתחילת הסקירה כיום היחסים הפיננסים נמצאים בשימוש נרחב בבדיקה חיונית כדי לחזות ביצועי מניה החדש הוא שכיוון שזה היה תהליך שכיח עם הזמן פותחו יחסים חדשים כמו ערך ספר, יחס רווחי מחירים, יחס מחיר תזרימי מזומנים (Mubin et al., 2014), כמובן חשוב לציין שמשמעות ערכם של היחסים הפיננסים והמשתנים החשבונאים שונה בין ענף לענף שכן לכל סקטור הערכים המאפיינים אותו. Altman 1968)) שנחשב לחלוץ בתחום זה (פיתח את מדד אלטמן) , לקח סט של 5 יחסים פיננסים על מנת לחזות ביצועי חברות ולבדוק את חדלות הפירעון שלהם, יחסים אלו הם X1 = יחס ההון חוזר נטו לסך הנכסים. X2 = יחס יתרת העודפים לסך הנכסים. X3 = יחס הרווח נקי לפני הוצאות מימון ומס לסך הנכסים. X4 = יחס ההון עצמי להון הזר. X5 = יחס סך ההכנסות השנתיות לסך הנכסים. בשיטה של ניתוח עם מודל רב משתני נמצא המודל מובהק ,דיוק המודל היה סביב 94% , בחינת תנועות היחס באופן אינדיבידואלי לפני פשיטת רגל אימתה את תוצאות המודל ופירטה כי ניתן לחזות בהצלחה את חדלות הפירעון במשך שנתיים עוד לפני הכישלון בפועל , אם כי הדיוק דועך עם השנים שלאחר מכן. מלבד היותם מנבאים אמיתיים של כשלים תאגידיים ניתן להשתמש ביחסים הפיננסים גם לסיווג תעשיות לפי רמת סיכון Fama and French 1992)). הציעו כי ברחבי ארצות הברית מספר מחקרים הבחינו בקשר קיים בין תשואות מניות ומשתנים פיננסים. משתנים פיננסיים אלה מורכבים על בסיס ניתוח של יחסים פיננסיים שונים כמו תשואת רווחים, תשואת תזרים מזומנים, ערך ספר לשוק וגודל החברה והם הוכיחו את עצמם כאינדיקטורים משמעותיים בקביעת ביצועי המניות ונבדקו כמי שמפעילים השפעה מהותית על חיזוי תשואות המניות. מחקרים מבוססי אירופה זיהו גם ממצאים דומים. לעומת זאת (Öğütk et al,2009) הבין כי שיטות כריית נתונים כגון רשתות עצביות מלאכותיות מתאימות יותר לזהות מניפולציה של מחיר המניה בהשוואה לניתוח נתונים רב משתנית למשל מודל רגרסיה לוגיסטית, הסיבה לכך היא שטכניקות כריית נתונים מבצעות סיווגים טובים ומדויקים יותר ולא טכניקות מרובות משתנים. במחקרם של במחקרם של (Pavan & Balakesavareddy & Anand Kumar,2022) ניסו לבצע חיזוי בעזרת מודל AI כמו יערות רנדומיים וSVM. שיטת היערות הרנדומיים היא אסטרטגיית למידה קבוצתית והיא שיטה יעילה ביותר החוזרת על עצמה. תהליכים אלו מנוצלים על מנת לקבוע אם עלות המניה תהיה גבוהה יותר מעלותם ביום נתון כדי לבצע אסטרטגיות מסחר רווחיות, המטרה העיקרית של המחקר דומה מאוד לשאר המאמרים ששויכו לסקירת הספרות בעבודתנו ,והיא בעצם ניבוי מחירי מניה בעזרת AI וכריית נתונים. היכולת להעריך את עלות מניה החזוי משמעותית למטרות שונות. כמו למשל ירידת סיכון בהשקעת מיזם על ידי מומחים הפיננסיים ותמיכה בזיהוי דלתות פתוחות למשקיעים המנסים להפיק תועלת על ידי השקעת משאבים ברשומות המניות. שיטת יערות רנדומיים היא אסטרטגיה שבאמצעותה בונים מספר רב של עצי החלטה בזמן האימון. עבור משימות סיווג, הפלט של היער האקראי הוא המחלקה שנבחרה על ידי רוב העצים. עבור משימות רגרסיה, החיזוי הממוצע או הממוצע של העצים הבודדים מוחזר. יערות החלטה רנדומיים מתקנים את ההרגל של עצי החלטה להתאים יתר על המידה למערכת האימון שלהם. היא כוללת אסטרטגיה המכונה צבירת Bootstrap שבו יוצרים סטים חדשים של אימון על ידי ביצוע דגימות רנדומליות מתוך סט הנתונים המקורי ולאחר מכן מתבצעת דגימה חדשה שוב ושוב כלומר בהינתן סט נתונים מקורי כל פעם נעשית דגימה של תת סט עם חזרות ואז מתוך סט אחד ייווצר שפע של סטים. אשר גורמת לביצוע ייצוג משופר על ידי צמצום הסטיות . במחקר זה השתמשו בשני מודלים משמעותיים של בינה מלאכותית, מודל יער אקראי SVM כדי לתת ניבוי מתאים על מידע שנאסף. בהתבסס על התוצאות שהתקבלו, טוענים כי שני המודלים הציגו ביצועים בולטים בחיזוי שוקי המניות. מודל היער האקראי משתמש במודל רב משתנים שהפיק דיוק של 81.6 אחוזים ולעומת זאת SVM הפיק בדיוק של 85.5 אחוזים. מודל היער האקראי עולה בביצועיו על מודל SVM תוך שימוש בערכת הנתונים שסופקה. ייתכנו רכיבים שונים שיכולים להשפיע על ביצוע התחזית של המודלים המשמשים לתחזיות שוק המניות שעבורם יש כדאיות לבצע מחקר המשך. כפי שכבר ניתן להבין מסקירת הספרות שלנו חיזוי תשואות שוק המניות הוא משימה מאתגרת בשל שינוי עקבי של ערכי המניות התלויים במספר פרמטרים היוצרים דפוסים מורכבים. מערך הנתונים ההיסטורי הזמין באתר האינטרנט של החברה מורכב רק מתכונות מעטות כמו מחיר מקסימלי, מחיר מינימלי, מחיר פתיחה, וכולי, אשר לבדן אינן מספיקות לצורך חיזוי משביע רצון. כדי להתמודד עם מגוון הנתונים נדרש לבנות מודל יעיל שיכול לזהות את הדפוסים הנסתרים ואת היחסים המורכבים במערך נתונים גדול זה. לצורך כך נוצרו משתנים חדשים באמצעות המשתנים הקיימים שגילמו תפקיד מכריע במונחים של שיפור דיוק המודלים בחיזוי מחיר הסגירה של יום המחרת של חברה מסוימת. במחקרם של (2020 Mehar & Deeksha & Vinay & Arun,) השתמשו בשתי טכניקות ANN ו- RF) Randomforest) לצורך חיזוי מחיר הסגירה של 5 חברות, הנתונים ההיסטוריים שלהם נלקחו מ Yahoo Finance ,הדאטה מכיל 10 שנים מתאריך 4/5/2009 עד תאריך 4/5/2019. החברות שנלקחו הן נייק , גולדמן זאקס, ג'ונסון ,פייזר וג'יי.פי מורגן צ'ייס ושות'. הנתונים מכילים מידע על המניה כגון מחיר פתיחה, מחיר סגירה, מחיר מקסימלי לאותו יום מסחר ומחיר מינימלי לאותו יום מסחר, ונפח מסחר. רוב העבודות הקודמות בתחום זה משתמשות באלגוריתמים קלאסיים כמו רגרסיה ליניארית , תורת ההליכה האקראית (Random Walk Theory) לחיזוי מחירי המניות כפי שהזכרנו עוד קודם בסקירה. מחקרים קודמים מראים כי ניתן לשפר את חיזוי שוק המניות באמצעות למידת מכונה אשר הוכיחו את עצמן משפרות את היעילות כ60%-86% בהשוואה לשיטות קודמות. טכניקות כגון מכונת וקטור תמיכה (SVM), יער אקראי (RF) וכולי. כמה טכניקות המבוססות על רשתות עצביות כגון רשת עצבית מלאכותית (ANN- Artificial Neural Network) הראו גם הן תוצאות מבטיחות. ANN מסוגלת למצוא תכונות נסתרות באמצעות תהליך למידה עצמית. ובנוסף מסוגלת למצוא את יחסי הקלט והפלט של מערך נתונים מורכב גדול מאוד. לפיכך, ANN מוכיחה את עצמה כבחירה טובה לחיזוי מחיר המניה עבור ארגון. יער אקראי (RF) לעומת זאת היא טכניקת למידת מכונה של אנסמבל. היא מסוגלת לבצע הן רגרסיה והן משימות סיווג. הרעיון הוא לשלב עצי החלטה מרובים על מנת לקבוע את התפוקה הסופית במקום להסתמך על עצי החלטה בודדים על מנת להפחית את השונות במודל. המשתנים החדשים שנוצרו הרכיבו סט אימון של כל עץ החלטה אשר בתורו קובע את ההחלטה בצמתים של העץ. הרעש בנתוני שוק המניות הוא בדרך כלל גבוה בגלל גודלו העצום ויכול לגרום לעצים לגדול בצורה שונה לחלוטין בהשוואה לצמיחה הצפויה. מטרתו למזער את טעות החיזוי על ידי התייחסות לניתוח שוק המניות כבעיית סיווג ועל סמך משתני אימון לחזות את מחיר הסגירה של מניה של יום למחרת עבור חברה מסוימת. הערכת האפקטיביות של המודלים נעשתה באמצעות מדדים כמו שורש השגיאה הריבועית הממוצעת ( RMSE - Root Mean Square Error), אחוז השגיאה הממוצעת בערך מוחלט (MAPE) ושגיאת הטיה ממוצעת (MBE - Mean Bias Error).

בסופו של דבר לפי המדדי הערכה שנלקחו נראה כי ANN הציג תוצאות חיזוי טובות יותר עבור מחירי המניות. התוצאות מראות כי הערכים הטובים ביותר המתקבלים על ידי מודל ANN נותנים RMSE (0.42), MAPE (0.77) ו- MBE (0.013). עבור עבודה עתידית, ניתן לפתח מודלים של למידה עמוקה אשר מתחשבים במאמרי חדשות פיננסיות יחד עם פרמטרים פיננסיים כגון מחיר סגירה, נפח נסחר, דוחות רווח והפסד וכו ', לקבלת תוצאות טובות יותר. מטרת עבודתנו היא להפחית את הסיכון בהשקעה של מניות ולהקל על חיזוי ביצועיהן. ההנחה הרווחת היא שרוב המשקיעים הינם שונאי סיכון אשר חוששים להשקיע במניות בעלות תשואה גבוה בגלל היותן מסוכנות. בעזרת המודלים שנבנה נוכל להפיג את מסך הערפל סביב הנושא הנ"ל ולהעלות את רמת הוודאות בהשקעות בניירות ערך מסוכנים.

