**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 המטבעות הקריפטוגרפיים הוא שוק מבוסס הייפ...

Google Trends הוא כלי של גוגל שבעזרתו ניתן לקבל מידע על כמות חיפושים בגוגל לאורך זמן לפי מילת מפתח...

הרעיון הכללי מאחורי האסטרטגיה שלנו היה שימוש בדאטה של חיפושים בגוגל, אותו לקחנו מ-Google Trends כאינדיקטור להייפ.

**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.  
שני הקשיים הראשונים היו קשיים טכניים משום שנבעו בעיקר מהיכולת של המכשירים שברשותנו, ישנם שני דרכים (שאנחנו מכירים) שדרכם ניתן להשיג את הנתונים אותם אנחנו צריכים, הראשונה היא ישירות מהאתר של Google Trends, ניתן להוריד כקובץ csv את כמויות החיפושים המנורמלות לכל תקופה מסויימת המוגדרת על ידי המשתמש באתר, דרך שנייה היא באמצעות api לא רשמי של Google Trends אשר מוריד את אותם קבצי csv ושומר אותם כdata frame בפייטון לשימוש בקוד.

קושי ראשון נבע מתוך בעיות עם השירות עצמו שGoogle Trends נתן (השירות הפסיק לעבוד באופן זמני, חסימה זמנית מחשש לתקיפות סייבר...).

קושי שני נבע מתוך מחסור בחיבור אינטרנט יציב שיאפשר לפנות ולקבל מאותו שירות שGoogle Trends נותן באופן עקבי.

הדרך שלנו להתמודד שני אלו הייתה יצירת פונקציה אשר בנוסף לשימוש בapi שומרת מקומית את טבלאות הנתונים הרצויות, והפעלת הפונקציה הזו ברווחי זמן סבירים כל זאת על מחשב אשר מחובר מקומית לאינטרנט (מחובר בכבל LAN לנטב האינטרנט), בכך כלל הנתונים שהצטרכנו נעשו נגישים לנו בטווח זמן סביר ויחסית קצר.

קושי נוסף אשר מתבטא בשני חלקים היו קושי שיטתי שנבע בעיקרו מתוך הפורמט בו הנתונים שייבאנו מGoogle Trends מוגדר ע"י שירות זה.

החלק הראשון בקושי זה היה שכאשר מייבאים טבלה של כמויות חיפושים של מילה מסויימת או צירוף מילים מסויים מGoogle Trends הטבלה שמתקבלת מכילה עד 270 דגימות והפרשיי הזמן בין כל דגימה קבוע מראש, במילים אחרות, ניתן היה להגדיר תקופת זמן שמתחילה מתאריך מסוים ונגמרת בתאריך אחר כלשהו, אך לא ניתן היה להגדיר את הפרשי הזמן בין הדגימות שמקבלים (כשקובעים תקופת זמן של שנה 🡨 הדגימות בטבלה המוחזרת יהיו שבועיות, כשקובעים תקופת זמן של 10 שנים 🡨 הדגימות בטבלה המוחזרת יהיו חודשיות), לאחר קצת ניסוי וטעייה מצאנו כי כדי לקבל את כמות הנתונים היומיים המקסימלית צריך להגדיר תקופת זמן של 38 שבועות.

חלק שני בקושי היה שאותם נתונים המוחזרים בכל טבלה מנורמלים לערכים בין 0 עד 100 בשיטה פנימית של Google שלא יכולנו למצוא מקור מהימן מספיק המביע איך הוא מחושב ובנוסף אותו נרמול ייחודי לכל טבלה בנפרד, בכך לא הייתה לנו דרך אשר בוודאות נכונה שתאפשר לנו לחבר כמה טבלאות אחת אחרי השנייה.

כדי להתגבר על הקושי השתמשנו בשירות נוסף שמצאנו הנקרא Glimpse, Glimpse הוא תוסף לדפדפן Google Chrome אשר "מתלבש" על האתר והשירות שGoogle Trends נותן, מה שGlimpse נותן לנו בעצם זה שיערוך של כמות החיפושים המלאה והלא מנורמלת שמתבטאת בGoogle Trends ובהתאם גם כן מאפשרת להוריד את הנתונים הללו כקובץ csv.

לכל מטבע שנכלל במודל שלנו הורדנו מGlimpse טבלה בטווח זמן של 5 השנים האחרונות, משמע ייבאנו לכל מטבע טבלה עם 262 דגימות שבועיות אשר כל דגימה מהווה כמות חיפושים משוערכת, מלאה ולא מנורמלת של שבוע ימים.

על בסיס הטבלה הזו וטבלאות עם דגימות יומיות מGoogle Trends, ביצענו לכל טבלה של דגימות יומיות שיערוך של כמות החיפושים המלאה והלא מנורמלת של כל דגימה יומית, ותחת ההנחה כי השיערוך שיצא הוא מדוייק מספיק לכל הדגימות היומיות ובפרט בדגימות שבקצוות של הטבלאות היומיות 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 סופם מיובאות בתקופות זמן של 38 שבועות ופעם שנייה כאשר כל הטבלאות מתחילת 5 השנים ועד סופם 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 הערכים שמתקבלים בנתונים פחות שונים אחד מהשני בנקודות התפר של הטבלאות שחיברנו.

יש לציין שדרך שיטה 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 ל100 וכדי למנוע טעויות חישוב בהמשך כמו חלוקה ב0 הצבנו בכל דגימה בה הערך הוא 0 (בדרך כלל רק דגימה אחת) את הערך 0.1.

**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).