



Maisteritutkielma
Datatieteen maisteriohjelma

Markkinaliikkeiden ennustaminen ohjatun oppimisen ja teknisen analyysin kautta

Mika Huttunen

13.5.2021

Ohjaaja(t): Professori Jussi Kangasharju

Tarkastaja(t): Professori Kangasharju
Apulaisprofessori Arto Klami

HELSINGIN YLIOPISTO
MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA

PL 68 (Pietari Kalmin katu 5)
00014 Helsingin yliopisto

| | | | |
|---|--|---|--|
| Tiedekunta — Fakultet — Faculty | | Koulutusohjelma — Utbildningsprogram — Degree programme | |
| Matemaattis-luonnontieteellinen tiedekunta | | Datatieteen maisteriohjelma | |
| Tekijä — Författare — Author | | | |
| Mika Huttunen | | | |
| Työn nimi — Arbetets titel — Title | | | |
| Markkinaaliikkeiden ennustaminen ohjatun oppimisen ja teknisen analyysin kautta | | | |
| Työn laji — Arbetets art — Level | | Aika — Datum — Month and year | Sivumäärä — Sidantal — Number of pages |
| Maisteritutkielma | | 13.5.2021 | 44 |
| Tiivistelmä — Referat — Abstract | | | |
| <p>Arvopaperin tulevan hinnanmuodostuksen ennustaminen on mielenkiintoista niin sijoittajan kuin aktiivisesti kauppaa tekevän markkinatoimijan näkökulmasta. Tarpeeksi hyvällä tarkkuudella arvopaperin tulevaa hinnanmuodostusta ennustamalla voi markkinatoimija ostaa arvopaperia ennen sen mahdollista markkinahinnan nousua, tai suojata salkkuaan sitä jo omistaessaan, mikäli on vaara, että arvopaperin markkinahinta laskee ajan mittaan merkittävästi.</p> <p>Tutkielmassani käsittelen koneoppimisen soveltamista tekniseen analyysiin. Tutkin, voidaanko tekniseen analyysiin pohjautuen markkinan tai arvopaperin tulevaa hinnanmuodostusta ennustaa lyhyellä aikavälillä tarpeeksi hyvällä tarkkuudella. Selvitän arvopaperimarkkinoiden toimintaa ja käyn läpi, miten tarkasteltavan markkinan tulevaa kysynnän ja tarjonnan suhdetta voidaan teknistä analyysiä hyödyntäen pyrkiä ennustamaan. Taustoitin myös omassa tutkimuksessa käyttämäni teknisen analyysin indikaattorien sekä koneoppimisen menetelmien toimintaa ja esitän aiempaa tutkimusta ongelman parissa.</p> <p>Havaitsin, että markkinoiden tulevan hinnanmuodostuksen ennustaminen on haastavaa. Käyttämälläni ohjatun oppimisen menetelmillä en onnistunut generoimaan mallia, joka olisi osannut ennustaa <i>S&P</i> 500-osakeindeksille, onko tarkasteltavaa ajanhetkeä seuraavan lyhyen aikavälin päätteeksi markkinahinta korkeammalla vai enintään yhtä korkealla kuin tarkasteluajankohtana. Opetetut mallit saavuttivat parhaimmillaan vain 50.8 – 51.4 % ennustetarkkuuden, kun taas naiivi luokittelija, joka ennustaa jokaisen aikavälin päätteeksi markkinahinnan kohonneen saavuttaa 53.0 %:n tarkkuuden. Vehnäfutuurisopimusmarkkinalle saamani tulokset olivat lupaavampia ja opetetut mallit saavuttivat edellä mainitun ongelmanratkaisuun parhaimmillaan 51.7 – 52.5 % ennustetarkkuuden, joka ylitti naiivin luokittelijan 50.9 % tarkkuuden. Analysoin saamiani tuloksia ja esitin jatkotutkimusmahdollisuuksia mallien tehostamiseksi.</p> <p>ACM Computing Classification System (CCS): Computing methodologies → Machine learning Social and professional topics → Professional topics → Computing and business → Automation</p> | | | |
| Avainsanat — Nyckelord — Keywords | | | |
| osakemarkkinat, tekninen analyysi, koneoppiminen | | | |
| Säilytyspaikka — Förvaringsställe — Where deposited | | | |
| Muita tietoja — Övriga uppgifter — Additional information | | | |

Sisältö

| | | |
|----------|---|-----------|
| 1 | Johdanto | 1 |
| 2 | Tekninen lähestyminen markkinoihin | 3 |
| 2.1 | Taustoitusta markkinoiden toimintaan | 3 |
| 2.2 | Indikaattoreita trendin suunnan havaitsemiseen | 8 |
| 2.2.1 | Liukuvat keskiarvot | 8 |
| 2.2.2 | Kynttilämuodostelmat | 11 |
| 2.2.3 | Momentum-indikaattorit | 13 |
| 2.3 | Yhteenveto | 18 |
| 3 | Koneoppimisen soveltaminen tekniseen markkina-analyysiin | 21 |
| 3.1 | KNN-luokittelija | 21 |
| 3.2 | Satunnainen metsä | 23 |
| 3.3 | Aiempia koneoppimisen sovellutuksia | 25 |
| 3.4 | Yhteenveto | 28 |
| 4 | Tutkimus | 29 |
| 4.1 | Toteutus | 29 |
| 4.2 | Tulokset | 32 |
| 4.3 | Jatkotutkimusmahdollisuuksia | 37 |
| 4.4 | Yhteenveto | 39 |
| 5 | Johtopäätökset | 41 |
| | Kirjallisuutta | 43 |

1. Johdanto

Kysynnän ja tarjonnan suhde määrää pääosin markkinoilla kaupattavien tuotteiden hinnat. Kun kysyntä on suurta, on myyjillä mahdollisuus pyytää kauppaamistaan tuotteista korkeaa hintaa, kun taas kysynnän puuttuessa joutuvat myyjät taipumaan alhaisempaan hintaan löytääkseen tuotteilleen ostajat. Sama kysynnän ja tarjonnan laki pätee arvopaperimarkkinoilla. Kysyntä ja tarjonta kohtaavat pääsääntöisesti arvopaperin oikean markkinahinnan läheisyydessä, mutta hinnanvaihtelua sen ympärillä tapahtuu jatkuvasti. Tällaisesta lyhyelläkin aikavälillä tapahtuvasta hinnanvaihtelusta voisi aktiivinen kaupankävijä hyötyä ostamalla arvopapereita silloin, kun niille on vain vähän kysyntää suhteessa tarjontaan ja myymään, kun kysynnän määrä on kasvanut merkittävästi ja arvopapereiden markkinahinnat kohonneet. Tällainen kaupankäyntistrategia vaatisi tulevan kysynnän ja tarjonnan määrän ennustamista.

Tehokkaiden markkinoiden hypoteesin mukaan kysynnän ja tarjonnan suhde muuttuu vain uuden informaation saapuessa ja kysyntä ja tarjonta kohtaavat aina oikealla markkinahintatasolla. Hypoteesia on haastettu useilla tutkimuksilla [23] ja empiirisesti on havaittu, että kysyntä ja tarjonta muuttuvat myös erilaisten teknisten ajurien seurauksena. Näiden ajurien tunnistamista ja tulevan kysynnän ja tarjonnan suhteen ennustamista historiallisen markkinadatan perusteella kutsutaan *tekniseksi analyysiksi* tai *tekniseksi markkina-analyysiksi*.

Tutkielmassani käsittelen teknisen analyysin näkökulmasta koneoppimisen soveltamista markkinan tulevan hinnanmuodostuksen ennustamiseen. Selvitin, voidaanko lyhyellä aikavälillä tarkasteltavalla markkinalla ennustaa, onko markkinahinta aikavälin päätteeksi suurempi, vai enintään yhtä suuri kuin sen alussa. Aikavälin pituudeksi valitsin yhden pörssipäivän. Tällöin aikaväli on tarpeeksi lyhyt, jotta hinnanmuodostusta ei voi selittää fundamenttitekijöiden muutoksilla, kuten tavallisten sijoittajien tietoon tulevalla uudella informaatiolla.

Tutkielmani rakenne on seuraavanlainen. Luvussa 2 tarkastelen arvopaperimarkkinoiden toimintaa ja sitä, miten hinnanmuodostus markkinoilla tapahtuu. Esittelen myös teknisen analyysin indikaattoreita, joita hyödyntäen voidaan pyrkiä ennustamaan markkinan tulevaa hinnanmuodostusta erimittaisilla aikaväleillä.

Luvussa 3 käsittelen koneoppimisen menetelmiä, joita olen sekä itse soveltanut,

että on aiemmassa tutkimuksessa sovellettu markkinan tulevan hinnanmuodostuksen ennustamiseen teknistä analyysiä hyödyntäen. Tarkastelen myös aiempia teknisen analyysin ja koneoppimisen sovellutuksia.

Luvussa 4 käsittelen omaa tutkimustani, jonka toteutin kahta koneoppimisen menetelmää ja markkinadatajoukkoa hyödyntäen. Havaitsin, että markkinan tulevan hinnanmuodostuksen ennustaminen on lyhyellä aikavälillä haastavaa. Tutkimani vehnäfutuurisopimusdatan osalta tulokset olivat silti lupaavia ja indikoivat siitä, että lyhyillä aikaväleillä voisi tulevaa markkinahinnan muodostusta olla mahdollista ennustaa tarpeeksi hyvällä tarkkuudella, että ennusteita voisi hyödyntää esimerkiksi aktiiviseen kaupankäyntiin ja tuoton tavoitteluun. Oletusta tukee myös luvussa 3 esitetetyin aieman tutkimuksen tulokset. Käyn läpi myös jatkotutkimusmahdollisuuksia tutkimukseni parissa, ja miten ennustetarkkuuksia voisi pyrkiä parantamaan. Luku 5 kattaa tutkielmani yhteenvedon ja johtopäätökset.

2. Tekninen lähestyminen markkinoihin

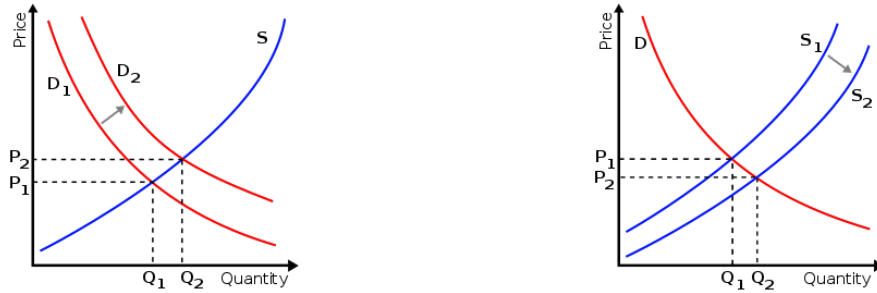
Tässä luvussa taustoitan arvopaperimarkkinoiden toimintaa ja esittelen erilaisia indikaattoreita, joiden avulla tulevia markkinahinnan muutoksia voi pyrkiä ennustamaan. Alaluvussa 2.1 käsittelen kysynnän ja tarjonnan vaikutusta kulloinkin voimassaolevaan markkinahintaan. Esitän myös teknisen analyysin oletuksia ja käyn esimerkinomaisesti läpi, miten muutamia teknisen analyysin indikaattoreita hyödyntäen voidaan arvioida tulevaa kysynnän ja tarjonnan suhdetta eräälle pörssiosakkeelle.

Alaluvussa 2.2 tarkastelen erilaisia teknisen analyysin indikaattoreita ja syitä siihen, miksi ne toimivat. Olen hyödyntänyt niitä omassa tutkimuksessani ja osaa on käytetty myös aiemmissa teknisen analyysin ja koneoppimisen sovellutuksissa. Tutkimustani ja aiempia sovellutuksia käsittelen luvuissa 3 ja 4.

2.1 Taustoitusta markkinoiden toimintaan

Kaupankäynti arvopapereilla tapahtuu erilaisten välittäjien kautta. Halutessaan ostaa tiettyä arvopaperia, ostaja määrittää haluamansa kappalemäärän, ostohinnan sekä toimeksiannon voimassaoloajan. Kauppa syntyy, jos voimassaoloajan puitteissa löytyy myyjä, joka on halukas myymään arvopaperia ostajan tarjoamaan hintaan. Ostajalla on myös mahdollista ostaa arvopaperia suoraan alhaisimmalla pyydetyllä myyntihinnalla, jolloin kauppa tapahtuu välittömästi. Vastaavasti myyjänä on mahdollista jättää myyntitoimeksianto markkinalle haluamallaan kappalemäärällä, myyntihinnalla ja voimassaoloajalla, tai toteuttaa myynti välittömästi korkeimmalla tarjotulla ostohinnalla.

Jos tarkastellaan jonkin yhtiön osaketta eli arvopaperia, joka vastaa osuutta tästä yhtiöstä, osakkeen hinta määräytyy useiden kuukausien ja vuosien ajanjaksolla enimmäkseen yhtiön liiketoiminnan perusteella. Mikäli yhtiö pystyy kasvattamaan liikevaihtoaan ja tulostaan vuosittain, näkyy tämä väistämättä osakekurssin nousuna. Osakekurssin nousu itsessään ei kuitenkaan johdu fundamentaalisista muutoksista koskien yhtiön liiketoimintaa, vaan kysynnän ja tarjonnan suhteiden muuttumisesta. Jos osakkeelle on nimittäin enemmän kysyntää kuin tarjontaa, niin myyjillä on mahdollista



(a) Kysynnän kasvu johtaa kauppahintojen nousuun

(b) Tarjonnan kasvu johtaa kauppahintojen laskuun

Kuva 2.1: Kysynnän ja tarjonnan muutosten vaikutus kauppahintoihin [1]

pyytää osakkeesta korkeampaa hintaa. Vastaavasti tilanteessa, jossa tarjonta ylittää kysynnän määrän, on ostajilla mahdollista tarjota entistä vähemmän ja silti markkinalta löytyy todennäköisesti halukkaita myyjiä tekemään kauppvoja ostajien tarjoamaan hintaan. Luonnollisesti jatkuvasti tuottavampi liiketoiminta poikii kuitenkin enemmän kysyntää kuin tarjontaa ja osakekurssi pitkällä aikavälillä nousee.

Kuva 2.1 havainnollistaa kysynnän ja tarjonnan suhdetta suhteessa kauppahintoihin. Kuvassa 2.1a S kuvastaa tarjonnan määrää, D_1 alkuperäistä kysynnän määrää ja D_2 kasvanutta kysynnän määrää. Käyrien D_1 ja S leikkauspiste on tasapainotilanne, jonka hintatasolla kaupat tapahtuvat alussa keskimäärin. Keskimääräistä kauppahintaa on kuvattu P_1 :llä. Kysynnän määrän kasvaessa D_1 :stä D_2 :een, siirtyy tasapainotilanne tarjontakäyrällä S ylemmäs ja keskimääräinen kauppahinta nousee P_1 :stä P_2 :een. Vastaavasti kuvassa 2.1b tarjonnan määrän kasvaessa S_1 :stä S_2 :een, siirtyy tasapainotilanne kysyntä kuvaavalla käyrällä D alemmas ja keskimääräiset kauppahinnat laskevat.

Pitkällä aikavälillä kysyntään ja tarjontaan vaikuttavat siis osakkeen tapauksessa yhtiön liiketoiminnan fundamenteissa tapahtuvat muutokset. Tuottavuuden kasvu johtaa pitkällä aikavälillä osakekurssin nousuun ja heikkenevä liiketoiminta vastaavasti osakekurssin laskuun. Kaupankäyntihinnat ja siten kysynnän ja tarjonnan määrä vaihtelevat kuitenkin paljon myös lyhyellä aikavälillä. Lyhyellä aikavälillä tarkoitetaan aikaa, jossa esimerkiksi yksittäisen yhtiön liiketoiminnassa eikä maailmantalouden tilanteessa tapahdu muutoksia, jotka selittäisivät kyseisen yhtiön osakkeen kaupankäyntihintojen vaihtelua.