# **שיטות**

## **הדאטה מהו וכיצד נאסף:**

נתונים עליהם נעבוד בפרויקט הינם מאתר KAGGLE 5 - קבצים מסוג CSV משנת 2014 עד שנת 2018. בכל אחד מטפסי האקסל יש מעל ל-250 עמודות שמייצגות את הפיצ'רים למודל שלנו ויותר מ-4000 רשומות שכל אחת מהן מציגה נתונים פיננסים של מניה אחרת. הנתונים מתבססים על טפסי 10-K של חברות הנסחרות בבורסה. טפסים אלה כוללים מאזן מפורט, דו״ח רווח והפסד ותיאור פעילות במהלך השנה. בנוסף מצורפים מדדים מחושבים על בסיס הדוחות כמו יחס חוב, גידול ברווחים וכד'. את הטפסים החברות המונפקות בבורסה האמריקאית מגישות לרשות ניירות ערך ברבעון הראשון של השנה העוקבת. בדאטה שלנו נוספו עמודות של חישוב גידול בערך המניה-עמודת PRICE VAR [%]. עמודה זו מפרטת את אחוזי שינוי המחיר של כל מניה לשנה, מיום המסחר הראשון בינואר ועד ליום המסחר האחרון בדצמבר של אותה שנה, והאם היא עלתה או ירדה. את הנתונים האלה בונה הדאטה לקח מממשק אינטרנטי שנקרא Financial Modeling Prep API זהו בעצם ממשק חינמי המספק מידע על דוחות היסטוריים כספיים ועל מחירי מניות בזמן אמת. לידע כללי נסחרות שם גם חברות ישראליות, כמו סלקום, דלק החזקות, כיל, אלביט וטבע. זאת ועוד, בנתונים שלנו ישנם מגזרים שונים כגון שירותי תקשורת, אנרגיה, שירותים פיננסיים, בריאות, תעשייה, נדל"ן, טכנולוגיה ושירותים שמאפשר לנו לבצע ניתוחים והשוואות לפי מגזר. את עמודה זו הפכנו למשתנה דמה (**נספח ג**). 2 עמודות מתוך כלל הנתונים הן מחרוזות: אחת מהן היא שם המניה והשנייה היא המגזר אליו היא משתייכת. שאר הנתונים הם מספריים כאשר 222 עמודות הינן FLOAT ו-2 INT. כמו כן, בכל אחד מהטפסים שקיבלנו יש מספר שונה של מניות (שורות) אך מספר זהה של עמודות (פיצ'רים). השוני בכמות המניות יכול לנבוע מכניסתן של חברות חדשות או מיציאתן ממסחר בבורסה. 2 עמודות חשובות שיש לשים לב אליהן אלו עמודת ה"Class" שמקבלת ערכי 0 או 1 היא מציגה האם במהלך השנה החולפת המניה עלתה או ירדה ביחס לסוף השנה שעברה. עמודה חשובה נוספת היא עמודת ה - "PRICE VAR [%]", אם ערך "PRICE VAR [%]" חיובי אז בעמודת הכיתה יסומן 1 ובמקרה שהוא שלילי יסומן 0. פרט ששמנו לב אליו במהלך העבודה על הנתונים הוא שישנם פערים יחסית גדולים בכל אחת מהשנים בין מניות שעלו לכאלו שירדו דבר אשר בהמשך עלול לפגוע בדיוק המודלים לכן איחדנו את הקבצים והגענו לנתונים יחסית מאוזנים (**נספח ב**) .בנוסף גילינו כי בכל רשומה בקובץ ישנם ערכים חסרים. בעמודות כמו "מחזור תפעולי" ו "מחזור המרה במזומן" גילינו כי קיים אחוז גבוה (99%) של ערכים ריקים. בנתונינו קיימים הרבה ערכי 0 במקומות לא הגיוניים כמו הוצאות מנהלה וכלליות או הוצאות מס הכנסה שכן לחברות הנסחרות בבורסה יש בוודאות הוצאות כאלו. ממצא זה גרם לנו לחקור לעומק את כל אחד מהפיצ'רים ולהבין באיזה מקרים ערך 0 נרשם במקום ערך חסר ובאיזה מקרים הוא הגיוני (דיבידנד, גידול ברווחים וכד'). כחלק מתהליך הכנת הדאטה עבור הרצת המודלים החלפנו את ערכי ה-0 הלא הגיוניים לערכי NULL ומילאנו את כל הערכים החסרים ע"י שימוש ב-KNN, מצאנו כי ע"י שכן אחד מילוי הערכים הינו אופטימלי (**נספח ד**). במחברת נוספת מילאנו ערכים חסרים באמצעות ערך הממוצע זאת על מנת לבדוק את ההבדלים. מילוי ערכים ע"י ממוצע מושפע מערכים קיצוניים ומהיר לביצוע לעומת KNN אשר לוקח זמן ומשאבים. שתי השיטות בעלות היגיון. רוב הדאטה שקיבלנו מורכב מנתונים שהם ערכי קיצון. ניסינו להרחיב את הIQR מ-1.5 ל-3 על מנת לתת טווח רחב יותר לדאטה שלא ייחשב כערך קיצון- הכמות לא קטנה בצורה משמעותית. לאחר מכן, בדקנו את ההתפלגות של כל אחד מהפיצ'רים. ביצענו מניפולציה מתמטית לפי השיטה הבאה: במקרה בו ההתפלגות הזדנבה לימין שינינו את הערך ע"י הפעלת לוג על 1 + הערך המקורי. במקרה אחר בו ההתפלגות הזדנבה לשמאל שינינו את הערך ע"י הפעלת לוג על 1 – הערך המקורי. כך שבשני המצבים התפלגות הפיצ'רים התקרבה יותר להתפלגות נורמלית. פיצ'רים בעלי התפלגות נורמלית לא עברו טרנספורמציה כלשהי. לאחר כל זאת עדיין נשארנו עם כמו גדולה של אאוטליירים ובחרנו לבסוף להמשיך איתם. הדאטה הסופי שלנו להרצת המודלים מורכב מ-48 פיצ'רים, 19,757 רשומות (מניות), 19,542 אאוטליירים, 10,790 מניות עולות, 8,967 מניות יורדות וללא ערכים ריקים.

## **האלגוריתמים והשיטות הסטטיסטיות:**

במהלך עבודתנו השתמשנו במגוון אלגוריתמים של למידת מכונה מפוקחת אשר עוסקת במצבים בהם אנו יכול לכמת או לתייג כל דוגמה, אותה למדנו במסגרת קורס מתקדם במדעי הנתונים. אלגוריתמים אשר עונים על בעיית קלסיפיקציה ענו על השאלה האם ניתן לחזות עלייה או ירידה בערך המניה, כלומר שאלה אשר התשובה עליה מתויגת 1 – ערך המניה יעלה, 0 – ערך המניה ירד. ענינו על שאלה זו באמצעות רגרסיה לוגיסטית, KNN, עצי החלטה ויערות רנדומיים.   
**רגרסיה לוגיסטית** הינה מודל קלסיפיקציה מפוקח בה ערך החיזוי אינו רציף אלא בדיד. כמו במקרה שלנו הכרעה בין שני מצבים עליית ערך או ירידת ערך. שמה של הרגרסיה הלוגיסטית אינו מקרי והוא נובע מצורת החישוב שלה. הפונקציה המתארת את היא הפונקציה הלוגיסטית או הסיגמואיד : 🡨

פונקציית העלות של הרגרסיה הלוגיסטית

הרגרסיה הלוגיסטית משתמשת באלגוריתם Gradient Descent למציאת הטטות שהם המקדמים של הפיצ'רים. היתרון בשימוש באלגוריתם הנ"ל הוא שניתן לבצע אותו בלי בעיה גם אם יש הרבה מאוד דוגמאות. יש לשים לב כי לא בוחרים α גדולה או קטנה מדי, פרמטר α קבוע הוא די טוב כאשר אנו מתקרבים למינימום המקומי ובזמן ריצה סביר. נוסחת ה Gradient Descent : , מטרתו היא למצוא את הפרמטרים המיטביים כך שיביאו את פונקציית העלות למינימום. זהו אלגוריתם חיפוש שהולך ומשתפר כל הזמן – מצמצם את הערך של פונקציית העלות. **Random forest** הינו אלגוריתם מתקדם של ML בו התחזית הסופית היא לפי החלטת הרוב. אלגוריתם זה מפחית את השגיאה בסט המבחן על ידי הפחתת שונות הניבוי וכל תת-קבוצה בו מחושבת הרבה יותר מהר מהקבוצה המקורית. לאלגוריתם זה מספר רב של יתרונות , הוא מתבסס על גישות שונות, מונע overfitting, טוב גם לרגרסיה וגם לקלסיפיקציה ומבוסס על יתרונות עצי ההחלטה. ניתן לשפר את ביצועי המודל ע"י בחירה אופטימלית של מספר עצים ועומק עץ מקסימלי. כברירת מחדל אלגוריתמים של קלסיפיקציה מסווגים את הדוגמא כשערך הסף הוא 0.5 כלומר, כאשר הציון שניתן לדוגמא הינו 0.6 היא תסווג 1 וכאשר הוא נמוך מ0.5 יסווג אוטומטית 0. ניתן לשנות ערך סף זה בהתאם לדרישותינו מהמודל. **KNN אלגוריתם למציאת שכנים**, אחד האלגוריתמים הפשוטים ביותר. אלגוריתם קלסיפיקציה מהיר שמבוסס על מרחק בין דוגמאות (מפוקח ודיסקרטי). הרעיון: לשנן את סט האימון ולאחר מכן לחזות את התיוג של דוגמה חדשה. הפרמטר החשוב ביותר הינו K, מספר השכנים לפיהם אנו מתייגים. בניגוד לאלגוריתמים אחרים של ML, אין השערה. ב-KNN, התחזית מבוססת אך ורק על סט האימון. האלגוריתם מקבל: סט אימון, ווקטור של איקסים, דוגמאות חדשות לתיוג ו-K שמייצג את מספר השכנים. בתהליך האלגוריתם מתייג את הדוגמאות ע"י שימוש בפונקציית המרחק/דמיון של הדוגמא החדשה משאר הדוגמאות בסט האימון ובוחרת את התיוג לפי כמות השכנים ע"פ כלל החלטת הרוב. הדיוק בדרך כלל משתפר עם יותר נתונים. KNN הוא גם אחד הפתרונות להתמודדות עם ערכים חסרים. **רגרסיה לינארית רבת משתנים** רגרסיה מציגה את היחסים בין ה-features (משתנים ב"ת) למשתנה המטרה – Y(המשתנה התלוי) ומשמשת לניבוי ערך של דוגמא חדשה. לאלגוריתם יתרונות רבים כגון יכולת למידול קשרים בין שני משתנים או יותר, שימוש בקורלציה ליניארית, והוא פשוט מאוד. ברגרסיה לינארית תמיד שווה 1. בדומה לרגרסיה לוגיסטית גם ברגרסיה רבת משתנים אפשר להשתמש באלגוריתם ה Gradient Descent. עלינו לוודא שהפרמטרים בעלי אותו קנה מידה (ע"י נרמול). דבר העשוי להביא לתוצאות מדויקות יותר, משפר את הביצועים ואת יכולת ההסבר של התוצאות. לכל תכונה (פיצ'ר) יש בטא משלה. רגרסיה ליניארית לא תמיד משתמשת במודלים ליניאריים, לעיתים מסוימים מודל מתואר בצורה טובה יותר כפונקציות פולינום. בחרנו לעבוד עם 3 מחברות במקביל, סיקרן אותנו לגלות את ההבדלים והפערים בין שיטה כזו לאחרת. המחברת עם השינויים הגדולים ביותר היא זו אשר שינינו בה את ערכי הפיצ'רים שלא התפלגו נורמלית. בגלל שרוב הפיצ'רים שלנו לא התפלגו נורמלית, ביצענו מניפולציה מתמטית לפי השיטה הבאה: במקרה בו ההתפלגות הזדנבה לימין שינינו את הערך ע"י הפעלת לוג על 1 + הערך המקורי. במקרה אחר בו ההתפלגות הזדנבה לשמאל שינינו את הערך ע"י הפעלת לוג על 1 – הערך המקורי. כך שבשני המצבים התפלגות הפיצ'רים התקרבה יותר להתפלגות נורמלית. פיצ'רים בעלי התפלגות נורמלית לא עברו טרנספורמציה כלשהי. ההפעלה של LOG על חלק מהמשתנים גורמת להם להיות מסבירים טובים יותר של ערך המטרה (נראה זאת גם בהמשך בעיקר ברגרסיה הלינארית). השתמשנו בשיטה זו במטרה להיפטר מהכמות האדירה של ערכי קיצון בנתונים, לאחר שבכל זאת נשארנו עם ערכי קיצון החלטנו להמשיך איתם אבל לראות בכל זאת את ההשפעה של שינוי ערכי הפיצ'רים על תוצאות המודלים.

