

TALLINNA TEHNIKAÜLIKOOL  
Infotehnoloogia teaduskond

Kaido Sõmera 212336IADB

# **Eramajapidamise energiatarbimise ennustamine**

Bakalaureusetöö

Juhendaja: Vahur Kotkas  
M.Sc.

Tallinn 2025

## **Autorideklaratsioon**

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

Autor: Kaido Sõmera

31.12.2024

## **Annotatsioon**

Käesoleva töö üldine eesmärk on prognoosida eramajapidamise energiatarbimist selleks, et efektiivselt juhtida eramajapidamise energiasalvestuse seadet ning seeläbi saavutada majanduslik kasu. Konkreetseteks eesmärkideks, mida töö käigus lahendatakse, on prognoosida eramajapidamise energiatarbimist olukorras, kus varasemad tarbimisandmed puuduvad või ei ole piisavalt, millal lugeda tarbimisandmete pikkus piisavaks ja kuidas prognoosida tarbimise tippe.

Tarbimisandmete pikkusega seotud probleemide lahendamiseks pakutakse välja neli lahendust, millest kasutati sarnase mudeli lahendust. Energiatarbimise tippude prognoosimiseks pakuti välja kaks lahendust, millest kasutati prognoosi modifitseerimise lahendust. Prognooside tegemiseks kasutatakse töös XGBoost masinõppealgoritmi.

Töö tulemusena valmis Python keeles loodud programm, mis simuleerib eramajapidamist olukorras, kus tarbimisandmed puuduvad või ei ole neid piisavalt. Töö Lisas 3 on toodud ühe kuu simulatsiooni väljundid. Simulatsiooni tulemused näitavad autori pakutud lahenduse toimimist.

Lõputöö on kirjutatud Eesti keeles ning sisaldab teksti 35 leheküljel, 5 peatükki, 18 joonist, 1 tabelit.

## **Abstract**

### **Predicting Single Household Energy Consumption**

The general objective of the thesis is to forecast private household energy consumption to efficiently manage energy storage devices and thereby achieve economic benefits. The specific objectives addressed in the thesis include predicting household energy consumption when historical consumption data is unavailable or insufficient, determining when the length of consumption data is adequate and forecasting consumption peaks. The thesis provides an overview of what has been previously done in this field in Estonia and globally.

To solve the outlined problems, the author chose the XGBoost machine learning algorithm. The algorithm was selected initially based on a literature review of similar solutions and then validated through practical experimentation.

To address the issue of missing or insufficient consumption data, the thesis proposes four solutions: training a universal model, clustering households based on certain characteristics and training a model for each cluster, using a model trained for a similar household, or employing a trivial model. The work analyses and utilizes solutions involving a similar model.

According to the algorithm for the similar model solution, the system waits until the first 24-hour consumption data is available for the household. Then, this period is forecasted using all available models, and the model with the most accurate prediction is identified. This identified model is then used for subsequent forecasts. When a minimal amount of consumption data becomes available, a new model is trained with the new household's data, the model will be stored in the database and the model will start competing with other models. The application is governed by parameters that determine how often a new model is selected, how often a new model is trained, the length of data required for training, etc.

For forecasting energy consumption peaks, the thesis proposes two solutions: using two separate models where one forecasts base consumption and the other forecasts peak consumption, or adjusting the base consumption forecast afterward by amplifying trends identified by the model. The thesis analysis the trend amplification solution.

When training a new model, features with a correlation to consumption greater than 0,055 are selected from 31 features. Hyperparameters for the XGBoost model are calculated with correlated features, irrelevant features are removed using a recursive feature elimination method, the model is trained and written to the database.

As a result of the thesis, a program written in Python was developed to simulate a household in situations where consumption data is missing or is insufficient. The outputs of a one-month simulation are presented in Appendix 3, demonstrate that the proposed solution works effectively in practice, thereby validating the functionality of the algorithm. The Python source code created during the work is publicly available and can be adapted for practical use with minor modifications.

The thesis is in Estonian and contains 35 pages of text, 5 chapters, 18 figures, 1 table.

## Lühendite ja mõistete sõnastik

ARIMA	<i>Auto Regressive Integrated Moving Average</i>
CPU	<i>Central Processing Unit</i>
IDE	<i>Integrated Development Environment</i>
LSTM	<i>Long Short-Term Memory</i>
MAE	<i>Mean Absolute Error</i>
MSE	<i>Mean Squared Error</i>
PoC	<i>Proof Of Concept</i>
RAM	<i>Random Access Memory</i>
RF	<i>Random Forest</i>
RFE	<i>Recursive Feature Elimination</i>
RFECV	<i>Recursive Feature Elimination Cross-Validation</i>
SARIMA	<i>Seasonal AutoRegressive Integrated Moving Average</i>
SQL	<i>Structured Query Language</i>
SVM	<i>Support Vector Machine</i>
SVR	<i>Support Vector Regression</i>
UTC	<i>Coordinated Universal Time</i>
XGBoost	<i>eXtreme Gradient Boosting</i>

## Sisukord

1 Sissejuhatus .....	11
1.1 Probleemi püstitus .....	11
1.2 Töö skoop .....	12
1.3 Analogsed uurimused Eestis ja Maailmas .....	13
2 Masinõppe algoritmid.....	15
2.1 Masinõppe algoritmi valimine.....	15
2.1.1 Algoritmi valimine kirjanduse põhjal.....	15
2.1.2 Algoritmi eksperimentaalne valimine .....	17
2.2 XGBoost algoritmi kirjeldus .....	18
2.3 Meetrika.....	20
3 Analüüs.....	21
3.1 Energiatarbimise prognoosimine puuduvate andmete olukorras.....	21
3.2 Energiatarbimise tippude prognoosimine .....	24
3.3 Alternatiivsed lahendused.....	26
3.3.1 Üks mudel kõigi jaoks .....	26
3.3.2 Mudelite grupeerimine .....	27
3.3.3 Triviaalne mudel.....	27
4 Eksperiment.....	29
4.1 Majapidamiste kirjeldused.....	29
4.2 Kasutatud riist- ja tarkvara .....	29
4.3 Tarkvara ettevalmistamine .....	30
4.4 Tarbimisandmete analüüs .....	32
4.5 Mudeli treenimine ja prognoosimine.....	35
4.5.1 Andmete ettevalmistamine .....	35
4.5.2 Korrelatsioonide leidmine .....	36
4.5.3 Hüperparameetrite leidmine .....	37
4.5.4 Oluliste tunnuste valimine.....	39
4.5.5 Mudeli treenimine .....	40
4.5.6 Mudeliga prognoosimine.....	40

4.6 Uue majapidamise tarbimise prognoosimise eksperiment .....	41
5 Kokkuvõte .....	44
Kasutatud kirjandus .....	46
Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks .....	51
Lisa 2 – Majapidamiste tunnuste vahelised korrelatsioonid.....	52
Lisa 3 – MK majapidamise energiatarbimise prognoosimise simulatsiooni väljund.....	55



## Jooniste loetelu

Joonis 1 Majapidamise energiavõrk. ....	11
Joonis 2 Üldistatud XGBoost algoritm.....	19
Joonis 3 Mudeli valik, treenimine ja prognoosimine. ....	22
Joonis 4 Prognoosi modifitseerimine. ....	25
Joonis 5 MK majapidamise reaalne tarbimine, prognoos ja modifitseeritud tarbimine. 26	
Joonis 6 MK majapidamise tarbimise prognoosimine triviaalse mudeliga. ....	28
Joonis 7 Andmemudel. ....	31
Joonis 8 MK majapidamise energiatarbimine aastate lõikes.....	33
Joonis 9 TH majapidamise energiatarbimine aastate lõikes.....	34
Joonis 10 VH majapidamise energiatarbimine aastate lõikes. ....	34
Joonis 11 Mudeli treenimine ja prognoosimine tervikuna. ....	35
Joonis 12 Andmete ettevalmistamine. ....	35
Joonis 13 Võimalike tunnuste ja tarbimise vahel korrelatsioonide leidmine. ....	37
Joonis 14 Hüperparameetrite leidmine. ....	37
Joonis 15 Oluliste tunnuste valimine.....	39
Joonis 16 Mudeli treenimine ja andmebaasi salvestamine. ....	40
Joonis 17 Mudeli lugemine andmebaasist ja prognoosimine. ....	41
Joonis 18 MK, TH ja VH majapidamiste prognooside MAE seos treeningu perioodi pikkusega.....	43

## **Tabelite loetelu**

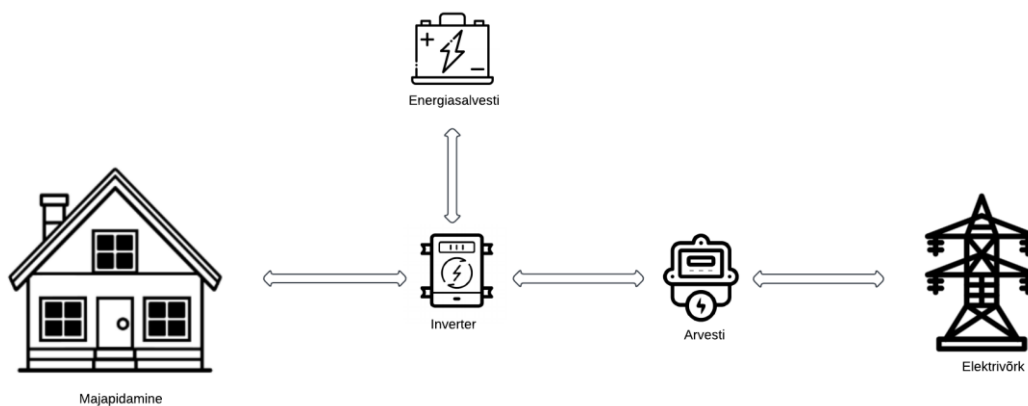
Tabel 1 Algoritmide eksperimentaalne võrdlus MK majapidamise andmetega.....	18
---	----

# 1 Sissejuhatus

Hetkel on käimas laiemalt tehisintellekti ja kitsamalt masinõppe buum ning masinõppe meetoditega püütakse lahendada erinevaid probleeme, sealhulgas energiatarbimise prognoosimine. Masinõppe meetodid võimaldavad suurte andmehulkade analüüsimist, mustrite tuvastamist ja prognooside tegemist, mille pealt on võimalik vastu võtta otsuseid. Rakendades masinõppe meetodeid olemasolevate energiatarbimise-, ilma- ja energia hinnaandmete peal, on võimalik optimeerida ressursikasutust ja seeläbi vähendada kulusid.

## 1.1 Probleemi püstitus

Käesoleva töö üldiseks eesmärgiks on eraldiseisva majapidamise energiavõrgus asuva energiasalvestusseadme juhtimise efektiivsuse suurendamine majandusliku kasu saamise eesmärgil (Joonis 1) [1], [2].



Joonis 1 Majapidamise energiavõrk.

Energiasalvestist saadava kasu võib laias laastus jagada kaheks:

- Majanduslik kasu, mis tekib võrguenergia tarbimise nihutamisel odavamale ajale.
- Majanduslik kasu, mis tekib salvestatud energia võrku tagasi müümisel.

Tervikvaates jaguneb käsitletav probleem kaheks:

- Kuidas võimalikult täpselt prognoosida tuleviku energiatarbimist?
- Kuidas prognoosi põhjal võimalikult efektiivselt juhtida energiasalvestusseadet?

Käesolev töö tegeleb probleemi esimese osaga, ehk kuidas võimalikult täpselt prognoosida tuleviku energiatarbimist. Konkreetselt otsib autor oma töös vastuseid neljale küsimusele:

- Kuidas prognoosida tuleviku energiatarbimist olukorras, kus majapidamisel puudub eelnev energiatarbimise ajalugu?
- Kuidas prognoosida energiatarbimist olukorras, kus majapidamisel on juba tekkinud energiatarbimise ajalugu, aga see ei ole piisavalt pikk aktsepteeriva täpsusega prognoosi tegemiseks?
- Milliste kriteeriumite alusel otsustada, et energiatarbimise aegrida on piisavalt pikk selleks, et aegrea pealt treenitud mudeli prognoosi täpsus on piisav?
- Kuidas prognoosida majapidamise energiatarbimise tippe?

Autor ei käsitle töös olukordasid, kus majapidamise energiatarbimise prognoosimisega tegeletakse mingil muul, kui majanduslikel põhjustel.

Töö käigus luuakse Python keeles süsteem, mille eesmärgiks on tõestada kontseptsiooni toimimist (*Proof Of Concept*). Python kood on avalikult kättesaadavad Bitbucket [3] koodihalduskeskkonnas aadressil: [https://bitbucket.org/kaidos12/predicting\\_single\\_household\\_energy\\_consumption/src/main/](https://bitbucket.org/kaidos12/predicting_single_household_energy_consumption/src/main/)

## 1.2 Töö skoop

Töö fookuse täpsustamiseks ja üldiste piirangute seadmiseks sõnastas autor töö üldise skoobi järgmiselt:

- Töö käigus luuakse kaks algoritmi, mis lahendavad püstitatud probleemid.
- Töö käigus valmib Python keeles kirjutatud PoC, mis simuleerib lahendusena pakutud algoritmi toimimist, eesmärgiga kinnitada või ümber lükata autori

pakutud lahendus. Simulatsioon käsitleb kolme esimest töös püstitatud probleemi: puuduvaid ja lühikesi energiatarbimise aegridu ning piisavuse kriteeriumit.

- Töö pakub lahenduse neljandale püstitatud probleemile, energiatiipude prognoosimisele, aga ei valideeri lahenduse toimimist praktikas.
- Töös implementeeritud PoC eeldab, et kõik vajalikud andmed on andmebaasis ja ei käsitle seda, kuidas vajalikud andmed andmebaasi jõuavad. Töös kasutatud andmemudel vastab PoC vajadustele ja ei pea olema praktikas täiendusteta rakendatav.
- Töös pakutud lahendus käsitleb uuritavaid majapidamisi musta kastina, mille puhul on teada ainult tunnipõhine energiatarbimine.
- Töö ei käsitle lahendust, kuidas prognoos jõuab energiasalvestit juhtivasse süsteemi. Samuti on energiasalvesti käesoleva töö kontekstis must kast, mille juhtimist töö ei käsitle.

### **1.3 Analoogsed uurimused Eestis ja Maailmas**

Eesti keeles ei ole välja kujunenud ühest masinõppega seotud sõnavara ning erinevad autorid kasutavad sama nähtuse tähistamiseks erinevaid väljendeid. Käesolevas töös kasutas autor spetsiifiliste väljendite eestindamisel sõnastikke [4], [5], Tartu Ülikooli masinõppe kursuste materjale [6], [7] ja kaitstud eestikeelseid bakalaureuse ning magistritöid [8] [9].

Töö käigus vaatas autor läbi Taltech'is ja Tartu Ülikooli Arvutiteaduste Instituudis kaitstud energiatarbimise prognoosimise teemadel kaitstud tööd [10], [11]. Mitte ükski autori leitud töödest ei käsitlenud energiatarbimise prognoosimist käesoleva töö kontekstis. Kaitstud töödest on bakalaureuse tasemel kolm tööd [12], [13], [14], magistritasemel kolm tööd [15], [16], [17] ja doktoritasemel üks töö [18].

Maailmas on tehisintellekti buumi tuules avaldatud palju energiatarbimise prognoosimist puudutavaid teadusartikleid ja raamatuid. Samas on vähe uuritud eraldiseisva majapidamise energiatarbimise prognoosimist. Tehes Taltech'i raamatukogus päringu „machine learning energy consumption prediction single house“ [19], tuleb vastuseks 32 artiklit, mis on avaldatud aastatel 2011 - 2024. Leitud artiklitest ainult kaks omavad otsust

seost käesoleva tööga, milleks on energiatarbimise prognoosimine eramajapidamises [20], [21].

Tehes sama päringu Google's, tuleb vastuseks hulk energia tulevikutarbimise uurimisega seotud artikleid ja blog tüüpi lehekülgi, mis on küll kasulikud, kuid ei seostu otseselt käesoleva töö kontekstiga. Enamik Google poolt leitud vasted on seotud kortermaja energiatarbimise prognoosimisega, mis ei ole käesoleva töö kontekst.

## 2 Masinõppe algoritmid

Autor otsustas töös püstitatud probleemide lahendamiseks kasutada masinõppe algoritmi. Masinõppealgoritmi valiku tegi autor esmalt kirjanduse põhjal ning valideeris valikut eksperimendiga.

### 2.1 Masinõppe algoritmi valimine

Püstitatud probleemide lahendamiseks sobiliku masinõppealgoritmi valimiseks otsis autor energia tulevikutarbimise prognoose puudutavaid uurimusi ning praktilisi töid. Otsimiseks kasutati Taltechi e-raamatukogu kaudu kättesaadavaid materjale ja Google otsingut. Läbiviidud uuringu tulemusena pani autor kokku nimekirja algoritmidest, mida võiks kasutada töös püstitatud ülesande lahendamiseks [22], [23], [24], [25], [26], [27].

#### 2.1.1 Algoritmi valimine kirjanduse põhjal

Läbitöötatud kirjanduse põhjal valis autor välja ARIMA (*Auto Regressive Integrated Moving Average*), juhuslik mets (*Random Forest*), tugivektori regressioon (*Support Vector Regression*), XGBoost (*eXtreme Gradient Boosting*) ja pika lühiajalise mälu (*Long Short-Term Memory*) masinõppe algoritmid. Võrdlemaks, kas püstitatud ülesannet on võimalik lahendada lihtsamalt, ilma masinõppe meetoditeta, lisas autor võrdluseks triviaalse (*naïve*) mudeli.

ARIMA algoritmi juured pärinevad 1970-ndatel välja pakutud statistilistest meetoditest ning ARIMA on laialdaselt kasutatud aegriidade prognoosimisel [28].

ARIMA algoritmi saab kirjeldada kolme parameetriga ARIMA(p, d, q) [12]:

- p, ehk auto-regressiivne (AR) osa, mis kasutab kombinatsioone varasematest väärtustest („lag“).
- d: integreerimine (I), kus mittestatsionaarsetest algandmete põhjal arvutatakse uued, statsionaarsed väärtused [29].
- q: libisev keskmine (MA), mida kasutatakse eelnevate vigade parandamiseks.

