

TALLINNA TEHNIKAÜLIKOOL  
Infotehnoloogia teaduskond

Markus Somp 154763IABB

**AKTSIAINDEKSI HINNALIIKUMISTE  
PROGNOOSIMINE MASINÕPPEGA  
ÜLLATAVATE  
MAJANDUSINDIKAATORITE PÕHJAL**

Bakalaureusetöö

Juhendaja: Martin Rebane

MSc

Kaasjuhendaja: Kristofer Vähi

Tallinn 2018

## **Autorideklaratsioon**

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Markus Somp

21.05.2018

## Annotatsioon

Käesoleva töö eesmärk on leida seaduspärasusi S&P500 aktsiaindeksi ebatavalistes hinnaliikumistes peale üllatava majandusindikaatori avaldamist ja võimalusi, kuidas investor või kaupleja saaks tekkivaid olukordi kasumlikult ära kasutada. Selleks realiseeritakse käesolevas töös kaks kauplemisalgoritmi – deterministlik ja masinõppel põhinev algoritm – ning võrreldakse nende tulemusi.

Deterministlik algoritm realiseeriti eeldusel, et aktsiaturg kordab ajaloolisi hinnaliikumisi vastavalt töös käsitletud olukordadele. Sügav-masinõppel põhineva algoritmi koostamisel kasutas autor rekurrentseid närvivõrke ja GRU meetodit. Algoritmide parameetrid leitakse optimeerimise tulemusena.

Kauplemisalgoritmide tulemusi võrreldakse viie USA kõige tähtsama majandusindikaatori raames. Deterministliku kauplemisalgoritmi, mis arvestas üllatavate majandusindikaatoritega, kogutootlus oli kõigi indikaatorite puhul kõrgem kui masinõppel põhineva kauplemisalgoritmi kogutootlus. See kinnitab asjaolu, et aktsiaindeksi ebatavaline reageerimine üllatavale majandusindikaatori avaldamisele ei ole juhuslik ning, et selliseid olukordi on võimalik investoril või kauplejal ära kasutada. Autor täheldas, et eelnevalt kirjeldatud ebatavalised hinnareaktsioonid võivad olla algused 6- kuni 7-päevasele korrektsioonile.

Kauplemisalgoritmide on realiseeritud programmeerimiskeeles Python ning sügavate närvivõrkude implementeerimiseks on kasutatud TensorFlow'd ja Kerast. Kogu koodibaas on avalikustatud autori GitHubi veebilehel, aadressil <https://github.com/markussomp/Kauplemisalgoritmide>.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 39 leheküljel, 6 peatükki, 20 joonist, 1 tabelit.

## **Abstract**

### **Predicting stock index price movements after surprising economic indicators using machine learning**

The purpose of this thesis is to find regularities in unusual price movements of the S&P500 Stock Index after the publication of a surprising economic indicator and ways how an investor or trader can profitably exploit emerging situations. As a result, the author creates two trading algorithms – a deterministic and machine learning based algorithm – and compares their results.

The deterministic algorithm was realized on the assumption that stock market repeats historical price movements in accordance with the situations discussed in the work. In developing a deep machine learning algorithm, the author used recurrent neural networks and the GRU method. Algorithm parameters were found as a result of optimization.

The results of trading algorithms are compared to the five most important economic indicators in the US. Deterministic trading algorithm, which takes into account surprising economic indicators, generated a higher total return than machine learning trading algorithm for all indicators. This confirmed the fact that the unusual response of the stock index to the publication of a surprising economic indicator is not coincidental, and such situations can be exploited by an investor or a trader. The author observed that unusual price responses described above may be the beginning of a 6- to 7-day correction.

Trading algorithms are realized in the Python programming language, using TensorFlow and Keras to implement deep neural networks. The codebase is published on the author's GitHub page at <https://github.com/markussomp/Kauplemisalgorithmid>.

The thesis is in Estonian and contains 39 pages of text, 6 chapters, 20 figures, 1 table.

## Lühendite ja mõistete sõnastik

Aken	<i>Sliding Window</i> , eelnevate üllatavate majandusindikaatorite avaldamiste hulk
API	<i>Application Programming Interface</i> , arvuti operatsioonisüsteemiga või rakendusprogrammiga määratud reeglistik, mille alusel rakendusprogramm kasutab operatsioonisüsteemi või teise rakendusprogrammi teenuseid
CPI	<i>Consumer Price Index</i> , tarbijahinnaindeks
CREDIT	<i>Consumer Credit Index/Report</i> , tarbijakrediidiindeks
ETT	Efektiivse turu teooria
GRU	<i>Gated Recurrent Unit</i> , olekuvärvatega varustatud rekurrentne/tagasisidestatud üksus
Hüppe päev	Päev, kus aktsiaindeks on teinud märkimisväärse hüppe hinnaliikumises
LSTM	<i>Long Short Term Memory</i> , pikk lühiajaline mälu
MA	<i>Moving Average</i> , hinnaliikumise libisev keskmine
Majandusindikaator	Statistiline informatsioon, mida kasutatakse majanduse hetkeolukorra kvantitatiivseks kirjeldamiseks [47]
NFP või NFPAY	<i>Nonfarm Payrolls</i> , mittepõllumajanduse tööhõive raport
PMI	<i>Purchasing Managers Index</i> , ostujuhtide indeks
RSXAUT	<i>Retail Sales Excluding Auto</i> , jaemüügiraport ilma automüügita
S&P500	Aktsiaindeks <i>Standard &amp; Poors</i> 500
Üllatus	Majandusindikaatori reaalkäitumise ja prognooside konsensusvahe

## Sisukord

1 Sissejuhatus .....	11
2 Informatsiooni roll finantsturgudel.....	13
2.1 Efektiivse turu teooria (ETT) .....	14
2.1.1 Volatiilsus ja ETT.....	15
2.2 Fundamentaalne analüüs ja majandusindikaatorite olulisus.....	16
2.2.1 Varasemad uuringud majandusindikaatorite seostest finantsturu hindadega	17
2.3 Olulisemad majandusindikaatorid .....	19
2.3.1 Majandusindikaatorite liigitamine.....	21
2.3.2 Mittepõllumajanduse tööhõive raport (NFP või NFPAY) .....	22
2.3.3 Jaemüügiraport ilma automüügita (RSXAUT) .....	22
2.3.4 Tarbijakrediidiindeks (CREDIT).....	22
2.3.5 Tarbijahinnaindeks (CPI) .....	23
2.3.6 Ostujuhtide indeks (PMI) .....	23
2.4 Tehniline analüüs.....	24
2.4.1 Trend ja libisev keskmine.....	25
3 Masinõppe teoreetilised alused.....	26
3.1 Masin-sügavõpe.....	27
3.2 Rekurrentsed närvivõrgud .....	28
3.2.1 LSTM .....	28
3.2.2 GRU.....	29
4 Algoritmide realiseerimine .....	31
4.1 Deterministlik algoritm.....	32
4.2 Masinõppel põhinev algoritm.....	34
4.2.1 Arhitektuur .....	34
4.2.2 Hüperparameetrid .....	37
4.3 Python.....	38
4.3.1 Numpy .....	38
4.3.2 Pandas.....	39
4.3.3 TensorFlow.....	40

4.3.4 Keras.....	40
4.3.5 Scikit-learn ja h5py.....	40
5 Analüüs ja arutelu.....	42
5.1 Deterministliku algoritmi tulemused.....	42
5.1.1 NFP.....	42
5.1.2 RSXAUT.....	43
5.1.3 CREDIT.....	43
5.1.4 CPI.....	44
5.1.5 PMI.....	44
5.1.6 Järeldus.....	45
5.2 Masinõppel põhineva algoritmi tulemused.....	46
5.3 Algoritmide omavaheline võrdlus.....	48
5.3.1 Järeldus.....	49
6 Kokkuvõte.....	50
Kasutatud kirjandus.....	51
Lisa 1 – S&P500 turuhinna liikumine peale üllatava majandusindikaatori avaldamist.....	55
Lisa 2 – Masinõppel põhineva kauplemisalgoritmi tulemusi illustreerivad joonised....	56
Lisa 3 – Deterministlik kauplemisalgoritm.....	58
Lisa 4 – Masinõppel põhinev algoritm (preparation, training, plotting).....	59

## Jooniste loetelu

Joonis 1. S&P500 volatiilsusindeksi VIX graafik. Punane trendijoon tähistab aastate 1990–2008 ajaloolist keskmist [49].	16
Joonis 2. Loetelu majandusindikaatoritest ja nende lühenditest [18].	19
Joonis 3. Suhe hüppe päeva ja majandusindikaatorite avaldamise päevade vahel [18].	20
Joonis 4. Fundamentaalse ja tehnilise analüüsi osakaal sõltuvalt investeerimisperioodist [40].	24
Joonis 5. Masinõppe liigitamine koos kategooriale vastavate algoritmide näidetega [13].	27
Joonis 6. LSTM-i illustratsioon, kus $i$ , $f$ ja $o$ on sisendi (Input Gate), unustamise (Forget Gate) ja väljundi (Output Gate) väravad. $C$ ja $c$ katusega on vastavalt mäloraku sisu ja uue mäloraku sisu [2].	29
Joonis 7. GRU illustratsioon, kus $r$ on taastav värav (Reset Gate) ja $z$ on uuenduse värav (Update Gate), $h$ ja $h$ katusega on mälorakk ja kandidaat mälorakk [2].	29
Joonis 8. Vaatluse all olevate sündmuste lahterdamine erinevatesse olukordadesse.	33
Joonis 9. Deterministliku kauplemisalgoritmi üldistatud loogika. Täpsemalt Lisas 3.	33
Joonis 10. Masinõppel põhineva algoritmi arhitektuur.	35
Joonis 11. Masinõppel põhineva algoritmi tegevusdiagramm.	36
Joonis 12. Andmemassiivi läbitöötamise aja võrdlus puhta Pythoni ja NumPy kasutamisel [14].	39
Joonis 13 Deterministliku kauplemisalgoritmi parim tulemus NFP puhul. Vasakpoolne graafik iseloomustab positsiooni avamise ja sulgemise kohti S&P500 aktsiaindeksi hinnaliikumise suhtes. Parempoolne graafik iseloomustab portfelli suurust ajas. Roheline ja must kolmnurk tähistavad vastavalt ostu- ja müügitehinguid.	43
Joonis 14. Deterministliku kauplemisalgoritmi parim tulemus RSXAUT puhul.	43
Joonis 15. Deterministliku kauplemisalgoritmi parim tulemus CREDIT-i puhul.	44
Joonis 16. Deterministliku kauplemisalgoritmi parim tulemus CPI puhul.	44
Joonis 17. Deterministliku kauplemisalgoritmi parim tulemus PMI puhul.	45



Joonis 18. Ennustusmudeli prognoositud hinnaliikumine vahemikus juuli 2014 kuni detsember 2016. Must joon tähistab S&P500 aktsiaindeksi tegelikku hinda ning värvilised jooned on aktsiaindeksi hinnaliikumise 10 päeva prognoosid.....	47
Joonis 19. Tegelik aktsiaindeksi hinna (must) ja 10 päeva prognoositud hinna (sinine) ühisjaotuse graafiline esitus.....	47
Joonis 20. Kauplemisalgoritmide tulemused. D ja M tähistavad vastavalt deterministlikku ja masinõppel põhinevat kauplemisalgoritmi.....	48

## **Tabelite loetelu**

Tabel 1. Optimeerimise käigus leitud parameetrite väärtused. .... 37

# 1 Sissejuhatus

Kuna finantsturgudel on kauplejate ja investorite eesmärk teenida tulu, siis jälgitakse pingsalt majandusindikaatorite avaldamisi. Majandusindikaatorid on aluseks tulevikuprognosile ja seeläbi ka positsioonide võtmisel turgudel. Samuti on üha levinum kauplemisalgoritmide kasutamine, mis baseeruvad nii tehnilisel analüüsil kui masinõppel.

Turud ei reageeri majandusindikaatori avaldamisele alati nii nagu praktikast lähtudes eeldada võiks. Töö eesmärk on leida seaduspärasusi S&P500 aktsiaindeksi ebatavalistes hinnaliikumistes peale üllatava majandusindikaatori avaldamist ja võimalusi, kuidas investor või kaupleja saaks tekkivaid olukordi kasumlikult ära kasutada. Selleks realiseeritakse käesolevas töös kaks kauplemisalgoritmi – deterministlik ja masinõppel põhinev algoritm – ning võrreldakse nende tulemusi. Eesmärgi saavutamiseks hõlmab töö vaheülesandeid:

- uurida majandusindikaatoreid, mis mõjutavad aktsiaturgu kõige enim;
- uurida, kuidas aktsiaturg käitub peale prognoosist märkimisväärselt erineva majandusindikaatori avaldamist;
- uurida, kuidas deterministliku kauplemisalgoritmi raames saab tehnilist ja fundamentaalset analüüsi ühendada;
- uurida, milliste tööriistadega on võimalik realiseerida sügaval närvivõrgul põhinevat kauplemisalgoritmi;
- leida parim viis aktsiaindeksi hinnaliikumise ennustumudeli realiseerimiseks, kasutades varasema perioodi hinnaliikumisi.

Analüüsitava teema on aktuaalne, sest S&P500 aktsiaindeksi turukapitalisatsioon on kõigi aegade tipus, käive on suur ning investorid ja kauplejad otsivad järjepidevalt signaale, mida oleks võimalik kasumlikult ära kasutada. Samuti on tehnoloogia arengu

tõttu nii institutsionaalsed kui ka personaalsed kauplemisalgoritmid muutumas aina populaarsemaks. Tänu arvutusvõimsuse kasvule on väga päevakohane tehislike närvivõrkude uurimine ja nende kasutamine kõikvõimalikele probleemidele lähenemisel.

Töö sisuline osa jaguneb neljaks. Alguses tutvustatakse lugejale informatsiooni rolli aktsiaturgudel ja majandusindikaatorite olulisust. Samuti esitletakse varasemaid uuringuid nende omavahelistest seostest. Seejärel selgitatakse masinõppe teoreetilisi aluseid ning selle liigitumist. Pärast seda keskendub autor algoritmide realiseerimisele ja tööriistade kirjeldamisele. Viimaks analüüsib autor arutelu käigus töö tulemusi, võrdleb algoritme, teeb järeldused ning annab sellele tuginedes soovitusel edasisteks uuringuteks. Masinõppel põhineva algoritmi puhul on autor silmas pidanud kitsamat tähendust, st masinõpet käsitletakse kui tehislikke närvivõrke.

Autor tänab juhendajaid mitmete ideede ja paranduste eest bakalaureuse töö valmimisel.

## 2 Informatsiooni roll finantsturgudel

Finantsturule tuleb pidevalt uut informatsiooni uudiste, makromajanduse näitajate, majandussektorite väljavaadete ja ettevõtete finantsolukordade näol. Uue informatsiooni ilmumisel hakkab kujunema alusvara hind. Selle kujunemise kiirus ja ulatus oleneb informatsiooni olulisusest, turu likviidsusest ja sellest, kas ja kui ruttu see investoriteni jõuab. Viimase kriteeriumi järgi saab informatsiooni liigitada avalikuks ja mitteavalikuks. Mitteavalik informatsioon on teave, mida ei ole avalikkusele avaldatud. Näiteks ettevõtte finantsaruanne võib olla kokku pandud ja kooskõlastatud, aga avaldamise kuupäev on millalgi tulevikus. Mitteavaliku informatsiooni omamist nimetatakse ka siseinfo omamiseks. Sellest on kujunenud üks viimase aja olulisemaid probleeme börsil, kuna siseinfot omavad kauplejad ostavad või müüvad alusvara enne info avalikustamist. Siseinformatsiooni valdamine on küll legaalne, kuid reguleeritud turgudel, näiteks aktsiabörsil, on selle teabe kasutamine alusvara ostuks või müügiks keelatud.

Lisaks majanduslikele näitajatele avaldavad finantsturgudele suurt mõju ka riikide poliitiline olukord ja tegevus, looduskatastroofid, loodushoid, toorainete kättesaadavus jpm. Informatsiooni hulk ja selle momentaanne kajastumine alusvara hindades teeb finantsturud maailma kõige kiiremini muutuvaks keskkonnaks. Naljatades öeldakse, et kui soovite teada Florida ilma, siis pole mõtet vaadata ilmateadet, vaid uurida, kui palju maksavad apelsinid New Yorgi börsil [26].

Kuna uus informatsioon tekitab inimestes emotsioone, siis võivad investorid viia turuhinna ajutiselt eemale turu tasakaalust. Nii võivad tekkida ülereageerimised – liigne optimism ja eufooria põhjustavad liiga kõrgeid hindu või ülemäärane pessimism viib hinnad liiga madalale. Selle tulemusena arutlevad majandusteadalased jätkuvalt, kas finantsturud on üldiselt efektiivsed.

## 2.1 Efektiivse turu teooria (ETT)

Efektiivse turu teooria (edaspidi ETT) looja on Eugene Fama, kes tutvustas seda 1964. aastal oma doktoritöös. Algselt defineeris ta efektiivset turgu kui turgu, kus väärtpaberi hind on igal hetkel enda sisemise väärtuse hea hinnang. Aastal 1969 kohandas Fama efektiivse turu teooria sõnastust, väites, et turg on efektiivne, kui ta kohandub kiiresti uuele informatsioonile. Hiljem, 1970. aastal, parandas Fama veelgi sõnastust, öeldes, et turg on efektiivne, kui vara hind kajastab täielikult kogu kättesaadavat informatsiooni. Sellega tõi Fama kasutusele mõiste „informatsiooniline efektiivsus“ [48].

ETT puhul eristatakse kolme vormi:

1. nõrgalt efektiivne turg – väärtpaberite hinnad kajastavad igal ajahetkel kogu olemasolevat informatsiooni väärtpaberite hindade kohta minevikus;
2. keskmiselt efektiivne turg – väärtpaberite hinnad kajastavad igal ajahetkel kogu kättesaadavat avalikku informatsiooni selle väärtpaberi kohta;
3. tugevalt efektiivne turg – väärtpaberi hind kajastab igal ajahetkel lisaks avalikule ka mitteavalikku teavet selle väärtpaberi kohta.

1970. aastal täpsustas Eugene Fama eelduste sõnastust, mis tagasid ETT piisavad tingimused:

1. tehingukulude puudumine;
2. informatsiooni tasuta kättesaadavus kõigile turuosalistele;
3. turuosaliste ühise seisukoha kajastumine nii hetkel kehtivale kui ka tulevastele hindadele [48].