**השיטות הסטטיסטיות**  
החלטנו להשתמש ב-2 מבחנים סטטיסטיים. האחד טי טסט והשני קורלציה. מטרתו של הטי טסט הינה השוואה בין שתי קבוצות בלתי תלויות. את מבחן זה עושים למדגמים בלתי תלויים. יכולה להיות הטיה אם המנייה חוזרת על עצמה- לכן לאחר שחיברנו את 5 קבצי הדאטה הוספנו שנה לשם המנייה וכך יצרנו ערך ייחודי. המבחן בעצם עונה על השאלה האם יש הבדל מובהק בין 2 קבוצות – קבוצה ראשונה המניות שעלו וקבוצה שנייה המניות שירדו. במבחן זה משתנים מובהקים הם אלו שערך ה- P VALUE שלהם נמוך מ- 0.05. במחקרנו נמצא הבדל מובהק בין קבוצת המניות העולות לקבוצת המניות היורדות. כמו כן, נמצאו 82 משתנים מובהקים ביניהם: הכנסות, רווחים, הוצאות מחקר ופיתוח, הוצאות מנהלה וכלליות, חלקות דיבידנדים, השקעות לטווח קצר, מזומן, מלאי, סך הנכסים השוטפים ועוד. ואילו מספר המשתנים הלא מובהקים עמד על כ- 138 ביניהם: חייבים, הוצאות מס הכנסה, רווח למניה, חובות, סך נכסים, סה"כ נכסים לא שוטפים, סה"כ התחייבויות, השקעות , הון חוזר ועוד. מבחן זה עזר לנו להבין אילו פיצ'רים משמעותיים ביותר עבור החיזוי אותו אנו רוצים לבצע על המשתנה התלוי. מטרתה של הקורלציה הינה בדיקת קשרים לא ליניאריים בין משתנים כלומר השוואה בין המשתנים הבלתי תלויים (הפיצ'רים) לעצמם וגם לערך המטרה במקרה של רגרסיה לינארית. על מנת לבצע זאת הרצנו את מבחן ספירמן ובדקנו האם יש מולטיקולינאריות גבוהה בין פיצ'רים. הגדרנו מולטיקולינאריות גבוהה כאשר הקשר בין המשתנים הבלתי תלויים הינו מעל 0.85. במניעת המוליטיקולינאריות הורדנו מספר רב של פיצ'רים שהקורלציה ביניהם הייתה גבוהה תוך התחשבות בהשארת הפיצ'ר בעל הקורלציה הגבוהה ביותר למשתנה התלוי ("PRICE VAR"). יש לציין כי קיימת סבירות ששני משתנים בלתי תלויים יחד (סינרגיה) עשויים להסביר טוב יותר את משתנה המטרה.

# **תוצאות**

## **סטטיסטיקה תיאורית והסקה סטטיסטית:**

בעזרת סטטיסטיקה אנחנו יכולים להתבונן במידע כראוי, להסיק מסקנות ממערכי נתונים גדולים, ולבצע תחזיות אמינות. סטטיסטיקה תיאורית מציגה את הנתונים על ידי חישוב ממוצע, חציון, שכיח, סטיית תקן וכד'. כפי שניתן ללמוד מ**נספח ג** על התפלגות משתנה "סקטור".

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

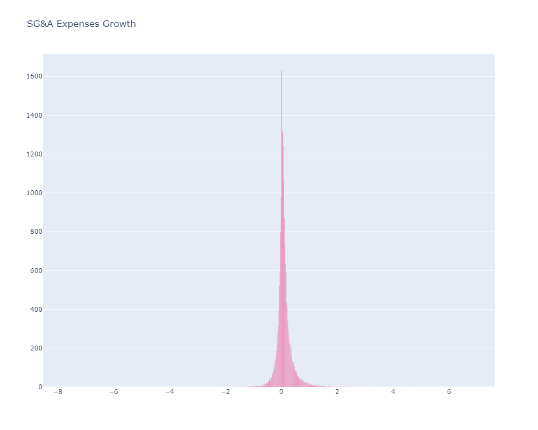
כפי שניתן לראות מספר הערכים בכל אחד מהמשתנים אינו זהה דבר הנובע מהימצאותם של ערכים חסרים בנתונים. זאת ועוד, סטיית התקן בכל אחד מהמשתנים הבלתי תלויים גדולה מאוד דבר המעיד על פיזור גדול בנתונים כלומר, מספר גדול של אאוטליירים. כמו כן, ניתן לראות כי קיימים פערים עצומים בין הערך המינימאלי וערך הממוצע לערך המקסימלי.

התפלגות המשתנים לפי ההיסטוגרמות:

בעיית האאוטליירים גרמה לנו לבדוק את התפלגות המשתנים. המטרה הייתה להיפטר מכמות נכבדת של אווטליארים ע"י ביצוע מניפולציה מתמטית עבור משתנים שלא מתפלגים נורמלית.

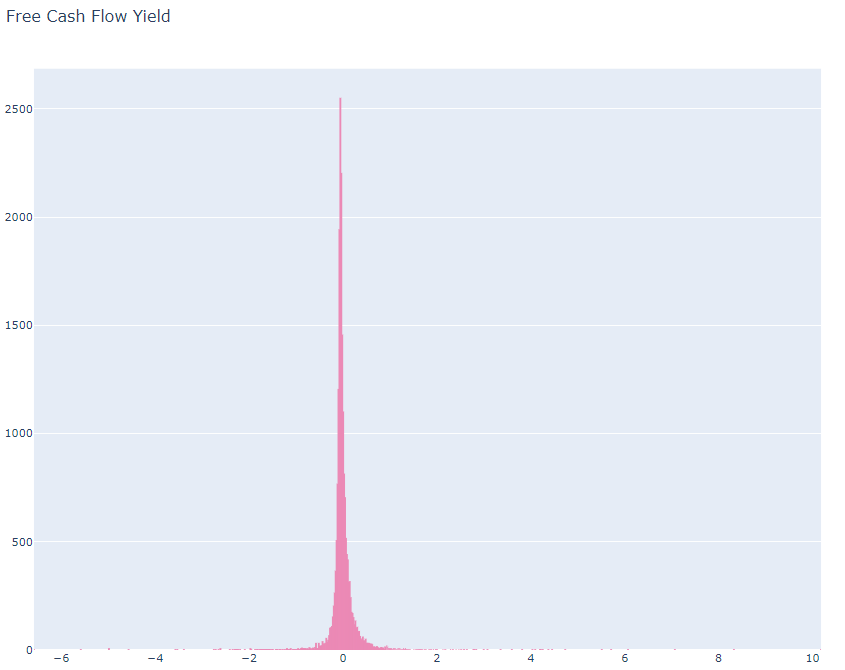
דוגמאות של משתנים בעלי התפלגות מזדנבת לימין הינם: גידול בהוצאות מכירות הנהלה וכלליות, גידול בחובות, גידול בחייבים ועוד. בדוגמא שלפנינו התפלגות של משתנה " גידול בהוצאות מכירות הנהלה וכלליות ".

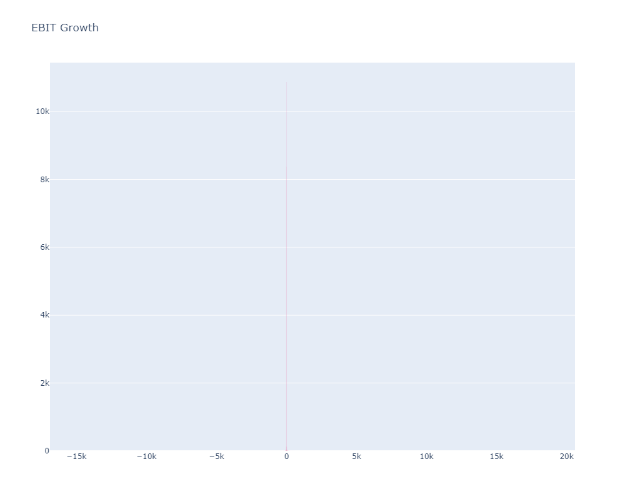
תמונה שמכילה שולחן

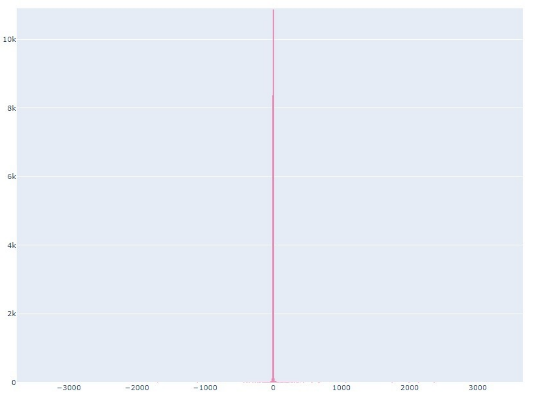
התיאור נוצר באופן אוטומטי

דוגמאות של משתנים בעלי התפלגות מזדנבת לשמאל הינם: תשואת תזרים מזומנים חופשי, הוצאות הוניות למניה ועוד. בדוגמא שלפנינו התפלגות של משתנה " תשואת תזרים מזומנים חופשי ".

תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי

דוגמאות של משתנים בעלי התפלגות נורמאלית הינם: גידול ברווח תפעולי, גידול רווח נוכחי למניה ב12 החודשים האחרונים, איכות הכנסה ועוד. בדוגמא שלפנינו התפלגות של משתנה " גידול ברווח תפעולי ".

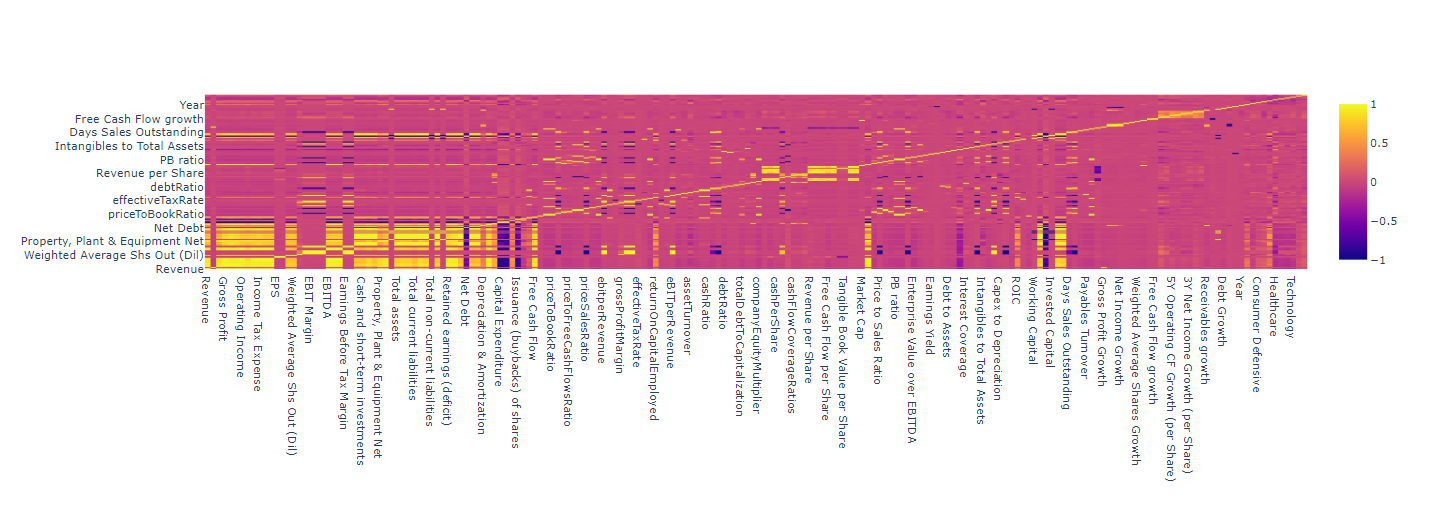


T TEST:

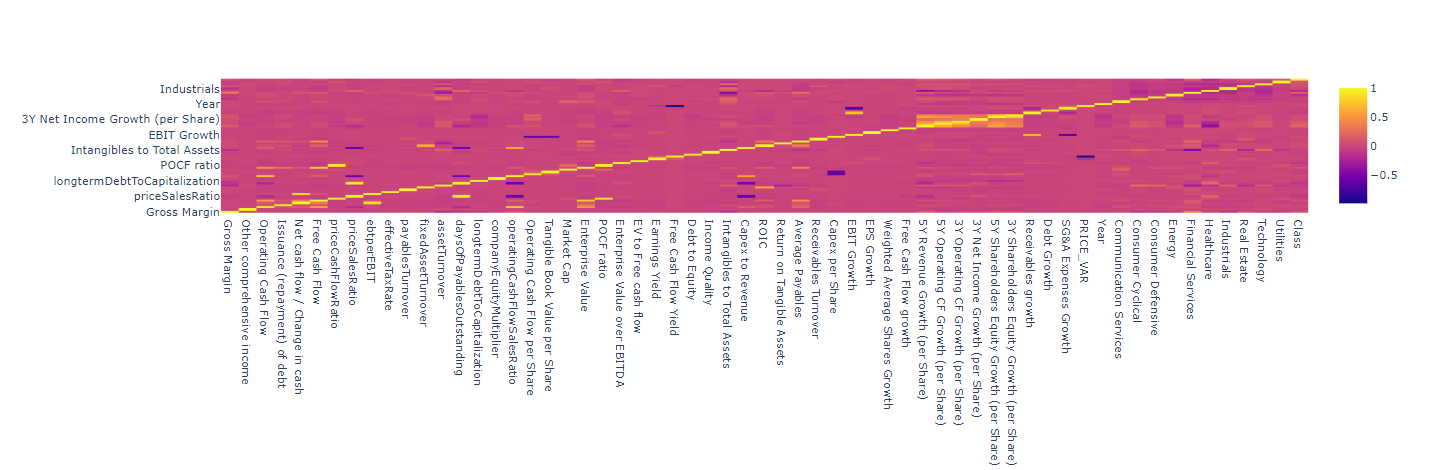
מבחן שמטרתו לבצע השוואה בין שתי קבוצות בלתי תלויות, נעשה למדגמים בלתי תלויים. המבחן עונה על השאלה האם יש הבדל מובהק בין 2 הקבוצות- קבוצת המניות שעלו וקבוצת המניות שירדו. במבחן זה קיימות מספר הנחות: 1. התפלגויות האוכלוסייה הן נורמליות 2.לדוגמאות יש שונות שוות 3.שתי הקבוצות הנבדקות אינן תלויות.

קורלציה:

מפת חום לפני הורדת המשתנים בעלי קורלציה גבוהה ביניהם (קורלציה גבוה הוגדרה כגדולה מ0.85)



מפת חום אחרי הורדת המשתנים בעלי קורלציה גבוהה ביניהם תוך התחשבות בהשארת המשתנה בעל הקורלציה הגבוהה ביותר למשתנה המטרה



## **תוצאות הרצת האלגוריתמים ו-evaluation:**

על מנת להביא לתוצאות טובות יותר ולהימנע מOVERFITTING במודלים שענו על קלסיפיקציה ביצענו חלוקה של הנתונים בצורה הבאה: 60% סט אימון, 20% סט ולידציה ו20% סט מבחן. את כל הבדיקות ביצענו על סט הוולידציה זאת על מנת שנוכל לשפר את המודלים מבלי לגעת בקבוצת המבחן, רק בסוף תהליך האופטימיזציה הרצנו את המודלים על סט המבחן. ניתן לראות שבכל אחד מהמודלים הRECALL והF1 טובים יותר בלחזות 1 מאשר 0 זה ככל הנראה מהסיבה שבהרצת המודלים השתמשנו בדאטה עם מספר רב יותר של מניות שעלו מאשר שירדו. לכן מתוך הנחה זו בדקנו את אחוז המניות שעלו וירדו בכל אחד מהסטים ומצאנו כי הפערים לא גדולים עד כדי כך שיסבירו בצורה טובה את הפערים במדדים בין 2 הקבוצות (קבוצת המניות שעלו לעומת קבוצת המניות שירדו), היינו ממליצות בעתיד לבדוק כיוונים נוספים לפער הגדול בתוצאות.

כפי שהזכרנו לעיל בנינו 3 מחברות:

1. מחברת MAIN- מילוי ערכים חסרים באמצעות KNN מתוך ההנחה כי זוהי השיטה הטובה ביותר למילוי ערכים חסרים כאשר מדובר בנתונים על מניות.(**נספח ד**)
2. מחברת FILL\_MEAN – מילוי ערכים חסרים באמצעות ממוצע
3. מחברת HIST בה ערכי הפיצ'רים הינם ערכים לוגריתמים.

להערכת ביצועי הקלסיפיקציה נשתמש במדד Accuracy – המעיד על הדיוק בחיזוי.

תוצאות רגרסיה לוגיסטית: (**נספח ה**)

במחברת 3 קיבלנו את התוצאות הכי טובות ברגרסיה הלוגיסטית עם Accuracy 0.61 בסט המבחן. כמו כן בכל המחברות תוצאות סט הוולידציה דומות לתוצאות סט המבחן, דבר המעיד על כך שאין לנו אוברפיטינג וביצענו אופטימיזציה טובה בכל אחד מהמודלים. המחברת עם הביצועים הגרועים ביותר ברגרסיה הלוגיסטית הינה מחברת 2 עם Accuracy של 0.58.

תוצאות יערות רנדומיים לפני דירוג פיצ'רים (**נספח ו**)

לפני דירוג הפיצ'רים במחברת 1 קיבלנו את התוצאות הטובות ביותר עם Accuracy 0.62 בסט המבחן. במחברות האחרות Accuracy זהה שווה ל-0.6.

תוצאות cross validation:

כחלק מתהליך האופטימיזציה של מודל יערות הרנדומיים ביצענו cross validation על המודל במטרה לקבל Accuracy גבוה. Cross validation מבוסס על פיצולים אקראיים מרובים של הנתונים לסט אימון וסט מבחן. ה-Accurancy הגבוה ביותר שהתקבל בשיטה זו הינו 0.5567 במחברת 3.כמו כן, מודלים אחרים הציגו ביצועים טובים יותר.

תוצאות יערות רנדומיים אחרי בחירת פיצ'רים:

ביצענו feature selection (**נספח יד**) במטרה לשפר את ביצועי המודל שנמצא כטוב ביותר משאר המודלים. לכן החלטנו לבצע במודל זה אופטימיזציה נרחבת יותר וניסינו אף לשנות את ערך הסף לערך אופטימלי. לאחר בדיקה גילינו כי שינוי ערך הסף לא מועיל בחיזוי והסטיות מינוריות ביותר לכן החלטנו להשאיר אותו על ערך ברירת מחדל - 0.5. בהרצת המודל עם feature selection התייחסנו רק לפיצ'רים אשר הקשר שלהם לפונקציית המטרה גבוהה מ-0.005 ומצאנו כי התוצאות במחברת 2 השתפרו מ-=Accuracy0.6 ל Accuracy=0.61 לעומת זאת במחברת 1   
ה Accuracy-ירד ב0.01.

