

Saulės fotovoltinių elektrinių valandinės energijos prognozavimo strategijų ir modelių palyginimas

Ratmir Semionov, Lukas Voveris, Jolita Bernatavičienė

Vilniaus universitetas, Matematikos ir informatikos fakultetas,
Duomenų mokslo ir skaitmeninių technologijų institutas,
Akademijos g. 4, Vilnius, Lietuva
*ratmir.semionov@mif.stud.vu.lt, lukas.voveris@mif.stud.vu.lt,
jolita.bernataviciene@mif.vu.lt*

Santrauka. Šiame tyrime nagrinėjamas valandinis saulės fotovoltinių elektrinių energijos prognozavimas Lietuvoje, lyginant tiesioginio ir dviejų pakopų prognozavimo strategijas. Analizei taikyti kelių grupių modeliai: klasikiniai mašininio mokymosi modeliai – atraminių vektorių modelis (SVM), atsitiktinių miškų modelis (RF) ir ekstremalaus gradientinio stiprinimo modelis (XGBoost), bei gilieji neuroniniai tinklai – ilgos trumpalaikės atminties (LSTM) ir laikinis suliejimo transformeris (TFT). Tiesioginio prognozavimo strategijoje (kai modelis tiesiogiai prognozuoja energijos gamybą, naudodamas meteorologinius duomenis) geriausi dieninės paklaidos rezultatai gauti naudojant XGBoost modelį šiaurės ir centrinėje elektrinėse, kur dienos svorinė vidutinė absoliuti procentinė paklaida (DWMAPE) siekė atitinkamai 13,24 % ir 16,25 %, o pietinėje elektrinėje geriausius rezultatus pasiekė TFT modelis, pasiekęs 12,68 % DWMAPE. Dviejų pakopų strategijos pirmoje pakopoje bendros dienos energijos prognozėms tiksliausi modeliai skyrėsi tarp elektrinių: šiaurės elektrinėje geriausius rezultatus pasiekė RF modelis, centrinėje – SVM, o pietinėje – TFT. Antroje pakopoje valandinių proporcijų prognozėse visose elektrinėse geriausi rezultatai gauti naudojant RF modelį. Galutiniai rezultatai parodė, kad tiesioginio prognozavimo strategija visose trijose elektrinėse buvo tikslesnė už dviejų pakopų strategiją. Tiesioginio prognozavimo strategijoje XGBoost pasižymėjo mažiausia rezultatų sklaida ($\sigma = 1,56$) pagal DWMAPE tarp visų elektrinių. Dviejų pakopų prognozavimo strategijos pirmoje pakopoje mažiausia rezultatų sklaida pasižymėjo SVM modelis ($\sigma = 1,66$), o antroje pakopoje – XGBoost modelis ($\sigma = 1,12$).

Raktiniai žodžiai: PV energijos prognozavimas, saulės elektrinės, prognozavimas, mašininis mokymasis, SVM, RF, XGBoost, LSTM, TFT.

1 Įvadas

Saulės energija yra vienas sparčiausiai augančių atsinaujinančios energijos šaltinių. Technologijų pažanga ir mažesnės eksploatacijos sąnaudos lėmė platų taikymą tiek buitiniuose, tiek pramoniniuose sprendimuose. Viena pagrindinių saulės energetikos plėtros priežasčių yra galimybė mažinti klimato kaitą, nes, palyginti su iškastiniu kuru, sumažinamos šiltnamio efektą sukeliančių dujų emisijos [1]. Saulės elektrinių energijos išgavimo prognozavimas yra svarbus dėl kelių priežasčių, pirmiausia dėl gamybos kintamumo valdymo ir elektros tinklo stabilumo palaikymo. Kadangi saulės generacija priklauso nuo oro sąlygų ir kitų veiksnių, galimi staigūs gamybos kritimai. Tikslus prognozavimas padeda mažinti rizikas, susijusias su šiais svyravimais [2]. Taip pat jis leidžia operatoriams efektyviau planuoti ir optimizuoti veiklą, suderinant energijos pasiūlą su rinkos paklausa [3],[4]. Iššūkis yra tas, kad saulės generaciją stipriai veikia orai ir sezonai. Šiaurinėse šalyse (pvz., Baltijos ir Šiaurės šalyse) prognozavimą papildomai apsunkina trumpos dienos žiemos sezonu, žemas saulės aukštis, greitai judantys debesys, sniegas ir apledėjimas, taip pat dideli sezoniniai svyravimai. Saulės elektrinių galia sparčiai didėja, o atsinaujinančių išteklių dalis turėtų pasiekti 40 % iki 2030 m. ir 85 % iki 2050 m. Didėjant saulės generacijos daliai, kasdienis ir valandinis prognozavimas tampa reikšmingesnis, todėl tikslūs ir praktiškai pritaikomi prognozavimo įrankiai laikomi būtinais [5].