On selge, et selline lähenemine on reaalistest finantsturgudest üsna kaugel – turuosalised maksavad tehingukulud, kusjuures tehingukulud võivad investoritel erineda lausa sadades kordades; kõikide investoriteni ei jõua informatsioon samal ajahetkel ehk tihtipeale saab tasulisi kanaleid pidi informatsiooni kiiremini kätte; turuosalised ei tõlgendada uut informatsiooni samasuguse ajaga ning kogemuse puudumise tõttu kujuneb teatul grupil ebaadekvaatne arvamus alusvara hinna suhtes hetkel ja tulevikus. Sellele vaatamata arvavad finantsturgude uurijad, et turud liiguvad

aina rohkem tugevama efektiivsuse suunas. Sellele võivad kaasa aidata ka masinkauplejad, kes on kogu aeg avatud uuele informatsioonile ja selle analüüsimisele, ning kes avavad turul positsioone momentaanselt. Sellega koos suureneb anomaaliate risk, mis võib väljenduda näiteks informatsiooni valesti tõlgendamises või valeinformatsiooni võimendamises börsil. Ka viimased nn *Flash Crash*'id on just masinkauplejate õlgadel.

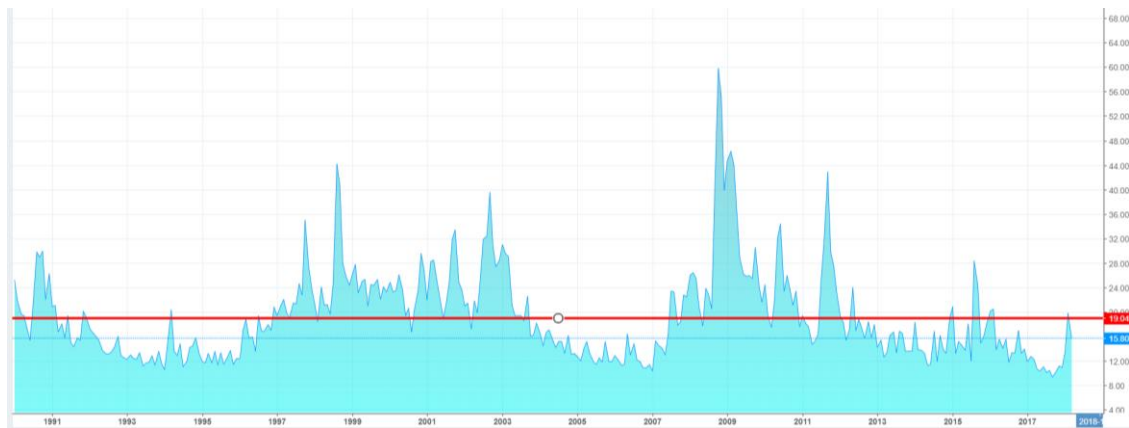
### **2.1.1 Volatiilsus ja ETT**

Numbriliseks viisiks, kuidas väljendada minevikus esinenud määramatust ja ekstrapoleerida seda kuidagi ka tulevikku, hakati kasutama „volatiilsust“ ehk hinnamuutuste standardhälvet [25]. Lihtsamalt öeldes on volatiilsus alusvara hinna kõikumine – mida rohkem alusvara hind kõigub, seda suurem on volatiilsus. Professor John C. Hull on kirjutanud, et volatiilsus on alusvara tuleviku tootluse määramatuse mõõdik [20].

Volatiilsuse tõlgendamisel saab huvitava märkusena välja tuua vastuolu ETT looja Eugene Fama arusaama ja informatsiooniteooria vahel. Viimane väidab, et informatsiooni saab vaadelda kui määramatuse vähenemist. Seega peaks informatsiooniteooria kohaselt informatsiooni hulga suurenemine vähendama määramatust ja tooma kaasa madalama volatiilsuse.

Eugene Fama on öelnud, et ETT ei tähenda madalat volatiilsust. Vastupidiselt, ta väidab, et teooria tähendab kõrget volatiilsust. Hinnad peavad muutuma kiiresti vastavalt uuele informatsioonile. Kui ebakindlus on kõrge, peab ka volatiilsus olema kõrge [11].

Nagu näha, on kaks teooriat sisulises vastuolus. Ükskõik kummalt poolt ETT-le läheneda, ei ole turg kogu aeg efektiivne. Kui järgida informatsiooniteooriat, siis on turg efektiivne madala volatiilsuse korral, kui aga Eugene Famat, siis on turg efektiivne kõrge volatiilsuse korral (Joonis 1. X-teljel on kujutatud aega, y-teljel volatiilsusindeksi VIX väärtust).



Joonis 1. S&P500 volatiilsusindeksi VIX graafik. Punane trendijoon tähistab aastate 1990–2008 ajaloolist keskmist [49].

Käesoleva töö raamides võib siiski teha järelduse, et finantsturud on nõrgalt efektiivsed, sest vaatluse all olevad aktsiabörsid on maailma suurimad. S&P500 indeksit sisaldavad ettevõtted kauplevad New Yorgi ja Nasdaq'i aktsiabörsil, mis hõivavad turukapitalisatsiooni poolest vastavalt esimese ja teise koha. Võib mõistlikult eeldada, et sellistel turgudel on piisavalt palju osalisi, kes tagavad ETT suures plaanis toimimise. Hinnad sisaldavad vähemalt enamikku kogu olemasolevast teabest ning uus informatsioon peegeldub koheselt hindades.

## 2.2 Fundamentaalne analüüs ja majandusindikaatorite olulisus

Fundamentaalseks analüüsiks nimetatakse ettevõtte analüüsimist ja hindamist, keskendudes finants- ja majandusaruannetele, turusituatsioonile, juhtkonnale ja konkurentsile eesmärgiga leida näiteks turu poolt alahinnatud või ootustest kiiremini kasvavaid ettevõtteid [36].

Fundamentaalse analüüsi otstarbekuse kinnitamiseks tuleb teha kaks olulist eeldust. Esiteks võib ettevõtte turuhind erineda tema õiglasest väärtusest (ettevõtte võib olla üle- või alahinnatud). Teiseks tuleb eeldada, et pikemas perspektiivis parandab turg oma vea ning turuhind liigub õiglase väärtuse suunas [21].

Kuna selle töö eesmärk on vaadata laiemat pilti, siis ei ole üksikute ettevõtete analüüsimisega vaja tegeleda. Ettevõtete käekäiku üldistavad fundamentaalsed näitajad on kõikidele turuosalistele samad – majanduskeskkond ja turusituatsioon. Kõige paremini sobivad üldise majanduskeskkonna kirjeldamiseks riiklikud või regionaalsed majandusindikaatorid, näiteks statistika tööhõive, sisemajanduse koguprodukti,



ettevõtete kasumi ja käibe, investeeringute, hindade, laenude, tootmise, kaubanduse jms kohta. Majandusindikaatoreid avaldavad riiklikud institutsioonid kokku lepitud perioodide järel, milleks võib olla näiteks kuu, kvartal või aasta. Seda teavet kasutavad turuosalisel nii üldise majanduskeskkonna hindamiseks, trendide mõistmiseks, analüüsitudel täiendamiseks, minevikus tehtud otsuste kontrolliks ja positsioonide juhtimiseks. Loomulikult püütakse prognoosida tulevikusündmusi ja leida parimaid rahapaigutusvõimalusi. Saadud informatsiooni abil sooritavad investorid ja kauplajad nii pikaajalisi investeerimisotsuseid kui ka lühiajalisi, päevasiseid tehinguid. Poliitikud kasutavad indikaatoreid langetamaks poliitilisi otsuseid. Kui majandusindikaatori tegelik väärtus erineb turu oodatud väärtusest, siis kohanduvad hinnad koheselt indikaatori avaldamisel [47].

### **2.2.1 Varasemad uuringud majandusindikaatorite seostest finantsturu hindadega**

On tehtud palju uuringuid, et kinnitada või ümber lükata seosed majandusindikaatorite ja aktsia hindade vahel. Roger Waud'i 1970. aasta uuring oli üks esimesi ning näitab kokkuvõtlikult, et aktsia hinnad reageerivad koheselt muutustele intressimäärades. Kuigi see uuring ei pööranud tähelepanu oodatavatele ja mitte oodatavatele muutustele eraldi, oli see ikkagi üks esimesi kvantitatiivseid uuringuid näitamaks seost aktsia hindade ja majandusindikaatorite vahel [33].

Aastal 1979 uurib Richard Castanias veelgi täpsemalt eelmainitud seost ning leiab, et aktsia hinnad on palju muutlikumad päevadel, mil avaldatakse teatud majandusindikaatorid. Ta lähtus hüpoteesist, et aktsia hinnad sisaldavad endas adekvaatselt kogu avalikku informatsiooni ning tegi järelduse, et kui majandusindikaator erineb prognoosist, siis see peaks aktsia hindasid mõjutama [33].

Pearce ja Roley olid esimesed, kes uurisid otseselt ootamatute ja üllatavate majandusindikaatorite mõju aktsiaturgudele. 1983. aasta uuringus leiti, et monetaarpoliitilised otsused mõjutavad aktsia hindasid ainult siis, kui nende otsuste ulatus või suund on ettearvamatud. Samuti jõuti järelduseni, et selle tulemusena toimunud turu liikumised on järgmise kauplemispäeva hommikuks täielikult sisse arvestatud ja lõppenud [33].

1986. aastal Chen, Roll ja Ross uurisid süvendatult üllatavate majandusindikaatorite mõju turule. Ka nemad leidsid, et üllatavad majandusindikaatorid mõjutavad aktsiate

oodatavat tootlikkust. Kuid eelnevad uuringud ei ole üllatuste mõju standardiseerinud. Balduzzi, Elton ja Green demonstreerisid aastal 2001, et majandusindikaatorite üllatuste ulatuste standardiseerimine on hädavajalik, et oleks võimalik võrrelda erinevate näitajate üllatuste osakaalu tähtsust turule. Uuringus jõuti järelduseni, et võlakirjaturul kohanduvad võlakirjade hinnad uuele majandusindikaatori avalikustamisele juba esimese minuti jooksul pärast avaldamist. Käive ja volatiilsus tõusevad momentaanselt ja püsivad kõrgemal tasemel kuni tund aega. Kuigi tegu on uuringuga võlakirjaturul, kinnitab see Eugene Fama ETT hüpoteesi kontekstis, et majandusindikaatorite avaldamisel kohanduvad finantsturud uuele informatsioonile koheselt [33].

McQueen ja Roley tulid 1993. aastal välja uurimusega, kus tutvustati majandusstsüklite mõju majandusindikaatorite avaldamisel aktsiabörsile. Uuring viidi läbi ajavahemiku 1977–1988 andmete põhjal ning uuriti majandusindikaatoreid, mis hõlmasid tööstustoodangut, töötuse määra, mittepõllumajanduse tööhõivet, kaubandusdefitsiiti, tootjahinnaindeksit, tarbijahinnaindeksit, M1 teadaandeid ja inflatsioonimäära. Avastati, et majandusindikaatoritel on märkimisväärne mõju aktsiabörsile, kuid nende mõju oleneb majanduse üldisest seisust. Väga positiivne üllatus majanduse kõrgperioodil toob kaasa aktsiate hinna languse ning sama üllatus majanduse madalseisu ajal toob kaasa aktsiate hinna tõusu. See on seletatav asjaoluga, et majanduslikus madalseisus annab positiivne üllatus märku sellest, et majandusel hakkab paremini minema. Täpselt sama uudis majanduse õitsengu ajal võib aga tähendada rahapoliitika karmistumist, mille tulemuseks on kõrgemad intressimäärad ja madalamad aktsiate hinnad. Järelikult on majandusindikaatori üllatuse lahterdamine „positiivseks“ või „negatiivseks“ sõltuvuses avaldamise ajastamisega [30].

Kokkuvõtvalt saab öelda, et turuhind on enne majandusindikaatorite avaldamist kujunenud vastavalt kõigi turuosaliste informatsioonile ja ootustele. Majandusindikaatori avaldamisel mängib rolli just ootamatuse komponent, mis tähendab erinevust prognoosidest, turuosaliste ja analüütikute konsensusest. Mida täpsem on teadaandele eelnenud informatsioon, seda vähem mõjutab teadaande avaldamine turuhinna liikumist. Investorid, kellel on enne teadaande avaldamist rohkem ja täpsemat informatsiooni, aimavad suurema tõenäosusega ära teadaande tegeliku väärtuse. Seetõttu ei pea need investorid pärast teadaande avaldamist oma positsioone suurel määral korrigeerima, sest nende jaoks ei muutunud olukord märkimisväärselt. Samas said investorid, kes omasid enne teadaande avaldamist vähem ja ebatäpsemat

informatsiooni, teadaande avaldamisel rohkem uut informatsiooni ning olid kohustatud korrigeerima oma positsioone turul. Selle tulemusena kujuneb pärast teadaande avaldamist turuhind. Järelikult ei turuhind kujune puhtalt informatsiooni enda põhjal, vaid selle põhjal, kui palju on teadaandes uut informatsiooni võrreldes turuosaliste keskmise varasema informatsiooniga, mille põhjal oli hinnatasakaal kujunenud enne teadaannet [47].

### 2.3 Olulisemad majandusindikaatorid

Huang on oma uuringus „*Macroeconomic News Announcements, Financial Market Volatility and Jumps*“ [18] muu hulgas välja toonud USA aktsiaturu ühe suurima ja tähtsaima indeksi S&P500 ja USA võlakirjaturu kõige rohkem mõjutavad majandusindikaatorid, mille avaldamisel turuhind reageerib koheselt ja märgatavalt tõuseb ka volatiilsus. Indeksi S&P500 aktiivsust ja hinnamuutusi vaadeldakse perioodil 1994–2005, kokku oli vaatluse all 2857 kauplemispäeva viieminutilise hinnagraafikul. Iga kauplemispäev jagatakse kas uudise avaldamise päevaks või ilma uudiseta päevaks. Joonis 2 esitab uuringus hõlmatud 26 majandusindikaatorit.

BUSINV:	Business Inventories	ICLM:	Initial Unemployment Claims
CAPA:	Capacity Utilization	INDPRD:	Industrial Production
CCONF:	Consumer Confidence	LDERS:	Leading Economic Indicators
CONST:	Construction Spending	NAPM:	National Association of Purchasing Managers
CPI:	Consumer Price Index	NFPAY:	Nonfarm Payroll Employment
CREDIT:	Consumer Credit	NHOMES:	New Home Sales
DGORD:	Durable Goods Orders	PCE:	Personal Consumption Expenditures
FACORD:	Factory Orders	PERINC:	Personal Income
FFR:	Average Fed Funds Rate	PPI:	Producer Price Index
GDPADV:	GDP Advance	RETLS:	Retail Sales
GDPFIN:	GDP Final	RSXAUT:	Retail Sales excluding Auto
GDPPRE:	GDP Preliminary	TRDBAL:	Trade Balance
HSTART:	Housing Starts	TREBUD:	Treasury Budget

Joonis 2. Loetelu majandusindikaatoritest ja nende lühenditest [18].

Visualiseerimaks turu tüüpilist reageeringut üllatavale majandusindikaatori avaldamisele on Lisas 1 välja toodud graafik kuupäevaga 07.06.1996, kui mittepõllumajandus tööhõive raporti reaalkväärtus oli 340, kuid turu konsensus oli enne teadaannet 170 standardhälbega 56,6.

Uuringus eristatakse üllatavaid teadaandeid, kus majandusindikaatori reaalkväärtus erines oluliselt prognoositud väärtusest, ja teadaandeid, mis vastasid üldiselt prognoosidele. Üllatused standardiseeritakse vastavalt Balduzzi, Eltoni ja Greeni ning Anderseni,

Bollerslevi, Dieboldi ja Vega uuringutele. Kõige selle taustal pandi uuringu tulemusena kokku tabel, mis on käesoleva uurimistöo raames oluline (Joonis 3). Lihtsustamaks järgnevat kirjeldust märgib autor, et antud peatükis mõistetakse „hüppe päeva“ all päeva, kus aktsiaindeks on teinud märkimisväärse hüppe hinnaliikumises.

Announcement	SP
BUSINV	0.181<116>( 1.633)[0.051]*
CAPA	0.162<136>( 1.250)[0.106]
CCONF	0.139<137>( 0.593)[0.277]
CONST	0.104<134>(-0.545)[0.707]
CPI	0.232<138>( 2.982)[0.001]**
CREDIT	0.244<131>( 3.181)[0.001]**
DGORD	0.150<133>( 0.924)[0.178]
FACORD	0.170<135>( 1.476)[0.070]*
GDPADV	0.143< 42>( 0.414)[0.339]
GDPFIN	0.217< 23>( 1.123)[0.131]
GDPPRE	0.091< 22>(-0.468)[0.680]
HSTART	0.176<136>( 1.639)[0.051]*
ICLM	0.179<581>( 3.068)[0.001]**
INDPRD	0.162<136>( 1.250)[0.106]
LDERS	0.094<138>(-0.953)[0.830]
NAPM	0.119<134>(-0.021)[0.509]
NFPAY	0.356<132>( 5.482)[0.000]**
NHOMES	0.118<136>(-0.081)[0.532]
PCE	0.126<127>( 0.189)[0.425]
PERINC	0.125<128>( 0.159)[0.437]
PPI	0.228<136>( 2.873)[0.002]**
RETLS	0.243<136>( 3.200)[0.001]**
RSXAUT	0.250<136>( 3.361)[0.000]**
TRDBAL	0.091< 33>(-0.569)[0.715]
TREBUD	0.138<130>( 0.573)[0.283]
FFR	0.108<120>(-0.386)[0.650]
News	0.169<1949>
No-news	0.120< 908>
Total	0.154<2857>

Joonis 3. Suhe hüppe päeva ja majandusindikaatorite avaldamise päevade vahel [18].

Joonise vasakpoolne veerg koosneb majandusindikaatorite lühenditest. Parempoolne veerg koosneb neljast elemendist, vasakult loetledes on need:

1. hüppe päevade suhe majandusindikaatori avaldamise päevadega;

2. majandusindikaatori avaldamiste päevade arv (või päevade arv, millal ühtegi majandusindikaatorit ei avaldatud), mis on märgitud sümboolite „<“ ja „>“ vahel;
3. sulgudes on t-statistik. T-statistiku nullhüpoteesi aluseks on, et majandusindikaatorite avaldamised ei too kaasa aktsiaindeksi hüppelist hinnaliikumist. Seega on nullhüpotees väide, et hüppe päevade toimumise tõenäosus on võrdne mõlematel päevadel – nendel, kui avaldatakse majandusindikaator, kui ka nendel, millal ei avaldata ühtegi majandusindikaatorit. Kaks täрни tähistab, et olulisuse nivoo on 5%, üks tärn tähistab 10%. Ääremärkusena toob autor välja, et kui t-statistiku väärtus jääb kvantiilide ( $\alpha=5\%$ ,  $\alpha=10\%$ ) vahele, mis on arvatavad  $\alpha/2$  ja  $1-\alpha/2$ , siis tuleb jääda nullhüpoteesi juurde, vastavalt 95% või 90% valimite korral. Kui t-statistiku väärtus jääb kvantiilidest välja, siis võib lugeda tõestatuks alternatiivhüpoteesi. Antud töö raames on alternatiivhüpoteesi alus, et majandusindikaatorite avaldamised toovad kaasa aktsiaindeksi hüppelise hinnaliikumise;
4. Kandilistes sulgudes on ühepoolne p-väärtus. See näitab, et kui nullhüpotees on tõene (ehk kahe grupi vahel pole mingeid erinevusi), siis juhusliku valimi korral saaksime p-väärtusele vastava erinevuse ikkagi. Näiteks kui p-väärtus on 0,02, siis eeldades, et majandusindikaatorite avaldamine ei mõjuta hinnaliikumist, oleks 2% juhtudel ikkagi efekt olemas, aga see efekt tuleks valimi veast. Tegelik eksimine on aga p-väärtusest veelgi suurem.

