

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Kadri-Liis Kutser 211720IABM

Mitme sagedusega finantsandmete aegrea prognoosimise meetodite võrdlusanalüüs

Magistritöö

Juhendaja: Olga Dunajeva
Phd

Tallinn 2024

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: Kadri-Liis Kutser

08.05.2024

Annotatsioon

Käesolev magistritöö uurib finantsvarade sulgemishindade aegridade prognoosimist, keskendudes mitme ajasagedusega andmete kasutamisele. Finantsprognoosimise olulisus peitub selle võimes anda väärtuslikku teavet tuleviku trendide ja mustrite kohta, mis omakorda mõjutavad otsuste tegemist mitmete huvigruppide jaoks.

Selles magistritöös on eesmärk ehitada ja võrrelda statistilisi, masinõppe- ja sügavõppemudeleid, mis võtavad arvesse finantsaegridade tunni- ja päevatäpsusega andmeid ning prognoosivad nende tuleviku väärtusi tunni täpsusega. Olulisimaks küsimuseks on, kuidas mõjutab mitme ajaintervalli kasutamine prognoosimise täpsust ja milline meetod osutub kõige täpsemaks.

Töö tulemused näitavad, et mitme ajaintervalli kasutamine parandas statistiliste meetodite, otsustusmetsa ja mitmekihilise närvivõrgu meetodi puhul prognoosi tulemusi. Tugivektorregressiooni ja rekurrentsete närvivõrkude meetodite puhul ei pakkunud täiendav ajaintervall lisandväärtust. Üldiselt osutusid kõige täpsemateks meetoditeks tugivektorregressioon ja sügavõppemeetoditest värvatega rekurrentne üksus.

Kokkuvõttes annab see uurimus olulise panuse finantsprognoosimise valdkonnas, andes ülevaate mitmesuguste meetodite sobivusest selle ülesande täitmisel ja avades uksi edasistele uurimissuundadele.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 53 leheküljel, 7 peatükki, 14 joonist, 4 tabelit.

Abstract

Comparative analysis of time series forecasting methods for multi-frequency financial data

This master's thesis investigates the forecasting of financial asset closing prices, focusing on the use of multi-frequency data. The importance of financial forecasting lies in its ability to provide valuable insights into future trends and patterns, which in turn influence decision-making for various stakeholders.

The aim of this thesis is to build and compare statistical, machine learning, and deep learning models that consider both hourly and daily frequency financial time series data, forecasting their future values at an hourly granularity. The main question addressed is how the use of multiple time intervals affects the accuracy of forecasting and which method proves to be the most accurate.

The results of the study show that employing multiple time intervals improves forecast accuracy for statistical methods, random forests, and multilayer perceptron models. However, for support vector regression and recurrent neural networks, the additional time interval did not provide added value. Overall, support vector regression and the gated recurrent unit from recurrent neural networks methods emerged as the most accurate techniques.

In conclusion, this research makes a significant contribution to the field of financial forecasting by providing insight into the suitability of various methods for this task and opening doors to further research avenues.

The thesis is in Estonian and contains 53 pages of text, 7 chapters, 14 figures, 4 tables.

Lühendite ja mõistete sõnastik

ANN	<i>Artificial Neural Network</i> , tehisnärvivõrk
AR	Autoregressiivne
ARIMA	<i>Autoregressive Integrated Moving Average</i> , autoregressiivne integreeritud libiseva keskmise meetod
biLSTM	<i>Bidirectional Long Short-Term Memory</i> , kahe-suunaline pikk lühiajaline mälu
CNN	<i>Convolutional Neural Network</i> , konvolutsiooniline närvivõrk
DFT	<i>Discrete Fourier Transform</i> , diskreetne Fourieri teisendus
DT	<i>Decision Tree</i> , otsustuspuu
ETS	<i>Exponential Smoothing</i> , eksponentsiaalne silumine
FFNN	<i>Feedforward Neural Networks</i> , edasisuunalised närvivõrgud
GRU	<i>Gated Recurrent Unit</i> , väravatega rekurrentne üksus
IFT	<i>Inverse Fourier Transform</i> , Fourieri pöördteisendus
LSTM	<i>Long short-term memory</i> , pikk lühiajaline mälu
M3 võistlus	Empiirilised uuringud, mis võrdlevad peamiste aegriidade meetodite suutlikkust erialaspetsialistide abil
MA	<i>Moving Average</i> , libisev keskmine
MAE	<i>Mean Absolute Error</i> , keskmine absoluutviga
MAPA	<i>Multi Aggregation Prediction Algorithm</i> , mitme agregatsiooni prognoosimisalgoritm
MAPE	<i>Mean Absolute Percentage Error</i> , keskmine suhteline absoluutviga
MLP	<i>Multilayer Perceptron</i> , mitmekihiline närvivõrk
MSE	<i>Mean Squared Error</i> , keskmine ruutviga
RF	<i>Random Forest</i> , otsustusmets
RMSE	<i>Root Mean Squared Error</i> , ruutkeskmine viga
RNN	<i>Recurrent Neural Network</i> , rekurrentne ehk korduv närvivõrk
SFM	<i>State Frequency Memory</i> , olekufrekvents mälu
SGBM	<i>Stochastic Gradient Boosting Machine</i> , stohhastilise gradiendi võimendamise masin
SMAPE	<i>Symmetric Mean Absolute Percentage Error</i> , sümmeetriline keskmine suhteline absoluutviga

SVM *Support Vector Machine*, tugivektorklassifitseerija
SVR *Support Vector Regression*, tugivektorregressioon
XGBoost *Extreme Gradient Boosting*, äärmuslik gradienttõstmine

Sisukord

1 Sissejuhatus	11
2 Taust	13
2.1 Kirjanduse ülevaade	14
2.1.1 Finantsandmete aegridade prognoosimise meetodid.....	15
2.1.2 Hübriidmudelid.....	17
2.1.3 Mitme sagedusega finantsandmete aegrea prognoosimise meetodid.....	18
3 Metoodika ja andmed	21
3.1 Traditsioonilised statistilised aegrea meetodid.....	21
3.1.1 Autoregressiivne integreeritud libiseva keskmise mudel ARIMA.....	21
3.1.2 Eksponentsiaalne silumine	22
3.2 Masinõppe meetodid aegridade prognoosimiseks.....	23
3.2.1 Otsustusmets.....	23
3.2.2 Tugivektorregressioon.....	23
3.3 Sügavõppemeetodid aegridade prognoosimiseks.....	24
3.3.1 Mitmekihiline närvivõrk.....	25
3.3.2 Rekurrentne võrk.....	25
3.3.3 Pikk lühiajaline mälu.....	27
3.3.4 Väravatega rekurrentne üksus	29
3.4 Finantsandmed.....	30
3.5 Kasutatud tarkvara.....	32
4 Analüüs.....	33
4.1 Andmete eeltötlus analüüsiks	33
4.1.1 Treening- ja testandmete valik	35
4.2 Analüüsi arhitektuur	36
4.2.1 Mitme sagedusega mudelite meetodid	37
4.3 Analüüsis kasutatavad veamõõdikud.....	38
4.4 Mudelite seadete valik.....	40
4.4.1 Eksponentsiaalne silumine	40
4.4.2 ARIMA.....	40

4.4.3 Masinõppe ja sügavõppe meetodid	40
5 Tulemused	44
5.1 Madala, päevase sagedusega analüüs	44
5.2 Kõrge, tunnise sagedusega analüüs	48
5.3 Mitme sagedusega andmete analüüs.....	51
6 Järeldused ja edasine arutelu	60
7 Kokkuvõte	63
Kasutatud kirjandus	64
Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks	69
Lisa 2. Statistilised testid	70
Lisa 3. Teisendatud andmestik	71
Lisa 4. Madala sagedusega prognoosi veamõõdikud finantsvara ja meetodi lõikes	72
Lisa 5. Kõrge sagedusega prognoosi veamõõdikud finantsvara ja meetodi lõikes	74
Lisa 6. Mitme sagedusega prognoosi veamõõdikud finantsvara ja meetodi lõikes.....	76
Lisa 7. Mitme sagedusega prognoosi veamõõdikud finantsvara ja meetodi lõikes, kasutades hierarhilist ansambelõpet	78
Lisa 8. Mitme sagedusega prognoosi veamõõdikud finantsvara ja meetodi lõikes, kasutades ansambelõpet.....	80
Lisa 9. Ühe ajasagedusega mudelite lõplikud seadistused	82
Lisa 10. Mitme sagedusega mudelite lõplikud seadistused.....	84

Jooniste loetelu

Joonis 1. Otsustusmetsa arhitektuur [27].....	23
Joonis 2. Tugivektroregressiooni arhitektuur [29].	24
Joonis 3. Mitmekihilise närvivõrgu arhitektuur [31].....	25
Joonis 4. Rekurrentse võrgu (RNN) arhitektuur [33].....	26
Joonis 5. Pika lühiajalise mälu (LSTM) võrgu arhitektuur [35].....	28
Joonis 6. Väravatega rekurrentse üksuse (GRU) struktuur [37].....	29
Joonis 7. Finantsandmed [38] [39].	33
Joonis 8. Analüüsi aluseks võetud võrreldavate mudelite seadistused ja arhitektuur. ...	36
Joonis 9. Aegridade prognoosimise veamõõdikud [47].	39
Joonis 10. S&P 500 tunni ja mitme intervalliga mudelite prognoositulemuste võrdlus.	52
Joonis 11. Dow Jones tunni ja mitme intervalliga mudelite prognoositulemuste võrdlus.	54
Joonis 12. Nasdaq 100 tunni ja mitme intervalliga mudelite prognoositulemuste võrdlus.	55
Joonis 13. Bitcoin tunni ja mitme intervalliga mudelite prognoositulemuste võrdlus. .	56
Joonis 14. Ethereum tunni ja mitme intervalliga mudelite prognoositulemuste võrdlus.	57

Tabelite loetelu

Tabel 1. Finantsandmete päevase aegrea kirjeldav statistika.	31
Tabel 2. Finantsandmete tunnise aegrea kirjeldav statistika.	31
Tabel 3. S&P 500 indeksi aegrea normaaljaotuse ja statsionaarsuse testi tulemused. ...	34
Tabel 4. Treening- ja testandmete ajavahemikud.	35

1 Sissejuhatus

Tuleviku ennustamine on aegade algusest olnud inimkonna jaoks huvipakkuv ülesanne. Tänapäeval on tuleviku prognoosimine vähem ennustajate ja pigem statistikute või majandusteadlaste töö. Sarnaselt tegeleb käesolev magistritöö just prognoosimisega, keskendudes finantsturgude andmetele.

Aegridade prognoosimine on statistiline tehnika, mida on aastaid kasutatud konkreetse aegrea tulevaste väärtuste ennustamiseks selle varasemate vaatluste põhjal. Aegridade prognoosimist on rakendatud mitmetes erinevates valdkondades nagu rahandus, majandus, meteoroloogia, aga ka tehnika ja meditsiin. Aegridade prognoosimise olulisus põhineb selle võimel anda väärtuslikku teavet tuleviku trendide ja mustrite kohta. See on kasulik tegevus abistamiseks nii avaliku sektori kui ka eraettevõtete ja organisatsioonide otsustusprotsessis.

Aegridade prognoosimine on siiski üsna keeruline ülesanne, kuna see hõlmab andmete aluseks olevate mustrite ja trendide modelleerimist, mida mõjutavad paljud tegurid, nagu tsüklid, trendid ja hooajalisus. Lisaks võivad aegridade andmed olla mürarikkad ja sisaldada kõrvalekaldeid, mis raskendavad täpsete prognooside tegemist. Aegridade prognooside täpsus on väga oluline, kuna sellel võib olla otsene mõju äriotsustele. Näiteks finantsturgudel võib prognooside ebatäpsus tuua kaasa nõrgad investeerimisotsused, mis võivad omakorda põhjustada märkimisväärsed kahjusid. Finantsandmete aegridade prognoosimise kontekstis on täpsuse olulisus eriti märgatav ja see on ka käesoleva lõputöö fookus.

Aegridade prognoosimiseks on läbi aegade kasutatud mitmeid meetodeid, sealhulgas traditsioonilisi statistilisi meetodeid, nagu autoregressiivne integreeritud libiseva keskmise mudel (ARIMA) aga ka masinõppel põhinevaid meetodeid. Meetodi valik sõltub peamiselt andmete olemusest ja ennustushorisonidist. Lihtsad meetodid, nagu eksponentsiaalne silumine ja libisev keskmine, võivad olla sobivamad lühiajaliseks prognoosimiseks, samas kui keerukamad meetodid, nagu ARIMA ja masinõppel põhinevad mudelid, sobivad paremini pikaajaliseks prognoosimiseks. Eelnimetatud

meetodite kõrval on viimastel aastatel populaarsemaks muutunud ka sügavõppel põhinevad meetodid, mis annavad paljulubavaid tulemusi [1]. Kõige tähelepanuväärsemad aegridade ennustamiseks kasutatavad sügavõppe meetodid on rekurrentne närvivõrk (RNN) ja pika lühiajalise mälu (LSTM) võrk [1]. Need on osutunud kasulikuks aegrea andmete keerukate mustrite modelleerimisel ja on tõestanud, et need on paljudel juhtudel paremad kui traditsioonilised meetodid.

Lõputöö eesmärk on aidata kaasa finantsandmete aegridade prognoosimise ja modelleerimise edendamisele, kasutades ära selliste andmete mitme ajasageduse olemust. Töö lõppeesmärgiks on ehitada statistiliste, masinõppe ja sügavõppe meetoditega mudelid, mis võtavad arvesse finantsaegridade tunni- ja päevatäpsusega andmeid korruga ning prognoosivad aegridade tulevikuväärtusi tunni täpsusega. Väljatöötatud lähenemisviise võrreldakse enamlevinud finantsaegridade ühele ajasagedusega andmetele rakendatud prognoosimise meetoditega, et selgitada, kas ja mil määral parandab täiendava ajaintervalli lisamine prognoositäpsust. Mudelite võrdlemiseks kasutatakse erinevaid veamõõdikuid, mis kirjeldavad nende meetodite prognoosimise täpsust. Järgnev nimekiri kirjeldab konkreetsemalt töös lahendatavaid ülesandeid:

1. Kuidas muutub finantsandmete prognoosivõime kasutades statistiliste aegridade meetodite asemel masinõppe ja sügavõppe meetodeid?
2. Mil määral paraneb finantsandmete prognoosivõime kasutades ühe ajasageduse asemel mitut?
3. Kui täpselt suudavad mitme ajaintervalliga andmete põhjal prognoosida erinevad algoritmid?
4. Milline on täpseim meetod finantsandmete prognoosimiseks? Kas see on kõige optimaalsem arvestades vajalikku lisatööd ja kuluvat aega?

Töö on jaotatud seitsmesse sektsiooni. Käesolevas esimeses osas antakse ülevaade töö sisust, eesmärgist ja lahendatavatest ülesannetest. Teises osas tehakse täpsem tutvustus probleemi olemusest ja taustast ning antakse ülevaade mitmesugustest varasemalt kasutusele võetud prognoosimise meetoditest, mida kajastatud olemasolevas kirjanduses. Kolmas osa keskendub töös kasutatavatele meetoditele ja andmetele ning nende täpsemale kirjeldusele. Töö neljas sektsioon võtab fookusesse töö käigus tehtud analüüsid ning viiendas osas järgneb analüüside põhjal saadud tulemuste ülevaade. Kuuendas osas on töö põhilised järeldused ja arutelu. Töö lõpetab seitsmes, kokkuvõttev osa.

2 Taust

Aegridade prognoosimise ajalugu ulatub tagasi 20. sajandi esimesse poole, kui Briti statistik Udeny Yule võttis päikeselaikude analüüsis esmakordselt kasutusele autoregressiivse mudeli [2]. See oli omakorda algus keerukama autoregressiivse integreeritud libiseva keskmise (ARIMA) mudeli väljatöötamisel 1970ndatel [2]. Tänapäeval on aga kiire masinõppe ja sügavõppe meetodite areng võimaldanud aegridade prognoosimisel võtta kasutusele veelgi keerukamaid ja nutikamaid lahendusvõimalusi, muutes selle maastiku väga mitmekesiseks. Töö käesolev osa annab ülevaate mitmesugustest meetoditest, mida varasemas kirjanduses on aegridade prognoosimisel rakendatud.

Finantsaegread on olnud paljude teadlaste, ökonomistide, aga ka aktiivsete investorite ja börsimaaklerite fookuses. Kuid aegread on iseenesest keerulised andmekogumid ja seetõttu on nende käitlemisel ja mineviku väärtuste põhjal tulevikuprognoside tegemisel mitmeid takistusi. Finantsturge iseloomustab nende mittelineaarne, dünaamiline ja ettearvamatu iseloom, mistõttu on varade hindade ja turusuundumuste ennustamine keeruline ülesanne. Finantsturge mõjutavad paljud erinevad tegurid – poliitilised ja majanduslikud sündmused, investorite tegevus, viimasel ajal ka sotsiaalmeedia jne – nii makro- kui ka mikromajanduslikud mõjutajad. Nende tegurite mõju finants-aegridade andmetele ei ole ilmselge – seetõttu on finantsandmete prognoosimine nende väliste tegurite ebakindluse ja sisemise keerukuse tingimustes kompleksne [3]. Finantsandmete aegridade täpne prognoosimine on olnud kvantitatiivse rahanduse valdkonnas kriitiline väljakutse aastakümneid ja jääb selleks ka tänapäeval. Teema on pälvinud suurt huvi, kuna on tehtud mitmeid pingutusi finantsturgude hinnakujunduse ja erinevate finantsvarade volatiilsuse prognoosimiseks. Täpsemalt on aktsiahinna ennustamine juba aastaid pälvinud suurt huvi selle otsese rahalise kasu tõttu.

Eelnevalt mainitud rahanduslike aegridade ja ka laiemalt võttes aegridade keerukus on andnud põhjuse ka selle lõputöö kirjutamiseks. Kuigi aktsiahinna ennustamist on paljud teadlased ja majandusteadlased aastaid uurinud, on see teema endiselt kompleksne.

Probleem seisneb selles, et investeerimise või börsil kauplemise ajahorisont on üsna erinev. Kui investor soovib teha passiivseid toiminguid ja tal on kümneaastane või veelgi pikem horisont, et tema investering oleks kasumlik, siis börsimaakler soovib tulusaid tulemusi väga lühikeste ajavahemike tagant ja on valmis oma portfellis aktiivselt muudatusi tegema. Seetõttu on finantsturgudele tehtavad prognoosid väga erinevad lähtuvalt ajahorisondist, iseloomust ja uuritavast kasumlikkuse eesmärgist. Lõputöö on koostatud eesmärgiga võtta neid erinevusi arvesse finants-aegridade mitme sagedusega prognoosimudelites.

2.1 Kirjanduse ülevaade

Finantsandmete ennustamiseks on kasutatud mitmesuguseid meetodeid. Temaatiline kirjandus kajastab meetodikate pidevat arengut. Traditsioonilised statistilised mudelid, nagu autoregressiivse integreeritud libiseva keskmise mudelid (ARIMA) ja nende erinevad versioonid, on oma lihtsuse ja tõlgendatavuse tõttu selle ülesande jaoks pikka aega kasutust leidnud [4].

Võimalus võtta kasutusele suurema arvutusvõimega arvutid on aidanud kaasa täiustatud meetodite populaarsuse kasvule. Masinõppe algoritmid, sealhulgas otsustusmetsad (*Random Forests*, RF) ja tugivektorklassifitseerija (*Support Vector Machine*, SVM), on populaarsust kogunud tänu nende võimele tabada andmetes mittelineaarseid seoseid [5]. Veelgi enam, sügavõppe tehnikad nagu tehisnärvivõrgud (*Artificial Neural Networks*, ANN) ja mitmekihiline närvivõrk (*Multilayer Perceptron*, MLP), eriti aga rekurrentsed närvivõrgud (*Recurrent Neural Networks*, RNN) ja pika lühiajalise mälu (*Long Short-Term Memory*, LSTM) võrgud, on näidanud paljutootavat edu aegridade andmete järjestikuste sõltuvuste käsitlemisel [6]. Kõik need masinõppel (*Machine Learning*) ja sügavõppel (*Deep Learning*) põhinevad meetodid ei ole enam pelgalt mudelipõhised andmeanalüüsi lähenemisviisid, vaid pigem andmepõhised meetodid, mis tähendab, et nende rakendamiseks ja väga täpsete prognooside tootmiseks on vaja suurt hulka andmeid, mis seeläbi suurendab analüüsiprotsessi keerukust [7].

Kuna finantsturud toimivad mitmel ajaskaalal, on oluline seda arvesse võtta. Viimasena mainitud sügavõppe tehnikad – RNN ja LSTM – võivad olla selliste omadustega andmetega töötamisel kõige kasulikumad. Selle lõputöö eesmärk on aga minna kaugemale, kaasates ennustustesse erinevad ajasagedused. Lõputöö tulemuseks luuakse

erinevatel meetoditel leitud kombineeritud prognoosimudelid, mille ehitamisel on kasutatud erinevaid ajasagedusi. Neid võrreldakse ühe ajaintervalliga andmetel loodud mudelite tulemustega, et selgitada, kas täiendava ajasageduse lisamine muudab ennustused täpsemaks.

2.1.1 Finantsandmete aegridade prognoosimise meetodid

Varasemast temaatilisest kirjandusest ilmneb, et üha enam uuritakse erinevate meetodite võimalusi mitmesuguste finantsaegridade tulevaste väärtuste ennustamiseks ja võrreldakse nende lähenemisviiside täpsust. Nende tööde fookus on olnud erinevat tüüpi finantsandmestikel – alustades üksikutest aktsiahindadest või nende rühmadest, aktsiaindeksitest, kulla hinnast ja lõpetades loomulikult ühe populaarseima valdkonnaga, milleks on krüptovaluutade hinna ennustamine. Järgnevad lõigud annavad lühikese ülevaate varem kirjutatud töödest seoses erinevate statistiliste, masinõppe ja sügavõppe meetoditega finantsaegridade prognoosimisel ja peamistest tulemustest nendes töedes.

Kulla hinna prognoosimisel on ühes varasemas artiklis võetud kasutusele tugivektorklassifitseerija (*Support Vector Machine*, SVM) [8]. Saadud prognooside jaoks leitud ruutkeskmise viga (*Root Mean Square Error*, RMSE) ja keskmine suhteline absoluutviga (*Mean Absolute Percentage Error*, MAPE) näitasid selgelt, et SVM ületas märkimisväärselt ARIMA meetodi täpsust. Selle uurimistöo tulemuste põhjal on lihtne mõista, miks SVM meetodid on oma täpsusega muutunud hinnaennustuste jaoks populaarseks.

Aktsiaturu indeksite prognoosimise osas keskendub üks viimaste aastate töö lihtsama statistilise ARIMA meetodi võrdlemisele keerukamate pika lühiajalise mälu võrgu (*Long Short-Term Memory*, LSTM) ja äärmusliku gradienttõstmise (*Extreme Gradient Boosting*, XGBoost) meetoditega [9]. Selles töös annavad üldiselt just viimased keerukamad mudelid parema täpsusega tulemuse, kuid kahel juhul suudab ARIMA anda isegi täpsemad ennustused. Nende kahe olukorra puhul tuleb aga silmas pidada, et nende andmekogumite väärtused on palju madalamad kui teistel juhtudel, mis võib omakorda tulemusi mõjutada. Enamikul juhtudel annab siiski LSTM kõige madalama veaväärtuse võrreldes teistega.

Mõnel juhul on autorid otsustanud minna edasi LSTM-i baasmeetodist ja lisada meetodile juurde teise suuna, muutes selle kahesuunaliseks (*Bidirectional LSTM*, biLSTM), nagu

see oli Namini jt [10] töös, kus aktsiaturuindeksite prognoosimiseks võrreldi ARIMA-t, LSTM-i ja BiLSTM-i. Eesmärk oli mõista, kuidas täiendavate treenimiskihtide lisamine analüüsi muudab selle väljundit. Selgelt andis just BiLSTM kõige täpsemad prognoosid. Kuid LSTM-i mudelid saavutavad treeningprotsessis stabiilse oleku palju kiiremini kui nende kahesuunalised variandid ja on seetõttu palju hõlpsamini kasutatavad.

Siiski kasutatakse aktsiaindeksite prognoosimisel ka kauem kasutusel olnud mudeleid. Näiteks võeti vaatluse alla SVM, MLP, RF ja stohhastilise gradiendi võimendamise masin (*Stochastic Gradient Boosting Machine*, SGBM), et võrrelda nende prognoosivõimet populaarsete aktsiaindeksite prognoosimisel [11]. Erinevate meetoditega leitud tulemuste võrdlemiseks kasutati MAPE veamõõdikut. Üldiselt oli võrdlemise tulemusena parima täpsusega SGBM, kuid teatud andmestike peal näitas parimaid tulemusi hoopis SVM või MLP. Kõige nõrgema täpsusega oli RF, mis oli kasutusel baasmudelina.

