

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Ilja Samoilov, 155407IAPB

**BÖRSIL KAUBELDAVATE FONDIDE
HINNAPROGNOOS OTSUSTUSMETS
ABIL MAJANDUSLIKE INDIKAATORITE
AJALIST NIHET KASUTADES**

Bakalaureusetöö

Juhendaja: Martin Rebane
MSc

Tallinn 2019

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: Ilja Samoilov

21.05.2019

Annotatsioon

Investeeringisotsuste tegemine on tehniliselt keeruline ning vajab teadmisi fundamentaalsest ja tehnilisest analüüsist, kuna analüüsis figureerib suures koguses andmeid. Neid probleeme saab lahendada masinõppe abil.

Antud töö eesmärk on välja selgitada võimalikkust ansambelõppe meetodite abil ennustada börsil kaubeldavate fondide hindu eelmise kuu majandusindikaatorite põhjal. Ansambelõppe meetodina kasutatakse käesolevas töös otsustusmetsa.

Töö eesmärkide täitmiseks on arendatud rakendus, mis automatiseeriks finantsandmete laadimist ja vajalikule kujule viimist ning ennustusmudelite treenimiseks. Lõppkasutajale antud rakendus pakub börsil kaubeldavate fondide hindade ennustusi ja toetab läbi selle tema investeerimisvalikuid.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 25 leheküljel, 8 peatükki, 5 joonist, 10 tabelit.

Abstract

Forecasting exchange-traded funds prices using random forest and time shifts of economic indicators

Investment decision making is difficult and requires knowledge in fundamental and technical analysis, as the analysis is well-defined by a lot of data, these difficulties can be solved by machine learning.

The government's plans to allow the transfer of 2 pension pillars to an investment account add actuality to the given work. After changing the law, every person can control where and what amounts of their own money to invest. Exchange-traded funds are a good choice to build their future pension because they hedge the risks of investing in a single share, but they can also generate visible returns.

The aim of this work is to identify the potential to predict the prices of exchange-traded funds using last month's economic indicators with an ensemble learning method. As an ensemble learning method, the random forest algorithm is used in this work.

To meet the goals of the work, an application was developed that would automate the downloading and transforming of financial data to the required shape and to train prediction models. The end-user application provides price predictions for exchange-traded funds and, through this, supports its investment options.

The thesis is in Estonian and contains 25 pages of text, 8 chapters, 5 figures, 8 tables.

Lühendite ja mõistete sõnastik

API	<i>Application programming interface</i> . Kirjeldus viisidest, kuidas üks arvutiprogramm võib suhelda teise programmiga
EMA	<i>Exponential moving average</i> . Eksponentsiaalne libisev keskmine
ETF	Exchange-Traded fund (Börsil kaubeldav fond)
RSME	<i>Root square mean error</i> . Ruutkeskmine viga
SMA	<i>Simple Moving Average</i> . Tavaline libisev keskmine
Ülesobitamine	<i>Overfitting</i> analüüsi tulemus, mis vastab liiga lähedalt või täpselt konkreetse andmestikule ja võib valesti sobitada uued või ennustada tuleviku andmeid. [1]

Sisukord

1 Sissejuhatus	10
2 Töö majanduslik taust.....	12
2.1 Passiivne investeerimine.....	12
2.2 Börsil kaubeldavad fondid.....	13
2.3 Investeerimise alused.....	15
2.3.1 Efektiivse turu teooria	15
2.3.2 Tehniline analüüs.....	16
2.3.3 Fundamentaalne analüüs	16
2.4 Majandusindikaatorid	17
2.4.1 Volatiilsuse indeks VIX	17
2.4.2 Libisev keskmine.....	17
2.4.3 RSI.....	18
2.5 Rakenduse tähtsus tänapäeva olukorras	19
3 Masinõpe	20
3.1 Otsustusmets.....	21
3.1.1 Otsustuspuu	23
4 Tehniline lahendus.....	24
4.1 Kasutatud teegid ja tehnoloogiad	24
4.1.1 Spring boot	24
4.1.2 MongoDB	24
4.1.3 Scikit-learn	25
4.1.4 Flask	25
4.1.5 NumPy.....	25
4.1.6 Pandas.....	26
4.2 Treeningandmete kirjeldus	26
4.3 Arhitektuur.....	28
4.4 Mudelite testimise meetodika.....	29
5 Tulemuste analüüs ja optimeerimine	30
5.1 Ennustusmudelite optimeerimine	31

5.1.1 Hüperparameetrite optimeerimine.....	31
5.2 Optimeerimisvõimalused tulevikus	33
5.2.1 Suuna klassifitseerimine	34
5.2.2 Lisatunnused.....	34
6 Kokkuvõte	35
Kasutatud kirjandus	36
Lisa 1 – 50 puuga otsustusmetsa ennustamistulemused.....	38
Lisa 2 – 100 puuga otsustusmetsa ennustamistulemused.....	39
Lisa 3 – 150 puuga otsustusmetsa ennustamis tulemused.....	40
Lisa 4 – 200 puuga otsustusmetsa ennustamistulemused.....	41
Lisa 5 – hüperparameetritega otsustusmetsa ennustamistulemused.....	42

Jooniste loetelu

Joonis 1. Masinõppe liigid [10]	20
Joonis 2. Otsustusmets [12]	22
Joonis 3. Rakenduse arhitektuur. Autori joonis.....	28
Joonis 4. Algsete mudelite tunnuste tähtsuse graafikud. Autori joonis.....	31
Joonis 5. Optimeeritud otsustusmetsa tunnuste tähtsused. Autori joonis.....	33

Tabelite loetelu

Tabel 1. Kasutatud fondide nimikiri.....	14
Tabel 2. API-st saadud hinnaseeria kirjeldus	27
Tabel 3. Treening andmete kuju	27
Tabel 4. Optimeerimata ennustusmudelite parimad tulemused.....	30
Tabel 5. Hüperparameetrite tabel	32
Tabel 6. Optimeeritud otsustusmetsa tulemuste võrdlus algsetega	33
Tabel 7. 15. mai ETF hindade ennustamise tulemus 50 puuga otsustusmetsa puhul.....	38
Tabel 8. 15. mai ETF hindade ennustamise tulemus 100 puuga otsustusmetsa puhul...	39
Tabel 9. 15. mai ETF hindade ennustamise tulemus 150 puuga otsustusmetsa puhul...	40
Tabel 10. 15. mai ETF hindade ennustamise tulemus 200 puuga otsustusmetsa puhul.	41

1 Sissejuhatus

Suuremate finantsvõimalustega kasvab inimeste huvi passiivse investeerimise vastu, kuid investeerimise valikute tegemine on raske ja teadmisi nõudev. Fundamentaalse ja tehnilise analüüsi abil on võimalik valikute tegemist kergemaks teha, aga nad vajavad hulga teadmisi ja kogemust. Analüüsi on võimalik kergendada masinõppe abil, mis on antud töö üks eesmärkidest.

Majandusindikaatorid mängivad põhirolli investeerimisel. Nendele toetudes teevad kogenud investorid ostu- ja müügiotsuseid. Masinõppe ennustab tunnuste alusel soovitud tulemusi. Sellele toetudes seab autor hüpoteesi, et on võimalik ennustada majandusindikaatorite abil väärtpaberite hindu.

Valitsus plaanib võimaldada üle viia 2. pensionisamba raha investeerimiskontole. Peale seaduse muutmist igal inimesel on võimalik ise juhtida seda, kuhu ja mis kogustes oma raha investeerida. Börsil kaubeldavad fondid on hea valik oma tulevase pensioni ehitamiseks, kuna maandavad üksikaktsiasse investeerimise riske, aga suudavad ka tuua nähtava tulu.

Antud töö eesmärk on välja selgitada võimalikkust ansambelõppe meetodi abil ennustada börsil kauplevate fondide hindu kasutades eelmise kuu majandusindikaatoreid. Ansambelõppe meetodina kasutatakse käesolevas töös otsustusmetsa.

Töö eesmärkide täitmiseks on arendatud rakendus, mis automatiseeriks finantsandmete laadimist ja vajalikule kujule viimist ning ennustusmudelite treenimiseks. Lõppkasutajale antud rakendus pakub börsil kaubeldavate fondide hindade ennustusi ja läbi selle toetab tema investeerimisvalikuid.

Käesolevas töös piirduakse börsil kaubeldavate fondidega ning ei vaadelda teisi finantsinstrumente. Lisaks ei kasutata ka kõrgsageduslikku kauplemist, vaid keskendutakse passiivsele investeerimisele.

Töö jaguneb kuueks osaks. Esimeses tutvustakse lugejale passiivset investeerimist ning selle populaarseimat vahendit, fonde. Edasi uuritakse investeerimise analüüsi aluseid ja seda, mis rolli selles mängivad majandusindikaatorid. Kolmandas osas tutvustab autor rakenduse tähtsust Eesti tänapäeva poliitilises ja majanduslikus olukorras. Pärast seda tutvustab autor masinõppe aluseid ja käesolevas töös kasutava masinõppe algoritmi - otsustusmetsa. Eelviimases osas antakse ülevaade tehnilistest valikutest ja tulemuste valideerimise meetodikatest. Viimases osas autor analüüsib töötulemusi ning selgitab valitud mudeli optimeerimise meetodeid.