### **2.3.1 Majandusindikaatorite liigitamine**

Enne töös kasutatavate majandusindikaatorite lähemat vaatlemist, tuleb vaadelda majandusindikaatorite liigitamist. Majandusindikaatoreid saab liigitada kolmeks vastavalt indikaatori tavapärasele ajastusele seoses majandustsüklitega:

1. ennetavad indikaatorid prognoosivad tulevikusündmuse ja tavaliselt muutuvad enne, kui majandus tervikuna muutub;
2. viivitusega indikaatorid muutuvad alles peale seda, kui majandus tervikuna on juba muutunud;
3. kattuvad indikaatorid liiguvad ajalisel samas taktis kui majandus tervikuna.

Järgnevad majandusindikaatorid on valitud Huangi uuringu põhjal, valides sealt viis kõige rohkem S&P500 aktsiaindeksit mõjutavat majandusindikaatorit. Kuna tootjahinnaindeksi (PPI) andmeid ei olnud autoril võimalik piisavas koguses saada, siis autor vahetas varasema kogemuse ja soovitude põhjal tootjahinnaindeksi (PPI) ostujuhtide indeksi (PMI) vastu.

### **2.3.2 Mittepõllumajanduse tööhõive raport (NFP või NFPAY)**

Mittepõllumajanduse tööhõive raport (*Nonfarm Payrolls*) on USA Tööhõivestatistika Büroo uuring, mille eesmärk on kajastada USA töötajate koguarvu igas sektoris, v.a keskvalitsuse töötajad, eramajapidamiste töötajad, mittetulundusühingute töötajad ja põllumajandustöötajad. NFP hõlmab ligikaudu 80% USA sisemajanduse koguprodukti moodustavatest töötajatest. NFP indikaator avaldatakse igakuiselt kuu esimesel reedel [47].

NFP puhul on tegemist kattuva indikaatoriga ning seda peetakse oluliseks majandustingimuste indikaatoriks, sest selles avaldatud uute töökohtade arv liigub väga sarnaselt üldise majandusega ja seda näitajat avaldatakse igakuiselt. Föderaalreserv hindab raportit kõrgelt ka selle detailsuse poolest, sest järeldusi on võimalik teha nii majandussektorite kui ka piirkondade kaupa. Autori arvates on NFP kõige suurem eelis asjaolu, et see avaldatakse kohe kuu alguses esimesel reedel ja seega ilmneb enne paljusid teisi näitajaid.

### **2.3.3 Jaemüügiraport ilma automüügita (RSXAUT)**

Jaemüügiraport ilma automüügita (*Retail Sales Excluding Auto*) on igakuiselt avaldatav indikaator, mis näitab eelmisel kuul müüdud kaupade väärtust USA dollarites jaekaubanduses, v.a automüügis. Enamasti jälgitakse, kui suur protsentuaalne muutus on toimunud võrreldes eelmise kuuga [47].

Tegemist on kattuva indikaatoriga, mis näitab majanduse hetkeseisu. Indikaator on tähtis seetõttu, et avab pildi laiemale majanduslikule olukorrale, sest tarbijate kulutused moodustavad tavaliselt umbes kaks kolmandikku kogu sisemajanduse koguproduktist.

### **2.3.4 Tarbijakrediidiindeks (CREDIT)**

Tarbijakrediidiindeks (*Consumer Credit Index/Report*) on majandusindikaator, mis näitab eelmisel kuul tekkinud võlga, mida Ameerika Ühendriikide inimesed toote või

teenuse ostmisel tekitasid. Põhimõtteliselt on tegemist tarbimislaenuga, millest kõige levinum viis on krediitkaart [6].

Majandusindikaatorina liigitatakse tarbijakrediit viivitustega indikaatorite alla. Liigitus põhineb eeldusel, et tarbijad ei suurenda oma krediidimäära enne, kui nende isiklik sissetulek on piisavalt suur, et suurema laenu teenindamine on jõukohane. Selle eelduse kohaselt suureneb tarbijakrediidimäär alles pärast seda, kui majandus on madalseisust tõusule pööranud. Järelikult kasutatakse tarbijakrediidiindeksit trendi kinnitamiseks [12].

### **2.3.5 Tarbijahinnaindeks (CPI)**

Tarbijahinnaindeks (*Consumer Price Index*) on indikaator, mis iseloomustab tarbekaupade ja -teenuste, nagu transport, toit ja arstiabi, hindade muutust võrreldes eelmise perioodiga. Ameerika Ühendriikides kogub andmeid ja väljastab igakuiselt informatsiooni tarbijahinnaindeksi kohta Tööhõivestatistika Büroo (*The U.S. Bureau of Labor Statistics*) [7].

Tarbijahinnaindeks on kõige levinum inflatsiooni mõõtmise vahend ja seda kasutatakse tihtipeale ka valitsuse majanduspoliitika tõhususe hindamiseks. Sarnaselt tarbijakrediidiindeksiga on ka tarbijahinnaindeksi puhul tegemist viivitusega majandusindikaatoriga, seega kasutatakse seda majandustrendi või -tsükli kinnitamiseks.

### **2.3.6 Ostujuhtide indeks (PMI)**

Ostujuhtide indeks (*Purchasing Managers Index*) on Ameerika Ühendriikide Hankejuhtimise Instituudi (*Institute of Supply Management*) poolt väljastatud äriraport, mis valmib ettevõtetele ja nende juhtidele suunatud küsimustiku tulemusena. Juhtidel küsitakse erinevate ettevõtlast mõjutavate tegurite kohta, kas nende ettevõttes on vastav näitaja võrreldes eelmise kuuga kõrgem, madalam või sama. Uuringu vastused summeeritakse ning koostatakse ostujuhtide indeks [47].

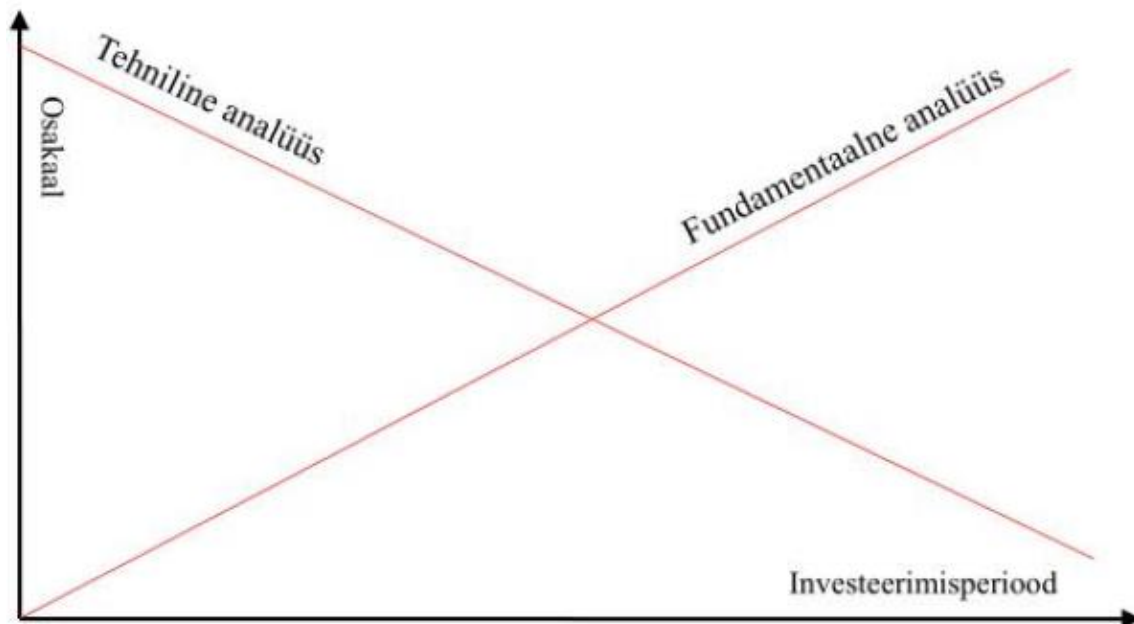
Kuigi eristatakse tootmise ja teeninduse ostujuhtide indeksit, on käesolevas töös kasutatud just esimest. See avaldatakse tavaliselt kuu esimesel tööpäeval, samas kui teeninduse ostujuhtide indeks avaldatakse alles kolmandal tööpäeval. Just see on üks peamisi PMI eeliseid – see väljastatakse teistest indikaatoritest varem. See on

indikaator, mis kujundab investorite ootusi ka teiste majandusindikaatorite suhtes. PMI peetakse ennetavaks majandusindikaatoriks, sest ostujuhid avaldavad muuhulgas informatsiooni ka tuleviku tellimuste suhtes [27].

## 2.4 Tehniline analüüs

Lisaks fundamentaalsele analüüsile kasutatakse laialdaselt ka tehnilist analüüsi. Tehnilise analüüs on meetod, mille eesmärk on prognoosida alusvara hinda tulevikus, võttes aluseks graafilised ja statistilised vahendid alusvara ajalooliste andmete kohta [21]. Tehnilise analüüsi raames ei arvestata eraldi fundamentaalsete näitajatega ehk finants- ja majandusaruannetega, ettevõtte juhtkonnaga jms [21]. Investor usub, et alusvara hindasi mõjutab lõppkokkuvõttes ainult turupsühholoogia.

Reaalsuses kasutatakse parima tulemuse saavutamiseks tehnilist ja fundamentaalset analüüsi koos. Üldlevinud hoiak on, et mida pikem on investeerimishorisont, seda rohkem peab investor arvestama fundamentaalsete näitajatega, sest pikemas perspektiivis kujundab aktsia hind ettevõtte majandustegevuse põhjal (Joonis 4).



Joonis 4. Fundamentaalse ja tehnilise analüüsi osakaal sõltuvalt investeerimisperiodist [40].

Lühiajaliselt on aktsiate hinnad rohkem kauplejate emotsioonidega seotud. See tähendab, et lühiajaliselt pööratakse tehnilisele analüüsile rohkem tähelepanu [21].



#### 2.4.1 Trend ja libisev keskmine

Olenemata sellest, kas tehniline analüüs töötab tuleviku hindade prognoosimiseks või mitte, saab tehnilise analüüsi tulemusena lisainformatsiooni turul toimunud ja turul hetkel toimuva suhtes. Erinevad tööriistad aitavad lihtsustada, visualiseerida ja tõlgendada alusvara hinna ja statistiliste näitajate koosmõju investorile. Eelnevalt välja toodud varasemate uuringute põhjal on majandustsüklitel oluline mõju majandusindikaatorite avaldamisele aktsia- ja võlakirjaturu hinnaliikumise suhtes. Üks lihtsamaid viise, kuidas majandustsüklitest aimu saada, on trendid.

Trend näitab turu üldist suunda ja meelestatust. Aktsiaindeksi pikaajaliste trendide põhjal saab teha järeldusi majanduse üldise käekäigu kohta. Kuidagi tuleb aga trende defineerida, eriti siis kui tegemist on programmiga ja kasutaja nõuab programmitrendist arusaamist. Trendide defineerimiseks võetakse appi alusvara hinna libisevad keskmised (*Moving Average*, edaspidi MA) [50].

Üks võimalus, kuidas trendi määrata, on ühe MA kasutamine. Kui MA tõus on positiivne, siis on tegemist tõusutrendiga, vastasel juhul on tegemist langustrendiga [28]. Sellise lihtsa käsitlemise miinus on asjaolu, et ei arvestata olukordadega, kus reaalselt märkimisväärset trendi ei ole. Kasutades ühe MA asemel kahte MA-d mõlemal erinev periood, on võimalik tuvastada ka olukorrad, kus ei ole tõusu- ega langustrendi. Sellisel juhul on tegemist tõusutrendiga (langustrendiga), kui alusvara hind on lühikese perioodiga MA-st kõrgemal (madalamal). Lühikese perioodiga MA on omakorda kõrgemal (madalamal), kui pika perioodiga MA. Kõikidel muudel juhtudel on tegemist neutraalse trendiga. Deterministliku algoritmi realiseerimisel on kasutatud trendi defineerimiseks kahte MA-d [5].

### 3 Masinõppe teoreetilised alused

Masinõppe on arvutiteaduse valdkond, mille eesmärk on välja töötada algoritme, mis oskaksid teha empiiriliste andmete põhjal otsuseid ja prognoose. Arthur Samuel oli üks esimesi, kes kasutas efektiivselt ära heuristilisi otsingumeetodeid ehk ajalise vahega õppimist [29]. Masinõppe terminite eestikeelsete vastete aluseks on võetud Eesti andmeteaduse masinõppe sõnastik, mis on ainus teemakohane eestikeelne sõnastik [8].

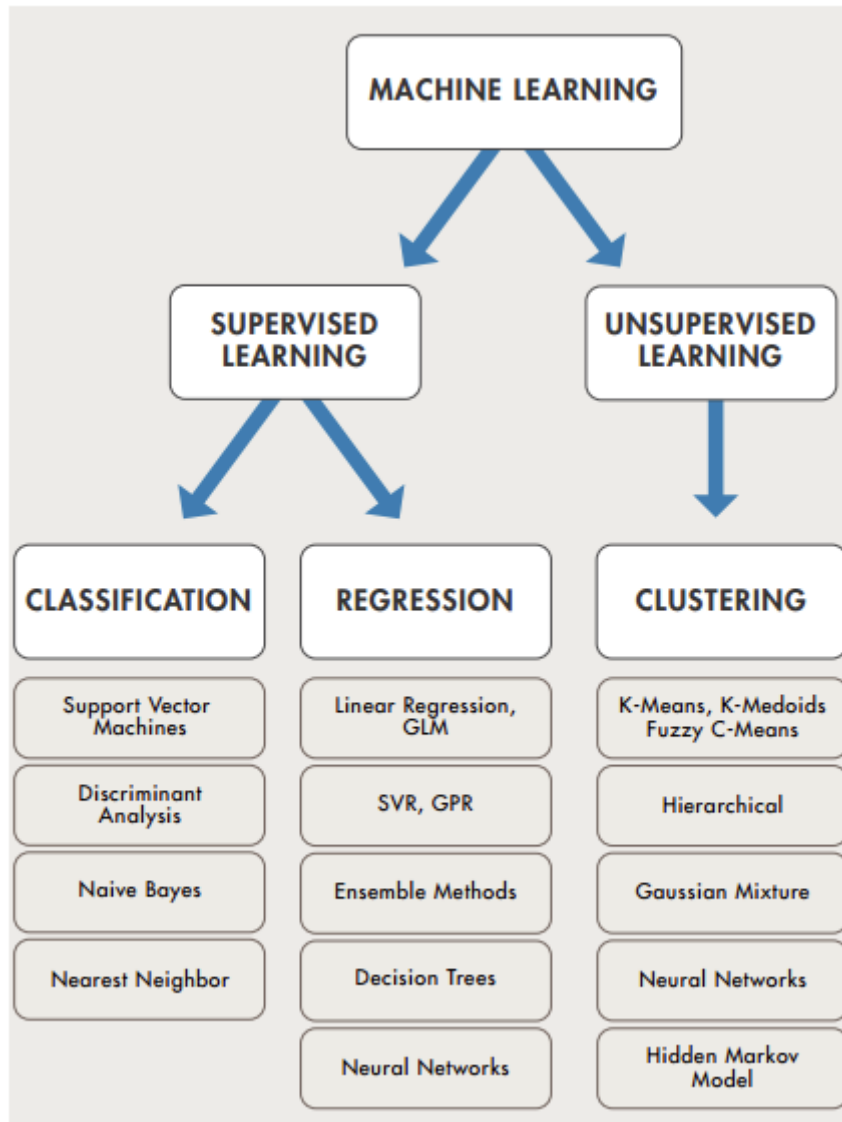
Masinõppe kiiret arengut soodustab asjaolu, et arvutite arvutusvõime on plahvatuslikult kasvanud. Seetõttu kasutatakse praeguseks masinõpet väga paljudes valdkondades. Masinõppel põhinevad tehnoloogiad on näiteks:

- otsingumootorid, mis õpivad, kuidas tuua inimesteni paremaid tulemusi ja suunatud reklaame;
- spämmivastane tarkvara, mis õpib, kuidas filtreerida e-maile;
- automaatsed näotuvastus tarkvarad;
- häälkäskluste tundma õppimise tarkvarad (näiteks Apple Siri);
- autode avarii ärahoidmise süsteemid [46].

Lauri Ülpner on oma magistritöös [46] öelnud: „Enamik masinõppe meetodeid põhineb ideel, et arvuti õpib ülesannet sooritama läbi eelneva õppimise või treeningu.“

Masinõppe liigitatakse peamiselt kaheks suureks grupiks (Joonis 5):

- juhendamisega õpe (*Supervised Learning*) – ennustatakse väärtust, mis põhineb nii sisend- kui ka väljundvektori väärtuste kogumil [35][35].
- juhendamata õpe (*Unsupervised Learning*) – võimeline häälestama oma kaalukoefitsiente lähtudes ainult sisendvektori väärtusest [35][35].



Joonis 5. Masinõppe liigitamine koos kategooriale vastavate algoritmide näidetega [13].

Selles töös arendatav S&P500 aktsiaindeksi hindasid prognoosiv meetod kuulub juhendamisega õppe valdkonda (*Supervised Learning*) regressiooni kategooriasse (*Regression*) ning ennustusmudeli loomiseks kasutatakse tehislikke närvivõrke (*Neural Networks*). Tehisnärvivõrk on arvutuslik arhitektuur, mille aluseks on bioloogilise närvivõrgu toimimise mehhanism. Veel täpsemalt keskendutakse käesolevas töös masin-sügavõppele (*Deep Learning* või *Deep Neural Networks*).

