

TARTU ÜLIKOOL
Arvutiteaduse instituut
Andmeteaduse õppekava

Kirill Grjaznov
Päikesepaneelide tootlikkuse ennustamine
Magistritöö (15 EAP)

Juhendaja: Meelis Kull
Kaasjuhendaja: Janika Aan

Tartu 2023

Päikesepaneelide tootlikkuse ennustamine

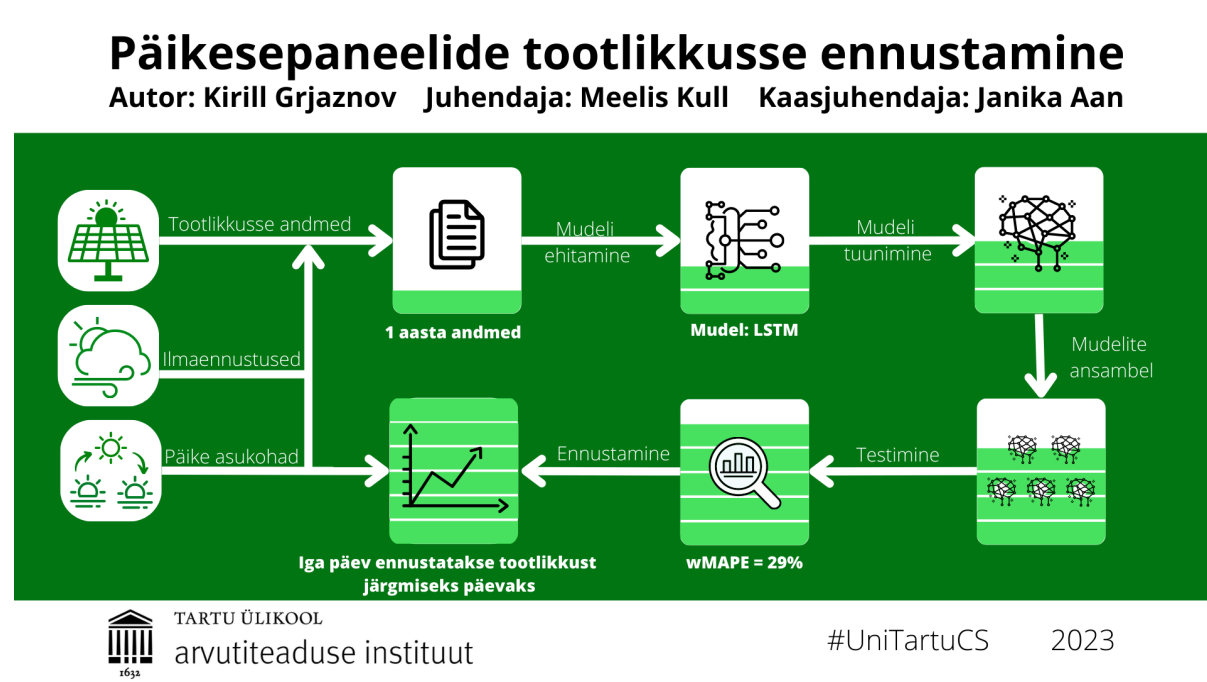
Lühikokkuvõte:

Eesti on osa Nord Pool elektribörsist, seega on vajalik elektrienergia tarbimise prognoosimine järgmiseks päevaks. Eesmärk on hoida elektrisüsteemi bilanssi, tagades, et ostetud elektrienergia kogus vastab tegelikule tarbimisele. Taastuvenergia, nagu päikeseenergia, on kõikuv, seega selle tootlikkuse ennustamine võimaldab paremini planeerida elektrienergia tarnet järgmiseks päevaks. Magistritöö käigus loodi masinõppe mudel, mis ennustab päikesepaneelide pargi elektritootlikkust tunni täpsusega järgmise päeva jooksul. Mudeli treenimisel kasutati 1-aastase ajaloolise ilmaennustuste andmeid ja päikesepaneelide tootlikkuse andmeid ning arvutati ka päikese ja paneelide vaheliste nurkade väärtused igal ajahetkel. Ehitati ja võrreldi kolme mudelit: lineaarne regressioon, XGBoost ja LSTM-ansambel. Parimaks osutus LSTM-ansambel, mille wMAPE-testväärtus oli terve kalendriaasta jooksul 29%.

Võtmesõnad:

Päikesepaneelide tootlikkus, taastuvenergia ennustamine, tehisintellekt

CERCS: P176



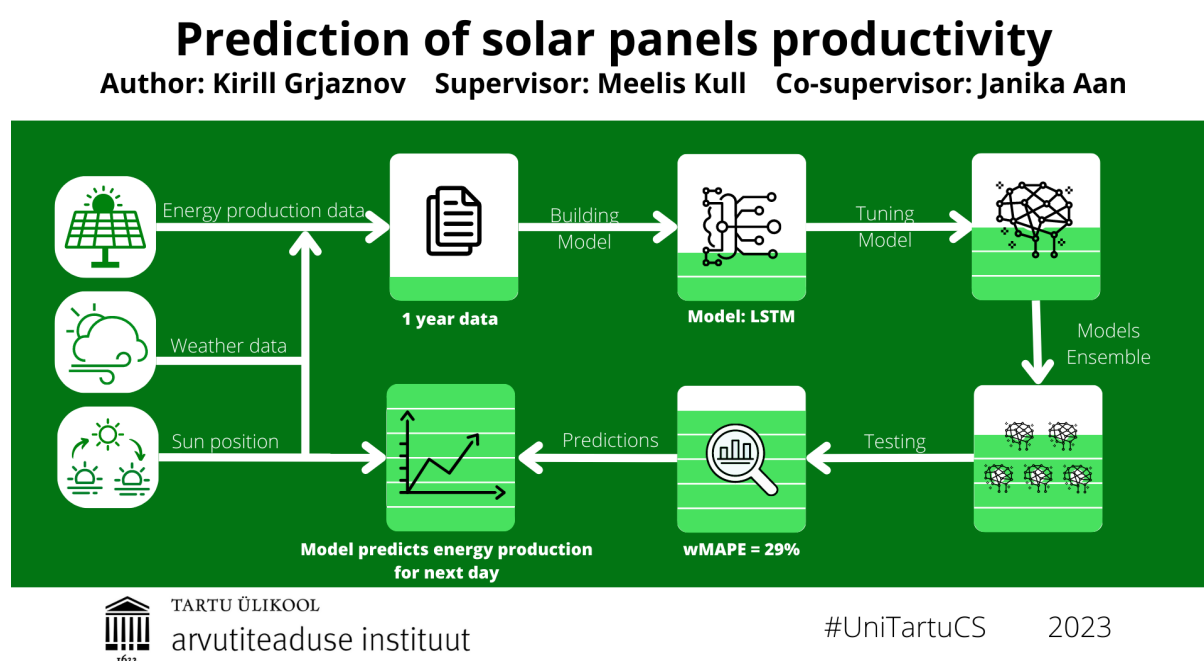
Prediction of solar panels productivity

Abstract:

Estonia is part of the Nord Pool electricity exchange, therefore it is necessary to forecast the electricity consumption for the next day. The goal is to maintain the balance of the power system by ensuring that the purchased amount of electricity matches the actual consumption. Renewable energy sources, such as solar energy, are variable, so predicting their productivity allows for better planning of electricity supply for the next day. During the master's thesis, a machine learning model was created that predicts the electricity production of the solar panel park with hourly accuracy for the next day. The model was trained using 1-year historical weather forecast data, solar panel productivity data, and the values of angles between the sun and the panels at each time. Three models were built and compared: linear regression, XGBoost, and LSTM ensemble. The best performance was achieved by the LSTM ensemble, whose wMAPE test score was 29% throughout the entire calendar year.

Keywords: Solar panels energy production, renewable energy forecasting, artificial intelligence

CERCS: P176



Sisukord

1. Sissejuhatus	5
2. Taustainfo	7
3. Mõisted ja terminid	9
4. Andmete kirjeldus	12
4.1. Päikesepaneelide tehnilised andmed	12
4.2. Ilmaennustuste andmed	13
4.3. Astronoomilised andmed	13
5. Andmete ettevalmistamine	14
5.1. Andmete täiendamine	14
5.2. Päikese suhteline asukoht	14
5.3. Horisondi tunnus	15
5.4. Skaleerimine	15
5.5. Aegrida vorm	16
5.6. Treenimis-, valideerimis- ja testandmed	18
6. Mudeli ehitamine	19
6.1. Lineaarne regressioon	19
6.2. XGBoost	20
6.3. LSTM	22
6.4. LSTM modifikatsioon	24
6.5. Ansambel	25
7. Tulemused	27
7.1. Üldine ja spetsiifiline testid	27
7.2. Aastaajad	27
8. Kokkuvõte	29
Viidatud kirjandus	30
Lisad	32
Kasutatud töövahendid	32

1. Sissejuhatus

Eesti on Nord Pool energia börsi osa, mis tähendab, et Eesti impordib elektrienergiat ning samas ka ekspordib. Iga riik peab Nord Pooli süsteemi raames edastama iga päev aruannet, kui suur on elektri nõudlus järgmiseks päevaks (tunni täpsusega). Lisaks riik peab edastama elektri tootlikkuse prognoosi järgmiseks päevaks, millest lähtuvalt arvutatakse elektri hind järgmiseks päevaks. Eestis vastutab selle eest Elering. Elektrienergiat püütakse hoida bilansis, tagades, et ostetud elektri kogus vastaks reaalsele elektritarbimisele. Juhul kui seda ei ole, siis tekib eabilanss (eabilanssi mõiste Eleringi käsiraamatust [1]) ehk etteostetud elektrikogus osutus liiga suureks või liiga väikseks. Sellisel juhul elektri ülejääk müüakse maha ning elektri puuduse korral ostetakse juurde. Eabilanssi korral elektri hinnad on teistsugused ning toovad lisaraha kulu. Antud protsessi reguleerib Eleringi poolt koostatud dokument “Baltic CoBA Imbalance Settlement Rules” [2].

Tänapäeval on roheline energia muutunud üha aktuaalsemaks teemaks. Taastuvenergia allikate, nagu tuule- ja päikeseenergia, kasutamine on oluliselt kasvanud. Kuid nende energiaallikate kasutamisel on peamine probleem nende raskesti prognoositavus, sest nad sõltuvad otseselt ilmastikutingimustest, mitte inimtegevusest. Selline kõikumine toob kaasa energia eabilanssi, mis võib omakorda põhjustada elektrienergia hinnakõikumisi ja elektrivõrgu stabiilsuse probleeme.