במודל יערות רנדומיים ביצענו אופטימיזציה של מס העצים האופטימלי ועומק העץ המירבי כקומבינציה של שניהם יחד ולא כל אחד בנפרד. במחברת 1 התבססנו על 33 עצים ועומק עץ 11, במחברת 2 על 28 עצים ועומק עץ 18 ובמחברת 3 התבססנו על 36 עצים ועומק עץ 18. (**נספח טז**)

תוצאות KNN (**נספח ז**)

בכל אחת מהמחברות הרצנו את המודל לפי כמות השכנים האופטימלית עבורו ע"פ כלל המרפק. במחברת 1 התבססנו על 31 שכנים(**נספח יא**), במחברת 2 על 35 שכנים (**נספח יב**) ובמחברת 3 על 19(**נספח יג**). כפי שניתן לראות בכל מחברת בחרנו מספר אי זוגי ויחסית גבוה. ממצא מעניין הוא שלמרות שבמחברת 3 התבססנו על מס השכנים הקטן ביותר התוצאות בה דווקא הכי טובות עם Accuracy של 0.6 לעומת 0.59 במחברת 1 ו0.58 במחברת 3.

תוצאות רגרסיה לינארית

במודל זה השתמשנו ב2 קבוצות סט האימון וסט המבחן. הגענו למסקנה שהמודל שלנו חוזה טוב יותר מניות שהשינוי השנתי שלהן קטן יחסית, זאת בגלל שרב הפיזור הוא סביב ה-0 (**נספח ח**). ומנגד מתקשה מאוד לחזות שינויים גבוהים יותר. ניתן להסביר את הפיזור והתוצאות של הרגרסיה הלינארית לפי ההתפלגות של המשתנה התלוי "PRICEVAR" הפיזור שלו מתרכז סביב ערך ה-0 (**נספח ט**) .

ברגרסיה זו נתייחס למחברת 3 בה קיבלנו את הנתונים הטובים ביותר בפער עצום משאר המחברות. הסיבה לכך היא שבמחברת 3 הרגרסיה היא בעצם רגרסיה לינארית פולינומית רבת משתנים דבר המדייק יותר את המודל. ברגרסיה זו ה- RMSE (שורש הסטיות) של סט האימון קטן פי 17 מזה של סט המבחן דבר המעיד על אוברפיטינג. כמו כן המודל יצא לא מדויק הן בסט האימון והן בסט המבחן R2 0.0691 בסט האימון ו0.0918 בסט המבחן. תוצאות אלה עולות בקנה אחד עם השערותינו והמסקנות מסקירת הספרות שקשה מאוד לחזות במדויק ביצועי מניות על בסיס נתונים פיננסיים בלבד באמצעות מודלים פשוטים כמו רגרסיה.

**הRMSE ל סטיית התקן של פונקציית המטרה** (PRICE VAR)

הוא מדד סטטיסטי המשמש ברגרסיה ליניארית המעיד על כמה מהשונות של המשתנה התלוי (y) מוסברת על ידי המשתנים הבלתי תלויים(X). ככל שהוא יותר גדול – המודל מסביר טוב יותר את סט המבחן.

**MAE -** מחשב את המרחקים הממוצעים של כל השגיאות.זהו ציון ליניארי: כל ההבדל משוקלל באופן שווה בממוצע.

**MSE-** אחת הגישות השימושיות ביותר, מחשבת את המרחקיםהריבועיים הממוצעים של כל השגיאות זהו ציון לא ליניארי: outliers יגדילו דרמטית את הציון.

**RMSE -** שווה ערך ל-MSE, לוקח את השורש הריבועי של הניקוד

במודלים טובים **RMSE < STD** (השונות הפנימית צריכה להיות גבוהה יותר משגיאות החיזוי).   
בנוסף האימון RMSE ~= המבחן RMSE (דומה)

ברגרסיה הפולינומית אכן סטיית התקן גדולה משורש ריבועי הטעויות דבר המעיד על מודל טוב יחסית לעומת שתי הרגרסיות הלינאריות האחרות שהרצנו בהן סטיית התקן קטנה בהרבה משורש ריבועי הטעויות.

# **מסקנות ודיון**

## **מסקנות ותובנות:**

לאורך השנים משקיעים וחוקרים רבים מחפשים את התשובה לשאלה - כיצד ניתן לחזות ולהעריך את התשואה על נייר ערך מסוים. השתמשנו במגוון אלגוריתמים על מנת לענות על שאלות המחקר שהעלנו בעבודה. לאחר קריאה מעמיקה של מחקרים קודמים והרצת המודלים הגענו למספר תובנות. ראשית, קיימים הרבה מדדים פיננסיים אשר חופפים במשמעותם, ובשל כך בעלי מולטיקולינאריות גבוהה ביניהם. ע"י בדיקת קורלציה מנענו פגיעה באיכות המודל.

מסקנה מעניינת שעלתה בעבודתנו היא ששווה יותר להשקיע במניות במגזר השירותים ואילו במניות של חברות האנרגיה הכי פחות כדאי להשקיע מפני ש-60% מהן ירדו (**נספח ג**).

רוב המודלים בהם השתמשנו אינם מספקים תוצאות מדויקות מספיק לחיזוי מניות כי אינם מסוגלים לדמות בעיות מורכבות. מודל היערות הרנדומיים אשר נחשב כמודל יותר מתוחכם הגיע לתוצאות טובות יותר. הסקירה הספרותית מראה כי על מנת להצליח בחיזוי מניות עדיף להשתמש במודלים מורכבים כגון רשתות נויירוניות , SVM וכד'.

בנוסף, לאחר ביצוע אופטימיזציה למודל היערות הרנדומיים הצלחנו לשפר במעט את תוצאות המודל כלומר, בדקנו מהי הטעות המינימלית עבור קומבינציה של מספר העצים ועומקם המירבי.

זאת ועוד, בעזרת feature selection הצלחנו לאתר את הפיצ'רים הרלוונטיים למודל שלנו והגענו למסקנה אילו מהם יכולים להרכיב את המודל שלנו בצורה האופטימילית ביותר (**נספח יד**). במהלך הרצת היערות הרנדומליים החלטנו להשתמש ב-cross validation אך לאחר קבלת התוצאות הגענו למסקנה שחלוקה זו לא שיפרה את התוצאות בצורה משמעותית. זאת ועוד, ניתן לראות בתוצאות שהתקבלו כי מודל הרגרסיה הפולינומית מניב ביצועים טובים בהרבה מאשר רגרסיה לינארית מרובת משתנים. בנוסף , הגענו למסקנה כי בעת שימוש במודל רגרסיה לינארית מרובת משתנים אנו נצליח לחזור בצורה טובה מניות שאחוז השינוי שלהן קטן (**נספח ח**).

## **מה למדתי מהעבודה?**

השקעה במניות יכולה להיות משתלמת במיוחד, אבל היא עלולה להיות גם מסוכנת. שוק המניות הינו נושא בעל מספר רחב של רבדים. על מנת לחזות האם המניה בה אנו רוצים להשקיע תרד או תעלה יש הרבה משתנים שיש לקחת בחשבון. אך חשוב להבין בורסה היא לא קזינו, ומניות מתנהגות, בטווח הארוך בהתאם לביצועים הכספיים של החברות - אם החברה מציגה שיפור עסקי מתמשך, גידול ברווחים ובהכנסות צפוי שזה יתבטא במחיר המניה. בשלב הראשוני של המחקר שביצענו היה לנו חשוב להיכנס לעומק הנתונים ולהבין מהם המשתנים שמופיעים בדוחות הכספיים? מתי החברות המונפקות בבורסה מגישות אותם לרשות? עד כמה הנתונים אמינים? והאם אפשר להסתמך עליהם לצורך חיזוי השינוי במניה בשנה העוקבת? כל זה תרם לנו רבות הן להבנה מהם דוחות ה-10K וכיצד אפשר לבנות על-פיהם מודלים. בשלב השני של המחקר נתקלנו בשני קשיים עיקריים שעל מנת לטפל בהם היינו צריכות ללמוד עוד יותר לעומק את נושא מחקרנו. הראשון ערכי 0 רבים והשני – ערכים קיצוניים. גילינו כי בדוחות הכספיים מופיעים ערכי 0 במקומות אינם הגיוניים כמו הוצאות מנהלה וכלליות או הוצאות מס הכנסה שכן לחברות הנסחרות בבורסה יש בוודאות הוצאות כאלו. בנוסף כאשר ראינו שיש בנתונים מספר עצום של ערכי קיצון בדקנו האם מצב זה הגיוני לשוק המניות. גילינו כי בתחום המניות קיים מושג- תנודתיות- תכונה של נכס פיננסי, שנמדדת באופן סטטיסטי דרך פיזור ערכיו על פני ציר זמן. ככל שערכיו משתנים בקצב מהיר יותר כלפי מעלה או מטה - כך מתחזקת התנודתיות שלו. התנודתיות באה לידי ביטוי בסטיית התקן של הנכס. ככל שגדלה סטיית התקן של המניה כך גם התנודתיות שלה עולה. סטיית תקן של נכס משמשת לעתים קרובות מדד לרמת הסיכון שלו. לכן יכול להיות מצב בו המניות שנתנו לנו בדאטה הן מסוכנות יותר והתנודתיות שלהן גבוהה יותר – לכן ישנם הרבה הערכים קיצוניים. זאת ועוד, בעזרת feature selection וקורלציה למדנו מהם המשתנים מתוך הדוחות הכספיים המשפיעים ביותר על חיזוי המניה. במהלך כתיבת המחקר חידדנו את הידע שלנו בכתיבת קוד ובספריות הקיימות, גילינו עולם עצום שנרצה להרחיב בו את אופקינו גם בעתיד. על אף שתוצאות המחקר לא נתנו תשובה חד משמעית האם המניה תעלה או תרד למדנו מהמחקר כמה שתחום המניות קשה לחיזוי הן בגלל שהוא רגיש לשינויים כגון מגפות, טעם משתנה של האוכלוסייה (טרנדים), מלחמות ועוד. והן בגלל שהוא מושפע ממגוון משתנים , מה שדורש שימוש במודלים מתוחכמים שנרצה ללמוד בהמשך. חשוב לציין כי המחקר תרם לנו לפיתוח יכולת למידה וחשיבה עצמית.

## **יתרונות וחסרונות וכיצד היה ניתן להתגבר עליהם:**

למחקר שביצענו יש מספר יתרונות אך גם מספר חסרונות.

## **מבט לעתיד:**

אנו מאמינות שניתן להמשיך לפתח ולשפר את המודל על ידי שימוש במאגר מידע גדול יותר, שימוש בנתונים עדכניים יותר וחיבור לנתונים בזמן אמת שיחזקו את תוצאות המחקר. בנוסף לכך, היינו ממליצות לאסוף מידע לא רק על נתוני דוחות 10K אלא גם לערוך שאלון עבור משקיעים ועל ידי כך לבדוק את הציפיות שלהם. ממחקרים רבים גילינו שלעיתים רק ניתוח פונדמנטלי שמסתמך על לימוד מצבה הכלכלי של חברה מסוימת על ביצועיה הכספיים וחוזקה הפיננסי לא מספיק כיוון ששוק המניות הוא מורכב והתנהגות השווקים מושפעת גם מפסיכולוגיה רבה ודפוסי עבר ,ולבסוף היינו ממליצות גם לאסוף מידע על נתונים מאקרו כלכליים של המשק. שילוב של כל אלה יכול לשפר את אמינות המודל.

