



UNIVERZITET U BANJOJ LUCI
ELEKTROTEHNIČKI FAKULTET



Primjena neuronskih mreža za generisanje sintetičkih podataka o solarnoj iradijansi

MASTER RAD

Mentor:
prof. dr. Vladimir Risojević

Kandidat:
Romanela Lajić

Banja Luka, 2024.



UNIVERSITY OF BANJA LUKA
FACULTY OF ELECTRICAL ENGINEERING



Generating synthetic solar irradiance data using neural networks

MASTER THESIS

Mentor:
prof. dr. Vladimir Risojević

Candidate:
Romanela Lajić

Banja Luka, 2024.

Informacije o mentoru i disertaciji

Mentor:

Dr. Vladimir Risojević, vanredni profesor
Univerzitet u Banjoj Luci, Elektrotehnički fakultet

Naslov master rada:

Primjena neuronskih mreža za generisanje sintetičkih podataka o solarnoj iradijansi

Rezime:

U posljednjih nekoliko decenija došlo je do rapidnog porasta udjela obnovljivih izvora energije u elektroenergetskim sistemima, pri čemu solarna energija prednjači kada je u pitanju instalisana snaga. Snaga nekog solarnog sistema direktno zavisi od solarne iradijanse, čija je vrijednost stohastičke prirode. Povećanjem udjela solarne energije, povećava se potreba za nekom vrstom planiranja i kratkoročne predikcije rada ovakvih sistema. Planiranje rada solarne elektrane zahtijeva veliku količinu podataka o solarnoj iradijansi za duži vremenski period, dok kratkoročna predikcija zahtijeva obučavanje modela za predikcijom, takođe korišćenjem veće količine podataka o solarnoj iradijansi. S obzirom na to da je mjerenje solarne iradijanse na nekom području dugačak i skup proces, u ovom radu razmatrana je mogućnost generisanja sintetičkih podataka o solarnoj iradijansi uz pomoć neuronskih mreža. Pokazano je da je određenim tipovima neuronskih mreža moguće generisati sintetičke podatke koji su u stanju da realistično modeluju rad solarne elektrane, i koji se mogu upotrijebiti za obučavanje modela kratkoročne predikcije solarne iradijanse. Upotrebljivost ovakvih podataka testirana je na modelu solarnog sistema, kao i na različitim tipovima kratkoročnih prediktora.

Ključne riječi: solarna iradijansa, predikcija solarne iradijanse, generisanje vremenskih podataka, generativni modeli, generativne suprostavljajuće mreže

Naučna oblast: Inženjerstvo i tehnologija

Naučno polje: Elektrotehnika, elektronika i informacione tehnologije

Klasifikaciona oznaka: T 121

Tip odabrane licence Kreativne zajednice: CC BY-NC

Information about mentor and thesis

Mentor:

Dr. Vladimir Risojević, professor
University of Banja Luka, Faculty of Electrical Engineering

Title of master thesis:

Generating synthetic solar irradiance data using neural networks

Abstract:

In the last few decades there has been a rapid growth of renewable energy sources in electric power systems, with solar energy as the most common source. Power of a solar system depends directly on the value of solar irradiance, which is of a stochastic nature. With the increase of the number of solar systems, there is a need for a certain type of both long-term planning and short-term prediction of the behaviour of these systems. Long-term planning requires a large amount of solar irradiance data taken during a longer time span, while short-term prediction models need to be trained, also using a large amount of irradiance data. Considering the fact that measurement of solar irradiance is a long and expensive process, this paper considers the possibility of using neural network based models to generate realistic synthetic solar irradiance data. It is shown that this data can be used to realistically model long-term behaviour of a solar system as well as train short-term prediction models. The data is validated on a model of a solar system, as well as different types of predictor, and its performance compared to real, measured data.

Keywords: solar irradiance, solar irradiance prediction, time series generation, generative models, generative adversarial networks

Scientific area: Engineering and technology

Scientific field: Electrical engineering, electronics and information engineering

Classification code: T 121

Creative Commons license type: CC BY-NC

Mojim roditeljima, braći
i Milanu, uvijek posvećenom čitaocu.

Sadržaj

Informacije o mentoru i disertaciji	i
Information about mentor and dissertation	ii
Posveta	iii
Spisak tabela	vii
Spisak slika	viii
Lista skraćenica	1
1 Uvod	3
1.1 Doprinos rada	6
1.2 Organizacija rada	7
2 Potreba za sintetičkim podacima o solarnoj iradijansi	8
2.1 Dugoročno planiranje rada solarne elektrane	9
2.2 Kratkoročno planiranje rada solarne elektrane	11
2.2.1 AR model	11
2.2.2 MA model	11
2.2.3 ARMA model	12
2.2.4 Neuronske mreže	12
3 Mjerenje i reprezentacija podataka o solarnoj iradijansi	15
3.1 Uređaji za mjerenje solarne iradijanse	15

3.2	Indeks vedrine i indeks vedrog neba	17
3.3	Solarna iradijansa van atmosfere	18
3.4	Modeli vedrog neba	19
4	Algoritmi za generisanje sintetičkih podataka	21
4.1	Markovljev lanac	21
4.2	Generativne suprostavljajuće mreže	23
4.2.1	Kontinualna rekurentna generativna suprostavljajuća mreža	24
4.2.2	Rekurentna uslovna generativna suprostavljajuća mreža	25
4.2.3	Vremenska generativna suprostavljajuća mreža	26
4.2.4	Uslovna Sig-Wasserstein generativna suprostavljajuća mreža	29
4.3	Maskirani autoenkoder	30
5	Metodologija	32
5.1	Baza podataka o solarnoj iradijansi	32
5.1.1	Meteorološka stanica Davis Vantage Pro 2	32
5.1.2	Priprema podataka	33
5.2	Obučavanje modela	35
5.3	Validacija sintetičkih podataka	36
5.3.1	tSNE metoda	37
5.3.2	Analiza glavnih komponenti	38
5.3.3	Validacija podataka pomoću prediktora	39
5.3.4	Kalkulator energije koju proizvede solarna elektrana	39
6	Eksperimentalni rezultati	41
6.1	PCA i tSNE dijagrami	44
6.2	Rezultati prediktora	47
6.3	Rezultati kalkulatora	55
7	Zaključak	56

Spisak tabela

6.1	Rezultati prediktora obučenih na podacima dužine sekvence 7	48
6.2	Rezultati prediktora obučenih na podacima dužine sekvence 25	48
6.3	Generisana energija fotonaponskog sistema	55
6.4	Generisana energija fotonaponskog sistema u MWh po mjesecu	55

Spisak slika

1.1	Fotonaponski panel	4
1.2	Izolovan sistem fotonaponskih panela	5
1.3	Rangiranje država svijeta prema ukupnoj instalisanoj snazi fotonaponskih panela	6
1.4	Globalna prosječna cijena električne energije dobijene iz fotonaponskih panela	6
2.1	LSTM ćelija	14
3.1	Piranometar	16
3.2	Pirheliometar	17
3.3	Ugao deklinacije	18
4.1	C-RNN-GAN	26
4.2	TimeGAN	27
5.1	Meteorološka stanica Davis Vantage Pro 2	33
5.2	Primjer nevalidnih podataka u bazi	34
6.1	Stvarne vrijednosti indeksa vedrog neba	42
6.2	Sintetičke vrijednosti indeksa vedrog neba generisane putem TimeGAN mreže	42
6.3	TimeGAN sintetička sekvenca, mjesec jun	43
6.4	TimeGAN sintetička sekvenca, mjesec decembar	43
6.5	PCA i tSNE dijagrami LSTM-GAN podataka dužine 7	44
6.6	PCA i tSNE dijagrami C-RNN-GAN podataka dužine 7	44
6.7	PCA i tSNE dijagrami TimeGAN podataka dužine 7	45

6.8	PCA i tSNE dijagrami MAE podataka dužine 7	45
6.9	PCA i tSNE dijagrami LSTM-GAN podataka dužine 25	46
6.10	PCA i tSNE dijagrami C-RNN-GAN podataka dužine 25	46
6.11	PCA i tSNE dijagrami TimeGAN podataka dužine 25	46
6.12	PCA i tSNE dijagrami MAE podataka dužine 25	47
6.13	Koeficijenti AR modela obučenog na stvarnim podacima	49
6.14	Koeficijenti AR modela obučenog na TimeGAN25 podacima	49
6.15	PACF stvarnih podataka	50
6.16	PACF TimeGAN25 podataka	50
6.17	Predikcije dobijene AR prediktorom obučenom na MAE podacima dužine 25	51
6.18	Predikcije dobijene LSTM prediktorom obučenom na MAE podacima dužine 25	52
6.19	Predikcije dobijene AR prediktorom obučenom na TimeGAN podacima dužine 25	53
6.20	Predikcije dobijene LSTM prediktorom obučenom na TimeGAN podacima dužine 25	54

Lista skraćenica

AR – FNN Autoregressive Feed-Forward Neural Network

AR Autoregressive

ARMA Autoregressive Moving Average

BERT Bidirectional Encoder Representations from Transformers

C – RNN – GAN Continuous Recurrent Neural Network Generative Adversarial Network

CDD Cooling Degree Days

EIA Energy Information Administration

GAN Generative Adversarial Networks

HDD Heating Degree Days

LCOE Levelized Cost of Energy

LSTM Long-Term Short-Term Memory

MA Moving Average

MACC Monitoring Atmosphere Composition and Climate

MAE Masked Autoencoder

MAE Mean Absolute Error

MSE Mean Squared Error

NOCT Nominal Operating Cell Temperature

NREL National Renewable Energy Laboratory

NWP Numerical Weather Prediction

PCA Principal Component Analysis

PCAF Partial Autocorrelation Function

RCGAN Recurrent Conditional Generative Adversarial Network

RGAN Recurrent Generative Adversarial Network

RNN Recurrent Neural Network

SNE Stochastic Neighbour Embedding

SVD Singular Value Decomposition

THSW Temperature-Humidity-Sun-Wind

THW Temperature-Humidity-Wind

TMY Typical Meteorological Year

Glava 1

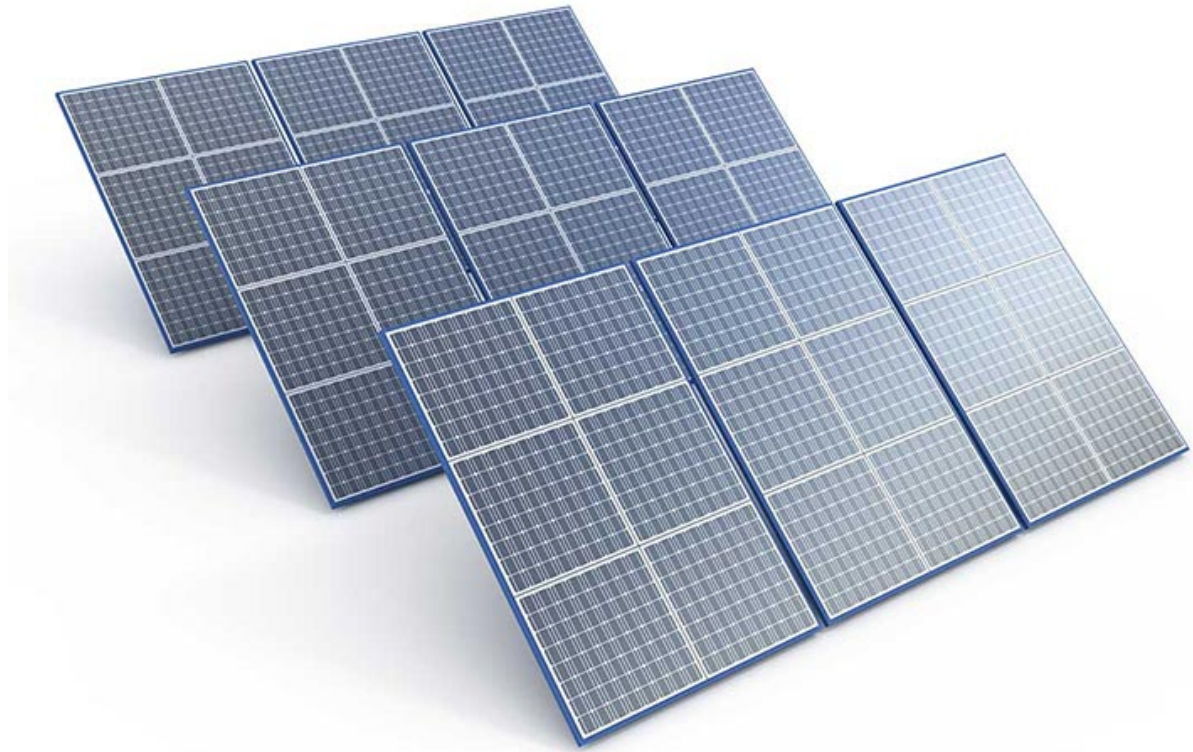
Uvod

Zaštita životne sredine i nedostatak energenata predstavljaju neke od najznačajnijih problema savremenog društva. Povećana emisija gasova staklene bašte dovodi do efekata kao što su povećanje temperature, topljenje leda u polarnim krajevima, povećanje nivoa mora, ekstremne oluje i mnogi drugi. Sagorijevanje fosilnih goriva u energetsom sektoru predstavlja najveći pokretač klimatskih promjena. Čak 30% emisija gasova staklene bašte potiče od proizvodnje struje i grijanja. Jedini način usporavanja ovih negativnih efekata je smanjenje emisije gasova, te već nekoliko decenija postoji težnja većeg broja svjetskih država ka dekarbonizaciji i u tu svrhu pokrenute su različite političke direktive [1]. Pariškim sporazumom iz 2015. godine, koji je potpisalo 195 zemalja odlučeno je da će porast temperature biti ograničen na vrijednost manju od 2 °C u odnosu na predindustrijsko doba [2]. Sem ekoloških problema koje izazivaju, količina fosilnih goriva je ograničena i postavlja se pitanje održivosti ovakvog sistema.

Kao jedna od osnovnih mjera borbe protiv klimatskih promjena javlja se inicijativa zamjene klasičnih metoda proizvodnje električne energije kao što su termo i nuklearne elektrane, obnovljivim izvorima energije. Obnovljivi izvori energije kao što su vjetro i solarna energija predstavljaju jeftin i održiv izvor energije, i u posljednjih nekoliko godina doživjeli su rekordan porast koji se predviđa i u skorijoj budućnosti. Američka agencija EIA (eng. *Energy Information Administration*) predviđa da će do 2050. godine udio obnovljivih izvora u globalnoj proizvodnji električne energije iznositi čak 59% [3].

Kao najbrže rastući obnovljivi izvor ističe se solarna energija. Pod solarnom energijom podrazumijevamo energiju koja potiče od Sunčevog zračenja, a koja se potom konvertuje u termalnu ili električnu energiju. Konvertovanje solarne u električnu energiju zasnovano je na principu fotoelektričnog efekta, otkrivenog još 1839. godine. Poluprovodnički materijali kao što je silicijum pri izlaganju Sunčevom zračenju otpuštaju elektrone i na taj način mogu da generišu električnu energiju. Današnji fotonaponski paneli, koji se koriste za proizvodnju električne energije dobijaju se međusobnim povezivanjem nekoliko stotina, ili čak hiljada ovakvih poluprovodničkih ćelija kao na Slici 1.1 [4]. Korištenjem solarne energije moguće je napajati različite vrste potrošača, tako da srećemo širok spektar solarnih postrojenja, od jako malih, koji napajaju isključivo jedan potrošač ili jedno domaćinstvo,

do velikih, megavatnih, postrojenja za proizvodnju električne energije. Solarna energija se jednostavno može iskoristiti i za napajanje ruralnih ili teško dostupnih područja gdje priključenje na elektroenergetsku mrežu nije ekonomski isplativo. Slika 1.2 [5] prikazuje izolovan sistem solarnih panela koji se koristi za napajanje pumpe za vodu u afričkom selu na lokaciji koja je previše udaljena da bi se mogla priključiti na elektroenergetsku mrežu.



Slika 1.1: Fotonaponski panel

Solarna postrojenja tokom rada ne izazivaju emisiju štetnih gasova, i predstavljaju ne samo ekološki prihvatljiv već i jedan od "najčistijih" izvora električne energije. Prosječno postrojenje ima životni vijek od otprilike 20 godina, tokom kojih ne zahtijeva gotovo nikakvo održavanje s obzirom na to da u postrojenju ne postoje pokretni dijelovi koji bi se mogli istrošiti s vremenom. Sem toga, solarna energija istinski predstavlja nepresušan izvor energije sa jako velikim potencijalom. Energija koja potiče od Sunca, a koja dopiye na Zemlju svaki sat iznosi 100 000 TW, što je dovoljno energije da opskrbi čitavu planetu na godinu dana, pri čemu svega mali procenat ove energije biva iskorišten.