Antud magistritöö on tehtud “IT tööstusmagistrantuuri” raames ning eesmärgiks oli ehitada masinõpe mudel, mis ennustab päikesepaneelide pargi tootlikkust järgmiseks päevaks tunni täpsusega, selleks et ennustuse abil võimaldada efektiivsemat energia kasutust eabilanssi minimeerimise teel. Päikesepaneelide park asub Eestis Madismäe külas ning selle omanikuks on Eesti Gaas. Magistritöö oli tehtud Telia initsiatiivil Eesti Gaasi jaoks.

Käesolev ülesanne on aegrida ennustamine, mida on varasemalt uuritud, näiteks, artiklis [3] oli uuritud “feed-forward” närvivõrku kahes implementatsioonis, kus esimeses implementatsioonis algul ennustati päikese radiatsiooni, mille järel klassifitseeriti päeva kahte gruppi: “pilvine” ja “päikseline”. Iga grupi jaoks treeniti eraldi närvivõrk, mis ennustas tootlikkust järgmiseks päevaks. Teises implementatsioonis kasutati nii nimetatud Physical Hybrid Artificial Neuron Network (PHANN), kus treeniti üks närvivõrk, mille sisendiks olid lisaks ilmaennustusele ka teoreetiline päikese radiatsiooni tase puhta taevas tingimustel [3]. Hiljuti tehtud uurimus [4] näitab, et lühiajalisel (4 kuni 72 tunni) ennustamisel, kus on üheks sisendiks ilmaennustused, kasutati tihti Numerical weather prediction (NWP) mudelit.

Juhtudel, kus ei ole teada ilma andmeid kasutati statistilisi meetodeid ning LSTM närvivõrku. Varasemalt kasutati ka CNN närvivõrku, millega algul klassifitseeriti järgmist päeva viite kategooriasse: vihmane, tugevalt pilvine, pilvine, nõrgalt pilvine ning päikseline [5]. Peale seda treeniti CNN regressori iga tüüpi jaoks, selleks et ennustada tootlikkust järgmiseks päevaks. Lisaks, artiklis [6] võrreldi LSTM, XGBoost ning ARIMA mudeleid aegrida ennustamisel, kus LSTM näitas paremad tulemused.

Töö esimeses peatükis kirjeldatakse andmestikku, mille alusel oli ehitatud mudel. Teises peatükis räägitakse andmete ettevalmistamisest ning tunnuste konstrueerimisest. Kolmandas peatükis käsitletakse mudeli loomise protsessi ning läbiviidud eksperimente. Neljandas peatükis on näidatud töö tulemused. Viimases peatükis on toodud töö kokkuvõte.

2. Taustainfo

Päikesepaneelide tootlikkuse ennustamise kasu seisneb selles, et see võimaldab paremini aru saada, kui palju toodetakse energiat järgmisel päeval. Selle abil saab paremini planeerida, kui palju on vaja osta energiat juurde, ning tulemusena vähendada eabilanssi, mis omakorda säästab raha. Elering on süsteemihaldur ehk vastutab selle eest, et terve süsteem (antud juhul Eesti) oleks bilansis. Töö eesmärk on aga ennustada ühe konkreetse päikesepaneelide pargi tootlikkust järgmiseks päevaks, mille omanikuks on Eesti Gaas. Eesti Gaas hangib müüdava elektri börsilt Nord Pool ning müüb seda oma klientidele edasi [7]. Teades päikesepaneelide pargi tootlikkust järgmiseks päevaks, saab Eesti Gaas paremini planeerida, kui palju energiat on vaja hankida, et vähendada eabilanssi. Dokument [2] reguleerib eabilansi hindu, vaadates terve süsteemi eabilanssi. Eabilansi hind kujuneb:

$$P_{imb} = P_{bal} - P_{ntc} \text{ juhul, kui on energiaülejääk}$$

$$P_{imb} = P_{bal} + P_{ntc} \text{ juhul, kui on energia defitsiit, kus}$$

P_{imb} on eabilansi hind, P_{bal} on bilansi hind ning P_{ntc} on neutraalsuse komponent, mis on konstantne ühe kuu jaoks. Kõik ülaltoodud liikmed on EUR/MWh ühikutes. Kui me vaatame antud olukorda tootlikkuse ennustamise kontekstis, siis võib juhtuda kolm stsenaariumi:

- 1) Mudel ennustas tootlikkust ideaalselt. Sellisel juhul toetudes ennustuse peale me ostime ideaalse koguse elektrit ning selle tõttu rahakulu on:

$$R_{ideaalne} = (N - T)P_{bal} = (N - E)P_{bal}, \text{ kus}$$

N on elektri nõudlus (MWh), T on päikesepaneelide reaalne tootlikkus (MWh) ning E on ennustatud tootlikkus (MWh).

- 2) Mudel ennustas suuremat tootlikkust, kui reaalsuses oli. Sellisel juhul toetudes ennustuse peale me ostime vähem energiat, kui oli vaja ehk tegu on energia defitsiidiga. Järelikult peame ostma elektrit juurde. Meie raha kulu on:

$$R_{defitsiit} = (N - E)P_{bal} + (E - T)P_{imb}, \text{ mis erineb ideaalsest olukorrast:}$$

$$\begin{aligned} R_{erinevus} &= R_{defitsiit} - R_{ideaalne} = (N - E)P_{bal} + (E - T)P_{imb} - (N - T)P_{bal} = \\ &= T * P_{bal} - E * P_{bal} + E * P_{imb} - T * P_{imb} = (T - E)P_{bal} - (T - E)P_{imb} = \\ &= (T - E)(P_{bal} - P_{imb}) = (E - T)P_{ntc}, \end{aligned}$$

see on lisa rahakulu, mis oli põhjustatud mudeli ebatäpsusega.

- 3) Mudel ennustas väiksemat tootlikkust, kui reaalsuses oli. Sellisel juhul toetudes ennustuse peale me ostime rohkem energiat, kui oli vaja ehk tegu on energia ülejäägiga. Järelikult peame üleliigset elektrit maha müüma. Meie raha kulu on:

$R_{\text{ülejääk}} = (N - E)P_{bal} - (T - E)P_{imb}$, mis erineb ideaalsest olukorrast:

$$\begin{aligned} R_{\text{erinevus}} &= R_{\text{ülejääk}} - R_{\text{ideaalne}} = (N - E)P_{bal} - (T - E)P_{imb} - (N - T)P_{bal} = \\ &= T * P_{bal} - E * P_{bal} - T * P_{imb} + E * P_{imb} = (T - E)P_{bal} - (T - E)P_{imb} = \\ &= (T - E)(P_{bal} - P_{imb}) = (T - E)P_{ntc} = -(E - T)P_{ntc} \end{aligned}$$

Kokkuvõtteks, lisaraha kulu võrdub $\pm (E - T)P_{ntc} = |E - T| * P_{ntc}$ vastavalt sellele, kas ennustatud tootlikkus oli väiksem või suurem reaalsest tootlikkusest. Kuna P_{ntc} on konstantne ühe kuu jooksul, siis ennustamisel saame minimiseerida ainult $|E - T|$ tegurit. Lisaks, antud mudeli viga käib ühe ennustuse kohta ehk ühe ennustatud tunni kohta. Kuna töö eesmärgiks on ennustada tootlikkust järgmiseks päevaks, siis kokkuvõttes need vead summeeritakse, seega antud magistritöö autor otsustas masinõpe mudeli loomisel optimeerida just **MAE** meetrikut ehk keskmist absoluutviga, mida tutvustatakse järgmises peatükis.

3. Mõisted ja terminid

Antud töös ehitati masinõpe mudel, mis ennustas päikesepaneelide tootlikkust järgmiseks päevaks, mille ühikuks oli kWh ehk sihttunnus on mitte negatiivne reaalarv. Mudeli ennustuste vea mõõtmiseks on olemas mitmeid meetrikuid, nagu keskmine absoluutviga (mean absolute error ehk MAE) ja keskmine ruutviga (mean squared error ehk MSE). MAE matemaatiline valem on:

$$MAE = \frac{\sum_{i=1}^n abs(y_i - p_i)}{n}, \text{ kus}$$

n on andmepunktide koguarv, y_i on tegelik tootlikkus i -ndal ajahetkel ning p_i on ennustatud tootlikkus i -ndal ajahetkel. MSE matemaatiline valem on:

$$MSE = \frac{\sum_{i=1}^n (y_i - p_i)^2}{n}$$

Eelmises peatükis oli järeldatud, et antud töös valitakse mudeli optimeerimiseks MAE, aga selleks, et tulemus oleks paremini interpreteeritav, oli võetud kasutusele ka protsendiline MAE ehk MAPE, mille valem on:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n abs\left(\frac{y_i - p_i}{y_i}\right)$$

MAPE meetrik on problemaatiline antud ülesandes, sest päikesepaneelide reaalne tootlikkus on tihti 0, mis toob ebakvaliteetset protsendilist hinnangut, seega oli otsustatud kasutada MAPE modifikatsiooni ehk kaalutud keskmist protsendilist absoluutviga (wMAPE), kus kaalud (w_i) on reaalse tootlikkuse väärtused. wMAPE lahendab probleemi, mis on seotud 0-ga, ning selle valem on:

$$\begin{aligned} wMAPE &= \frac{\sum_{i=1}^n w_i * \frac{|y_i - p_i|}{|y_i|}}{\sum_{i=1}^n w_i} 100\% = \frac{\sum_{i=1}^n |y_i| * \frac{|y_i - p_i|}{|y_i|}}{\sum_{i=1}^n |y_i|} 100\% = \frac{\sum_{i=1}^n |y_i - p_i|}{\sum_{i=1}^n |y_i|} 100\% = \\ &= \frac{\frac{\sum_{i=1}^n |y_i - p_i|}{n}}{\frac{\sum_{i=1}^n |y_i|}{n}} 100\% = \frac{MAE}{\text{Keskmine tootlikkus}} 100\% \end{aligned}$$

