**Google Trends as a Sentiment Indicator in Algorithmic Trading**

מגישים:

גיא גלאט, ת.ז. 206745440

טום גבע, ת.ז. 318334901

הלל שרביט, ת.ז. 318757572

**הרעיון הכללי**

שוק הקריפטו, הוא במובנים רבים שוק מבוסס הייפ.

בניגוד לסקטורים אחרים, בהם ערכם של נכסים נקבע בעקבות שינויים כלכליים או טכנולוגיים, פעמים רבות, הערך של מטבעות קריפטו עולה ויורד בהתאם לאירועים חברתיים ולתגובות של אנשים ברשתות החברתיות. דוגמה בולטת לכך היא מטבע ,Shiba Inu מטבע שהושק כבדיחה ומתחרה ב- .Dogecoin הערך של Shiba Inu עלה משמעותית בעקבות ציוצים של דמויות ציבוריות כמו אילון מאסק.

התמיכה הפומבית של מאסק והאזכורים ברשתות החברתיות יצרו גל של התעניינות ורכישות, אשר הובילו לעלייה חדה בערכו של המטבע, למרות שאין לו ערך פנימי ברור או שימוש ממשי.

מגמה זו באה לידי ביטוי גם בחיפושים בגוגל, כאשר ניתן לראות גידול משמעותי בנפח החיפושים סביב Shiba Inu. החיפושים בגוגל מתפקדים כסוג של ברומטר למידת ההייפ וההתעניינות הציבורית במטבעות מסוימים, ועלולים להעיד על עלייה עתידית בערכו של המטבע המחופש.

Google Trends הוא כלי של גוגל שבעזרתו ניתן לקבל מידע על כמות חיפושים בגוגל לאורך זמן לפי מילת מפתח. הרעיון הכללי שלנו היה בניית אסטרטגיה המבוססת על מעקב אחר תנודות בנפח החיפושים של מונחי קריפטו שונים, במטרה לזהות מגמות בשוק לפני שהן מתבטאות בשינויי מחירים. לדוגמה, עלייה חדה בחיפושים של מונח מסוים יכולה לרמז על גל של התעניינות שעשוי להוביל לעלייה בערך המטבע.

**מקורות נתונים**

[Google Trends](https://trends.google.com/trends?geo=VI&hl=en-US)

[Glimpse](https://trends.google.com/trends/?glimpse=true)

[Binance](https://www.binance.com/en)

**סטטיסטיקה תיאורית**

להוסיף גרפים

**הסבר על תהליכי ניקוי וטיוב של הנתונים**

בתהליך הניקוי והטיוב של הנתונים נתקלנו ברוב הקשיים שלנו במהלך הפרוייקט.  
טיוב הנתונים נעשה באמצעות api לא רשמי של Google Trends אשר מוריד דאטה מGoogle Trends- כקובץ csv.

כתבנו פונקציה המקבלת תקופת זמן ומילת חיפוש ומשתמשת ב-api כדי להפיק קובץ csv של הנתונים הרלוונטיים, אותו היא שומרת כ-data frame שאיתו יכלנו לעבוד.

קושי אשר בו נתקלנו היה הפורמט בו הנתונים שייבאנו מGoogle Trends מוגדר ע"י שירות זה – כמות הימים המיוצגים על ידי כל data point הוא פונקציה של כמות הימים בתקופת הזמן המבוקשת, כך שכל קובץ מכיל עד 270 data points. כך, לדוגמה, אם נרצה דאטה על פני שנה קלנדרית, נקבל מ-Google Trends קובץ דאטה בו כל נקודה מייצגת שבוע, לעומת ייצוג יומי שנקבל אם נרצה דאטה על פני חצי שנה.

בנוסף, נתונים המוחזרים בכל טבלה מנורמלים לערכים בין 0 עד 100 בשיטה פנימית של Google שלא יכולנו למצוא מקור מהימן מספיק אשר נותן את החישוב המדוייק.

נרמול זה פגע ביכולת שלנו "לחבר" שתי תקופות זמן לתקופה אחת, מפני שכל תקופה הייתה מנורמלת אחרת.

כדי להתגבר על הקושי השתמשנו בשירות נוסף שמצאנו הנקרא Glimpse. Glimpse הוא תוסף לדפדפן Google Chrome אשר "מתלבש" על האתר והשירות שGoogle Trends- נותן. Glimpse נתן לנו את האפשרות להוריד כקובץ csv את כמות החיפושים בערכים מוחלטים במקום דאטה מנורמל.

לכל מטבע שנכלל במודל שלנו הורדנו מGlimpse- טבלה בטווח זמן של 5 השנים האחרונות, כלומר טבלה עם 262 דגימות שבועיות עבור על מטבע, כאשר כל דגימה מהווה כמות חיפושים משוערכת, מלאה ולא מנורמלת של שבוע ימים.

