

Energiatehnoloogia instituut

**KAUGJAHUTUSE TARBIJATE
JAHUTUSKOORMUSE PROGNOOSIMINE,
KASUTADES MASINÕPPE ALGORITMI
MAGISTRITÖÖ**

Üliõpilane: Kiur Kalme

Üliõpilaskood: 211546MASM

Juhendaja: Kertu Lepiksaar, Doktorant-nooremteadur

Tallinn 2024

AUTORIDEKLARATSIOON

Olen koostanud lõputöö iseseisvalt.

Lõputöö alusel ei ole varem kutse- või teaduskraadi või inseneridiplomit taotletud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on viidatud.

Autor: Kiur Kalme

/ allkirjastatud digitaalselt /

Töö vastab magistritööle esitatud nõuetele

Juhendaja: Kertu Lepiksaar

/ allkirjastatud digitaalselt /

Kaitsmisele lubatud

".....".....2024

Kaitsmiskomisjoni esimees

/ allkirjastatud digitaalselt /

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

Mina, Kiur Kalme (15.06.1999),

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Kaugjahutuse tarbijate võimsuse prognoosimine, kasutades masinõppe algoritmi“, mille juhendaja on Kertu Lepiksaar,
 - 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.

¹*Lihtlitsents ei kehti juurdepääsupiirangu kehtivuse ajal, välja arvatud ülikooli õigus lõputööd reprodutseerida üksnes säilitamise eesmärgil.*

/ allkirjastatud digitaalselt /

LÕPUTÖÖ ÜLESANNE

Üliõpilane: Kiur Kalme, 211546MASM

Õppekava, peaariala: MASM02/18 Energiatehnoloogia ja soojusenergeetika

Juhendaja(d): Kertu Lepiksaar, Doktorant-nooremteadur, +3726203908

Lõputöö teema:

Kaugjahutuse tarbijate jahutuskoormuse prognoosimine, kasutades masinõppe algoritmi

Prediction of district cooling consumers' cooling load using machine learning

Lõputöö põhieesmärgid:

1. Koostada jahutuskoormuse ennustus, kasutades regressioonanalüüsi
2. Koostada jahutuskoormuse ennustus, kasutades juhendatud masinõpet
3. Võrrelda regressioonanalüüsi ning juhendatud masinõppe ennustusi ning analüüsida tulemusi

Lõputöö etapid ja ajakava:

Nr	Ülesande kirjeldus	Tähtaeg
1.	Ülesande püstitus, magistritöö struktuuri loomine	01.02.2024
2.	Algandmete kogumine ja töötlemine ettevõttest AS Utilitas Tallinn	01.03.2024
3.	Regressioonanalüüsi läbiviimine ning jahutuskoormuse ennustuse tegemine	15.03.2024
4.	Juhendatud masinõppe läbiviimine ning jahutuskoormuse ennustuse tegemine	15.04.2024
5.	Ennustuste võrdlus ning järelduste tegemine	01.05.2024
6.	Vormistuse ja tekstinõuete kontroll	19.05.2024
7.	Magistritöö esitamine kaitsmiseks	20.05.2024

Töö keel: eesti keel

Lõputöö esitamise tähtaeg: 20. mai 2024

Üliõpilane: Kiur Kalme

/ allkirjastatud digitaalselt /

Juhendaja: Kertu Lepiksaar

/ allkirjastatud digitaalselt /

Programmijuht: Alar Konist

/ allkirjastatud digitaalselt /

SISUKORD

JOONISTE SISUKORD	6
TABELITE SISUKORD	8
EESSÕNA	9
LÜHENDITE JA TÄHISTE LOETELU	10
SISSEJUHATUS	11
1. KAUGJAHUTUS	13
1.1 Hoonete sisekliima	13
1.2 Hoone jahutus lokaaljahutusega	14
1.3 Hoone jahutus kaugjahutusega	15
1.4 Kaugjahutussüsteem	16
2. KAUGJAHUTUSE VÕIMSUSKOORMUSE ENNUSTAMINE REGRESSIOONANALÜÜSIGA	18
2.1 Tarbija jahutuskooormuse sõltuvus välistemperatuurist	20
2.2 Tarbija jahutuskooormuse sõltuvus jahutuskraadtundidest	22
2.3 Tarbija jahutuskooormuse ennustamine jahutuskraadtundidest sõltavana	25
3. MASINÕPE	29
3.1 Juhendatud masinõpe	31
3.2 Masinõppe algoritmi teegi valik	33
3.3 Otsustusmetsa mudel	34
3.4 Otsustusmetsa mudeli puude õpetamine	36
4. KAUGJAHUTUSE VÕIMSUSKOORMUSE ENNUSTAMINE MASINÕPPEGA	40
4.1 Sisendparameetrite valik hoonete mudelite jaoks	41
4.2 Hoonete mudelite treeningparameetrid	41
4.3 Hoonete mudelite treenimine	43
4.4 Sisendparameetrite tähtsus mudelis	45
4.5 Treenitud mudeli kasutamine ennustuste tegemiseks	46
5. REGRESSIOONANALÜÜSI JA JUHENDATUD MASINÕPPE VÕRDlus JA TULEMUSED	49
5.1 Järeldused	51
KOKKUVÕTE	53
SUMMARY	55
KASUTATUD KIRJANDUSE LOETELU	57
LISA 1	60
LISA 2	69
LISA 3	70

JOONISTE SISUKORD

Joonis 1. Jahutussüsteemide liigutus külmakandjate järgi	15
Joonis 2. Kaugjahutussüsteemi tööpõhimõte	16
Joonis 3. Hoone A 2022. aasta jahutuse koormusgraafik	18
Joonis 4. Hoone B 2022. aasta jahutuse koormusgraafik	19
Joonis 5. Hoone C 2022. aasta jahutuse koormusgraafik	19
Joonis 6. Hoone A 2022. aasta tarbimisvõimsuse sõltuvus välisõhu temperatuurist ...	21
Joonis 7. Hoone B 2022. aasta tarbimisvõimsuse sõltuvus välisõhu temperatuurist ...	21
Joonis 8. Hoone C 2022. aasta tarbimisvõimsuse sõltuvus välisõhu temperatuurist ...	21
Joonis 9. Hoone A optimaalne tasakaalu temperatuur.....	23
Joonis 10. Hoone A 2022. aasta tarbimisvõimsuse sõltuvus jahutuskraadtundidest ...	23
Joonis 11. Hoone B optimaalne tasakaalu temperatuur.....	23
Joonis 12. Hoone B 2022. aasta tarbimisvõimsuse sõltuvus jahutuskraadtundidest ...	24
Joonis 13. Hoone C optimaalne tasakaalu temperatuur.....	24
Joonis 14. Hoone C 2022. aasta tarbimisvõimsuse sõltuvus jahutuskraadtundidest ...	24
Joonis 15. Hoone A regressioonanalüüsi meetodil leitud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskooormusega	26
Joonis 16. Hoone B regressioonanalüüsi meetodil leitud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskooormusega	26
Joonis 17. Hoone C regressioonanalüüsi meetodil leitud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskooormusega	27
Joonis 18. Masinõppe protsess	29
Joonis 19. Masinõppe meetodite ja peamiste tehnikate klassifitseerimine.....	30
Joonis 20. Regulariseerimistermini ja treeningkaotuse funktsiooni illustreeriv näide ..	32
Joonis 21. CART mudeli lihtsa ülesande näide.....	34
Joonis 22. Kahe otsustuspuu ennustuskoori summeerimine	35
Joonis 23. Parameetrite mõju determinatsiooni kordaja suurenemisele.....	40
Joonis 24. Hoone A üks otsustuspuu mudeli treenimisel.....	43
Joonis 25. Hoone A masinõppe mudeli treeningtsükkel 2022. aasta andmete põhjal..	44
Joonis 26. Hoonete A, B ja C sisendparameetrite tähtsus mudelis	45
Joonis 27. Hoone A juhendatud masinõppe meetodil leitud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskooormusega	46
Joonis 28. Hoone B juhendatud masinõppe meetodil leitud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskooormusega	46
Joonis 29. Hoone C juhendatud masinõppe meetodil leitud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskooormusega	47

Joonis 30. Hoone A kaugjahutuse tarbimiskoormuse ennustused regressioonanalüüsi ja masinõppe meetoditega, võrrelduna tegelikkusega	49
Joonis 31. Hoone B kaugjahutuse tarbimiskoormuse ennustused regressioonanalüüsi ja masinõppe meetoditega, võrrelduna tegelikkusega	49
Joonis 32. Hoone C kaugjahutuse tarbimiskoormuse ennustused regressioonanalüüsi ja masinõppe meetoditega, võrrelduna tegelikkusega	50

TABELITE SISUKORD

Tabel 1. Hoonete tasakaalu temperatuurid ja determinatsioonikordajad	25
Tabel 2. Valitud 2023. aasta nädalad, mille jooksul töös valitud hoonete jahutuskoormust ennustatakse.....	26
Tabel 3 Regressioonmeetodil leitud ennustusväärtuste ruutkeskmised vead	28
Tabel 4. Erinevate juhendatud masinõppe algoritmi teekide hindamine	33
Tabel 5. Parimad XGBoost treeningparameetrite vaikeväärtused.....	42
Tabel 6. Hoonete A, B ja C treenitud masinõppe mudelite 2022. aasta testandmete keskmised ruutvead.....	44
Tabel 7. Hoonete A, B ja C treenitud mudeli sobivus 2022. aasta andmetele	44
Tabel 8. Juhendatud masinõppe meetodil leitud ennustusväärtuste ruutkeskmised vead	48
Tabel 9. Kahe ennustusmeetodi ruutkeskmiste vigade võrdlus erinevatel perioodidel	50

EESSÕNA

Käesoleva magistritöö teema sõnastati töö autori Kiur Kalme algatusel ning täpsustati koostöös lõputöö juhendaja Kertu Lepiksaarega. Teema on aktuaalne tõhusa kaugjahutuse arendamise kontekstis.

Töö jaoks vajalikud sisendandmed, mida on kasutatud regressioonanalüüsiks ning juhendatud masinõppeks, on kogutud AS Utilitas Tallinna andmebaasist ning Keskkonnaagentuuri Tallinn-Harku ilmajaama ajaloolistest ilmaandmetest. Kuna sisendandmeteks on ka Tallinna kaugjahutusvõrguga liitunud klientide andmed, siis andmekaitse eesmärgil hoonete aadresse ei avaldada, vaid nimetatakse tähestiku järjekorras. AS Utilitas Tallinna poolt on sisendandmete õiguspärase kasutamise järelevalve ning sisendandmete väljastaja Tanel Kirs, kaugjahutuse osakonna juhataja.

Töö raames on läbi viidud regressioonanalüüsi ennustused 2023. aasta valitud ajaperioodide kohta; juhendatud masinõppe teegi ehk infoobjektide kogumi XGBoost regressiooni meetodil kaugjahutuse jahutuskoormuse ennustused 2023. aasta valitud ajaperioodide kohta; viimaks võrreldi saadud tulemusi tegelike jahutuskoormuse andmetega.

Koostatud võrdluse eesmärk on hinnata erinevate ennustusmeetodite efektiivsust ning teha järeldusi ühe või teise meetodi eelistest ja puudustest. Läbiviidud analüüsi põhjal on võimalik hinnata, kas kaugjahutuse kliendi hoone jahutuskoormuse ennustamine on võimalik, kasutades suhteliselt väheseid andmeid ning saavutades mõistlik täpsus, ning kuidas oleks võimalik suurendada täpsust just masinõppe meetodi puhul.

Töö autor soovib Raestipendiumi määramise eest tänada Tallinna Strateegiakeskust ja Tallinna Linnavalitsuse õppe- ja teaduskoostöö komisjoni ning „Puhta energia magistriõppe stipendiumi“ määramise eest tänada ettevõtet AS Utilitas Tallinn.

Kaugjahutus, regressioonanalüüs, juhendatud masinõppe, magistritöö

LÜHENDITE JA TÄHISTE LOETELU

CDH	jahutuskraadtund
CART	klassifikatsiooni ja regressiooni otsustuspuud (<i>classification and regression trees</i>)
DBT	kuivtermomeetri näit (<i>dry bulb temperature</i>)
F	hulk, millesse kuuluvad kõik võimalikud CART funktsioonid f_k , mis on võimalikud lähteandmete x_i korral
f_k	otsustuspuu funktsiooni valem, mis kuulub hulka F
K	otsustuspuude arv otsustusmetsas
L	treeningkaotuse funktsioon
$obj(\theta)$	eesmärgifunktsioon
θ	koefitsient või kaal muutujale ennustusväärtuse leidmiseks
Ω	regulariseerimisterminite funktsioon
RMSE	ruutkeskmine viga
R^2	determinatsioonikordaja
S/V	hoone pindala ja ruumala suhe
T	otsustuspuu lehtede arv
w	otsustuspuu lehtede ennustuskooride vektor
$\omega(f_k)$	treenitud otsustusmetsa keerukus
x_i	treeningandmestiku muutuja, näiteks välistemperatuur
y_i	otsitav väljund või väärtus, tarbitav jahutusvõimsus
\hat{y}_i	ennustusväärtus masinõppe mudeli poolt otsitavale väljundile

SISSEJUHATUS

Kaugjahutussüsteem koosneb üldistatult kolmest osast: kaugjahutusjaamad, kaugjahutusvõrk ning kaugjahutussõlmed, mis asuvad kaugjahutusega liitunud klientide hoonetes. Kaugjahutusega liitunud kliendid loovutavad soojuslevi vaatenurgast liigse soojuse kaugjahutussüsteemi ehk teisisõnu tarbija hoone vaatest tarbitakse jahutusenergiat. Kaugjahutuse süsteemide levikut soodustas 2018. aasta Euroopa Liidu taastuvenergia direktiiv EU/2018/2001, milles tutvustati esmakordselt spetsiifilisi eesmärke taastuvenergia arenguks kütte- ja jahutussektorites [1]. Uuendatud Euroopa Liidu direktiivi EU/2023/2413 kohaselt on artiklis 24 suurendatud taastuvenergia eesmärke veelgi [2] ning tugevalt ergutatakse liikmesriike enda kaugenergia süsteeme laiendama.

Kaugjahutussüsteemid on arenemas Eestis kolmes linnas: Tartus ja Pärnus ettevõtte Gren eestvedamisel [3] ning käesoleva töö sisendandmete allikaks olevas Tallinnas, mis areneb ettevõtte AS Utilitas Tallinn eestvedamisel. Tallinn on sõnastanud säästva energiamajanduse ja kliimamuutustega kohanemise kavas 2030, et kaugjahutuse võrgu rajamisega peab alustama ning tänu jahutusvajaduse kasvu prognoosile on muutumas aina olulisemaks ligipääs kaugjahutusele [4].

Kaugjahutussüsteemi operaator peab tagama klientidega kokkulepitud teenuse taseme ehk jahutusenergia olemasolu siis, kui kliendid seda vajavad. Seetõttu on kaugjahutussüsteemi operaatoril õigustatud huvi saada iga hetk tarbimisandmeid jahutussõlmes kaugjahutussüsteemi poolelt, et oma süsteemi toimimiseks vajalikke andmeid koguda ja süsteemi seeläbi parendada ning toota vastavas koguses jahutusenergiat. Kaugjahutussüsteemi operaatoril pole aga õigust saada igal ajahetkel infot kliendi sekundaarse ehk majasisese tehnosüsteemi, hoone konstruktsiooni ja toimimise kohta. See tekitab probleemi, sest hoone jahutuskooormuse ennustamiseks on kaugjahutussüsteemi valdajal kasutada ainult ilmaandmed, eelnev jahutusenergia tarbimise ajalugu ning hetkeandmed.

Kasutades saadaolevaid andmeid, saab kaugjahutussüsteemi operaator kasutada regressiooni, aegrea (inglise keeles *time-series*) või erinevate masinõppe analüüsimeetodeid, et ennustada jahutuskooormust kaugjahutussüsteemis. Käesolevas töös on võrreldud regressioonanalüüsi ja juhendatud masinõppe teegi ehk infoobjektide kogumi XGBoost regressiooni meetodeid.

Jahutuskoormuse ennustamine annab kaugjahutussüsteemi valdajale võimaluse planeerida oma tootmist ette ning arvestada jahutusenergia tootmiseks vajaliku elektrienergia hinnaga, vajadusel jahutades kaugjahutusvõrku eelnevalt mõned kraadid külmemaks või vastupidi – mitte toota kõrge elektri hinnaga, kui nõudlust antud ajahetkel on vähe. Selline tootmise nihutamine aitab vähendada koormust elektrisüsteemile ning ühtlustab kaugjahutussüsteemi piirkonna elektritarvet [5].

Jahutuskoormuse ennustuse rakendamiseks peab kaugjahutusvõrgu valdajal olema väga hea ülevaade ja arusaam torustiku hüdraulikast, sest kaugjahutussüsteemil on suur inerts ning muutused ei jõua tarbijateni koheselt. Käesolevas töös ei ole hüdraulika teemat käsitletud.

Käesolev magistritöö koosneb viiest osast. Esimeses osas antakse ülevaade kaugjahutuse süsteemist ning süsteemi jahutuskoormust mõjutavatest teguritest, millest olulisimad on eelkõige hoonete sisekliima nõuded ja välisõhu temperatuur.

Töö teises osas antakse ülevaade kaugjahutuse tarbija jahutuskoormuse ja välistemperatuuri sõltuvusest. Seejärel antakse ülevaade jahutuse valdkonnas kasutatava jahutuskraadtunni mõistest ning viimaks kirjeldatakse kaugjahutuse võimsuskoormuse ennustamist regressioonanalüüsiga, kasutades eelnevalt kirjeldatud jahutuskraadtunde.

Töö kolmandas osas käsitletakse masinõppe valdkonna tutvustust, kuidas juhendatud masinõppe töötab ning kuidas otsustusmetsa (inglise keeles *decision tree ensemble*) mudelit õpetatakse.

Töö neljandas osas kirjeldatakse detailselt juhendatud masinõppe teegi XGBoost mudeli treenimist ning jahutuskoormuse ennustuse tegemist.

Töö viiendas osas vaadeldakse ja võrreldakse töö teise ja neljanda osa tulemusi, analüüsitakse erinevusi ning tehakse järeldused kaugjahutuse operaatori vaatest.

1. KAUGJAHUTUS

Peatüki eesmärk on anda ülevaade kaugjahutuse süsteemist ning jahutuskoormust mõjutavatest teguritest, millest olulisimad on eelkõige hoonete sisekliima nõuded ja välisõhu temperatuur.

Hoonete sisekliima nõuded on olulised, kuna need määravad, millises ulatuses ja millist tüüpi jahutust vajavad hooned. Erinevad hoonetüübid võivad omada erinevaid sisekliima nõudeid. Näiteks hoone energiatõhususe miinimumnõuete määruse kohaselt korterelamu suvise ruumitemperatuuri nõuded loetakse täidetuks, kui elamus ei ületata ajavahemikul 1. juunist 31. augustini jahutuse seadeväärtust 27°C rohkem kui 150 kraadtunni võrra. Ärihoone puhul aga samal perioodil ei tohi jahutuse seadeväärtust, mis on 25°C, ületada rohkem kui 100 kraadtundi [6].

Teine oluline mõjutegur on välisõhu temperatuur. Välisõhu temperatuur mõjutab otseselt kaugjahutussüsteemi efektiivsust ja süsteemi koormust. Kõrgemad välisõhu temperatuurid suurendavad jahutuskoormust, kuna siseruumides oleva temperatuuri hoidmiseks tuleb rohkem jahutusenergiat tarbida.

1.1 Hoonete sisekliima

Eesti Riigi Kinnisvara Aktsiaseltsi juhendi kohaselt [7] mõjutab sisekliima otseselt inimeste töövõimet ja tervist. Seetõttu on äärmiselt oluline tagada töökohtadel ja eluruumides tervislik ja mugav sisekliima. Sisekliima parameetrid on:

- õhu temperatuur;
- operatiivne temperatuur;
- suhteline niiskus;
- õhu liikumiskiirus
- õhu puhtus;
- müra;
- valgustatus jm.

Kõige tähtsam neist on käesoleva töö mõistes soojuslik mugavus. Vastavalt energiatõhususe miinimumnõuete määrusele [6] kasutatakse vastavalt ruumi otstarbest, kuid samas pidevalt viibitavates siseruumides kütteperioodil üldjuhtudel 21°C piirtemperatuuri, alla mille hoone sisetemperatuur ei tohi minna.

Jahutusperioodil ehk suvisel ajal ei tohi hoonete sisetemperatuur ületada mittelehuhoonetes 25°C piirtemperatuuri. Vastavalt määrusele ei tohi elamuhooned ületada piirtemperatuuri 150 kraadtundi ning mitte-elamuhoonetes 100 kraadtundi vahemikus 1. juuni kuni 31. august. [6] Kui hoonesse projekteeritakse ja ehitatakse jahutussüsteem, ei ole vaja teha suvise ruumitemperatuuri arvutust.

Jahutuskoormus tuleb arvutada dünaamilise arvutusprogrammiga, mis vastab Majandus- ja taristuministri 5. juuni 2015. a määruses nr 58 „Hoone energiatõhususe arvutamise meetodika“ § 29 nõuetele [8].

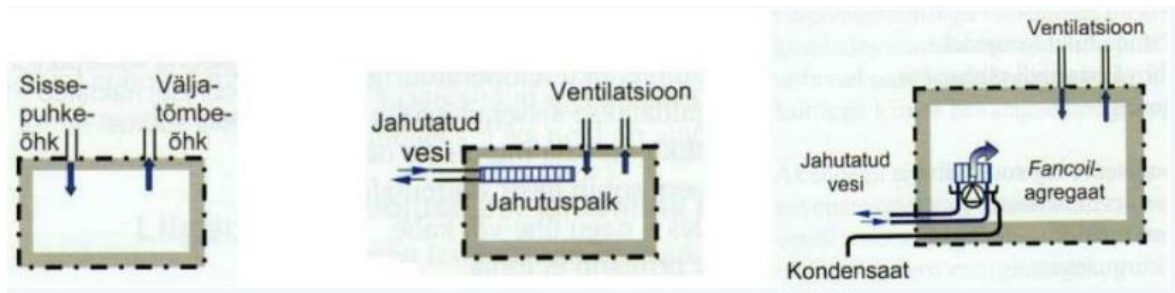
Hoone jahutamise põhieesmärgiks on vältida ruumi temperatuuri tõusmist üle piirtemperatuuri, ehk tagada inimestele soojuslik mugavus ja/või kindlustada tehnoloogilise protsessi jaoks vajalikud tingimused. Ruumi temperatuuri tõusmine välisõhu temperatuurist kõrgemaks on eelkõige põhjustatud kütteta ruumis eralduvast vabasoojusest, kuna see on suurem ruumi soojuskadudest. Mõned näited passiivsetest jahutuse lahendustest [9] on:

- Paigutuse valik ilmakaarte suhtes – päike peaks paistma võimalikult vähesele akna pinnale päeva jooksul;
- Optimaalne soojusisolatsioon, et vältida ülekuumenemist suvel;
- Päikesekiirguse mõju vähendavad päikesekatted – kardinad, välimised päikesekatted;
- Akende päikese läbivusteguri vähendamine – spetsiaalsed klaasid, kiled jne;
- Ruumist tuleneva soojuseralduse (valgustus, elektriseadmed) vähendamine ning energiatõhusamate seadmete kasutamine;
- Öisel ajal ventilatsiooni kaudu jahedama välisõhu ära kasutamine hoone maha jahutamiseks.

Kui aga selgub, et passiivsetest jahutuse lahendustest ei piisa, et tagada hoonete vastavust energiatõhususe miinimumnõuete määrusele [6], siis tuleb kasutada jahutusseadmeid. Hoonele vajalikku jahutust võib toota nii lokaalselt kui tsentraalselt ehk kaugjahutusena [10].

1.2 Hoone jahutus lokaaljahutusega

Lokaaljahutus tähendab seda, et jahutussüsteem on projekteeritud ja ehitatud hoonele selliselt, et külmaallikaks on külmamasin ning külmakandjaks võib olla õhk, vedelik või külmaaine (joonis 1).



Joonis 1. Jahutussüsteemide liigutus külmakandjate järgi [9]

Jahutussüsteem võib lokaaljahutuse puhul olla ehitatud tervele hoonele ühtselt ehk on tsentraliseeritud (vee või külmaaine kaudu) või on igas hoone osas jahutus lahendatud erinevalt ehk detsentraliseeritud. Arendatava kaugjahutussüsteemiga saavad eelkõige liituda need hooned, mille hoone jahutussüsteem on ehitatud terve hoone teenindamiseks [9].

Õhu kui külmakandja puhul transporditakse jahe õhk ventilatsiooniga või konditsioneeriga ruumidesse ning soe õhk eemaldatakse ruumist ventilatsiooni väljatõmbesüsteemiga. Õhu kui külmakandja puhul võib iga toa jahutussüsteem olla ehitatud erinevalt või olla integreeritud ühtsesse hoone jahutussüsteemi [9].

Vedeliku kui külmakandja puhul transporditakse külm vesi ruumides paiknevatesse vesi-õhk soojusvahetitesse (milleks on *fan coil*-agregaat ehk konvektorpuhurid või jahutuspaigid), kus soe õhk suunatakse üle külma soojusvahetuspinna ning soe vesi suunatakse tagasi jahutusseadme juurde [9].