Seuraavaksi tarkastelen esimerkkinä Sammon osakkeen hinnanvaihtelua vuonna 2020 toukokuun 14. päivästä alkaen syyskuun 1. päivään saakka. Osakkeen hinnanvaihtelua kuvaava kurssigraafi on esitetty kuvassa 2.2 ja graafin datapisteet ovat yksittäisten pörssipäivien päätöskursseja.

Sampo julkaisi 5.8.2020 vuoden toisen kvartaalin osavuositarkastuksena, jonka tulos oli positiivinen, mutta kurssireaktio oli kyseisen päivän aikana hyvinkin negatiivi-



Kuva 2.2: Sammon osakkeen kurssikehitys kesällä 2020 [2]

nen ja osakkeen hinta laski edeltävän pörssipäivän päätöskurssistaan noin 1,30 euroa. Elokuun 10. päivä alkaen osakkeen hinta lähti taas voimakkaaseen nousuun. Hinnan nousu ja sitä edeltänyt raju lasku havaitaan graafilta ”W-muotoisen” kuvion oikeana puolena. Fundamenttien perusteella sijoittaja voisi perustella tällaiset kurssireaktiot sillä, että tulos otettiin vastaan negatiivisesti, mutta myöhemmin tajuttiin, että osake on varsin edullisesti hinnoiteltu, mikä kasvatti kysyntää ja osakkeen hinta alkoi nousta.

Tarkastellaan seuraavaksi osakkeen *kynttilägraafia* samalta aikaväliltä, joka on esitetty kuvassa 2.3. Toisin kuin ylempältä kurssigraafilta, kynttilägraafilta voidaan havaita päivittäisten päätöskurssien lisäksi päivittäiset avauskurssit sekä matalimmat että korkeimmat noteeraukset. Graafin vihreät kynttilät kuvastavat pörssipäiviä, joiden päätöskurssit ovat korkeampia kuin niiden avauskurssit. Avauskurssit löytyvät kynttilöiden paksujen osien eli *kehojen* alareunoista ja päätöskurssit yläreunoista. Piikit kynttilöiden ylä- ja alapuolilla kertovat siitä, millä hinnoilla päivän aikana tehtiin korkeimmillaan ja matalimmillaan kauppaa. Yleensä vihreä kynttilä indikoi myös markkinahinnan kohoamisesta edeltävän pörssipäivän päätöskurssiinsa nähden, mutta näin ei välttämättä tapahdu, jos päivän avauskurssi on matalampi kuin edeltävän pörssipäivän päätöskurssi. Tällainen vihreä kynttilä havaitaan graafilta mm. kesäkuun 15. päivänä.

Punaiset kynttilät kuvaavat pörssipäiviä, joiden päätöskurssit ovat matalampia kuin niiden avauskurssit. Avauskurssit löytyvät kynttilöiden kehojen yläreunoista ja päätöskurssit alareunoista. Piikit kynttilöiden ylä- ja alapuolilla tarkoittavat samaa kuin vihreiden kynttilöidenkin tapauksessa. Graafille on piirretty myös 20. (punaisella) ja 50. (sinisellä) päivän *liukuvat keskiarvot*. Nämä kuvastavat jokaiselle pörssipäivälle keskimääräisen päätöskurssin nykyisen ja sitä edeltävien yhteensä kahdenkymmenen



Kuva 2.3: Sammon osakkeen kurssikehitys kesällä 2020 kynttilägraafina kuvattuna [2]. Graafilla on esitetty myös 20. (punaisella) ja 50. (sinisellä) päivän liukuvat keskiarvot.

ja viidenkymmenen pörssipäivän ajalta. Esimerkiksi pörssipäivälle 1.9.2020, jota kuvaa graafin viimeinen kynttilä, 20. päivän liukuvan keskiarvon arvo on 32.62 ja se saadaan laskemalla 1.9.2020 ja yhdeksäntoista edeltävän pörssipäivän päätöskurssit yhteen ja jakamalla saatu summa luvulla 20.

Syy miksi liukuvat keskiarvot ovat mielenkiintoisia on se, että verratessa nykyhetken osakekurssia tiettyjen paljon seurattujen liukuvien keskiarvojen kuten 20. päivän, 50. päivän ja 200. päivän arvoihin, voidaan päätellä, onko kurssi tarkastellulla aikavälillä nousevassa vai laskevassa *trendissä* [18]. Liukuvat keskiarvot myös *tukevat* tai *vastustavat* kurssinousua. Esimerkiksi 16. ja 17. kesäkuuta 20. päivän liukuva keskiarvo vastusti kurssinousua markkinahinnalla 31.55, ja koska kysyntää ei ollut silloin riittävästi nostamaan markkinahintaa ylöspäin, hinta lähti lyhyellä aikavälillä laskuun. Hinta laski 50. päivän liukuvan keskiarvon läheisyyteen noin 29.20:een saakka, kunnes liukuva keskiarvo tuki markkinahinnan säilymistä sen yläpuolella. Käytännössä voidaan olettaa, että markkinahinnan siirtyminen merkittävästi seurattun liukuvan keskiarvon läheisyyteen luo arbitraarista kysyntää tai tarjontaa liukuvia keskiarvoja seuraavien tietokoneidenkin toimesta riippuen siitä, toimiiko liukuva keskiarvo tuki- vai vastustasona.

Heinäkuun alussa osakkeen markkinahinta kapusi jälleen 20. päivän liukuvan keskiarvon yläpuolelle luoden lisää kysyntää ja hinta nousi 31.10:stä noin 33.50:een saakka. Kesäkuun korkeimmat noteeraukset tapahtuivat hintatasolla 33.50–33.80, mikä on merkattu graafille vaakasuorilla mustalla ja violetilla janoilla. Hintataso toimi heinäkuussa vastustasona ja käynnisti omalta osaltaan suurehkon *korjausliikkeen* alaspäin.

Markkinahinta laskee aina elokuun 5. ja 7. päivään asti, kunnes myöhemmin elokuussa markkinahinta ylittää edellä mainitun hintatason.

Kuten yllä olevasta esimerkistä havaitaan, on olemassa teknisiä ajureita, jotka vaikuttavat markkinoilla kulloinkin voimassaolevan kysynnän ja tarjonnan määrään. Toukokuun 14. päivästä alkaen syyskuun 1. päivään saakka Helsingin pörssin yleisindeksin OMXHPI:n arvo kasvoi noin 22 %, mikä selittää osittain Sammon osakekursin nousevaa trendiä. Lyhyen aikavälin hinnanvaihtelua kesäkuun huipusta elokuun alun pohjalukemiin ei voi kuitenkaan yleisindeksin arvonmuutoksilla selittää. Teknisen markkina-analyysin keinoin hinnanvaihtelua voidaan tosin perustella ja sitä olisi saattanut pystyä reaaliajassa ennustamaankin.

J. Larsen käsitteli tutkielmassaan [18] teknisen analyysin oletuksia. Teknisessä analyysissä oletetaan, että kaikki fundamentaalinen informaatio on aina hinnoiteltu kurssiin. Tällä tarkoitetaan sitä, että esimerkiksi yksittäisen yrityksen osakkeen tapauksessa yrityksen viimeisten tilinpäätöstietojen ohella odotukset tulevaisuuden kasvavirroista sekä maailman yleisen taloustilanteen kehittymisestä ovat jo osana osakkeen hintaa. Oletetaan myös, että hinta liikkuu trendeissä, ja siksi teknisen analyysin indikaattoreita kuten em. liukuvia keskiarvoja pyritään hyödyntämään meneillä olevan trendin tunnistamiseen ja potentiaalisen trendin suunnan muutoksen identifiointiin. Teknisessä analyysissä tavoitteena on tunnistaa trendi tai suunta, johon markkinahinta pääsääntöisesti liikkuu. Pääsääntöisen trendin aikana havaitaan kuvioita, jotka voivat indikoida joko trendin jatkumisesta tai potentiaalisesta suunnan kääntymisestä. Saataessa signaali potentiaalisesta suunnan kääntymisestä, voidaan myydä tuotteita, joiden arvo alenisi suunnan kääntyessä, tai ostaa sellaisia, jotka hyötyisivät vastakkaisuuntaisesta trendistä.

Kolmas tärkeä teknisen analyysin oletus liittyy markkinoiden taipumukseen toistaa historiallista käyttäytymistä. Esimerkiksi kuvio, joka on nousevassa trendissä johtanut historiallisesti suurella todennäköisyydellä trendin kääntymiseen aiheuttaa riskin, että sellaisen havaitseminen nousevassa trendissä myöhemmällä ajanhetkellä johtaa myös trendin kääntymiseen. Sen sijaan esimerkiksi kuvan 2.3 graafilla havaittava ”W-muotoinen” kuvio pidemmän aikavälin nousevassa trendissä johtaa merkittävällä todennäköisyydellä arvonnousuun pian kuvion toteutumisen jälkeen [3].

Markkinan tulevaa hinnanmuodostusta teknisen analyysin kautta voidaan tarkastella myös bayes-lähtöisesti olettamalla, että on olemassa tuntematon jatkuvasti muuttuva jakauma, jonka perusteella tarkasteltavalla ajanjaksolla kynttilät generoidaan pääsääntöisesti historiallisen hintatansa suhteen. Indikaattoreita käyttäen pyritään approksimoimaan tuntematonta jakaumaa tarkasteltavalla ajanhetkellä ja analysoimaan, millä todennäköisyyksillä markkinahinta muuttuu seuraaviksi ja millä tavoin.

2.2 Indikaattoreita trendin suunnan havaitsemiseen

Haluttaessa ennustaa, kohtaavatko kysyntä ja tarjonta jatkossa korkeammalla vai matalammalla markkinahinnalla kuin tietyllä ajanhetkellä, riittää tunnistaa pääsääntöinen markkinatrendi ja odottaa hinnan siirtyvän trendin osoittamaan suuntaan. Tässä alaluvussa käsittelemme indikaattoreita, joilla pääsääntöisen markkinatrendin suunta voidaan tunnistaa, ja miten trendin suunnan vaihtuminen voidaan havaita.

2.2.1 Liukuvat keskiarvot

Alaluvussa 2.1 käsittelemme Sammon osakkeen kurssikehitystä kesällä 2020 ja sen ohessa esittelimme 20. ja 50. päivän liukuvat keskiarvot. Liukuviin keskiarvoihin viitataan jatkossa myös englanninkielisellä lyhenteellä SMA ("Simple moving average") ja n . päivän liukuvaan keskiarvoon lyhenteellä SMA_n . Matemaattisesti i :n päivän päätöskurssille x_i , SMA_n :n arvo määritellään lausekkeella

$$SMA_{x_i,n} = \frac{1}{n} \cdot \sum_{j=0}^{n-1} x_{i-j}.$$

Voidaan havaita, että $i + 1$:n päivän päätöskurssille $SMA_{x_{i+1},n}$ määritetään muutoin samoin kuin $SMA_{x_i,n}$, paitsi että summalausekkeessa yhteenlaskettavista termeistä termi x_{i-n+1} korvautuu termillä x_{i+1} . Tästä seuraa, että

$$x_{i+1} > x_{i-n+1} \iff SMA_{x_{i+1},n} > SMA_{x_i,n}.$$

Jos siis valitulle pörssipäivälle i , SMA_n on sitä seuraavina pörssipäivinä kasvava, ovat niiden päätöskurssit vähintään yhtä suuria kuin niitä edeltäneiden n pörssipäivien päätöskurssit ja päinvastoin, jos SMA_n on i :tä seuraavina pörssipäivinä laskeva.

Koska teknisessä analyysissä oletetaan markkinahinnan liikkuvan trendeissä, voidaan halutun mittaisen ajanjakson pääsääntöinen markkinatrendi havaita tarkasteltavan liikkuvan keskiarvon suunnasta ja markkinahinnan suhteesta sen arvoon, eli onko hinta liikkuvan keskiarvon ylä- vai alapuolella. Esimerkiksi kuvassa 2.3 markkinahinta pysyy lähes aikavälin alusta alkaen SMA_{50} :n (sinisellä) yläpuolella ja sen arvo nousee päivittäin, kunnes pörssipäivän 30.7.2020 päätöskurssi on ensimmäistä kertaa sen alapuolella. Tuohon 30.7.2020 saakka SMA_{50} indikoi markkinahinnan nousevasta trendistä. Hinnan ollessa SMA_{50} :n alapuolella markkina on pääsääntöisesti laskevassa trendissä 50. pörssipäivän mittaisella ajanjaksolla tarkasteltuna varsinkin, jos liikkuvan keskiarvon suunta kääntyy selvästi laskevaksi. Kuvan graafilla tätä ei tapahdu ja

elokuun 11. ja 12. päivien aikana hinta ylittää jälleen $SMA_{50:n}$, mikä indikoi trendin kääntymisestä nousevaksi.

Yksinkertainen yhteen liukuvaan keskiarvoon perustuva tulevien markkinaliikkeiden ennustustapa voisi olla sellainen, että kun markkinahinta on liukuvan keskiarvon yläpuolella ja liukuvan keskiarvon suunta on ylöspäin, odotettaisiin myös jatkossa nousevia markkinahintoja ja päinvastoin markkinahinnan ollessa sen alapuolella. Ennustustapaa käytettäessä aikaväli, jonka aikana markkinaliikkeitä pyrittäisiin ennustamaan pitäisi suhteuttaa liukuvan keskiarvon pituuteen. Esimerkiksi 20. päivän liukuvaa keskiarvoa hyödynnettäessä voisi olla mielekästä yrittää ennustaa, onko markkinahinta viiden pörssipäivän päästä nykyhintaansa korkeammalla ja 50. päivän liukuvan keskiarvon osalta samaa kymmenelle seuraavalle pörssipäivälle.

Ongelma pitkien ajanjaksojen liukuvien keskiarvojen käytössä on se, että ne indikoivat trendin kääntymisestä erittäin myöhään tai eivät ollenkaan. Edellä mainitun kaltaisella ennustustavalla ja kuvan 2.3 pörssidatalla ei 50. päivän liukuvaa keskiarvoa hyödyntäen olisi saatu lainkaan myyntisignaalia, kun Sammon osake oli 33.50:n hinnassa kesäkuun alussa. Osaketta olisi sen sijaan pidetty kesäkuun alusta kesäkuun puoleen väliin, missä välissä se olisi ehtinyt laskea noin hintaan 29.20. Lisäksi kun hinta lähti jälleen nousuun heinäkuun 21. päivän huippuun saakka, ei myöskään tuolloin olisi saatu myyntisignaalia, vaan vasta heinäkuun 30. päivänä, jolloin myynti olisi voinut tehdä suunnilleen hinnalla 31.70. Lyhyemmän ajanjakson liukuva keskiarvo antaa taas enemmän osto- ja myyntisignaaleja keskiarvon ylityksistä ja alituksista. Mitä lyhyempää ajanjaksoa käytetään, sitä todennäköisemmin signaalit ovat toisaalta myös virheellisiä.

Eksponentiaalista liukuvaa keskiarvoa, mihin viitataan lyhenteellä EMA ("Exponential moving average"), voidaan SMA :n ohella käyttää pääsääntöisen markkinatrendin tunnistamiseen. Eksponentiaalinen liukuva keskiarvo painottaa tuoreimpia päätöskursseja vanhimpiin nähden ja i :n päivän päätöskurssille x_i , n . päivän eksponentiaalisen liukuvan keskiarvon arvo määritellään rekursiivisesti lausekkeella

$$EMA_{x_i,n} = \alpha \cdot x_{i-1} + (1 - \alpha) \cdot EMA_{x_{i-1},n}, \quad (2.1)$$

jossa $\alpha = \frac{\beta}{n+1}$ ja β *tasoituskertoimen*. Usein valitaan $\beta = 2$ ja tällöin esimerkiksi 20. päivän EMA :lle edeltävän pörssipäivän päätöskurssin paino on kaavan 2.1 nojalla

$$\alpha = \frac{2}{20 + 1} \approx 0.095$$

eli noin 9.5 %. Vertailun vuoksi SMA_{20} :lle kunkin pörssipäivän päätöskurssin paino on $\frac{1}{20} = 5.0$ %. Tasoituskertoimen ohella tulee kaavan 2.1 soveltamiseksi määrittää $EMA_{x_1,n}$ ja yleensä valitaan $EMA_{x_1,n} = x_1$.



