



Frederikke Malou Larsen

Professionsbachelor i dataanalyse

CPHBusiness

Bachelorprojekt

StudieID: cph-fl103

Hold: CPH-l-dal-da-0921a

Vejleder: Thorbjørn Baum

04-01-2023

Anslag: 47.615

Indholdsfortegnelse:

Resume	3
Indledning	5
Baggrund	5
Problemformulering	5
Underspørgsmål	6
Rammeafsnit	7
Beskrivelse af virksomheden	7
Kryptovaluta	7
Twitter	8
Afgrænsning	9
Kryptoevents i slutningen af 2022	9
Twitter	10
Bitcoin hashtag	11
Metode	12
Data retrieval	12
Twitter API	12
Databaser	13
Firis database	13
Opbevaring af indhentede tweets	14
Modeller	15
Sentiment analyse	15
R-Pakker	15
Rtweet	16
Httr	16
Mongolite	16
Syuzhet	17
Analyse	18
Hvordan kan Twitters API anvendes til at indhente brugbart data vedrørende stemningen omkring krypto?	18
Delkonklusion	23
Hvilke forholdsregler skal man gøre sig, når man bruger ekstern data fra Twitter?	23
Bots	23
Manglende data:	24
Delkonklusion	25
Kan machine learning modeller anvendes til at finde en korrelation mellem sentimentet på Twitter og Firis handelsaktivitet?	26
Multipel Lineær Regression	26
Træmodeller	30
Delkonklusion	33

Hvordan kommunikeres eventuelle findings fra undersøgelsen bedst til virksomheden samt dennes stakeholders?	33
Konklusion	35
Perspektivering	35
Enden på (Krypto)Twitter?	36
Kryptos fremtid	36
Yderlige og fremtidige analyser	38
Langsigtet analyse	38
Automatisering	38
Analyse af andre kryptovalutaer	38
Bibliografi	39
Bøger	39
Hjemmesider	39
Bilag	42

Resume

This final bachelor's project in data analysis investigates the relationship between what is written on the social media platform Twitter about cryptocurrency, and the trade history of cryptocurrencies in Norway on the Norwegian crypto exchange Firi, specifically in November and December 2022.

Twitter's API was used to collect tweets daily from mid-November to the end of December 2022, focusing on the keywords "crypto", "krypto" (in Norwegian) and "Bitcoin". Bitcoin was the first cryptocurrency, and still holds the largest market cap. Furthermore, Bitcoin is, on average, the most traded coin on Firi's exchange. For these reasons, Bitcoin was selected as a separate keyword.

This resulted in approximately 270.000, 600, and 174.000 tweets about crypto, krypto, and Bitcoin respectively. All tweets, together with the timestamp of when they were written were stored in a local database.

Natural language processing was then used to subtract the sentiment, which is whether the text is negative or positive, from each tweet. Each tweet was then allocated a score between -10 and +10. After assigning each tweet a sentiment score, the mean score for each day was calculated for each keyword.

This analysis showed that during the biggest events in the crypto industry, which was FTX CEO Sam Bankman-Fried's arrest and subsequent release on bail, the sentiment score shifted in both directions. First negatively during the arrest, but later positively during his release. Additionally, over the Christmas holiday, the score shifted in a positive direction again.

A multiple linear regression was then used to find correlations between the trade volume in Norway in the same period, and the sentiment scores of the three keywords. This model showed that the sentiment of crypto is significant to the volume of trade on Firi's platform. The model also showed a R^2 score of 0,3773, meaning that around 38% of the trade volume can be explained by the independent variables (sentiment of the keywords).

The same model was then used to find correlations between the trade of only Bitcoin, with the same independent variables. This showed similar results, but with weaker significance than the previous iteration.

Another machine learning model, regression trees, was then used to predict the trade volume with the same independent variables as above. This model was chosen because the results are more easily interpreted than those of the linear regression, especially for employees of a company or their stakeholders who have little or no knowledge of statistics or machine learning.

The analysis shows that trade volume in Norway on Firi's platform decreases when there is a positive sentiment score on tweets about cryptocurrency in November and December 2022.

It is then discussed how to expand the analysis to include longer time periods or other cryptocurrencies for better yielding results. Additionally, the future of both cryptocurrency, Twitter as a social media platform to discuss and share news about cryptocurrency, and blockchain is reviewed.

Indledning

Baggrund

Jeg var i efteråret 2022 i praktik hos den norske kryptovalutabørs Firi, i deres nystartede danske afdeling, som dataanalytiker. Her fik jeg blandt andet mulighed for at anvende mine færdigheder inden for Natural Language Processing, til at lave sentiment analyse af størstedelen af Firis danske artikler på deres hjemmeside. Dette valgte jeg at gøre, da der gennem praktikforløbet blev lagt meget vægt på SEO, og at oversætte norske artikler til dansk.

Ideen til analysen kom da jeg skrev ud til alle medarbejdere vedrørende ideer til mit bachelorprojekt på Slack¹. Jeg ønskede at lave en analyse, hvor jeg kunne anvende machine learning modeller til at forudsige en relevant problemstilling for virksomheden. Firis CEO, Thug Hoang, vendte tilbage med en del ideer til projektet².

Et af Hoangs forslag var “Hvor tidlig kan man se på markedssentimentet (på f.eks Twitter) før markedet faktisk faller eller at prisene går opp.”

Gennem min tid hos Firi har jeg lært en masse omkring både kryptovalutabranschen og blockchain. Jeg har derigennem fået et indblik i, hvor stor en rolle sociale medier og online fællesskaber betyder for branchen. Jeg valgte derfor at undersøge kryptosegmentet på Twitter nærmere. Dog viste dette sig at være mere besværligt end først antaget, grundet Twitters APIs begrænsninger. Jeg endte derfor med følgende problemformulering, samt underspørgsmål til yderligere at undersøge emnet:

Problemformulering

Hvordan er stemningen omkring kryptovaluta i slutningen af 2022?

¹ Slack er en kommunikationsplatform anvendt af virksomheden.

² Se Bilag I

Underspørgsmål

- Hvordan kan Twitters API anvendes til at indhente brugbart data vedrørende stemningen omkring krypto?
- Hvilke forholdsregler skal man gøre sig, når man bruger ekstern data fra Twitter?
- Kan machine learning modeller anvendes til at finde en korrelation mellem sentimentet på Twitter og Firis handelsaktivitet?
- Hvordan kommunikeres eventuelle findings fra undersøgelsen bedst til virksomheden samt dennes stakeholders?

Udgangspunktet for opgaven er en undersøgelse af en formodet sammenhæng mellem hvad der bliver skrevet om kryptovaluta generelt, samt kryptovalutaen Bitcoin specifikt, på det sociale medie Twitter, og handelsaktiviteten af kryptovalutaer hos den norske kryptobørs Firi. Fokus for iagttagelserne er at klarlægge hvordan stemningen er omkring kryptovaluta i slutningen af 2022. Opgaven analyserer sentimentet omkring kryptovaluta, specifik på Twitter, i de sidste måneder af 2022, og om hvorvidt dette korrelerer med køb og salg af kryptovaluta i Norge på Firis handelsplatform.

Formålet med analysen er at kunne forudsige handelsaktiviteten på Firis handelsplatform, med henblik på at kunne effektivisere og målrette virksomhedens markedsføring af de forskellige kryptovalutaer.

I et forsøg på at svare på ovenstående, vil jeg komme ind på følgende underspørgsmål i løbet af analysen: Hvordan kan Twitters API anvendes til at indhente brugbart data vedrørende stemningen omkring krypto? Hvilke forholdsregler skal man gøre sig når man bruger ekstern data fra Twitter? Kan machine learning modeller anvendes til at finde en korrelation mellem sentimentet på Twitter og Firis handelsaktivitet? Samt hvordan kommunikeres eventuelle findings fra undersøgelsen bedst til virksomheden samt dennes stakeholders.

Rammeafsnit

Beskrivelse af virksomheden

Firi (tidligere MiraiEx) er en norsk kryptovalutabørs baseret i Oslo, der blev grundlagt i 2017 af Thuc Tuan Hoang (nuværende CEO) og Øyvind Kvanes. Formålet med Firi var at introducere en lokal kryptovalutabørs, for at gøre det nemt og sikkert at handle med kryptovaluta i Norge med norske kroner.

Virksomheden har i dag omkring 35 medarbejdere og over 160.000 verificerede brugere, hvilket gør den til den største kryptovalutabørs i Norden.

Firi blev i 2019 registreret hos det norske Finanstilsyn som tjenesteudbyder af veksling og opbevaring af virtuel valuta, og hos det danske Finanstilsyn i januar 2022 (Firi, n.d., "Om os," para. 1). I marts 2022 ansatte Firi deres danske landechef og lancerede officielt i Danmark i juni 2022. Derudover er der planer om at lancere i Sverige i 2023.

Kryptovaluta

Kryptovaluta er en digital valuta bygget på blockchain teknologi. En blockchain er en ny type database med en specifik netværksarkitektur. Databasen er distribueret over flere computere, og er dermed decentraliseret.

I forhold til kryptovaluta, fungerer en blockchain som en digital regnskabsbog, hvorpå alle oplysninger omkring kryptovalutaen opbevares. Oplysninger på blockchainen er sikret gennem kryptografi, deraf navnet kryptovaluta.

Idet en blockchain er decentral, er kryptovaluta ligeledes decentral, og der er derfor ingen central enhed som udsteder eller verificerer valutaerne. Kryptovaluta gør det derfor også muligt for alle aktører at være anonyme.

Kryptovaluta kan anvendes som et alternativ til fiat-valutaer³, idet den kan bruges til at købe varer og tjenester, men den har nogle unikke egenskaber, der adskiller den fra andre valutaer. En anden forskel er, at kryptovaluta ofte anvendes som investeringsobjekt, da den kan svinge i værdi, blandt andet på grund af udbud og efterspørgsel på markedet. Dette kan føre til høje afkast for dem, der investerer i kryptovaluta, men det kan også medføre store tab, hvis værdien falder.

Der findes mange forskellige typer af kryptovaluta, heriblandt Bitcoin, Ethereum og XRP, hver med deres egne unikke egenskaber. Kryptovaluta kan købes og sælges på kryptobørser eller via peer-to-peer-transaktioner, og den kan også opbevares i en digital wallet, der fungerer på samme måde som en traditionel bankkonto.

Selv om kryptovaluta stadig er et relativt nyt fænomen, har den vundet stor popularitet de senere år, og er blevet anerkendt af mange store virksomheder og institutioner som et legitimt betalingsmiddel. Der er dog stadig mange uafklarede spørgsmål omkring regulering og lovgivning af kryptovaluta, og det er vigtigt at være opmærksom på de risici, der følger med at investere i denne form for valuta (Holten, Å. 2022).