### 3.1 Masin-sügavõpe

Masin-sügavõpe on tehnoloogia, mis kasutab tehisnärvivõrke, et õpetada arvutitele massiivsete andmehulkade kiiret läbitöötamist ja mustrite otsimist [46]. Kõige lihtsamal tehisnärvivõrgul on kokku kolm kihti, mis koosnevad neuronitest. Sellise võrgu

moodustavad varjatud, sisend- ja väljundikiht. Kui varjatud kihte on mitu, siis on tegemist masin-sügavõppel põhineva tehisnärvivõrguga. Närvivõrkude varjatud kihtide neuronites (elementides) asuvad aktivatsioonifunktsioonid, mille ülesandeks on väljundi andmine konkreetsete sisendite ja kaalude puhul [23], [42].

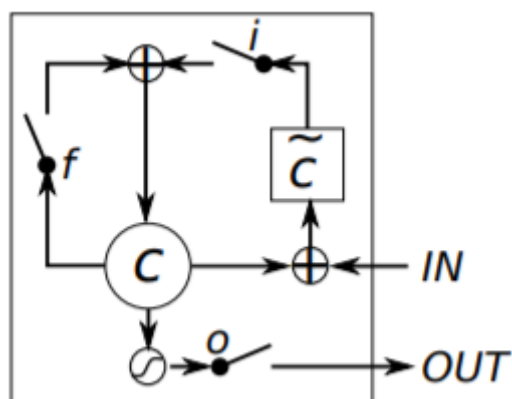
Tüüpiliselt liiguvad andmed sügavate närvivõrkude puhul suunaga sisendkihist väljundkihti ilma eelmiste varjatud kihtide poole pöördumata. Selle tulemusena ei ole tavalisel sügaval närvivõrgul pikaajalisi sõltuvusi ehk puudub mälu. See tähendab, et aegridade ennustamisel ei ole tavalise sügava närvivõrgu kasutamine kõige efektiivsem. Aegridadel põhinevate andmete ennustamisel, kuid ka teksti ja kõne tõlkimisel ja ennustamisel, kasutatakse rekurrentseid närvivõrke (*Recurrent Neural Network*) [9].

## 3.2 Rekurrentsed närvivõrgud

Rekurrentsed närvivõrgud, ka tagasisidestatud närvivõrgud, on masin-sügavõppe alamvariant, milles andmete liikumise suund ei ole rangelt määratud. Rekurrentses närvivõrgus levib signaal sisendist väljundi poole, üles ja alla, kuid ka vastassuunas [35]. Rekurrentne kiht on varjatud kiht, milles iga neuroni väljundid on seotud kõigi selle kihi neuronite sisenditega [35]. Siiski väheneb kiht kihilt neuronite poolt arvatud kaalude mõju ning lõpuks need haihtuvad (*Vanishing Gradient Problem*). See tähendab, et närvivõrk arvestab minimaalselt asjaoludega, mis toimusid kaua aega tagasi. Kaalude säilitamiseks on loodud kontrollitud olekuväravad (*Gate* või *Gated Memory*). LSTM (*Long Short Term Memory*) ja GRU (*Gated Recurrent Unit*) on meetodid, mis kasutavad selliseid väravaid [32], [39].

### 3.2.1 LSTM

Aastal 1997 tutvustasid Hochreiter ja Schmidhuber meetodit nimega LSTM. See tähendab pikka lühiajalist mälu, mis on suuteline õppima ka pikaajalisi sõltuvusi. Erinevalt tavalisest rekurrentsest närvivõrgust, mis kirjutab sisuliselt igal astmel oma eelmise mälu üle, suudab LSTM meetod läbi kontrollitud oleku värava otsustada, kas olemasoleva mälu edasikandmine on vajalik või mitte (Joonis 6) [2], [19].

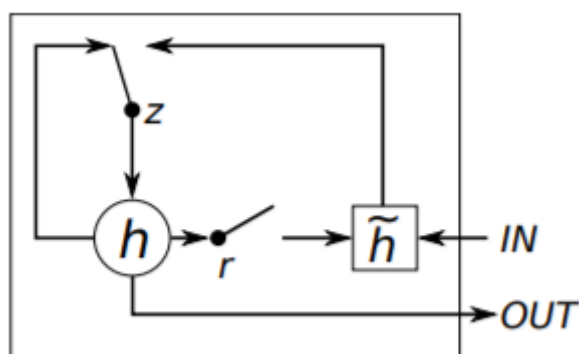


Joonis 6. LSTM-i illustratsioon, kus  $i$ ,  $f$  ja  $o$  on sisendi (*Input Gate*), unustamise (*Forget Gate*) ja väljundi (*Output Gate*) väravad.  $C$  ja  $c$  katusega on vastavalt mäloraku sisu ja uue mäloraku sisu [2].

Jooniselt on näha, et mälu uueneb osade kaupa, olemasoleva mälu unustamise ja uue mälu lisamise tulemusena. Olemasoleva mälu unustamise ulatust moduleerib unustamise värav ning uue mälu lisamist kontrollib sisendi värav. Väljundi värav reguleerib läbi sigmoidfunktsiooni järgmisesse tehisneuronite alamvõrku suunatava mälu osakaalu [2].

### 3.2.2 GRU

GRU tähendab olekuvärvatega varustatud rekurrentset/tagasisidestatud üksust, mille mõtles välja Cho aastal 2014. GRU eesmärk on, et iga rekurrentne üksus suudaks adaptiivselt meelde jätta erinevatel aegridadel põhinevaid sõltuvusi. Sarnaselt LSTM-ile on GRU varustatud olekuvärvatega, mis reguleerivad andmete voogu, kuid GRU-l on üks olekuvärav vähem (Joonis 7) [2], [19].



Joonis 7. GRU illustratsioon, kus  $r$  on taastav värav (*Reset Gate*) ja  $z$  on uuenduse värav (*Update Gate*),  $h$  ja  $h$  katusega on mälorakk ja kandidaat mälorakk [2].

GRU sisaldab kahte mälorakku. Mäloraku sisu (ajahetkel  $t$ ) on lineaarne interpolatsioon eelmise mäloraku sisu (ajahetkel  $t-1$ ) ja kandidaat mäloraku sisu (ajahetkel  $t$ ) vahel. Uuenduse värav otsustab, kui palju mäloraku sisu uuendatakse. Taastav värav otsustab,

kui palju kandidaat mälurakk võtab arvesse kogu GRU närvivõrguga seotud ja edasikanduvat mäluraku sisu [2].

GRU on võrdlemisi uus viis rekurrentsetele närvivõrkudele lähenemiseks ning sellel on mitmeid eeliseid LSTM-i ees. Need tulevad peamiselt sellest, et GRU on LSTM-iga võrreldes vähem kompleksne. Peamine erinevus LSTM-i ja GRU vahel on asjaolu, et LSTM kontrollib mäluraku olekut ise, kuid GRU puhul on kogu mäluraku sisu avatud teistele võrgu sõlmedele. See tähendab, et GRU ei vaja eraldi sisendi ja unustamise väravat (*Input* ja *Forget Gate*), vaid sellega saab hakkama taastav värav (*Reset Gate*) [44]. Üldiselt näitavad uuringud ja arvamused, et GRU sooritus ning arvutuslik efektiivsus on mõnevõrra parem kui traditsioonilisel LSTM-il [3], [10]. Sellest tulenevalt on autor otsustanud S&P500 aktsiaindeksi ennustusmudel loomisel kasutada just GRU-d. GRU väravate aktivatsioonifunktsioonidena kasutab autor Keras poolt vaikimisi määratud hüperboolset tangensfunktsiooni (*tahn Function*) ja lihtsustatud sigmoidfunktsiooni (*Hard Sigmoid Function*).

## 4 Algoritmide realiseerimine

Viimastel aastatel on masinkauplemine muutunud väga populaarseks, moodustades USA aktsiaturu tehingutest enam kui poole, volatiilsetel aegadel isegi kuni 90% [22]. Enamik sellest moodustab kõrgsageduskauplemine (*High-Frequency Trading*), kuid üha rohkem kasutavad ka individuaalsed investorid oma koduarvuteid algoritmkauplemiseks. Seetõttu on ka üha rohkem võimalusi kauplemisalgoritmide realiseerimiseks. Antud töö raames on programmeerimiskeeleks valitud Python, sest selle kohta on internetis õpetavaid materjale kauplemisalgoritmide teostamiseks kõige rohkem. Samuti on Pythonile väga mitmeid erinevaid laiendusi ja teeke, mis on just kauplemisalgoritmide realiseerimiseks mõeldud. Ka masinõppe keerulisemate vormide implementeerimiseks on arendatud Pythonile suhteliselt uued ja võimekad raamistikud, mida autor masinõppel põhineva algoritmi realiseerimisel kasutab.

Käesolevas töös realiseeritakse kaks eraldiseisvat kauplemisalgoritmi. Kuna eesmärgiks on uurida, kas peale üllatava majandusindikaatori avaldamist tekkinud ebatavapärasest aktsiaindeksi hinnareaktsiooni on investoril või kauplejal antud olukorda võimalik ära kasutada, siis deterministlik kauplemisalgoritm keskendub selliste olukordade leidmisele ja positsiooni avamisele. Selleks, et kontrollida, kas deterministliku algoritmi tulemused on olulised, koostab autor masinõppel põhineva kauplemisalgoritmi. Erinevalt deterministlikust algoritmist ei arvesta masinõppel põhinev kauplemisalgoritm eraldi majandusindikaatoritega, analüütikute konsensussega, trendidega jms. Masinõppel põhinev kauplemisalgoritm võtab aluseks ainult ajaloolise aktsiaindeksi hinnaliikumise. See on vajalik, et kahe algoritmi tulemuste võrdlemisel saaks teha järelduse, kas investori või kaupleja tulemuslikkusele aitab kaasa ajalooliste seoste arvestamine olukordadega, kus peale üllatava majandusindikaatori avaldamist on toimunud ebatavaline aktsiaindeksi hinnaliikumine.

## 4.1 Deterministlik algoritm

Algoritm on üles ehitatud eeldusel, et ajalooliste andmete põhjal on võimalik teha järeldusi tulevikus toimuvate „sarnaste sündmuste“ kohta. See eeldab, et „sarnased sündmused“ on teatud kriteeriumite alusel määratud. Mida rohkem on kirjeldavaid kriteeriumeid, seda täpsemini saab sarnaseid sündmuseid määrata. Liiga täpne kirjeldamine jõuab ühel hetkel kriitilise piirini, mille tulemusena tõlgendatakse kõiki mineviku sündmuseid erinevatena. Sellisel juhul ei saa statistilisi järeldusi tuleviku seoste kirjeldamiseks teha. Antud töö raames saab üsna vähe kriteeriumeid sündmuste kirjeldamiseks lisada, sest analüüsi aluseks on üllatavad majandusindikaatorid, mis tähendab, et:

1. kuna majandusindikaatoreid avaldatakse enamjaolt igakuiselt, siis ühe majandusindikaatori kohta on aasta peale 12 vaatlust, kümne aasta peale 120 vaatlust jne;
2. olenevalt üllatuse määramise meetodist väheneb vaatluste arv veel vähemalt poole võrra.

Töös on kasutatud tasulisi vahendeid ajalooliste andmete hankimiseks, sest vastasel juhul oleks andmete hulk liiga väike.

Nagu eelnevalt mainitud, siis varasemates uuringutes on tehtud kindlaks, et majandustsüklitel on kindel mõju üllatava majandusindikaatori avaldamisele ja aktsiate hinnaliikumistele. Seetõttu oleks põhjendamatu nendega mitte arvestada. Vastasel juhul hakkaks keskmine aktsiaindeksi hinnaliikumine üllatava majandusindikaatori avaldamise järel lähenema nullile, eeldusel, et majandustõusu- ja majanduslangusfaase on vaatluse all ajaliselt võrdselt.

Algoritmis eristatakse majandustõusu- ja majanduslangusfaase trendidega, trendi määramise aluseks kasutatakse lühikese ja pika hinnaliikumise libisevate keskmiste ja hinnaliikumise enda kombinatsiooni. Positiivse (negatiivse) trendi korral järeldame, et majandus on antud hetkel tõusufaasis (langusfaasis). Eraldi vaatleme ka juhtumeid, kui trend ei ole märkimisväärselt positiivne ega negatiivne ehk kui trend on neutraalne. Positiivne (negatiivne) trend on siis, kui lühike MA on kõrgemal (madalamal) kui pikk



MA ning kui aktsiaindeksi hind ise on omakorda kõrgemal (madalamal) kui lühike MA. See tähendab sisuliselt seda, et sündmused jagatakse selle kriteeriumi alusel kolmeks.

Samuti on oluline ka asjaolu, kas majandusindikaatori üllatus on positiivne või negatiivne. Ei saa eeldada, et positiivsetel ja negatiivsetel üllatustel ei ole vahet, sest positiivsele üllatusele reageerib turg erinevalt kui negatiivsele üllatusele. Kõige selle tulemusena eristatakse omavahel kuut (6) „sarnaste sündmuste“ komplekti (Joonis 8).

Trend	Positiivne üllatus	Negatiivne üllatus
Positiivne	Olukord 1	Olukord 4
Neutraalne	Olukord 2	Olukord 5
Negatiivne	Olukord 3	Olukord 6

Joonis 8. Vaatluse all olevate sündmuste lahterdamine erinevatesse olukordadesse.

Igale olukorrale leiab algoritm 1, 2, 3, 5 ja 10 kauplemispäeva keskmise tootluse peale üllatava majandusindikaatori avaldamist. Algoritm tunneb ära, millise olukorraga on järgmisena tegemist ning leiab sellele olukorrale kõige suurema keskmise tootlusega perioodi. Vastavalt perioodi pikkusele ja suunale avab algoritm positsiooni.

Joonis 9 esitab deterministliku kauplemisalgoritmi üldistatud loogika. Täpsem pseudokood on esitatud Lisas 3.

```

teekide ja data importimine
muutujate väärtustamine;
data sobivale kujule teisendamine;
ebatavaliste hinnareaktsioonide välja sorteerimine;
iga perioodi (1, 2, 3, 4, 5, 10) korral:
    keskmise ebatavalise hinnareaktsiooni arvutamine vastavalt igale
    olukorrale (1-6);
järgmise tehingu pikkus on vastava olukorra periood, mille korral
keskmise ebatavalise hinnareaktsiooni absoluutväärtus on kõige suurem;
tulemuste kuvamine;

```

Joonis 9. Deterministliku kauplemisalgoritmi üldistatud loogika. Täpsemalt Lisas 3.

Deterministliku kauplemisalgoritmi realiseerimiseks piisab Pythoni tekidest NumPy, pandas ja matplotlib kasutamisest. Pythonit ja kasutatavaid teeke tutvustatakse peatükis 4.3.

## 4.2 Masinõppel põhinev algoritm

Masinõppel põhineva kauplemisalgoritmi realiseerimiseks on kasutatud rekurrentseid tehisnärvivõrke, täpsemalt GRU-d. Järgnevalt kirjeldatakse algoritmi andmestikku, tööpõhimõtet, arhitektuuri ja hüperparameetrite valikut.

Sisendandmete saamiseks CSV formaadis on kasutatud Yahoo Finance'i portaali ning S&P500 aktsiaindeksi hinnaliikumised ulatuvad kuni aastani 1970. Enne mudelisse sisendandmete suunamist tuleb andmestik sobivasse formaati viia. Andmestik tuleb normaliseerida, et mudel oskaks ka ajaloolisi seoseid tänapäevaga kohandada, sest andmestiku skaala on ajas muutuv – aastal 1975 on aktsiaindeksi hind 65\$ tasemel ning aastal 2018 on hind 2700\$. Seejärel koostatakse hulk erinevad massiive, mis kirjeldavad sisend- ja väljundandmeid vastavalt nende dimensioonidele. Massiivid talletatakse binaarses HDF5 formaadis, mis muuhulgas võimaldab suurte massiivide kirjutamist ja lugemist kiiremini.

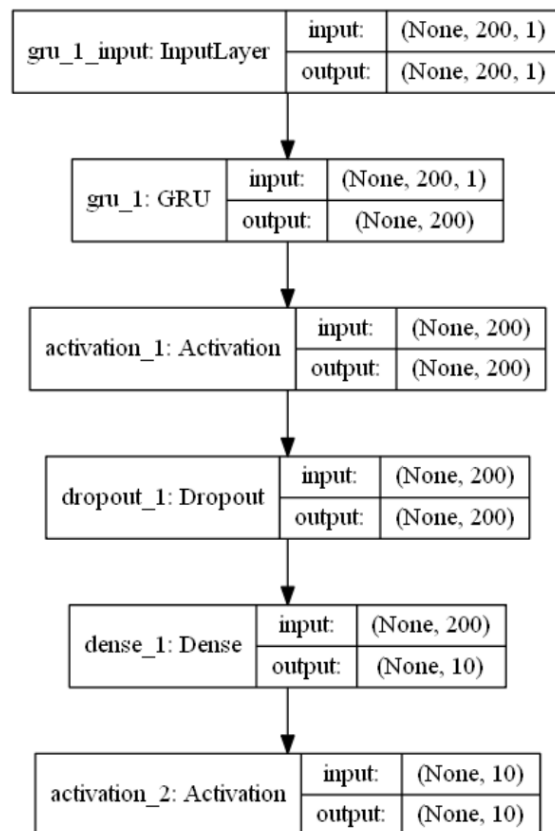
Ennustusmudelisse laetakse treenimise käigus valideerimisandmestikul kõige paremini toimunud kaalud ning mudel arvutab testandmete peal prognoositud aktsiaindeksi hinnaliikumised. Testandmed on treeningprotsessist täielikult eraldatud ning kaalude arvutamisel nendest ei lähtuta. Tulemus visualiseeritakse joon- ja hajuvusdiagrammil, lisaks arvutatakse keskmine ruutviga ja keskmine absoluutne protsentuaalne viga (*Mean Absolute Percentage Error*).

Algoritm realiseeritakse Pythonis ning kasutatavad keskkonnad ja teegid on NumPy, pandas, matplotlib, TensorFlow, Keras, scikit-learn ja h5py. Pythonit ja kasutatavaid teke tutvustatakse peatükis 4.3.

