

Janne Kleemola

**Koneoppiminen sijoitussalkun optimoinnissa ja sen
ongelmat**

Tietotekniikan kandidaatintutkielma

25. toukokuuta 2022

Jyväskylän yliopisto

Informaatioteknologian tiedekunta

Tekijä: Janne Kleemola

Yhteystiedot: `janne.i.kleemola@student.jyu.fi`

Ohjaaja: Tytti Saksa

Työn nimi: Koneoppiminen sijoitussalkun optimoinnissa ja sen ongelmat

Title in English: Machine learning in Portfolio Optimization and the problems with it

Työ: Kandidaatintutkielma

Sivumäärä: 18+0

Tiivistelmä: Tämän tutkielman tarkoituksena on selvittää sijoitussalkun optimointia koneoppimisen näkökulmasta, sekä siihen liittyviä ongelmia.

Avainsanat: Koneoppiminen, sijoitussalkun optimointi, aikasarja, hahmontunnistus, ulottuvuuksien vähentäminen

Abstract: The purpose of this study is to find out about portfolio optimization from the perspective of machine learning, as well as the related problems.

Keywords: Machine learning, portfolio optimization, time series, pattern recognition, dimensionality reduction

Kuviot

Kuvio 1. Mahdollinen skenaario 'osta halvalla, myy kalliilla' -menetelmässä (Iba ja Aranha 2012).	2
Kuvio 2. Yksinkertainen esimerkki omistuksien hajauttamisesta (Iba ja Aranha 2012). ...	3
Kuvio 3. Aikasarja Helsingin pörssin indeksoidusta kehityksestä (Nyberg ja Vaihekoski 2010).	4
Kuvio 4. Koneoppimisen monitieteisyys (Alzubi, Nayyar ja Kumar 2018).	6
Kuvio 5. Havainnollisesti tärkeät kohdat (Fu ym. 2008).	8
Kuvio 6. Esimerkki kvantitatiivisen sijoittamisen toteutuksesta (Ta, Liu ja Tadesse 2020).	9

Sisällys

1	JOHDANTO	1
2	FINANSSIMARKKINAT	2
	2.1 Sijoitussalkun optimointi (Portfolio optimization)	2
	2.2 Aikasarja (engl. Time series)	3
	2.3 Tekninen analyysi ja fundamenttianalyysi	4
3	KONEOPPIMINEN SIJOITUSSALKUN OPTIMOINNISSA	6
	3.1 Hahmontunnistus (engl. Pattern recognition)	7
	3.2 Ulottuvuuksien vähentäminen (engl. Dimensionality reduction)	7
	3.3 Riskien ja tuoton monimutkainen suhde	8
4	KONEOPPIMISEN ONGELMAT SIJOITUSSALKUN OPTIMOINNISSA	9
	4.1 Markkinoiden arvaamattomuus	9
	4.2 Opetusdatan ongelmat.....	10
5	YHTEENVETO.....	11
	LÄHTEET	12

1 Johdanto

Tekoälyn käyttö finanssieteessä voidaan yleisesti jakaa kolmeen osa-alueeseen. Nämä osa-alueet ovat sijoitussalkun optimointi, tulevien hintojen ja trendien ennustaminen, sekä yrityksiä ja resursseja koskevien uutisten ja reaktioiden tunneanalyysi (engl. sentiment analysis) (Ferreira, Gandomi ja Cardoso 2021). Suuri osa nykypäivän arvopaperikaupasta käydään bottien avulla, jotka perustuvat koneoppimisen luomiin laskelmiin. Emerson ym. (2019) avaavat konferenssijulkaisussaan tarkemmin koneoppimisen suosion kasvua finanssimarkkinoilla seuraavalla tavalla: ”Kvantitatiivinen lähestymistapa markkina-analyysiin saavutti suosiota tietotekniikan kehittyessä ja sen mahdollistaessa suurten datamäärien keräämisen sekä analysoinnin. Tämä mahdollisti markkinamallien kehittämisen ja todentamisen mittakaavassa, joka ei aikaisemmin ollut mahdollista.”

Tässä tutkielmassa sijoitussalkun optimointia tutkitaan koneoppimisen näkökulmasta, joka on tekoälyn osa-alue. Sijoitussalkun optimoinnilla tarkoitetaan parhaan omaisuuseräjakouman määrittämistä tiedossa olevien prioriteettien mukaisesti. Tutkielman tavoitteena on avata sekä koneoppimisen luomia mahdollisuuksia sijoitussalkun optimoinnissa, että sen ongelmakohtia. Tutkielma suoritetaan kirjallisuuskatsauksena, ja sen lähteet ovat tutkimuslaitosten tuottamia julkaisuja, tieteellisiä kirjoja, konferenssijulkaisuja sekä alan vertaisarvioituja aikakauslehtiartikkeleita.