2 Töö majanduslik taust

2.1 Passiivne investeerimine

Passiivne investeerimine on investeerimisstrateegia, mille eesmärk on maksimeerida kasumit minimeerides ostmist ja müüki. [2]

Passiivse investeerimise meetodite eesmärk on vältida kõrgsagedase kauplemisega kaasneva võivaid tasusid ja piiratud tulemuslikkust. Passiivse investeerimise eesmärk on suurendada investeerija rikkust järk-järgult. Ka ostu- ja hoiustamisstrateegiana tuntud passiivne investeerimine tähendab väärtpaberi ostmist kavatsusega omada seda pikaajaliselt. Erinevalt aktiivsetest kauplejatest ei püüa passiivsed investorid kasu saada lühiajalistest hinnakõikumistest ega turu ajastamisest. Passiivse investeerimisstrateegia aluseks on see, et turg annab aja jooksul positiivset tulu. [2]

Passiivne investeerimine sobib väga hästi inimestele, kellel ei ole väga suurt huvi finantsmaailma vastu või kellel pole vastavaid teadmisi sellest, kuid on olemas huvi oma raha kasvatada ning selle võrra tõsta oma elukvaliteeti.

ETFi (inglise keeles ETF = **Exchange Traded Fund**) investeringud loodi 1990. aastate alguses, et pakkuda kasutajatele ühtset tagatist indeksi jälgimiseks ning võimekust kogu päeva jooksul oma varasid osta ning müüa. ETFi investeringuid kasutades on investorid teoreetiliselt võimelised ostma ja müüma väärtpabereid, mis moodustavad ühtses kaubanduses terve turu. Selle tõttu antakse investoritele suur paindlikkus osta või müüa igal ajal kogu päeva jooksul. See on eelis investoritele, kes eelistavad osta ja hoida või eelistavad oma investeringut ise hallata. Ostmise ja hoidmise investeerimisstrateegiat peetakse passiivseks, kuid siiski efektiivseks. Passiivse lähenemise abil saavad kauplejad jälgida börsiindeksit ja teenida indeksi teenimise ajal. See lähenemine tähendab investoritele madalamaid tasusid, kuid vähem kontrolli. Tänu sellele ETFi investeringutele on indekseerimise või passiivse juhtimise rakendamiseks mugav ja madal kuluvuus. [3]

2.2 Börsil kaubeldavad fondid

Investeeringifond on mitme osaniku raha kogum, mida kasutatakse väärtipaberite ostuks. Iga investor säilitab omandi ja kontrolli oma osa üle. Fondide mõte on suurendada investorite raha ja leevendada tavalisi väärtipaberitesse investeerimise riske.

Käesolevas töös vaatleme ainult börsil kaubeldavaid fonde, kuna nende tulemusi on võimalik mõõta börsi hinna muutumise ja tehnilise analüüsi järgi.

Börsil kaubeldavad fondid (inglise keeles ETF = **Exchange Traded Fund**) jagunevad kaheks- börsil kaubeldavad kinnised fondid (closed end ETF), mis lähtuvad investeerimisel fondijuhi otsustest ning börsil kaubeldavad indeksfondid (index ETF), mille portfelli koosseis vastab mõne indeksi koosseisule ja mille hind järgib seetõttu indeksi liikumist ning ei sõltu fondijuhi otsustest. Fondi kauplemine börsil annab sellele tavalise aktsia omadused ja eelised. Erinevalt tavalisest fondist saab börsil kaubeldavat fondi lühikeseks müüa ja osta võimendusega, kaubelda päevasiseselt ning kasutada erinevaid ostutüüpe. [4]

Lisaks sellele, et börsil kaubeldavad fondid pakuvad reaalses hinnakujundust, võimaldavad nad investoril kasutada ka keerukamaid ostutüüpe, mis annavad investorile kõige rohkem kontrolli oma hinna üle. [5]

Käesolevas töös kasutatakse 10.05.2019 seisuga LHV panga poolt pakutavaid fonde kasvukontos.

Fondi nimi	Sümbol
iShares Automation & Robotics UCITS ETF	2B76
iShares Ageing Population UCITS ETF	2B77
iShares Healthcare Innovation UCITS ETF	2B78
Amundi ETF MSCI India UCITS ETF	CI2
Xtrackers MSCI Russia Capped Swap UCITS ETF	DBXV
iShares NASDAQ-100® UCITS ETF	EXXT
Amundi S&P Global Luxury UCITS ETF	GLUX
iShares \$ Corp Bond UCITS ETF	IBCD
iShares Global Corp Bond EUR Hedged UCITS ETF	IBCQ
iShares Global Clean Energy UCITS ETF	IQQH
iShares Global Infrastructure UCITS ETF USD (Dist)	IQQI
iShares Global Water UCITS ETF	IQQQ
iShares STOXX Global Select Dividend 100 UCITS ETF	ISPA
iShares Dow Jones Global Sustainability Screened UCITS ETF	IUSL
SPDR® MSCI Europe Consumer Staples UCITS ETF	SPYC
SPDR® S&P® U.S. Dividend Aristocrats UCITS ETF	SPYD
SPDR® Dow Jones Global Real Estate UCITS ETF	SPYJ
SPDR® S&P® Euro Dividend Aristocrats UCITS ETF	SPYW
iShares Core MSCI Pacific ex-Japan UCITS ETF	SXR1
iShares Core S&P 500 UCITS ETF	SXR8
Lyxor MSCI EMU Value UCITS ETF	VAL
Vanguard FTSE All-World UCITS ETF	VWRL
Xtrackers CSI300 Swap UCITS ETF	XCHA
MSCI USA Index UCITS ETF (DR) 1C	XD9U
MSCI World Index UCITS ETF	XDWD
Xtrackers MSCI World Health Care UCITS ETF 1C	XDWH
Russell 2000 UCITS ETF	XRS2

Tabel 1. Kasutatud fondide nimikiri

2.3 Investeerimise alused

Informatsioon mängib investeerimises põhirolli ning on aluseks fundamentaalsele ja tehnilisele analüüsile. Informatsiooni ja väärtpaperite hindade seose võtab kokku Eugene Fama 1970. aastal loodud efektiivse turu teooria.

2.3.1 Efektiivse turu teooria

Efektiivse turu teooria on finantsmajanduslik teooria, milles väidetakse, et väärtpaperite hinnad peegeldavad täielikult kogu olemasolevat informatsiooni. Otsene tagajärg on see, et ei ole võimalik järjekindlalt "turgu edestada", sest turuhinnad peaksid reageerima vaid uuele informatsiooniline. [6]

Efektiivse turu hüpotees seostub "juhusliku ekslemise" ideega, mis on finantskirjanduses kasutatud termin hinnaseeria iseloomustamiseks, kus kõik järgnevad hinnamuudatused esindavad juhuslikke varasematest hindadest kõrvalekaldeid. Juhusliku ekslemise idee loogika on see, et kui infovool on takistamatu ja info kajastub kohe aktsiahindades, siis hinnamuutus peegeldab vaid homseid uudiseid ja on tänasest hinnamuutustest sõltumatu. Aga uudised on definitsiooni kohaselt ettearvamatud ja seega peavad hinnamuudatused olema ettearvamatud ja juhuslikud. Selle tulemusena peegeldavad hinnad täielikult kogu teadaolevat teavet ning isegi informeerimata investorid, kes ostavad turu poolt antud hinnatabelis hajutatud portfelli, saavad sama helde tasuvuse kui eksperdid saavutasid. [7]

Efektiivse turu teooria järgi on turul kolm vormi: nõrk, pooltugev ja tugev efektiivsuse vorm, millest igaühel on erinevad mõjud turgude toimimisele.

- Nõrgas vormis ei saa tulevikuhindu ennustada mineviku hindade analüüsimisega. Ülemäärast tootlust ei saa pikas perspektiivis teenida, kasutades investeerimisstrateegiaid, mis põhinevad ajaloolistel aktsiahindadel või muudel ajaloolistel andmetel. Tehnilise analüüsi abil ei ole võimalik järjekindlalt toota ülemääraseid tootlusi, kuigi mõned fundamentaalse analüüsi vormid võivad ikka anda ülemäärast tootlust.
- Pooltugev vorm eeldab, et aktsiahinnad kohanduvad avalikult kättesaadava uue teabega väga kiiresti ja erapooletult nii, et mingit ülemäärast tulu ei saa teenida

selle teabe kauplemisega. Fundamentaalse ega tehnilise analüüsi tehnika ei suuda usaldusväärselt toota ülemääraseid tootlusi.

- Tugevas vormis aktsiahinnad peegeldavad kogu infot – avalikku ja eraõiguslikku – ning keegi ei saa teenida ülemäärast tootlust.

Antud töös tugineme selle, et turg on nõrgalt efektiivne, kuna on ilmselge, et investorite emotsioonid mängivad rolli nende otsuste tegemisel ja kõigil pole sama hea info kättesaamise võimalus. Kuna nõrga efektiivse turu seisus ei ole võimalik teenida tulu ainult tehnilise analüüsi abil, siis lisame tehnilisele analüüsile ka fundamentaalse, volatiilsuse indeksi näol.

2.3.2 Tehniline analüüs

Tehniline analüüs on meetod, mis prognoosib väärtpaberi hinna muutumist tulevikus toetudes minevikus toimunud aktsiahinna liikumisele. Tehniline analüüs erineb fundamentaalanalüüsist, mille järgi analüüsitakse ettevõtte majandusnäitajaid, tegutsemiskeskonda jms ettevõtte tegevust mõjutavaid tegureid. Tehniline analüüs on tihti ainult statistiline ning vaatleb aktsia hinda ja käivet, nõudlust ning pakkumist, seejuures ei ole olulised fundamentaalsed põhjendused. [4]

Tehniline analüüs põhineb efektiivse turu teoorial, mille kohaselt sisaldavad aktsiate hinnad juba olemasolevat informatsiooni, sealhulgas ka investorite ootusi. Efektiivne turg reageerib kohe igale uuele infole ning ühelgi investoril ei ole võimalik sama riski juures teenida rohkem kui teistel. Kui turg on efektiivne, võib eeldada, et aktsiate hinnad on õiglastel tasemetel ning peegeldavad tegelikku väärtust. Reaalselt pole ükski turg täielikult efektiivne, sest investorid ei oma võrdset hulgal informatsiooni või on turg informatsiooni valesti hinnanud. Mitteefektiivse turu korral toimuvad uudise järel aktsia hindades suured liikumised ning ülereaktsioonid. [4]

2.3.3 Fundamentaalne analüüs

Fundamentaalne analüüs püüab mõõta väärtpaberi olemuslikku väärtust uurides sellega seotud majanduslikke ja finantstegureid, mis võivad olla nii kvalitatiivsed kui ka kvantitatiivsed. Fundamentaalsed analüütikud uurivad kõike, mis võib mõjutada väärtpaberi väärtust, sealhulgas makromajanduslikke tegureid (nt majanduse ja tööstuse tingimused) ja mikromajanduslikke tegureid (nt finantstingimused ja ettevõtte juhtimine).