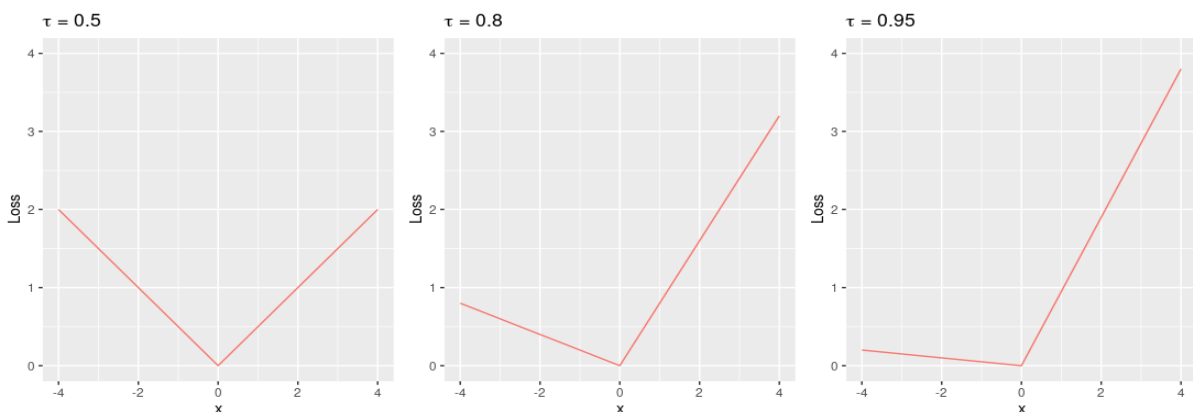
Aegrea prognoosimisel mudeli väljundiks on tavaliselt üks ennustatud väärtus, mille põhjal tehakse, näiteks, mingeid ärilisi otsused. Ennustatud väljund ei anna mingit infot selle kohta, kui tõenäoline on antud väärtus, ning võib mõnikord see väärtus olla üllatavalt suur või väike. Otsuste tegemisel oleks kasulik teada mitte ainult ennustatud väärtust, kuid ka ennustusintervalli, kuhu langeb tegelik väärtus mingi eelsätestatud tõenäosusega [8]. Ennustusintervall võimaldab hinnata määramatust ning kaaluda võimalikud riskid otsuste tegemisel. Päikesepaneelide tootlikkuse ennustamise kontekstis on määramatus suuresti tingitud ilmaennustuste ebatäpsusega. Antud töös autor otsustas lisaks tootlikkuse ennustamisele treenida masinõpe mudelit nii, et ta väljastaks ka ennustusintervalli, et hinnata määramatust. Esimene samm on panna paika, kui laia intervalli me soovime ennustada ehk mis tõenäosusega reaalse tootlikkuse väärtus peab langema ennustusintervalli. Tavaline praktika on valida 90% tõenäosusega intervall, mis tähendab, et on vaja ennustada määramatuse kaks kvantiili, mis katavad 90% määramatuse väärtustest. Antud töös autor valis sümmeetrilise intervalli ehk ülemiseks kvantiiliks on 0.95 ning alumiseks kvantiiliks 0.05, mis näitavad päikesepaneelide ennustatud tootlikkuse intervalli ülemist ja alumist piire. Mudeli testimisel on vaja mõõta, kui tihti reaalse tootlikkuse väärtus langeb ennustatud intervalli sisse. Ideaalis see peaks olema 90% juhtudest.

Masinõpe mudeli treenimisel selleks, et saada väljundiks mingi konkreetse kvantiili väärtuse, kasutatakse nn “*pinball loss*” funktsiooni, mille valem on [9]:

$$L(y, p) = (y - p)\tau, \text{ kui } y \geq p$$

$$= (p - y)\tau, \text{ kui } y < p, \text{ kus}$$

y on reaalne väärtus, p on ennustatud väärtus ning τ on sihtkvantiil. Edaspidi töös nimetame seda funktsiooni *pinball*(τ)-ks. Joonisel 1 on näidatud *pinball* kaofunktsioon, juhul kui τ võrdub 0.5, 0.8 ja 0.95.



Joonis 1. Pinball kaofunktsiooni graafik, kui $\tau = \{0.5, 0.8, 0.95\}$ [10].

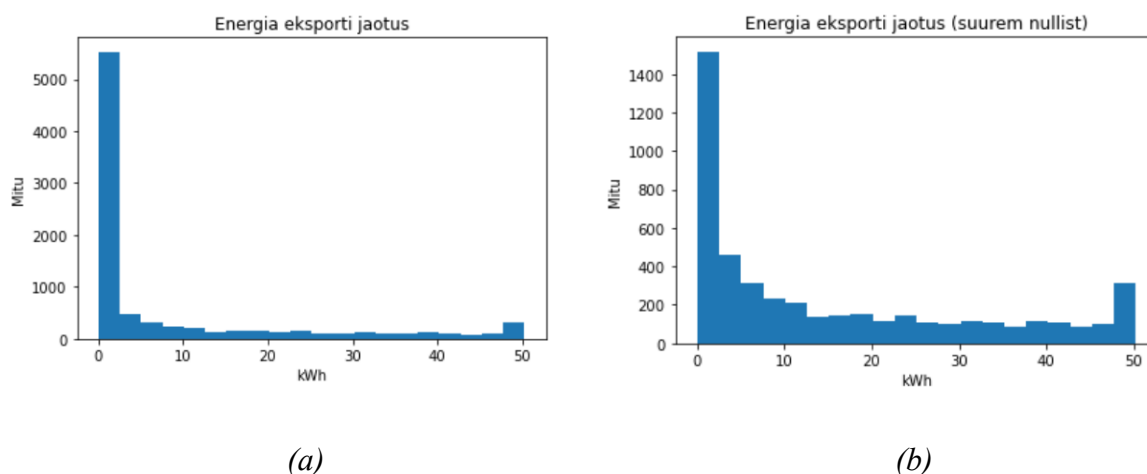
Antud töö üldine eesmärk on treenida masinõpe mudeli, mis ennustaks päikesepaneelide pargi energia tootlikkust mingi konkreetse tunni jaoks, andes sisendiks vastavad tunnused. Seejärel on võimalik ennustada tootlikkust terveks järgmiseks päevaks, ennustades iga tunni tootlikkust eraldi. Lõplik mudel peaks lahendama aga spetsiifilist ülesannet, nimelt, iga päev kell 11:00 ennustada tootlikkust järgmiseks päevaks (alustades kellast 00:00 kuni 23:00). Ülesanne on põhjustatud äriliste vajadustega. Vastavalt sellele antud töös oli läbi viidud mudeli kaks lõplikku testi: **üldine**- ja **spetsiifiline** test. Spetsiifilisel testil testandmete loomisel võeti ilmaennustused, mis olid tehtud kell 11:00, ning hinnati mudeli ennustusi järgmiseks päevaks ehk alates 00:00 kuni 23:00. Kuna mudelit treeniti kasutades kõik ilmaennustuse hetki (mitte ainult kell 11:00), selleks et oleks rohkem treenimise andmeid, siis töös tehti ka üldist testi, kus testimisandmed olid samast jaotusest nagu treenimis- ja valideerimisandmed.

4. Andmete kirjeldus

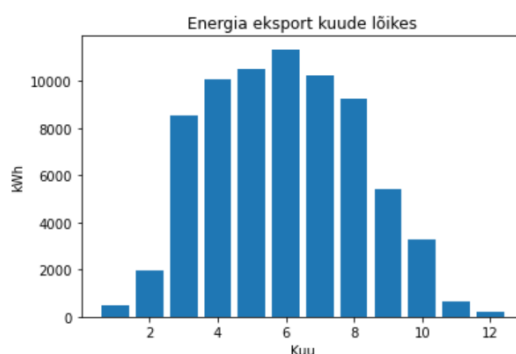
Antud peatükis kirjeldatakse töös kasutatud andmed.

4.1. Päikesepaneelide tehnilised andmed

Päikesepaneelide ajalooline tootlikkus on saadud Eesti Gaasi käest. Andmed kirjeldavad päikesepaneelide terve pargi energia eksporti (kWh) perioodil 04.01.2022 kuni 31.12.2022 ning olid agregeeritud tunni järgi. Päikesepaneelide park koosnes 200-st paneelist, pargi täpne asukoht on Madismäe küla laiuskraadiga 59.445° , paneelide kaldenurk on 35° maapinna suhtes ning paneelid olid suunatud lõuna. Energia tootlikkuse jaotus on näidatud Joonisel 2a. Joonisel 2b on näidatud sama jaotus, kust on välja võetud 0 väärtused ehk hetked, millal päikesepaneelide park ei tootnud energiat. Kui vaadata energia tootlikkust kuude lõikes (Joonis 3), siis on näha, et suurim tootlikkus oli märtsist augustini ning väiksem tootlikkus septembrist veebruarini.



Joonis 2 (a,b). a) Päikesepaneelide pargi energia tootlikkuse jaotus.
b) Sama jaotus, kust on võetud välja 0 väärtused ehk hetked, millal päikesepaneelide park ei tootnud energiat.



Joonis 3. Päikesepaneelid tootlikkus kuude lõikes

4.2. Ilmaennustuste andmed

Päikesepaneelide tootlikkus sõltub päikese radiatsioonist, mis omakorda sõltub ilmast, näiteks pilved takistavad kiirguste maandumist paneelide peale, mille tõttu energia tootlikkus väheneb. Kvaliteetne ilmateave on oluline päikesepaneelide tootlikkuse ennustamisel. Kuna töö eesmärk on ennustada energia tootlikkust järgmiseks päevaks, siis on vaja teada ka ennustatud ilma järgmiseks päevaks. Järelikult, mudeli testimisel ja rakendamisel kasutame sisendiks ilmaennustusi, seega ka mudeli treenimisel kasutame ajaloolisi ilmaennustusi, mitte teadaolevat tegelikku ilmateadet. Ilmaennustuste andmed olid kogutud Telia poolt Norra Meteoroloogia Instituudi portaalist [11]. Iga tund koguti ilmaennustusi järgmiseks 48ks tunniks. Andmed hõlmasid endas järgmisi tunnuseid: õhurõhk, õhutemperatuur (koos 0.1 ja 0.9 kvantiilidega), pilvisus (üldine ning kolmel erineval kõrgusel), katte kujunemise temperatuur, udu osakaal, suhteline niiskus, ultraviolet indeks, tuulekiirus (koos 0.1 ja 0.9 kvantiilidega) ning maksimaalne tuulekiirus. Lisaks iga ennustuse küljes oli info: millal tehti ennustust ning mis ajahetkeks ennustatakse. Kokku oli 16 ilma kirjeldavat tunnust. Ilmaennustuste andmeid koguti perioodil 17.01.2022 kuni 31.12.2022, seega ka lõplik andmestik katab ainult seda perioodi.