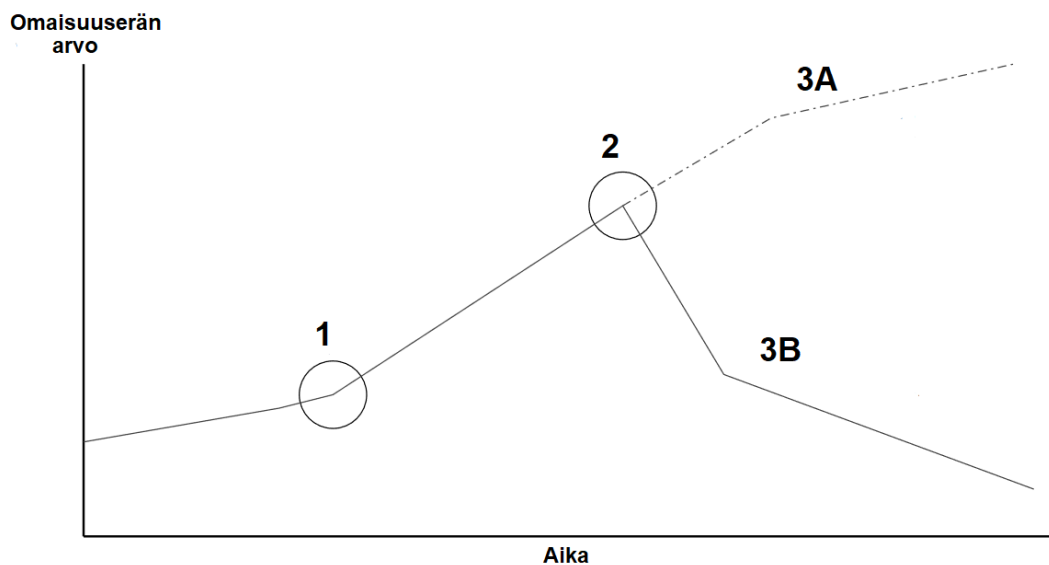
Tutkielma sisältää kokonaisuudessaan viisi lukua. Johdannon tavoitteena on avata nopeasti tutkielman tarkoitusperiä ja rajata sen aihetta selkeällä tavalla. Luvussa 2 tuodaan ilmi finanssimarkkinoihin liittyviä termejä ja käsitteitä, jotka ovat tutkielman ymmärtämisen kannalta tärkeitä. Luku 3 käsittelee koneoppimisen avaamia mahdollisuuksia sijoitussalkun optimoinnissa ja luku 4 puolestaan siihen liittyviä ongelmia. Viimeinen luku 5 vetää tutkielman yhteen ja pohtii mahdollisia jatkokysymyksiä tuleville tutkimuksille.

2 Finanssimarkkinat

Tutkiakseen ongelmia, joita sijoitussalkun optimointiin koneoppimisen avulla liittyy, on tunnettava heti alkuun muutama tärkeä finanssimarkkinoihin liittyvä keskeinen käsite. Jos käsitteeseen ei löydy vastinetta suoraan suomen kielestä, se avataan tarkemmalla kuvauksella.

2.1 Sijoitussalkun optimointi (Portfolio optimization)

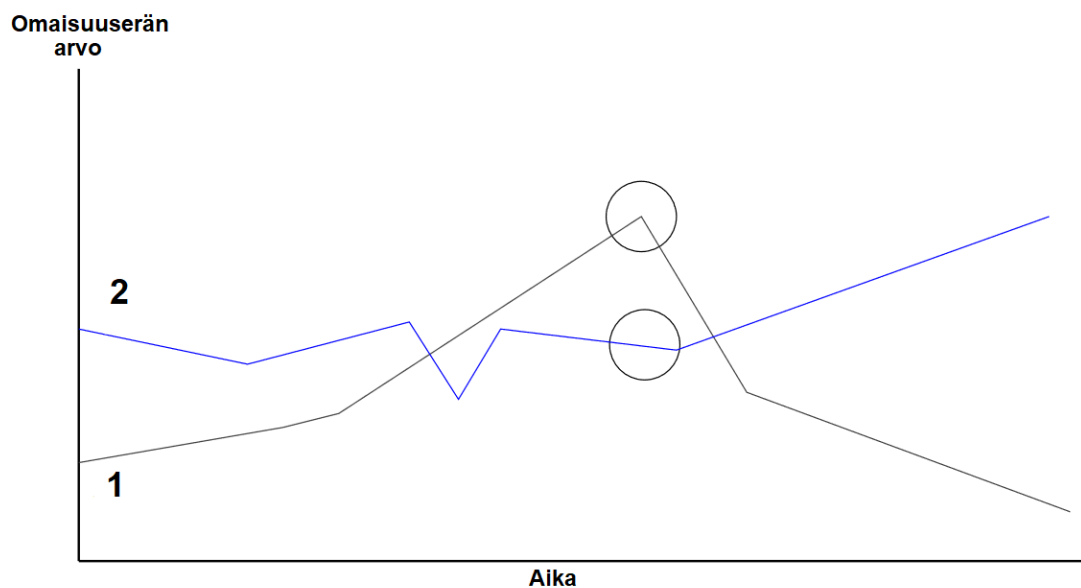
Sijoitussalkun optimointi tarkoittaa yksinkertaisuudessaan parhaan omaisuuseräjakauman löytämistä tiettyjen tavoitteiden mukaisesti. Näitä tavoitteita voivat olla muun muassa yksittäisen sijoittajan henkilökohtainen suhtautuminen ympäristöarvoihin sekä sijoituksen riskeihin.