1.3 Hoone jahutus kaugjahutusega

Kaugjahutus on sama põhimõttega nagu kaugküte, kuid vastupidiselt kaugjahutus mitte ei vii sooja tarbijani, vaid annab tarbijale väljundi sooja ära juhtimiseks. Kaugjahutust kasutatakse peamiselt büroo-, kaubandus- ja ärihoonetes.

Kaugjahutusega hoone jahutamisel peab hoonel olema tsentraalne süsteem, mis on võimeline ära andma hoonest liigset soojust kaugjahutussüsteemi läbi tsentraalsete kaugjahutussõlme(de). Lokaaljahutusega võrreldes väheneb hoone elektritarbe vajadus kaugjahutuse kui külmaallika kasutamisega, kuna hoonele vajalik jahutusenergia toodetakse kaugjahutusjaamas, mitte hoones endas ning seetõttu ka ööpäevane elektritarve on stabiilsem [9].

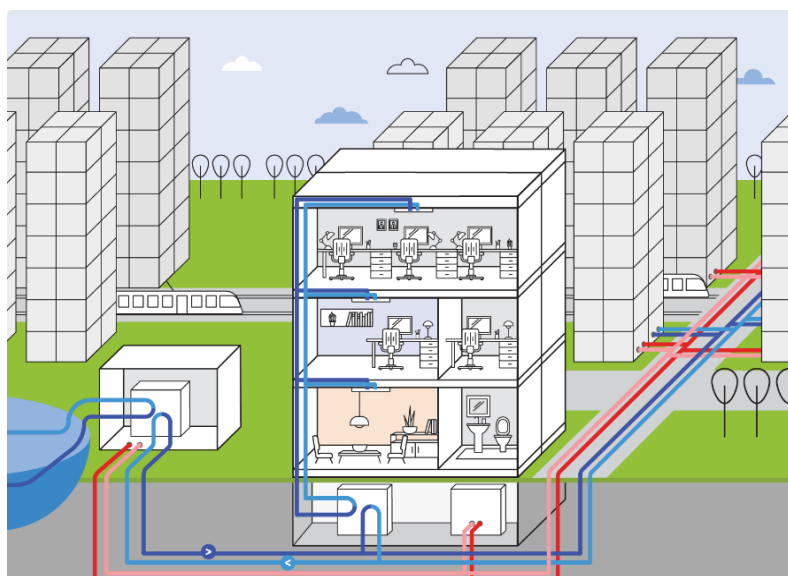
Kaugjahutus on sobilik lahendus tingimustes, kus jahutust vajavaid hooneid on ühel väiksemal alal palju, nagu näiteks linnades. Sobivate tingimuste olemasolul suudab kaugjahutussüsteem kasutada ära lokaaljahutusest palju efektiivsemaid looduslikult esinevaid jahutusallikaid, nagu järved, jõed ja mered, ning kokkuvõttes olla aasta lõikes tõhusam kui igas hoones eraldi jahutusseadmete kasutamine.

Samuti aitab kaugjahutussüsteemi kasutamine linnades vältida soojussaare efekti tekke [11]. Soojussaare efekt on nähtus, mil linna temperatuur on jahutusseadmete heitsoojuse atmosfääri paiskamise tõttu jahutusperioodil kõrgem kui seda linna ümbritsevate alade temperatuur.

Kopenhaageni energiatõhususe keskuse juhendmaterjalide sõnul võib kaugjahutussüsteemi ülesehitamise tagajärjel väheneda linnas suurima elektritarbe perioodidel elektrisüsteemi koormus 30 kuni 35% [12]. See tähendab, et linnad peavad vähem investeerima oma elektrisüsteemi vabaneva elektrilise koormuse tõttu. See saab toimuda tänu tõhusamale jahutusenergia tootmisele ning kaugjahutussüsteemi inertsusest tuleneva jahutusenergia tootmise ajalise nihutamise võimaluse tõttu [13].

1.4 Kaugjahutussüsteem

Kaugjahutus on tsentraalne viis toota hoonete jahutamiseks vajalikku jahutusenergiat. Harilikult on kaugjahutuse tarbijad hotellid, büroohooned, kaubanduskeskused ja muud jahutuse võimekusega varustatud elu- ja ärihooned. Kaugjahutussüsteem koosneb kolmest põhikomponendist: kaugjahutusjaam, kaugjahutusvõrk ning kaugjahutusvõrku ühendatud tarbijad (joonis 2).



Joonis 2. Kaugjahutussüsteemi tööpõhimõte [14]

Kaugjahutusjaamas jahutatakse kaugjahutusvõrgus ringlevat vett järgnevalt [15]:

- Looduslikust allikast ehk vabajahutusega;
- elektriliste kompressorjahutitega ja soojuspumpadega;
- absorptsioonjahutitega, mis kasutavad jahutuse tootmiseks heitsoojust (näiteks kaugkütte süsteemist koostootmisjaamade heitsoojust).

Kaugjahutusjaamas toodetud jahutusenergia viiakse tarbijateni läbi kaugjahutusvõrgu, mis on sarnane kaugküttesüsteemiga. Kaugjahutusvõrgus on pealevoolu ehk tarbijani jõudva vee temperatuur harilikult 4 kuni 7°C ning tagasivoolu vee temperatuur 12 kuni 17°C. Soojem tagasivoolu vesi suunatakse kaugjahutusjaama, kus see uuesti maha jahutatakse [15].

Kaugjahutus vähendab suvisel ajal linnades esinevat soojussaare efekti, kus tumedaid pindu soojendab päike rohkem, millele annab hoogu juurde lokaalsete jahutusseadmete poolt heitsoojuse õhku paiskamine. Lokaalseadmete asendamine kaugjahutusega viib heitsoojuse hoonetest eemale kaugjahutusjaamadesse, mis suunavad tekkinud soojuse veekogudesse, kaugküttesse või õhku. Kaugjahutussüsteemid vähendavad ka linnade elektrivajadust tipukoormuste ajal [5].

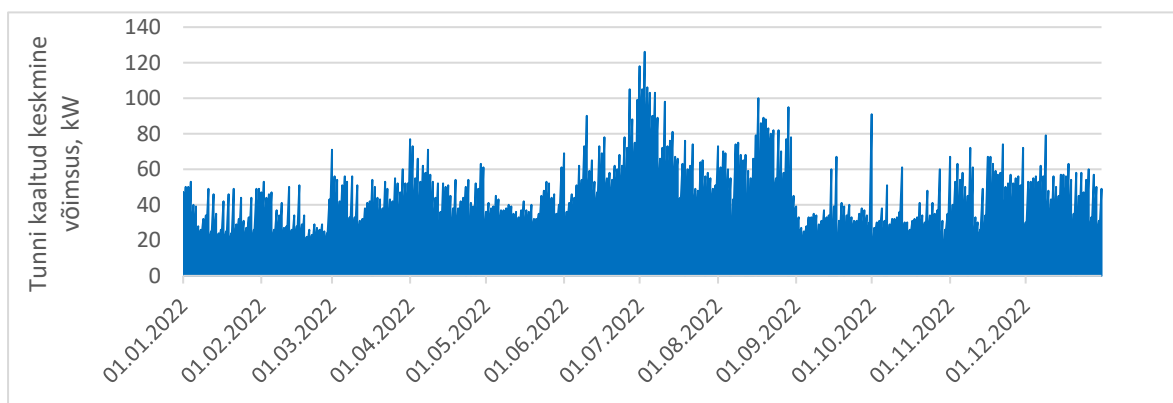
2. KAUGJAHUTUSE VÕIMSUSKOORMUSE ENNUSTAMINE REGRESSIOONANALÜÜSIGA

Kaugjahutussüsteemi valdajal on õigustatud huvi saada iga hetk tarbimisandmeid jahutussõlmes kaugjahutussüsteemi poolelt, et oma süsteemi toimimiseks vajalikke andmeid koguda ja süsteemi seeläbi parendada. Küll aga pole õigust kaugjahutussüsteemi valdajal saada infot kliendi sekundaarse ehk majasisese tehnosüsteemi, hoone konstruktsiooni ja toimimise kohta igal ajahetkel infot.

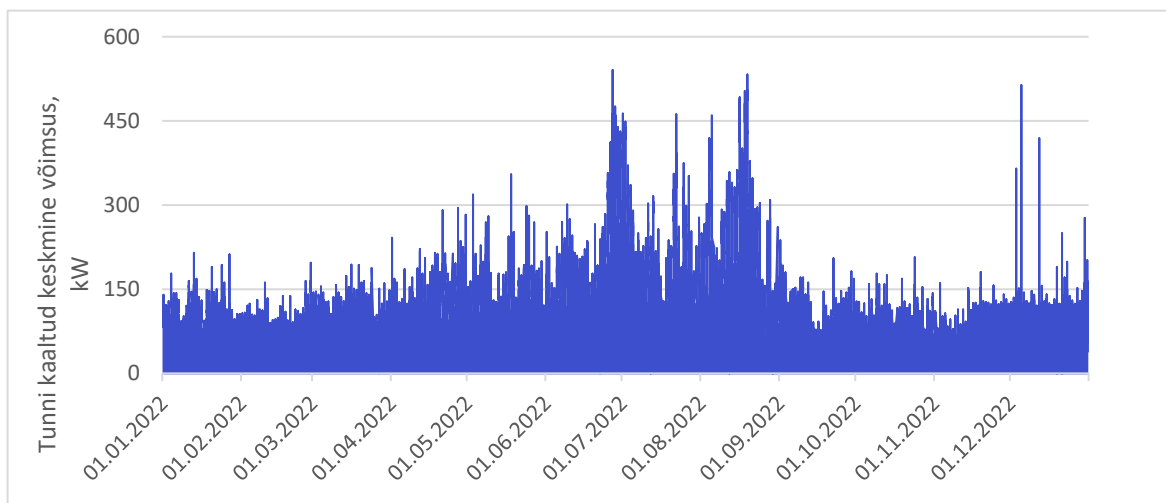
See tekitab probleemi, sest hoone jahutuskoormuse ennustamiseks on kaugjahutussüsteemi valdajal kasutada ainult ilmaandmed, eelnev jahutusenergia tarbimise ajalugu ning hetkeandmed. Seetõttu on hoone jahutuskoormuse ennustamine ebatäpne. Jahutuskoormuse ennustamine annab kaugjahutussüsteemi valdajale võimaluse planeerida oma tootmist ette ning arvestada jahutusenergia tootmiseks vajaliku elektrienergia hinnaga, vajadusel jahutades kaugjahutusvõrku eelnevalt mõned kraadid külmemaks või vastupidi mitte tootes kõrge elektri hinnaga, kui nõudlust antud ajahetkel on vähe. Sellise ennustuse rakendamiseks peab kaugjahutusvõrgu valdajal olema väga hea ülevaade torustiku hüdraulikast, sest kaugjahutussüsteemi suure inertsitõttu ei jõua muutused tarbijateni kohe.

Käesolevas töös on kasutatud kolme AS Utilitas Tallinn kaugjahutusvõrgus oleva hoone 2022. ja 2023. aasta tarbimise tunniandmeid, mille põhjal tuuakse edasised näited ning sooritatakse analüüs. Hoonete aadresse ei avaldata, need on esindatud lõputöös kui Hoone A, Hoone B ja Hoone C. Peab rõhku panema, et kasutatud võimsus on tunni keskmine jahutusvõimsus.

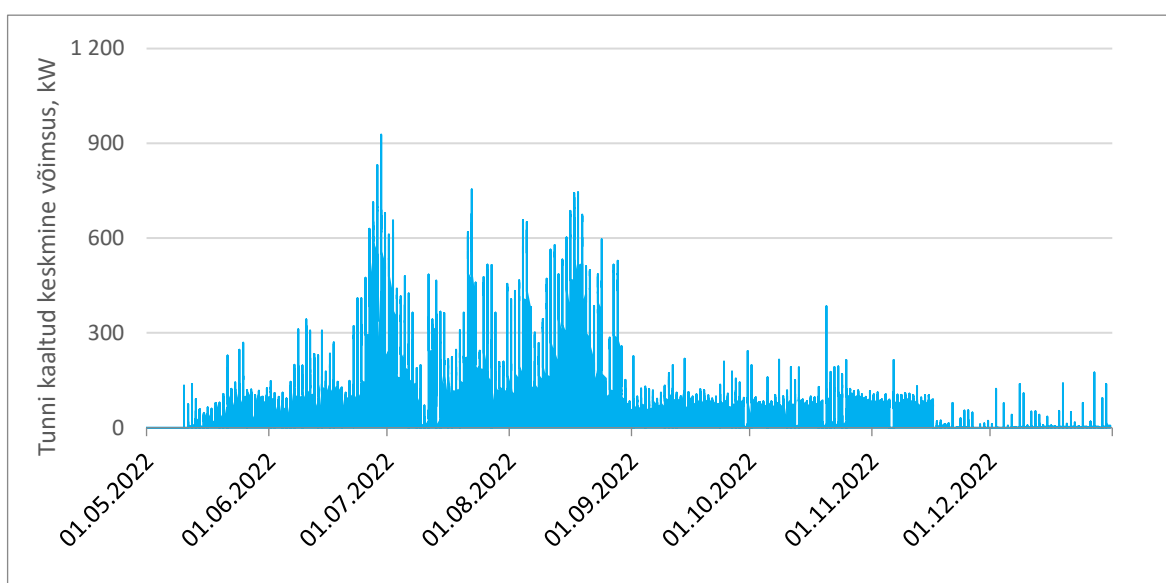
Järgnevalt on toodud ülalmainitud kolme hoone 2022. aasta jahutuse koormusgraafikud (joonised 3 kuni 5):



Joonis 3. Hoone A 2022. aasta jahutuse koormusgraafik



Joonis 4. Hoone B 2022. aasta jahutuse koormusgraafik



Joonis 5. Hoone C 2022. aasta jahutuse koormusgraafik

Ülaltoodud graafikutel saame järeldada, et peamine tarbimine on suvel, kui ilmad on soojemad. Küll aga on tarbimise profiilid ning minimaalse ja maksimaalse tarbimise amplituudid erinev igal hoonel. Hoone C ei tarbi jahedamatel kuudel kaugjahutust, samas kui hooned A ja B tarbivad. Hoone B võimsuste amplituud on suurem kui Hoone A oma. Peab esile tõstma, et Hoone C 2022. aasta andmeid on alates esimesest maist kaugjahutussüsteemiga hilisema liitumise tõttu.

2.1 Tarbija jahutuskoormuse sõltuvus välistemperatuurist

Enamus teadusartikleid ning uurimistöid, mis keskenduvad jahutusenergia vajadusele suurtes süsteemides, kasutavad peamiselt välistemperatuuri ning jahutuskraadtunde, et ennustada süsteemi koormust [16]. Suurele süsteemile võib taoline ennustamine olla piisavalt täpne, et sellega üldiseid ennustusi teha, kuid konkreetsetele hoonetele ennustuste tegemisest ainult välistemperatuurist ei piisa.

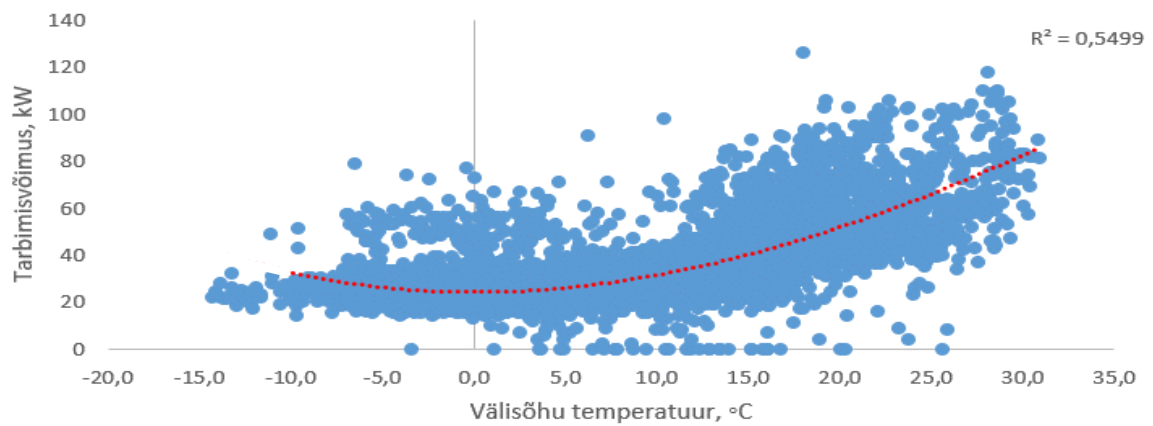
Järgnevalt on toodud kõige kolme hoone kohta joonised (joonised 6 kuni 8), kus on kasutatud tarkvara Microsoft Exceli regressioonanalüüsi funktsionaalsust „*Trendline*“, mis leiab parima andmeid kirjeldava funktsiooni.

Ülaltoodud funktsionaalsusega saame analüütilise valemi iga hoone jaoks ning hinnangu antud valemi sobivust algandmete ennustamisel ehk determinatsioonikordaja R^2 , mis ütleb kui suure osa uuritava tunnuse varieeruvusest mudel ära kirjeldab. Täpsemalt näitab determinatsioonikordaja, kui suure osa sõltuva tunnuse koguhajuvusest moodustab regressioonhajuvus (läbi valemi muutujate). Determinatsioonikordaja on regressioonhajuvuse ja koguhajuvuse ruutude summade jagatis. Seetõttu jääb vahemikku $0 \leq R^2 \leq 1$, kus 1 on ideaalselt sobiv mudel ja 0 tähistab väärtust, kus leitud valem ei selgita koguhajuvust. Tavaliselt tuuakse R^2 väärtus protsendina [17].

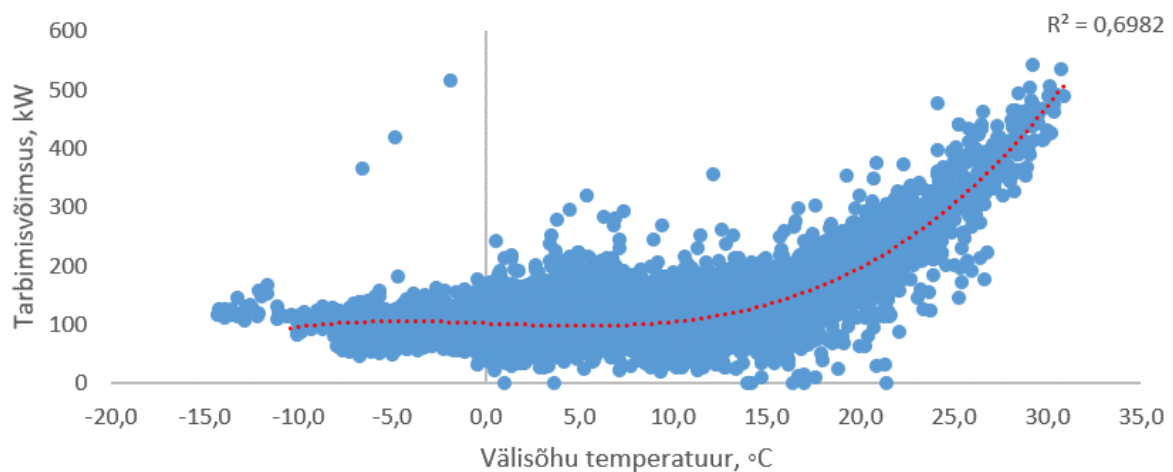
Mida suurem on R^2 , seda parem. Kõige paremini sobituva ehk suurima determinatsioonikordaja väärtusega funktsioon on Hoonete A, B ja C puhul kolmanda astme polünoom. Neljanda, viienda ja kuuenda astme polünoomid suurendavad determinatsioonikordaja väärtust marginaalselt ning lisavad funktsiooni keerukust ilma täpsust oluliselt suurendamata.

Iga hoone kolmanda astme polünoomi funktsiooni determinatsioonikordaja R^2 on toodud järgmiselt, et oleks võimalik võrrelda tulemusi:

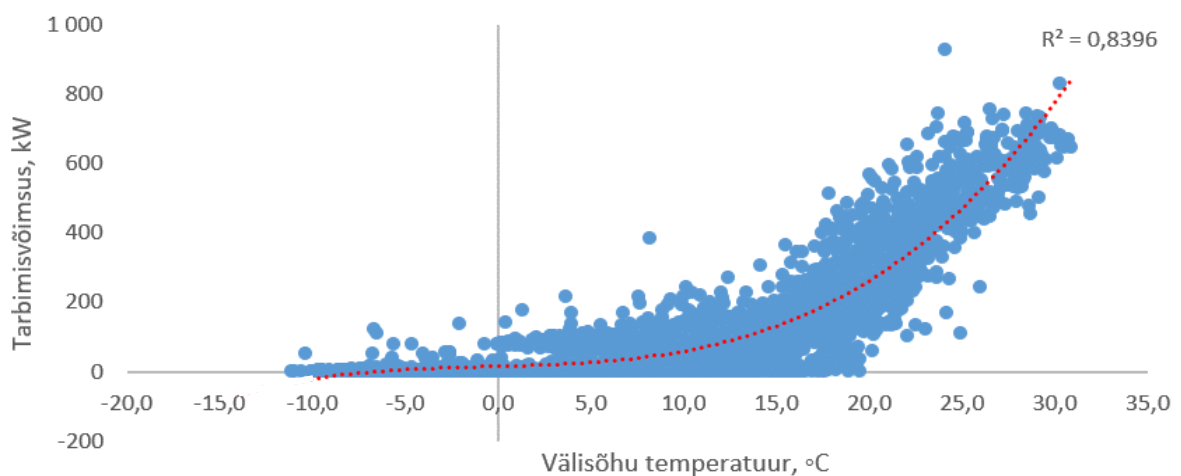
- Hoone A: 54,99%
- Hoone B: 69,82%
- Hoone C: 83,96%



Joonis 6. Hoone A 2022. aasta tarbimisvõimsuse sõltuvus välisõhu temperatuurist



Joonis 7. Hoone B 2022. aasta tarbimisvõimsuse sõltuvus välisõhu temperatuurist



Joonis 8. Hoone C 2022. aasta tarbimisvõimsuse sõltuvus välisõhu temperatuurist

Nagu ülaltoodud joonistelt näha, on iga hoone jahutuskoormuse iseloom isemoodi. Kõige rohkem mõjutab funktsiooni kuju baaskoormuse olemasolu hoones ehk

välitemperatuurist sõltumatu jahutuskoormus. Baaskoormuse mõju tuleb kõige paremini välja madalatel välitemperatuuridel, sest hoone jahutuskoormus muutub sisemiste soojusallikate tõttu stabiilsemaks. Baaskoormust võivad hoonele anda serverid, toitlustusfunktsioonid, jahutusprotsessid või külmakambrid.

Selgelt on Hoones A suur baaskoormus, millele järgneb Hoone B ning seejärel Hoone C, mis ei tarbinud jahutust joonise 3 järgi kuni maikuuni 2022 ning lõpetas tarbimise novembri lõpus 2022. Hoone A ja B tarbivad jahutusenergiat aastaringelt.

2.2 Tarbija jahutuskoormuse sõltuvus jahutuskraadtundidest

Jahutuskoormuse paremaks iseloomustamiseks saab eeskujuks võtta hoone soojusvajaduse määramisel tuntud mõiste kraadpäev ning sellest täpsema, jahutuse valdkonnas tuntud kraadtunni analoogi mõiste – jahutuskraadtunni (CDH, inglise keeles *cooling degree hour*). Üks jahutuskraadtund väljendab 1°C erinevust arvestusliku sisetemperatuuri ja tunni keskmise välisõhu temperatuuri vahel [18].

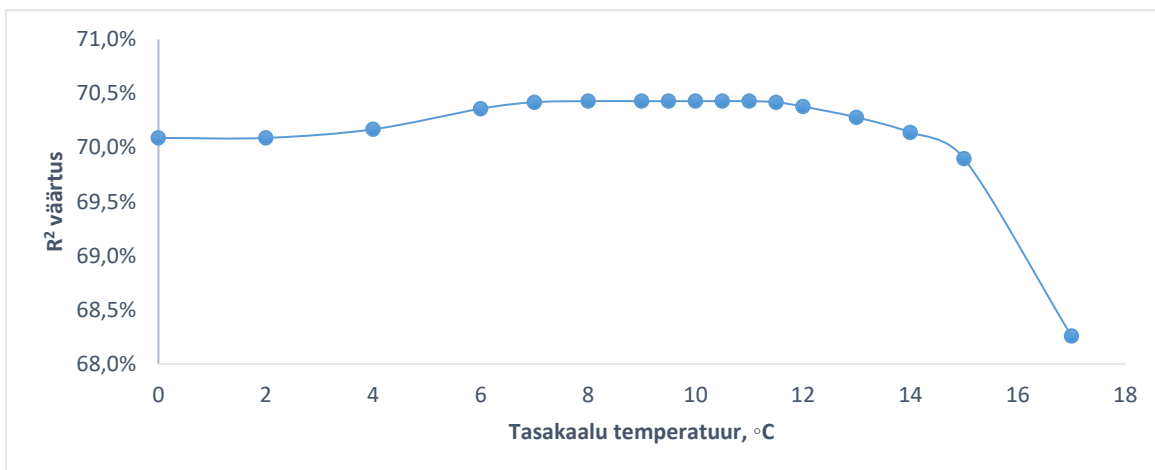
Juhul, kui välitemperatuur on madalam kui arvestuslik sisetemperatuur, siis jahutuskraadtundide arv võrdub nulliga. Arvestuslik sisetemperatuur on igal hoonel erinev ning sõltub hoone ehitusest ja jahutussüsteemist. Seetõttu peab võrguvaldaja leidma korduva arvutusega parima hoonele kohalduva arvestusliku sisetemperatuuri, mis väljendab konkreetse hoone tarbimisprofiili.