Algoritmist on olemas ka variatsioon, mis arvestab sesoonsusega: ARIMA (p, d, q) (P, D, Q), kus

- (p, d, q) on mudeli mitte-sesoonne osa
- (P, D, Q) on mudeli sesoonne osa

Algoritmi nimetakse SARIMA (*Seasonal AutoRegressive Integrated Moving Average*) [30].

RF algoritmi näol tegemist juhendatud õppega algoritmiga, mille käigus luuakse hulk otsustuspuud, mille agregeerimisel saadakse mudel, mille põhjal omakorda tehakse prognoos [31]. RF algoritm kasutab kolme üldist parameetrit [32]: puu sõlmede arv, puude arv ja kasutatud tunnused (*feature*). Treeningandmetest valitakse juhuslikult etteantud arv andmekomplekti, milles olevad andmed võivad olla kattuvad. Andmekomplektide pealt luuakse otsustuspuud, mis hiljem agregeeritakse kokku. Mainitud tehnikat nimetatakse pakkimiseks (*bagging*, ehk *bootstrap aggregating*) [31], [32].

SVR algoritm on variatsioon tugivektor masinate (*Support Vector Machine*) algoritmist ja kasutatakse regressioonianalüüsil. Algoritmi eesmärgiks on leida võimalikult optimaalne N-mõõtmeline hüperatasand, mis kirjeldaks võimalikult täpselt treeningandmeid. Nii SVR, kui ka SVM mõlemad, on juhendatud õppega algoritmid [33], [34].

XGBoost on gradiendi võimendamise (*gradient boosting*) algoritmide perekonda kuuluv juhendatud õppega algoritm [35]. Kuna XGBoost algoritmist räägib käesolev töö edaspidi pikemalt, siis siinkohal autor algoritmi olemust rohkem ei ava.

LSTM mudel kasutab tehisnärvivõrke, mis on disainitud leidma pika ajahorisondiga seoseid. Erinevalt traditsionaalsetest närvivõrkude mudelitest, kus sõlmedeks on neuronid, kasutab LSTM läbi kihtide ühendatud mälurakke. LSTM näol on tegemist juhendatud õppe algoritmiga [16].

Selgitamaks, kuivõrd keeruliste masinõppe algoritmide kasutamine annab paremaid tulemusi, pakkus autor välja triviaalse algoritmi. Pakutud algoritmi kohaselt on järgmise



tunni energiatarbimine keskmine eelmise tunni, eelmise päeva sama tunni ja eelmise nädala sama päeva ning sama tunni tarbimistest.

Uuringute põhjal võib välja tuua:

- Kõik toodud algoritmid sobivad energiatarbimise prognoosimiseks [22], [23], [25], [26], [27].
- XGBoost ja SVR on mõlemad kiired ja prognoosimise täpsus on hea [22].
- Kommunaalteenuste tarbimise prognoosimisel olid XGBoost-i tulemused sarnased ARIMA-ga, samas oli XGBoost kiirem [23].
- Aktsiate liikumise aegridade prognoosimisel suutsid XGBoost ja LSTM enamikel juhtudel prognoosida täpsemalt, kui ARIMA [24].
- XGBoosti ja RF mudelite prognooside tulemused olid sarnased. Samuti olid mõlema mudeli tulemused paremad, kui teistel võrdluses olnud mudelitega. Teisteks võrreldud algoritmideks olid lineaarne regressioon, Ridge regressioon, otsuste puude regressioon ja gradiendi võimendamine [25].
- XGBoost prognoosib energiatarbimist täpsemalt, kui LSTM ja SARIMA. Samuti saab XGBoost paremini hakkama mittelineaarsete muustritega. Sõltuvalt valitud konfiguratsioonist tulenevast arvutusmahust, oli XGBoost SARIMA-st kuni 48x ja LSTM-st kuni 24x kiirem [26].
- XGBoost-i ja RF prognooside täpsused olid sarnased [27].

### 2.1.2 Algoritmi eksperimentaalne valimine

Autori läbitöötatud allikate põhjal võiks kõikidel eeltoodud algoritmidel olla piisavalt potentsiaali lahendamaks püstitatud ülesannet. Algoritmi lõplikuks valikuks viis autor läbi praktilise eksperimendi, milleks kirjutas Pyhonis mudeli treenimise programmid kõikide eelpoolmainitud algoritmide jaoks. Kuna algoritmi valik ei ole käesoleva töö põhiteemaks, siis ei häälestanud autor mudeleid eelnevalt, vaid kasutas neid vaikeväärtustega.

Mudelite treenimisel kasutati MK majapidamise, mille kirjelduse leiab töö eksperimentaalse osa peatükist 4.1, tunnitäpsusega tarbimise andmeid ajavahemikus

01.01.2023 00:00:00 kuni 31.12.2023 23:00:00 ja Eesti Keskkonnaagentuuri ajaloolisi ilmaandmeid [36]. Kasutatud tunnusteks olid: elektri tarbimine, päikese intensiivsus, õhurõhk, sademete hulk, õhuniiskus, temperatuur, tuule suund, tuule kiirus, elektri tarbimise aasta, elektri tarbimise kuu, elektri tarbimise päev, elektri tarbimise tund, nädalapäev, päev aastas, nädal aastas ja kvartal. Triviaalse mudeliga prognoosimisel kasutati ainult tarbimise andmeid. Tulemuse hindamisel kasutati keskmist absoluutviga (*Mean Absolute Error*) ja keskmist ruutviga (*Mean Squared Error*), mis on kirjeldatud peatükis 2.3.

Tulemuste (Tabel 1) tõlgendamisel on oluline meeles pidada, et mudeleid kasutati vaikeväärtustega ja konfiguratsiooni muudatuste või andmete teistsuguse ettevalmistamisega on võimalik prognooside tulemusi oluliselt täpsemaks muuta.

Tabel 1 Algoritmide eksperimentaalne võrdlus MK majapidamise andmetega.

Algoritm	MAE (kW)	MSE (kW <sup>2</sup> )	Töötlemise aeg (s)
SARIMA	1,039	1,517	36,71
RF	0,629	0,677	4,61
SVR	0,816	1,131	2,05
XGBoost	0,651	0,744	0,17
LTSM	0,716	0,833	3,20
Triviaalne	1,415	3,504	0,09

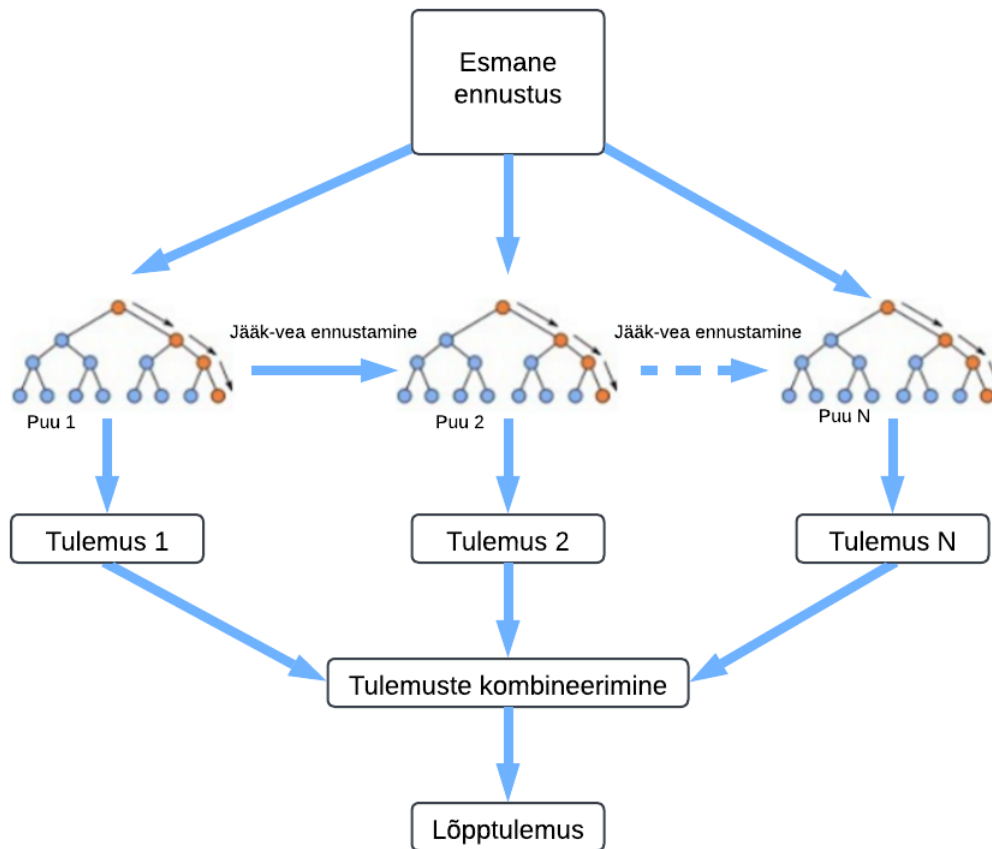
Toetudes läbitöötatud kirjandusele ja läbiviidud praktilistele eksperimentidele, otsustas autor prognooside tegemiseks valida XGBoost algoritmi.

## 2.2 XGBoost algoritmi kirjeldus

XGBoost algoritmi loojateks on Tianqi Chen ja Carlos Guestrin Washingtoni Ülikoolist ning seda presenteeriti esmakordselt aastal 2016 [37], [38].

XGBoost algoritm (Joonis 2) [39] on juhendatud õppega masinõppealgoritm ja on gradiendi võimendamise algoritmide edasiarendus. Kasutades ansambelmeetodit kombineerib XGBoost algoritm mitu nõrga otsustusvõimega puud üheks suure otsustusvõimega mudeliks. Algoritm loob iteratiivselt hulga otsustuspuid eesmärgiga vähendada iga iteratsiooniga eelmise puu poolt tehtud viga. Puid itereeritakse seni, kuni

uus loodud puu viga rohkem ei vähenda, mille järel iga iteratsiooni tulemus kombineeritakse lõpptulemuseks [35], [38], [40], [41].



Joonis 2 Üldistatud XGBoost algoritm.

Mudeli ülesobitamise vähendamiseks lisatakse täiendavat teavet (*regularization*), milleks peamisteks on L1 (Lasso, parameeter:  $\alpha$ ), L2 (Ridge, parameeter:  $\lambda$ ), varajane peatumine (parameeter: `early_stopping_rounds`), õppimise kiiruse vähendamine (parameeter: `learning_rate`, ehk  $\eta$ ), võimenduse iteratsioonide arv (parameeter: `n_estimators`) ja puude maksimaalne sügavus (parameeter: `max_depth`) [40], [42].

Lisaks pakkusid XGBoost algoritmi autorid välja mitmeid süsteemseid parandusi, kuidas tõsta mudeli kiirust ning realiseerisid need ka XGBoost teekides [38], [40], [43]:

- Paralleelsus. Tervikülesanne jagatakse tükkideks vastavalt protsessori (*Central Processing Unit*) tuumade arvule ja CPU koormatakse maksimaalselt.

- CPU vahemälu kasutamine. Gradientidega seotud statistikat hoitakse CPU vahemälu, mistõttu väheneb vajadus küsida andmeid juhupöördusega mälust (*Random Access Memory*) ning seega suureneb mudeli kiirust.
- Andmete pakkimine. Olukorras, kus süsteemil ei ole piisavalt RAM-i selleks, et laadida kõik vajalikud andmed mällu, pakitakse andekandjalt loetud andmed enne RAM-i salvestamist kokku. Kuna väliselt andmekandjalt andmete laadimine võtab rohkem aega, kui andmete lugemine RAM-st ja nende lahti pakkimine, siis selline liiasusega lahendus suurendab XGBoost mudeli üldist kiirust.

## 2.3 Meetrika

Praktiliste eksperimentide kirjeldamise meetrikaks valis autor keskmise absoluutvea (*Mean Absolute Error*) ja keskmise ruutvea (*Mean Squared Error*). Autor lähtus oma valikus asjaolust, et neid meetrikaid kasutatakse enamikes käesolevas töös viidatud teadusartiklites, kus on läbi viidud praktilisi katseid. Toodud meetrikate kasutamine võimaldab käesoleva töö tulemusel paigutada laiemasse konteksti.

Keskmine absoluutviga leitakse valemiga [44]:

$$\text{MAE} = \frac{1}{N} \sum_{k=1}^N |y_k - \hat{y}_k|$$

Kus  $N$  on andmepunktide arv,  $y_k$  tarbimise tegelik väärtus ja  $\hat{y}_k$  tarbimise prognoositud väärtus.

Keskmine ruutviga leitakse valemiga [44]:

$$\text{MSE} = \frac{1}{N} \sum_{k=1}^N (y_k - \hat{y}_k)^2$$

Kus  $N$  on andmepunktide arv,  $y_k$  tarbimise tegelik väärtus ja  $\hat{y}_k$  tarbimise prognoositud väärtus.

## 3 Analüüs

Kolme esimese, tarbimisandmete puudumisega seotud, probleemi lahendamiseks pakub autor välja neli lahendust, millest kolme analüüsib pealiskaudselt ja ühte põhjalikult. Neljanda, tarbimise tippude prognoosimisega seotud, probleemi lahendamiseks pakub autor välja kaks lahendust, milles ühte analüüsib põhjalikult.

### 3.1 Energiatarbimise prognoosimine puuduvate andmete olukorras

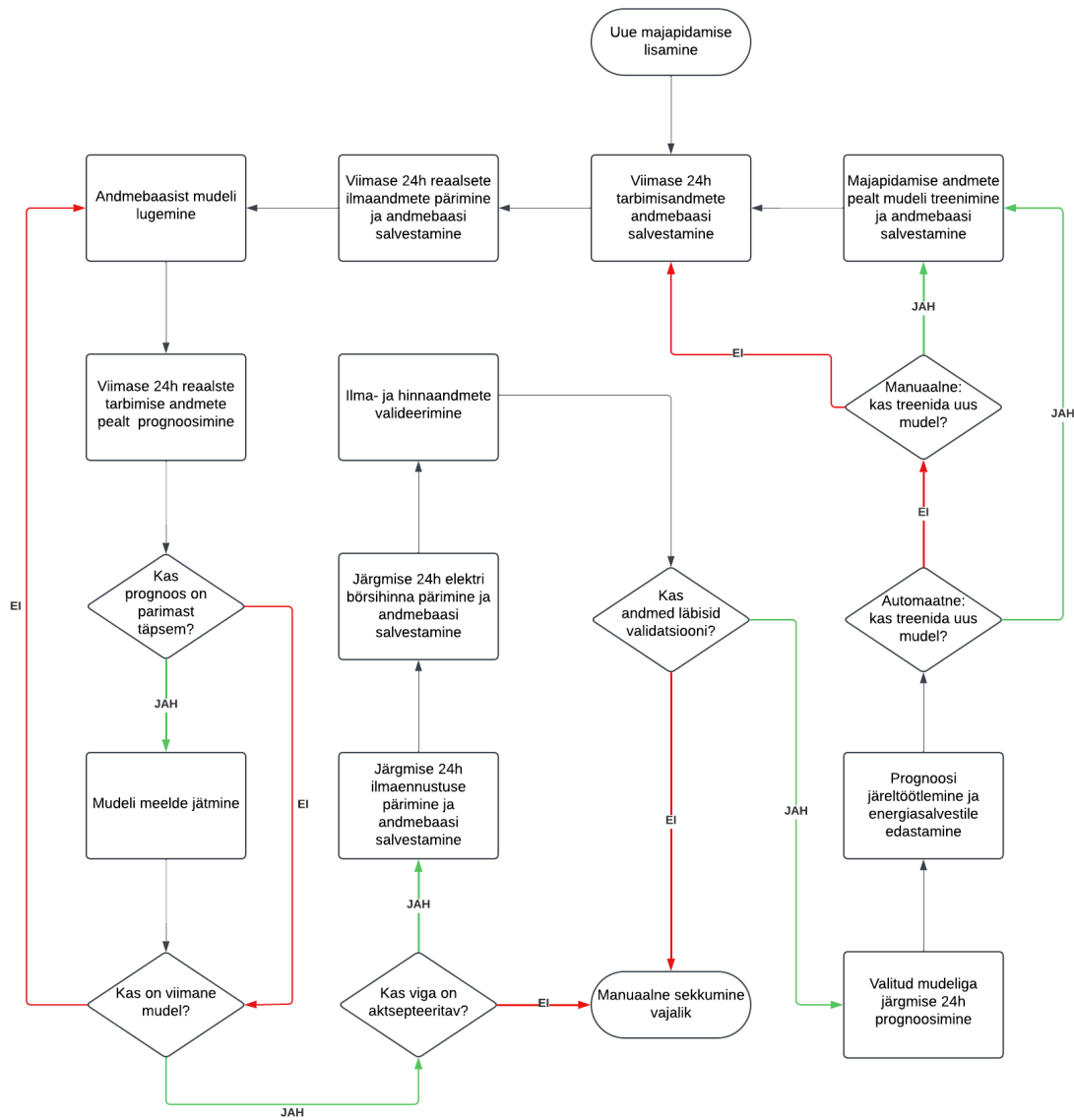
Autor pakub välja neli võimalikku lahendust, kuidas prognoosida majapidamise energiatarbimist olukorras, kus energiatarbimise andmed puuduvad või ei ole piisavalt pikad:

- Sarnane mudel
- Üks mudel kõigi jaoks
- Mudelite grupeerimine
- Triviaalne mudel

Töös kasutab autor sarnase mudeli lahendust, analüüsib ja implementeerib lahenduse Python koodis. Teisi lahendusi autor küll kirjeldab ja arutleb lahenduste üle, aga ei realiseeri praktikas. Kuigi sarnaste mudelite lahendus annab eelise ennekõike puuduvate ja väheste andmete olukorras, saab lahendust edukalt kasutada piiramata aja jooksul.