Kuvio 1. Mahdollinen skenaario 'osta halvalla, myy kalliilla' -menetelmässä (Iba ja Aranha 2012).

Otetaan esimerkiksi tilanne, jossa lentokoneita valmistava yhtiö aloittaa uuden tehtaan rakentamisen ja lupaa tehtaan valmistuttua tehostaa huomattavasti tuotantoansa. Tilanteessa 1 (katso kuvio 1) sijoittaja reagoi positiiviseen uutiseen ostamalla yhtiön osakkeita ja olettaa kasvun jatkuvan tasaisena. Globaali pandemia kuitenkin keskeyttää lentoliikenteen kuvion

1 tilanteessa 2 ja vaikuttaa ratkaisevalla tavalla lentokoneiden kysyntään, mikä osaltaan romahduttaa yhtiön arvostuksen. Tältä varautuakseen sijoittaja hajauttaa omistuksiaan myös aloihin, jotka hyötyvät matkailualan ahdingosta (Katso Kuvio 2). Hajauttaminen tarkoittaa finanssimarkkinoiden käsitteenä käytäntöä, jossa sijoitukset jaetaan eri toimialoille niin, että yksittäisen omaisuuserän painoarvo ei nouse liian suureksi (Iba ja Aranha 2012).



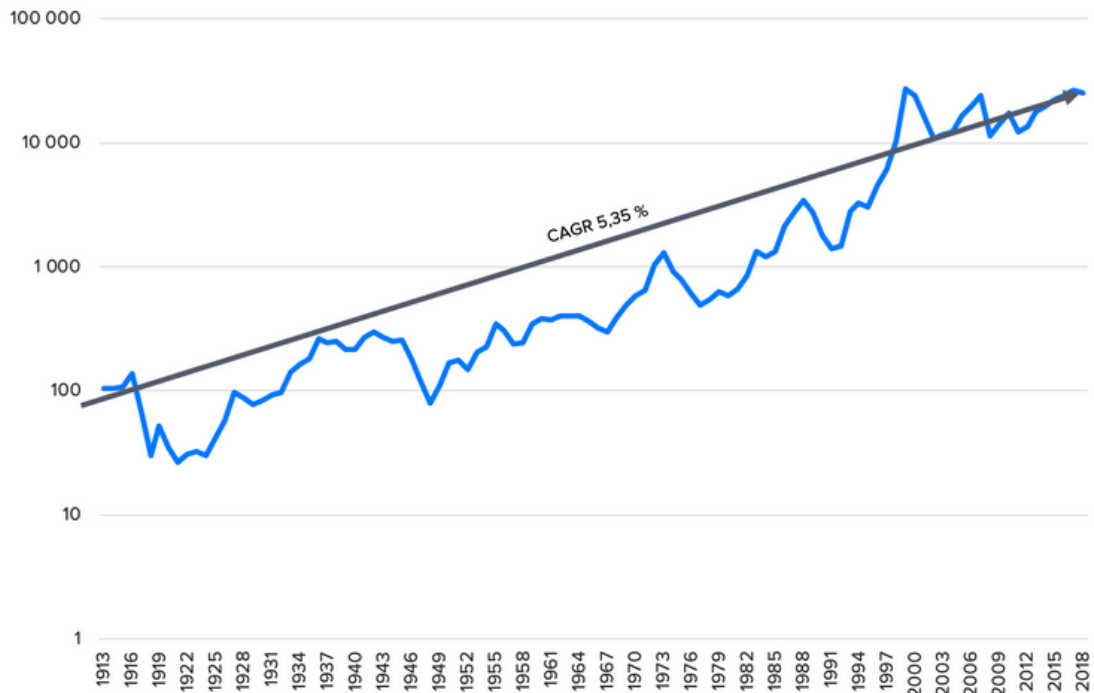
Kuvio 2. Yksinkertainen esimerkki omistuksien hajauttamisesta (Iba ja Aranha 2012).

Näitä useille aloille ulottuvia sekä useiden omaisuuseräjakaumien kokonaisuuksia kutsutaan portfolioiksi eli sijoitussalkuiksi. Sijoitussalkun optimoinnissa pyritään minimoimaan näiden portfolioiden riskejä mahdollisimman tehokkaasti sijoituskohteiden valinnalla sekä valittujen sijoituskohteiden painoarvoilla (Zhou ym. 2021).

2.2 Aikasarja (engl. Time series)

Aikasarjalla tarkoitetaan kokoelmaa peräkkäisistä havainnoista jonkin tietyn ajan kuluessa. Näiden havaintojen ei kuitenkaan tarvitse olla tehty säännöllisin väliajoin, vaan myös epäsäännöllisesti kerätty data voi muodostaa aikasarjan. (Palma 2016). Aikasarjojen avulla voidaan esimerkiksi tutkia helposti, miten erilaiset osatekijät vaikuttavat suurempiin kokonaisuuksiin eri aikajaksoilla. Kuvion 3 aikasarjasta Helsingin pörssin indeksoidusta kehityksestä voidaan huomata miten pörssi on reagoanut itsenäisen Suomen historian aikana erilaisiin

kriiseihin sekä talouden nousukausiin.