Fundamentaalse analüüsi lõppeesmärk on toota kvantitatiivset väärtust, mida investor saab võrrelda väärtpaberi praeguse hinnaga, näidates seega, kas väärtpaber on alahinnatud või ülehinnatud. [2]

2.4 Majandusindikaatorid

Majandusindikaatorid on majandusandmed, tavaliselt makromajanduslikust skaalast, mida analüütikud kasutavad praeguste või tulevaste investeerimisvõimaluste tõlgendamiseks. Need näitajad aitavad hinnata ka majanduse üldist tervist. Majandusindikaatorid võivad olla kõik, mida investor valib, kuid laialdaselt on järgitud konkreetseid tükke valitsuse ja kasumiühingute poolt avaldatud andmetest. [2]

2.4.1 Volatiilsuse indeks VIX

Chicago optsioonibörsi volatiilsuse indeks näitab turuosaliste 30 päevase volatiilsuse ootust. See indeks on koostatud kasutades S&P500 indeksi *put*- ja *call* optsioone. VIX indeksit kasutatakse turusentimendi mõõtmiseks. Mida kõrgem on volatiilsuse indeks, seda ebakindlamad on investorid. [4]

Volatiilsuse indeksi tähtsus antud töös on tehnilise analüüsile lisada üldise turuseisundi näitajat, kuna börsil kauplevad fondid (ETF) on tundlikud majanduse muudatuste vastu. Eeldame, et tänapäeva turg on nõrgas efektiivsuse vormis ja ilma fundamentaal analüüsi, mida volatiilsuse indeks pakub, ülemäärast tootlust ei ole võimalik saavutada.

2.4.2 Libisev keskmine

Moving average (MA) e libisev keskmine arvutab teatud arvu viimaste päevade aktsiahinna keskmise. Kuna iga uue päevaga arvutamisel kasutatud päevad muutuvad (uue päeva hind lisatakse ning kõige vanema päeva hind võetakse maha), siis liigub ka keskmine. Mida vähem päevi kasutatakse, seda aktiivsemalt MA aktsiahinna muutustele reageerib, kuid igal juhul kõigub MA aktsiahinnast vähem. [4]

Finantsmaailmas on tavaline libisev keskmine (SMA) ehk varasemate n hinnaseeriade päeva lõpu hindade kaalumata keskmine. Tavalise libiseva keskmist arvutakse järgnevalt:

$$SMA(t) = \frac{1}{n} \sum_{i=0}^{n-1} Price(i)$$

Kus:

$$n = \text{päevade arv SMA} - s$$

$$t = \text{täna}$$

Eksponentsiaalses libisevas keskmises (EMA) on suurem tähtsus hilisematel hinnaseeriatega päevade lõpu hindadel.

$$EMA(t) = Price(t) * k + EMA(y) * (1 - k)$$

Kus:

$$t = \text{täna}$$

$$y = \text{eile}$$

$$n = \text{päevade arv EMA} - s$$

$$k = \frac{2}{n + 1}$$

Käesolevas töös kasutakse 50 ja 200 päevade libisevaid keskmisi, mida arvutakse hinnaseeriatega päeva lõpphindadest. Sellised perioodid olid valitud, sest börsil kauplevate fondide (ETF) hinnad on üsna stabiilsed hajutatud aktsiaportfellide tõttu. Pikemad perioodid annavad võimalust leevendada investorite impulsiivsust, majanduslike üllatusi ja teisi segavaid faktoreid.

2.4.3 RSI

Relative Strength Index (RSI) võrdleb hüplemist plusspäevade arvu miinuspäevade arvuga ning väljendab saadud tulemust arvuga 1st 100ni.

Saadud arvu saab kasutada ostu- või müügisoovitustena: Kui RSI on alla 30ne, siis peetakse aktsiat üle muuduks ning ostmiseks on õige aeg. Kui RSI on üle 70ne, siis peetakse aktsiat üle ostmiseks ning müümiseks on õige aeg. [4]

2.5 Rakenduse tähtsus tänapäeva olukorras

2019. aastal moodustatud koalitsioon tuli välja plaaniga muuta 2020. aastal 2. pensionisammas vabatahtlikuks. Seoses sellega inimesed otsivad võimalusi oma raha investeerida teistesse kohtadesse eesmärgiga kindlustada oma tulevik.

Teise pensionisambaga liitunud ja tulevikus liituvad inimesed saavad õiguse teise pensionisamba skeemist avalduse alusel väljuda. Väljunud inimesi koheldakse analoogselt nendega, kes ei ole teise sambaga liitunud. Selleks valitsus töötab välja väljamaksmise tingimused ja korra. Väljamaksed teise samba pensionifondidest tehakse pärast avalduse esitamist kahe aasta jooksul ning neilt tasutakse tulumaks. [8]

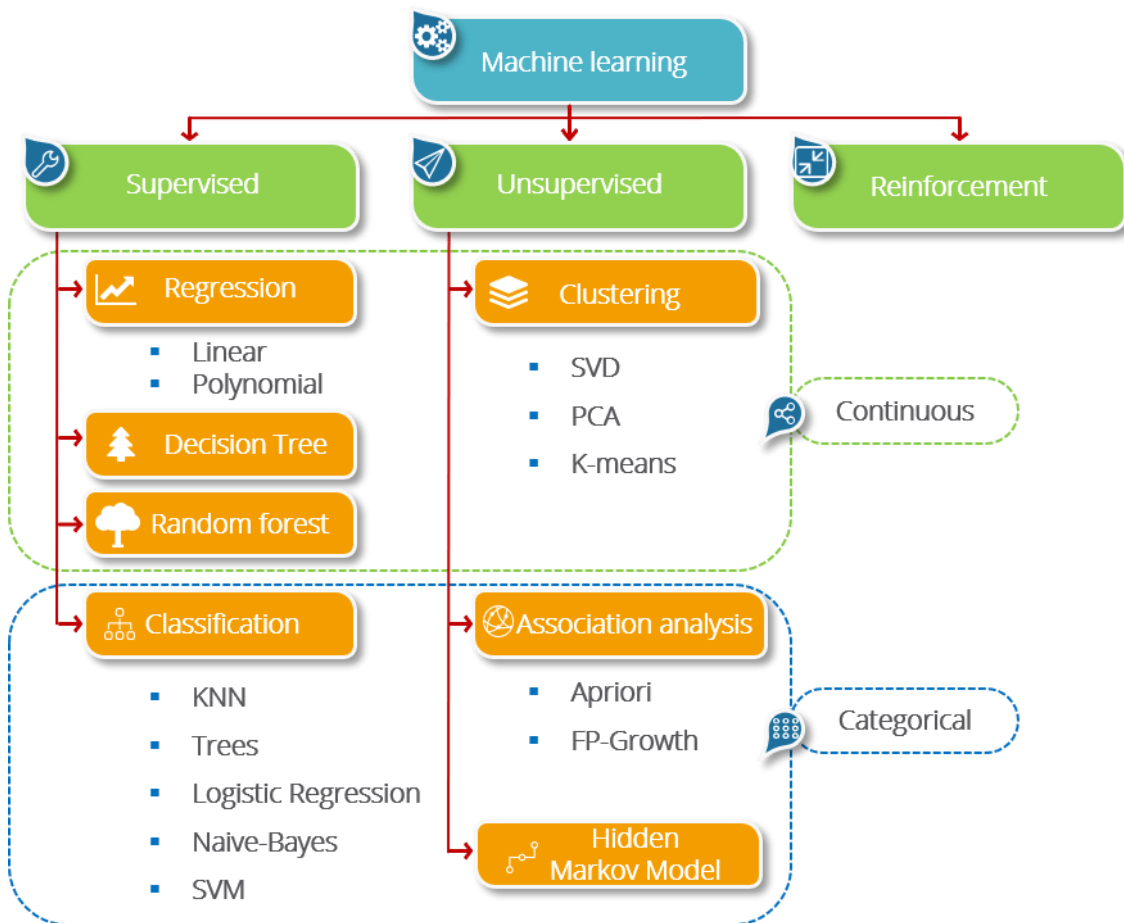
2015. Saar Poll OÜ viis läbi Rahandusministeeriumi tellimusel uuringu „Finantsalase kirjaoskuse uuring Eesti elanike seas“, kust tuleneb, et üle poolte vastajatest (56%) hindab oma üldiseid rahaga ümberkäimise oskuseid võrreldes teiste täiskasvanutega Eestis keskmiseks. Väga kõrgeks või küllaltki kõrgeks hindab oma finantsalaseid oskuseid 30%, samas kui küllaltki madalaks või väga madalaks 11% elanikest. [9]

Sellest võime järeldada, et inimesed ei ole üldiselt teadlikud tehnilisest ja fundamentaalsest analüüsist, kuid aastal 2020 on võimalik enda kogutud raha uuesti investeerida endale sobivasse suunda. Börsil kaubeldavad fondid sobivad just kõige paremini inimestele, kes on huvitatud raha kasvatamisest ja kel pole piisavalt aega ega teadmisi aktiivseks aktsiate kauplemiseks. Parema valiku analüüsimiseks oli välja töötatud antud töö raames rakendus.