על בסיס הטבלה הזו וטבלאות עם דגימות יומיות מGoogle Trend-s, ביצענו לכל טבלה של דגימות יומיות שיערוך של כמות החיפושים המלאה והלא מנורמלת של כל דגימה יומית, תחת ההנחה כי השיערוך שיצא הוא קירוב טוב מספיק לכל הדגימות היומיות.

לאחר מכן חיברנו את הטבלאות היומיות יחד להיות טבלה אחת רציפה של דגימות יומיות לאורך שנים.

היות וההנחה כי השיערוך שיצא הוא מדוייק מספיק היא הנחה יחסית חלשה, ביצענו את התהליך הזה פעמיים, פעם אחת כאשר כל הטבלאות מתחילת 5 השנים ועד סופן מיובאות בתקופות זמן של 38 שבועות ופעם שנייה כאשר כל הטבלאות מתחילת 5 השנים ועד סופן מיובאות בתקופות זמן של 38 שבועות, פרט לטבלה הראשונה אשר מיובאת בתקופת זמן של 19 שבועות. את התוצאות של שני התהליכים הללו מיצענו לכל תאריך בהתאמה, כך שהערכים שהתקבלו בנתונים קרובים יותר אחד לשני בנקודות התפר של הטבלאות שחיברנו.

יש לציין כי בשיטה זו היה ניתן לדייק את השיערוך עוד יותר, ואפילו להקטין את פערי הזמן בין דגימה לדגימה, אך לצורך פרוייקט זה החלטנו להשתמש ברמת דיוק כזו.

לסיום נרמלנו את הערכים שיצאו לנו לטווח שבין 0 ל-100 לפי המקסימום והמינימום בכל טבלה בצורה הבאה- לקחנו בכל טבלה את המרחק בין המקסימום למינימום, ובמקום כל ערך בטבלה הצבנו את המרחק שלו מהמינימום חלקי המרחק של המקסימום מהמינימום.

כדי למנוע טעויות חישוב בהמשך כמו חלוקה ב-0, הצבנו בכל דגימה בה הערך הוא 0 (בדרך כלל רק דגימה אחת) את הערך 0.1 (ערך שעל סמך הסתכלות שלנו על הדאטה קבענו כי הוא מספיק קטן ביחס לשאר הערכים, אבל לא יגרום לבעיות נומריות).

את הנתונים על מחירי המטבעות ייבאנו מ-Binance.

הדאטה מ-Binance כלל מחירי פתיחה וסגירה יומיים, את המחיר הכי גבוה והכי נמוך לכל יום ואת נפך המסחר בכל יום.

היות ואותנו עניינו התשואות היומיות של כל מטבע, חישבנו עבור כל מטבע את ה-log returnes היומי, כלומר לקחנו את המנה של ערך הסגירה של המטבע ביום מסויים חלקי ערך הסגירה של אותו מטבע ביום וקודם, והוצאנו לה log.

**תיאור מפורט של המודל והאסטרטגיה**

האסטרטגיה שלנו מבוססת כאמור על אינדיקציות שמבוססות על דאטה מחיפושים בגוגל, לכן, האינדיקציה שלנו לקנות מטבע מסויים הייתה קפיצה בחיפושים של אותו מטבע, והאינדיקציה שלנו למכור הייתה ירידה בחיפושים...

**Backtesting**

המודל שלנו התבסס על Granger Causality בין הדאטה מ-Google Trends לגבי חיפושי שם של מטבע לבין log returns של אותו מטבע, דאטה המופק מ-Binance.

תיאור המודל: בהתחלה חילקנו את הדאטה שלנו ל-train ו-test, מצאנו עבור ה-train Granger Causality עבור כל אחד מהמטבעות שבדקנו והחלטנו באילו מטבעות אנחנו סוחרים.

לאחר מכן, ביצענו greed search עבור ההיפר-פרמטרים הבאים: (היפרפרמטרים)

על מנת למצוא מקסימום ל(מדד)

לאחר מכן בדקנו עבור ההיפר-פרמטרים שמצאנו מה התוצאות שלהם עבור ה-test.

בדקנו את אותם המדדים גם עבור תוצאות המודל שלנו ב-test.

ערכנו השוואה בין האסטרטגיה שלנו לבין אסטרטגיית buy & hold על אותם המטבעות כדי לבודד את השפעת האסטרטגיה שלנו מהגידול בערך המטבעות שבהם סחרנו.

**תוצאות**

(להכניס מדדים של שתי האסטרטגיות)

כפי שניתן לראות האסטרטגיה שלנו לא עבדה.

(להכניס רעיונות לאיך לגרום לא אולי לעבוד בעתיד).