Kuvio 3. Aikasarja Helsingin pörssin indeksoidusta kehityksestä (Nyberg ja Vaihekoski 2010).

Finanssimarkkinoihin liittyvä data muodostaa luonnostaan useita eri aikasarjoja esimerkiksi indeksien hinnanvaihteluiden sekä kaupankäyntivolyymien ansiosta. Aikasarjojen avulla on myös mahdollista tutkia miten eri raaka-aineet, arvopaperit tai jopa taloudelliset muuttujat kuten köyhyys ovat vaihdelleet vuosien varrella. Aikasarjojen selvä vahvuus on siis erilaisten trendien tunnistaminen datasta.

2.3 Tekninen analyysi ja fundamenttianalyysi

Kun puhutaan finanssimarkkinoiden analysoinnista, kaksi koulukuntaa erottuu selvästi joukosta (Bonga 2015). Vaikka tekninen analyysi ja fundamenttianalyysi suoritetaan täysin eri tavoilla, niiden päämäärä on sama. Tarkoituksena on tulkita osakkeiden sekä erilaisten hyödykkeiden nykyisen arvon oikeellisuutta sekä luoda valistuneita arvauksia markkinoiden tulevaisuuden liikkeistä.

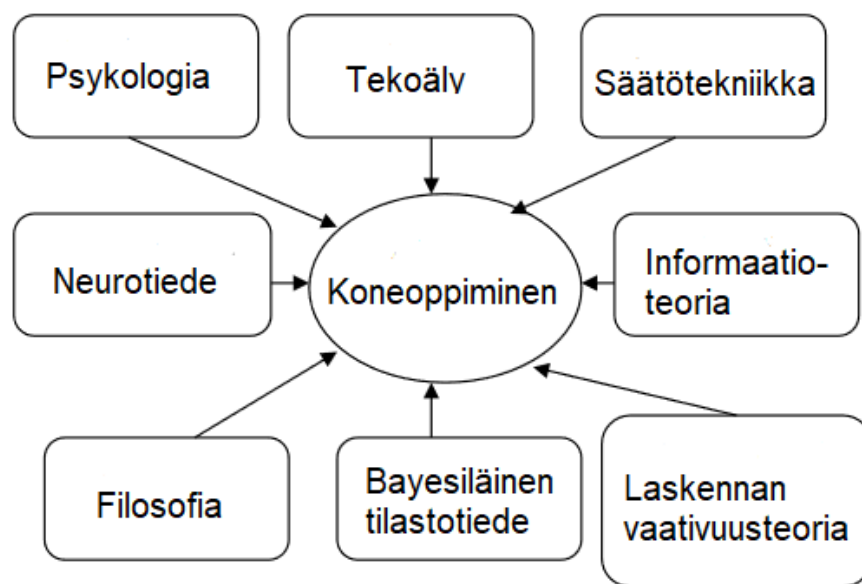
Fundamenttianalyysissä arvopaperin todellinen arvo yritetään määrittää tutkimalla siihen liittyviä taloudellisia ja rahoituksellisia tekijöitä (Abarbanell ja Bushee 1997). Näitä tekijöitä ovat esimerkiksi yrityksen tilinpäätökset, voittomarginaalit sekä tutkittavaan arvopaperiin liittyvät rahoituspoliittisten ratkaisut. Fundamenttianalyysin perimmäisenä tarkoituksena on siis selvittää, onko arvopaperi ali- vai yliarvostettu erilaisten tunnuslukujen avulla ja tuoda näkökulma osto- sekä myyntihetken mahdollisimman tehokkaaseen ajoittamiseen.

Tekninen analyysi ei fundamenttianalyysin tavoin perustu yritysten tunnuslukuihin ja rahoituspoliittisiin ratkaisuihin, vaan sen avulla pyritään tulkitsemaan finanssimarkkinoihin liittyvää dataa ja löytämään siitä erilaisia kaavioita, malleja sekä trendejä, mitkä osaltaan auttavat oikeellisten sijoituspäätösten tekemisessä. Esimerkkejä edellä mainituista kaavioista ja malleista ovat pörssikurssikaavioissa ilmenevät sonni- ja karhulippukuviot. Lipuilla tarkoitetaan volatiliteetin ja arvopaperin tai hyödykkeen hinnanvaihtelun pörssikurssikaavioon muodostamaa kuviota, jolla voidaan ennustaa tulevaa hintakehitystä. Chenin (2010) mukaan teknisen analyysin suosio on kasvanut viime vuosikymmeninä huomattavasti, koska se tarjoaa konkreettisen, loogisen ja tehokkaan lähestymistavan suurten rahoitusmarkkinoiden hallintaan.

Kvantitatiivinen sijoittaminen tarkoittaa sijoitusstrategiaa, jossa pyritään kvantitatiivisen analyysin avulla löytämään finanssimarkkinoiden säännönmukaisuuksia ja trendejä (Li, Wu ja Bu 2016). Kvantitatiivisen analyysin ja teknisen analyysin väliltä voidaan löytää yhtäläisyyksiä, mutta ne eivät kuitenkaan tarkoita samaa asiaa. Kvantitatiivinen sijoittaminen tukeutuu suuriin tietokantoihin sekä kompleksisiin matemaattisiin malleihin etsiessään optimaalisia vastauksia erilaisiin finanssimarkkinoilla esiintyviin tilanteisiin. Kvantitatiivisten sijoitusstrategioiden luonne sopii tästä syystä erityisen hyvin yhteen koneoppimisen algoritmien luomien mahdollisuuksien kanssa.