4.3. Astronoomilised andmed

Üheks vajalikuks faktoriks päikesepaneelide tootlikkuse kontekstis on päikese asukoht igal hetkel, millal soovime ennustada tootlikkust. Selleks, et arvutada vajalikud nurgad, olid võetud andmed astronoomia veebiportaalist [12]. Andmetest on kättesaadav päikesetõusu, loojangu ning kulminatsiooni ajahetked iga päeva jaoks.

5. Andmete ettevalmistamine

5.1. Andmete täiendamine

Ilmaandmetes esinesid puudulikud read. Selleks, et säilitada andmete terviklikkust aegrea kontekstis, autor otsustas täita puudulikud read “forward-fill” meetodiga, kus lünka täitmiseks võetakse eelmise rea väärtusi [13]. Antud meetod näitas paremaid tulemusi võrreldes keskmise väärtuste võtmisega juhul kui korrelatsioonid on positiivsed.

5.2. Päikese suhteline asukoht

Päikesetõusu, loojangu ja kulminatsiooni ajahetkede abil arvutati järgmised tunnused iga tunni jaoks: päike ja horisondi vaheline nurk, päike praeguse asukoha ja päike kulminatsiooni asukoha vaheline nurk, päike ja Maa telje vaheline nurk (deklinatsiooni nurk) ning päike kiirgede ja paneelide vaheline nurk. Vajalikud füüsika valemid olid võetud Soteris A. Kalogirou raamatutest [14,15]. Alguses oli välja arvutatud päikese deklinatsiooni nurk andmestiku iga päeva jaoks. Päikese deklinatsioon saavutab maksimaalse nurga 23 kraadi 17 minutit 21. juunil (põhjapoolkeral suvine pööripäev) ning minimaalse nurga -23 kraadi 27 minutit 20. detsembril (põhjapoolkeral talvine pööripäev). Iga päeva jaoks saame hinnata päikese deklinatsiooni nurga järgneva valemiga:

$$\delta = (23 + \frac{27}{60}) * \sin(\frac{360d}{365.25}),$$

kus d on möödunud päevade arv pärast kevadist pööripäeva põhjapoolkeral ehk pärast 21. märtsi.

Järgmiseks tunnuseks oli arvutatud päike praeguse asukoha ja kulminatsiooni asukoha vaheline nurk, mida nimetatakse ka *tunni nurgaks*. Päike liigub oma trajektooriga iga minut 0.25 kraadi (täispöörde 24*60 minuti jooksul). Teades, millal on kulminatsioon ning praegust ajahetke, arvutame ajanihe Δt minutites ning korrutame 0.25-ga:

$$h = \Delta t * 0.25 \text{ kraadi}$$

Päike ja horisondi vahelist nurga iga tunni jaoks saame arvutada, kui teame antud päeva päikese deklinatsiooni nurga, antud ajahetkel h ning päikesepaneelide pargi asukoha laiuskraadi L :

$$\sin(\alpha) = \sin(L) * \sin(\delta) + \cos(L) * \cos(\delta) * \cos(h)$$

Lõpuks, arvutame päikese kiirgede ja paneelide vahelist nurga iga tunni jaoks:

$$\begin{aligned}
\cos(\theta) = & \sin(L)\sin(\delta)\cos(\beta) - \\
& - \cos(L)\sin(\delta)\sin(\beta)\cos(A) + \\
& + \cos(L)\cos(\delta)\cos(h)\cos(\beta) + \\
& + \sin(L)\cos(\delta)\cos(h)\sin(\beta)\cos(A) + \\
& + \cos(\delta)\sin(h)\sin(\beta)\sin(A),
\end{aligned}$$

kus β on paneelide kaldenurk maa suhtes ning A on paneelide asimuutnurk lõuna suhtes, kus lääne suund on määratud positiivseks.

Autori eeldusel päike ja horisondi vaheline nurk peaks aitama mudelil saada aru, millal algab ja lõpeb energia tootlikkus päeva jooksul. Päike praeguse asukoha ja päike kulminatsiooni asukoha vaheline nurk peaks aitama leidma tootlikkuse tipphetke. Päike ja Maa telje vaheline nurk aitab paremini võtta arvesse aastaajad. Päikese kiirgede ja paneelide vaheline nurk on vajalik selleks, et teada, kui palju valgust võib teoreetiliselt langeda paneelide peale.

5.3. Horisondi tunnus

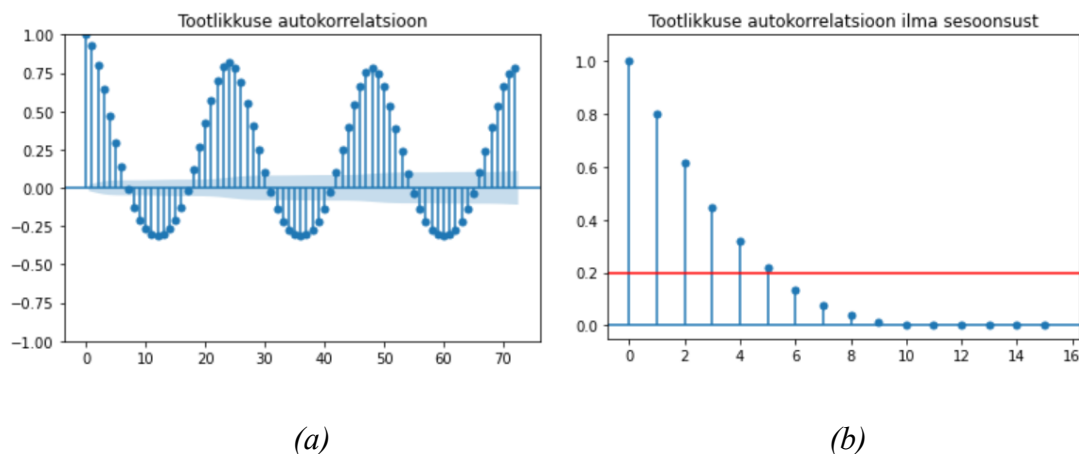
Töö eesmärgiks oli ennustada päikesepaneelide tootlikkust järgmiseks päevaks, selleks kasutatakse ilmaennustust järgmiseks päevaks, mille andmed on rohkem ebatäpsemad, mida kaugemale me ennustame. Selleks, et treeningandmed oleksid samast jaotusest nagu testandmed, oli lisatud tunnus “horisont”, mis näitab mitu tundi tagasi oli ennustatud ilm vaadeldava ajahetke jaoks. Antud tunnus võimaldab mudelil aru saada, kui kaugele tahetakse ennustada tootlikkust.

5.4. Skaleerimine

Andmeteaduses hea praktika on tunnuste skaleerimine juhul, kui tunnuste väärtused on erinevatel skaaladel. Antud töös see on vajalik, sest tunnuste skaalad on erinevad, näiteks õhurõhu väärtused on vahemikus 960 kuni 1050, õhutemperatuuri väärtused -20 kuni 32 ning tuulekiiruse väärtused 0 kuni 13. Skaleerimiseks on teada mitut meetodit nagu MinMax, standardne ja robustne (ing.k *robust*) skalareemine [16]. Iga tunnuse jaoks on vaja valida sobiv skaleerimine sõltuvalt tunnuse jaotusest. Standardset skaleerimist kasutatakse siis, kui tunnuse jaotus on sarnane normaaljaotusega. MinMax skaleerimist kasutatakse siis, kui tunnusel ei ole ekstreemselt suuri või väikseid väärtusi. Vastupidavat skaleerimist kasutatakse, kui ekstreemseid väärtusi on suhteliselt palju.

5.5. Aegrida vorm

Töös kasutatud andmed on aegrida, seega on vajalik otsustada, kui palju samme enne ennustuse hetke me võtame arvesse.



Joonis 4 (a,b). a) Päikesepaneelide tootlikkuse autokorrelatsioon.
b) Päikesepaneelide tootlikkuse autokorrelatsioon ilma perioodilist mõju.

Aegrea kontekstis autokorrelatsioon on tunnuse sõltuvus (korrelatsioon) tema eelmistest väärtustest. Vaadates sihttunnuse (energia tootlikkus) autokorrelatsiooni (Joonis 4a), on näha, et eksisteerib perioodiline sõltuvus (24 tundi). Antud autokorrelatsioon võib olla tingitud päikese asukohaga, näiteks tootlikkus kell 12:00 on sarnane eelmise päeva tootlikkusega kell 12:00, kui ilm on sarnane. Varasemalt peatükis 5.2 me arvutasime nurgad, mis juba omavad perioodilisuse konteksti, seega on mõistlik vaadata autokorrelatsiooni, kus on välja võetud perioodiline mõju, selleks et vähendada ebavajalikku andmete dimensionaalsust. Perioodilise mõju väljavõtmiseks lahutatakse aegrea väärtusest nihutatud väärtust. Seda nimetatakse esimese taseme sesooniliseks erinevuseks [17]:

$$D_t = y_t - y_{t+24}, \text{ kus}$$

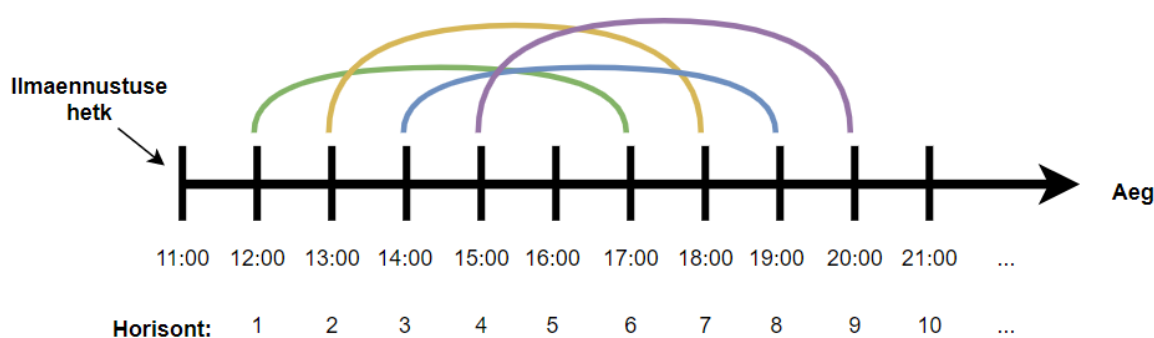
y_t on energia tootlikkus ajahetkel t ning D_t on energia tootlikkus ajahetkel t , millest lahutati energia tootlikkust ajahetkel $t+24$. Tulemuseks saame uue aegrea D_t , kust on välja võetud perioodiline mõju. Uurides D_t autokorrelatsiooni graafikut (Joonis 4b), autor otsustas, et võetakse arvesse 5 eelmist tundi ning praegune tund (kokku 6 sammu), sest peale seda autokorrelatsiooni väärtus on alla 0.2. Teaduses ei ole fikseeritud kindlat autokorrelatsiooni lävendit, mis liigitaks seda tugevale või nõrgale seosele, seega antud valik on subjektiivne.