Twitter

Twitter er en social medieplatform, der blev lanceret i 2006. Det er en mikroblogging tjeneste, hvor brugere kan sende og læse korte beskeder kaldet "tweets", der har en maksimal længde på 280 tegn. Tweets kan indeholde tekst, billeder, videoer og links til andre websteder.

Twitter er blevet en vigtig kilde til nyheder og information for mange mennesker, da tweets fra journalister, politikere, virksomheder, og andre offentlige figurer ofte går viralt og når en bred målgruppe. Det har også været et værdifuldt værktøj til at sprede budskaber, og organisere sociale bevægelser.

På samme tid har Twitter også været genstand for kritik for manglende handling på stødende og upassende indhold, herunder hadefulde tale, og spredning af misinformation.

Virksomheden har løbende arbejdet på at forbedre deres politikker og procedurer, for at bekæmpe denne type indhold, men der er stadig udfordringer.

³ Fiat er lovligt betalingsmidler udstedt af en centralbank f.eks danske kroner eller amerikanske dollar.

I dag har Twitter omkring 330 millioner månedlige aktive brugere på verdensplan, og er en af de mest populære sociale medieplatforme i verden (Twitter n.d).

Afgrænsning

Kryptoevents i slutningen af 2022

Analysen vil tage udgangspunkt i de sidste to måneder af 2022. Dette skyldes til dels projektets starttidspunkt, samt de begrænsninger der er ved brug af Twitters API med den adgang jeg har fået opsat. Derudover har jeg anvendt funktion `search_tweets` fra pakken “rtweet” til at indhente tweets ud fra bestemte søgeord. Denne funktion kan kun indhente data op til 6-9 dage tilbage i tiden (RDocumentation (n.d). rtweet). Dog gør mængden af tweets som indeholder mine udvalgte søgeord, at jeg kun har haft mulighed for at indhente data fra to dage af gangen.

Ydermere har de sidste måneder af 2022 budt på flere store begivenheder inden for kryptoverden. De makroøkonomiske udfordringer verden har været udsat for i 2022, har også sat sine spor i kryptoindustrien.

Vi har været vidne til flere konkurser (Celcius, FTX, Voyager m.fl) som ikke kun har været drevet af de makroøkonomiske forhold, men også været resultatet af manglende ansvarlighed, ikke bæredygtige forretningsmodeller i lavkonjunktur situationer, og angiveligt direkte svindel. Konkurser der på mange måder har sat fokus på industrien og den mangelfulde regulering på området.

Et af de største events i slutningen af 2022 var kryptobørsen FTXs konkurs den 11. November (De, N. 2022,), samt grundlæggeren Sam Bankman-Frieds efterfølgende arrestering i Bahamas d. 13. December (Aljazeera, 2022).

Twitter

Twitter har siden “The Bitcoin Block Size Wars” i 2016-17, været hyppigt anvendt inden for krypto- og blockchain verden. “The Bitcoin Block Size Wars” var en omstridt debat inden for Bitcoin-fællesskabet, der blev diskuteret bredt på sociale medieplatforme, inklusiv Twitter. Mange enkeltpersoner og organisationer involveret i debatten brugte Twitter til at dele deres meninger, opdateringer, og analyser om problemet.

Twitter spillede også en rolle i at forstærke debatten, og forme offentlighedens opfattelse af de forskellige forslag. For eksempel brugte nogle fremtrædende personer i Bitcoin-fællesskabet Twitter til at fremme deres synspunkter om blokstørrelse spørgsmålet, og til at mobilisere tilhængere. Dette bidrog til debattens polariserende karakter og de stærke følelser, den skabte blandt de involverede.

Det er dog værd at bemærke, at mens Twitter spillede en vigtig rolle i “The Bitcoin Block Size Wars” i 2016-2017, var det blot en af de mange platforme, der blev brugt til at diskutere spørgsmålet. Debatten fandt også sted på fora, blogs og andre sociale medieplatforme, samt i mere formelle rammer såsom konferencer og møder (Bitcoin. 2019).

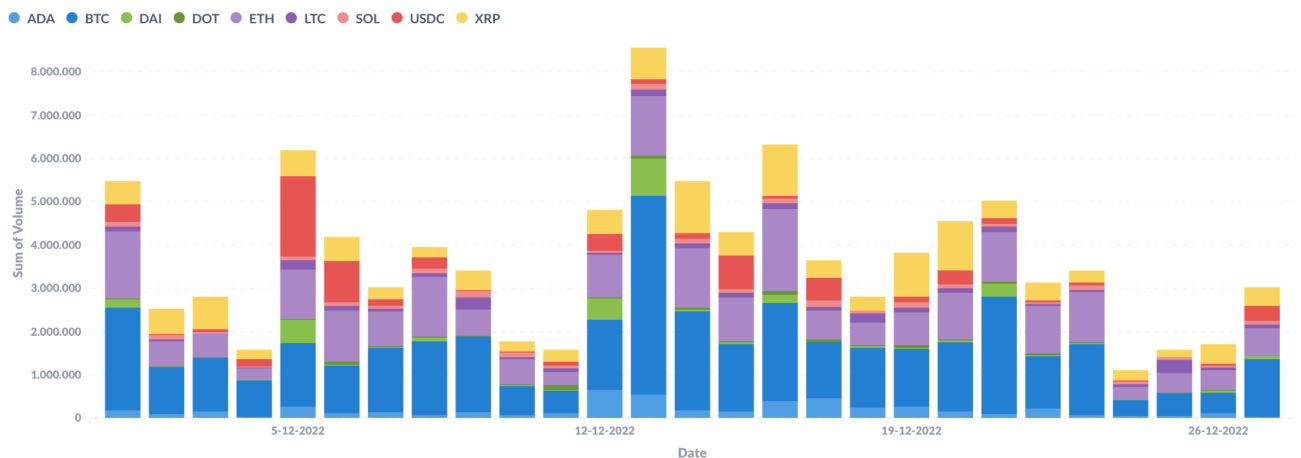
Twitter gør det altså muligt for brugerne at dele information og holdninger om forskellige kryptovalutaer, samt relaterede emner. Mange indflydelsesrige figurer inden for kryptoindustrien bruger Twitter til at dele deres tanker, indsigter, og opdateringer omkring miljøet. Dette gør Twitter til et værdifuldt værktøj til at holde sig opdateret omkring de seneste udvikler og tendenser i krypto- og blockchain verden.

Sociale medier som Reddit og Discord er også hyppigt anvendt inden for kryptomiljøet. Jeg har valgt ikke at bruge data fra disse platforme, da de til dels er mindre tilgængelige (Discord) eller organiseret i tråde (Reddit), sammenlignet med Twitter som har en API tilgængelig til indhentning af tweets (mere herom i metodeafsnittet).

Bitcoin hashtag

Udover at indhente alle tweets som indeholder ordet “crypto”, har jeg valgt at indhente alle tweets som indeholder ordet “bitcoin”.

Dette har jeg til dels valgt at gøre, da Bitcoin var den første kryptovaluta, og derfor delvis stadigvæk er synonym med kryptovaluta. Derudover har Bitcoin fortsat den største markedsværdi på tværs af alle kryptovalutaer (Coingecko. n.d), og på Firis platform.



Figuren viser handelsvolumen af de forskellige kryptovalutaer Firi tilbyder per dag i December (27/12). Som grafen viser er der et stort udsving i handelsvolumen fra dag til dag, men Bitcoin har den gennemsnitlige største volumen på tværs af måneden.

Metode

Data retrieval

Twitter API

For at få hentet tweets ned fra Twitter, har jeg anvendt Twitters API. En API (Application Programming Interface) er det værktøj, der gør en hjemmesides data tilgængelig for en computer. Gennem den kan computeren se og redigere data, ligesom en person kan ved at indlæse sider og indsende formularer.

Twitters API giver udviklere mulighed for at få adgang til, og interagere med, Twitters funktionaliteter. Med Twitters API kan udviklere bygge applikationer, der kan udføre forskellige opgaver, såsom:

- Søge efter og hente tweets baseret på søgeord eller hashtags
- Indhentning af brugeroplysninger og tidslinjer
- Sende tweets, svar og direkte beskeder
- Følge og blokere brugere
- Like og dislike tweets

Twitters API er en REST (Representational State Transfer) API, hvilket betyder, at den bruger HTTP-anmodninger til at hente og manipulere data.

For at anvende Twitters API skal udviklere registrere sig, hvilket giver adgang til adgangsnøgler og tokens, som kan bruges til at autentificere deres API-anmodninger. Twitter API'en har hastighedsgrænser, der begrænser antallet af anmodninger, som udviklere kan lave inden for en bestemt tidsperiode, så det er vigtigt at designe applikationer, der er effektive og kun laver det nødvendige antal anmodninger (Twitter Developer. n.d).

Databaser

De to hyppigst anvendte database typer er SQL og NoSQL databaser.

SQL-databaser (Structured Query Language) er baseret på en relationel model, hvilket betyder, at de gemmer data i tabeller, der er relateret til hinanden gennem nøgler.

SQL-databaser bruger et specifikt sprog kaldet SQL til at oprette, manipulere og forespørge dataene i databasen.

