

Twittien hyödyntäminen osakemarkkinoiden ennustamisessa

Using Tweets to Predict Stock Markets

Kandidaatintyö

Matias Rikkinen

TIIVISTELMÄ

Tekijä: Matias Rikkinen

Työn nimi: Twiittien hyödyntäminen osakemarkkinoiden ennustamisessa

Vuosi: 2019

Paikka: Lappeenranta

Kandidaatintyö. LUT-yliopisto, tuotantotalous.

30 sivua, 6 kuvaa

Tarkastaja(t): Lea Hannola

Hakusanat: Big data, ennustaminen, osakemarkkinat, Twitter, sosiaalinen media

Keywords: Big data, forecasting, predicting, stock market, Twitter, social media

Tämän kandidaatintyön tavoitteena on luoda katsaus siihen, onko mikroblogipalveluiden tuottama tekstidata ja sen sisältämä sentimentti hyödynnettävissä päätöksenteon tukena osakemarkkinoilla. Työssä selvitetään big datan analysointikeinot ja käyttökohteet. Työ on tehty kirjallisuuskatsauksena, sekä työssä esitellään yritys, jonka liiketoiminta perustuu sosiaalisen median sentimentin analysoimiseen.

Osakemarkkinoilla liikkuu valtavia määriä rahaa, mutta markkinoilla pärjätäkseen tarvitaan usein muita parempaa näkemystä. Hyödyntämällä erilaisia, toimiviksi todettuja analyysikeinoja, on mahdollista pärjätä markkinoilla muita paremmin. Työssä tehdään oletus, että osakemarkkinat eivät ole aina tehokkaita ja markkinoiden voittaminen on mahdollista.

Tiedon määrä maailmassa on kasvanut räjähdysmäisesti ja tulee kasvamaan alati kiihtyvällä tahdilla. Tämän seurauksena hyödyllistäkin tietoa voidaan haravoida paikoista, joista ei ensimmäiseksi sitä odottaisi löytävänsä. Tietokoneiden laskentatehon kasvaessa ja pilvipalveluiden käytön yleistyessä, voidaan hyödyllistä ja valtavien rahasummien arvoista tietoa erotella valtavista datavirroista.

Big datan hyödyntäminen ei kuitenkaan ole vaivatonta ja se edellyttää hyödyntäjältä perehtyneisyyttä sekä nokkeluutta. Useat tutkimukset ovat osoittaneet, että mikroblogiviestien sentimentti korreloi osakekurssien kanssa, mutta tutkimukset eivät ole yksimielisiä siitä, onko korrelaatio merkittävää etenkin pitkällä aikajänteellä. Osakekaupan lisäksi taloudellista lisäarvoa on mahdollisuus luoda esimerkiksi myymällä mikroblogiviestien sentimenttiin pohjautuvia analyysejä.

SISÄLLYSLUETTELO

1	Johdanto	3
2	Big data	4
2.1	Mitä on big data	4
2.2	eWOM.....	5
2.3	Tekstianalyysi	6
2.3.1	Tietokoneavusteiset tutkimusmenetelmät	7
2.3.2	Itsenäiset koneoppimisalgoritmit	7
2.4	Pilvipalvelut	8
3	Mikroblogipalvelu twitter	10
3.1	Avoimuus	11
3.2	Nopeus	12
3.3	Seuraaminen Twitterissä	13
3.4	StockTwits	13
4	Osakemarkkinat	14
4.1	Markkinoiden tehokkuus	15
4.2	Markkinoiden voittaminen.....	17
4.2.1	Ylituotot informaatioedulla.....	18
4.2.2	Aktiivinen kaupankäynti	18
4.3	Tunteet osana päätöksentekoa.....	18
4.4	Fundamenttianalyysi	20
4.5	Lyhyeksi myyminen.....	20
5	Big datan hyödyntäminen osakkeen hinnan kehityksen ennustamisessa.....	21
5.1	Big data- analytiikka	21
5.1.1	Sentimenttianalyysi	21
5.1.2	Älykkäät twiittaajat	23
5.1.3	Kritiikki sentimenttianalyysin tarkkuudesta.....	24
5.2	Lyhyen aikavälin ennustaminen.....	25
5.3	Osana fundamenttianalyysiä?	26
5.4	Case StockFluence	26
6	Johtopäätökset ja yhteenveto	29
7	Lähteet.....	33

LYHENTEET

MAPE	Mean Absolute percentage error, keskivirhe prosentteina
RMSE	Root-mean-square error, keskineliövirheen neliöjuuri
DJIA	Dow Jones Industrial Average -indeksi

KUVALUETTELO

Kuva 1 Big data, paljon nopeasti lisääntyvää monipuolista dataa (Salo 2014, s.28)

Kuva 2 Twitterin aktiiviset käyttäjät maailmanlaajuisesti 2014-2020

(<https://www.statista.com/statistics/303681/twitter-users-worldwide/>) Viitattu 10.4.2019

Kuva 3 Esimerkiksi Suomen BKT asukasta kohden on kasvanut merkittävästi vuosien 1975-2018 välillä

(https://www.tilastokeskus.fi/static/media/uploads/tup/suoluk/kuviot/suomi_kan002.gif)

Viitattu 12.4.2019

Kuva 4 Tehokkaiden ja ei-tehokkaiden markkinoiden reaktio positiivisesti hintaan vaikuttavaan informaatioon. (Knüpfer, Puttonen 2018 s. 166)

Kuva 5 Yksinkertaistettu kuvaus kurssiromahduksesta ja sijoittajien tunnetiloista.

Kuva 6 StockFluencen sentimenttianalyysin ennakoima hinnanvaihtelu Microsoftille on reaalihintakehityksen suuntaista, mutta voimakkaampaa.

(<https://www.stockfluence.com/>) Kuvakaappaus 27.4.2018

1 JOHDANTO

Tiedon määrä on lisääntynyt maailmassa räjähdysmäisesti ja tuotamme tietoa eli dataa yhä kiihtyvällä tahdilla. Informaation määrä kymmenkertaistuu viiden vuoden välein (Khan Nawsher et al., 2014) Teknologian kehittymisen myötä, yhä useammalla ihmisellä on mahdollisuus osallistua arvopaperikauppaan, käytännössä missä tahansa maailman arvopaperipörssissä. Ihmisillä on taipumus osallistua keskusteluun osakkeiden hinnankeskitykseen liittyen esimerkiksi sosiaalisessa mediassa tai sijoitusaiheisilla Internet-foorumeilla. Osallistumalla keskusteluun ihminen voi tiedostamattaan tai tiedostaen jättää mielipiteensä, tunteensa tai tavoitteensa alitajuisena viestinä viestinsä sisältöön. Yksittäisesti tarkasteltuna viestit voivat antaa täysin ristiriitaista informaatiota, mutta tarkasteltaessa valtavia määriä viestejä, voi jonkinlainen toistuma tai trendi, olla löydettävissä onnistuneen analyysin pohjalta. (Sul, Dennis & Yuan, 2017)

Osakemarkkinoiden liikehdinnän ennustaminen on erittäin vaikeaa, eikä välttämättä kenelläkään ole varmaa tietoa, kuinka markkinat kehittyvät. Osakemarkkinoilla liikkuvat valtavat rahamäärät kuitenkin ovat kannustaneet ihmisiä ja instituutioita kehittämään erilaisia analysointi- ja ennustamistyökaluja. Teknologian kehittyessä avautuu uusia mahdollisuuksia ennustaa ja analysoida, mitä tahansa ilmiötä. Uusien ennustamiskeinojen onnistunut käyttöönotto ja hyödyntäminen voi antaa sitä hyödyntävälle taholle selkeitä voittoja osakemarkkinoilta (Sul, Dennis & Yuan, 2017).

Tässä kandidaatin työssä tarkastellaan mahdollisuutta mikroblogipalvelu Twitterin sekä StockTwitsin tuottaman tekstidatan käyttämisestä osakemarkkinoiden hinnankeskityksen ennustamiseen. Työn tavoitteena on selvittää, onko mikroblogiviestien tulvasta löydettävissä trendejä tai tietoa, joita voitaisiin käyttää osakemarkkinoilla päätöksenteon tukena. Työn tärkeimmät tutkimuskysymykset ovat, voiko osakemarkkinoiden hinnankeskitystä ennustaa mikroblogikirjoitusten eli twiittien perusteella ja voidaanko tätä keinoa käyttäen voittaa markkinat eli tehdä enemmän tuottoa kuin markkinoiden yleiskehitystä mittaava indeksi. Kirjallisuuskatsauksella selvitetään, mitä aiempia tutkimuksia ja kokeiluja big datan käytöstä ennustamisen apuna on tehty. Teoriaosassa selvitetään, mitä big data on ja keinoja sen käyttämiseen päätöksenteon tukena.

2 BIG DATA

Termi ”Big Data” on noussut yhdeksi 2010-luvun kuumimmista trendisanoista. Käsite big data tuli tutuksi jo vuonna 2005, mutta nykyinen big data -hehkus lähti nousuun vuonna 2011. Ensimmäinen maininta big datan olemassaolosta lienee se, kun vuonna 2001 META Group-nimisen yrityksen työntekijä Doug Laney mainitsi raportissaan datamäärän vaihtelevuuden, määrän ja nopeuden kasvavan tulevaisuudessa. (Salo 2014, s.26)

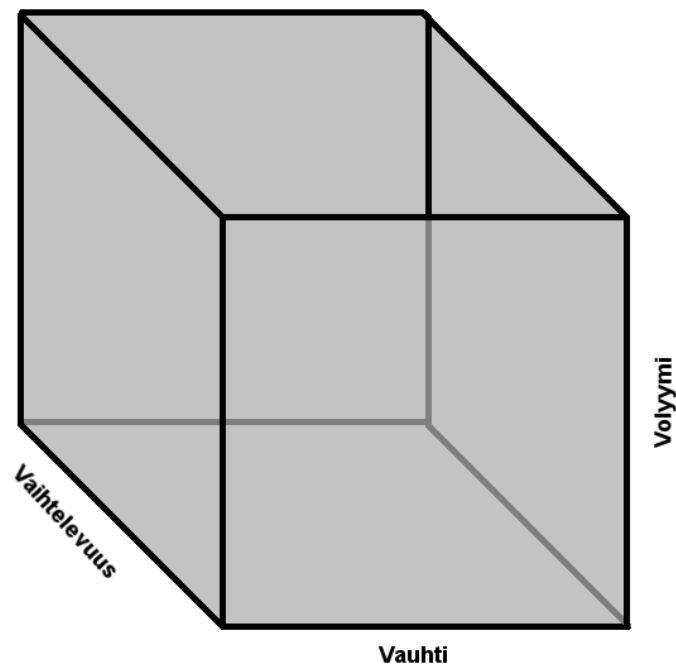
2.1 Mitä on big data

Salon (2014, s.30) mukaan big data on pilvipalveluiden kaltainen kattokäsite joukolle havaintoja, teorioita sekä konkreettisia ratkaisuja. Yleisimmin sillä viitataan siihen, että dataa maailmassa on paljon, ja sitä tulee kiihtyvällä tahdilla lisää ja se on muodoltaan yhä vaihtelevampaa. Toinen asia, mihin big datalla viitataan, on ne konkreettiset ratkaisumallit eli tuotteet, palvelut ja tekniikat, jotka pyrkivät vastaamaan datan määrän, kasvuvauhdin ja vaihtelevuuden asettamiin haasteisiin. (Salo 2014, s.30)

Piirteet, jotka erottavat big datan tavallisesta ovat sen volyyymi, vaihtelevuus ja vauhti. Englanniksi nämä ominaisuudet kuvataan kolmen V:n avulla; Volume, Variety ja Velocity (Rainer, Prince, 2017). Volyymi viittaa datan suureen määrään, vaihtelevuus datan laatuun ja vauhti sen virtaamisnopeuteen sekä siihen, että vauhti luo paineita reagoida uuteen dataan nopeasti (Kuva 1). Tavalliseen dataan verrattuna big data on laajempaa, sitä syntyy valtavalla vauhdilla ja se on koostumukseltaan vaihtelevaa. Big data voidaan jakaa laatunsa perustella kahteen eri luokkaan, strukturoituun ja strukturoimattomaan dataan. Strukturoitu data on perinteisempää dataa, esimerkiksi asiakastiedot, kun taas strukturoimaton data voi olla peräisin valvontakameran tuottamasta videokuvasta. (Salo 2014, s.26-27) On siis todettavissa, että big data asettaa paineita nykytekniikalle, sillä dataa syntyy valtavia määriä, yhä vaihtelevimmista lähteistä, kasvavalla nopeudella.