3 Koneoppiminen sijoitussalkun optimoinnissa

Koneoppimisessa kone pystyy itse muokkaamaan vastausmallejaan sille annetun datan perusteella. Kone pyrkii löytämään mahdollisimman optimaalisen vastauksen erilaisiin tilanteisiin ja oppii jatkuvasti uusia kaavoja sekä malleja sen saavuttamiseksi (Alpaydin 2014). Koneoppiminen on monitieteinen ala ja sen laajalle ulottuvat kuviossa 4 esitettävät tutkimusalat vahvistavat sen olemassaoloa.



Kuvio 4. Koneoppimisen monitieteisyys (Alzubi, Nayyar ja Kumar 2018).

Perinteisesti sijoitussalkun optimointi on perustunut Harry Markowitzin kehittämään moderniin portfolioteoriaan. Markowitzin teorian mukaan sijoittajat suosivat portfolioita, jotka tarjoavat mahdollisimman korkean tuotto-odotuksen siihen kohdistuviin riskeihin nähden (Jorion 1992). Teoria olettaa myös, että sijoittajat tekevät aina järkipäätöksiä, kun heillä on saatavilla kaikki näihin päätöksiin tarvittava informaatio. Tietokoneiden prosessointikyvyn jatkuvan kasvun sekä koneoppimisen kehitysaskleet ovat kuitenkin mahdollistaneet aivan uuden tavan käsitellä finanssimarkkinoihin liittyvää dataa. Tämän luvun tarkoituksena on avata näitä koneoppimisen luomia mahdollisuuksia sijoitussalkun optimoinnissa.

3.1 Hahmontunnistus (engl. Pattern recognition)

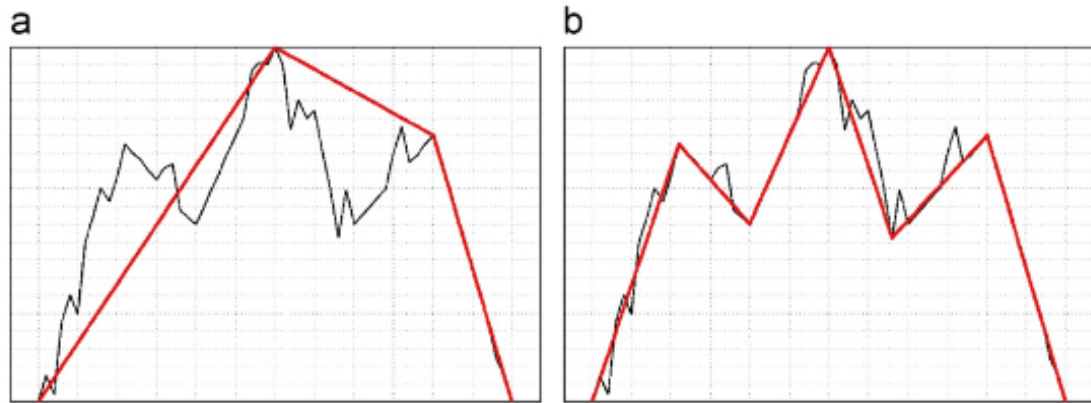
Hahmontunnistuksella tarkoitetaan koneoppimisen algoritmien käyttöä erilaisten kaavojen sekä mallien tunnistamiseen. Hahmontunnistuksen tavoitteena on luokitella käsittelyssä oleva malli tai kaava tunnetun datan perusteella omaan luokkaansa (Beyerer, Richter ja Nagel 2017). Käytettävissä olevan datan laajuus sekä laatu vaikuttavat suoraan hahmontunnistuksen onnistumiseen. Finanssimarkkinoista vuosikymmenien aikana kerätty mittava sekä korkeatasoinen data antavat hahmontunnistukselle hyvät lähtökohdat sijoitussalkun optimointiin liittyvien ongelmien ratkaisemiseen.

Finanssimarkkinoista aikasarjoihin kerätty data ja siinä ilmenevät kaavat sekä mallit voivat auttaa sekä omaisuuserien tehokkaassa valinnassa, että niiden painoarvojen säätämisessä. Hahmontunnistus voi myös käyttää useita eri keinoja kaavojen havaitsemiseen. Yksi osoitus näistä keinoista on painotetun ristikon (engl. weighted grid) hyödyntäminen. Painotetussa ristikossa jokainen solu saa oman arvonsa ja näiden solujen avulla hahmontunnistus pyrkii löytämään yhteensopivan kuvion sille annetusta tietoaaineistosta (Martins ja Neves 2020).