Baltijos šalių kontekste fotovoltinių elektrinių (PV) prognozėse dažniausiai naudoti rekurentinio tipo neuroniniai tinklai (RNN), taip pat visiškai sujungti neuroniniai tinklai (FCNN) ir konvoliuciniai neuroniniai tinklai (CNN), derinant su meteorologiniais požymiais ir laiko kintamaisiais [6],[7],[8],[9]. Greta giliųjų tinklų praktikoje dažnai taikomi ir klasikiniai mašininio mokymosi metodai, kurie naudojami kaip baziniai modeliai arba kaip greitesnės alternatyvos, ypač kai svarbūs diegimo ir permokymo kaštai. Tokiuose palyginimuose dažniausiai sutinkami atsitiktinių miškų (RF), ekstremalaus gradientinio stiprinimo (XGBoost), atraminių vektorių regresijos (SVR), kNN, tiesinės regresijos ar GAM modeliai [10],[11]. Pastaraisiais metais dažniau taikomi dėmesio mechanizmu paremti metodai, tarp jų laikinis suliejimo transformeris (TFT), kurie skirti daugiahorizontėms prognozėms ir leidžia įvertinti požymių svarbą. PV galios prognozavimo uždavinyje TFT pritaikytas valandinei dienos į priekį prognozei ir palygintas su keliais baziniais modeliais, pateikiant mažesnes paklaidas [12].

Praktikoje valandinis dienos į priekį prognozavimas gali būti formuluojamas skirtingomis metodologijomis. Vienas iš galimų darbo srautų yra

paprastas ir plačiai naudojamas laiko eilučių prognozavimo uždaviniuose. Mokymo dalyje modeliai išmoksta ryšius tarp meteorologinių ir kalendorinių požymių bei saulės elektrinės pagamintos energijos. Po mokymo modeliai taikomi testavimo aibei, kurioje prognozuojama valandos energijos gamyba. Alternatyvus darbo šrautas remiasi dviejų pakopų prognozavimo principu. Šios strategijos tikslas yra atskirti bendrą dienos energijos kiekio prognozę nuo energijos pasiskirstymo dienos viduje. Tokia struktūra leidžia modeliams lengviau išmokti skirtingo pobūdžio priklausomybes: pirmasis modelis orientuojasi į bendrą dienos energijos gamybą, o antrasis – į dienos valandinį profilį [13].

Šiame darbe lyginamos dvi prognozavimo strategijos: tiesioginis valandinis prognozavimas ir dviejų pakopų prognozavimas, kuriame atskirai modeliuojama dienos energijos suma ir jos pasiskirstymas valandomis. Prognozėms sudaryti taikomi keli modeliai: ekstremalaus gradientinio stiprinimo modelis (XGBoost), ilgos trumpalaikės atminties neuroninis tinklas (LSTM) ir laikinis suliejimo transformeris (TFT). Papildomai vertinami atraminių vektorių (SVM) su RBF branduoliu ir atsitiktinių miškų (RF) modeliai, siekiant įvertinti klasikinių mašininio mokymosi modelių tinkamumą ir palyginti jų rezultatus su sudėtingesnėmis architektūromis.

2 Duomenys

Darbe naudojami trys duomenų rinkiniai, sudaryti remiantis trijų prie elektros tinklo prijungtų saulės fotovoltinių elektrinių, esančių skirtinguose Lietuvos regionuose, duomenimis. Juose pateikiami valandiniai elektrinių veikimo ir meteorologinių sąlygų matavimai. Pirmoji elektrinė yra šiaurės Lietuvoje, jos duomenys apima laikotarpį nuo 2024 m. birželio 1 d. iki 2025 m. spalio 5 d. Šioje elektrinėje įrengti 7 inverteriai, o maksimali elektrinės galia siekia 300 kWp. Antroji elektrinė yra centrinėje Lietuvos dalyje, jos duomenys apima laikotarpį nuo 2025 m. sausio 10 d. iki 2025 m. spalio 5 d. Šioje elektrinėje įrengta 15 inverterių, o maksimali elektrinės galia siekia 3200 kWp. Trečioji elektrinė yra pietų Lietuvoje, jos duomenys apima laikotarpį nuo 2024 m. birželio 1 d. iki 2025 m. spalio 5 d. Šioje elektrinėje įrengti 8 inverteriai, o maksimali elektrinės galia siekia 500 kWp. Visuose trijuose duomenų rinkiniuose pateikiami identiški kintamieji: laiko žyma UTC ISO 8601 formatu, tiesioginė normalioji spinduliuotė (DNI), išsklaidytoji horizontalioji spinduliuotė (DHI), globalioji horizontalioji spinduliuotė (GHI), globalioji į modelio plokštumą pakrypusi spinduliuotė, apskaičiuota

taikant Liu–Jordan izotropinį apšvitos modelį pagal DNI, DHI, GHI reikšmes, aplinkos oro temperatūra Kelvino skalėje, santykinė oro drėgmė procentais, vėjo greitis metrais per sekundę, debesuotumas procentais, meteorologinės būklės kodas ir per valandos intervalą pagamintos PV energijos kiekis, išreikštas kilovatvalandėmis.