Big dataa voisi kuvailla enemmänkin raaka-aineeksi kuin hyödylliseksi lopputuotteeksi. Käsittelemätön data on yksinään hyödytöntä, siitä ei voi vetää johtopäätöksiä tai löytää trendiä.

Jotta raaka-aineesta eli datasta olisi hyötyä, tulee siitä louhia informaatiota ja siitä muodostaa tietoa. Tieto lisää ymmärrystä ja kumuloitunut tieto muodostaa ymmärrystä. (Salo 2014, s. 32)



Kuva 1 Big data, paljon nopeasti lisääntyvää monipuolista dataa (Salo 2014, s.28)

2.2 eWOM

Electronic word-of-mouth (eWOM) polveutuu alkuperäisestä word-of-mouthista (WOM) eli kuulopuheesta. WOM on vanha markkinointiviestinnän keino ja eWOM on nykyaikaan paremmin sopiva termi, sillä se käsittää esimerkiksi sosiaalisessa mediassa kirjoitetut tuote-arvostelut kuulopuheena. Word-of-mouth määritellään suulliseksi, ihmiseltä ihmiselle tapahtuvaksi viestinnäksi, jossa kuulija tulkitsee, että kertoja ei ole yrityksen, tuotteen tai brändin palkattu mainostaja. (Arndt, 1967) WOM:ia voidaan ajatella yhtenä vanhimmista markkinointiviestinnän keinoista ja sen teho perustuu ihmisten väliseen oletukseen toisen ihmisen rehellisyydestä.

eWOM:n vaikutusta ihmisten päätöksentekoon on aiemmin tutkittu. Chen (2015) perehtyi kuluttajien lomamatkojen valitsemisprosessiin ja heidän mukaansa etenkin ensimmäistä kertaa matkustavat altistuvat herkemmin eWOM:lle. Toisen saman toimialan tutkimuksen mukaan kuitenkin perinteinen WOM oli eWOM:ia tehokkaampaa (Hernández-Méndez, 2015).

Vaikka eWOM onkin markkinointiviestinnän keino ja sen tuloksista kiinnostuneita ovat lähinnä yritykset, joidenka tuotteet ovat sen kohteena, tehdään tässä tutkielmassa oletus, että eWOM vaikuttaa myös muihin asioihin kuin esimerkiksi kuluttajatuotteisiin tai brändeihin. Kuitenkin eWOM:n vaikutukset ovat nähtävissä muuallakin, kuten opiskelijoiden kurssivalinnoissa, muiden opiskelijoiden arvostelujen perusteella. Vuonna 2013 julkaistussa tutkimuksessa Cong Li totesi, että muiden opiskelijoiden professoreille antamalla arvosanalla oli vaikutusta opiskelijoiden kurssivalintoihin. Muiden opiskelijoiden antamiin arvosteluihin suhtauduttiin vähemmän kriittisesti. (Li, 2013)

Voimme siis tehdä oletuksen, että nämä samat ihmisten väliset sosiaaliset vuorovaikutukset voivat päteä myös mikroblogiviestien ja osakemarkkinoiden välisessä yhteydessä. Tämän seurauksena voisi olla mahdollista, että trendejä tunnistamalla on mahdollista toimia siten, että avointa dataa analysoimalla, voidaan luoda taloudellista lisäarvoa.

2.3 Tekstianalyysi

Jotta twiittien valtavasta massasta olisi löydettävissä etsijää hyödyttävää, analysoitavaa dataa, tulee twiiteistä jollakin keinolla erottaa sen sisältämä informaatio. Yleisin kvantitatiivinen tutkimusmenetelmä suurten joukkojen kommunikoimiseen liittyen on manuaalinen sisältöanalyysi. Manuaalinen sisältöanalyysi on prosessi, jossa ihmisten tuottamaa dataa kategorisoidaan, jotta saataisiin vastaus tutkimuskysymykseen kuten ”Mistä osakkeesta ihmiset twiittaavat aktiivisimmin?” Tämän perinteisen tutkimusmenetelmän hyviä puolia ovat se, että menetelmä perustuu ihmisten aiempiin aiheesta tehtyihin tutkimuksiin ja ihmisten ajattelutyöhön, mikä on edellytys esimerkiksi sarkasmin tunnistamiseen tekstidatasta. Menetelmän huonoja puolia ovat se, että menetelmä on kallis ja aikaa vievä, ihmisen tekemiä virheitä unohtamatta. (Guo, 2016)

2.3.1 Tietokoneavusteiset tutkimusmenetelmät

On luonnollista, että tietokoneet avustavat tutkijoita suurten datamassojen käsittelyssä. Tietokoneavusteiset menetelmät keventävät ihmisten taakkaa uuvuttavissa toimenpiteissä kuten suurten datamassojen keräämisessä ja lajittelussa, mutta menetelmän tulokset nojaavat vahvasti ihmisten luovuuteen koodaamisessa ja siinä, että ymmärretään etsiä massasta oikeita asioita. Sanastopohjainen tekstianalyysi on yksi uuvuttavimmista analyysikeinoista, joka tietokone voidaan valjastaa tekemään. Mutta kuten perinteisessä manuaalisessa sisältöanalyysissä, menetelmän tehokkuus on hyvin riippuvainen etukäteen tehdyistä oletuksista, jaotteluista ja teorioista. (Guo, 2016)

2.3.2 Itsenäiset koneoppimisalgoritmit

Nykyaikaisimmat tekstianalyysityökalut kykenevät kasvaneen laskentatehon ansiosta yhä monimutkaisempiin ja laajempiin analyyseihin. Jos analysoitava data on kooltaan suuri, esim. yli miljoona yksikköä, voi perinteinen etukäteen jaottelu kategorioihin vääristää analyysin kohdetta. (Guo, 2016) Itseoppivien algoritmien vahvuus piileekin siinä, että niille ei ole annettuna mitään subjektiivista esitietoa, vaan sen tehtävä on löytää datan joukosta kuvio tai toistuvuus, jota ei välttämättä etukäteen osata edes kuvitella. Yksi yleisin lähestymistapa on aiheen perusteella jaottelu (eng. Topic Modeling). Menetelmä toimii siten, että se tunnistaa, mihin aihepiiriin sana kuuluu, esimerkiksi ”osake”, ”toimeksianto” ja ”osinko” liittyvät vahvasti sijoittamissanastoon. Menetelmä tunnistaa sanojen esiintymisyleisyyden ja jaoteltuaan sanat tietyn aiheen piiriin. Tämän perusteella menetelmä kertoo, mitä aihetta dokumentti käsittelee. (Guo, 2016)

Eniten käytetty Topic Modeling- menetelmä on Latent Dirichlet Allocation (LDA)- malli. Malli soveltuu hyvin oikeinkirjoitettujen, yhteen aiheeseen keskittyvien, artikkelien analysointiin, mutta sen tehokkuus heikkenee merkittävästi esimerkiksi twiittien kohdalla, sillä twiitit ovat lyhyitä, ennen alle 140 ja nykyään alle 280 merkkiä pitkiä, tekstinpätkiä, jotka sisältävät lyhenteitä ja erikoismerkkejä sekä hyymiöitä. Kuitenkin menetelmän tehokkuutta voidaan parantaa ihmisen avulla. Ihminen jalostaa ja tarkentaa algoritmin tuottamaa tulosta

iteratiivisesti. (Guo, 2016) Tietokoneiden heikkous on se, että ne eivät kykene ajattelemaan kovinkaan luovasti tai subjektiivisesti.

2.4 Pilvipalvelut

Big datasta puhuttaessa ei voida olla mainitsematta pilvipalveluita osana big datan keräämistä, analysoimista ja hyödyntämistä. Kuten aiemmin on todettu, big data eroaa perinteisestä datasta oleellisesti kokonsa, vaihtelevuutensa ja virtausnopeutensa vuoksi. Tämä asettaa perinteisille tietoteknisille laitteille, ohjelmistoille ja toimintamalleille haasteita, sillä näistä voi muodostua pullonkauloja, kun perinteisen data-analytiikan sijasta toteutetaan big data-analytiikkaa. Haasteen ydin nivoutuu resurssien ympärille. Big data-analytiikkaa ei voida harjoittaa samoilla toiminta-ajatuksilla, kuin perinteistä data-analytiikkaa, mikä johtuu siitä, etteivät perinteiset alustat pysty tallentamaan, käsittelemään ja analysoimaan valtavia, vaihtelevia, datamääriä samaan aikaan nopeasti ja kustannustehokkaasti. (Salo 2014, s. 6) Pilvipalvelut (eng. Cloud Computing) on big datan kaltainen kattokäsite, jolla tarkoitetaan etäkäytettävää eli pilvipalvelussa tapahtuvaa toimintaa. ”Yhdysvaltojen elinkeinoministeriön alainen National Institute of Standards and Technology määritelmä pilvipalveluille: ”Cloud Computing on toimintamalli, joka mahdollistaa pääsyn vapaasti konfiguroitaviin ja skaalautuviin tietotekniikkaresursseihin, jotka voidaan ottaa käyttöön tai poistaa käytöstä helposti ja nopeasti.” (Salo 2014, s.93) Määritelmän tietotekniikkaresurssit viittaavat laskentatehon ja tallennustilan lisäksi sovelluskehitysalustoihin ja sovelluksiin. (Salo 2014, s. 93) Oleellisimmat erot perinteisen toimintamallin ja pilvipalveluiden hyödyntämisen välillä ovat ne, että käytettävien resurssien määrä on skaalautuva, niin ylös- kuin alaspäinkin, ja ettei pilvipalveluiden hyödyntäminen ole päätelaiteriippuvaista. Tämä mahdollistaa sen, että big data-analytiikka voi olla kustannustehokkaampaa ja ylipäätään mahdollista, kun toinen taho tarjoaa tarvittavan laskentatehon, säilytystilan ja analyysityökalut ja kaikki nämä ovat tarvittaessa skaalattavissa tarvittavaan mittasuhteeseen.