Kui välisõhu tunnikeskmine temperatuur oli tunnil X näiteks +19,5°C ning hoonele iseloomulik arvestuslik sisetemperatuur on +15°C, siis jahutuskraadtunde oli antud tunnil X: $19,5^{\circ}\text{C} - 15^{\circ}\text{C} = 4,5$ CDH.

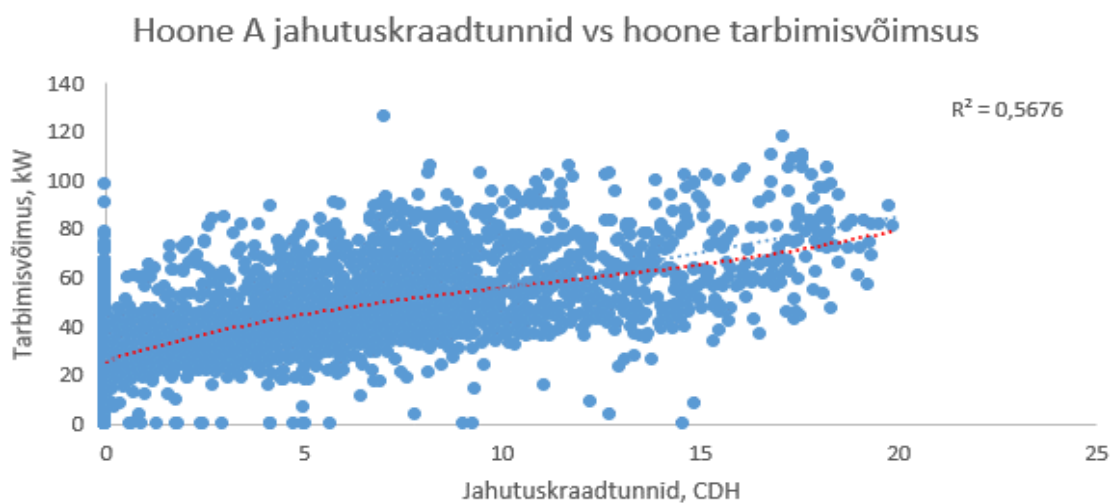
Jahutuskraadtundide leidmiseks pikema perioodi jooksul tuleb kõikide perioodi tundide kraadtunnid liita kokku, et saada vaadeldava perioodi kraadtundide koondarv.

Iga hoone puhul on leitud arvestuslik sisetemperatuur ehk tasakaalu temperatuur (joonised 9, 11 ning 13), mille puhul determinatsiooni kordaja on maksimaalne. Selline tasakaalu temperatuur iseloomustab iga hoone puhul välitemperatuuri, millest soojemad temperatuurid tekitavad jahutuskoormuse kasvule lisa. Leitud tasakaalutemperatuuridest on tekitatud jahutuskraadtundide ja tarbimisvõimsuse sõltuvusgraafikud (joonised 10, 12 ning 14). Tunnid, mille puhul kraadtundide arv on 0,

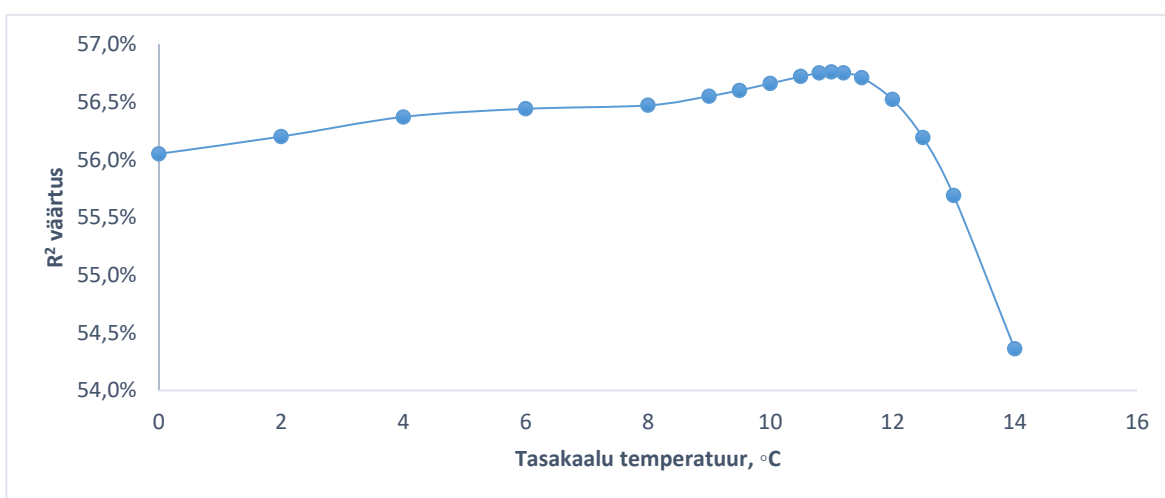
on baaskoormuse tunnid ning nende tundide jahutuskoormus ei ole välistemperatuurist nii tugevalt mõjutatud, kui üldse.



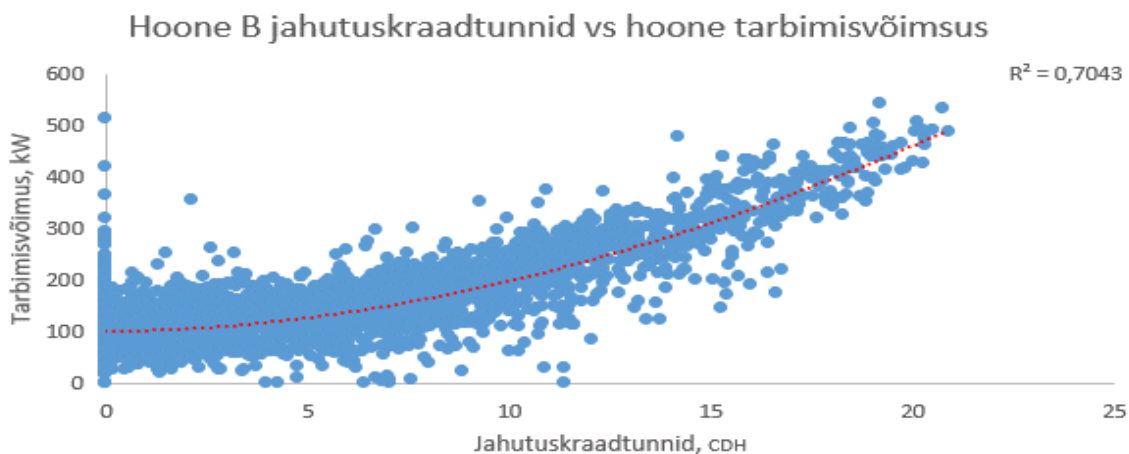
Joonis 9. Hoone A optimaalne tasakaalu temperatuur



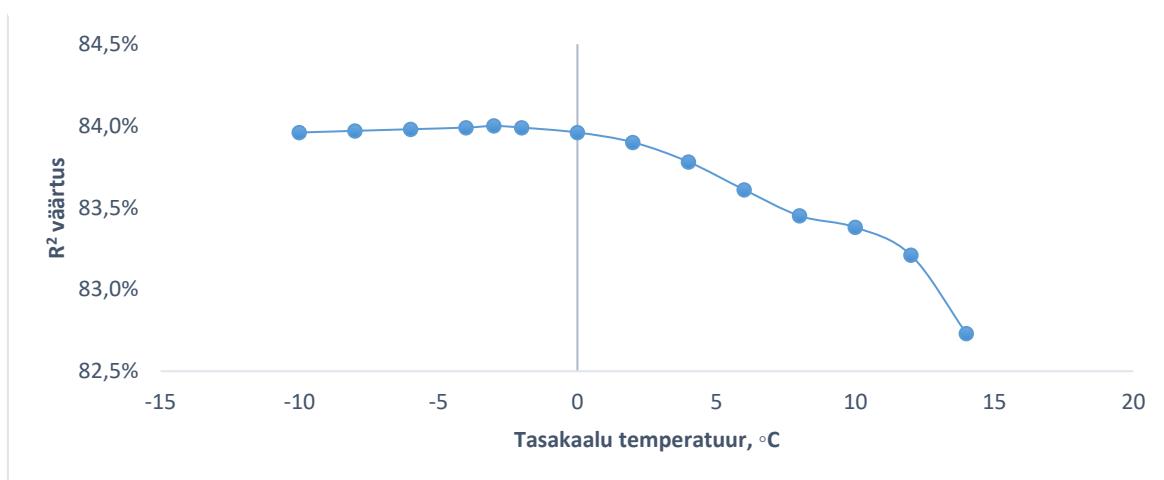
Joonis 10. Hoone A 2022. aasta tarbimisvõimsuse sõltuvus jahutuskraadtundidest



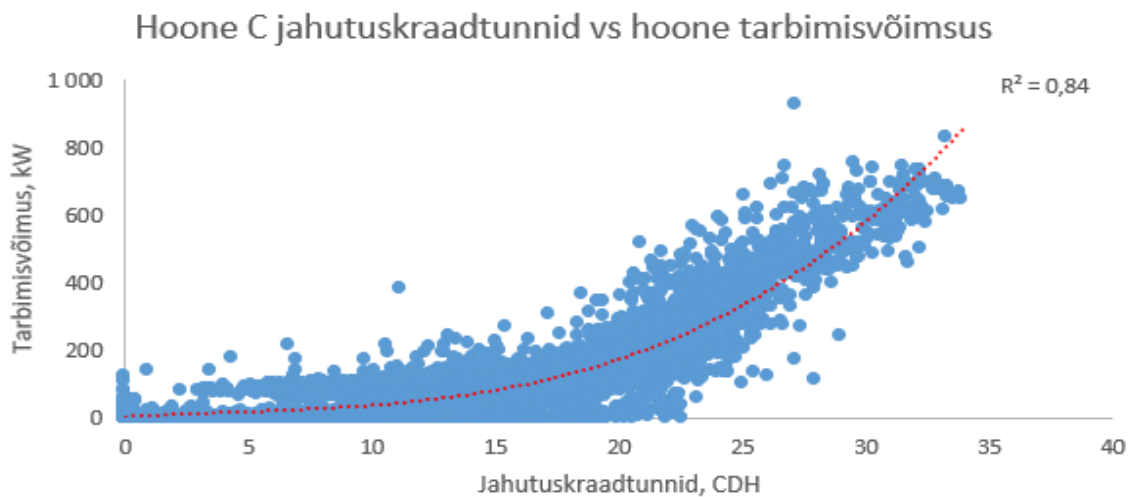
Joonis 11. Hoone B optimaalne tasakaalu temperatuur



Joonis 14. Hoone B 2022. aasta tarbimisvõimsuse sõltuvus jahutuskraadtundidest



Joonis 13. Hoone C optimaalne tasakaalu temperatuur.



Joonis 12. Hoone C 2022. aasta tarbimisvõimsuse sõltuvus jahutuskraadtundidest

Tabelis 1 on toodud hoonetele leitud tasakaalutemperatuurid ning vastav determinatsioonikordaja väärtus. Nagu R^2 täpsuse kasvu veerust on näha, siis jahutuskraadtundide kasutamine annab täpseima tulemuse hoonete puhul, kus on baaskoormuse osakaal suur.

Hoone C puhul ei anna jahutuskraadtundide kasutamine märgatavat efekti, mida selgitab baaskoormuse puudus. Seega antud hoone jahutuskooormuse korrelatsioon välistemperatuuri ja jahutuskooormuse vahel on väga tugev.

Tabel 1. Hoonete tasakaalu temperatuurid ja determinatsioonikordajad

Hoone	Tasakaalu temperatuur, °C	R^2 väärtus, % jahutuskraadtundidest sõltavana	R^2 väärtus, % Välistemperatuurist sõltavana	R^2 väärtuse muutus
Hoone A	11,0 °C	56,75 %	54,99 %	+1,76%
Hoone B	10,0 °C	70,43 %	69,82 %	+0,61 %
Hoone C	-3,0 °C	84,00 %	83,96 %	+0,04 %

2.3 Tarbija jahutuskooormuse ennustamine jahutuskraadtundidest sõltavana

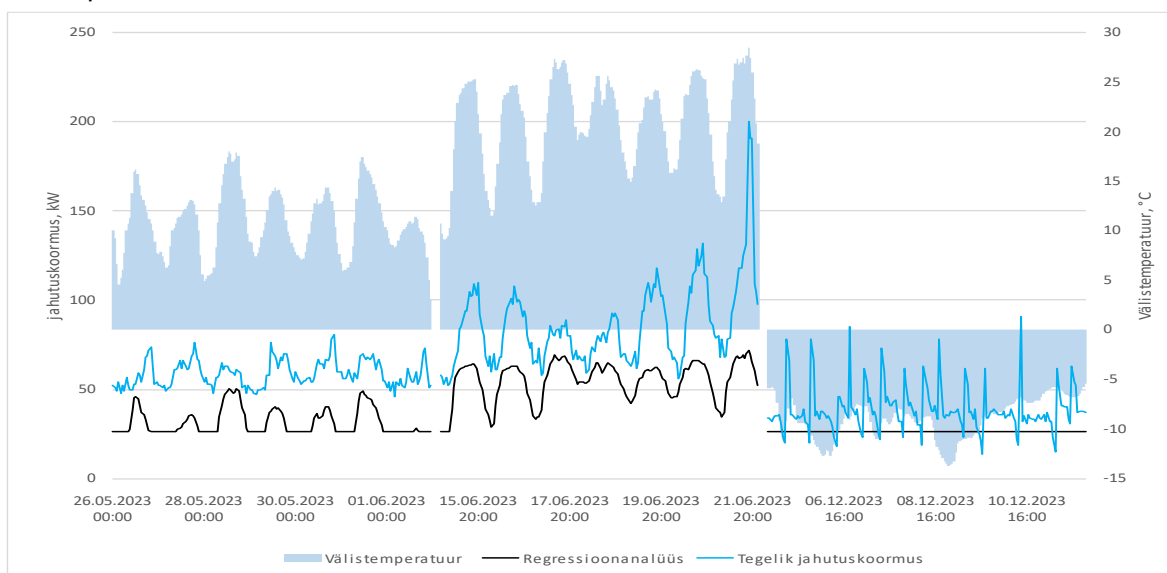
Kuna jahutuskraadtundide kasutamine annab väikse positiivse efekti determinatsioonikordajale ehk suurendab korrelatsiooni välistemperatuuri ja hoone tarbimisvõimsuse vahel, siis ennustuse tegemiseks on kasutatud joonistes 8, 10 ning 12 kuvatud Microsoft Excelisse sisse ehitatud regressioonanalüüsi kolmanda astme polünoomist funktsiooni (teisi tuntud kui trendijoone valem), et ennustada 2023. aasta valitud perioodide ilmaandmete põhjal tunnitarbimist.

Täpsemaks analüüsimiseks on töös valitud üks nädalane periood suvekuudest, mil jahutuskooormused on kõige suuremad. Üks nädalane periood on valitud talvekuudest, mil väljas oli kõige külmem ning välistemperatuurist tingitud jahutuskooormust ei tohiks olla või on minimaalne. Viimaseks on valitud üks nädalane periood kevadest, mil välistemperatuur jääb keskmiselt Hoone A ja B tasakaalu temperatuuride lähedale. Valitud perioodid ja iseloomustavad suurused on toodud tabelis 2. Sügiskuudest ei ole nädalast perioodi valitud, kuna välistemperatuurid on samad nagu kevadkuudel.

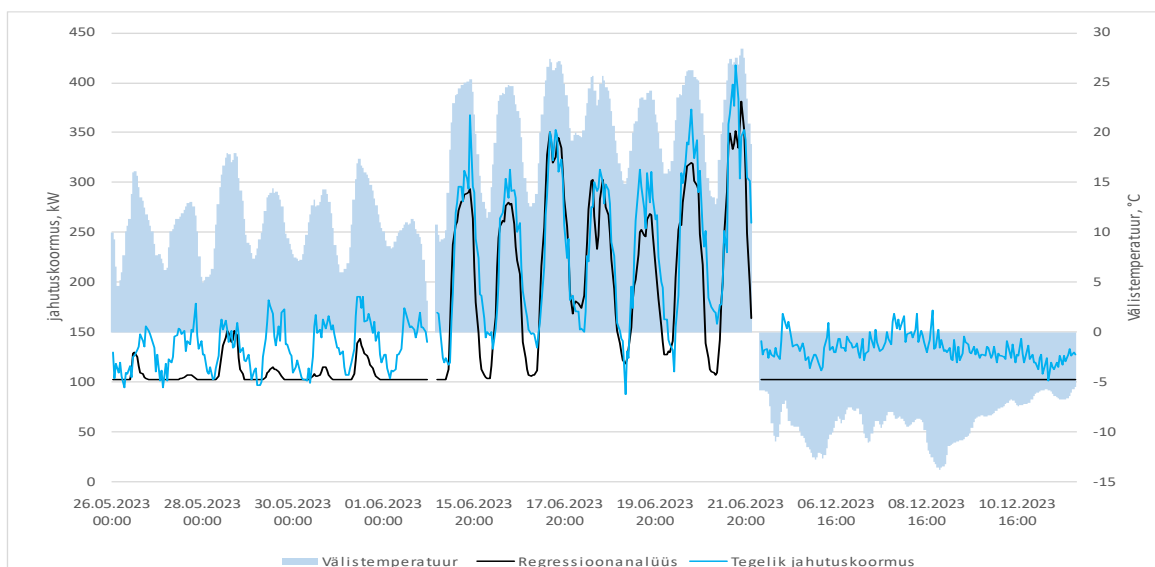
Tabel 2. Valitud 2023. aasta nädalad, mille jooksul töös valitud hoonete jahutuskoormust ennustatakse

Valitud nädal	Keskmine välistemperatuur, °C
15.06.2023 - 21.06.2023	21,0
05.12.2023 - 11.12.2023	-8,9
26.05.2023 - 01.06.2023	10,9

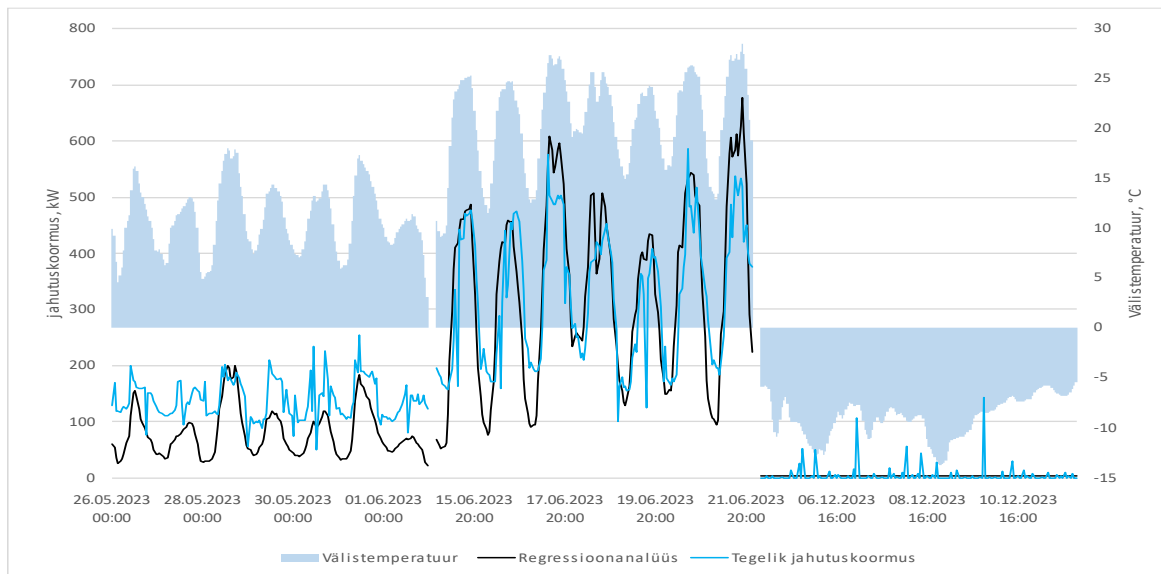
Regressioonanalüüsi meetodiga saadud ennustuste võrdlused tegeliku 2023. aasta jahutuskoormusega on toodud välja järgmistel joonistel ning lisa 1, kus eristuvad visuaalselt erinevused regressioonanalüüsi ennustuste ja tegelike jahutuskoormuste vahel paremini.



Joonis 15. Hoone A regressioonanalüüsi meetodil leitud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskoormusega



Joonis 16. Hoone B regressioonanalüüsi meetodil leitud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskoormusega



Joonis 17. Hoone C regressioonanalüüsi meetodil leitud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskoormusega

Nagu ülaltoodud joonistelt 15 kuni 17 on näha, siis Hoone A puhul ei ennusta regressioonanalüüsi meetodil leitud funktsioon eriti täpselt hoone jahutuskoormust võrreldes tegeliku jahutuskoormusega. Hoonete B ja C puhul on olukord parem, kuid kevadise ja talvise perioodi koormuse ennustamise jääb samuti puudulikuks. Lisaks võib välja tuua ka selle, et kuna regressioonanalüüs otsib funktsiooni, mis sobib kõige rohkemate andmepunktidega, siis meetod pigem alahindab tarbimiskoormuse ennustamisel.

Jahutuskoormuse alahindamine on kaugjahutussüsteemi operatori vaatest halvem kui ülehindamine, sest tootes jahutusenergiat vastavalt ennustatud graafikule, jääks kaugjahutuse klientidel jahutusenergiast puudu. See tähendab et kaugjahutussüsteemi operatuur ei täida oma kohustust igal ajahetkel hoone jahutusenergia vajadust rahuldada. Operatuur peab sellisel juhul suurendama pumpade tööd, olenemata elektri hinnast, ning täisvõimsusel hakkama jahutama kaugjahutusvõrku, mis võib kujuneda väga kalliks.

Hea viis hinnata ennustusmeetodi vea suurust võrreldes tegeliku jahutuskoormusega on ruutkeskmise vea arvutamine. Ruutkeskmise vea eelis seisneb selles, et see annab keskmise absoluutse vea suuruse. Ruutkeskmist viga, tuntud ka lühendiga RMSE (*root mean squared error*), arvutatakse valemiga 1:

$$RMSE = \sqrt{\frac{1}{n} \sum_i^n y_i - \hat{y}_i)^2} , \text{ kus} \quad (1)$$

n = vaadeldavas perioodis olevate tundide arv;

y_i = tegelik jahutuskoormus;

\hat{y}_i = ennustatud jahutuskoormus.

Joonistel 15 kuni 17 kujutatud ajavahemike ennustatud jahutuskoormuste ruutkeskmised vead, lühendina RMSE (inglise keeles *root mean squared error*), on toodud tabelis 3. Peatükis 5 on täpsemalt võrreldud regressioonanalüüsiga ennustatud jahutuskoormuste ruutkeskmisi vigu masinõppe meetodil tehtud ennustuste ruutkeskmiste vigadega.

Lisas 2 on toodud 2022. aasta algandmed regressioonanalüüsiks, 2023. aasta sisendandmed ennustuste tegemiseks ning tegelike jahutuskoormuste ja regressioonanalüüsi ennustuste tulemused.

Tabel 3 Regressioonmeetodil leitud ennustusväärtuste ruutkeskmised vead

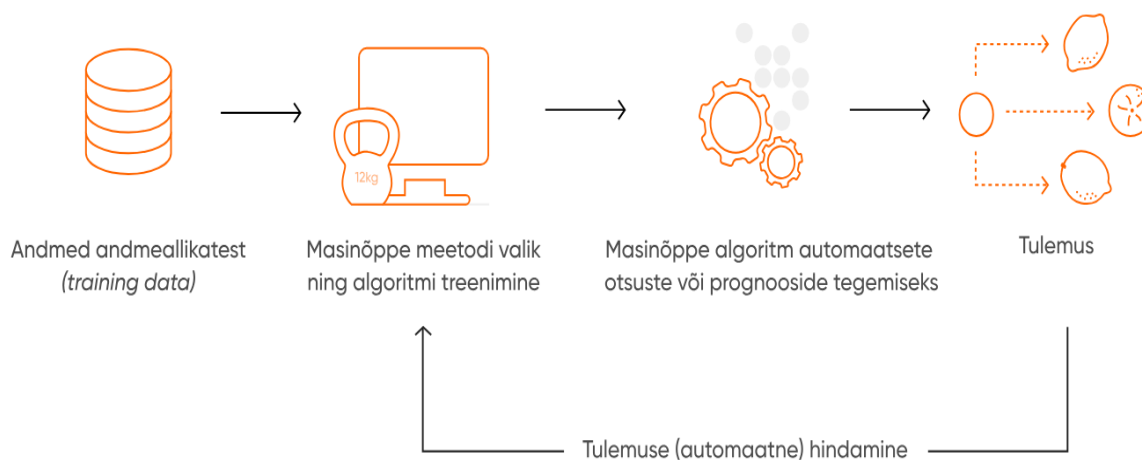
	Hoone A	Hoone B	Hoone C
26.05.2023 - 01.06.2023			
RMSE	28,44	34,03	67,15
15.06.2023 - 21.06.2023			
RMSE	36,25	40,56	88,14
05.12.2023 - 11.12.2023			
RMSE	17,83	33,62	16,05

3. MASINÕPE

Masinõppimine (inglise keeles *machine learning*) on infotehnoloogia valdkond, mille eesmärgiks on olemasolevate andmete põhjal välja töötada ennustusi või otsuseid tegev algoritm.