Kuva 2.4: Trendin tunnistaminen kahden liukuvan keskiarvon leikkauspisteiden perusteella [5]

Koska eksponentiaalinen liukuva keskiarvo painottaa tuoreimpia päätöskursseja, on se herkempi hinnanmuutoksille kuin *SMA* ja tunnistaa potentiaalisen trendin muutoksen nopeammin. Mutta samoin kun lyhyen ajanjakson *SMA*, antaa *EMA* herkästi virheellisiä osto- ja myyntisignaaleja lyhytaikaisen nopean hinnanvaihtelun seurauksena. Eksponentiaalinen liukuva keskiarvo toimii hyvin samankaltaisesti kuin lyhyemmälle ajanjaksolle sovellettava liukuva keskiarvo [4, 18].

Käytettäessä useampaa kuin yhtä (eksponentiaalista) liukuvaa keskiarvoa, voidaan pääsääntöinen markkinatrendi pyrkiä tunnistamaan liukuvien keskiarvojen leikkauspisteiden perusteella. Kuvassa 2.4 on havainnollistettu tällainen tapa tunnistaa pääsääntöinen markkinatrendi *S&P* 500-indeksifutuurille syyskuussa 2020. Kuvan graafissa jokainen kynttilä on kolmenkymmenen minuutin mittainen ja graafille on piirretty 50. (sinisellä) ja 200. (vihreällä) kynttilän liukuvat keskiarvot, jotka kuvaavat 50. ja 200. viimeisen kolmenkymmenen minuutin mittaisen ajanjakson päätöskurssien keskihintaa.

Pääteltäessä markkinatrendiä liukuvien keskiarvojen leikkauspisteiden perusteella oletetaan, että markkina on nousevassa trendissä, kun lyhyemmän ajanjakson liukuva keskiarvo on pidemmän ajanjakson liukuvan keskiarvon yläpuolella. Vastaavasti kun lyhyemmän ajanjakson liukuva keskiarvo on pidemmän ajanjakson liukuvan keskiarvon alapuolella, päätellään markkinan olevan laskevassa trendissä [18]. Kuvassa 2.4 on esitetty tilanteet, jolloin tällaista trendin tunnistustapaa hyödyntäen saadaan signaali ostaa indeksifutuurin arvon noususta hyötyvää tuotetta, ja milloin tuote kannattaa myydä pois. Sinisistä suorakaiteista taas nähdään, kuinka suuri määrä tappioita strategialla vältettäisiin siihen nähden, että graafin osoittaman laskun aikana oltaisiin

jatkuvasti markkinalla mukana. Kyseisten suorakaiteiden sisällä voitaisiin myös *myydä lyhyeksi* indeksifutuurin noususta hyötyvää tuotetta, jolloin olisi mahdollista tehdä tuottoa laskevassa trendissä. Lyhyeksi myynnissä tarkoitus on lainata tuote kolmannelta osapuolelta, myydä se pois tarkoituksena ostaa myöhemmin takaisin halvemmalla ja palauttaa kolmannelle.

Kuvan 2.4 graafista havaitaan, että käytettäessä kahden liukuvan keskiarvon leikkauspisteitä osto- ja myyntipäätösten tekemiseen, vältetään useilta virhesignaaleilta, joita saataisiin yksittäistä liukuvaa keskiarvoa käytettäessä. Hinta nimittäin saavuttaa varsinkin 50. kynttilän liukuvan keskiarvon useisiin otteisiin siten, että se tekee vain lyhyitä liikkeitä liukuvan keskiarvon ylä- ja alapuolella. Havaitaan myös, että graafin liukuvat keskiarvot reagoivat trendin muutoksiin melko myöhään ja sekä ensimmäisen että toisen myyntisignaalin kohdalla markkina on ehtinyt jo laskea reilusti optimaalisesta edeltävästä myyntihetkestään. Vastaavasti myös ostopäätökset tehtäisiin strategialla huomattavasti optimaalisia osto hetkiä myöhemmin ja korkeampaan hintaan. Tällaisella strategialla näin nopeiden laskujen aikana eksponentiaaliset liukuvat keskiarvot saattaisivatkin toimia liukuvia keskiarvoja paremmin.

2.2.2 Kynttilämuodostelmat

Kun liukuvat keskiarvot signaloivat trendin kääntymisestä jälkikäteen, antavat tietynlaiset Japanissa keksityt *kynttilämuodostelmat* taas välittömän signaalin potentiaalisesta markkinatrendin muutoksesta. Toiset muodostelmat taas vahvistavat signaalia markkinatrendin jatkumisesta tiettyyn suuntaan [21].

Markkinatrendin kääntymistä tai jatkumista signaloivat kynttilämuodostelmat koostuvat pääasiassa 1–5 kynttilästä muodostelmasta riippuen. Kynttilöiden nimentäperiaate pohjautuu 1500- ja 1600-luvun Japaniin, jolloin maassa oli runsaasti lääninaapurien välisiä sotia. Kynttilöiden nimennässä korostuukin taistelukenttäterminologia, mutta oman kokemuksen mukaan myös niiden visuaalinen ilme. Seuraavaksi esittelen pienen joukon yksinkertaisia muodostelmia, jotka johtavat merkittävällä todennäköisyydellä markkinatrendin kääntymiseen. Kynttilöiden englanninkieliset yleiskäyttöiset vastineet on esitelty kunkin kynttilän osalta suluissa suomenkielisen nimen ohessa.

Vasara ("Hammer") on kynttilä, joka antaa potentiaalisen signaalin laskutrendin päättymisestä ja trendin kääntymisestä nousevaksi. Kynttilälle ominaista on piikki kynttilän alapuolella ja sen tulee esiintyä selvästi laskevassa trendissä. Kynttilän kehon tulee olla joko vihreä, tai punainen, jolloin sen olisi oltava pieni suhteessa alapuolella olevan piikin pituuteen. Kynttilän alapuolella oleva piikki kertoo siitä, että kynttilän muodostumisen aikana markkinahinta kävi liiankin alhaalla, kysynnän suhde tarjontaan nähden kasvoi suureksi ja hinta lähti nousuun antaen signaalin kysynnän



Kuva 2.5: Trendin kääntymisestä signaloivat vasara- ja pyrstötähti-kynttilät

potentiaalisesta kasvamisesta tarjontaan nähden jatkossakin. Vahvistus trendin kääntymisestä saadaan, jos seuraavien kynttilöiden aikana päätöshinta muodostuu selvästi vasara-kynttilän päätöshinnan yläpuolelle. Vasara on esitetty kuvassa 2.5a toisena kynttilänä oikealta katsottuna.

Pyrstötähti ("Shooting star") on vasara-kynttilän vastakohta, joka signaloi potentiaalisesta laskutrendiin siirtymisestä, kun kynttilä havaitaan nousevassa markkinatrendissä. Pyrstötähdelle ominaista on piikki kynttilän yläpuolella ja punainen keho, tai vihreä keho pitkällä yläpuolella olevalla piikillä. Kuvassa 2.5b on nähtävillä kaksi pyrstötähteä - ensimmäinen sekä kolmas punainen kynttilä vasemmalta katsottuna. Kuvan viimeinen kynttilä on myös muodoltaan pyrstötähti, mutta se esiintyy laskevassa trendissä ja jätetään huomioimatta.

Kuvan ensimmäinen pyrstötähti antaa samalla lailla signaalin potentiaalisesta trendin kääntymisestä kuin toinenkin. Kuvan markkinatilanteessa trendi jatkuu kuitenkin ylöspäin ja trendin kääntymisen vaatima vahvistus seuraavien kynttilöiden päätöshintojen jäämisestä pyrstötähti-kynttilän päätöshinnan alapuolella jää saamatta.

Yksi kynttilä, tai kynttilämuodostelma itsessään ei anna yleensä vahvaa signaalia trendin kääntymisestä. Usein trendin kääntymistä ennen havaitaan kuitenkin jo muita tekijöitä, jotka niin ikään signaloivat siitä, että trendi on potentiaalisesti pian kääntymässä. Kuvassa 2.5a toinen kynttilä on *käännetyn vasaran* kaltainen ("Inverted hammer"), joka antaa myös potentiaalisen signaalin laskevan trendin päättymisestä. Sille ominainen piikki kynttilän yläpuolella kertoo siitä, että kysyntä ylitti tarjonnan kynttilän muodostumisen aikana, mutta korkeammalla hinnalla tarjontaa oli kuitenkin liikaa kysyntään nähden ja hinta palautui uudestaan alemmas. Tämän esiintyminen juuri ennen vasara-kynttilää tuo lisävahvistusta potentiaalisesta trendin kääntymisestä.

Kuvassa 2.5b taas toisena punaisena kynttilänä vasemmalta havaitaan vasaran mallinen kynttilä ("Hanging man"), jonka esiintyminen nousevassa markkinatrendissä indikoi trendin heikkoudesta ja potentiaalisesta kääntymisestä kynttilän alla olevan piikin vuoksi. Pyrstötähden muodostuessa tämän jälkeen vahvistaa tämä kynttilä pyrstötähden antamaa signaalia potentiaaliselle trendin muutokselle.



(a) Härkämäinen peittävä kynttilä



(b) Karhumainen peittävä kynttilä

Kuva 2.6: Trendin kääntymisestä signaloivat edeltävän kynttilän täysin peittävät kynttilät

Kuva 2.6 esittää lisää trendin potentiaalista kääntymistä signaloivia kynttilämuodostelmia. Kuvassa 2.6a havaitaan kolmantena kynttilänä oikealta katsottuna *härkämäinen peittävä kynttilä* ("Bullish engulfing"). Tälle kynttilälle on ominaista on vihreä keho, joka peittää täysin sitä edeltävän punaisen kynttilän. Tämä tarkoittaa sitä, että sen avauskurssin tulee olla alhaisempi kuin edeltävän kynttilän päätöskurssi ja sen päätöskurssin tulee olla korkeampi kuin edeltävän kynttilän avauskurssi. Härkämäisen peittävän kynttilän tulee esiintyä laskevassa markkinatrendissä.

Karhumainen peittävä kynttilä ("Bearish engulfing") toimii päinvastoin kuin härkämäinen peittävä kynttilä. Sen keho on punainen ja peittää täysin sitä edeltäneen vihreän kynttilän. Karhumaisen peittävän kynttilän tulee esiintyä nousevassa markkinatrendissä, jotta se signaloi trendin potentiaalisesta kääntymisestä. Tällainen kynttilä havaitaan kuvassa 2.6b kolmantena kynttilänä oikealta. Sekä härkäisen että karhumaisen peittävän kynttilän muodostumiseen vaaditaan ero sen avauskurssin sekä sitä edeltävän kynttilän päätöskurssin välillä. Hintaero muodostuu, mikäli ensimmäinen kauppa peittävän kynttilän avausajan jälkeen syntyy eri hinnalla kuin millä edeltävän kynttilän aikavälin puitteissa tehty viimeinen kauppa tapahtuu.

2.2.3 Momentum-indikaattorit

Pääsääntöisen markkinatrendin suunnan vaihtumista voidaan ennustaa myös erilaisia *momentum-indikaattoreita* hyödyntäen. Momentum-indikaattoreilla voidaan mitata markkinatrendin *vahvuutta* nousu- tai laskutrendeissä. Vahvassa nousevassa markkinatrendissä kysyntä ylittää tarjonnan jatkuvasti, mikä heijastuu kurssigraafilla vihreinä kynttilöinä kerta toisensa perään. Jos nousevassa markkinatrendissä on heikkoutta, kasvaa tarjonta ajoittain kysyntää suuremmaksi ja kurssigraafilla tämä havaitaan korjausliikkeinä, jolloin markkinahinta siirtyy ajoittain alemmas, kunnes jatkaa taas nousua ylöspäin. Laskevassa markkinatrendissä vahvuus taas nähdään alituiseen punaisia kynttilöitä ja heikkous heijastuu kurssigraafilla korjausliikkeinä ylöspäin.

Pääsääntöisen markkinatrendin suunnan vaihtuminen vaatii yleensä sen, että



Kuva 2.7: Momentumin heikkeneminen ja markkinatrendin kääntyminen

trendissä voidaan havaita heikkoutta. Täten momentum-indikaattoreita hyödyntäen voidaan olettaa trendin jatkuvan, mikäli se on vahva, ja ne taas antavat ennenaikaisen signaalin potentiaalisesta trendin kääntymisestä, mikäli se on heikko. Kuva 2.7 esittää markkinatrendin kääntymistä *S&P* 500-indeksissä viikon mittaisilla kynttilöillä finanssikriisin pohjalukemista vuosina 2008 - 2009.

Kuvan graafin vasemmassa laidassa havaitaan enimmäkseen punaisia kynttilöitä signaloiden laskutrendin vahvuudesta. Seuraavaksi havaitaan pidempiaikainen jakso, jolloin kysyntä ja tarjonta kohtaavat kutakuinkin ja hinta liikkuu osan ajasta myös ylöspäin. Tätä heikkoa laskutrendiä seuraa graafin oikeassa laidassa pääsääntöisen markkinatrendin kääntymistä indikoiva nousu, joka käynnistyy trendin kääntymistä indikoivalla vasara-kynttilällä. Nousu on vahva, josta kertoo kerta toisensa jälkeen muodostuneet vihreät kynttilät.

RSI

RSI ("Relative strength index") [6] on eräs markkinatrendin vahvuutta mittaava indikaattori ja sen arvo on laskettavissa kullekin kurssigraafin kynttilälle erikseen. *RSI* määritetään lausekkeella

$$RSI_n = 100 - \frac{100}{1 + \frac{AvgGain_n}{AvgLoss_n}}, \quad (2.2)$$

jossa $AvgGain_n$ on keskimääräinen, painotettu hinnan nousu kynttilöille, joiden päätöshinta on korkeampi kuin edeltävän kynttilän päätöshinta ja $AvgLoss_n$ vastavasti keskimääräinen, painotettu hinnan lasku kynttilöille, joiden päätöshinta on alhaisempi kuin edeltävän kynttilän päätöshinta. Lauseke 2.2 palauttaa desimaaliluvun

| Päivämäärä | Päätöshinta | Muutos | $AvgGain_5$ | $AvgLoss_5$ | RSI_5 |
|------------|-------------|--------|--------------------------------|-------------------------------------|---------|
| 30.10.2020 | 10.55 | - | - | - | - |
| 2.11.2020 | 10.55 | 0 | - | - | - |
| 3.11.2020 | 10.58 | 0.03 | - | - | - |
| 4.11.2020 | 10.59 | 0.01 | - | - | - |
| 5.11.2020 | 10.60 | 0.01 | - | - | - |
| 6.11.2020 | 10.57 | -0.03 | $\frac{0.05}{5} = 0.01$ | $\frac{0.03}{5} = 0.006$ | 62.50 |
| 9.11.2020 | 10.56 | -0.01 | $\frac{4*0.01}{5} = 0.008$ | $\frac{4*0.006+0.01}{5} = 0.0068$ | 54.05 |
| 10.11.2020 | 10.56 | 0 | $\frac{4*0.008}{5} = 0.0064$ | $\frac{4*0.0068}{5} = 0.00544$ | 54.05 |
| 11.11.2020 | 10.52 | -0.04 | $\frac{4*0.0064}{5} = 0.00512$ | $\frac{4*0.00544+0.04}{5} = 0.0124$ | 29.30 |

Taulukko 2.1: Kuvitteellinen markkinadata ja RSI_5 :n laskentaesimerkki

väliltä $]0, 100[$. *Painokertoimeksi* valitaan yleensä $n = 14$.

Tarkastellaan seuraavaksi $AvgGain_n$:n ja $AvgLoss_n$:n ja sitä kautta myös RSI_n :n arvonmäärittystä. Taulukko 2.1 sisältää päivittäistä markkinadataa kuvitteelliselle arvopaperille sekä näyttää, miten RSI_5 määritellään kutakin pörssipäivää kuvaavalle riville. Taulukon sarakkeet *Päivämäärä* ja *Päätöshinta* ovat julkista markkinadataa ja sarake *Muutos* kuvaa arvopaperin päätöshinnan muutosta edeltävän päivän päätöshintaan nähden. Sarakkeet $AvgGain_5$, $AvgLoss_5$ ja RSI_5 määritellään kuten on yllä kuvattu.