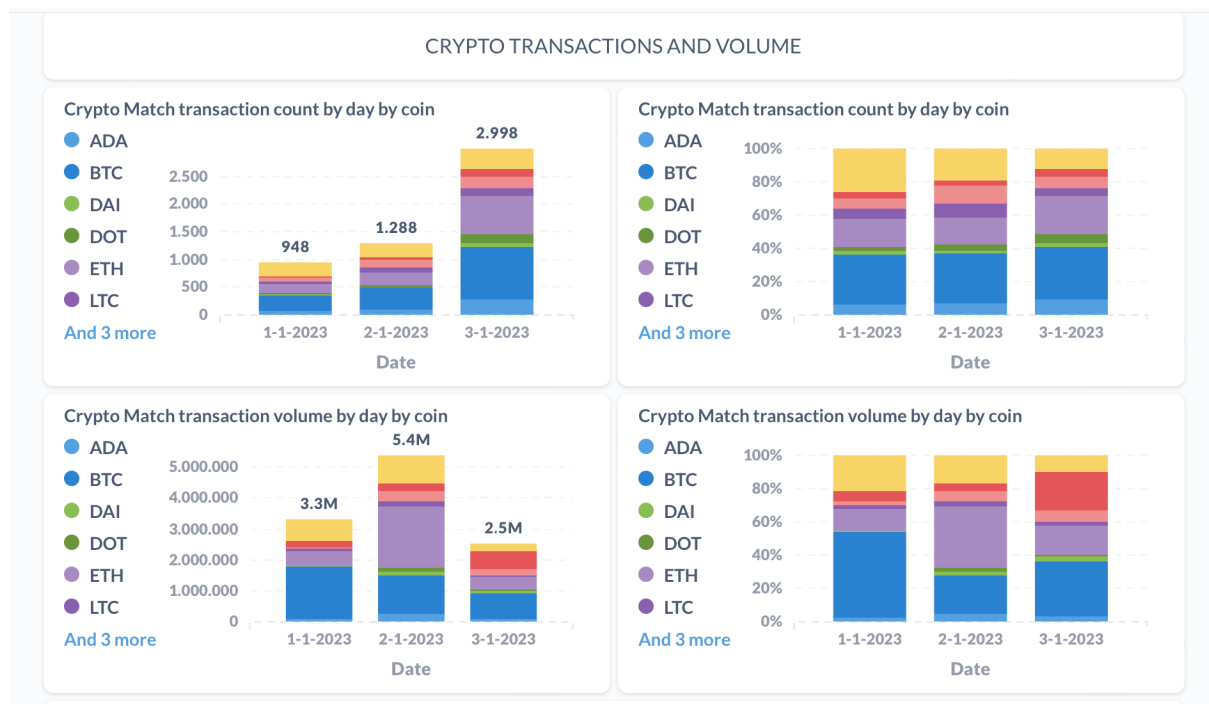
NoSQL-databaser er en nyere type database, der ikke bruger en relationel model. I stedet anvendes en række forskellige datamodeller, såsom nøgleværdi, dokument, kolonnefamilie og graf. NoSQL-databaser er generelt mere fleksible og skalerbare end SQL-databaser, og de er bedre egnede til at håndtere store mængder ustrukturerede data.

Firis database

Alt data fra Firi vedrørende deres kunder, handelsaktivitet osv. er samlet i én database. Firis database er en PostgreSQL database, som administreres af Metabase. PostgreSQL anses som værende den mest avancerede open source-database, og et mere robust

SQL-databasealternativ til MySQL. Postgres er open source, og alle funktioner er tilgængelige uden en betalt licens (Linode, 2022). Metabase kan anvendes til flere typer databaser, men har den fordel, at det ikke kræver, at brugerne kender SQL for at lave queries. På Metabase er det også muligt at lave grafer og dashboards (Metabase. n.d).

Nedenfor ses et eksempel på et dashboard i Firis database:



Opbevaring af indhentede tweets

Jeg har behov for at kunne opbevare de mange tweets jeg indhenter et nemt tilgængeligt sted. Udover at opbevare dem lokalt på min computer, valgte jeg at opbevare dem i en MongoDB database. MongoDB er en dokumentdatabase, der gemmer data som JSON-dokumenter, for at give mere fleksibilitet til skalering og data forespørgsler. MongoDB er en løsning til opbevaring af ikke-relational data(NoSQL).

Denne tilgang giver mig mulighed for at begynde at oprette dokumenter, uden først at skulle etablere en dokumentstruktur.

Modeller

Sentiment analyse

Sentimentanalyse, eller opinion mining, er en teknik inden for machine learning, der gør det muligt at analysere menneskers følelser og holdninger i forhold til et bestemt emne. Dette kan gøres ved at scanne store mængder tekst, som for eksempel indlæg på sociale medier, nyhedsartikler eller anmeldelser, for at finde ud af, hvad folk generelt mener om et produkt, en virksomhed, eller et emne (Silge, J. & Robinson, D. 2022).

Der findes mange forskellige måder at udføre sentimentanalyse på, men den mest almindelige metode er ved hjælp af naturlig sprogbehandling (Natural Language Processing, NLP).

NLP-teknikker gør det muligt at identificere de følelser og holdninger, der ligger bag teksten, ved at analysere ordvalg og sætningsstruktur.

Ved at analysere følelser og holdninger kan virksomheder og organisationer forbedre deres produkter, tjenester og kommunikation for at tiltrække og tilfredsstille deres kunder og stakeholders. Det kan også være nyttigt for organisationer, der ønsker at måle effekten af deres kommunikation eller marketingstrategier.

Selv om sentimentanalyse kan være et brugbart værktøj til at måle og forstå følelser og holdninger, er det vigtigt at være opmærksom på, at denne teknik ikke er perfekt og kan have begrænsninger. Det er for eksempel muligt at sentimentanalysen ikke altid kan skelne mellem ironi, sarkasme, og andre former for figurativt sprog. Derfor bør en sentiment analyse altid bruges som et supplement til andre metoder, og ikke som en eneste kilde til viden og beslutninger.

R-Pakker

Rtweet

Pakken rtweet bruges til at forbinde til Twitters API i R. Med rtweet kan brugere søge efter og hente tweets, brugeroplysninger og andet data fra Twitter, samt poste tweets, svar og direkte beskeder .

For at kunne anvende rtweet skal man have en Twitter-udviklerkonto, og oprette en ny applikation. Dette giver adgangsnøgler og tokens, som kan bruges til at godkende API-anmodninger. Derefter bruges funktioner fra rtweet-pakken til at få adgang til Twitters API, og hente eller manipulere data⁴.

Httr

httr-pakken er en R-pakke, der giver funktioner til at arbejde med HTTP-anmodninger og -svar. HTTP (HyperText Transfer Protocol) er en protokol, der bruges til kommunikation mellem klienter (f.eks webbrowsere) og servere (såsom webservere). HTTP giver klienter mulighed for at sende anmodninger til servere, hente eller manipulere data, og for servere at sende svar tilbage til klienter⁵.

httr-pakken bruges ofte i kombination med andre R-pakker, der giver funktioner på højere niveau til at interagere med specifikke API'er eller tjenester, såsom rtweet-pakken til at få adgang til Twitters API.

Mongolite

Mongolite-pakken er en R-pakke, giver en nem interface til MongoDB-databasen. Det er nødvendigt at anvende en pakke, som f.eks Mongolite, for at f.eks kunne indsætte data direkte fra R til MongoDB⁶.

⁴ <https://github.com/ropensci/rtweet>

⁵ <https://github.com/r-lib/httr>

⁶ <https://www.rdocumentation.org/packages/mongolite/versions/2.7.0>

Syuzhet

Syuzhet-pakken er en R-pakke, der giver værktøjer til at udtrække følelser fra tekstdata.

Syuzhet-pakken indeholder flere funktioner til at udtrække sentiment fra tekstdata ved hjælp af forskellige sentiment-ordbøger og machine learning klassifikationer. Pakken gør det blandt andet muligt at udvinde følelser ved hjælp af machine learning klassifikatorer trænet på forskellige datasæt, normalisering og skalering af sentiment score, samt plotte og visualisere følelser over tid⁷.

Syuzhet-pakken kan lave sentiment analyse på både engelske og norske tekster, samt give en numerisk score frem for kun at rangere sætninger som positiv/negativ, hvilket er årsagen til at jeg har valgt at anvende denne model til at lave sentiment analyse med.

⁷ <https://github.com/mjockers/syuzhet>

Analyse

Hvordan kan Twitters API anvendes til at indhente brugbart data vedrørende stemningen omkring krypto?

Det første skridt for at kunne analysere data fra Twitter, var at oprette en udviklerkonto i Twitter Developer Portal, for at få adgang til deres API. For at kunne indhente et større antal tweets, ansøgte jeg om at få Elevated Access, hvilket krævede at jeg indsendte en skriftlig ansøgning som beskrev projektets formål.

Efter at have fået Elevated Access til Twitters API, oprettede jeg forbindelse til API'en i R ved hjælp af httr og rtweet-pakkerne. Herefter kunne jeg indhente tweets direkte i R ved at søge på bestemte søgeord. Jeg valgte her at opsætte tre forskellige forespørgsler (search queries):

- 1: Indhenter alle tweets som indeholder ordet “crypto” og er på engelsk. Ekskluderer retweets.
- 2: Indhenter alle tweets som indeholder ordet “krypto” og er på norsk. Ekskluderer retweets.
- 3: Indhenter alle tweets som indeholder ordet “Bitcoin” og er på engelsk. Ekskluderer retweets.

Disse search queries returnerer en dataframe med 43 forskellige variabler. Da jeg ikke er interesseret i andet end selve teksten og datoen, fjernede jeg de resterende 41 variabler inden jeg gemte dem i en lokal MongoDB database.

Der blev i November og December skrevet ca. 10.000 tweets dagligt på engelsk som indeholder ordene “Crypto” eller “Bitcoin”, og kun mellem 10-20 på norsk som indeholder “krypto”. Twitters API kan dermed, hvis man har udvalgt nogle populære søgeord, returnere en stor mængde data.

Det er også muligt at indhente tweets som kun indeholder bestemte hashtags, frem for søgeord. Jeg har dog valgt ikke at gøre dette, da ikke alle brugere på Twitter anvender hashtags, og dette derfor muligvis ikke ville give alle Tweets som indeholder de ønskede søgeord.

De indhentede tweets indeholder som nævnt ovenfor en variabel med dato for hvornår tweetet var skrevet i dansk lokal tid. Da handelsaktiviteten fra Firis platform er fordelt på dage, valgte jeg at fjerne klokkeslættet fra datasættet.

Kryptovalutamarkedet lukker ikke på samme måde, som traditionelle aktiemarkeder gør. Mens traditionelle aktiemarkeder, såsom New York Stock Exchange og London Stock Exchange, opererer efter en fast tidsplan og lukker for handel i slutningen af hver hverdag, er kryptovalutamarkedet åbent 24/7.

Dette skyldes, at kryptovalutaer, såsom Bitcoin og Ethereum, handles på decentraliserede børser, der opererer på et globalt netværk af computere. Disse børser har ikke en central placering, så de er i stand til at facilitere handel døgnet rundt.

En anden grund er, at kryptovalutaer er decentraliserede, hvilket betyder, at de ikke kontrolleres af en central myndighed såsom en regering eller en bank. Denne mangel på centralisering giver mulighed for mere fleksibilitet i, hvordan og hvornår handel kan foretages.

Samlet set giver den kontinuerlige handel med kryptovalutaer mulighed for større likviditet og tilgængelighed, sammenlignet med det traditionelle aktiemarkeder, men det betyder også, at handlende skal være opmærksomme på de risici, der er forbundet med handel på et marked i konstant forandring (Firi (n.d). Kjøp og salg kryptovaluta enkelt med markedsordre).