Požymių inžinerija buvo grindžiama saulės PV sistemų fizikiniais veikimo principais ir laiko eilučių analizės praktika [14],[15]. Kadangi PV energijos gamyba pasižymi ryškiais paros ir sezoniniais ciklais, priklauso nuo saulės apšvitos intensyvumo ir jautriai reaguoja į trumpalaikius debesuotumo pokyčius, požymių rinkinys sudarytas siekiant aiškiai užkoduoti šiuos dėsninumus. Taikyti cikliniai laiko požymiai, apibūdinantys paros valandą ir metų dieną, kalendoriniai kintamieji, nusakantys savaitės ir mėnesio struktūrą, bei dienos šviesos indikatorius. Apšvitos režimui apibūdinti naudoti pagrindiniai meteorologiniai kintamieji ir iš jų sudaryti santykiniai rodikliai, tokie kaip tiesioginės spinduliuotės dalis ir pakrypusios į plokštumą ir horizontalios apšvitos santykis. Taip pat įtraukti trumpalaikius pokyčius aprašantys pirmieji skirtumai, įvairių horizontų vėlavimai ir slenkantieji vidurkiai, leidžiantys modeliams įvertinti trumpalaikę meteorologinių sąlygų ir energijos gamybos atmintį. Papildomai buvo skaičiuojami slenkantieji apšvitos kintamumo rodikliai, paprastos sąveikos tarp apšvitos ir meteorologinių veiksnių bei kai kurie netiesiniai nariai, skirti geriau atspindėti temperatūrinius ir atmosferinius poveikius energijos išgavimui. Diskretūs meteorologinių būklių kodai buvo transformuoti į indikatorinius požymius, kad modeliai galėtų atskirti skirtingus orų režimus.

3 Metodologija

Šiame darbe taikomos dvi prognozavimo strategijos: tiesioginis prognozavimas ir dviejų pakopų strategija. Abiejose strategijose turimas valandinis duomenų rinkinys suskirstomas į dvi dalis chronologine tvarka: pirmieji 80 % stebėjimų naudojami modelio mokymui, o likę 20 % – testavimui.

Tiesioginio prognozavimo strategijoje valandinis dienos į priekį prognozavimo uždavinys sprendžiamas vienu modeliu, kuris tiesiogiai prognozuoja kiekvienos valandos energijos gamybą. Šiuo atveju naudojamas pradinis valandinis duomenų rinkinys, kuriame kiekvienas stebėjimas atitinka atskirą valandą, o prognozuojamas tos valandos pagamintos energijos kiekis. Mokymo metu modelis nustato priklausomybes tarp įvesties požymių ir valandinės energijos gamybos reikšmių, o testavimo metu, remiantis tais

pačiais požymiais, sudaromos valandinės energijos prognozės. Tokia strategija leidžia tiesiogiai įvertinti, kaip tiksliai modelis atkuria valandinį energijos gamybos profilį. Gautos prognozės vėliau lyginamos su tikromis valandinėmis reikšmėmis, o modelio tikslumas vertinamas pasirinktomis paklaidos metrikomis.

Dviejų pakopų strategijos pirmoje pakopoje valandiniai požymiai yra agreguojami į dieninius požymius. Energijos valandinės reikšmės yra sumuojamos pagal dienas, taip pat sumuojami ir spinduliuotės požymiai. Meteorologiniai požymiai, tokie kaip temperatūra, debesuotumas, vėjo greitis ir drėgmė yra vidurkinami. Taip pat įtraukiami kalendoriniai požymiai, apibūdinantys sezoninius ir meteorologinius santykius. Požymių vėlavimai, sąveikos ir kvadratinės reikšmės perskaičiuojami pagal dienas. Naudojant šiuos dieninius požymius pirmoje pakopoje apmokomas modelis, kuris prognozuoja bendrą dienos energijos kiekį. Antroje pakopoje naudojamas valandinis duomenų rinkinys, tačiau prognozuojama energijos reikšmė transformuojama. Vietoje valandinių energijos reikšmių prognozuojamos valandinės energijos proporcijos nuo bendros tos dienos energijos. Kitaip tariant, kiekvienos valandos energija padalijama iš tos dienos bendros gamybos ir gaunamos proporcijos. Taigi, apmokomas antras modelis, kuris prognozuoja valandines dienos energijos proporcijas. Galiausiai, valandinė energijos prognozė apskaičiuojama padauginus pirmojo modelio prognozuotą dienos energiją iš antrojo modelio prognozuotos valandinės proporcijos. Taip gaunama kiekvienos valandos energijos prognozė, kuri lyginama su tikromis reikšmėmis.