Havaitaan, että päivämäärän 6.11.2020 rivin $AvgGain_5$:n arvo lasketaan taulukossa esitetyin tavoin ko. rivin ja sitä edeltäneen neljän rivin positiivisten muutosten perusteella ja vastaava pätee $AvgLoss_5$:lle kyseisten rivien negatiivisille muutoksille. Lisäksi päivämäärästä 9.11.2020 alkaen $AvgGain_5$:n ja $AvgLoss_5$:n arvot riippuvat vain edeltävän päivämäärän arvoistaan, ko. päivämäärän muutoksesta ja valitusta painokertoimesta.

Kuten taulukon 2.1 esimerkistä havaitaan, kun 6.11.2020 viimeisen viiden pörssipäivän aikana markkinahinta oli noussut enemmän kuin se oli laskenut, oli $RSI_5 = 62.50 > 50$. Kun taas 11.11.2020 pörssipäivän päätöshinta oli jo tuota merkittävästi alhaisempi, oli $RSI_5 = 29.30 < 50$. Yleisesti ottaen pätee, että selkeässä nousutrendissä RSI_n on selvästi yli 50 ja selkeässä laskutrendissä taas selvästi alle 50 ja RSI_n :n arvo kertoo näin trendin vahvuudesta. Jos markkinahinta kuitenkin laskee laskutrendissä tai nousee nousutrendissä hyvin voimakkaasti, voi RSI -indikaattori signaloida *ylimyydyttä* ("Oversold") tai *yliostetusta* ("Overbought") tilasta. Tällöin sen sijaan, että voidaan odottaa markkinahinnan liikkuvan lyhyellä aikavälillä trendin suuntaisesti, on varsin todennäköistä, että nähdään myös korjausliikkeitä trendin vastaiseen suuntaan, jolloin markkinan momentum normalisoituu ja RSI_n palautuu lähemmäksi arvoa 50.

Kuvassa 2.8 on havainnollistettu tilanne, jossa laskevassa trendissä momentum heikkenee ja markkinatrendi kääntyy. Kuvan markkina ajautuu aluksi suuren hinnan laskun seurauksena ylimyytyyn tilaan, jossa RSI_{14} :n arvo on alle 30. Tämä havaitaan kuvan alalaidasta RSI_{14} :n käyrän piirtymisenä violetin suorakaiteen alapuolelle. Markkinahinta korjautuu seuraavaksi ylöspäin ja samanaikaisesti RSI_{14} palautuu lähemmäksi arvoa 50. Lopulta markkinatrendi kääntyy ja tätä ennen graafille piirtyvät myös positiiviset *RSI-divergenssit* ("RSI divergence"). Positiivinen RSI-divergenssi tarkoittaa sitä, että markkinahinta muodostaa matalamman pohjan samanaikaisesti, kun RSI_{14} muodostaa korkeamman pohjan. Tämä kertoo momentumin heikkenemisestä ja signaloi trendin potentiaalista kääntymistä. Negatiivinen RSI-divergenssi tapahtuisi nousevassa markkinatrendissä päinvastoin.

MACD

MACD ("Moving average convergence divergence") [7] on *RSI*:n ohella toinen yleisesti käytetty indikaattori markkinatrendin vahvuuden mittaamiseen ja trendin suunnan vaihtumisen ennustamiseen. *MACD* voidaan määrittää valitulla ajanjaksolla kullekin kynttilälle erikseen niille 12. ja 26. viimeisimmän kynttilän päätöskurssien eksponentiaalisten liukuvien keskiarvojen erotuksena eli lausekkeella

$$MACD = EMA_{x_i,12} - EMA_{x_i,26},$$

jossa $EMA_{x_i,n}$ on määritelty lausekkeessa 2.1.

MACD-käyrä saadaan laskemalla valitun ajanjakson kynttilöiden *MACD*-arvot. Tämän *signaalikäyränä* toimii *MACD*-käyrän datapisteille niistä yhdeksän viimeisimmän havainnon eksponentiaalinen liukuva keskiarvo. *MACD*-käyrän ja sen signaalikäyrän leikkauksia voidaan hyödyntää potentiaalisina signaaleina markkinatrendin suunnan muutoksesta. Suunnan muutosta voidaan ennustaa myös *MACD*-arvojen ja markkinan vastaavien päätöskurssien välillä muodostuvien divergenssien kautta, kuten kuvassa 2.8 on *RSI*:n arvoille ja markkinahinnalle havainnollistettu.

Kuvan 2.9 graafi esittää *S&P 500*-indeksifutuurin yhden tunnin mittaisia tunti-kynttilöitä tammikuun 2020 lopulta. Graafiin on myös piirretty tuntikynttilöiden päätöskurssille laskettu *MACD*-käyrä (sinisellä) ja sen signaalikäyrä (oranssilla). Havaitaan, että indeksifutuurin noin 3330 pisteen lukemassa *MACD*-käyrä leikkaa sen signaalikäyrän siirtyen sen alapuolelle. Tämä on potentiaalinen myyntisignaali ja indeksifutuurin arvo jatkaa leikkauksen jälkeen alaspäin noin 3240 pisteen lukemaan saakka. Seuraava *MACD*-käyrän ja sen signaalikäyrän leikkaus tapahtuu markkinatrendin kääntymisen jälkeen nousevaksi noin 3250 pisteen lukemassa. Tämä toimii potentiaalisena ostosignaalina ja indeksifutuurin arvo nouseekin välissä tapahtuvan korjausliik-



Kuva 2.8: RSI_{14} :n arvonmuutos markkinatrendin kääntyessä. Positiiviset RSI-divergenssit janoilla kuvattuna.

keen jälkeen.

MACD antaa myös monesti virheellisiä osto- ja myyntisignaaleja *MACD*- ja signaalikäyrien leikkausten myötä. Näitä tapahtuu varsinkin silloin, kun markkinahinnan muutokset tasaantuvat selkeämmän nousun tai laskun jälkeen ja siirrytään ajanjaksoon, jonka aikana hinta liikkuu tiettyjen raamien sisällä, kunnes aikanaan jatkaa joko trendin suuntaisesti tai trendi kääntyy [8]. *MACD*, *RSI* ja muita indikaattoreita kannattaakin tarkastella yhtäaikaa, ja mikäli useampi niistä antaa samankaltaisen signaalin, voidaan signaalia pitää luotettavampana trendin jatkon ennustamisen suhteen.

2.3 Yhteenveto

Alaluvussa 2.1 tarkasteltiin markkinoiden toimintaa ja sitä, miten kysynnän ja tarjonnan suhde vaikuttaa kaupankäynnin kohteena olevien arvopaperien hinnoitteluun. Havaittiin, että lyhyellä aikavälillä arvopaperien hinnat eivät selity yksin niiden taustalla olevien fundamenttien perusteella, vaan hintoihin vaikuttavat myös monet tekniset tekijät. Näihin tekijöihin liittyen tarkastelin teknisen analyysin oletuksia, joihin alaluvussa 2.2 käsitellyt teknisen analyysin indikaattorit pohjautuivat. Indikaattoreilla pyrittiin tulevan pääsääntöisen markkinatrendin tunnistamiseen olettaen, että markkinahinta siirtyy jatkossa trendin osoittamaan suuntaan.

Teknisen analyysin indikaattoreista esitin liukuvat keskiarvot sekä strategioita niiden hyödyntämiseen trendin tunnistuksessa, trendin kääntymistä signaloivia kynttilämuodostelmia ja trendin vahvuutta mittaavia momentum-indikaattoreita. Momentum-indikaattoreilla saatiin monien kynttilämuodostelmien tapaan signaaleja potentiaalisesta trendin kääntymisestä. Lopulta todettiin, että mitä useampi indikaattori signaloi trendin jatkumisesta tai kääntymisestä, sitä luotettavampana signaalia voidaan pitää markkinahinnan tulevan hinnanmuodostuksen ennustamisen suhteen.



Kuva 2.9: S&P 500-indeksifutuurin tuntikynttilöitä tammikuun 2020 lopulta. Kuvan alalaidassa kynttilöiden päätöskursseille laskettu sininen *MACD*-käyrä ja sen oranssi signaalikäyrä.

3. Koneoppimisen soveltaminen tekniseen markkina-analyysiin

Tässä luvussa käsittelen teknisen analyysin näkökulmasta koneoppimisen soveltamista markkinan tulevien hintaliikkeiden ennustamiseen. Alaluvuissa 3.1 ja 3.2 esittelen koneoppimisen menetelmät, joita olen käyttänyt omassa tutkimuksessani; kNN-luokittelijan ja satunnaisen metsän. Alaluvussa 3.3 käsittelen aiempaa tutkimusta koneoppimisen ja teknisen analyysin parissa.

Aiempi tutkimus pohjautuu kNN-luokittelijan ja satunnaisen metsän sovellutuksiin sekä muutamiin muihin ohjatun oppimisen menetelmiin. Tutkimuksen tulokset osoittavat, että markkinan tulevaa hinnanmuodostusta voidaan ennustaa lyhyellä aikavälillä hyvällä tarkkuudella.

3.1 KNN-luokittelija

Artikkelissaan [16] Thomas Cover ja Peter Hart esittelivät *lähimmän naapurin säännön* ("Nearest neighbor rule"), jota voidaan hyödyntää uusien havaintovektorien luokitteluun. Sääntöä käytettäessä oletetaan, että käsitellään (x, y) muotoista dataa, missä x on metrisen avaruuden X vektori ja y sitä vastaava luokka. Määritellään lisäksi, että avaruudessa X on käytössä metriikka d . Tällöin uudelle havainnolle (x, y) , jossa y ei ole tiedossa pätee, että \tilde{x} on x :n lähin naapuri avaruudessa $X = \{x_1, x_2, \dots, x_n\}$, jos

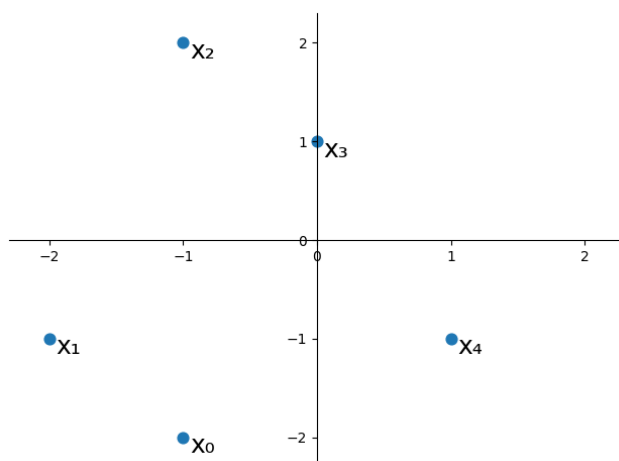
$$d(x, \tilde{x}) = \min d(x, x_i), \quad i = 1, 2, \dots, n.$$

Sääntöön perustuva yhden lähimmän naapurin luokittelija (jatkossa "1NN-luokittelija") luokittelee tällöin vektorin x luokaksi \tilde{y} , missä \tilde{y} on vektoria \tilde{x} vastaava luokka. Luokittelija tekee virheen, jos $\tilde{y} \neq y$.

Cover ja Hart vertasivat myös 1NN-luokittelijan *luokitteluvirhettä* ("Generalization error") geneeriseen optimaaliseen luokittelijaan, jolle uusien havaintojen yhteisjakaumat niin vektoreille $x \in X$ kuin niiden luokillekin ovat ennalta tiedossa. He todistivat, että jos pisteitä (x, y) voidaan generoida jakaumilla identtisesti ja toisistaan

| X | Y |
|------------------|-----------|
| $x_1 = (-2, -1)$ | $y_1 = 0$ |
| $x_2 = (-1, 2)$ | $y_2 = 1$ |
| $x_3 = (0, 1)$ | $y_3 = 1$ |
| $x_4 = (1, -1)$ | $y_4 = 0$ |

Taulukko 3.1: KNN-luokittelijan opetusdata



Kuva 3.1: Taulukon 3.1 joukko X sekä uusi havaintovektori x_0 havainnollistettu koordinaatiston pisteinä

riippumattomasti ja avaruus X on tarpeeksi suuri, 1NN-luokittelija tekee alle kaksi kertaa niin paljon virheitä kuin yo. optimaalinen luokittelija. Lisäksi he osoittivat, että mikäli X on suuri, uudet havaintovektorit saattavat olla perusteltuja luokitella niiden $k \in N$ lähimpien naapurien perusteella, vaikka yhteen lähimpään naapuriin perustuva menetelmä on jo itsessään kelvollinen. Tällaista kNN-luokittelijaa käytettäessä vektorin x luokaksi määritetään se, jonka frekvenssi sen k :n lähimmän naapurin luokkien joukossa on suurin.

Seuraava esimerkki havainnollistaa kNN-luokittelijan toimintaa. Oletetaan, että käsitellään avaruuden R^2 vektoreita, metriikaksi d valitaan euklidinen etäisyysmitta ja jokaiselle vektorille voidaan assosoida luokka 0 tai 1. Luokittelijalle annetaan opetusdatana taulukon 3.1 joukko X sekä sen vektoreita vastaavien luokkien joukko Y . Valitaan lisäksi $k = 3$.

Halutaan luokitella uusi havaintovektori $x_0 = (-1, -2)$, joka on kuvattu joukon X ohessa kuvassa 3.1. Vektorin x_0 euklidiset etäisyydet joukon X vektoreihin ovat

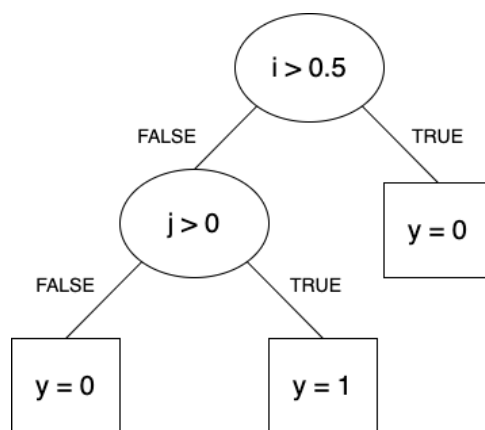
$$\begin{aligned}
 d(x_0, x_1) &\approx 1.41, \\
 d(x_0, x_2) &= 4, \\
 d(x_0, x_3) &\approx 3.16 \text{ ja} \\
 d(x_0, x_4) &\approx 2.24
 \end{aligned}$$

ja kolmen lähimmän naapurin joukko on siis $\{x_1, x_3, x_4\}$. Näiden luokkien joukko on $\{y_1, y_3, y_4\} = \{0, 1, 0\}$, ja koska luokalla 0 on tässä suurempi frekvenssi kuin luokalla 1, luokittelija pääättelee vektorin x_0 luokaksi $y_0 = 0$.

3.2 Satunnainen metsä

L. Breiman esitteli artikkelissaan [13] luokittelijat, jotka perustuvat useilla saman ohjatun oppimisen menetelmällä opettujen mallien yhteiskäyttöön. Havaintojen luokittelu tällaisella joukolla eri malleja tapahtuu äänestyksellä ja valiten luokaksi se, jonka frekvenssi yksittäisten mallien palauttamista luokista on suurin. Syyt siihen, miksi tällaiset luokittelijat ovat mielenkiintoisia joitakin ohjatun oppimisen menetelmiä hyödynnettäessä perustuvat Breimanin aiemmassa raportissaan [14] esittämiin havaintoihin. Hän esitti, että ohjatun oppimisen menetelmä A on *epästabiili*, jos pätee, että datajoukoilla X ja \tilde{X} , missä \tilde{X} vastaa X :ä pienellä muutoksella, A :lla opetetut mallit voivat poikeata uusien havaintojen luokittelutavan perusteella merkittävästi toisistaan. Samalla hän osoitti, että epästabiilien menetelmien tarkkuutta voidaan tehostaa hyödyntäen menetelmillä opettuja malleja luokittelussa yhteiskäyttöisesti äänestämällä. Stabiilien menetelmien kuten lähimmän naapurin sääntöön perustuvien luokittelijoiden hyödyntäminen yhteiskäytössä johtaa sen sijaan heikompaan suorituskykyyn, kuin mitä menetelmillä itsessään saavutetaan [13].