Tidsforskelle spiller derfor ikke nødvendigvis en væsentlig rolle for hvordan det norske markedet klarer eller udvikler sig, idet Firis kunder har mulighed for at handle døgnet rundt.

Jeg endte med at indhente tweets fra d. 17/11-2022 og frem til d. 31/12-2022.

Antallet af tweets endte med at være henholdsvis:

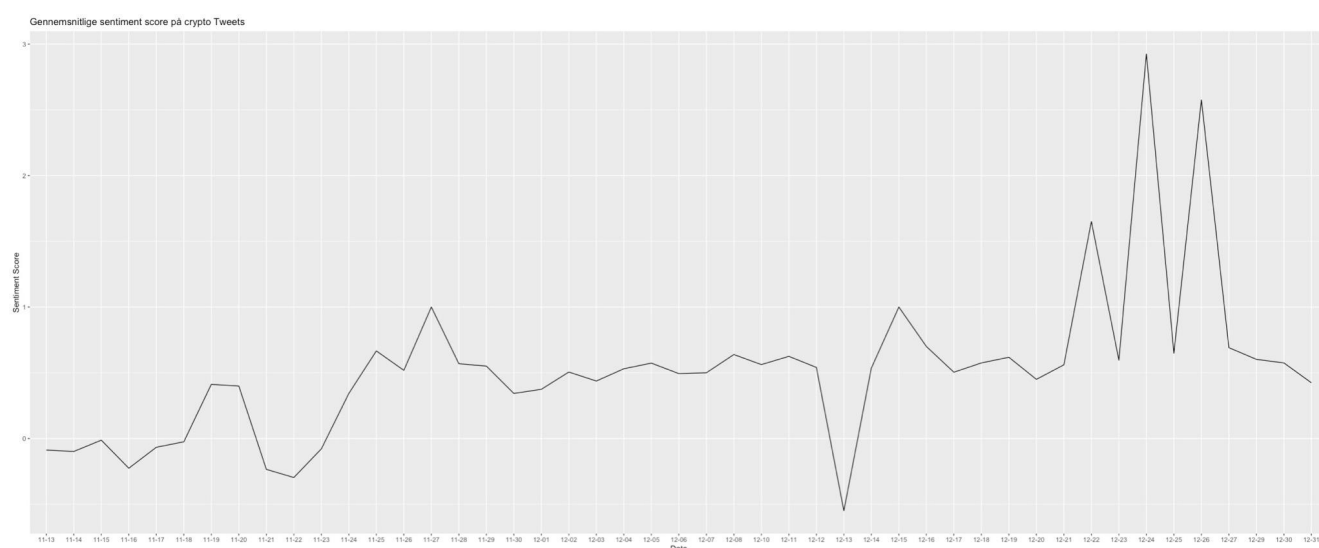
Keyword	Antal
Crypto	270.141
Krypto (norsk)	598
Bitcoin	174.423

Efter at have fjernet tidspunktet fra dato kolonnen, brugte jeg Syuzhet-pakken til at udregne en sentiment-score for hvert enkelt tweet. Dette returnerede en numerisk score mellem +10 og -10, hvoraf + 10 repræsenterer den mest positive score, og -10 den mest negative score. Jeg lavede herefter en gennemsnitlig score per dag for hvert af de tre keywords. Jeg opdagede her at jeg manglede data for en række datoer, især for tweets omkring Bitcoin, men også en data for en enkelt dag (d. 9/12) mangler for crypto. Denne udfordring uddybes i det næste afsnit.

Herefter plottede jeg den gennemsnitlige score for at kunne se udviklingen over tid for hver af de tre keywords.

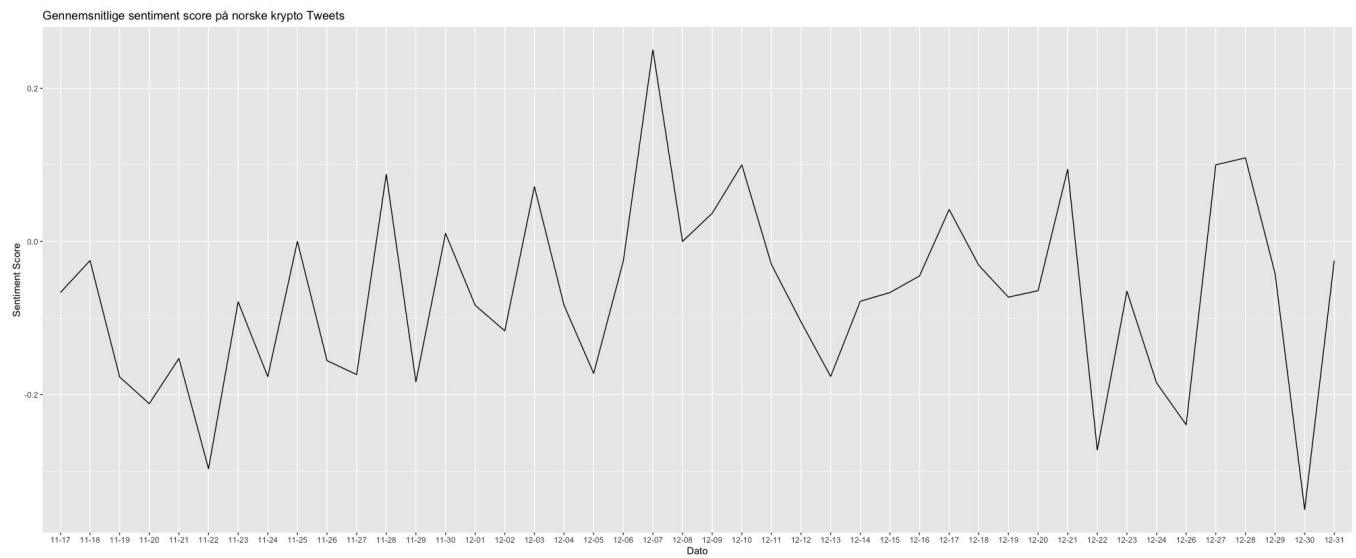
Gennemsnitlige sentiment score på crypto tweets:

Figur I:



Gennemsnitlige sentiment score på norske krypto tweets:

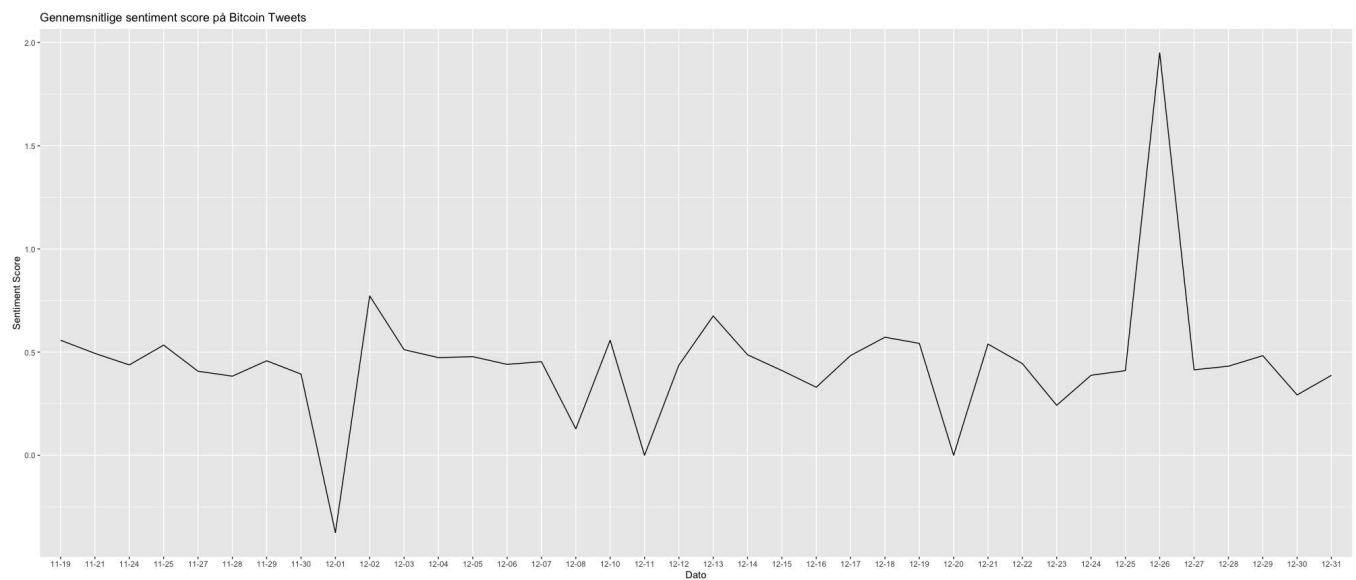
Figur II:



9

Gennemsnitlige sentiment score på Bitcoin tweets:

Figur III:



10

⁹ Egen tilblivelse

¹⁰ Egen tilblivelse

For at undersøge om de største begivenheder på kryptomarkedet afspejles i stemningen på Twitter, har jeg samlet alle tre sentiment scorer i én graf.

De største begivenheder på kryptomarkedet i tidsperioden 19. November til og med 31.

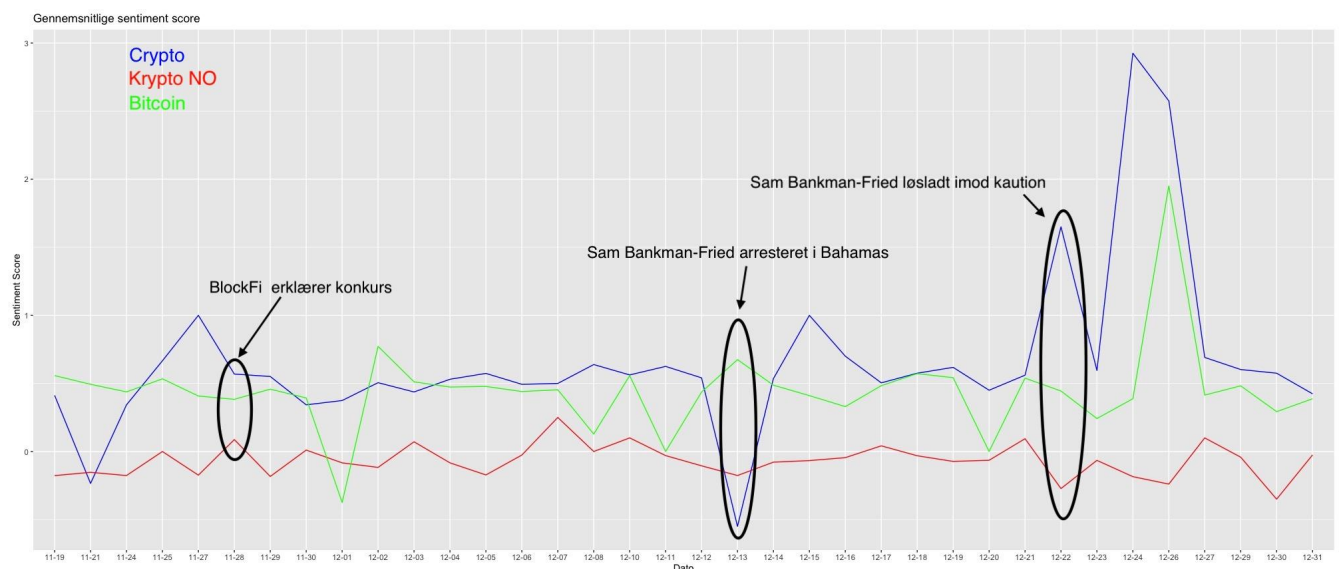
December 2022 var følgende:

28/11: Kryptovirksomheden BlockFi erklærer konkurs, som konsekvens af FTXs konkurs tidligere på måneden (Sigalos, M. 2022).