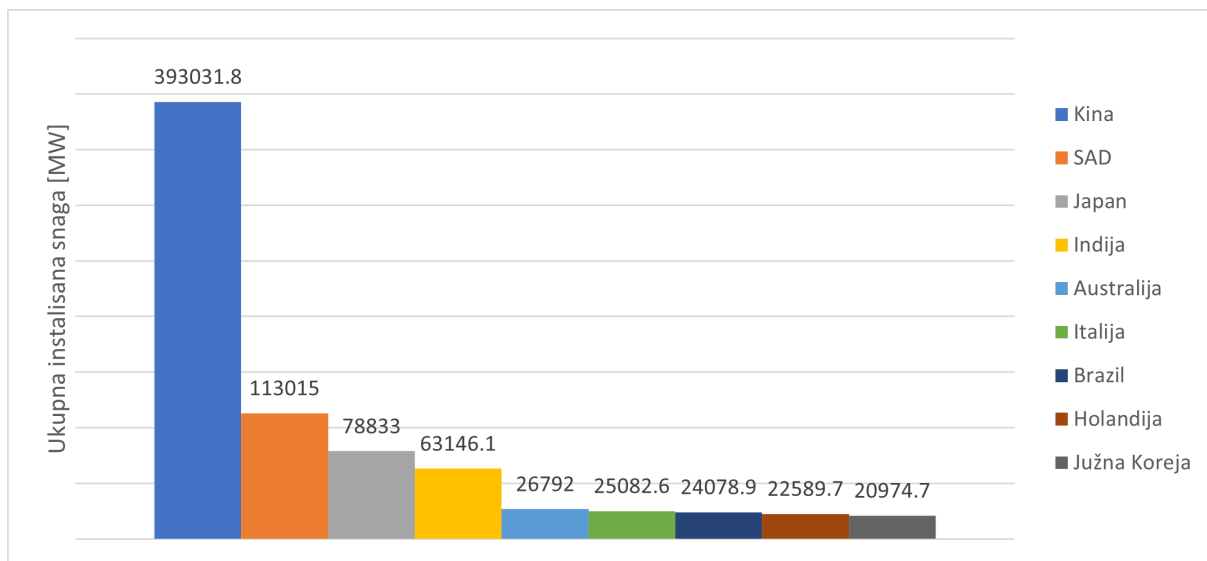
Od 2010. do 2022. godine, količina globalne instalisane snage solarnih proizvođača povećala se sa 41 593 MW na vrijednost od 1 053 115 MW. Solarna energija takođe predstavlja jedan od najjeftinijih izvora električne energije. Prosječna cijena električne energije koja potiče iz solarnih jedinica u 2010. godini iznosila je 0,35 €/kWh, a do 2019. ova brojka



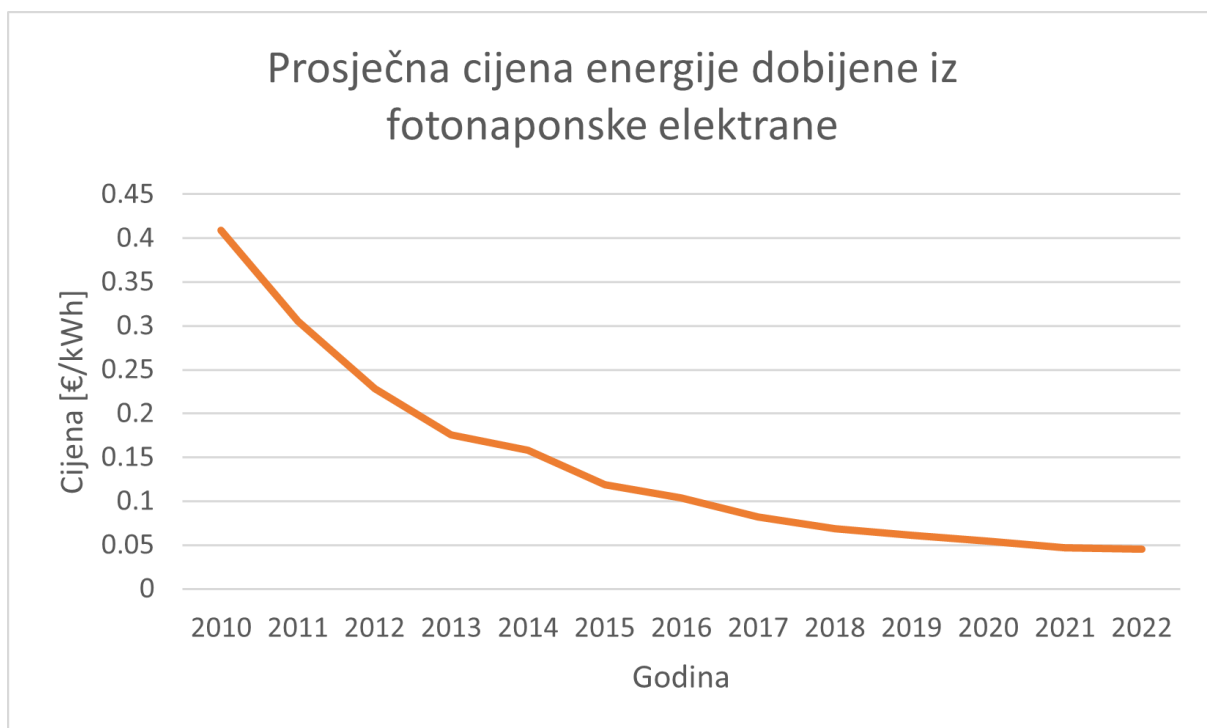
Slika 1.2: Izolovan sistem fotonaponskih panela

opala je na iznos od 0,062 €/kWh [6]. Od svih zemalja prema instalisanoj snazi prednjači Kina sa instalisanih 393 031,8 MW, a potom slijede SAD sa 113 015 MW, Japan sa 78 833 MW i Njemačka sa 66 554 MW [7]. Na Slici 1.3 prikazan je dijagram na kom je prikazano 10 vodećih zemalja svijeta prema ukupnoj instalisanoj snazi fotonaponskih panela i ukupna snaga za svaku od zemalja. Na Slici 1.4 prikazana je promjena vrijednosti LCOE (eng. *Levelized Cost of Energy*) solarne energije, koja predstavlja prosječan trošak proizvodnje električne energije iz nekog izvora kada se uzima u obzir njen životni vijek.

Sa ovakvim porastom upotrebe solarne energije porasla je i potreba za mehanizmima za optimizaciju rada postojećih solarnih postrojenja, kao i za alatima za planiranje novih postrojenja. Rad solarne elektrane, odnosno energija koju elektrana generiše, zavisi od jačine sunčevog zračenja. Jačina sunčevog zračenja kvantifikuje se uz pomoć solarne iradijance, koja predstavlja snagu sunčevog zračenja po jedinici površine. Poznavanjem istorijskih podataka o solarnoj iradijansi na lokaciji od interesa, tokom dužeg vremenskog perioda, moguće je izvesti zaključke o solarnom potencijalu lokacije i varijacijama izlazne snage elektrane koje se mogu desiti. Mjerenje solarne iradijance na većini lokacija za duži vremenski period veoma je teško, prema tome, javlja se potreba za sintetičkim podacima koji bi uspješno simulirali realistične scenarije sunčevog zračenja. Cilj ovog rada je ispitivanje mogućnosti korištenja neuronskih mreža u svrhu generisanja takvih podataka.



Slika 1.3: Rangiranje država svijeta prema ukupnoj instaliranoj snazi fotonaponskih panela



Slika 1.4: Globalna prosječna cijena električne energije dobijene iz fotonaponskih panela

1.1 Doprinos rada

U ovom radu biće razmatrana mogućnost upotrebe različitih arhitektura neuronskih mreža namijenjenih za generisanje sintetičkih vremenskih nizova na problemu generisanja podataka o solarnoj iradijansi za područje grada Banjaluke. S obzirom na dugačak i skup proces mjerenja solarne iradijanse, cilj je ispitati da li je moguće na manjoj količini po-

dataka obučiti model koji će potom biti u stanju da generiše velike količine podataka u svrhu upotrebe u primjenama kao što su predikcija rada solarne elektrane, kao i predikcija budućih vrijednosti samog zračenja. U radu će biti dat pregled dosadašnjih značajnijih modela za generisanje vremenskih nizova, potom opisana priprema baze podataka o sunčevom zračenju i meteorološkim parametrima za teritoriju grada Banjaluke. Uz pomoć ove baze podataka biće obučena četiri modela za generisanje sintetičkih podataka o solarnoj iradijansi. Takođe će biti ispitan uticaj dužine vremenskog niza na performanse modela i validnost i upotrebljivost dobijenih sintetičkih podataka. Performanse sintetičkih podataka u ciljnim primjenama će potom biti upoređene sa performansama stvarnih, mjerenih podataka o solarnoj iradijansi.

1.2 Organizacija rada

Nakon uvoda, ostatak rada organizovan je u šest glava. Glava 2 opisuje potrebu za sintetičkim podacima o solarnoj iradijansi, i primjenu ovih podataka u praksi. Glava 3 opisuje instrumente i procedure za mjerenje podataka o solarnoj iradijansi. Sem toga opisani su načini reprezentacije podataka koji ih čine pogodnim za primjenu u svrhu treniranja neuronskih mreža. U Glavi 4 dat je pregled literature i opisani su modeli koji su do sada pokazali uspjeh u generisanju različitih tipova vremenskih nizova. Glava 5 opisuje korištenu metodologiju, a Glava 6 prikazuje dobijene rezultate. Posljednja, Glava 7 sadrži zaključak i potencijalne pravce daljeg istraživanja.

Glava 2

Potreba za sintetičkim podacima o solarnoj iradijansi

Prije izgradnje solarne elektrane potrebno je poznavati solarni kapacitet geografske lokacije na kojoj će elektrana biti smještena. Od solarnog kapaciteta zavisi proizvodnja elektrane, a samim tim i ekonomska isplativost njene izgradnje. Jačina sunčevog zračenja je stohastičke prirode, zavisi od doba dana, godine i oblačnosti koja je korelisana sa različitim meteorološkim parametrima kao što su temperatura, vlažnost vazduha i slično. Poznavanje tipičnih vrijednosti solarne iradijanse za duži vremenski period omogućava da se pomoću različitih modela koji simuliraju rad elektrane procijeni potencijalna količina električne energije koja može da bude proizvedena, kao i uticaj elektrane na ostatak sistema na koji je povezana. Podaci o solarnoj iradijansi za vremenski period dovoljno dugačak da bi se analiziralo ponašanje elektrane najčešće nisu dostupni. Solarna iradijansa varira, ne samo na kratkoročnom nivou, već i na nivou nekoliko godina, tako da čak ni podaci o dužem vremenskom periodu, kao što je 20 ili 30 godina, ponekad ne mogu realistično da prikažu sve potencijalne scenarije [8]. Za prikupljanje podataka o solarnoj iradijansi bilo bi potrebno vršiti mjerenja tokom nekoliko decenija uz pomoć posebnih uređaja o kojima će kasnije biti riječ u Glavi 3. Da bi se izbjegao ovaj proces potrebne su metode koje bi na osnovu ograničenog broja podataka o prethodnim vrijednostima solarne iradijanse mogle da generišu veći broj sintetičkih podataka koji bi bili u stanju da uspješno simuliraju potencijalne scenarije.

Solarne elektrane mogu funkcionisati kao dio postojećeg elektroenergetskog sistema, pri čemu se njihovo priključenje može vršiti na prenosnu ili čak direktno na distributivnu mrežu. Druga opcija je ostrvski rad elektrane. Potrošači na udaljenim lokacijama, čije bi priključenje na distributivnu mrežu bilo skupo ili čak tehnički neizvodljivo, često ovaj problem rješavaju napajanjem iz fotonaponskih panela, ili nekog drugog obnovljivog izvora energije kao što su vjetroelettrane, ili kombinacijom više različitih izvora kako bi se obezbijedila rezerva napajanja.

Ukoliko je elektrana priključena na mrežu, korisno je vršenje i kratkoročne predikcije njenog ponašanja. Stohastička priroda solarne energije dovodi do variranja snage koju

elektrana predaje u mrežu, što u toj mreži može da izazove poremećaje ukoliko ne postoji adekvatno planiranje. Poremećaji u mreži mogu narušiti rad mreže, uticati na potrošače, pa čak dovesti i do oštećenja, kvarova i smanjenja životnog vijeka opreme. Sem toga, proizvođači električne energije zainteresovani su za predikciju kako bi procijenili količinu fosilnih goriva koju mogu da uštede korištenjem obnovljivih izvora, kako bi bolje organizovali rad instaliranih elektrana, i kako bi olakšali integraciju solarnih postrojenja u električnu mrežu. Na osnovu proizvodnje solarnih elektrana i drugih obnovljivih izvora i na osnovu trenutnih potreba potrošača u sistemu potrebno je planirati rad baznih elektrana čija snaga ne zavisi od stohastičkih elemenata i koje je moguće kontrolisati (termo, hidro, nuklearne elektrane). Povećanjem udjela obnovljivih izvora energije, mnoge zemlje uvele su i regulacije koje govore o monitoringu i planiranju.

Ostrvski sistemi nisu povezani na elektroenergetsku mrežu i napajaju se u potpunosti iz solarnih izvora energije. Stohastička priroda solarne energije u ovakvim sistemima može izazvati značajan disbalans između proizvodnje i potrošnje, odnosno do toga da solarna elektrana proizvodi premalo ili previše energije za trenutne potrebe potrošača. Kako bi se obezbijedio pravilan rad potrebno je planiranje i neka vrsta predikcije snage elektrane na različitim vremenskim skalama.

Proizvodnja solarne elektrane zavisi direktno od jačine sunčevog zračenja. U skladu s tim, kako bi se ostvarila predikcija snage, u navedenim sistemima potrebno je vršiti kratkoročnu predikciju vrijednosti solarne iradijance, pri čemu vremenska rezolucija može da bude između 15 minuta i 24 sata. Poznavanje iradijance omogućava adekvatno planiranje rada tog sistema. Postoji više različitih metoda za predikciju. Ono što je zajedničko većini metoda je potreba za prethodnim vrijednostima solarne iradijance, na osnovu kojih se određuju parametri modela, koji su potom u stanju da vrše predikciju budućih vrijednosti. I u ovom slučaju, u svrhu određivanja parametara modela za predikciju, potreban je velik broj podataka o vrijednostima solarne iradijance. Radi povećanja obima podataka korisno je posjedovati mehanizam za generisanje realističnih sintetičkih podataka.

U nastavku rada biće opisano nekoliko metoda za dugoročno planiranje i kratkoročnu predikciju rada solarne elektrane, čija je zajednička karakteristika potreba za podacima o solarnoj iradijansi za područje od interesa.

2.1 Dugoročno planiranje rada solarne elektrane

Tipična meteorološka godina, TMY (eng. *Typical Meteorological Year*) već duže vrijeme predstavlja osnovni standard modelovanja solarnog kapaciteta nekog područja. TMY predstavlja mehanizam pomoću kojeg se podaci o solarnoj iradijansi nekog područja, dobijeni za duži niz godina mogu sažeti, i predstaviti u vidu jedne tipične meteorološke godine. Nastala je u periodu kada ograničena procesorska moć računara nije dozvoljavala obradu veće količine podataka. Danas se koristi za procjenu prosječnog solarnog kapaciteta nekog područja i određivanje prosječne potencijalne proizvodnje koju bi solarna elektrana mogla da dostigne na toj lokaciji.

Ulazni podaci za formiranje TMY su podaci o solarnoj iradijansi, dobijeni na osnovu mjerenja na površini Zemlje ili na osnovu satelitskih podataka, za duži vremenski period, najčešće više od deset godina. TMY se formira od odabranih podataka iz ulaznog skupa. Postoji više različitih procedura formiranja TMY i više različitih kriterijuma za odabir podataka koji ulaze u formiranje TMY.