### 4.2.1 Arhitektuur

Masinõppel põhineva algoritmi puhul on tegemist kahekihilise ennustusmudeliga – 200 varjatud üksusega GRU kiht ja täissidus kiht (*Dense Layer*), milles asub 10 (väljundi mõõtmed) neuronit. Kihtide arvu valikul on arvestatud varasemate uuringutega ja lähtutud arusaamast, et kahekihiline mudel on aegriksel põhineva hinnamuutuste ennustamiseks piisav [16], [17]. Samuti toetavad antud lähenemist Stone-Weierstrassi teoreem ja Sontag'i teoreem [35]. Sisendsõlm on osa GRU kihist ning peale seda kasutatakse hüperboolset tangensfunktsiooni (*tahn Function*) x- ja y-telje sõltuvuste kirjeldamiseks. Seejärel rakendatakse dropout-meetodit (*Dropout Technique*), et vältida

ülesobitumist (*Overfitting*) [43]. Dropout-meetod lülitab juhuslikult treenimise faasis neuroneid välja. Rekurrentsete närvivõrkude puhul soovitatakse üldiselt dropout-meetodi väärtus hoida 20% ja 40% vahel [4]. Autor leiab täpse väärtuse optimeerimise käigus, kuid lähtub soovituselt hoida see väärtus eelmainitud vahemikus. Edasi ühendatakse esimesest kihist tulevad andmed teise kihi 10-ne neuroniga. Peale varjatud kihti rakendatakse mittenegatiivset lineaarfunktsiooni (*Rectified Linear Unit/Function*) ja andmed suunatakse väljundsõlme. Ennustusmodeli kompileerimisel kasutatakse optimeerijana Adam-it (*Adaptive Moment Estimation*) ja veafunktsioonina keskmist ruutviga (*Mean Squared Error*). Joonis 10 esitab masinõppel põhineva algoritmi arhitektuuri, mille genereerimiseks on kasutatud teeki pydot ja graphviz.



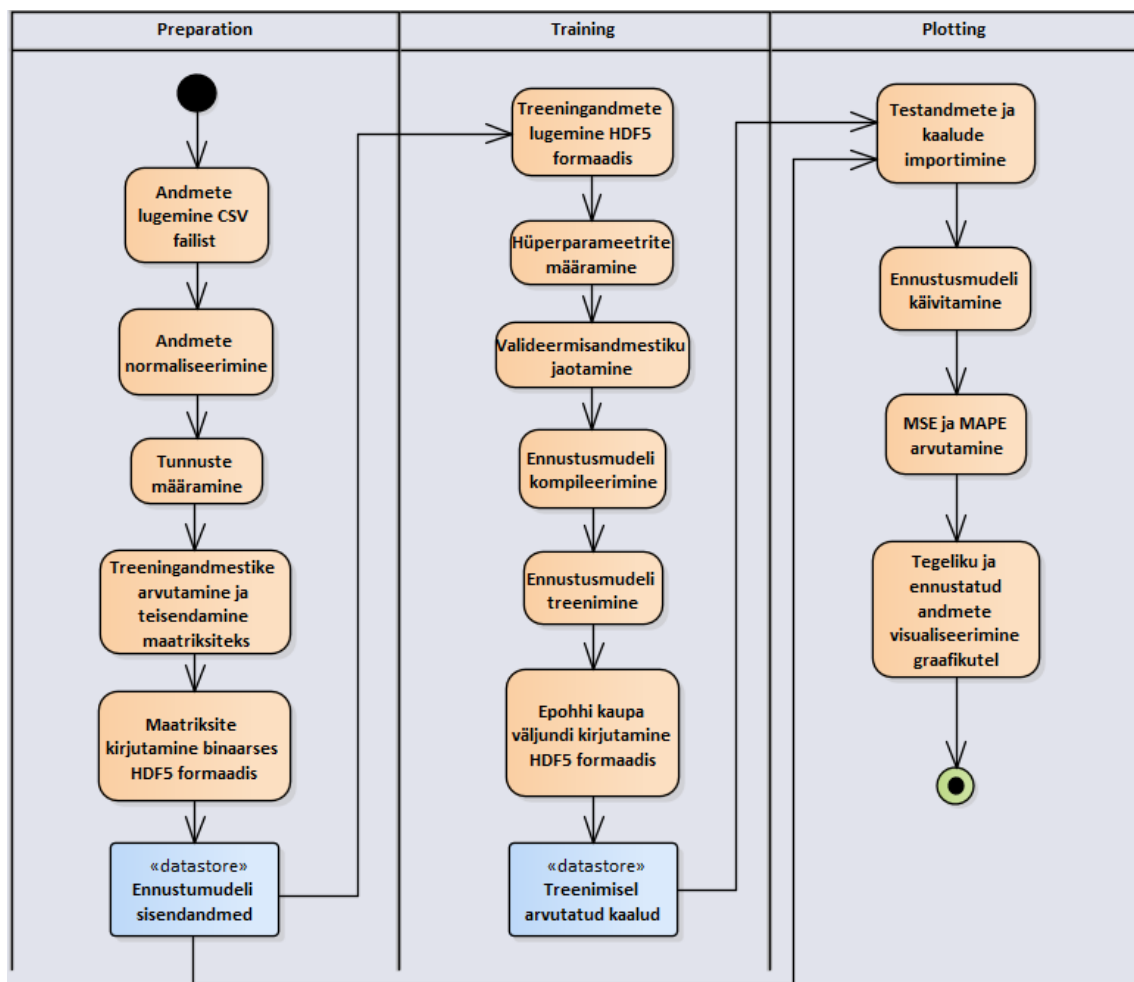
Joonis 10. Masinõppel põhineva algoritmi arhitektuur.

Algoritm on realiseeritud kolmes osas:

1. *preparation* – ajalooliste aktsiaindeksi hindade lugemine CSV failist, andmete normaliseerimine (skaleerimine) scikit-learn teegi abiga [45], tunnuste määramine, treeningandmestike arvutamine ja teisendamine maatriksiteks, ennustusmodeli sisendandmete kirjutamine binaarses HDF5 formaadis;

2. *training* – ennustusmodeli sisendandmete lugemine HDF5 formaadis, GRU üksuste, treeningploki suuruse (*Batch*) ja epohhide (*Epoch*) määramine, valideerimisandmestiku jaotamine, ennustusmodeli kompileerimine ja treenimine, epohhi kaupa väljundi kirjutamine HDF5 formaadis, millele on lisatud ennustusmodeli täpsus valideerimisandmestiku suhtes;
3. *plotting* – ennustusmodelisse testandmete ja treenimisel arvatud kaalude importimine, ennustusmodeli käivitamine, tegeliku ja ennustatud andmete visualiseerimine graafikutel.

Joonis 11 esitab masinõppel põhineva algoritmi tegevusdiagrammi.



Joonis 11. Masinõppel põhineva algoritmi tegevusdiagramm.

Masinõppel põhineva kauplemisalgoritmi pseudokood on esitatud Lisas 4.

## 4.2.2 Hüperparameetrid

Sügav-masinõppel põhineva kauplemisalgoritmi hüperparameetrid leiti optimeerimise tulemusena. Optimeerimisel muudetavad suurused olid sisendi suurus, GRU kihi varjatud üksuste arv, dropout-meetodi väärtus, treeningploki suurus ja epohhide arv. Kõikide suuruste muutmine toimus autori personaalarvuti võimete piires. Epohhide arvu valimisel lähtus autor varajase lõpetamise tehnikast (*Early Stopping Technique*), et vältida ala- ja ülesobitumist. Treeningploki suurus soovitatakse hoida number kahe astmena, sest CPU ja GPU mälu arhitektuur tavaliselt organiseerib mälu kahendastmena. Autor katsetab optimeerimisel väärtusi 128, 256 ja 512, sest liiga väike treeningploki suurus võib põhjustada suurel määral müraga arvestamise ning liiga suur treeningploki suurus vähendab ennustusmodeli kvaliteeti [43]. Sisendi ja GRU kihi varjatud üksuste arvu suuruse peamiseks piiranguks on autori personaalarvuti arvutusvõimsus.

Ennustusmodeli aluseks võetakse kaalud, mis treenimise käigus valideerimisandmestiku peal kõige paremat tulemust näitasid. Tulemusi mõõdeti keskmise ruutveaga ning optimeerimise käigus leiti, milliste mudeli parameetritega keskmine ruutviga on valideerimisandmestiku peal kõige väiksem. Optimeerimise tulemusel selgus, et vähima keskmise ruutveaga tulemus saadi parameetrite kasutamisel, mis on välja toodud järgnevas tabelis (Tabel 1).

Tabel 1. Optimeerimise käigus leitud parameetrite väärtused.

Parameeter	Väärtus/Suurus
Sisendi suurus	200
GRU kihi varjatud üksuste arv	200
Dropout-meetodi väärtus	30%
Treeningploki suurus	128
Epohhide arv	128

Optimeerimine näitas, et antud mudeli täpsust mõjutavad eelkõige treeningploki suurus ja GRU kihi varjatud üksuste arv. Treeningploki suurus määrab treeningvalimite arvu, mille järel toimub kaalude ümberarvutamine. Seeläbi määrab treeningploki suurus ka iteratsioonide arvu, mis on vajalik ühe epohhi täitmiseks. GRU kihi varjatud üksuste arv on otseselt seotud närvivõrgu õppevõimega, mis avaldub õpitud kaalude massiivi

suuruses. Mida rohkem üksusi kasutada, seda paremini suudab närvivõrk sõltuvusi meelde jätta. Ka üksuste arvu kasutamisel on oma piir, et vältida ülesobitumist. Antud töö raames oli piiriks hoopis treenimise ajaline kulu, mis on otseselt seotud autori personaalarvuti sooritusvõimega. Mida rohkem üksusi kasutada, seda suurem on treenimise ajakulu ning seda rohkem arvutusvõimsust on vaja. Dropout-meetodi väärtuse muutmine vahemikus 30%–40% ei avaldanud liiga suurt mõju tulemuste täpsusele, kuid avaldas nähtavat mõju treenimise ajalisele kulule – mida suurem on dropout-meetodi väärtus, seda kiirem on treenimise protsess. Sisendi suurus mõjus mudeli täpsusele negatiivselt, kui selle väärtus oli väiksem kui 200. Sisendi suuruste 200–600 korral püsis mudeli täpsus üsna sarnasel tasemel. Sisendi suurendamine mõjus treenimise ajakulule samuti negatiivselt, väärtusi üle 600 ei olnud autori personaalarvuti peal enam mõistliku ajakuluga võimalik katsetada.

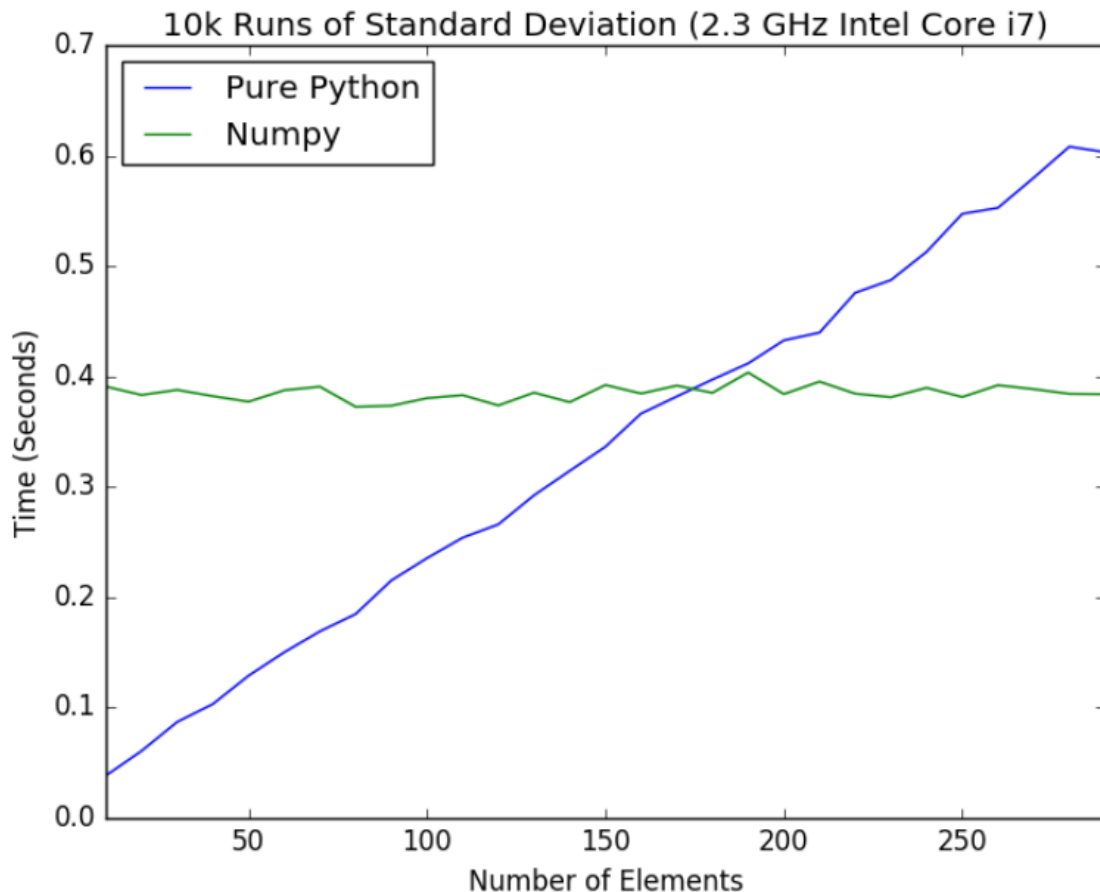
### 4.3 Python

Python on üldotstarbeline interpreteeritav programmeerimiskeel, mis võimaldab mitut programmeerimisstiili, näiteks objektorienteeritud, protseduraalset või funktsionaalset programmeerimist. Pythonit peetakse küllaliski lihtsaks keeleks ja seda on soovitatud programmeerimise õppimisel esimeseks keeleks [37].

Antud töö raames võib probleemiks osutada, et Python on suhteliselt aeglane ja kohmakas suurte andmemasside töötlemisel, kuid selle lahendavad Pythoni teegid NumPy ja pandas [31].

#### 4.3.1 Numpy

NumPy (*Numeric Python*) on programmeerimiskeele Python laiendus, mis on mõeldud andmemassiivide haldamiseks ja töötlemiseks. NumPy kirjutatud programmeerimiskeeles C, sest tegu on kiireima andmemassiive töötleva programmeerimiskeelega. NumPy kasutajate jaoks on oluline andmemassiiv (*Array*), mis sarnaneb Pythonisse sisseehitatud järjendi (*List*) andmestruktuuriga. Andmemassiiv võib olla n-dimensiooniline, näiteks ühedimensiooniline (1D) või kahedimensiooniline (2D). Puhta Pythoni ja NumPy andmemassiivide läbitöötamise aja sõltuvust andmemassiivi suurusest iseloomustab allolev joonis (Joonis 12. X-teljel on andmemassiivi kuuluvate elementide number ja y-teljel andmemassiivi läbitöötamise aeg) [34].



Joonis 12. Andmemassiivi läbitöötamise aja võrdlus puhta Pythoni ja NumPy kasutamisel [14].

Koos NumPyga kasutatakse andmete visualiseerimiseks populaarset ja võimalusterohket teeki nimega matplotlib. matplotlib moodul nimega pyplot pakub MATLAB-ile sarnaseid võimalusi ja liidest. Erinevalt MATLAB-ist on Python ja matplotlib avatud lähtekoodiga ja kasutamine neid saab kasutada tasuta [34].

#### 4.3.2 Pandas

Pythonit kasutatakse edukalt andmete korrastamiseks ja ettevalmistamiseks töötluse eel, kuid selle kasutamine andmeanalüüsiks ja modelleerimiseks on raskendatud. Tihtipeale tuleb ülesannete lahendamiseks kasutada vahendeid, mis peegeldavad antud valdkonna iseärasusi, näiteks programmeerimiskeel R. Pandas (*Python Data Analysis Library*) likvideerib suures ulatuses antud puuduse ning annab võimaluse viia läbi paljud andmeanalüüsi etapid Pythonis. Pandas on üles ehitatud NumPy, millest eelnevalt juttu oli [38].

Pandas võimaldab kasutada spetsiaalseid andmestruktuure ja operatsioone, mis on mõeldud andmetabelite ja aegridade töötlemiseks. Pandase võimalus ja mugavus

aegridade analüüsimiseks ja manipuleerimiseks teeb teegist sisuliselt kohustusliku osa kauplemisalgoritmide kasutamisel programmeerimiskeeles Python [38].

### 4.3.3 TensorFlow

TensorFlow on avatud lähtekoodiga tarkvarateek, mis on suunatud suure jõudlusega numbrilisteks arvutusteks. Selle paindlik arhitektuur võimaldab viia läbi arvutusi erinevatel ja mitmel CPU-l ja GPU-l või ka serveris ja nutiseadmes. TensorFlow on arendatud Google teadlaste ja inseneride poolt, kes töötasid Google Brain Teamis masinintelligentsi osakonnas eesmärgiga viia läbi masinõppe ja sügav-närvivõrkude uuringut, kuid süsteem on piisavalt üldine, mistõttu saab seda kasutada ka teistes valdkondades [1].

Vaikimisi kasutab TensorFlow arvutamiseks CPU-d, kuid autor proovis TensorFlow'd seadistada ka GPU jaoks. Selleks tuli installida Nvidia CUDA ja cuDNN (Nvidia *CUDA Deep Neural Network Library*). Selle tulemusena autor veendus, et antud probleemi lahendamisel GPU kasutamisest kasu ei ole, CPU peal viib TensorFlow oluliselt kiiremini arvutused läbi. GPU kasutamist võib kaaluda, kui on tegu väga spetsiifiliste ja keeruliste närvivõrkude või ka suurte dimensioonidega andmestikega.

### 4.3.4 Keras

Keras on kõrghetasemeline närvivõrkude API. See on kirjutatud programmeerimiskeeles Python ning kasutab taustaprogrammina kas Theanot, CNTK-d või TensorFlow'd. Kerase eesmärk võimaldada kiiret eksperimenteerimist sügavate närvivõrkudega ning fookus on kasutajasõbralikkusel, modulaarsusel ja laiendamise lihtsusel. Keras sisaldab erinevaid närvivõrgu osi implementeerivaid tükke nagu kihid, eesmärgid, aktivatsioonifunktsioonid, optimeerijad ning hulgaliselt muid tööriistu, mis on abiks pildi- ja tekstitöötlemisel [24].

### 4.3.5 Scikit-learn ja h5py

Scikit-learn on Pythoni teek, mis on keskendunud masinõppe arendamisele, sisaldades erinevaid regressiooni, klassifitseerimise ja klasterdamise algoritme, samuti tugivektorklassifitseerijaid ja otsustusmetsasid. Antud töö raames kasutab autor scikit-learn teeki masinõppe algoritmis andmete skaleerimise ja normaliseerimise eesmärgil ning tulemuse keskmise ruutvea arvutamisel [41].



H5py on Pythoni binaarse HDF5 andmevormingu liides, mis võimaldab salvestada tohutul hulgal numbrilisi andmeid. Neid samu andmeid on võimalik töödelda ja käsitleda NumPyga. Samuti võimaldab h5py salvestada tuhandeid andmekogusid ühte faili, kus neid saab vajadusel liigitada ja märgistada. H5py omab veel eeliseid, kuid käesoleva töö raames kasutatakse h5py teeki just eelnevalt mainitud põhjustel [15].