3.2 Ulottuvuuksien vähentäminen (engl. Dimensionality reduction)

Ulottuvuuksien vähentäminen koneoppimisessa tarkoittaa tekniikkaa, jolla karsitaan tietoaaineistojen monimutkaisuuksia vähentämällä niissä olevien tulosuureiden määrää. Käytännössä ulottuvuuksien vähentäminen auttaa huomattavasti koneoppimisen algoritmien käyttöönotossa, sekä vähentää niiden taakkaa myös tämän jälkeen (Reddy ym. 2020). Tämä on tärkeää, sillä se auttaa hallitsemaan koneoppimisen tarvitsemaa prosessointikykyä ja antaa sille elintilan mahdollisimman tehokkaaseen työskentelyyn.

Ulottuvuuksien vähentämisen suuri haaste on tunnistaa tietoaaineiston havainnollisesti tärkeät kohdat. Kuvion 5 ensimmäisessä kuvassa havainnollisesti tärkeitä pisteitä ei olla rekisteröity tarpeeksi ja se on muuttanut tietoaaineiston dataa merkitsevällä tavalla. Jos tilannetta vertaa kuvion 5 toiseen kuvaan, jossa havainnollisesti tärkeitä kohtia on tarpeeksi, voi tästä huomata miksi kyseessä on niin tärkeä asia ulottuvuuksien vähentämisestä puhuttaessa.



Kuvio 5. Havainnollisesti tärkeät kohdat (Fu ym. 2008).

3.3 Riskien ja tuoton monimutkainen suhde

Sijoitusmaailmassa riski-tuottosuhte on mittari, joka vertailee potentiaalista voittoa sijoituksen mahdollisiin riskeihin. Koneoppimisen tuomat mahdollisuudet ovat luoneet myös riski-tuottosuhteen analysointiin uusia ovia. The Journal of Finance and Data Science - lehdessä julkaistussa tutkimuksessa avataan riskien ja tuoton monimutkaista suhdetta markkinoiden volatiliteetin avulla (Pinelis ja Ruppert 2022). Artikkelin kirjoittajien Michael Pineliksensä sekä David Ruppertin mukaan tuottoon liittyvien riskien hallinta riippuu sekä siitä ajoituksesta, jolloin tuottoa haetaan, että markkinoiden yleisestä volatiliteetista sijoitushetkellä.

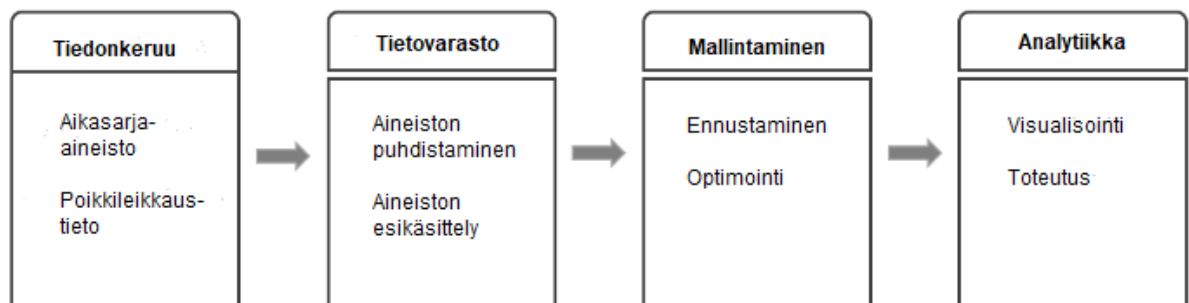
Riski-tuottosuhdetta sekä siihen liittyvien tekijöiden hallintaa voidaan tehostaa koneoppimisen algoritmien avulla. Yksi esimerkki tällaisesta lähestymistavasta on vahvistusoppiminen (engl. reinforcement learning). Vahvistusoppiminen on koneoppimisen osa-alue, joka pyrkii toimimaan vuorovaikutuksessa ympäristönsä kanssa erilaisten tavoitteiden saavuttamiseksi. Vahvistusoppimisen tehokkaiden algoritmien ansiosta se voi oppia datavetoisesti hallitsemaan riski-tuottosuhteen monimutkaista luonnetta (Yu ym. 2019).

4 Koneoppimisen ongelmat sijoitussalkun optimoinnissa

Tämän tutkielman sisällön on ollut tähän mennessä tarkoitus selittää loogisesti sekä ymmärrettävästi koneoppimisen sekä sijoitussalkun optimoinnin välisiä yhteyksiä. Koneoppimista ja sen luomia mahdollisuuksia sijoitussalkun optimoinnissa on kuvattu tutkielmassa tuomatta esiin siihen liittyviä riskejä sekä ongelmia. Tämän luvun tarkoitus on nostaa esiin myös myös aiheen tätä puolta.

4.1 Markkinoiden arvaamattomuus

Kvantitatiivinen sijoittaminen ja sen strategiat liittyvät vahvasti markkinoiden sen hetkisiin trendeihin ja volatilitettiin sekä pyrkimykseen ajoittaa omaisuuserän hankinta ja myynti mahdollisimman optimaalisella tavalla. Markkinoiden arvaamattomuutta ja sijoitussalkun optimointiin liittyvien ennustusmallien kehittämistä koneoppimista hyödyntäen on tutkittu laajasti, mutta se koetaan edelleen yhdeksi finanssimarkkinoiden suurimmista haasteista. Tan, Liun ja Tadessen (2020) mukaan moniulotteiset aikasarja-attribuutit ja finanssimarkkinoista kerätyn datan monimutkaisuus sekä runsaus ovat kaikki osaltaan vaikuttamassa tehokkaiden ennustusmallien luomisen vaikeuteen. Toisaalta Wen, Lin ja Nie (2020) tuovat esiin myös muita ennustusmallien luomiseen liittyviä ongelmia. Poliittisten, makrotaloudellisten sekä oikeudellisten tekijöiden aiheuttamiin muutoksiin on äärimmäisen vaikeaa varautua. Näiden muutosten tapahtuessa, myös finanssimarkkinat kokevat suurta epävarmuutta, minkä johdosta osakemarkkinoilla tapahtuvat liikkeet eivät välttämättä aina vastaa markkinoiden todellista tilaa.