- Najjednostavniji pristup je računanje prosječne vrijednosti solarne iradijanse za svaki mjesec u ulaznom skupu podataka. Ove prosječne vrijednosti se potom porede sa ukupnim prosjekom odgovarajućeg mjeseca na cijelom skupu podataka. Kao podaci za formiranje meteorološke godine biraju se oni mjeseci čiji su prosjeci najpribližniji ukupnom prosjeku za odgovarajući mjesec. TMY se formira konkatencijom odabranih mjeseci.
- Drugi potencijalni metod je odbacivanje mjeseci iz ulaznog skupa podataka u koracima. U prvom koraku odbacuju se svi mjeseci čija kumulativna raspodjela prosječnih dnevnih vrijednosti iradijanse previše odstupa od ukupne raspodjele. U drugom koraku biraju se mjeseci koji nekoliko dana zaredom imaju vremenske uslove koji su identifikovani kao tipični u originalnom skupu podataka. U posljednjem koraku biraju se oni mjeseci čija je prosječna vrijednost solarne iradijanse najpribližnija ukupnom prosjeku odgovarajućeg mjeseca u ulaznom skupu podataka.
- Još jedan moguć pristup je korištenje klizećeg prozora dužine 365 dana na originalnom skupu ulaznih podataka, pri čemu se za svaki prozor određuje prosječna vrijednost solarne iradijanse, i kao TMY bira onaj prozor koji ima prosječnu vrijednost najpribližniju ukupnom prosjeku cijelog skupa podataka.

Sem navedenih metoda, postoji i pristup u kojem se za formiranje TMY koristi kombinacija izmjerenih podataka i podataka modelovanih na osnovu statističkih svojstava stvarnih vrijednosti iradijanse. Dobijeni modeli se najčešće kategorišu prema nivou povjerenja, tako da razlikujemo TMY P50, TMY P75, TMY P90 itd. Oznake 50, 75 i 90 nam govore da će stvarna energija sunčevog zračenja biti veća od one procijenjene na osnovu TMY modela, 50, 75 ili 90% vremena [9].

Iako TMY daje dobru sliku o prosječnom solarnom potencijalu nekog područja, ona ne može da prikaže potencijalne varijacije koje se dešavaju u toku životnog vijeka jedne solarne elektrane. Zbog načina na koji je TMY formirana, ona ne sadrži ekstremne vrijednosti solarne iradijanse koje mogu da se pojave, i samim tim ne prikazuje realističnu sliku o varijabilnosti koju solarna iradijansa može da ispoljava. Ovo predstavlja poseban problem kada je potrebno simulirati najgore moguće uslove u kojima elektrana može da se nađe. Za detaljniju analizu dugoročnog ponašanja potreban nam je veći broj podataka, koji bi prikazivali ne samo prosječne vrijednosti zračenja, već i krajnje vrijednosti koje zračenje može dostići tokom životnog vijeka elektrane. U ovu svrhu potrebni su nam realistični podaci o solarnoj iradijansi za područje od interesa, koji bi omogućili simulaciju stvarnog rada elektrane i njen uticaj na sistem tokom dužeg vremenskog perioda.

2.2 Kratkoročno planiranje rada solarne elektrane

Kao što je već objašnjeno u prethodnim odjeljcima, kratkoročna predikcija, sa vremenskom rezolucijom od nekoliko minuta ili nekoliko sati je ključna za planiranje rada solarne elektrane, kao i za planiranje rada samog elektroenergetskog sistema. Na osnovu ulaznih podataka koji se koriste razlikujemo dvije osnovne grupe modela za predikciju vrijednosti solarne iradijance.

Prvu grupu čine metode koje vrše predikciju na osnovu stanja atmosfere i meteoroloških uslova. U ovoj grupi razlikujemo numeričku predikciju vremena, NWP (eng. *Numerical Weather Prediction*) i metode zasnovane na analizi satelitskih snimaka. NWP se zasniva na numeričkom rješavanju sistema parcijalnih diferencijalnih jednačina koje opisuju stanje u atmosferi, pri čemu su početni uslovi jednačina prethodna stanja atmosfere. Kao što i sam naziv govori, metode zasnovane na analizi satelitskih snimaka koriste satelitske snimke područja od interesa i na osnovu informacija koje snimci daju, unaprijed predviđaju, ne samo vrijednost solarne iradijance, već i druge meteorološke parametre [10].

Drugu grupu čine modeli koji buduće vrijednosti solarne iradijance predviđaju na osnovu poznavanja prethodnih vrijednosti iradijance. U ovu grupu spadaju statistički modeli, kao što su AR, MA i opštiji ARMA model, i u skorije vrijeme sve popularnije neuronske mreže.

2.2.1 AR model

AR (eng. *Autorregressive*) model je regresivni model koji predviđa narednu vrijednost nekog vremenskog niza na osnovu poznatih vrijednosti niza u prethodnim vremenskim trenucima. Kao posljedica, preduslov korištenja ovog modela je postojanje značajne korelacije trenutne vrijednosti vremenskog niza sa vrijednostima u prethodnim trenucima. Ukoliko su prethodne vrijednosti vremenskog niza poznate, AR model računa predikciju narednog odmjerka kao linearnu kombinaciju prethodnih odmjeraka, odnosno kao sumu odmjeraka pomnoženih odgovarajućim težinama. Težine kojima se množi svaki odmjerak određuju se na osnovu trening skupa podataka. Prema AR modelu vrijednost t -tog odmjerka nekog vremenskog niza iznosi:

$$x_t = \sum_{i=1}^p \phi_i x_{t-i} + \epsilon_t, \quad (2.1)$$

gdje je p red modela, odnosno broj prethodnih odmjeraka na osnovu kojih se vrši izračunavanje, ϕ_i su konstante koje je potrebno odrediti pri obučavanju modela, a ϵ_t greška između stvarne vrijednosti odmjerka i predikcije dobijene AR modelom.

2.2.2 MA model

MA (eng. *Moving Average*) model oslanja se na sličan koncept kao i AR model. U ovom slučaju pretpostavlja se korelacija odmjerka vremenskog niza koji se posmatra sa greškom predikcije prethodnih odmjeraka. Vrijednost vremenskog niza od interesa predviđa

se kao linearna kombinacija grešaka predikcije prethodnih vrijednosti, odnosno linearna kombinacija razlika između stvarnih vrijednosti niza i njihovih predikcija. Težine kojima se množe vrijednosti grešaka i u ovom slučaju određuju se na osnovu trening skupa podataka. Matematička definicija MA modela je data kao:

$$x_t = \mu + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t, \quad (2.2)$$

gdje je μ srednja vrijednost niza koji se modeluje, θ_i konstante koje je potrebno odrediti, a ϵ_t predstavlja razliku između stvarne vrijednosti t -tog odmjerka i predikcije dobijene modelom.

2.2.3 ARMA model

ARMA (eng. *Autoregressive Moving Average*) model predstavlja linearnu kombinaciju AR i MA modela. Samim tim izračunavanje se vrši sabiranjem dvije sume, od kojih svaka odgovara jednom od dva ranije opisana modela. Odmjerak vremenskog niza na osnovu ARMA modela izračunava se kao:

$$x_t = \sum_{i=1}^p \phi_i x_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t, \quad (2.3)$$

pri čemu prva suma u jednačini predstavlja AR a druga suma MA komponentu modela.

Statistički modeli imaju nekoliko prednosti. Ovakvi modeli su fleksibilni i zavisno od izabranog reda mogu raditi sa različitim tipovima vremenskih podataka. Takođe je pokazano da imaju sposobnost da izdvoje linearne korelacije i statistička svojstva unutar podataka. Osnovni problem sa ovakvim modelima je činjenica da ne mogu da uspješno vrše predikciju vremenskih nizova koji imaju nelinearna svojstva. Iz tog razloga ARMA modeli često se koriste u kombinaciji sa nekim tipom nelinearnih modela kao što su neuronske mreže, te se na ovaj način dobijaju tzv. hibridni modeli [11].

2.2.4 Neuronske mreže

Drugu podgrupu modela koji za predikciju koriste prethodne vrijednosti solarne iradijance čine neuronske mreže, čija je upotreba dobila poseban značaj u prethodnih nekoliko godina. Za razliku od statističkih modela, neuronske mreže predstavljaju nelinearan model za predikciju i samim tim mogu da prepoznaju nelinearne šablone koji se kriju u podacima. Sem toga, neuronske mreže mogu da predikciju vrše ne samo na osnovu prethodnih vrijednosti iradijance, već i na osnovu drugih poznatih podataka u prethodnom periodu koji su korelisani sa jačinom sunčevog zračenja. Najjednostavnija opcija je korištenje neuronskih mreža bez povratnih veza (eng. *Feedforward Neural Network*), u kojima izlaz neurona zavisi isključivo od izlaza neurona u prethodnom sloju. Za rad sa sekvencijalnim podacima mnogo je povoljnije korištenje rekurentnih neuronskih mreža, RNN (eng. *Recurrent Neural Network*). Neuroni rekurentnih neuronskih mreža sadrže povratne veze, tako

da izlaz pojedinog neurona zavisi ne samo od izlaza neurona prethodnog sloja, već i od izlaza posmatranog neurona u prethodnom koraku. Zahvaljući ovoj osobini ovakve mreže imaju memorijska svojstva i mogu bolje da detektuju sekvencijalne zavisnosti podataka. Osnovni problem rekurentnih mreža je javljanje iščezavanja ili eksploziranja gradijenta. Pokazano je da se signal greške u rekurentnim mrežama, za vrijeme propagacije unazad, značajno povećava ili smanjuje kao posljedica postojanja povratnih veza, te se problem javlja već nakon 5 do 10 koraka [12].

LSTM (eng. *Long-Term Short-Term Memory*) ćelija [13] nastala je kao rješenje problema nestajućeg ili eksplozirajućeg gradijenta koji se javlja pri treniranju rekurentnih mreža. Za razliku od rekurentnih mreža, LSTM koristi kombinaciju vrijednosti izlaza u prethodnoj iteraciji (kratkotrajna memorija) i dugotrajne memorije koja zavisi od izlaza ćelije na većem broju prethodnih iteracija. Originalna verzija LSTM ćelije dopunjena je takozvanim gejtom za zaboravljanje [14] i danas se ova verzija smatra kao standardni oblik LSTM ćelije. Šema LSTM ćelije prikazana je na Slici 2.1. Izlaz LSTM ćelije u vremenskom koraku t , h_t , zavisi od trenutnog ulaza ćelije x_t , izlaza ćelije u prethodnom koraku h_{t-1} , što ujedno predstavlja i kratkoročnu memoriju i tzv. stanja ćelije u prethodnom koraku c_{t-1} . Stanje ćelije zavisi od prethodnih nekoliko vremenskih koraka i predstavlja dugoročnu memoriju ćelije.

Sama LSTM ćelija sastoji se od nekoliko dijelova i ažuriranje svakog od dijelova u određenom vremenskom koraku odvija se na sljedeći način:

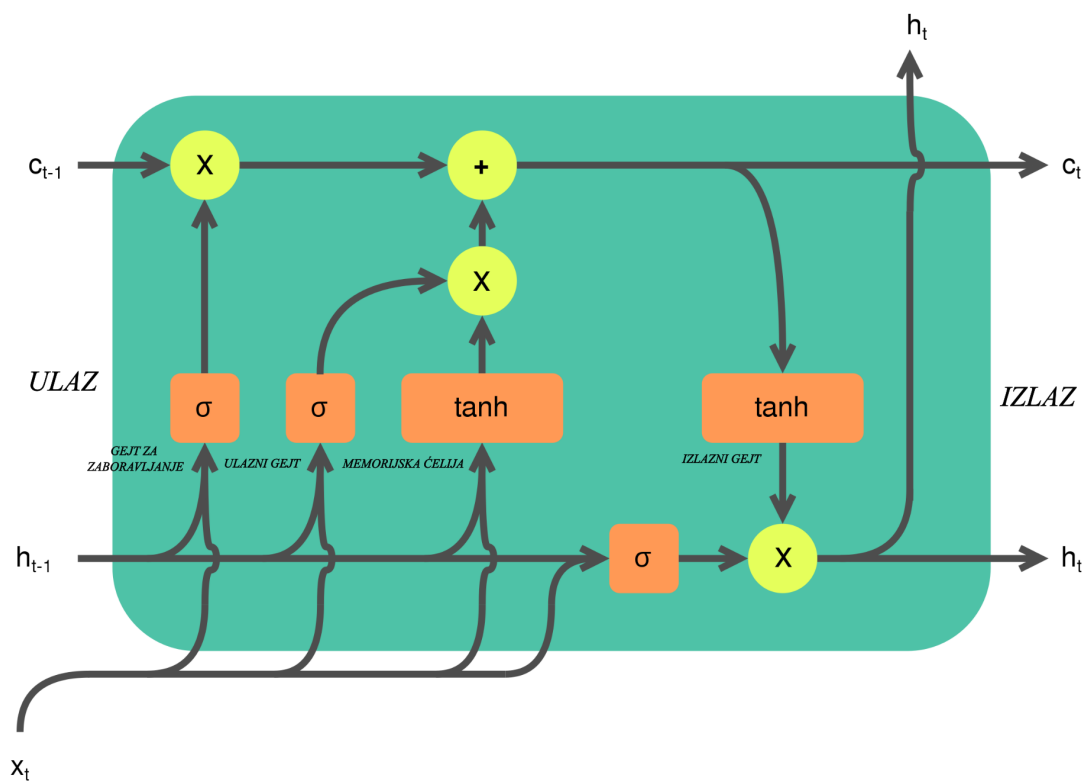
1. Ulazni blok: njegovo ažuriranje se vrši kombinovanjem trenutne vrijednosti ulaza ćelije i vrijednosti izlaza ćelije u prethodnoj iteraciji,
2. Ulazni gejt: ažuriranje se vrši kombinovanjem ulaza, izlaza u prethodnoj iteraciji i vrijednosti memorijske ćelije u prethodnoj iteraciji,
3. Gejt za zaboravljanje: u ovom koraku odlučuje se koje vrijednosti iz memorijske ćelije će biti izbrisane,
4. Memorijska ćelija: ažuriranje se vrši na osnovu vrijednosti ulaznog gejta, gejta za zaboravljanje i vrijednosti memorijske ćelije u prethodnoj iteraciji,
5. Izlazni gejt: ažuriranje se vrši kombinovanjem ulaza, izlaza u prethodnoj iteraciji i vrijednosti memorijske ćelije u prethodnoj iteraciji,
6. Izlazni blok: ažuriranje se vrši kombinacijom trenutne vrijednosti memorijske ćelije i izlaznog gejta.

Pri navedenim koracima primjenjuju se nelinearne aktivacije, pri čemu je aktivaciona funkcija gejtova uvijek sigmoid, dok je u blokovima česta aktivacija hiperbolički tangens.

Jedna od najznačajnijih primjena LSTM-a je predikcija vremenskih nizova [15]. Pokazano je da je superiorna za primjene kao što su predikcija finansijskog tržišta [16], cijene nafte [17] i medicinskih podataka, kao što je širenje virusa COVID-19 [18]. Između ostalog,

2.2. *Kratkoročno planiranje rada solarne elektrane*

postoje primjeri korištenja LSTM mreža u svrhu kratkoročne predikcije solarne iradijanske, zarad štednje energenata i planiranja rada sistema. Takvi primjeri prikazani su u [19] i [20].



Slika 2.1: LSTM ćelija

Glava 3

Mjerenje i reprezentacija podataka o solarnoj iradijansi

Za rješenje problema dugoročne i kratkoročne predikcije, koji su pomenuti u prethodnim odjeljcima, potrebna nam je relativno velika količina podataka o solarnoj iradijansi na geografskom području na kom vršimo analizu. Kada je dugoročno planiranje u pitanju, podaci su nam potrebni kako bismo prije izgradnje same elektrane uspješno simulirali kako kratkoročne, tako i dugoročne varijacije njene proizvodnje. Što se tiče kratkoročne predikcije rada elektrane, modele koji koriste prethodne vrijednosti solarne iradijanse potrebno je na neki način obučiti na osnovu poznatih podataka o iradijansi u prošlosti. Iz tog razloga takođe je potrebno poznavati prethodne vrijednosti iradijanse za duži vremenski period. U ovoj glavi biće razmatrano mjerenje i modeli za predstavljanje izmjerene količine sunčevog zračenja.