Modelių tikslumas vertinamas naudojant kelias dažniausiai energijos prognozavimo uždaviniams literatūroje taikomas paklaidos metrikas, t. y. naudojamos vidutinės kvadratinės paklaidos šaknis (RMSE), vidutinė absoliuti paklaida (MAE) ir svorinė vidutinė absoliuti procentinė paklaida (WMAPE), kuri skaičiuojama valandinei ir dieninei energijai [16],[17]. WMAPE dažniau taikomas duomenims, kuriuose pasitaiko reikšmių arti nulio, nes tokiais atvejais standartinio MAPE vardiklis tampa labai mažas ir santykinė paklaida dirbtinai išauga. Dėl to paprastas MAPE gali iškreipti bendrą modelio vertinimą, kadangi net nedideli absoliutūs nuokrypiai prie mažų tikrųjų reikšmių virsta neproporcingai didelėmis procentinėmis paklaidomis. Modelių prognozavimo stabilumui vertinti buvo apskaičiuotas dieninės paklaidos (DWMAPE) imties standartinis nuokrypis (σ) tarp trijų elektrinių, t. y. kuo mažesnis paklaidų standartinis nuokrypis tarp elektrinių, tuo laikoma, kad

modelio prognozavimo tikslumas yra stabilesnis ir mažiau priklausomas nuo duomenų rinkinio ar geografinės vietos savybių [18].

Modelių hiperparametrų optimizavimas atliekamas naudojant laiko eilučių kryžminės validacijos metodą su slenkančiu mokymo langu. Duomenys yra padalijami į 5 nuoseklius segmentus, kuriuose kiekvienoje iteracijoje modelis apmokomas su ankstesniais duomenimis, o jo veikimas vertinamas prognozuojant tolimesnį laikotarpį. Vertinimas vyksta pagal valandinio nuokrypio metriką. Taip gaunami tikslumo rezultatai kiekvienam segmentui pagal valandinį nuokrypį ir gauti rezultatai yra vidurkinami, taip įvertinant bendrą parametro rinkinio efektyvumą. Galiausiai parenkamas tas hiperparametrų rinkinys, kuris pasižymi mažiausia vidutine prognozavimo paklaida [19]. TFT modelio hiperparametrų optimizavimui buvo naudojama Optuna biblioteka, kadangi tai yra sudėtingesnis modelis, palyginti su kitais šiame tyrime naudojamais modeliais, kuris pasižymi dideliu architektūrinių ir mokymosi parametrų skaičiumi, kurių tarpusavio sąveika gali stipriai paveikti prognozavimo tikslumą. Naudoti kryžminės validacijos metodą su slenkančiu mokymo langu TFT modeliui būtų neefektyvu laiko atžvilgiu dėl didelio hiperparametrų skaičiaus. Optuna biblioteka leidžia automatiškai ieškoti optimalios kombinacijos nustatytuose hiperparametrų intervaluose [20].

Hiperparametrų optimizavimo metu kiekvienam modeliui buvo apibrėžtos parametrų gardelės, skirtos sisteminei geriausio parametro rinkinio paieškai. SVM: *C*: [1, 10, 100], *gamma*: ["scale", 0,01, 0,05, 0,1, 1], *epsilon*: [0,1, 1, 2, 5, 10]. RF: *n_estimators*: [200, 500, 800], *max_features*: [2, 3, 4, 5, 10, 20], *max_depth*: [None, 10, 20, 30, 50]. XGBoost: *eta*: [0,05, 0,1, 0,2], *max_depth*: [4, 6, 8], *subsample*: [0,5, 0,6, 0,7, 0,8, 0,9], *colsample_bytree*: [0,5, 0,6, 0,7, 0,8, 0,9], *num_boost_round*: [100, 150, 200]. LSTM: *hidden_size*: [16, 32, 48], *num_layers*: [1, 2, 3], *dropout*: [0, 0,2], *lr*: [0,001, 0,003], *epochs*: [20, 30]. TFT: *max_encoder_length*: [168, 336, 504], *hidden_size*: [64, 128], *hidden_continuous_size*: [32, 64], *attention_head_size*: [4, 8], *lstm_layers*: [1, 2], *dropout*: [0,05–0,40], *learning_rate*: [0,00005–0,01], *gradient_clip_val*: [0,05, 0,1, 0,2], *batch_size*: [64, 128].

4 Eksperimentai ir rezultatai

Po hiperparametrų optimizavimo modeliams buvo parinkti šie parametrai tiesioginio prognozavimo strategijai. SVM: *C* = 1, *gamma* = "scale", *epsilon* = 5. RF: *n_estimators* = 200, *max_features* = 20, *max_depth* = 30. XGBoost: *eta* = 0,1, *max_depth* = 6, *subsample* = 0,5, *colsample_by_tree* = 0,5, *num_boost_round* =

200. LSTM: *hidden_size* = 48, *num_layers* = 3, *epochs* = 30, *lr* = 0.003, *dropout* = 0. TFT: *max_encoder_length* = 504, *hidden_size* = 128, *hidden_continuous_size* = 64, *attention_head_size* = 4, *lstm_layers* = 2, *dropout* = 0,178, *learning_rate* = 0,000281, *gradient_clip_val* = 0,05, *batch_size* = 64. Dviejų pakopų strategijai hiperparametrai buvo parinkti tokiu pačiu principu, naudojant laiko eilučių kryžminės validacijos metodą su slenkančiu mokymo langu bei optimizavimą naudojant Optuna biblioteką TFT modeliui.