Krüptovaluuta prognoosimise tööd on olnud üsna ulatuslikud, eriti seoses Bitcoin'i hinnaprognosidega. Üks populaarseim tehnika on prognoosida Bitcoin'i hinda selle enda ajajoone põhjal ehk ühemõõtmeliselt [12]. Artiklis võrreldakse ARIMA mudelit paljude erinevate masin- ja sügavõppe algoritmidega, nagu tugivektorregressioon (*Support Vector Machine*, SVR), pärilevivõrk (*Feedforward Neural Network*, FFNN), konvolutsiooniline närvivõrk (*Convolutional Neural Network*, CNN) ja LSTM. Tulemused kalduvad selgelt nii LSTM-i kui ARIMA poole, kusjuures viimane annab veelgi parema täpsuse. CNN ei töötanud üldse nii hästi, nagu oli eeldatud.

Samuti on varasemaid töid Bitcoin'i ennustamise kohta, milles võrreldakse ARIMA-t, LSTM-i ja väravatega korduvat üksust (*Gated Recurrent Unit*, GRU) [13]. Kuigi GRU on töösse valitud oma lihtsama arhitektuuri tõttu võrreldes LSTM-iga, osutub kõige edukamaks just ARIMA meetod nii ennustuste täpsuse kui ka kiireima tasakaalustamise osas. GRU jõudlus on küll parem kui LSTM, kuid kaugelt mitte ARIMA lähedal.

Lisaks on töid, kus Bitcoin'i hinna ennustamiseks on omavahel võrreldud kahte populaarsemat ARIMA ja LSTM lähenemisviisi [14] [15]. Esimene uuring näitab, et mõlemad meetodid võivad olla tõhusad. See sõltub aga sellest, kui palju andmeid analüüsiks võetakse – LSTM on efektiivsem, kui varasemaid andmeid kasutatakse vähem – ja milline on ajahorisont – ARIMA suudab anda täpseid tulemusi lühiajaliselt, kuid

pikaajalised prognoosid tekitavad suuremaid vigu. Teine uuring korraldati veidi teistlaadi, kasutades kahte erinevat treenimis- ja testimisandmestikku. Lisaks kasutasid nad järgmise päeva staatilist prognoosimeetodit, kus igal sammul kas siis hinnati mudel ümber või mitte. Kuigi esimese treeningkomplektiga oli edukam ARIMA, suutis LSTM siiski teises treeningkomplektis ARIMA-t edestada. Jõuti järeldusele, et LSTM abiga on võimalik paremini modelleerida nii tegelike hindade lähedale jõudvaid prognoose kui ka hinnamuutusi, samas kui ARIMA meetod järgib lihtsalt trendi. Nende töös oli Bitcoin hinnal selge tõusutrend, mis oli ARIMA-le abiks ja seetõttu olid ennustused sarnased tegelikkusega, kuid kui hinnad oleksid langenud, siis ARIMA seda nii hästi ennustada ei oleks osanud.

Üks uurimistööst kasutab Bitcoin tavapäraseid ja kõrge sagedusega hindu, et ennustada hinda erinevatel ajasagedustel [16]. Nad kasutavad Bitcoin päevahinna prognoosimiseks nii statistilisi kui ka masinõppe lähenemisviise ja võrdlevad Bitcoin 5-minutilise hinna ennustamise meetodeid. Võrdlus näitab, et esimese eesmärgi puhul töötavad paremini statistilised lähenemisviisid, millest logistiline regressioon annab suurima täpsuse 66%. Kõrge sagedusega prognooside puhul on aga selgelt tõhusamad masinõppemeetodid, millest LSTM saavutab suurima täpsuse 67,2%.

2.1.2 Hübriidmudelid

Kuna mudeli valik võib olla keeruline ülesanne paljude mõjuteguritega ja mitmetes erinevates stsenaariumides pole ühte selget võitjat olnud, on kasutatud mitmete meetodite kombineerimist hübriidmudeliteks. Nende idee seisneb selles, et ARIMA ja masinõppe mudelid võivad iseseisvalt anda head tulemused vaid teatud olukordades, kuid mitte üldiselt. Kuna reaalmaailma andmed on keerulised ja nende omadused ei ole kergesti mõistetavad, on välja töötatud hübriidmetoodika, mis ühendab nii lineaarseid kui ka mittelineaarseid modelleerimise aspekte praktikas [17]. Mainitud uurimistöös töötas autor analüüsi jaoks välja hübriidse lähenemisviisi, mis ühendab ARIMA ja tehiskärvivõrgu (*Artificial Neural Network*, ANN). Analüüsi esimeses etapis rakendati aegre linearse osa analüüsimiseks ARIMA mudelit ja teises etapis rakendati ANN eelnevalt väljatöötatud ARIMA mudeli jääkide modelleerimiseks. Seetõttu kasutas hübriidne lähenemine suurepäraselt ära erinevate modelleerimismeetodite eelised, kuna ARIMA on teadaolevalt väga edukas lineaarsete andmetega ja ANN-l on oma plussid mittelineaarsete andmete analüüsimisel. Tulemused näitasid selgelt, et kombineeritud

model võib anda täpsemad prognoosid võrreldes eraldiseisvate meetoditega. Autor järeldas ka teistest kirjandusest pärit kombineerimismeetoditest, et kuigi paljud tööd sisaldavad hübriidmudeleid, mis kombineerivad sarnaseid meetodeid (st ainult lineaarsete mudelite kombinatsioon või kõigest närvivõrkude kombinatsioon), siis parimad tulemused annavad siiski need lähenemisviisid, mis ühendavad üksteisest kõige enam erinevaid meetodeid. Lisaks, kui rakendada andmetele esmalt ARIMA meetodit, on mudeli ülesobitamise, mis on närvivõrkude tavapärase puudus, tõenäosus palju väiksem.

Üks eespool mainitud töödest katsetas samuti hübriidmudelitega [12]. Kuna iseseisvalt andis ARIMA esmase analüüsi käigus parimad tulemused, ühendati see hübriidmudeliteks kombineerides seda erinevate masin- ja sügavõppemeetoditega. Kombineerides ARIMA mudelit FFNN, CNN, LSTM või SVR-iga olid kõik prognoosid sarnaselt täpsed. Ja kuigi iseseisvalt oli CNN täpsuse osas halvim, andis hübriidmudel, mis ühendas CNNi ARIMAGA, kokkuvõttes kõige väiksema vea.

2.1.3 Mitme sagedusega finantsandmete aegrea prognoosimise meetodid

Kuna käesoleva lõputöö eesmärk on välja töötada mudelid, milles oleks arvesse võetud aegridade erinevaid sagedusi, tasub vaadata varasemaid samal teemal tehtud uurimistöid. Ent kirjanduse hulk sellest vaatenurgast on palju hõredam kui pelgalt regulaarsete finantsandmete aegridade prognoosimise puhul.

Siiski on mõningaid tõendeid selle kohta, et selline lähenemine, milles käsitletakse finantsaegridade erinevat ajasagedust, võib anda kõige täpsema prognoosi, näiteks järgmises artiklis [18]. Autorid analüüsisid paljusid New Yorki börsi ettevõtteid, ennustades nende aktsiate järgmise päeva sulgemishinda. Nende eesmärk oli arvestada nii pika- kui ka lühiajalisi mõjusid, võttes arvesse päevaseid ja aastaseid aegridu. Selle tulemusena töötasid nad välja LSTM hierarhilise ansambelmeetodi, kus järgmise päeva sulgemishinna prognoosimiseks kombineeriti kaks ajasageduse sisendit. Analüüs hõlmas mitmemõõtmelisi andmeid, sealhulgas hindu ja päevaseid indikaatoreid ning aastaseid finantssuhtarve valitud ettevõtete kohta. Mudel töötati välja järjestikku – esimese sammuna töötati aastasuhtarvude alusel välja esimese õppija ja teise sammuna kasutati ennustust teise õppija sisendiks. Järgnev analüüs andis tulemused väljatöötatud ansambelmudeli täpsuse kohta, võrreldes seda eraldiseisva LSTM-iga ja tehisnärvivõrguga (ANN) nii päevaste kui ka aastaandmete puhul. Tulemused näitasid LSTM ansambelmeetodi puhul suurimat täpsust ja tehisnärvivõrgu puhul halvimat.

Teine teadlaste rühm ei leidnud varasemaid töid finantsaegridade prognoosimise teemal, mis käsitleksid kauplemise erinevaid sagedusmustreid [19]. Nende eesmärk oli pakkuda välja lühiajalise prognoosimise mudel, mis keskenduks kõrge ajasagedusega kauplemismustritele ja oleks suunatud lühiajalisele tootlusele, lisaks aga ka pikaajalise prognoosimise mudel, mis sõltuks madala sagedusega kauplemismustritest. Diskreetsest Fourieri teisendusest (*Discrete Fourier Transform*, DFT) inspireeritud töös pakutakse välja uus lähenemisviis – olekufrekvents mälu (*State Frequency Memory*, SFM). Meetod jagab mälu rakkude peidetud olekud mitmeks sageduskomponendiks, millest igaüks modelleerib aktsiahinna volatiilsuse aluseks olevat erinevat sagedust. Järgmise sammuna ennustatakse tulevasi aktsiahindu nende eelnevalt avastatud komponentide kombinatsiooni mittelineaarse kaardistamise abil Fourieri pöördteisenduses (*Inverse Fourier Transform*, IFT). Meetodit testitakse väga erinevate sektorite ettevõtete aktsiahindadel ja võrreldakse seejärel teiste tippasemel meetoditega. Kui kõige traditsioonilisem autoregressiivne (AR) mudel jääb kõige halvema täpsusega teistest maha, siis ka LSTM, mis väidetavalt arvestab nii lühi- kui ka pikaajalisi sõltuvusi, on SFMiga võrreldes suhteliselt ebatäpne. Väljapakutud lähenemisviis on kõige täpsem võimalus finantsaegridade erinevate sagedusmustrite avastamiseks, ületades täpsuse osas oluliselt teisi.

Finantsaegridadele omase suure volatiilsuse ja mittelineaarsuse tõttu on nende andmete prognoosimise raskust täheldanud paljud, seega on probleemi käsitlemise alternatiivid üsna mitmekesised. Järgmises artiklis [20] pakuti välja alternatiivne raamistik, mis ühendab ajalise koondamise ja prognoosimise. Nende idee oli eristada aegrea erinevad elemendid, koondades algandmed madalamalt agregeerimistasandilt kõrgemale. Samal ajal kui madalam agregeerimistasand kirjeldab andmete muutlikumat liikumist ja aitab tabada hooajalisi elemente, siis kõrgem agregeerimistasand kajastab andmete taset ja trendi. Prognoosimise etapis kohandasid autorid eksponentsiaalse silumise (ETS) meetodit igale ajaliselt agregeeritud andmestikule. ETS-i eeliseks on võimalus ekstraheerida kõik aegridade komponendid – tase, trend ja hooajalisus – otse. Kolmandas etapis ühendasid autorid prognoositud aegridade komponendid eraldi taseme-, trendi- ja sesoonsuse hinnangute jaoks, mida seejärel kasuti aegrea lõpliku prognoosi koostamiseks. Pakutud meetodit nimetatakse mitme agregatsiooni prognoosimisalgoritmiks (*Multi Aggregation Prediction Algorithm*, MAPA). Selle tehnika jõudlust võrreldakse ETSiga, mida rakendati otse M3-võistluse [21] ja St. Louisi

Föderaalreservi Panga [22] algandmestikule. Tulemusena oli selge, et kavandatud MAPA ületab ETS meetodi täpsust, eriti pikaajaliste prognooside puhul. Lühiajaliste prognooside korral, kus madala sagedusega komponentide mõju on väike või puudub üldse, ei ole olulisi erinevusi ja ETS toimib samuti hästi.

Varasema kirjanduse uurimine andis põhjaliku ülevaate mitmesugustest aegriidide prognoosimise meetoditest, mille hulgast valiti sobivamad välja käesoleva töö jaoks. Statistiliste meetodite hulgast valiti eksponentsiaalne silumine ja autoregressiivne integreeritud libiseva keskmise meetod, masinõppemeetoditest osutusid valituks otsustusmets ja tugivektorregressioon ning sügavõppemeetoditele pöörati suurimat tähelepanu, valides analüüsi mitmekihilise närvivõrgu, rekurrentse närvivõrgu, pika lühiajalise mälu ja väravatega rekurrentse üksuse.

Tänu varasematele töödele mitme ajasageduse ühendamise teemal otsustati rakendada samal eesmärgil hierarhilist ansambelmeetodit. Kuna käesoleva töö fookuses on hästituntud ja populaarsust kogunud meetodite võrdlus, siis teised mitme ajasagedusega andmete rakendamise töödes käsitletud tehnikad jäid antud töös kasutamata. Samas on nii MAPA kui ka SFM huvitavad ja omapärased tehnikad, mida on võimalik rakendada teema täiendaval uurimisel tulevikus.

3 Metoodika ja andmed

Järgmine lõik annab ülevaate järgnevas finantsandmete aegridade prognoosimise analüüsis kasutatud andmetest ja valitud analüüsimeetoditest.

3.1 Traditsioonilised statistilised aegria meetodid

3.1.1 Autoregressiivne integreeritud libiseva keskmise mudel ARIMA

Autoregressiivne integreeritud libiseva keskmise (ARIMA) mudel on võimas ja laialdaselt kasutatav aegridade prognoosimise tehnika. See ühendab kolm komponenti: autoregressiivsus (AR), integreeritud (I) ja libisev keskmine (MA). See meetod on mõeldud ajaliste seeriade modelleerimiseks ja ennustamiseks, võimaldades käsitada andmete trende, hooajalisi mõjusid ja autokorrelatsiooni.

ARIMA mudel on määratletud kolme peamise parameetriga: p , d ja q . Autoregressiivne parameeter (p) viitab sellele, kuidas praegune vaatlus sõltub eelnevatest vaatlustest. See määrab, kui palju varasemad ajaperioodid mõjutavad praegust. Integreeritud osa ehk diferentseeriv parameeter (d) andmete tasandamisele või diferentseerimisele, et muuta aegrida statsionaarseks. See võimaldab andmetest eemaldada trendi ja hooajalisi mõjusid. Libiseva keskmise parameeter (q) näitab libiseva keskmise akna suurust, jäädvustades seose vaatluse ja varasemate vaatluste jääkvea vahel. See võimaldab modelleerida juhuslikku kõikumist andmetes ja autokorrelatsiooni [23].

ARIMA mudelit väljendatakse tavaliselt kujul $ARIMA(p, d, q)$. ARIMA on eriti kasulik statsionaarsete aegridade andmete jaoks ning selle paindlikkus ja lihtsus muudavad selle aegridade analüüsi oluliseks tööriistaks. Mudeli tõhusus seisneb selle võimes pakkuda täpseid prognoose paljude ühemõõtmeliste aegridade muustrite kohta, muutes selle rakendatavaks erinevates valdkondades, alates majandusest kuni kliimaalüüsini, aga ka mujal. ARIMA mudeli valem on järgmine [24]:

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_p Y_{t-p} + \dots + \theta_1 \epsilon_{t-1} + \theta_q \epsilon_{t-q} + \epsilon_t, \text{ kus:} \quad (1)$$

- Y_t on vaadeldav väärtus ajahetkel t ,
- μ on aegrea keskmine,
- ϕ_1, ϕ_p on autoregressiivsed koefitsiendid,
- ϵ_t on valge müra veatermin,
- θ_1, θ_q on libiseva keskmise koefitsiendid,
- Y_{dt-1}, Y_{dt-p} on viitajaga vaatlused.

Eesmärk on hinnata $\phi_1, \phi_p, \theta_1, \theta_q$ väärtusi, mis minimeerivad erinevuse vaadeldavate väärtuste ja mudeli tehtud ennustuste vahel.

3.1.2 Eksponentsiaalne silumine

Eksponentsiaalne silumine on laialdaselt kasutatav aegridade prognoosimismeetod, mis on tuntud oma lihtsuse ja tõhususe poolest aluseks olevate trendide ja mustrite kirjeldamisel. See meetod määrab varasematele vaatlustele eksponentsiaalselt kahanevad kaalud, andes suurema tähtsuse hiljutistele andmetele, vähendades samas vanemate andmepunktide mõju. Protsess hõlmab korduvat prognoosi värskendamist, kombineerides praeguse vaatluse varem prognoositud väärtusega. Eksponentsiaalse silumise põhiparameeter on silumiskordaja, mida sageli tähistatakse tähega α , mis määrab kõige värskemale vaatlusele antud kaalu. α suuremad väärtused eelistavad hiljutisi andmeid tugevamalt, muutes mudeli lühiajalistele kõikumistele paremini reageerivaks. Ja teistpidi, α väiksemad väärtused rõhutavad rohkem ajaloolisi andmeid, mille tulemuseks on sujuvam prognoos. Eksponentsiaalne silumine on eriti tõhus aegridade andmete puhul, mis näitavad trendi ja hooajalisust. Selle kohanemisvõime, rakendamise lihtsus ja võime käsitleda erinevaid aegridade mustreid muudavad selle väärtuslikuks tööriistaks erinevate valdkondade rakenduste prognoosimisel. [25] Eksponentsiaalse silumise põhimudel on tähistatud järgmiselt:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t, \text{ kus:} \quad (2)$$

- \hat{Y}_{t+1} on ennustus järgmiseks perioodiks,
- Y_t on vaadeldav väärtus ajaperioodil t ,
- \hat{Y}_t on ennustatav väärtus perioodil t ,
- α on silumiskordaja.

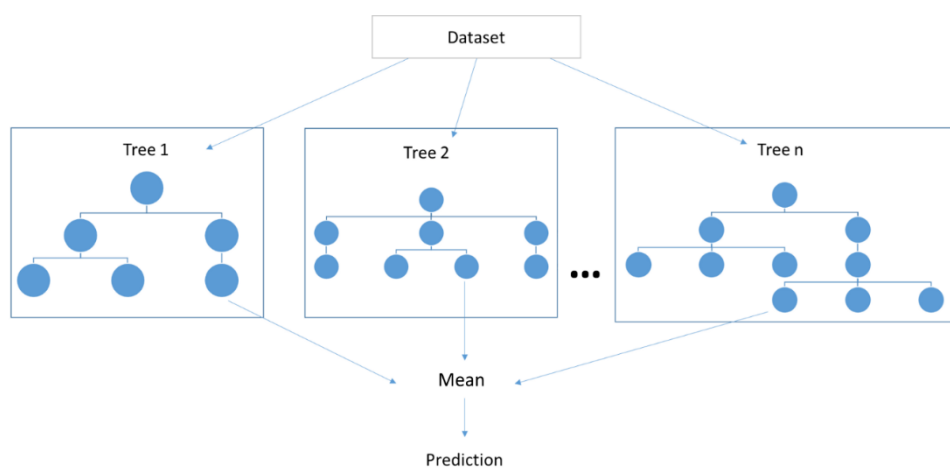
3.2 Masinõppe meetodid aegridade prognoosimiseks

Järgnevalt antakse teoreetiline ülevaade kahest masinõppe meetodist, mida on varasemalt edukalt kasutatud just aegridade prognoosimisel.

3.2.1 Otsustusmets

Otsustusmets (*random forest*, RF) on ansambelõppemeetod, mis koosneb mitmest otsustuspuust (*decision tree*, DT), mida treenitakse erinevate alamandmete osahulkadel. Iga otsustuspuu annab ennustuse ja lõplik ennustus tehakse, võttes arvesse kõigi otsustuspuude ennustusi. See meetod on tugev ja mitmekülgne, sobides hästi erinevate andmetüüpide ja probleemide lahendamiseks [26].

Aegridade prognoosimisel võib RF olla tõhus, eriti kui andmestik sisaldab mitmekesisaid tunnuseid ja keerulisi ajamustreid. RF võimaldab mudelil omandada andmetest keerukamaid seoseid ja vältida ülesobitamist, kuna see koosneb mitmest otsustuspuust [26]. Joonis 1 annab lihtsa ülevaate sellest, kuidas otsustusmets koosneb paljudest otsustuspuudest ning kuidas jõuab algoritm lõpliku ennustuseni.

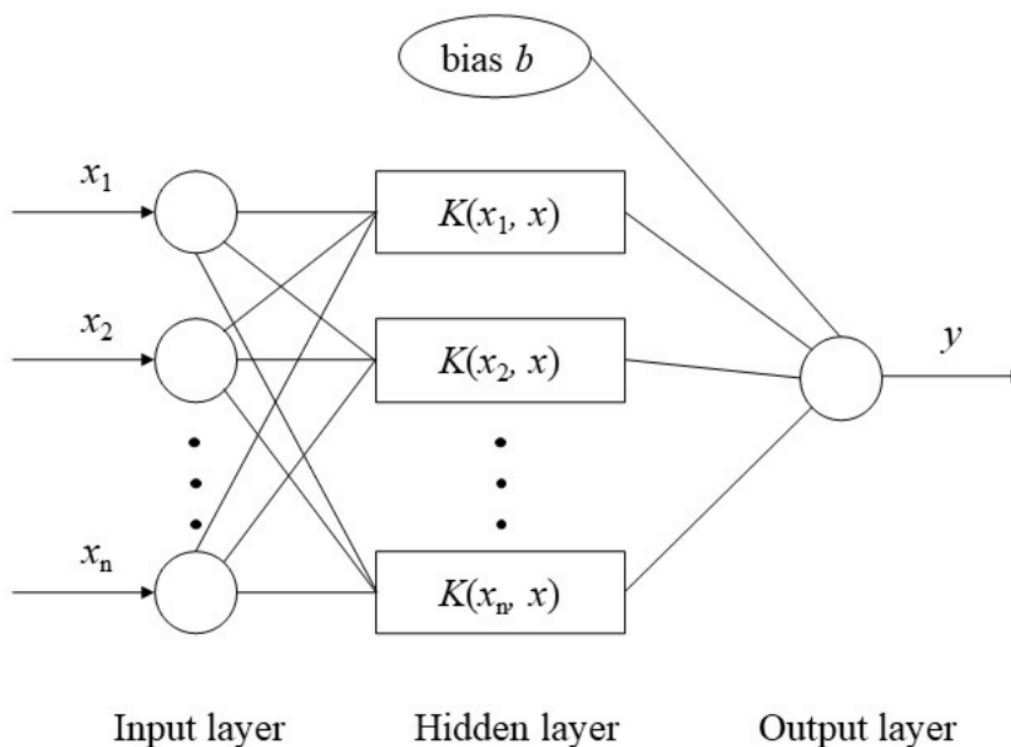


Joonis 1. Otsustusmetsa arhitektuur [27].

3.2.2 Tugivektorregressioon

Tugivektorklassifitseerija (*Support Vector Machine*, SVM) on masinõppe meetod, mis on mõeldud nii klassifitseerimiseks kui ka regressiooniks. Kuna käesolevas töös püstitatud eesmärk on peamiselt regressiooni ülesanne, siis edaspidi kajastatakse tugivektor-

regressiooni (*Support Vector Regression*, SVR) just sellest vaatenurgast. SVR-i eesmärk on leida otsustuspind andmepunktide vahel selliselt, et see maksimeeriks ennustatavate väärtuste vahelise marginaali. Kuigi SVR on eriti tõhus keerukate, mitmemõõtmeliste andmestike jaoks, on näidanud see efektiivsust ka aegridade prognoosimisel, eriti kui andmestik sisaldab keerukaid ajalisi sõltuvusi. Sealjuures võib SVR olla efektiivne nii lineaarsete kui ka mitte-lineaarsete seoste mudeldamisel [28]. Joonis 2 annab ülevaate tugivektorregressiooni tööpõhimõttest.



Joonis 2. Tugivektroregressiooni arhitektuur [29].

Tugivektorregressiooni peidetud kihis olev K tähistab mittelineaarse regressiooni Kernel funktsiooni, millest üks populaarsemaid on käesolevas töös kasutatud radiaalne baasfunktsioon (*radial basis function*).

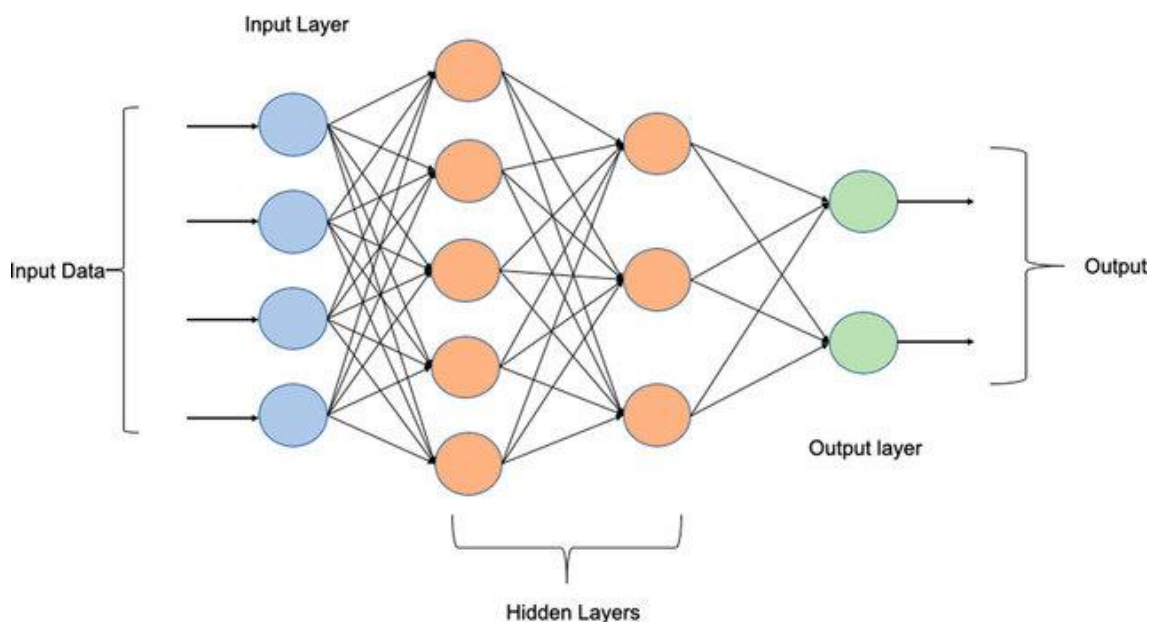
3.3 Sügavõppemeetodid aegridade prognoosimiseks

Viimasena tehakse ülevaade erinevatest sügavõppe algoritmidest, millega on varasemalt saavutatud väga häid tulemusi aegridade prognoosimise ülesannetes. Varasema kirjanduse põhjal on selgeks tehtud, et need meetodid töötavad hästi just väga volatiilsete finantsandmete ennustamisel.