Kuvio 6. Esimerkki kvantitatiivisen sijoittamisen toteutuksesta (Ta, Liu ja Tadesse 2020).

Markkinoiden arvaamattomuuden ja sen tulkitsemisen vaikeuden tueksi on esitetty myös useita teorioita. Satunnaiskulkuhypoteesin mukaan, joka on yksi tämän aihepiirin tunnetuimmista teorioista, finanssimarkkinoiden hinnanvaihteluita ei voida ennustaa, koska ne kehittyvät satunnaiskulun mukaan. Smidt (1968) avaa artikkelissaan satunnaiskulkuhypoteesiin liittyviä kysymyksiä seuraavalla tavalla: ”Satunnaiskulkuhypoteesin perusidea on, että vapailla kilpailluilla markkinoilla tuotteiden tai palveluiden sen hetkiset hinnat kuvaavat markkinoiden ajankohtaista tietoa kyseisestä tuotteesta tai palvelusta.” Teorian mukaan ainoastaan satunnaiskulun varassa olevat muuttujat, kuten kilpailijoiden epäonni, voivat todella heiluttaa vapaiden kilpailluiden markkinoiden hintatasoa. Toimivien ennustusmallien luominen on siksi mahdotonta.

4.2 Opetusdatan ongelmat

Vaikka finanssimarkkinoista on saatavilla paljon korkealaatuista tietoa, ei se poista kokonaan koneoppimisen kouluttamiseen käytettävän opetusdatan riskejä. Finanssimarkkinoiden luonteenomainen jatkuva liike ja nopeat suunnanvaihdokset luovat arvaamattomuudellaan jo valmiiksi haasteelliset olosuhteet, joiden navigointi saattaa osoittautua hyvin vaikeaksi. Satunnaiskulkuhypoteesin esiin nostamat haasteet ovat tästä hyvä esimerkki. (Smidt 1968)

Opetusdatan mahdolliset ongelmat eivät myöskään lopu tähän. Opetusdatan heikko tiheys voi aiheuttaa ongelmia, joita ei pahimmassa tapauksessa saada edes koneoppimisen testivaiheessa kiinni. Yksi esimerkki tämän tapaisesta ongelmasta voi olla muun muassa tarkkuuden puute hahmontunnistuksen aikana mallien ja kuvioiden tunnistamisessa, mikä voi johtaa datan väärään luokittelutulokseen (Devarakota ja Mirbach 2008).

5 Yhteenveto

Tässä tutkielmassa tarkasteltiin sijoitussalkun optimointia koneoppimisen näkökulmasta. Tutkielma suoritettiin kirjallisuuskatsauksena, jossa koneoppimisen todettiin tarjoavan useita eri mahdollisuuksia sijoitussalkun optimoinnin ja yleisemmin myös finanssimarkkinoiden tueksi. Toisaalta myös ongelmakohtia löydettiin. Koneoppimisen hyödyt sijoitussalkun optimoinnissa rajattiin tässä tutkielmassa kolmeen pääkohtaan:

- Hahmontunnistus
- Ulottuvuuksien vähentäminen
- Riskien ja tuotto-odotusten suhde

Hahmontunnistuksen edut sijoitussalkun optimoinnissa keskittyvät sen kykyyn luokitella käsitellyssä oleva malli tai kaava tunnetun datan perusteella omaan luokkaansa. Ulottuvuuksien vähentämisen avulla pystytään puolestaan yksinkertaistamaan tietoaineistojen monimutkaisuuksia vähentämällä niissä olevien tulosuureiden määrää ja riskien ja tuotto-odotusten suhdetta on mahdollista hallita vahvistusoppimisen avulla. Koneoppimisen ongelmat sijoitussalkun optimoinnissa keskittyvät sen sijaan vahvasti ennustamisen vaikeuteen liittyviin ongelmiin:

- Markkinoiden arvaamattomuus
- Opetusdatan ongelmat

Markkinoiden arvaamattomuuden vuoksi ennustusmallien kehittäminen on vaikeaa. Myös opetusdatan ongelmat tuovat omat haasteensa. Jos valitun opetusdatan tiheys on heikko, se voi aiheuttaa tarkkuuden puutetta hahmontunnistuksessa ja johtaa datan väärään luokitteluun. Opetusdatan valintaan liittyviä kysymyksiä olisi syytä tutkia lisää.

Lähteet

Abarbanell, Jeffrey S., ja Brian J. Bushee. 1997. “Fundamental Analysis, Future Earnings, and Stock Prices”. *Journal of Accounting Research* 35 (1): 1–24. Viitattu 25. huhtikuuta 2022.