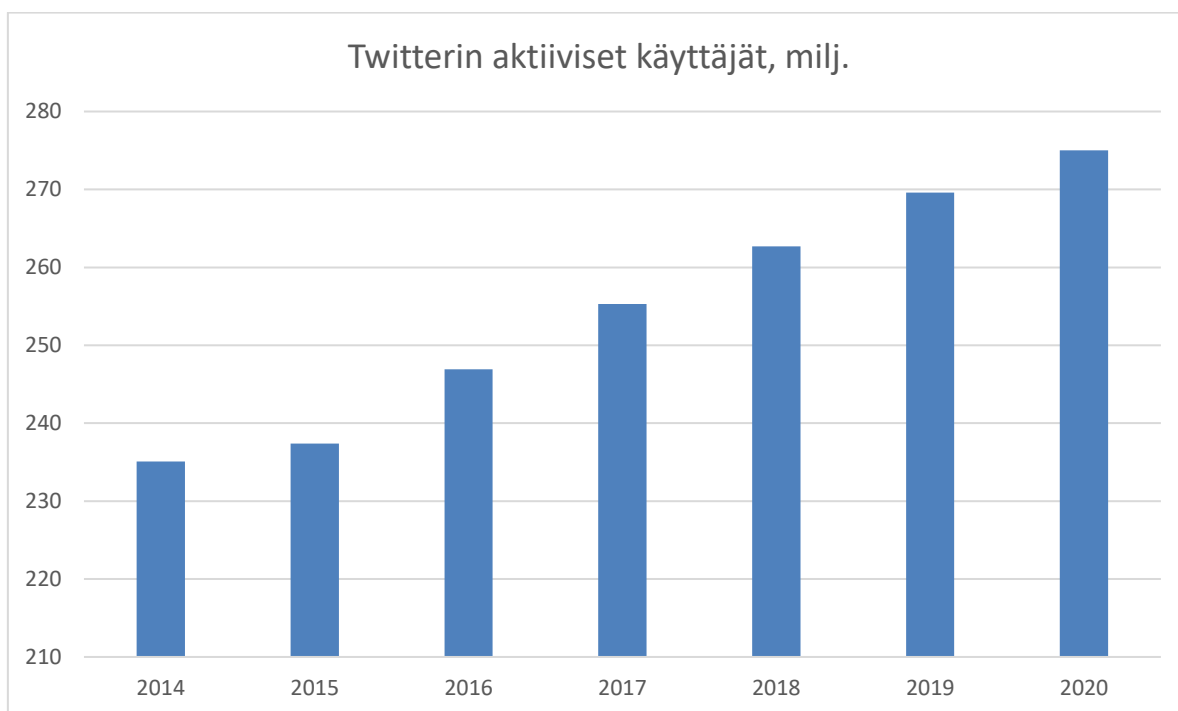
Pilvipalveluihinkin sisältyy myös omat riskinsä. Pilvipalveluiden käyttö lisää niitä käyttävän organisaation riippuvuutta palveluntarjoajaan, mikä altistaa organisaation erinäisille itsestään riippumattomille riskeille kuten palveluntarjoajan konkurssiriskille tai tietovuotoriskille. Etäkäyttöön liittyy luonnollisesti se, että on mahdollista jokin taho voi urkkia tietoa asiakkaan

ja palveluntarjoajan välistä. Riskejä voidaan hallita valitsemalla luotettava palveluntarjoaja, mielellään vakaasta ja luotettavasta maasta sekä hajauttamalla riski käyttämällä useampaa palveluntarjoajaa. (Rajaraman, 2014)

3 MIKROBLOGIPALVELU TWITTER

Vuonna 2006 podcast-yrityksen työntekijöiden harrasteprojekti kokeili siipiään, sen alkuperäinen tarkoitus oli toimia kommunikaatiovälineenä, ikään kuin korvatakseni SMS-viestit. Näin twtr eli Twitter sai alkunsa. Alun vaikeuksien jälkeen yritys koki valtavan kasvupyrähdyksen ja lennähti nopeasti ladatuimpien sovellusten listoille. (MacArthur, 2018)

Vuonna 2017 Twitter nosti mikroviestien eli twiittien merkkirajoituksen 140 merkistä 280 merkkiin eli Twitteriä voidaan edelleenkin kutsua mikroblogipalveluksi. Vuonna 2018 Twitterillä oli keskimäärin 262,7 miljoonaa aktiivista käyttäjää (Kuva 2). Twitteriä voidaan käyttää uutisten, ihmisten ja organisaatioiden julkaisujen seuraamiseen.



Kuva 2 Twitterin aktiiviset käyttäjät maailmanlaajuisesti 2014-2020 (<https://www.statista.com/statistics/303681/twitter-users-worldwide/>)

Twitterissä kommunikoidaan pääsääntöisesti tekstin avulla, mutta kuvien ja videoidenkin jakaminen on mahdollista. Oletuksena jokaisen henkilökohtainen tili on julkinen eli potentiaalisesti kaikki 270 miljoonaa muuta aktiivista käyttäjää voivat nähdä käyttäjän twiitit.

Profiilin voi kuitenkin muuttaa yksityiseksi eli ainoastaan käyttäjän seuraajat näkevät twiitit. Kuitenkin lähes 90 % profiileista on julkisia (Patel, 2017), mikä on sinänsä intuitiivista, koska mikroblogin tarkoituksena useasti on jakaa henkilökohtaisia ajatuksia myös tuntemattomille ihmisille. Keskimäärin päivässä twiitataan 500 miljoonaa kertaa ja trendi on kasvava (Live stats, 2019).

3.1 Avoimuus

Mikroblogipalvelu Twitterin missio on toimia alustana, jotta kaikilla olisi mahdollisuus luoda ja jakaa ideoita ja informaatiota välittömästi, ilman esteitä (TwitterInc., 2019). Twitter on pysynyt yhtenä suosituimmista ja merkittävimmistä sosiaalisista medioista varmasti juuri tästä syystä. Mikään muu sosiaalinen media ei ole niin keskittynyt välittömään ja vaivattomaan tiedon jakamiseen ja levittämiseen. Vaikka käyttäjämäärät eivät ole samaa luokkaa kuin Facebookilla tai Instagramilla, tiedonvälityskykynsä ansiosta Twitter on edelleen relevantti julkaisualusta. Twitterin tavaramerkkinä voidaan pitää sen #-merkkiä eli hashtagia, joilla liitetään twiitti johonkin kontekstiin esimerkiksi johonkin tapahtumaan. Hashtagien avulla käyttäjät automaattisesti kategorisoivat viestinsä, mikä helpottaa samasta aiheesta kiinnostuneita löytämään lisää twiittejä aiheesta. Tämän avulla myös analysointi helpottuu, kun kaikki data on lähtökohtaisesti korvamerkittyä. Yrityksien osakkeista keskusteltaessa käytetään #-merkin sijaan \$-merkkiä, esim. \$GOOG Googlen osakkeen kohdalla, (Calore, 2009) mikä on mielenkiintoista tämän kandidaatin tutkielman kannalta sillä, jos twiitit ovat jo alustavasti kategorisoitu, niiden analysoiminen helpottuu merkittävästi.

Kun informaation välittäminen on vaivatonta ja välitöntä, varsinkin kun julkaisija on yksityishenkilö, voi hänen välittämäänsä viestiin sisältyä enemmän informaatiota, kuin julkaisija välttämättä haluaisi kertoa. Viesteihin voi sisältyä alitajuinen eli subliminaalinen viesti eli esimerkiksi sentimentin eli julkaisijan tunnetilan paljastava tieto. Sentimentti voi myös välittyä twiiteistä täysin avoimesti ja tarkoituksellisesti.

Twitterin avoimuuden ansiosta, kun käytännössä kaikki twiitit ovat avoimesti kaikkien luettavissa, voi twiittejä käyttää analyysin raaka-aineena. Dataa voidaan pitää strukturoimattomana tekstidatana, kun käsitellään vain twiittien tekstiosuutta. Analyysiä

keventävä seikka on se, että #- tai \$-merkkien avulla voidaan jo karsia analysoitavaa dataa mielekkääksi. Analyysi voisi esimerkiksi keskittyä yhden osakkeen ympärille kehkeytyneeseen avoimeen keskusteluun. Tutkimuskysymys voisi olla ”Mikä tunnetila osakkeeseen liittyy?”, jonka vastausta voitaisiin käyttää päätöksenteon tukena osakemarkkinoilla. Aiemmat tutkimukset antavat syyn olettaa, että Twitterin sentimentti korreloi osakemarkkinoiden kanssa. Bollen (2011) totesi, että päivittäiset muutokset sentimentissä korreloivat tilastollisesti merkittävästi Dow Jones Industrial Average (DJIA:n) päivän päätöskurssin kanssa.

3.2 Nopeus

Twitterin tuottama datavirta on valtava, keskimäärin yli 500 miljoonaa twiittiä päivässä (Live stats, 2019). Tutkimuksen kannalta siitä tekee mielenkiintoisen se, että twiittien pyrkimys on lähtökohtaisesti olla mahdollisimman lyhyitä ja tiiviitä informaation sisällöltään. Twitterin tuottama data on tästä huolimatta volyymiltaan, vaihtelevuudeltaan ja vauhdiltaan big datan kriteerit täyttävää.

Twitterin datavirran voisi ajatella olevan kuin vesiputous, jonka virtaus on niin voimakasta ja suurta, ettei veden kerääminen talteen ole käytännössä järkevää tai edes mahdollista. Veden, tai tässä tapauksessa datan, kerääminen talteen veisi valtavasti resursseja ja keräämisen jälkeen kerätyn datan merkittävyys vähenisi radikaalisti, sillä varsinkin päätöksenteossa osakemarkkinoilla oikea ajoitus on äärimmäisen tärkeää. Siksi on tärkeää, että analyysi tapahtuisi lähes reaaliajassa, ikään kuin datavirrasta siivilöitäisiin informaatio datan syntymishetkellä ja analyysin muodostamaa tietoa käytettäisiin päätöksenteon tukena.

Datamäärien räjähdysmäisen kasvun seurauksena on syntynyt kysyntää uudentyyppisille analyysikeinoille. Kuten osakemarkkinoilla, oikeaan aikaan toimiminen on elintärkeää, jos motiivina on pärjätä muita paremmin. Tämän kandidaatintutkielman tutkimuskysymykset nivoutuvat juuri reaaliaikaisen analyysin ympärille, sillä tarkoitus tutkia ennustamisen mahdollisuutta, eikä niinkään tehdä tutkimusta jo aiemmin tapahtuneesta.

Reaaliaikainen analyysi asettaa suuremmat odotukset tiedonsiirrolle ja laskentateholle kuin säilytystilalle. Käytännössä tämä tarkoittaa pilvipalveluiden hyödyntämistä analyysin

työstämisessä. Pilvipalveluiden arkkitehtuurin skaalautuvuus sopiikin oleellisesti reaaliaikaiselle analyysille, sillä sekin on luonteeltaan skaalautuvaa (Das et al., 2018).

3.3 Seuraaminen Twitterissä

Twitterin yksi keskeisimpiä ominaisuuksia on mahdollisuus seurata muita käyttäjiä omien preferenssien mukaan. Seuraamalla muita käyttäjiä saa etusivulle eli feedilleen näkyviin seurattujen käyttäjiensä tuoreimmat julkaisut. Näin seuraaja pysyy ajan tasalla seurattujensa toimista. Seuraamistoiminnon lisäksi jakamistoiminto on merkittävä informaation leviämisen kannalta. Jaettaessa julkaisun levikki kasvaa suuresti sillä jakajan seuraajat näkevät julkaisun ja niin edelleen. Käyttäjillä, joilla on paljon seuraajia, on lähtökohtaisesti suurempi todennäköisyys saada ajatuksilleen huomiota.

3.4 StockTwits

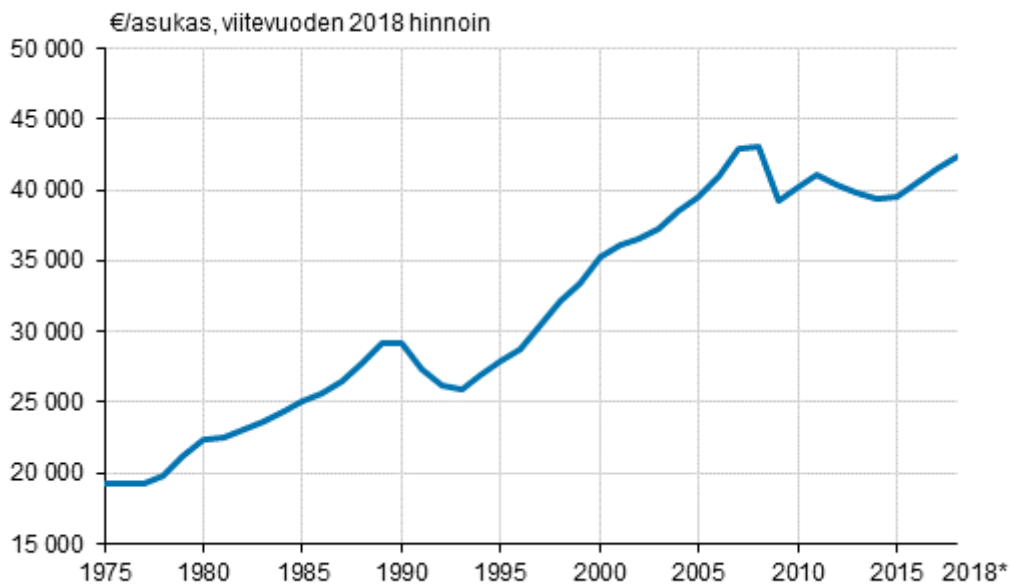
StockTwits on vuonna 2009 perustettu, Twitterin kaltainen, mikroblogipalvelu, jonka tarkoitus oli toimia kuten Twitter, mutta käyttäjien käymä keskustelu liittyisi lähinnä osakemarkkinoihin. Twitter väitetyesti kaappasi StockTwitsin toiminnan ydinajatuksen vuonna 2012 lisäämällä StockTwitsin omintakeisen \$-tagin omaan sovellukseensa. StockTwitsin tuottamaa aineistoa on käytetty lukuisissa tutkimuksissa, kun on tutkittu sosiaalisen median ja osakemarkkinoiden välistä yhteyttä. (Ryan, 2013) StockTwits vaikuttaa jääneen täysin Twitterin jalkoihin, mutta StockTwitsistä tutkimuksissa tehdyt johtopäätökset todennäköisesti pätevät myös Twitterissä.