# **ביבליוגרפיה:**

1. Stock selection with random forest: An exploitation of excess return in the Chinese stock market (Zheng Tan a,\* , Ziqin Yan a , Guangwei Zhu b)

<https://www.cell.com/heliyon/pdf/S2405-8440(19)35970-5.pdf>

1. Stock Price Prediction Using K-Nearest Neighbor (kNN) Algorithm (Khalid Alkhatib1 Hassan Najadat2 Ismail Hmeidi 3 Mohammed K. Ali Shatnawi)

<https://www.ijbhtnet.com/journals/Vol_3_No_3_March_2013/4.pdf>

1. Prediction of stock performance by using logistic regression model: evidence from Pakistan Stock Exchange (PSX)( Syed Shahan Alia , Muhammad Mubeenb, Irfan Lalc, Adnan Hussain) <https://archive.aessweb.com/index.php/5004/article/view/4221/6554>

# Stock price prediction methodology using random forest algorithm and support vector machine ([Pavan Kumar Illa](https://www.sciencedirect.com/science/article/pii/S2214785321069674" \l "!) [Balakesavareddy Parvathala](https://www.sciencedirect.com/science/article/pii/S2214785321069674#!) [Anand Kumar Sharma](https://www.sciencedirect.com/science/article/pii/S2214785321069674#!))

<https://www.sciencedirect.com/science/article/pii/S2214785321069674>

# Stock Closing Price Prediction using Machine Learning Techniques ([Mehar Vijh](https://www.sciencedirect.com/science/article/pii/S1877050920307924" \l "!) [Deeksha Chandola](https://www.sciencedirect.com/science/article/pii/S1877050920307924#!) [Vinay Anand Tikkiwal](https://www.sciencedirect.com/science/article/pii/S1877050920307924#!) [Arun Kumar](https://www.sciencedirect.com/science/article/pii/S1877050920307924#!))

https://www.sciencedirect.com/science/article/pii/S1877050920307924

1. מדד אלטמן -המכלול

<https://www.hamichlol.org.il/%D7%9E%D7%93%D7%93_%D7%90%D7%9C%D7%98%D7%9E%D7%9F>

1. Bank Profitability and Inflation Valuation of Bank's Shares / הערכת מניות הבנקים המסחריים

<https://www.jstor.org/stable/23753054>

# **נספחים:**

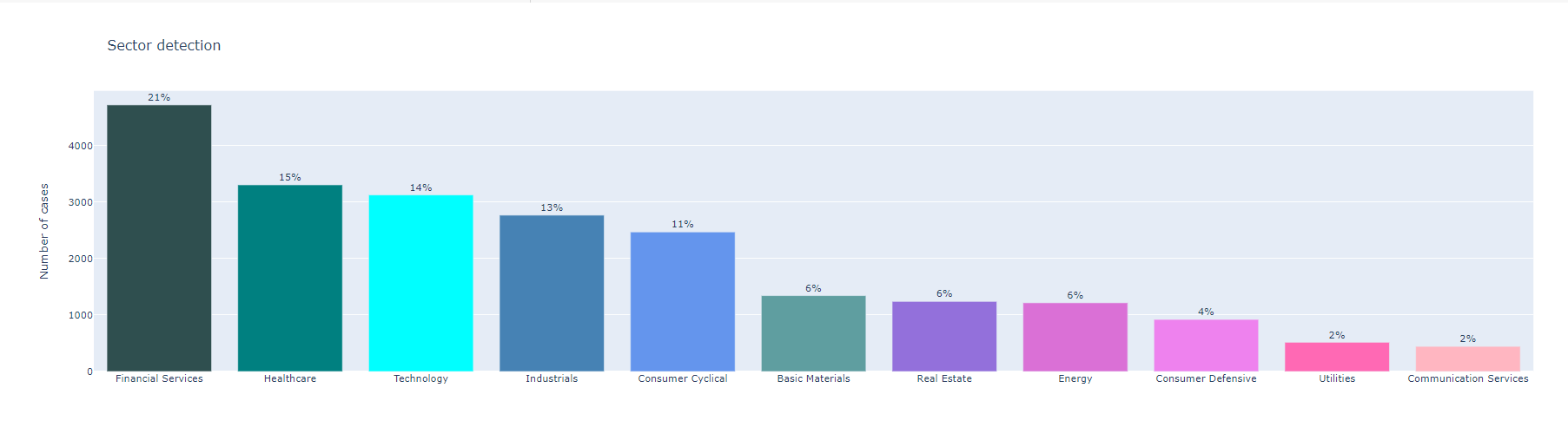
### **נספח א**

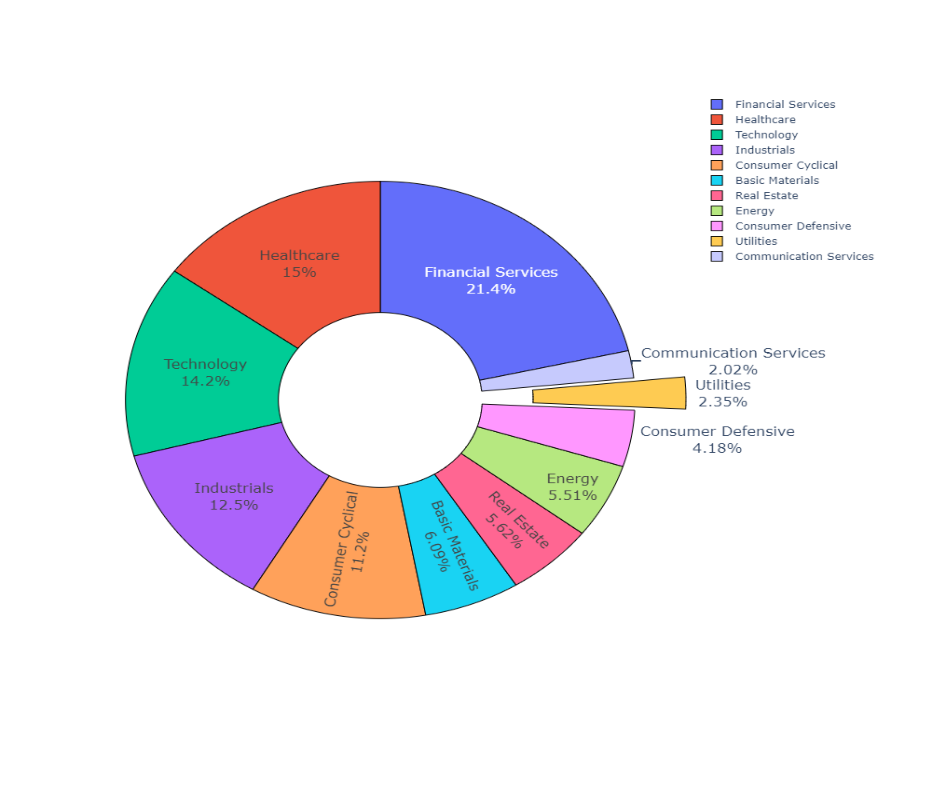
|  |  |
| --- | --- |
| Gross Margin | רווח גולמי |
| Other comprehensive income | הכנסה כוללת אחרת |
| Operating Cash Flow | תזרים מזומנים תפעולי |
| Issuance (repayment) of debt | הנפקה (החזר) חוב |
| Net cash flow / Change in cash | תזרים מזומנים נטו / שינוי במזומן |
| Free Cash Flow | תזרים מזומנים חופשי |
| price Cash Flow Ratio | מחיר יחס תזרים מזומנים |
| price Sales Ratio | מחיר יחס מכירה |
| ebt per EBIT | רווח לפני מסים לכל רווח תפעולי |
| effectiveTaxRate | שיעור מס אפקטיבי |
| payables Turnover | חובות מחזור |
| fixed Asset Turnover | מחזור נכסים קבועים |
| asset Turnover | מחזור נכסים |
| days Of Payables Outstanding | ימי תשלומים |
| longterm Debt To Capitalization | חוב להיוון טווח ארוך |
| company Equity Multiplier | מכפיל הון של חברה |
| operating Cash Flow Sales Ratio | יחס מכירות תזרים מזומנים תפעולי |
| Operating Cash Flow per Share | תזרים מזומנים תפעולי למניה |
| Tangible Book Value per Share | ערך ספר מוחשי למניה |
| Market Cap | שווי שוק |
| Enterprise Value | ערך ארגוני |
| POCF ratio | יחס מכפיל מזומנים מפעילות שוטפת |
| Enterprise Value over EBITDA | ערך ארגוני מעל הEBITDA |
| EV to Free cash flow | שווי פעילות החברה מול תזרים מזומנים חופשי |
| Earnings Yield | תשואת רווחים |
| Free Cash Flow Yield | תשואת תזרים מזומנים חופשי |
| Debt to Equity | חוב להון עצמי |
| Income Quality | איכות הכנסה |
| Intangibles to Total Assets | נכסים בלתי מוחשיים לסך הנכסים |
| Capex to Revenue | הוצאות הוניות מול ההכנסה |
| ROIC | תשואה להון מושקע |
| Return on Tangible Assets | החזר על נכסים מוחשיים |
| Average Payables | תשלומים ממוצעים |
| Receivables Turnover | מחזור חובות |
| Capex per Share | הוצאות הוניות למניה |
| EBIT Growth | גידול ברווח תפעולי |
| EPS Growth | גידול רווח נוכחי למניה ב12 החודשים האחרונים |
| Weighted Average Shares Growth | צמיחת מניות ממוצעת משוקללת |
| Free Cash Flow growth | צמיחה בתזרים מזומנים חופשי |
| 5Y Revenue Growth (per Share) | גידול בהכנסות במשך 5 שנים למניה |
| 5Y Operating CF Growth (per Share) | צמיחת תקבול שנתי תפעולי של 5 שנים (למניה) |
| 3Y Operating CF Growth (per Share) | צמיחת תקבול שנתי תפעולי של 3 שנים (למניה) |
| 3Y Net Income Growth (per Share) | גידול בהכנסה נטו של שלוש שנים |
| 5Y Shareholders Equity Growth (per Share) | צמיחה במניות בעלי המניות של 5 שנים למניה |
| 3Y Shareholders Equity Growth (per Share) | צמיחת הון בעלי המניות של שלוש שנים למניה |
| Receivables growth | גידול בחייבים |
| Debt Growth | גידול בחובות |
| SG&A Expenses Growth | גידול בהוצאות מכירות הנהלה וכלליות |