Tiesioginio prognozavimo strategijoje vidutinė modelių apmokymo trukmė skyrėsi ryškiai: XGBoost modelio apmokymas truko 2,81 sekundės, SVM 5,54 sekundės, RF 20,13 sekundės, LSTM 106,89 sekundės, o TFT apie 220 minučių. Šie laiko matavimai bei modelių rezultatai buvo gauti ant kompiuterio su Intel Core i5-6300U procesoriumi (2,40 GHz, 2 branduoliai, 4 loginiai procesoriai) ir 16 GB operatyviosios atminties, o galutiniai TFT rezultatai, kurie yra lentelėse, buvo gauti naudojant superkompiuterį su NVIDIA Tesla V100-SXM2-32GB grafiniu procesoriumi, kuris buvo naudojamas siekiant pagreitinėti skaičiavimus šiam modeliui. Kadangi RMSE ir MAE reikšmės priklauso nuo elektrinės generacijos masto, šios metrikos tinkamos modeliams lyginti tik tos pačios elektrinės ribose, o skirtingų elektrinių palyginimui svarbesnės yra valandos svorinė vidutinė absoliuti procentinė paklaida (HWMape) ir dienos svorinė vidutinė absoliuti procentinė paklaida (DWMape).

Tiesioginio prognozavimo strategijoje geriausi dienos paklaidos rezultatai šiaurės ir centrinėje elektrinėse gauti naudojant XGBoost modelį, kur DWMape siekė atitinkamai 13,24 % ir 16,25 %. Pietinėje elektrinėje geriausią dienos paklaidos rezultatą pasiekė TFT modelis, kurio DWMape sudarė 12,68 %. Mažiausia valandinė paklaida tarp visų tiesioginio prognozavimo rezultatų nustatyta pietinėje elektrinėje naudojant RF modelį, kur HWMape siekė 24,13 %. Šie rezultatai rodo, kad tiesioginio prognozavimo strategijoje geriausius rezultatus pasiekė XGBoost, RF ir TFT modeliai, o LSTM daugeliu atvejų nusileido kitiems modeliams. Gauti imties standartinio nuokrypio rezultatai parodė skirtingą modelių jautrumą duomenų rinkiniams: SVM ($\sigma = 1,89$), RF ($\sigma = 2,04$), XGBoost ($\sigma = 1,56$), LSTM ($\sigma = 5,17$) bei TFT ($\sigma = 5,19$). Mažiausia rezultatų sklaida tiesioginio prognozavimo strategijoje pasižymėjo XGBoost modelis, o tai rodo didžiausią šio modelio prognozių stabilumą nepriklausomai nuo geografinės vietos. Tuo tarpu giliųjų neuroninių tinklų modeliai (LSTM ir TFT) pasižymėjo didžiausiu jautrumu duomenų rinkiniams, kas indikuoja mažesnį jų stabilumą lyginant su klasikiniiais mašininio mokymosi modeliais.

1 lentelė. Prognozavimo rezultatai naudojant tiesioginio prognozavimo strategiją. Čia HWMape – valandos svorinis MAPE, DWMape – dienos svorinis MAPE.

Elektrinė	SVM	RF	XGBoost	LSTM	TFT
Šiaurės	RMSE: 34,65 MAE: 24,1 HWMape: 47,95 % DWMape: 18,23 %	RMSE: 33,09 MAE: 22,97 HWMape: 45,7 % DWMape: 14,57 %	RMSE: 33,07 MAE: 22,52 HWMape: 44,8 % DWMape: 13,24 %	RMSE: 35,61 MAE: 22,59 HWMape: 44,94 % DWMape: 23,31 %	RMSE: 33,14 MAE: 21,68 HWMape: 43,21 % DWMape: 15,48 %
Centrinė	RMSE: 251,1 MAE: 141,41 HWMape: 31,37 % DWMape: 18,77 %	RMSE: 232,13 MAE: 122,48 HWMape: 27,17 % DWMape: 17,97 %	RMSE: 230,06 MAE: 125,52 HWMape: 27,84 % DWMape: 16,25 %	RMSE: 302,86 MAE: 166,3 HWMape: 36,89 % DWMape: 22,28 %	RMSE: 280,19 MAE: 141,11 HWMape: 31,36 % DWMape: 22,73 %
Pietinė	RMSE: 44,57 MAE: 24,9 HWMape: 33,4 % DWMape: 21,74 %	RMSE: 33,96 MAE: 17,99 HWMape: 24,13 % DWMape: 14,33 %	RMSE: 34,86 MAE: 18,53 HWMape: 24,85 % DWMape: 14,04 %	RMSE: 40,75 MAE: 22,42 HWMape: 30,06 % DWMape: 13,88 %	RMSE: 36,42 MAE: 18,34 HWMape: 24,58 % DWMape: 12,68 %

2 lentelė. Prognozavimo rezultatai dviejų pakopų prognozavimo strategijos pirmoje pakopoje. Čia DWMape – dienos svorinis MAPE.