Artikkelissaan [13] Breiman esitti tavan muodostaa edellä mainitun kaltaisen luokittelijan valitulla ohjatun oppimisen menetelmällä A ja käytössä olevalla opetusdatajoukolla X . Ensin joukosta X poimitaan $n:n$ alkion kokoisia satunnaisotoksia valitsemalla jokainen alkio sattumanvaraisesti koko joukosta ("Sampling with replacement"), jolloin poimitut alkioit ovat toisistaan riippumattomia. Tällöin satunnaisotokset on generoitu jakaumasta, joka toimii approksimaationa jakaumalle, jonka perusteella X on generoitu [13]. Tämän jälkeen jokaista satunnaisotosta kohti menetelmällä A opetetaan yksittäinen malli, ja malleja hyödynnetään yhteiskäyttöisesti luokitteluun. Breimanin esittämä perinteinen *satunnaisen metsän* ("Random forest") sovellutus on tällainen luokittelija, jossa luokittelu perustuu yhteiskäyttöisesti *päätöspuiden* ("Decision tree") kautta opettujen mallien hyödyntämiseen [13].



Kuva 3.2: Binäärinen malli, joka saatettaisiin generoida päätöspuulla käyttäen opetusdatana taulukossa 3.1 esitettyä dataa

Päätöspuu on epästabiili ohjatun oppimisen menetelmä. Käytettäessä päätöspuita luokitteluun, voidaan menetelmällä generoida malleja luokittelemaan uusia havainnoita syötetyn opetusdatan ominaisuuksiin perustuen. Generoidut mallit ovat puurakenteisia ja ne koostuvat *solmuista*, *lehdistä* ja *haaroista*. Jokaisesta solmusta lähtee puun alemmalle tasolle kaksi tai useampaa haaraa ja haaroihin liittyy sääntöjä, joiden perusteella malli osaa päätellä, missä alemman tason solmussa tai lehdessä luokiteltavaa havaintoa tulee tarkastella. Yksittäisestä solmusta lähtevien haarojen säännöt ovat yksiselitteisiä tarkoittaen, että täsmälleen yksi sääntö pätee mille tahansa luokiteltavalle havainnolle. Uuden havainnon luokittelu aloitetaan puun ylimmän tason juuri-solmusta. Puu haarautuu lopulta lehtiin ja luokiteltavan havainnon luokka määräytyy lehdestä, jossa havaintoa lopulta tarkastellaan.

Kuvassa 3.2 on esimerkkinä havainnollistettu binäärinen malli, joka olisi ainakin teoriassa mahdollista generoida päätöspuulla käyttäen opetusdatana taulukon 3.1 datajoukkoa. Puurakenteisen mallin solmut on kuvattu ellipseinä ja lehdet suorakulmioina. Haaroihin liittyvät säännöt on kirjattu osittain solmuihin, joista haarat lähtevät, sillä tämä on binääriselle mallille mahdollista. Malli palauttaa uudelle havainnolle $x = (i, j)$, missä $i, j \in R$ luokan $y = 1$, jos $i \leq 0.5$ ja $j > 0$. Muutoin malli palauttaa havainnolle luokan $y = 0$.

Uudemmassa tutkimuspaperissaan [15] Breiman tarkasteli satunnaisen metsän suorituskykyä. Hän osoitti, että *suurten lukujen lakiin* ("Law of large numbers") perustuen menetelmän luokitteluvirhe suppenee, kun puiden lukumäärää kasvatetaan. Tällöin *liiallista sopeutumista* ("Overfitting") opetusdatan pisteiden luokitteluun ei myöskään tapahdu. Breiman myös todisti, että satunnaisen metsän luokitteluvirhe muodostuu vain sen yksittäisten puiden välisistä korrelaatioista sekä puiden ennustetarkkuuksista uusien havaintojen luokitteluun.

Aiemmin esittämässään satunnaisen metsän sovellutuksessa [13] Breiman käytti

puiden opettamiseen perinteistä päätöspuualgoritmia, jolla generoitavat mallit pyrkivät oppimaan sääntöperusteisen strategian luokitella uusia havaintoja tuntemattoman jakauman pohjalta, mistä niille annetun opetusdatan pisteet ovat peräisin. Tällöin mikäli opetettavien puiden opetusdatojen välillä on merkittävästi korrelaatiota, muodostuu korrelaatiota myös opetettujen puiden välille.

Paperissaan [15] Breiman tarkasteli tapoja muuttaa perinteistä päätöspuualgoritmia, jotta hyödynnettäessä sitä satunnaisen metsän toteutuksessa saataisiin generoitujen mallien välinen korrelaatio minimoitua. Hän havaitsi, että tämä onnistuu muuttamalla algoritmia siten, että yksittäistä mallia generoitaessa puurakenteen jokaisessa solmussa valitaan satunnaisotoksena yksi tai muutama opetusdatan ominaisuus ja sääntöperusteinen haarautuminen solmusta alemmalle tasolle määritetään datan valittujen ominaisuuksien perusteella. Breimanin tulokset osoittavat, että tällaisten mallien joukko on yhteiskäyttöisesti tarkempi uusien havaintojen luokitteluun kuin perinteiseen päätöspuualgoritmiin perustuvien mallien joukko, vaikka perinteiseen algoritmiin perustuvat mallit ovatkin itsessään todennäköisesti tarkempia.

3.3 Aiempia koneoppimisen sovellutuksia

Osto- ja myyntitoimeksiantojen ajoittaminen kNN-luokittelijan avulla

L. A. Teixeira ja A. L. I. de Oliveira esittelivät tutkimuspaperissaan [23] tavan hyödyntää teknistä analyysiä ja kNN-luokittelijaa osakkeilla kaupankäyntiin. He osoittivat, että heidän luokittelijansa saavuttaa halutulla tarpeeksi pitkällä aikavälillä merkittävästi ”osta ja pidä” -strategiaa paremman markkinatuoton. ”Osta ja pidä” -strategialla viitataan strategiaan, jossa osakkeet ostetaan aikavälin ensimmäisenä päivänä ja pidetään aikavälin viimeiseen päivään saakka. Tämä on mielekäs vertailustrategia siitä syystä, että tehokkaiden markkinoiden hypoteesi nojaa siihen, että tulevia markkinaliikkeitä on mahdoton ennustaa ja markkinoilla on järkevää olla näin jatkuvasti mukana.

Tutkimusdatanaan Teixeira ja de Oliveira käyttivät viittätoista brasilialaisen pörssin eri osaketta, joista jokaiselle oli tiedossa päiväkohtaisesti päätöshinta, ylin ja alin notteeraus sekä osakevaihtomäärä. Heidän menetelmässään kukin em. tiedoista koostunut datapiste muunnettiin ensin vektoriksi, joka koostui noin kahdestakymmenestä ominaisuudesta sekä luokasta *osta*, *myy*, tai *pidä*. Vektoreita käytettiin kNN-luokittelijan opetus- ja testidatana ja niiden ominaisuudet määritettiin suoraan teknisen analyysin indikaattoreita hyödyntäen. Ominaisuudet perustuivat mm. päivittäisten päätöskurssien sekä osakevaihtomäärien kautta muodostettuihin liukuviin keskiarvoihin sekä RSI_{14} :n arvoihin. Ominaisuuksia laskettiin myös päätöskurssien ja tietyille aikaväleille valittujen *stokastisen oskillaattorin* (”Stochastics oscillator”) sekä *bollinger-nauhojen*

| Stock | Classifier precision (%) | Difference to buy-and-hold (%) | Buy operations per year |
|--------|--------------------------|--------------------------------|-------------------------|
| AMBV4 | 40.78 | +12.20 | 27.85 |
| ARCZ6 | 34.47 | +108.43 | 44.57 |
| BBAS3 | 36.12 | +45.34 | 47.28 |
| BBDC4 | 36.81 | -36.78 | 44.14 |
| CMIG4 | 36.99 | +1.10 | 48.28 |
| CRUZ3 | 36.86 | -8.09 | 38.57 |
| CSNA3 | 39.16 | +18.86 | 48.28 |
| ELET 6 | 37.34 | -10.90 | 54.28 |
| ITAU4 | 37.40 | +115.99 | 44.85 |
| ITSA4 | 37.20 | +76.83 | 37.14 |
| NETC4 | 37.87 | +51.98 | 52.00 |
| PETR4 | 34.59 | +8.13 | 42.14 |
| TNLP4 | 35.96 | +95.20 | 37.71 |
| USIM5 | 40.00 | +47.18 | 43.28 |
| VALE5 | 36.55 | +33.12 | 40.14 |

Kuva 3.3: KNN-luokittelijan tarkkuus ja suorituskyky ”osta ja pidä” -strategiaan verrattuna keskimäärin vuosittain tarkastelluille osakkeille [23]

(”Bollinger bands”) ylä- ja alanauhaa verraten. En käsittele näitä indikaattoreita tutkielmassani tarkemmin.

Tutkimuksessa oli käytössä tili, jonka saldo alussa oli 100,000 brasilian realia ja yhden toimeksiannon kustannus määritettiin olevan vakio 5 realia. Tutkimuksessa jokaisen osto- ja myyntitoimeksiannon yhteydessä käytössä oli aina koko tilin saldo. Luokittelijan opetusdatana käytettiin kolmen edeltävän vuoden päivittäisen markkinadatan perusteella määritettyjä ominaisuuksia ja näitä seuraavan vuoden markkinadatan perusteella määritettyjä ominaisuuksia käytettiin testidatana. Koe toistettiin seitsämälle testivuodelle väliltä 2002 - 2008 ja kullekin viidelletoista osakkeelle. Luokittelijaa lisäksi optimoitiin mm. määrittämällä eksplisiittisesti suurin mahdollinen tappio yksittäiselle kaupalle myymällä omistus heti, jos se oli laskenut yli 3 prosenttia.

Kuvan 3.3 taulukko esittää Texeiran ja de Oliveiran saamia keskimääräisiä vuosittaisia tuloksia. Havaitaan, että luokittelija pärjää paremmin kuin ”osta ja pidä” -strategia kahdelletoista osakkeelle ja kolmelle näistä tuotto tai pienempi tappio on luokittelijan hyväksi jopa yli 95 %. Taulukon oikean sarakkeen esittämistä ostooperaatioiden määristä voidaan taas päätellä, että luokittelija tekee vain muutamia kymmeniä edestakaisia kauppia eri osakkeille vuosittain, mikä puoltaa sen käyttöä yleisesti kaupankäyntiin.

Mielenkiintoinen seikka, joka kuvasta havaitaan on luokittelijan heikko, keskimäärin alle 40 %:n ennustetarkkuus sille, pitäisikö kunakin päivänä osaketta ostaa, myydä vai pitää. Luokittelija on silti merkittävästi tehokkaampi kuin ”osta ja pidä” -strategia, mutta suoraa selitystä Texeira ja de Oliveira eivät tähän anna. Oma oletukseni on, että luokittelija osannee luokitella uuden havainnon luokkaan *myy*, jos markkina on siirtynyt selvästi laskevaan trendiin ja tekee *osta* luokituksen uudelle havainnolle vasta, kun osaketta pidettäessä on jälleen hyvä tuottoriskisuhde. Tällä tarkoitan si-

tä, että tietyn aikavälin yli osakkeen pitämällä voidaan ennustaa tuottopotentiaalin olevan tappiopotentiaalia korkeampi. Myös rajattu kauppakohtainen kolmen prosentin maksimitappio johtaa siihen, että välttyään merkittävältä tilin saldon alenemiselta tilanteessa, jossa luokittelija itse olisikin sitä mieltä, että osaketta kannattaa vielä pitää eikä myydä.

Markkinan suunnan ennustaminen lyhyellä aikavälillä

Artikkelissaan [22] J. Patel tutkimusryhmineen tarkastelivat, kuinka tarkasti eri koneoppimisen menetelmillä voidaan ennustaa, nouseeko vai laskeeko markkinahinta lyhyellä aikavälillä seuraavaksi. Tutkimusdataksi ryhmä valitsi vuosilta 2003 - 2012 markkinadataa kahdelta Intian pörssin osakkeelta sekä kahdelta ko. pörssin indeksiltä eli yhteensä neljä datajoukkoa. Markkinadata koostui pörssipäivien avaus- ja päätöshinnoista sekä ylimmistä ja alimmista noteerauksista.

Tutkimuksessaan J. Patel et al. muunsivat kymmentä teknisen analyysin indikaattoria hyödyntäen markkinadatan vektoreiksi, joita oli mielekästä käyttää ohjatun oppimisen menetelmien syötteinä. Vertailtaviksi menetelmiksi he valitsivat *eteenpäin syöttävän neuroverkon* ("Artificial feed-forward neural network") [20], *tukivektorikoneen* ("Support vector machine") [17], satunnaisen metsän ja *naive bayes -luokittelijan* ("Naive Bayes classifier") [19]. Hyödynnetyt menetelmät palauttivat syötevektorille ennusteen siitä, laskeeko (-1) vai nouseeko (1) markkinahinta vektoriin liittyvää ajanhetkeä seuraavalla lyhyellä aikavälillä. Lyhyellä aikavälillä tutkimusryhmä tarkoitti todennäköisesti 1 – 5 pörssipäivää perustuen tutkimuksessa teknisinä indikaattoreina hyödynnettyihin 10. päivän liukuviin keskiarvoihin.

Ryhmä tutki myös, miten ennustetarkkuuksiin vaikutti se, muunnettiin markkinadata normalisoiduiksi syötevektoreiksi, joiden ominaisuudet olivat liukulukuja väliltä $[-1, 1]$ vai käytettiin syötevektorien ominaisuuksina diskreettejä kokonaislukuja joukosta $\{-1, 1\}$. Tutkimus perustui siihen, että ryhmä havaitsi, että diskreetit kokonaisluvut sisällyttivät ominaisuuksille niitä vastaavien indikaattorien havaitseman markkinatrendin, mikä voisi heijastua menetelmien tekemiin ennusteisiin.

Taulukko 3.2 esittää ryhmän saamia keskimääräisiä tuloksia hyödynnetyille menetelmille ja datajoukoille. Menetelmien ennustetarkkuudet on esitetty sarakkeissa *Liukulukuominaisuudet* ja *Kokonaislukuominaisuudet* erikseen tilanteissa, joissa syötevektorien ominaisuudet ovat edellä mainitun kaltaisia liukulukuja tai diskreettejä kokonaislukuja. Ennustetarkkuudet määritetään lausekkeella

$$Accuracy = \frac{T_1 + T_{-1}}{T_1 + T_{-1} + F_1 + F_{-1}},$$

joka on yleisesti binäärisessä luokittelussa käytetty tarkkuuden mittaustapa.

| Menetelmä | Liukuluku- ominaisuudet | Kokonaisluku- ominaisuudet |
|--------------------------------|----------------------------|-------------------------------|
| Eteenpäin syöttävä neuroverkko | 0.7494 | 0.8669 |
| Tukivektorikone | 0.7871 | 0.8933 |
| Satunnainen metsä | 0.8359 | 0.8998 |
| Naive bayes -luokittelija | 0.7331 | 0.9019 |

Taulukko 3.2: Ennustetarkkuudet eri menetelmille käytettäessä syötevektoreiden ominaisuuksina liukulukuja tai diskreettejä kokonaislukuja

Symboli T_1 kuvastaa oikein luokiteltujen vektoreiden määrää, kun hinnan odotettiin seuraavaksi nousevan ja T_{-1} oikein luokiteltujen vektoreiden määrää, kun hinnan odotettiin seuraavaksi laskevan. Symboli F_1 kuvastaa vektoreiden lukumäärää, joiden osalta ennustettiin virheellisesti hinnan seuraavaksi nousevan ja F_{-1} taas niiden määrää, joiden osalta ennustettiin virheellisesti hinnan seuraavaksi laskevan.