4 OSAKEMARKKINAT

Ihmiset ovat käyneet kauppaa kautta aikojen, milloin milläkin kaupankäyntivälineillä. Osakemarkkinoilla tarkoitetaan pääomamarkkinoiden osaa, johon kuuluu primäärimarkkinat eli emissiomarkkinat ja sekundääri- eli jälkimarkkinat. Emissio tarkoittaa arvopaperien liikkeelle laskemista ja jälkimarkkinoilla tarkoitetaan sijoittajien välistä kauppaa arvopapereilla. Modernin ajan osakemarkkinat perustuvat lähes poikkeuksetta sähköiseen pörssiin, jossa kaupankäynti on paljon vaivattomampaa kuin aikana, jolloin osakkeet ja muut arvopaperit olivat fyysisiä asiakirjoja.

Yksi suurimmista syistä miksi ihmiset tai instituutiot osallistuvat osakemarkkinoille on se, että osakemarkkinat tarjoavat keinon vaurastua. Suurin osa ihmisistä sijoittaa osakemarkkinoille lisätuottojen toivossa. (Garvey, 2005) Kun kyseessä on keino luoda vaurautta, syntyy luonnollisesti sen ympärille kilpailua. Osakemarkkinoita voidaan pitää äärimmäisen kilpailtuina ja markkinoiden voittamista säännöllisesti pitkällä aikavälillä mahdottomana, ainakin teorian mukaan. Osakemarkkinoiden on väitetty olevan nollasummapelejä eli kun joku saa suuremman palan kakusta, niin se automaattisesti tarkoittaisi sitä, että joku muu saa pienemmän palan jaettavasta kakusta. Nollasummapelejä-teoria ei kuitenkaan historian valossa ole pitänyt paikkansa ainakaan pitkän aikavälin sijoittajille, sillä jaettavan kakun koko on keskimääräisesti kasvanut (Kuva 3) ja sijoittajat saavat sijoitukselleen esimerkiksi osinkoja, mikä ei ole pois muilta sijoittajilta. Huomion arvoista on kuitenkin se, että talouskasvu on edellytys kasvavalle kakulle.

Osakemarkkinat yleisesti ovat äärimmäisen kompleksi kokonaisuus, mikä tekee niiden täysimääräisestä ymmärtämisestä todella haastavaa sekä herättää kysymyksen, onko niiden ennustaminen vain valistunutta arvailua. Osakemarkkinat ovat riippuvaisia niin poliittisista päätöksistä, makrotaloudesta, ilmastosta sekä muista itsestään riippumattomista tekijöistä. Osakemarkkinoihin liittyy siis paljon epävarmuutta, joka lisääntyy ennustushorisontin kasvaessa.



Kuva 3 Esimerkiksi Suomen BKT asukasta kohden on kasvanut merkittävästi vuosien 1975-2018 välillä

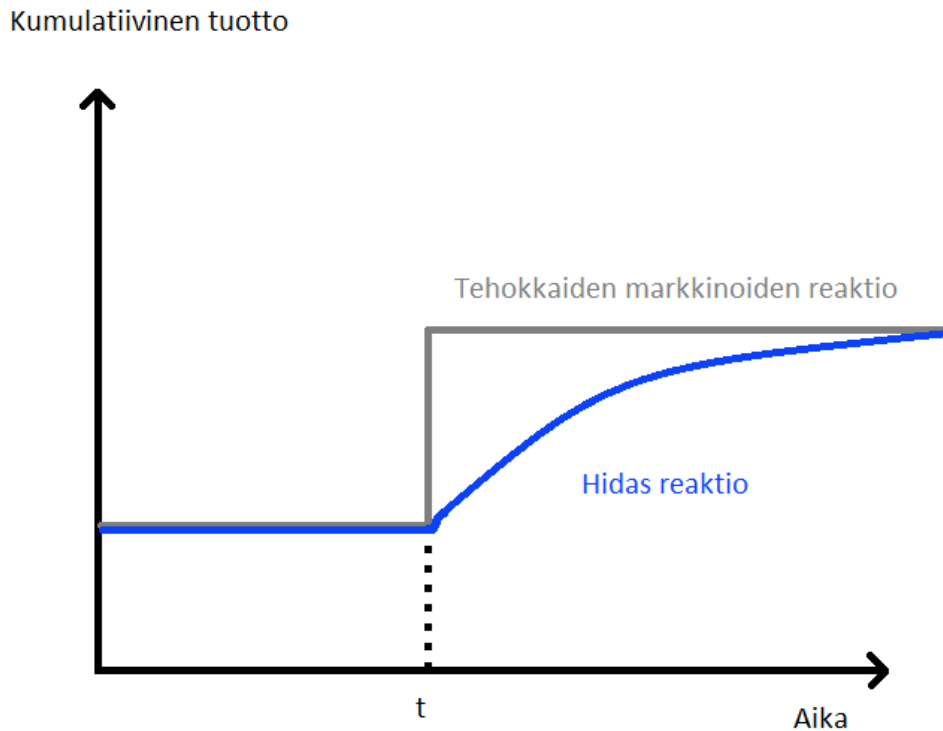
(https://www.tilastokeskus.fi/static/media/uploads/tup/suoluk/kuviot/suomi_kan002.gif)

Viitattu 12.4.2019

4.1 Markkinoiden tehokkuus

Rahoitusmarkkinoista puhuttaessa ei voi olla törmäämättä tehokkaiden markkinoiden -teoriaan (eng. Efficient-market Hypothesis, EMH). EMH on amerikkalaisen ekonomistin, Eugene Faman, esittelemä teoria rahoitusmarkkinoiden tehokkuudesta. Rahoitusmarkkinoilla liikkuu päivittäin valtavasti rahaa, josta desimaalin kymmenesosatkin ovat miljoonia euroja (Knüpfer & Puttonen, 2018). Ihmisten ikaikainen haave on ollut saada edes murto-osa pelipöydältä putoavista pelimerkeistä. Rahoitusteoriassa yleisenä perusoletuksena on se, että rahoitusmarkkinat ovat tehokkaat. ”Käytännössä tällä tarkoitetaan sitä, että kaikki julkinen ja relevantti tieto heijastuu osakkeiden hintoihin välittömästi, koska sijoittajia on valtavasti ja kaikki julkinen tieto on silmänräpäyksessä kaikkien saatavilla.” (Knüpfer & Puttonen, 2018 s. 166) Teoria siis olettaa, että reaali maailmassakin relevantti informaatio aiheuttaa hintaan reaktion, joka tapahtuu välittömästi, on oikean suuntainen ja suuruinen. Teoria tekee myös oletuksen, että sijoittajat toimivat aina rationaalisesti, kaiken informaation hyväksi käyttäen, ilman tunteiden tai ennakkoasenteiden vaikutusta päätöksentekoon. Alla olevassa kuvaajassa

(Kuva 4) yksinkertaistettu ero hinnan reagoinnista uuteen tietoon, tehokkaiden ja vähemmän tehokkaiden markkinoiden välillä.



t = Hetki, jolloin hintaan vaikuttavaa relevanttia informaatiota tulee kaikkien saataville

Kuva 4 Tehokkaiden ja ei-tehokkaiden markkinoiden reaktio positiivisesti hintaan vaikuttavaan informaatioon. (Knüpfer & Puttonen 2018 s. 166)

Rahoitusmarkkinat ovat varmasti tehokkaammat kuin esimerkiksi kuluttajamarkkinat, sillä informaatiota on laajalti ja helposti saatavilla. Kuitenkin jo rahoitusteorian määritelmän perusteella voi intuitiivisesti olettaa, että rahoitusmarkkinat eivät ole ainakaan aina täysin tehokkaat, sillä markkinoilla päätöksiä tekevät pääsääntöisesti ihmiset.

”Markkinoiden tehokkuus ei kuitenkaan edellytä, että markkinahinta on koko ajan sama kuin sijoituksen todellinen arvo. Markkinoiden tehokkuus tarkoittaa sitä, että markkinahinnan poikkeamat todellisesta arvosta ovat satunnaisia, ts. markkina-arvo voi olla oikean arvon ylä-

tai alapuolella, mutta näiden poikkeamien tulee olla sattumanvaraisia ja siten ennustamattomia.” (Knüpfer & Puttonen, 2018 s. 170)

Yhdenkään sijoittajan tai sijoittajaryhmän ei pitäisi pystyä löytämään toistuvasti ali- tai yliarvostettuja osakkeita, jos markkinahinta poikkeaa satunnaisesti yhtiön oikeasta arvosta. Tehokkailla markkinoilla ei yksikään sijoittaja pysty millään investointistrategialla saavuttamaan ylisuuria tuottoja. Ylisuurilla tuotoilla tarkoitetaan tuottoja, jotka ovat suurempia kuin mitä pelkkä sijoituksen riski antaisi olettaa. (Knüpfer & Puttonen, 2018) Yleisen rahoitusteorian mukaan, kukaan ei voi säännöllisesti voittaa indeksiä eikä aktiivisella kaupankäynnillä tai osakepoiminnalla voi voittaa indeksiä, sillä teorian mukaan saadut hyödyt kumoutuisivat kaupankäyntikuluilla ja analyysityöhön käytetyillä resursseilla.

Jotta tutkimus aiheesta ei olisi täysin ristiriidassa rahoitusteorian kanssa, täytyy tehdä oletus, ettei rahoitusteorian esittämä tehokkaiden markkinoiden hypoteesi pidä täydellisesti paikkansa. Osakemarkkinat ovat varmasti yksi tehokkaimmista markkinoista, johon tavallisilla ihmisillä on keino osallistua, mutta markkinoilla on oletuksen mukaan myös toistuvaa tehottomuutta. Onnistuneella analyysillä voisi siis olla vaikutusta osakemarkkinoilla pärjäämiseen. Kuitenkin huomionarvoista on se, että tehokkailla markkinoilla voi olla toimijoita, jotka toistuvasti voittavat markkinat, mutta näiden taustalla ei olisi teorian mukaan mikään muu kuin hyvä tuuri.

4.2 Markkinoiden voittaminen

Kun puhutaan osakemarkkinoiden voittamisesta, tarkoitetaan tilannetta, jossa sijoittajan tuotot ovat enemmän kuin markkinoiden yleiskehityksen eli markkinoiden vertailuindeksin tuotto. Osakeindeksin tarkoitus on kuvata tietyn markkinan yleistä kehitystä, indeksi muodostetaan ottamalla esimerkiksi 25 markkina-arvoltaan suurinta yritystä ja painotetaan niiden vaikutus indeksiin niiden markkina-arvojen perusteella. Näin esimerkiksi OMX Helsinki 25-indeksi on muodostettu.