Elektrinė	SVM	RF	XGBoost	LSTM	TFT
Šiaurės	RMSE: 246,67 MAE: 202,35 DWMape: 16,81 %	RMSE: 243,24 MAE: 192,25 DWMape: 15,97 %	RMSE: 250,25 MAE: 205,15 DWMape: 17,05 %	RMSE: 417,5 MAE: 340,42 DWMape: 28,28 %	RMSE: 237,03 MAE: 193,75 DWMape: 16,1 %
Centrinė	RMSE: 2568,15 MAE: 2120,27 DWMape: 19,85 %	RMSE: 3094,65 MAE: 2354,13 DWMape: 22,04 %	RMSE: 3052,93 MAE: 2205,72 DWMape: 20,65 %	RMSE: 5684,15 MAE: 4524,19 DWMape: 42,35 %	RMSE: 2940,97 MAE: 2417,64 DWMape: 22,63 %
Pietinė	RMSE: 383,66 MAE: 309,9 DWMape: 17,19 %	RMSE: 388,01 MAE: 312,54 DWMape: 17,34 %	RMSE: 403,8 MAE: 321,81 DWMape: 17,86 %	RMSE: 735,94 MAE: 588,4 DWMape: 32,65 %	RMSE: 374,59 MAE: 297,77 DWMape: 16,52 %

Dviejų pakopų strategijos pirmoje pakopoje, prognozuojant bendrą dienos energijos kiekį, geriausi rezultatai gauti naudojant RF modelį šiaurės elektrinėje, kur DWMape siekė 15,97 %, SVM modelį centrinėje elektrinėje, kur DWMape siekė 19,85 %, ir TFT modelį pietinėje elektrinėje, kur DWMape siekė 16,52 %. Vertinant pirmosios pakopos prognozių stabilumą, nustatyti

3 lentelė. Prognozavimo rezultatai dviejų pakopų prognozavimo strategijos antroje pakopoje. Čia HWMape – valandos svorinis MAPE, DWMape – dienos svorinis MAPE.

Elektrinė	SVM	RF	XGBoost	LSTM	TFT
Šiaurės	RMSE: 0,03 MAE: 0,02 HWMape: 45,55 % DWMape: 9,67 %	RMSE: 0,03 MAE: 0,02 HWMape: 44,8 % DWMape: 3,36 %	RMSE: 0,03 MAE: 0,02 HWMape: 43,5 % DWMape: 4,66 %	RMSE: 0,03 MAE: 0,02 HWMape: 42,9 % DWMape: 7,61 %	RMSE: 0,03 MAE: 0,02 HWMape: 38,76 % DWMape: 5,24 %
Centrinė	RMSE: 0,02 MAE: 0,02 HWMape: 37,05 % DWMape: 14,75 %	RMSE: 0,02 MAE: 0,01 HWMape: 24,25 % DWMape: 6,33 %	RMSE: 0,02 MAE: 0,01 HWMape: 24,72 % DWMape: 6,57 %	RMSE: 0,02 MAE: 0,01 HWMape: 24,72 % DWMape: 8,09 %	RMSE: 0,02 MAE: 0,01 HWMape: 24,86 % DWMape: 9,12 %
Pietinė	RMSE: 0,02 MAE: 0,01 HWMape: 33,82 % DWMape: 10,3 %	RMSE: 0,02 MAE: 0,001 HWMape: 21,89 % DWMape: 3,39 %	RMSE: 0,02 MAE: 0,001 HWMape: 22,44 % DWMape: 4,59 %	RMSE: 0,02 MAE: 0,01 HWMape: 29,69 % DWMape: 9,92 %	RMSE: 0,02 MAE: 0,01 HWMape: 25,77 % DWMape: 6,52 %

šie paklaidų imties standartinio nuokrypio rodikliai: SVM ($\sigma = 1,66$), RF ($\sigma = 3,18$), XGBoost ($\sigma = 1,89$), LSTM ($\sigma = 7,20$) bei TFT ($\sigma = 3,65$). Mažiausia rezultatų sklaida dviejų pakopų prognozavimo strategijos pirmoje pakopoje pasižymėjo SVM modelis, o didžiausia – LSTM modelis. Tai reiškia, kad SVM modelis stabiliai prognozuoja dienos energijos kiekius. Antroje pakopoje, prognozuojant valandines energijos proporcijas, visose elektrinėse geriausi dieninės paklaidos rezultatai gauti naudojant RF modelį, kur DWMape siekė 3,36 %, 6,33 % ir 3,39 %. Vertinant antrosios pakopos prognozių stabilumą, nustatyti šie paklaidų imties standartinio nuokrypio rodikliai: SVM ($\sigma = 2,77$), RF ($\sigma = 1,71$), XGBoost ($\sigma = 1,12$), LSTM ($\sigma = 1,22$) bei TFT ($\sigma = 1,98$). Mažiausia rezultatų sklaida dviejų pakopų prognozavimo strategijos antroje pakopoje pasižymėjo XGBoost modelis, o didžiausia – SVM modelis. Tai reiškia, kad XGBoost modelis stabiliai prognozuoja valandines dienos energijos proporcijas. Sujungus geriausius abiejų pakopų modelius, gautos šios galutinės dviejų pakopų strategijos kombinacijos: šiaurės elektrinėje – RF ir RF, centrinėje – SVM ir RF, pietinėje – TFT ir RF. Jų galutiniai DWMape rezultatai atitinkamai sudarė 15,59 %, 20,42 % ir 15,90 %, todėl visose trijose elektrinėse dviejų pakopų strategija nusileido tiesioginiam prognozavimui.