3.3.1 Mitmekihiline närvivõrk

Mitmekihiline närvivõrk (*Multilayer Perceptron*, MLP) on sügavõppe meetod, mis koosneb mitmest peidetud kihist ja on üks levinumaid tehisnärvivõrkude arhitektuure. Sellel on mitu sisendit ja väljundit, mida ühendavad peidetud kihid. Igas peidetud kihis on mitu neuronit, mis on täielikult ühendatud eelmise ja järgmise kihiga [30].

Mitmekihilist närvivõrku kasutatakse laialdaselt mitmesugustes rakendustes, sealhulgas ka just aegridade prognoosimisel. Tänu oma võimele õppida andmete keerukaid ja mitmekesiseid mustreid on MLP meetod osutunud efektiivseks vahendiks keerukate andmekogumite analüüsimisel ja modelleerimisel. MLP algoritm on võimeline käsitlema nii lühiajalisi kui pikaajalisi, regulaarseid kui ebaregulaarseid ning ühe- ja mitmemõõtmelisi aegridu. Samuti on nad näidanud täpset prognoosivõimet nii mittelineaarsete kui ka mittestatsionaarsete andmete analüüsimisel ning ajaliste sõltuvuste ja mustrite dünaamika õppimisel [30]. Joonisel 3 on näidatud mitmekihilise närvivõrgu arhitektuur.

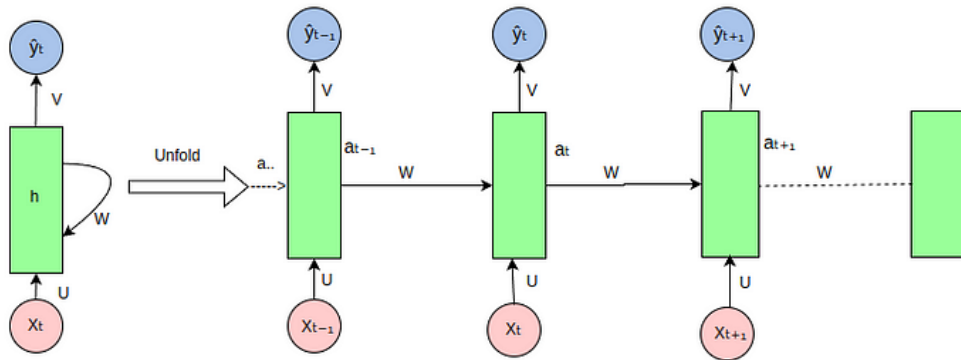


Joonis 3. Mitmekihilise närvivõrgu arhitektuur [31]

3.3.2 Rekurrentne võrk

Korduvad närvivõrgud (*Recurrent Neural Network*, RNN) esindavad tehisnärvivõrkude spetsiaalset klassi, mis on loodud ajalistest sõltuvustest tingitud järjestikuste andmete modelleerimiseks, pakkudes ainulaadset võimet jäädvustada ajalisi sõltuvusi [32]. Erinevalt traditsioonilistest pärilevivõrkudest (*feedforward networks*, FFNN) on

RNNidel korduvad ühendused, mis võimaldavad teabel aja jooksul võrgus püsida. See võimaldab neil töödelda erineva pikkusega järjestikuseid sisendandmeid, muutes need eriti tõhusaks selliste ülesannete puhul nagu loomuliku keele töötlemine, aegridade ennustamine ja kõnetuvastus. RNNide põhiomadus on peidetud olek, mis säilitab eelmiste sisendite mälu ja mõjutab järgnevate sisendite töötlemist. Meetodi arhitektuur on kujutatud Joonisel 4.



Joonis 4. Rekurrentse võrgu (RNN) arhitektuur [33].

Vastavalt RNN arhitektuurile saab järgnevalt panna kirja peidetud kihi valemi [33]:

$$a_t = f(U \times X_t + W \times a_{t-1} + b), \text{ kus:} \quad (3)$$

- a_t tähistab võrgu peidetud kihi loodud väljundit ajahetkel t ,
- f on aktiveerimisfunktsioon,
- U on kaalumaatriks, mis reguleerib ühendusi sisendist peidetud kihini,
- X_t on võrgu sisend ajahetkel t ,
- W on kaalumaatriks, mis reguleerib peidetud kihi ühendusi iseendaga (korduvad ühendused),
- a_{t-1} on peidetud kihi väljund ajahetkel $t-1$,
- b on peidetud kihi nihkevektor.

Peidetud kihi väljund on omakorda komponendiks võrgu väljundi valemis [33]:

$$\hat{y}_t = f(V \times a_t + c), \text{ milles:} \quad (4)$$

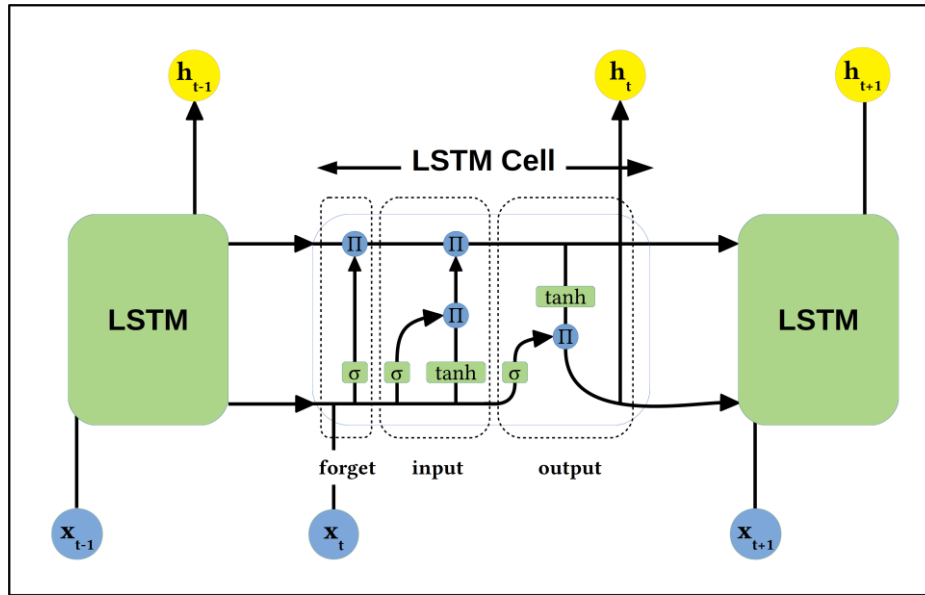
- \hat{y}_t on võrgu väljund ajahetkel t ,
- V tähistab peidetud kihi ja väljundkihi vahelise ühendusega seotud kaalu,
- c on väljundkihi nihkevektor.

Nagu eelnevast näha, koosneb RNN korduvatest moodulitest, mida nimetatakse ajurakkudeks, mis võimaldavad infot liikuda läbi ahela, võimaldades sellega mudelil käsitada ajas muutuvaid sisendeid. Kuigi RNN-id on tõhusad ajalistes seeriates ja järjestikustes andmetes, võivad nad aga kogeda probleeme, näiteks gradientide kadumist või plahvatamist, eriti pikemate sõltuvuste korral [34]. Selle probleemi lahendamiseks võeti kasutusele pika lühiajalise mälu (LSTM) võrgu ja väravatega korduva üksuse (GRU) arhitektuurid, mis hõlmavad spetsiaalseid mälulahtreid ja väravamehhanisme, et paremini püüda ja hallata pikaajalisi sõltuvusi.

3.3.3 Pikk lühiajaline mälu

Pika lühiajalise mälu (*Long short-term memory*, LSTM) närvivõrk on sügava õppimise algoritm, mis omakorda kuulub korduva närvivõrgu (RNN) perekonda. See on spetsiaalselt loodud lahendama järjestikuste andmete pikaajaliste sõltuvuste tabamise ja õppimise väljakutseid [34].

LSTM närvivõrkude oluline komponent on nn mäluplokk, mis võimaldab sellel efektiivselt töötada pikaajaliste sõltuvuste korral [34]. Mäluplokil on kolm peamist elementi: sisendvärav (*input gate*), unustamisvärav (*forget gate*) ja väljundvärav (*output gate*), millest kõiki juhib sigmoidne aktiveerimisfunktsioon. Sisendvärav reguleerib uue teabe liikumist raku olekusse, unustamisvärav juhib teabe säilitamist või eemaldamist raku olekust ja väljundvärav juhib teabe liikumist raku olekust väljundisse. Nende elementidega arhitektuur võimaldab LSTM mudelitel teavet valikuliselt meelde jätta või unustada, leevendades tõhusalt traditsiooniliste RNN algoritmidega seotud kaduvaid ja plahvatavaid gradiendiprobleeme. LSTM võrgud on sel põhjusel väga võimsad sellistes rakendustes nagu loomuliku keele töötlemine, kõnetuvastus ja aegridade prognoosimine. Meetodi arhitektuur on kujutatud Joonisel 5.



Joonis 5. Pika lühiajalise mälu (LSTM) võrgu arhitektuur [35].

Pika lühiajalise mälu olekute struktuuri ajasammul t saab väljendada ka järgmiselt [35]:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (5)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (6)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (7)$$

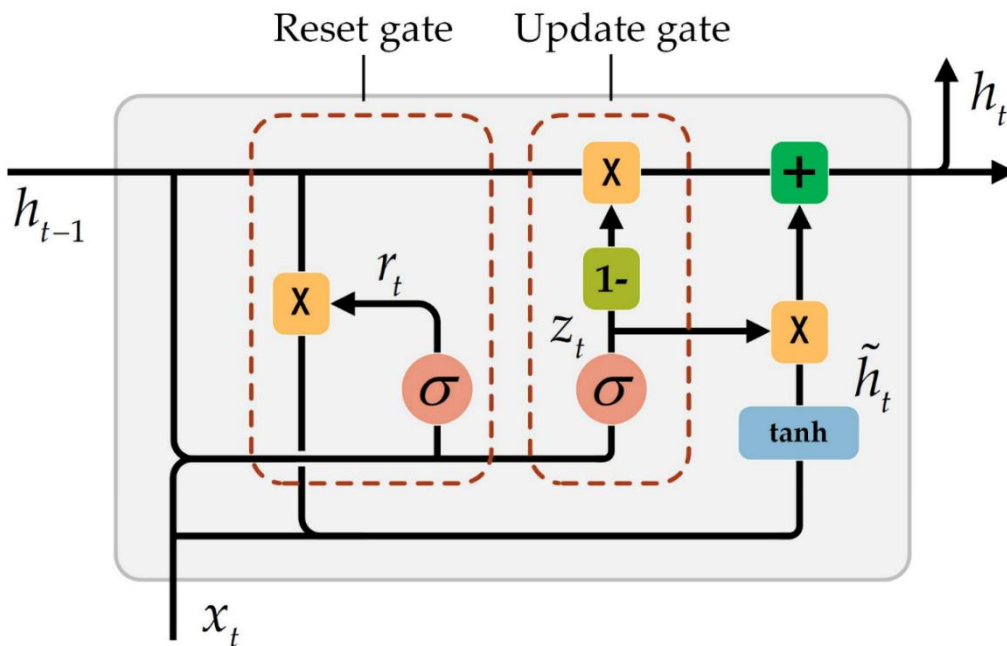
$$c_t = f_t \circ c_{t-1} + i_t \circ \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (8)$$

$$h_t = o_t \circ \tanh(c_t), \text{ kus:} \quad (9)$$

- x_t on sisendvektor,
- f_t on unustusvärava aktiveerimise vektor ajahetkel t ,
- i_t on sisendvärava aktiveerimisvektor ajahetkel t ,
- o_t on väljundvärava aktiveerimise vektor ajahetkel t ,
- c_t on mälu rakkude aktiveerimise vektor ajahetkel t ,
- h_t on peidetud oleku aktiveerimise vektor ajahetkel t ,
- W_f , W_i ja W_o on vastavate väravate sisendtuumad,
- U_f , U_i ja U_o on vastavate väravate korduvad tuumad,
- b_f , b_i , b_o ja b_c on nihked,
- σ on logistiline sigmoidfunktsioon,
- \circ on maatrikskorrutis,
- \tanh on hüperboolse puutuja aktiveerimise funktsioon.

3.3.4 Väravatega rekurrentne üksus

Väravatega rekurrentne üksus (*Gated Recurrent Unit*, GRU) on RNN arhitektuuri tüüp, mis on loodud traditsioonilise RNN mõningate piirangute ületamiseks, nagu selle suutmatust tabada pikaajalisi sõltuvusi. GRU sisaldab väravamehhanisme, mis sarnanevad pika lühiajalise mälu (LSTM) võrguga, kuid neil on sujuvam arhitektuur [36]. GRU-del on kaks väravat: lähtestamisvärav ja värskendusvärav. Lähtestamisvärav kontrollib, mil määral tuleks minevikuteave unustada, värskendusvärav aga reguleerib mämlahtrisse salvestatava uue teabe hulka. See sujuvam struktuur võimaldab GRU-del tõhusamalt jäädvustada sõltuvusi pikematel ajajadadel kui tavapärased RNN-id suudavad, olles samas arvutuslikult efektiivsemad kui LSTM-id. GRU-d on näidanud tõhusust mitmesugustes järjestikuste andmetega ülesannetes, sealhulgas loomuliku keele töötlemisel ja aegridade analüüsil. GRUde lihtsus ja jõudlus on muutnud need populaarseks valikuks rakendustes, kus ajaliste sõltuvuste püüdmine on ülioluline. GRU struktuur on näidatud järgmisel Joonisel 6.



Joonis 6. Väravatega rekurrentse üksuse (GRU) struktuur [37].

Meetodi struktuuri saab väljendada ka järgmiste valemitega [37]:

$$z_t = \sigma(w_{zx}x_t + u_{zh}h_{t-1}) \quad (10)$$

$$r_t = \sigma(w_{rx}x_t + u_{rh}h_{t-1}) \quad (11)$$

$$\tilde{h}_t = \tan(w_{hx}x_t + r_t \odot u_{hh}h_{t-1}) \quad (12)$$

$$h_t = (1 - z_t)\odot\tilde{h} + \odot z_t \odot h_{t-1}, \text{ kus} \quad (13)$$

- z_t on värskendusvärav,
- r_t on lähtestusvärav,
- \tilde{h}_t on hetkel peidetud sõlme peidetud olek,
- h_t on praegune peidetud olek,
- x_t on praeguse närvivõrgu sisend
- h_{t-1} on peidetud olek eelmisel hetkel
- σ on sigmoidne aktiveerimisfunktsioon,
- \odot on Hadamardi maatriksi korrutis,
- w ja u on kaalumaatriksid, mida peab õppima.

3.4 Finantsandmed

Käesolevas magistritöös on võetud finantsandmete aegrea analüüsiks kasutusele valitud aktsiaindeksid ning krüptovaluutad, mis on laialdaselt tuntud. Nendeks on kolm erinevat aktsiaindeksit S&P 500, Dow Jones Industrial Index, Nasdaq 100 ning kaks enamtuntud krüptovaluutat Bitcoin ja Ethereum, eesmärgiga võtta arvesse erineva skaala ja volatiilsusega andmestikke. Tunni täpsusega andmed pärinevad tasulisest allikast First Rate Data [38] ning päevased andmed pärinevad Yahoo Finance'i andmebaasist [39].

Andmete osas oli võimalik valida vastava sageduse, kas siis päeva või tunni, avamishind, kõrgeim hind, madalaim hind ja sulgemishind. Käesoleva magistritöö eesmärkide saavutamiseks valiti analüüsi aluseks just viimane, seega on analüüsitavateks aegridadeks just vastavate aktsiaindeksite ja krüptovaluutade päeva- ja tunniintervalliga sulgemishinnad. Kuna indeksid on turul kaubelnud erineva ajaperioodi jooksul, erinevad seetõttu ka kättesaadavate andmete ajavahemikud. Täpsemalt on paigutatud päeva resolutsiooniga andmestiku ülevaade Tabelisse 1. Selgelt on aktsiaindeksite aeg olnud börsil oluliselt pikem, mistõttu kasutatav aegrida algab 2008. aasta algusest. Krüptovaluutade puhul on tegu värskema varaklassiga, mistõttu ka periood on oluliselt lühem – Bitcoin'i puhul on andmed kättesaadavad alates 2013. aastast, Ethereum'i puhul alates 2016. aastast.

Tabel 1. Finantsandmete päevase aegrea kirjeldav statistika.

	S&P 500	Dow Jones	Nasdaq 100	Bitcoin	Ethereum
Alguse kuupäev	02.01.2008	02.01.2008	02.01.2008	01.04.2013	11.03.2016
Lõpu kuupäev	26.01.2024	26.01.2024	26.01.2024	27.01.2024	27.01.2024
Aritmeetiline keskmine	2 364,23	20 331,97	6 175,63	12 873,93	1 008,93
Standardviga	17,67	134,31	70,39	254,03	20,56
Mediaan	2 088,89	17 864,39	4 487,50	6 616,62	410,31
Mood	1 447,16	9 509,28	1 949,15	126,78	10,55
Standardhälve	1 122,91	8 536,70	4 473,75	15 959,38	1 102,10
Valimi dispersioon	1 260 918	72 875 293	20 014 398	254 701 759	1 214 619
Järsakuskordaja	-0,83	-1,13	-0,52	0,81	0,63
Asümmeetriakordaja	0,59	0,41	0,88	1,35	1,19
Ulatus	4 217,63	31 562,38	16 480,48	67 490,37	4 805,14
Miinumum	676,53	6 547,05	1 036,51	61,35	6,76
Maksimum	4 894,16	38 109,43	17 516,99	67 551,72	4 811,90
Summa	9 551 495	82 141 143	24 949 550	50 813 413	2 897 661
Arv	4 040	4 040	4 040	3 947	2 872

Kõrgema sagedusega ehk tunni resolutsiooniga andmete ülevaade on paigutatud Tabelisse 2. Nii päeva- kui tunniintervalliga andmestike ülevaatest selgub, et lisaks erinevale ajaperioodile on valitud finantsvarad ka muude statistiliste näitajate poolest väga erinevad. Varad on läbi aja kaubelnud väga erinevates hinnaklassides ja erinevad oluliselt nii keskmise väärtuse kui ka ulatuse poolest. Lisaks on suured erinevused varade standardhälbes ja dispersioonis, eriti silmnähtav on see S&P 500 indeksi ning Bitcoin'i võrdlemisel.

Tabel 2. Finantsandmete tunnise aegrea kirjeldav statistika.

	S&P 500	Dow Jones	Nasdaq 100	Bitcoin	Ethereum
Alguse tund	02.01.2008 09:00	02.01.2008 09:00	02.01.2008 09:00	01.04.2013 00:00	11.03.2016 05:00
Lõpu tund	26.01.2024 16:00	26.01.2024 16:00	26.01.2024 16:00	26.01.2024 23:00	26.01.2024 23:00
Aritmeetiline keskmine	2 419,25	20 841,91	6 364,12	12 876,21	1 009,90

	S&P 500	Dow Jones	Nasdaq 100	Bitcoin	Ethereum
Standardviga	6,37	47,80	25,53	51,88	4,20
Mediaan	2 110,48	18 129,66	4 604,88	6 630,18	410,00
Mood	1 414,98	12 738,90	1 808,51	118,00	11,00
Standardhälve	1 118,19	8 457,70	4 484,15	15 951,26	1 101,43
Valimi dispersioon	1 250	71 532	20 107	254 442	1 213
	353,50	770,66	624,04	846,16	144,94
Järsakuskordaja	-0,88	-1,16	-0,62	0,81	0,62
Asümmeetriakordaja	0,53	0,32	0,82	1,35	1,19
Ulatus	4 238,22	31 711,82	16 635,37	68 584,50	4 843,02
Miinumum	667,79	6 482,30	1 024,94	55,70	6,02
Maksimum	4 906,01	38 194,12	17 660,31	68 640,20	4 849,04
Summa	74 609 745	652 435 005	196 282 175	1 217 445 734	69 373 330
Arv	30 840	31 304	30 842	94 550	68 693

Valitud finantsvarade suured erisused on aga omaette huvipakkuv aspekt järgnevas analüüsis. Sealt võib selguda, millised meetodid on parimad indeksitele, mis oma olemuselt on veidi vähem volatiilsed, ja mis tehnikad toimivad paremini just kõrge volatiilsusega krüptovaluutade puhul.

3.5 Kasutatud tarkvara

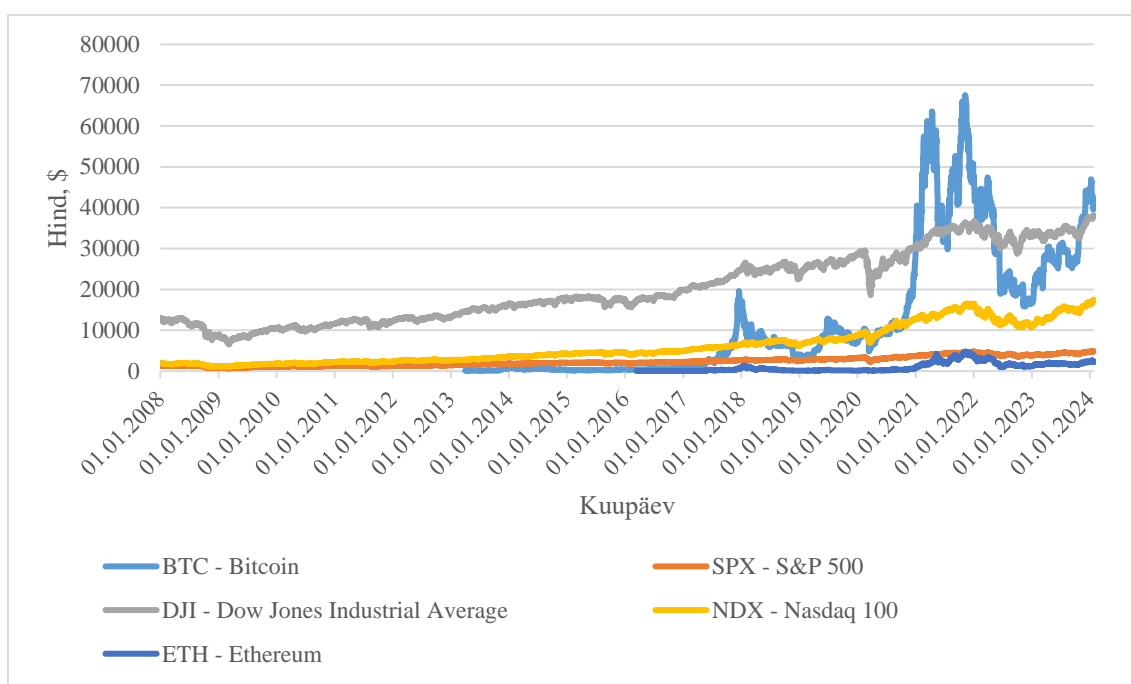
Käesolevas töös on esialgseks andmete ülevaateks kasutusel Microsoft Excel. Edasise analüüsi, mudelite treenimise ja võrdlemise jaoks oli peamiseks töövahendiks Visual Studio Code ja Python 3.11.7.

4 Analüüs

Analüüsi ülesehitamisel oli esialgse uurimise all S&P 500 aktsiaindeksi andmestik ja kõiki meetodeid katsetati esimesena just selle aegrea peal. Seega oli S&P 500 indeksi analüüs justkui illustreeriv näide erinevate analüüside teostamiseks ja mudelite ehitamiseks. Ülejäänud aktsiaindeksite ja krüptovaluutade aegridasid kasutati läbiviidud analüüsides testimiseks ja võrdlemiseks, et hinnata mudelite robustsust ja nende üldistusvõimet.

4.1 Andmete eeltöötlus analüüsiks

Analüüsis kasutatud aegridade puhul ilmneb selgelt nii trend kui ka sesoonsus ja ka graafilisel vaatlusel selgub, et tegu on mittestatsionaarsete andmetega (Joonis 7).



Joonis 7. Finantsandmed [38] [39].

Lisaks näitasid mittestatsionaarsust ka esialgsetel andmetel läbiviidud statistilised testid. Aegridade normaaljaotuse analüüsimiseks kasutati Jarque-Bera testi [40] ning stabiilsuse uurimiseks võeti kasutusele Dickey-Fulleri test [41]. Jarque-Bera testi tulemused viitasid

selgelt normaaljaotuse puudumisele ning Dickey-Fulleri test näitas aegrea mittestatsionaarsust, nagu näha Tabelis 3, kuhu on koondatud tulemused S&P 500 andmestiku kohta. Ülejäänud andmeridade testitulemused on lisatud tabelisse Lisas 2.