4.2.1 Ylituotot informaatioedulla

Pankkiiriliike Nordnetin #rahapodi-podcastissa talousasiantuntija Martin Paasi havainnollisti hyvin sen, että ylituotot osakemarkkinoilta ovat kiven alla. Hän muistutti, että markkinoilla on aina muita ihmisiä seuraamassa samoja yrityksiä, varsinkin yrityksen koon kasvaessa analysoijien määrä lisääntyy. Informaatioedun saavuttamiseksi tulee tietää tai ymmärtää jotain mitä muut eivät vielä ole tajunneet sekä informaation tulee olla merkittävää ja julkista. Kuitenkaan oikeassa oleminen ei ole tae siitä, että sijoittaja tekisi ylituottoa sillä sijoittajan täytyy myös toimia oikeaan aikaan sekä hän on luonnollisesti riippuvainen markkinoiden muiden sijoittajien toiminnasta. Jotta joku voi ostaa tai myydä osakkeen ali- tai ylihintaan, on jonkun toisen myytävä tai ostettava sama kyseinen osake ali- tai ylihinnalla. (Luukkonen & Paasi, 2019)

4.2.2 Aktiivinen kaupankäynti

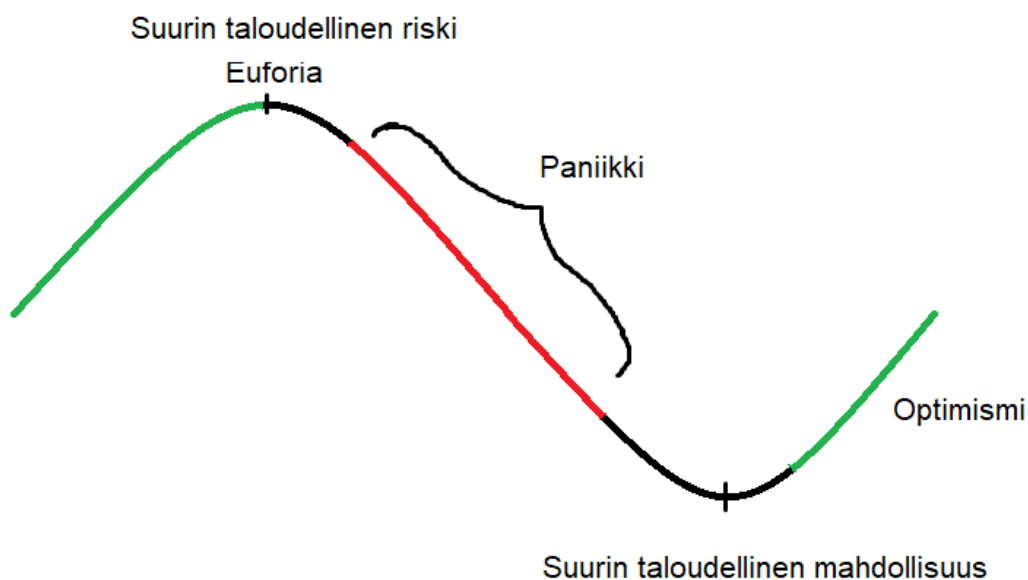
Käytännössä ylituottojen saavuttaminen on mahdollista aktiivisella kaupankäynnillä, olettaen että markkinat eivät toimi tehokkaasti. Markkinoiden tehottomuudesta ei pääse samalla tavalla hyötymään passiivisella osta & pidä- sijoitusstrategialla. Aktiivista kaupankäyntiä harjoittavat lähinnä ammattilaissijoittajat sekä instituutiot, ja siten yksityissijoittajan lähtökohdat näille edellä mainituille tahoille pärjäämiselle ovat heikot.

4.3 Tunteet osana päätöksentekoa

Tehokkaiden markkinoiden teorian mukaan ihmiset toimivat aina rationaalisesti markkinoilla, perustaen päätöksenteon faktoihin ja todennäköisyyksiin. Aiheesta on tehty lukemattomasti tutkimusta ja näiden tulokset viittaavat siihen, että ihmiset eivät aina toimi rationaalisesti. Mamun et al. (2015) totesivat artikkelissaan, ettei voi olettaa, että sijoittajat toimisivat aina rationaalisesti ja varsinkin irrationaalisuuden osuutta markkinakuplien syntymisessä tulisi tutkia enemmän. Irrationaalinen optimismi, euforia, markkinoilla on edesauttanut kuplien syntymistä. Voi olla, että tunteet näkyvät selkeämmin lyhyellä aikajänteellä ja pidemmällä ajanjaksolla tunteiden aiheuttama kurssiheilahtelu normalisoituu. Kuten tehokkaiden markkinoiden teoriassa todettiin, kurssi voi poiketa osakkeen todellisesta arvosta ja se on

normaalia, mutta poikkeamat ovat täysin satunnaisia. Voisi kuitenkin olettaa, että nämä lyhyen ajan kurssiheilahtelut eivät ole täysin satunnaisia eli ne voisivat olla ennustettavissa. Tämän seurauksena on hyvä pohtia, onko markkinoiden sentimentin ennustaminen mahdollista esim. Twitteristä.

Sijoittajien toiminnassa on huomattu samanlaista kaavaa toistavaa laumakäyttäytymistä. Optimistinen sentimentti tarttuu sijoittajiin ja kurssi nousee toisinaan liiankin korkeaksi. Kurssin kääntyessä laskuun seuraa lyhyt kieltämisen aika siitä, ettei kurssi voi oikeasti myös laskea, mutta jatkaessaan laskua, syntyy paniikki, jolloin arvopaperia myydään jopa alihintaan. Suurin taloudellinen mahdollisuus piileekin juuri kurssin pohjalla, mutta oston ajoittaminen on haastavaa. (Kuva 5)



Kuva 5 Yksinkertaistettu kuvaus kurssiromahduksesta ja sijoittajien tunnetiloista.

4.4 Fundamenttianalyysi

Fundamenttianalyysi on analyysityökalu, joka yrityksen arvonmäärittämiseen reaalimaailman tekijöiden kautta. Tämä tarkoittaa siis yrityksen tilinpäätöksen, taseen, osinkopolitiikan ja tulostilauksien tarkastelua. (Nikkinen, Rothovius & Sahlström, 2008) Fundamenttianalyysin voidaan siis ajatella pitkälti faktoihin perustuva analyysityökalu. Erään määritelmän mukaan fundamenttianalyysi perustuu 90 prosenttisesti järkeen ja 10 prosenttisesti psykologiaan (Malkiel, 1999).

Tämän puhtaasti lukuihin ja tosiasioihin perustuvan analyysimenetelmän rinnalle voisi kuvitella analyysityökalun, joka ennustaa yrityksen näkymiä toisella perusteella, kuten sosiaalisen median sentimenttiin perustuvan analyysin näkökulmasta.

4.5 Lyhyeksi myyminen

Osakemarkkinoilla rahan tekeminen ei aina edellytä sitä, että osakekurssit nousisivat. Enimmäkseen kokeneempien sijoittajien työkalupakkiin kuuluu lyhyeksi myyminen, tai tarkemmin suomennettuna, vajaaksi myyminen eli shorttaus. Lyhyeksi myymisellä tarkoitetaan tilannetta, jossa sijoittaja myy osakkeita, joita sijoittaja ei omista vaan lainaa joltain muulta taholta ja jotka sijoittaja joutuu ennen pitkää ostamaan takaisin. (Knüpfer & Puttonen, 2018) Shorttaaminen tulee siis kysymykseen silloin, kun sijoittaja uskoo kurssin laskevan, ja tällöin hän tekee voittoa lainattujen osakkeiden myyntihinnan ja takaisin ostohinnan erotuksen verran. Lyhyeksi myyminen ei kuitenkaan ole riskitöntä, sillä teoreettisesti mahdollinen tappio, osakekurssin noustessa, on ääretön. Teoreettinen maksimivoitto taas on vain 100 % sillä osakkeen arvo ei voi laskea nollan alapuolelle.

5 BIG DATAN HYÖDYNTÄMINEN OSAKKEEN HINNAN KEHITYKSEN ENNUSTAMISESSA

Kun puhutaan osakemarkkinoista, puhutaan automaattisesti myös rahasta. Maailman suurimman, New Yorkin pörssin, markkina-arvo oli vuonna 2018 lähes 30 biljoonaa eli 30 000 miljardia Yhdysvaltain dollaria, mikä on enemmän kuin Yhdysvaltojen bruttokansantuote (Coyne, 2017). Ei siis mikään ihme, että kilpailijoita edes tuottojen murto-osista riittää. Osakemarkkinat ovat äärimmäisen kilpailtu ja kompleksi kokonaisuus, johon vaikuttaa valtava määrä toisistaan riippuvia ja riippumattomia tekijöitä.

Tietotekniikan sovellusten kehittyessä, on syntynyt uusia informaation jakamisalustoja, kuten esimerkiksi sijoituskeskustelupalstat ja Twitter. Ihmiset jakavat ajatuksiaan ja mielipiteitään melko avoimesti, joko avoimesti tai anonyymisti. Twitterin päivittäinen twiittivirta on jo sen verran valtava, että sitä voidaan käyttää aineistona pienienkin muutosten havaitsemiseen esimerkiksi markkinoiden sentimentissä.

5.1 Big data- analytiikka

Big data-analytiikka on termi, joka varmasti jää historiaan yhtenä 2010-luvun hehkutetuimmista trendisanoista. Big data-analytiikka ei kuitenkaan ole pelkkää puhetta, vaan sen hyödyntämistä testataan ja sovelletaan mitä erilaisimmissa teollisuuden ja liike-elämän sovelluksissa. Big data-analytiikka eroaa tavallisesta data-analytiikasta juuri volyyminsa, vaihtelevuutensa ja vauhtinsa vuoksi. Tämä asettaa uuden tason sille, miten kasvavasta, monipuolistuvasta ja kiihtyvästä datavirrasta saadaan siivilöityä hyödyllistä informaatiota. Tässä kappaleessa on käsitelty tutkimuksia, jotka ovat keskittyneet Twitterin lisäksi, sen kaltaisen StockTwits-palvelun tuottaman datan analysoimiseen.

5.1.1 Sentimenttianalyysi

Yksi keino selvittää yleinen tunnetila jostakin tietystä aiheesta tai asiasta twiittimassasta on sentimenttianalyysi. Yleisimmin aiheesta tehdyissä tutkimuksissa on käytetty Bag-of-words-

mallia. Mallin tarkoitus on yksinkertaistaa ja kiteyttää luonnollisen kielen sanojen merkitys tai sisältämä tunnetila. Sanat pisteytetään esimerkiksi siten, että positiivissävytteiset sanat pisteytetään +1 pisteellä, neutraalit sanat ovat 0 pistettä ja negatiiviset sanat ovat -1 pistettä. Juuri edellä mainitulla tavalla Oh määritteli tutkimuksessaan vuonna 2011 StockTwits-palvelusta ladatut osakkeita koskevat twiitit. Oh määrittelivät ensin manuaalisesti n. 10 prosenttia aineistosta kolmeen eri kategoriaan twiittien välittämän sentimentin perusteella. Kategoriat olivat ”bullish” eli optimistinen sentimentti, neutraali ja ”bearish” eli pessimistinen sentimentti markkinoilla. Jaottelun tarkoituksena on toimia koneoppimisalgoritmin perustana. Tutkimuksessa käytettiin myös sanastollista analyysiä (eng. Lexical analysis) sekä näiden kahden yhdistelmää. BOW-malli osoittautui kuitenkin tarkimmaksi, vaikka ero olikin pieni menetelmien välillä. (Oh, 2011)

Zhang (2011) tutki aihetta myös vuonna 2011, hänen tutkimuksensa keskittyi enemmän pessimistisiin twiitteihin ja näiden vaikutuksiin seuraavan päivän osakekursseihin. Twiitit jaoteltiin sen perusteella, oliko niissä tunnetilaa kuvaavia sanoja kuten iloinen, toiveikas tai pelokas. Heidän tutkimustulostensa perusteella, kun ihmiset ilmaisevat paljon tunteita, positiivisia tai negatiivisia, DJIA-indeksi (Dow Jones Industrial Average) laskee seuraavana päivänä. Kun taas vähemmän tunnepitoisen päivän jälkeen indeksillä on taipumus nousta. Zhangin tutkimus oli suppeahko, mutta korrelaatio oli selkeää etenkin tunnepitoisilta päiviltä.