## 5 Analüüs ja arutelu

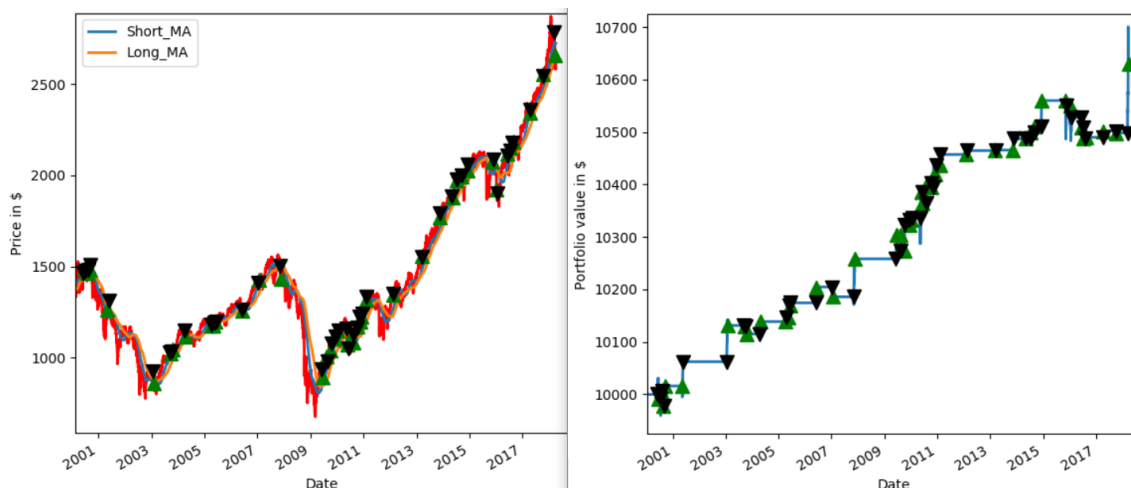
Käesolevas peatükis tutvustatakse kõigepealt algoritmide tulemusi eraldi ning peale seda võrreldakse kahte kauplemisalgoritmi omavahel. Portfellis on alghetkel 10000\$ ja positsiooni avamisel ostetakse või müüakse üks ühik S&P500 aktsiaindeksit.

### 5.1 Deterministliku algoritmi tulemused

Deterministliku kauplemisalgoritmi eesmärk on kasutada ära olukordi, kus peale üllatava majandusindikaatori avaldamist on toimunud ebatavapärane S&P500 aktsiaindeksi hinnaliikumine. Uurimise aluseks on võetud majandusindikaatorid, mis varasemate empiiriliste uuringute tulemusena mõjutavad S&P500 aktsiaindeksi hinnaliikumist enim. Nendeks on mittepõllumajanduse tööhõive raport (NFP, NFPAY), jaemüügi raport, v.a automüük (RSXAUT), tarbijakrediidiindeks (CREDIT), tarbijahinnaindeks (CPI) ja ostujuhtide indeks (PMI). Vastavalt igale majandusindikaatorile on viidud läbi kauplemisalgoritmi parameetrite optimeerimine, tulemused on välja toodud iga indikaatori jaoks eraldi. Järgnevatel joonistel on vasakpoolse ja parempoolse graafiku x-teljel kujutatud aega, y-teljel on vastavalt S&P500 aktsiaindeksi hind ja portfelli väärtus.

#### 5.1.1 NFP

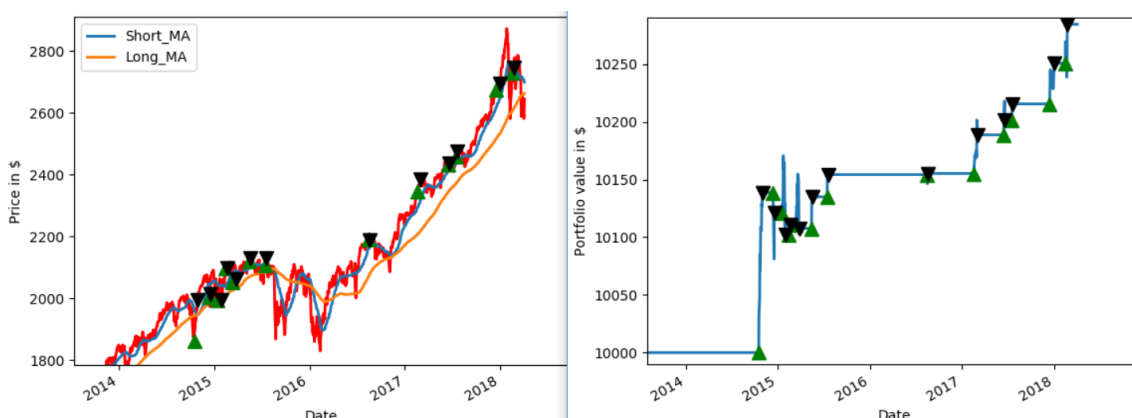
Optimeerimise tulemusena on selgelt näha, et lõplik portfelli seis on seda parem, mida väiksem on eelnevate üllatavate NFP avaldamiste hulk, edaspidi aken (*Sliding Window*), keskmise aktsiaindeksi hinnareaktsiooni arvutamisel. NFP puhul on tegemist kattuva indikaatoriga ning kauplemisalgoritm annab parimad tulemused siis, kui akna väärtus oli 20, mis on ühtlasi ka kõige väiksem väärtus, mida optimeerimisel kasutati. Lühikese ja pika MA väärtused on vastavalt 70 ja 130 päeva. Kokku tehti 42 tehingut, millest ligi 65% olid lühikeseks müügid ning keskmine tehingu tootlus oli 1,1%. Keskmine tehingu pikkus oli 6,2 päeva. Joonis 13 esitab tulemuse graafikutel.



Joonis 13 Deterministliku kauplemisalgoritmi parim tulemus NFP puhul. Vasakpoolne graafik iseloomustab positsiooni avamise ja sulgemise kohti S&P500 aktsiaindeksi hinnaliikumise suhtes. Parempoolne graafik iseloomustab portfelli suurust ajas. Roheline ja must kolmnurk tähistavad vastavalt ostu- ja müügitehinguid.

### 5.1.2 RSXAUT

RSXAUT puhul on tegemist kattuva indikaatoriga, mis andis kauplemisalgoritmis parimad tulemused, kui akna suuruseks oli 60, lühikese ja pika MA periood oli vastavalt 30 ja 120–180 päeva. Kusjuures parimate tulemuste portfelli lõppseis on sama, see tähendab, et kauplemisalgoritm käitus tehingute tegemisel täpselt sama moodi, isegi kui pikk MA periood muutus suhteliselt suure ulatuses (120–180 päeva). Kokku tehti 13 tehingut, millest ükski ei olnud lühikeseks müük ning keskmine tehingu tootlus oli rohkem kui 1%. Keskmine tehingu pikkus oli 5,8 päeva. Joonis 14 esitab tulemuse graafikutel.

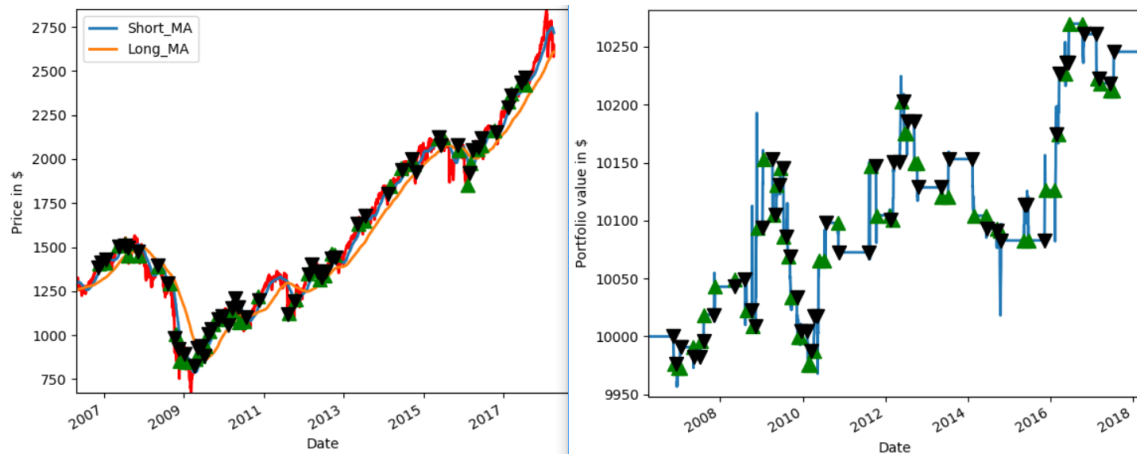


Joonis 14. Deterministliku kauplemisalgoritmi parim tulemus RSXAUT puhul.

### 5.1.3 CREDIT

CREDIT puhul on tegemist viivitusega indikaatoriga, mis andis kauplemisalgoritmis parimad tulemused, kui akna suurus oli 40, lühike ja pikk MA olid vastavalt 50 ja 180–

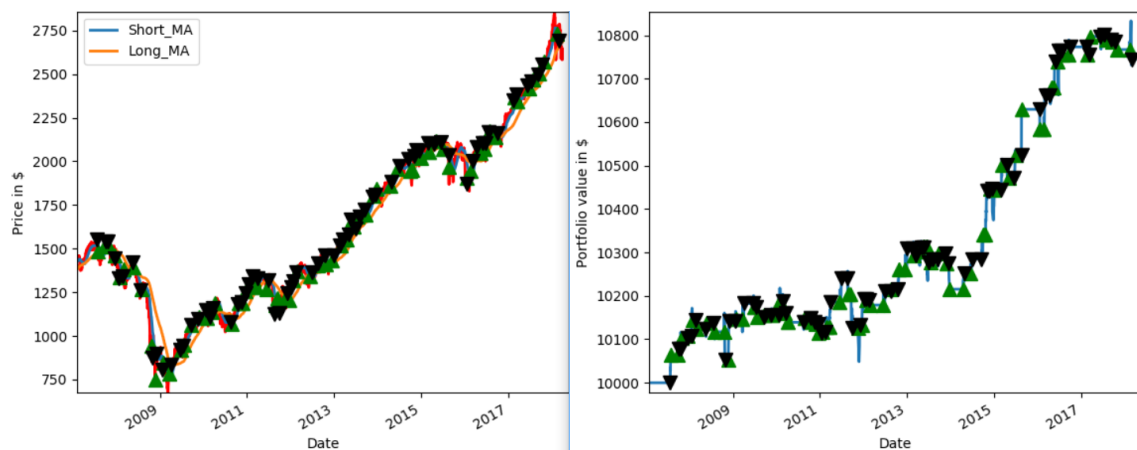
200 päeva. Kokku tehti 53 tehingut, millest 60% olid lühikeseks müügid ning keskmine tehingu tootlus oli 0,3%. Keskmine tehingu pikkus oli 6,1 päeva. Joonis 15 esitab tulemuse graafikutel.



Joonis 15. Deterministliku kauplemisalgoritmi parim tulemus CREDIT-i puhul.

### 5.1.4 CPI

CPI puhul on tegemist viivitusega indikaatoriga, mis andis kauplemisalgoritmis parima tulemuse, kui akna suurus oli 80, mis on ühtlasi ka kõige suurem väärtus, mida optimeerimisel kasutati. Lühike ja pikk MA olid vastavalt 30 ja 110 päeva. Kokku tehti 70 tehingut, millest ligi 70% olid lühikeseks müügid ning keskmine tehingu tootlus oli 0,65%. Keskmine tehingu pikkus oli 7,1 päeva. Joonis 16 esitab tulemuse graafikutel.

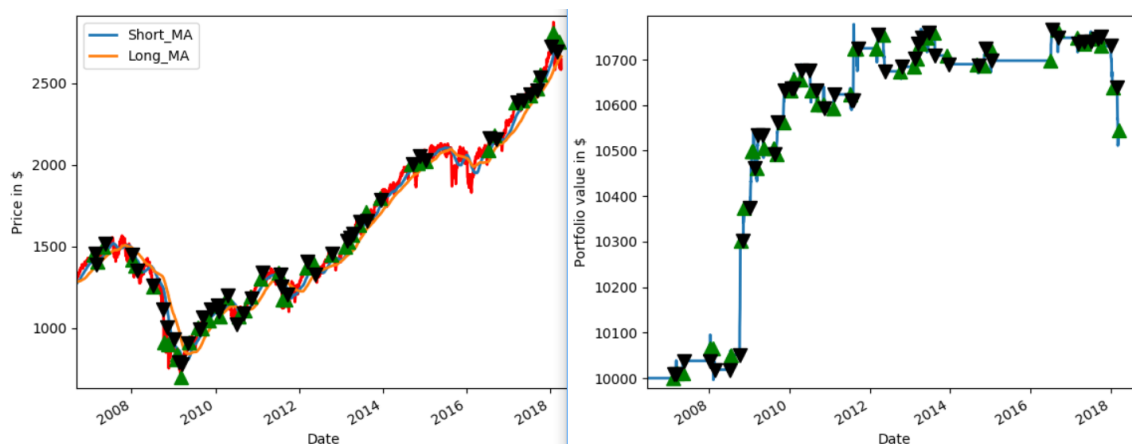


Joonis 16. Deterministliku kauplemisalgoritmi parim tulemus CPI puhul.

### 5.1.5 PMI

PMI puhul on tegemist ennetava indikaatoriga, mis andis kauplemisalgoritmis parima tulemuse, kui akna suurus oli 50, lühike ja pikk MA olid vastavalt 60 ja 130 päeva. Kokku tehti 46 tehingut, millest ligikaudu 35% olid lühikeseks müügid ning keskmine

teingu tootlus oli rohkem kui 1,3%. Keskmise teingu pikkus oli 8,2 päeva. Joonis 17 esitab tulemuse graafikutel.



Joonis 17. Deterministliku kauplemisalgoritmi parim tulemus PMI puhul.

### 5.1.6 Järeldus

Tulemustest selgub, et lühikeseks müüke toimus rohkem kui ostmisi ning teades, et S&P500 aktsiaindeks on enamik ajast ülestrendis, saame teha kaks järeldust:

1. turu „valesti“ reageerimine üllatavale majandusindikaatori avaldamisele ei ole juhuslik ning olukorda saab ka trendisuuna vastu ära kasutada;
2. turu „valesti“ reageerimine üllatavale majandusindikaatori avaldamisele võib tähendada signaali, mis viitab korrigeerimisele, millest vähemalt lühiajaliselt on võimalik kasu lõigata. Näiteks olukord, kui turg ei reageeri üllatavalt heale majandusindikaatorile enam tõusuga, vaid hoopis langusega, siis see annab signaali, et tuleb korrigeerimine. Keskmisest teingu pikkusest võib ka järeldada, et keskmine korrigeerimise kestus on 6–7 päeva.

Tulemustest hakkas ka silma, et NFP puhul on akna pikkus minimaalne (20) ning CPI puhul maksimaalne (80). Erinevus võib olla tingitud sellest, et NFP on kattuv indikaator ning CPI on viivitusega indikaator. Kattuvad indikaatorid liiguvad ajaliselt samas taktis kui majandus tervikuna ning antud kauplemisalgoritmi tulemuste järgi on kõige tõhusam arvutada keskmine hinnaliikumine just lähiajal toimunud üllatavate NFP avaldamiste põhjal. See sisuliselt tähendab, et kattuva indikaatori puhul ei ole mõistlik vaadata kaugeid ajaloolisi seoseid, vaid tuleks lähtuda pigem majanduse hetkeseisust ja selle lähiajaloost. Viivitusega indikaatorid muutuvad alles peale seda, kui majandus tervikuna on muutunud ning neid kasutatakse trendide kinnitamiseks. Ilmneb, et

viivitusega indikaatori puhul on mõistlik võtta vaatluse alla ka kaugemad ajaloolised seosed üllatava indikaatori avaldamise ja aktsiaindeksi hinnaliikumise vahel. Järelikult üllatavate trende kinnitavate indikaatorite ja sellest tingitud aktsiaindeksi hinnaliikumist on võimalik kirjeldada teatud mustrite abil, mis põhinevad pikematel ajaloolistel seostel. Sama väide aga kattuvate indikaatorite korral ei kehti.

Tuleb meeles pidada, et akna suurus, lühikese ja pika MA perioodid on optimeerimise tulemusel leitud. See tähendab, et kauplemisalgoritm ei pruugi olla robustne ning ei ole kindel, et samad parameetrid ka tulevikus töötavad.

Käesolevaid tulemusi on võimalik tulevastel uurimustes mitmel moel edasi arendada. Võimalusel võiks uurida pikemat ajaperioodi, kui tekib võimalus pääseda ligi rohkematele andmetele seoses majandusindikaatorite avaldamistega. Samuti võiks sama meetodikaga uurida ka teisi finantsturge, näiteks valuuta-, toorme- ja võlakirjaturge. Lisaks saab analüüsida teisi majandusindikaatoreid ja teisi aktsiaindekseid, ka regionaalselt. Veel rohkem süvitsi minnes on võimalik uurida, milline on teatud perioodi vältel avaldavate majandusindikaatorite koosmõju alusvara hinnaliikumisele ja kas investoril või kauplejal on võimalik sellest kasu lõigata.

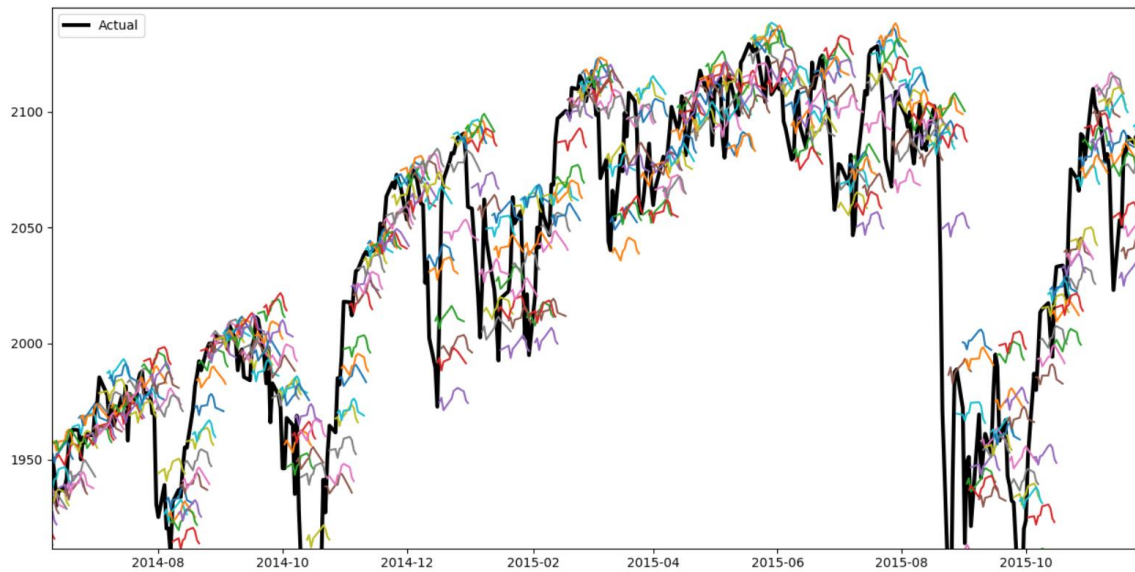
## **5.2 Masinõppel põhineva algoritmi tulemused**

Käesolevas alapeatükis ei tooda eraldi välja masinõppel põhineva kauplemisalgoritmi tulemusi tehingute tootlikkuse põhjal, vaid tutvustatakse algoritmi ennustuse täpsust kirjeldavaid tulemusi.