Tabel 3. S&P 500 indeksi aegrea normaaljaotuse ja statsionaarsuse testi tulemused.

	Test statistik	Kriitiline väärtus	p-väärtus
Jarque-Bera	350,01	5,99	9,90E-77
Dickey-Fuller	0,80	-2,86	0,99

Kahe testi tulemuste uurimisel on oluline silmas pidada, et esimese puhul on nullhüpoteesiks aegrea vastamine normaaljaotusele. Kuna p-väärtus on oluliselt madalam 5% usaldus tasemest, tuleb nullhüpotees tagasi lükata. Seetõttu ei vasta S&P 500 indeksi sulgemishindade aegrida normaaljaotusele. Vastupidiselt, teisel testil on p-väärtus oluliselt kõrgem usaldustasemest 5%. See näitab, et nullhüpoteesi ühikruudu olemasolu kohta ei saa tagasi lükata, mis omakorda tähendab, et andmestik on mittestatsionaarne.

Eelnevate testide tulemusel selgunud aktsiahindade ebanormaalne jaotus ja mittestatsionaarne olemus näitavad väga edukalt prognoosimise keerukust ning on ka oluliseks takistuseks, rakendamaks aegrea modelleerimisel statistilisi meetodeid. Sel põhjusel tuleb andmetele rakendada mitmesuguseid eeltötluse meetodeid, et neid oleks võimalik statistiliste meetoditega modelleerida.

Selleks, et stabiliseerida aegridade varieeruvust, võeti kasutusele andmestiku logaritmimeine. Lisaks rakendati andmetele diferentseerimist. Selle abil on võimalik stabiliseerida aegrea keskmine väärtus, eemaldades aegrea taseme muutused ja seega vähendades oluliselt nii trendi kui ka sesoonsust [42]. Töödeldud andmestike graafikud asuvad Lisas 3. Läbiviidud töötlus andmetele aitab muuta finantsandmete aegread statsionaarseks ja seega ka sobivaks, et rakendada neile traditsiooniliste meetoditega prognoosimist. Kõige olulisem on see just ARIMA mudeli rakendamisel. Siiski otsustati töödeldud andmestikku kasutada ka eksponentsiaalse silumise ning otsustusmetsa meetodi rakendamisel. Teisendatud andmetega treeniti mudelid ja tehti prognoosid, kuid tulemuste võrdlemiseks rakendati prognoosidele pöördteisendust, et tulemusi saaks võrrelda algandmetega.

Analüüsi osas on aga suurem roll just sügavõppe meetoditel, mille puhul võeti kasutusele erinev andmete teisendamise meetodid. Kuna töö analüüsi jaoks oli valitud Python ja

sellega seotud raamistikud, siis andmete standardiseerimiseks ja normaliseerimiseks oli kasutusel scikit-learn raamistik. Selles raamistikus on mitmeid erinevaid skaleerijaid, kuid oma lihtsuse ja mõistetavuse tõttu osutus valituks just MinMaxScaler [43]. See skaleerija normaliseerib andmed kindlasse vahemikku, mis käesolevas töös jäeti selliseks, nagu vaikimisi oli määratud ehk [0,1]. MinMaxScaler on nii nimetatud just seetõttu, et kasutab minimaalset ja maksimaalset väärtust. Seega ei vähenda MinMaxScaler erindite mõju, vaid normaliseerib need lineaarselt kindlaksmääratud vahemikku.

Standardiseeritud andmete peal treeniti nii tugivektorregressiooni kui ka kõigi sügavõppemeetodite mudelid. Prognoosid tehti seejärel normaliseeritud andmetega, kuid siis kasutati pöördteisendust, et võrrelda prognoositud väärtuseid algandmetega.

4.1.1 Treening- ja testandmete valik

Mudelite treenimise jaoks oli vaja andmestik jaotada treening- ja testandmeteks. Kuna tegu on aga aegriudadega, ei ole võimalik teha juhuslikku treening- ja testandmestiku jaotust. Uurimise all on ka mitme erineva finantsvara andmed, mille kättesaadavuse ajavahemik erineb oluliselt. Seetõttu otsustati käesolevas töös lähtuda kõige lihtsamast aegrea andmestiku jaotamisest [44], mille alusel 80% aegrea esimesest osast võeti kasutusele mudeli treenimisel ning 20% aegrea lõpust oli testandmestik. Testandmestiku järgi prognoositi aegrea tulevikuväärtusi ning seejärel võrreldi tegelike väärtustega, et leida mudelite veamõõdikud ja neid omavahel võrrelda. Vastavalt 80:20 jaotusele said andmestikud järgmise jaotuse treening- ja testandmete vahel, mis näidatud Tabelis 4.

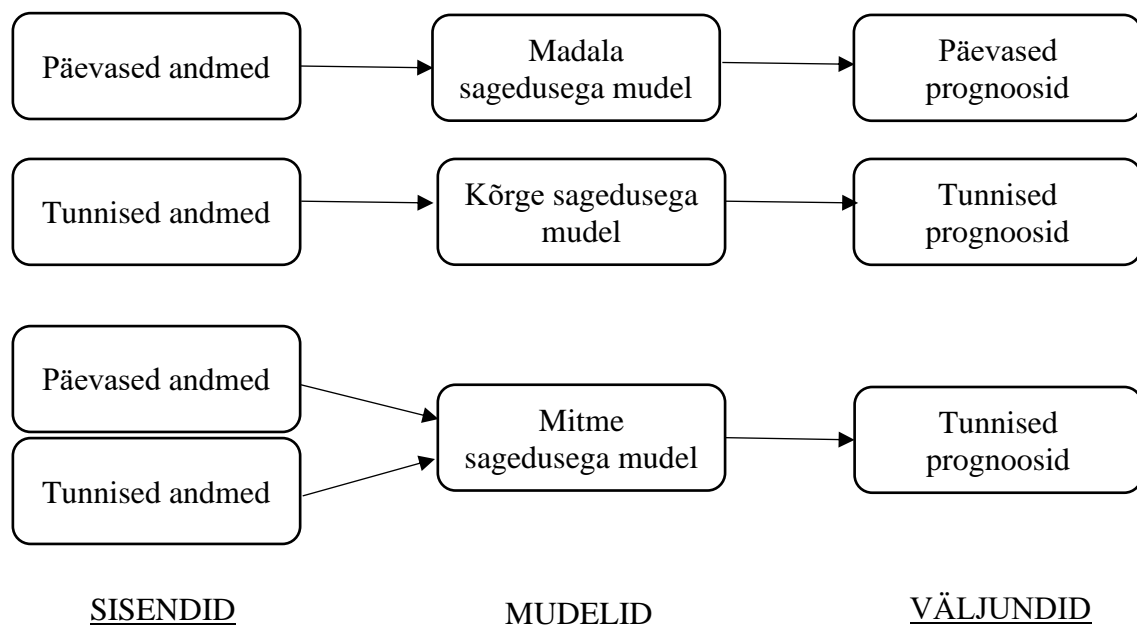
Tabel 4. Treening- ja testandmete ajavahemikud.

Finantsvara	Aja intervall	Treeningandmed	Testandmed
S&P 500	Päev	02.01.2008–05.11.2020	06.11.2020–26.01.2024
	Tund	02.01.2008 09:00–06.11.2020 10:00	06.11.2020 11:00–26.01.2024 16:00
Dow Jones	Päev	02.01.2008–05.11.2020	06.11.2020–26.01.2024
	Tund	02.01.2008 09:00–03.11.2020 10:00	03.11.2020 11:00–26.01.2024 16:00
Nasdaq 100	Päev	02.01.2008–05.11.2020	06.11.2020–26.01.2024
	Tund	02.01.2008 09:00–23.10.2020 12:00	23.10.2020 13:00–26.01.2024 16:00

Finantsvara	Aja intervall	Treeningandmed	Testandmed
Bitcoin ¹	Päev	01.04.2013–28.11.2021	29.11.2021–27.01.2024
	Tund	01.01.2019 00:00–21.01.2023 14:00	21.01.2023 15:00–26.01.2024 23:00
Ethereum ²	Päev	11.03.2016–01.07.2022	02.07.2022–27.01.2024
	Tund	01.01.2019 00:00–21.01.2023 17:00	21.01.2023 18:00–26.01.2024 23:00

4.2 Analüüsi arhitektuur

Käesoleva magistritöö peamine analüüs koosneb erinevate aegrea prognoosimise meetodite võrdlusest. Analüüsi ettevalmistamisel lähtus autor olemasolevatest tunni ja päeva resolutsiooniga andmetest. Analüüsi plaanitav arhitektuur on lihtsal kujul näidatud alljärgneval Joonisel 8.



Joonis 8. Analüüsi aluseks võetud võrreldavate mudelite seadistused ja arhitektuur.

¹ Bitcoinini tunni intervalliga andmete suure mahu tõttu vähendati töö käigus andmestiku ajavahemikku. Sel põhjusel on mitme sagedusega mudelites vähendatud sarnaselt ka päeva intervalliga aegrea pikkust.

² Ethereumi tunni intervalliga andmete suure mahu tõttu vähendati töö käigus andmestiku ajavahemikku. Sel põhjusel on mitme sagedusega mudelites vähendatud sarnaselt ka päeva intervalliga aegrea pikkust.

Analüüsi esimeses faasis kasutati esialgseid päevaseid ja tunniseid andmeid, millega loodi vastavalt madala ja kõrge sagedusega mudelid. Nende mudelite abil ennustati sama resolutsiooniga tuleviku andmed ja saadi esialgsete mudelite täpsuse andmed. Analüüsi teises faasis rakendati aga mitut ajasagedust üheaegselt ning loodi edasiarenenud mudelid.

4.2.1 Mitme sagedusega mudelite meetodid

Seoses valitud algoritmide erinevustega ei olnud võimalik kasutada ainult ühte meetodit mitme erineva ajasageduse ühendamiseks mudelites. Sel põhjusel lahendati ülesanne vastavalt olemasolevatele võimalustele kolmel erineval viisil.

Esialgne lähenemisviis oli kasutada samaaegselt päeva- ja tunniintervalliga andmestikke ühes mudelis nii, et päevaandmed on liigendatud tundideks ja on mudelisse lisatud eksogeense muutujana. Sel viisil saab andmetele rakendada erinevaid algoritme ja eelduslikult leida parema prognoosivõimega mudelid, mis ei ole liialt mõjutatud kõrgema ajasageduse hüpetest ning arvestavad madalama ajaintervalli määratud pikaajalisi trende.

Teine variant mitut ajaintervalli korraga rakendada pärineb ühest viidatud teadustööst [18]. Kuigi sealses töös olid kasutusel mitmemõõtmelised andmed kahel ajaintervallil, on võimalik sama loogikat käesolevas töös rakendada. Esiteks kasutatakse valitud algoritmi madalama sagedusega andmetel – artiklis olid nendeks aastased andmed, antud töös aga päevased – mille abil prognoositakse sulgemishindasid. Teise sammuna võetakse need prognoositud väärtused täiendava muutujana kõrgema sagedusega andmete mudelisse – artiklis olid nendeks päevased hinna andmed, käesolevas töös aga tunnihinnad. Sel viisil hierarhilise ansambelmeetodiga ajaintervalle kombineerides saavutati viidatud artiklis parema prognoosivõimega mudel kui üksikutel andmestikel ja võib osutada edukaks ka käesolevas magistritöös.

Viimaseks lähenemiseks on eri ajaintervallidega mudelid ühendada klassikalise ansambelõppe abil. Ansambelõppe on masinõppe lähenemine, kus mitme mudeli kombineerimisel luuakse üks ühendatud mudel, mis on võimeline andma paremaid ennustusi kui üksikud mudelid eraldi. Selle meetodi peamine eesmärk on vähendada üldist ennustusviga, suurendades mudelite mitmekesisust ja ühendades nende ennustused [45]. Ansambelõppe tehnikad erinevad keerukuse tõttu oluliselt ja lisaks saab neid ka rakendada eri moodi – kas paralleelselt või järjestikusest. Siinkohal on fookuses vaid

paralleelne rakendusviis. Põhimõtteliselt võib ansambelõppe tehnikad jagada kaheks suureks grupiks [46]:

1. Põhilised ansambeltehnikad

- a. Keskmistamine – arvutatakse mitme mudeli poolt loodud ennustuste keskmine, mis on lihtne ja efektiivne ning kasutatakse sageli regressiooni probleemide lahendamisel.
- b. Kaalutud keskmistamine – sarnane meetod nagu eelmine, kuid iga mudeli ennustused kaalutakse vastavalt selle usaldusväärsusele ehk see annab rohkem kaalu parematele mudelitele ning vähendab nõrgemate mõju.
- c. Maksimaalne hääletus (*Max Voting*) – lõplik ennustus on mitme mudeli poolt esitatud ennustustest selgunud kõige sagedam klass või väärtus. Sel põhjusel sobib see hästi just klassifitseerimisprobleemidele.

2. Täiustatud ansambelõppe tehnikad

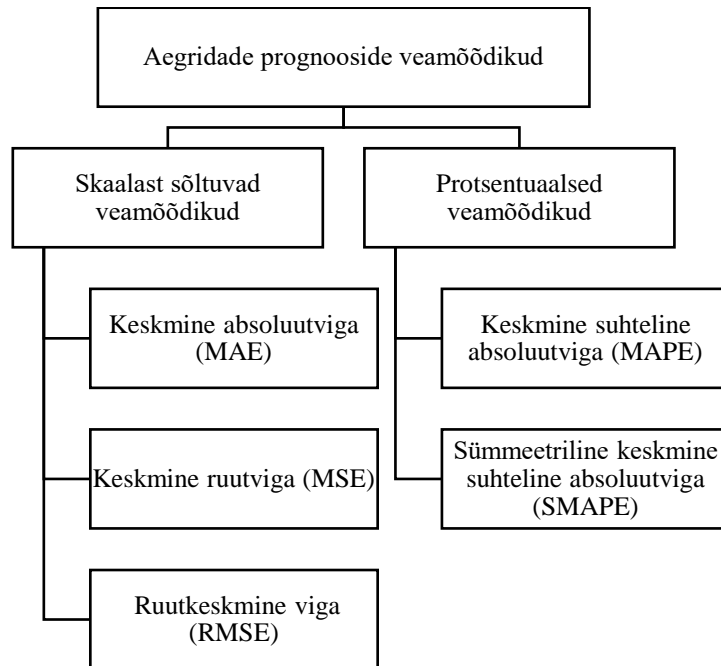
- a. Kuhjamine (*Stacking*) – kombineeritakse mitme mudeli ennustused, luues uue mudeli, mis õpib ennustama esialgsete mudelite ennustusi.
- b. Segamine (*Blending*) – sarnane meetod eelmisega, kuid erinevus seisneb selles, et andmestik jagatakse kaheks osaks – treeningandmeteks esialgsetele mudelitele ja valideerimisandmeteks niiõelda meta-mudeli jaoks.
- c. Pakkimine (*Bagging*) – luuakse mitu mudelit, kasutades treeningandmete erinevaid alamkogumeid, mis on juhuslikult valitud.
- d. Hoogustamine (*Boosting*) – mudelid treenitakse järjestikku, kus iga järgnev mudel keskendub eelnevate mudelite vigadele.

Käesoleva magistritöö eesmärgi täitmise seisukohast pole niivõrd oluline, millise ansambelõppe tehnika kasuks otsustada. Eesmärk on lihtsalt selgitada, kas ja mil määral muudab täiendava ajaintervalli kaasamine prognoosimudelisse selle ennustusi täpsemaks. Seega võeti mitme ajasagedusega andmete analüüsis kasutusele just kõige lihtsam ansambelõppe tehnika ehk keskmistamine.

4.3 Analüüsis kasutatavad veamõõdikud

Selleks, et analüüsi käigus arendatud meetodite prognoositud tulemuste täpsust kontrollida ja omavahel võrrelda, on valitud teatud hulk veamõõdikuid, mis sobivad just aegridade prognoosimise jaoks. Mõõdikute väljavalimisel lähtuti viiest kõige

populaarsemast mõõdikust, mille võib omakorda jaotada kahte gruppi [47] nagu näha Joonisel 9:



Joonis 9. Aegriade prognoosimise veamõõdikud [47].

Valitud veamõõdikute leidmiseks on kasutusel konkreetsed valemid [48]:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (14)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (15)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (16)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100 \quad (17)$$

$$SMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{(|\hat{y}_i| + |y_i|)/2} \times 100, \text{ milles:} \quad (18)$$

\hat{y}_i esindab i-nda ajaperioodi või vaatluse prognoositud väärtust,

y_i esindab i-nda ajaperioodi või vaatluse tõelist väärtust,

n esindab ajaperioodide või vaatluste summat.

Veamõõdikute võrdlemisel näitab kõige madalam väärtus kõige väiksemat prognoosiviga ehk kõige täpsemat prognoosi.

4.4 Mudelite seadete valik

Kui üldiselt viib mudelite treenimine välja mudeli parameetrite väljaselgitamiseni, siis mudeli seaded peab paika panema juba enne vastava meetodi rakendamist treeningandmetel. Teisisõnu võib mudeli seadeid nimetada ka mudeli hüperparameetriteks (*hyperparameters*). Statistiliste mudelite puhul oli sobiva parima mudeli valik suhteliselt kerge ülesanne, kuid sügavõppe meetodite rakendamine tõi endaga kaasa ka palju keerukama protsessi, et valida algoritmidele sobilikud hüperparameetrid. Järgnevalt on toodud ülevaade peamistest abivahenditest, mida kasutati mudelite parimate tehniliste seadete valimiseks, et loodud mudelid ei oleks alasobituvad ega ka ülesobituvad ning annaks võimalikult täpsed ja realistlikud tulemused nii treening- kui testandmestiku peal.

4.4.1 Eksponentsiaalne silumine

Eksponentsiaalse silumise puhul kasutati kõige lihtsamat meetodi varianti. Teisisõnu võibki seda nimetada kas lihtsaks (*Simple*) või üksikuks (*Single Exponential Smoothing*) eksponentsiaalseks silumiseks. Sellisel juhul võetakse mudeli treenimisel arvesse vaid ühte silumiskoeffitsienti alfa, mis ongi eksponentsiaalse silumise puhul peamine hüperparameeter [49].

4.4.2 ARIMA

ARIMA meetodi puhul on juba metoodika osas detailselt kirjeldatud selle peamisi seadeid, mida tuleb andmete treenimisel arvesse võtta. Need on p - autoregressiivne komponent, d - integreeritud komponent ja q - libiseva keskmise komponent. Kuna käesolevas töös oli aga erinevaid andmestikke mitu ning ka mitu erinevat ajaresolutsiooni, võeti parimate komponentide valikuks kasutusele automaatse ARIMA pakett, mille leiab Pythonis raamistikust nimega 'pmdarima'. Vastava raamistiku funktsioon nimega 'auto_arima' võimaldas kiirelt katsetada läbi mitmed võimalused ARIMA meetodi komponentidest ning andis tulemuseks kõige täpsema, mida siis ka analüüsis rakendati.

4.4.3 Masinõppe ja sügavõppe meetodid

Sügavõppemeetodite puhul oli hüperparameetrite valimise protsess suhteliselt sarnane, seega saab teha sellest meetodite ülese kokkuvõtte. Esimesed mudelid valmisid standardsete hüperparameetrite paikapanemisega, aga ootuspäraselt ei andnud need kõigi

andmestike ja meetodite puhul parimaid tulemusi. Lihtsalt katse-eksituse meetod poleks olnud mudelite mahtu arvestades kõlblik. Seetõttu võeti Pythoni abiga kasutusele juba olemasolevad tööriistad.

Võrgu otsing (*Grid Search*) - Võrgu otsing on lihtne võimalus erinevate hüperparameetrite kombinatsioonide treenimiseks treeningandmestikul ning testimiseks valideerimisandmestikul. Selle peamine eeldus on, et sobivad seadete kombinatsioonid on juba domeeni valitud. Ja suurimaks puuduseks on see, et valideerimisel on tegu väga aja- ja ressursimahuka meetodiga, kuna läbi on vaja testida kõik võimalikud kombinatsioonid [50]. Siiski osutus meetod väga kasulikuks masinõppe meetodite parameetrite häälestamisel. Tugivektorklassifitseerija sobivaimate mudeli seadete valimisel rakendati scikit-learn paketi funktsiooni GridSearchCV, mille abil sai luua võimalike parameetrite variantide võrgu, sobitada nende erinevaid kombinatsioone treeningandmetele ja treenida siis SVR algoritmi ning leida kõige väiksema veaga mudeli parameetrid, mida siis testandmestikul rakendada [51]. Otsustusmetsa mudeli puhul tuli aga appi paketi skforecast funktsioonid ForecasterAutoreg ning grid_search_forecaster [52]. Esimene neist aitas otsustusmetsa meetodil prognoosimudelit ehitada ning teise abil viidi läbi mudeli seadete häälestamine leidmaks parimad hüperparameetrid testandmestikul rakendamiseks.

Juhuslik otsing (*Randomized Search*) - Eelmise meetodi mõningad puudused ajendasid kasutusele võtma juhusliku otsingu. Kui võrgu otsing peab läbi katsetama kõik võimalikud variandid ja seetõttu jõuab oma otsingutes ka täiesti ebatõenäoliste kombinatsioonideni, siis juhuslik otsing õpib juba otsingu keskkonda ja proovib leida kõige lähedasema kombinatsiooni oluliselt väiksema arvu iteratsioonidega [50].

Lõplike hüperparameetrite valikuks otsustati siiski kasutusele võtta KerasTuner raamistik. Kuna ka mudelite ehitamisel oli kasutusel just Keras raamistik, oli väga loogiline samm võtta hüperparameetrite häälestamise jaoks kasutusele KerasTuner, mille sees omakorda rakendati RandomSearch klassi [53].

KerasTuner on raamistikega TensorFlow ja Keras integreeritud hüperparameetrite häälestamise tööriist, mis võimaldab automatiseerida mudeli seadete optimeerimist. KerasTuner pakub lihtsat ja tõhusat viisi mudelite seadete optimeerimiseks, võimaldades määrata hüperparameetrite otsinguruumi ja seejärel leida parimad seaded automaatselt.

Koos otsinguruumi määramisega võimaldab KerasTuners määrata võimaliku õppimiskiiruse, võrgu arhitektuuri, aktiveerimisfunktsioonid ja muud parameetrid. Seejärel kasutab KerasTuner erinevaid optimeerimisalgoritme, et leida parimad hüperparameetrid. Selle abil saab ka mitmeid võimalikke mudeli versioone treenida ja hinnata, mille tulemusena valitakse parim mudel vastavalt määratud hindamiskriteeriumitele. Lisaks on tööriista kasutamine väga lihtne ja aega kokkuhoidev parandades omakorda mudelite jõudlust ja täpsust [54].

Sügavõppe meetoditega rakendatud mudelite puhul võib seadete valikut detailsemalt vaadelda, et mõista vajaliku analüüsi hulka. Töö analüüsi esimeses faasis rakendati kõigile sügavõppe meetoditele standardsed hüperparameetrid:

- Neuronite hulk kihis = 50
- Aktiveerimisfunktsioon = 'relu'
- Optimeeriija = Adam
- Kihid = 1

Lihtsuse huvides oleks saanud jätta mudelite seaded kõigil algoritmidel ühesuguseks ja teha lihtsustatud mudelite tulemuste pealt ka järeldused. Kuna KerasTuner on väga hea abivahend meetodite hüperparameetrite häälestamiseks ja RandomSearchi abil parimate seadete leidmiseks, siis otsustati ikkagi iga mudeli osas seda rakendada ja leida sobivaimad ja parima üldistusvõimega hüperparameetrid vastavalt valitud meetodile ja ka andmestikule.

Valitud kihtide osas siiski muudatusi ei tehtud ja igale sügavõppemeetodile jäeti ainult üks kiht. Küll aga katsetati erineva arvu neuronitega ehk sõlmedega vastavas kihis. Kui esialgsete seadete järgi oli ühes kihis kokku vaid 50 neuronit (*units*), siis KerasTuneri abil valiti sellised võimalikud variandid, millest leiti parim:

```
units = hp.Int('units', min_value=32, max_value=256, step=32) (19)
```

Aktiveerimisfunktsioonid on sügavõppemeetodites olulisel kohal, kuna toovad mudelisse sisse mittelineaarsuse, mis võimaldab mudelitel õppida andmete keerukaid mustreid [55]. Võimalikke aktiveerimisfunktsioone on aga mitmeid – tuntuimad just sigmoidne, hüperboolse tangensi (lüh *tanh*) ja parandatud lineaarse ühiku (*Rectified Linear Unit*, ReLU) aktiveerimisfunktsioonid. Käesolevas töös otsustati jääda aktiveerimisfunktsioonide valikul võimalikult lihtsaks ja lähtuda varasematest töödest,

kus on analüüsitud, milline neist sobib kõige paremini sarnase aegrea prognoosimise ülesande lahendamiseks [56]. Seetõttu anti KerasTunerile valimiseks kaks varianti – tanh ja relu, millest iga meetodi ja andmerea puhul valiti sobivam variant.