3 Masinõpe

Masinõpe uurib algoritme ja statistilisi mudeleid, mida arvutisüsteemid kasutavad konkreetse ülesande tõhusaks täitmiseks ilma selgesõnaliste juhusteta, tuginedes hoopis mustritele ja järeldamisele. Seda peetakse tehisintellekti alagrupiks. Masinõppe algoritmid ehitavad prooviandmetel põhineva matemaatilise mudeli, mida tuntakse kui "treeningandmeid", et teha ennustusi või otsuseid, ilma et oleks otseselt programmeeritud ülesande täitmiseks.

Masinõppe algoritme kasutatakse väga erinevates rakendustes, näiteks meilifiltreerimises ja arvutinägemises, kus on võimatu töötada välja konkreetse ülesande täitmiseks vajalike juhiste algoritm. Masinõpe on tihedalt seotud arvutusstatistikaga, mis keskendub arvutite abil ennustuste tegemisele.



Joonis 1. Masinõppe liigid [10]

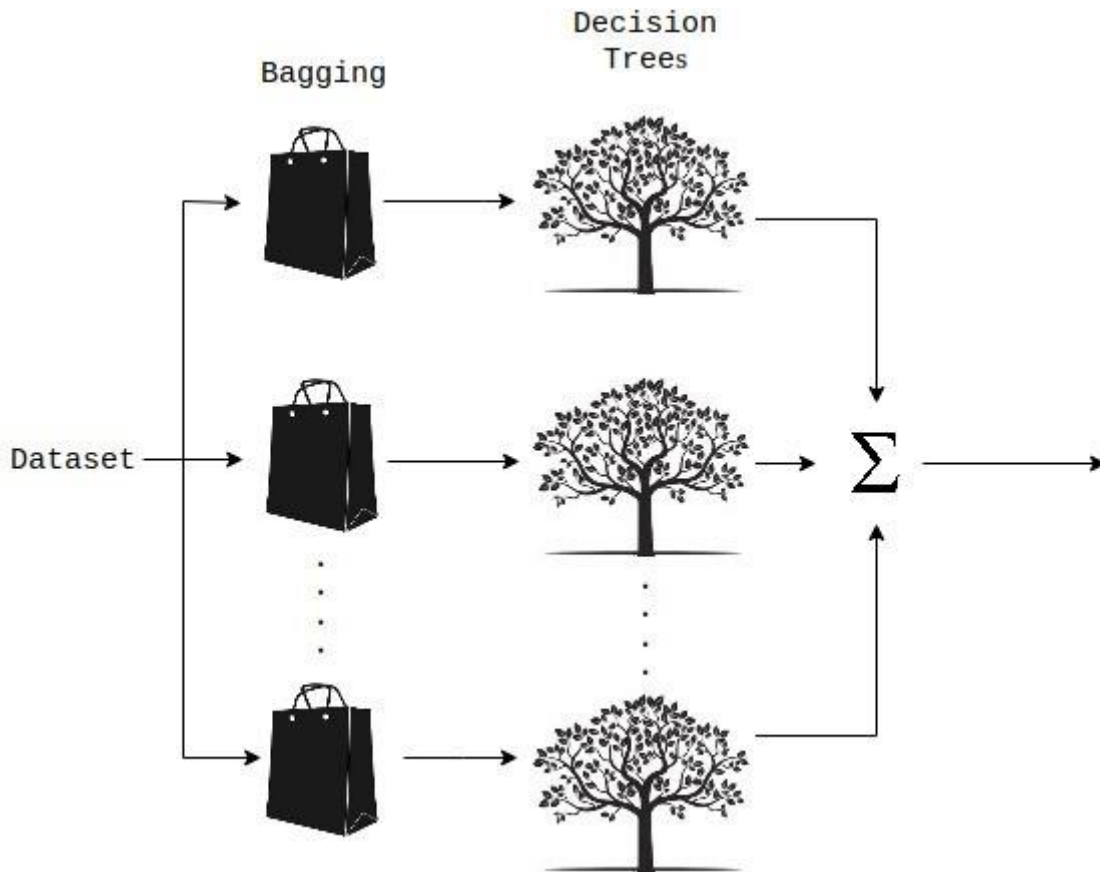
Masinõpe tavaliselt jagatakse kolmeks liigiks:

- juhendatud õppe (*supervised learning*) algoritmid ehitavad matemaatilise mudeli andmekogu abil, mis sisaldab nii sisendeid kui ka soovitud väljundeid;
- juhendamata õppe (*unsupervised learning*) jaoks kasutakse andmeid, mis sisaldavad ainult sisendeid. Leitakse andmetes struktuuri, nagu näiteks andmete punktide grupeerimine või klasterdamine;
- stiimulõppe (*reinforcement learning*) tegeleb tarkvaraagentidega, mis peaksid tegutsema keskkonnas, et maksimaalselt ära tunda mingisugust ühistulist tasu.

Antud töös kasutakse börsil kauplevate fondide hindade ennustamises ennustusmodelina otsustusmetsa, mis kuulub juhendatud õppe valdkonda. Otsustusmetsa abil on võimalik realiseerida nii regressiooni kui ka klassifikatsiooni, kuid selles töös kasutakse ainult regressiooni osa eesmärgiga genereerida konkreetseid numbrilisi väärtuseid.

3.1 Otsustusmets

Otsustusmets on ansambelõppe meetod, mida kasutakse klassifikatsiooniks, regressiooniks ja teiste ülesannete jaoks, konstrueerides eri puudest koosneva ansambli. Väljundiks on individuaalsete puude poolt kõige ennustatum klass või keskmine ennustus. [11]



Joonis 2. Otsustusmets [12]

Otsustusmetsa algoritmi plussid:

- saab kasutada nii regressiooni kui ka klassifikatsiooni jaoks;
- ülesobitamine on minimaalne, kui puude arv on piisav;
- suudab töötada ka puudevate andmetega;
- oskab välja tuua tähtsamad tunnused andmestikust;
- saab kasutada ka kategoorilisi väärtuseid. [13]

Otsustusmetsa algoritmi miinused:

- otsustusmets vajab suurt arvutusvõimekust;
- vajab rohkem treenimisaega võrreldes sarnaste algoritmidega. [13]

Otsustusmetsa treenimise pseudokood:

Iga puu loomise korral:

Kuni otsustuspuus on loodud **l** arv tippu:

Juhuslikult vali **K** tunnust koguarvu **M**-st, kus $K \leq M$

K tunnuste seas arvuta puu tipp **D**, kasutades parimat lahknemispunkti

Jaota tipp tütartippudeks, kasutades parimat lahknemispunkti. [13]

3.1.1 Otsustuspuu

Otsustuspuu loob puustruktuuri kujuga regressiooni- või klassifikatsioonimudeli. See jaotab andmekogumi väiksemateks ja väiksemateks alamkogumiteks, samal ajal kui seostuv otsusepuu on järk-järgult välja töötatud. Lõpptulemuseks on otsustussõlmede ja lehesõlmedega puu. Otsusesõlmel on kaks või enam haru, millest igaüks esindab testitud atribuudi väärtusi. Lehesõlm kujutab endast otsust arvulise sihtmärgi kohta. Puu kõige ülemist otsusesõlme, mis vastab parimale ennustajale, nimetatakse juursõlmeks. Otsustuspuud saavad käsitseda nii kategoorilisi kui ka numbrilisi andmeid. [14]

Otsustuspuud on iseenesest nõrgad õppijad ning nad kipuvad väga sügavale kasvanud puu puhul õppima väga ebaregulaarseid mustreid: need ülesobitavad oma treeningkomplektidele, st neil on madal kallutatatus, kuid väga suur varieeruvus. Otsustusmets on meetod, kus kombineeritakse mitme sama treeningkomplekti eri osades treenitud otsustuspuu ennustused, et vähendada varieeruvust. [15]

4 Tehniline lahendus

Rakenduse eesmärk on aidata lõppkasutajal valida börsil kauplevaid фонде, kuhu investeerida LHV kasvukonto põhjal, ennustades järgmise kuu fondide hindu. Antud töös masinõppe aluseks on hüpotees, et fondi hindu on võimalik ennustada eelmise kuu majandusindikaatorite põhjal.

Antud töös autor nimetab rakenduseks mikroteenuste kogumit, mis on kirjeldatud alapeatükis „Arhitektuur“.

4.1 Kasutatud teegid ja tehnoloogiad

4.1.1 Spring boot

Spring Boot on Spring framework'i *convention-over-configuration* lahendus autonoomsete tootmisvalmis Java rakenduste loomiseks. See on eelkonfigureeritud Spring meeskonna poolt, et oleks võimalikilustada minimaalse vaevaga. Enamik Spring Boot rakendusi vajab väga vähe Spring konfiguratsiooni. Selle põhi eesmärgid on:

- Võimaldada kiire ja laialdaselt kättesaadavat alustamiskogemust Spring raamistikuga;
- Olla “karbitoode” - töötada kohe, kuid vajadusel pakkuda kerget võimalust muuta sätteid, kui nõuded hakkavad vaikesätetest kõrvale kalduma;
- Pakkuda suurt erinevatele projektitüüpidele vajalikku kogumit mittefunktsionaalseid võimalusi;
- Olla vaba koodi genereerimisest ning vajadusest XML konfiguratsiooni järele.