Alpaydin, Ethem. 2014. *Introduction to Machine Learning*. The MIT Press.

Alzubi, Jafar, Anand Nayyar ja Akshi Kumar. 2018. “Machine Learning from Theory to Algorithms: An Overview”. *Journal of Physics: Conference Series* 1142:012012. <https://doi.org/10.1088/1742-6596/1142/1/012012>.

Beyerer, Jürgen, Matthias Richter ja Matthias Nagel. 2017. *Pattern Recognition*. De Gruyter.

Bonga, Wellington Garikai. 2015. “The Need for Efficient Investment: Fundamental Analysis and Technical Analysis”. *Social Science Research Network (SSRN)*, <https://doi.org/10.2139/ssrn.2593315>.

Chen, James. 2010. *Essentials of Technical Analysis for Financial Markets*. John Wiley & Sons, Incorporated.

Devarakota, Pandu, ja Bruno Mirbach. 2008. “Reliability Estimation of a Statistical Classifier”. *Pattern Recognition Letters*, https://www.academia.edu/1559559/Reliability_estimation_of_a_statistical_classifier.

Emerson, Sophie, Ruairí Kennedy, Luke O’Shea ja John O’Brien. 2019. “Trends and Applications of Machine Learning in Quantitative Finance”. *8th International Conference on Economics and Finance Research (ICEFR 2019)* (toukokuu).

Ferreira, Fernando, Amir Gandomi ja Rodrigo Cardoso. 2021. “Artificial Intelligence Applied to Stock Market Trading: A Review”. *IEEE Access* 9:30898–30917.

Fu, Tak-chung, Fu-lai Chung, Robert Luk ja Chak-man Ng. 2008. “Representing financial time series based on data point importance”. *Engineering Applications of Artificial Intelligence* 21 (2): 277–300. ISSN: 0952-1976. <https://doi.org/10.1016/j.engappai.2007.04.009>.

- Iba, Hitoshi, ja Claus Aranha. 2012. *Practical Applications of Evolutionary Computation to Financial Engineering: Robust Techniques for Forecasting, Trading and Hedging*. Springer Berlin Heidelberg.
- Jorion, Philippe. 1992. "Portfolio Optimization in Practice". *Financial Analysts Journal* 48 (1): 68–74.
- Li, Yelin, Junjie Wu ja Hui Bu. 2016. "When quantitative trading meets machine learning: A pilot survey". Teoksessa *2016 13th International Conference on Service Systems and Service Management (ICSSSM)*, 1–6. <https://doi.org/10.1109/ICSSSM.2016.7538632>.
- Martins, Tiago Mousinho, ja Rui Ferreira Neves. 2020. "Applying genetic algorithms with speciation for optimization of grid template pattern detection in financial markets". *Expert Systems with Applications* 147:113191. ISSN: 0957-4174. <https://doi.org/10.1016/j.eswa.2020.113191>.
- Nyberg, Peter, ja Mika Vaihekoski. 2010. "A New Value-Weighted Total Return Index for the Finnish Stock Market 1912-1969". *Research in International Business and Finance* 24 (3): 267–283.
- Palma, Wilfredo. 2016. *Time Series Analysis*. John Wiley & Sons, Incorporated.
- Pinelis, Michael, ja David Ruppert. 2022. "Machine learning portfolio allocation". *The Journal of Finance and Data Science* 8:35–54. ISSN: 2405-9188. <https://doi.org/10.1016/j.jfds.2021.12.001>.
- Reddy, G. Thippa, M. Praveen Kumar Reddy, Kuruva Lakshmana, Rajesh Kaluri, Dharmendra Singh Rajput, Gautam Srivastava ja Thar Baker. 2020. "Analysis of Dimensionality Reduction Techniques on Big Data". *IEEE Access* 8.
- Smidt, Seymour. 1968. "A New Look at the Random Walk Hypothesis". *Journal of Financial and Quantitative Analysis* 3 (3): 235–261. <https://doi.org/10.2307/2329812>.
- Ta, Chuan-Ming Liu ja Tadesse. 2020. "Portfolio Optimization-Based Stock Prediction Using Long-Short Term Memory Network in Quantitative Trading". *Applied Sciences* 10 (tammi-kuu): 437. <https://doi.org/10.3390/app10020437>.

Wen, Yulian, Peiguang Lin ja Xiushan Nie. 2020. "Research of Stock Price Prediction Based on PCA-LSTM Model". *IOP Conference Series: Materials Science and Engineering* 790, numero 1 (maaliskuu): 012109. <https://doi.org/10.1088/1757-899x/790/1/012109>.

Yu, Pengqian, Joon Sern Lee, Ilya Kulyatin, Zekun Shi ja Sakyasingha Dasgupta. 2019. *Model-based Deep Reinforcement Learning for Dynamic Portfolio Optimization*. <https://doi.org/10.48550/arxiv.1901.08740>.

Zhou, Zhongbao, Meng Gao, Helu Xiao, Rui Wang ja Wenbin Liu. 2021. "Big data and portfolio optimization: A novel approach integrating DEA with multiple data sources". *Omega* 104:102479.