Üks andmepunkt koosnes kuuest sammust, mis tähendab, et ajahetkel t meil on teada reaalne energia tootlikkus y_t , sellele vastavad tunnuste komplektid $X_t, X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}$ ja X_{t-5} . Tunnuste komplekt X_t koosnes ilmaennustusest antud hetkeks (õhutemperatuur, õhurõhk jne, kokku 16 tunnust), arvutatud nurkadest antud hetkeks (kokku 4 tunnust) ning “horisont” tunnus. Lisaks, kaks tunnust representeerisid antud ajahetke siinuse ja koosinuse formaadis:

$$t_{sin} = \sin\left(\frac{t}{23} * 2\pi\right)$$

$$t_{cos} = \cos\left(\frac{t}{23} * 2\pi\right)$$

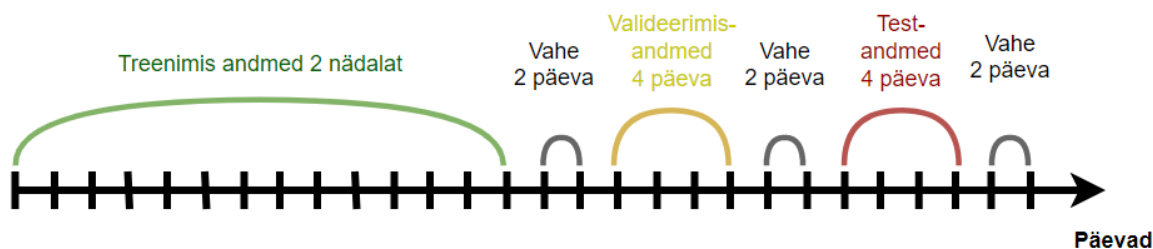
Kokkuvõtteks, üks tunnuste komplekt koosnes 23-st tunnusest, seega ka lõpliku andmepunkti dimensionaalsus oli (6x23) ehk 138 tunnust, millele vastas reaalne tootlikkus y_t . Tähtis on see, et ilmaennustused nende kuue sammu jaoks olid tehtud samal ajahetkel. Näiteks, ilmaennustused olid tehtud kell 11:00 järgmise 48-ks tunniks. Esimese andmepunkti loomisel võtame ilmaennustused alustades kellast 12:00 kuni 17:00, kusjuures sellele vastav sihttunnus y on reaalne tootlikkus kell 17:00. Teise andmepunkti loome nihutades sellist akent ühe tunni võrra nagu näidatud Joonisel 5. Jätkame seda protseduuri kasutades kõik 48 ennustatud tundi. Joonisel võime näha, kuidas “horisont” tunnus kirjeldab seda, kui kaugele me tahame ennustada tootlikkust konkreetse ilmaennustuse hetke suhtes. Peale seda võetakse järgmine ilmaennustuse hetk ning korratakse protseduuri.



Joonis 5. Andmete kokkupanemise meetod.

5.6. Treenimis-, valideerimis- ja testandmed

Päikesepaneelide tootlikkus sõltub aastajast, lisaks tegu on aegreaga, seega andmete jagamine treenimis-, valideerimis- ja testandmeks oli otsustatud teha järgmiseks (Joonis 6): võetakse 2 nädalat mudeli treenimiseks, 2 päeva jäetakse vahele, 4 päeva valideerimiseks, 2 päeva jäetakse vahele, 4 päeva testimiseks, 2 päeva jäetakse vahele, 2 nädalat treenimiseks jne kuni aasta lõpuni. Üks tsükkel koosneb 28st päevast.



Joonis 6. Andmete jagamine treenimis-, valideerimis- ja testimis andmeteks.

Antud jaotus võimaldab treenida/valideerida/testida mudeli võttes arvesse kõik aastaajad. Autokorrelatsiooni graafik näitas, et kaks kõrvuti olevat päeva võivad olla väga sarnased, siis otsustasime, et treening/valideerimis/testimis andmete vahel peavad olema vahed, selleks et minimeerida andmete kattumist, mis vähendab ülesobitamise võimalust. Otsustasime panna vaheks 2 päeva, sest suuremad vahed vähendaksid treenimise andmete kogust.

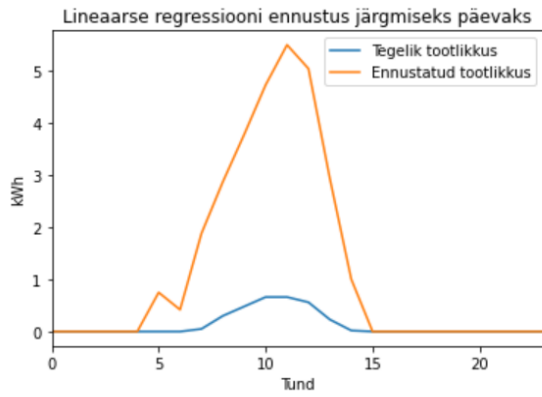
Antud jagamise viisil on kaks probleemi, esiteks me ei kasuta 6/28 andmetest (vahed), teiseks valideerimis- ja testandmetesse võivad juhuslikult sattuda päikseline päevad, kus ennustamine on lihtsam. See võib tuua testimisel ebarealistlikult head tulemust. Seega otsustasime teha mitu jagamist, nimelt kui me alustame jagamist 2 päeva hiljem, siis treenimis-, valideerimis- ja testimisandmed hakkavad olema teistsugused võrreldes esialgse jagamisega. Sellisel moel autor tegi 14 erinevat jagamist (nimetame seda kaustaks). Neid kasutatakse hiljem, kui viiakse läbi lõpliku mudeli testimist. Mudeli ehitamiseks ning tuunimiseks oli kasutatud ainult üks fikseeritud kaust, kus treeningandmed sisaldasid 175 päeva, valideerimisandmed sisaldasid 50 päeva, testimisandmed – 50 päeva ning 74 päeva jäid kasutamata. Andmed katsid perioodi alates 17.01.2022 kuni 31.12.2022 ehk 349 päeva. Igas päevas on 24 tundi ning igal tunnil tehti ilmaennustused järgmiseks 48-ks tunniks. Kokkuvõtteks treeningandmetes oli umbes 200 tuhat andmepunkti, valideerimisandmetes oli 57 tuhat ning testandmetes oli 57 tuhat andmepunkti.

6. Mudeli ehitamine

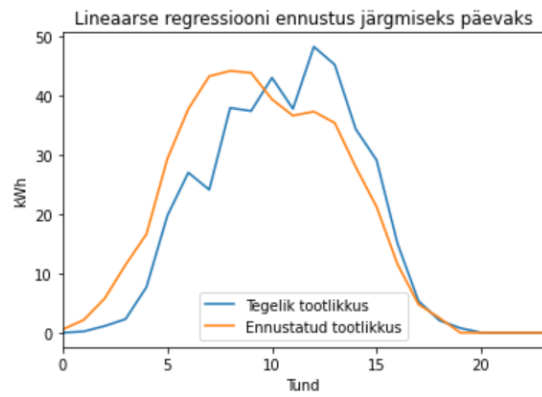
Antud peatükis kirjeldatakse mudelite loomise protsessi. Masinõpe mudeli loomisel esimene samm on luua lihtsa baasmudeli, selleks et aru saada minimaalset võimalikku tulemust, millega hakatakse võrdlema järgmisi mudeleid. Regressiooni ülesande lahendamisel baasiks võib valida erinevaid naiivseid mudeleid nagu keskmise väärtuse ennustamine või lihtsa masinõpe mudeli nagu lineaarne regressioon [18]. Peale baasmudeli võrreldi omavahel XGBoost ja LSTM mudeleid, nagu see oli argumenteeritud uuringus [6]. Kõik eksperimendid olid viidud läbi ühe ja sama kausta peal, selleks et saadud tulemused oleksid võrreldavad.

6.1. Lineaarne regressioon

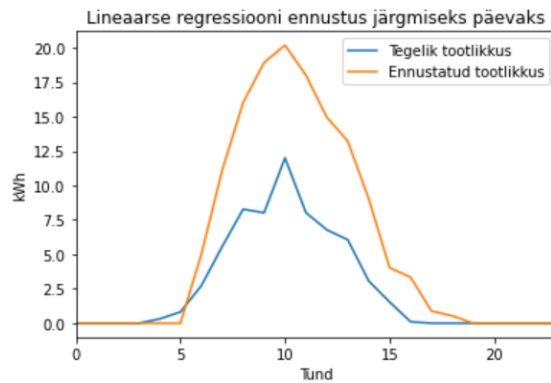
Antud töös baasmudeliks oli valitud lineaarne regressioon vaikimisparameetritega. Kasutati Python'i *sklearn.linear_model* paketti. Mudeli sisendiks on ühe dimensiooniga vektor, mis sisaldab 5 eelmise ning praeguse sammu tunnuste komplekti. Kokku sisend koosnes $(5+1)*23=138$ tunnusest, mis olid kirjeldatud peatükis 5.5. Lineaarse regressiooni MAE treenimisandmetel oli **3.414** kWh ning valideerimisandmetel **3.934** kWh. Joonisel 7 on näidatud mudeli ennustused järgmiseks päevaks testandmete peal.



(a)



(b)



(c)

Joonis 7 (a,b,c). Lineaarse regressiooni mudeli ennustused terveks järgmiseks päevaks.
Näidatud kolm päeva: a) 1. veebruar, b) 22. juuni, c) 14. september.