5 Išvados

Šiame darbe buvo išanalizuotas valandinis saulės fotovoltinių elektrinių energijos prognozavimas Lietuvoje, lyginant dvi skirtingas strategijas – tiesioginį valandinį prognozavimą ir dviejų pakopų prognozavimo strategiją. Tyrime taikyti įvairių grupių modeliai – nuo klasikinių mašininio mokymosi iki giliųjų neuroninių tinklų ir transformerių architektūrų.

Tyrimo rezultatai parodė, kad tiesioginio prognozavimo strategija visose trijose elektrinėse buvo tikslesnė už dviejų pakopų strategiją. Geriausi galutiniai tiesioginio prognozavimo rezultatai gauti šiaurės elektrinėje naudojant XGBoost modelį, kur DWMAPE siekė 13,24 %, centrinėje elektrinėje taip pat naudojant XGBoost modelį, kur DWMAPE siekė 16,25 %, o pietinėje elektrinėje naudojant TFT modelį, kur DWMAPE siekė 12,68 %.

Tiesioginio prognozavimo strategijoje XGBoost pasižymėjo stabiliausiais ($\sigma = 1,56$) rezultatais tarp visų elektrinių, o TFT parodė, kad tam tikrais atvejais jis gali pranokti klasikinius modelius. Mažiausia valandinė paklaida tiesioginio prognozavimo atveju nustatyta pietinėje elektrinėje naudojant RF modelį, kur HWMape siekė 24,13 %. LSTM modelis daugeliu atvejų nusileido kitiems modeliams.

Dviejų pakopų prognozavimo strategijos pirmoje pakopoje dienos energijos prognozėms geriausi modeliai priklauso nuo elektrinės: šiaurės elektrinėje tiksliausias buvo RF modelis, centrinėje – SVM, o pietinėje – TFT; jų DWMAPE atitinkamai siekė 15,97 %, 19,85 % ir 16,52 %. Antroje pakopoje valandinių proporcijų prognozėse geriausi rezultatai visose elektrinėse gauti naudojant RF modelį, kurio DWMAPE šiaurės, centrinėje ir pietinėje elektrinėse sudarė atitinkamai 3,36 %, 6,33 % ir 3,39 %. Mažiausia rezultatų sklaida dviejų pakopų prognozavimo strategijos pirmoje pakopoje pasižymėjo SVM modelis ($\sigma = 1,66$), o antroje pakopoje – XGBoost modelis ($\sigma = 1,12$). Vis dėlto net ir parinkus geriausias modelių kombinacijas, galutinis dviejų pakopų strategijos tikslumas išliko mažesnis nei tiesioginio prognozavimo.

Rezultatai rodo, kad klasikiniai mašininio mokymosi modeliai išliko tinkami ir aktualūs valandinės PV energijos prognozavimo uždavinyje bei daugeliu atvejų pasiekė panašų arba geresnį tikslumą nei sudėtingesni modeliai. Tuo pačiu TFT rezultatai parodė, kad sudėtingesnės architektūros taip pat gali veikti efektyviai, tačiau jų taikymas reikalauja gerokai didesnių skaičiavimo sąnaudų. Be to, nustatyta, kad prognozavimo tikslumas priklauso ir nuo elektrinės geografinės vietos: mažiausios dieninės paklaidos buvo pasiektos šiaurės ir pietų Lietuvoje esančiose elektrinėse, o didžiausios buvo pasiek-

tos centrinėje Lietuvos dalyje esančioje elektrinėje. Šie skirtumai tikėtina yra susiję tiek su duomenų rinkinių savybėmis, tiek su vietovės ypatumais. Centrinės elektrinės atveju mažesnę tikslumą galėjo lemti trumpesnis stebėjimo laikotarpis ir dėl to mažesnė sezoninė aprėptis. Papildomą įtaką galėjo turėti techninės elektrinių charakteristikos. Nors Lietuvoje reljefo skirtumai nėra dideli, tam tikrą įtaką prognozavimo tikslumui galėjo turėti su reljefu susiję mikroklimato ir vietinių meteorologinių sąlygų skirtumai.