Tom M. Mitchell, Ameerika Foundersi ülikooli professor, on defineerinud masinõppimise järgnevalt: „Õeldakse, et arvutiprogramm õpib kogemusest E seoses teatud ülesannete T klassiga ja jõudluse mõõtmisega P, kui selle jõudlus ülesannetes T, nagu mõõdetud P-ga, paraneb kogemuse E-ga“ [19].

Masinõppe algoritmid ehitavad ise automaatselt matemaatilise mudeli, kasutades näidisandmeid ehk "õppeandmeid", et teha otsuseid või ennustusi, kasutades ehitatud matemaatilist mudelit. Väljendit „masinõpe“ kirjeldas esimest korda 1959. aastal ettevõtte IBM töötaja Arthur Samuel, kes leiutas programmi, mis arvutab kabe mängus võiduvõimaluse, arvestades mängu seisu [20]. Masinõppe protsessi illustreerib hästi joonis 18:

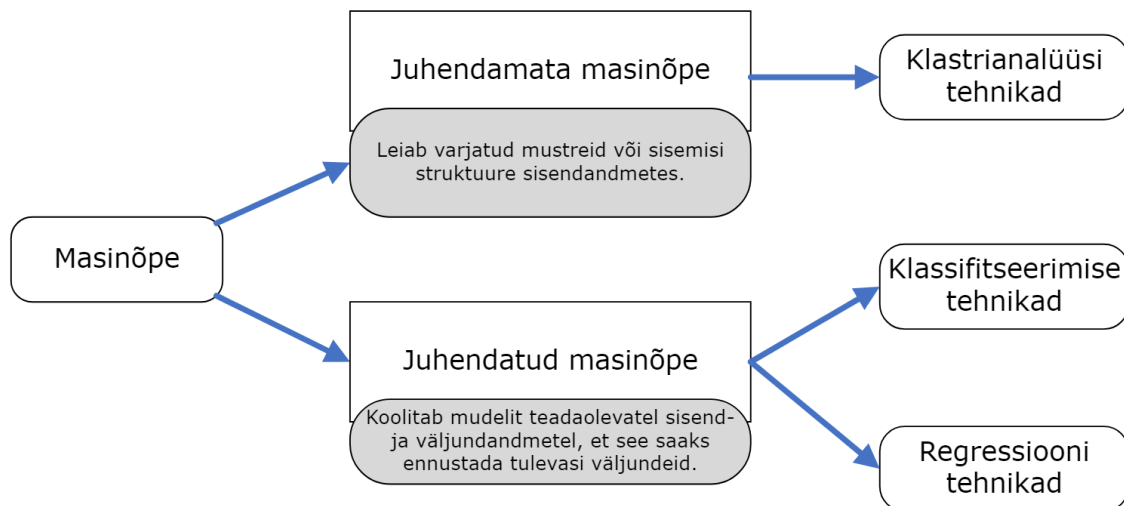


Joonis 18. Masinõppe protsess [21]

Masinõppe algoritme ja meetodeid on tänapäeval väga palju ning kõigil neist on oma nüansid. Üldiselt võib masinõppe meetodeid jagada suure pildis kaheks [22]:

- klastrianalüüs juhendamata masinõppest
- regressiooni ja klassifitseerimistehnikad juhendatud masinõppest.

Sellist jaotust illustreerib hästi joonis 19.



Joonis 19. Masinõppe meetodite ja peamiste tehnikate klassifitseerimine [22]

Juhendamata masinõpe leiab varjatud mustreid või sisemisi struktuure andmetes. Seda meetodikat kasutatakse järelduste tegemiseks andmehulkadest, mis koosnevad sisendandmetest ilma sildistatud vastusteta [22].

Klastrianalüüs on kõige levinum juhendamata masinõppe tehnika. Seda kasutatakse avastuslikuks andmeanalüüsiks, et leida varjatud mustreid või rühmitusi andmetes. Klastrianalüüsi rakendusteks on näiteks geneetilise järjestuse analüüs, turu-uuringud või objektide tuvastamine.

Juhendatud masinõpe loob mudeli, mis teeb ennustusi. Juhendatud masinõppe algoritm võtab teadaoleva sisendandmete kogumi ja teadaolevad vastused nendele andmetele (väljundid) ning koolitab mudeli tegema mõistlikke ennustusi uute sisendandmete mudelisse sisestamisel. Juhendatud masinõpet kasutatakse, kui on teada sisendandmed väljundi kohta, mida üritad ennustada; kuid reeglid, kuidas sisendandmed ja väljundandmed omavahel seonduvad, on teadmata. Juhendatud masinõpe kasutab klassifitseerimise või regressiooni tehnikaid masinõppe mudelite arendamiseks [22].

Klassifitseerimistehnikad ennustavad diskreetseid vastuseid – näiteks, kas e-kiri on ehtne või rämpspost, kas kasvaja on pahaloomuline või healoomuline. Klassifitseerimismudelid liigitavad sisendandmed kategooriatesse. Tüüpilised rakendused on kõne tuvastus või krediitdiskoori hindamine. Klassifitseerimist kasutatakse, kui andmeid saab märgistada, kategoriseerida või eraldada konkreetsetesse rühmadesse või klassidesse.

Regressioonitehnikad ennustavad pidevaid vastuseid - näiteks raskesti mõõdetavaid füüsikalisi suurusid, nagu aku laetuse tase, elektritarbimine võrgus või finantsvarade hinnad. Tüüpilised rakendused hõlmavad virtuaalset andurit, elektritarbimise prognoosimist ja algoritmilist kauplemist. Regressioonitehnikaid kasutatakse, kui töötatakse andmehulga või vastustega, mis on reaalsed numbrid, nagu temperatuur või seadme riknemise aeg. [22]

3.1 Juhendatud masinõpe

Juhendatud masinõpet kasutatakse probleemide jaoks, kus kasutatakse mitme muutujaga x_i treeningandmeid otsitava muutuja või väljundi y_i ennustamiseks. Käesolevas töös on otsitavaks väljundiks y_i hoonete A, B ja C tunni jahutuskoormus.

Juhendatud masinõppe aluseks on mudel ehk matemaatiline struktuur, mille abil tehakse ennustus \hat{y}_i sisendist x_i . Üks levinud näide on lineaarne mudel, kus ennustus antakse kui lineaarne kombinatsioon kaalutud sisendomadustest. Lineaarse mudeli korral (valem 2) on ülesanne leida koefitsiendid või kaalud θ_j sisendväärtustele selliselt, et ennustusväärtus \hat{y}_i oleks võimalikult lähedal tegelikule väljundväärtusele y_i , mis hiljem on mõõdetav näiteks jahutuskoormusena antud tunnil.

$$\hat{y}_i = \sum_j \theta_j * x_{ij} \quad (2)$$

Ennustusväärtus \hat{y}_i võib erinevates ülesannetes olla erinevalt tõlgendatav, näiteks regressiooni või klassifitseerimise korral [23].

Masinõppe aluseks oleva mudeli täpseks ennustamiseks on vaja mudelit treenida. Mudeli treenimise ülesanne seisneb parimate parameetrite θ leidmises, mis sobivad kõige paremini treeningandmete x_i ja etteantud väljunditega y_i . Mudeli treenimiseks on vaja määratleda eesmärgifunktsioon – see mõõdab, kui hästi sobib treenitud mudel treeningandmetega.

Eesmärgifunktsiooni oluline omadus on, et see koosneb kahest osast: treeningkaotusest ja regulariseerimisterminist [23]:

$$Obj(\theta) = L(\theta) + \Omega(\theta), \text{ kus} \quad (2.1)$$

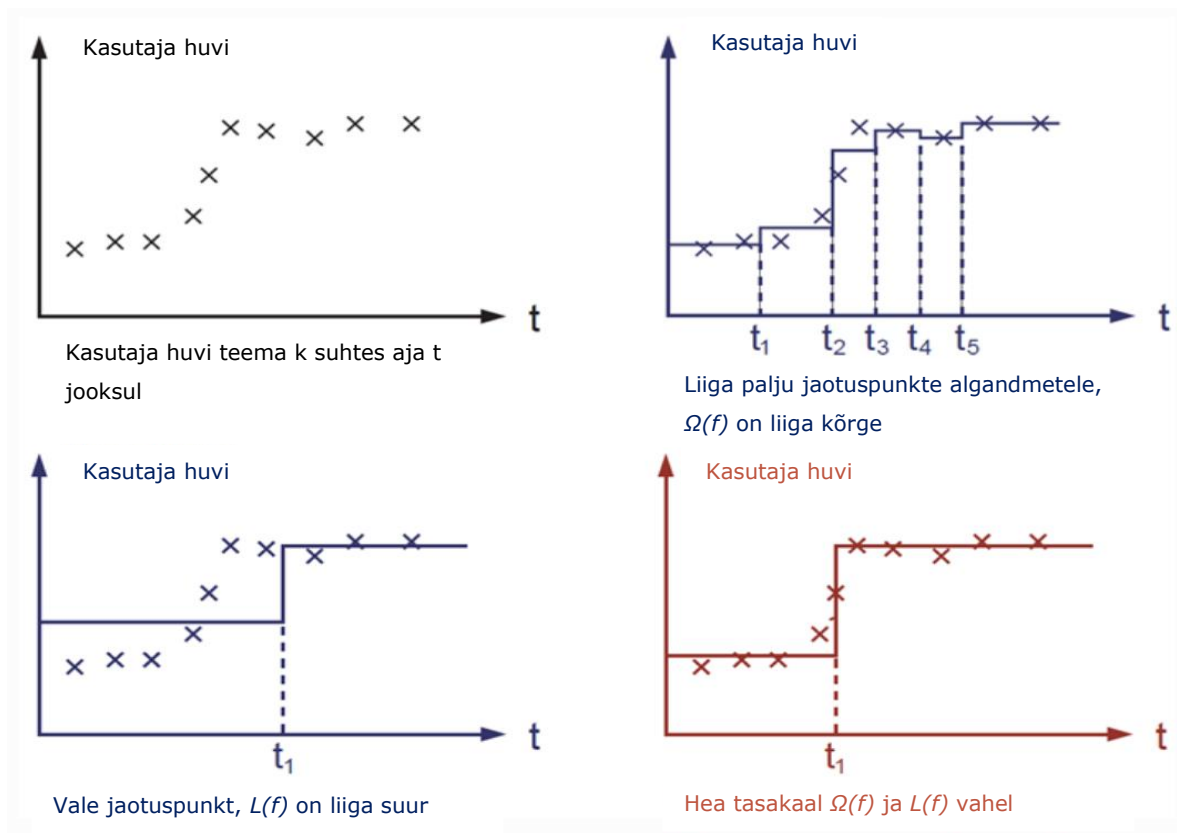
L on treeningkaotuse funktsioon;

Ω on regulariseerimisterminite funktsioon.

Treeningkaotuse funktsioon $L(\theta)$ mõõdab, kui ennustav on meie mudel treeningandmete suhtes. Tavaline treeningkaotuse valik on keskmine ruutviga, mida arvutatakse järgmiselt:

$$L(\theta) = \sum_i (y_i - \hat{y}_i)^2 \quad (2.2)$$

Regulariseerimisterminitega kontrollitakse mudeli keerukust, mis aitab vältida mudeli üle õppimist ehk liigset sobitumist treeningandmetega. Hea näide selgitamiseks visuaalselt, mida eesmärgifunktsiooni komponendid teevad, on toodud joonisel 20, kus ülesanne on visuaalselt sobitada sammufunktsioon antud andmepunktidele:



Joonis 20. Regulariseerimistermini ja treeningkaotuse funktsiooni illustreeriv näide

3.2 Masinõppe algoritmi teegi valik

Leidub erinevaid juhendatud masinõppe teeke ehk infoobjektide kogumeid, seega on tähtis leida oma probleemi lahendamiseks sobivaim. Energeetika valdkonnas on koormuste ja tarbimise ennustamine olnud artiklite ja uurimistööde teemaks juba kümneid aastaid, seega hoonete jahutusenergia kasutust on varasemalt erinevate meetoditega ennustatud. 2021. aastal ilmus ajakirjas *Applied Energy* artikkel „*Space cooling energy usage prediction based on utility data for residential buildings using machine learning methods*“ [24] ehk jahutuse energiakasutuse ennustamine eluhoonetele tuginedes kommunaalteenuste andmetele masinõppe meetodite abil.

Artiklis võrreldi erinevaid masinõppe algoritmide teeke, kasutades erinevaid hindamiskriteeriume, ning jõuti järeldusele, et masinõppe algoritm XGBoost on valimist kõige täpsem. Käesolevas töös on mudeli ennustamise võrdlemiseks kasutuses determinatsioonikordaja R^2 , mis hindab, kui suure osa uuritava tunnuse varieeruvusest mudel ära kirjeldab. Vaadeldes artiklis toodud võrdlustabelit ning võttes hindamise aluseks selles toodud determinatsioonikordaja R^2 ja keskmise ruutvea kriteeriumid (tabel 4), saame hinnata, milline masinõppe meetod võiks sobida käesolevas töös. Antud juhul – mida suurem on determinatsioonikordaja väärtus ning mida väiksem ruutkeskmise viga, seda sobilikum on masinõppe algoritmi teek.

Tabel 4. Erinevate juhendatud masinõppe algoritmi teekide hindamine [24]

Meetod	R^2	Ruutkeskmise viga
polynomial	0.818	0.519
glm	0.802	0.566
XGBoost	0.922	0.221
ranger	0.852	0.420
cforest	0.829	0.489
knn	0.806	0.553
svm	0.836	0.470
gamboost	0.833	0.478
glmboost	0.804	0.562

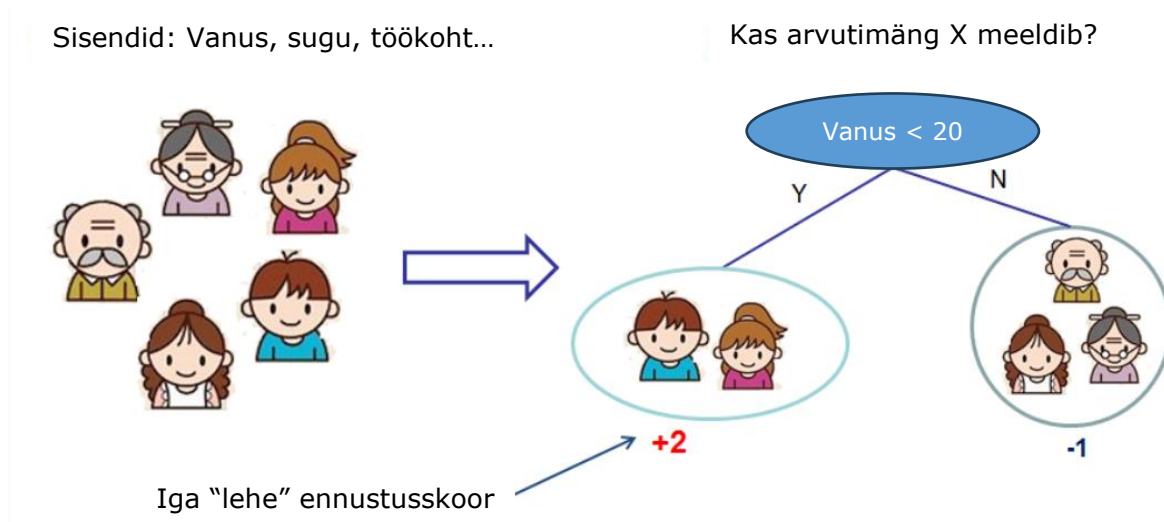
Selgelt on ülaltoodud kriteeriumite alusel juhendatud masinõppe parim meetod XGBoost ning seetõttu käesolevas töös on kasutatud seda juhendatud masinõppe tehnikat.

XGBoost on lühend ingliskeelsele väljendile *Extreme Gradient Boosting*, mis on oma alge saanud artiklist „*Greedy function approximation: A gradient boosting machine*” Jerome H. Friedmani poolt [25], kus autor lõi väljendi *gradient boosting* [23], mis on masinõppe tehnikate valdkond, kus mudelid koolitatakse järk-järgult ja kombineeritakse, et saavutada parem ennustusvõime. *Gradient boosting* põhineb nõrkade mudelite kooskõlastamisel ja iga järgneva mudeli treenimisel eelnevate mudelite poolt tehtud vigade korrigeerimisel.

XGBoost kasutab otsustuspuude kogumite ehk otsustusmetsa mudelit CART (inglise keeles *classification and regression trees*), mis koosneb nii klassifikatsiooni kui regressiooni puudest ehk hargnevatest otsustuslausetest [23].

3.3 Otsustusmetsa mudel

Lihtne näide probleemist, mida CART mudel võib lahendada, on küsimus: „Kas antud isikule meeldib arvutimäng X?” Oletame, et meil on viis inimest (nagu toodud joonisel 21), kelle kohta peame ennustama, kas neile arvutimäng X meeldib. Selle jaoks on meile antud inimeste kohta info nende soo, vanuse, töökoha jm kohta ehk treeningandmestiku muutuja x_i .

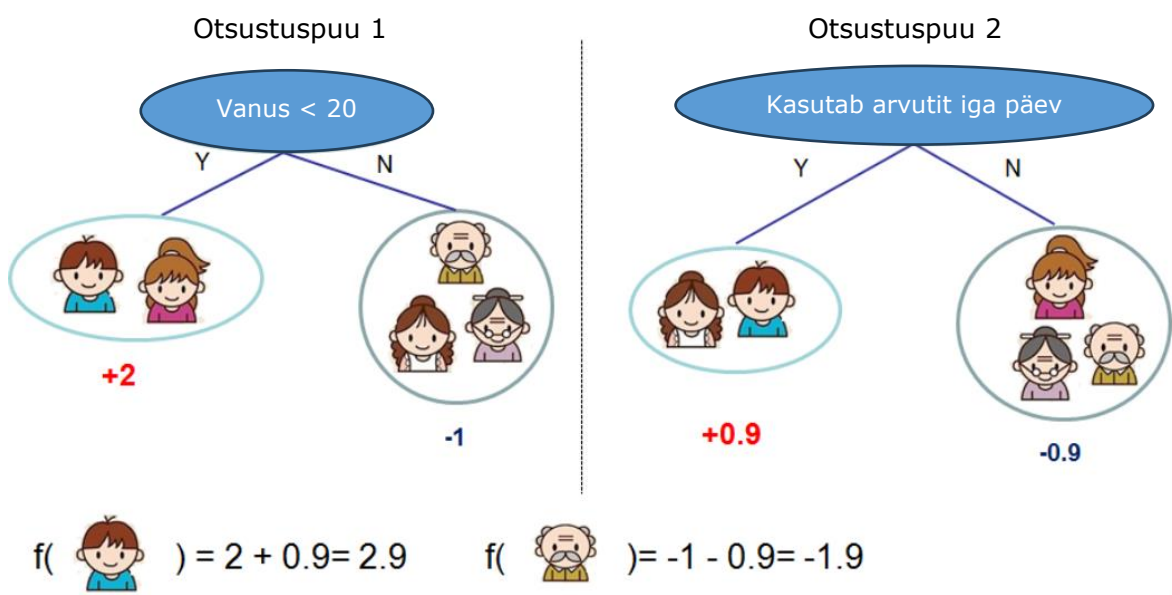


Joonis 21. CART mudeli lihtsa ülesande näide [23]

Mudel klassifitseerib inimesed ehk y_i otsitavad väljundid erinevatesse „lehtedesse” ning annab suvaliselt ennustusskoori ehk kaalu igale lehele. Ülaltoodud joonisel 21 on otsustuspuu, mis koosneb kahest „lehest”.

Kui otsustusmets on piiratud ennustusskoori andmisega, siis CART mudelis on iga lehega seotud tegelik skoor, mis annab meile sügavama tõlgenduse lehtede kaaludes, mis võimaldab minna kaugemale klassifitseerimisest. See võimaldab ka põhimõttelist ja ühtset lähenemist optimeerimisele. [23]

Tavaliselt ei piisa ühest ainsast otsustuspuust, et teha otsuseid või ennustusi. Seetõttu kasutatakse otsustuspuude kogumeid ehk summeeritakse otsustuspuude ennustused kokku üheks otsustusmetsa ennustuseks \hat{y}_i . Joonisel 22 [23] on toodud näide kahe puu kogumist, kus mõlema puu ennustusskoorid summeritakse lõplikuks skooriks ehk kaaluks. Tähtis on siinkohal rõhutada, et mõlemad puud üritavad teineteist täiendada. [23]



Joonis 22. Kahe otsustuspuu ennustusskoori summeerimine [23]

Matemaatiliselt saame kirjeldada joonisel 22 toodud funktsioone valemiga 3.1 [23]:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F, \text{ kus} \quad (3.1)$$

K on puude arv;

f_k on puu funktsiooni valem, mis kuulub hulka F ;

F on hulk, millesse kuuluvad kõik võimalikud CART funktsioonid f_k , mis on võimalikud lähteandmete x_i korral.

Nagu kirjeldatud peatükis 3.1, on mudeli treenimiseks ehk optimeerimiseks vaja määratleda eesmärgifunktsioon $Obj(\theta)$, mis mõõdab, kui hästi treenitud mudel sobib treeningandmetega.

Ülaltoodud joonise 22 näite üldistatud eesmärgifunktsiooni saame kirjeldada valemi 2.1 alusel [23]:

$$Obj(\theta) = L(\theta) + \Omega(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^n \omega(f_k) , \text{ kus} \quad (3.2)$$

$\omega(f_k)$ on iga otsustuspuu keerukus, mis on defineeritud valemis 3.6.

3.4 Otsustusmetsa mudeli puude õpetamine

Nagu kirjeldatud peatükis 3.1 valemiga 2.1, peab otsustusmetsa mudeli treenimiseks kirjeldama eesmärgifunktsiooni ning selle treeningkaotuse ja regulariseerimisterminite komponente [23]:

$$obj = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^t \omega(f_i) \quad (3.3)$$

Esmalt peab mudeli treenimiseks kirjeldama puude parameetrid ehk leidma otsustuspuude funktsioonide valemid f_k , mis kirjeldavad otsustuspuude struktuuri ja „lehtede“ ennustuskoorid. On tähtis rõhutada, et kõiki koostatud otsustuspuud korruga lahendada ning optimeerida on väga keeruline või võimatu. Seetõttu otsustusmetsa mudelite puhul kasutatakse summeerimise strateegiat: fikseeritakse eelnevalt õpitu ning lisatakse iga kord juurde üks uus otsustuspuu [23].

Kirjutades ennustuskoori $\hat{y}_i^{(t)}$ sammul t , saame summeerida otsustuspuude f_i järgmiselt [23]:

$$\begin{aligned} \hat{y}_i^{(0)} &= 0 \\ \hat{y}_i^{(1)} &= f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i) \\ \hat{y}_i^{(2)} &= f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i) \\ &\dots \\ \hat{y}_i^{(t)} &= \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \end{aligned}$$

Ülaltoodud valemiga saame tulemi, kuid see pole kindlasti optimaalne. Selleks, et optimeerida ja saada parem ennustus, võtame arvesse ainult need otsustuspuud, mis parandavad treeningkaotuse funktsiooni ehk teisisõnu vähendavad vahet ennustuse ja otsitava väljundi või väärtuse vahel. Seega eesmärgi funktsiooni valemit 1.3 saame teisendada järgmiselt [23]:

$$\begin{aligned} obj^{(t)} &= \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^t \omega(f_i) \\ &= \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \omega(f_t) + \textit{konstant} \end{aligned}$$

Kasutades ruutkeskmist viga (valem 2.2) treeningkaotuse funktsiooniks [23]:

$$\begin{aligned} obj^{(t)} &= \sum_{i=1}^n (y_i - (\hat{y}_i^{(t-1)} + f_t(x_i)))^2 + \sum_{i=1}^t \omega(f_i) \\ &= \sum_{i=1}^n [2(\hat{y}_i^{(t-1)} - y_i)f_t(x_i)^2] + \omega(f_t) + \textit{konstant} \end{aligned}$$

Ruutkeskmise vea valem andmepunkti jaoks on lihtne, sisaldades esimese astme tegurit (mida tavaliselt nimetatakse jäägiks) ja teise astme tegurit. Üldjuhul võtame kaotusfunktsiooni Tayloriga rea kuni teise astmesse [23].

Tayloriga rida on matemaatilise funktsiooni esitus astmereana, mille kordajateks on funktsiooni tuletis etteantud argumendi väärtuse juures. 0-järku tuletis on funktsioon ise.

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x - a)^n, \text{ kus}$$

$n!$ on faktoriaal;

$f^{(n)}(a)$ on n -dat järku tuletis punktis a .