13/12: FTX CEO Sam Bankman-Fried arresteres i Bahamas.

22/12: Sam Bankman-Fried løslades mod kaution på 250 millioner USD (ABP News Bureau 2022)

Figur IV:



11

På ovenstående graf ses det hvordan der er store udsving i sentimentet omkring kryptovaluta netop på de to dage hvor Bankman-Fried henholdsvis arresteres og løslades. Dette giver en indikation af, at stemningen på Twitter, ihvertfald omkring kryptovaluta generelt, til en vis grad afspejler de begivenheder som forekommer i kryptoindustrien.

Det er dog vigtigt at huske på at kryptovaluta er uafhængige af en centralbank og er ikke direkte bundet til traditionelle fiat-valutaer, såsom den amerikanske dollar eller euro, dog med undtagelse af stablecoins¹². I stedet bestemmes værdien af en kryptovaluta af faktorer

¹¹ Egen tilblivelse

¹² Stablecoins er kryptovalutaer som følger værdien på Fiat eller andre kryptovalutaer, f.eks USDC.

som udbud og efterspørgsel, nyheder og markedsstemning, og den overordnede ydeevne af kryptovalutamarkedet.

Priserne på kryptovalutaer kan være volatile, og de er ikke nødvendigvis bundet til præstationen på det traditionelle aktiemarked. Faktisk kan priserne på kryptovalutaer nogle gange bevæge sig i den modsatte retning af aktiemarkedet. Der er dog nogle faktorer, der kan påvirke priserne på både traditionelle aktier og kryptovalutaer, såsom økonomiske forhold, regeringspolitikker og markedsspekulation.

Delkonklusion

Twitters API kan anvendes til at indhente store mængder tweets omhandlende bestemte keywords, eksempelvis crypto eller Bitcoin. Dog har API'en sine begrænsninger, især når det kommer til tweets på andre sprog, hvorfor datamængden af tweets på norsk er mindre. Ovenstående analyse viser at stemningen på Twitter svinger ved større begivenheder i kryptoindustrien. Dog følger de ikke nødvendigvis hinanden, som det for eksempel ses i Figur IV, falder sentimentet omkring crypto d. 13/12 hvor Bankman-Fried arresteres, hvorimod det stiger omkring Bitcoin samme dag.

Hvilke forholdsregler skal man gøre sig, når man bruger ekstern data fra Twitter?

Bots

Twitter har et ry for at have mange bots på deres platform, hvilket også har været genstand for kritik, især ved Musks overtagelse.

Bots kontrolleres typisk af enkeltpersoner eller organisationer med en specifik agenda, som ønsker at manipulere den offentlige mening og påvirke den politiske diskurs.

Twitter forsøger dog at bekæmpe spambots, blandt andet gennem brugen af algoritmer og machine learning til at identificere og fjerne dem fra platformen. Dette er en stor udfordring for Twitter at løse blandt andet grundet den hurtige udvikling af spambot-taktik og vanskeligheden ved at skelne mellem legitime og ondsindede konti.

Enkeltpersoner, såsom Elon Musk, har medhjulpet til at forstærke rækkevidden og indflydelsen af spambots ved at promovere deres indhold på platformen. Det tyder på, at spredningen af spambots på Twitter har haft en negativ indflydelse på kvaliteten af information og debat på platformen, og at det har bidraget til udbredelsen af misinformation og propaganda (Frenkel, S. 2022).

Der er potentielle konsekvenser ved at bruge bots på platformen, herunder konto suspendering eller opsigelse. Twitter tager en stærk holdning imod brugen af bots til ondsindede formål, såsom spamming eller manipulation af den offentlige mening, og vil fortsætte med at arbejde for at identificere og fjerne bots, der overtræder deres politikker (Roth, Y. & Pickles, N. 2020)

For at undgå at medtage bots så vidt som muligt i mine datasæt, da de muligvis kan give udsving i sentiment scoren, har jeg valgt at fjerne alle dupliserede tweets, altså steder hvor den samme tekst forekommer mere end én gang.

Manglende data:

Efter at have lavet den gennemsnitlige sentiment score per dag på alle tre keywords, opdagede jeg at der manglede en del datoer fra Bitcoin-datasættet. Ved yderligere gennemgang viste det sig at følgende datoer mangler fra datasættet:

17/11
18/11
20/11
22/11
23/11
26/11
9/12

Derudover mangler d. 9/12 også fra crypto-datasættet.

Twitters API gør det desværre kun muligt at hente data 7 dage tilbage i tiden, medmindre man som virksomhed tilkøber enterprise adgang dertil. Enterprise adgang inkluderer API'er til søgning og analyse af historisk Twitter-data, samt API'er til streaming af Twitter-data i realtid. Prisen på Twitter Enterprise API afhænger af den specifikke use-case og mængden af data, virksomheden skal have adgang til.

Årsagen til den manglende data er ukendt, men skyldes muligvis at jeg ikke har haft mulighed for at indhente tilstrækkeligt med data med den API adgang jeg har opsat. Der har ingen afbrydelser været i Twitters API¹³, så manglen på data skyldes min dataindsamling.

Det er muligt at anvende data augmentation til at generere data for de manglende datoer. Dette kræver dog både tid og ressourcer, som jeg ikke har haft adgang til gennem dette forløb. Ydermere ses det i Figur IV, hvordan der kan forekomme udsving i sentiment scoren,

¹³ <https://api.twitterstat.us/>

hvis der for eksempel er en større begivenhed inden for kryptoindustrien. Det kan derfor være risikabelt at gøre sig nogle antagelser i forhold til det manglende data, da kryptomarkedet og -industrien konstant er under udvikling.

Derudover viser den senere analyse at sentimentet omkring netop Bitcoin muligvis ikke er signifikant.

Delkonklusion

Twitter er desværre kendt for at have mange bots på deres platform, som kan manipluere eller disrupte den indsamlede data. Dette er yderst vigtigt at være opmærksom på, når man anvender data fra Twitter.

Dette har jeg forsøgt at tage højde for ved at fjerne duplikerede tweets fra min datasæt. Dette er dog ingen garanti for at bots ellers spam ikke er blevet medtaget i datasættet.

Twitters API gør det, med elevated access, kun muligt at indhente op til 7 dage tilbage i tiden. Dette er essentielt at tage højde for løbende, da man ellers ikke kan få adgang til evt. mistet data.

Kan machine learning modeller anvendes til at finde en korrelation mellem sentimentet på Twitter og Firis handelsaktivitet?

Machine Learning kan anvendes til at finde sammenhænge mellem variabler. En måde er at bruge lineær regression, som er en metode til at modellere den lineære sammenhæng mellem en afhængig variabel, og en eller flere uafhængige variabler.

Endelig kan machine learning også bruges til at identificere mønstre og relationer i data ved hjælp af teknikker som clustering eller dimension reduction. Disse metoder kan bruges til at identificere grupper af korrelerede variabler, eller til at reducere antallet af variabler i et datasæt, samtidig med at de vigtige relationer mellem disse bevares.

Machine learning er i stand til at håndtere store mængder data mere effektivt end mennesker, og kan identificere mønstre og relationer, som kan være svære for en person at opdage. Dette kan især være nyttigt, når man arbejder med datasæt, der er for store eller komplekse til, at en person kan analysere manuelt.

Ydermere kan det være mere objektivt at bruge machine learning til at identificere sammenhænge, end at stole på en persons subjektive fortolkning af dataen. Dette kan især være vigtigt, når man arbejder med følsomme eller kontroversielt data, hvor det er vigtigt at undgå bias eller forudantagelser.

Multipel Lineær Regression

Jeg har valgt at bruge en multipel lineær regression til at estimere hældningskoefficienterne til de uafhængige variabler, sådan at man kan forudsige værdien på den afhængige variabel. Multipel lineær regression er en udvidelse af simpel lineær regression, som bruges til at modellere forholdet mellem en enkelt uafhængig variabel og en afhængig.

Min afhængige variabel, y , er den samlede handelsvolumen i Firi Norge. Denne data har jeg indhentet fra Firis database.

Dataen kommer fra tabellen “Trades Day”. Jeg har derefter filtreret dataen så jeg kun får fra de relevante dage (17/11-31/12 2022), kun norske kunder, samt kun de handler som ikke er Fiat (altså kun kryptovaluta).