Coynen, Madirajun ja Coelhon vuonna 2017 julkaistussa konferenssijulkaisussa oli verrattain enemmän viiteaineistoa verrattuna vuoteen 2011. Heidän tutkimusaineistonsa oli samasta lähteestä, StockTwitsistä, kuten Oh:n vuoden 2011 tutkimuksessa. Tässä tuoreimmassa tutkimuksessa otettiin kantaa aiemmin huomiotta jääneen seikkaan, siihen että jaoteltaessa twiitit esimerkiksi juuri positiivisiin ja negatiivisiin kategorioihin twiitit, joilla ei ole mitään todellista sentimentaalista informaatioarvoa voivat vääristää tutkimustulosta. (Coyne, 2017) He nimittävät näitä twiittejä kohinaksi, jota on valtavasti tutkimusaineistossa. Tutkijoiden mukaan, mitä paremmin tai perusteellisemmin ensisijainen lajittelu kategorioihin tehdään, sitä tarkempi lopullisen analyysin tulos usein on. Jaotellessaan twiittejä tutkijat tulivat tulokseen, että suurin osa twiiteistä on joko sentimentiltään merkityksettömiä tai neutraaleja. Toinen löydös oli se, että twiittaaajat julkaisivat joko täyttä kohinaa tai täysin ennakoivia twiittejä. (Coyne, 2017) Tästä herää ajatus, että suodattamalla kohina pois analysoitavasta aineistosta,

voitaisiin saavuttaa paljon tehokkaampia ja tarkempia ennusteita lähitulevaisuuden osakemarkkinoista. Tulevaisuudessa tulemme varmasti näkemään sofistikoituneempia, itseoppivia algoritmeja, jotka pystyvät tehokkaasti erottelamaan relevantit twiitit kohinasta.

Vuonna 2011 tehdyssä tutkimuksessa Bollen tutki suurta Twitteristä haalittua twiittijoukkoa. Tutkimuksen tarkoitus oli selvittää, onko Twitterin yleisellä sentimentin ja DJIA-indeksin välillä korrelaatiota. He käyttivät sentimentin selvittämiseen kahta eri tunnetilan selvittämiseen tarkoitettua työkalua, OpinionFinderia ja Google-Profile of Mood States:ia (GPOMS). OpinionFinder luokittelee mielipiteet positiiviseksi tai negatiiviseksi, kun taas GPOMS oli moniulotteisempi sen luokittellessa twiitit kuuden tunneulottuvuuden perusteella. Nämä ulottuvuudet olivat rauhallinen, valpas, varma, vitaali, ystävällinen ja iloinen. Tutkimuksessa tutkittiin korrelaatiota validoimalla menetelmät ristiin, mutta huomattiin, että ainoastaan GPOMS:n tunne ”rauhallinen” oli ennakoiva DJIA-indeksin lähipäivien kehitykselle. Parhaimmillaan saatiin 86,7 prosentin ennustustarkkuuden seuraavan päivän indeksikehitykselle, mutta tutkimusryhmä tiedosti analyysinsä heikkoudet. Analyysi ei ottanut huomioon twiittaajien maantieteellistä sijaintia, vaan otti mukaan kaikki englannin kieliset twiitit. (Bollen, 2011) Nykyään tätä ongelmaa ei ole, sillä twiitit ovat jaoteltavissa sijainnin perusteella. Tutkimuksessa esitettiin myös, että sosiaalinen media ei välttämättä kata täysin koko markkinan tunnetilaa.

5.1.2 Älykkäät twiittaajat

Coynen tutkimuksessa tehtiin oletus, etteivät kaikki twiitit ole painoarvoltaan samanarvoisia. Tämä oletus on hyvin intuitiivinen, sillä on helppo uskoa, että huippusijoittajan statuksesta nauttivan sijoittajan mielipide aiheuttaa suuremman reaktion kuin keskivertokäyttäjän.

He nimittivät tutkimuksessaan uutta analyysikeinoaan älykkään käyttäjän -malliksi. Malli perustuu yksinkertaisuudessaan siihen, että se pisteyttää käyttäjät sen perusteella, kuinka usein käyttäjän twiittien sisältämä ennustesentimentti osoittautuu oikeaksi. Coynen tutkimuksessa älykkään käyttäjän statuksen sai, kun käyttäjä julkaisi 6-8 sisällöllisesti merkittävää twiittiä yhdeksän kuukauden tarkastelujakson sisällä ja ollessaan vähintään 80 prosenttia ajasta

oikeassa. (Coyne, 2017) Kun tätä huippujoukkoa tutkittiin, huomattiin paljon suurempaa osumatarkkuutta siinä, kuinka usein tämän huippujoukon twiittien välittämä sentimentti ennustaa seuraavien päivien osakkeen hintakehityksen oikein.

Kuitenkin tälläkin mallilla on omat rajoitteensa, sillä otantajoukko on pieni ja älykkäät käyttäjät eivät julkaise mielipiteitään usein ja säännöllisesti, joten päiviä jolloin osakkeiden kurssin kehitystä ei edes voida yrittää ennustaa, tulee paljon. Malli vaatii myös sen, että twiitit lajitellaan ainakin aluksi manuaalisesti, joko optimistiseksi tai pessimistiseksi, jotta voidaan todeta, onko käyttäjän twiitti ennustanut oikein vai väärin osakkeiden hintakehityksen. Luonnollisesti ainoastaan äärimmäisen yleisiä osakkeita, joista on valtavasti keskustelua, voidaan käyttää tässä mallissa, sillä suurin osa twiiteistä suodattuu pois niiden kohinapitoisuuden takia. Pienemmissä ja vähemmän keskustelluista osakkeista on muutenkin vaikeampi löytää niistä twiittaavia älykkäitä käyttäjiä. (Coyne, 2017)

Käytännössä tämä tarkoittaa sitä, että tämän hetkistä mallia voidaan kuvitella käytettäväksi ainoastaan suurten yritysten osakkeiden seurantaan ja ennustamiseen. Luonnollisesti juuri suurten yritysten osakkeita seurataan ja analysoidaan jatkuvasti suuren yleisön ja suurten analytikkotalojen sekä instituutioiden, kuten eläkerahastojen, toimesta. Näiden osakkeiden markkinoiden voitaisiin olettaa olevan todella tehokkaat, joten ylisuurten tuottojen saaminen on todennäköisempää pienemmistä yrityksistä, joita ei seurata niin tarkasti. Älykkään käyttäjän malli on kuitenkin lupaava, sillä jos käyttäjät, jotka ovat keskimäärin ennustuksissaan useammin oikeassa, voidaan valjastaa päätöksenteon tueksi, voidaan saavuttaa kilpailuetua markkinoihin nähden.

5.1.3 Kritiikki sentimenttianalyysin tarkkuudesta

Kaikki tutkimukset eivät ole täysin linjassa jo esiteltyjen tutkimuksien kanssa. Oliveira, Cortez ja Areal pyrkivät vuonna 2013 tekemässään tutkimuksessaan selvittämään aikaisempien tutkimuksien paikkansapitävyyden.

He tekivät aiempia tutkimuksia vankemman tutkimuksen, jossa StockTwitsistä saatu pohja-aineisto oli aiempaa laajempi ja heidän tutkimuksessaan osakkeen hinnan kehityksen lisäksi

seurattiin osakkeiden tuoton keskihajontaa eli volatilitteettia ja vaihtovolyymien suhdetta twiitteihin. Heidän tutkimusmenetelmänsä olivat heidän mielestään kattavammat, ja menetelmässä oli yhden virhemittarin, MAPE, lisäksi toinen, RMSE. Tuloksiansa perusteella osakemarkkinoiden ennustaminen sosiaalisen median kautta on paljon monimutkaisempaa kuin aiemmissa tutkimuksissa annetaan ymmärtää. Heidän mielestään tulokset eivät ole yllättäviä, sillä rahoitusteorian mukaan tuottojen ennustaminen on mahdotonta. Heidän mielestään mielenkiintoinen tutkimustulos oli se, että he löysivät twiittien määrän ja osakkeiden vaihdantamäärien väliltä korrelaatiota. Tämä löydös oli merkittävä siksi, että he käyttivät aiempia tutkimuksia suurempaa otoskokoja ja tiukempaa analysointia. (Oliveira, Cortez & Areal, 2013)

Oliveiran, Cortezin ja Arealin löydökset ovat pitkälti linjassa rahoitusteorian kanssa siitä, ettei markkinoita voida ennustaa etenäkään pitkällä aikajänteellä. Aiemmissakin tutkimuksissa huomioitiin kylläkin se, että ennustettavuus oli korkeimmillaan vain 1-3 päivän päähän. Tästä voidaan tehdä johtopäätös, että sosiaalisen median sentimenttianalyysi tällä hetkellä, voi parhaimmillaan ennustaa korkeintaan lähitulevaisuutta.

5.2 Lyhyen aikavälin ennustaminen

Ekonomit ovat kehittäneet nykyhetken ja lähitulevaisuuden ennustamiselle termin nowcasting, mikä tulee englannin kielen sanojen, ”now” eli sillä viitataan nykyhetkeen ja sanan ”forecasting” eli ennustaminen, yhdistelmästä. Suomessa aiheesta puhutaan lyhyen aikavälin ennustamisena.

Aiheesta tehdyn tutkimuksien valossa sosiaalisen median viesteihin pohjautuva sentimenttianalyysi voi tarjota ennustustyökalun vain lähitulevaisuuteen. Tämä tarkoittaa sitä, että teoriassa ylituottoja markkinoilta on mahdollista saavuttaa, mutta se edellyttäisi aktiivista kaupankäyntiä.

Lyhyen aikavälin ennustamisen toimivuutta tukee se ajatus, että informaation epäsymmetrinen jakautuminen tasoittuu ajan kanssa, joten mitä aikaisemmin sijoittaja tiedostaa oikean tiedon, sitä suurempi taloudellinen mahdollisuus tällä on. Aktiivisessa kaupankäynnissä ajoitus on

ratkaiseva, minkä vuoksi aktiivista kaupankäyntiä käyvät lähinnä osakemarkkinoita todella tiukasti seuraavat sijoittajat.

5.3 Osana fundamenttianalyysiä?

Fundamenttianalyysi on yksi yleinen analyysityökalu, joka on vakiinnuttanut paikkansa sijoittajien keskuudessa. Fundamenttianalyysin vahvuutena on se, että se perustuu reaali maailman tapahtumiin, tilinpäätöksiin, yrityksen taloudesta kertoviin lukuihin ja ennusteisiin.

Kuten fundamenttianalyysin määritelmässäänkin todetaankin, fundamenttianalyysissä muiden ihmisten toiminta eli psykologia on osa yhtälöä. Sentimenttianalyysi yksinään ei välttämättä ole riittävän tarkka ennustamistyökalu, varsinkaan pitkällä aikavälillä, eikä se riittäne yksinään päätöksentekoon. Kuitenkin soveltamalla näitä kahta edellä mainittua analyysityökalua samanaikaisesti, voitaisiin mahdollisesti saavuttaa parempia tuloksia kuin näitä kahta erikseen käyttämällä. Fundamenttianalyysi ei kuitenkaan ole ainoa analyysityökalu, jonka tukena sentimenttianalyysi voisi toimia.