Lahenduse (Joonis 3) peamiseks ideeks on leida treenitud mudelite hulgast uue majapidamise jaoks parima prognoosimisvõimega mudel ja kasutada mudelit parameetriga määratud aeg. Pärast parameetriga määratud aja möödumist, protsess kordub ning leitakse uus, hetkel parim mudel. Parima mudeli leidmiseks prognoositakse kõikide olemasolevate mudelitega viimase 24 tunni tarbimist, võrreldakse prognoosi reaalse tarbimisega, leitakse viga (MAE) ja valitakse mudel, mille viga oli väikseim. Parima mudeliga prognoositakse järgmise 24 tunni tarbimist, viiakse läbi prognoosi modifitseerimine ja edastatakse energiasalvestile. Parameetriga määratud aja möödumise järel treenitakse uue majapidamise andmete pealt mudel, lisatakse andmebaasi ning treenitud mudel hakkab konkureerima kõikide olemasolevate mudelitega.

Toodud algoritm lahendab ülesande püstituses toodud neljast probleemist kolm esimest. Viimase püstitatud probleemiga, ehk kuidas prognoosida energiatarbimise tippe, tegeletakse prognoosimise järgselt prognoosi modifitseerides ja seega ei otseselt tegemist kirjeldatud algoritmi osaga.



Joonis 3 Mudeli valik, treenimine ja prognoosimine.

Uue majapidamise mudeli treenimisel ja tarbimise prognoosimisel kasutatud parameetrid:

- Uuele majapidamisele mudeli treenimise regulaarsus päevades. Ehk, mitme päeva möödudes treenitakse olemasolevate andmete pealt majapidamisele uus mudel ja

salvestatakse andmebaasi. Läbiviidud eksperimentide põhjal pakub autor parameetri vaikeväärtuseks 7 päeva.

- Ühe mudeliga ennustamise kestvus päevades. Parima mudeli valimine võib suure arvu mudelite korral osutada arvutusmahukaks tegevuseks. Samuti ei pruugi igapäevaselt uue mudeli leidmine suurendada prognoosi täpsust piisavalt selleks, et tegevust õigustada. Läbiviidud eksperimentide põhjal pakub autor parameetri vaikeväärtuseks 7 päeva.
- Vea suurus protsentides, mida loetakse normaalseks. Vea leidmiseks liidetakse prognoositava päeva tarbimised ja korrutatakse parameetriga määratud protsendiga. Läbiviidud eksperimentide põhjal pakub autor parameetri vaikeväärtuseks 10%.
- Minimaalne andmete hulk andmeid päevades, mille pealt tohib uuele majapidamisele mudeli treenida. Läbiviidud eksperimentide põhjal pakub autor parameetri vaikeväärtuseks 7 päeva.
- Maksimaalne ajavahemik päevades, mille pealt tohib uuele majapidamisele mudelit treenida. Parameeter on mõeldud kasutamiseks olukorras, kus majapidamisele on tekkinud pikk tarbimise ajalugu. Parameetri eesmärk on vältida mudeli treenimisel liiga vanade andmete kasutamist, mis võib viia mudeli prognoosimisvõime vähenemiseni. Läbiviidud eksperimentide põhjal pakub autor parameetri vaikeväärtuseks 365 päeva.
- Ainult majapidamise X mudeli kasutamine. Parameetri eesmärk on sundida süsteemi kasutama ainult ühe konkreetse majapidamise mudelit. Parameetri vaikeväärtuseks peaks autor arvates olema *false* ja seda peaks kasutama pigem erandolukorras.
- Majapidamise nimi, millele tarbimist prognoositakse.

Autori arvates on toodud parameetrid minimaalne komplekt, millega on võimalik süsteemi efektiivselt juhtida. Süsteemi praktilise kasutamise käigus saadud uute teadmiste põhjal on mõistlik parameetrite kasutus üle vaadata.

Autori pakutud parameetrite vaikeväärtustesse tuleks suhtuda ettevaatlikult, sest need sobivad PoC käigus uuritud kolme majapidamisega. Reaalses kasutuses, eriti kui mudeleid on vähe ning arvutusressurss ei ole piiravaks teguriks, tasub kaaluda majapidamisele mudeli treenimist ja parima mudeli valimist kord ööpäevas. Parameetrite väärtused on mõistlik regulaarselt üle vaadata ning otsustada nende väärtused vastavalt hetkeolukorrale.

Lisaks automaatikat reguleerivatele parameetritele on algoritmi järgi alati võimalik alkatada majapidamise mudeli treenimine ja andmebaasi salvestamine manuaalselt. Manuaalset mudeli treenimise ja andmebaasi lisamisega PoC ei tegele.

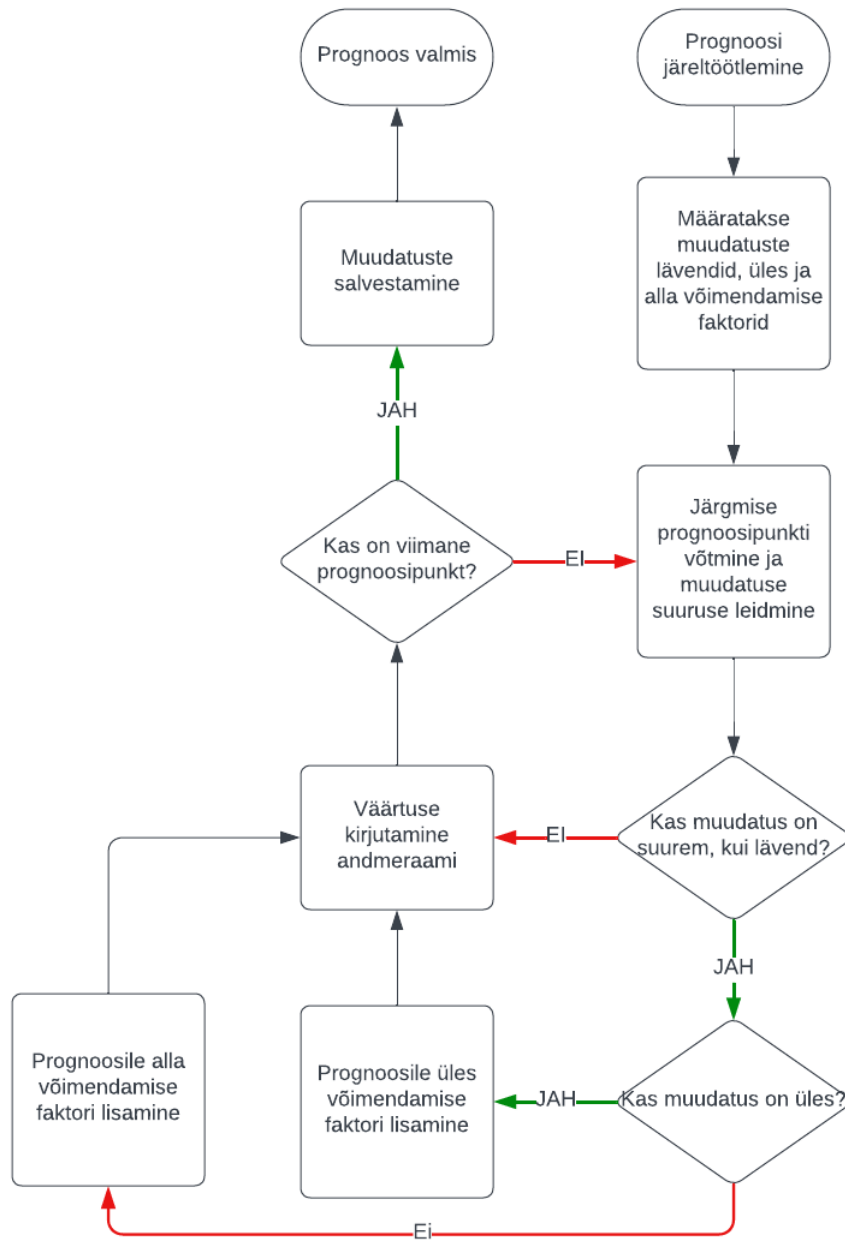
Käesolev töö ei käsitle olukorda, kus olemasolevate mudelite prognoosimisevõime on ajas vähenenud ning mõistlik oleks olemasolevad mudelid uuesti treenida.

### **3.2 Energiatarbimise tippude prognoosimine**

Andmete uurimise ja eksperimentide käigus jõudis autor järeldusele, et loodud XGBoost mudel töötab hästi olukorras, kus ei ole suuri, pealtnäha juhuslikke, muutusi energiatarbimises. Autori järeldust jagavad ka teised autorid, kes on uurinud eramajapidamiste elektritarbimist [45], [46]. Probleemi lahendamiseks pakub autor välja kaks lahendust:

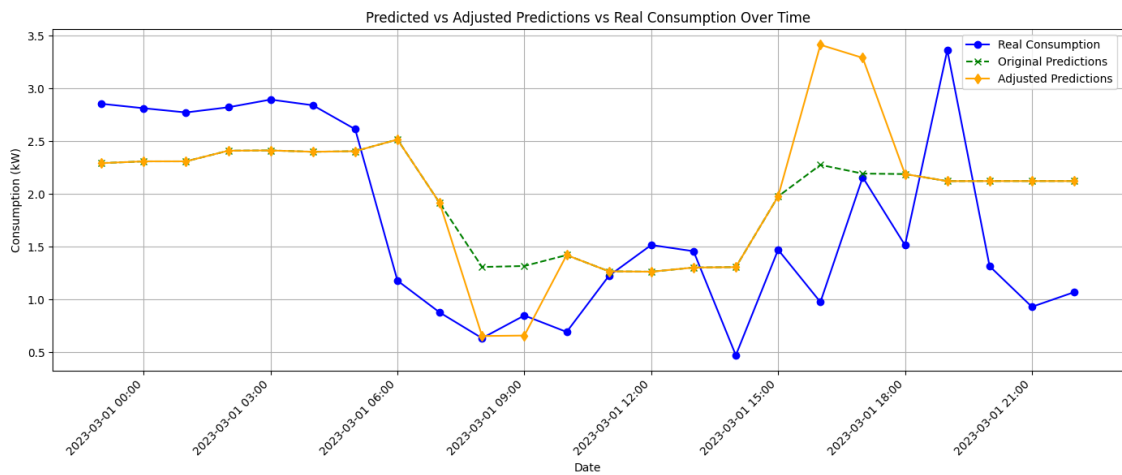
- Tegeleda probleemiga mudeli treenimisel. Lahenduseks treenitakse kaks mudelit, millest üks prognoosib baastarbimist ja teine prognoosib tippe. Lõpptulemuses pannakse kahe mudeli prognoosid kokku [46].
- Tegeleda probleemiga prognoosimise järel (Joonis 4). Lahendusena treenitakse baastarbimist prognoosiv mudel, mille tehtud prognoosi modifitseeritakse prognoosimise järel [47]. Antud lahendust, küll triviaalsel kujul, on kasutatud ka käesolevas töös.





Joonis 4 Prognnoosi modifitseerimine.

Lahenduse toimimise illustreerimiseks treenis autor MK majapidamise jaoks mudeli ajavahemiku 01.01.2022 – 28.02.2023 andmete pealt, salvestas mudeli andmebaasi, luges mudeli andmebaasist ja prognoosis 01.03.2024 24 tunni elektritarbimist ning modifitseeris prognoosi (Joonis 5).



Joonis 5 MK majapidamise reaalne tarbimine, prognoos ja modifitseeritud tarbimine.

Lahenduse eeliseks on selle lihtsus, ilma keerulist masinõppemudelit rakendatama on võimalik saada praktika seisukohalt piisavalt hea tulemus. Prognoosi seisukohalt on suured muudatused olulised, sest nende pealt tehakse otsus energiasalvesti laetuse taseme kohta. Praktilisest seisukohast ei ole kõige olulisem saada täpselt kokkulangev prognoos, vaid saada üldjoontes aru, millal on elektritarbimine tipus ja millal madalal.

Lahenduse puuduseks on prognoosi täpsuse vähenemine, sest mitte kõik trendid, üles või alla, ei viita tegelikult toimuvatele elektritarbimise liikumistele.

### 3.3 Alternatiivsed lahendused

Autori arvates on töös püstitatud esimest kolme probleemi võimalik lahendada veel kolmel erineval viisil.

#### 3.3.1 Üks mudel kõigi jaoks

Lahendusena luuakse üks üldine mudel, millega on võimalik prognoosida kõikide majapidamiste tarbimist. Lahenduse eeliseks on lihtne rakendamine ja kasutamine. Lahenduse puudusteks on mudeli loomise keerukus ja oletatavalt madal prognoosimise võime. Lisaks peab mudelit regulaarselt ajakohastama selleks, et mudeli prognoosimise võime ei langeks.

Autor uuris Maailma praktikat ühe universaalse mudeli loomise osas ning sellist lahendust on kasutatud marginaalsel määral [48]. Samuti suurendab autori skeptitsismi ühe mudeli lahenduse toimimise osas töös kasutatud majapidamistega läbiviidud

eksperimendid. Majapidamiste erinevus energiatarbimise põhjuste osas seab ühe mudeli lahenduse väljatöötamisele väga suure väljakutse.

### **3.3.2 Mudelite grupeerimine**

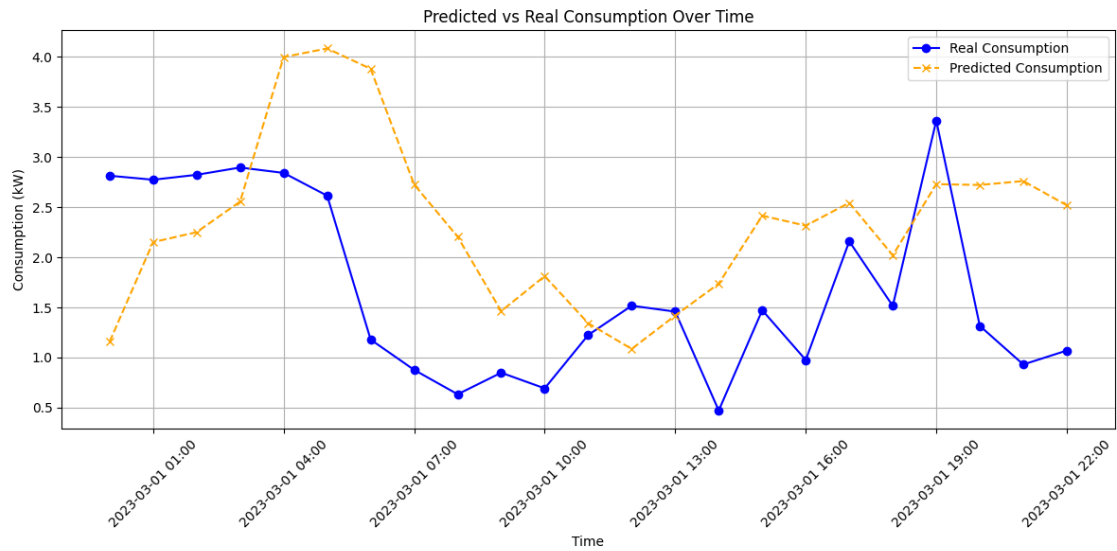
Lahendusena luuakse mitu mudelit, mis kirjeldavad sarnaseid majapidamisi. Süsteemi rakendamisel määratakse, millisesse gruppi uus majapidamine kuulub, kasutatakse grupi jaoks treenitud mudelit ning prognoositakse tuleviku energiatarbimist. Lahenduse peamine keerukus seisneb toimivate gruppide väljatöötamises, iga grupi jaoks mudeli treenimine ja uue majapidamise gruppi määramise keerukus on väiksem. Prognoosi täpsuse tõstmiseks võib parameetrite järgi samasse gruppi määratud mudelid kombineerida üheks mudeliks. Lahendusele lisandub hilisem haldamise keerukus, mida tingib vajadus hoida gruppide parameetreid ajakohasena ja majapidamise gruppi kuuluvuse regulaarne ülevaatamine.

Grupi tunnusparameetriteks võivad olla näiteks majapidamise soojapidavus, küttesüsteemi tüüp, sauna elektrikerise olemasolu, avatus tuulele, lisaks muud majapidamises enim elektrit tarbivad seadmed. Kuna tegemist on alternatiivse lahendusega, siis gruppide parameetreid autor siinkohal rohkem lahti ei kirjuta. Autor uuris Maailma praktikat mudelite grupeerimise osas ning seda ideed on rakendatud mitmetes uuringutes [49].

### **3.3.3 Triviaalne mudel**

Triviaalse mudeli korral leitakse tuleviku energiatarbimine lähimeneviku energiatarbimisest arvutuslikult. Autori pakutud triviaalse lähenemise korral on tulevikutarbimise prognoosiks prognoositavale tunnile eelneva tunni, prognoositava tunni väärtus 24 tundi tagasi ja prognoositava tunni väärtus üks päev tagasi keskmine väärtus. Kuna tegemist on alternatiivse lahendusega, mille välja toomise ainsaks eesmärgiks on näidata võimalikke alternatiive autori valitud lahendusele, siis ei arutle autor kuidas triviaalse mudeli prognoosi täpsust tõsta.

Illustreerimaks triviaalse mudeli prognoosimise võimet, prognoosis autor MK majapidamise, mille täpse kirjelduse leiab töö eksperimentaalse osa peatükist 4.1, energiatarbimist kasutades triviaalset mudelit. Triviaalse mudeli prognoos (Joonis 6) ei ole täpne, aga mudel suudab tabada üldist trendi ja võiks olla olukorras, kus muid prognoose ei ole, täiesti kasutatav.



Joonis 6 MK majapidamise tarbimise prognoosimine triviaalse mudeliga.

Triviaalsel mudelil on mitmeid eeliseid, mille pärast tasub triviaalset lähenemist kaaluda:

- rakendamine on jõukohane igäühele ning prognoosi võib arvutada mõnda tabelarvutusprogrammi kasutades.
- kasutamine nõuab äärmiselt vähe arvutusvõimsust.

## 4 Eksperiment

Autor pidas läbiviidud eksperimentide, loetud artiklite ning raamatute kohta päevikut, kuhu pani lühidalt kirja, mida katsetas ja millised olid tulemused. Samuti pani autor kirja loetu põhjal tekkinud ideed või arusaamised, mis lõppkokkuvõttes aitasid jõuda parema lahenduseni.