Optimeerija on teisisõnu algoritm, mida kasutatakse närvivõrgu kaalude värskendamiseks treenimise ajal, eesmärgiga vähendada mudeli kadu. Optimeerija ülesanne on leida selliste parameetrite kogum, mis minimeeriks kaofunktsiooni, mis mõõdab mudeli prognoositud väljundi erinevust tegelikest väärtustest. Seetõttu on optimeerija kasutamine sügavõppemeetodite puhul väga oluline, aidates parandada mudelite viljakust [57]. Populaarsematest optimeerijatest on kindlasti tuntuim stohhastiline gradientlaskumine (*Stochastic Gradient Descent*, SGD). See on lihtne ja efektiivne optimeerija ja on väga laialdaselt kasutusel, kuid mitte kõigi probleemidega sobiv variant. Teisalt on Adam (*Adaptive Moment Estimation*) kõrgetasemeline optimeerija, mis kombineerib mitme erineva optimeerija head omadused, mistõttu on tuntud kui üks efektiivsemaid optimeerijaid. Lisaks peetakse Adam optimeerijat sobilikuks just aegrea prognoosimise ülesannetes, kuna see on kohanemisvõimeline õppimismäärade valiku osas ning võimaldab robustset õppimismäärade häälestamist. Seetõttu jäädi käesolevas analüüsis samale optimeerijale kindlaks kõigi sügavõppe meetodite puhul.

Igale optimeerijale vastab oma reegel uuendamise jaoks ja lisaks on neil erinevad hüperparameetrid, mis vastava optimeerija käitumist juhivad. Adam optimeerija puhul ongi selleks just õppimismäär. Standardne määr on 0,01, kuid KerasTuneri abil sai määrata alternatiivseks variandiks ka 0,001 ning sel viisil selgus, kumma õppimismääraga on tulemused täpsemad ja suurema üldistusvõimega.

Kõigi mudelite lõplikud seadistused leiti seega häälestamise tulemusel. Ühe ajasagedusega mudelite hüperparameetrid on paigutatud Lisasse 9. Mitme ajaintervalliga mudelite lõplikud seaded on kirjas aga Lisas 10. Sealjuures pole märgitud keskmistatud ansambelmeetodi tulemusi, kuna selle jaoks on kasutusel ühe ajasagedusega mudelite väljundid ja eraldi hüperparameetreid polnud vaja leida.

5 Tulemused

Vastavalt eelnevalt kirjeldatud magistritöö analüüsi arhitektuurile algas andmetega töö päevaste andmete analüüsist. Eesmärk oli selgitada erinevate finantsandmete aegridade näitel valitud meetodite tõhusust päevaste prognooside tegemisel ja leida ka konkreetseid veamõõdikud, mille abil prognoosivõimet hinnata.

Vastavalt läbitöötatud varasemale kirjandusele said välja valitud lihtsad statistilised meetodid, mis käitusid analüüsis kui mõõdupuud, millega keerukamaid algoritme võrreldi. Nendeks kaheks referentsmudeliks olid eksponentsiaalne silumine (ETS) ja ARIMA, mille kohta ka eelmises peatükis lühike ülevaade tehti.

Masinõppemeetodid olid analüüsis kasutusel kui keskmise keerukusega mudelid. Need olid justkui samm edasi traditsioonilistest meetoditest, kuid mitte veel nii keerukad kui järgnevas faasis rakendatavad sügavõppemeetodid. Kaheks masinõppe mudeli näiteks valiti otsustusmets (RF) ja tugivektorregressioon (SVR).

Sügavõppe meetoditest valiti välja need algoritmid, mida varasemas kirjanduses kõige enam kajastati. Nende kõigi kohta on eelnevalt nähtud suuri edusamme just aegridade prognoosimisel ja eriti võrreldes nende produktiivsust traditsiooniliste statistiliste meetoditega. Sel põhjusel osutusid sügavõppe meetoditest analüüsi valituks mitmekihiline närvivõrk (MLP), rekurrentne närvivõrk (RNN), pikk lühiajaline mälu (LSTM) ja värvatega rekurrentne üksus (GRU).

5.1 Madala, päevase sagedusega analüüs

Esimene osa analüüsist keskendub madala sagedusega ehk päevaste andmete analüüsile. Analüüsi võeti finantsandmete päevased sulgemishinnad vastavalt olemasolevale ajavahemikule. Päevaste andmete analüüs on vaheetapp enne mitme ajaintervalli rakendamist üheaegselt, seega neid tulemusi ei saa otseselt võrrelda tunniintervalliga andmetel leitud tulemustega. Tehtud mudelite prognoosivõime ja täpsus testandmetel on välja toodud Lisas 4.

Tabeli esimeses veerus paiknevad uuritavad finantsvarad, mille tulemused on vastavates ridades kirjas. Teises veerus on välja toodud kasutatud meetodid, millele igal real vastavad veamõõdikute väärtused. Igas järgmises veerus on konkreetse veamõõdiku tulemus, millele viitab ka tabeli päis. Iga finantsvara puhul on esile tõstetud kõige täpsemate veamõõdikutega meetodid, mille tulemused on kirjas rasvasena.

Esmasel vaatlusel on selge, et statistilised meetodid ei suuda võtta arvesse finantsaegridade mittelineaarset iseloomu, mistõttu on prognoosid väga ebatäpsed ja veamõõdikud hoopis kõrgemal skaalal kui keerukamate mudelite puhul. Siiski ei ole järeldus traditsiooniliste meetodite osas täiesti must-valge. Näiteks S&P 500 indeksi puhul jäävad eksponentsiaalse silumise ja ARIMA tulemused väga sarnaseks. Seetõttu pole võimalik täiesti veendunult öelda, kumb meetod on täpsem, kuna näiteks keskmise absoluutvea MAE puhul oleks eelistatud esimene, ruutkeskmise vea RMSE puhul aga teine. Seega oleneb otsus sellest, millise veamõõdiku tulemust eelistada. Teiste finantsvarade puhul on erinevus kahe statistilise meetodi vahel selgem, aga sellegipoolest on näiteks Dow Jones indeksi puhul täpsem eksponentsiaalne silumine, Nasdaq 100 puhul aga ARIMA meetod. Minnes edasi krüptovaluutade poole, on selgelt ARIMA meetod eduseisus, eksponentsiaalse silumisega võrreldes.

Masinõppe meetoditest on otsustusmetsa prognoosid sarnased pigem statistiliste meetoditega, kuna prognoosid on tehtud lihtsuse huvides kasutades vaid viitaegasid. Meetodi puhul oleks saanud kasutada ka lükandakna lähenemist (*sliding window approach*), kuid selle tehniline teostus ei õnnestunud käesolevas töös ja otsustati jääda lihtsa lahenduse juurde. Siiski pakub otsustusmetsa lähenemine teatud juhtudel parema täpsusega prognoose võrreldes statistiliste meetoditega, näiteks S&P 500, Dow Jones ja Ethereum. Teiste finantsvarade puhul jääb meetod aga ühele või mõlemale traditsioonilisele meetodile siiski alla.

Tugivektorklassifitseerija on aga märgatavalt edukam täpsete prognooside pakkumisel ja selle prognoosivõimet võib juba võrrelda sügavõppemeetoditega. Kahel juhul on SVR mudeli prognoositud tulemused isegi kõige täpsemad – S&P 500 ja Ethereumi puhul. Esimese puhul on ennustatud väärtuste võrdlusele tegelikkusega kõik veamõõdikud madalamad teistest meetoditest. Kõige lähemale küündivad GRU meetodiga prognoositud väärtused, mille prognoosivõime on vaid veidi madalam – kui SVR puhul jääb keskmine suhteline absoluutviga MAPE 1,1% juurde, siis GRU puhul on see 1,3%.

Samasse suurusjärku jääb MAPE ka LSTM meetodi puhul – 1,5%. MLP ja RNN puhul on viga juba märgatavalt suurem, vastavalt 2,1% ja 3,0%.

Kuna S&P 500 indeksi on ka mitmetes varasemates töödes põhjalikult uuritud, võiks võrrelda prognoosi tulemusi teiste autoritega. Ühes töös on vaatluse all just SVR ja MLP meetodid [11], mille S&P 500 indeksi prognoosimine annab tulemuseks MAPE vastavalt 2,91 ja 2,35. Käesoleva töö tulemustest on need seega küll kõrgemad, kuid suurusjärg on siiski sarnane. Tuleb ka silmas pidada, et kasutatud ajavahemik on erinev, mistõttu ei saakski tulemused identsed olla.

Mõnes töös on sarnaselt vaatluse all olnud palju erinevaid baasmudeleid koos mitmekesiste sügavõppemeetoditega, näiteks järgnev töö [58], millest võib leida prognoosimise tulemused S&P500 indeksi kohta nii SVR, MLP, RNN kui ka LSTM meetodiga. Kui SVR ja LSTM meetodite prognoositulemused ehk MSE väärtused 4272 ja 4905 on võrreldavad käesoleva töö tulemustega 3090 ja 5574, siis teiste osas on käesoleva töö tulemused olulised ebatäpsemad. Samas tuleb ka silmas pidada, et võrreldava artikli ajaperiood on 2000-2020, mistõttu neid täielikult võrrelda ei saagi.

Nagu mainitud, on ka Ethereum puhul parimad prognoosid loodud just SVR mudeli poolt. Tegelikult on ka GRU mudeli prognoositud väärtused väga täpsed, jäädes minimaalselt SVR tulemustele alla. Kui SVR meetod annab tulemuseks MAPE väärtusega 2,2%, siis GRU puhul on see vaid veidi kõrgem 2,3%. Natuke kõrgema tulemuse annab MLP – 2,6% ja märgatavalt suurema veaga on juba LSTM ja RNN, vastavalt 3,1% ja 3,7%.

Ethereumi andmetel on varasemates töödes uuritud erinevate meetodite prognoosivõimet, näiteks ARIMA, LSTM ja viimase baasil leitud hübriidmudelit [59]. Neid tulemusi saab teatud piirideni ka käesoleva tööga võrrelda. Artiklis pole küll vaatluse all ühemõõtmelise aegrea prognoosimine ja ajaperiood on erinev, kuid tulemused on siiski võrreldavad. Pikaajalise ARIMA prognoosi puhul on artiklis tulemused RMSE 561,28 ja MAPE 21,16%. Mõlemad veamõõdikud on seega sarnases suurusjärgus käesoleva töö tulemustega – vastavalt 477,51 ja 21,0%. Lihtsa LSTM meetodi puhul on artiklis küll tulemused veidi tagasihoidlikumad kui käesolevas töös – MAPE on seal 8,95% ja käesolevas töös 3,1%.

Dow Jones indeksi puhul annab üllatuslikult kõige kehvema tulemuse lihtne RNN meetod, mil keskmine suhteline absoluutviga on 2,7%, kuid see sarnaneb ka S&P 500 andmestikul saadud tulemustega. Samas on rekurrentse võrgu keerukamad edasiarendused LSTM ja GRU mõlemad palju parema prognoosivõimega, sest mõlemal juhul on MAPE 0,8%. Esimene on küll teisest täpsem, mida näitavad madalamad väärtused teiste veamõõdikute osas. Kõige täpsema tulemuse annab aga hoopis MLP, mille MAPE väärtus on küll täpselt sama nagu eelmistel meetoditel, kuid teiste veamõõdikute poolest on MLP täpsem. Siinkohal jääb SVR meetodi tulemus keskpäraseks ning MAPE näitab 1,6% suurust viga.

Ülejäänud finantsvarade puhul on GRU meetodiga leitud prognoosid kõige täpsemad. Nasdaq 100 indeksi osas on olukord veidi sarnane teiste aktsiaindeksite tulemustes nähtuga – kõige viletsama tulemusega on sügavõppemeetoditest jälle RNN, mille keskmine suhteline absoluutviga on 6,1%. SVR prognoositud ennustused saavutavad kolm korda väiksema vea 2,1%, kuid jäävad siiski teistest meetoditest maha. Veelgi parema tulemuse annavad nii MLP kui ka LSTM, mille MAPE on vastavalt 1,5% ja 1,2%. Kuid nagu öeldud, on kõige täpsema tulemusega GRU poolt saadud tulemus, mil vea suurus vaid 1,1%.

Bitcoin'i puhul on veamõõdikute võrdlus samuti kirju ja parima keskmise suhtelise absoluutvea saavutavad võrdselt kaks erinevat meetodit – SVR ja GRU – mil veaprotsendiks on 2,0%. Vaadates tulemuste tabelit detailsemalt, on aga märgata, et keskmise absoluutvea puhul on täpsust võimalik eristada – kui SVR puhul on selleks 605,7, siis GRU puhul on viga väiksem väärtusega 601,1. Täpsuse poolest jääb kolmandale kohale LSTM, mille MAPE 2,4% jääb veel sarnasesse suurusjärku. Suurema vea annab aga lihtne rekurrentne närvivõrk RNN, mille MAPE on juba 3,9% ning kõige ebatäpsema tulemuse annab MLP, mil veaprotsent 7,2%.

Kuna krüptovaluutadest on just Bitcoin'i kõige enam varasemates töödes uuritud, tasub tulemusi varasematega võrrelda. Bitcoin'i päevast sulgemishinda prognoosinud artikkel [12] on võtnud vaatluse alla mitu erinevat meetodit, kaasa arvatud SVR ja LSTM. Veamõõdikutest on võetud kasutusele RMSE ja MAPE, kuid kuna ajavahemik erineb oluliselt, siis on raske mõlemat mõõdikut käesoleva töö tulemustega üks-ühele võrrelda. MAPE võib aga anda kinnitust selles osas, kas tulemused on loogilised võrreldes varasemaga. Võrreldavas artiklis on MAPE tulemused vastavad: SVR – 2,86 ja LSTM –

2,59. Võrreldes käesoleva magistritöö tulemusi 2,0% ja 2,4% varasematega, võib veenduda, et suurusjärgud on sarnased ja tulemused loogilised.

Päeva sagedusega andmeridade analüüsi üheselt kokku võtta on küll keeruline, kuna tulemusena ei selgunud ainsat teisi variante edestavad meetodit. Küll aga selgusid kaks parimat SVR ja GRU. Esimese puhul on tegu masinõppe meetodiga, mille arhitektuur on sügavõppe meetoditest lihtsam ja seega töötab ka kiiremini ning seega võib eeldada, et ka töö järgmises faasis võib see anda häid tulemusi. GRU puhul on tegu rekurrentse närvivõrgu edasiarendusega, mis sisaldab sarnaselt LSTM meetodiga mälu säilitamist ja kontrollimist, kuid tänu väiksemale hulgale värvatele on see lihtsam ja kiirem meetod.

5.2 Kõrge, tunnise sagedusega analüüs

Analüüsi teine osa keskendus finantsvarade tunniintervalliga hindade aegriidade prognoosimisele. Eesmärk oli võtta olemasoleva ajaperioodi tunniintervalliga andmed, millele rakendada samuti erinevaid prognoosimise meetodeid, leida veamõõdikud ja siis neid omavahel võrrelda. Ühtlasi on tunniandmete analüüs aluseks järgnevale mitme ajaintervalli kombineerimisele.

Tunniandmete analüüsi alguses oli selge, et andmemaht läheb väga suureks ja seetõttu võivad olla vajalikud teatud mõõndused, et hoida kokku väärtuslikku tööaega ja mitte kulutada seda liigselt mudelite jooksmisele. Kuna aktsiaindeksid kauplevad vaid tööpäevadel ja sealjuures ainult tavapärasel töötundidel, siis suurenes andmemaht vaid kaheksa korda, millega oli veel võimalik mõistlikkuse piirides töötada. Krüptovaluutade puhul on olemasolev ajavahemik oluliselt lühem, kuna need on turgudel kaubelnud palju lühemat aega. Siiski muudab krüptorahade analüüsi väga mahukaks see, et nendega kaubeldakse igapäevaselt ja igal kellaajal, mistõttu päevaste andmetega võrreldes kasvas andmemaht 24-kordselt. See tähendas, et kui aktsiaindeksite andmerea pikkus oli tunniandmete analüüsis 32 295 tundi, siis näiteks Ethereumi puhul oli andmeid kaks korda rohkem – 68 694 tundi – Bitcoin'i puhul aga pea kolm korda rohkem – 94 550 tundi. See raskendas oluliselt mudelitega töötamist, kuna ainuüksi mudelite jooksmine võttis ebamõistlikult kaua aega, rääkimata veel hüperparameetrite häälestamisest, millega läks veelgi kauem aega. Sel põhjusel võeti vastu otsus lühendada krüptovaluutade andmestikku oluliselt, võttes algus punktiks 01.01.2019. Sel viisil oli võimalik mõlema krüptoraha aegrea puhul vähendada andmestiku pikkust 44 tuhande tunnini ja mõistliku

ajaga analüüse katsetada ilma, et tulemused seetõttu oluliselt kannataks. Kõrge sagedusega andmetele rakendatud mudelite tulemused on koondatud Lisas 5.

Tunniintervalliga andmetega aegridadel läbiviidud analüüs näitab samuti, et statistilised meetodid jäävad edasiarenenud meetodite kõrval lahjaks, kuna nad ei suuda võtta arvesse mittelineaarseid seoseid aegrea elementide vahel. Siiski on üldiselt ARIMA meetodi abil leitud prognoosid täpsemad kui eksponentsiaalse silumise puhul. Vaid Dow Jones indeksi puhul on olukord vastupidine. Bitcoin'i puhul üllatab ARIMA meetodi poolt prognoositud väärtuste suhteliselt madalad veamõõdikute tulemused. Kuigi need on ikkagi mitu korda kõrgemad kui sügavõppemeetodite puhul, on tulemus siiski paljulubav arvestades ARIMA meetodi lihtsust ja selle rakendamise mugavust. See on ka sarnane olukord ühe varasema artikliga [14], kus tänu tõusvale trendile oli ARIMA prognoosivõime väga hea.

Masinõppemeetoditest jääb otsustusmetsa algoritm jälle sarnasele tasemele statistiliste meetoditega andes mitmete varade puhul isegi kõige kehvemad prognoosid. See paistab eredalt välja just krüptovaluutade puhul, mis on erakordselt volatiilsed ning ajalooliste andmete põhjal teeb otsustusmetsa mudel tegelikkusest kaugemale kalduvad ennustused. Samas, S&P 500 ja Nasdaq 100 puhul suudab RF anda statistilistest meetoditest täpsemad prognoosid.

Järgmiste meetodite puhul on veamõõdikud juba oluliselt madalamad, mistõttu tasub neile detailsemalt tähelepanu pöörata. Nii S&P 500 kui ka Dow Jones indekseid puhul prognoosib testandmete peal kõige täpsemalt juba eelnevas alapeatükis täpsusega silma paistnud GRU. S&P 500 puhul on GRU ennustuste keskmine suhteline absoluutviga 0,3%. Tegelikult on väga täpse tulemusega ka SVR meetod, mille prognoosi MAPE on 0,4%. Pea kaks korda suurema MAPE annavad RNN meetodi poolt prognoositud tulemused – 0,7%. Järgnevate meetodite, LSTM ja MLP prognoositulemused on ebatäpsemad kui eelmised, vastavalt 1,2% ja 1,5%.

Dow Jones indeksi puhul on GRU meetodi abil saadud prognooside täpsust näitav MAPE 0,4% ja teiste veamõõdikute abil saab olla kindel, et tegu on antud juhul kõige täpsema meetodiga. Suhteliselt lähedase tulemuse annab ka LSTM, mille MAPE on sama, kuid teiste veamõõdikute poolest jääb nõrgemaks. Järgnevad sarnaste prognoosivigadega SVR ja RNN, mille MAPE on vastavalt 0,5% ja 0,6%. Sügavõppemeetoditest jääb prognoositäpsuse poolest viimasele kohale MLP, mille puhul on MAPE väärtus 1,8%.

Järgmisel kahel finantsvaral – Nasdaq 100 ja Bitcoin – annab suurima täpsusega prognoosid RNN, mis päevaintervalliga analüüsis jäi pigem tahaplaanile. Nasdaq 100 indeksi puhul annab RNN tulemuseks prognoosid, mille MAPE on 1,4%. Sellele järgneb täpsuse poolest LSTM, mille veamõõdik on 1,6%. Järgmiste meetodite, GRU, SVR ja MLP puhul on MAPE väärtus juba kõrgemal tasemel, vastavalt 2,1%, 2,5% ja 2,6%.

Bitcoin'i puhul annavad RNN meetodi prognoositud väärtused tulemuseks MAPE 0,3% ja see osutub ka teiste veamõõdikute järgi kõige täpsemaks. Tegelikult on sarnaselt täpsed ka pea kõik teised meetodid – SVR, LSTM ja GRU, mille prognoosivõime veamõõdikutena kasutusel olev MAPE jääb vahemikku 0,3-0,4%. Ainsana on halvema prognoosivõimega MLP, mille MAPE on mitu korda suurem teistest, jäädes 1,4% juurde.

Bitcoin'i sulgemishinda on samuti uuritud lühema ajaintervalliga ning võrreldud kahte enamtuntud meetodit ARIMA ja LSTM [60]. Kuigi antud juhul oli autorite fookuses 10-minuti sulgemishindade prognoosimine, võib protsentuaalsete veamõõdikute poolest tulemusi käesoleva töö omadega võrrelda. ARIMA andis mainitud artiklis väga täpse tulemuse tänu sellele, et kasutati ühe perioodi etteprognoosimist igal sammul – MAPE oli 1,8%. LSTM oli siiski oluliselt täpsem andes MAPE väärtuseks 0,3% ja on suurusjärgu poolest väga sarnane käesolevas töös leitud 0,4%-ga.

Ethereumi puhul on väga täpse prognoosivõimega meetodeid mitu tükki. Kui vaadelda ainult MAPE väärtusi, ongi kolme meetodi – LSTM, RNN ja SVR – tulemused identsed. Tegelikult aga on meetodite prognoosivõime veidi erinev. Keskmine absoluutviga MAE on kõige madalam LSTM meetodil – 6,0 ja sellele järgnevad RNN 6,2-ga ja SVR 6,5-ga. Kaks esimest näitavad selgelt, kui hästi sobivad korduvad närvivõrgud väga volatiilsete finantsaegridade prognoosimiseks. Kolmele esimesele järgneb samuti rekurrentse närvivõrgu edasiarendus GRU, mille veamõõdik on 0,4%. Kõige ebatäpsemad prognoosid annab antud juhul MLP, mille keskmine suhteline absoluutviga on mitu korda suurem – 1,5%.

Tunniintervalliga andmestike prognoosimismudelite tulemused ei selgitanud samuti välja ühte konkreetset meetodit, mis kõigis olukordades kõige paremini toimib. Küll aga näitas tunnisagedusega analüüs selgelt, et korduvad närvivõrgud on põhjusega leidnud kasutust just aegridade prognoosimise ülesannete puhul. Analüüsi tulemusena saab parimate

meetoditena tuua välja just korduvate närvivõrkude baasmeetodi RNN ja selle edasi arendused pikk lühiajaline mälu LSTM ning väravatega rekurrentne üksus GRU.

5.3 Mitme sagedusega andmete analüüs

Analüüsi kolmanda osa fookuses olid mitme sagedusega andmete alusel tehtud prognoosid. Seega kasutati üheaegselt nii aegrea päevaseid kui ka tunniseid andmeridu, mille abil ennustati tunniseid väärtuseid.

Esmajärgus lähtuti lihtsast loogikast, milles lahendati aegrea prognoosimine tavapärase ühemõõtmelise aegrea asemel hoopis mitme muutujaga aegrea analüüsis. Selle jaoks oli vaja täiendavat andmestikku, kus mudeli väljundiks on küll endiselt tunnised sulgemishinnad, kuid sõltumatuid muutujaid mudeli jaoks ühe asemel kaks – kõigepealt tunnised sulgemishinnad ja seejärel päevased sulgemishinnad iga tunni kohta. Selle jaoks tehti andmestikule täiendus, milles on päevaandmestik liigendatud tunniintervallile ehk iga päeva tundidele vastab konkreetse päeva sulgemishind. Nii viidi läbi prognoosid pea kõigi meetodite puhul, kasutades mitme muutujaga aegrea prognoosimise tehnikat. Nende prognooside tulemused on paigutatud Lisasse 6.

Sellest analüüsi osast jäid välja nii statistilised meetodid eksponentsiaalne silumine ja ARIMA kui ka otsustusmets. Eksponentsiaalse silumise puhul oli takistuseks täiendava eksogeense muutuja lisamise võimatus. Lisaks, kuna ARIMA ja otsustusmetsa meetodite puhul ei kasutata libiseva akna lähenemist, ei saa samal viisil teise ajaintervalliga algandmeid mudelisse eksogeensete muutujatena lisada. Seetõttu kasutati teise sammuna mitme ajaintervalli ühendamiseks niiõelda hierarhilist ansambelõpet, mis on ka varasemalt edukalt kasutatud [18].