Seega võttes ruutkeskmise vea kaotusfunktsiooni teise astme Tayloriga rea, saame [23]:

$$obj^{(t)} = \sum_{i=1}^n [l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \omega(f_t) + \textit{konstant}, \text{ kus}$$

g_i ja h_i on defineeritud järgnevalt:

$$g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$$

$$h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)})$$

Eemaldades konstandid, saame konkreetse eesmärgi sammul t [23]:

$$\sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + w(f_t) \quad (3.4)$$

Valem 1.4 saab meie otsustuspuu optimeerimise eesmärgiks. Üks oluline eelis sellise definitsiooni puhul on see, et eesmärgifunktsiooni väärtus sõltub ainult kahest funktsioonist, g_i -st ja h_i -st. Nii toetab XGBoost kohandatud kaotusfunktsioone ning läbi selle juhendatud masinõpet [23].

Üldistatud eesmärgifunktsioonil on ka teine komponent, mida peame arvesse võtma - regulariseerimisterminite funktsioon. Et seda teha, peame defineerima otsustuspuu keerukuse $w(f_k)$. Täpsustame otsustuspuu definitsiooni valemiga 3.5 [23]:

$$f_t(x) = w_{q(x)}, \text{ kus } w \in R^T, q : R^d \rightarrow \{1, 2, \dots, T\}, \text{ kus} \quad (3.5)$$

w on lehtede ennustuskooride vektor;

q on funktsioon, mis määrab iga andmepunkti vastavale lehele;

T on „lehtede“ arv.

XGBoost defineerib otsustusmetsa keerukust valemiga 3.6 [23]:

$$\omega(f_t) = \gamma T + \frac{1}{2} \gamma \sum_{j=1}^T w_j^2 \quad (3.6)$$

XGBoosti dokumentatsioon toob välja, et kompleksust saab ka teisiti defineerida, kuid praktikas on antud valemiga 1.6 toodud definitsioon hästi kasutatav. Regulariseerimisterminite funktsiooni komponent eesmärgifunktsioonis on komponent, mida paljud juhendatud masinõppe algoritmid kas ignoreerivad või ei käsitle süvitsi. See oli tingitud sellest, et traditsiooniline otsustusmetsa õpetamine rõhutas ainult kaotusfunktsiooni parandamist, samas kui keerukuse kontroll jäeti teoretikutele. Defineerides keerukuse funktsiooni formaalselt, saame parema ettekujutuse sellest, mida otsustuspuud õpivad ja nii saavutame mudelid, mis töötavad hästi realses maailmas [23].

Viies kokku valemid 1.6 ja 1.4, saame täieliku eesmärgifunktsiooni valemi (3.7) otsustusmetsa jaoks puude arvuga t:

$$obj^{(t)} = \sum_{i=1}^n [g_i w_{q(x_i)} + \frac{1}{2} h_i w_{q(x_i)}^2] + \gamma T + \frac{1}{2} \gamma \sum_{j=1}^T w_j^2, \text{ kus} \quad (3.7)$$

j on konkreetne „leht“ otsustuspuus;
q(x) on „lehtede“ struktuur.

Kokkuvõttes, iga antud puu struktuuri korral suuname statistilised numbrid g_i ja h_i vastavatele lehtedele, summeerime statistilised numbrid kokku ja kasutame valemit 3.7, et arvutada, kui „hea“ otsustuspuu on. See skoor on sarnane mõiste entroopiaga, mis mõõdab andmete ebamäärasust, välja arvatud nüansiga, et see võtab arvesse ka mudeli keerukust. Algoritm püüab leida sellise küsimuse, millele vastamine vähendaks entroopiat ehk teisisõnu annaks uut informatsiooni võimalikult palju [23].

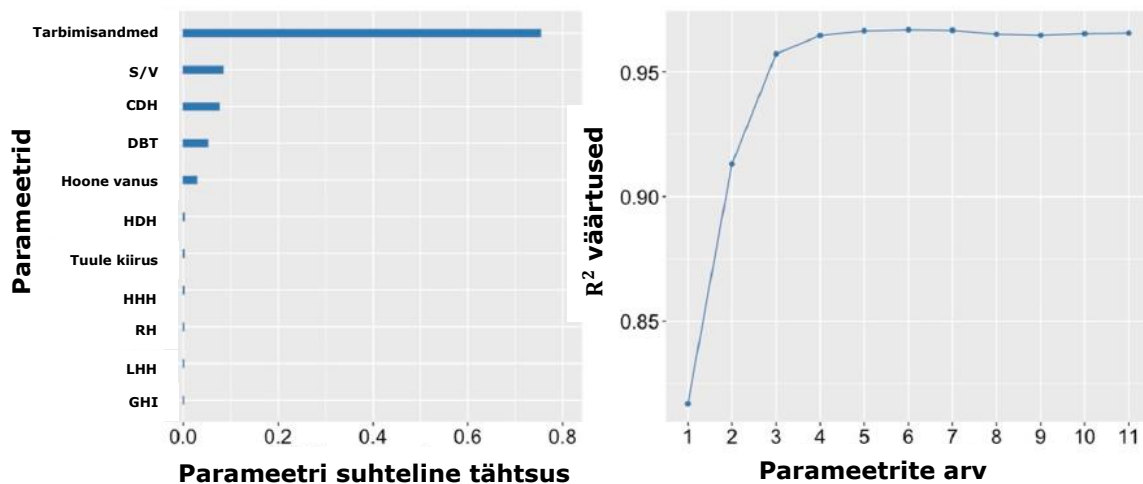
Nüüd, kui on olemas viis mõõta, kui hea on üks otsustuspuu, oleks ideaalis loetletud kõik võimalikud puud ja valitud neist parim isend. Praktikas on see lahendamatu ülesanne, seega optimeerime puu ühte taset korraga. Konkreetselt proovime jagada ühte olemasolevat „lehte“ kaheks uueks „leheks“, jagades algses „lehes“ olevad andmepunktid suvaliselt uute lehtede vahel ning arvutades „lehtede“ ennustuskoorid. Seejärel võrreldakse, kas ennustus läks võrreldes oodatud tulemiga täpsemaks või mitte. „Lehed“, mille otsustuspuule lisamine halvendas eesmärgifunktsiooni täpsust, eemaldatakse ning „lehed“, mis parandasid täpsust jäetakse alles [23].

4. KAUGJAHUTUSE VÕIMSUSKOORMUSE ENNUSTAMINE MASINÕPPEGA

Nagu toodud joonisel 18, siis esimene samm masinõppe protsessis on alati andmete kogumine. Peame tegema valiku andmetest, mis on saadaval ning mis mõjutavad lõppmudeli täpsust piisaval määral.

Artiklis „*Space cooling energy usage prediction based on utility data for residential buildings using machine learning methods*” on analüüsitud 11 erinevat sisendparameetrit eluhoonete andmete põhjal, millest neljal on oluline mõju determinatsioonikordaja suurendamisele ehk ennustuse täpsuse tõusule (joonis 23). Nendeks neljaks valitud sisendparameetriteks on artiklis vastavalt [24]:

- Eelnevad tarbimisandmed;
- Hoone pindala ja ruumala suhe ehk S/V (inglise keeles *surface to volume ratio*);
- CDH ehk jahutuskraadtunnid;
- Kuivtermomeetriga mõõdetud temperatuur DBT ehk välisõhu temperatuur, °C.



Joonis 23. Parameetrite mõju determinatsiooni kordaja suurenemisele

Samuti on artiklis toodud tähtsa parameetrina hoone vanus, kuid selle lisamine ei tõstnud oluliselt mudeli täpsust ning peamiselt tegi mudelit keerukamaks, seega otsustati seda käesolevas töös mitte kasutada.

Joonise 23 parempoolselt graafikult näeme ka seda, et parameetrite arv võiks jääda nelja ja kuue vahele, sest rohkemate muutujate lisandumisel mudeli täpsus oluliselt ei parane.

4.1 Sisendparameetrite valik hoonete mudelite jaoks

Käesolevas töös kasutan XGBoost masinõppe mudeli treenimiseks peatükis 2 toodud Hoonetele A, B ja C arvutatud 2022. aasta jahutuskraadtunde, 2022. aasta jahutuskooormuse andmeid ning 2022. aasta Tallinn-Harku ilmajaama tunni keskmise välisõhu temperatuuri andmeid.

Lisaks ülaltoodud artiklis toodud tähtsamatele sisendparameetritele toon mudelisse ka ilmasteenistuse ilmaandmetest 2022. tunni keskmise summaarse kiirguse Tallinn-Harku ilmajaama põhjal ning vastavalt päeva tunni numbri (0 kuni 24) igale tunnile aastas.

Hoonete A, B ja C pindala ja ruumala ehk S/V suhet ei ole eraldi hoonete hindamisel vajalik tuua, kuna need on konstandid iga hoone puhul. Juhul kui treeniks XGBoost mudelit, kus on korruga kõik kolm hoonet, siis muutub S/V suhe vajalikuks sisendparameetriks. Tunni number annab indikatsiooni, kas on päev või öö, ning sellest lähtuvalt peaks tunni number andma mudelile sisendi, kas parasjagu hoones on inimesed sees või mitte.

Andmed, mille vastu mudelit kontrollin, on 2023. aasta hoonete jahutuskooormuse andmed tabelis 2 toodud nädalatel.

Kokkuvõetult kasutan masinõppe mudeli treenimiseks järgmisi sisendparameetreid:

- Päeva tunni number igale tunnile 2022. aastast;
- 2022. aasta Tallinn-Harku ilmajaama tunni keskmised välisõhu temperatuurid;
- 2022. aasta hoonete A, B ja C-le arvutatud jahutuskraadtunnid;
- 2022. aasta Tallinn-Harku ilmajaama tunni keskmine summaarne kiirgus;
- 2022. aasta jahutuskooormuse tunni andmeid mudeli treenimiseks.

4.2 Hoonete mudelite treeningparameetrid

Enne, kui saame hakata treenima lõplikku mudelit iga hoone jaoks, peame otsima parimaid suurusjärgnevalt toodud põhilistele treeningparameetritele:

- Otsustuspuude arv (XGBoost dokumentatsioonis `n_estimators`);
- õppimise kiirus (XGBoost dokumentatsioonis `learning_rate`);
- otsustuspuude sügavus (XGBoost dokumentatsioonis `max_depth`) ehk kui mitu kihti otsustuspuu lehti võib tekkida;
- minimaalne kaotusfunktsiooni parendus, et otsustuspuu „leht“ uuesti jaotada kaheks leheks (XGBoost dokumentatsioonis `gamma`);

- regulariseerimistermini $L1$ kaal mudelis (XGBoost dokumentatsioonis alpha), tuntud ka kui *Lasso* regressiooni regulariseerimise meetod [26];
- regulariseerimistermini $L2$ kaal mudelis (XGBoost dokumentatsioonis lambda), tuntud ka kui *Ridge* regressiooni regulariseerimise meetod [27].

Parimaid treeningparameetrite suurusi saame otsida, kasutades Pythoni moodulit scikit-learn [28], kus on toodud funktsionaalsus GridSearchCV. See aitab otsida erinevate parameetrite kombinatsioonide parimat tulemust eelnevalt defineeritud kriteeriumi alusel. Meie mudeli puhul on etteantud kriteeriumiks ruutkeskmine viga. Iga parameetri puhul anname programmile ette vahemiku ning funktsionaalsus GridSearchCV leiab parimad väärtused.

Ühe kuu jooksul korratud treeningparameetrite otsingu tulemusel osutusid algandmete vähesuse tõttu parimateks treeningparameetriteks XGBoost masinõppe mooduli vaikeväärtused, mis on toodud tabelis 5 [29]:

Tabel 5. Parimad XGBoost treeningparameetrite vaikeväärtused

XGBoost Parameeter	Vaikeväärtus	XGBoost Parameeter	Vaikeväärtus
N_estimators	100	Gamma	0
Learning_rate	0,3	Alpha	0
Max_depth	6	Lamda	1

Taoline mõneti üllatav tulemus on selgitatav järgmiste asjaoludega:

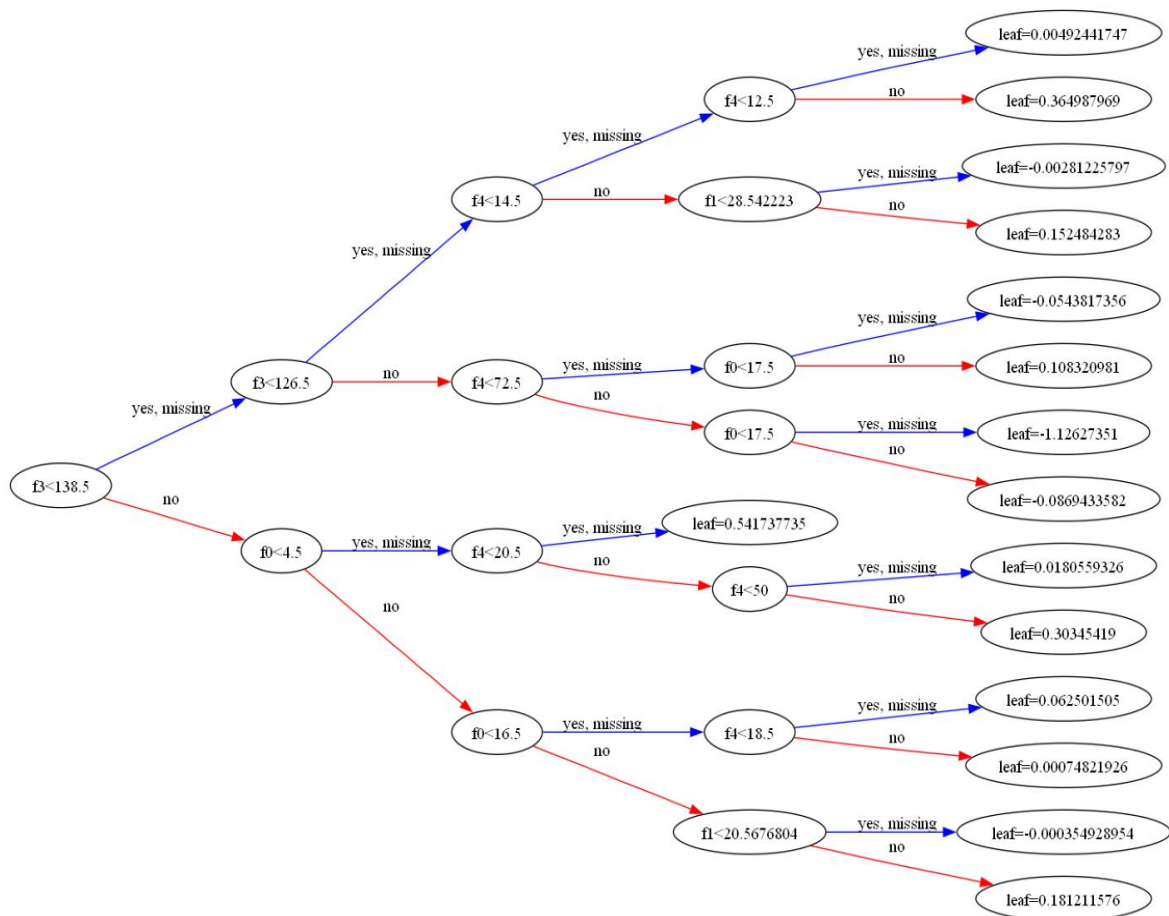
- 1) Algandmeid on vaid ühe aasta kohta, mis tähendab, et treenitav mudel saab algandmeid ainult ühe suvise, kevadise, sügisese ja talvise hooaja kohta. Paremaks ennustamiseks on tarvis andmeid mitme aasta kohta, et tuvastada hooajalisi trende täpsemalt.
- 2) Jahutuskoormuse algandmete ajasamm on ühe tunni kohta, mis on väga pikk periood. Juhendatud masinõppe mudelid on väga head tuvastamiseks peidetud mustreid algandmetes, kuid ühe tunni sees võib hoones väga palju muutuda:
 - a. Automaatika muudab jahutusrežiimi öisest päevaseks;
 - b. toimub suurem koosolek kontoriruumis;
 - c. pilv tuleb päikese eest ära jpm.

Juhendatud masinõppe töötab paremini väiksemate ajasammudega, ideaalis sekundi-mastaabis, aga realistlikult minuti-skaalal, sest muutused sammude vahel ei ole nii drastilised.

- 3) XGBoosti arendajad on pühendanud aastaid vaikeväärtuste optimeerimiseks – alates 13. augustist 2016 [30], mistõttu on väikeste andmekogude treenimiseks vaikeväärtused väga asjalikud.

4.3 Hoonete mudelite treenimine

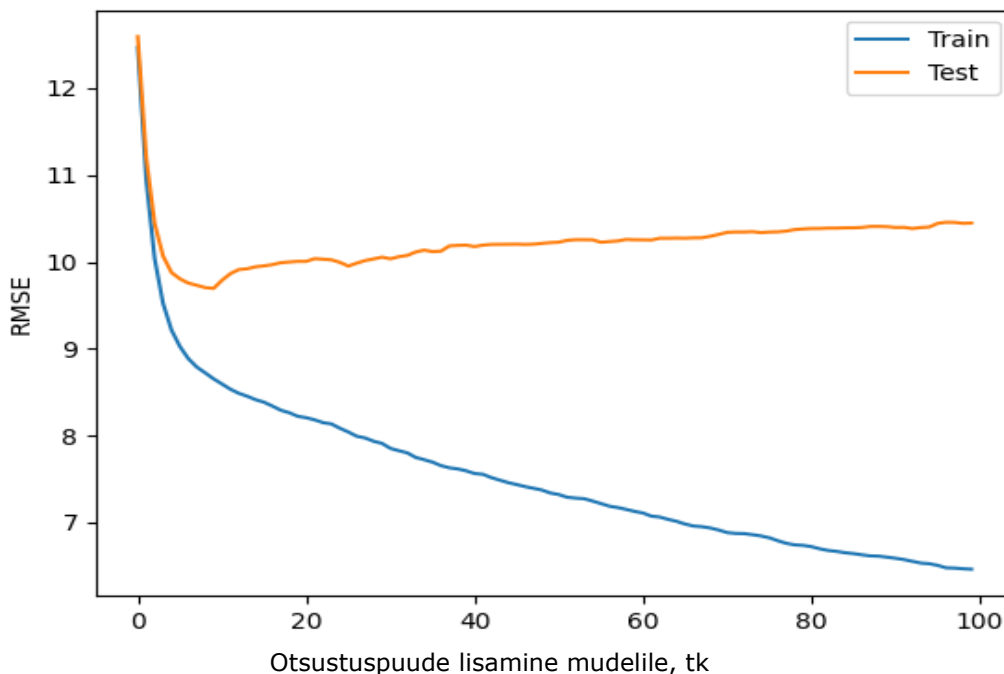
Kasutades tabelis 5 toodud treeningparameetreid, saame treenida iga hoone jaoks eraldi mudeli 2022. aasta jahutuskoormuse andmete põhjal. Treenimisprotsessi iseloomustab hästi joonis 24, kus on kujutatud mudeli poolt treenitud Hoone A ühte otsustuspuud.



Joonis 24. Hoone A üks otsustuspuu mudeli treenimisel

Treenimise efektiivsust saame hinnata läbi keskmise ruutvea (RMSE) paranemise mudelite treenimise jooksul – nii kaua, kuni lisaotsustuspuude lisamine enam mudeli täpsust ei paranda. See treenimise efektiivsus on näidatud joonisel 25, Hoone A mudeli treenimisel. Joonisel 25 võrreldakse treenimise käigus otsustuspuude lisamisel ruutkeskmise vea vähenemist, mida näitab sinine joon (*Train*). Samal joonisel on toodud ka algandmetest ennustuse testimiseks eraldatud andmekogumi ennustuse täpsuse kasv või kahanemine otsustuspuude lisamisel mudelile, mida kirjeldab oranž joon

(Test). Joonisest saab järeldada, et peale 10. kuni 15. otsustuspuu lisamist hakkab testimise andmekogu ennustustäpsus vaikselt langema, samas kui treeningandmete ennustustäpsus suureneb. Seda selgitab algandmete vähesus (üks aasta) ning jahutuskoormuse suur varieeruvus tunnist tundi.



Joonis 25. Hoone A masinõppe mudeli treeningtsükkel 2022. aasta andmete põhjal

Teine väljund, mida treenimise käigus leitakse, on treenitud mudeli sobivus 2022. aasta algandmetest lahku löödud testandmete kogumile läbi determinatsioonikordaja R^2 ning keskmise ruutvea, mis on toodud tabelites 6 ja 7.

Tabel 6. Hoonete A, B ja C treenitud masinõppe mudelite 2022. aasta testandmete keskmised ruutvead

	Hoone A	Hoone B	Hoone C
keskmise ruutvea	10,45	31,45	74,73

Tabel 7. Hoonete A, B ja C treenitud mudeli sobivus 2022. aasta andmetele

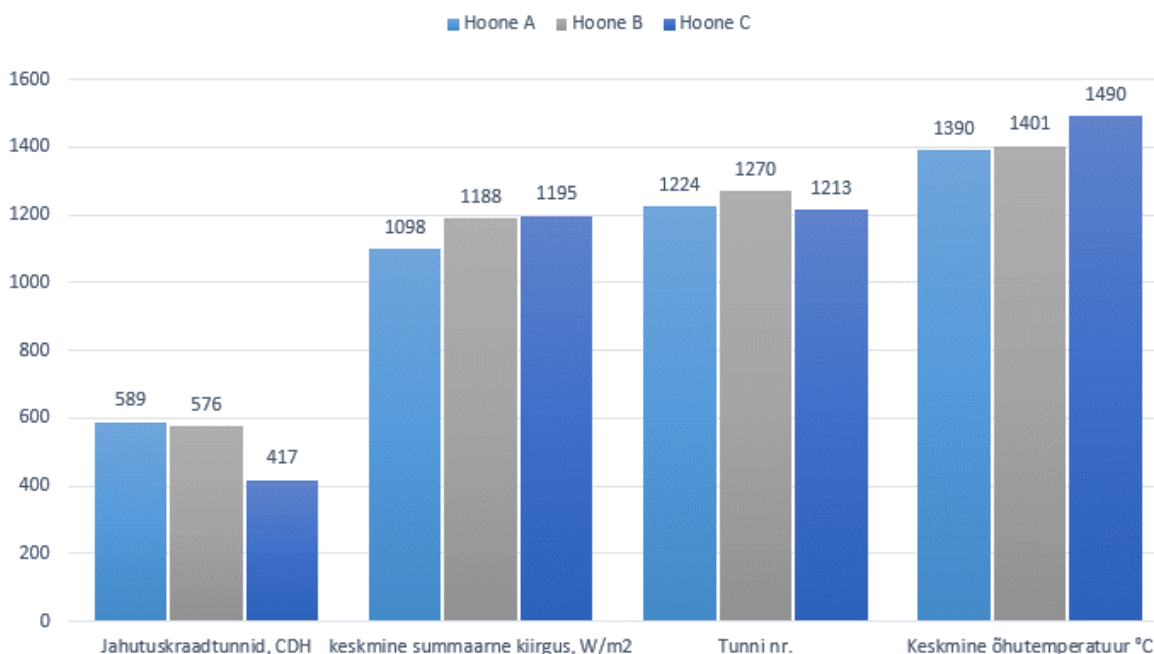
	Hoone A	Hoone B	Hoone C
XGBoost R^2	52,32 %	75,65 %	76,00%
Regressioonanalüüsi R^2	56,75 %	70,43 %	84,00 %

Tabeli 7 põhjal näeme, et Hoone B puhul on võrreldes peatükis 2.2 toodud determinatsioonikordajatega XGBoost mudeli determinatsioonikordaja üle 5 protsendipunkti võrra parem, mis on märgatav täpsuse tõus. Küll aga hoonete A ja C

puhul on determinatsioonikordaja isegi mõned protsendipunktid langenud. Seetõttu saab järeldada, et Hoone C jahutuskoormuse sõltuvus on küllaltki hästi ennustatav ainult välisõhu temperatuuriga ning Hoone A puhul ei leidnud mudel sisendparameetritest piisavalt hästi seost jahutuskoormusega. Seda selgitab algandmete vähesus ning hoone jahutuskoormuse sõltumatuse määr välisteguritest.

4.4 Sisendparameetrite tähtsus mudelis

XGBoosti üheks funktsionaalsuseks on sisendparameetrite tähtsuse arvutamine treenitud mudelis. Sellest lähtuvalt saame hinnata, kui tähtis üks või teine parameeter oli. Kasutades peatükis 4.1 kirjeldatud sisendparameetreid, saame iga hoone jaoks joonisel 26 kujutatud sisendparameetrite tähtsused 2022. aasta andmete põhjal treenitud mudelis. Joonisel toodud suurused määrab XGBoost arvutades, kui palju iga sisendparameeter osaleb otsustuskohtades ning kui tähtsad on otsused kaotusfunktsiooni vähendamiseks [31]. Numbrite enda väärtus pole tähtis, kuna suurus sõltub otsustuspuude arvust. Tähtis on numbrite võrdlus omavahel.