### 4.1 Majapidamiste kirjeldused

Mudeli treenimiseks, testimiseks ja prognooside tegemiseks on töös kasutatud kolme majapidamist, mida anonüümsuse tagamiseks tähistakse koodnimedega:

- MK: majapidamine Tallinna lähedal. Suuremad energiatarbijad: maakütteseade maja kütmiseks ning sooja vee tootmiseks ja sauna keris.
- TH: majapidamine Tallinna linnas. Suuremad energiatarbijad: elektriboiler vee soojendamiseks.
- VH: majapidamine Tallinna lähedal. Suuremad elektritarbijad: õhk-soojus pump ja elektriboiler vee soojendamiseks. Päikesepaneelid elektrienergia tootmiseks.

Käesolevas töös kasutatud andmed on ühe tunni täpsusega aegread, kus tarbimise-, hinna- ja ilmaandmete kellaaeg on universaalses (*Coordinated Universal Time*) ajatsoonis. Energia tarbimise ja hindade allikateks on Elering [50], ilmaandmete allikaks on Ilmateenistus [36].

### 4.2 Kasutatud riist- ja tarkvara

Töös väljatöötatud lahendus on implementeeritud programmeerimiskeeles Python versiooniga 3.13. Programmeerimise keele valik on tehtud praktilistel kaalutlustel, sest Python on enimlevinud keel masinõppealgoritmide rakendamisel [51]. Alternatiivina kaalus autor ka R ja Java keeli, aga nende puhul sai takistuseks olemasolevate teekide vähesus võrreldes Python'ga [51]. Lahenduse loomisel on kasutatud Python teeke: *datetime*, *matplotlib*, *numpy*, *os*, *pandas*, *pickle*, *pmdarima*, *scipy*, *seaborn*, *sklearn*, *skopt*, *sqlite3*, *statistics*, *statsmodels*, *tensorflow*, *time*, *warnings*, *xgboost*.

Masinõppe rakenduste loomisel loetakse parimateks keskkondadeks (*Integrated Development Environment*) Jupyter, Spyder ja PyCharm [52]. Kuna autor kasutab oma igapäevatoos JetBrains firma keskkondi, siis otsustas autor subjektiivsetel kaalutlustel kasutada Python koodi loomiseks IDE-na PyCharm 2024.2. Lisaks kasutas autor:

- Andmebaas: SQLite 3.40.0 [53].
- Andmebaasi klientrakendus: DBeaver Community Edition Version 24.2.2.202410061621 [54].
- Andmete esmane võrdlemine ja ülevaade: Microsoft Excel 365 versioon 2411.
- Andmemudeli joonis ja andmebaasi loomise struktureeritud päringute keele (*Structured Query Language*) laused: Vertabelo [55].
- Voodiagrammid: LucidChart [1].
- Programmeerimise tugi: ChatGPT [56].

Mudelite loomisel ja prognoosimisel kasutatud riistvara:

- CPU AMD Ryzen 7 PRO 5850U, 8 tuuma, 16 lõime; 16 GB DDR4 RAM; SSD 3,4 GB/s lugemine, 1,2 GB/s kirjutamine; CPU võrdlusskoor: 16953 [57].

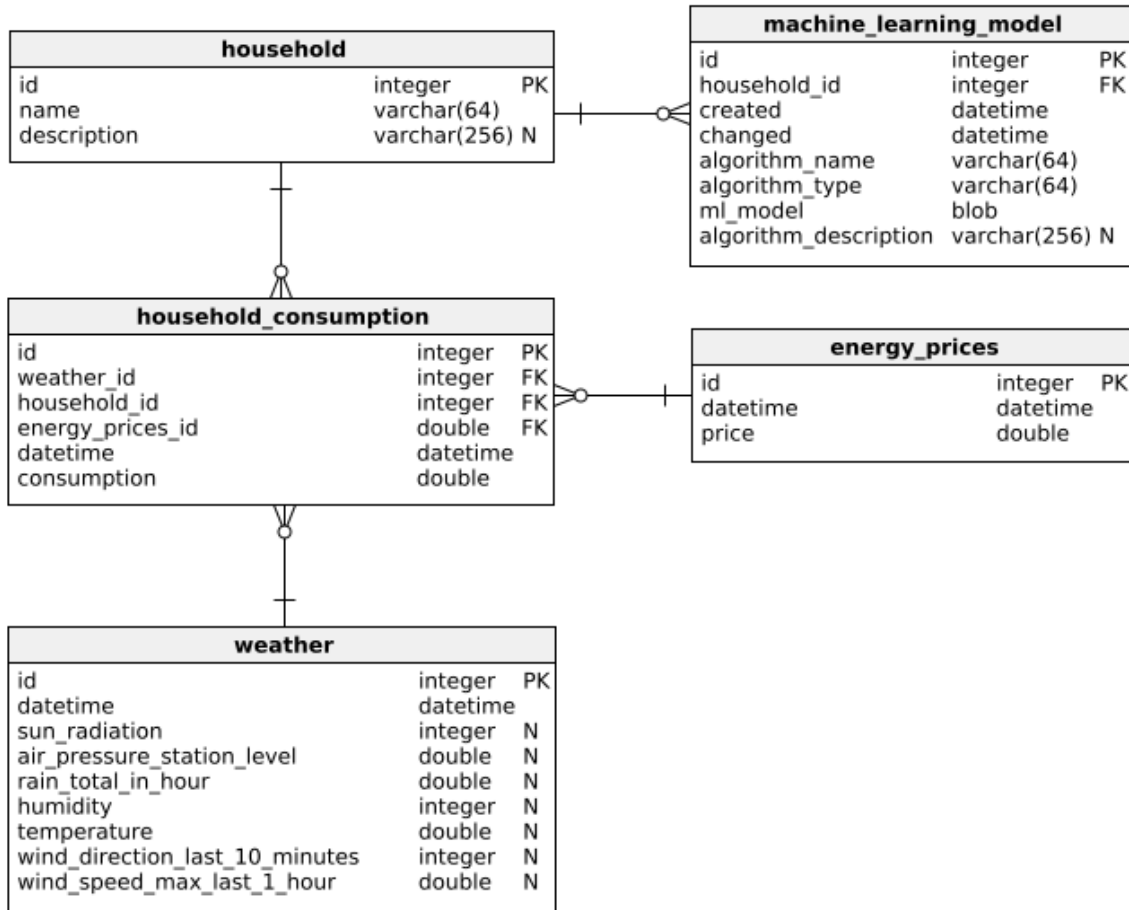
### 4.3 Tarkvara ettevalmistamine

Andmebaasi mootoriks valis autor SQLite [53] andmebaasi. Lahenduse jaoks vajalike andmete hoidmiseks lõi autor Vertabelo [55] keskkonnas andmemudeli.

Andmemudeli (Joonis 7) juures on oluline tähele panna, et andmemudel vastab PoC vajadustele, aga praktiliseks kasutamiseks on vaja andmemudelit täiendada. Näiteks ei võimalda kasutatud andmemudel:

- Laadida andmeid tabelitesse nende tekkimise hetkel, vaid kindlas järjekorras.
- Lisada teiste piirkondade ilma-andmeid.
- Lisada erinevate ilmajaamade andmeid.
- Hoida ette-arvutatud tarbimise või ilmaandmetega seotud tunnuseid.

Toodud piirangute nimekiri ei ole lõplik ja praktilise rakendamise osas on andmemudelil parendamise ruumi. Loodud andmemudeli pealt genereeriti Vertabelo [55] keskkonnas andmebaasi loomise SQL laused, mida kasutati Python koodis andmebaasi tabelite loomisel.



Joonis 7 Andmemudel.

Python koodi käsitlemiseks kasutas autor Bitbucket [3] GIT keskkonda. Autor lõi Bitbucket keskkonda kaks eraldi koodi repositooriumit:

- Privaatne repositoorium eksperimentide jaoks.
- Avalik repositoorium lõpptulemuse jaoks.

Juhul, kui eksperiment õigustas ennast, kopeeriti korrastatud kood lõpptulemuse repositooriumisse. Samuti on kõik töös andmeid ja prognoosimist kujutavad joonised genereeritud privaatses repositooriumis oleva koodi abil. Muudatuste paremaks jälgimiseks laadis autor iga eksperimendi valmimisel koodi koos põhjendusega üles Bitbucket [3] keskkonda.

Enne töö alustamist uuendas autor Python interpretaatorid ja PyCharm keskkonnad. Vältimaks võimalikke segadusi Python koodi organiseerimisega, lõi autor vajalikud kataloogid, mis jagasid koodi loogilisteks tükkiideks. Tagamaks, et GIT keskkonda laetakse ainult oluline osa koodist, täiendas autor mõlemas repositooriumis .gitignore faili. Koodi toimimiseks vajalikud teegid laadis ja uuendas autor koodi kirjutamise käigus.

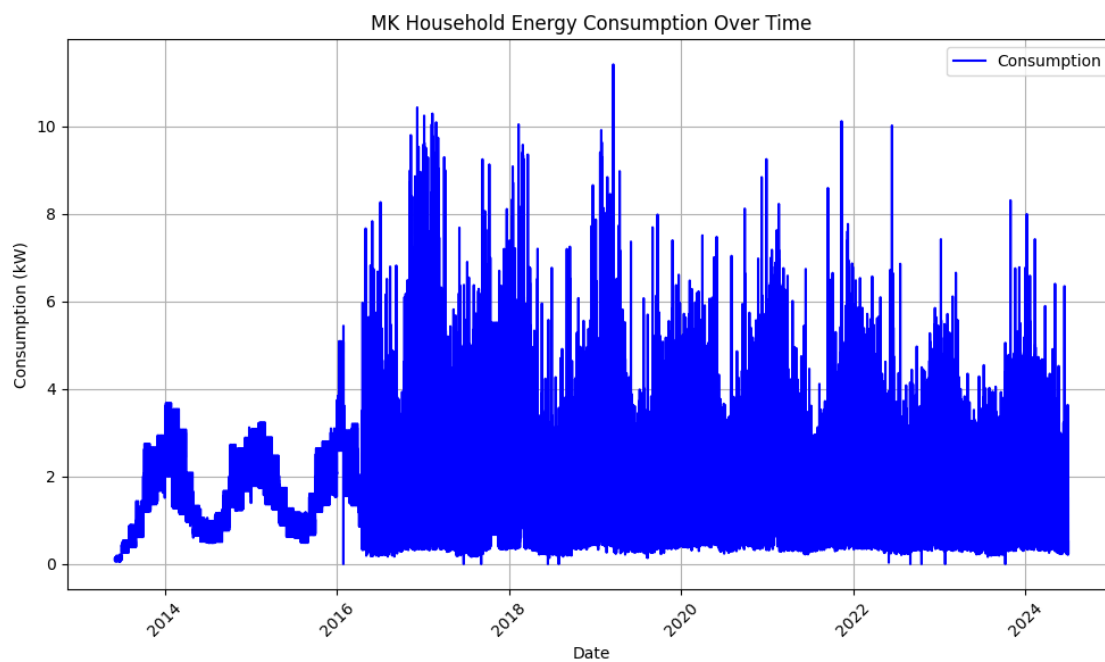
#### **4.4 Tarbimisandmete analüüs**

Masinõppe meetodite rakendamine algab alusandmetest aru saamisest, sest madala kvaliteediga andmete pealt tehtud prognoosid on samuti madala kvaliteediga. Masinõppes kasutatakse tihti väljendit „prügi sisse, prügi välja“ (*garbage in, garbage out*) [6]. Majapidamiste tarbimiste seostest aru saamiseks uuris autor enne mudeli loomist MK, TH ja VH majapidamiste andmeid. Uurimiseks kirjutas autor Python keeles funktsioonid, mis:

- Joonistasid etteantud andmete põhjal graafikuid eesmärgiga anda andmetest visuaalne ülevaade.
- Näitasid erinevaid tarbimisega seotud statistilisi seoseid, nagu tarbimise jaotumine [58], standardhälve, korrelatsioon, keskmine, mediaan, jne.

MK majapidamise energiatarbimine (Joonis 8) on kuni 2016 aastani väga ühtlane ja seda peab lugema anomaaliks, millest tulenevalt ei tohi mainitud ajavahemikku kasutada mudeli treenimisel. Ülejäänud energiatarbimine tundub Joonis 8 järgi normaalne ning võiks olla kasutatav masinõppe mudeli treenimiseks.



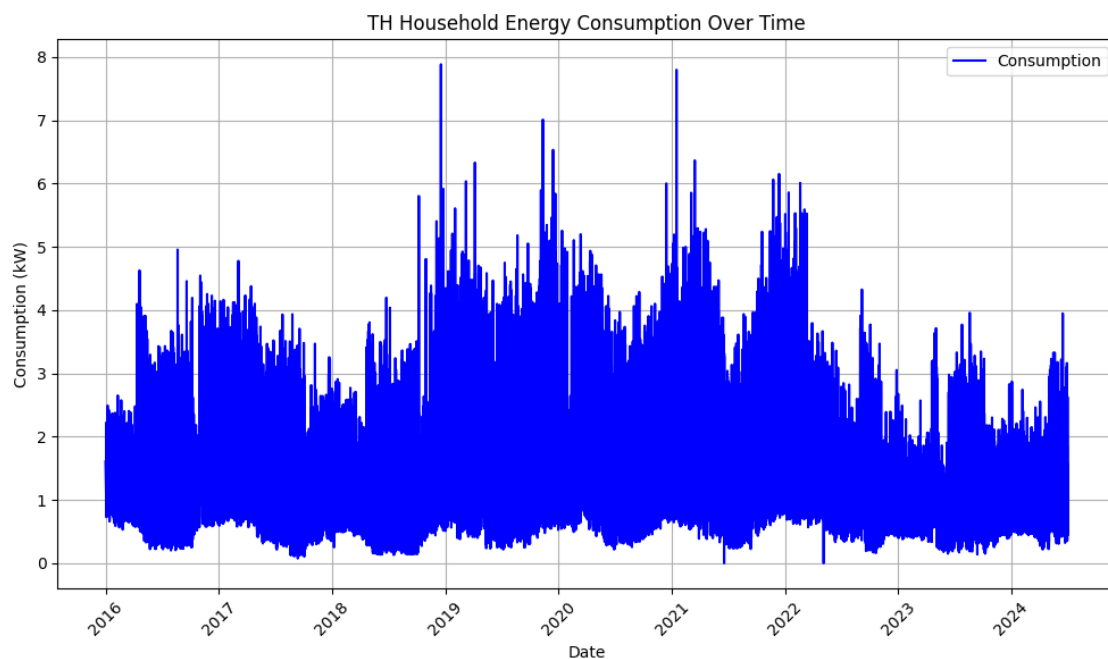


Joonis 8 MK majapidamise energiatarbimine aastate lõikes.

TH majapidamise energiatarbimise (Joonis 9) võib visuaalse vaatluse põhjal jagada kolme ossa:

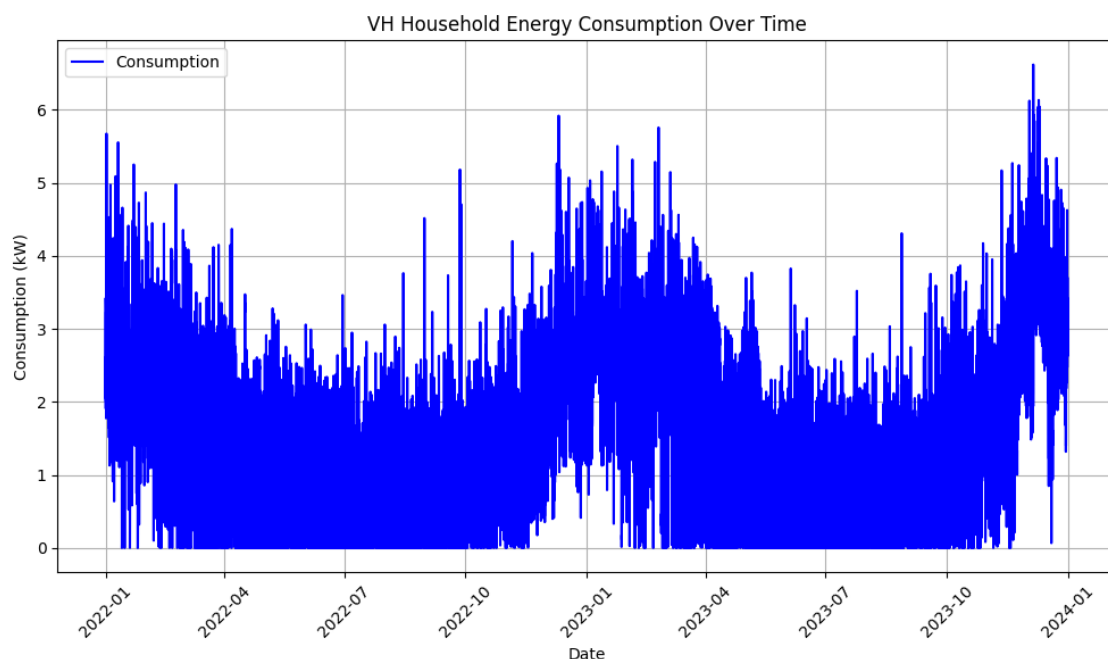
- 2016 – 2019.
- 2019 – 2022.
- 2022 kuni praeguse hetkeni.

Autoril ei ole teada, millised muutused toimusid TH majapidamises kahel esimesel perioodil, mistõttu peab TH majapidamise andmeid enne 2022 aastat käsitlema ettevaatlikult.



Joonis 9 TH majapidamise energiatarbimine aastate lõikes.

VH majapidamises (Joonis 10) toodetakse osa kasutatavat energiat päikesepaneelidega, mis selgitab väikest ja null tarbimist suvekuudel.