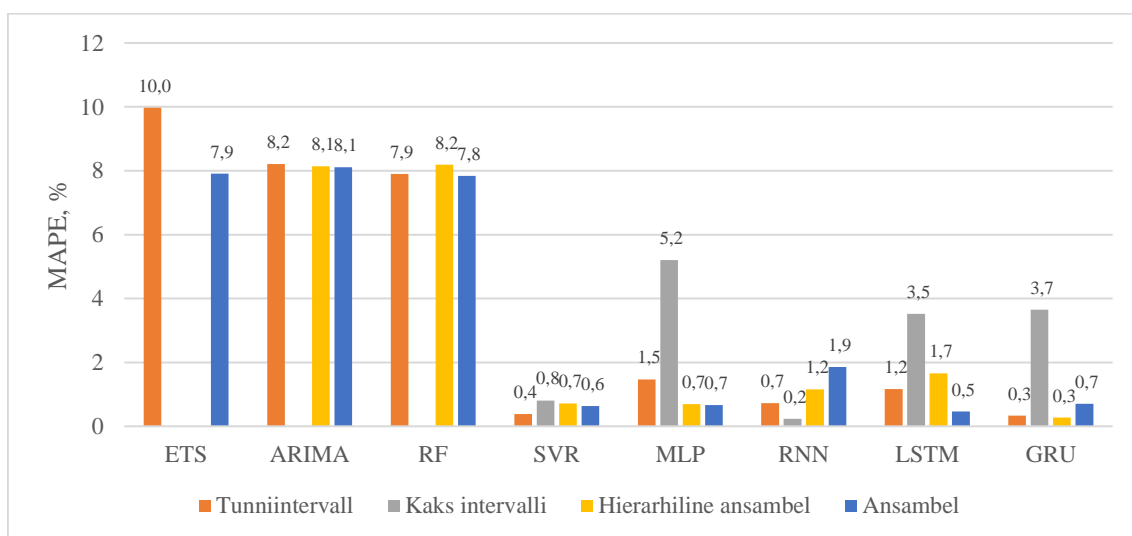
Hierarhilise ansambelõppe puhul võeti kõigi meetodite puhul päevaintervalliga andmetel loodud mudelite ennustused treeningandmetel kasutusse tunni intervalliga andmetel loodava mudeli eksogeense muutujana. Seetõttu jäi ka selles osas välja ETS meetod. See viis aga võttis mudeli trenimise faasis arvesse mõlemat ajasagedust – tunniintervalliga sulgemishindasid ning päevaintervalliga sulgemishindade ennustusi. Testimise faasis valiti prognooside perioodi pikkus vastavalt testandmestiku pikkusele ning eksogeenseks muutujaks valiti eelnevalt leitud päevaintervalliga mudelite prognoosid testandmetel, mis

olid siis samamoodi liigendatud tunniintervallile. Kõik analüüsi tulemused on koondatud Lisas 7.

Viimasena võeti mitme ajaintervalli ühendamiseks kasutusele ansambelõpe selle klassikalisel viisil. Kuna eksponentsiaalne silumine jäi välja mõlemast eelmisest meetodist, siis selle jaoks oligi ainuke variant kahe ajaintervalli kombineerimiseks kasutada lihtsat ansambelõpet ehk keskmistamist. Meetodit kasutati kõigi valitud algoritmide puhul ning prognoosi tulemused on Lisas 8.

Kõigi meetodite tulemused on paigutatud ülevaatlike graafikutena joonistele, et muuta võrdlemine lihtsamaks. Kõigil joonistel on vastava finantsvara prognoosi tulemused tunni intervalliga andmetel ning kahe ajaintervalliga mudelite tulemused kasutades eelpool mainitud kombineeritud tehnikaid. Edaspidi keskendub analüüs sellele, kas ja mil määral on kahe ajaintervalli kombineeritud meetodite prognoosid täpsemad kui ainuüksi tunniintervalliga andmetel läbiviidud mudelite ennustused.

Esimesena saab vaadelda S&P500 aktsiaindeksi sulgemishindade prognoositäpsuse võrdlust tunni sagedusega ja mitme ajaintervalli kombineerimisel Joonisel 10, kus on visualiseeritud kõigi meetodite keskmine suhteline absoluutviga MAPE.



Joonis 10. S&P 500 tunni ja mitme intervalliga mudelite prognoositulemuste võrdlus.

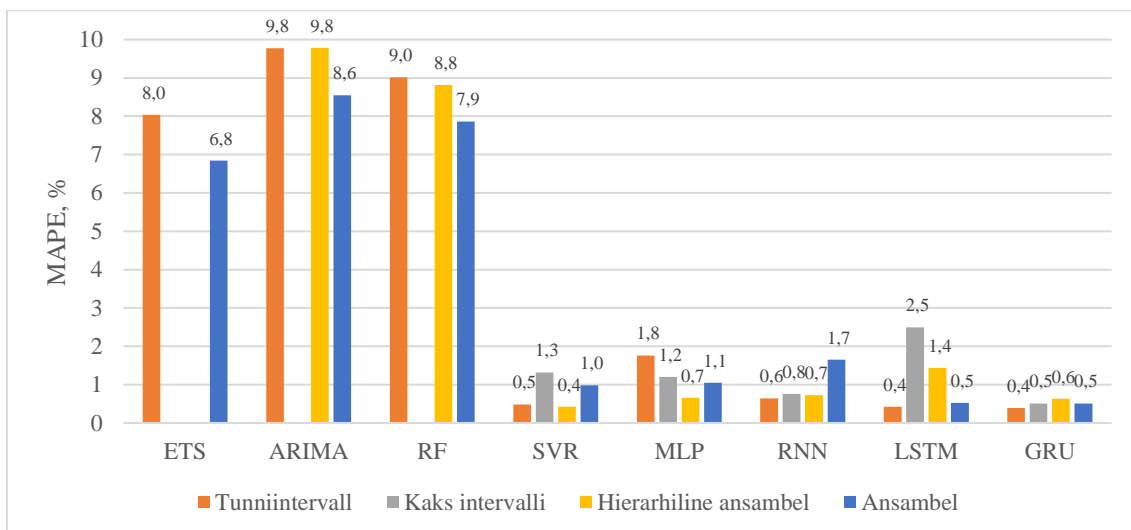
Alustades S&P 500 aktsiaindeksist, on selgelt näha, et ainuüksi kahe ajaintervalliga mudelid ei paranda prognooside täpsust võrreldes mudelitega, mis kasutasid sisendina vaid tunniintervalliga andmeid eraldi. Need tulemused on graafikul märgitud halli värviga

ja sealjuures on protsentuaalne viga oluliselt suurem nii SVR, MLP, LSTM kui ka GRU meetodite puhul. Vaid RNN meetodi puhul töötab kahe ajaintervalli kombineerimine paremini kui tunnisagedusega mudel – vastavad MAPE väärtused on 0,7% ja 0,2%, seega parandab teise intervalli lisamine oluliselt prognoosi täpsust.

Hierarhilise ansambelõppe puhul ei ole tulemus nii lihtsalt järeldatav. Joonisel 10 on need tulemused tähistatud kollase värviga. ARIMA puhul parandab teise intervalli lisamine mudelisse prognoosi täpsust esialgse 8,2% pealt 8,1% peale, kuid RF puhul on olukord vastupidine ning prognoosiviga läheb 7,9% pealt 8,2% peale. Prognoositäpsus väheneb ka SVR, RNN ja LSTM puhul, sest märgata on MAPE väärtuse suurenemist võrreldes tunniintervalliga mudelite tulemustele. Oluliseks erandiks on aga MLP, mille puhul viga väheneb esialgse 1,5% pealt 0,7% peale, kui kasutatud on hierarhilist ansambelõpet. Viimasena on näha vähest prognoosivõime paranemist ka GRU puhul. Graafikul on küll MAPE võrdne nii tunniintervalli kui ka hierarhilise mudeli puhul, kuid kontrollides lisades paiknevatest tulemuste tabelitest teiste veamõõdikute tulemusi, peegeldub, et GRU puhul muutub keskmine absoluutviga MAE 13,6 pealt 11,5 peale.

Viimasena rakendatud ansambelõppega keskmistamine võimaldas analüüsi kaasata ka eksponentsiaalse silumise, mis eelmistes etappides välja jäi. Need tulemused on graafikul tähistatud sinise värviga. Selgelt parandab kahe mudeli kombineerimine keskmistamisega oluliselt eksponentsiaalse silumise prognoositäpsust – 10% pealt 7,9% peale. ARIMA osas jääb tulemus sarnaseks eelmise meetodiga, parandades 0,1 protsendipunkti võrra esialgset tulemust. Sama on ka RF puhul, kus esialgse 7,9% vea suurus jõuab ansambelõppe abil 7,8% juurde. SVR, RNN ja GRU puhul ei suuda ansambelõppega leitud tulemused parandada prognoose, mis saadi vaid tunniintervalliga andmetega saadud mudelitest. Samas annab üllatavalt täpse tulemuse MLP meetodi kasutamine – MAPE on küll sama nagu eelmise, hierarhilise ansambelõppe puhul, kuid muude veamõõdikute põhjal (Lisad 7, 8) on keskmistamise tulemus veel täpsem. Lisaks on üllatavalt ansambelõppega keskmistamine muutnud oluliselt täpsemaks LSTM meetodil leitud prognoositäpsust – kui tunni mudeli tulemusel oli MAPE 1,2%, siis viimase meetodiga leitud MAPE on kõigest 0,5%.

Teisena saab võrrelda Dow Jones aktsiaindeksi sulgemishindade prognoositäpsuse võrdlust tunniintervalliga andmetel ning mitme ajasageduse kombineerimisel Joonisel 11.



Joonis 11. Dow Jones tunni ja mitme intervalliga mudelite prognoositulemuste võrdlus.

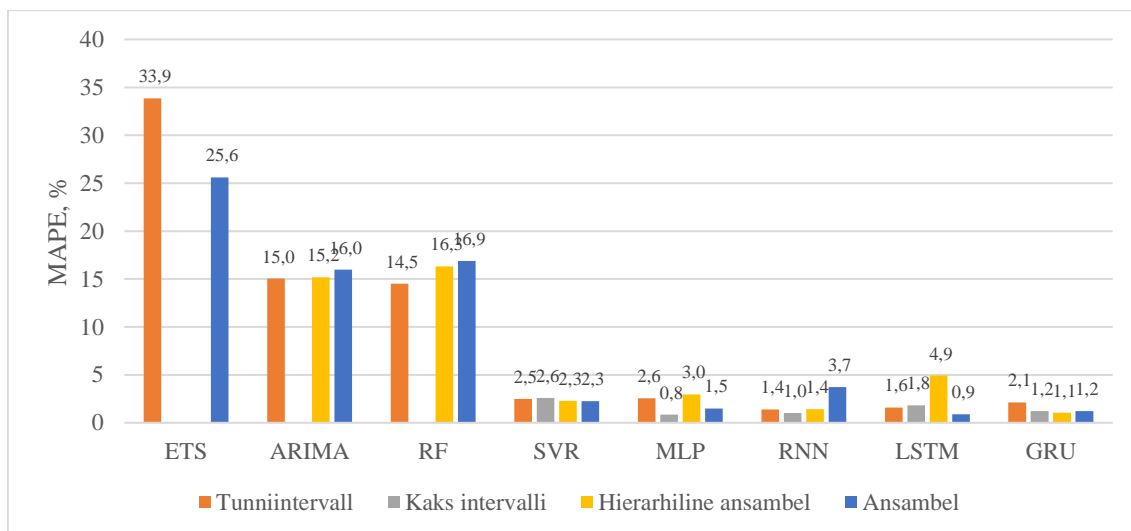
Dow Jones indeksi puhul on kahe intervalliga mudelite puhul samuti näha, et prognoosi täpsus ei parane oluliselt võrreldes sellega, kui kasutusel on vaid tunnisagedusega andmed. See kehtib nii SVR, RNN, LSTM kui ka GRU puhul. Ainus erand on MLP, mille puhul tunniintervalliga andmete kasutamisel andis mudel tulemuseks keskmise suhtelise absoluutvea 1,8% ning teise ajasageduse lisamisel on viga oluliselt väiksem – 1,2%.

Hierarhilise ansambli puhul on aga märgata prognoositäpsuse paranemist nii MLP kui ka SVR puhul. MLP meetodiga annab hierarhiline ansambel veelgi täpsema tulemuse kui eelneva meetodiga – MAPE on 0,7%. SVR puhul on esialgse tunniintervalliga andmetel loodud mudeli keskmine suhteline absoluutviga 0,5% ning hierarhilise ansambliga paraneb see 0,4%-ni. Kõigi korduvate närvivõrgu mudelite RNN, LSTM ja GRU puhul ei anna hierarhilise ansambelõppe kasutamine oluliselt parema täpsusega prognoose, kuigi RNN puhul erineb see tunniintervalliga mudeli täpsusest vaid 0,1% võrra, seega jääb prognoositäpsus samale tasemele. ARIMA ja RF puhul ei anna hierarhiline ansambelõppe samuti suurt erinevust, vähendades vaid RF puhul MAPE väärtust 9%-lt 8,8%-ni.

Viimasena rakendatud lihtne ansambelõppe keskmistamise kaudu töötab väga edukalt just statistiliste meetodite puhul, parandades prognoositäpsust ETS puhul 8%-lt 6,8%-ni ja ARIMA puhul 9,8%-lt 8,6%-ni. Ka RF puhul annab ansambelõppe kasutamine täpsemad prognoosid, mille MAPE on 7,9%. Ülejäänud algoritmidega ei tööta ansambelõppe sellisel kujul nii edukalt. Nii SVR, RNN, LSTM kui ka GRU puhul on prognooside täpsus tegelikult halvem kui tunnisagedusega mudelitel, kuigi viimase kahe puhul erineb vea suurus kõigest 0,1% võrra. MLP meetodi puhul annab ansambelõppe siiski täpsemad

prognoosid kui tunnisagedusega mudel – esialgne MAPE on 1,8% ning ansambelõppega on MAPE 1,1%, mis jääb küll hierarhilisest ansambelõppest ebatäpsemaks.

Kolmanda finantsvara Nasdaq 100 erinevate meetodite prognoositäpsuse võrdlus on paigutatud Joonisele 12.



Joonis 12. Nasdaq 100 tunni ja mitme intervalliga mudelite prognoositulemuste võrdlus.

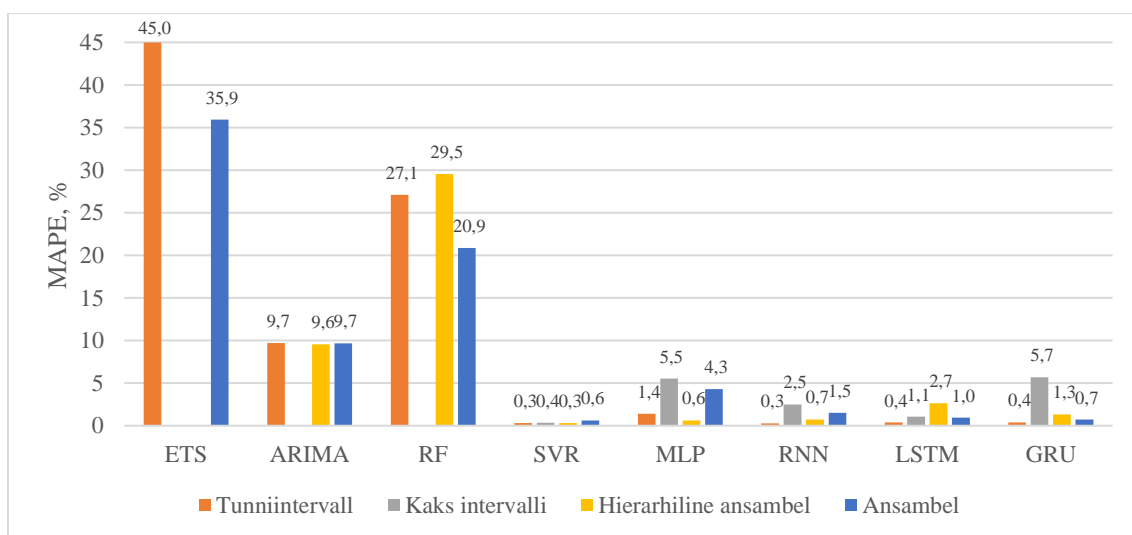
Nasdaq 100 indeksi puhul on märgata, et mitmel juhul parandab prognooside täpsust just teise ajaintervalli lisamine mudelisse (tähistatud halli värviga). Kui SVR ja LSTM puhul ei ole prognoosivea osas erilist muutust ja kaks intervalli korraga muudavad MAPE pigem suuremaks, siis teiste puhul paraneb MAPE märgatavalt. MLP puhul on tunniintervalliga mudeli MAPE 2,6% ning kahe intervalliga paraneb see 0,8%-ni. Suur on erinevus ka GRU puhul, kus MAPE liigub 2,1% pealt 1,2% peale. Ka RNN puhul on märgata prognoosivõime paranemist, kuna esialgne MAPE on 1,4% ning kahe intervalli kombineerimine annab tulemuseks 1,0%.

Hierarhilise ansambli puhul on näha, et nii ARIMA kui ka RF puhul ei anna meetod tulemuseks väiksemaid prognoosivigu. Sarnane on olukord ka MLP ja LSTM puhul, kuna mõlema algoritmi kasutamisel on tulemuseks suurem MAPE väärtus võrreldes tunniintervalliga mudelitega. RNN puhul aga jääb MAPE hierarhilise ansambelõppe korral täpselt samasuguseks kui vaid tunnisagedusega andmeid kasutades. Samas on aga märgata hierarhilise ansambelõppe edu GRU ja SVR puhul. Esimese puhul on hierarhilise ansambelõppe kasutamine kõige täpsema prognoosivõimega, sest MAPE on vaid 1,1%.

SVR puhul paraneb MAPE 2,5% pealt ainult tunnisagedusega andmeid kasutades 2,3% peale, seega näitab see väikest edumaad.

Viimasena on vaatluse all ansambelõpe keskmistamise abil, mis näitab märgatavalt paremat tulemust eksponentsiaalse silumise puhul, kus MAPE muutub 33,9% pealt 25,6% peale. Samas annab lihtsa ansambelõppe rakendamine mitmel juhul ka kehvema tulemuse võrreldes eelnevatega – seda nii ARIMA, RF kui ka RNN puhul. Teiste meetodite puhul on aga märgata viimase tehnika olulist edumaad nii tunniintervalliga mudelite tulemuste ees kui ka mõne teise tehnika ees. SVR puhul jääb MAPE sarnasele tasemele hierarhilise ansambelõppega ehk 2,3%, kuid tegelikult edestab seda võrreldes teisi veamõõdikuid. MLP puhul annab ansambelõppe rakendamine tulemuseks MAPE 1,5%, mis on küll oluliselt madalam tunnisagedusega mudeli tulemusest 2,6%, kuid siiski pea kaks korda kõrgem võrreldes kahe intervalliga mudeliga. LSTM puhul on ansambelmeetodil leitud mudeli prognoosiviga kõige väiksem jäädes 0,9% juurde. GRU puhul jääb tulemus tunniintervalliga mudelist oluliselt väiksemaks kuid sarnaseks teiste tehnikatega, milles mitu ajasagedust ühendati.

Järgnevalt võrreldakse erinevate meetodite rakendamist mitme ajaintervalliga krüptovaluutade andmetel ja võrreldakse tulemusi tunnitäpsusega andmete prognoosidega. Kõigepealt on vaatluse all Bitcoin'i tulemused, mida näeb Joonisel 13.



Joonis 13. Bitcoin'i tunni ja mitme intervalliga mudelite prognoositulemuste võrdlus.

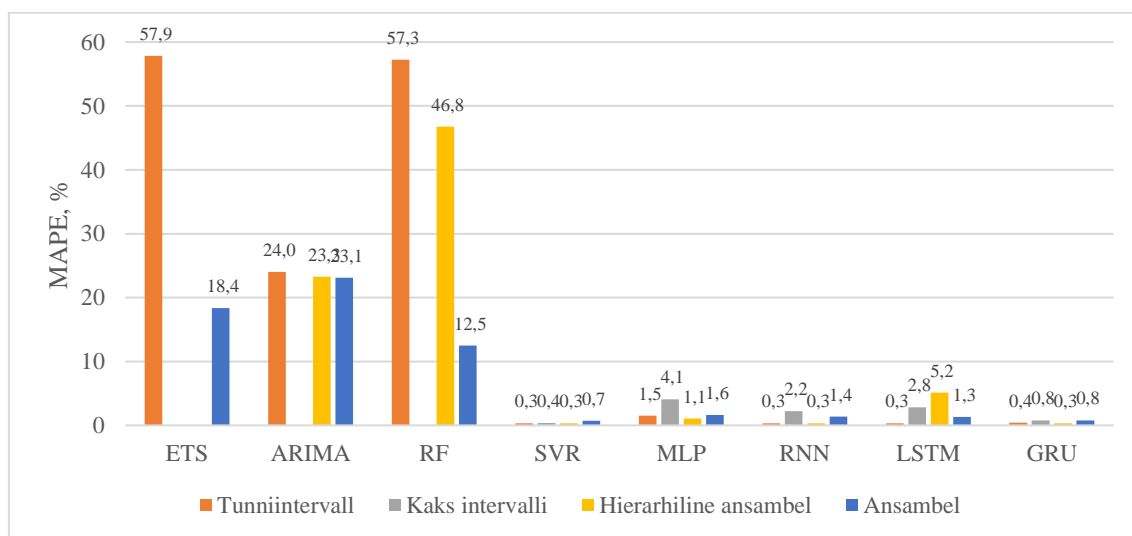
Vaadates Bitcoin'i puhul mitme ajaintervalli kasutust, mis tähistatud halliga, on selge, et see ei suuda tunni sagedusega mudelitest paremaid prognoose anda ja kõigil juhtudel on

veamõõdik suurem. Vaid SVR meetodi puhul on veamõõdiku erinevus vaevu märgatav, kuid siiski on see tunniintervalliga mudeli tulemustest 0,1% võrra suurem.

Hierarhilise ansambli rakendamisel ei ole tulemused nii erinevad tegelikkusest. ARIMA puhul jääb vea suurus sarnaseks tunnisagedusega mudeli tulemustega, samamoodi ka SVR puhul, kus MAPE ongi täpselt sama kui tunnisagedusega mudelil. Kuid RF puhul on hierarhilise ansambelmeetodi tulemus halvem. Sarnaselt on kehvem tulemus ka RNN ja selle edasiarenduste LSTM ja GRU puhul, kus hierarhiline ansambel annab ebatäpsemad prognoosid. MLP puhul aga toimib tehnika edukalt ja annab tulemuseks paremad prognoosid – MAPE on 0,6%.

Viimasena rakendatud lihtne ansambelõpe parandab oluliselt prognoosiviga eksponentsiaalse silumise ja otsustusmetsa meetodi puhul. Siiski on need vead oluliselt suuremad teiste algoritmide tulemustest. ARIMA puhul annab lihtne ansambelmeetod sarnase tulemuse tunnimudeliga, kuid teiste meetodite puhul jääb viimane meetod alla kas ainult tunniintervalliga andmetega leitud mudelite tulemustele või ka mõnele eelnevalt vaadatud mitme sagedusega meetodile.

Viimaseks analüüsitavaks finantsvaraks on Ethereum, mille puhul on erinevate meetodite prognoositulemused paigutatud Joonisele 14.



Joonis 14. Ethereumi tunni ja mitme intervalliga mudelite prognoositulemuste võrdlus.

Sarnaselt Bitcoin andmetel läbi viidud analüüsiga, selgub ka Ethereum tulemusi vaadates, et kaks intervalli korraga ei prognoosi tunni sagedusega andmeid eriti täpselt.

Seetõttu on otstarbekas eelistada tavalisi tunniintervalliga andmetega tehtud mudelite prognoose, mille prognoosivead on madalamad.

Hierarhilise ansambelmeetodi puhul on olukord teatud osas ikkagi parem. Nii ARIMA kui ka RF mudelite puhul annab hierarhilise ansambelõppega kombineeritud mudel suurema täpsusega prognoosid. ARIMA puhul on MAPE küll vähesel määral parem tunniintervalliga mudelist – vastavalt 23,3% võrreldes 24,0%-ga, kuid RF puhul on erinevus märgatav – 46,8% võrreldes 57,3%-ga. Teistest meetoditest paistavad silma MLP ja GRU, mille puhul hierarhiline ansambel annab madalama MAPE ja seega suurema prognoosivõime. SVR ja RNN puhul on valitud veamõõdiku MAPE tulemus täpselt sama nagu tunniandmetega mudelil, kuid vaadates täpsemalt Lisades 5 ja 7 paiknevaid tabeleid, on märgata, et teiste veamõõdikute järgi on siiski tunnisagedusega andmetel tehtud prognoosid täpsemad ja pole vajadust täiendavat intervalli lisada. Kõige halvem on tulemus LSTM puhul, kuna hierarhilise ansambliga loodud mudeli veamõõdik on mitmekümnekordselt suurem kui tunnimudelil saadud tulemused – vastavalt 5,2% ja 0,3%.

Viimasena rakendatud lihtsat ansambelõpet sai rakendada ka eksponentsiaalse silumise puhul, kus see andis oluliselt väiksema veaprotsendiga ennustused, isegi paremad kui ARIMA meetodil leitud tulemused – MAPE langes 57,9% pealt 18,4% peale. ARIMA puhul on samuti märgata veidi madalamat veaprotsenti 23,1%, kuid see pole nii suur erinevus. RF puhul on muutus aga väga suur ning ansambelmeetodil leitud prognooside veaprotsent väheneb lausa 12,5%-ni. Ülejäänud algoritmide puhul jäävad ansambelõppe tulemused keskpäraseks ja ükski ei suuda anda täpsemaid prognoose kui lihtsalt tunniintervalliga mudelitega.