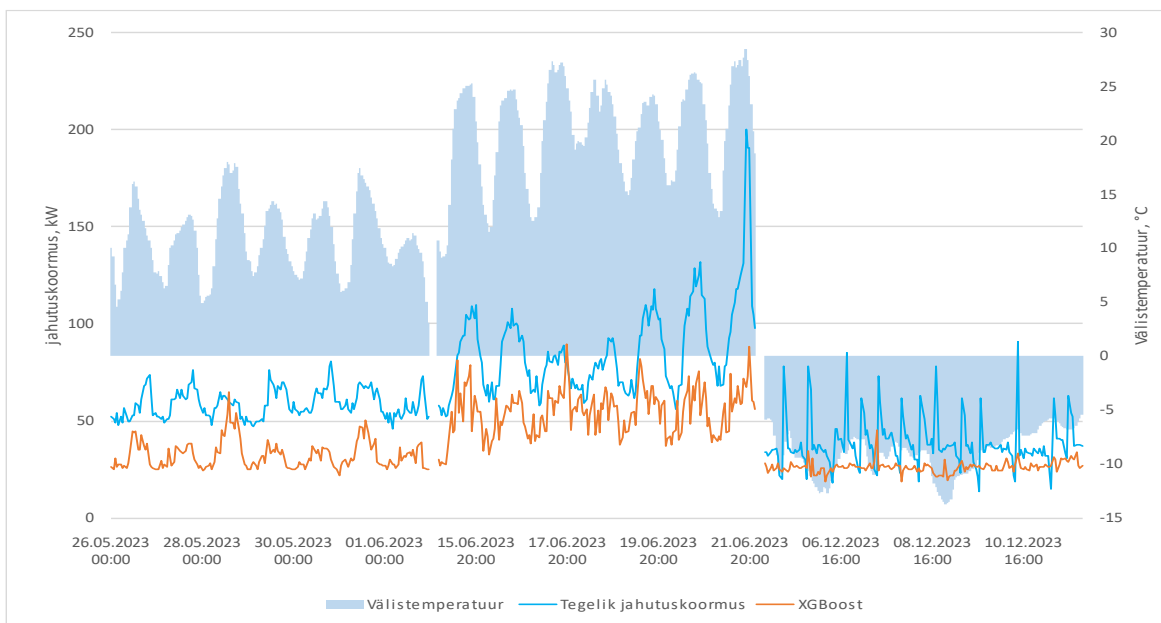


Joonis 26. Hoonete A, B ja C sisendparameetrite tähtsus mudelis

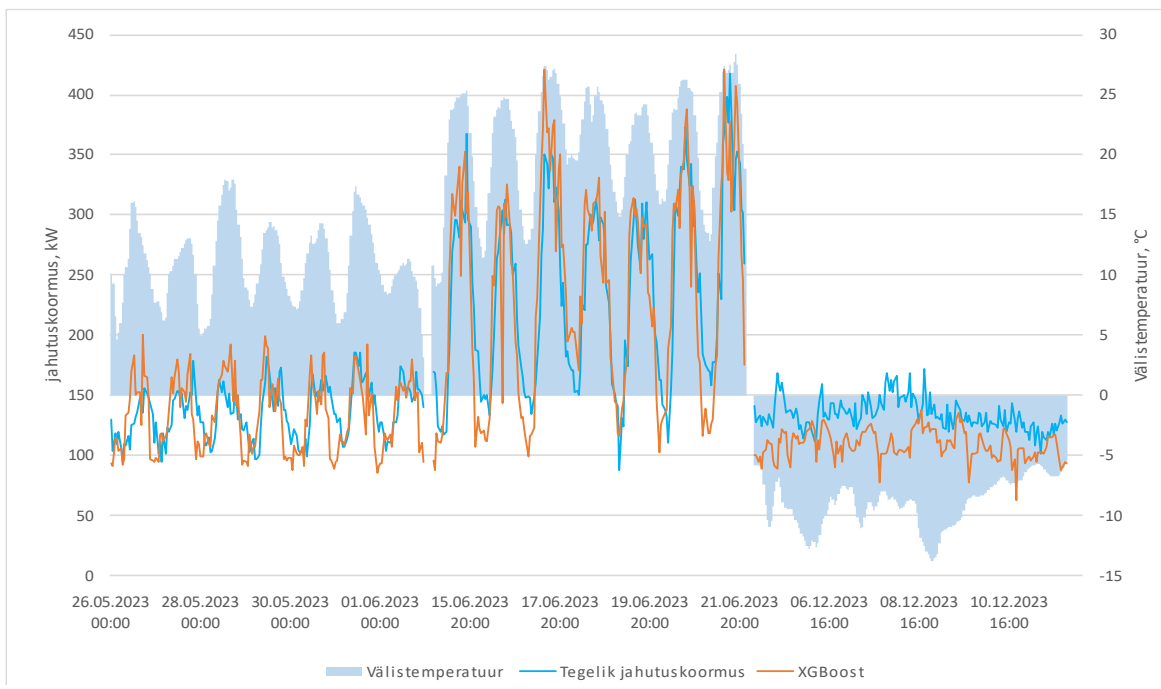
Joonisest 26 saame järeldada, et iga hoone iseloom on veidi erinev ning jahutuskoormus sõltub erinevatest sisendparameetritest. Hoone C puhul on selgelt suurim mõju välisõhutemperatuuril, kuid erinevalt Hoonest A ja B, on Hoone C puhul jahutuskraadtundide mõju väikseim.

4.5 Treenitud mudeli kasutamine ennustuste tegemiseks

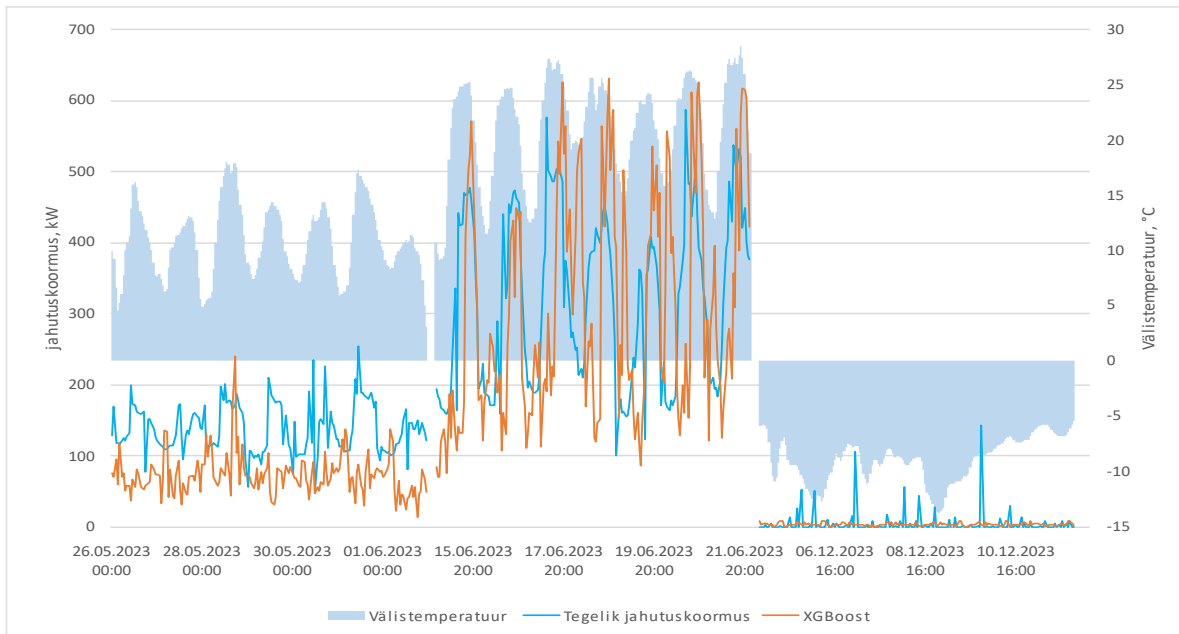
Ennustades punktis 2.3 toodud tabeli 2 kolme nädala perioodi jahutuskoormusi, saame võrrelda masinõppe meetodil koostatud ennustusi tegeliku jahutuskoormusega. Tulemused on toodud joonistel 27 kuni 29 ning lisas 1, kus eristuvad visuaalselt paremini erinevused masinõppe ennustuste ja tegelike jahutuskoormuse vahel.



Joonis 27. Hoone A juhendatud masinõppe meetodil leitud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskoormusega



Joonis 28. Hoone B juhendatud masinõppe meetodil leitud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskoormusega



Joonis 29. Hoone C juhendatud masinõppe meetodil leitud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskoormusega

Nagu ülaltoodud joonistelt 27 kuni 29 näha, siis Hoone A puhul ennustab masinõppe meetod jahutuskoormuseks väiksemat väärtust, kui tegelikult kujunes.

Hoone B puhul on olukord parem ning ennustusgraafik vastab päris täpselt tegelikule jahutuskoormusele, kuid talvise perioodi koormuse puhul ennustab mudel väiksemat jahutuskoormust, kui see tegelikkuses oli.

Hoone C puhul on mudeli ennustus väga hüplik ning palju suurema amplituudiga kui teiste hoonete puhul. Talvisel perioodil ei ennusta mudel üldse tarbimist ning kevadisel perioodil mudel alahindab tarbimist. On näha, et mudel toetub väga tugevalt välisõhu temperatuurile, sest suvisel perioodil ennustab see suuremat jahutuskoormust, kui tegelikult kujunes. Huvitav on see, et mudel ennustab jahutuskoormust kasvavalt, mil suurim jahutuskoormus on päeva lõpus. Seda võib selgitada hoone struktuuri soojenemisega päeva jooksul, mis päeva lõpuks käitub kui sisemine soojusallikas. Erinevalt hoonetest A ja B, on hoonel C palju klaaspinda, mis taolist käitumist soodustab.

Nagu on toodud välja peatükis 2.3, on jahutuskoormuse ülehindamine kaugjahutussüsteemi operatori vaatest parem kui alahindamine. Alahindamise puhul, tootes jahutusenergiat vastavalt ennustatud graafikule, jääks kaugjahutuse klientidel jahutusenergiast puudu, mis tähendab, et kaugjahutussüsteemi operاتور ei täida oma kohustust igal ajahetkel hoone jahutusenergia vajadust rahuldada. Ülehindamise puhul

on süsteemis pinget vähem ning operaator saab tootmist pikemalt ette planeerida, tootes jahutusenergiat tundidel, kui elekter on odavam.

Sarnaselt peatükile 2.3, toon tabelis 8 välja joonistel 27 kuni 29 ennustatud jahutuskooormuste väärtuste ruutkeskmised vead iga hoone ja perioodi kohta, kasutades valemit 1. Peatükis 5 on võrreldud saadud tulemusi regressioonanalüüsi meetodil tehtud ennustuste ruutkeskmise vigadega täpsemalt.

Lisas 2 on toodud 2022. aasta algandmed masinõppe mudeli treenimiseks, 2023. aasta sisendandmed ennustuste tegemiseks ning tegelike jahutuskooormuste ja masinõppe ennustuste tulemused.

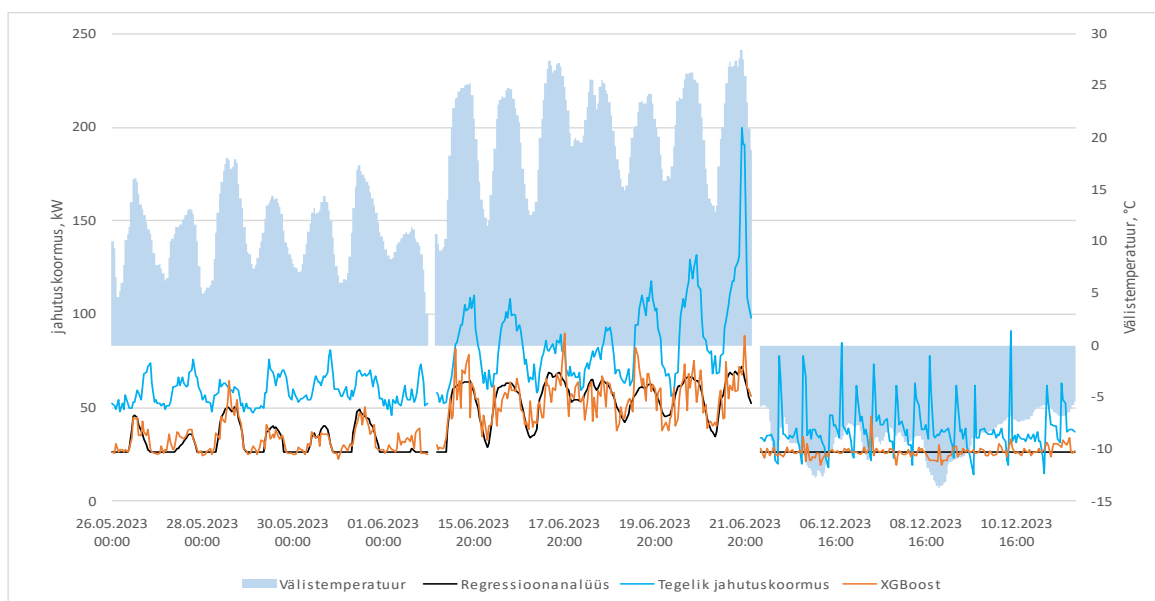
Lisas 3 on toodud masinõppe mudeli Pythoniga treenimise failid ning treenitud mudelid.

Tabel 8. Juhendatud masinõppe meetodil leitud ennustusväärtuste ruutkeskmised vead

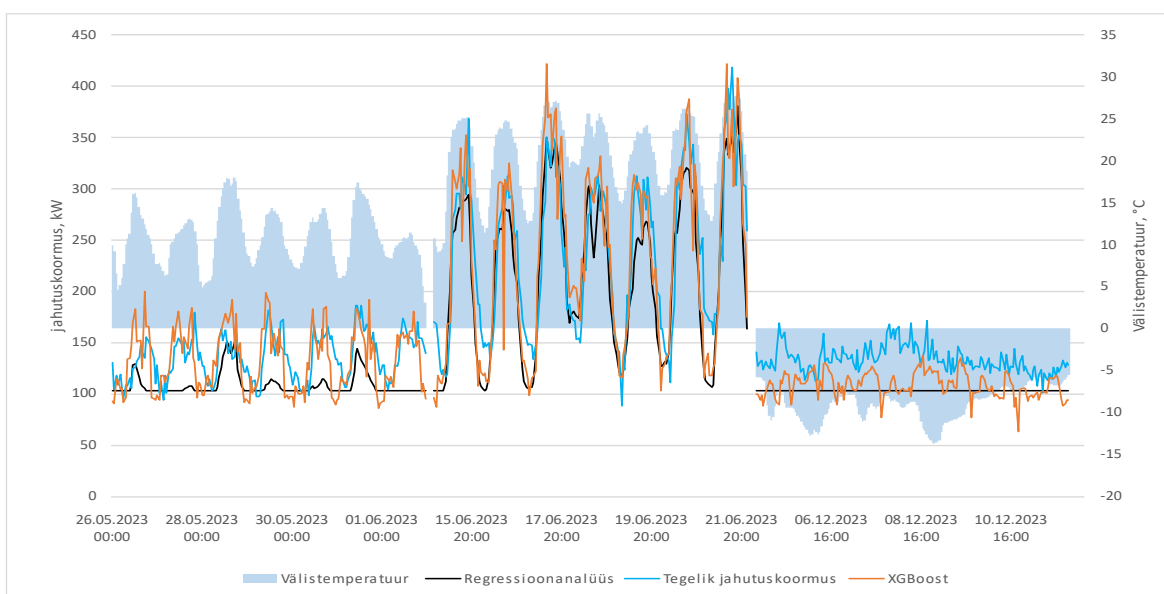
	Hoone A	Hoone B	Hoone C
26.05.2023 - 01.06.2023			
RMSE	36,85	45,48	144,61
15.06.2023 - 21.06.2023			
RMSE	17,95	29,75	16,39
05.12.2023 - 11.12.2023			
RMSE	27,63	24,80	79,68

5. REGRESSIOONANALÜÜSI JA JUHENDATUD MASINÕPPE VÕRDlus JA TULEMUSED

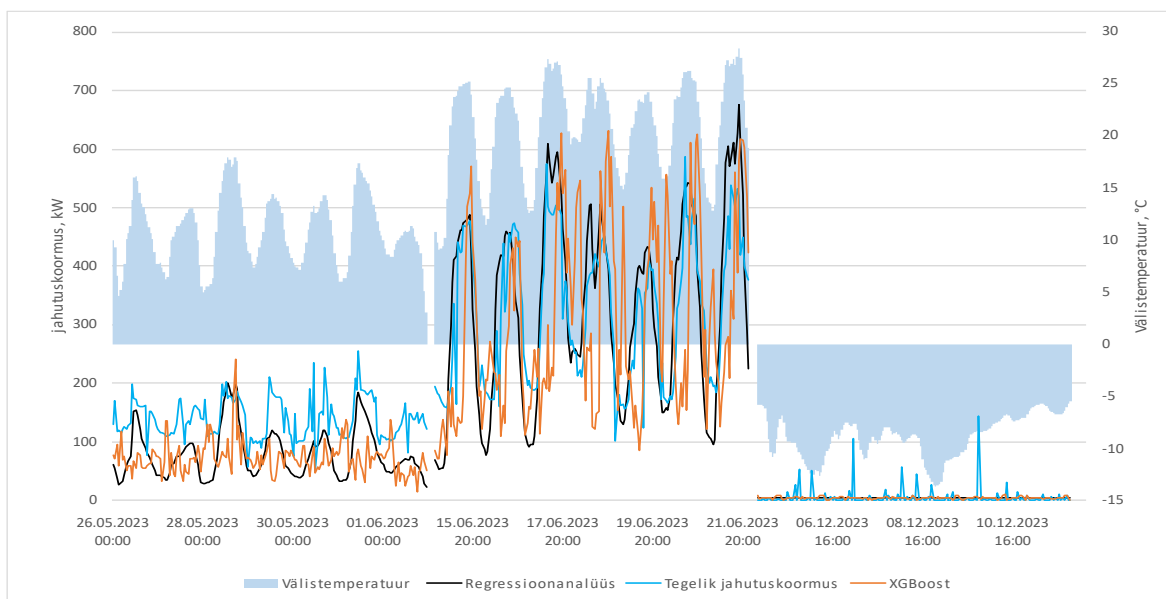
Käesolevas töös on ennustatud tarbijate jahutuskoormuste tulemused peatükis 2.3 regressioonanalüüsi meetodil ning peatükis 4.5 juhendatud masinõppe meetodil. Saadud tulemused on koos näidatud joonistel 30 kuni 32, mis on toodud suuremalt lisas 1, kus eristuvad visuaalselt erinevused regressioonanalüüsi ja masinõppe vahel paremini. Lisaks on leitud mõlema ennustusmeetodi jaoks igale hoonele kõigi kolme perioodi jooksul ruutkeskmiste vead, võrreldes tegeliku jahutuskoormusega, mis on toodud tabelis 9.



Joonis 30. Hoone A kaugjahutuse tarbimiskoormuse ennustused regressioonanalüüsi ja masinõppe meetoditega, võrrelduna tegelikkusega



Joonis 31. Hoone B kaugjahutuse tarbimiskoormuse ennustused regressioonanalüüsi ja masinõppe meetoditega, võrrelduna tegelikkusega



Joonis 32. Hoone C kaugjahutuse tarbimiskoormuse ennustused regressioonanalüüsi ja masinõppe meetoditega, võrrelduna tegelikkusega

Tabel 9. Kahe ennustusmeetodi ruutkeskmiste vigade võrdlus erinevatel perioodidel

26.05.2023 - 01.06.2023			
	Hoone A	Hoone B	Hoone C
RMSE Reg. analüüs	28,44	34,03	67,15
RMSE XGBoost	27,63	24,80	79,68
Vahe	-0,80	-9,23	12,53
15.06.2023 - 21.06.2023			
	Hoone A	Hoone B	Hoone C
RMSE Reg. analüüs	36,25	40,56	88,14
RMSE XGBoost	36,85	45,48	144,61
Vahe	0,60	4,92	56,48
05.12.2023 - 11.12.2023			
	Hoone A	Hoone B	Hoone C
RMSE Reg. analüüs	17,83	33,62	16,05
RMSE XGBoost	17,95	29,75	16,39
Vahe	0,12	-3,87	0,33

Kõigi kolme hoone puhul on näha, et regressioonanalüüsi meetodi ennustused pigem alahindavad tegelikku jahutuskoormust, eriti sügis-kevad ning talvistel perioodidel. Küll aga on ennustuste väärtused vähem varieeruvad ning eksimused on kokkuvõttes väiksemad kui masinõppe meetodi puhul, kuna üksikute tundide eksimused on väiksemad jahutuskoormuse alahindamise tõttu.

Juhendatud masinõppe meetodi puhul on näha, et meetodi ennustused suvel pigem kergelt ülehindavad tegelikku jahutuskoormust, küll aga on iga hoone puhul väga suur erinevus ning Hoone A puhul on suur alahindamine. See tuleneb sellest, et iga hoone tarbimine sõltub erineval määral välistest teguritest, nagu välistemperatuur, päikesekiirguse intensiivsus jne. Lisaks olid masinõppe meetodile ette antud ainult ühe aasta jahutuskoormuse andmed, mistõttu erinevate aastate isepärad ei jõudnud lõplikku mudelisse. Üksikute tundide eksimused võivad tulla suuremad kui regressioonanalüüsi meetodi puhul, mis tuleneb peamiselt algandmete vähesusest ning liiga suurest ajasammust.

Kaugjahutuse operaatori seisukohalt annavad paremaid tulemusi Hoone B puhul masinõppe ennustused. Hoone A ja C puhul on masinõppe väga suur varieeruvus küll negatiivne omadus, kuid ennustuste kalduvus suveperioodil tarbijate jahutuskoormust ülehinnata on väiksema finantsilise tagajärgjega kui alahindamine.

Võrdluse kokkuvõtteks toob autor välja, et masinõppe meetodi täpsus ajas suureneb lisaandmetega mudeli treenimisel, nagu on toodud peatükis 3.2 masinõppe mudeli valiku aluseks olevas artiklis [24], kuid regressioonanalüüsi meetod andmete lisandumisel läheneb keskmistele väärtustele, seega jahutuskoormuse alahindamine säilib.

5.1 Järeldused

Jahutuskoormuse alahindamise korral peab kaugjahutuse operaator rakendama kiiresti tootmisvõimsused ning suurendama pumpamise kiirust, et vajalik jahutusenergia võimsus jõuaks kiiremini tarbijani ning oleks kasutatav kokkulepitud teenustaseme piires. See tähendab, et kaugjahutuse operaator peab tootma jahutusenergiat ka siis, kui kaugjahutuse peamine sisend – elekter, on kallis. Operaatoril ei jää võimalust ajastada kaugjahutuse tootmist madalate elektrihindade perioodile – seega elektrisüsteemile pannakse juurde lisakoormust perioodil, kui elektritarve nõudlus on juba suur.

Kaugjahutuse tarbijate jahutuskoormuse ennustamise väärtus seisneb selles, nagu toodud ka peatükis 2, et jahutuskoormuse ennustamine annab kaugjahutussüsteemi

valdajale võimaluse planeerida oma tootmist ette ning arvestada jahutusenergia tootmiseks vajaliku elektrienergia hinnaga, vajadusel jahutades kaugjahutusvõrgu eelnevalt mõned kraadid külmemaks või vastupidi, mitte tootes kõrge elektrihinnaga, kui nõudlust antud ajahetkel on vähe. See tagab kaugjahutusvõrgu stabiilsemad töörežiimid ning võimaldab toota jahutusenergiat odavamalt.

Käesolevas töös treenitud masinõppe mudeli täpsus ei piisav, et seda üks-ühele koheselt kasutusele saaks võtta, kuid kasutatud meetodi edasisel arendamisel on jahutuskoormuse ennustamisel suur potentsiaal. Kui regressioonanalüüsi puhul on suuremad aja ühikud eelistatud, siis masinõppe puhul on käesolevas töös kasutatud ühe tunnine ajasamm liiga suur ning hoone reageerimine välistemperatuuri ja päikesekiirguse intensiivsuse muutusele tunnisiseselt jääb tabamata.

Microsoft 365 Exceli põhise regressioonanalüüsi ja juhendatud masinõppe võrdlemiseks ei ole autor kasutanud juhendatud masinõppe võimet teostada aegrea ennustamist. Osaliselt tuleneb see autori programmeerimisoskuste tasemest, kuid peamiselt sellest, et käesolev magistritöö võrdleks sama ennustuspõhimõtet – regressiooni.

Kui aegrea meetodiga ennustamisel on sisendparameetriteks eelmiste tundide sisendparameetrid, sest eelmiste ennustuste väärtused mängivad samuti rolli järgmise ennustuste tegemisel, siis regressiooni meetodil mitte. Regressiooni meetodil on alati ühed sisendparameetrid ning olenemata eelmistest ennustusväärtusest, on ainukesed tähtsad sisendparameetrid ennustatava hetke omad.

Tulevastes uurimistöodes soovitab autor läbi viia sarnane uuring paralleelselt regressiooni ja aegrea meetodil ning võrrelda neid tulemusi omavahel, kasutades erineva varieeruvusega ajasamme – ühest tunnist kuni ühe sekundini.

KOKKUVÕTE

Käesolevas magistritöös oli eesmärk uurida kaugjahutusega liitunud hoonete ehk tarbijate jahutusvõimsuste ennustamist nii regressioonanalüüsi kui juhendatud masinõppega. Ennustusi võrreldi eraldi hoonete tegelike jahutuskooormustega peatükis 2 toodud perioodidel ning ka omavahel.

Töös koguti kokku tunni täpsusega 2022. aasta jahutuskooormuste andmed Hoonete A, B ja C kohta ning tunni täpsusega 2022. aasta andmed summaarse päikesekiirguse, õhutemperatuuri ning igale hoonele vastavate jahutuskraadtundide kohta. Seejärel hinnati regressioonanalüüsi abil, kui hästi on võimalik ülaltoodud piiratud andmemahuga ennustada jahutuskooormusi.