5.4 Case StockFluence

Pohdittaessa, voidaanko mikroblogin viestien sentimentin selvittämisellä luoda taloudellista lisäarvoa, tulee myös huomioida muutkin kuin suoraan aktiiviseen kaupankäyntiin pörssissä liittyvät toimet. Arvopaperipörssien rinnalle on syntynyt valtavasti siihen liittyvää liiketoimintaa kuten sijoitusaiheisia uutispalveluita ja analyysijä tuottavia tahoja.

Yksi osakemarkkinoista analyysijä tuottava yritys on hollantilainen StockFluence. StockFluence on vuonna 2012 perustettu yritys, jonka liiketoiminta perustuu useiden eri sosiaalisten medioiden seuraamiseen ja niiden tuottaman datan analysoimiseen. Datan analysoimisen tarkoituksena on selvittää yksittäisiin osakkeisiin tai kokonaisiin markkinoihin liittyvät tunnetilat ja luoda näin pohjalta analyysi. StockFluence tarjoaa asiakkailleen rahaa vastaan ennusteen asiakkaan mieleisestä osakkeesta. (Nieland, 2017a) StockFluence tuottaa

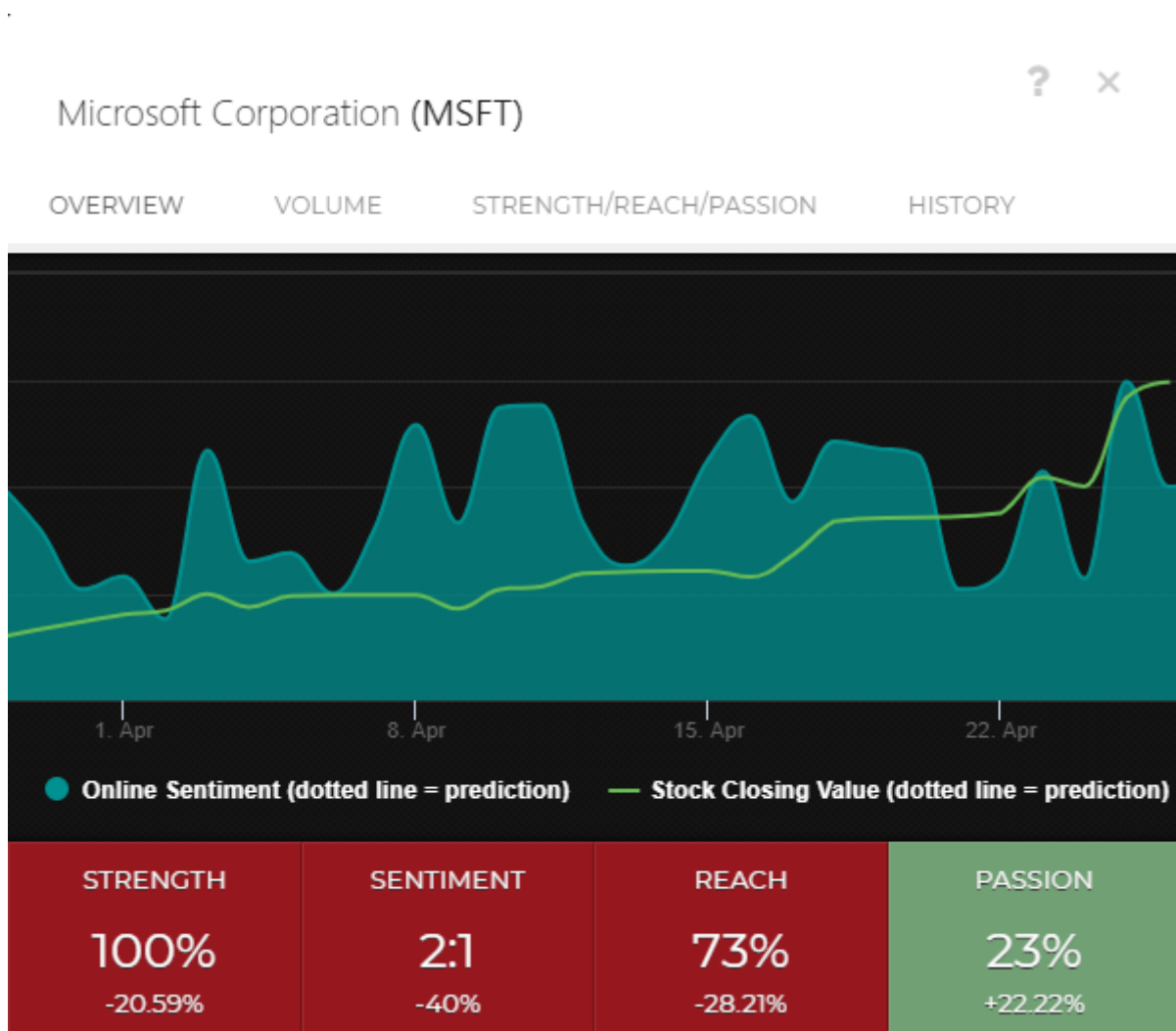
asiakkailleen kuvaajia, jotka kuvastavat osakkeen reaalihintakehitystä sekä pelkästään sentimentin indikoimaa hintakehitystä (Kuva 6). Maksavat asiakkaat saavat myös ennusteen lähitulevaisuuden hintakehityksestä.

StockFluence kertoo Internet-sivuillaan, että heidän analyysinsä ennakoivat osakkeiden arvot 70 prosentin tarkkuudella. Ennustetarkkuus on erittäin hyvä, kun ajatellaan, että yli 50 prosentin ennustetarkkuus kielii voitollisesta sijoitustoiminnasta pitkällä aikavälillä. Heidän tarjoamansa ennusteet ovat myös kohtuullisen hintaisia, sillä 400:llä Yhdysvaltain dollarilla saa vuodeksi rajoittamattoman pääsyn, jokaisen StockFluencen seuraaman osakkeen ennusteisiin (Nieland, 2017b).

Yrityksen liiketoimintamalli on nokkela, sillä heidän analyysinsä perustuvat täysin muiden ihmisten eli sosiaalisen median toiminnan seuraamiseen. Heidän ei itse tarvitse pohtia reaali maailmaan ja osakemarkkinoihin, suorasti tai epäsuorasti, liittyviä syy-seuraus-suhteita. Ainoastaan StockFluencen oma algoritmi on vastuussa sentimentin selvittämisessä.

Yrityksen verkkosivujen ja heidän lupaamiensa ennustetarkkuuksien perusteella kuitenkin herää viero epäily yrityksen uskottavuudesta. Yleisesti hyvänä pidetty nyrkkisääntö kuuluukin; ”Jos jokin asia kuulostaa liian hyvältä ollakseen totta, mitä todennäköisimmin se myös on liian hyvää ollakseen totta.” Vaikkei StockFluence ei olisi täysin yhtä hyvä kuin se mainospuheissaan antaa ymmärtää, on heillä mielenkiintoinen liiketoimintamalli, jonka yleisyys tulee todennäköisesti kasvamaan.

Taloudellisen lisäarvon sentimenttianalyysillä on houkutteleva liiketoimintaidea, sillä se perustuu puhtaasti muiden ihmisten sentimentin analysoimiseen. Liiketoiminnan voisi olettaa hyvin kevyttä organisaatiollisesti sekä kulurakenteeltaan ja sitovan vähän pääomaa. Analyysienkään ei tarvitse olla läheskään aina oikeassa, sillä analyysit perustuvat suurten massojen tunnetilojen seurantaan ja analysoimiseen. Ei ole mitenkään poikkeuksellista, että toisinaan suurtenkin massojen sentimentti on väärässä, suhteessa osakkeiden hintojen reaali kehitykseen.



Kuva 6 StockFluencen sentimenttianalyysin ennakoima hinnanvaihtelu Microsoftille on reaalihintakehityksen suuntaista, mutta voimakkaampaa.

(<https://www.stockfluence.com/>) Kuvakaappaus 27.4.2018

6 JOHTOPÄÄTÖKSET JA YHTEENVETO

Sosiaalisessa mediassa ja muissa Internetin avoimissa kanavissa jaetaan valtavasti, alati kasvavalla vauhdilla, tietoa. Tämän työn tarkoituksena oli pohtia, onko näistä valtavista informaatiovirroista, ensinnäkään siivilöitävissä tietoa populaation sentimentistä, saati voidaanko tätä tietoa käyttää eduksi osakemarkkinoilla.

Aiheesta on tehty aiemmin runsaasti tutkimuksia ja suuri osa näiden tutkimuksien tuloksista puoltaa väitettä, että sosiaalisista medioista, kuten Twitteristä tai StockTwitsistä, on mahdollista kvantitatiivisella analyysillä selvittää suurten massojen tunnetila eli sentimentti. Tutkimuksia on lukuisia juuri osakemarkkinoiden ja mikroblogipalveluiden käyttäjien sentimentin välisestä korrelaatiosta. Osakemarkkinat kiinnostanevat tutkijoita juuri niissä piilevän taloudellisen arvonluontipotentiaalinsa takia. Useissa tutkimuksissa todettiin, että tunnetila korreloi jollakin asteella vertailtavan indeksin kehitykseen, kun historiallisesti verrattiin tietyn ajanhetken sentimenttiä sen hetken kurssikehitykseen. Tutkimuksia kuitenkin kritisoiitiin siitä, että otosmäärä ja -aika olivat liian suppeita, ja että pitkällä aikavälillä korrelaatiota ei ole, koska kaikki vaihtelu on satunnaista, kuten tehokkailla markkinoilla tulisikin rahoitusteorian mukaan tapahtua.

Jotta analysoiminen ja halutun arvopaperin hinnankehityksen ennustaminen olisi ylipäätään mahdollista, vaaditaan määrällisesti valtavasti pohja-aineistoa eli toisin sanoen oikeiden ihmisten tulisi keskustella tästä nimetystä arvopaperista riittävän paljon. Tämä käytännössä asettaa rajoitteita sille, minkä kokoluokan yritysten kurssikehitystä voidaan tällä menetelmällä ennustaa. Pienistä ja keskisuurista yrityksistä ei tällä hetkellä, sosiaaliseen mediaan yksinään pohjautuvalla analyysillä, voida kovin luotettavasti ennustaa sitä, miten kurssi kehittyy, koska pohja-aineistoa on niukasti. Epäedullista ylituottoja etsivän sijoittajan kannalta on se, että ylituotot perustuvat informaation jakautumiseen epäsymmetrisesti, mitä tapahtuu todennäköisemmin juuri pienien ja keskisuurten pörssiyritysten kohdalla. Suuret yritykset ovat erityisen seurattuja ja analysoituja ja suuret pelurit, kuten eläkerahastot, ovat hyvin perillä osakkeen todellisesta arvosta. Nykyisen tiedon valossa on siis hyvin haasteellista saavuttaa ylituottoja osakepoiminnalla, joka perustuisi pelkästään tässä tutkimuksissa esitettyihin, ainoastaan sentimenttiin perustuviin, ennustamismenetelmiin.

Useissa tutkimuksissa vertailuindeksi oli pörssin yleisindeksi, joka kuvaa arvopaperimarkkinoiden kokonaistilannetta. Indeksiin vaikuttaa luonnollisesti useita eri yrityksiä eri toimialoilta, joten tämän indeksin ennustamiseen otettaisiin käytännössä kaikki kyseiseen arvopaperimarkkinaan koskeva spekulatio esim. Twitterissä. Bollenin tutkimuksessa saavutettiin 86,7 prosentin tarkkuus DJIA-indeksin seuraavan päivän muutoksen ennustamiselle, mikä tukee väitettä, että ennustaminen Twitterin kautta on mahdollista.

Coynen tutkimus toi uuden näkökulman ennustamiseen, sillä he profiloivat twiittaajat, jotka olivat historiallisesti olleet tarpeeksi usein oikeassa. Tulevaisuudessa tämä aspekti tulee saamaan lisähuomiota, sillä nämä niin sanotut älykkäät käyttäjät ovat usein oikeassa ennusteissaan. Älykkäiden käyttäjien valjastamisen analyysin tekemiseksi voisi tuottaa merkittävää taloudellista lisäarvoa. Toinen seikka mikä nousi profiloinnin seurauksena esille, on se, että sen avulla voidaan tunnistaa ns. maksettuja mainostajia tai botteja, joiden tarkoituksena on vaikuttaa pienten yritysten osakekursseihin mielipidevaikuttamisen kautta.