Eelnev analüüs näitas selgelt, et üldistatud valemit mitme ajasagedusega andmete rakendamiseks ühes mudelis ei ole. Küll aga on võimalik tulemuste abil teatud järeldused paika panna:

1. Traditsiooniliste meetodite puhul annab kahe ajasageduse üheaegne rakendamine palju lisandväärtust. Eksponentsiaalse silumise puhul on võimalik rakendada vaid ansambelmeetodit läbi keskmistamise ning see töötab väga edukalt vähendades prognoosiviga ka üle 50% (nt Ethereum'i puhul). ARIMA meetodi puhul ei olnud

erinevused küll nii suured, kuid võrreldes hierarhilise ansambelõppega andis paremad tulemused ansambelmeetod läbi keskmistamise.

2. Otsustusmetsa puhul on teise ajaintervalli lisamine analüüsi andnud pigem paremad prognoositulemused, kuigi on ka erandeid. Meetodi osas on eelistatud samuti lihtsal kujul ansambelõpe läbi keskmistamise, mis käesolevas töös oli edukam.
3. Tugivektorregressiooni kasutamine näitas väga täpseid tulemusi juba ainuüksi madala ja kõrge ajasagedusega andmetel. Tegelikult võib kokkuvõtlikult järeldada, et tugivektorregressioon ongi vaid ühe ajaintervalliga andmetel kasutatuna niivõrd täpne, et täiendava sageduse lisamine ei muuda seda täpsemaks. Seetõttu on see esimene soovituslik meetod, mille poole pöörduda finantsvarade aegride prognoosihuvi korral. Tegu on meetodiga, mis ei vaja mudeli ehitamise faasis palju lisatööd andmete ettevalmistamiseks ja hüperparameetrite paika panemiseks ning mudelite jooksmisele kulub mõistlikult aega ka siis, kui tegu on väga kõrge sagedusega andmetega, mis muudavad aegread mitmekümne tuhande pikkuseks.
4. Mitmemõõtmeline närvivõrk oli sügavõppemeetoditest niiõelda kõige lihtsam. Varasema kirjanduse ülevaatest jäi see silma kui hea meetod aegride prognoosimise jaoks. Siiski olid selle tulemused nõrgemad võrreldes korduvate närvivõrkudega. Mitme sageduse koostoimel aga tulemused paranesid. Soovides rakendada just MLP mudelit aegride prognoosimisel tuleb kasuks täiendava ajasageduse lisamine mudelisse ja seda võiks pigem teha hierarhilise ansambelmeetodiga, kuigi edu näitas ka tavaline ansambelõpe läbi keskmistamise.
5. Nii rekurrentne närvivõrk kui ka selle edasiarendused pikk lühiajaline mälu ja väravatega rekurrentne üksus suudavad üldiselt vaid ühe ajasageduse kasutamisel võtta väga edukalt arvesse aegrea mustreid ja jäädvustada ajalisi sõltuvusi. Seetõttu on need aegride prognoosimisel ka nii palju tuntust kogunud. Käesolevas töös läbiviidud analüüsid näitasid samuti selgelt, et üldiselt ei too täiendava ajasageduse lisamine selliste korduvate närvivõrkude puhul mudelisse lisandväärtust ja prognoosid ei parane oluliselt. Üksikutest eranditest oli eespool juba juttu ja neid leidis nii mitme intervalli koostoimel, hierarhilise ansambelõppe puhul kui ka keskmistamise abil, kuid suures plaanis suudavad korduvad närvivõrgud püüda ühe sagedusega andmetest vajalikud pikaajalised sõltuvused, nende põhjal õppida ja seejärel kindlustada väga täpsed prognoosid. Lisaks saab nende meetodite puhul prognoositäpsust veelgi parandada läbi täiendava hüperparameetrite häälestamise.

6 Järeldused ja edasine arutelu

Käesolev magistritöö andis esmase ülevaate sellest, kuidas rakendada finantsandmete aegridade prognoosimisel samaaegselt mitme ajaintervalliga andmeid kasutades mitmesuguseid aegridade meetodeid, sealhulgas nii traditsioonilisi statistilisi mudeleid kui ka masinõppemeetodeid ja edasiarenenud sügavõppe algoritme. Peamised järeldused töö tulemustest said kirja eelmises peatükis, kuid kokkuvõtlikult saab väita, et mitme ajaintervalli kasutamine on mõistlik statistiliste meetodite ja ka lihtsama mitmemõõtmelise närvivõrgu puhul. Edasiarenenud korduvate närvivõrgu arhitektuurides suudetakse juba suurel määral võtta arvesse ajaliste andmete sõltuvusi pikaajaliselt, mistõttu prognoositäpsus on väga hea ka vaid ühe ajasageduse kasutamisel. Üllatavalt edukaks osutus ka tugivektorregressioon, mis andis võrdväärselt häid tulemusi nii ühe kui ka mitme ajaintervalli kasutamisel ning tulemused olid sarnased sügavõppemeetoditega.

Töö peamises fookuses olid just sellised statistilised, masinõppe ja sügavõppe meetodid, mis on vastavalt varasemale kirjandusele tõestanud oma otstarbekust just aegridade modelleerimisel ja on laialdast kasutust leidnud just volatiilsete andmestike puhul nagu finantsaegread. Töö tulemuseks on lai ülevaade erinevate meetodite prognoosivõimest kahel ajaintervallil aga ka nende kombineerimisel mitmesuguste meetoditega. Töö maht ja sellele kulutatud aeg võimaldas teemasse süveneda vaid teatud piirideni, kuid mahukama ja pikaajalisema projekti puhul oleks võimalik teemasse palju sügavamalt sisse vaadata. Kindlasti ei ole kaetud kõikvõimalikud meetodid ja algoritmid, mistõttu saaks teemat veelgi laiemalt uurida täiendades valitud meetodite hulka või arendades välja mitmesuguse keerukusastmega kombineeritud hübriidmudeleid, mis lõikavad kasu nii mitmest ajasagedusest kui ka statistilistest lineaarsetest ning sügavõppe mittelineaarsetest mudelitest üheaegselt.

Mitme ajaintervalli kombineerimisel oli töö eesmärk prognoosida tunni intervalliga andmeid kasutades abivahendina päevase resolutsiooniga andmeid. Tegelikult oleks võimalik teha analüüs ka vastupidi ning prognoosida päevaseid või veel madalama

sagedusega andmeid, näiteks pikaajalise majandusliku prognoosihuvi korral. Sellises analüüsis oleks huvitav kasutada pikemate ajaintervallide prognoosimisel täiendava faktorina kõrgema sagedusega andmete alusel leitud hinnamustreid ja volatiilsuse näitajaid. Seega oleks üks võimalik töö edasiarendus muuta prognoosiaken pikaajaliseks ning kasutada lühiajalisi andmeid pikaajaliste prognooside täpsustamiseks.

Kui käesoleva töö fookuses oli just hindade prognoosimine kasutades selle jaoks regressioonmudeleid, siis rakendusliku poole pealt võiks olla mõistlik läheneda probleemile ka kui klassifitseerimisülesandele. Sellisel juhul oleks analüüsi eesmärk tuvastada hinnaliikumise suunda – kas siis kahe või kolme klassina ehk positiivne, negatiivne ja neutraalne. Klassifitseerimise puhul oleksid analüüsi tulemused tõenäoliselt lihtsamini rakendatavad kas äriotsuste vastuvõtmisel, investeerimisel või kauplemisel.

Tegelikult on võimalus muuta analüüs majanduslikult lihtsamini rakendatavaks ja muuta fookus hindade prognoosimise asemel hoopis tootluste prognoosimiseks. Sellisel juhul annaks analüüs konkreetsemat infot huvilistele, kes soovivad analüüsi tulemusi lihtsa vaevaga praktiliselt kasutusse võtta ja nende abil majanduslikku kasu lõigata.

Lisaks on oluline märkida, et käesoleva magistr töö andmestikeks valiti justnimelt ühemõõtmelised aktsiaindeksite ja krüptovaluutade sulgemishinnad. Töö eesmärk oli nende andmete peal analüüsida erinevate meetodite prognoosivõimet ühemõõtmeliste aegriidade puhul. See näitab aga, et fookuses on olnud vaid üks aspekt finantsvarade analüüsist. Tuleb silmas pidada, et finantsandmetele avaldavad mõju mitmed tegurid – nii sisesed nagu ettevõtete finantsaruanded või ka finantsvarasid puudutavad suhtarvud, aga ka välised nagu makromajanduslikud tegurid või hoopis meediategevus. Nende lisamine muudaks täiendava analüüsi mitmemõõtmeliseks ja annaks kindlasti palju infot juurde. Käesolevas töös olid need töö mahtu silmas pidades sihilikult välja jäetud, et töö suund ei muutuks liiga laiaks ja vähem hoomatavaks.

Tehtud analüüs võib aga anda võimaluse valida välja parimaid tulemusi pakkunud meetodid ning kombineerida neid teistlaadi finantsturgu mõjutavate teguritega järgnevates analüüsides. Sealhulgas on paljud finantsvarade hindasid mõjutavad tegurid väga erineva ajaintervalliga, näiteks kvartaalsetest aruannetest pärinevad ettevõtte majandusandmed, finantsvarade suhtarvud, makromajanduslikud näitajad jne. Kuna käesolev magistr töö tegeles erineva ajasagedusega andmete kombineerimisega, saaks

sarnaseid tehnikaid kasutada ka selliste tegurite analüüsi lisamisel. Mõned võimalused edasisteks analüüsideks on järgmised:

- Sentimendianalüüs näiteks sotsiaalmeedia, uudiste, turundustegevuste, geopoliitiliste sündmuste andmete põhjal, mida saaks lisada kasutades loomuliku keele töötamise vahendeid;
- Makromajanduslike indikaatorite lisamine analüüsi;
- Tehniliste indikaatoritega analüüsi täiendamine;
- Fundamentaalse andmete lisamine analüüsi jne.

Kokkuvõtlikult võib öelda, et käesolev magistritöö on oluline samm finantsprognoosimise valdkonnas, tuues kaasa uusi teadmisi mitme ajaintervalli kasutamise mõjust ning erinevate prognoosimismeetodite sobivusest finantsvarade hinnaprognosimisel. Töö tulemused näitavad, et mitme ajaintervalli kasutamine võib parandada prognoosimistäpsust statistiliste meetodite ja lihtsamate sügavõppe mudelite puhul, samal ajal kui edasijõudnud sügavõppemeetodite puhul võib see lisandväärtus olla piiratud. Tugivektorregressioon osutus ootamatult edukaks nii ühe kui ka mitme ajaintervalli kasutamisel, pakkudes tulemusi, mis on võrreldavad sügavõppemeetoditega. Käesoleva uurimuse tulemused annavad väärtuslikku teavet finantsprognoosimise meetodite kohta ja pakuvad alust edasisteks uurimissuundadeks, mis võivad kaasa tuua täpsemad ja rakenduslikult olulisemad prognoosimisstrateegiad.

7 Kokkuvõte

Käesoleva magistritöö eesmärk oli edendada finantsvarade sulgemishindade aegridade prognoosimist, kasutades mitme ajasagedusega andmeid. Analüüsiks valiti erinevad meetodid, sealhulgas statistilised, masinõppe- ja sügavõppemeetodid. Andmestike hulka kuulusid populaarsed aktsiaindeksid S&P 500, Dow Jones ja Nasdaq 100 ning kõige enam tuntud krüptovaluutad Bitcoin ja Ethereum. Analüüs hõlmas prognoose madala ja kõrge sagedusega andmetel ning kahe ajaintervalli kombineerimist.

Püstitatud uurimisküsimused said analüüsi käigus vastatud ja võib kokkuvõtlikult esitada:

1. Statistilised meetodid on võrreldes sügavõppemeetoditega palju vähem tõhusad.
2. Mitme ajaintervalli kasutamine parandas ARIMA, eksponentsiaalse silumise, otsustusmetsa ja mitmekihilise närvivõrgu prognoosivõimet. Keerukamate, korduvate närvivõrgu arhitektuuride ja tugivektorregressiooni puhul ei aidanud see kaasa.
3. Prognoosi täpsus täiendava ajaintervalli kasutamisel jäi edasijõudnud, keerukamatel meetoditel sarnaseks või halvemaks. Statistiliste meetodite puhul parandati prognoosivõimet kuni 50%.
4. Ühtset kõige täpsemat meetodit ei selgunud, kuid kõige täpsemaid tulemusi andsid tugivektorregressioon ja sügavõppemeetoditest väravatega rekurrentne üksus. Esimene neist on lihtsam ja kiirem, samas kui teine on efektiivne pikaajaliste sõltuvuste arvestamisel ja infokogumisel, kuid on teiste korduvate närvivõrkudega võrreldes kiirem.

Kokkuvõttes annab see uurimus olulise panuse finantsprognoosimise valdkonnas, andes ülevaate mitmesuguste meetodite sobivusest selle ülesande täitmisel ja avades uksi edasistele uurimissuundadele.

Kasutatud kirjandus

- [1] Y. Tang, Z. Song, Y. Zhu, H. Yuan, M. Hou, J. Ji, C. Tang ja J. Li, „A survey on machine learning models for financial time series forecasting,“ *Neurocomputing*, kd. 512, pp. 363-380, 2022.
- [2] D. A. Macias, „A Brief History on Time Series Analysis & Forecasting,“ 14 August 2022. [Võrgumaterjal]. Available: <https://medium.com/@deniseamacias1/a-brief-history-on-time-series-analysis-forecasting-f5a22bbd0641>. [Kasutatud 13 January 2024].
- [3] E. F. Fama, „Efficient Capital Markets: A Review of Theory and Empirical Work,“ *The Journal of Finance*, kd. 25, nr 2, pp. 383-417, 1970.
- [4] G. E. P. Box ja G. M. Jenkins, *Time series analysis: forecasting and control*, San Francisco, USA: Holden-Day, 1970.
- [5] Z. Chen ja B. Liu, *Lifelong Machine Learning*, 2nd edition, Morgan & Claypool, 2018.
- [6] S. Hochreiter ja J. Schmidhuber, „Long short-term memory,“ *Neural Computation*, kd. 9, nr 8, pp. 1735-1780, 1997.
- [7] V. I. Kontopoulou, A. D. Panagoupoulos ja I. M. G. K. Kakkos, „A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks,“ *Future Internet*, kd. 15, nr 8, p. 255, 2023.
- [8] D. Makala ja Z. Li, „Prediction of gold price with ARIMA and SVM,“ *Journal of Physics: Conference Series*, kd. 1767, nr 012022, 2022.
- [9] I. Paliari, A. Karanikola ja S. Kotsiantis, „A comparison of the optimized LSTM, XGBOOST and ARIMA in Time Series forecasting,“ %1 *2021 12th International Conference on Information, Intelligence, Systems & Applications*, Chania, Crete, Greece, 2021.
- [10] S. Siami-Namini, N. Tavakoli ja A. S. Namin, „Cornell University arxiv,“ 21 November 2019. [Võrgumaterjal]. Available: <https://arxiv.org/abs/1911.09512>. [Kasutatud 14 January 2024].
- [11] V. Derbentsev, A. Matviychuk, N. Datsenko, V. Bezkorovainyi ja A. Azaryan, „Machine learning approaches for financial time series forecasting,“ %1 *Machine Learning for Prediction of Emergent Economy Dynamics 2020*, Odessa, Ukraine, 2020.
- [12] D.-T. Nguyen ja H.-V. Le, „Predicting the price of bitcoin using hybrid ARIMA and machine learning,“ %1 *In Proceedings of the Future Data and Security Engineering - 6th International Conference*, Nha Trang City, Vietnam, 2019.
- [13] P. T. Yamak, L. Yujian ja P. K. Gadosey, „A comparison between arima, lstm, and gru for time series forecasting,“ %1 *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*, Sanya, China, 2019.

- [14] Y. Hua, „Bitcoin price prediction using ARIMA and LSTM,“ %1 2020 *International Symposium on Energy, Environmental Science and Engineering (ISEESE 2020)*, 2020.
- [15] N. Latif, J. Selvam, M. Kapse, V. Sharma ja V. Mahajan, „Comparative Performance of LSTM and ARIMA for the Short-Term,“ *Australasian Accounting Business and Finance Journal*, kd. 17, nr 1, pp. 256-276, 2023.
- [16] Z. Chen, C. Li ja W. Sun, „Bitcoin price prediction using machine learning: An approach to sample dimension engineering,“ *Journal of Computational and Applied Mathematics*, kd. 365, 2020.
- [17] G. P. Zhang, „Time series forecasting using a hybrid ARIMA and neural network model,“ *Neurocomputing*, kd. 50, pp. 159-175, 2003.
- [18] Z. Lanbouri ja S. Achhab, „A new approach for trading based on long short-term memory technique,“ *International Journal of Computer Science Issues*, 2019.
- [19] L. Zhang, C. Aggarwal ja G.-J. Qi, „Stock Price Prediction via Discovering Multi-Frequency Trading Patterns,“ %1 *KDD '17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017.
- [20] N. Kourentzes, F. Petropoulos ja J. R. Trapero, „Improving forecasting by estimating time series structural components across multiple frequencies,“ *International Journal of Forecasting*, kd. 3, nr 2, pp. 291-302, 2014.
- [21] S. Makridakis ja M. Hibon, „The M3-Competition: results, conclusions and implications,“ *International Journal of Forecasting*, kd. 16, pp. 451-476, 2000.
- [22] „Federal reserve economic data (FRED),“ 2012. [Vörgumaterjal]. Available: <https://fred.stlouisfed.org/>. [Kasutatud 27 January 2024].
- [23] R. J. Hyndman ja G. Athanasopoulos, „Forecasting: Principles and Practice, ARIMA models,“ April 2018. [Vörgumaterjal]. Available: <https://otexts.com/fpp2/arima.html>. [Kasutatud 29 February 2024].
- [24] D. Abugaber, „Chapter 23: Using ARIMA for Time Series Analysis,“ [Vörgumaterjal]. Available: <https://ademos.people.uic.edu/Chapter23.html>. [Kasutatud 01 05 2024].
- [25] R. J. Hyndman ja G. Athanasopoulos, „Forecasting: Principles and Practice, Exponential smoothing,“ April 2018. [Vörgumaterjal]. Available: <https://otexts.com/fpp2/expsmooth.html>. [Kasutatud 29 February 2024].
- [26] B. Goehry, H. Yan, Y. Goude, P. Massart ja J.-M. Poggi, „Random Forests for Time Series,“ *REVSTAT-Statistical Journal*, kd. 21, nr 2, 2023.
- [27] S. Haines, „Reading Tea Leaves: Principles of Predictive Analytics and the Path to Time-Series Predictions,“ ODSC, 31 January 2019. [Vörgumaterjal]. Available: <https://opendatascience.com/principles-of-predictive-analytics-and-the-path-to-time-series-prediction/>. [Kasutatud 14 April 2024].
- [28] H. Muthiah, U. Sa'adah ja A. Efendi, „Support Vector Regression (SVR) Model for Seasonal Time Series Data,“ %1 *Proceedings of the Second Asia Pacific International Conference on Industrial Engineering and Operations Management*, Surakarta, Indonesia, 2021.
- [29] Y. Zhang, Q. Wang, X. Chen, Y. Yan, R. Yang, Z. Liu ja J. Fu, „The Prediction of Spark-Ignition Engine Performance and Emissions Based on the SVR Algorithm,“ *Processes*, kd. 10, nr 2, 2022.

- [30] M. Shiblee, P. K. Kalra ja B. Chandra, „Time Series Prediction with Multilayer Perceptron (MLP): A New Generalized Error Based Approach,“ %1 *Advances in Neuro-Information Processing*, Auckland, New Zealand, 2008.
- [31] H. A. Afan, A. I. A. Osman, Y. Essam, A. A. N. Ahmed, Y. F. Huang, O. Kisi, M. Sherif, A. Sefelnasr, K. W. Chau ja A. El-Shafie, „Modeling the fluctuations of groundwater level by employing ensemble deep learning techniques,“ *Engineering Applications of Computational Fluid Mechanics*, kd. 15, nr 1, pp. 1420-1439, 2021.
- [32] „What are recurrent neural networks?,“ IBM, [Võrgumaterjal]. Available: <https://www.ibm.com/topics/recurrent-neural-networks>. [Kasutatud 28 January 2024].
- [33] M. Nabil, „Recurrent Neural Network (RNN) Architecture Explained,“ 28 August 2023. [Võrgumaterjal]. Available: <https://medium.com/@poudelsushmita878/recurrent-neural-network-rnn-architecture-explained-1d69560541ef>. [Kasutatud 28 January 2024].
- [34] S. Hochreiter ja J. Schmidhuber, „Long short-term memory,“ *Neural Computation*, kd. 9, nr 8, pp. 1735-1780, 1997.
- [35] A. Shenfield ja M. Howarth, „Novel Deep Learning Model for the Detection and Identification of Rolling Element-Bearing Faults,“ *Sensors*, kd. 20, nr 18, 2020.
- [36] J. Chung, C. Gulcehre, K. Cho ja Y. Bengio, „Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling,“ 11 December 2014. [Võrgumaterjal]. Available: <https://doi.org/10.48550/arXiv.1412.3555>. [Kasutatud 30 January 2024].
- [37] P. Li, A. Luo, J. Liu, Y. Wang, J. Zhu, Y. Deng ja J. Zhang, „Bidirectional Gated Recurrent Unit Neural Network for Chinese Address Element Segmentation,“ *ISPRS International Journal of Geo-Information*, kd. 9, nr 11, 2020.
- [38] „Complete Intraday Bundle - Stock, Futures, ETF, Index, Crypto,“ FirstRate Data, [Võrgumaterjal]. Available: <https://firstratedata.com/cb/1/complete-us-stocks-index-etf-futures>. [Kasutatud 28 January 2024].
- [39] „finance.yahoo.com,“ [Võrgumaterjal]. Available: <https://finance.yahoo.com/quote/%5ESPX/history?period1=1199145600&period2=1706313600&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true>. [Kasutatud 28 January 2024].
- [40] C. M. Jarque ja A. K. Bera, „Efficient tests for normality, homoscedasticity and serial independence of regression residuals,“ *Economic Letters*, kd. 6, nr 3, pp. 255-259, 1980.
- [41] D. A. Dickey ja W. A. Fuller, „Distribution of the Estimators for Autoregressive Time Series With a Unit Root,“ *Journal of the American Statistical Association*, kd. 74, nr 366, pp. 427-431, 1979.
- [42] R. J. Hyndman ja G. Athanasopoulos, „Forecasting: Principles and Practice, 2nd edition - Stationarity and differencing,“ April 2018. [Võrgumaterjal]. Available: <https://otexts.com/fpp2/stationarity.html>. [Kasutatud 5 March 2024].
- [43] „sklearn.preprocessing.MinMaxScaler,“ [Võrgumaterjal]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>. [Kasutatud 31 March 2024].

- [44] M. Filho, „How to do time series cross-validation in Python,“ Forecastegy, 12 July 2023. [Võrgumaterjal]. Available: <https://forecastegy.com/posts/time-series-cross-validation-python/>. [Kasutatud 6 April 2024].
- [45] J. Murel ja E. Kavlakoglu, „What is ensemble learning?,“ IBM, 18 January 2024. [Võrgumaterjal]. Available: <https://www.ibm.com/topics/ensemble-learning>. [Kasutatud 20 April 2024].
- [46] I. Nti, A. Adekoya ja B. Weyori, „A comprehensive evaluation of ensemble,“ *Journal of Big Data*, kd. 7, nr 1, 2020.
- [47] K. Rink, „Time Series Forecast Error Metrics You Should Know,“ Towards Data Science, 21 October 2021. [Võrgumaterjal]. Available: <https://towardsdatascience.com/time-series-forecast-error-metrics-you-should-know-cc88b8c67f27>. [Kasutatud 10 March 2024].
- [48] D. Andres, „Error Metrics for Time Series Forecasting,“ 22 June 2023. [Võrgumaterjal]. Available: <https://ml-pills.dev/time-series/error-metrics-for-time-series-forecasting/>. [Kasutatud 1 April 2024].
- [49] „Exponential Smoothing: Formula, Method, and Solved Examples,“ testbook, 16 June 2023. [Võrgumaterjal]. Available: <https://testbook.com/maths/exponential-smoothing>. [Kasutatud 20 April 2024].
- [50] „Tuning the hyper-parameters of an estimator,“ scikit-learn, [Võrgumaterjal]. Available: https://scikit-learn.org/stable/modules/grid_search.html. [Kasutatud 10 April 2024].
- [51] „sklearn.model_selection.GridSearchCV,“ scikit-learn.org, [Võrgumaterjal]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html. [Kasutatud 10 April 2024].
- [52] „Quick start skforecast,“ skforecast.org, [Võrgumaterjal]. Available: https://skforecast.org/0.8.1/user_guides/quick-start-skforecast#backtesting-forecaster-validation. [Kasutatud 10 April 2024].
- [53] „RandomSearch Tuner,“ keras.io, [Võrgumaterjal]. Available: https://keras.io/api/keras_tuner/tuners/random/. [Kasutatud 10 April 2024].
- [54] „KerasTuner,“ Keras, [Võrgumaterjal]. Available: https://keras.io/keras_tuner/. [Kasutatud 10 April 2024].
- [55] N. Buhl, „Activation Function in Neural Networks: With 15 examples,“ Encord, 25 July 2023. [Võrgumaterjal]. Available: <https://encord.com/blog/activation-functions-neural-networks/>. [Kasutatud 6 April 2024].
- [56] H. M. Sami, K. A. Ahshan ja P. N. Rozario, „Determining the Best Activation Functions for Predicting Stock Prices,“ *Australian Journal of Engineering and Innovative Technology*, kd. 5, nr 2, pp. 63-71, 2023.
- [57] „Optimizers,“ ML Glossary, 2017. [Võrgumaterjal]. Available: <https://ml-cheatsheet.readthedocs.io/en/latest/optimizers.html>. [Kasutatud 6 April 2024].
- [58] C. Mendoza, W. Kristjanpoller ja M. Minutolo, „Market index price prediction using Deep Neural Networks with a Self-Similarity approach,“ *Applied Soft Computing*, kd. 146, 2023.
- [59] M. Samin-Al-Wasee, P. Kundu, I. Mahzabeen, T. Tamim ja G. Dr. Alam, „Time-Series Forecasting of Ethereum Price Using Long Short-Term Memory (LSTM) Networks,“ %1 *Proc. of the 8th International Conference on Engineering and Emerging Technologies (ICEET)*, Kuala Lumpur, Malaysia, 2022.