Taulukosta 3.2 havaitaan, että satunnainen metsä toimii huomattavasti muita menetelmiä paremmin, jos vektoreiden ominaisuudet ovat liukulukuja. Jos taas ominaisuudet ovat diskreettejä kokonaislukuja, on jokainen menetelmä hyvinkin tarkka, vaikka eteenpäin syöttävä neuroverkko jääkin tarkkuuden osalta noin 3 prosenttiyksikköä muille menetelmille. Havaittavaa kuitenkin on, että vektoreiden diskreeteillä kokonaislukuominaisuuksilla pystytään opettamaan tehokkaampia malleja kuin liukulukuominaisuuksilla. Tämä perustuu oletettavasti siihen, että tällöin ominaisuudet sisällyttävät tiedon niitä vastaavien indikaattoreiden havaitsemasta markkinatrendistä.

3.4 Yhteenveto

Tässä luvussa käsittelin koneoppimisen soveltamista tekniseen markkina-analyysiin. Tarkastelin kahta ohjatun oppimisen menetelmää, kNN-luokittelijaa ja satunnaista metsää, niiden toimintaperiaatteiden ja luokittelutarkkuuksien osalta. Näitä menetelmiä olen myös soveltanut omassa tutkimuksessani, jota käsittelen luvussa 4. Esitin myös aiempaa tutkimusta teknisen analyysin ja koneoppimisen sovellutuksista. Aiempi tutkimus pohjautui kNN-luokittelijan ja satunnaisen metsän sovellutusten lisäksi muutamiin muihin ohjatun oppimisen menetelmiin.

Aiemman tutkimuksen tulokset osoittavat, että lyhyellä aikavälillä voidaan tarkasteltavasta markkinasta yli 80 % tarkkuudella ennustaa, onko sen hinta korkeampi vai enintään yhtä korkea aikavälin päätteeksi kuin sen alussa. Tuloksista selviää myös, että maltillisella vuosittaisella kaupankäyntimäärällä voidaan koneoppimista hyödyntäen saavuttaa arvopapereiden ”osta ja pidä” -strategiaa korkeampi markkinatuotto.

4. Tutkimus

Tässä luvussa käsittelen omaa tutkimustani koneoppimisen ja teknisen analyysin yhdistämisestä markkinan tulevan hinnanmuodostuksen ennustamiseen. Pyrin selvittämään, voidaanko ennustaa, mihin suuntaan markkinahinta valittua ajanhetkeä seuraavalla yhden vuorokauden mittaisella ajanjaksolla siirtyy. Markkinahinnan suunta voi olla ylöspäin, jolloin hinta nousee, tai alaspäin, jolloin hinta laskee. Hinta voi myös säilyä ajanjakson ajan ennallaan.

Alaluvussa 4.1 esittelen valitsemani tutkimusdatan, joka koostui *S&P* 500-indeksin ja vehnäfutuurisopimusten päivittäisestä hintadatasta. Kerron myös datalle tehdystä esiprosessoinnista teknisen analyysin indikaattoreita hyödyntäen sekä valitsemistani koneoppimisen menetelmistä, joiden syötteinä tätä esiprosessoitua dataa käytin. Alaluvussa 4.2 käsittelen saamiani tuloksia.

Havaitsin, että toteuttamani datan esiprosessointi yhdessä valitsemieni koneoppimisen menetelmien kanssa eivät tuoneet lisäinformaatiota *S&P* 500-indeksin tulevan hinnanmuodostuksen ennustamiseen siihen nähden, että markkinahinnan suunnan ennustettaisiin olevan aina ylöspäin. Vehnäfutuurisopimusten hinnanmuodostuksen ennustamiseen toteuttamani menetelmät toimivat siinä määrin, että niillä saavutettiin 0.8 – 1.6 % korkeampi ennustetarkkuus kuin valitsemalla markkinahinnan seuraavaksi suunnaksi se, joka oli ollut frekvenssiltään historiallisesti todennäköisempi. Kumpaankaan datajoukon tapauksessa ei menetelmillä kuitenkaan päästy kuin parhaimmillaan 50.8 – 52.5 % ennustetarkkuuksiin. Alaluvussa 4.3 käsittelen jatkotutkimusmahdollisuuksia tuloksiini pohjautuen, ja kuinka ennustettavuutta saattaisi pystyä parantamaan.

4.1 Toteutus

Lähteäkseeni ratkaisemaan yllä määrittämäni tutkimuskysymystä, valitsin kaksi erilaista markkinadatajoukkoa tarkasteluun ja laskin niiden datavektoreille teknisen analyysin indikaattoreita hyödyntäen ominaisuuksia, joita pystyin hyödyntämään koneoppimisen menetelmien syötteinä. Koneoppimisen menetelmiksi valitsin kNN-luokittelijan ja satunnaisen metsän. Päädyin näihin menetelmiin, sillä ne tuntuivat potentiaalisilta

valinnoilta ongelman ratkaisuun. Satunnainen metsä soveltuu monenlaisiin luokitteluongelmiin. Lisäksi koska markkinoilla samankaltaiset kuviot toistuvat uudelleen ja uudelleen [18] ja myös luokittelu kNN-algoritmia hyödyntäen perustuu vektoreiden samankaltaisuuteen, oli ko. luokittelija ongelmanratkaisuun mielestäni perusteltu.

Tutkimusdatani koostui kahdesta datajoukosta; *S&P* 500-osakeindeksin ja vehnäfutuurisopimusten (jatkossa ”vehnä”) päivittäisestä hintadatasta. *S&P* 500-indeksin data on Yahoo Financen julkaisemaa [9] ja se sisältää kullekin pörssipäivälle tiedot päivän avaushinnasta, päätöshinnasta, matalimmasta ja korkeimmasta noteerauksesta sekä kaupankäyntivolyymista ko. päivänä. Markkinadataa hyödynsin aikaväliltä 30.12.1927 - 1.6.2020 ja tässä oli yhteensä n. 23200 riviä. Vehnän hintadata on Quandl:n julkaisemaa [10] ja se sisältää samat tiedot kuin *S&P* 500-indeksin hintadatakin, tosin hieman erilaisessa formaatissa. Hyödyntämäni markkinadata oli aikaväliltä 1.7.1959 - 1.6.2020 ja rivejä tässä oli n. 15300 kpl.

Tutkimuksen ensimmäisessä vaiheessa esiprosessoin molemmat datajoukot yhtenevään muotoon. Sillä molemmat datajoukot olivat CSV-formaatissa ja *S&P* 500-indeksin data oli eheää, en muokannut paljonkaan alkuperäistä dataa. Kaupankäyntihinnat muutin kaksidesimaaliseksi, sillä datan perusteella oli selvää, että tämä esitystapa oli oikea. Poistin myös kaupankäyntivolyymisarakkeen, sillä volyyymia ei ollut koko markkinadatalle saatavilla ja hyödyntämäni teknisen analyysin indikaattorit eivät ottaneet kantaa kaupankäyntivolyymien muutoksiin.

Vehnädatan tapauksessa järjestin datan uudelleen ja poistin sarakkeita, joita ei *S&P* 500-indeksin markkinadatassa ollut. Tein myös tietoisesti valinnan poistaa datasta rivit, joilta hintatiedot puuttuivat ainakin osittain. Näitä rivejä oli 19 kpl sattumanvaraisissa kohdissa dataa, ja oletin, että ko. rivien poistaminen ei vaikuttaisi merkittävästi tuloksiini. Poistettujen rivien tapauksessa dataa tarkasteltiin, kuten poistetusta rivistä seuraava rivi olisi ollut poistetun rivin tilalla edeltävästä katsottuna.

Yllä mainittujen toimenpiteiden jälkeen lisäsin molempien datajoukkojen riveille lisätietoina muutoksen rivin ja sitä seuraavan rivin päätöshintojen välillä, ja onko tuo muutos positiivinen. Tavoitteenani tutkimuksessa oli juurikin pyrkiä ennustamaan tuon jälkimmäisen sarakkeen arvoja, eli onko seuraavan pörssipäivän päätöshinta korkeampi kuin edeltäjänsä vai ei. Esiprosessoidut datajoukot olivat tässä tilanteessa taulukon 4.1 esittämässä formaatissa. Taulukosta voi havaita, että jos pörssipäivän päätöshinta on sama kuin edeltäjänsä, annetaan edeltävää pörssipäivää kuvaavalle riville luokka 0.

Taulukon 4.1 kaltaisille riveille määritin niiden päätöshintoihinsa perustuen seuraavaksi ominaisuuksia eri teknisen analyysin indikaattoreita hyödyntäen. Trendiä seuraaviksi indikaattoreiksi valitsin liukuvat keskiarvot SMA_3 , SMA_7 sekä SMA_{20} ja momentum-indikaattoreiksi $RSI_{5:n}$ ja $RSI_{14:n}$, jotka painottavat tarkastelluille pörssipäiville viimeaikaisten pörssipäivien päätöshintojen muutoksia eri kertoimilla. Jot-

| Päivämäärä | Avaus | Korkein | Alin | Päätös | Muutos seuraavaan | Luokka |
|------------|-------|---------|-------|--------|-------------------|--------|
| 24.1.1978 | 89.24 | 89.80 | 88.67 | 89.25 | 0.14 | 1 |
| 25.1.1978 | 89.25 | 89.94 | 88.83 | 89.39 | -0.81 | 0 |
| 26.1.1978 | 89.39 | 89.79 | 88.31 | 88.58 | 0.00 | 0 |
| 27.1.1978 | 88.58 | 89.10 | 88.02 | 88.58 | 0.76 | 1 |

Taulukko 4.1: *S&P* 500-indeksin esiprosessoitua markkinadataa

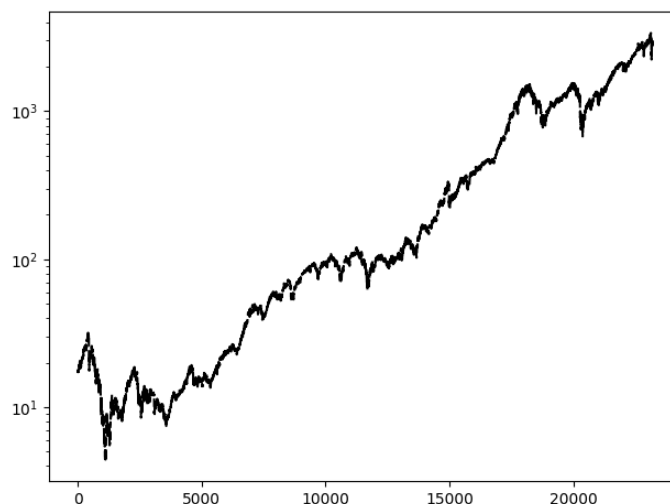
ta tällaista dataa oli mielekästä hyödyntää valitsemieni koneoppimisen menetelmien syötteinä, skaalasin ominaisuuksien arvot välille $[-1, 1]$. Skaalausperusteena liukuvien keskiarvojen tapauksessa oli määrittää kunkin rivin päätöshinnan suhteellinen etäisyys ko. liukuvasta keskiarvosta. Jos päätöshinta oli liukuvan keskiarvon yläpuolella, suhteellinen etäisyys oli positiivinen ja muutoin negatiivinen. Ääriarvot -1 ja 1 annettiin riveille, joiden päätöshinnan suhde tarkastellusta liukuvasta keskiarvosta oli suurin koko datajoukon yli. RSI_5 :n ja RSI_{14} :n arvojen skaalauksen toteutin lausekkeella

$$RSI_n^{sc} = \frac{RSI_n - 50}{50}, n \in N. \quad (4.1)$$

Näiden indikaattorien ohessa määritin jokaisen rivin toteuttavan jonkin seuraavista kynttilämuodostelmista: vasara, pyrstötähti, härkämäinen peittävä kynttilä, karhumainen peittävä kynttilä, tai ei määritelty, mikäli rivi ei toteuttanut yhtäkään näistä muodostelmista edeltävän rivin määrittämään päiväkynttilään suhteutettuna. Nämä alkiot kuvasin myös ominaisuuksiksi välille $[-1, 1]$ määritellen jokaiselle niistä oman vakioarvonsa ja tuoden riveille asetettaviin arvoihin hieman satunnaisuutta.

Tutkimustulokset keräsin ajamalla sekä Scikit-learn ohjelmiston kNN-luokittelijaa [11] että satunnaista metsää [12] molemmille datajoukoille erikseen eri parametrikombinaatioilla. KNN-luokittelijalle testasin sen vakioparametrien ohessa, miten eri k :n arvot, valitut ominaisuudet sekä opetusdatan koko aikajärjestyksessä ensimmäisestä datapisteestä alkaen vaikuttivat ennustetarkkuuteen testidatalla. Toteutin myös mahdollisuuden mallin uudelleenopettamiseen halutun iteraatiomäärän $i \in N$ jälkeen, eli aina kun i kpl seuraavia testidatan pisteitä oli luokiteltu.

Satunnaista metsää ajoin eri ominaisuusjoukoilla, opetusdatan ko'oilla ja muutoin menetelmän vakioparametreilla, paitsi että testasin myös puiden lukumäärän ja päätöspuualgoritmiin liittyvien puiden maksimisyvyyksien ja jakokriteerifunktion vallinnan muuttamista. Jakokriteerifunktion muuttaminen ei joskaan vaikuttanut tutkimustuloksiini ja jatkossa esitettävät tulokset perustuvat sen osalta vakioparameterina olevan gini-indeksin käyttöön. Testidataksi valitsin molempien menetelmien osalta loppuosan datasta, jota ei käytetty opetusdatana. Tulostin myös erilaisia kuvaajia tulosten tulkitsemista varten.



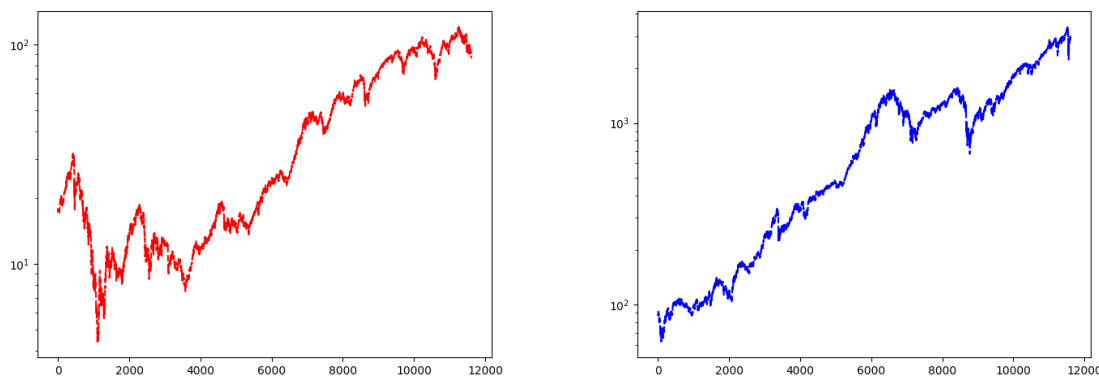
Kuva 4.1: *S&P* 500-indeksin hinnankehitys logaritmisella asteikolla vuosina 1928 - 2020

4.2 Tulokset

Tutkimuksessa testasin joukon erilaisia parametrikombinaatioita kumpaakin datajoukkoa ja koneoppimisen menetelmää käyttäen. *S&P* 500-indeksin osalta havaitsin, että kNN-luokittelijalla voitiin suorittamalla datan esikäsittelyllä opettaa parhaimmillaan malli, joka ennustaa uudelle datapisteelle 51.4 % tarkkuudella, onko sitä seuraavan pörssipäivän päätöskurssi korkeampi (1), tai enintään yhtä suuri (0) kuin ko. päivän päätöskurssi. Parhaimman mallin parametreiksi valitsin $k = 3$, opetus- ja testidatan jaoin ajan perusteella järjestettynä suhteessa 1 : 1 ja määritin, että malli uudelleenopetetaan jokaisen sadan testidatan pisteen havaitsemisen jälkeen. Datajoukoksi rajasin esikäätelyn datan rivien ET_SMA_3 , ET_SMA_{20} ja RSI_5^{sc} sarakkeet, joissa ET_SMA_n , $n \in N$ kuvaa päätöskurssin suhteellista etäisyyttä n . päivän liukuvasta keskiarvostaan. RSI_5^{sc} on määritelty lausekkeessa 4.1.