Ukupna iradijansa koja dopiye na horizontalnu površinu na Zemlji naziva se globalna solarna iradijansa horizontalne podloge i sastoji se od dvije osnovne komponente: direktne i difuzne. Direktna komponenta predstavlja dio sunčevih zračenja koji na površinu dopiyeva direktno iz pravca trenutnog položaja Sunca. Pri prolasku kroz atmosferu, dio sunčevih zraka rasipa se usljed interakcije sa česticama vazduha, aerosola i drugih gasova u atmosferi. Ovakvi zraci na površinu Zemlje dolaze iz svih pravaca i čine drugu, difuznu komponentu zračenja. Poznavanjem ovih dvaju komponenata globalna iradijansa na horizontalnu podlogu može se odrediti kao njihova suma.

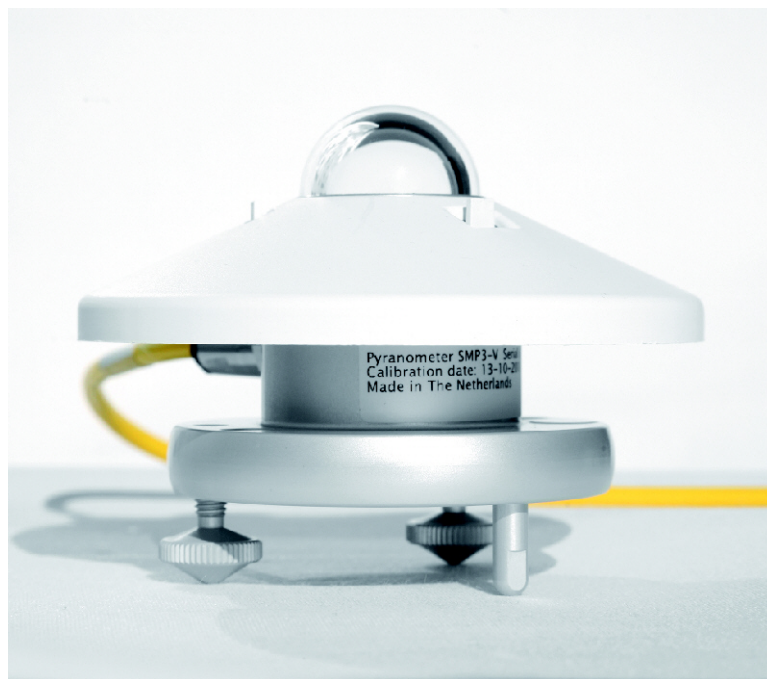
3.1 Uređaji za mjerenje solarne iradijanse

Baza podataka o solarnoj iradijansi formira se višegodišnjim mjerenjem vrijednosti iradijanse na području od interesa. Postoji više vrsta uređaja za mjerenje iradijanse, pri čemu ti uređaji mogu da mjere globalnu iradijansu ili neku od opisanih komponenti.

Na Slici 3.1 [21] prikazan je uređaj koji se naziva piranometar. Piranometri su uređaji koji služe za mjerenje globalne solarne iradijanse, odnosno ukupne iradijanse koja dopiye na površinu Zemlje. Oni mjere ukupnu solarnu iradijansu na prostornom uglu od 2π

3.1. Uređaji za mjerenje solarne iradijance

steradijana iznad senzora. Osnovna podjela uređaja je prema tipu senzora. Prvi tip senzora koji se ugrađuje zasnovan je na principu termoelektričnog efekta, odnosno realizovan je kao termopar, spoj dva metala na čijim krajevima se javlja napon koji zavisi od njihove temperaturne razlike. Drugi tip senzora zasnovan je na principu fotoelektričnog efekta, odnosno realizovan je kao poluprovodnik koji pod uticajem svjetlosnih zraka generiše električnu struju. Termoparski senzori su obično precizniji i ugrađuju se u skuplje uređaje. Mjerenje piranometrom tipično se vrši na nivou sekunde, nakon čega se vrši usrednjavanje i čuvanje podataka sa željenom vremenskom rezolucijom.



Slika 3.1: Piranometar

Pirheliometar, prikazan na Slici 3.2 [22], je uređaj koji služi za mjerenje direktne komponente solarne iradijance. Sadrži cijev za usmjeravanje i mehanizam koji omogućava praćenje Sunca. Na ovaj način osigurava se to da su zrake direktne komponente zračenja uvijek usmjerene ka uređaju. Pirheliometar se obično koristi u kombinaciji sa piranometrom kada je potrebno preciznije određivanje direktne i difuzne komponente zračenja. Difuzna komponenta dobija se kao razlika globalne iradijance izmjerene piranometrom i direktne komponente izmjerene pirheliometrom [23].



Slika 3.2: Pirheliometar

3.2 Indeks vedrine i indeks vedrog neba

Ranije je napomenuto da potreba za predikcijom solarne iradijance potiče od njene stohastičke prirode. Ukupna količina solarne iradijance koja dopiye na površinu Zemlje zavisi od atmosferskih parametara i oblačnosti. Sa druge strane, solarne iradijance sadrži i determinističku komponentu koju je moguće egzaktno odrediti. Ukoliko bismo razmatrali solarne iradijance na nekim geografskim koordinatama iznad atmosfere, njenu vrijednost bismo mogli odrediti u svakom trenutku. Ova vrijednost zavisi od trenutnog položaja Sunca u odnosu na posmatranu lokaciju, što dalje zavisi od geografskih koordinata te lokacije, posmatranog dana u godini i sata. Jedan način za uklanjanje determinističke komponente i sezonskih trendova iz podataka o iradijanci je dijeljenje stvarne vrijednosti globalne solarne iradijance na horizontalnu površinu sa odgovarajućom vrijednošću van atmosfere svedenom na horizontalnu površinu. Na ovaj način dobijamo tzv. indeks vedrine (eng. *clearness index*), koji zavisi isključivo od stanja u atmosferi. Indeks vedrine može da poprime vrijednost iz intervala $[0, 1]$, te je njegovim uvođenjem izvršena normalizacija podataka.

Čak i pri idealno vedrim uslovima sva količina zračenja van atmosfere ne dopiye na Zemlju. Čestice vodene pare i aerosola izazivaju raspršivanje sunčevih zraka i, zavisno od koncentracije, čak do 25% zračenja ne dopiyeva na površinu Zemlje. Sem modela kojima je moguće odrediti količinu zračenja van atmosfere, postoje tzv. modeli vedrog neba, kojima se modeluje količina sunčevog zračenja koja dopiye na površinu Zemlje pri idealno vedrim uslovima. Dijeljenjem stvarne vrijednosti globalne iradijance sa modelom vedrog neba dobija se tzv. indeks vedrog neba (eng. *clear sky index*). Indeks vedrog neba, kao i indeks vedrine, vrši normalizaciju podataka i uklanjanje determinističke komponente iz izmjerenih vrijednosti iradijance [24, 25].

Radi uklanjanja determinističke komponente iz podataka o solarnoj iradijansi i normalizacije podataka, podaci korišteni za obučavanje generativnih modela korištenih u ovom radu svedeni su na domen indeksa vedrog neba korištenjem jednog od modela vedrog neba opisanih u narednom odjeljku. U skladu s tim, i sintetički podaci generisani od strane ovih modela biće generisani u vidu sekvenci indeksa vedrog neba. Na ovaj način omogućava se obučavanje modela za generisanje isključivo stohastičke komponente podataka.

3.3 Solarna iradijansa van atmosfere

Da bismo na osnovu podataka o solarnoj iradijansi formirali vrijednosti indeksa vedrine, potrebno je poznavati model solarne iradijanse ukoliko ne bi postojao nikakav uticaj atmosfere. Solarna iradijansa na ulazu u Zemljinu atmosferu zavisi isključivo od udaljenosti Zemlje od Sunca, koja se mijenja zavisno od dana u godini. Za određeni dan u godini, n , moguće je odrediti vrijednost iradijanse na ulazu u atmosferu kao:

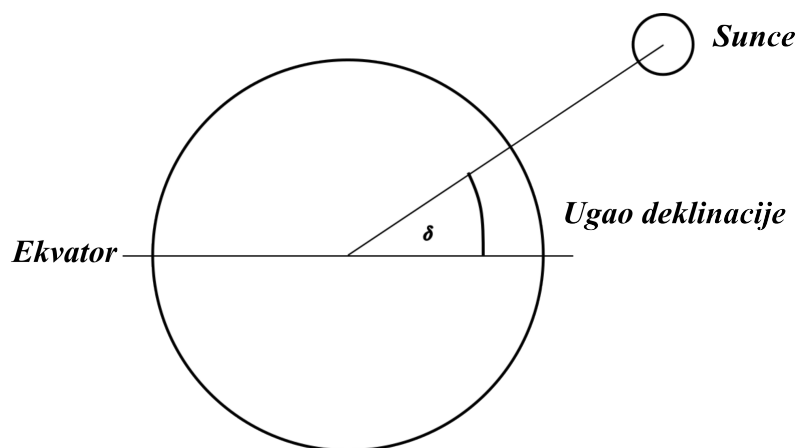
$$G_0 = G_{SC} \left[1 + 0,034 \cdot \cos\left(\frac{360n}{365}\right) \right], \quad (3.1)$$

gdje je G_{SC} srednja vrijednost solarne iradijanse van atmosfere i iznosi 1366 W/m^2 .

Na osnovu vrijednosti iradijanse van atmosfere moguće je izračunati iradijansu van atmosfere svedenu na površinu paralelnu sa površinom Zemlje na određenoj lokaciji na sljedeći način:

$$G_{0H} = G_{SC} \left[1 + 0,034 \cdot \cos\left(\frac{360n}{365}\right) \right] [\cos L \cdot \cos \delta \cdot \cos \omega + \sin L \cdot \sin \delta], \quad (3.2)$$

gdje L predstavlja geografsku širinu posmatrane lokacije, δ predstavlja ugao deklinacije, odnosno ugao između ekvatorske ravni i prave koja prolazi kroz centre Sunca i Zemlje (Slika 3.3).



Slika 3.3: Ugao deklinacije

Veličina ω naziva se satni ugao i daje informaciju o tome za koji ugao Zemlja treba da se zarotira kako bi pozicija Sunca bila tačno iznad meridijana lokacije od interesa. Definiše se kao:

$$\omega = 15T_{dsp}, \quad (3.3)$$

gdje T_{dsp} predstavlja broj sati preostalih do solarnog podneva.

3.4 Modeli vedrog neba

Model vedrog neba za cilj ima procjenu globalne solarne iradijance na površini Zemlje pri uslovima vedrog neba, odnosno pri uslovima kada ne postoji oblačnost, a u svrhu određivanja indeksa vedrog neba. Ova vrijednost će se razlikovati od maksimalne moguće vrijednosti iradijance van atmosfere, jer i pri idealno vedrim uslovima postoji interakcija sunčevih zraka sa česticama atmosferskih gasova i određena količina rasipanja. Model vedrog neba uveliko zavisi od količine i vrste aerosola i vodene pare u atmosferi, i zapravo estimira maksimalnu moguću jačinu zračenja pri uslovima koji u datom trenutku vladaju u atmosferi.

Postoji više različitih modela vedrog neba. Najjednostavniji modeli, kao što je Daneshyar–Paltridge–Proctor (DPP) [26] [27] model, koriste isključivo podatak o uglu zenita, i relaciju da pri manjim uglovima zenita sunčev zrak prolazi kroz veću količinu vazdušne mase, pri čemu je rasipanje sunčevog zraka veće. Malo kompleksniji modeli koriste i osnovne podatke o atmosferskim uslovima kao što su temperatura, vlažnost vazduha, pritisak i količina aerosola. Primjeri ovakvih modela su Kastenov model [28] i Inieichen i Perez model [29]. Najkompleksniji modeli, kao što je Bird model [30], koriste složenije atmosferske parametre poput apsorpcije vodene pare, apsorpcije ozona, i apsorpcije uniformo miješanih gasova.

McClear model [31] predstavlja jedan od najnovijih modela vedrog neba koji pokazuje dobre rezultate pri različitim primjenama. Ovaj model koristi podatke o stanju atmosfere prikupljene u sklopu MACC projekta (eng. *Monitoring Atmosphere Composition and Climate*). MACC projekat bavi se nadzorom atmosferskih parametara koji govore o stanju aerosola, reaktivnih gasova i gasova staklene bašte. Ideja za ovakav model potiče upravo iz značajne korelacije koja postoji između modela vedrog neba i trenutnog stanja aerosola i vodene pare u atmosferi.

McClear model koristi look-up tabele zajedno sa interpolacijskim funkcijama kako bi na osnovu ulaznih parametara odredio model vedrog neba. Ulazne parametre McClear modela predstavljaju:

- ugao zenita,
- albedo površine Zemlje,

- ukupni sadržaj ozona i vodene pare u atmosferi,
- profil temperature, pritiska, gustine i procenta zapremine gasova u funkciji nadmorske visine,
- optička dubina aerosola,
- tip aerosola i Angstrom koeficijent koji govori o veličini čestice aerosola,
- geografska dužina, širina i nadmorska visina,
- vremenski period koji se posmatra,
- vremenska rezolucija modela.

U kontekstu formiranja modela vedrog neba indeks vedrine definiše se kao odnos globalne solarne iradijance pri uslovima vedrog neba i globalne iradijance van atmosfere. Na isti način može se definisati indeks vedrine za direktnu i difuznu komponentu pri uslovima vedrog neba, dijeljenjem svake od ovih veličina sa globalnom iradijansom van atmosfere. Look-up tabele McClear modela sadrže vrijednosti indeksa vedrine i indeksa vedrine direktne komponente za izabrane vrijednosti sljedećih ulaznih parametara: ugla zenita Sunca, nadmorske visine, elevacije u odnosu na horizontalnu površinu, albeda površine, sadržaja ozona i vodene pare, optičke dubine aerosola, Angstrom koeficijenta, tipa aerosola i profila temperature, pritiska, gustine i zapremine vazduha. Na osnovu interpolacijskih funkcija datih uz tabele, pri generisanju podataka, interpoliraju se vrijednosti indeksa vedrine za konkretne ulazne parametre. Nakon generisanja indeksa podaci se vraćaju o domen solarne iradijance množenjem sa vrijednostima globalne iradijance van atmosfere.

Korištenjem McClear modela dobijaju se precizni podaci o solarnoj iradijansi pri uslovima vedrog neba. Osnovna mana modela je vrijeme potrebno za generisanje look-up tabela, koje može potrajati nekoliko nedelja, dok je generisanje podataka na osnovu već postojećih modela prilično brz i jednostavan proces.

Glava 4

Algoritmi za generisanje sintetičkih podataka

Većina modela za dugoročno i kratkoročno planiranje rada solarne elektrane zahtijeva veliki broj kvalitetnih historijskih podataka o solarnoj iradijansi, kao i druge meteorološke podatke kao što su temperatura, vlažnost vazduha, brzina i pravac vjetra itd. Kao što je ranije pomenuto, sekvence solarne iradijanse dostupne su za samo određene lokacije. Takođe nije neuobičajeno da u izmjerenim podacima nedostaje čak do 40% vrijednosti kao posljedica grešaka u sistemu za prenos podataka ili šuma [32]. Jedan od osnovnih problema bilo kakvog tipa predikcije solarne iradijanse ili proizvodnje solarne elektrane leži u činjenici da je jako teško snimiti podatke za dovoljan broj godina kako bi ti podaci bili statistički značajni. Kako bi se otklonio ovaj problem i omogućili dugoročno i kratkoročno planiranje sa manjim brojem dostupnih podataka razvijeni su različiti modeli koji bi omogućili generisanje realističnih sintetičkih podataka o solarnoj iradijansi na osnovu poznavanja manjeg broja stvarnih, izmjerenih podataka.

Jedan od prvih pokušaja generisanja sintetičkih podataka koristio je, slično kao i kod problema predikcije, AR i ARMA modele. Ovakvi modeli nisu pokazali dobre rezultate kada je u pitanju preslikavanje raspodjele originalnih podataka, odnosno, ovakvi modeli ne garantuju da će sintetički podaci pripadati istoj raspodjeli kao i originalni. Još jedan veliki nedostatak ovakvih modela je taj da oni nisu u stanju da generišu veliku količinu sintetičkih podataka. Kao što se može vidjeti iz definicije AR i ARMA modela u prethodnom odjeljku, za generisanje jednog novog odmjerka potrebni su podaci o prethodnim vrijednostima odmjeraka što otežava generisanje veće količine podataka. Sem navedenih, tokom godina formirani su mnogi drugi, kompleksniji modeli za generisanje podataka o solarnoj iradijansi. U ovom odjeljku biće opisani neki od najznačajnijih takvih modela.