Kaugjahutuse tarbijate jahutuskooormuse ennustamise väärtus seisneb selles, et jahutuskooormuse ennustamine annab kaugjahutussüsteemi valdajale võimaluse planeerida oma tootmist ette ning arvestada jahutusenergia tootmiseks vajaliku elektrienergia hinnaga. Vajadusel saab operaator madala elektrihinna korral jahutada võrgu eelnevalt mõned kraadid külmemaks või vastupidi kõrge elektrihinna puhul tootmise pausile panna, kui nõudlust on antud ajahetkel vähe. See tagab kaugjahutusvõrgu stabiilsemad töörežiimid ning võimaldab toota jahutusenergiat odavamalt.

Töö esimeses osas on antud ülevaade kaugjahutuse süsteemist ning süsteemi jahutuskooormust mõjutavatest teguritest, millest olulisimad on eelkõige hoonete sisekliima nõuded ja välisõhu temperatuur.

Töö teises osas on kirjeldatud kaugjahutuse tarbija jahutuskooormuse sõltuvust välistemperatuurist ning antud ülevaade jahutuse valdkonnas kasutatava jahutuskraadtunni mõistest. Peatükis kirjeldati ka kaugjahutuse võimsuskooormuse ennustamist regressioonanalüüsiga. Lõpuks on võrreldud regressioonanalüüsi käigus leitud valemi ennustusi tegeliku jahutuskooormusega Hoonetele A, B ja C.

Töö kolmandas osas on tutvustatud masinõppe valdkonda, kuidas juhendatud masinõppe töötab ning kuidas otsustusmetsa mudelit õpetatakse matemaatiliste selgituste ja lihtsate näidetega. Lisaks on kirjeldatud kuidas töö autor jõudis kõikide juhendatud masinõppe variantidest teegini XGBoost.

Töö neljandas osas on kirjeldatud detailselt juhendatud masinõppe teegi XGBoost mudeli treenimist ning jahutuskooormuse ennustuse tegemist. On selgitatud, milliseid

sisendparameetreid juhendatud masinõppe mudeli treenimiseks autor on kasutanud. Osas on välja toodud mudeli treenimise käigus tehtavad tegevused ning sisendparameetrite tähtsuse võrdlus koos lõpliku ennustuse tegemisega. Lõpuks on võrreldud mudeli poolt tehtud ennustusi tegeliku jahutuskoormusega Hoonetele A, B ja C.

Töö viiendas osas on võrreldud töö teise ja neljanda osa tulemusi ning tehtud tulemuste võrdlus ja analüüs.

Töö tulemusena võib väita, et nii regressioonanalüüs kui ka juhendatud masinõppe saab ülesandega hakkama rahuldusväärselt, kuid kindlasti on tarvis jahutuse valdkonna kiire iseloomu tõttu väiksema ajasammuga andmeid kui üks tund. Võrreldes regressioonanalüüsi ja juhendatud masinõppet, siis regressioonanalüüs annab stabiilsemaid tulemusi, kuid alahindab tegelikku tarbimist rohkem kui masinõppe. Masinõppega treenitud mudelis on ennustamises rohkem varieeruvust, kuid see ennustab suuremaid koormusi kui regressioonanalüüs ning kohati ülehindab tegelikku koormust.

Kaugjahutuse operaatori seisukohast on väike ülehindamine rohkem eelistatud viga kui alahindamine, sest alahindamise korral peab kaugjahutuse operaator rakendama kiirelt tootmisvõimsused ning suurendama pumpamise kiirust, et vajalik jahutusenergia võimsus jõuaks kiiremini tarbijani ning oleks kasutatav kokkulepitud teenustaseme piires. See tähendab, et kaugjahutuse operaator peab tootma jahutusenergiat ka siis, kui kaugjahutuse peamine sisend – elekter, on kallis. Operaator ei saa ajastada tootmist madalate elektrihindade perioodile ning elektrisüsteemile lisatakse koormust perioodil, kui elektritarve nõudlus on juba suur.

Autor soovib tänada ettevõtet AS Utilitas Tallinn vastutulelikkuse ja koostöö eest. Autor on pälvinud käesoleva magistr töö kirjutamise eest Tallinna Raestipendiumi ning Utilitase Puhta Energia magistriõppe stipendiumi.

SUMMARY

This master's degree thesis' aim was to investigate the prediction of cooling loads of buildings connected to district cooling systems using both regression analysis and supervised machine learning. Predictions were compared separately with the actual cooling loads of Buildings A, B, and C during the periods described in Chapter 2 and among themselves.

The thesis' author collected hourly cooling load data for the year 2022 for Buildings A, B, and C as well as hourly data for total solar radiation, air temperature, and cooling degree hours corresponding to each building. Using the regression analysis, an assessment was made of how well the limited dataset collected can be used to predict cooling loads.

The value of predicting cooling loads for district cooling consumers lies in the ability for the system operator to plan their production in advance and to consider the cost of electricity needed to produce cooling energy. This ensures more stable operation of the district cooling network and allows to produce cooling energy at a lower cost.

The first part of the thesis provides an overview of the district cooling system and the factors affecting the system's cooling load, with the most important being the requirements of building indoor climate and outdoor air temperature.

The second part describes the dependence of district cooling consumer's cooling loads on outdoor temperature and provides an overview of the concept of cooling degree hours used in the cooling field. The second part also describes the prediction of district cooling capacity loads with regression analysis.

In the third part, the field of machine learning is introduced, explaining how supervised machine learning works and how the decision tree model is trained, along with mathematical explanations and simple examples. Additionally, the section describes how the author selected the XGBoost model from all supervised machine learning variants.

The fourth part describes in detail the training of the supervised machine learning library XGBoost model and the making of the cooling load prediction. The section explains which input parameters the author used to train the supervised machine learning model. The section also highlights the activities performed during model training and compares the

importance of input parameters with the final prediction. Finally, it compares the model's predictions with the actual cooling load for Buildings A, B, and C.

In the fifth part, the results of the second and fourth parts of the thesis are compared, and the results are compared and analysed.

As a result of the thesis, it can be stated that both regression analysis and supervised machine learning can manage the task satisfactorily, but it is certainly necessary to have data with a smaller time step than one hour due to the fast nature of the cooling buildings. Comparing regression analysis and supervised machine learning, regression analysis provides more stable results, but underestimates actual consumption more than machine learning. The machine learning trained model has more variability in predictions but predicts higher loads than regression analysis and sometimes overestimates the actual load a small amount.

From the district cooling operator's point of view, a small overestimation is a preferred error over underestimation because in the case of underestimation, the district cooling operator must quickly deploy production capacities and increase pumping speed to ensure that the required cooling energy reaches the consumer faster and is usable within the agreed service level. This means that the district cooling operator must produce cooling energy even when the main input to district cooling, electricity, is expensive. Production cannot be timed to periods of low electricity prices, and additional load is added to the electricity system at a time when electricity demand is already high.

The author would like to thank the company AS Utilitas Tallinn for their cooperation and understanding. The author has been awarded the Tallinn City Government Scholarship and the Utilitas Clean Energy Master's Scholarship for writing this master's thesis.

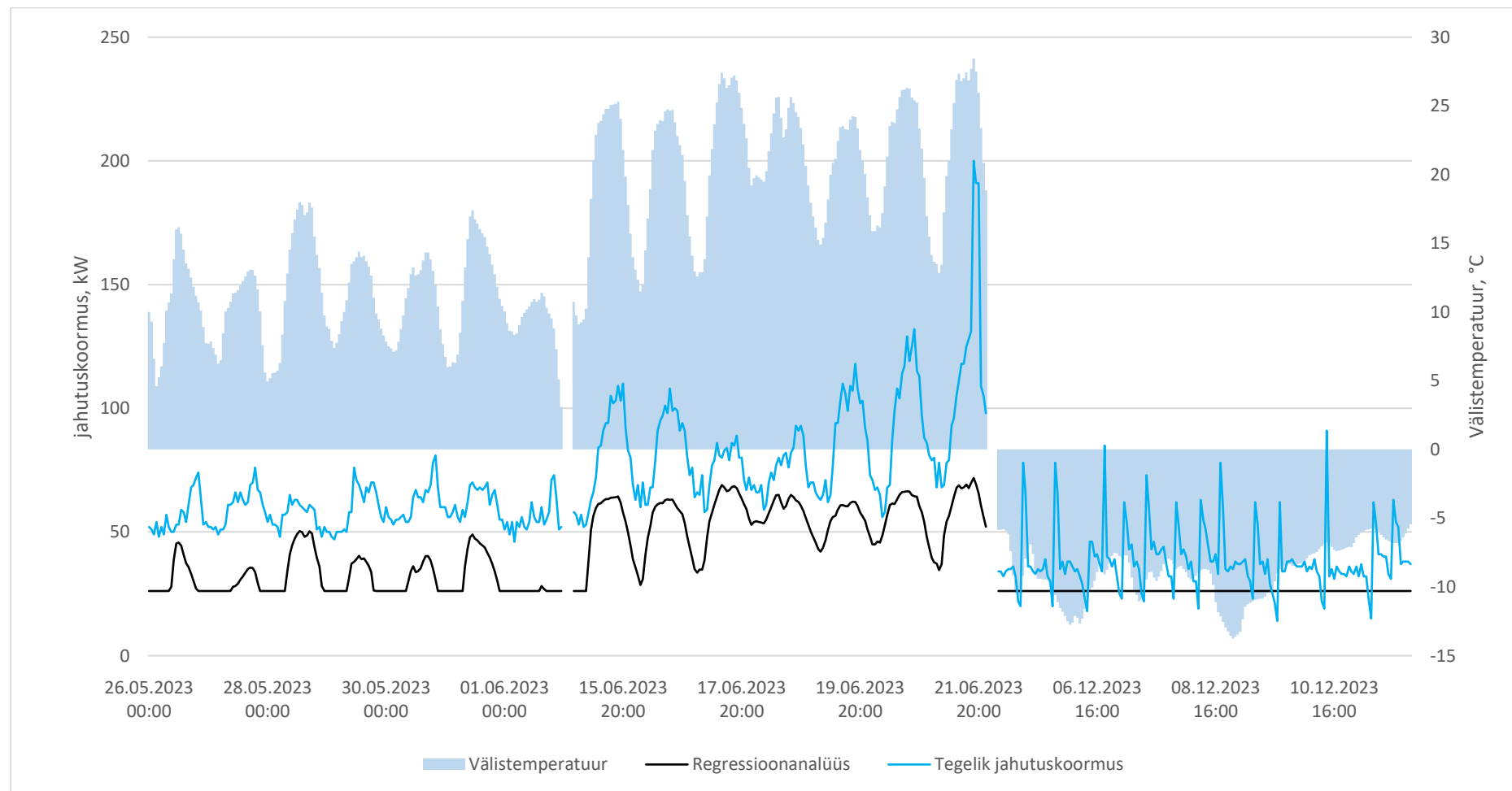
KASUTATUD KIRJANDUSE LOETELU

- [1] Euroopa parlament, „Directive - 2018/2001 - EN - EUR-Lex,“ 11 December 2018. [Võrgumaterjal]. Available: <https://eur-lex.europa.eu/eli/dir/2018/2001/oj>. [Kasutatud 20 04 2024].
- [2] Euroopa Parlament, „Directive - EU - 2023/2413 - EN - Renewable Energy Directive - EUR-Lex,“ 18 oktoober 2023. [Võrgumaterjal]. Available: <https://eur-lex.europa.eu/eli/dir/2023/2413/oj>. [Kasutatud 20 Aprill 2024].
- [3] Gren, „kaugjahutus,“ Gren Holding Company B.V, 2024. [Võrgumaterjal]. Available: <https://gren.com/ee/kaugjahutus/>. [Kasutatud 20 Aprill 2024].
- [4] Tallinna Linnavolikogu, „Tallinna kliimakava,“ 03 Juuni 2021. [Võrgumaterjal]. Available: <https://www.tallinn.ee/et/strateegia/tallinna-kliimakava>. [Kasutatud 20 Aprill 2024].
- [5] T. Kush ja P. Rajesh, „Techno-economic analysis of district cooling system: A case study,“ *Journal of Cleaner Production*, kd. 313, 1 September 2021.
- [6] Vabariigi Valitsus, Ettevõtlus- ja infotehnoloogiainister, „Hoone energiatõhususe miinimumnõuded,“ 8 Juuli 2023. [Võrgumaterjal]. Available: <https://www.riigiteataja.ee/akt/105092012004>. [Kasutatud 5 Mai 2024].
- [7] J. K. M. Thalfeldt, 2016. [Võrgumaterjal]. Available: https://www.rkas.ee/sites/default/files/public-uploaded-files/juhendid/Olemasolevate_hoonete_sisekliima_hindamise_juhend_2016.pdf. [Kasutatud 28 08 2023].
- [8] Majandus- ja taristuminister, „Hoone energiatõhususe arutamise metoodika,“ 05 06 2015. [Võrgumaterjal]. Available: <https://www.riigiteataja.ee/akt/107072020012>. [Kasutatud 28 08 2023].
- [9] Energiatalgud, „Jahutustehnoloogia | Energiatalgud,“ 05 Veebruar 2021. [Võrgumaterjal]. Available: <https://energiatalgud.ee/Jahutustehnoloogia#fn8>. [Kasutatud 28 August 2023].
- [10] Riigi Kinnisvara AS, „Tehnilised nõuded mitteeluhoonetele 2021,“ [Võrgumaterjal]. Available: <https://nouded.rkas.ee/kulmavarustus-ja-jahutus>. [Kasutatud 28 08 2023].
- [11] Tallinn Strategic Management Office, „Climate-neutral Tallinn Tallinn Sustainable Energy and Climate Action Plan 2030,“ [Võrgumaterjal]. Available: <https://www.tallinn.ee/et/media/309750>. [Kasutatud 01 09 2023].

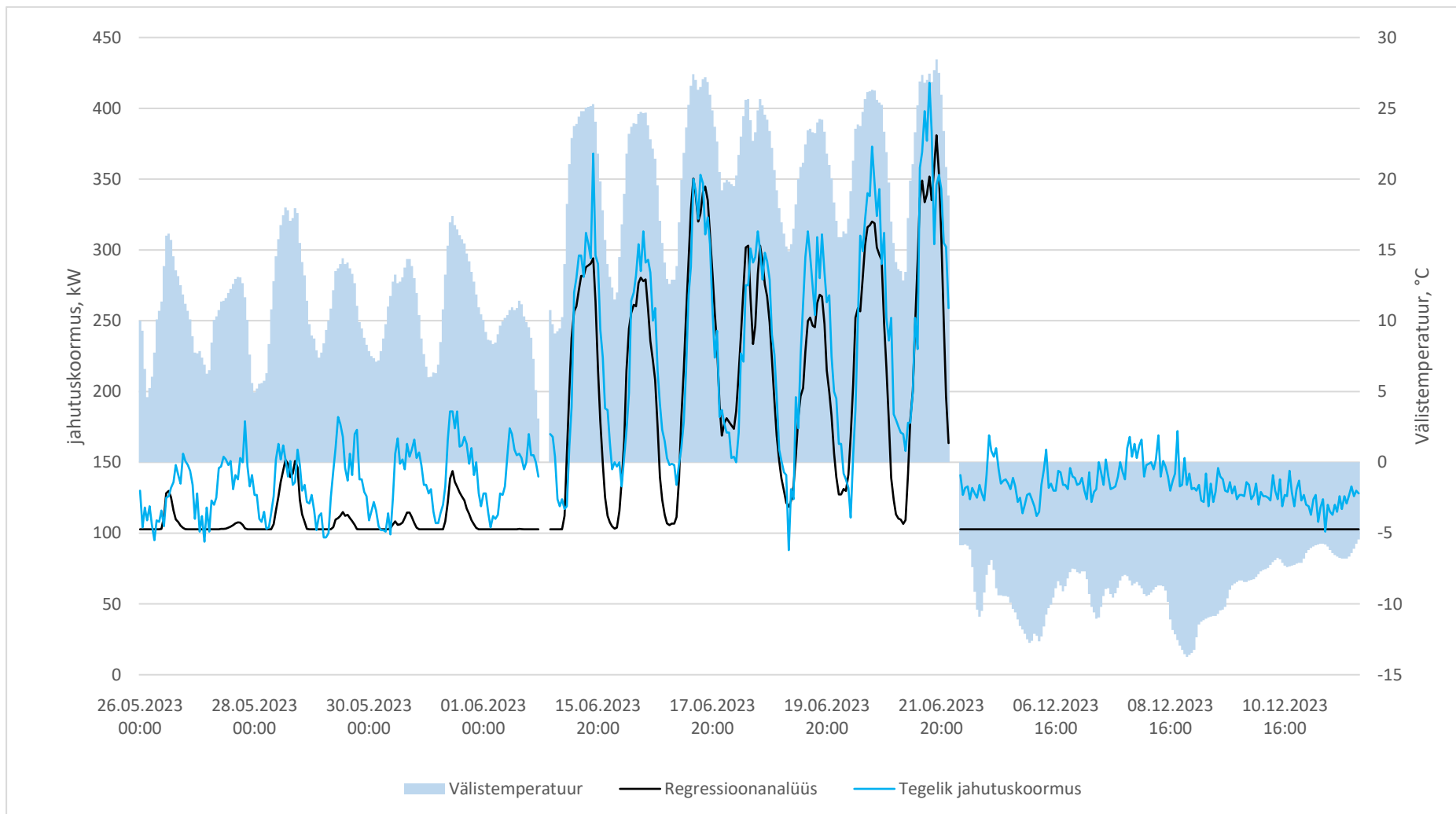
- [12] Copenhagen Centre on Energy Efficiency, „<https://c2e2.unepccc.org/collection/district-cooling-systems-etraining-india/>,“ 03 2022. [Võrgumaterjal]. Available: <https://c2e2.unepccc.org/wp-content/uploads/sites/3/2022/03/module-1-introduction-to-district-cooling.pdf>. [Kasutatud 17 03 2024].
- [13] UN Environment Programme, „DISTRICT ENERGY,“ UN Environment Programme, 2015.
- [14] Utilitas, „Jahutusenergia tootmine,“ Utilitas, 2022. [Võrgumaterjal]. Available: <https://www.utilitas.ee/energia-tootmine/jahutusenergia/#jahutuse-tootmine>. [Kasutatud 07 04 2024].
- [15] E. L. V. M. I. K. a. A. S. A. Volkova, „Kaugjahutus - Jätkusuutlik kaugküte,“ 2022. [Võrgumaterjal]. Available: <https://kaugkute.taltech.ee/kaugjahutus/>. [Kasutatud 01 09 2023].
- [16] N. B. David, P. S. Hannes ja B. Martin, „Estimating the daily peak and annual total electricity demand for cooling in Vienna, Austria by 2050,“ *Urban Climate*, kd. 28, June 2019.
- [17] A. Sauga, Statistika õpik majanduseriala üliõpilastele, Tallinn: TTÜ Kirjastus, 2017.
- [18] O3 Technology, „Ülekuumenemine,“ 2023. [Võrgumaterjal]. Available: <https://o3.ee/ulekuumenemine/>. [Kasutatud 01 09 2023].
- [19] T. M. Michell, Machine learning, New York: McGraw-hill, 1997.
- [20] A. L. Samuel, „Some studies in machine learning using the game of Checkers,“ *IBM JOURNAL OF RESEARCH AND DEVELOPMENT*, kd. 3, pp. 210-229, 1959.
- [21] Datafruit, „Masinõppe lahendused,“ Datafruit, 2021. [Võrgumaterjal]. Available: <https://datafruit.ee/masinoppe-lahendused/>. [Kasutatud 1 Aprill 2024].
- [22] MathWorks, „What Is Machine Learning?,“ The MathWorks, Inc., [Võrgumaterjal]. Available: <https://www.mathworks.com/discovery/machine-learning.html#:~:text=Machine%20learning%20algorithms%20use%20computational,specialized%20form%20of%20machine%20learning..> [Kasutatud 29 03 2024].
- [23] X. arendajad, „Introduction to Boosted Trees,“ 2022. [Võrgumaterjal]. Available: <https://xgboost.readthedocs.io/en/stable/tutorials/model.html>. [Kasutatud 29 03 2024].
- [24] F. Yanxiao, D. Qiuhua, C. Xi ja W. Julian, „Space cooling energy usage prediction based on utility data for residential buildings using machine learning methods,“ *Applied Energy*, kd. 291, 1 June 2021.

- [25] J. H. Friedman, „Greedy function approximation: A gradient boosting machine,” *Ann. Statist.*, kd. 29, nr 5, pp. 1189-1232, Oktoober 2001.
- [26] IBM, „What is lasso regression?,” IBM, 18 Jaanuar 2024. [Võrgumaterjal]. Available: <https://www.ibm.com/topics/lasso-regression>. [Kasutatud 20.04.2024].
- [27] M. Jacob ja K. Eda, „What is ridge regression?,” IBM, 21 November 2023. [Võrgumaterjal]. Available: <https://www.ibm.com/topics/ridge-regression#:~:text=Ridge%20regression%E2%80%94also%20known%20as,for%20multicollinearity%20in%20regression%20analysis..> [Kasutatud 20.04.2024].
- [28] Scikit-learn arendajad, „User Guide,” 2024. [Võrgumaterjal]. Available: https://scikit-learn.org/stable/supervised_learning.html. [Kasutatud 20.04.2024].
- [29] XGBoost arendajad, „XGBoost Parameters,” 2022. [Võrgumaterjal]. Available: <https://xgboost.readthedocs.io/en/stable/parameter.html>. [Kasutatud 21.04.2024].
- [30] C. Tianqi, „XGBoost: A Scalable Tree Boosting System,” %1 *KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California, 2016.
- [31] J. Brownlee, „Feature Importance and Feature Selection With XGBoost in Python,” 27 August 2020. [Võrgumaterjal]. Available: <https://machinelearningmastery.com/feature-importance-and-feature-selection-with-xgboost-in-python/>. [Kasutatud 20.04.2024].
- [32] XGBoost arendajad, „Python Package Introduction,” 2022. [Võrgumaterjal]. Available: https://xgboost.readthedocs.io/en/stable/python/python_intro.html#plotting. [Kasutatud 20.04.2024].