S&P 500-indeksin pörssipäivien päätöskurssit kaiken käytössä olleen markkinedatan osalta on havainnollistettu logaritmisella asteikolla kuvassa 4.1. Kuvat 4.2a ja 4.2b taas esittävät yllä kuvatun parhaimman mallin alkuperäistä opetus- ja testidataa vastaavien indeksin päätöskurssien hinnankehitystä.

Havaitaan, että markkina seuraa pitkällä aikavälillä tiettyjä säännönmukaisuuksia; laskutrendissä markkinahinta laskee yleensä voimakkaasti lyhyellä aikavälillä, kun taas nousutrendissä se nousee usein maltillisissa jaksoissa ylöspäin. Lisäksi havaitaan, että pitkällä aikavälillä tarkasteltuna opetus- ja testidataa vastaavien päätöskurssien muutokset vaikuttavat muutoin samankaltaisilta, paitsi että testidataa kohti ei tapahdu yhtä rajua kurssilaskua, mitä opetusdataa kohti datan alkupuoliskolta nähdään. Tämä huomioiden markkinedatan jako suhteessa 1 : 1 ja käyttäen ensimmäistä puolis-



(a) Parhaimman kNN-algoritmilla opetetun mallin alkuperäistä opetusdataa vastaavat päätöskurssit, *S&P* 500 vuosina 1928 - 1974

(b) Parhaimman kNN-algoritmilla opetetun mallin testidataa vastaavat päätöskurssit, *S&P* 500 vuosina 1974 - 2020

Kuva 4.2: *S&P* 500-indeksin markkinadatan jako suhteessa 1 : 1

| | Ennuste: 0 | Ennuste: 1 |
|----------|------------|------------|
| Oikea: 0 | 2659 | 2788 |
| Oikea: 1 | 2844 | 3306 |

Taulukko 4.2: Tarkkuusmatriisi *S&P* 500-indeksin testidatan luokittelusta kNN-algoritmia käyttäen

koa opetusdatana ja toista testidatana vaikuttaa kutakuinkin järkevältä.

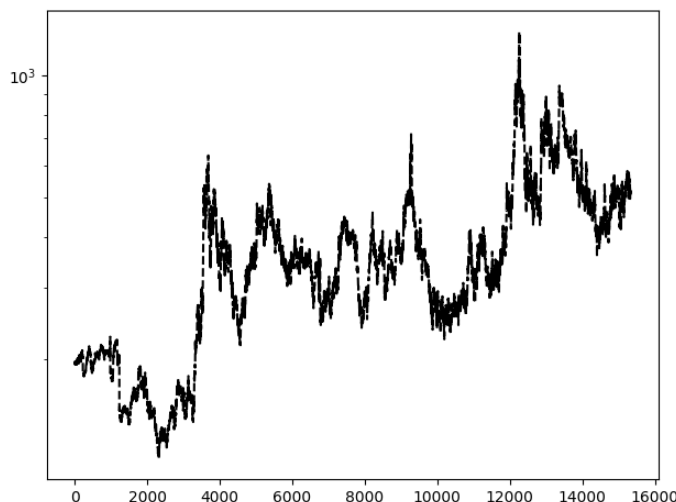
Taulukon 4.2 *tarkkuusmatriisi* ("Confusion matrix") esittää yllä kuvatun mallin tarkkuutta uusien datapisteiden luokitteluun. Matriisin diagonaalilta havaitaan datapisteiden määrä, jotka malli osasi luokitella oikein. Solussa (*Ennuste: 0, Oikea: 1*) on kuvattu määrä testidatan pisteitä, joille malli ennusti seuraavaksi laskupäivää, vaikka päätöshinta nousi edeltävään pörssipäivään suhteutettuna ja vastaavasti solussa (*Ennuste: 1, Oikea: 0*) on kuvattu pisteitä, joille malli ennusti seuraavaksi virheellisesti nousupäivää. Sekä tässä että jatkossa termillä "laskupäivä" tarkoitan pörssipäivää, jonka päätöshinta on enintään yhtä suuri kuin sitä edeltävän pörssipäivän päätöshinta.

Matriisista havaitaan, että jaettaessa *S&P* 500-indeksin hintadata em. tavoin suhteessa 1 : 1, päätyy testidataan nousupäiviä $\frac{2844+3306}{2659+2788+2844+3306} \approx 53.0\%$ koko testidatasta. Näin ollen naiivi luokittelija, joka luokittelisi kullekin testidatan pisteelle luokan 1 saavuttaisi korkeamman ennustetarkkuuden kuin edellä kuvattu paras mahdollinen malli. Malli ei tosin painota ennusteissaan luokkaa 1, sillä se luokittelee vain n. 52,5 % testidatan pisteistä tuohon luokkaan, mikä on vähemmän, kuin luokkaan kuuluvia pisteitä on testidatassa.

Satunnaista metsää hyödyntäen en myöskään onnistunut generoimaan mallia, joka olisi ollut kilpailukykyinen em. naiiviin luokittelijaan nähden. Menetelmällä saavutin parhaimmillaan 50.8 % ennustetarkkuuden tilanteessa, jossa luokittelu perustui suoraan päätöskurssien suhteellisiin etäisyyksiin 7. päivän liukuvista keskiarvoistaan eli

| | Ennuste: 0 | Ennuste: 1 |
|----------|------------|------------|
| Oikea: 0 | 1302 | 1390 |
| Oikea: 1 | 1461 | 1646 |

Taulukko 4.3: Tarkkuusmatriisi *S&P* 500-indeksin testidatan luokittelusta satunnaista metsää käyttäen



Kuva 4.3: Vehnäfutuurisopimusten hinnankelitys logaritmisella asteikolla vuosina 1959 - 2020

datajoukoksi valitsin vektorin ET_SMA_7 . Tämän ohella puiden lukumääräksi valitsin 100 kpl rajoittamatta niiden maksimisyvyksiä. Datan jaon opetus- ja testidataan tein suhteessa 3 : 1.

Taulukossa 4.3 on kuvattu edellä esitetyn mallin tarkkuutta uusien datapisteiden luokitteluun. Tämäkään malli ei painottanut kumpaakaan luokkaa erityisemmin ja 1-ennusteita testidatalle se teki vain $1390 + 1646 = 3036$ kpl, kun taas datapisteitä, joiden luokka on 1 oli joukossa $1461 + 1646 = 3107$ kpl. Tästä seuraa, että mallin yli 50 % ennustetarkkuus ei perustunut naiivisti frekvenssiltään todennäköisemmän luokan ennustamiseen. Sen sijaan parametrijoukolla, jossa puiden maksimisyvyys oli rajattu pieneksi, päästiin joitakin prosenttiyksikön kymmenyksiä korkeampiin ennustetarkkuuksiin. Tällaiset mallit kuitenkin luokittelivat pääsääntöisesti testidatan pisteitä luokkaan 1 ja eivät näin tuoneet lisäarvoa naiivin luokittelijan käyttöön nähden. Taustasyynä tässä saattoi olla se, että puiden maksimisyvyysien rajaaminen johti vähempään monimuotoisuuteen puiden välillä ja sääntöperusteet uusien datapisteiden luokitteluun saattoivat olla tämän myötä hyvinkin samankaltaiset.

Vehnän pörssipäivien päätöskurssien hinnankelitys käytössä olleen markkinadatan osalta on esitetty logaritmisella asteikolla kuvassa 4.3. Jos tätä verrataan kuvassa 4.1 esitettyyn *S&P* 500-indeksin pitkän aikavälin hinnankelitykseen, voidaan havaita,

| | <i>S&P</i> 500 | Vehnä |
|--------------------|--------------------|-------|
| $RSI_{14} < 30$ | 0.038 | 0.046 |
| $RSI_{14} < 50$ | 0.389 | 0.504 |
| $RSI_{14} \geq 50$ | 0.611 | 0.496 |
| $RSI_{14} > 70$ | 0.106 | 0.053 |

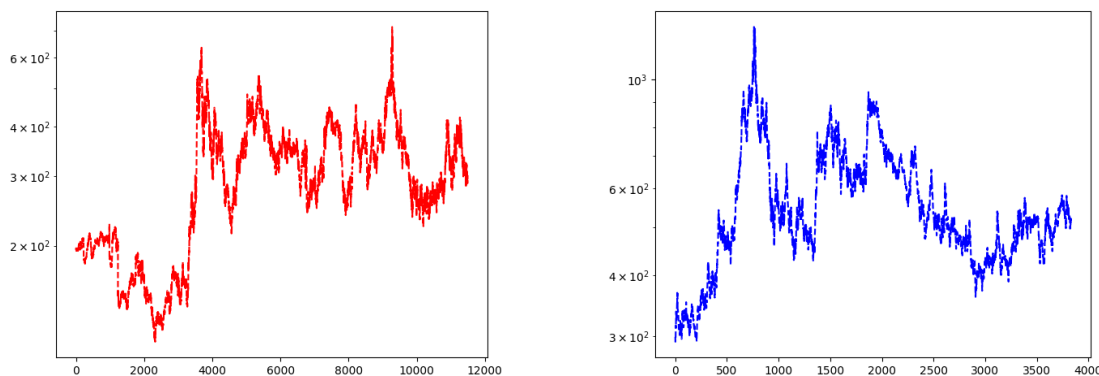
Taulukko 4.4: *S&P* 500-indeksin ja vehnäfutuurisopimusten markkinatrendien momentumvertailu painokertoimen 14 *RSI*:llä mitattuna

että *S&P* 500-indeksin arvo kasvaa pitkällä aikavälillä melko tasaisesti ja tulevaisuuden hinnankkehitystä saattaisi pystyä mielekkäällä tarkkuudella regressioanalyysillä ekstrapoloimaan. Vehnän osalta havaitaan taas huomattavasti suurempaa hinnanvaihtelua ja aaltomaista liikettä ylös ja alas, vaikka markkinahinta vaikuttaakin pitkällä aikavälillä nousevan. Näitä yksityiskohtia havainnollistaa myös taulukossa 4.4 esitetty datajoukkojen markkinatrendien välinen momentumvertailu.

Taulukon rivillä $RSI_{14} < 30$ on esitetty datapisteiden määrä, joille markkinatrendin momentum on vahvasti negatiivinen. Rivi $RSI_{14} \geq 50$ esittää datapisteitä, joille momentum on positiivinen ja $RSI_{14} > 70$ pisteitä, joille momentum on vahvasti positiivinen. Havaitaan, että *S&P* 500-indeksissä on vähemmän vahvasti negatiivisten ja negatiivisten momentumin pörssipäiviä kuin vehnädatassa. Samaan aikaan *S&P* 500-indeksissä on enemmän positiivisten ja vahvasti positiivisten momentumin pörssipäiviä kuin vehnässä. Markkinatrendien vahvuuserot ovat datajoukkojen välillä siis merkittäviä ja joukot ovat vahvuuserojenkin suhteen sopivan erilaisia tutkittaviksi. Erilaiset vahvuuserot voivat heijastua myös omalta osaltaan saamiini tuloksiin.

Vehnän parissa tutkin myös sekä kNN-luokittelijan, että satunnaisen metsän osalta erilaisia parametrikombinaatioita mahdollisimman hyvien mallien löytämiseen. Havaitsin, että käytössä olleen markkinadatan jako suhteessa 3 : 1 toimi ainakin lähes tulkoon parhaalla mahdollisella tavalla molemmille menetelmille. Lisäksi esikäsittelyn datan rivien ET_SMA_{20} ja RSI_5^{sc} sarakkeista koostuva datajoukko oli paras mahdollinen valinta kummankin menetelmän opetus- ja testidataksi. Käsittelenkin seuraavaksi nämä yhteneväisyydet huomioiden molempien menetelmien kautta saamiini tuloksia vehnän tulevan hinnamuodostuksen ennustamiseen.

Kuvissa 4.4a ja 4.4b on esitetty vehnän päätöskurssien hinnankkehitys datan aloituspäivästä alkaen menetelmissä käytetyllä opetus- ja testidatan jakosuhteella 3 : 1. Markkinadatan jako tällä tavoin opetus- ja testidataan vaikuttaa mielekkäältä, sillä molemmista kuvista on havaittavissa rajuja hinnan nousuja ja laskuja sekä samankaltaisia kuvioita kuten lyhytaikaisia piikkejä markkinahinnan käydessä merkittävästi lähiaikojen keskihintaansa korkeammalla. Tarkan koneoppimisen mallin luulisi siten



(a) Menetelmien opetusdataa vastaavat päätöskurssit, vehnän hinnankehitys vuosina 1959 - 2005

(b) Menetelmien testidataa vastaavat päätöskurssit, vehnän hinnankehitys vuosina 2005 - 2020

Kuva 4.4: Vehnäfutuurisopimusten markkinadatan jako suhteessa 3 : 1

| | Ennuste: 0 | Ennuste: 1 |
|----------|------------|------------|
| Oikea: 0 | 989 | 932 |
| Oikea: 1 | 888 | 962 |

Taulukko 4.5: Tarkkuusmatriisi vehnän testidatan luokittelusta kNN-algoritmia käyttäen

osaavan ennustaa seuraavaa päätöskurssin suuntaa testidatan pisteille opetusdataan peilaten.

Ajettaessa kNN-algoritmia yllä mainitulla vehnädatan jaolla päästään 51.7 % ennustetarkkuuteen valitsemalla $k = 1$ ilman mallin uudelleenopetusta testidatan pisteiden havaitsemisen yhteydessä. Malli on siis yhden lähimmän naapurin luokittelija ja sen tarkkuus on esitetty taulukon 4.5 tarkkuusmatriisissa. Matriisista voi laskea, että testidatassa on laskupäiviä 1921 ja nousupäiviä 1850 kpl. Naiivi yllä kuvattu luokittelija luokittelisi testidatan pisteet täten luokkaan 0 olettaen, että myös opetusdatassa on enemmän lasku- kuin nousupäiviä ja sen tarkkuus olisi $\frac{1921}{1921+1850} \approx 50.9$ %. KNN-algoritmilla opetettu malli on siis tarkempi kuin naiivi luokittelija. Se ei myöskään priorisoi frekvenssiltään todennäköisempää luokkaa 0, vaan luokittelee hieman suuremman joukon testidatan pisteitä luokkaan 1.

Satunnaista metsää käyttäen voidaan muodostaa kilpailukykyinen malli edellä kuvatulle kNN-algoritmilla opetetulle mallille. Jos metsän opetus- ja testidatat määritetään kuten yllä, parametreina valitaan puiden lukumääräksi 100 kpl ja ei rajoiteta niiden maksimisyvyyttä, saadaan keskimäärin malli, jonka ennustetarkkuus on 51.8 %. Erään tällaisen mallin tarkkuus on kuvattu taulukon 4.6 tarkkuusmatriisissa.

Matriisista voidaan havaita, että malli painottaa luokkaa 0 sopivasti luokkaan 1 nähden ja tekee 0-ennusteita vain 5 kpl enemmän kuin testidatassa on laskupäiviä. Samankaltaisia tuloksia sain myös toistaen algoritmin ajoa ko. parametreilla usei-

| | Ennuste: 0 | Ennuste: 1 |
|----------|------------|------------|
| Oikea: 0 | 1014 | 907 |
| Oikea: 1 | 912 | 938 |

Taulukko 4.6: Tarkkuusmatriisi vehnän testidatan luokittelusta satunnaista metsää käyttäen

siin otteisiin. Parhaimmillaan algoritmi onnistui opettamaan mallin, jonka ennuste-tarkkuus testidatalle oli 52.5 %. Tarkkuuserot eri ajoissa generoitujen mallien välillä selittyvät varsinkin eri puille opetusdataksi valituilla satunnaisotoksilla käytettävissä olleesta opetusdatasta. Satunnaisuutta muodostui myös siitä, että algoritmi määrittä puiden solmuissa tapahtuvat sääntöperusteiset haaraumat alemmalle tasolle tarkastelemalla opetettavan puun opetusdataa yksittäisen satuinnaisesti valitun ominaisuuden perusteella. Tämä perustui menetelmän vakioparametrin `max_features` asetukseen.

Ainoastaan yhtä tai kahta ominaisuutta tarkastelemalla saattoi satunnaista metsää käyttäen muodostua selvästi korrelaatiota opetettujen puiden välille [15]. Korrelaation minimoimiseksi olisikin saattanut olla parempi, jos olisin valinnut sekä *S&P* 500-indeksiä että vehnää tutkiessa menetelmän opetusdataan useampia ominaisuuksia. Näin tekemällä satunnaisen metsän generoimat mallit eivät kuitenkaan päässeet yhtä korkeisiin ennustetarkkuuksiin kuin yllä kuvatut.