- [60] N. Latif, J. Selvam, M. Kapse, V. Sharma ja V. Mahajan, „Comparative Performance of LSTM and ARIMA for the Short-Term Prediction of Bitcoin Prices,“ *Australasian Accounting Business and Finance Journal*, kd. 17, nr 1, pp. 256-276, 2023.

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

Mina, Kadri-Liis Kutser

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Mitme sagedusega finantsandmete aegrea prognoosimise meetodite võrdlusanalüüs“, mille juhendaja on Olga Dunajeva
 - 1.1. reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

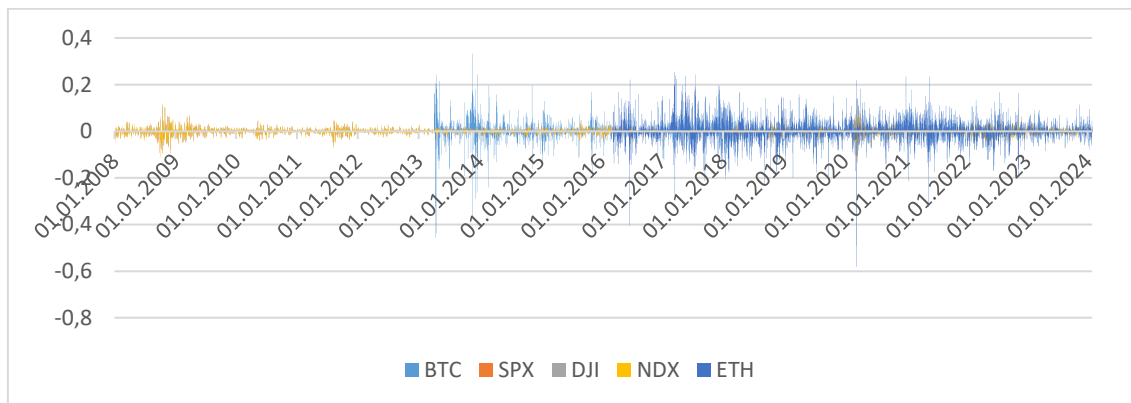
08.05.2024

¹ Lihtlitsents ei kehti juurdepääsupiirangu kehtivuse ajal vastavalt üliõpilase taotlusele lõputööle juurdepääsupiirangu kehtestamiseks, mis on allkirjastatud teaduskonna dekaani poolt, välja arvatud ülikooli õigus lõputööd reprodutseerida üksnes säilitamise eesmärgil. Kui lõputöö on loonud kaks või enam isikut oma ühise loomingu tegevusega ning lõputöö kaas- või ühisautor(id) ei ole andnud lõputööd kaitsvale üliõpilasele kindlaksmääratud tähtajaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktile 1.1. ja 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.

Lisa 2. Statistilised testid

Finantsvara	Test	Test statistik	Kriitiline väärtus	p-väärtus
S&P 500	Jarque-Bera	350,01	5,99	9,90E-77
	Dickey-Fuller	0,80	-2,86	0,99
Dow Jones	Jarque-Bera	325,14	5,99	2,49E-71
	Dickey-Fuller	0,21	-2,86	0,97
Nasdaq 100	Jarque-Bera	564,66	5,99	2,43E-123
	Dickey-Fuller	1,38	-2,86	0,99
Bitcoin	Jarque-Bera	1299,26	5,99	7,40E-283
	Dickey-Fuller	-1,35	-2,86	0,6
Ethereum	Jarque-Bera	719,39	5,99	6,10E-157
	Dickey-Fuller	-1,65	-2,86	0,46

Lisa 3. Teisendatud andmestik



Lisa 4. Madala sagedusega prognoosi veamõõdikud finantsvara ja meetodi lõikes

Veamõõdikud						
Finantsvara	Meetod	MAE	MSE	RMSE	MAPE	SMAPE
S&P500	ETS	345,7	195 867	442,6	8,0	8,5
	ARIMA	346,9	194 961	441,5	8,0	8,5
	RF	333,3	172 893	415,8	7,8	8,2
	SVR	44,9	3 090	55,6	1,1	1,1
	MLP	88,4	10 305	101,5	2,1	2,1
	RNN	130,6	20 686	143,8	3,0	3,1
	LSTM	63,0	5 574	74,7	1,5	1,5
	GRU	55,2	4 498	67,1	1,3	1,3
Dow Jones	ETS	2 388,9	9 247 874	3 041,0	7,0	7,4
	ARIMA	2 551,6	9 995 122	3 161,5	7,5	7,9
	RF	2 366,4	9 045 921	3 007,6	6,9	7,3
	SVR	539,2	393 672	627,4	1,6	1,6
	MLP	265,3	116 234	340,9	0,8	0,8
	RNN	920,0	990 940	995,5	2,7	2,7
	LSTM	266,7	118 658	344,5	0,8	0,8
	GRU	270,2	123 870	352,0	0,8	0,8
Nasdaq 100	ETS	2 268,5	7 095 345	2 663,7	17,2	15,6
	ARIMA	2 219,4	6 805 597	2 608,8	16,8	15,3
	RF	2 548,3	9 245 624	3 040,7	19,3	17,1
	SVR	309,1	147 428	384,0	2,1	2,2
	MLP	197,5	66 013	256,9	1,5	1,5
	RNN	876,0	924 453	961,5	6,1	6,3
	LSTM	161,7	43 422	208,4	1,2	1,2

	GRU	151,4	39 319	198,3	1,1	1,1
Bitcoin	ETS	127 406,7	21 368 935 784	146 181,2	472,9	126,6
	ARIMA	74 330,7	6 693 909 553	81 816,3	281,1	107,0
	RF	82 764,6	8 510 232 298	92 250,9	311,0	110,8
	SVR	605,7	851 928	923,0	2,0	2,0
	MLP	2 029,7	5 146 224	2 268,5	7,2	6,9
	RNN	1 048,5	1 622 387	1 273,7	3,9	3,8
	LSTM	682,6	932 392	965,6	2,4	2,4
	GRU	601,1	806 110	897,8	2,0	2,0
Ethereum	ETS	502,7	397 806	630,7	29,8	25,5
	ARIMA	368,0	228 015	477,5	21,0	19,1
	RF	295,6	125 556	354,3	16,4	18,5
	SVR	37,6	3 096	55,6	2,2	2,2
	MLP	44,8	3 693	60,8	2,7	2,6
	RNN	59,6	5 354	73,2	3,7	3,6
	LSTM	52,2	4 443	66,7	3,1	3,0
	GRU	38,1	3 151	56,1	2,3	2,3

Lisa 5. Kõrge sagedusega prognoosi veamõõdikud finantsvara ja meetodi lõikes

Veamõõdikud						
Finantsvara	Meetod	MAE	MSE	RMSE	MAPE	SMAPE
S&P500	ETS	419,6	210 292	458,6	10,0	10,0
	ARIMA	356,0	204 532	452,3	8,2	8,7
	RF	340,4	185 253	430,4	7,9	8,4
	SVR	16,0	461	21,5	0,4	0,4
	MLP	62,3	4 536	67,4	1,5	1,5
	RNN	31,5	1 294	36,0	0,7	0,7
	LSTM	49,6	2 743	52,4	1,2	1,2
	GRU	13,6	328	18,1	0,3	0,3
Dow Jones	ETS	2 705,8	9 698 212	3 114,2	8,0	8,3
	ARIMA	3 339,5	14 871 728	3 856,4	9,8	10,4
	RF	3 077,1	13 401 633	3 660,8	9,0	9,6
	SVR	163,4	38 092	195,2	0,5	0,5
	MLP	592,9	385 532	620,9	1,8	1,8
	RNN	217,2	59 474	243,9	0,6	0,6
	LSTM	142,2	29 106	170,6	0,4	0,4
	GRU	132,4	27 223	165,0	0,4	0,4
Nasdaq 100	ETS	4 540,8	33 316 643	5 772,1	33,9	26,8
	ARIMA	1 990,6	5 166 248	2 272,9	15,0	14,2
	RF	1 920,6	4 832 199	2 198,2	14,5	13,8
	SVR	368,5	216 178	465,0	2,5	2,5
	MLP	374,1	234 813	484,6	2,6	2,6
	RNN	192,6	43 643	208,9	1,4	1,4
	LSTM	219,0	52 387	228,9	1,6	1,6

	GRU	303,5	110 808	332,9	2,1	2,1
Bitcoin	ETS	14 909,4	321 638 540	17 934,3	45,0	63,6
	ARIMA	3 171,5	16 744 048	4 092,0	9,7	10,0
	RF	8 431,0	107 622 259	10 374,1	27,1	22,6
	SVR	96,5	24 143	155,4	0,3	0,3
	MLP	424,1	229 646	479,2	1,4	1,4
	RNN	90,4	22 573	150,2	0,3	0,3
	LSTM	114,4	28 290	168,2	0,4	0,4
	GRU	114,6	26 993	164,3	0,4	0,4
Ethereum	ETS	1 112,1	1 562 517	1 250,0	57,9	89,7
	ARIMA	440,2	287 678	536,4	24,0	20,4
	RF	1 100,4	1 521 757	1 233,6	57,3	88,0
	SVR	6,5	103	10,1	0,3	0,3
	MLP	28,3	1 006	31,7	1,5	1,5
	RNN	6,2	96	9,8	0,3	0,3
	LSTM	6,0	93	9,6	0,3	0,3
	GRU	7,5	152	12,3	0,4	0,4

Lisa 6. Mitme sagedusega prognoosi veamõõdikud finantsvara ja meetodi lõikes

Veamõõdikud						
Finantsvara	Meetod	MAE	MSE	RMSE	MAPE	SMAPE
S&P 500	SVR	35,1	1 775	42,1	0,8	0,8
	MLP	220,8	51 509	227,0	5,2	5,3
	RNN	9,7	244	15,6	0,2	0,2
	LSTM	151,2	25 542	159,8	3,5	3,5
	GRU	156,0	26 175	161,8	3,7	3,7
Dow Jones	SVR	454,8	263 283	513,1	1,3	1,3
	MLP	409,2	211 015	459,4	1,2	1,2
	RNN	256,6	76 339	276,3	0,8	0,8
	LSTM	849,3	806 607	898,1	2,5	2,5
	GRU	170,8	37 362	193,3	0,5	0,5
Nasdaq 100	SVR	381,7	233 945	483,7	2,6	2,6
	MLP	113,2	20 418	142,9	0,8	0,8
	RNN	140,4	23 549	153,5	1,0	1,0
	LSTM	254,2	71 756	267,9	1,8	1,8
	GRU	172,0	37 349	193,3	1,2	1,2
Bitcoin	SVR	106,3	24 326	156,0	0,4	0,4
	MLP	1 862,8	5 292 654	2 300,6	5,5	5,7
	RNN	707,3	565 687	752,1	2,5	2,6
	LSTM	352,5	218 778	467,7	1,1	1,1
	GRU	1 934,3	6 138 470	2 477,6	5,7	5,5
Ethereum	SVR	7,2	106	10,3	0,4	0,4
	MLP	78,9	11 256	106,1	4,1	4,1
	RNN	40,5	1 742	41,7	2,2	2,3

	LSTM	52,6	2 963	54,4	2,8	2,9
	GRU	14,5	359	19,0	0,8	0,8

Lisa 7. Mitme sagedusega prognoosi veamõõdikud finantsvara ja meetodi lõikes, kasutades hierarhilist ansambelõpet

Veamõõdikud						
Finantsvara	Meetod	MAE	MSE	RMSE	MAPE	SMAPE
S&P 500	ARIMA	352,7	201 067	448,4	8,1	8,7
	RF	354,9	202 208	449,7	8,2	8,7
	SVR	30,7	1 341	36,6	0,7	0,7
	MLP	29,4	1 382	37,2	0,7	0,7
	RNN	49,7	2 993	54,7	1,2	1,2
	LSTM	71,4	5 952	77,2	1,7	1,7
	GRU	11,5	272	16,5	0,3	0,3
Dow Jones	ARIMA	3 342,4	14 890 251	3 858,8	9,8	10,5
	RF	3 009,4	13 201 280	3 633,4	8,8	9,4
	SVR	141,4	31 196	176,6	0,4	0,4
	MLP	220,6	82 488	287,2	0,7	0,7
	RNN	245,1	77 391	278,2	0,7	0,7
	LSTM	480,9	262 888	512,7	1,4	1,4
	GRU	212,8	62 178	249,4	0,6	0,6
Nasdaq 100	ARIMA	2 011,3	5 186 976	2 277,5	15,2	14,3
	RF	2 163,3	5 890 597	2 427,1	16,3	15,3
	SVR	336,9	175 518	419,0	2,3	2,3
	MLP	410,7	229 185	478,7	3,0	3,0
	RNN	199,4	48 180	219,5	1,4	1,4
	LSTM	687,7	499 150	706,5	4,9	5,1
	GRU	142,6	26 753	163,6	1,1	1,1
Bitcoin	ARIMA	3 100,1	15 910 495	3 988,8	9,6	9,7
	RF	9 182,2	124 317 776	11 149,8	29,5	24,4

	SVR	101,3	25 769	160,5	0,3	0,3
	MLP	198,2	98 454	313,8	0,6	0,6
	RNN	227,1	77 226	277,9	0,7	0,7
	LSTM	807,9	744 252	862,7	2,7	2,6
	GRU	396,6	184 043	429,0	1,3	1,3
Ethereum	ARIMA	426,4	271 325	520,9	23,3	19,8
	RF	907,7	1 102 283	1 049,9	46,8	66,9
	SVR	6,5	104	10,2	0,3	0,3
	MLP	20,0	718	26,8	1,1	1,1
	RNN	6,4	101	10,1	0,3	0,3
	LSTM	96,2	9 450	97,2	5,2	5,0
	GRU	5,9	96	9,8	0,3	0,3

Lisa 8. Mitme sagedusega prognoosi veamõõdikud finantsvara ja meetodi lõikes, kasutades ansambelõpet

Veamõõdikud						
Finantsvara	Meetod	MAE	MSE	RMSE	MAPE	SMAPE
S&P500	ETS	336,0	166 084	407,5	7,9	8,2
	ARIMA	351,6	199 805	447,0	8,1	8,6
	RF	336,7	178 432	422,4	7,8	8,3
	SVR	26,6	1 075	32,8	0,6	0,6
	MLP	27,9	1 434	37,9	0,7	0,7
	RNN	80,1	7 695	87,7	1,9	1,9
	LSTM	19,2	669	25,9	0,5	0,5
	GRU	19,2	669	25,9	0,7	0,7
Dow Jones	ETS	2 316,0	8 563 531	2 926,4	6,8	7,2
	ARIMA	2 924,1	12 205 025	3 493,6	8,6	9,1
	RF	2 686,8	10 983 147	3 314,1	7,9	8,3
	SVR	336,5	149 925	387,2	1,0	1,0
	MLP	353,0	160 498	400,6	1,1	1,1
	RNN	561,8	363 629	603,0	1,7	1,7
	LSTM	174,8	45 890	214,2	0,5	0,5
	GRU	170,0	45 715	213,8	0,5	0,5
Nasdaq 100	ETS	3 411,3	17 341 340	4 164,3	25,6	21,6
	ARIMA	2 110,6	5 932 938	2 435,8	16,0	14,8
	RF	2 231,7	6 680 794	2 584,7	16,9	15,5
	SVR	331,0	172 699	415,6	2,3	2,3
	MLP	210,9	69 697	264,0	1,5	1,5
	RNN	532,1	333 444	577,5	3,7	3,8
	LSTM	120,8	21 925	148,1	0,9	0,9

	GRU	171,9	39 824	199,6	1,2	1,2
Bitcoin	ETS	11 954,6	208 701 129	14 446,5	35,9	46,8
	ARIMA	3 160,2	16 616 232	4 076,3	9,7	9,9
	RF	6 346,6	61 750 987	7 858,2	20,9	18,0
	SVR	191,4	84 181	290,1	0,6	0,6
	MLP	1 276,7	1 775 077	1 332,3	4,3	4,2
	RNN	441,6	246 629	496,6	1,5	1,5
	LSTM	283,1	124 785	353,3	1,0	0,9
	GRU	215,7	88 479	297,5	0,7	0,7
Ethereum	ETS	360,2	174 242	417,4	18,4	20,8
	ARIMA	423,6	270 187	519,8	23,1	19,7
	RF	230,5	75 813	275,3	12,5	11,9
	SVR	13,2	396	19,9	0,7	0,7
	MLP	29,8	1 241	35,2	1,6	1,6
	RNN	25,2	862	29,4	1,4	1,4
	LSTM	24,6	831	28,8	1,3	1,3
	GRU	14,6	489	22,1	0,8	0,8

Lisa 9. Ühe ajasagedusega mudelite lõplikud seadistused

Vara	Mudel	Päeva intervall	Tunniintervall
S&P500	ARIMA	ARIMA(1,0,0)	ARIMA(1,0,5)
	RF	n_estimators=500, max_depth=5, lags=20	n_estimators=100, max_depth=5, lags=80
	SVR	C: 100, epsilon: 0,01, gamma: auto	C: 100, epsilon: 0,01, gamma: auto
	MLP	units: 160, activation: relu, learning_rate: 0,01	units: 160, activation: relu, learning_rate: 0,001
	RNN	units: 192, activation: relu, learning_rate: 0,001	units: 192, activation: relu, learning_rate: 0,001
	LSTM	units: 128, activation: tanh, learning_rate: 0,01	units: 192, activation: relu, learning_rate: 0,001
	GRU	units: 192, activation: relu, learning_rate: 0,001	units: 160, activation: relu, learning_rate: 0,01
Dow Jones	ARIMA	ARIMA(1,0,0)	ARIMA(0,0,1)
	RF	n_estimators=500, max_depth=5, lags=20	n_estimators=500, max_depth=5, lags=80
	SVR	C: 100, epsilon: 0,01, gamma: auto	C: 100, epsilon: 0,01, gamma: auto
	MLP	units: 192, activation: relu, learning_rate: 0,001	units: 192, activation: relu, learning_rate: 0,001
	RNN	units: 192, activation: relu, learning_rate: 0,001	units: 160, activation: relu, learning_rate: 0,001
	LSTM	units: 128, activation: tanh, learning_rate: 0,01	units: 128, activation: tanh, learning_rate: 0,01
	GRU	units: 128, activation: tanh, learning_rate: 0,01	units: 160, activation: relu, learning_rate: 0,01
Nasdaq 100	ARIMA	ARIMA(0,0,1)	ARIMA(0,0,2)
	RF	n_estimators=500, max_depth=5, lags=20	n_estimators=500, max_depth=5, lags=80
	SVR	C: 10, epsilon: 0,01, gamma: auto	C: 1, epsilon: 0,01, gamma: auto
	MLP	units: 192, activation: relu, learning_rate: 0,001	units: 224, activation: tanh, learning_rate: 0,01
	RNN	units: 192, activation: relu, learning_rate: 0,001	units: 192, activation: relu, learning_rate: 0,001

	LSTM	units: 160, activation: tanh, learning_rate: 0,01	units: 224, activation: tanh, learning_rate: 0,01
	GRU	units: 128, activation: tanh, learning_rate: 0,01	units: 160, activation: relu, learning_rate: 0,01
Bitcoin	ARIMA	ARIMA(2,0,2)	ARIMA(0,0,2)
	RF	n_estimators=500, max_depth=10, lags=20	n_estimators=500, max_depth=5, lags=20
	SVR	C: 100, epsilon: 0,01, gamma: auto	C: 100, epsilon: 0,01, gamma: auto
	MLP	units: 192, activation: relu, learning_rate: 0,01	units: 192, activation: relu, learning_rate: 0,001
	RNN	units: 192, activation: relu, learning_rate: 0,001	units: 192, activation: relu, learning_rate: 0,001
	LSTM	units: 224, activation: tanh, learning_rate: 0,01	units: 192, activation: relu, learning_rate: 0,001
	GRU	units: 160, activation: relu, learning_rate: 0,01	units: 160, activation: relu, learning_rate: 0,01
Ethereum	ARIMA	ARIMA(2,0,1)	ARIMA(1, 0, 1)
	RF	n_estimators=100, max_depth=5, lags=20	n_estimators=500, max_depth=5, lags=20
	SVR	C: 1, epsilon: 0,01, gamma: auto	C: 100, epsilon: 0,01, gamma: auto
	MLP	units: 128, activation: tanh, learning_rate: 0,01	units: 160, activation: relu, learning_rate: 0,001
	RNN	units: 192, activation: relu, learning_rate: 0,001	units: 160, activation: relu, learning_rate: 0,01
	LSTM	units: 128, activation: tanh, learning_rate: 0,01	units: 128, activation: tanh, learning_rate: 0,01
	GRU	units: 160, activation: relu, learning_rate: 0,01	units: 160, activation: relu, learning_rate: 0,001

Lisa 10. Mitme sagedusega mudelite lõplikud seadistused

Vara	Mudel	Kaks intervalli	Hierarhiline ansambel
S&P 500	ARIMA	-	ARIMA(1,0,5)
	RF	-	n_estimators=500, max_depth=5, lags=80
	SVR	C: 100, epsilon: 0,01, gamma: auto	C: 100, epsilon: 0,01, gamma: auto
	MLP	units: 224, activation: relu, learning_rate: 0,001	units: 224, activation: relu, learning_rate: 0,001
	RNN	units: 224, activation: relu, learning_rate: 0,001	units: 224, activation: relu, learning_rate: 0,001
	LSTM	units: 160, activation: relu, learning_rate: 0,001	units: 160, activation: relu, learning_rate: 0,001
	GRU	units: 224, activation: relu, learning_rate: 0,001	units: 192, activation: relu, learning_rate: 0,001
Dow Jones	ARIMA	-	ARIMA(0,0,1)
	RF	-	n_estimators=500, max_depth=5, lags=80
	SVR	C: 100, epsilon: 0,01, gamma: auto	C: 100, epsilon: 0,01, gamma: auto
	MLP	units: 192, activation: relu, learning_rate: 0,001	units: 192, activation: relu, learning_rate: 0,001
	RNN	units: 224, activation: relu, learning_rate: 0,001	units: 160, activation: relu, learning_rate: 0,001
	LSTM	units: 160, activation: relu, learning_rate: 0,001	units: 160, activation: relu, learning_rate: 0,001
	GRU	units: 160, activation: relu, learning_rate: 0,001	units: 160, activation: relu, learning_rate: 0,01
Nasdaq 100	ARIMA	-	ARIMA(0,0,2)
	RF	-	n_estimators=500, max_depth=5, lags=80
	SVR	C: 1, epsilon: 0,01, gamma: auto	C: 1, epsilon: 0,01, gamma: auto

	MLP	units: 224, activation: relu, learning_rate: 0,001	units: 224, activation: relu, learning_rate: 0,001
	RNN	units: 224, activation: relu, learning_rate: 0,001	units: 192, activation: relu, learning_rate: 0,001
	LSTM	units: 160, activation: tanh, learning_rate: 0,001	units: 160, activation: relu, learning_rate: 0,001
	GRU	units: 192, activation: relu, learning_rate: 0,001	units: 160, activation: relu, learning_rate: 0,01
Bitcoin	ARIMA	-	ARIMA(0,0,2)
	RF	-	n_estimators=100, max_depth=10, lags=168
	SVR	C: 100, epsilon: 0.01, gamma: auto	C: 100, epsilon: 0,01, gamma: auto
	MLP	units: 192, activation: relu, learning_rate: 0,001	units: 192, activation: relu, learning_rate: 0,001
	RNN	units: 224, activation: relu, learning_rate: 0,001	units: 192, activation: relu, learning_rate: 0,001
	LSTM	units: 128, activation: tanh, learning_rate: 0,01	units: 192, activation: relu, learning_rate: 0,001
	GRU	units: 192, activation: relu, learning_rate: 0,001	units: 160, activation: relu, learning_rate: 0,01
Ethereum	ARIMA	-	ARIMA(1, 0, 1)
	RF	-	n_estimators=500, max_depth=5, lags=168
	SVR	C: 100, epsilon: 0,01, gamma: auto	C: 100, epsilon: 0,01, gamma: auto
	MLP	units: 192, activation: relu, learning_rate: 0,001	units: 160, activation: relu, learning_rate: 0,001
	RNN	units: 192, activation: relu, learning_rate: 0,001	units: 160, activation: relu, learning_rate: 0,01
	LSTM	units: 192, activation: relu, learning_rate: 0,001	units: 192, activation: relu, learning_rate: 0,001
	GRU	units: 192, activation: relu, learning_rate: 0,001	units: 160, activation: relu, learning_rate: 0,001