6.2. XGBoost

Teiseks mudeliks oli valitud XGBoost regressor. Kasutati Python'i `xgboost.XGBRegressor` moodulit. Sisendiks on ühe dimensiooniga vektor, mis sisaldab 5 eelmise ning praeguse sammu tunnuste komplekti. Kokku sisend koosnes $(5+1)*23=138$ tunnusest, mis olid kirjeldatud peatükis 5.5. Mudeli tuuniti valideerimisandmetel kasutades erinevaid parameetreid: `n_estimators`, `eta` ning `max_depth`. Parameeter `n_estimators` määrab puude arvu ansamblis, `eta` on õppimise samm ning `max_depth` määrab maksimaalset puu sügavust [19]. Treenimisel kasutati varase peatumise meetodit (ing. k. *early stopping*), mille abil lõpeb treenimise protsess, kui MAE valideerimisandmetel ei parane 50. epohhi jooksul.

n_estimators	100	200	400	1000	n_estimators	100	200	400	1000
max_depth					max_depth				
20	5.785211	4.383612	3.195527	3.020957	20	4.433062	3.233318	3.067585	3.067585
30	5.828861	4.436471	3.261585	3.108439	30	4.561870	3.411917	3.280515	3.280515
40	5.826997	4.449541	3.320659	3.181943	40	4.536516	3.402146	3.265321	3.265321
50	5.823227	4.432637	3.258241	3.095116	50	4.592035	3.537835	3.424119	3.424119
60	5.816292	4.470633	3.272135	3.121844	60	4.615907	3.542850	3.460514	3.460514
70	5.816292	4.470633	3.307986	3.173732	70	4.615907	3.526410	3.437107	3.437107

(a) (b)

n_estimators	100	200	400	1000
max_depth				
20	3.651086	3.051850	3.051850	3.051850
30	3.728299	3.143352	3.143352	3.143352
40	3.793235	3.242925	3.242925	3.242925
50	3.816292	3.336719	3.336719	3.336719
60	3.835541	3.384372	3.384372	3.384372
70	3.835541	3.366678	3.366678	3.366678

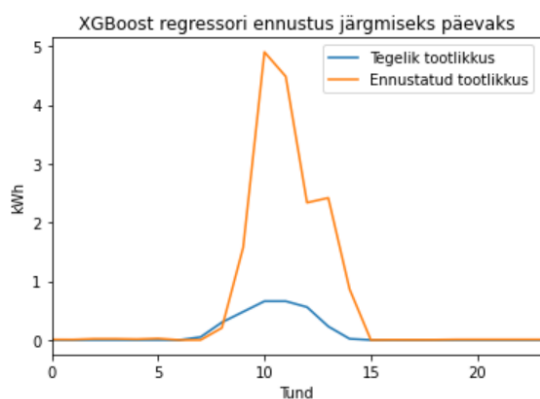
(c)

Joonis 8 (a,b,c). XGBoost mudeli tuunimise tulemus (MAE valideerimis andmetel).

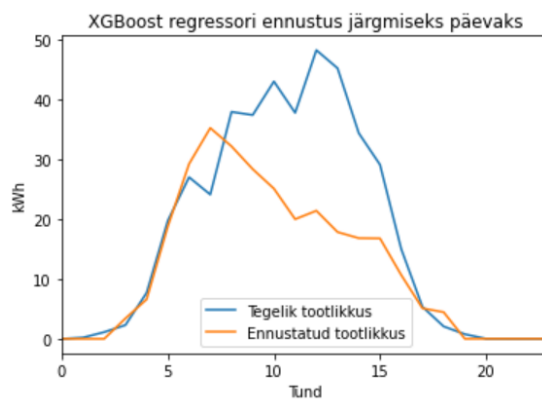
Kolmel alam joonisel oli kasutatud erinev eta parameeter:

a) Ülemisel vasakul joonisel eta=0.1 b) Ülemisel paremal joonisel eta=0.2 c) Alumisel joonisel eta=0.3

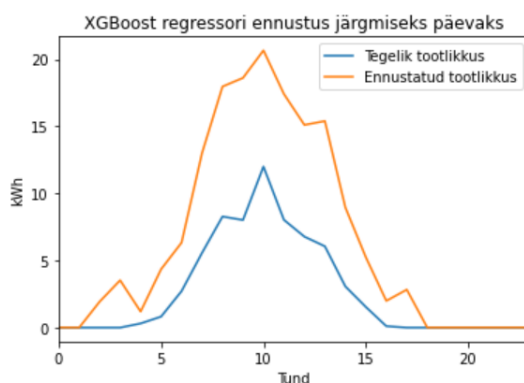
Tuunimine valideerimisandmetel näitas, et paremad parameetrid on $\eta=0.1$, $\max_depth=20$ ning $n_estimators=1000$ (Joonis 8). Nende parameetritega parim MAE valideerimisandmetel oli **3.021** kWh. Joonisel 9 on näidatud mudeli ennustused järgmiseks päevaks testandmete peal.



(a)



(b)



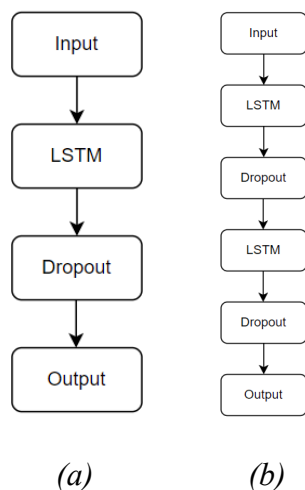
(c)

Joonis 9 (a,b,c). XGBoost regressiooni mudeli ennustused terveks järgmiseks päevaks. Näidatud kolm päeva: a) 1. veebruar, b) 22. juuni, c) 14. september.

6.3. LSTM

Kolmandaks mudeliks oli valitud rekurrentne närvivõrk LSTM. Närvivõrgu ehitamiseks kasutati Python'i *Tensorflow* teeki. Töö eesmärgiks oli ennustada tootlikkust järgmiseks päevaks ning lisaks ka 90% ennustuse intervall ehk 0.05 ja 0.95 kvantiilid. Eesmärgi saavutamiseks LSTM mudel oli ehitatud nii, et närvivõrgu väljundiks oli 3 neuronit, millel olid omad kaofunktsioonid. Kvantiilide leidmiseks olid kasutatud nn “*pinball loss*” funktsioonid ning kokkuvõttes oli 3 väljundneuronit kaofunktsioonidega: *pinball*(0.05), *pinball*(0.5) ehk MAE, *pinball*(0.95). Aktiveerimise funktsiooniks oli valitud ReLu, sest sihttunnus on mittenegatiivne reaalarv. Selleks, et võrrelda seda mudelit XGBoost mudeliga, alguses oli läbi viidud tuunimine. Nimelt, võrreldi kaht arhitektuuri, kus oli kasutatud üks rekurrentne kiht ning kaks rekurrentset kihti (Joonis 10). Treenimisel kasutati varase peatumise meetodit, mille abil lõpeb treenimise protsess, kui MAE valideerimise andmetel ei

parane 5. epohhi jooksul. Kusjuures treenimise samm (ing. k. *learning rate*) oli pandud võrduma 0.0001-ga. Tuuniti neuronite arvu rekurrentsetes kihtides (*LSTM units*) ning “dropout” parameetrit.



Joonis 10. a) LSTM mudeli ühekihiline arhitektuur.
b) LSTM mudeli kahekihiline arhitektuur.

Dropout	0.100000	0.200000	0.300000
First layer units			
20	2.629570	2.597873	2.698735
30	2.663921	2.720747	2.735698
40	2.718025	2.703018	2.710319
50	2.657072	2.638870	2.700931
60	2.712981	2.728766	2.715385

(a)

Dropout	0.100000	0.200000	0.300000	
Second layer units				
10	2.616178	2.669653	2.766223	
20	2.584369	2.761882	2.675196	
30	2.737464	2.574957	2.711990	
40	2.533058	2.587577	2.626685	
50	2.499590	2.616210	2.620205	
60	2.512066	2.522039	2.629471	

(b)

	Dropout	0.100000	0.200000	0.300000
Second layer units				
10	2.670057	2.757229	2.868378	
20	2.442245	2.661599	2.628484	
30	2.590364	2.553509	2.659248	
40	2.504480	2.526992	2.607194	
50	2.633464	2.639379	2.526869	
60	2.752017	2.588789	2.616835	

(c)

	Dropout	0.100000	0.200000	0.300000
Second layer units				
10	2.599407	2.637930	2.771192	
20	2.790731	2.608637	2.773964	
30	2.656947	2.568535	2.530674	
40	2.722613	2.590352	2.645180	
50	2.549803	2.602723	2.502159	
60	2.603574	2.516638	2.602532	

(d)

Joonis 11 (a,b,c,d). LSTM mudeli tuunimise tulemus (MAE valideerimis andmetel)

a) Ühekihilise arhitektuuri tuunimise tulemus

b) Kahekihilise arhitektuuri tuunimise tulemus, kus esimese kihi ühikute arv on 5

c) Kahekihilise arhitektuuri tuunimise tulemus, kus esimese kihi ühikute arv on 10

d) Kahekihilise arhitektuuri tuunimise tulemus, kus esimese kihi ühikute arv on 15

Parimaks osutus kahekihiline arhitektuur järgmiste parameetritega (Joonis 11): esimese kihi ühikute arv 10, teise kihi ühikute arv 20 ning *dropout* 0.1. Mudeli MAE valideerimise andmetel oli **2.442** kWh. LSTM saavutas parimad tulemused valideerimise andmetel, seega edaspidistes eksperimentides kasutati ainult seda mudelit.

6.4. LSTM modifikatsioon

LSTM mudeli loomisel tekkis hüpotees, et ühe pea lisamisel, mille kadu funktsiooniks on MSE, väheneb mudeli viga ehk siht meetriku (MAE) väärtus. Hüpoteesi kontrollimiseks olid ehitatud uued mudelid, mille arhitektuurid erinesid ainult väljundneuronite arvuga ja nende kaofunktsioonidega. Ülejäänud arhitektuur oli sama, mis tuunimise käigus osutus parimaks ehk kahekihiline arhitektuur parameetritega: esimese kihi ühikute arv 10, teise kihi ühikute arv 20 ning *dropout* 0.1. Iga mudel oli treenitud 5 korda ning võrdlemiseks oli võetud keskmine tulemus valideerimise andmetel. Nimelt olid ehitatud mudelid järgmiste väljund neuronitega:

- 1) 3 neuronit kaofunktsioonidega: *pinball*(0.5) ehk MAE, *pinball*(0.05), *pinball*(0.95). Nimetame antud arhitektuuri *A1*-ks.
- 2) 3 neuronit kaofunktsioonidega: MSE, *pinball*(0.05), *pinball*(0.95). Nimetame antud arhitektuuri *A2*-ks.
- 3) 4 neuronit kaofunktsioonidega: *pinball*(0.5) ehk MAE, MSE, *pinball*(0.05), *pinball*(0.95). Antud mudeli treeniti ka erinevate kaaludega MSE kadu jaoks. Nimetame antud arhitektuuri *A3*-ks.

A1 esimene neuron optimeerib MAE-t, mille väljundiks on ennustatud tootlikkus. Teine ja kolmas neuronid optimeerivad *pinball* kaofunktsioonid, mille väljunditeks on ennustatud tootlikkuse 0.05 ja 0.95 kvantiilid. Mudeli ennustatud tootlikkuse keskmine MAE valideerimisandmetel oli **2.5432** kWh.