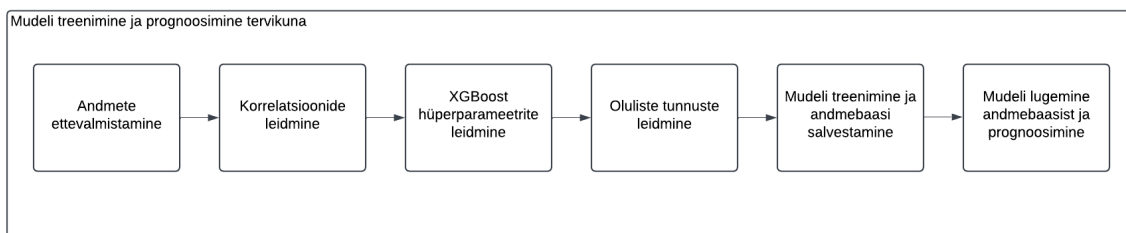
Joonis 10 VH majapidamise energiatarbimine aastate lõikes.

Jooniste põhjal võib väita, et kõikide majapidamiste tarbimises on olemas muster ning seega võiks nende majapidamiste tuleviku energiatarbimine olla masinõppe mudeliga prognoositav.

Energiatarbimise prognoosimisel on oluline välja tuua 25.03.2020 alanud COVID-19 pandeemiast põhjustatud eriolukord [59], mis ei paista MK (Joonis 8) ja TH (Joonis 9) majapidamiste energiatarbimiste pealt otseselt välja, aga muutis inimeste käitumist ja sellest tulenevalt tarbimise mustreid [15]. Näiteks on ajavahemikus 01.01.2020 – 31.12.2020 TH majapidamise energiatarbimise prognoosi viga MAE 0,582 kW ja MSE 0.571 kW<sup>2</sup>. Samas on ajavahemikus 01.01.2023 – 31.12.2023 TH majapidamise energiatarbimise prognoosi viga MAE 0,249 kW ja MSE 0,148 kW<sup>2</sup>. Suurem prognoosiviga on seletatav treenimise ja testimise ajavahemikku jääva COVID-19 eriolukorra algusega tekkinud tarbimismustri kiire muutumisega.

## 4.5 Mudeli treenimine ja prognoosimine

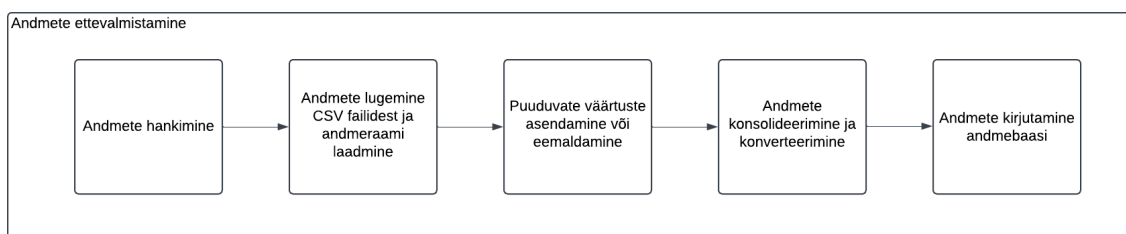
Masinõppemudeliga prognoosimiseks tuleb läbi teha rida samme (Joonis 11) ning käesolev peatükk kirjeldab vajalikke samme.



Joonis 11 Mudeli treenimine ja prognoosimine tervikuna.

### 4.5.1 Andmete ettevalmistamine

Andmete ettevalmistamisel (Joonis 12) on kõige keerulisem osa andmete hankimine, täpsemalt piisava koguse korrektsete andmete hankimine ning autoril ei ole siinkohal häid soovitusi.



Joonis 12 Andmete ettevalmistamine.

Kõik töös kasutatud sisendandmed olid CSV formaadis, mis loeti Pandas andmeraamidesse, mille järel need töödeldi:

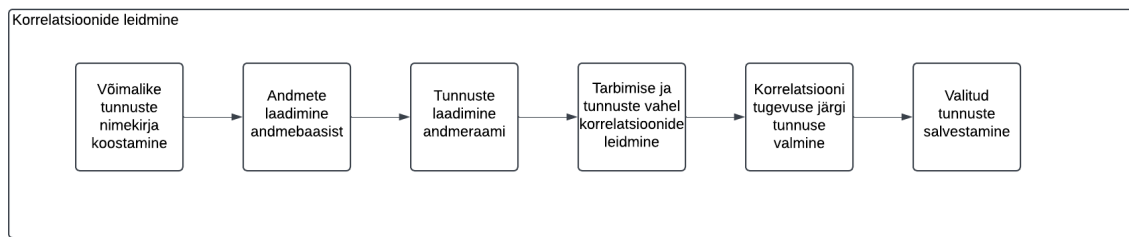
- Puuduvate väärtuste asendamine keskmiste väärtustega. Kuigi XGBoost algoritm suudab toime tulla puuduvate väärtustega [60], peaks puuduvad väärtused asendama keskmiste väärtustega. Eleringi [50] tarbimise- ja hinnaandmetes ei olnud puuduvaid väärtusi, aga Ilmateenistuse ilmaandmetes [36] oli arvestatav kogus puuduvaid väärtusi, mis asendati ümbritsevate andmete keskmiste väärtustega.
- Andmeveergude konsolideerimine. Ilmateenistuse ilmaandmetes [36] olid aasta, kuu, päev ja tund jagatud erinevatesse veergusse, mis konsolideeriti üheks kuupäev-kellaeg tüüpi veeruks.
- Ajaveergude konverteerimine UTC ajatsooni.
- Kümnenndkohtade eraldajana asendatakse koma punktiga. Eleringi [50] tarbimise- ja hinnaandmed olid kümnenndkoha eraldajana kasutusel koma, mis konverteeriti enne andmebaasi salvestamist punktiks.

Pärast andmete töötlemist kirjutati puhastatud andmed andmebaasi.

#### **4.5.2 Korrelatsioonide leidmine**

Prognoosi täpsuse suurendamiseks on vaja masinõppe mudelile lisaks tarbimise väärtustele ette anda tarbimisega seotud võimalikult lähedalt seotud tunnused. Majapidamised on erinevad ja tunnused, mis toimivad ühe majapidamise jaoks, ei pruugi toimida teise majapidamisega. Autori eksperimendid näitasid, kui anda XGBoost mudelile ette liiga palju tunnuseid, suureneb mudeli prognoosi viga ja mudeli töötamise aeg pikeneb iga lisatud tunnusega. XGBoost mudel väljastab pärast treenimise ja testimise läbiviimist kasutatud tunnuste olulisuse ja kõikides eksperimentides töös kasutatud majapidamistega jäi oluliste tunnuste arv vahemikku 3 kuni 5. Siinkohal on sobilik meenutada autori valitud täiskomplekti kuuluvate tunnuste arvu, milleks on 31. Selgitamaks välja, milliseid tunnuseid võiks prognoosimisel kaaluda, peab leidma seoste tugevused tarbimise ja konkreetse tunnuse vahel. Seoste tugevuste, ehk korrelatsioonide, leidmine on arvutuslikult vähenõudlik tegevus ja võimaldab elimineerida ebaolulised seosed [61] (Joonis 13). Kirjanduses on kasutatud valiku piirina korrelatsiooni 0,055 [61], mis sobis hästi ka autori läbiviidud eksperimentidega. Kui tunnuse ja tarbimise

korrelatsioon oli absoluutväärtuses alla 0,055, ei osunud tunnus rekursiivse tunnuste elimineerimise meetodi kasutamise järel mitte kunagi piisavalt oluliseks.



Joonis 13 Võimalike tunnuste ja tarbimise vahel korrelatsioonide leidmine.

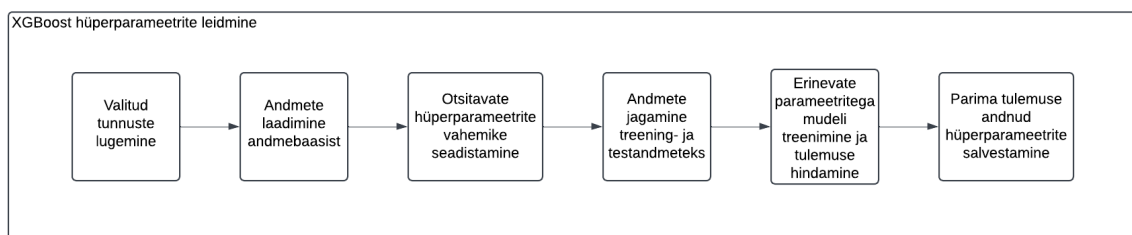
Läbi viidud eksperimentide järel pakkus autor välja nimekirja tunnustest, mille hulgast leitakse olulised: tarbimine, elektri hind, temperatuur, tund, tund koosinus, tund siinus, nädalapäev, nädalapäev koosinus, nädalapäev siinus, päev samas kuus, päev samas kuus koosinus, päev samas kuus siinus, päev aastas, päev aastas koosinus, päeva aastas siinus, nädal aastas, nädal aastas koosinus, nädal aastas siinus, kuu, kvartal, aasta, kas on talv, kas on kevad, kas on suvi, kas on sügis, päikese intensiivsus, õhuniiskus, viimase 10 minuti tuule suund, õhurõhk, vihma keskmiselt ühes tunnis ja tuule maksimaalne kiirus viimase tunni jooksul.

MK, TH ja VH majapidamiste korrelatsiooni 0,055 ületavad tunnused ja korrelatsioonide väärtused ajavahemikul 01.01.2023 00:00:00 – 31.12.2023 23:00:00 on toodud Lisas 2.

#### 4.5.3 Hüperparameetrite leidmine

XGBoost algoritm on paindlik ja selle käitumist juhitakse hüperparameetritega, mis määratavad mudeli käitumise õppimisel ja prognoosimisel [62], [63], [64], [65].

Kõikidele hüperparameetritele optimaalsete väärtuse leidmine (Joonis 14) on arvutusmahukas tegevus ja praktikas on mõistlik leida väärtused kõige olulisematele tunnustele.



Joonis 14 Hüperparameetrite leidmine.

Praktikas piisavalt hea mudeli jaoks on piisav, kui leida optimaalsed väärtused kõige olulisematele hüperparameetritele [62], [63], [64], [65]:

- `max_depth`: kasutatud puude maksimaalne sügavus. Liiga madalad väärtused põhjustavad alasobitamist (*underfit*) ja liiga sügavad puud ülesobitamist (*overfit*). Võimalike väärtuste vahemik  $0 \dots \infty$ , vaikeväärtusega 6.
- `min_child_weight`: minimaalne kaalude summa, mis on vajalik selleks, et toimuks sõlme jagunemine. Koos `max_depth` parameetriga määravad tasakaalu mudeli keerukuse ja üldistamisvõime vahel. Võimalike väärtuste vahemik  $0 \dots \infty$ , vaikeväärtusega 1.
- `subsample`: protsent, mis määrab ridade arvu iga puu treenimisel. Võimalik väärtuste vahemik  $0 \dots 1$ , vaikeväärtusega 1.
- `colsample_bytree`: protsent, mis määrab tunnuste arvu, mis valitakse juhuslikult puu konstrueerimisel. Koos `subsample` parameetriga lisavad juhuslikkust, mis vähendab mudeli ülesobitumist ja muudab mudeli robustsemaks. Võimalike väärtuste vahemik  $0 \dots 1$ , vaikeväärtusega 1.
- `learning_rate`, alias `eta`: vähendab igal treenimise ringil puude kaalu. Väiksem väärtus muudab mudeli robustsemaks ülesobitumise suhtes, samas muudab treenimise protsessi aeglasemaks. Võimalik väärtuste vahemik  $0 \dots \infty$ , vaikeväärtusega 0.3.

Ülejäänud hüperparameetrite väärtused võib jätta vaikeväärtusteks [64]. Juhul, kui toodud parameetrid ei anna praktikas piisavalt head tulemust, siis võib kaaluda järgmiste hüperparameetrite vaikeväärtuste muutmist [62], [64]:

- `gamma`: suurem väärtus muudab mudeli konservatiivsemaks, aitab vältida ülesobitumist. Võimalike väärtuste vahemik  $0 \dots \infty$ , vaikeväärtusega 0.
- `reg_alpha` ja `reg_lambda`: vähendavad mudeli kompleksust ja aitavad vähendada ülesobitumist. Võimalike väärtuste vahemik  $0 \dots \infty$ , `reg_alpha` vaikeväärtusega 0 ja `reg_lambda` vaikeväärtusega 1.

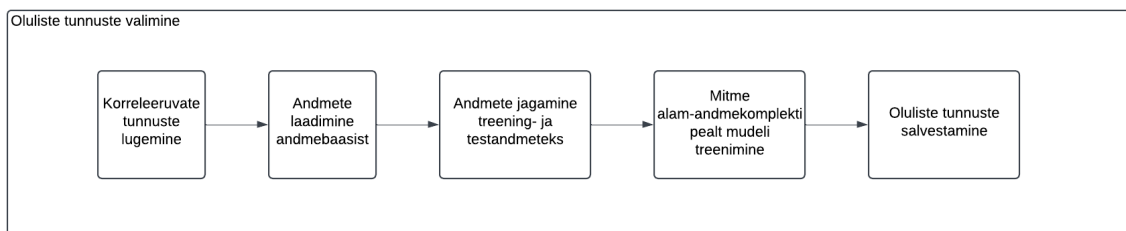
- `scale_pos_weight`: kasutatakse mudeli tasakaalustamiseks olukorras, kus negatiivsete väärtuste hulk andmetes on oluliselt suurem positiivsetest väärtustest. Positiivsetele väärtustele antakse suurem kaal. Võimalike väärtuste vahemik  $0 \dots \infty$ , vaikeväärtusega 1.
- `n_estimators`: määrab puude arvu mudelis. Väärtuse suurendamine suurendab enamasti ka mudeli täpsust, aga suurendab vajaminevat arvutusvõimsust ja seega ka töötamise koguaega. Parameetrite otsimisel on eesmärgiks leida selline puude arv, pärast mida uude puude lisamine mudeli täpsust enam ei tõsta. Võimalike väärtuste vahemik  $1 \dots \infty$ , vaikeväärtusega 100.

Eksperimentide põhjal otsustas autor mudelite treenimisel kasutada `colsample_bytree`, `gamma`, `learning_rate`, `max_depth`, `n_estimators`, `subsample` ja `early_stopping_rounds` hüperparameetreid.

#### 4.5.4 Oluliste tunnuste valimine

XGBoost mudeli prognoosi täpsus ei sõltu prognoosimisel kasutatavate tunnuste arvust, vaid nende olulisusest [66], millest tulenevalt on mudeli kiiruse tõstmiseks oluline valida väike arv olulisi tunnuseid (Joonis 15). Eelnevalt kasutas autor korrelatsioonanalüüsi valimaks välja tarbimisega enimkorreleeruvad tunnused. Siinkohal on oluline rõhutada, et korrelatsiooni ei tohi segamini ajada kausaalsusega [67]. Kui X korreleerub Y'ga, siis ei pruugi seos tähendada seda, et X põhjustab Y'i, aga võib.

Mudeli treenimiseks ja prognoosimisel kasutatavate tunnuste valimisel kasutas autor rekursiivset tunnuste elimineerimise (*Recursive Feature Elimination*) meetodit, täpsemalt selle kordustega valideerimise (*Recursive Feature Elimination Cross-Validation*) variatsiooni [68] [69].



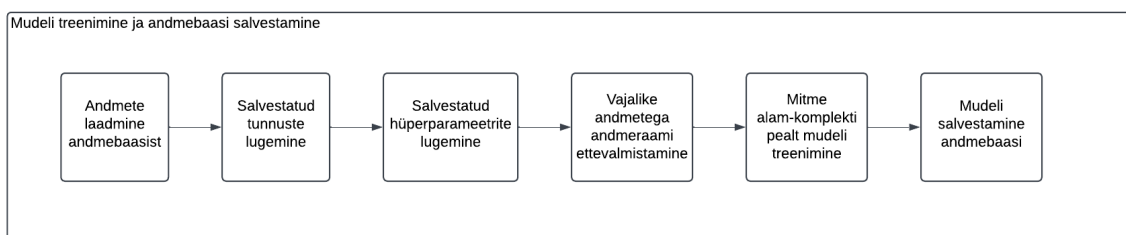
Joonis 15 Oluliste tunnuste valimine.

RFECV korral viiakse läbi järgmised sammud [68]:

- Samm 1: etteantud tunnuste pealt treenitakse baas-mudel, kus hüperparameetritena kasutatakse eelmises peatükis leitud optimaalseid parameetreid.
- Samm 2: iga tunnuse jaoks arvutatakse olulisus.
- Samm 3: kasutades kordustega aegridade valideerimist (*TimeSeriesSplit*), leitakse mudeli jaoks kõige olulisemad tunnused iga korduse jaoks.
- Samm 4: sammude 1 – 3 kordamine seni, kuni kõikide tunnuste olulisus on välja arvutatud.

#### 4.5.5 Mudeli treenimine

Kasutades optimaalseid hüperparameetreid ja olulisi tunnuseid, treenitakse mudel ja salvestatakse andmebaasi (Joonis 16).



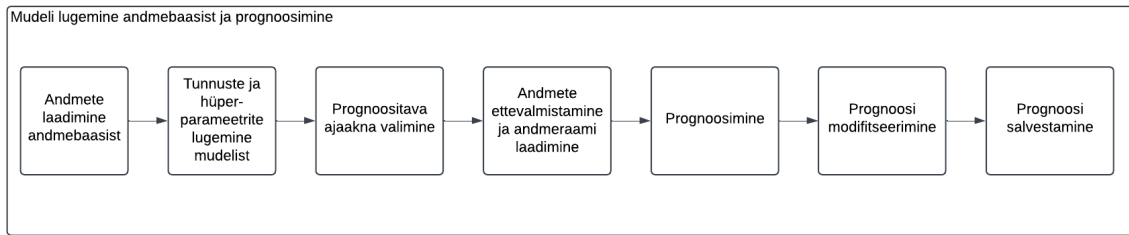
Joonis 16 Mudeli treenimine ja andmebaasi salvestamine.