Trades Days, Filtered by Date

Data

Trades Day

Filter

Date between November 17, 2022 December 31, 2022

Is Fiat is false

Nationality is NO

+

14

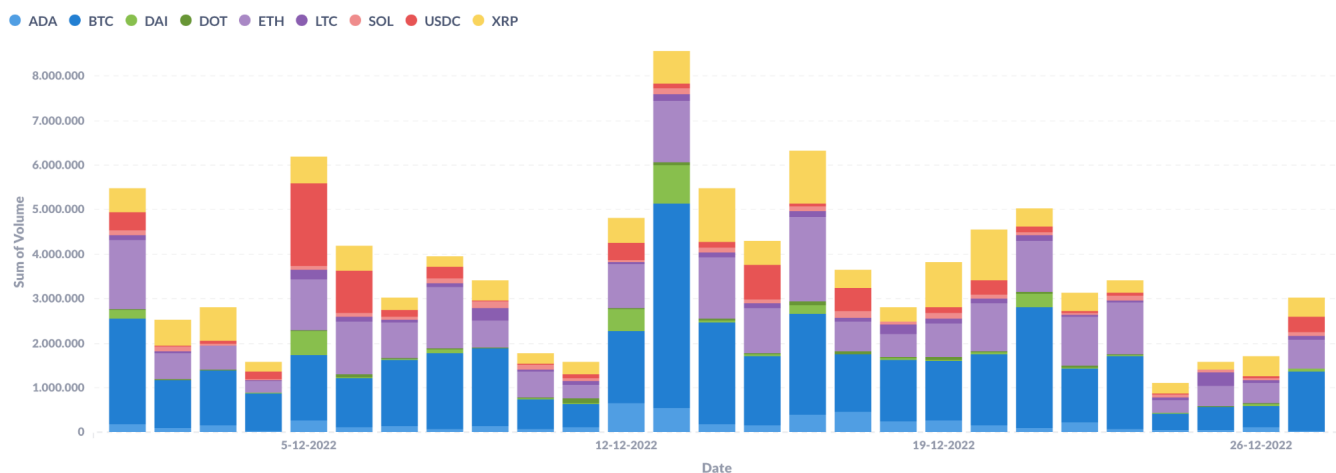
Denne query ser således ud i SQL:

```
SELECT "reports"."trades_day"."date" AS "date",
"reports"."trades_day"."currency_id" AS "currency_id",
"reports"."trades_day"."cost_type" AS "cost_type",
"reports"."trades_day"."cost_nok" AS "cost_nok",
"reports"."trades_day"."cost_volume" AS "cost_volume",
"reports"."trades_day"."is_fiat" AS "is_fiat",
"reports"."trades_day"."nationality" AS "nationality"
FROM "reports"."trades_day"
WHERE ("reports"."trades_day"."date" >= timestamp with time zone '2022-11-17
00:00:00.000+01:00'
AND "reports"."trades_day"."date" < timestamp with time zone '2023-01-01
00:00:00.000+01:00' AND "reports"."trades_day"."is_fiat" = FALSE AND
"reports"."trades_day"."nationality" = 'NO')
LIMIT 1048575
```

Dataen fra databasen er opdelt i de forskellige kryptovalutaer Firi tilbyder, som set herunder:

Figur V:

¹⁴ Screenshot fra Metabase



15

Jeg har derfor samlet alle valutaer, for at få én samlet volumen per dag, som jeg har defineret som min afhængige variabel.

Dernæst har jeg defineret mine uafhængige variabler, x , som værende henholdsvis:

x_1 = Crypto Sentiment

x_2 = Krypto NO Sentiment

x_3 = Bitcoin Sentiment

Disse variabler anvendte jeg til at lave en multipel lineær regression, for at undersøge om der er en sammenhæng mellem sentimentet på Twitter og handelsvolumen i Norge.

Resultatet af den multiple lineære regression (se Bilag II) viser at følgende p-værdier for variablerne:

Variabel	P-værdi
Intercept	$2,65 \cdot 10^{-12}$
x_1	0,000302
x_2	0,054320
x_3	0,9750515

¹⁵ Egen tilblivelse

Intercept er hvis alle de uafhængige variabler er 0, altså hvis sentimentet på alle tre keywords er fuldstændig neutral.

P-værdierne for de uafhængige variabler x_1 (crypto sentiment) og x_2 (krypto NO sentiment) er mindre end 0,05, altså er disse to variabler signifikante for udfaldet af den afhængige variabel (y), som er den daglige handelsvolumen.

Modellen har en R^2 på 0,3773. Dette betyder at omkring 38% af variansen i den afhængige variabel kan forklares af de uafhængige variabler.

Koefficienter for modellens uafhængige variabler er:

Variabel	Koefficient (NOK)
x_1	-1962054
x_2	-4597630
x_3	32841

Dette betyder at hvis x_1 , sentimentet omkring crypto på Twitter, stiger med 1, altså bliver mere positivt, falder handelsvolumen med 1.962.054 norske kroner.

Næste analyse er lavet kun på handelsvolumen på Bitcoin. Igen er handelsvolumen den afhængige variabel (y), og de uafhængige variabler (x) x_1 = Crypto Sentiment, x_2 = Krypto NO Sentiment, x_3 = Bitcoin Sentiment.

Resultatet af regressionen kan ses i Bilag III.

Denne model har følgende p-værdier:

Variabel	P-værdi
Intercept	$2,18 \cdot 10^{-9}$
x1	0,000829
x2	0,124097
x3	0,824780

Igen kan vi se at p-værdien for intercept er meget lille.

I denne regression er det kun x1, Crypto Sentimentet, som er signifikant for den afhængige variabel, handelsvolumen af Bitcoin.

Koefficienter for modellens uafhængige variabler er:

Variabel	Koefficient (NOK)
x1	-1021143
x2	-2077480
x3	112471

Ens med den forrige model, falder handelsvolumen hvis sentimentet omkring crypto stiger.

Modellen har en R^2 på 0,3231. Dette betyder at omkring 32% af variansen i den afhængige variabel kan forklares af de uafhængige variabler.

Træmodeller

Træbaserede modeller er en type machine learning algoritme, der bruges til både klassifikations- og regressionsanalyser. Som navnet antyder, er disse modeller baseret på decisions trees, som er grafiske repræsentationer af en række beslutninger og deres mulige konsekvenser.

Decision trees kan være tilbøjelige til høj varians, hvilket betyder, at de muligvis ikke er stabile og kan give forskellige resultater afhængigt af trænings dataen. Regressionstræer har den fordel, at de er nemme at forklare og fortolke, og at de kan håndtere kvalitative prædiktorer uden behov for dummy-variable. De har dog muligvis ikke samme niveau af

prædiktiv nøjagtighed som andre metoder og kan være følsomme over for små ændringer i dataene (James, G. Et. al. 2021).

Jeg har lavet et regressionstræ med samme variabler som ved den multiple lineære regression, i et forsøg på at få et nemmere fortolkeligt svar.

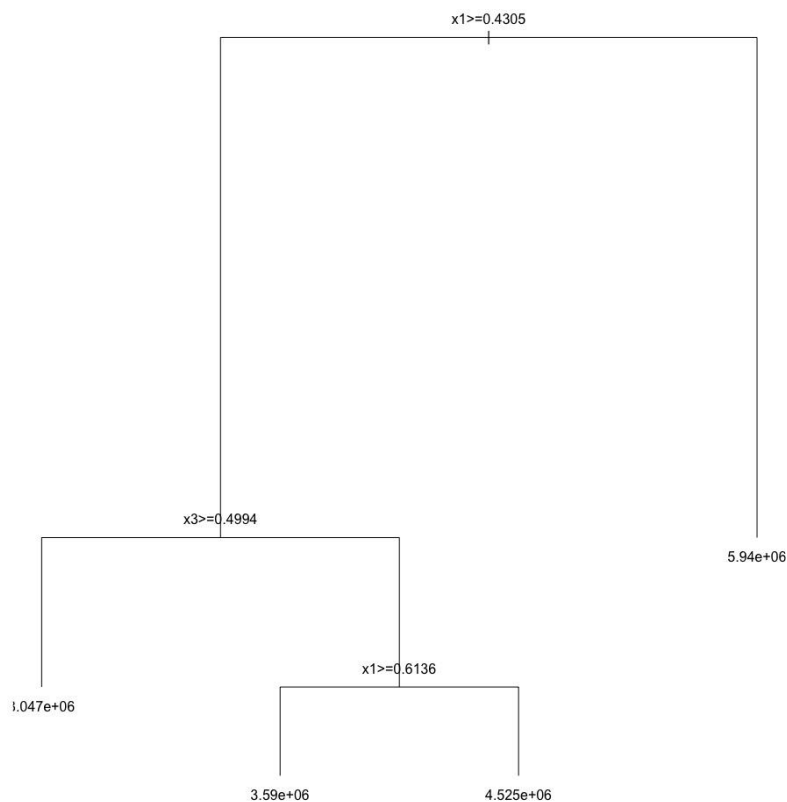
Det første regressionstræ er lavet med den afhængige variabel, y, som hele handelsvolumen. Igen er de uafhængige variabler, x, følgende:

x1 = Crypto Sentiment

x2 = Krypto NO Sentiment

x3 = Bitcoin Sentiment

Figur VI¹⁶:



17

¹⁶ Det har desværre ikke været muligt at ændre fra videnskabelig notation i modellen. Alle tal er millioner af NOK.

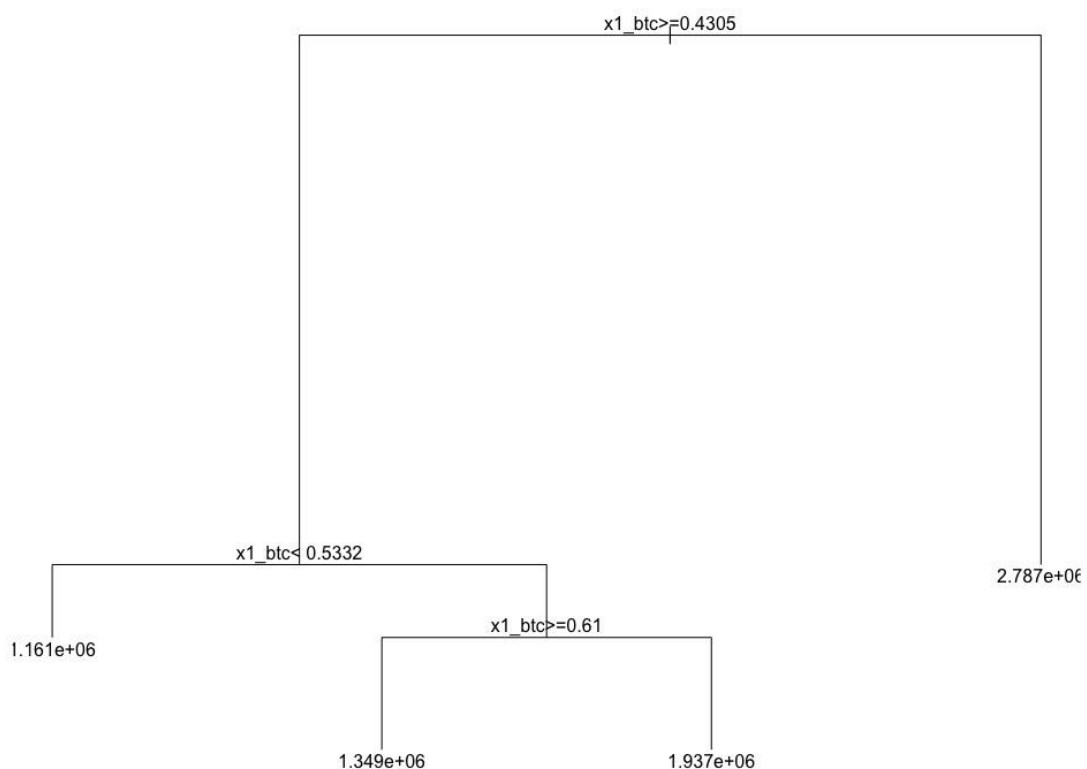
¹⁷ Egen tilblivelse

Det første split i dette decision tree sker ved variablen x_1 , som den lineære regression også viste var den mest signifikante. Vi kan her se at hvis sentimentet omkring crypto er højere end eller lig med 0,4305 vil handelsvolumen være 5.940.000 NOK.