[16]

4.1.2 MongoDB

MongoDB on platvormisõltumatu ja avatud lähtekoodiga dokumendikeskne andmebaas, omamoodi NoSQL-andmebaas. NoSQL-i andmebaasina kahandab MongoDB

relatsioonandmebaasi tabelipõhist struktuuri, et kohandada JSON-laadseid dokumente, millel on dünaamilised skeemid, mida nimetatakse nimetab BSONiks.

See muudab teatud tüüpi rakenduste andmete integreerimise kiiremaks ja lihtsamaks. MongoDB on loodud skaleeruvuse, suure käideldavuse ja jõudluse tagamiseks alates ühest serveri juurutamisest kuni suurte ja keerukate mitme saidi infrastruktuurideni. [17]

MongoDB valiti, kuna andmemahud on suured, kuid puudub vajadus teha raskeid ja pesastatud päringuid, mida pakuvad SQL andmebaasid.

4.1.3 Scikit-learn

Scikit-learn on Pythoni programmeerimiskeelele valmistatud tasuta masinõppe teek. Sellel on erinevad klassifitseerimis-, regressiooni- ja klasterdamisalgoritmid, sealhulgas tugivektorklassifitseerijad, otsutusmetsad, gradiendi võimendamine ja DBSCAN, ning see on ette nähtud koostöömiseks Pythoni arvuliste ja teaduslike teekidega NumPy ja SciPy. [18]

Scikit-learn on hästi optimeeritud ja annab häid võimalusi otsutusmetsa ehitamiseks ja treenimiseks, seetõttu osutus see valituks.

4.1.4 Flask

Flask on Pythonis kirjutatud mikroveebiraamistik. See kuulub mikroveebiraamistike liigituse alla, kuna see ei vaja kindlaid tööriistu ega teeke. Sellel pole andmebaasi abstraktsioonikihti, vormi valideerimist ega muid komponente, kus olemasolevad muude tootjate teegid pakuvad ühiseid funktsioone. [19]

Kuna erilisi nõudeid ennustamisveebirakenduse jaoks pole, sobib Flask ideaalselt.

4.1.5 NumPy

NumPy on Pythoni programmeerimiskeele teek, mis lisab tuge mitmemõõtmeliste massiividele ja matriksitele koos suure hulga kõrgtasemeliste matemaatiliste funktsioonidega, mis toimivad nendel massiividel. [20]

Ilma NumPy-ta ei ole võimalik kergelt ette valmistada masinõppe treeningandmeid ning NumPy on mõeldud koostöötamiseks Scikit-learn masinõppe teekiga.

4.1.6 Pandas

Pandas on avatud lähtekoodiga, BSD-litsentseeritud teek, mis pakub Pythoni programmeerimiskeele jaoks suure jõudlusega hõlpsasti kasutatavaid andmestruktuure ja andmeanalüüsi tööriistu. Pythonit on juba kaua andmete ettevalmistamise jaoks suurepäraseks peetud, kuid vähem andmeanalüüsi ja modelleerimise jaoks. Pandas aitab seda lünka täita, võimaldades kogu andmeanalüüsi töövoogu Pythonis teostada, ilma et peaks lülituma domeenikohasemale keelele, nagu R. [21]

4.2 Treeningandmete kirjeldus

Otsustusmetsa treenimiseks kasutatakse börsil kaubeldavate fondide päeva lõpu hinnaseeriaid, mille majanduslikud indikaatorid on nihutatud ühe kuu võrra minevikku.

Hinnaseeria on järjestikuste numbriliste andmepunktide jada. Investeerimisel jälgib hinnaseeria valitud andmepunktide liikumist, näiteks väärtpaberi hinda määratud aja jooksul, kusjuures andmepunktid registreeritakse korrapäraste ajavahemike järel. Ei ole miinimum- ega maksimumaega, mis peab sisalduma, võimaldades andmeid koguda viisil, mis pakub õiget andmete tihedust, sõltuvalt investori või analüütiku eesmärkidest. [2]

Antud töös kasutatakse päevaseid hinnaseeriaid, siis andmete tihedus on suur ja võimalik saada piisavalt andmeid, et masinõpe oleks võimalik. Lisaks on päevased seeriad kõige enim levinud ja just nende peal on kõige mugavam arvutada majandusindikaatoreid. Kasutatav API on <https://eodhistoricaldata.com>.

Väli	Kirjeldus
Date	Hinnaseeria kuupäev.
Open	Aksia hind antud ajaperioodi alguses.
High	Aksia kõige kõrgem hind antud ajaperioodi jooksul.
Low	Aksia kõige madalam hind antud ajaperioodi jooksul.
Close	Sulgemishind. Aksia hind antud ajaperioodi lõpus.

Adjusted close	Korrigeeritud sulgemishind muudab aktsia sulgemishinda, et kajastada täpselt selle aktsia väärtust pärast ettevõtte mis tahes tegevusearvestamist. Seda peetakse selle aktsia tõeliseks hinnaks ja seda kasutatakse sageli ajaloolise tulu uurimisel või ajaloolise tuluüksikasjalikul analüüsimisel. [2]
Volume	Tervel turul kauplevate aktsiate arv antud ajaperioodi jooksul.

Tabel 2. API-st saadud hinnaseeria kirjeldus

Enne hinnaseeriade kasutamist masinõppes arvutab rakendus majanduslikud indikaatorid ning nihutab neid 30 päeva võrra. Lisaks jagatakse kuupäeva sõne kolmeks eraldiseisavaks numbriks.

Väli	Kirjeldus
Day	Hinnaseeria päev
Month	Hinnaseeria kuu
Year	Hinnaseeria aasta
Adjusted close	Korrigeeritud sulgemishind
Volume	Kauplevate aktsiate arv
LastPeriodPrice	Eelmise kuu sulgemishind.
SMA50	Tavaline libisev keskmine 50 päevade jooksul
SMA200	Tavaline libisev keskmine 200 päevade jooksul
EMA50	Ekspponentsiaalne libisev keskmine 50 päevade jooksult
EMA200	Ekspponentsiaalne libisev keskmine 200 päevade jooksult
RSI	<i>Relative strength index</i>
VIX	Volatiilsusendeks

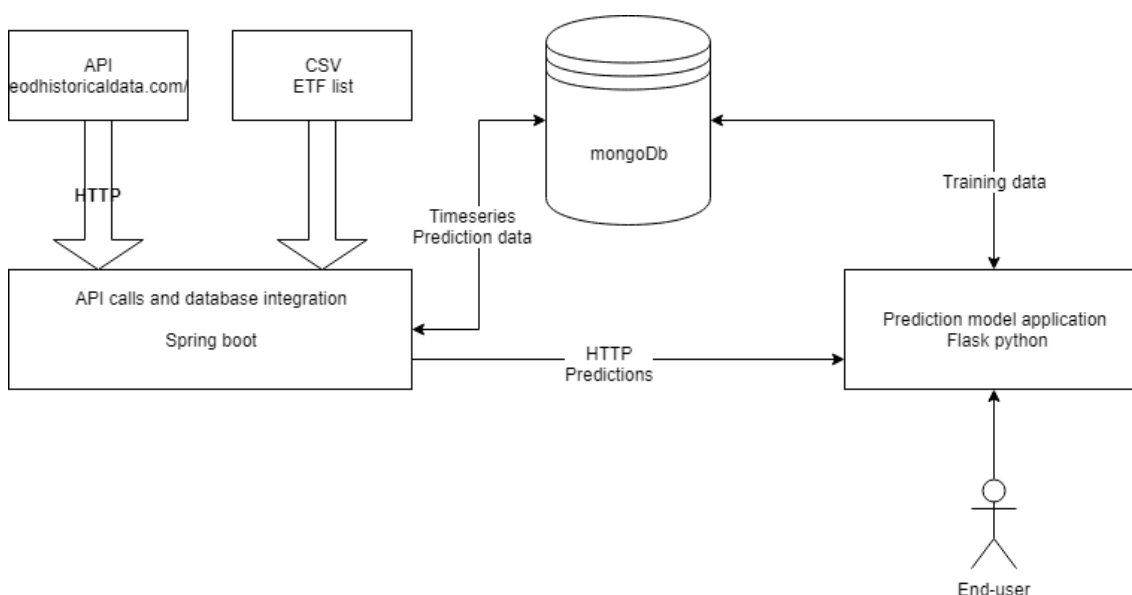
Tabel 3. Treening andmete kuju

Korrigeeritud sulgemishind on ennustatav arv ning ülejäänud väljad on tunnused.

4.3 Arhitektuur

Rakenduse jaoks on valitud andmebaasikeskne arhitektuur, kuna masinõppe abil ennustamine on rakenduse eesmärk. Antud arhitektuurile räägib kasuks ka andmete staatiline struktuur ning nende olemasolu ja kasutatavuse prioriseerimine.

Lisaks on rakendus realiseeritud mikroteenuste kogumina, kuna Java baasil mikroteenus saab paremini hakkama andmete liigutamise ja ettevalmistamisega, sest realiseerib objektorienteeritud põhimõtteid. Python on parim valik masinõppe jaoks ning valdav osa andmetealasi kasutavad just seda. Otsustusmetsa treenimiseks ja selle abil ennustamiseks on valitud just eelnevalt mainitud põhjuste tõttu see programmeerimise keel.



Joonis 3. Rakenduse arhitektuur. Autori joonis.

Spring boot rakendus (nimega Cobolt koodibaasis) tegeleb järgneva:

1. Loeb .csv failist sisse, milliste fondide hinnaseeriaid rakendus hakkab kasutama mudeli treenimiseks ja ennustuste tegemiseks.
2. Pärib eodhistorical API kaudu nende fondide ajaloolised päevahinnaseeriad.
3. Hinnaseeriade abil arvutab majandusindikaatoreid ning salvestab andmed andmebaasi.
4. Valmistab ette objektid ja salvestab andmebaasi, mida kasutakse järgmise kuu ennustuste tegemiseks.