Tolimesniuose tyrimuose tikslinga išplėsti duomenų imtį, analizuoti atskirus sezonus ir vertinti modelių veikimą ilgesniuose prognozavimo horizontuose.

Literatūra

- [1] Karatayev, M., & Clarke, M. L. (2016). A review of current energy systems and green energy potential in Kazakhstan. *Renewable and Sustainable Energy Reviews*, 55, 491–504.
- [2] Zhu, T., Zhou, H., Wei, H., Zhao, X., Zhang, K., & Zhang, J. (2019). Inter-hour direct normal irradiance forecast with multiple data types and time-series. *Journal of Modern Power Systems and Clean Energy*, 7(5), 1319–1327.
- [3] Wang, F., Zhen, Z., Liu, C., Mi, Z., Shafie-khah, M., & Catalão, J. P. (2018). Time-section fusion pattern classification based day-ahead solar irradiance ensemble forecasting model using mutual iterative optimization. *Energies*, 11(1), 184.
- [4] Sun, Y., Wang, F., Wang, B., Chen, Q., Engerer, N. A., & Mi, Z. (2016). Correlation feature selection and mutual information theory based quantitative research on meteorological impact factors of module temperature for solar photovoltaic systems. *Energies*, 10(1), 7.
- [5] Gecevičius, G. (2021). Development of renewable energy in Lithuania: Experience, state and trends. *Environmental Research, Engineering and Management*.
- [6] Mohammadi Lanbaran, N., Naujokaitis, D., Kairaitis, G., & Radziukynas, V. (2025). Hybrid Hourly Solar Energy Forecasting Using BiLSTM Networks with Attention Mechanism, General Type-2 Fuzzy Logic Approach: A Comparative Study of Seasonal Variability in Lithuania. *Applied Sciences*, 15(17), 9672.
- [7] Narkutė, V. (2022). Forecast of energy produced by solar power plants (Doctoral dissertation, Vilniaus universitetas.).
- [8] Shabbir, N., Kütt, L., Raja, H. A., Jawad, M., Allik, A., & Husev, O. (2022). Techno-economic analysis and energy forecasting study of domestic and commercial photovoltaic system installations in Estonia. *Energy*, 253, 124156.
- [9] Nikulins, A., Sudars, K., Edelmers, E., Namatevs, I., Ozols, K., Komasilovs, V., ... & Reinhardt, A. (2024). Deep learning for wind and solar energy forecasting in hydrogen production. *Energies*, 17(5), 1053.
- [10] Ledmaoui, Y., El Maghraoui, A., El Aroussi, M., Saadane, R., Chebak, A., & Chehri, A. (2023). Forecasting solar energy production: A comparative study of machine learning algorithms. *Energy Reports*, 10, 1004–1012.
- [11] Obiora, C. N., Ali, A., & Hasan, A. N. (2021, August). Implementing extreme gradient boosting (XGBoost) algorithm in predicting solar irradiance. In 2021 IEEE PES/IAS PowerAfrica (pp. 1–5). IEEE.

- [12] López Santos, M., García-Santiago, X., Echevarría Camarero, F., Blázquez Gil, G., & Carrasco Ortega, P. (2022). Application of temporal fusion transformer for day-ahead PV power forecasting. *Energies*, 15(14), 5232.
- [13] Alanazi, M., Mahoor, M., & Khodaei, A. (2017, September). Two-stage hybrid day-ahead solar forecasting. In 2017 North American Power Symposium (NAPS) (pp. 1–6). IEEE.
- [14] Huld, T., & Gracia Amillo, A. M. (2015). Estimating PV module performance over large geographical regions: The role of irradiance, air temperature, wind speed and solar spectrum. *Energies*, 8(6), 5159–5181.
- [15] De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American statistical association*, 106(496), 1513–1527.
- [16] Barhmi, K., Heynen, C., Golroodbari, S., & Van Sark, W. A. (2024). A review of solar forecasting techniques and the role of artificial intelligence. *Solar*, 4, 99–135.
- [17] Chen, S., Wan, H., Peng, B., Quan, R., Chang, Y., & Derigent, W. (2026). Accurate multi-step wind and solar power forecasting based on multi-scale convolutional Kolmogorov-Arnold network and improved Lemming-optimized attention fusion. *Engineering Applications of Artificial Intelligence*, 163, 112832.
- [18] Wang, J. C., Chang, C. T., & Ke, K. C. (2026). Data-driven representation learning and stability validation for multi-quality prediction in smart injection molding. *The International Journal of Advanced Manufacturing Technology*, 1–15.
- [19] Cerqueira, V., Torgo, L., & Mozetič, I. (2020). Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning*, 109(11), 1997–2028.
- [20] da Silva, E. C., Finardi, E. C., & Stefenon, S. F. (2024). Enhancing hydroelectric inflow prediction in the Brazilian power system: A comparative analysis of machine learning models and hyperparameter optimization for decision support. *Electric Power Systems Research*, 230, 110275.