Ennustusmudel prognoosib igal sammul (päev) järgmise 10 päeva hinnaliikumist, mis tuleneb sellest, et deterministlikus algoritmis kasutati maksimaalse tehinguperioodina 10 päeva. See tähendab, et ka masinõppel põhineval kauplemisalgoritmis on maksimaalne tehinguperiood 10 päeva. Prognoosi aluseks võetakse sisendi suurus ehk 200 päeva hinnaliikumine. Optimeerimise käigus leitud hüperparameetrid on esitatud peatükis 4.2.2 (Tabel 1).

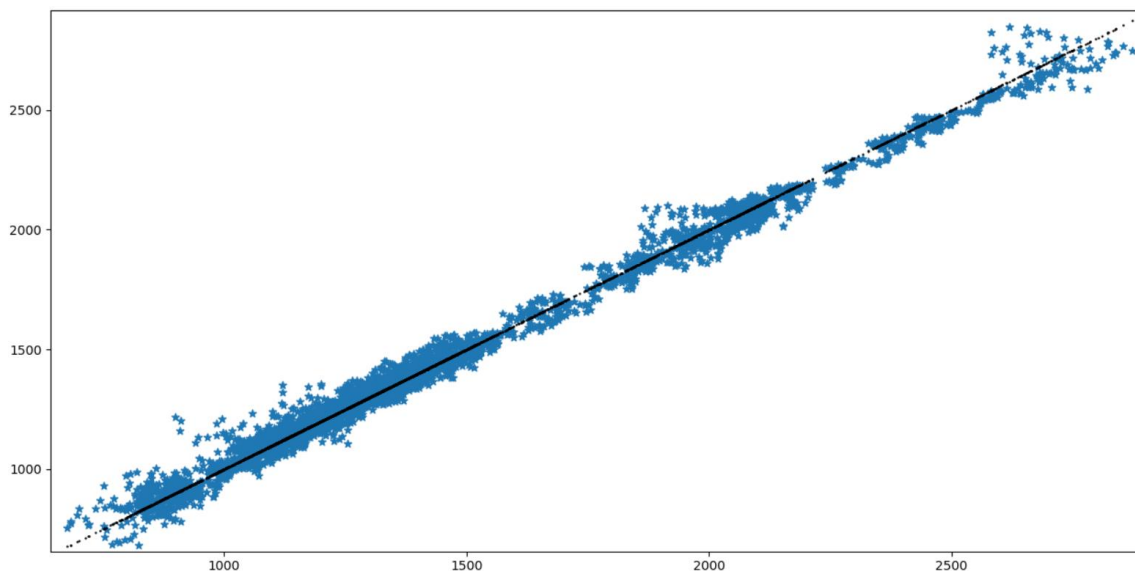
Ennustusmudeli treenimiseks kasutati 7728 rida andmeid, millest 20% moodustas valideerimisandmestik. Parameetrite arv oli 123210 ning kõik neist olid treenitavad. Testandmeid, mille peal ennustusmudel hinnaliikumisi prognoosis, oli 4382 rida.

Aktsiaindeksi hinnaliikumise ennustamine on visualiseeritud järgneval joonisel (Joonis 18. X-teljel on kujutatud aega, y-teljel on S&P500 aktsiaindeksi hind).



Joonis 18. Ennustumudeli prognoositud hinnaliikumine vahemikus juuli 2014 kuni detsember 2016. Must joon tähistab S&P500 aktsiaindeksi tegelikku hinda ning värvilised jooned on aktsiaindeksi hinnaliikumise 10 päeva prognoosid.

Viimase prognoositud päeva hinna ja tegeliku hinna kattuvust iseloomustab järgnev hajuvusdiagramm (Joonis 19. X-teljel on kujutatud S&P500 aktsiaindeksi tegelikku hinda, y-teljel on S&P500 aktsiaindeksi prognoositud hind).



Joonis 19. Tegeliku aktsiaindeksi hinna (must) ja 10 päeva prognoositud hinna (sinine) ühisjaotuse graafiline esitus.

Testandmetel puhul ja eelnevalt välja toodud parameetrite korral olid ennustumudeli täpsust kirjeldavad keskmine ruutviga (MSE) ja keskmine absoluutne protsentuaalne

viga (MAPE) vastavalt 1849 ja 2,33%, mis on arvatud tegeliku hinna ja 10 päeva prognoositud hinna järgi.

Allar Karu on oma magistritöös [23] jõudnud täpsuseni 1,85% (MAPE), kuid väljundi suurus oli 1 periood (prognoositi ainult järgmise perioodi hinnaliikumist). Käesoleva töö väljundi suurus on 10 perioodi, mis on tõenäoliselt üks peamisi faktoreid, miks ennustusmudeli prognoosi täpsus on veidi madalam.

### 5.3 Algoritmide omavaheline võrdlus

Masinõppel põhinevasse kauplemisalgoritmi imporditi sisse tehingute kuupäevad, mis pärinevad deterministlikust kauplemisalgoritmist. See tähendab, et masinõppel põhinev kauplemisalgoritm sooritas tehinguid täpselt samadel kuupäevadel nagu deterministlik, kuid positsiooni suund ja tehingu pikkus sõltuvad juba masinõppe ennustusmudeli prognoosidest. Joonis 20 esitab kokkuvõtlikud tulemused.

Kauplemisalgoritm	NFP		RSXAUT		CREDIT		CPI		PMI	
	D	M	D	M	D	M	D	M	D	M
Tehinguid	42		13		53		70		46	
Lühikeseks müükide osakaal	65%	12%	0%	31%	60%	9%	70%	11%	35%	15%
Keskmine tehingu tootlus	1,1%	-0,24%	1%	0,83%	0,3%	0,37%	0,65%	0,03%	1,3%	-0,59%
Keskmine tehingu pikkus	6,2	3,7	5,8	4,8	6,1	3,8	7,1	3,9	8,2	4,2

Joonis 20. Kauplemisalgoritmide tulemused. D ja M tähistavad vastavalt deterministlikku ja masinõppel põhinevat kauplemisalgoritmi.

Kuigi CREDIT puhul oli keskmine tehingu tootlus masinõppel põhineva algoritmil natuke kõrgem, siis lõplik portfelli seis oli ikkagi deterministliku algoritmi kasuks (vt. peatükk 5.1.3 ja Lisa 2). See tähendab, et deterministlik algoritm näitas paremat tulemuslikkust kõikide majandusindikaatorite korral. Samuti tuleb välja, et masinõppel põhineva algoritmi lühikeseks müükide osakaal on palju väiksem, mis ilmselt tuleneb sellest, et S&P500 aktsiaindeks on enamik ajast olnud ülestrendis. Sellele tuginedes prognoosib masinõppel põhinev algoritm, et keskmiselt peaks aktsiaindeksi hind ülespoole liikuma. Lisaks ilmneb, et masinõppel põhineva algoritmi keskmine tehingu pikkus on paar päeva lühem kui deterministliku puhul.

Masinõppel põhineva kauplemisalgoritmi tulemusi illustreerivad joonised on eraldi välja toodud Lisas 2.



### 5.3.1 Järeldus

Deterministliku kauplemisalgoritmi kogutootlus oli kõrgem kui masinõppel põhineva kauplemisalgoritmi kogutootlus. See kinnitab taas kord, et turu „valesti“ reageerimine üllatavale majandusindikaatori avaldamisele ei ole juhuslik ning antud olukordi on võimalik investoril või kauplejal ära kasutada. Masinõppel põhineva kauplemisalgoritmi sisendiks oli ainult varasem aktsiaindeksi hinnaliikumine, mis tähendab, et teoorias ei tohtinud antud algoritm efektiivselt ära kasutada olukordi, kus turg reageerib „valesti“ üllatavale majandusindikaatori avaldamisele. Ka tulemused kinnitavad seda, sest lühikeseks müükide osakaalude võrdlusest on näha, et masinõppe algoritmi puhul oli antud osakaal palju väiksem.

Kuigi masinõppel põhinev kauplemisalgoritm korrektsioone korrektselt ei prognoosinud, siis algoritm leidis, et kõige suurema tootluse teenib prognooside järgi siis, kui positsiooni hoida avatuna keskmiselt 4 päeva. See tähendab seda, et kuna masinõppe algoritm ei teadnud, et turu „valesti“ liikumise on ilmselt põhjustanud üllatusliku majandusindikaatori avalikustamine, siis algoritm arvas, et tegemist on „tavalise“ hinnakõikumisega ning leidis, et tegemist on olukorraga, kus saab aktsiaindeksit odavalt osta ning suurima kasumi teenida keskmiselt 4 päevaga. See kinnitab asjaolu, et turu „valesti“ reageerimine üllatavale majandusindikaatori avaldamisele on lisandväärtust pakkuv informatsioon, millega oleks mõistlik investoritel, kauplejatel ja ka kauplemisalgoritmidel arvestada.

Käesoleva töö tulemusi on võimalik tulevastel uurimustel edasi arendada. Masinõppel põhineva kauplemisalgoritmi tulemusi tuleks analüüsida teiste sisendite ja tunnuste kombinatsioonidena, et näha, millised sisendid ja tunnused kõige rohkem lisandväärtust, kasulikku informatsiooni endas sisaldavad. Näiteks majandusindikaatoritega seotud informatsioon, aktsiaindeksi päevasised hinnaliikumised, globaalselt suure tähtsusega uudised või ka sama päeva, kuid mõne teise aktsiaindeksi hinnaliikumine (Tokyo aktsiabörs ja Nikkei 225).

## 6 Kokkuvõte

Töö eesmärk oli leida seaduspärasusi S&P500 aktsiaindeksi ebatavalistes hinnaliikumistes peale üllatava majandusindikaatori avaldamist ja võimalusi, kuidas investor või kaupleja saaks tekkivaid olukordi kasumlikult ära kasutada. Eesmärgi täitmiseks realiseeriti kaks kauplemisalgoritmi – deterministlik ja masinõppel põhinev algoritm – ning võrreldi nende tulemusi.

Deterministlik algoritm realiseeriti eeldusel, et aktsiaturg kordab ajaloolisi hinnaliikumisi vastavalt töös käsitletud olukordadele. Masinõppel põhineva algoritmi koostamisel kasutas autor rekurrentseid närvivõrke ja GRU meetodit, sest tegemist on võrdlemisi uue, vähe uuritud meetodiga. Mõlema algoritmi parameetrid leiti optimeerimise tulemusena. Deterministliku algoritmi parimate tulemuste tehingupäevad olid aluseks masinõppel põhinevale algoritmile ehk mõlemad kauplemisalgoritmid sooritasid tehinguid täpselt samal ajal. Algoritmid pidid otsustama, kui kaua positsiooni hoida ja millises suunas seda avada.

Töö teine fookus oli kauplemisalgoritmide tulemuste võrdlemisel viie USA kõige tähtsama majandusindikaatori raames. Deterministliku kauplemisalgoritmi, mis arvestas üllatavate majandusindikaatoritega, kogutootlus oli kõigi indikaatorite puhul kõrgem kui masinõppel põhineva kauplemisalgoritmi kogutootlus. See kinnitab asjaolu, et aktsiaindeksi ebatavaline reageerimine üllatavale majandusindikaatori avaldamisele ei ole juhuslik ning et selliseid olukordi on võimalik investoril või kauplejal ära kasutada.

Töö tulemusena realiseeritud algoritme on võimalik kauplemisplatvormidega integreerida, kasutades vastavaid laiendusi. On oluline märkida, et parameetrid leiti optimeerimise tulemusena. See tähendab, et deterministlik kauplemisalgoritm ei pruugi olla robustne ning pole kindel, et samad parameetrid ka tulevikus töötavad. Edaspidi saab sama meetodikaga uurida ka teisi finantsturge, näiteks valuuta-, toorme- ja võlakirjaturge. Käesolevat tööd jätkavates uurimistes soovib autor keskenduda masinõppel põhineva kauplemisalgoritmi täiustamisele, mille eesmärk oleks anda kasumlikumaid tulemusi kui deterministlik algoritm.

## Kasutatud kirjandus

- [1] About TensorFlow. TensorFlow, 2018. [Võrgumaterjal] <https://www.tensorflow.org/> (09.04.2018)
- [2] Bengio, Y., Chung, J., Gulcehre, C. ja Cho, K. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. Universite de Montreal, Montreal, 2014.
- [3] Britz, D. Implementing a GRU/LSTM RNN with Python and Theano. WildML, 2015. [Võrgumaterjal] <http://www.wildml.com/2015/10/recurrent-neural-network-tutorial-part-4-implementing-a-grulstm-rnn-with-python-and-theano/> (06.04.2018)
- [4] Brownlee, J. How to Use Dropout with LSTM Networks for Time Series Forecasting. Machine Learning Mastery, 2017. [Võrgumaterjal] <https://machinelearningmastery.com/use-dropout-lstm-networks-time-series-forecasting/> (03.04.2018)
- [5] Chapman, A. Moving averages determine trends. Dynamic Trader, 2015. [Võrgumaterjal] <https://dynamictraders.pro/moving-averages-determine-trends-5525d284e8a1> (31.03.2018)
- [6] Consumer Credit. Investopedia. [Võrgumaterjal] <https://www.investopedia.com/terms/c/consumercredit.asp> (10.03.2018)
- [7] Consumer Price Index – CPI. Investopedia. [Võrgumaterjal] <https://www.investopedia.com/terms/c/consumerpriceindex.asp> (12.03.2018)
- [8] Data Science Estonia, Närvivõrkude ja masinõppe sõnastik. Data Science Estonia. [Võrgumaterjal] <http://datasci.ee/masinoppe-sonastik/> (05.05.2018)
- [9] Deep learning – Artificial neural networks. Wikipedia, 2018. [Võrgumaterjal] [https://en.wikipedia.org/wiki/Deep\\_learning#Artificial\\_neural\\_networks](https://en.wikipedia.org/wiki/Deep_learning#Artificial_neural_networks) (02.04.2018)
- [10] Dey, R. ja Salem, F. Gate-Variants of Gated Recurrent Unit (GRU) Neural Networks. Michigan State University, East Lansing, 2017.
- [11] Dimitrakopoulos, K. Examining the financial crisis with professor Eugene Fama – *Q&A with the “father of modern finance”*, Chicago, 2011.
- [12] Economic Indicators: Consumer Credit Report. Investopedia. [Võrgumaterjal] <https://www.investopedia.com/university/releases/consumercreditreport.asp> (10.03.2018)
- [13] Filiz, F. Artificial Intelligent Algorithms. Towards Data Science, 2017. [Võrgumaterjal] <https://towardsdatascience.com/4-1-artificial-intelligent-algorithms-aff1a1ca910a> (12.04.2018)
- [14] Fowler, M. Speeding up Python and NumPy: C++ing the Way. A Medium

- Corporation, 2016. [Võrgumaterjal] <https://medium.com/coding-with-clarity/speeding-up-python-and-numpy-c-ing-the-way-3b9658ed78f4> (22.04.2018)
- [15] HDF5 for Python. h5py, 2018. [Võrgumaterjal]. <https://www.h5py.org/> (19.04.2018)
- [16] Heaton, J. The Number of Hidden Layers. Heaton Research, 2017. [Võrgumaterjal] <http://www.heatonresearch.com/2017/06/01/hidden-layers.html> (14.04.2018)
- [17] How to choose the number of hidden layers and nodes in a feedforward neural network? StackExchange, 2017. [Võrgumaterjal] <https://stats.stackexchange.com/questions/181/how-to-choose-the-number-of-hidden-layers-and-nodes-in-a-feedforward-neural-netw> (14.04.2018)
- [18] Huang, X. Macroeconomic News Announcements, Financial Market Volatility and Jumps. Duke University, Durham, 2008.
- [19] Hui, J. RNN, LSTM and GRU tutorial. DisQus, 2017. [Võrgumaterjal] <https://jhui.github.io/2017/03/15/RNN-LSTM-GRU/> (14.04.2018)
- [20] Hull, J. C. Options, Futures, And Other Derivatives. United States of America: Pearson Education, 2012, 303.
- [21] Investeerimisõpik. AS LHV Pank. [Võrgumaterjal] <https://fp.lhv.ee/academy/investmentguide/347> (25.03.2018)
- [22] Isidore, C. Machines are driving Wall Street's wild ride, not humans. CNN, 2018. [Võrgumaterjal] <http://money.cnn.com/2018/02/06/investing/wall-street-computers-program-trading/index.html> (20.03.2018)
- [23] Karu, A. Tehislike närvivõrkude rakendamise aktsia hinna liikumise prognoosimisel : magistritöö. Tartu Ülikool, Tartu, 2012.
- [24] Keras: The Python Deep Learning library. Keras, 2018. [Võrgumaterjal] <https://keras.io/> (14.04.2018)
- [25] Kress, M. Investeerimisriskide mõõtmisest puust ja punaseks. *Äripäev*, 2017, 20. oktoober.
- [26] Krumm, K. Investeeringute alused. Tallinn: Tallinna Tehnikaülikool Tallinna Kolledž, 2011.
- [27] Kuepper, J. What is the Purchasing Managers Index (PMI). The Balance, 2018. [Võrgumaterjal] <https://www.thebalance.com/what-is-the-purchasing-managers-index-pmi-1978996> (12.03.2018)
- [28] Larsen, J. I. Predicting Stock Prices Using Technical Analysis and Machine Learning : magistritöö. Norwegian University of Science and Technology, Trondheim, 2010.
- [29] Masinõppimine. Vikipeedia, 2018. [Võrgumaterjal] <https://et.wikipedia.org/wiki/Masin%C3%B5ppimine> (14.04.2018)
- [30] McQueen, G. ja Roley, V. Stock Prices, News, and Business Conditions. National Bureau Of Economic Research, Cambridge, Massachusetts, 1990.
- [31] Mindfire Solutions, Advantages and Disadvantages of Python Programming Language A Media Corporation, 2017. [Võrgumaterjal]