4.1 Markovljev lanac

Metoda zasnovana na korištenju Markovljevih lanaca je statistička metoda koja se koristi već duže vrijeme i dugo je predstavljala jedan od osnovnih načina za generisanje sekvenci solarne iradijanse.

Markovljev lanac ili Markovljev proces predstavlja stohastički model koji opisuje sekvencu mogućih događaja, pri čemu vjerovatnoća narednog događaja zavisi isključivo od trenutnog. Promjenu iz jednog stanja u drugo nazivamo tranzicija. Cilj ove metode je određivanje vjerovatnoća $P_{i,j}$ da će se desiti događaj i ukoliko mu je prethodio događaj j . Ove vjerovatnoće zapisuju se u matrice koje se nazivaju Markovljevim matricama tranzicija.

U [33] predloženo je korištenje Markovljevih matrica za generisanje prosječnih dnevnih vrijednosti indeksa vedrine K_t . Zaključeno je da dnevni prosjek indeksa vedrine ima veliku korelaciju isključivo sa vrijednošću ovog indeksa prethodnog dana, dok za sve vrijednosti koje su se desile prije toga korelacija drastično opada, čime je opravdano korištenje ove metode. Isto tako, ustanovljeno je da postoji značajna korelacija dnevnih vrijednosti indeksa vedrine sa mjesečnim prosjekom indeksa, i ova činjenica je iskorištena za pojednostavljenje modela, odnosno za smanjenje broja potrebnih matrica.

Kao prvi korak potencijalne vrijednosti indeksa vedrine podijeljene su na intervale. Markovljeve matrice formirane su tako da sadrže vjerovatnoće $P_{i,j}$ da će se indeks narednog dana nalaziti u određenom intervalu indeksiranom sa i ukoliko se trenutna vrijednost iradijansa nalazi u intervalu indeksiranom sa j . Prvobitna ideja je da se odgovarajuća matrica tranzicija formira za svaki dan u godini što bi rezultovalo sistemom od 365 matrica.

Kako bi se umanjila dimenzionalnost sistema, odnosno potreban broj matrica, iskorištena je činjenica da postoji korelacija između dnevnih vrijednosti indeksa vedrine i mjesečnog prosjeka indeksa. Potencijalne vrijednosti mjesečnog prosjeka takođe su podijeljene na intervale i za svaki od ovih intervala formira se odgovarajuća matrica tranzicija. Pri generisanju sintetičkih podataka potrebno je poznavanje mjesečnih prosjeka vrijednosti solarne iradijansa na osnovu kojih se bira odgovarajuća matrica tranzicija koja se potom koristi za generisanje dnevnih vrijednosti.

Prelazak sa dnevnih na satne vrijednosti najčešće se realizuje upotrebom AR modela prvog reda. Jedan od takvih modela predložen je u [34]. Uvedena je pretpostavka da satne vrijednosti indeksa vedrine k_t takođe predstavljaju Markovljev proces, odnosno da vrijednost satnog indeksa zavisi isključivo od vrijednosti prethodnog sata. Sem toga iskorištena je činjenica da postoji korelacija između satnih vrijednosti indeksa vedrine i dnevnog prosjeka. U tom slučaju, indeks vedrine za određeni sat u danu h za koji je poznat dnevni prosjek indeksa vedrine K_t moguće je odrediti kao:

$$k_t(h) = \Phi_1 k_t(h-1) + r, \quad (4.1)$$

gdje je Φ_1 koeficijent koji zavisi od vrijednosti K_t za posmatrani dan, i za svaki interval potencijalnih vrijednosti K_t određuje se odgovarajuća vrijednost ovog koeficijenta. Promjenljiva r je slučajna varijabla uzeta iz Gausove raspodjele nulte srednje vrijednosti i standardne devijacije $\sigma = (1 - \Phi_1^2)^{0.5}$

Određivanjem indeksa vedrine prelazak u domen solarne iradijancije je jednostavan i zahtijeva prosto množenje sa modelom iradijancije van atmosfere opisanom u Glavi 3.

4.2 Generativne suprostavljajuće mreže

Generativne suprostavljajuće mreže, GAN (eng. *Generative Adversarial Networks*) predstavljaju oblik nenadgledanog učenja gdje je cilj da model sam nauči određena pravila ili šablone na osnovu ulaznih podataka i da generiše sintetičke podatke za koje možemo da pretpostavimo da pripadaju istoj raspodjeli kao originalni podaci [35]. Sama ideja GAN-a potiče od teorije igara. GAN se sastoji od dvije mreže koje se treniraju istovremeno, i ukoliko ove mreže posmatramo kao igrače, cilj treniranja je dostizanje Nashovog ekvilibrijuma. Dva igrača su dostigli Nashov ekvilibrijum ukoliko je svaki od igrača donio najbolju moguću odluku, uzevši u obzir odluku protivnika. U suštini, svaka od mreža, pri procesu obučavanja, treba da dostigne minimum neke funkcije cijene sa parametrima koji su joj dostupni, a koji zavise od stanja u drugoj mreži.

Dvije mreže koje čine GAN nazivaju se generator i diskriminator. Cilj generatora je da od ulaznih slučajnih vektora, koji se uzimaju iz određene (najčešće normalne) raspodjele, generiše realistične podatke. Diskriminator najčešće predstavlja binarni klasifikator kome se predaju stvarni trening podaci i podaci generisani od strane generatora. Zadatak diskriminatora je da klasifikuje podatke koji su mu dati kao stvarne ili lažne. Generator nema nikakav pristup originalnim podacima i uči isključivo kroz interakciju sa diskriminatorom. Cilj istovremenog treniranja je da diskriminator postane dovoljno dobar u klasifikaciji podataka, ali da istovremeno generator formira realistične podatke, tako da ih diskriminator ne može prepoznati. U idealnom slučaju, tačnost istreniranog diskriminatora bi bila 50% što bi značilo da ne postoji razlika između generisanih i stvarnih podataka. Rad generatora G i diskriminatora D može biti sumiran kao:

$$G : G(z) \rightarrow R^{|x|}, \quad (4.2)$$

$$D : D(x) \rightarrow \{0,1\}, \quad (4.3)$$

odnosno generator može biti opisan kao preslikavanje iz prostora nasumičnih vektora z u prostor generisanih podataka x , dok diskriminator prostor podataka preslikava u skup čiji su elementi 0 i 1.

Veoma je značajno da podaci generisani GAN mrežom imaju istu raspodjelu kao originalni podaci, i ova osobina je ključna za dalju primjenu sintetičkih podataka. Prema tome, diskriminator se trenira tako da se maksimizuje njegova sposobnost klasifikacije, dok se generator trenira tako da se minimizuje razlika između raspodjele originalnih i generisanih podataka. Ovakav rad generatora i diskriminatora može biti sumiran kao:

$$\max_D \min_G V(G,D), \quad (4.4)$$

pri čemu $V(G,D)$ predstavlja funkciju cijene:

$$V(G,D) = E_{p_{data}(x)} \log D(x) + E_{p_g} \log(1 - D(G(z))), \quad (4.5)$$

gdje oznaka E predstavlja matematičko očekivanje, a p_{data} i p_g funkcije raspodjele stvarnih i generisanih podataka respektivno.

Podaci koje GAN generiše potiču od nasumičnih vektora i korisnik nema nikakvu kontrolu nad njima. Ubrzo nakon nastanka GAN-a on je proširen, tako da je omogućeno generisanje podataka koji ispunjavaju neki uslov definisan od strane korisnika i na ovaj način nastaje uslovni GAN [36]. Neka je dodatni uslov koji je zadat y , pri čemu y može biti bilo kakva dodatna informacija, kao što je npr. oznaka klase. U tom slučaju, dopuna GAN-a na uslovnu verziju vrši se dodavanjem dodatnih ulaznih slojeva generatoru i diskriminatoru na koje će se dovoditi y . Funkcija cijene u tom slučaju postaje:

$$V(G,D) = E_{p_{data}(x|y)} \log D(x) + E_{p_g} \log(1 - D(G(z)|y)). \quad (4.6)$$

Prva upotreba GAN-ova bila je generisanje realističnih slika [37] [38], dok danas GAN mreže imaju različite upotrebe, kao što su obrada jezika i teksta [39], generisanje muzike [40] i u skorije vrijeme generisanje sintetičkih vremenskih nizova kao što finansijski [41], medicinski [42] i energetske podaci [43]. U nastavku će biti opisani neki od najnovijih i najznačajnijih GAN algoritama za generisanje vremenskih sekvenci.

4.2.1 Kontinualna rekurentna generativna suprostavljajuća mreža

Kontinualna rekurentna generativna suprostavljajuća mreža, C-RNN-GAN (eng. *Continuous Recurrent Neural Network Generative Adversarial Network*) je generativni model za generisanje sekvencijalnih kontinualnih podataka koji je nastao u svrhu generisanja muzike [40]. Prije nastanka ovog modela muzika je generisana korištenjem različitih simboličkih reprezentacija melodije kao što je npr. diskretan niz tonova. Za razliku od takvih modela C-RNN-GAN muziku generiše kao kontinualan vremenski niz, pri čemu je svaki odmjerak predstavljen grupom obilježja, konkretno vrijednostima frekvencije, dužine, intenziteta tona i vremena koje je proteklo od početka prethodnog tona do početka narednog. Ovakav način reprezentacije muzike dozvoljava polifoniju, odnosno pojavljivanje više tonova istovremeno. S obzirom na ovakav način generisanja, gdje je muzika predstavljena kao kontinualni sekvencijalni niz, C-RNN-GAN je povoljan i za rad sa različitim tipovima vremenskih sekvenci.

Prema tipičnoj strukturi GAN-a, C-RNN-GAN sastoji se od generatora i diskriminatora, uz uvedene izmjene koje odgovaraju konkretnom problemu generisanja vremenskih nizova. Struktura C-RNN-GAN-a prikazana je na Slici 4.1. Generator je realizovan kao LSTM mreža čiji ulaz predstavlja nasumični vektor konkateniran sa izlazom generatora

u prethodnom vremenskom koraku. Diskriminator C-RNN-GAN-a je realizovan kao bidirekciona LSTM mreža. Bidirekzione LSTM mreže sastoje se od dvije LSTM mreže od kojih jedna obrađuje ulazni vektor u smjeru unaprijed a druga u smjeru unazad. Ovakav pristup je čest u jezičkim modelima i logika iza ovog pristupa je da je ovakav model u stanju da bolje nauči zavisnosti unutar samih sekvenci, što ga čini povoljnim za upotrebu u vremenskim nizovima.

Još jedna modifikacija u odnosu na standardni GAN je funkcija cijene. Jedan od glavnih ciljeva je da sintetički podaci imaju raspodjelu koja je što približnija raspodjeli stvarnih podataka. Kako bi se ovo postiglo, funkcija cijene generatora se minimizuje tako da reprezentacija sintetičkih podataka u posljednjem sloju diskriminatora što bolje odgovara reprezentaciji stvarnih podataka. Nova funkcija cijene može se opisati jednačinom

$$L_G = \frac{1}{m} \sum_{i=1}^m (R(x^{(i)}) - R(G(z^{(i)})))^2, \quad (4.7)$$

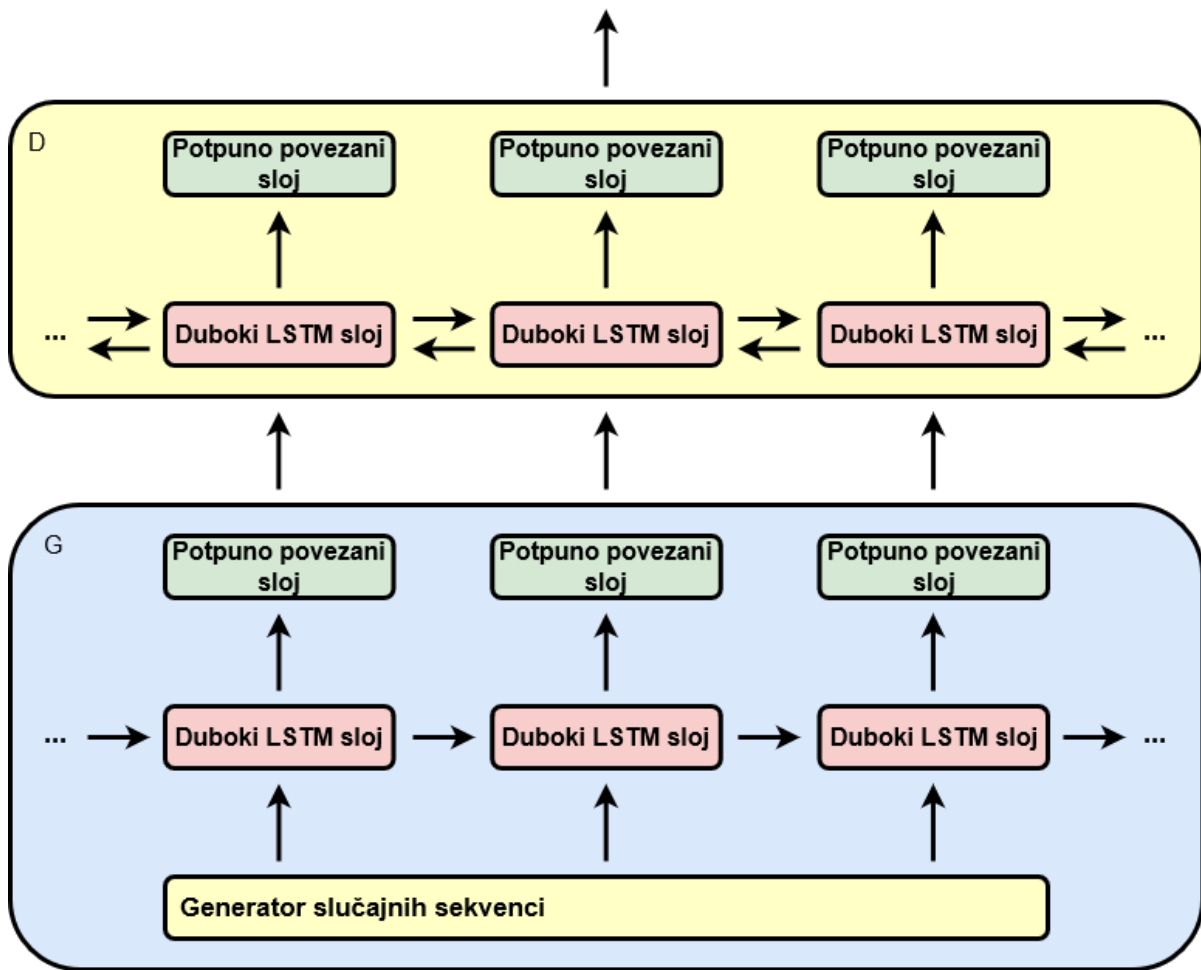
gdje oznaka R označava reprezentaciju podataka, x predstavlja stvarni niz podataka, a $G(z)$ podatke generisane od strane generatora.

Radi ostvarivanja boljih performansi pri treniranju ove mreže u originalnom radu uvedeno je nekoliko mehanizama. Na generator i diskriminator primijenjena je L2 regularizacija. Prije početka samog treniranja, model je pretreniran korištenjem kraćih sekvenci podataka, nakon čega se dužina sekvenci koja se predaje mreži postepeno povećava. Takođe se primijenjuje zamrzavanje diskriminatora. Često se dešava da diskriminator postane dominantan, tako da se gradijent ne može iskoristiti za adekvatno ažuriranje generatora. U tom slučaju zaustavlja se ažuriranje diskriminatora i ažurira se samo generator.

4.2.2 Rekurentna uslovna generativna suprostavljajuća mreža

Rekurentni GAN, RGAN (eng. *Recurrent Generative Adversarial Network*) i rekurentni uslovni GAN, RCGAN (eng. *Recurrent Conditional Generative Adversarial Network*) [42] predstavljaju sljedeći značajniji pokušaj generisanja kontinualnih vremenskih podataka uz pomoć GAN arhitekture. Ovaj model formiran je u svrhu generisanja medicinskih podataka koji su privatni i teško dostupni, a u svrhu obučavanja zaposlenih na odjelima intenzivne njege.