### **נספח ב**



### **נספח ג**





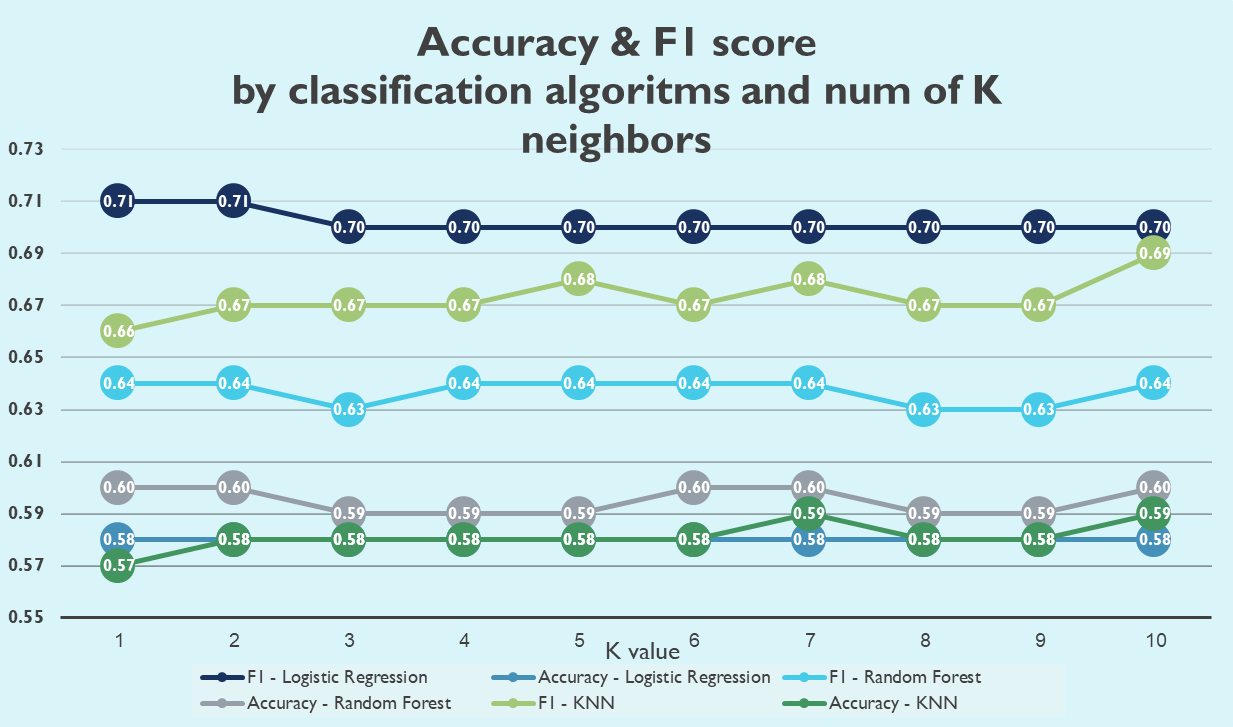
תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי

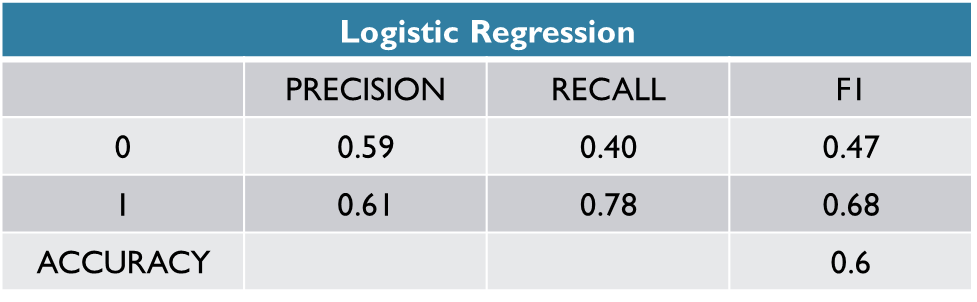
### **נספח ד**

תמונה שמכילה שולחן

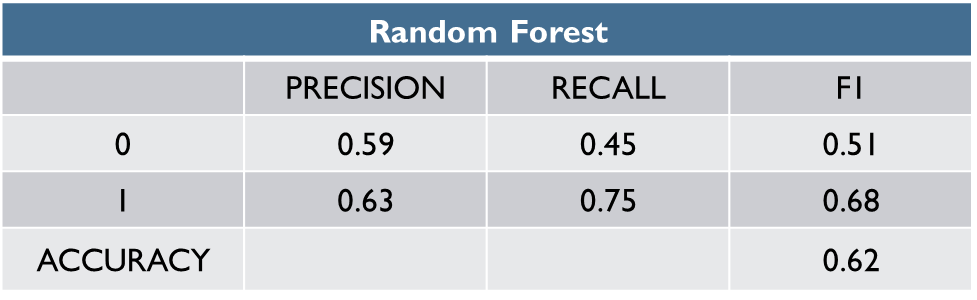
התיאור נוצר באופן אוטומטי



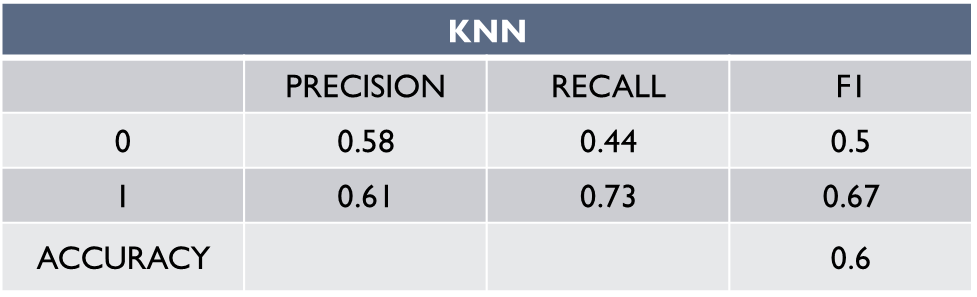
### **נספח ה**



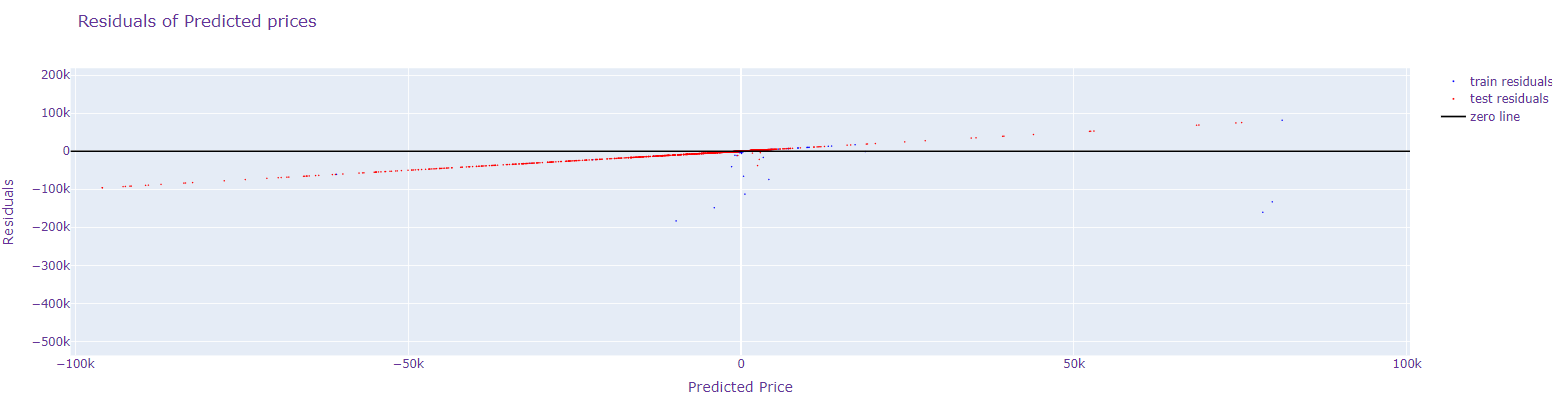
### **נספח ו**



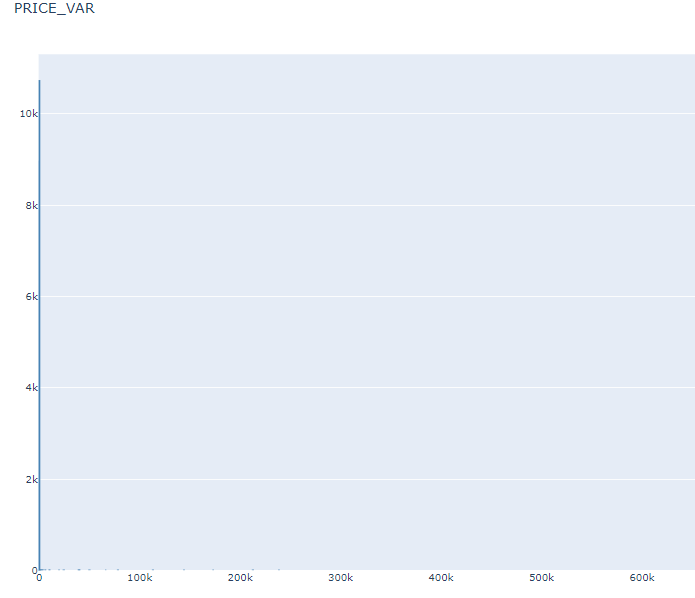
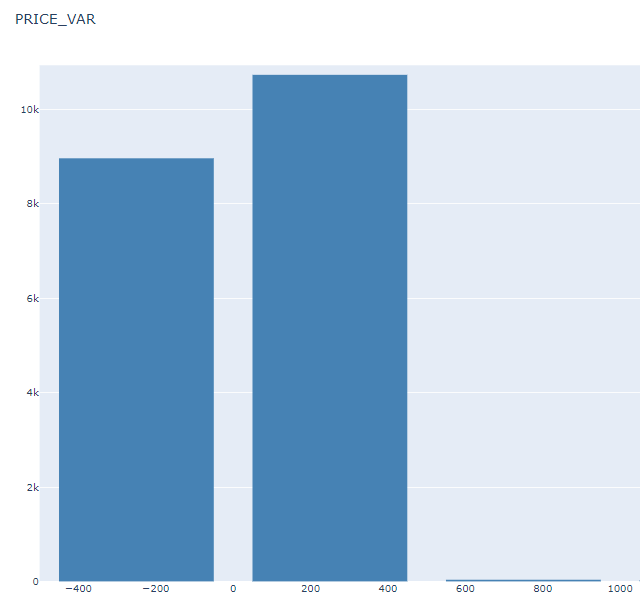
### **נספח ז**