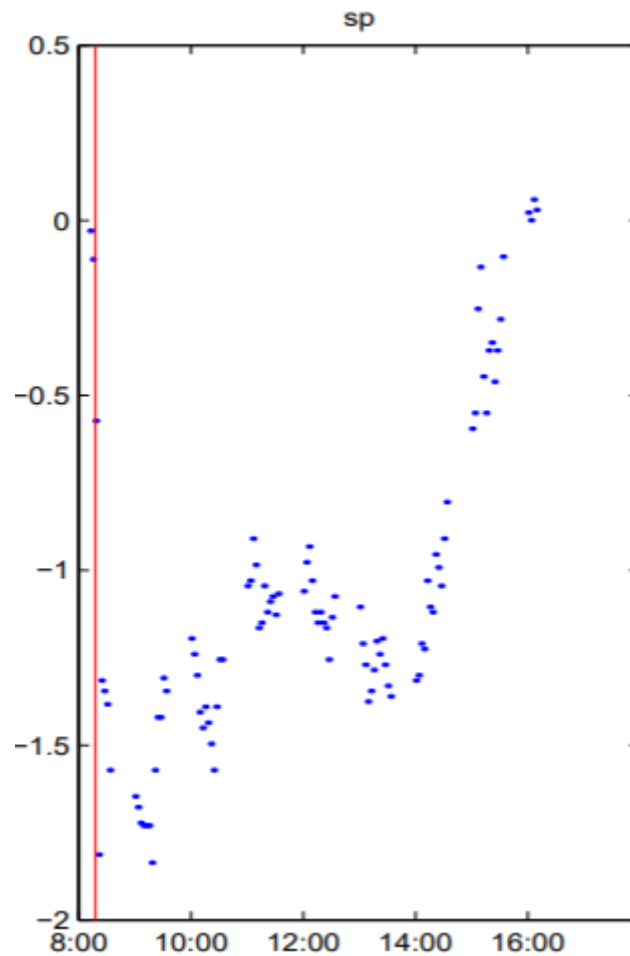
- <https://medium.com/@mindfiresolutions.usa/advantages-and-disadvantages-of-python-programming-language-fd0b394f2121> (14.03.2018)
- [32] Mitchell, T. DNN, CNN, RNN, and Representation Learning. Carnegie Mellon University, 2017. [Võrgumaterjal] <https://mr-why.com/machinelearning/tom-mitchell-machine-learning-04-dnn-cnn-rnn> (02.04.2018)
- [33] Mshomba, A. Macroeconomic surprise, forecast uncertainty, and stock prices : diplomitöö. University of Richmond, Richmond, 2014.
- [34] NumPy. Vikipeedia, 2017. [Võrgumaterjal] <https://et.wikipedia.org/wiki/NumPy> (15.03.2018)
- [35] Petlenkov, E. Tehisnärivõrgud ja nende rakendused. Tallinna Tehnikaülikool, Tallinn, 2004.
- [36] Pisang, S. Fundamentaalne analüüs. *LHV Tartu seminar*, Tartu, 2008.
- [37] Python. Vikipeedia, 2018. [Võrgumaterjal] [https://et.wikipedia.org/wiki/Python\\_\(programmeerimiskeel\)](https://et.wikipedia.org/wiki/Python_(programmeerimiskeel)) (14.03.2018)
- [38] Pythoni moodul pandas. Vikipeedia, 2018. [Võrgumaterjal] [https://et.wikipedia.org/wiki/Pythoni\\_moodul\\_pandas](https://et.wikipedia.org/wiki/Pythoni_moodul_pandas) (15.03.2018)
- [39] Recurrent neural network. Wikipedia, 2018. [Võrgumaterjal] [https://en.wikipedia.org/wiki/Recurrent\\_neural\\_network](https://en.wikipedia.org/wiki/Recurrent_neural_network) (02.04.2018)
- [40] Roosalu, A. Tehniline analüüs. AS LHV pank, Tallinn, 2007.
- [41] scikit-learn – Machine Learning in Python. scikit-learn, 2018. [Võrgumaterjal] <http://scikit-learn.org/stable/> (19.04.2018)
- [42] Sharma, A. Understanding Activation Functions in Neural Networks. A Medium Corporation, 2017. [Võrgumaterjal] <https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0> (14.04.2018)
- [43] Sharma, S. Epoch vs Batch Size vs Iterations. Towards Data Science, 2017. [Võrgumaterjal] <https://towardsdatascience.com/epoch-vs-iterations-vs-batch-size-4dfb9c7ce9c9> (14.04.2018)
- [44] Shen, T. GRUs vs. LSTMs. A Media Corporation, 2017. [Võrgumaterjal] <https://medium.com/paper-club/grus-vs-lstms-e9d8e2484848> (05.04.2018)
- [45] sklearn.preprocessing.MinMaxScaler. scikit-learn, 2017. [Võrgumaterjal] <http://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html> (14.04.2018)
- [46] Ülpner, L. USA ja Tallinna börsi ettevõtete aktsiate sulgemishindade ennustamine ennustusmudeliga : magistritöö. Tallinna Tehnikaülikool, Tallinn, 2017.
- [47] Vähi, K. Majandusloogika vastased valuutakursi muutused majandusindikaatorite avaldamisel valuutapaari EUR/USD näitel : bakalaureusetöö. Tartu Ülikool, Tartu, 2014.
- [48] Veskimägi, M. Efektiivse Turu Hüpooteesi Empiiriline Testimine Tallinna Börsil :

magistritöö. Tartu, Tartu Ülikool, 2006.

[49] Volatility S&P 500 index. TradingView. [Võrgumaterjal]  
<https://www.tradingview.com/> (23.02.2018)

[50] Zhu, Y. ja Zhou, G. Technical Analysis: An Asset Allocation Perspective on the Use of Moving Averages. Washington University, St. Louis, 2008.

## Lisa 1 – S&P500 turuhinna liikumine peale üllatava majandusindikaatori avaldamist

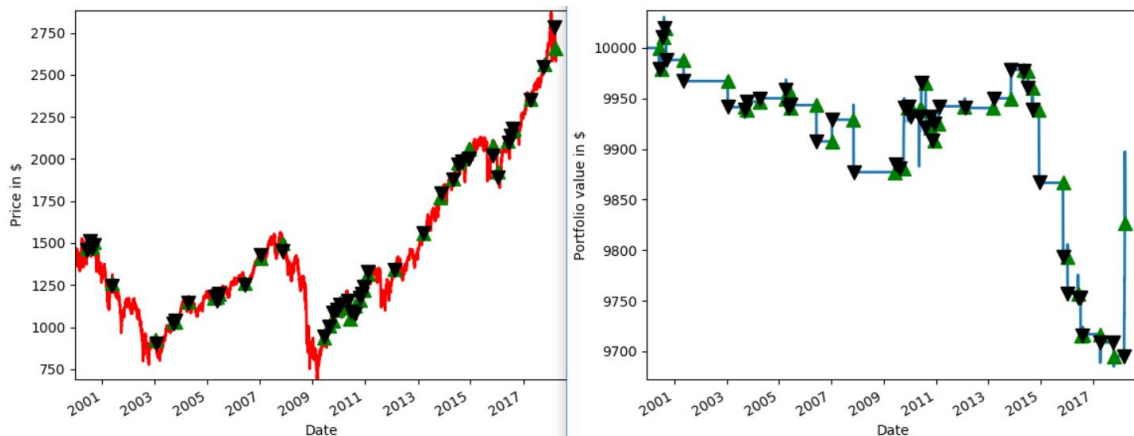


Joonis 21. S&P500 turuhinna liikumist järgiv punktdiagramm [18].

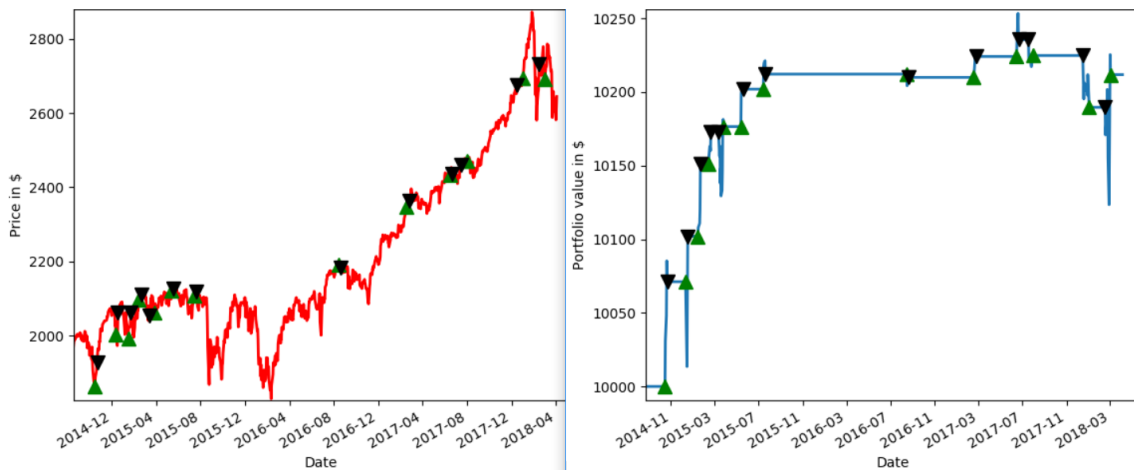
Logaritmiline S&P500 turuhinna liikumist järgiv punktdiagramm kuupäevaga 07.06.1996. Mittepõllumajanduse tööhõive raport: teadaande reaalväärtus on 340, turu konsensus/proгноos/oodatav väärtus on 170 standardhälbega 56,6. Punktid on hinnad erinevatel ajahetkedel, punane joon tähistab majandusindikaatori avaldamise hetke, kell 08.30. X-teljel on kujutatud aega, y-teljel on S&P500 aktsiaindeksi hinna protsentuaalne muutus.

## Lisa 2 – Masinõppel põhineva kauplemissalgoritmi tulemusi illustreerivad joonised

Järgnevatel joonistel on vasakpoolse ja parempoolse graafiku x-teljel kujutatud aega, y-teljel on vastavalt S&P500 aktsiaindeksi hind ja portfelli väärtus.

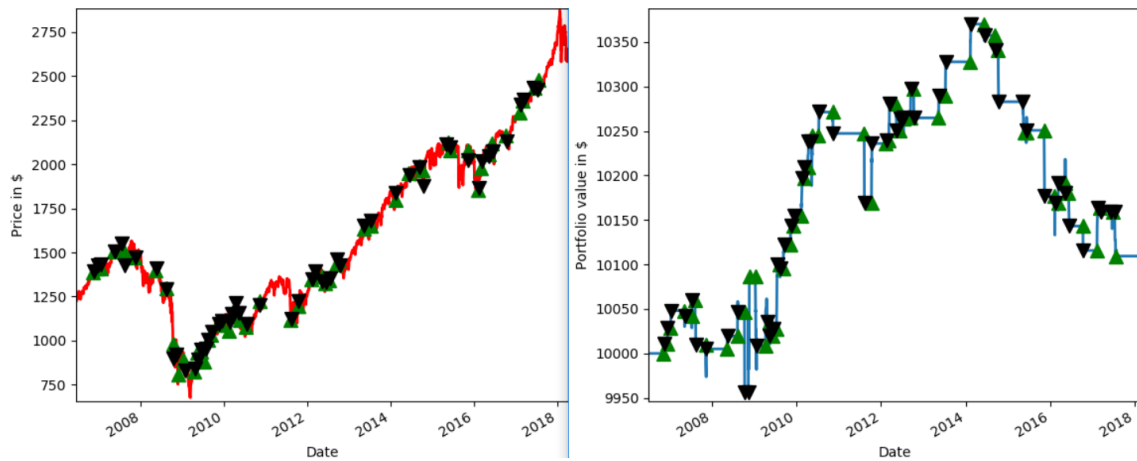


Joonis 22. Masinõppel põhineva kauplemissalgoritmi tulemus NFP puhul.

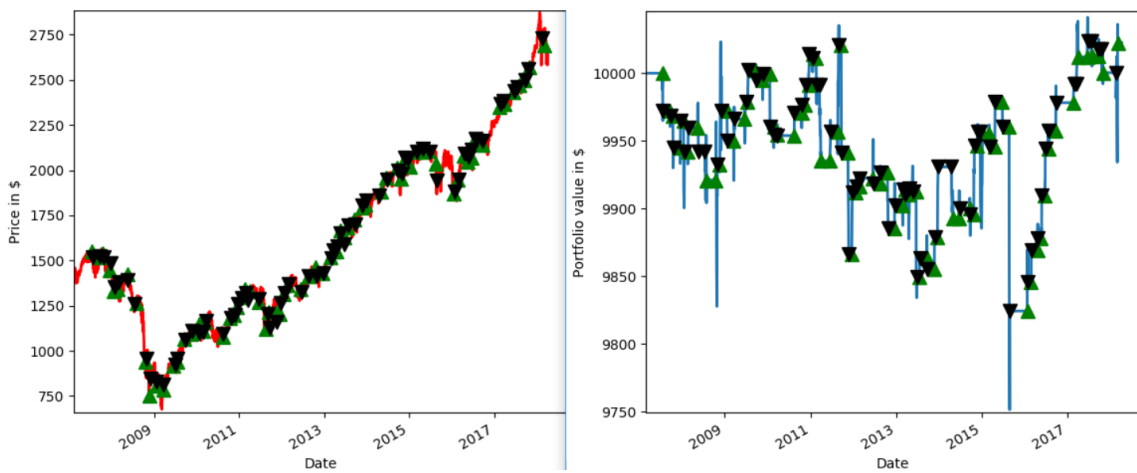


Joonis 23. Masinõppel põhineva kauplemissalgoritmi tulemus RSXAUT puhul.

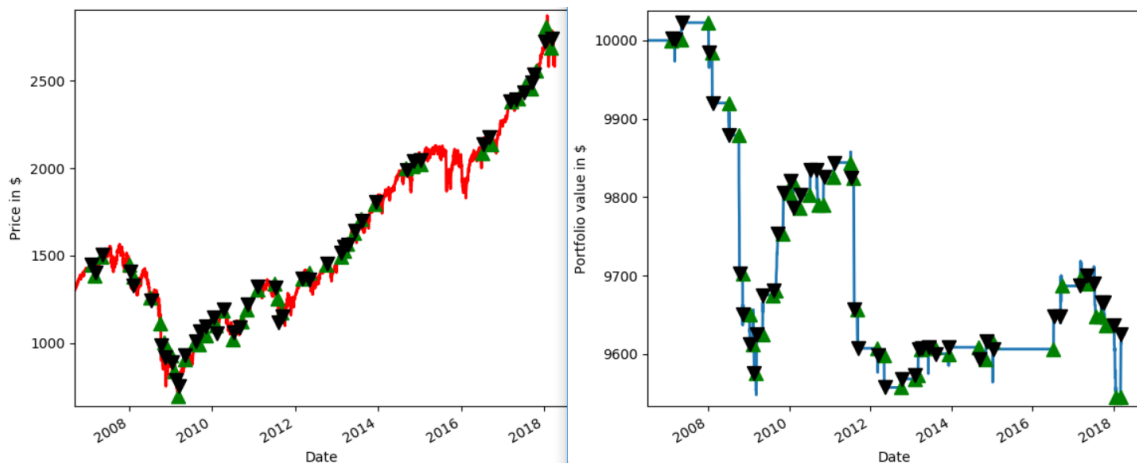




Joonis 24. Masinõppel põhineva kauplemissalgoritmi tulemus CREDIT puhul.



Joonis 25. Masinõppel põhineva kauplemissalgoritmi tulemus CPI puhul.



Joonis 26. Masinõppel põhineva kauplemissalgoritmi tulemus PMI puhul.

## Lisa 3 – Deterministlik kauplemisalgoritm

```
teekide ja data importimine;
muutujate väärtustamine;
data sobivale kujule teisendamine;
iga üllatava majandusindikaatori avaldamise päeva korral:
  kui window on piisavalt suur siis:
    kui üllatus on positiivne siis:
      arvuta keskmine hinnareaktsioon iga trendi korral;
      ebatavaliste hinnareaktsioonide massiivide loomine iga trendi jaoks;
    iga trendi korral:
      kui keskmine hinnareaktsioon on positiivne siis:
        massiivi lisatakse negatiivse hinnareaktsiooniga juhtum;
      vastasel juhul: massiivi lisatakse positiivse reaktsiooniga juhtum;
    iga perioodi (1,2,3,5,10) korral:
      arvuta keskmine ebatavaline hinnareaktsioon;
      arvuta positsiooni lahti hoidmise pikkus;
    kui üllatus on negatiivne siis:
      #loogika on sama nagu positiivse üllatuse puhul;
portfelli dünaamika arvutamine ajas vastavalt ostu ja müügi signaalidele;
ostu- ja müügitehingute kuvamine S&P500 hinnagraafikul;
portfelli väärtuse kuvamine graafikul koos ostu- ja müügitehingutega;
ostu ja müügi signaalide eksportimine vastavalt majandusindikaatorile;
```

Joonis 27. Deterministliku kauplemisalgoritmi pseudokood.

Programmikood on avalikustatud autori GitHubi veebilehel, aadressil <https://github.com/markussomp/Kauplemisalgoritmid>.

## Lisa 4 – Masinõppel põhinev algoritm (*preparation, training, plotting*)

andmete **importimine** (CSV) ja veergude (tunnuste) **määramine**;  
andmete **normaliseerimine**;  
andmete **teisendamine** manipuleerimiseks sobivale maatrikskujule;  
mineviku ja tuleviku valimite **määramine**;  
treeningandmestiku arvutamine;  
ennustusmudeli sisendandmete **eksportimine** HDF5 formaadis;

Joonis 28. *Preparation* pseudokood.

ennustusmudeli treeningandmestike **importimine** HDF5 formaadis;  
hüperparameetrite **väärtustamine**;  
treeningandmestiku ja valideerimisandmestiku **jaotamine**;  
ennustusmudeli **seadistamine** ja **kompileerimine**;  
ennustusmudeli **treenimine**;  
väljundi (kaalude) **eksportimine** epohhi kaupa HDF5 formaadis;

Joonis 29. *Training* pseudokood.

originaalandmestiku ja kaalude **importimine** HDF5 formaadis;  
testandmestiku eraldamine;  
ennustusmudeli **käivitamine**;  
ennustusmudeli täpsuse **hindamine**;  
tulemuste **visualiseerimine** joon- ja hajuvusdiagrammil;  
portfelli dünaamika arvutamine ja **kuvamine**;

Joonis 30. *Plotting* pseudokood

Programmikood on avalikustatud autori GitHubi veebilehel, aadressil <https://github.com/markussomp/Kauplemisalgoritmid>.