Tutkimuksissa ei suuremmin keskitytty kysymykseen, tarjoaako sosiaalinen media mitään uutta informaatiota, vai onko se vain yksi suodatin lisää uutisten ja muiden reaali maailman tapahtumien välille. Sijoittajien sentimentin voidaan olettaa pohjautuvan reaali maailman tapahtumiin, yleiseen markkinatilanteeseen ja julkisesti saatavaan informaatioon. Mielenkiintoista olisi tietää, onko tulevaisuudessa, laskentatehon ja tekoälyn kehittyessä, mahdollista ennustaa se, miten ihmismassat tulevat reagoimaan ja miten sentimentti tulee muuttumaan reaali maailman tapahtumien summan seurauksena, reaaliajassa. Tämä käytännössä olisi siis sentimentin ennustamista, eikä sentimentin mittaamista. Sentimentin ennustaminen tehokkaasti vähintään vähentäisi tarvetta mitata sentimenttiä. Sentimentin ennustaminen voisi tarjota tehokkaamman, syvemmän ja laajemman näkemyksen markkinoiden käyttäytymiseen, mikä tekisi siitä, sentimentin mittaamista tehokkaamman analyysityökalun.

Lyhyen aikavälin ennustaminen ja aktiivinen kaupankäynti ovat teoriassa ainoat keinot, joilla voidaan saavuttaa ylisuuria voittoja markkinoilta, sosiaalisen median analysoinnin pohjalta. Pidemmällä aikajänteellä ennustamisen tehokkuus heikkenee merkittävästi. Passiivisen

sijoitusstrategian näkökannalta on täysin merkityksetöntä, mitä lyhyellä aikavälillä tapahtuu, sillä strategia olettaa, että heilahtelut normalisoituvat ajan kanssa ja yleiskehitys osakkeilla on kasvava.

Aktiivisella kaupankäynnillä taloudellisen lisäarvon luominen perustuisi siihen, että sosiaaliseen mediaan pohjautuva sentimenttianalyysi antaisi indikaation siitä, että osake ei ole hinnoiteltu oikein, minkä seurauksena taloudellista lisäarvoa voitaisiin luoda kaupankäynnillä. Osakkeen ollessa alihinnoiteltu sitä siis ostettaisiin lisää ja odotettaisiin hinnannousua ja vastaavasti osakkeen ollessa ylihinnoniteltu osaketta myytäisiin pois tai myytäisiin lyhyeksi. Kuitenkaan keinottelu osakkeilla ei ole niin yksinkertaista kuin edellä esitetty. Kaupankäyntiin liittyy aina kaupankäyntikulut, voitoista maksetaan oletusarvoisesti verot sekä analyysit eivät aina osu kohdalleen. Huomionarvoista on myös se, että osakkeiden hinnanmuutokset ovat pääsääntöisesti marginaalisia, päivittäisen vaihtelun ollessa yleensä korkeintaan muutaman prosentin luokkaa. Tämä siis tarkoittaa sitä, että tehdäkseen lyhyen aikavälin kaupankäynnillä merkittävästi rahaa, pienillä marginaaleilla, on kaupankäyntiin sitoutuvien pääomien oltava suuria. Tämä useimmiten rajaa aktiivisen kaupankäynnin ulkopuolelle pienet ja yksityissijoittajat.

Kun ajatellaan tutkimuskysymystä taloudellisen lisäarvon luomisesta, tulee pohtia myös muitakin keinoja, kuin arvopapereilla voitokkaasti kaupankäymistä. Kuten katsaus StockFluence-nimiseen yritykseen osoitti, osakeanalyysien myyminen kolmansille osapuolille on validi keino luoda taloudellista lisäarvoa. Sosiaalisen median sentimentin analysointi ei vaadi suuria investointeja tai resursseja yritykseltä, sillä ainakin mikroblogipalveluiden tuottama data on avoimesti saatavilla ja itse analysointityö perustuu pitkälti ihmisten asiantuntijuuteen ja tietokoneiden laskentatehoon. Liiketoiminnan riskitkin ovat maltillisemmat kuin siinä, että käytäisiin kauppaa arvopaperimarkkinoilla. Liiketoiminnan jatkuvuus kuitenkin edellyttäisi sitä, että ennustetarkkuus olisi ja pysyisi hyväksyttävällä tasolla.

Omat rajoitteensa sentimentin löytämiselle twiiteistä ja analyysin luomiselle, asettaa se, että twiitit ovat kirjoitettu eri kielillä. Arvopaperipörssit ovat kansainvälisiä, minkä seurauksena arvopaperimarkkinoille osallistuu eri kielisiä toimijoita. Eri kieliset twiitit hankaloittavat analyysin tekijöiden työtä, sillä useiden luonnollisten kielten analysointi ja seuraaminen vaatii

enemmän resursseja. Kirjoitettuun tekstiin pohjautuvan sentimenttianalyysin voisi siis olettaa olevan tehokkaampaa kielellisesti homogeenisimmissä arvopaperipörssiä.

Mikroblogipalvelut ja muut sosiaaliset mediat tuottavat kasvavalla tahdilla dataa. Valtavissa datavirroissa liikkuu paljon merkityksetöntä kohinaa, mutta oikein kohdennetuilla ja kehittyneillä analyysimenetelmillä tästä virrasta voidaan siivilöidä merkityksellistä tietoa lähes reaaliajassa. Sentimentin löytäminen ei yksinään riitä siihen, että voitaisiin sanoa, mihin suuntaan osakekurssit liikkuvat, vaan sentimentistä pitää pystyä erottamaan sen indikoima muutos analyysin avulla. Tutkimukset osoittavat, että mikroblogiviestien sentimentillä ja lähitulevaisuuden osakekurssien kehityksen välillä on korrelaatiota.

Ennustamiseen liittyy aina epävarmuutta, ilman sitä ennustaminen olisi tietämistä. Osakemarkkinoiden ennustamisessa, mikroblogipalveluiden viestien sentimentin avulla, on samoja piirteitä kuin esimerkiksi sään ennustamisessa. Ennustetarkkuus laskee merkittävästi, mitä pidemmän ajan päähän pyritään ennustamaan. Säätä voidaan verrata osakemarkkinoiden mikrotason tapahtumiin, ja yleistä taloustilannetta, makrotaloutta ja sen lainalaisuuksia, ilmastoon. Yleinen taloustilanne asettaa pitkälti tietyt raamit kurssikehitykselle, kuten ilmasto asettaa lämpötilojen ja sateiden vaihteluvälit sääennusteille. Tietokoneiden laskentatehon kasvaessa ja yhä älykkäämpien algoritmien kehittyessä ennustamisesta tulee tarkempaa, minkä seurauksena osakemarkkinoita voidaan kyetä ennustamaan tarkemmin. Tulevaisuudessa saatetaan kuitenkin keskittyä enemmän siihen, voidaanko ennustamista aikaistaa, siten että ennustettaisiin osakemarkkinoiden sentimentin muutoksia, reaali maailman tapahtumiin ja uutisiin perustuvilla sentimenttiennusteilla, sentimentin mittaamisen sijaan.

7 LÄHTEET

- Arndt, J. 1967, *Role of product-related conversations in the diffusion of a new product*, 4th edn, Journal of Marketing Research.
- Bollen, J. 2011, "Twitter mood predicts the stock market", *Journal of Computational Science*, vol. 2, no. 1, pp. 1-8.
- Calore, M. 2009, Feb 18-last update, *On Twitter, \$ Is the New #*. Available: <https://www.wired.com/2009/02/on-twitter-is-t/>.
- Chen, C. 2015, "Exploring Electronic Word-of-Mouth (eWOM) in The Consumer Purchase Decision-Making Process: The Case of Online Holidays - Evidence from United Kingdom (UK) Consumers", *Journal of Travel & Tourism Marketing*, vol. 32, no. 8, pp. 953.
- Coyne, S. 2017, "Forecasting Stock Prices Using Social Media Analysis", *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)IEEE*, Orlando, FL, USA, 6-10 Nov. 2017.
- Das, S., Behera, R.K., kumar, M. & Rath, S.K. 2018, *Real-Time Sentiment Analysis of Twitter Streaming data for Stock Prediction*.
- Garvey, R. 2005, "The Profitability of Active Stock Traders", *Journal of Applied Finance*, vol. 15, no. 2, pp. 93-100.
- Guo, L. 2016, "Big Social Data Analytics in Journalism and Mass Communication: Comparing Dictionary-Based Text Analysis and Unsupervised Topic Modeling", *Journalism & Mass Communication Quarterly*, vol. 93, no. 2, pp. 332-359.
- Hernández-Méndez, J. 2015, "The influence of e-word-of-mouth on travel decision-making: consumer profiles", *Current Issues in Tourism*, vol. 18, no. 11, pp. 1001-1021.
- Khan Nawsher, Ibrar Yaqoob, Hashem Ibrahim Abaker Targio, Inayat Zakira, Ali Waleed Kamaleldin Mahmoud, Alam Muhammad, Shiraz Muhammad & Gani Abdullah 2014, "Big data: survey, technologies, opportunities, and challenges", .
- Knüpfer, S. & Puttonen, V. 2018, *Moderni rahoitus*, 10th edn, Sanoma Pro Oy.
- Li, C. 2013, "The power of eWOM: A re-examination of online student evaluations of their professors", *Computers in Human Behavior*, vol. 29, no. 4.
- Live stats, I. 2019, , *Twitter Usage Statistics*. Available: <http://www.internetlivestats.com/twitter-statistics/>.

- Luukkonen, M. & Paasi, M. 2019, *#rahapodi 154 | Ylituottoa informaatioedulla*.
- MacArthur, A. 2018, Nov 02-last update, *The Real History of Twitter, In Brief*. Available: <https://www.lifewire.com/history-of-twitter-3288854>.
- Malkiel, B.G. 1999, *A Random walk Down Wall Street*, .
- Mamun, A., Syeed, A. & Yasmeen, F. 2015, "Are investors rational, irrational or normal?", , pp. 1-15.
- Nieland, B. 2017a, , *About Us*. Available: <https://www.stockfluence.com/about>.
- Nieland, B. 2017b, , *Pricing*. Available: <https://www.stockfluence.com/pricing>.
- Nikkinen, J., Rothovius, T. & Sahlström, P. 2008, *Arvopaperisijoittaminen*, WSOY.
- Oh, C. 2011, "Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement", .
- Oliveira, N., Cortez, P. & Areal, N. 2013, "On the predictability of stock market behavior using StockTwits sentiment and posting volume", , pp. 355.
- Patel, V. 2017, Jan 12,-last update, *How many Twitter users have a private profile? How many have a public profile? *. Available: <https://www.quora.com/How-many-Twitter-users-have-a-private-profile-How-many-have-a-public-profile>.
- Rainer, K. & Prince, B. 2017, *Introduction to Information Systems*, Wiley Etextbooks.
- Rajaraman, V. 2014, "Cloud computing", *Resonance*, vol. 19, no. 3, pp. 242-258.
- Ryan 2013, , *About Howard Lindzon*. Available: <https://zurb.com/soapbox/howard-lindzon-s-soapbox>.
- Sul, H.K., Dennis, A.R. & Yuan, L.(. 2017, "Trading on Twitter: Using Social Media Sentiment to Predict Stock Returns", *Decision Sciences*, vol. 48, no. 3, pp. 454-488.
- TwitterInc. 2019, , *What is Twitter's mission statement?*. Available: <https://investor.twitterinc.com/contact/faq/default.aspx>.
- Zhang, X. 2011, "Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear"", *Procedia - Social and Behavioral Sciences*, vol. 26, no. C, pp. 55-62.