Mudeli treenimiseks on vajalik relevantsete andmete ja eelmistes sammudes leitud parimate tunnuste ning optimaalsete hüperparameetrite olemasolu. Kuigi Joonis 16 järgi tunnused ja parameetrid laetakse, ei toimu otseselt laadimist, sest vastavad andmed on juba muutujates olemas. Mudeli treenimisel ja testimisel kasutatakse aegridade kordustega valideerimist tagamaks andmete ajaline järjestus valideerimisprotsessis. Valideerimise järel salvestatakse loodud mudel andmebaasi.

#### 4.5.6 Mudeliga prognoosimine

Prognoosimiseks laetakse andmebaasist mudel, prognoosimiseks vajalikud ilma- ja hinnaandmed, võetakse laetud mudelist tunnused ning hüperparameetrid, arvutatakse tuletatud tunnused, transformeeritakse andmed sobivale kujule ja viiakse läbi prognoosimine (Joonis 17). Prognoosi tulemus tehakse kättesaadavaks energiasalvestile, probleem, mille lahendamine ei kuulu käesoleva töö skoopi.





Joonis 17 Mudeli lugemine andmebaasist ja prognoosimine.

Prognoosimise järel modifitseeritakse prognoosi selleks, et tuua paremini välja tarbimise tipud ning võimaldada energiasalvestit tõhusamalt juhtida.

## 4.6 Uue majapidamise tarbimise prognoosimise eksperiment

Tõestamaks või lükkamaks ümber autori poolt lahendusena pakutud sarnase mudeli kontseptsiooni toimimist, lõi autor Python keeles PoC, mis simuleerib Joonis 3 kujutatud algoritmi rakendamist ühe kuu jooksul. Simulatsioon ei sisalda Joonis 4 kirjeldatud modifitseerimist, sest tegemist on eraldiseisva sammuga, mille toimimise verifitseerimist tuleb läbi viia koos energiasalvestusseadmega ja ei kuulu seega käesoleva töö skooopi.

Enne simulatsiooni käivitamist laadis autor andmebaasi vajalikud ilmaandmed, hinnaandmed, ajavahemiku 30.04.2022 – 30.04.2023 andmetega treenitud TH ja VH mudelid ja MK majapidamise reaalsed tarbimisandmed. Globaalseteks parameetriteks määras autor:

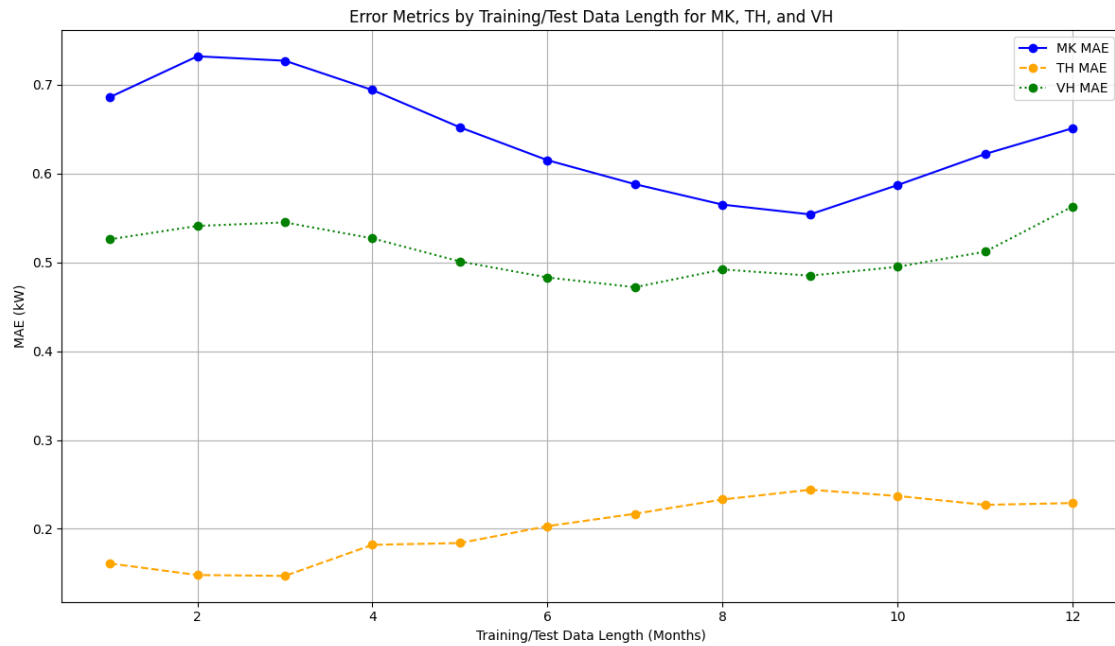
- Uue mudeli treenimise intervalliks 7 päeva.
- Ühe mudeliga prognoosimise kestvuseks 1 päev.
- Lubatud veaks 10%.
- Minimaalseks andmete pikkuseks, millega lubatakse mudel treenida, 7 päeva.
- Maksimaalseks andmete pikkuseks, millega lubatakse mudel treenida, 365 päeva.
- Ainult uue majapidamise mudeli kasutamise väärtuseks väär.
- Majapidamise nimeks, mille tarbimist prognoositakse, MK.
- Prognoosimise alguse kuupäevaks 01.05.2023
- Prognoosimise lõpukuupäevaks 31.05.2023

Täielik simulatsiooni väljund on toodud Lisas 3 ning siinkohal teeb autor lühikese kokkuvõtte tulemustest.

Keskmiselt kulus prognoosimiseks vajaliku mudeli leidmiseks 0,2 sekundit, prognoosimiseks 0,15 sekundit ja uue mudeli treenimiseks 68,73 sekundit. Prognoosimiseks kasutati ainult CPU arvutusressurssi ning prognoosimiseks kasutatud CPU skoor on 17011 [57]. Võrdluseks on Raspberry Pi 5 CPU (BCM2712) skooriks 3446 [70], mis on matemaatiliselt 5x aeglasem, aga praktikas on vahe väiksem, sest Raspberry Pi 5 CPU ei tegeleks samal ajal sama suures mahus kõrvaliste tegevusetega. Mis tähendab, et lahendus töötaks aktsepteeritava kiirusega Raspberry Pi 5 klassi arvutil.

Kokku prognoositi 31 päeva ja selleks kasutati TH mudelit 19 korral ja MK mudelit 12 korral. Keskmiseks prognoosimise koguveaks MAE kujunes 0,557 kW. VH mudel ei osutunud kasutatud ajavahemikus valituks mitte kordagi, mis võiks olla selgitatav VH majapidamises kasutatavate päikesepaneelidega. Eksperimentide, mida siinkohal ei kirjeldata, põhjal prognoosib VH mudel täpsemalt külmal ajal, näiteks valiti talvisel ajaperioodil läbiviidud eksperimendis MK majapidamise prognooside tegemiseks TH majapidamise asemel VH majapidamine.

Kuigi põhjapanevate järelduste tegemiseks on kasutatud mudelite arv väike, on ka kolme majapidamisega läbiviidud simulatsiooni põhjal näha autori pakutud lahenduse toimimist. Prognoosimiseks valitud parima majapidamise mudelid vaheldusid ning alati ei olnud parim MK majapidamise andmete pealt treenitud mudel. Olukorras, kus prognoosimiseks kasutatud mudeleid on palju, hakkab prognoositava majapidamise mudel võistlema teiste mudeliga ning ei pruugi alati parimaks osutada. Sõltuvalt majapidamise eripäradest, võib majapidamise mudeli treenida juba üsna lühikese ajavahemiku andmete pealt. Joonis 18 kirjeldab MK, TH ja VH majapidamiste treening ja testandmete pikkust ja prognoosimise viga MAE, kus mudelit treeniti erineva pikkusega andmetega ajavahemikus 01.01.2023 – 31.12.2023.



Joonis 18 MK, TH ja VH majapidamiste prognooside MAE seos treeningu perioodi pikkusega

Joonis 18 kohaselt saab näiteks TH majapidamise kohta üsnagi täpse prognoosi teha juba ühe kuu või isegi veel lühemate andmete pealt. Teisalt näitas reaalse andmetega läbiviidud simulatsioon, et ainult prognoositava majapidamise andmete pealt trenitud mudel ei olnud alati parim.

## 5 Kokkuvõte

Käesoleva töö üldiseks eesmärgiks oli prognoosida eramajapidamise energiatarbimist selleks, et efektiivselt juhtida majapidamise energiasalvestusseadet ning seeläbi saavutada majanduslik kasu. Konkreetseteks eesmärkideks on prognoosida majapidamise energiatarbimist olukorras, kus varasemad tarbimisandmed puuduvad või ei ole piisavalt, millal lugeda tarbimisandmete pikkus piisavaks ja kuidas prognoosida energiatarbimise tippe.

Tarbimisandmete vähesusega seotud probleemide lahendamiseks pakkus autor välja neli lahendust – treenida üks universaalne mudel, grupeerida majapidamised mingite tunnuste alusel ning treenida iga grupi jaoks üks mudel, kasutada sarnase majapidamise jaoks treenitud mudelit või kasutada triviaalset mudelit. Töös analüüsitakse ja kasutatakse sarnase mudeli lahendust.

Sarnase mudeli lahenduse algoritmi järgi oodatakse seni, kuni majapidamisel on tekkinud esimese 24 tunni tarbimisandmed, mille järel prognoositakse seda ajavahemikku kõikide olemasolevate mudelitega ning tehakse kindlaks kõige täpsemalt prognoositud mudel. Parimat mudelit kasutatakse edaspidi reaalsete prognooside tegemiseks. Minimaalse hulga tarbimisandmete tekkimisel treenitakse uue majapidamise andmetega mudel, kirjutatakse andmebaasi ning mudel hakkab võistlema teiste mudelitega. Rakendust juhitakse parameetritega, mis määravad, kui tihti uus mudel valitakse, uus mudel treenitakse, uue mudeli trenimiseks vajalike andmete pikkus, jne.

Energiatarbimise tippude prognoosimiseks pakkus autor välja kaks lahendust – kasutada kahte eraldi mudelit, milles üks prognoosib baas- ja teine tippude tarbimist või teise lahendusena korrigeerida prognoosimise järel baastarbimist, võimendades mudeli leitud trende. Autor kirjeldab töös trendide võimendamise lahendust, aga ei vii läbi praktilisi eksperimente lahenduse valideerimiseks, sest valideerimine eeldab sisendi andmist energiasalvesti juhtseadmele, mis ei ole käesoleva töö skoobi osa.

Püsitatud probleemide praktiliseks lahendamiseks valis autor XGBoost masinõppealgoritmi. Algoritm valiti esmalt kirjanduse põhjal, uurides analoogseid lahendusi ning seejärel valideeriti valik praktilise eksperimendiga.

Uue XGBoost mudeli treenimisel leitakse esmalt 31 tunnuse seast tunnused, mille korrelatsioon tarbimisega on suurem, kui 0,055, arvutatakse XGBoost mudeli hüperparameetrid piisavalt korreleerunud tunnustega, eemaldatakse rekursiivse tunnuste elimineerimise meetodiga ebaolulised tunnused, treenitakse mudel ning kirjutatakse treenitud mudel andmebaasi.

Kontrollimaks lahendust praktikas, kirjutas autor Python keeles programmi, mis simuleerib uut majapidamist olukorras, kus tarbimisandmed puuduvad või ei ole neid piisavalt. Ühe kuu simulatsiooni väljundid on toodud Lisas 3.

Simulatsiooni tulemused näitavad, et tuleviku energiatarbimist on võimalik prognoosida olukorras, kus olemasolevaid tarbimisandmeid ei ole piisavalt ning kinnitavad autori pakutud lahenduse toimimist. Samuti näitab läbiviidud simulatsioon XGBoost algoritmi väikest ressursivajadust, millest tulenevalt on võimalik pakutud lahendus piisava kiirusega tööle panna mõnel väikearvutil, näiteks Raspberry Pi 5.

Autori jaoks üllatuslikult prognoosivad triviaalsed mudelid paremini, kui esmapilgul võiks arvata. Energiatarbimise prognoosimise lahenduste väljatöötamisel peaks triviaalseid mudeleid kaaluma võimaliku lahendusena, ennekoike nende rakendamise kiiruse ja lihtsuse tõttu. Triviaalse mudeli rakendamine on jõukohane igaihele, kasutades näiteks tabelarvutustarkvara.

Järgmiste sammudena on autoril plaanis töö käigus valminud PoC kirjutada ümber realselt kasutatavaks tarkvaraks, lisades graafilise kasutajaliidese, parendada andmemudelit ning lihtsustada tarkvara kasutamise protseduure. Lisaks plaanib autor loodud tarkvarale lisada energiasalvesti majanduslikku tasuvuse hindamise funktsionaalsuse.

## Kasutatud kirjandus

- [1] “Flowchart drawings,” Lucid, [Online]. Available: <https://lucid.app/>. [Accessed 25 12 2024].
- [2] “Icons and Photos For Everything,” [Online]. Available: <https://thenounproject.com/>. [Accessed 26 11 2024].
- [3] “Code and CI/CD, powered by the Atlassian platform,” Atlassian, [Online]. Available: <https://bitbucket.org/>. [Accessed 16 10 2024].
- [4] “Andmekaitse ja Infoturbe Portaali,” Cybernetica AS, [Online]. Available: <https://akit.cyber.ee/>. [Accessed 20 12 2024].
- [5] “English-Estonian Dictionary,” [Online]. Available: <https://www.linguee.com/>. [Accessed 24 12 2024].
- [6] *LTAT.TK.013. Tehisintellekti algkursus*, Tartu Ülikool, 2020.
- [7] *LTAT.TK.015. Andmeteaduse võimalused äriettevõttes*, Tartu Ülikool, 2020.
- [8] K.-C. Kallaste, *Ülikoolist väljalangemise ennustamine masinõppe mudelite abil. Bakalaureusetöö*, Tartu Ülikool, 2023.
- [9] K.-L. Kutser, *Mitme sagedusega finantsandmete aegrea prognoosimise meetodite võrdlusanalüüs. Magistritöö*, Taltech, 2024.
- [10] “Taltech lõputööde register,” Taltech, [Online]. Available: <https://digikogu.taltech.ee/et/Search/Items?ItemTypes=9&ItemTypes=18&ItemTypes=19>. [Accessed 6 9 2024].
- [11] “Tartu Ülikooli Arvutiteaduste Instituudi lõputööde register,” Tartu Ülikool, [Online]. Available: [https://comserv.cs.ut.ee/ati\\_thesis/?year=all](https://comserv.cs.ut.ee/ati_thesis/?year=all). [Accessed 6 9 2024].
- [12] V. Tšetšnev, *Short-term load forecasting using artificial neural network. Bachelor's thesis*, Taltech, 2019.
- [13] I. Grets, *Energy load disaggregation approach based on heuristic optimization algorithms. Bachelor's thesis*, Taltech, 2019.
- [14] M. Maliarov, *Predicting residential electricity consumption using machine learning algorithms. Bachelor's thesis*, Taltech, 2020.
- [15] M. Arif-Uzzaman, *Forecast of the electricity consumption during pandemic time. Master's thesis*, Taltech, 2021.
- [16] G. Abdullayeva, *Application and Evaluation of LSTM Architectures for Energy Time-Series Forecasting. Master's Thesis*, University of Tartu, 2019.
- [17] M. Hatamian, *Predicting Location-Based Green Energy Availability in Smart Buildings. Master's Thesis*, University of Tartu, 2022.
- [18] T. Häring, *Research and Development of Energy Storage Control Strategies for Residential Area Microgrids. Doctoral thesis*, Taltech, 2022.
- [19] “Taltech Raamatukogu päring,” Taltech, [Online]. Available: <https://tutl-primo.hosted.exlibrisgroup.com/primo-explore/search?query=any,contains,machine%20>

- learning%20energy%20consumption%20prediction%20 single%20 house&tab=default\_tab&search\_scope=default\_scope& vid=372TUTL\_VU1&lang=en&offset=0. [Accessed 23 9 2024].
- [20] Z. Xia, R. Zhang, H. Ma and T. K. Saha, “Day-Ahead Electricity Consumption Prediction of Individual Household—Capturing Peak Consumption Pattern,” *IEEE TRANSACTIONS ON SMART GRID, VOL. 15, NO. 3*, pp. 2971-2984, May 2024.
- [21] N. Shabbir, R. Ahmadiyahangar, A. Rosin, M. Jawad, J. Kilter and J. Martins, “XgBoost based Short-term Electrical Load Forecasting Considering Trends & Periodicity in Historical Data,” in *IEEE International Conference on Energy Technologies for Future Grids (ETFG)*, 2023.
- [22] Z. Eddaoudia, Z. Aaraba, K. Boudmena, A. Elghazia and M. D. Rahmani, “A Brief Review of Energy Consumption Forecasting Using Machine Learning Models,” *Procedia Computer Science 236*, p. 33–40, 2024.
- [23] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos and G. K. Matsopoulos, “A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks,” *Future Internet 15 255*, pp. 1-31, 2023.
- [24] I. Paliari, A. Karanikola and S. Kotsiantis, “A comparison of the optimized LSTM, XGBOOST and ARIMA in Time Series forecasting,” in *12th International Conference on Information, Intelligence, Systems & Applications (IISA)*, 2021.
- [25] S. Saluja and S. Palwe, “Energy Consumption Prediction Using Machine Learning Algorithms,” in *IEEE Pune Section International Conference (PuneCon)*, 2023.
- [26] S. Hadri, M. Najib, M. Bakhouya, Y. Fakhri and M. El Arroussi, “Performance Evaluation of Forecasting Strategies for Electricity Consumption in Buildings,” *Energies 14 5831*, pp. 1-16, 2021.
- [27] H. Rathore, H. K. Meena and P. Jain, “Prediction of EV Energy consumption Using Random Forest And XGBoost,” in *International Conference on Power Electronics and Energy (ICPEE)*, 2023.
- [28] “What are ARIMA models?,” IBM, [Online]. Available: <https://www.ibm.com/think/topics/arima-model>. [Accessed 17 9 2024].
- [29] R. J. Hyndman and G. Athanasopoulos, “Forecasting: Principles and Practice. 8.1 Stationarity and differencing,” Monash University, Australia, [Online]. Available: <https://otexts.com/fpp2/stationarity.html>. [Accessed 16 11 2024].
- [30] R. J. Hyndman and G. Athanasopoulos, “Forecasting: Principles and Practice. 8.9 Seasonal ARIMA models,” Monash University, Australia, [Online]. Available: <https://otexts.com/fpp2/seasonal-arima.html>. [Accessed 16 11 2024].
- [31] N. Donges, “Random Forest: A Complete Guide for Machine Learning. All you need to know about the random forest model in machine learning.,” [Online]. Available: <https://builtin.com/data-science/random-forest-algorithm>. [Accessed 10 11 2024].
- [32] “What is random forest?,” IBM, [Online]. Available: <https://www.ibm.com/think/topics/random-forest>. [Accessed 9 11 2024].
- [33] S. Bangar, “Understanding Support Vector Regression: A Guide to Machine Learning’s Powerful Tool,” [Online]. Available: <https://medium.com/@siddheshb008/understanding-support-vector-regression-a-guide-to-machine-learnings-powerful-tool-4ada7081934b>. [Accessed 6 11 2024].