A2 esimene neuron optimeerib MSE-t, mille väljundiks on ennustatud tootlikkus. Kusjuures mudeli lõpliku vea arvutatakse ikkagi MAE kaudu, sest see on meie sihtmeetrik. Teine ja kolmas neuronid on samasugused nagu *A1*-l. Mudeli ennustatud tootlikkuse keskmine MAE valideerimisandmetel oli **2.7344** kWh.

A3 esimene neuron optimeerib MAE-t, mille väljundiks on ennustatud tootlikkus. Kolmas ja neljas neuronid on samasugused nagu *A1*-l ja *A2*-l. Teine neuron optimeerib MSE-t ning on lisatud selleks, et suurendada mudeli kadu, kui mudel eksib tootlikkuse tipphetkedel.

Mudeli keskmine MAE valideerimisandmetel oli **2.659** kWh. Kuna teine neuron on abistav, siis siinkohal prooviti anda erinevad kaalud (λ) selle neuroni kaofunktsioonile selleks, et uurida kas niisugune arhitektuuri täiendamine võib aidata mudelit, kui vähendame selle neuroni mõju (Tabel 1).

Tabel 1. A3 MAE valideerimis andmetel erinevate kaaludega teise neuroni jaoks.

λ	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
MAE	2.66	2.685	2.741	2.767	2.595	2.612	2.73	2.691	2.588	2.738

Tabel 2. A1, A2 ja A3 arhitektuuride võrdlemine

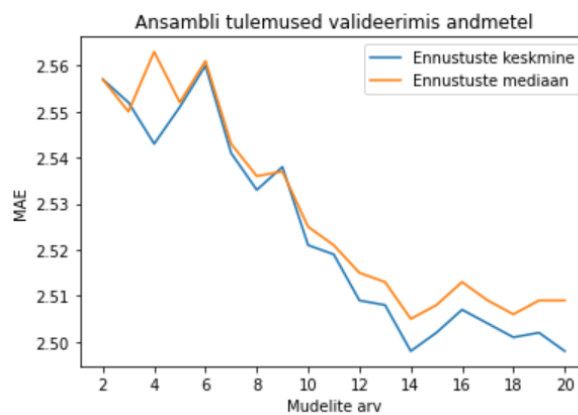
Arhitektuur	MAE
A1	2.5432
A2	2.7344
A3 ($\lambda=1$)	2.659
A3 ($\lambda=0.2$)	2.588

Kokkuvõtteks, neuroni, mille kaofunktsioon on MSE, lisamine ei aita mudelit saavutada paremat MAE-t (Tabel 2), seega arhitektuur, mis oli valitud lõpliku mudeli jaoks, on A1.

6.5. Ansambel

Varasematest uuringutest on teada, et mudelite ansambel väljastab täpsemaid ennustusi võrreldes ühe ainsa mudeliga. Mudelite ansambli tugev külg on mitmekesisus, mida võib saavutada mitmel erineval viisil, näiteks, andmete-, parameetrite- või arhitektuuriline mitmekesisus [20]. Artiklis [21] kasutati kaalutud LSTM ansamblit aegrida prognoosimiseks, kus mitu sarnase arhitektuuriga mudelid väljastasid ennustusi, mis olid hiljem agregeeritud kohandatud kaaludega. Antud töös autor otsustas kasutada sarnast lähenemist, aga lihtsuse mõttes kõik ansambelis olevate mudelite ennustused olid võrdsete kaaludega. Sellel lähenemisel kasutatakse parameetrilist mitmekesisust, mis tähendab, et

luuakse mitu sarnase arhitektuuriga mudelit, mis on treenitud sama andmete peal, aga omavad erinevaid parameetreid. Närvivõrku treenimisel esialgsed kaalud olid juhuslikud, seega ka kõik mudelid peale treenimist ei anna identseid ennustusi. Ansambli ennustusi on vaja agregeerida, selleks et saada üks lõplik väljund. Antud töös kasutati kaht agregeerimisviisi: keskmise võtmine ning mediaani võtmine. Neid kahte võrreldi omavahel, et otsustada lõpliku agregeerimisviisi. Ansambli loomisel on vaja otsustada mitmest mudelist hakkab koosnema ansambel, selleks oli viidud läbi eksperiment, kus võrreldi ansambli keskmist viga valideerimisandmetel erineva arvu mudelitega.



Joonis 12. Ansambli MAE valideerimis andmetel erineva arvu mudelitega.

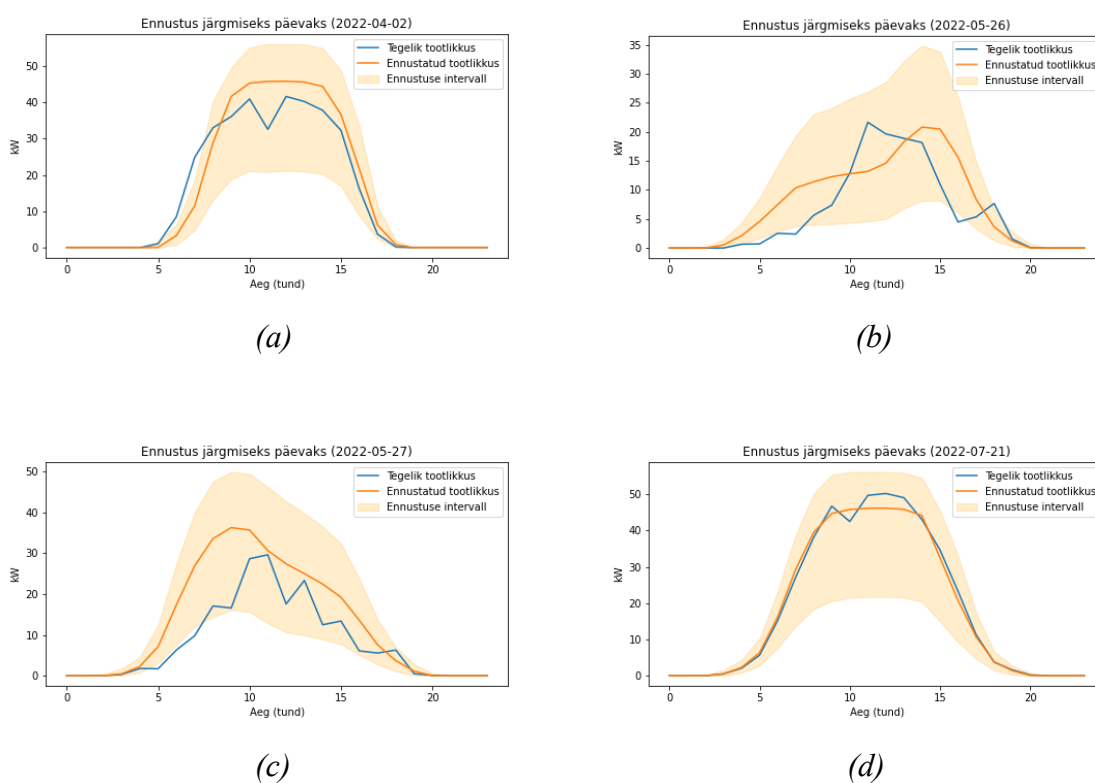
Eksperimentide tulemusena oli otsustatud kasutada agregeerimise viisina ennustuste keskmist ning jääda 14 mudeli juurde ühes ansambelis (Joonis 12), mille MAE valideerimisandmetel oli **2.498 kWh**.

7. Tulemused

7.1. Üldine ja spetsiifiline testid

Mudeli lõplik hindamine oli viidud läbi kahel viisil: üldise- ning spetsiifilise testiga. Mõlemal juhul treeniti ja testiti mudel erinevate kaustade peal ning lõplik tulemus on keskmine MAE üle kõikide kaustade.

Üldisel testil (Lisa 1) mudeli keskmine MAE oli **2.477 kWh**, mis vastab 30% kadule (wMAPE), ning 92.1% juhtudest reaalne tootlikkus jäi ennustusintervalli sisse. Spetsiifilisel testil (Lisa 2) mudeli keskmine MAE oli **2.399 kWh**, mis vastab 29% kadule (wMAPE), ning 91.8% juhtudest reaalne tootlikkus jäi ennustusintervalli sisse. Kuna mudel oli treenitud ennustama 90%-st intervalli, siis kokkuvõttes mudel väljastas natuke laiemat vahemikku kui oodati. Joonisel 13 on näidatud ennustatud tootlikkus koos ennustus intervalliga järgmiseks päevaks neljal erineval päeval testandmetest.