Arhitekture generatora i diskriminatora RGAN-a predstavljaju rekurentne LSTM mreže. U odnosu na C-RNN-GAN, RGAN ne koristi bidirekcionu diskriminator, i izlazi generatora se ne koriste kao ulazi u narednom koraku. Sem toga uvedena je i uslovna varijanta ove mreže koja omogućava da se generišu podaci koji ispunjavaju neki dodatni uslov, pri čemu se uslov dovodi na ulaze mreže pri svakom vremenskom koraku. Za generisanje medicinskih podataka, dodatni uslov je informacija o tome da li stanje pacijenta postaje kritično nakon 4 sata provedena na odjelu intenzivne njege.



Slika 4.1: C-RNN-GAN

4.2.3 Vremenska generativna suprostavljajuća mreža

Model za generisanje vremenskih nizova treba da bude u stanju da generiše takve podatke da u svakoj tački niza postoji odgovarajuća zavisnost između pojedinih obilježja, ali da sem toga bude ispunjena i vremenska zavisnost između pojedinih tačaka. Prethodno korišteni nenadgledani generativni modeli pri obučavanju ne uzimaju u obzir vremensku zavisnost niza kao cjeline. Sa druge strane, autoregresivni modeli, koji mogu da modeluju vremensku zavisnost niza nisu istinski generativni, jer nisu u stanju da generišu sintetičke podatke bez dodatnih, eksternih podataka. Vremenska generativna suprostavljajuća mreža, TimeGAN [44] pokušava da objedini ove dvije osobine. Ova mreža treba da bude u stanju da generiše sintetičke podatke bez eksternog uslovljavanja, na osnovu nasumičnih vektora, što je tipično za GAN mreže, ali tako da je sačuvana vremenska dinamika originalnog niza korištenog za obučavanje modela.

TimeGAN mreža sastoji se od četiri dijela, kao što je prikazano na Slici 4.2. Sem generatora i diskriminatora tipičnih za GAN arhitekturu, TimeGAN sadrži još tzv. *embedding* i *recovery* dijelove. Ova dva dijela zadužena su za kodovanje stvarnih podataka i podataka generisanih od strane generatora iz prostora obilježja u niskodimenzioni prostor kodova-

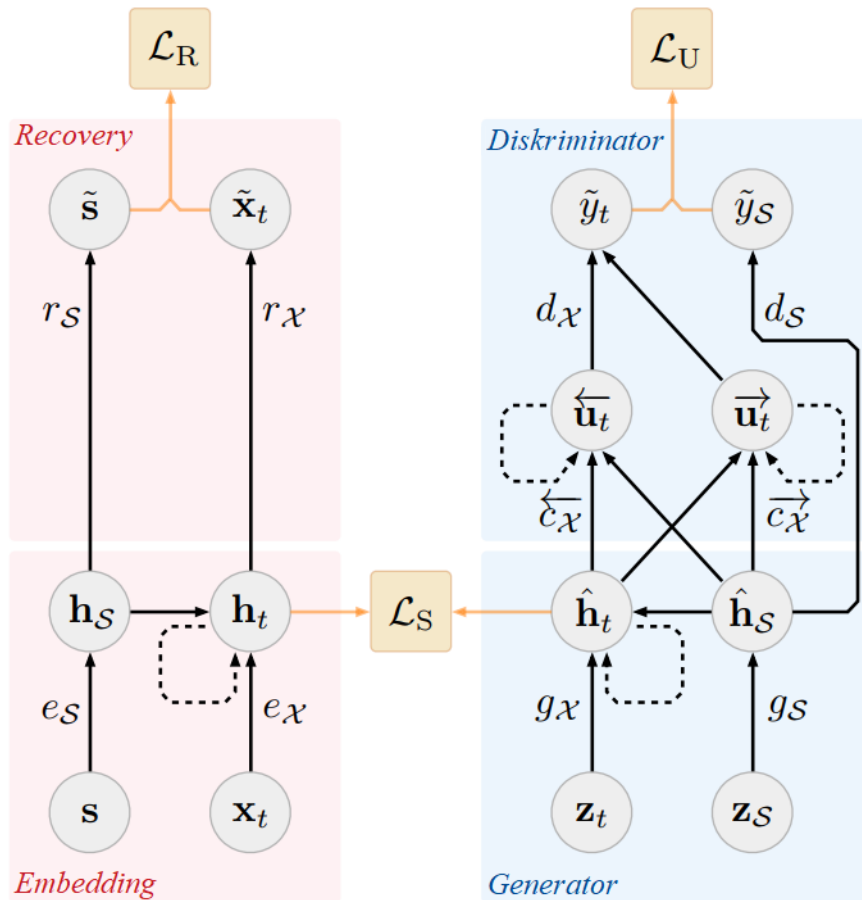
nih reprezentacija podataka koji nazivamo latentnim prostorom, i prevođenje nazad u prostor obilježja. Na ovaj način mreža je u stanju da iz niskodimenzionih reprezentacija uči vremenske zavisnosti unutar samih vremenskih nizova. *Embedding* i *recovery* dijelovi mreže obučavaju se istovremeno sa generatorom i diskriminatorom.

Princip rada TimeGAN mreže takođe je ilustrovan na Slici 4.2. Neka se u opštem slučaju vremenski niz sastoji od dva tipa obilježja, statičkih s , koji ne zavise od vremena, i dinamičkih x_t koji su vremenski zavisni. Zadatak *embedding* dijela je da koduje stvarne statičke i vremenske podatke u njihove reprezentacije h_s i h_t

$$h_s = e_s(s), h_t = e_x(h_s, h_{t-1}, x_t), \quad (4.8)$$

gdje su e_s i e_x *embedding* dijelovi za statička i dinamička obilježja. *Embedding* je realizovan kao rekurentna mreža. Slično tome, zadatak *recovery* dijela r je da koduje reprezentacije iz latentnog prostora nazad u prostor obilježja, i takođe je realizovan kao rekurentna mreža:

$$\tilde{s} = r_s(h_s), \tilde{x}_t = r_x(h_t). \quad (4.9)$$



Slika 4.2: TimeGAN

Zadatak generatora je generisanje sintetičkih reprezentacija statičkih i dinamičkih obilježja \hat{h}_s i \hat{h}_t , na osnovu ulaznih vektora z_t i z_s :

$$\hat{h}_s = g_S(z_s), \hat{h}_t = g_\chi(\hat{h}_s, \hat{h}_{t-1}, z_t), \quad (4.10)$$

gdje su g_S i g_χ generatori statičkih i dinamičkih obilježja respektivno. Konačno, diskriminator uzima reprezentacije stvarnih podataka i generisane reprezentacije i klasifikuje ih kao stvarne i lažne, pri čemu su izlazi diskriminatora statičkih i dinamičkih obilježja označeni kao \tilde{y}_s i \tilde{y}_t . Diskriminator je realizovan kao bidirekciona rekurentna mreža i ukoliko sa $\vec{u}(t)$ i $\overleftarrow{u}(t)$ označimo izlaze skrivenog sloja diskriminatora unaprijed i unazad, a sa \tilde{h}_s i \tilde{h}_t reprezentacije obilježja, nezavisno od toga da li odgovaraju stvarnim ili generisanim podacima, možemo predstaviti izlaz diskriminatora kao:

$$\tilde{y}_s = d_S(\tilde{h}_s), \tilde{y}_t = d_\chi(\vec{u}(t), \overleftarrow{u}(t)). \quad (4.11)$$

Princip treniranja TimeGAN mreže svodi se na minimizaciju tri različite funkcije cijene. Kao prvo, potrebno je obučiti *embedding* i *recovery* dijelove da vrše preciznu rekonstrukciju kodovanih podataka h_s i $h_{1:T}$ iz latentnog prostora nazad u prostor obilježja, odnosno potrebno je osigurati invertibilnost funkcije kodovanja. U tu svrhu definiše se rekonstrukcijska funkcija cijene:

$$\mathcal{L}_R = E_{S, X_{1:T} \sim P} \left[\|s - \tilde{s}\|_2 + \sum_t \|x_t - \tilde{x}_t\|_2 \right]. \quad (4.12)$$

Generator naizmjenično prihvata dva tipa ulaza na osnovu kojih generiše nove reprezentacije sintetičkih podataka. Prvi tip ulaza koji se dovodi na generator su reprezentacije generisane u prethodnom vremenskom koraku \tilde{h}_s i \tilde{h}_{t-1} . U slučaju ovakvog ulaza minimizuje se standardna funkcija cijene GAN-a opisana jednačinom 4.5, koju tada nazivamo nenadgledana funkcija cijene i označavamo sa \mathcal{L}_U . S obzirom na to da povratna informacija diskriminatora nije dovoljna kako bi generator naučio vremenske zavisnosti prisutne u originalnim podacima, kao drugi tip ulaza generatoru se dovode reprezentacije stvarnih podataka. Tada se vrši minimizacija nove funkcije cijene koju nazivamo nadgledana funkcija, a koja je definisana kao:

$$\mathcal{L}_S = E_{S, X_{1:T} \sim P} \left[\sum_t \|h_t - g_\chi(h_s, h_{t-1}, z_t)\|_2 \right]. \quad (4.13)$$

Na ovaj način postizemo poređenje reprezentacije stvarnog vektora i reprezentacije vektora generisanog od strane generatora u svakom vremenskom koraku. Ukratko, nenadgledana funkcija cijene obezbjeđuje generisanje realističnih podataka, dok nadgledana funkcija cijene omogućuje da generisani podaci imaju odgovarajuću vremensku zavisnost.

4.2.4 Uslovna Sig-Wasserstein generativna suprostavljajuća mreža

Rad GAN mreže moguće je posmatrati kao preslikavanje podataka iz raspodjele nasumičnih podataka koji se dovode na ulaz generatora μ u raspodjelu koja je u idealnom slučaju jednaka raspodjeli trening podataka ν . Kada su u pitanju vremenske sekvence, koje su često nestacionarne i predstavljaju dugačak tok podataka, od ukupne raspodjele mnogo su značajnije uslovne raspodjele $\nu(x_{t+1:t+q}|x_{t-p:t})$ koje suštinski opisuju vremensku zavisnost niza x . Dugačak priliv podataka, odnosno dugačke vremenske sekvence mogu da značajno povećaju dimenzionalnost generativnog modela, pogotovo ukoliko se uzima u obzir navedene uslovne raspodjele. Uslovni Sig-Wasserstein GAN [45] objedinjuje GAN arhitekturu sa matematičkom ekstrakcijom obilježja, pod nazivom potpis putanje, koja omogućava smanjenje dimenzionalnosti originalnih podataka. U svrhu smanjenja dimenzionalnosti osmišljena je metrika, nazvana Sig-W1, koja se koristi kao funkcija cijene diskriminatora. Ova metrika uzima informaciju o vremenskim zavisnostima nizova i predstavlja ih na apstraktan i univerzalan način (kao tzv. potpis putanje). Sem smanjenja dimenzionalnosti, ova metrika omogućava ekstrakciju informacija o uslovnim raspodjelama originalnog vremenskog niza. Takođe, predložena je nova arhitektura generatora koja je u stanju da nauči i uspješno generiše niz sa autoregresivnom prirodom originalnog niza.

Ukoliko posmatramo vremensku sekvencu, odnosno putanju X na vremenskom intervalu J , tada možemo označiti potpis putanje X kao $S(X_J)$. Radi smanjenja dimenzionalnosti definiše se skraćeni potpis putanje $S_M(X_J)$ u prostoru manje dimenzionalnosti M . Funkcija cijene diskriminatora, zasnovana na novoj Sig-W1 metrici, u skladu sa prethodno uvedenim oznakama, u tom slučaju definisana je kao:

$$L(\Theta) = \sum_t |E_\mu [S_M(X_{t+1:t+q})|X_{t-p+1:t}] - E_\nu [S_M(\hat{X}_{t+1:t+q}^{(t)})|X_{t-p+1:t}]|, \quad (4.14)$$

gdje \hat{X} označava putanju generisanu od strane generatora. U [45] predloženi su mehanizmi za estimaciju očekivanja E_μ i E_ν prije izračunavanja Sig-W1 metrike. Predloženi mehanizmi su linearna regresija za estimaciju E_ν i Monte Carlo metoda za estimaciju E_μ .

U [45] takođe je predstavljena nova arhitektura generatora, pod nazivom uslovna, autoregresivna, propagirajuća neuronska mreža, AR-FNN (eng. *Autoregressive Feed-Forward Neural Network*). Generator je realizovan kao mreža bez povratnih veza, sa rezidualnim vezama i ReLU aktivacionom funkcijom. Mreža koristi putanju u prethodnom vremenskom koraku $X_{t-p+1:t} = x$ i nasumični vektor Z_{t+1} da u narednom koraku generiše varijablu čija je uslovna vjerovatnoća što približnija $P(X_{t+1}|X_{t-p+1:t} = x)$. Ovakva arhitektura omogućava učenje autoregresivne prirode dugačkih vremenskih nizova i očuvanje vremenske zavisnosti originalnog niza.

4.3 Maskirani autoenkoder

Maskirani autoenkoder, MAE (eng. *Masked autoencoder*) proistekao je iz enkodera za uklanjanje šuma [46] i spada u grupu algoritama nenadgledanog obučavanja. Prva značajna upotreba maskiranih autoenkodera je BERT (eng. *Bidirectional Encoder Representations from Transformers*) model reprezentacije jezika [47]. BERT je korišten kao način pretreniranja jezičkih modela. Sam model sastoji se od tri dijela, pri čemu je prvi dio zadužen za maskiranje. On maskira nasumične dijelove podataka, i za svaku masku formira se odgovarajući token. Ovakvi podaci se potom predaju enkoderu, koji ujedno predstavlja drugi dio. Enkoder koduje podatke i preslikava ih u latentni prostor. Treći dio je dekoder, čiji je osnovni zadatak da na osnovu konteksta nemaskiranih dijelova podataka dopuni maskirane segmente i potom sve prevede nazad u originalni prostor podataka. Kada je jezički model u pitanju enkoder je realizovan kao bidirekcionni transformator.

Nakon primjene u jezičkim modelima, maskirani autoenkoderi pronalaze primjenu u oblasti računarskog vida, takođe u svrhu pretreniranja modela [48]. Konvolucione neuronske mreže, koje se tipično koriste u radu sa slikama, mijenjaju se transformatorima i maskiraju se nasumični dijelovi slike. Enkoder koduje nemaskirane dijelove slike, dok ih dekoder dopunjava i vraća u originalni prostor. S obzirom da je dokazano da algoritam radi bolje pri maskiranju većeg dijela slike, on ne samo da je efikasan u svrhu pretreniranja, već i smanjuje potrebnu procesorsku moć, s obzirom na to da u tom slučaju enkoder radi samo sa malim procentom slike.

Ovakav model proširen je tako da radi sa videom [49]. Primjena maskiranih autoenkodera na podatke kao što su jezički modeli i video, u kojima postoji sekvencijalna zavisnost znači da se ovakav model može proširiti i na upotrebu na vremenski zavisnim podacima, odnosno na bazama podataka vremenskih nizova kao što su podaci o solarnoj iradijansi.

U [50] predložena je jedna MAE arhitektura za generisanje vremenskih nizova nazvana ExtraMAE koja sem enkodera i dekodera sadrži dodatni blok, ekstrapolator. Princip rada je sljedeći:

- Originalni vremenski niz dijeli se na niz kraćih vremenskih sekvenci.
- Vrš se isjecanje svakog niza na segmente određene dužine, tako da između njih ne postoji preklapanje. Dužina segmenta, koju nazivamo i veličina maske, predstavlja hiperparametar.
- Vrš se maskiranje određenog broja ovih segmenata. Maskirani segmenti se uklanjaju iz niza, dok se ostali segmenti konkatenuiraju. Pamte se indeksi maskiranih i nemaskiranih segmenata. Broj maskiranih segmenata, koji nazivamo i broj maski takođe predstavlja hiperparametar.
- Enkoder radi isključivo sa nemaskiranim segmentima vremenskog niza, i njegov zadatak je da odredi latentnu reprezentaciju ovih segmenata.
- Ekstrapolator predviđa latentne reprezentacije maskiranih dijelova niza na osnovu reprezentacija nemaskiranih segmenata.

- Dekoder rekonstruiše signal iz latentne reprezentacije nazad u domen vremenskog niza.