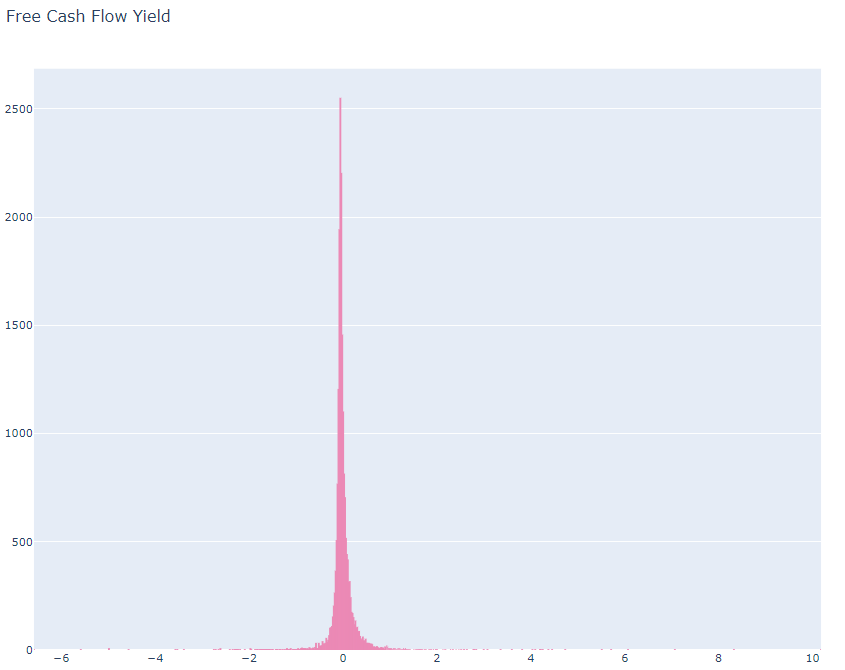
### **נספח ח**



### **נספח ט**



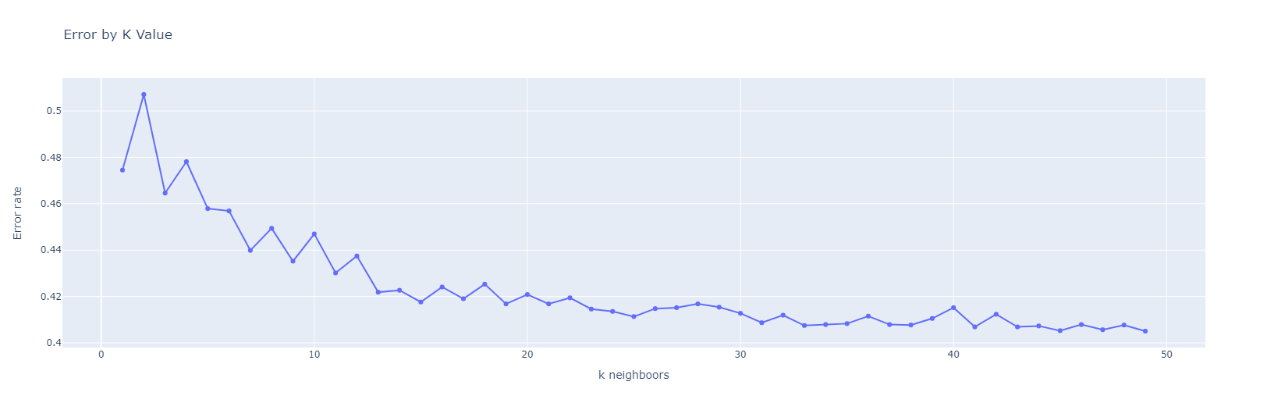
### **נספח י**

תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי

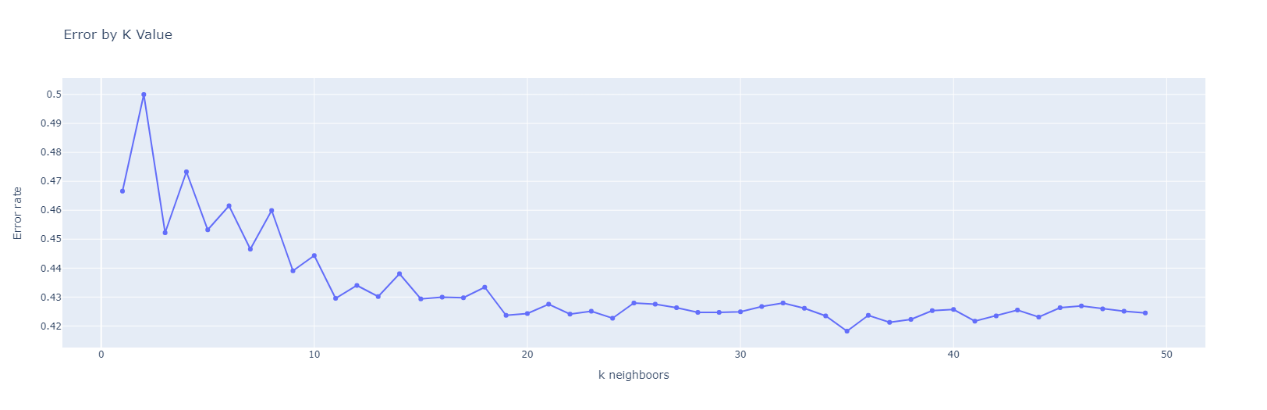
### **נספח יא**

main



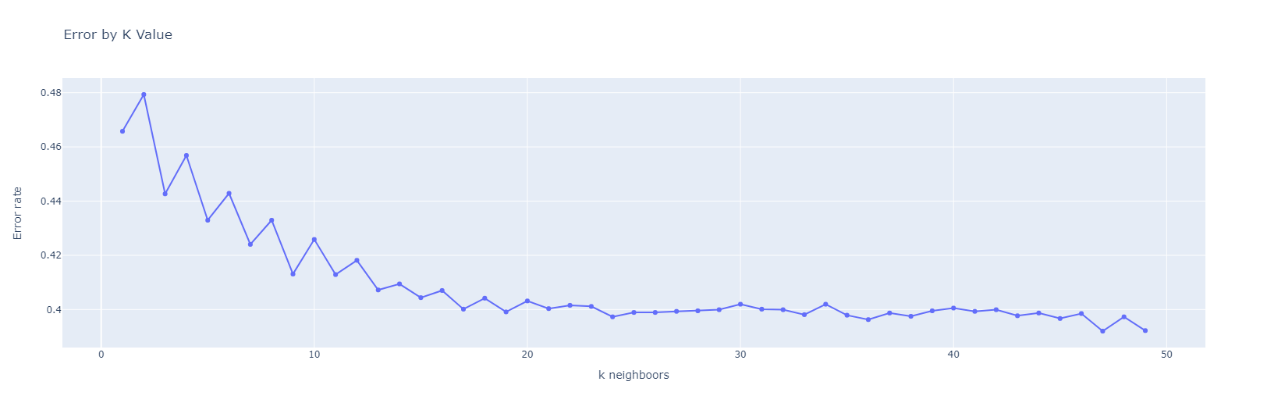
### **נספח יב**

Fill mean

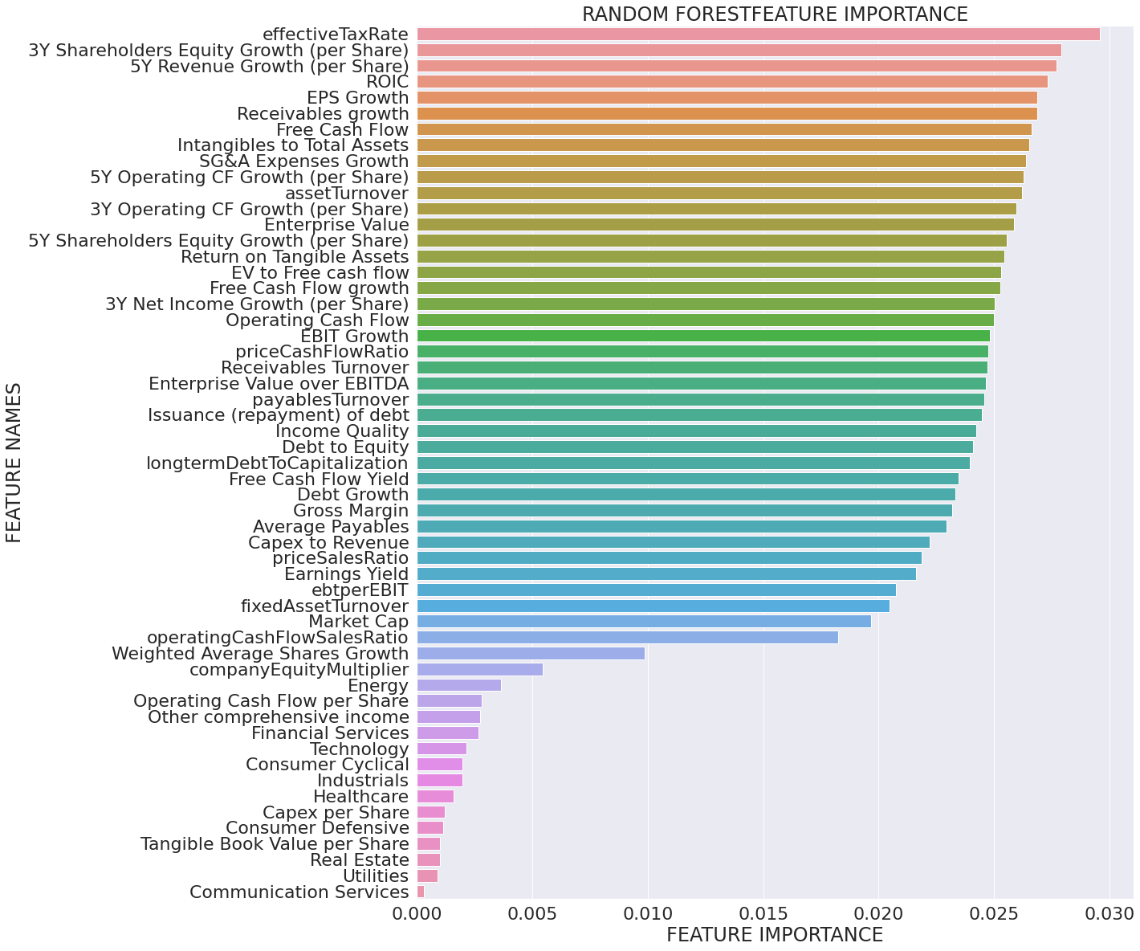


### **נספח יג**

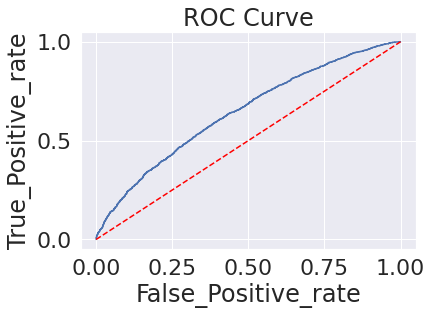
HIST



### **נספח יד**



### **נספח טו**



נספח טז

תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי

### **נספח טז**