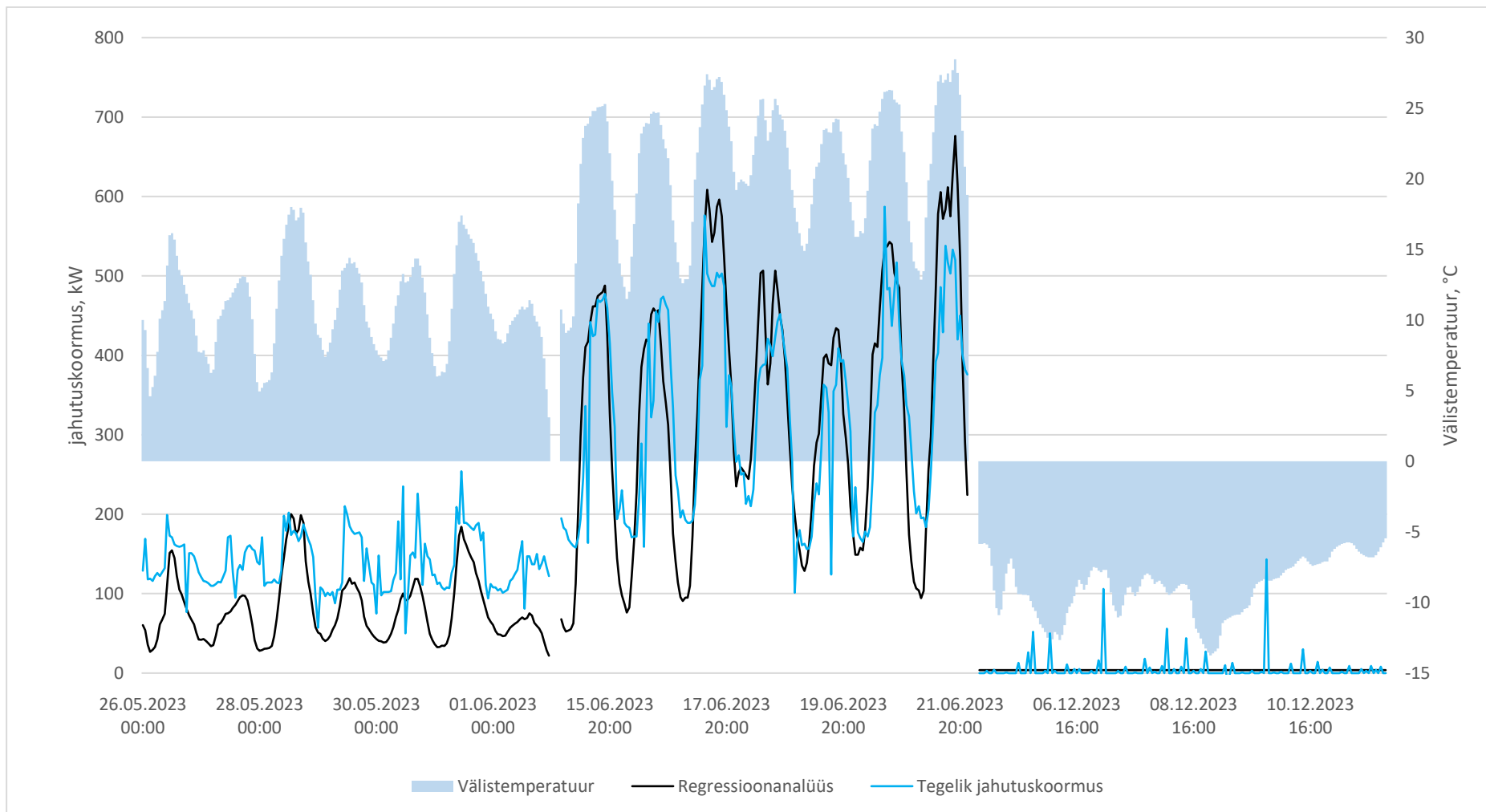
LISA 1



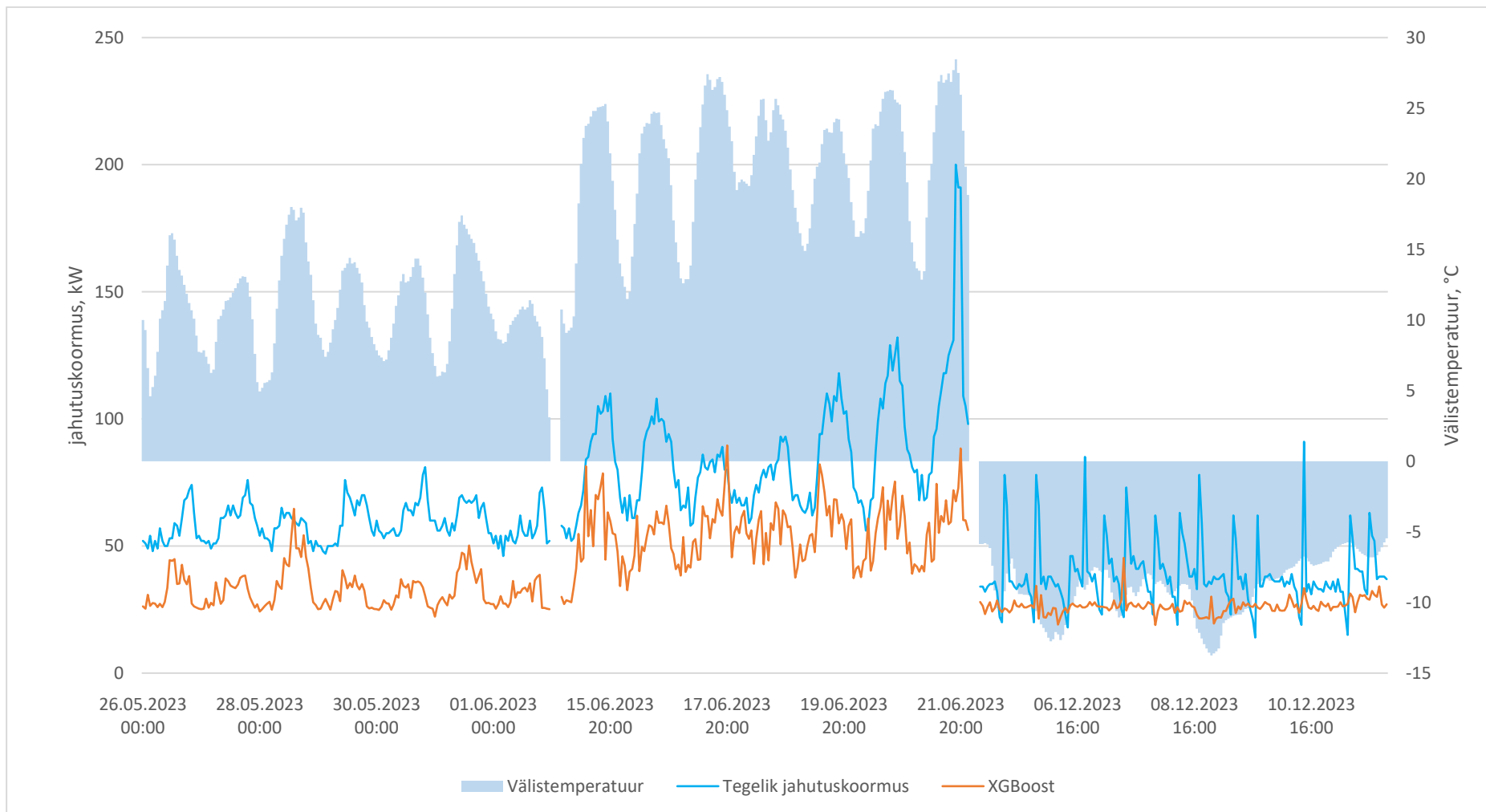
Joonis 15. Hoone A regressioonanalüüsi meetodil leitud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskoormusega



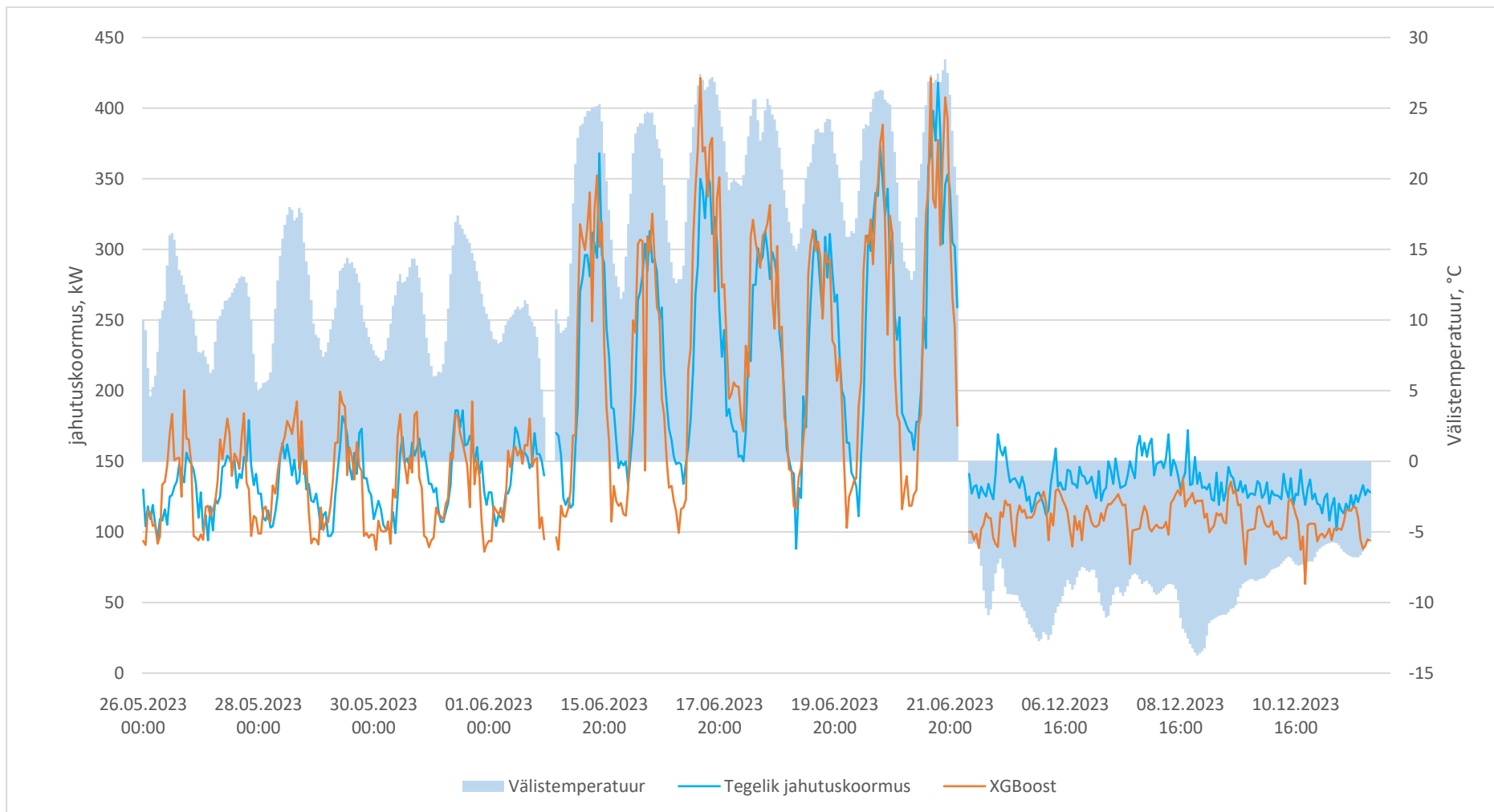
Joonis 16. Hoone B regressioonanalüüsi meetodil leitud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskoormusega



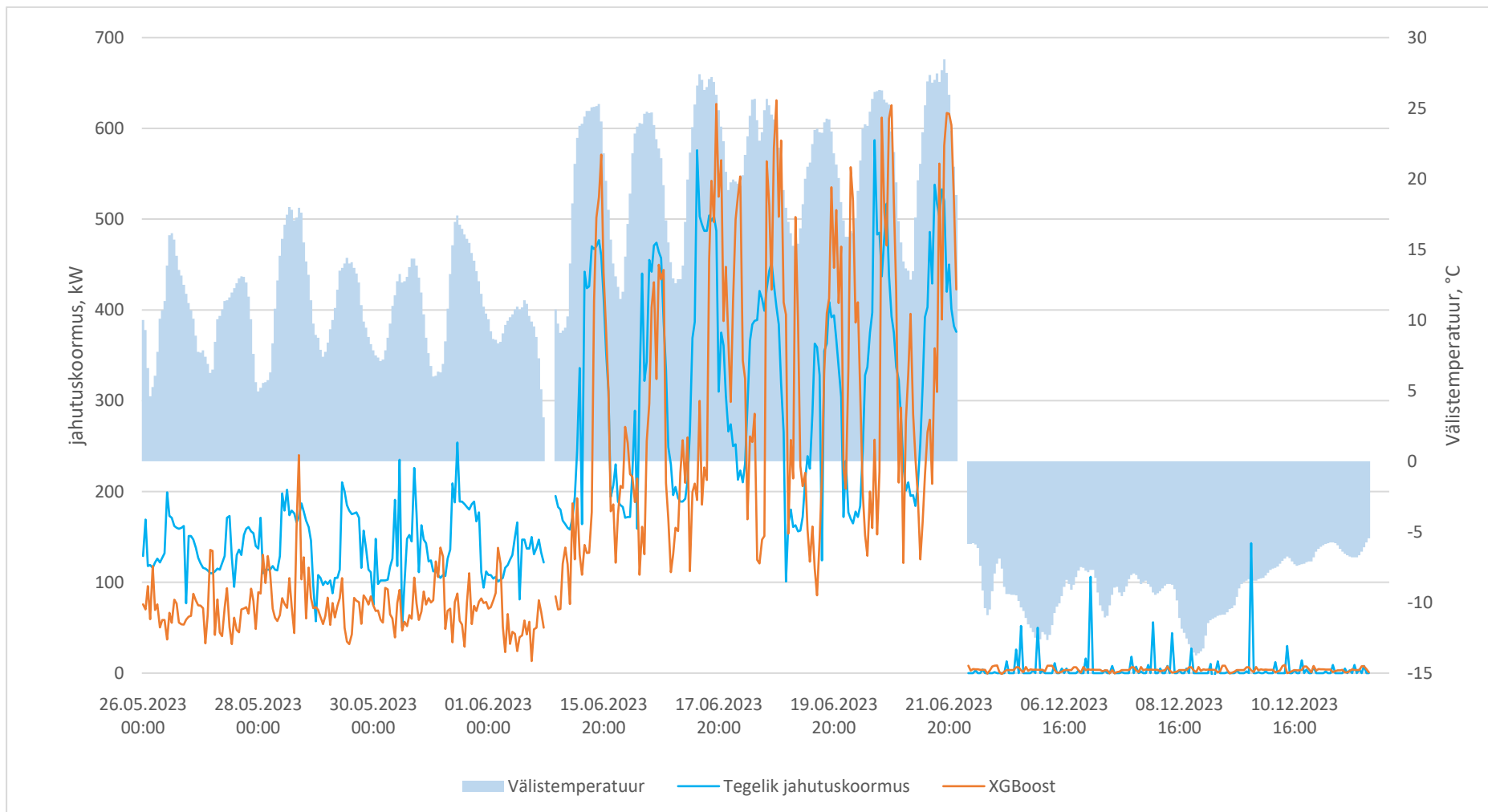
Joonis 17. Hoone C regressioonanalüüsi meetodil leidud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskoormusega



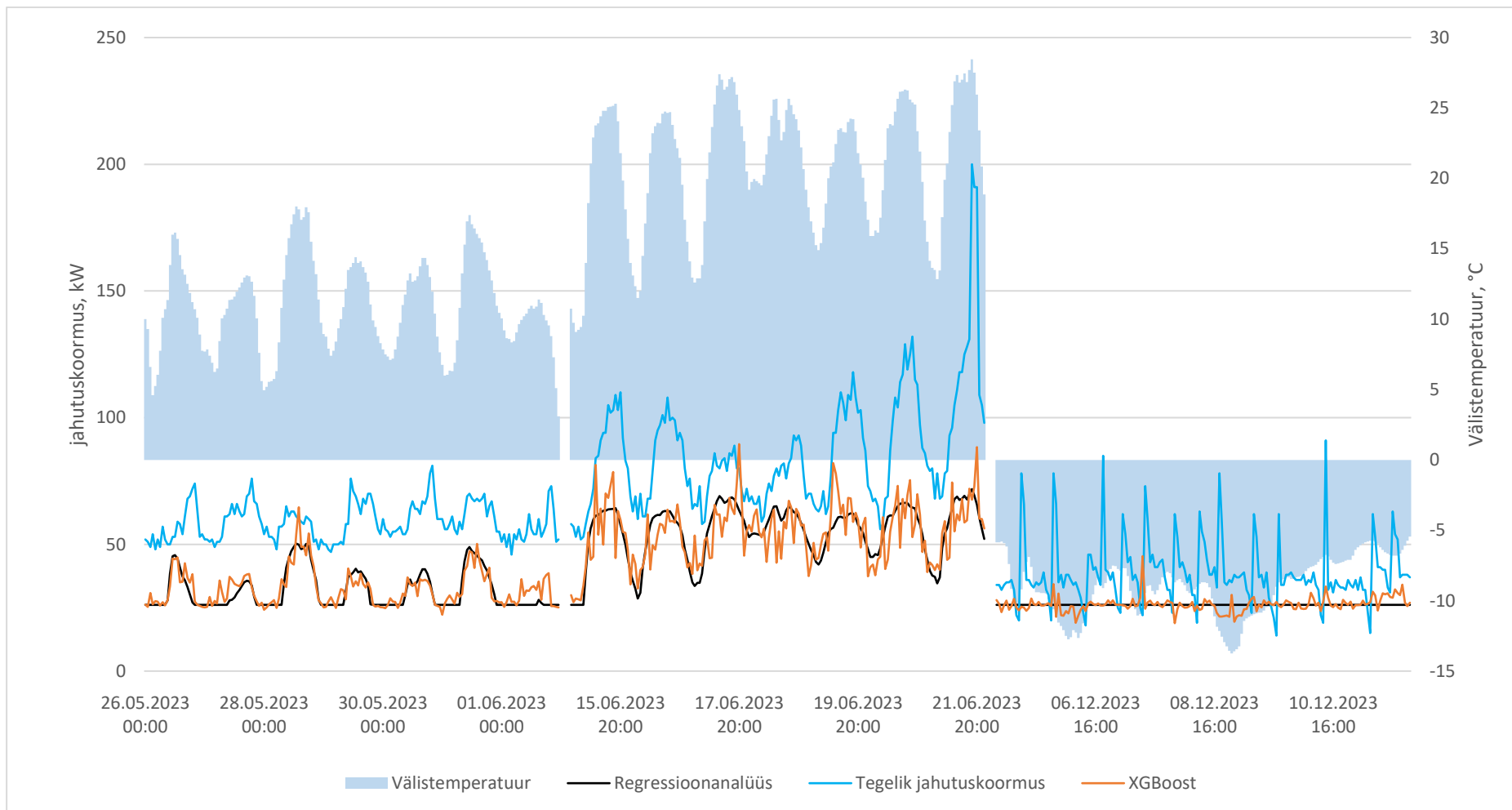
Joonis 27. Hoone A juhendatud masinõppe meetodil leitud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskoormusega



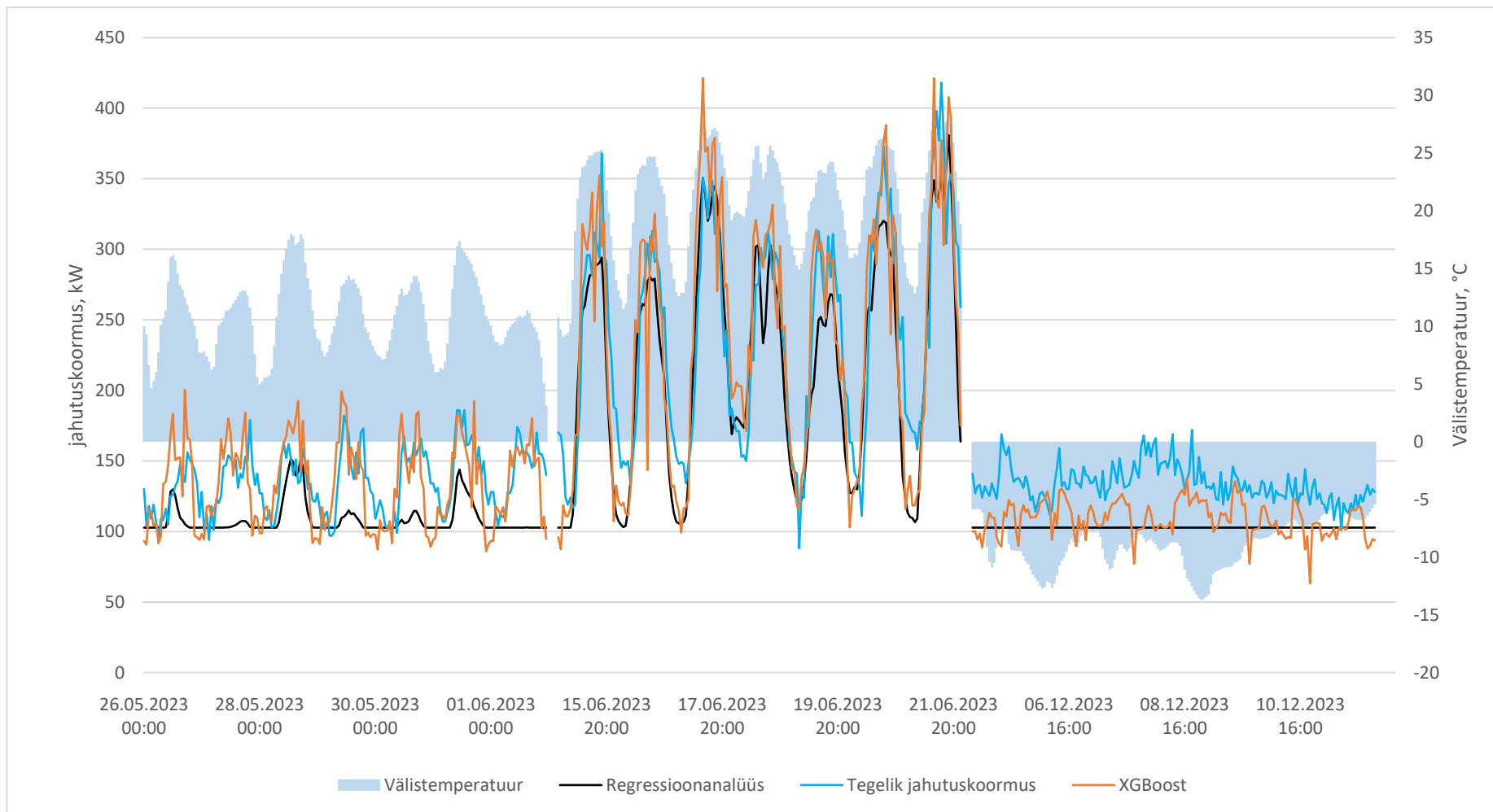
Joonis 28. Hoone B juhendatud masinõppe meetodil leidud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskoormusega



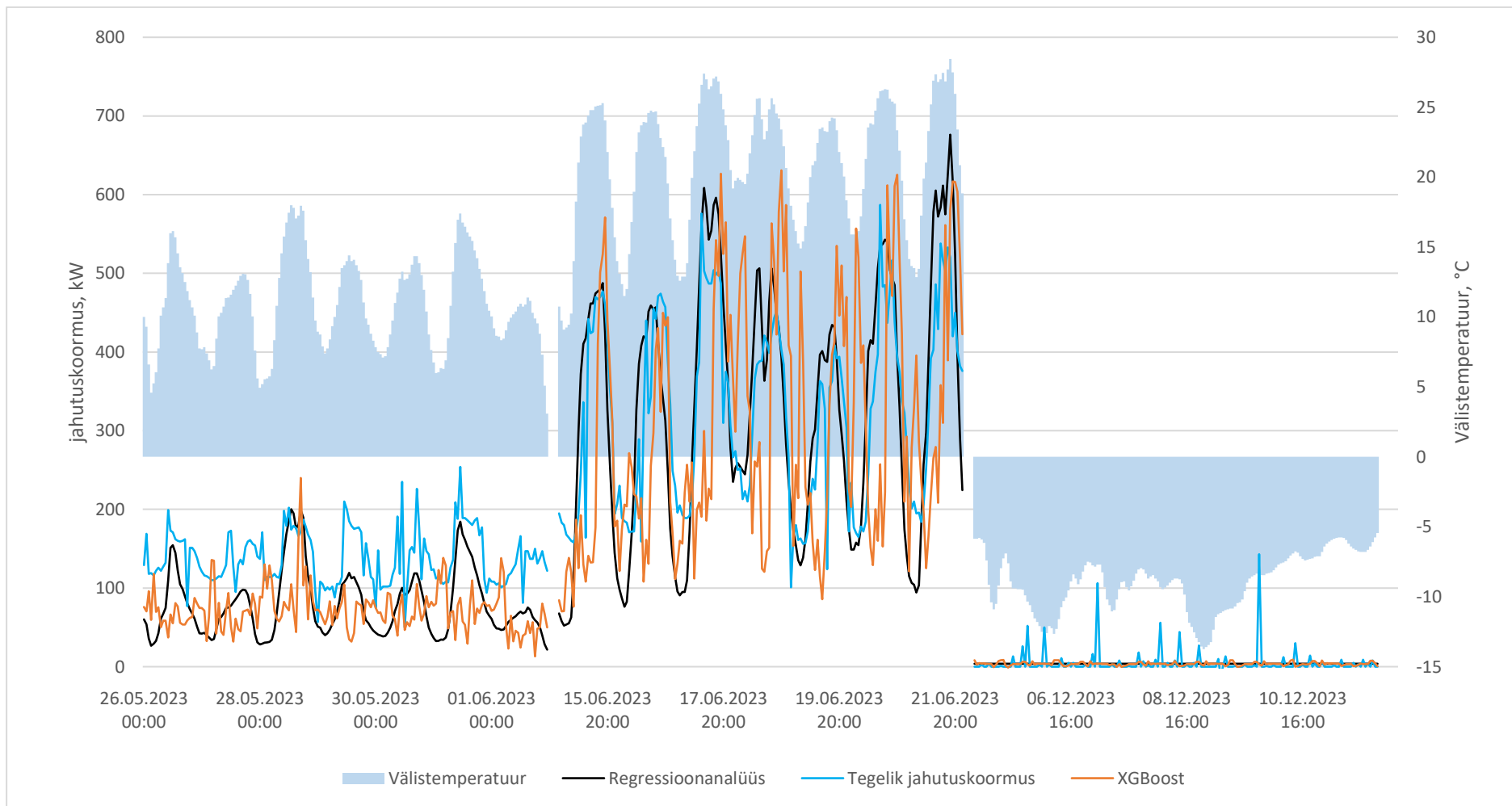
Joonis 29. Hoone C juhendatud masinõppe meetodil leitud funktsiooni ennustus võrreldes tabelis 2 toodud perioodide tegeliku jahutuskoormusega



Joonis 30. Hoone A kaugjahutuse tarbimiskoormuse ennustused regressioonanalüüsi ja masinõppe meetoditega, võrrelduna tegelikkusega



Joonis 31. Hoone B kaugjahutuse tarbimiskoormuse ennustused regressioonanalüüsi ja masinõppe meetoditega, võrrelduna tegelikkusega



Joonis 32. Hoone C kaugjahutuse tarbimiskoormuse ennustused regressioonanalüüsi ja masinõppe meetoditega, võrrelduna tegelikkusega

LISA 2

Hoonete analüüs 2023.xlsx

Hoonete analüüs 2022.xlsx

Cleaned_data_C_2023.xlsx

Cleaned_data_C.xlsx

Cleaned_data_B_2023.xlsx

Cleaned_data_B.xlsx

Cleaned_data_A_2023.xlsx

Cleaned_data_A.xlsx

LISA 3

Hoone_A.joblib

Hoone_B.joblib

Hoone_C.joblib

Hoone_A_Model_Train.py

Hoone_B_Model_Train.py

Hoone_C_Model_Train.py

Prediction_Hoone_A.py

Prediction_Hoone_B.py

Prediction_Hoone_C.py