- [34] “What are SVMs?,” IBM, [Online]. Available: <https://www.ibm.com/think/topics/support-vector-machine>. [Accessed 17 9 2024].
- [35] “Introduction to Boosted Trees,” [Online]. Available: <https://xgboost.readthedocs.io/en/stable/tutorials/model.html>. [Accessed 6 11 2024].
- [36] “Ajaloollised ilmaandmed,” Keskkonnaagentuur, [Online]. Available: <https://www.ilmateenistus.ee/kliima/ajaloollised-ilmaandmed/>. [Accessed 8 9 2024].
- [37] C. Wade, “Chapter 5: XGBoost Unveiled,” in *Hands-On Gradient Boosting with XGBoost and scikit-learn*, vol. Chapter 5: XGBoost Unveiled, Packt Publishing, 2020, p. 310.
- [38] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
- [39] W. Wang, G. Chakraborty and B. Chakraborty, “Predicting the Risk of Chronic Kidney Disease (CKD) Using Machine Learning Algorithm,” *Applied Sciences 11* 202, pp. 1-17, 2021.
- [40] J. Choudhary, “Mastering XGBoost: A Technical Guide for Machine Learning Practitioners,” [Online]. Available: <https://medium.com/@jyotsna.a.choudhary/mastering-xgboost-a-technical-guide-for-intermediate-machine-learning-practitioners-f7ad167c6865>. [Accessed 1 11 2024].
- [41] J. Brownlee, “Making You Awesome At XGBoost. What is XGBoost,” [Online]. Available: <https://xgboosting.com/what-is-xgboost/>. [Accessed 4 11 2024].
- [42] J. Brownlee, “Making You Awesome At XGBoost. XGBoost Regularization Techniques,” [Online]. Available: <https://xgboosting.com/xgboost-regularization-techniques/>. [Accessed 4 11 2024].
- [43] J. Brownlee, “Making You Awesome At XGBoost. What is XGBoost Fast?,” [Online]. Available: <https://xgboosting.com/what-is-xgboost-fast/>. [Accessed 4 11 2024].
- [44] A. Chugh, “MAE, MSE, RMSE, Coefficient of Determination, Adjusted R Squared — Which Metric is Better?,” [Online]. Available: <https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e>. [Accessed 7 11 2024].
- [45] X. M. Zhang, K. Grolinger, M. A. M. Capretz and L. Seewald, “Forecasting Residential Energy Consumption: Single Household Perspective,” in *17th IEEE International Conference on Machine Learning and Applications*, 2018.
- [46] W. Yang, J. Shi, S. Li, Z. Song, Z. Zhang and Z. Chen, “A combined deep learning load forecasting model of single household resident user considering multi-time scale electricity consumption behavior,” *Applied Energy* 307 118197, pp. 1-17, 2022.
- [47] V. Andrade, H. Morais and L. Pereira, “Industrial kitchen appliance consumption forecasting: Hour-ahead and day-ahead perspectives with post-processing improvements,” *Computers and Electrical Engineering* 116 109145, pp. 1-21, 2024.



- [48] L. S. Saoud, H. Al-Marzouqi and R. Hussein, "Household Energy Consumption Prediction Using the Stationary Wavelet Transform and Transformers," *Digital Object Identifier 10 1109 / ACCESS 3140818*, pp. 5171 - 5183, 2022.
- [49] J. García-López, S. Domínguez-Amarillo and J. J. Sendra, "Clustering Open Data for Predictive Modeling of Residential Energy Consumption across Variable Scales: A Case Study in Andalusia, Spain," *Buildings 14 2335*, pp. 1-23, 2024.
- [50] "Elering elektriandmed," Elering, [Online]. Available: <https://elering.ee/>. [Accessed 5 9 2024].
- [51] "10 Best Language for Machine Learning," [Online]. Available: <https://www.geeksforgeeks.org/best-language-for-machine-learning/>. [Accessed 16 9 2024].
- [52] "Best IDEs For Machine Learning," [Online]. Available: <https://www.geeksforgeeks.org/best-ides-for-machine-learning/>. [Accessed 8 9 2024].
- [53] "SQLite," [Online]. Available: <https://www.sqlite.org/>. [Accessed 15 9 2024].
- [54] "DBeaver Universal Database Tool," [Online]. Available: <https://dbeaver.io/>. [Accessed 16 9 2024].
- [55] "ERD drawings," Vertabelo, [Online]. Available: <https://vertabelo.com/>. [Accessed 17 12 2024].
- [56] "ChatGPT," [Online]. Available: <https://chatgpt.com/>. [Accessed 31 12 2024].
- [57] "CPU Benchmarks: AMD Ryzen 7 PRO 5850U," [Online]. Available: <https://www.cpubenchmark.net/cpu.php?cpu=AMD+Ryzen+7+PRO+5850U&id=4198>. [Accessed 28 11 2024].
- [58] "Skewness Formula," [Online]. Available: <https://www.geeksforgeeks.org/skewness-formula/>. [Accessed 12 10 2024].
- [59] I. Lutsar, "Covid-19 pandeemia kirjeldav analüüs ja õppetunnid," 2023. [Online]. Available: <https://s3.documentcloud.org/documents/24172929/ta-covid-19-pandeemia-analuus-2023.pdf>. [Accessed 16 10 2024].
- [60] A. Krishnan, "How XGBoost Handles Missing Data: A Guide to Optimized Decision Trees and Robust Predictions," [Online]. Available: <https://medium.com/@abhilashkrish/how-xgboost-handles-missing-data-a-guide-to-optimized-decision-trees-and-robust-predictions-0db33363e855>. [Accessed 11 11 2024].
- [61] C. Li, D. Zhu, C. Hu, X. Li, S. Nan and H. Huang, "ECDX: Energy consumption prediction model based on distance correlation and XGBoost for edge data center," *Information Sciences 643 119218*, pp. 1-13, 2023.
- [62] C. Wade, "Chapter 6: XGBoost Hyperparameters," in *Hands-On Gradient Boosting with XGBoost and scikit-learn*, Packt Publishing, 2020, p. 310.
- [63] P. Banerjee, "A Guide on XGBoost hyperparameters tuning," [Online]. Available: <https://www.kaggle.com/code/prashant111/a-guide-on-xgboost-hyperparameters-tuning>. [Accessed 14 11 2024].
- [64] J. Brownlee, "Making You Awesome At XGBoost. Most Important XGBoost Hyperparameters to Tune," [Online]. Available: <https://xgboosting.com/most-important-xgboost-hyperparameters-to-tune/>. [Accessed 4 11 2024].
- [65] "XGBoost Parameters," [Online]. Available: <https://xgboost.readthedocs.io/en/latest/parameter.html>. [Accessed 4 11 2024].

- [66] “Feature Importance With XGBoost in Python,” Machine Learning Expedition, [Online]. Available: <https://www.machinelearningexpedition.com/feature-importance-with-xgboost-in-python/>. [Accessed 2 11 2024].
- [67] R. J. Hyndman and G. Athanasopoulos, “Forecasting: Principles and Practice. 5.9 Correlation, causation and forecasting,” Monash University, Australia, [Online]. Available: <https://otexts.com/fpp2/causality.html>. [Accessed 12 11 2024].
- [68] K. Shi, R. Shi, T. Fu, Z. Lu and J. Zhang, “A Novel Identification Approach Using RFECV–Optuna–XGBoost for Assessing Surrounding Rock Grade of Tunnel Boring Machine Based on Tunneling Parameters,” *Applied Sciences* 14 2347, pp. 1-23, 2024.
- [69] J. Brownlee, “Making You Awesome At XGBoost. XGBoost Feature Selection with RFE,” [Online]. Available: <https://xgboosting.com/xgboost-feature-selection-with-rfe/>. [Accessed 4 11 2024].
- [70] “CPU Benchmarks: Raspberry Pi 5 (BCM2712),” [Online]. Available: <https://www.cpubenchmark.net/cpu.php?cpu=BCM2712&id=6054>. [Accessed 28 11 2024].

## **Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks<sup>1</sup>**

Mina, Kaido Sõmera

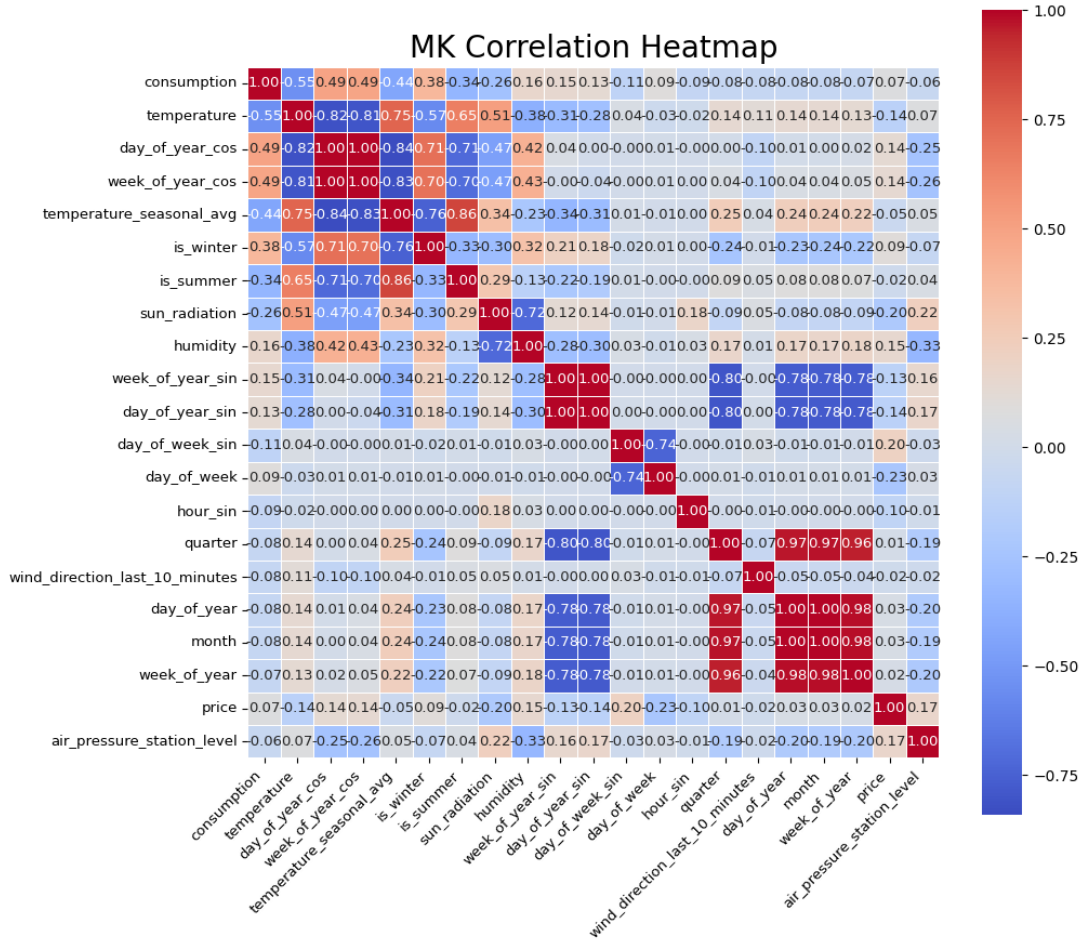
1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Eramajapidamise energiatarbimise ennustamine“, mille juhendaja on Vahur Kotkas
  - 1.1. reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise 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.

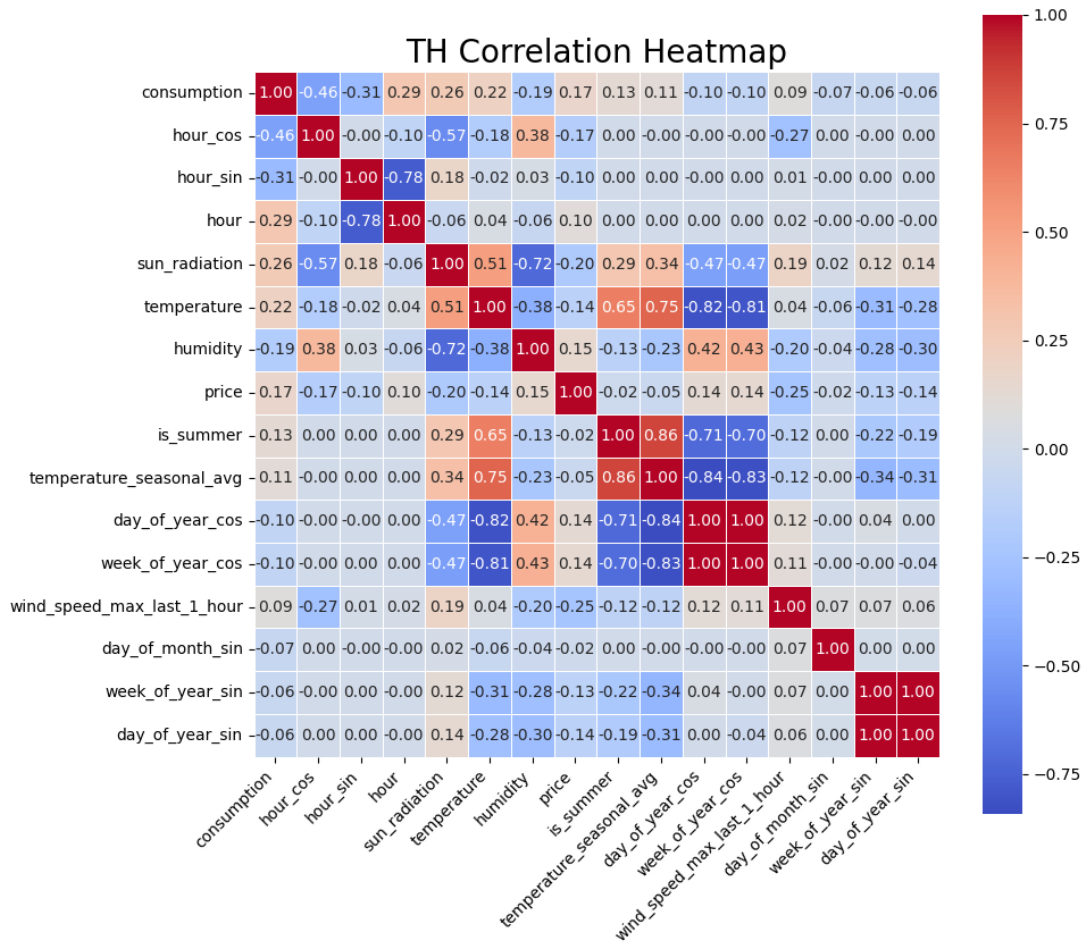
02.12.2024

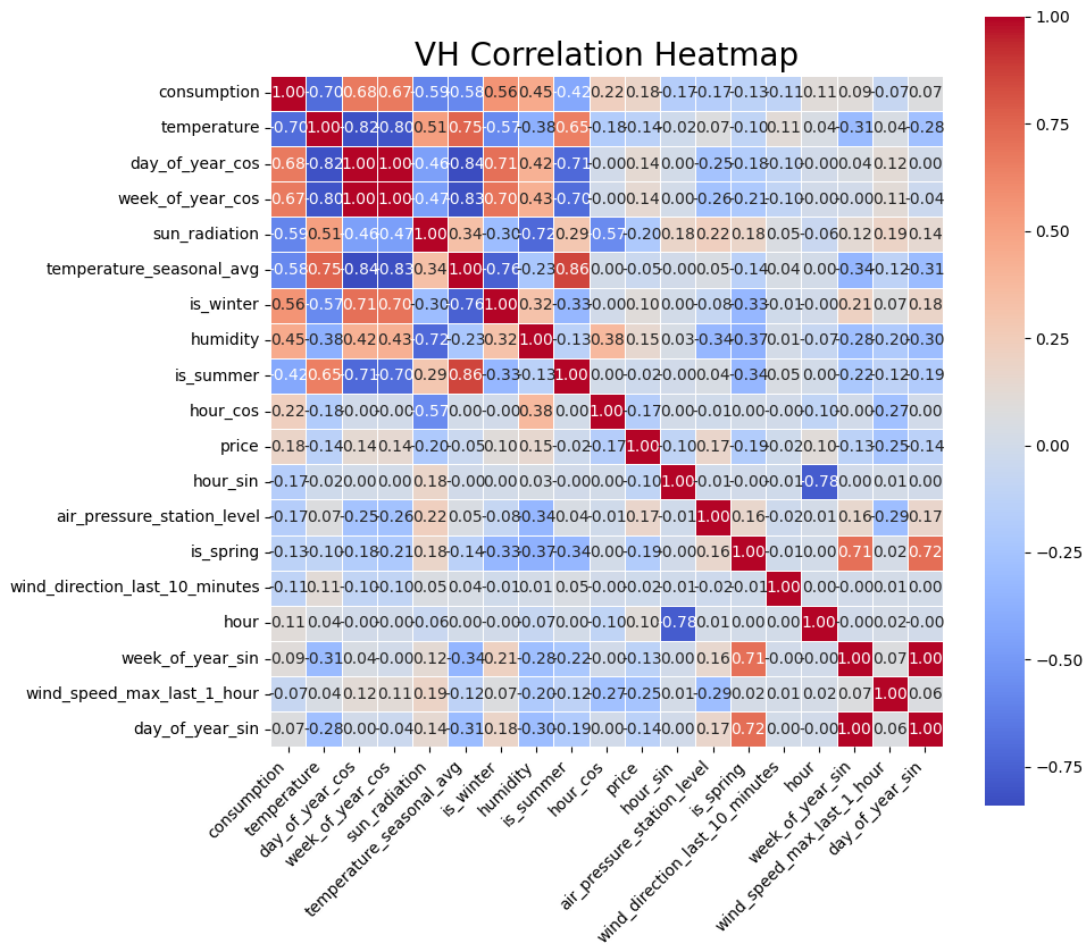
---

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

## Lisa 2 – Majapidamiste tunnuste vahelised korrelatsioonid







## **Lisa 3 – MK majapidamise energiatarbimise prognoosimise simulatsiooni väljund**

Household vh features: ['sun\_radiation', 'hour\_sin', 'hour\_cos']