Skaalattujen liukuvien keskiarvojen sekä RSI_5^{sc} ja RSI_{14}^{sc} arvojen ohella tutkin kynttilämuodostelmien toimivuutta seuraavan pörssipäivän päätöskurssin suunnan ennustamiseen. Havaitsin, että alaluvussa 4.1 mainitsemani tapa, jolla määritin datajoukkojen rivit toteuttamaan tiettyjä markkinatrendin potentiaalista kääntymistä signaloivia kynttilämuodostelmia, ei tuonut lisäarvoa tuloksiini käyttäessäni tätä ominaisuutta osana menetelmien opetusdataa. Pikemminkin ko. ominaisuuden hyödyntäminen muun datan ohessa johti hieman heikompiin malleihin niin *S&P* 500-indeksin kuin vehnänkin osalta. Seuraavaksi käsittelen jatkotutkimusmahdollisuuksia kynttilämuodostelmien hyödyntämiseen ja muihinkin implementaatiotietoihin liittyen.

4.3 Jatkotutkimusmahdollisuuksia

Vaikka varsinkin vehnän osalta yllä esittämäni tulokset olivat mielenkiintoisia, jäivät ne huomattavasti vaatimattomammiksi kuin vastaavanlaisen ongelman ratkaisuun perustuneet J. Patelin ja tutkimusryhmän tulokset [22]. Huomattavaa on, että J. Patel tutkimisryhmineen eivät määritelleet lyhyen aikavälin pituutta, jonka päähän he mittasivat ennustetarkkuutta eri koneoppimisen menetelmin, mihin suuntaan markkinahinta seuraavaksi liikkuu. Onkin mahdollista, että jos olisin lähtenyt tarkastelemaan päivittäiseen hintadataan perustuen, voidaanko esimerkiksi tulevilla viiden päivän mittaisel-

la aikavälillä ennustaa, onko markkinahinta korkeammalla vai matalammalla aikavälin päätteeksi kuin sitä ennen, olisivat käyttämäni menetelmät altistuneet vähemmän kohinalle ja tuloksista olisi saattanut saada hieman parempia.

Mille tahansa ajanhetkelle voidaan nimittäin muodostaa *konteksti* ajanhetkellä voimassaolevaan pääsääntöiseen markkinatrendiin ja trendin vahvuuteen perustuen. Trendin vahvuuteen vaikuttavat ajanhetkellä mm. markkinahinnan sijainti selkeimpiin tuki- ja vastustasoihinsa nähden ja tarkastellulla ajanjaksolla lähimenneisyudessa havaittavat kynttilämuodostelmat sekä momentum-indikaattoreilla mitattavissa oleva vahvuus. Mikäli tällaiseen kontekstiin liittyen on tavoite ennustaa, onko ajanjakson seuraavan kynttilän päätöshinta korkeammalla vai matalammalla kuin edeltäjensä, oletan, että havaitaan kohinaa, joka ei ole merkityksellistä hieman pidemmän aikavälin ennusteen suhteen. Esimerkiksi kuvassa 2.6a sekä kolmas että neljäs kynttilä vasemmalta katsottuna signaloivat molemmat potentiaalisesta trendin kääntymisestä nousevaksi ja jatkossa korkeammista hinnoista. Molemmissa tapauksissa yhden ajanjakson oikea ennuste on päätöshinnan lasku, vaikka useamman ajanjakson oikea ennuste onkin päinvastainen. Vastaavasti kuvassa 2.5b toinen pyrstötähti vasemmalta katsottuna yhdistettynä kahden edeltävän kynttilänsä signaloimaan heikkouteen antaa potentiaalisen signaalin trendin kääntymisestä, mutta tuota seuraavan kynttilän päätöshinta on kuitenkin hieman tätä korkeampi ennen kuin markkinahinta lähtee laskuun. Jos taas markkina on vahvassa nousu- tai laskutrendissä ei ole hieman pidemmän aikavälin ennusteen suhteen merkityksellistä, jos markkinahinnassa tapahtuu muutaman kynttilän mittainen liike trendin vastaiseen suuntaan, kuten mm. kuvasta 2.3 ”W-muotoisen” kuvion jälkeen havaitaan.

Kohinalle altistumisen ohessa tapa, miten markkinadataa käsittelin, loi uskoakseni liian vajavaisen kontekstin merkittävien ennustetarkkuuksien saavuttamiseen. Esimerkkinä tapa, jolla skaalasin datajoukkojen rivien päätöshintojen etäisyydet liukuviin keskiarvoihinsa, johtaa siihen, että rivit, joiden päätöshinnat ovat samojen liukuvien keskiarvojen läheisyydessä ovat niiden ominaisuuksien suhteen hyvinkin samankaltaisia, vaikka ne esiintyisivät päinvastaisissa markkinatrendeissä. Näin ollen uudelle havainnolle, jonka päätöshinta on menetelmän parametriksi valitun liukuvan keskiarvon läheisyydessä, saadaan tuskin RSI_5^{sc} ja RSI_{14}^{sc} arvojen perusteella tarpeeksi informaatiota, jotta markkinahinnan tulevaa nousua tai laskua voitaisiin merkittäväällä tarkkuudella ennustaa. Yllä esittämissäni tuloksissanihan vain kNN-luokittelijan paras opettamani malli datanaan *S&P* 500-indeksi sai parametreina etäisyyksiä useammasta kuin yhdestä liukuvasta keskiarvosta. Jos parametreiksi sisällyttäisi myös mm. kynttilämuodostelmat, saataisiin selkeämpi signaali siitä, miten hinnanmuodostus tapahtuu liukuvan keskiarvon ja sen määrittämän trendin suhteen. Tällä tarkoitan sitä, että tarkasteltaisiin, havaitaanko liukuvan keskiarvon hintatasolla selkeitä nousu- tai laskutrendiin

siirtymisestä tai sen jatkumisesta signaloivia kynttilämuodostelmia. Näiden eri suuntaa signaloivien muodostelmien kautta rivit, jotka ovat etäisyydeltään samoista liukuvista keskiarvoista lähellä toisiaan, mutta joiden kynttilämuodostelmat ovat hyvin erilaisia, voitaisiin tämän ominaisuuden suhteen erottaa ja menetelmien ennustetarkkuudet saattaisivat sitä kautta parantua.

Alaluvun 4.2 lopussa totesin, että en saanut kynttilämuodostelmien hyödyntämisestä lisäinformaatiota, vaan ne pikemminkin heikensivät tuloksia menetelmien syöteinä muiden ominaisuuksien ohessa. Tavassa, jolla määritin datajoukkojen rivien toteuttavan jonkin alaluvussa 4.1 mainituista muodostelmista en näe ongelmaa, mutta itse muodostelmien kuvaustapani toimivuudesta välille $[-1, 1]$ en ole vakuuttunut. Kuvaustavan mielekkyyden validointi olisi yksi mahdollinen jatkokehityskohde. Toisaalta yksittäisten rivien tarkastelu ei ollut välttämättä muutenkaan riittävä ratkaisu siihen, että niille olisi saatu talletettua informaatiota trendin potentiaalisesta heikkoudesta siltä osin, että se olisi tuonut lisäinformaatiota tuloksiin. Yhtenä jatkotutkimusmahdollisuutena voitaisiin lähteä tarkastelemaan sitä, että kynttilämuodostelmatiedon muodostaisi jokaiselle riville sen ja useamman sitä edeltäneen rivin kokonaisuutena ja kokonaisuudet kuvattaisiin menetelmien kannalta käyttökelpoiseen muotoon.

Olin alun perin suunnitellut hyödyntäväni myös *MACD*:tä *RSI*-indikaattorien ohessa markkinatrendin vahvuuden mittaamiseen. Tutkimus olisi tällöin paisunut tosin varsin suureksi tutkielman laajuuteen suhteutettuna, ja koska muitakin kehityskohteita oli olemassa, jätin *MACD*:n soveltamisen jatkotutkimusmahdollisuuksien piiriin.

Yhteenvetona todettakoon, että muutamien teknisen analyysin indikaattorien kautta määritettyjen ominaisuuksien hyödyntäminen ei vaikuta koneoppimisen menetelmien parametreina riittävältä, jotta menetelmillä voitaisiin opettaa malleja, jotka osaisivat merkittäväällä tarkkuudella ennustaa markkinan tulevaa hinnanmuodostusta. Muutamilla ominaisuuksilla opetettu yksinkertainen mallikin vaikuttaa tosin toimivan jossain määrin ainakin tietyillä markkinoilla, kun taas mm. laajamittaisen kaupankäynnin kohteena olevan osakeindeksin tulevan hinnanmuodostuksen ennustaminen voi olla sellaisella lähes mahdotonta. Hyödynnettäessä tarpeeksi suurta määrää erilaisia indikaattoreita sekä tiimin asiantuntemusta markkinoihin ja koneoppimiseen, tämä vaikuttaisi aiempaan tutkimukseen nojaten kuitenkin onnistuvan [22].

4.4 Yhteenveto

Tässä luvussa käsittelin omaa tutkimustani markkinan tulevan hinnanmuodostuksen ennustamiseen koneoppimista ja teknistä analyysiä hyödyntäen. Tavoitteeni oli markkinan historiallisen hintadatan perusteella selvittää, voidaanko merkittäväällä tarkkuudella ennustaa, mihin suuntaan markkinahinta siirtyy valitulla ajanhetkellä seuraavaksi.

Tutkimusdataksi valitsin kahden eri markkinan päivittäistä hintadataa ja koneoppimisen menetelmiksi kNN-luokittelijan ja satunnaisen metsän.

Havaitsin, että toteuttamani tapa määrittää tutkimusdatajoukkojen riveille informaatiota markkinatrendistä ja momentumista ei vaikuttanut riittävältä ongelman ratkaisemiseen. Toiselle tarkastelluista datajoukoista onnistuin silti generoimaan malleja, jotka saavuttivat korkeamman ennustetarkkuuden kuin naiivi luokittelija, joka ennusti seuraavan markkinahinnan suunnan olevan aina sen, joka oli ollut historiallisesti todennäköisempi. Käsittelin myös jatkotutkimusmahdollisuuksia ongelman parissa. Pohdin kohinan vaikutusta generoimieni mallien ennustetarkkuuksiin ja esitin ideoita, joita olisi mahdollista kokeilla, että esiprosessoitu data soveltuisi paremmin ongelmanratkaisuun ja menetelmillä generoiduista malleista tulisi tehokkaampia.

5. Johtopäätökset

Markkinoilla kaupankäynnin kohteina olevien arvopaperien hinnoittelu perustuu pääosin kysynnän ja tarjonnan suhteeseen. Kysynnän kasvaessa suhteessa tarjontaan hintoja nostetaan ja hintoja vastaavasti lasketaan, jos tarjonta kasvaa kysyntään nähden. Empiirisesti on havaittu, että kysyntä ja tarjonta eivät aina kohtaa oikealla markkinahintatasolla ja monet tekniset ajurit vaikuttavat näiden suhteeseen. Koska kysynnän ja tarjonnan suhde muuttuu jatkuvasti, voi sijoittaja tai aktiivinen kaupankävijä pyrkiä ostamaan arvopaperia silloin, kun kysynnän määrä suhteessa tarjontaan on sille vähäinen ja halutessaan myymään, kun kysynnän määrä on merkittävästi kasvanut.

Teknisessä analyysissä pyritään tunnistamaan teknisiä ajureita, joiden voidaan olettaa jatkossa vaikuttavan tarkasteltavan markkinan kysynnän ja tarjonnan suhteeseen. Tekninen analyysi perustuu erilaisten indikaattorien hyödyntämiseen, joilla tarkastellaan markkinan historiallista hintadataa. Indikaattorien avulla on tavoite ennustaa, miten kysynnän ja tarjonnan suhde todennäköisimmin tulevaisuudessa muuttuu.

Tutkimuskysymyksenäni lähdin selvittämään, voidaanko lyhyellä aikavälillä ennustaa, käydäänkö tarkasteltavalla markkinalla kauppaa ajanjakson päätteeksi korkeampaan vai enintään yhtä korkeaan hintaan kuin sen alussa. Lähestyin ongelmaa koneoppimista ja teknistä analyysiä hyödyntäen tarkastelemalla kahden eri markkinan, *S&P 500*-indeksin ja vehnäfutuurisopimusten päivittäistä hintadataa. Havaitsin, että käyttämilläni ohjatun oppimisen menetelmillä ja teknisen analyysin indikaattoreilla ei merkittäväällä tarkkuudella pystynyt ennustamaan markkinoiden tulevaa hinnanmuodostusta. Menetelmillä opettamani mallit toimivat jossain määrin vehnäfutuurisopimusdatalle, mutta laajamittaisen kaupankäynnin kohteena olevan *S&P 500*-osakeindeksin tulevan hinnanmuodostuksen ennustaminen historiallisen hintadatansa perusteella ei niillä onnistunut.

Esitin myös aiempaa tutkimusta ongelmanratkaisuun ohjattua oppimista ja teknistä analyysiä hyödyntäen. Aiemman tutkimuksen tuloksiin nojaten vaikuttaa siltä, että myös osakeindeksien hinnanmuodostuksen ennustaminen on merkittäväällä tarkkuudella hyvin lyhyelläkin aikavälillä mahdollista. Tarkastelin myös jatkotutkimusmahdollisuuksia omaan tutkimukseeni liittyen. Arvioin, että jos tarkasteltua lyhyttä aikaväliä pidennettäisiin hieman, altistuisivat menetelmillä generoimani mallit kohinalle vähem-

män ja niistä saattaisi tulla tehokkaampia. Mallien tarkkuuksia voisi pyrkiä kasvattamaan myös datajoukkojen esiprosessointitapoja jatkokehittäen.

Kirjallisuutta

- [1] https://fi.wikipedia.org/wiki/Kysyntä_ja_tarjonta.
- [2] <https://www.tradingview.com/symbols/OMXHEX-SAMPO/>.
- [3] <https://www.investopedia.com/terms/d/doublebottom.asp>.
- [4] <https://www.investopedia.com/terms/e/ema.asp>.
- [5] <https://www.tradingview.com/symbols/CURRENCYCOM-US500/>.
- [6] <https://www.investopedia.com/terms/r/rsi.asp>.
- [7] <https://www.investopedia.com/terms/m/macd.asp>.
- [8] <https://www.investopedia.com/trading/continuation-patterns-introduction/>.
- [9] <https://finance.yahoo.com/quote/%5EGSPC/history>.
- [10] https://www.quandl.com/data/CHRIS/CME_W1-Wheat-Futures-Continuous-Contract-1-W1-Front-Month.
- [11] <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>.
- [12] <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>.
- [13] L. Breiman. Bagging predictors. *Mach. Learn.*, 24(2):123–140, Aug. 1996.
- [14] L. Breiman. Heuristics of instability and stabilization in model selection. *Annals of Statistics*, 24:2350–2383, 1996.
- [15] L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, Oct. 2001.
- [16] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

- [17] S. Gunn. Support vector machines for classification and regression. Project report, 1998. Address: Southampton, U.K.
- [18] J. Larsen. Predicting stock prices using technical analysis and machine learning. 11 2010.
- [19] D. D. Lewis. Naive (bayes) at forty: The independence assumption in information retrieval. In C. Nédellec and C. Rouveirol, editors, *Machine Learning: ECML-98*, pages 4–15, Berlin, Heidelberg, 1998. Springer Berlin Heidelberg.
- [20] R. Lippmann. An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4(2):4–22, 1987.
- [21] S. Nison. Japanese candlestick charting techniques : a contemporary guide to the ancient investment techniques of the far east. 1991.
- [22] J. Patel, S. Shah, P. Thakkar, and K. Kotecha. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1):259 – 268, 2015.
- [23] L. A. Teixeira and A. L. I. de Oliveira. A method for automatic stock trading combining technical analysis and nearest neighbor classification. *Expert Systems with Applications*, 37(10):6885 – 6890, 2010.