Ligeledes kan vi ud fra modellen se at hvis x_1 er mindre en 0,4305, afhænger handelsvolumen af x_3 , sentimentet omkring bitcoin.

Samme decision tree kan laves på handelsvolumen af Bitcoin

Figur VII¹⁸:



¹⁹

¹⁸ Det har desværre ikke været muligt at ændre fra videnskabelig notation i modellen. Alle tal er millioner af NOK.

¹⁹ Egen tilblivelse

Igen kan vi se at hvis sentimentet omkring crypto er højere end eller lig med 0,4305 vil handelsvolumen være 2.787.000 NOK.

Delkonklusion

Machine learning modeller som lineær regression og decision trees, kan anvendes til at finde sammenhæng mellem handelsaktiviteten i Norge på Firis platform, og stemningen på Twitter omkring crypto og Bitcoin i november og december 2022.

Den lineære regression for både hele handelsvolumen og isoleret på Bitcoin, viste at sentimentet omkring crypto på Twitter er signifikant for handelsvolumen. Modellen viste også en forklaringsgrad på henholdsvis 38% og 32%.

Begge træmodeller viste at en sentiment score omkring crypto på 0,4305 havde største betydning for handelsvolumen, på både hele markedet og isoleret på Bitcoin.

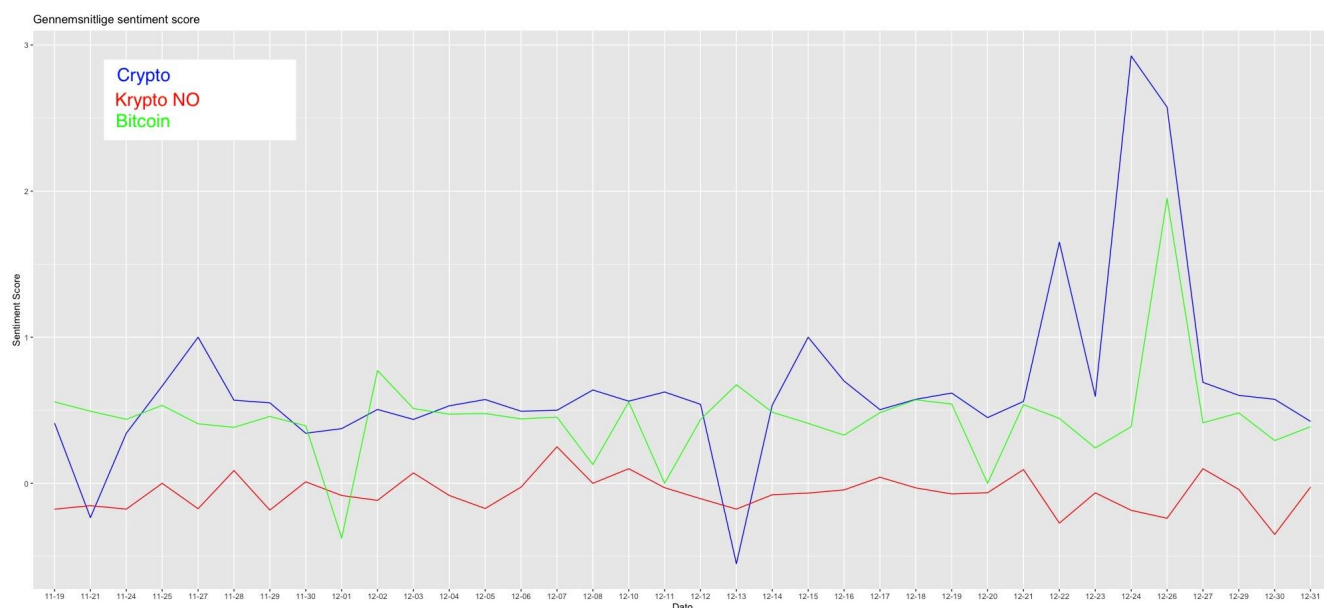
Hvordan kommunikeres eventuelle findings fra undersøgelsen bedst til virksomheden samt dennes stakeholders?

Ovenstående machine learning modeller kan være svære at kommunikere til en virksomhed og dens stakeholders. Det er derfor vigtigt, først og fremmest, at gøre problemet som er forsøgt løst klart for de involverede. Dernæst skal resultaterne af analysen præsenteres på en klar og kortfattet måde, ved hjælp af visualiseringer til at illustrere de væsentligste pointer. Konsekvenserne af resultatet skal forklares for virksomhedens og dens stakeholders. Dette kan eventuelt indebære anbefalinger til handling baseret på resultatet af analysen.

Det første skridt i at kommunikere resultatet er at gøre det forståeligt hvad en sentiment analyse er.

Den følgende figur viser den gennemsnitlige sentiment score fordelt på de tre forskellige udvalgte keywords per dag.

Figur VIII



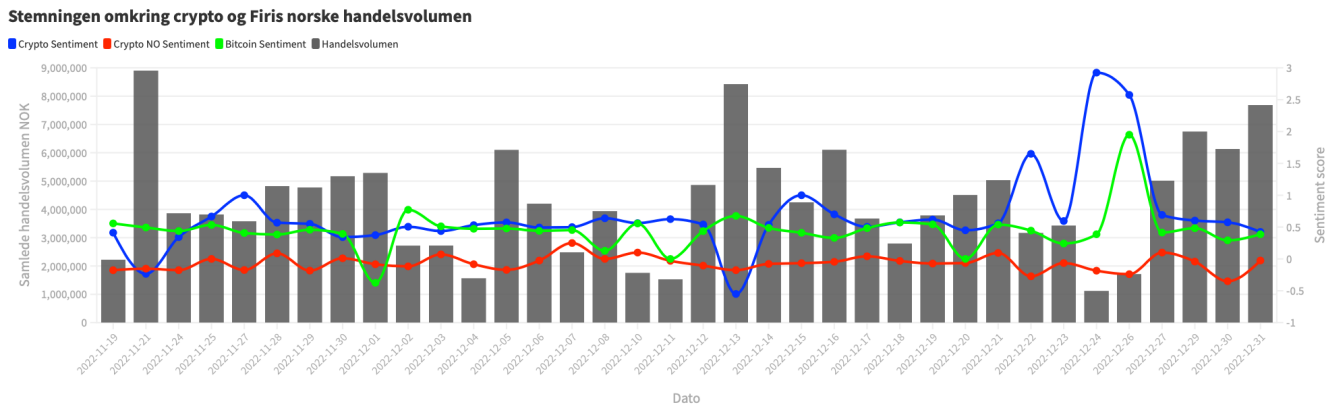
20

Denne figur giver et godt overblik over hvordan stemningen ændrer sig fra dag til dag, og hvor forskellig den kan være alt efter hvilket ord der er i fokus.

²⁰ Egen tilblivelse

For at vise resultatet fra den multiple lineære regression, kan ovenstående figur kombineres med den samlede handelsvolumen:

Figur IX:



21

Denne graf viser visuelt resultatet af den lineære regression: at når stemningen omkring crypto stiger, falder handelsvolumen.

²¹ Egen tilblivelse

Konklusion

Det sociale medie Twitter spiller en vigtig rolle i kryptoindustrien, da den gør det muligt for brugerne og investorer af kryptovaluta og blockchain at diskutere og dele informationer herom. Twitter er dog også en kontroversiel platform, især siden Elon Musks overtagelse har platformen været plaget af bots, som kan disrupte stemningen.

Stemningen omkring kryptovaluta og Bitcoin har i de sidste to måneder af 2022 været plaget af nogle store begivenheder i branchen, som kan ses i sentiment analyserne. Især FTXs Sam Bankman-Frieds arrestering og senere løsladelse har givet store udsving i sentimentet omkring crypto.

Jeg har ved brug af machine learning modeller som multipel lineær regression og decision trees, fundet en korrelation mellem sentimentet på Twitter omkring crypto i november og december 2022 og handelsvolumen af kryptovaluta i Norge på kryptobørsen Firi. Analysen viser at når stemningen omkring kryptovaluta stiger på Twitter, falder handelsvolumen i Norge hos Firi.

Ydermere varetager Firi også handel på vegne af virksomheder og institutioner i Norge, som ønsker at investere i kryptovaluta. Disse kunder påvirkes ikke nødvendigvis på samme grad af stemningen eller nyheder i samme grad som privatkunder.

Afslutningsvis er det vigtigt at påpege at kryptovaluta, og værdier heraf, ikke er isoleret fra hverken aktiemarkedet, verdensøkonomien eller fiat-valutaer. Prisen på kryptovalutaer er yderst volatil, og afgøres af flere faktorer som udbud- og efterspørgsel, politiske regulativer, opfattelse af værdi m.m.

Perspektivering

Enden på (Krypto)Twitter?

Twitters fremtid har været diskuteret siden Elon Musk overtog virksomheden i oktober 2022. Opkøbet er blevet mødt med en blanding af begejstring og skepsis, hvoraf nogle brugere udtrykte bekymring over potentialet for at Musk vil anvende platformen til at fremme sine egne interesser. Det diskuteres også om hvorvidt de ændringer, som Musk har implementeret siden overtagelsen som ejer af Twitter, herunder introduktionen af nye funktioner og fjernelse af bestemt indhold fra platformen, har fået nogle brugere til at migrere til alternative sociale medieplatforme, da de føler, at Twitter ikke længere opfylder deres behov (Bossio, D. 2022).

Det er værd at bemærke, at platformen, som engang var et knudepunkt for diskussion og debat om kryptovaluta industrien, er blevet mere fokuseret på at promovere forskellige produkter og tjenester. Dette skift har ført til et fald i kvaliteten af diskussioner og debat på platformen, samt bidraget til udbredelsen af misinformation og spam. Dette skift har bidraget til et fald i den samlede værdi af platformen for brugerne, og det er dermed op til brugerne af platformen at beslutte, om de vil fortsætte med at bruge den i sin nuværende form eller at opsøge alternative platforme til at diskutere og debattere kryptovaluta-relaterede spørgsmål (Casey, M. 2022)

Flere Twitter brugere er siden Musks overtagelse flyttet til det decentraliserede sociale medie Mastodon. Mastodon er bygget på open source-teknologi, og ejes og drives af et netværk af uafhængige servere, frem for af et enkelt firma. Denne decentraliserede struktur giver brugerne mulighed for at have mere kontrol over deres data og det indhold, der deles på platformen (Ilyshina, N. 2022).