Flaski rakenduse abil saab:

- Treenida ennustusmudeleid
- Ennustada börsil kaubeldavate fondide hindu

4.4 Mudelite testimise metoodika

Mudelite testimiseks kasutame 3 mõõdikut - RSME, vahede RMSE ja hinna ennustusekvaliteedi protsenti.

RMSE on ennustuste vea ruutkeskmise. Ruutkeskmise on valitud, kuna viga jääb samadesse ühikutesse, kui ennustatav tulemus.

$$\text{RMSE}(y, \hat{y}) = \sqrt{\frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} (y_i - \hat{y}_i)^2}$$

Kus:

$$\begin{aligned} y & - \text{tõelised väärtused} \\ \hat{y} & - \text{ennustatud väärtused} \\ n_{\text{samples}} & - \text{valimis suurus} \end{aligned}$$

Vahede RMSE arvutakse sarnaselt, kuid ennustatud ja tõeliste väärtuste asemel kasutatakse eelmise perioodi sulgemishinna ja ennustatud/tõelise väärtuse vahe absoluutväärtust.

Hinnaennustuse kvaliteedi protsent (HEKP) näitab kui palju ennustatud hinnad liikuvad samas suunas kui tõelised hinded.

$$\text{HEKP} = 100 * (\text{õigesti ennustatud suundade arv} / \text{koguarv})$$

Absoluutset viga ei saa kasutada, kuna treeningandmetes on tunnuste väärtused tavaliselt lähedased ennustatavale väärtusele, mistõttu kasutame ruutkeskmist viga.

Parima mudeli iseloomustab minimaalsed RSME ja vahede RSME ning maksimaalne hinnaennustuse kvaliteedi protsent.

5 Tulemuste analüüs ja optimeerimine

Hüpoteesi tõestamiseks oli algselt treenitud 4 erinevat ennustusmudelit, mille treeningandmeteks oli LHV kasvukontos saadavate ning 100 kõige aktiivsema kauplemisega börsil kaubeldavate fondide hinnaseeriad kuni 15.04.2019. Mudelite loetelu:

- 50 puuga otsustusmets
- 100 puuga otsustusmets
- 150 puuga otsustusmets
- 200 puuga otsustusmets

Parimad treenimise tulemused (iteratsioonide arv 50) vastavalt enne välja toodud testimise metoodikale:

Mudeli nimi	RSME	Vahe RSME	Suuna täpsus
50 puuga otsustusmets	1,5135	1,1840	93,2%
100 puuga otsustusmets	1,6540	1,1250	93,3%
150 puuga otsustusmets	1,4060	1,1470	93,4%
200 puuga otsustusmets	1,6146	1,1241	93,5%
Keskmine	1,547025	1,145025	93,35%

Tabel 4. Optimeerimata ennustusmudelite parimad tulemused

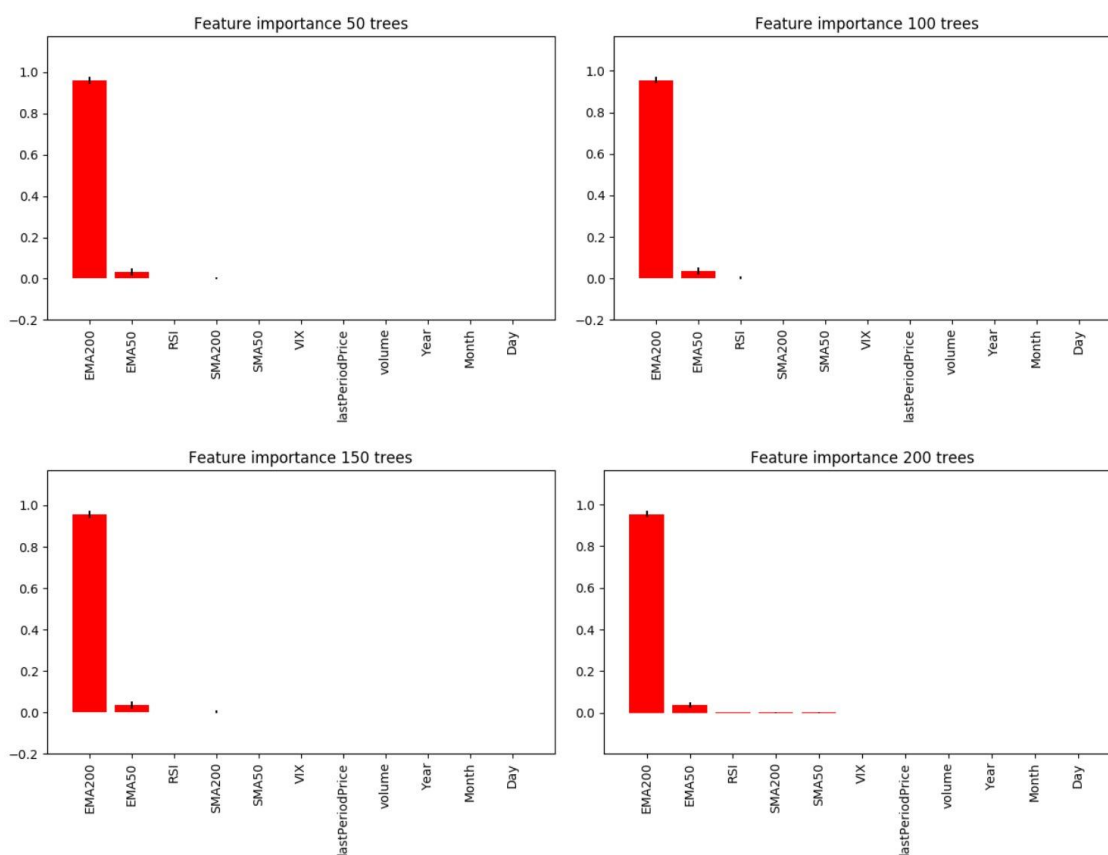
Näeme, et puude arvu kasvuga tulemused ei muutu eriti paremaks, millest võime järeldada, et puude arvust antud parameetritega tulemused ei parane, kuid õpetamise parameetrite muutmise on võimalik tulemusi parandada.

Lisades 1-4 on välja toodud 15.05.2019 ennustamistulemused ning võrdlused reaalsete hindadega 15.04.2019 andmete põhjal. Antud tulemuste põhjal kinnitavad üles seatud hüpoteesi, et fondi hindu on võimalik ennustada eelmise kuu majandusindikaatorite põhjal.

5.1 Ennustumudelite optimeerimine

Eelnevalt leitud mudelid ei ole kuidagi optimeeritud ning olid loodud hüpoteesi tõestamiseks. Esialgu ei ole muudetud ühtegi otsustusmetsa parameetri õpetamise muutmiseks. Parimate õpetamise parameetrite otsimiseks rakendatakse hüperparameetrite optimeerimise algoritmi.

Otsustusmets pakub võimalusi muuta, kui tähtis mudelis on iga talle ette antud tunnus. Selleks genereeritakse mudelite põhjal tunnuste tähtsuse graafikuid.



Joonis 4. Algsete mudelite tunnuste tähtsuse graafikud. Autori joonis

Graafikutel näeme, et kõige tähtsamad parameetrid on EMA200 ja EMA50, lisaks võib sisse jätta ka RSI, SMA50 ja SMA200. Ülejäänud parameetrid võib välja lõigata ning selle tulemusel teha mudeli kiiremaks ja vähendada failisuurust sarnaseid tulemusi säilitades.

5.1.1 Hüperparameetrite optimeerimine

Masinõppes on hüperparameetrite optimeerimine või häälestamine optimaalsete parameetrite kogumi valimise probleem õppimisalgoritmi jaoks. Hüperparameeter on

parameeter, mille väärtust kasutatakse õppeprotsessi juhtimiseks. Seevastu on teiste parameetrite väärtused (näiteks sõlmkaalud) treenitavad masinõppe jooksul. [20]

Sama tüüpi masinõppe mudel võib erinevate andmemustrite üldistamiseks nõuda erinevaid piiranguid, ka alusid või õppimismäärasid. Neid meetmeid nimetatakse hüperparameetriteks ja neid tuleb kohandada, et mudel saaks masinõppe probleemi optimaalselt lahendada. Hüperparameetrite optimeerimine leiab optimaalse mudeli abil hüperparameetrite arvu, mis minimeerib antud sõltumatute andmete puhul eelmääratletud kahjumifunktsiooni. [22]

Antud töös kasutatakse hüperparameetrite optimeerimiseks *randomized search* algoritmi. Algoritm teeb n iteratsiooni, iga iteratsiooni puhul valib ta suvalised parameetrid, arvutab tulemust ning võrdleb hetke parima tulemusega. Algoritm väljastab parimad parameetrid. Hüperparameetrite otsimiseks oli kasutatud scikit-learn teegis olev klass RandomSearchCV.

Parameetri nimi	Kirjeldus	Proovitud väärtused	Tulemus
n_estimators	Otsustuspuude arv otsustusmetsas.	50, 75, 100, 125, 150, 175, 200	150
max_features	Tunnuste arv, mida tuleks parima jaotuse otsimisel arvesse võtta.	Auto – kõik tunnused Sqrt – ruut juur tunnuste arvust	sqrt
max_depth	Puu maksimaalne sügavus.	10-110 ja puudub	30
min_samples_split	Minimaalne näidiste arv sõlme jaotuseks.	2, 5, 10	2
min_samples_leaf	Minimaalne näidiste arv, et tekitada puu lehe.	1, 2, 4	1
bootstrap	Kas asendada osa treening andmetest keskmiste väärtustega?	True, False	False

Tabel 5. Hüperparameetrite tabel

Tabelis välja toodud parimate parameetritega oli treenitud 30 otsustusmetsa ning sellest valitud parim üleval toodud ennustamise testimise metoodika järgi.