Dodatna izmjena ExtraMAE modela u odnosu na prethodne maskirane autoenkodere je realizacija enkodera i dekodera u vidu rekurentnih neuronskih mreža. U poređenju sa transformatorima koji su prethodno korišteni, obučavanje ovakvih mreža mnogo je brže i zahtijeva manje računarske moći. Ekstrapolator je realizovan kao potpuno povezan sloj. Za obučavanje ExtraMAE modela, kao funkcija cijene definisana je srednjekvadratna greška između rekonstruisane vremenske sekvence \hat{X} i stvarne vremenske sekvence X :

$$L_{recon} = E_X ||X - \hat{X}||^2, \quad (4.15)$$

pri čemu se funkcija cijene računa za cijelu vremensku sekvencu, ne samo za maskirani segment.

Dodavanje ekstrapolatora omogućava razdvajanje dva zadatka koja ima dekodeer, dopunjavanje maskiranih dijelova i vraćanje u domen vremenskog niza. Ovakva arhitektura maskiranog autoenkodera povoljna je za kopiranje vremenskih zavisnosti originalnog niza. Takođe, u odnosu na generativne modele, maskirani autoenkoder ima jednostavniju arhitekturu i njegovo obučavanje traje značajno kraće.

Glava 5

Metodologija

5.1 Baza podataka o solarnoj iradijansi

U svrhu treniranja modela potrebni su izmjereni podaci o solarnoj iradijansi za lokaciju od interesa. Podaci treba da budu snimljeni tokom dovoljnog broja godina kako bi baza podataka sadržala sve potencijalne scenarije koji se mogu desiti tokom dužeg vremenskog perioda. U nastavku će biti opisan način mjerenja i obrada podataka korištenih za treniranje i testiranje modela.

5.1.1 Meteorološka stanica Davis Vantage Pro 2

Mjerenje podataka o solarnoj iradijansi korištenih za obučavanje modela vršeno je pomoću meteorološke stanice Davis Vantage Pro 2. Meteorološka stanica predstavlja integrisano okruženje koje se sastoji od senzora koji prikupljaju podatke o vremenskim parametrima. Podaci se sa stanice šalju ka konzoli putem kabla, ili radio veze ukoliko je u pitanju bežična verzija stanice. Na Slici 5.1 prikazan je primjer bežične stanice i konzole.

Kada su senzori u pitanju, stanica je opremljena kolektorom kišnice, temperaturnim sensorom, sensorom vlage, sensorima sunčevog i UV zračenja i anemometrom. Na osnovu mjerenja stanica daje podatke o velikom broju parametara kao što su: prosječna, maksimalna i minimalna temperatura; vlažnost vazduha; prosječna i maksimalna brzina vjetra i smjer vjetra; prisustvo kiše i količina padavina; indeks UV zračenja; energija sunčevog zračenja; HDD (eng. *Heating Degree Days*) indeks koji govori o količini energije potrebnoj za grijanje za vrijeme sezone grijanja; CDD (eng. *Cooling Degree Days*) indeks koji govori o količini energije potrebnoj za hlađenje tokom sezone hlađenja; THW (eng. *Temperature-Humidity-Wind*) i THSW (eng. *Temperature-Humidity-Sun-Wind*) indeksi koji kvantifikuju subjektivni osjećaj temperature uzimajući u obzir i vlažnost vazduha, sunčevo zračenje i vjetar. Bežična verzija stanice sadrži i solarni panel koji vrši napajanje stanice u kombinaciji sa baterijama kao rezervom napajanja.

Podaci o solarnoj iradijansi koji su korišteni za obučavanje modela snimljeni su na području Banjaluke. Meteorološka stanica montirana je na krovu Akademije nauka i umjet-



Slika 5.1: Meteorološka stanica Davis Vantage Pro 2

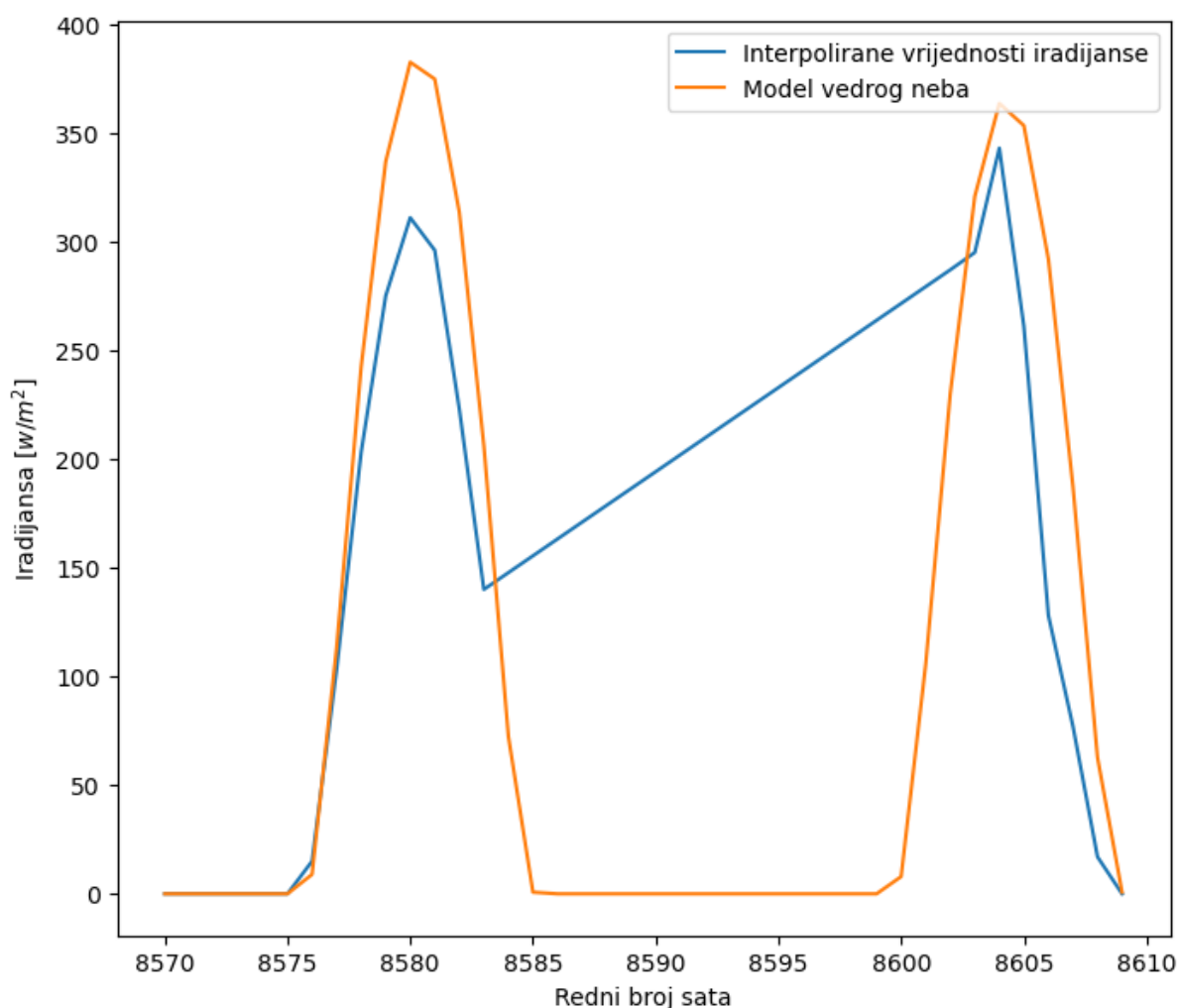
nosti Republike Srpske, smještene na koordinatama 44.7723 °N, 17.1935 °E. Podaci su mjereni na svakih pola sata u vremenskom periodu od 6.12.2012. godine u 18:30 časova, do 29.8.2018. godine u 15 časova, prema lokalnom vremenu, sa uzimanjem u obzir ljetnjeg i zimskog računanja vremena.

5.1.2 Priprema podataka

Zbog kvarova na meteorološkoj stanici, koji su se u periodu 2017. godine desili više puta, mjerenja vršena u tom periodu u velikom broju tačaka nisu validna ili ne postoje. Prema tome, mjerenja vršena u ovom vremenskom periodu su odbačena i nisu upotrebljena za treniranje modela. U preostalom vremenskom intervalu takođe postoje određene tačke u kojima mjerenja nedostaju, zbog kvarova ili prestanka napajanja stanice. Da bi se popunili nedostajući podaci, na cijelu bazu podataka primijenjena je linearna interpolacija. Odabrano je da generatori sintetičkih sekvenci generišu podatke na satnom nivou, te su u skladu s tim korištene samo tačke izmjerene na nivou jednog sata. Mjerenja vršena između satnih vrijednosti iskorištena su za precizniju interpolaciju tačaka koje nedostaju, nakon čega su ova mjerenja uklonjena iz baze. Finalna baza podataka se sastoji od svih mjerenja na satnom nivou u periodu od 1.1.2013. godine u 00:00 časova do 31.12.2016. godine u 23:00 časa.

Radi uklanjanja determinističke komponente iz podataka o solarnoj iradijansi odlučeno je da se na osnovu njih odredi indeks vedrog neba i da se takvi podaci predaju neu-

ronskim mrežama. Za proračun ovog indeksa na osnovu izmjerenih podataka potreban je model vedrog neba. U ovu svrhu korišten je McClear model opisan u Glavi 3. Nakon dijeljenja originalnih podataka vrijednostima modela vedrog neba pojavljuju se određene vrijednosti koje se teorijski ne bi mogle desiti, kao što je npr. indeks vedrog neba u vrijednosti nekoliko stotina. Jedan od potencijalnih razloga je interpolacija u dijelovima baze podataka u kojima nedostaje veliki broj uzastopnih tačaka, gdje interpolacija daje nerealistične vrijednosti. Jedan od takvih slučajeva prikazan je na Slici 5.2. Interpolacija podataka izvršena je između dvije tačke koje su dostupne u bazi podataka, a između kojih postoji velik broj tačaka koje nisu izmjerene. Primjenom interpolacije dobijaju se nenulte vrijednosti u noćnim satima, kao i jako velike vrijednosti iradijanse u satima u kojima se dešava izlazak i zalazak Sunca, što fizički nije moguće. Dijeljenjem ovih vrijednosti sa modelom vedrog neba, dobijaju se vrijednosti indeksa vedrog neba koje su značajno veće od 1.



Slika 5.2: Primjer nevalidnih podataka u bazi

Iz ovog razloga, sve vrijednosti indeksa koje su veće od vrijednosti 1,1 postavljene su na ovu vrijednost koja je usvojena kao maksimalna.

Za obučavanje mreža nisu potrebne tačke snimljene u noćnom periodu, kada solarna iradijansa ima vrijednost jednaku nuli. Ove vrijednosti su uklonjene, uklanjajući svih tačaka u kojima model vedrog neba ima nultu vrijednost. Ovakav scenario moguć je isključivo noću, jer se pri uslovima vedrog dana ne može desiti da iradijansa poprimi nultu vrijednost tokom dnevnih sati.

Pored indeksa vedrog neba, neuronske mreže za generisanje sintetičkih podataka zamišljene su tako da generišu i dodatne podatke koji bi mogli biti korelisani sa jačinom sunčevog zračenja. Izabrani parametri su redni broj dana u godini u kom je vršeno mjerenje, temperatura, vlažnost vazduha, podatak o postojanju padavina i brzina vjetera. Baza podataka je normalizovana tako što je na svaki od parametara $X^{(k)}$ primijenjena MinMax normalizacija definisana kao:

$$x^{(k)} = \frac{X^{(k)} - X_{min}^{(k)}}{X_{max}^{(k)} - X_{min}^{(k)}}, \quad (5.1)$$

gdje $x^{(k)}$ predstavlja normalizovanu vrijednost parametra a $X_{min}^{(k)}$ i $X_{max}^{(k)}$ minimalnu i maksimalnu vrijednost parametra u bazi podataka, respektivno.

Kao posljednji korak baza je podijeljena na trening i testni skup. Posljednjih 6 mjeseci 2016. godine određeno je kao testni skup kako bi se rezultati mogli testirati na širokom spektru vremenskih uslova. S obzirom na to da su iz baze podataka izbačene sve tačke izmjerene u toku noći, testni skup obuhvata 2350 tačaka. Ostatak podataka, koje čini 16708 tačaka, koristi se kao trening skup.

5.2 Obučavanje modela

U svrhu generisanja sintetičkih podataka o solarnoj iradijansi obučena su četiri modela. Prvi model koji ćemo nazvati LSTM-GAN inspirisan je arhitekturom RGAN-a. LSTM-GAN sastoji se od po jednog LSTM sloja u diskriminatoru i generatoru i kao funkcija cijene odabrana je standardna funkcija cijene GAN-a opisana jednačinom 4.5. Njegov princip rada i algoritam obučavanja takođe su identični algoritmu standardnog GAN-a koji je opisan u Glavi 4. Ostala tri modela koji su korišteni su C-RNN-GAN, ExtraMAE i TimeGAN čiji su principi rada takođe opisani u Glavi 4.

Svakom od modela predaju se vremenske sekvence određene dužine dobijene isjecanjem baze podataka. Dužina sekvence bira se kao hiperparametar. S tim u vezi, i izlaz svakog od modela predstavljaće vremensku sekvencu iste dužine kao i ulaz. Isjecanje sekvenci vrši se tako da između njih postoji preklapanje, odnosno susjedne sekvence razlikuju se isključivo za po jedan odmjerak. Na ovaj način dobijaju se sekvence koje imaju neophodnu vremensku zavisnost, dok se istovremeno povećava broj sekvenci koje se mogu iskoristiti za trening modela. Performanse mreže zavise od dužine sekvence. Radi poređenja performansi odabrane su dvije različite dužine, 7 i 25 odmjeraka.

LSTM-GAN treniran je u grupama od 64, na 3000 epoha. Odabrani koeficijent učenja je 0,001. Dimenzija nasumičnog vektora koji predstavlja ulaz generatora je 5.

Treniranje C-RNN-GAN modela vrši se na 100 epoha u grupama od 16. Koeficijent učenja je jednak za generator i diskriminator i iznosi 0,0001. Pri treniranju ovog modela primijećen je problem koji pravi velika razlika između funkcija cijene generatora i diskriminatora. Funkcija cijene diskriminatora poprima vrijednost koja je mnogo manja od funkcije cijene generatora, što onemogućava odgovarajuće ažuriranje generatora. Kako bi se ovaj efekat umanjio primijenjeno je zamrzavanje, pri čemu se ažuriranje diskriminatora zaustavlja kada je njegova tačnost veća od 70%.

ExtraMAE model treniran je tako da su veličina i broj maski postavljen na vrijednost 1. Enkoder i ekstrapolator su prvo odvojeno trenirani na po 5000 iteracija nakon čega cjelokupni model prolazi 50000 iteracija na grupama veličine 64.

Treniranje TimeGAN modela se vrši u tri dijela. Prvo se vrši treniranje samog *embedding* bloka, nakon čega se mreža trenira uzimanjem u obzir samo nadgledane funkcije cijene, i kao posljednji korak vrši se treniranje mreže kao cjeline. Što se tiče parametara mreže, generator i diskriminator imaju po tri sloja, sa po 24 čvora u svakom sloju. Treniranje se vrši u grupama veličine 64. Prva dva dijela treniranja vrše se na po 1000 iteracija, dok se treniranje cijele mreže vrši na 10000 iteracija.

5.3 Validacija sintetičkih podataka

Nakon što su sintetički podaci generisani, potrebno je prije upotrebe na neki način provjeriti njihovu validnost. U literaturi, mnogi pristupi se zasnivaju na provjeri da li originalni i sintetički podaci potiču iz iste raspodjele, korištenjem metrika kao što su Kolmogorov-Smirnov test i Kullback-Leibler divergencija. Osnovni nedostatak ovakvih metoda je velika osjetljivost čak i na male razlike u raspodjelama [51]. Za sada još uvijek ne postoji standardna procedura prema kojoj se provjerava validnost sintetičkih podataka, stoga je teško odabrati metrike koje treba da pokazuju kvalitet generatora sintetičkih podataka.

Jedna od najznačajnijih provjera je da li su generisani podaci upotrebljivi kada su u pitanju praktične primjene. Na početku rada pomenute su dvije osnovne primjene sintetičkih podataka o solarnoj iradijansi, a to su dugoročna i kratkoročna predikcija rada elektrane. Potrebno je na neki način, sem provjere raspodjele podataka, provjeriti performanse na ovim primjenama, odnosno uporediti performanse alata za planiranje, u slučajevima kada su njihovi ulazni podaci stvarne, izmjerene vrijednosti, i kada su ulazni podaci sintetički.