Household th features: ['hour\_cos', 'hour', 'hour\_sin', 'day\_of\_year\_cos']

##### Current datetime: 2023-05-01

##### New mk model training START

Data length of new household is NOT enough -> will NOT train model

Model training section running time: 0.00 seconds

##### Selecting best model START

All currently available models: ['vh', 'th']

vh household prediction MAE: 1.1730 kW

th household prediction MAE: 0.9167 kW

Best model: th

Best model selection section running time: 0.18 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-02

Prediction MAE: 0.7808 kW

Maximum allowed MAE: 3.3915 kW

Prediction section running time: 0.13 seconds

##### Current datetime: 2023-05-02

##### New mk model training START

Data length of new household is NOT enough -> will NOT train model

Model training section running time: 0.00 seconds

##### Selecting best model START

All currently available models: ['vh', 'th']

vh household prediction MAE: 1.0264 kW

th household prediction MAE: 0.7808 kW

Best model: th

Best model selection section running time: 0.18 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-03

Prediction MAE: 0.8057 kW

Maximum allowed MAE: 3.1527 kW

Prediction section running time: 0.15 seconds

##### Current datetime: 2023-05-03

##### New mk model training START

Data length of new household is NOT enough -> will NOT train model

Model training section running time: 0.00 seconds

##### Selecting best model START

All currently available models: ['vh', 'th']



vh household prediction MAE: 0.9067 kW

th household prediction MAE: 0.8057 kW

Best model: th

Best model selection section running time: 0.18 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-04

Prediction MAE: 0.7709 kW

Maximum allowed MAE: 3.3593 kW

Prediction section running time: 0.16 seconds

##### Current datetime: 2023-05-04

##### New mk model training START

Data length of new household is NOT enough -> will NOT train model

Model training section running time: 0.00 seconds

##### Selecting best model START

All currently available models: ['vh', 'th']

vh household prediction MAE: 0.9408 kW

th household prediction MAE: 0.7709 kW

Best model: th

Best model selection section running time: 0.17 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-05

Prediction MAE: 0.7667 kW

Maximum allowed MAE: 3.0993 kW

Prediction section running time: 0.16 seconds

##### Current datetime: 2023-05-05

##### New mk model training START

Data length of new household is NOT enough -> will NOT train model

Model training section running time: 0.00 seconds

##### Selecting best model START

All currently available models: ['vh', 'th']

vh household prediction MAE: 0.9247 kW

th household prediction MAE: 0.7667 kW

Best model: th

Best model selection section running time: 0.17 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-06

Prediction MAE: 0.7641 kW

Maximum allowed MAE: 4.1110 kW

Prediction section running time: 0.15 seconds

##### Current datetime: 2023-05-06

##### New mk model training START

Data length of new household is NOT enough -> will NOT train model

Model training section running time: 0.00 seconds

##### Selecting best model START

All currently available models: ['vh', 'th']

vh household prediction MAE: 1.1506 kW

th household prediction MAE: 0.7641 kW

Best model: th

Best model selection section running time: 0.16 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-07

Prediction MAE: 0.8056 kW

Maximum allowed MAE: 3.6823 kW

Prediction section running time: 0.15 seconds

##### Current datetime: 2023-05-07

##### New mk model training START

Data length of new household is NOT enough -> will NOT train model

Model training section running time: 0.00 seconds

##### Selecting best model START

All currently available models: ['vh', 'th']

vh household prediction MAE: 1.0271 kW

th household prediction MAE: 0.8056 kW

Best model: th

Best model selection section running time: 0.17 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-08

Prediction MAE: 0.8033 kW

Maximum allowed MAE: 3.5644 kW

Prediction section running time: 0.15 seconds

##### Current datetime: 2023-05-08

##### New mk model training START

Minimal amount of data in DB

Model is not in DB -> will train model

Important correlations: ['sun\_radiation', 'hour\_cos', 'humidity', 'temperature', 'wind\_speed\_max\_last\_1\_hour', 'hour\_sin', 'day\_of\_week\_cos', 'day\_of\_week\_sin', 'rain\_total\_in\_hour']

Best hyperparameters: ['colsample\_bytree: 1.00', 'gamma: 3.00', 'learning\_rate: 0.05', 'max\_depth: 9.00', 'n\_estimators: 1000.00', 'subsample: 0.99']

Important features: ['humidity', 'sun\_radiation', 'temperature', 'wind\_speed\_max\_last\_1\_hour', 'hour\_sin', 'hour\_cos', 'rain\_total\_in\_hour']

Overall MAE: 0.6225 kW

Model training section running time: 79.27 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.9601 kW

th household prediction MAE: 0.8033 kW

mk household prediction MAE: 0.5912 kW

Best model: mk

Best model selection section running time: 0.23 seconds

##### Prediction START

Household model: mk

Day to predict: 2023-05-09

Prediction MAE: 0.5522 kW

Maximum allowed MAE: 2.6250 kW

Prediction section running time: 0.14 seconds

##### Current datetime: 2023-05-09

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.02 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.8240 kW

th household prediction MAE: 0.4947 kW

mk household prediction MAE: 0.5522 kW

Best model: th

Best model selection section running time: 0.23 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-10

Prediction MAE: 0.6146 kW

Maximum allowed MAE: 2.7943 kW

Prediction section running time: 0.15 seconds

##### Current datetime: 2023-05-10

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.02 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.7559 kW

th household prediction MAE: 0.6146 kW

mk household prediction MAE: 0.5220 kW

Best model: mk

Best model selection section running time: 0.23 seconds

##### Prediction START

Household model: mk

Day to predict: 2023-05-11

Prediction MAE: 0.7165 kW

Maximum allowed MAE: 2.2282 kW

Prediction section running time: 0.15 seconds

##### Current datetime: 2023-05-11

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.01 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.7974 kW

th household prediction MAE: 0.6157 kW

mk household prediction MAE: 0.7165 kW

Best model: th

Best model selection section running time: 0.23 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-12

Prediction MAE: 0.5164 kW

Maximum allowed MAE: 2.4801 kW

Prediction section running time: 0.15 seconds

##### Current datetime: 2023-05-12

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.01 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.7546 kW

th household prediction MAE: 0.5164 kW

mk household prediction MAE: 0.5338 kW

Best model: th

Best model selection section running time: 0.23 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-13

Prediction MAE: 0.5482 kW



Maximum allowed MAE: 2.5644 kW

Prediction section running time: 0.14 seconds

##### Current datetime: 2023-05-13

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.01 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.7508 kW

th household prediction MAE: 0.5482 kW

mk household prediction MAE: 0.6783 kW

Best model: th

Best model selection section running time: 0.23 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-14

Prediction MAE: 0.2798 kW

Maximum allowed MAE: 2.7306 kW

Prediction section running time: 0.15 seconds

##### Current datetime: 2023-05-14

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.02 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.9461 kW

th household prediction MAE: 0.2798 kW

mk household prediction MAE: 0.5777 kW

Best model: th

Best model selection section running time: 0.23 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-15

Prediction MAE: 0.5994 kW

Maximum allowed MAE: 2.4826 kW

Prediction section running time: 0.14 seconds

##### Current datetime: 2023-05-15

##### New mk model training START

Minimal amount of data in DB

Enough time has passed since last training -> will train model

Important correlations: ['temperature', 'sun\_radiation', 'humidity', 'day\_of\_month\_cos', 'day', 'day\_of\_year', 'day\_of\_year\_cos', 'day\_of\_year\_sin', 'week\_of\_year\_cos', 'week\_of\_year', 'week\_of\_year\_sin', 'hour\_cos', 'rain\_total\_in\_hour', 'day\_of\_week\_cos', 'price', 'hour\_sin']

Best hyperparameters: ['colsample\_bytree: 0.60', 'gamma: 1.00', 'learning\_rate: 0.15', 'max\_depth: 10.00', 'n\_estimators: 239.00', 'subsample: 1.00']

Important features: ['sun\_radiation', 'hour\_cos', 'day\_of\_month\_cos', 'humidity', 'rain\_total\_in\_hour', 'hour\_sin', 'price']

Overall MAE: 0.4422 kW

Model training section running time: 47.84 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.7517 kW

th household prediction MAE: 0.5994 kW

mk household prediction MAE: 0.4438 kW

Best model: mk

Best model selection section running time: 0.19 seconds

##### Prediction START

Household model: mk

Day to predict: 2023-05-16

Prediction MAE: 0.5664 kW

Maximum allowed MAE: 1.9283 kW

Prediction section running time: 0.14 seconds

##### Current datetime: 2023-05-16

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.00 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.7522 kW

th household prediction MAE: 0.4522 kW

mk household prediction MAE: 0.5664 kW

Best model: th

Best model selection section running time: 0.19 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-17

Prediction MAE: 0.4949 kW

Maximum allowed MAE: 2.9267 kW

Prediction section running time: 0.14 seconds

##### Current datetime: 2023-05-17

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.00 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.7722 kW

th household prediction MAE: 0.4949 kW

mk household prediction MAE: 0.4953 kW

Best model: th

Best model selection section running time: 0.20 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-18

Prediction MAE: 0.7901 kW

Maximum allowed MAE: 3.1203 kW

Prediction section running time: 0.15 seconds

##### Current datetime: 2023-05-18

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.01 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.8975 kW

th household prediction MAE: 0.7901 kW

mk household prediction MAE: 0.6763 kW

Best model: mk

Best model selection section running time: 0.20 seconds

##### Prediction START

Household model: mk

Day to predict: 2023-05-19

Prediction MAE: 0.3889 kW

Maximum allowed MAE: 2.7361 kW

Prediction section running time: 0.15 seconds

##### Current datetime: 2023-05-19

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.00 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.7737 kW

th household prediction MAE: 0.4970 kW

mk household prediction MAE: 0.3889 kW

Best model: mk

Best model selection section running time: 0.19 seconds

##### Prediction START

Household model: mk

Day to predict: 2023-05-20

Prediction MAE: 0.4969 kW

Maximum allowed MAE: 3.0013 kW

Prediction section running time: 0.15 seconds

##### Current datetime: 2023-05-20

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.01 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.8900 kW

th household prediction MAE: 0.4593 kW

mk household prediction MAE: 0.4969 kW

Best model: th

Best model selection section running time: 0.18 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-21

Prediction MAE: 0.3832 kW

Maximum allowed MAE: 2.5154 kW

Prediction section running time: 0.15 seconds

##### Current datetime: 2023-05-21

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.01 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.9021 kW

th household prediction MAE: 0.3832 kW

mk household prediction MAE: 0.3961 kW

Best model: th

Best model selection section running time: 0.19 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-22



Prediction MAE: 0.5634 kW

Maximum allowed MAE: 2.3135 kW

Prediction section running time: 0.15 seconds

##### Current datetime: 2023-05-22

##### New mk model training START

Minimal amount of data in DB

Enough time has passed since last training -> will train model

Important correlations: ['temperature', 'day\_of\_month\_cos', 'week\_of\_year\_cos', 'sun\_radiation', 'week\_of\_year', 'day\_of\_year\_cos', 'day', 'day\_of\_year', 'week\_of\_year\_sin', 'day\_of\_year\_sin', 'hour\_cos', 'humidity', 'price', 'day\_of\_month\_sin', 'wind\_direction\_last\_10\_minutes', 'hour\_sin', 'rain\_total\_in\_hour']

Best hyperparameters: ['colsample\_bytree: 0.78', 'gamma: 5.00', 'learning\_rate: 0.01', 'max\_depth: 6.00', 'n\_estimators: 269.00', 'subsample: 0.78']

Important features: ['sun\_radiation', 'hour\_cos', 'rain\_total\_in\_hour', 'price', 'day\_of\_year\_cos', 'temperature', 'day\_of\_month\_cos', 'hour\_sin']

Overall MAE: 0.5244 kW

Model training section running time: 56.64 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.6162 kW

th household prediction MAE: 0.5634 kW

mk household prediction MAE: 0.4708 kW

Best model: mk

Best model selection section running time: 0.19 seconds

##### Prediction START

Household model: mk

Day to predict: 2023-05-23

Prediction MAE: 0.4272 kW

Maximum allowed MAE: 1.9639 kW

Prediction section running time: 0.14 seconds

##### Current datetime: 2023-05-23

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.01 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.7214 kW

th household prediction MAE: 0.4242 kW

mk household prediction MAE: 0.4272 kW

Best model: th

Best model selection section running time: 0.20 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-24

Prediction MAE: 0.5050 kW

Maximum allowed MAE: 1.9823 kW

Prediction section running time: 0.14 seconds

##### Current datetime: 2023-05-24

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.00 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.7055 kW

th household prediction MAE: 0.5050 kW

mk household prediction MAE: 0.4832 kW

Best model: mk

Best model selection section running time: 0.19 seconds

##### Prediction START

Household model: mk

Day to predict: 2023-05-25

Prediction MAE: 0.4690 kW

Maximum allowed MAE: 1.9810 kW

Prediction section running time: 0.14 seconds

##### Current datetime: 2023-05-25

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.00 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.7013 kW

th household prediction MAE: 0.5151 kW

mk household prediction MAE: 0.4690 kW

Best model: mk

Best model selection section running time: 0.20 seconds

##### Prediction START

Household model: mk

Day to predict: 2023-05-26

Prediction MAE: 0.4612 kW

Maximum allowed MAE: 2.3188 kW

Prediction section running time: 0.14 seconds

##### Current datetime: 2023-05-26

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.01 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.6020 kW

th household prediction MAE: 0.6056 kW

mk household prediction MAE: 0.4612 kW

Best model: mk

Best model selection section running time: 0.20 seconds

##### Prediction START

Household model: mk

Day to predict: 2023-05-27

Prediction MAE: 0.6167 kW

Maximum allowed MAE: 3.2937 kW

Prediction section running time: 0.14 seconds

##### Current datetime: 2023-05-27

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.01 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 1.0208 kW

th household prediction MAE: 0.5617 kW

mk household prediction MAE: 0.6167 kW

Best model: th

Best model selection section running time: 0.19 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-28

Prediction MAE: 0.5047 kW

Maximum allowed MAE: 3.1133 kW

Prediction section running time: 0.14 seconds

##### Current datetime: 2023-05-28

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.01 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 1.0512 kW

th household prediction MAE: 0.5047 kW

mk household prediction MAE: 0.5280 kW

Best model: th

Best model selection section running time: 0.20 seconds

##### Prediction START

Household model: th

Day to predict: 2023-05-29

Prediction MAE: 0.5992 kW

Maximum allowed MAE: 2.6293 kW

Prediction section running time: 0.14 seconds

##### Current datetime: 2023-05-29

##### New mk model training START

Minimal amount of data in DB

Enough time has passed since last training -> will train model

Important correlations: ['temperature', 'week\_of\_year\_cos', 'week\_of\_year', 'day\_of\_year\_cos', 'week\_of\_year\_sin', 'day', 'day\_of\_year', 'day\_of\_year\_sin', 'sun\_radiation', 'day\_of\_month\_sin', 'hour\_cos', 'day\_of\_month\_cos', 'humidity', 'day\_of\_week\_sin', 'price', 'air\_pressure\_station\_level', 'wind\_direction\_last\_10\_minutes', 'rain\_total\_in\_hour', 'hour\_sin']

Best hyperparameters: ['colsample\_bytree: 0.61', 'gamma: 2.00', 'learning\_rate: 0.21', 'max\_depth: 10.00', 'n\_estimators: 409.00', 'subsample: 0.92']

Important features: ['humidity', 'sun\_radiation', 'day\_of\_year', 'temperature', 'day\_of\_week\_sin', 'price', 'day\_of\_year\_cos']

Overall MAE: 0.4392 kW

Model training section running time: 73.37 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.7749 kW

th household prediction MAE: 0.5992 kW

mk household prediction MAE: 0.4955 kW

Best model: mk

Best model selection section running time: 0.19 seconds

##### Prediction START

Household model: mk

Day to predict: 2023-05-30

Prediction MAE: 0.3543 kW

Maximum allowed MAE: 2.2716 kW

Prediction section running time: 0.14 seconds

##### Current datetime: 2023-05-30

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.01 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']



vh household prediction MAE: 0.6355 kW

th household prediction MAE: 0.4994 kW

mk household prediction MAE: 0.3543 kW

Best model: mk

Best model selection section running time: 0.20 seconds

##### Prediction START

Household model: mk

Day to predict: 2023-05-31

Prediction MAE: 0.5532 kW

Maximum allowed MAE: 2.9382 kW

Prediction section running time: 0.14 seconds

##### Current datetime: 2023-05-31

##### New mk model training START

Minimal amount of data in DB

NOT enough time has passed since last training -> will NOT train model

Model training section running time: 0.01 seconds

##### Selecting best model START

All currently available models: ['vh', 'th', 'mk']

vh household prediction MAE: 0.8262 kW

th household prediction MAE: 0.6276 kW

mk household prediction MAE: 0.5532 kW

Best model: mk

Best model selection section running time: 0.20 seconds

##### Prediction START

Household model: mk

Day to predict: 2023-06-01

Prediction MAE: 0.4765 kW

Maximum allowed MAE: 2.9898 kW

Prediction section running time: 0.14 seconds

Code overall running time: 267.98 seconds

End of simulation...