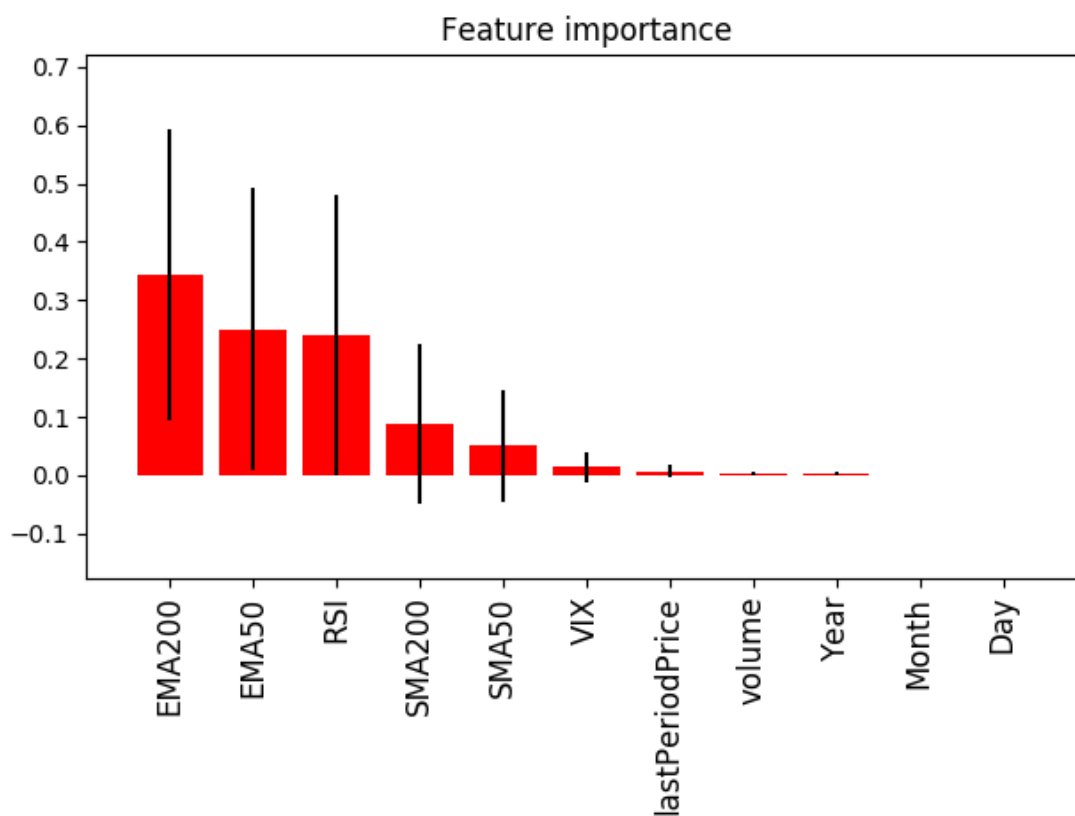
Mudeli nimi	RSME	Vahe RSME	Suuna täpsus
Hüperparameetritega otsustusmets	1,4627	1.0077	94,1%

Algsete otsustumetsade keskmine	1,547025	1,145025	93,35%
---------------------------------	----------	----------	--------

Tabel 6. Optimeeritud otsustumetsa tulemuste võrdlus algsetega

Antud tulemuste põhjal võib tõdeda, et tulemused on natuke paranenud.

Kuna parimate hüperparameetrite seas oli ka maksimaalsete tunnuste arvu muutmise parameeter, siis vaatleme ka tunnuste tähtsust.



Joonis 5. Optimeeritud otsustumetsa tunnuste tähtsused. Autori joonis.

Ilmnes, et mudel hakkas kasutama rohkem tunnuseid ennustuste tegemiseks ning selle abil suutis tulemusi parandada.

Lisa 5. on välja toodud 15.05.2019 ennustamistulemused ning võrdlused reaalse hindadega 15.04.2019 andmete põhjal.

5.2 Optimeerimisvõimalused tulevikus

Antud töös oli kasutatud ainult hüperparameetrite optimeerimise võimalus, kuid mudelit saaks ka edasi arendada ning tulemusi teha veel paremateks.

5.2.1 Suuna klassifitseerimine

Võrdlusest on küll välja toodud suuna ennustamise täpsus, kuid see ei olnud otseselt ennustamise eesmärgiks, seega mudel ei ürita suunda optimeerida.

Üks võimalustest on muuta ennustamise tulemust läbi korrigeeritud sulgemishinna asendamise börsil kaupleva fondi hinna muutmise suunaga ning uue välja, mille väärtusteks võiks näiteks olla [1;2], kus 1 oleks langev suund ja 2 oleks tõusev suund, lisamise. Arvutada iga hinnaseeria jaoks, kas selle hind kasvas või mitte ning asendada seda suunaga.

Selle tulemusel on kergem arvutada mudeli täpsust, kuna siis saab kasutada *confusion matrix*'it, mis näitab, millised klassifitseerimise tulemused on õiged ja mis mitte.

5.2.2 Lisatunnused

Lisatunnustustena saaks kasutada näiteks fondi käsitlevate riikide majandusindikaatoreid, üldiseid majandusnäitajaid ja fondides käsitlevate aktsiate heade uudiste osakaal.

6 Kokkuvõte

Töö eesmärgiks oli tõestada hüpoteesi, et on võimalik ennustada väärtpaberite hindu kasutades majandusindikaatoreid masinõppes. Selle jaoks loodi rakendus, mis suudab ette valmistada väärtpaberite andmeid ja nende baasil luua mudelid börsil kaubeldavate fondide hindade ennustamiseks otsustusmetsa abil. Rakendus on tehtud nii, et seda oleks võimalik ka kasutada teiste ennustuste perioodide jaoks, näiteks selgitada välja, mis on väärtpaberi hind nädala pärast. Lisaks väikeste muudatustega on võimalik vahetada börsil kauplevaid fonde teiste väärtpaberite tüüpide vastu ning teha nende põhjal ennustusi.

Põhiraskusteks oli mudelite ülesobituse vältimine, õigete andmete otsimine ning andmete masinõppe jaoks ettevalmistamine.

Algsed loodud mudelid näitasid häid tulemusi, ning suuna määramise täpsus oli 90% juures. Nende mudelite analüüsimisel tuli välja, et nad pidasid oluliseks ja tegid otsuseid ainult mõnede tunnuste põhjal. Nendeks tunnustusteks osutasid 200 ja 50 päeva eksponentsiaalsed liikuvad keskmised. Sellest tulenevalt oli kõige efektiivsem optimeerimise tee sundida mudelit kasutada nii palju tunnuseid kui võimalik.

Selle optimeerimise tee tõestamiseks oli kasutatud hüperparameetrite optimeerimise algoritmi. Algoritmi tulemuseks tuli mudel, kus tunnuste arv, mida tuleks parima jaotuse otsimisel arvesse võtta oli piiratud, mis andis väikse, aga nähtava täpsuse tõusu ning mudel hakkas kasutama peaaegu kõik tunnused.

Lõppmudeli on ka võimalik edasi optimeerida, lisades uusi tunnuseid või muutes hindade regressiooni suundade klassifikatsiooniks.

Töös realiseeritud rakendust saab kasutada igakuiste investeerimise otsuste tegemisel ning edasi optimeerides on võimalik saada veel paremaid tulemusi.

Kasutatud kirjandus

- [1] „Oxford dictionary,“ Oxford dictionary, [Võrgumaterjal]. Available: <https://en.oxforddictionaries.com/definition/overfitting>. [Kasutatud 10 05 2019].
- [2] „Investopedia,“ 23 04 2019. [Võrgumaterjal]. Available: <https://www.investopedia.com>. [Kasutatud 10 05 2019].
- [3] „investorplace,“ [Võrgumaterjal]. Available: <https://investorplace.com/how-to-invest/funds/etfs/passive-vs-active-etf/>. [Kasutatud 12 05 2019].
- [4] AS LHV Pank, „Investeerimisõpik,“ [Võrgumaterjal]. Available: <https://fp.lhv.ee/academy/investmentguide?locale=et>. [Kasutatud 11 05 2019].
- [5] „ETFs vs. mutual funds: A comparison,“ Vanguard. [Võrgumaterjal]. [Kasutatud 12 05 2019].
- [6] M. B.G., „Efficient Market Hypothesis,“ *The New Palgrave*, pp. 127-134, 1989.
- [7] B. G. Malkiel, „The Efficient Market Hypothesis and Its Critics,“ *Journal of Economic Perspectives*, nr 17 (1), pp. 59-82, 2003.
- [8] „Eesti Keskerakonna, Eesti Konservatiivse Rahvaerakonna ning Isamaa Erakonna valitsusliidu aluspõhimõtted 2019-2023“.
- [9] Saar Poll OÜ, „Finantsalase kirjaoskuse uuringu Eesti elanike seas,“ 2015.
- [10] „What is Machine Learning? Machine Learning For Beginners,“ [Võrgumaterjal]. Available: <https://www.edureka.co/blog/what-is-machine-learning/>. [Kasutatud 10 05 2019].
- [11] T. K. Ho, „Random decision forests,“ *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1995.
- [12] K. Hewa, „A Beginners Guide to Random Forest Regression,“ [Võrgumaterjal]. Available: <https://medium.com/datadriveninvestor/random-forest-regression-9871bc9a25eb>. [Kasutatud 10 05 2019].
- [13] S. Gu, „How Random Forest Algorithm Works in Machine Learning,“ Synced, 24 10 2017. [Võrgumaterjal]. Available: <https://syncedreview.com/2017/10/24/how-random-forest-algorithm-works-in-machine-learning/>. [Kasutatud 10 05 2019].
- [14] D. S. Sayad, „An Introduction to Data Science,“ [Võrgumaterjal]. Available: <https://www.saedsayad.com/>. [Kasutatud 12 05 2019].
- [15] T. Hastie, R. Tibshirani ja J. Friedman, *The Elements of Statistical Learning*, Springer, 2009.
- [16] „Spring,“ Pivotal software, [Võrgumaterjal]. Available: spring.io. [Kasutatud 12 05 2019].
- [17] „Technopedia,“ [Võrgumaterjal]. Available: techopedia.com. [Kasutatud 12 05 2019].
- [18] „Scikit-learn,“ [Võrgumaterjal]. Available: scikit-learn.org. [Kasutatud 12 05 2019].