Det er umuligt at forudsige om hvorvidt Musks opkøb af Twitter er enden på "Kryptotwitter" som vi kender det. Det tyder dog ikke på at brugerne stopper med at skrive omkring kryptovaluta og blockchain, men blot flytter til andre platforme.

Kryptos fremtid

Kryptovaluta markedet har oplevet et fald i priserne de seneste måneder. På trods af dette fortsætter ventureselskaber og andre investorer med at investere i kryptovaluta-relaterede startups og projekter.

Denne tendens kan være drevet af en tro på, at det langsigtede potentiale på kryptovalutamarkedet stadig er stærkt, og at den nuværende nedtur blot er et midlertidigt tilbageslag.

Værd at bemærke er også, at nogle investorer kan blive tiltrukket af kryptovaluta-relaterede projekter på grund af deres høje risiko/belønning profil og potentialet for store afkast.

Overordnet fortsætter investering i kryptovalutamarkedet, på trods af den seneste nedgang i priserne, og antyder, at denne tendens kan være drevet af en tro på markedets langsigtede potentiale og potentialet for høje afkast (Munster, B. 2022).

"Crypto Winter" er et udtryk, der bruges til at beskrive en periode med vedvarende fald i værdien af kryptovalutaer, især Bitcoin. Dette fald kan karakteriseres ved et fald i priserne på kryptovalutaer, samt et fald i aktivitetsniveauet og investeringerne på kryptovalutamarkedet.

“Crypto Winter” kan være drevet af en række faktorer, herunder regulatorisk usikkerhed, markedets spekulation og bredere økonomiske forhold. Det ses generelt som en udfordrende periode for kryptovaluta-industrien, da det kan føre til et fald i antallet af nye projekter og startups, der kommer ind på markedet, hvilket kan føre til et fald i antallet af mennesker, der deltager i markedet som investorer eller brugere.

Det er dog vigtigt at bemærke, at kryptovalutamarkedet er meget volatilt, og perioder med tilbagegang efterfølges ofte af perioder med vækst. Som sådan er det svært at forudsige, hvor længe en “Crypto Winter” kan vare, eller hvornår den næste vækstperiode kan begynde.

Yderlige og fremtidige analyser

Langsigtet analyse

Det ville være muligt at lave en langsigtet analyse, for eksempel ét helt år tilbage, såfremt virksomheden tilkøber enterprise adgang til Twitters API. En langsigtet analyse vil muligvis være mere præcis, da kryptoindustrien har været ramt af nogle store begivenheder i november og december 2022, som kan have haft indflydelse på resultatet.

Automatisering

Ønskes analysen af sentimentet/stemningen omkring kryptovaluta på Twitter at fortsætte, ville det eventuelt være muligt at opsætte automatisering til indhentning og opbevaring af tweets. Dette kan blandt andet gøres i R med pakken Cron. Cron gør det muligt at automatisere og planlægge opgaver i R (Cron. 2022)

Ved at automatisere indhentningen og opbevaringen af tweets, undgår man at mangle data, som der desværre skete i projektet her.

Analyse af andre kryptovalutaer

Bitcoin blev her i projektet valgt som et enkeltstående keyword, da det har den største marked cap. Det ville dog også være muligt at lave samme analyse på andre kryptovalutaer som Firi tilbyder handel med, f.eks Ethereum eller XRP. Især Ethereum, som i år har gennemgået nogle fundamentale ændringer, kunne være spændende at undersøge i fremtiden.

Bibliografi

Bøger

James, G. Et. al. (2021, 4. August). An Introduction to Statistical Learning with Applications in R. 2nd Edition.

Linode. (2022). Understanding Databases: Deploy High-Performance Database Clusters in Modern Applications. Linode.

Silge, J. & Robinson, D. (2022). Text Mining with R. O'Reilly,

Statista. (2022). Twitter Dossier. Statista.

Hjemmesider

ABP News Bureau. (2022, 23. December). What Caused The FTX Fiasco? A Timeline Of Events So Far. Abp live.

<https://news.abplive.com/business/crypto/ftx-crash-bankruptcy-sam-bankman-fried-arrest-bail-timeline-events-1571418>. Tilgået 2. Januar 2023.

Aljazeera. (2022, 13. December). FTX's Bankman-Fried arrested in the Bahamas as US files charges. Aljazeera.

<https://www.aljazeera.com/economy/2022/12/13/ftxs-bankman-fried-arrested-in-the-bahamas-as-us-unveils-charges>. Tilgået 16. December 2022.

Bitcoin. (2019, 24. April). Block size limit controversy. Bitcoin.

https://en.bitcoin.it/wiki/Block_size_limit_controversy. Tilgået 17. December 2022.

Bossio, D. (2022, 4. November). Why Elon Musk's first week as Twitter owner has users flocking elsewhere. The Conversation.

<https://theconversation.com/why-elon-musks-first-week-as-twitter-owner-has-users-flocking-elsewhere-193857>. Tilgået 27. December 2022.

Casey, M. (2022, 23. December). The End of Crypto Twitter as We Know It? Coindesk.
https://www.coindesk.com/layer2/2022/12/23/the-end-of-crypto-twitter-as-we-know-it/?utm_source=tldrnewsletter. Tilgået 27. December 2022.

Coingecko. (n.d). Cryptocurrency Prices by Market Cap. Coingecko.
<https://www.coingecko.com/>. Tilgået 22. December 2022.

Cron. (2022, 17. Februar). cronR. Github. <https://github.com/bnosac/cronR#readme>. Tilgået 23. November 2022.

De, N. (2022, 11. November). FTX Files for Bankruptcy Protection in US; CEO Bankman-Fried Resigns. Coindesk.
<https://www.coindesk.com/policy/2022/11/11/ftx-files-for-bankruptcy-protections-in-us/>.
Tilgået 16. December 2022.

Frenkel, S. (2022, 9. Juli). What Are Spam Bots and Why They're an Issue in Elon Musk's Twitter Deal. The New York Times.
<https://www.nytimes.com/2022/07/09/technology/elon-musk-twitter-spam-bots.html>. Tilgået 27. December 2022.

Firi (n.d). Kjøp og selg kryptovaluta enkelt med markedsordre. Firi.
<https://firi.com/no/om-oss/markedsordre>. Tilgået 30. November 2022.

Firi (n.d.). Om os. Firi. <https://firi.com/dk/om-os> Tilgået 15. November , 2022.

Holten, Å. (2022, 1. Marts). Kryptovaluta for dummies. Firi.
<https://firi.com/dk/artikler/kryptovaluta-for-dummies-guide>. Tilgået 17. November 2022.

Ilyshina, N. (2022, 12. November). Hvad er Mastodon, som Twitter-brugerne flygter til? Videnskab.dk.
<https://videnskab.dk/forskerzonen/kultur-samfund/hvad-er-mastodon-som-twitter-brugere-flygter-til>. Tilgået 25. December 2022.

Metabase. (n.d). Product. Metabase. <https://www.metabase.com/product/>. Tilgået 14. November 2022.

Munster, B. (2022, 24. December). Crypto Is Down, But VCs Keep Pouring Money In. Decrypt. <https://decrypt.co/117900/crypto-is-down-bad-but-vcs-keep-pouring-money-in>. Tilgået 25. December 2022.

RDocumentation (n.d). rtweet (version 1.0.2). RDocumentation. https://www.rdocumentation.org/packages/rtweet/versions/1.0.2/topics/search_tweets. Tilgået 24. November 2022.

Roth, Y. & Pickles, N. (2020, 18. Maj). Bot or not? The facts about platform manipulation on Twitter. Twitter. https://blog.twitter.com/en_us/topics/company/2020/bot-or-not. Tilgået 28. December 2022.

Sigalos, M. (2022, 28. November). Crypto firm BlockFi files for bankruptcy as FTX fallout spreads. CNBC. <https://www.cnbc.com/2022/11/28/blockfi-files-for-bankruptcy-as-ftx-fallout-spreads.html>. Tilgået 2. Januar 2023.

Twitter (n.d). About. Twitter. <https://about.twitter.com/en>. Tilgået 17. November 2022.

Twitter Developer. (n.d). Getting Started. Developer Platform. <https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api>. Tilgået 12. December 2022.

Bilag

Bilag I:



Thuc Hoang 3 months ago

Hvor tidlig kan man se på markedssentimentet (på f.eks Twitter) før markedet faktisk faller eller at prisene går opp.

Hva slags nyheter har størst påvirkningskraft på markedet (som igjen treffer volumet).

Var Ethereum's "The Merge" viktigere enn Bitcoin's "halvening" mtp aktivitet på plattformen?

Er det noe mønster (eller om det er sesongbasert) på hvilke kryptovaluta som kundene ønsker å holde igjennom stormene?

Bilag II:

```
> summary(lm_simple)
```

Call:

```
lm(formula = y ~ x1 + x2 + x3, data = df_total_handel)
```

Residuals:

Min	1Q	Median	3Q	Max
-3032862	-714808	-74288	959923	3150484

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	5238353	480653	10.898	2.65e-12	***
x1	-1962054	484164	-4.052	0.000302	***
x2	-4597630	2301563	-1.998	0.054320	.
x3	32841	881572	0.037	0.970515	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1575000 on 32 degrees of freedom

Multiple R-squared: 0.3773, Adjusted R-squared: 0.3189

F-statistic: 6.463 on 3 and 32 DF, p-value: 0.001517

Bilag III:

```

> summary(lm_btc)

Call:
lm(formula = y_BTC ~ x1_btc + x2_btc + x3_btc, data = df_btc_handel)

Residuals:
    Min       1Q   Median       3Q      Max
-1600353  -508076   -77624   369246  2873014

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2257524     274710   8.218 2.18e-09 ***
x1_btc       -1021143     276716  -3.690 0.000829 ***
x2_btc       -2077480    1315422  -1.579 0.124097
x3_btc         112471     503849   0.223 0.824780
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 900400 on 32 degrees of freedom
Multiple R-squared:  0.3231,    Adjusted R-squared:  0.2597
F-statistic: 5.093 on 3 and 32 DF,  p-value: 0.005393

```