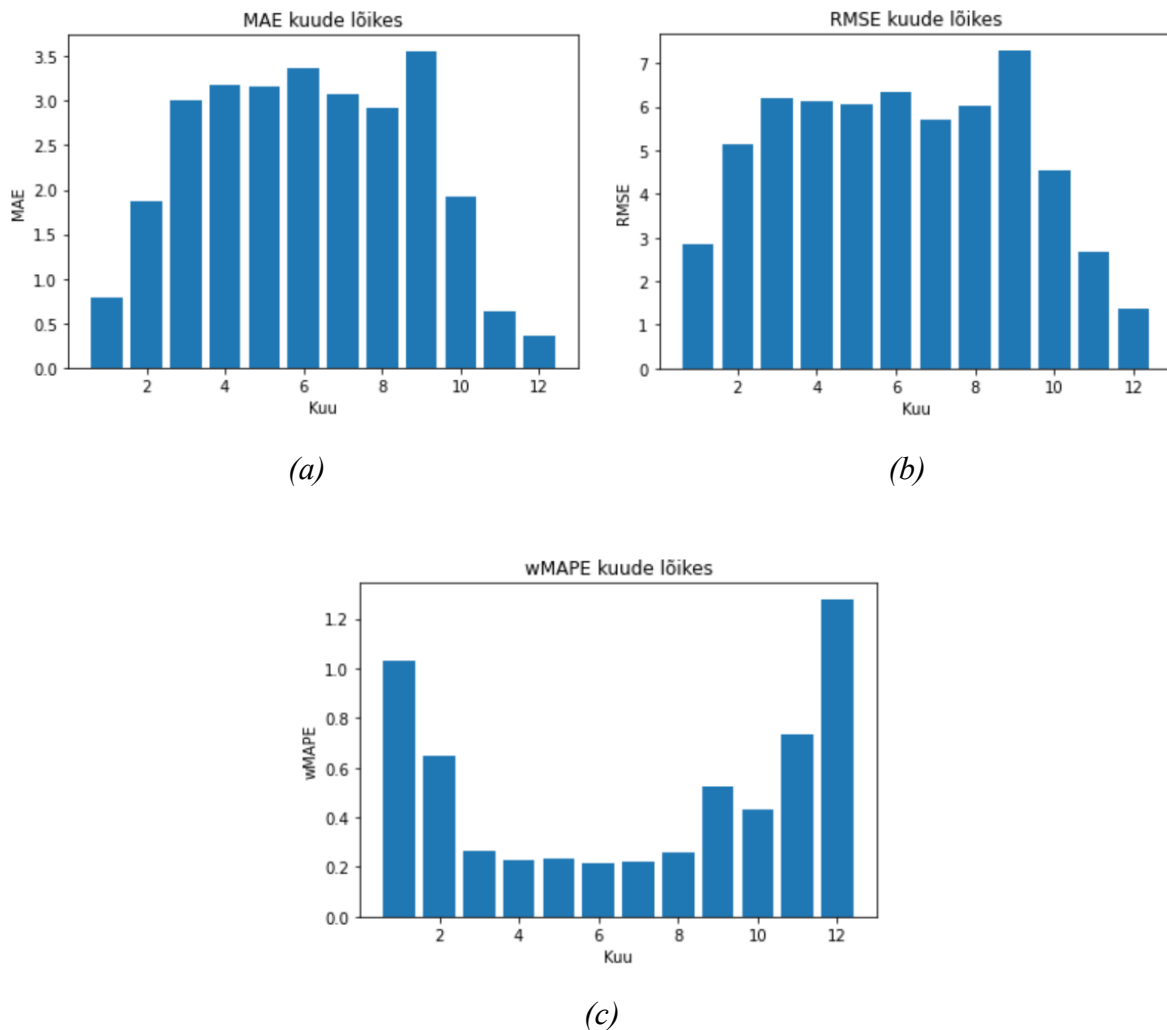


Joonis 13 (a,b,c,d). LSTM ansambli ennustused testandmete peal.

7.2. Aastaajad

Vaadates spetsiifilise testi MAE väärtusi kuude lõikes (Joonis 14a), võime näha, et mudeli absoluutvead on suuremad märtsist septembrini ning väiksemad oktoobrist

veebruari. Kui aga vaadata wMAPE väärtusi kuude lõikes (Joonis 14c), siis olukord on peaaegu vastupidine ehk protsendilised absoluutvead on väiksemad märtsist augustini (väärtused jäävad vahemikku 21.4% kuni 26.3%). Halvimad tulemused on jaanuaris ja detsembris (väärtused on vastavalt 103.3% ja 127.9%). Põhjus seisneb selles, et suvel on rohkem päikest ning tegelikud tootlikkuse väärtused on suuremad kui talvel.



Joonis 14. LSTM ansambli MAE, RMSE ja wMAPE kuude lõikes.

8. Kokkuvõte

Käesoleva magistritöö eesmärgiks oli ehitada masinõpe mudel, mis ennustab päikesepaneelide pargi energia tootlikkust koos ennustus intervalliga järgmiseks päevaks tunni täpsusega.

Töö esimeses etapis valmistati ette andmeid. Andmed olid kogutud perioodil 17.01.2022 kuni 31.12.2022 ning sisaldasid järgmist: päikesepaneelide pargi tootlikkus tunni täpsusega, ilmaennustused järgmiseks 48-ks tunniks iga tunni tagant ning päikese suhteline asukoht igal ajahetkel. Esiteks struktureeriti andmeid nii, et igal tunnil võeti ilmaennustusi järgmiseks 48-ks tunniks ning koostati nendest kõik võimalikud järjestatud kuue sammulised aegread. Siinkohal oli lisati tunnus “horisont”, mis näitas, kui mitu tundi tagasi oli tehtud antud ilmaennustus. Teiseks arvutati igale ajahetkele neli erinevat nurka: päike ja horisondi vaheline nurk, päike praeguse asukoha ja päike kulminatsiooni asukoha vaheline nurk, päike deklinatsiooni nurk ning päike ja päikesepaneelide vaheline nurk. Kolmandaks skaleeriti kõik tunnused MinMax, standardse ja robustse meetoditega.

Töö käigus ehitati ja võrreldi kolme mudelit: lineaarne regressioon, XGBoost regressioon ning LSTM. Valideerimisandmetel saavutati järgmised tulemused: lineaarse regressiooni MAE oli **3.934** kWh, XGBoost'i MAE oli **3.021** kWh ning LSTM'i MAE oli **2.543** kWh. Seejärel loodi LSTM ansambel, mis koosnes 14. mudelist. Ansambli ennustused agregeeriti, võttes ennustuste keskmised väärtused. Valideerimisandmetel oli LSTM ansambli MAE **2.498** kWh.

Lõplik testimine viidi läbi kahel viisil: üldine- ja spetsiifiline test. Üldisel testil saavutas ansambel MAE väärtuseks 2.477 kWh. Spetsiifilisel testil saavutas ansambel MAE väärtuseks **2.399** kWh, mis vastab **29%** (wMAPE) kadule üle terve kalendriaasta. Kuude lõikes saavutas mudel parima wMAPE juunis (**21.4%**) ning halvimal tulemusel olid jaanuaris ja detsembris (vastavalt **103.3%** ja **127.9%**).

Viidatud kirjandus

- [1] Elering. Käsiraamat. <https://elering.ee/elektrituru-kasiraamat/8-lisad/81-lisa-1-moisted>
- [2] Baltic CoBa Imbalance Settlement Rule. Elering AS. September 2021. https://elering.ee/sites/default/files/2022-04/Baltic_CoBA_Imbalance_Settlement_Rules_confirmed.pdf
- [3] A. Nespoli, E. Ogliari, S. Leva, A. M. Pavon, A. Mellit, V. Lughi, A. Dolara. Day-Ahead Photovoltaic Forecasting: A Comparison of the Most Effective Techniques. <https://doi.org/10.3390/en12091621>
- [4] Caicheng Liu, Ming Li, Yunjun Yu, Ziyang Wu, Hai Gong, Feier Cheng. A Review of Multitemporal and Multispatial Scales Photovoltaic Forecasting Methods. 2022. doi: 10.1109/ACCESS.2022.3162206.
- [5] H. Aprillia, H. Yang, C. Huang. Short-Term Photovoltaic Power Forecasting Using a Convolutional Neural Network–Salp Swarm Algorithm. Aprill 2020.
- [6] I. Paliari, A. Karanikola, S. Kotsiantis. A comparison of the optimized LSTM, XGBOOST and ARIMA in Time Series forecasting. Juuli 2021. DOI: 10.1109/IISA52424.2021.9555520
- [7] Eesti Gaas. Elektri pakettid. <https://www.gaas.ee/kodukliendile/elekter/>
- [8] C. Chatfield. Prediction Intervals for Time-Series Forecasting. DOI: 10.1007/978-0-306-47630-3_21
- [9] J. Vermorel. Pinball loss Funtion. Veebruar 2012. <https://www.lokad.com/pinball-loss-function-definition#:~:text=The%20pinball%20loss%20function%2C%20also,forecast%20is%20a%20subtle%20problem.>
- [10] B. Griffiths. Student Perspectives: An Introduction to QGAMs. November 2022. <https://compass.blogs.bristol.ac.uk/author/ben-griffiths/>
- [11] The Norwegian Meteorological Institute. <https://www.met.no/>
- [12] Astronoomia. <http://www.astronoomia.ee/tahistaevas/efemeriidid/>
- [13] F. Kamalov, H. Sulieman. Time series signal recovery methods: comparative study. 2021.
- [14] Soteris A.Kalogirou. Solar energy engineering. 2009.

- [15] Soteris A.Kalogirou. Comprehensive Renewable Energy (Second Edition). 2022.
- [16] Baijayanta Roy. All about feature scaling.
<https://towardsdatascience.com/all-about-feature-scaling-bcc0ad75cb35>
- [17] Swgheta. Introduction to Time Series Forecasting — Part 2 (ARIMA Models). Juuli 2021.
<https://towardsdatascience.com/introduction-to-time-series-forecasting-part-2-arima-models-9f47bf0f476b>
- [18] A. Nair. Baseline Models: Your guide for model building. 4 aprill 2022.
<https://towardsdatascience.com/baseline-models-your-guide-for-model-building-1ec3aa244b8d#:~:text=A%20baseline%20model%20is%20essentially,a%20necessity%20for%20many%20reasons.>
- [19] XGBoost'i dokumentatsioon.
https://xgboost.readthedocs.io/en/stable/python/python_api.html
- [20] Y. Ren, L. Zhang, P. N. Suganthan. Ensemble Classification and Regression-Recent Developments, Applications and Future Directions. Veebruar 2016.
DOI:10.1109/MCI.2015.2471235
- [21] J. Y. Choi, B. Lee. Combining LSTM Network Ensemble via Adaptive Weighting for Improved Time Series Forecasting. August 2018.

Lisad

Kasutatud töövahendid

Töö käigus kasutati Python'i programmeerimiskeelt ning Jupyter notebook'i KubeFlow keskkonnas. Lisaks, oli kasutatud OpenAI poolt loodud keelemudel ChatGPT, et paremini sõnastada töö abstrakti ning tõlkida seda inglise keelde.

Lisa 1. Üldise testi tulemused

Kausta number	MAE	Ennustuse intervalli sattumine
1	2.784	0.918
2	2.694	0.906
3	2.526	0.924
4	2.31	0.929
5	2.167	0.935
6	2.68	0.919
7	2.907	0.909
8	2.487	0.922
9	2.184	0.923
10	2.309	0.918
11	2.358	0.927
12	2.218	0.933
13	2.396	0.917
14	2.662	0.913
Keskmine:	2.477	0.921

Lisa 2. Spetsiifilise testi tulemused

Kausta number	MAE	Ennustuse intervalli sattumine
1	2.717	0.905
2	2.758	0.893
3	2.523	0.927
4	2.113	0.944
5	2.046	0.941
6	2.548	0.918
7	2.628	0.915
8	2.47	0.92
9	2.29	0.909
10	2.244	0.91
11	2.33	0.932
12	2.087	0.933
13	2.279	0.905
14	2.556	0.901
Keskmine:	2.399	0.918

I. Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Kirill Grjaznov,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose
Päikesepaneelide tootlikkuse ennustamine,

mille juhendajad on Meelis Kull ning Janika Aan,

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace
kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks
Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative
Commonsi litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost
reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja
kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.

3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.

4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega
isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Kirill Grjaznov

07.05.2023