- [19] A. Ronacher, „Flask,“ [Võrgumaterjal]. Available: flask.pocoo.org. [Kasutatud 12 05 2019].
- [20] „Wikipedia,“ [Võrgumaterjal]. Available: wikipedia.org. [Kasutatud 12 05 2019].
- [21] „pandas,“ [Võrgumaterjal]. Available: pandas.pydata.org. [Kasutatud 12 05 2019].
- [22] M. Claesen ja B. De Moor, „Hyperparameter Search in Machine Learning,“ 2015.
- [23] L. S. Sterling, The Art of Agent-Oriented Modeling, London: The MIT Press, 2009.

Lisa 1 – 50 puuga otsustusmetsa ennustamistulemused

Sümbol	Reaalne hind	Ennustatud hind	Eelmise kuu hind	Suund
2B76	6,558	6,94376	6,981	+
2B77	5,042	5,17176	5,181	+
2B78	5,466	5,7814	5,757	-
CI2	506,8364	523,964932	540,4394	+
DBXV	26,755	26,487604	26,675	-
EXXT	65,24	65,1914	65,58	+
GLUX	120,14	120,559806	123,16	+
IBCD	102,445	100,795366	101,055	-
IBCQ	101,525	100,607218	101,195	-
IQQH	5,139	5,11108	5,152	+
IQQI	25,6	25,422422	25,415	+
IQQQ	37,265	37,090766	37,685	+
ISPA	27,15	27,5933	27,685	+
IUSL	35,595	35,882542	36,22	+
SPYC	191,2	188,326	188,28	+
SPYD	48,79	49,09081	49,445	+
SPYJ	33,595	33,04691	33,245	-
SPYW	22,84	22,850844	23,01	+
SXR1	131,3	133,1098	133,88	+
SXR8	250,49	240,624448	252,28	+
VAL	108,5	112,551568	112,84	+
VWRL	65,18	65,09357	65,56	+
XCHA	11,3025	12,3288	12,2675	-
XD9U	68,914	69,166908	69,408	+
XDWD	54,532	55,061002	55,154	+
XDWH	28,455	29,143142	28,88	-
XRS2	191,88	194,85121	194,18	-

Tabel 7. 15. mai ETF hindade ennustamise tulemus 50 puuga otsustusmetsa puhul

Lisa 2 – 100 puuga otsustusmetsa ennustamistulemused

Sümbol	Reaalne hind	Ennustatud hind	Eelmise kuu hind	Suund
2B76	6,558	6,94933	6,981	+
2B77	5,042	5,16603	5,181	+
2B78	5,466	5,80145	5,757	-
CI2	506,8364	530,279855	540,4394	+
DBXV	26,755	26,5334395	26,675	-
EXXT	65,24	65,282863	65,58	+
GLUX	120,14	120,156501	123,16	+
IBCD	102,445	101,146068	101,055	+
IBCQ	101,525	101,029883	101,195	-
IQQH	5,139	5,1215	5,152	+
IQQI	25,6	25,322642	25,415	-
IQQQ	37,265	37,0907	37,685	+
ISPA	27,15	27,528	27,685	+
IUSL	35,595	35,979357	36,22	+
SPYC	191,2	188,5386	188,28	+
SPYD	48,79	49,285835	49,445	+
SPYJ	33,595	32,797666	33,245	-
SPYW	22,84	22,79397	23,01	+
SXR1	131,3	133,640813	133,88	+
SXR8	250,49	241,746989	252,28	+
VAL	108,5	112,511919	112,84	+
VWRL	65,18	65,24732	65,56	+
XCHA	11,3025	12,346625	12,2675	-
XD9U	68,914	69,140387	69,408	+
XDWD	54,532	55,009768	55,154	+
XDWH	28,455	28,7507975	28,88	+
XRS2	191,88	195,177873	194,18	-

Tabel 8. 15. mai ETF hindade ennustamise tulemus 100 puuga otsustusmetsa puhul

Lisa 3 – 150 puuga otsustusmetsa ennustamis tulemused

Sümbol	Reaalne hind	Ennustatud hind	Eelmise kuu hind	Suund
2B76	6,558	6,91886	6,981	+
2B77	5,042	5,171856667	5,181	+
2B78	5,466	5,796098333	5,757	-
CI2	506,8364	527,3136293	540,4394	+
DBXV	26,755	26,4846	26,675	-
EXXT	65,24	64,84012867	65,58	+
GLUX	120,14	120,1855953	123,16	+
IBCD	102,445	100,807408	101,055	-
IBCQ	101,525	101,117368	101,195	-
IQQH	5,139	5,134	5,152	+
IQQI	25,6	25,15076	25,415	-
IQQQ	37,265	37,31045933	37,685	+
ISPA	27,15	27,42526667	27,685	+
IUSL	35,595	35,92029467	36,22	+
SPYC	191,2	188,221442	188,28	-
SPYD	48,79	49,195598	49,445	+
SPYJ	33,595	32,85144533	33,245	-
SPYW	22,84	22,83851133	23,01	+
SXR1	131,3	133,529262	133,88	+
SXR8	250,49	242,5474567	252,28	+
VAL	108,5	112,4211273	112,84	+
VWRL	65,18	65,38868267	65,56	+
XCHA	11,3025	12,36159733	12,2675	-
XD9U	68,914	69,30173867	69,408	+
XDWD	54,532	55,00341933	55,154	+
XDWH	28,455	28,82538111	28,88	+
XRS2	191,88	194,5296247	194,18	-

Tabel 9. 15. mai ETF hindade ennustamise tulemus 150 puuga otsustusmetsa puhul

Lisa 4 – 200 puuga otsustusmetsa ennustamistulemused

Sümbol	Reaalne hind	Ennustatud hind	Eelmise kuu hind	Suund
2B76	6,558	6,922295	6,981	+
2B77	5,042	5,165425	5,181	+
2B78	5,466	5,8674175	5,757	-
CI2	506,8364	523,813897	540,4394	+
DBXV	26,755	26,560645	26,675	-
EXXT	65,24	64,503605	65,58	+
GLUX	120,14	118,8073215	123,16	+
IBCD	102,445	100,959139	101,055	-
IBCQ	101,525	100,954332	101,195	-
IQQH	5,139	5,136895	5,152	+
IQQI	25,6	25,2872515	25,415	-
IQQQ	37,265	37,05103617	37,685	+
ISPA	27,15	27,503725	27,685	+
IUSL	35,595	35,884755	36,22	+
SPYC	191,2	188,390589	188,28	+
SPYD	48,79	49,237214	49,445	+
SPYJ	33,595	32,9390495	33,245	-
SPYW	22,84	22,6236115	23,01	+
SXR1	131,3	133,7399935	133,88	+
SXR8	250,49	241,153664	252,28	+
VAL	108,5	112,588857	112,84	+
VWRL	65,18	65,2694675	65,56	+
XCHA	11,3025	12,3726945	12,2675	-
XD9U	68,914	68,9167185	69,408	+
XDWD	54,532	55,0232275	55,154	+
XDWH	28,455	28,89222833	28,88	-
XRS2	191,88	195,6101445	194,18	-

Tabel 10. 15. mai ETF hindade ennustamise tulemus 200 puuga otsustusmetsa puhul

Lisa 5 – hüperparameetritega otsustusmetsa ennustamistulemused

Sümbol	Reaalne hind	Ennustatud hind	Eelmise kuu hind	Suund
2B76	6,558	6,7594325	6,981	+
2B77	5,042	5,134	5,181	+
2B78	5,466	5,8003	5,757	-
CI2	506,8364	511,821894	540,4394	+
DBXV	26,755	26,65482867	26,675	-
EXXT	65,24	65,31599667	65,58	+
GLUX	120,14	121,9303167	123,16	+
IBCD	102,445	101,9471687	101,055	+
IBCQ	101,525	101,1275107	101,195	-
IQQH	5,139	5,251764	5,152	-
IQQI	25,6	25,69333867	25,415	+
IQQQ	37,265	37,76863867	37,685	-
ISPA	27,15	27,32868733	27,685	+
IUSL	35,595	36,60079267	36,22	-
SPYC	191,2	190,1016947	188,28	+
SPYD	48,79	49,17218	49,445	-
SPYJ	33,595	33,08540267	33,245	-
SPYW	22,84	22,97650133	23,01	+
SXR1	131,3	132,4724793	133,88	+
SXR8	250,49	251,035258	252,28	+
VAL	108,5	110,3322667	112,84	+
VWRL	65,18	65,90800667	65,56	-
XCHA	11,3025	12,63941	12,2675	-
XD9U	68,914	69,50762067	69,408	+
XDWD	54,532	55,19738933	55,154	+
XDWH	28,455	28,60104667	28,88	+
XRS2	191,88	195,2322173	194,18	-

Tabel 9. 15. mai ETF hindade ennustamise tulemus optimeeritud parameetritega otsustusmetsa puhul