Prije testiranja podataka uz pomoć alata za planiranje vrši se vizuelna provjera kvaliteta sintetičkih podataka, pri čemu se za vizuelizaciju podataka u niskodimenzionom prostoru koriste tSNE i PCA metode koje će biti opisane u nastavku. Potom, kako bi se izvršila validacija upotrebljivosti podataka, testira se njihovo ponašanje na problemima kratkoročne predikcije i planiranja. Kratkoročna predikcija testira se obučavanjem različitih prediktora korištenjem sintetičkih podataka, a potom testiranjem na predikciji stvarnih testnih podataka. Dugoročno planiranje testira se upotrebom sintetičkih podataka za proračunavanje energije koju solarno postrojenje može da generiše na dužem vremenskom periodu.

5.3.1 tSNE metoda

tSNE [52] predstavlja način vizuelizacije podataka sa više varijabli, tako što se svakoj tački dodjeljuje mjesto u dvodimenzionalnom ili trodimenzionalnom prostoru. Osnovni cilj je smanjenje dimenzionalnosti podataka, ali tako da se očuva što više njihove strukture u niskodimenzionalnoj mapi.

Metoda je nastala iz već postojeće SNE metode (eng. *Stochastic Neighbour Embedding*) [53]. Prethodne metode za smanjenje dimenzionalnosti, višedimenzionalnoj tački dodjeljuju tačno jednu tačku u niskodimenzionalnom prostoru. SNE omogućava da se višedimenzionalnoj tački dodijeli nekoliko niskodimenzionalnih slika koje dobro opisuju originalnu tačku. SNE metoda započinje tako što za svaki objekat i i svakog potencijalnog susjeda tog objekta j u višedimenzionalnom prostoru izračuna vjerovatnoću da je baš objekat j susjed objekta i prema formuli:

$$p_{i,j} = \frac{\exp(-d_{i,j}^2)}{\sum_{k \neq i} (\exp(-d_{i,k}^2))}, \quad (5.2)$$

gdje $d_{i,j}$ predstavlja Euklidovu distancu između tačaka i i j :

$$d_{i,j}^2 = \frac{\|x_i - x_j\|^2}{2\sigma_i^2}. \quad (5.3)$$

Parametar σ_i^2 bira se ručno ili binarnom pretragom. Ove vjerovatnoće definišu se i za tačke u niskodimenzionalnom prostoru:

$$q_{i,j} = \frac{\exp(-\|y_i - y_j\|^2)}{\sum_{k \neq i} (\exp(-\|y_k - y_i\|^2))}. \quad (5.4)$$

U odnosu na originalnu SNE metodu, tSNE uvodi izmjenu u raspodjeli koja se koristi za računanje vjerovatnoća. Za razliku od Gausove raspodjele u originalnoj metodi, tSNE koristi Studentovu - t raspodjelu. Korištenjem Studentove raspodjele vjerovatnoće $q_{i,j}$ postaju:

$$q_{i,j} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq i} (1 + \|y_k - y_i\|^2)^{-1}}. \quad (5.5)$$

Osnovni cilj metode je da se raspodjele vjerovatnoća $p_{i,j}$ i $q_{i,j}$ što bolje poklapaju, i traže se vrijednosti u niskodimenzionalnom prostoru y_i takve da je ovaj uslov ispunjen. Funkcija cijene minimizuje se primjenom metode gradijentnog spusta, po vrijednostima y_i . Kao funkcija cijene definiše se Kullback-Leibler divergencija, čiji je gradijent:

$$\frac{\partial C}{\partial y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j)(1 + \|y_i - y_j\|^2)^{-1}. \quad (5.6)$$

5.3.2 Analiza glavnih komponenti

Analiza glavnih komponenti, PCA (eng. *Principal Component Analysis*) predstavlja metodu za smanjenje dimenzionalnosti i ekstrakciju relevantnih informacija iz komplikovanih podataka sa većim brojem varijabli. Pearson prvi put opisuje ovaj metod kao pronalaženje linija i ravni koje najbolje odgovaraju sistemu tačaka u prostoru [54].

Recimo da je podatke sa N obilježja potrebno redukovati na dimenzionalnost M . Prema geometrijskom tumačenju, PCA metoda predstavlja projekciju tačaka podataka u M -dimenzioni prostor, pri čemu se taj prostor bira kao najbolja interpolacija između tačaka [55], odnosno pronalazi se linearna transformacija uzoraka u prostor manje dimenzionalnosti tako da se uzorci najbolje predstave u smislu srednjekvadratne greške.

Kao prvi korak potrebno je pronaći ortonormalnu bazu M -dimenzionog prostora u koji želimo da preslikamo podatke. Označimo bazu novog prostora sa $u = \{u_1, \dots, u_m\}$. Tada, N -dimenzionalne podatke $x^{(i)}$, čija je srednja vrijednost m_x želimo da aproksimiramo M -dimenzionalnim reprezentacijama:

$$\hat{x}^{(i)} = m_x + \sum_{j=1}^M z_j^{(i)} u_j, \quad (5.7)$$

tako da greška:

$$J = \sum_{j=1}^n \|x^{(i)} - \hat{x}^{(i)}\|^2 = \sum_{j=1}^n \|x^{(i)} - (m_x + \sum_{j=1}^M z_j^{(i)} u_j)\|^2, \quad (5.8)$$

bude minimalna. Glavne komponente podataka $z^{(i)}$ predstavljaju koordinate tačaka u novom prostoru [56]. Određivanje baze novog prostora nije jednostavan zadatak i najčešće se određuje singularnim razlaganjem, SVD (eng. *Singular Value Decomposition*) [57]. Ukoliko označimo matricu podataka, dimenzija $N \times n$, $\mathbf{X} = [\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}]$, tada možemo zapisati SVD ove matrice kao:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \quad (5.9)$$

gdje kolone \mathbf{U} predstavljaju lijeve a kolone \mathbf{V} desne singularne vektore matrice \mathbf{X} , a $\mathbf{\Sigma}$ predstavlja dijagonalnu matricu na čijoj se dijagonali nalazi $r = \min(N, n)$ singularnih vrijednosti. Tada se problem pronalaska baze novog prostora svodi na problem pronalaska lijevih singularnih vektora matrice podataka.

Primjena PCA i tSNE na sintetičke i stvarne podatke nam omogućava da pomoću grafičkog prikaza u dvodimenzionalnom prostoru poredimo koliko su slične raspodjele stvarnih i sintetičkih podataka.

5.3.3 Validacija podataka pomoću prediktora

Kako bi se potvrdila upotrebljivost podataka na problemu kratkoročne predikcije, podaci se testiraju na dva prediktora. Prvi je AR model kakav je opisan u Glavi 2, koji predviđa narednu tačku vremenskog niza kao linearnu kombinaciju prethodnih tačaka. Drugi tip prediktora je realizovan kao LSTM neuronska mreža. Mreža se sastoji od 2 skrivena sloja sa po 5 čvorova u svakom sloju. Korišten je Adam optimizator [58] sa koeficijentom učenja 0,0001 i kao funkcija cijene korištena je srednjekvadratna greška. Zadatak mreže je da predvidi vrijednost solarne iradijance jedan odmjerak, odnosno jedan sat unaprijed na osnovu prethodnih nekoliko odmjeraka koji se mreži dovode kao ulaz.

Oba prediktora testirana su na podacima sa dvije različite dužine sekvenci. Ukoliko su podaci generisani od strane modela koji kao izlaz daje sekvencu dužine 7 tada je prediktor obučen tako da na osnovu 6 prethodnih odmjeraka vrši predikciju sedmog odmjerka u sekvenci. Ukoliko su podaci generisani od strane modela koji kao izlaz daje sekvencu dužine 25 tada je prediktor obučen tako da na osnovu 24 prethodna odmjerka vrši predikciju 25. odmjerka u sekvenci.

Oba prediktora su testirana na testnom skupu podataka i za svaki pojedinačni slučaj računa se srednja apsolutna greška, MAE (eng. *Mean Absolute Error*) i srednjekvadratna greška, MSE (eng. *Mean Squared Error*).

5.3.4 Kalkulator energije koju proizvede solarna elektrana

Nakon testiranja podataka na prediktorima potrebno je verifikovati njihovu upotrebljivost na problemu dugoročne predikcije. Ideja je da se rad solarne elektrane na određenom vremenskom periodu simulira korištenjem sintetičkih podataka, a potom dobijeni rezultati uporede sa scenarijom dobijenim korištenjem testnog skupa stvarnih podataka.

Za simulaciju rada sistema solarnih panela korišten je kalkulator koji na osnovu podataka o solarnoj iradijansi računa energiju koju bi proizveo sistem solarnih panela na određenom vremenskom intervalu. Princip rada ovog kalkulatora zasnovan je na principu rada PVWatts kalkulatora američke Nacionalne laboratorije za obnovljivu energiju, NREL (eng. *National Renewable Energy Laboratory*) [59]. PVWatts servis daje podatke o ukupnoj generisanoj energiji solarnog sistema određenih karakteristika, koje se u kalkulator unose kao ulazni podaci zajedno sa lokacijom sistema. Kalkulatoru korištenom u ovom radu sem parametara sistema kao ulaz predaje se niz podataka o solarnoj iradijansi i temperaturi na lokaciji od interesa. Sam princip proračuna kalkulatora korištenog u ovom radu i koraci su su gotovo identični PVWatts kalkulatoru.

Sem podataka o solarnoj iradijansi i temperaturi ulazni podaci kalkulatora su:

- geografska širina i dužina lokacije na kojoj je smješten sistem panela,
- meridijan koji odgovara lokalnom vremenu pozicije na kojoj je smješten sistem panela,

- ugao azimuta panela,
- ugao nagiba panela u odnosu na horizontalnu površinu,
- ukupna snaga sistema panela,
- stepen iskorištenja invertora,
- odnos instalisanih snaga DC i AC dijelova sistema,
- ukupni gubici sistema izraženi u procentima,
- nominalna radna temperatura solarne ćelije, NOCT (eng. *Nominal Operating Cell Temperature*),
- koeficijent refleksije površine.

Na osnovu zadatih podataka kalkulator računa snagu na izlazu solarne elektrane u sljedećim koracima:

- za svaku vrijednost solarne iradijance u ulaznom fajlu računa se dan u godini i odgovarajući sat na osnovu čega se potom određuje položaj Sunca u odnosu na solarne panele,
- korištenjem Perezovog [60] modela vrši se svodenje solarne iradijance na površinu panela pod određenim uglom u odnosu na horizontalnu površ,
- vrši se izračunavanje radne temperature solarnih ćelija,
- određuje se snaga DC dijela sistema,
- na osnovu DC snage sistema i podatka o efikasnosti invertora vrši se izračunavanje krajnje izlazne snage sistema.

S obzirom na to da kalkulator kao ulazne vrijednosti prihvata vrijednosti iradijance u W/m^2 , prije upotrebe kalkulatora potrebno je svesti generisane vrijednosti indeksa vedrog neba na vrijednosti iradijance množenjem sa modelom vedrog neba. S obzirom na to da je za simulaciju rada elektrane potrebno poznavanje direktne i difuzne komponente iradijance na horizontalnu površinu, potrebno je na neki način, na osnovu vrijednosti globalne iradijance procijeniti udio ovih komponenti. Za razdvajanje globalne iradijance na direktnu i difuznu komponentu iskorištena je Boland metoda [61].

Kalkulator je podešen tako da računa energiju panela snage 1000 W, pod nagibom od 30° i uglom azimuta od 0° . Odnos instalisanih snaga DC i AC dijela zadat je kao 1,2, ukupni gubici 14%, a stepen iskorištenja invertora 0,96. Posljednja dva parametra, NOCT i koeficijent refleksije imaju vrijednosti $45^\circ C$ i 0,2 respektivno. Model se testira na testnom skupu baze podataka i na sintetičkim podacima koji odgovaraju istom vremenskom periodu u godini kao i testni skup.

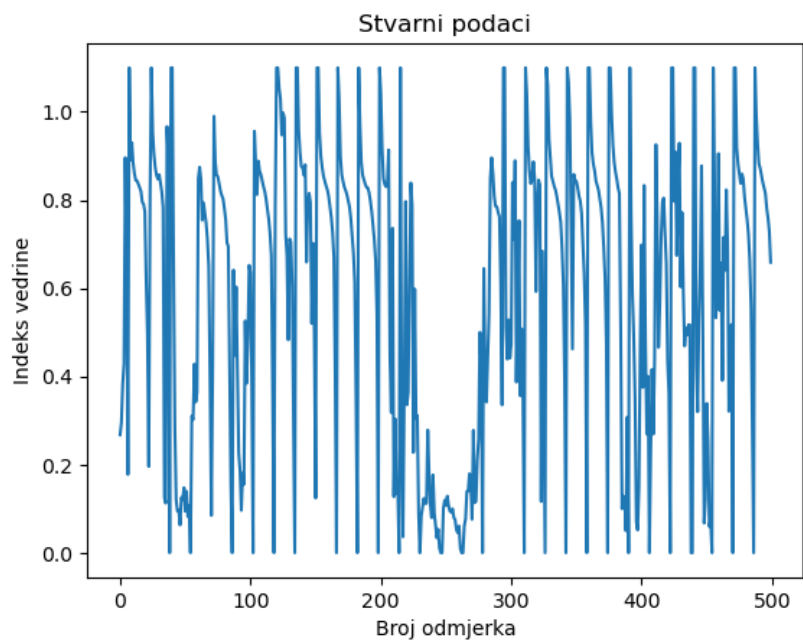
Glava 6

Eksperimentalni rezultati

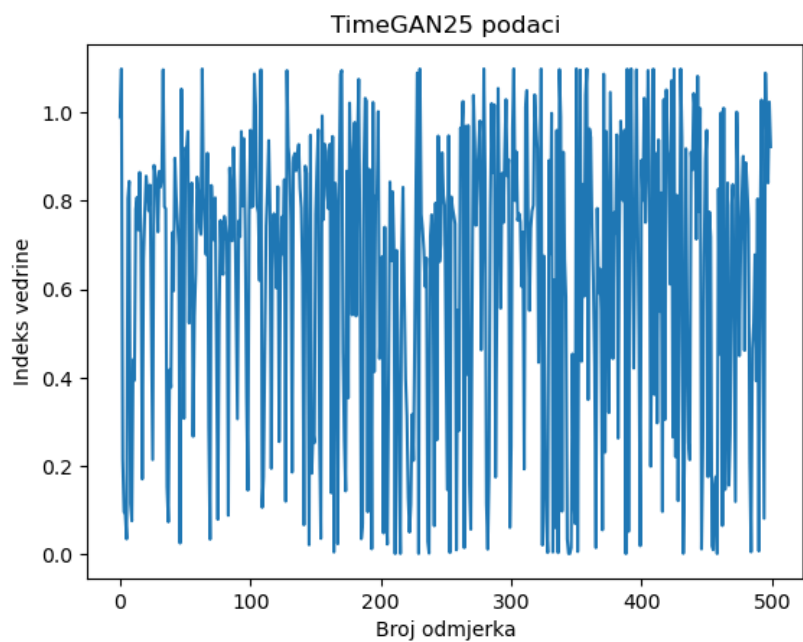
Svaki od modela koji su testirani: LSTM-GAN, C-RNN-GAN, TimeGAN i ExtraMAE, trenirani su na podacima različitih dužina vremenskih sekvenci kao što je to opisano u Glavi 5. Odabrane su dvije dužine sekvenci, 7 i 25. Nakon treniranja generisana je odgovarajuća količina sintetičkih podataka. Kako bi se performanse sintetičkih podataka mogle direktno porediti sa performansama stvarnih, generisana količina sintetičkih podataka jednaka je dužini trening skupa. Podaci o sunčevom zračenju generisani su u vidu indeksa vedrog neba, i pored njih generisani su i prethodno navedeni parametri: dan u godini, temperatura, brzina vjetra, vlažnost vazduha i prisustvo padavina.

Kao prvi korak u ocjeni kvaliteta sintetičkih podataka generisani su PCA i tSNE dijagrami koji bi trebali dati sliku o tome da li stvarni i sintetički podaci potiču iz iste raspodjele. Nakon toga vrši se obučavanje prediktora, prvo na stvarnim podacima, isječenim na dvije odgovarajuće dužine sekvence, a potom na sintetičkim. Posljednji korak je testiranje sintetičkih baza podataka sa najboljim performansama na kalkulatoru generisane energije solarne elektrane.

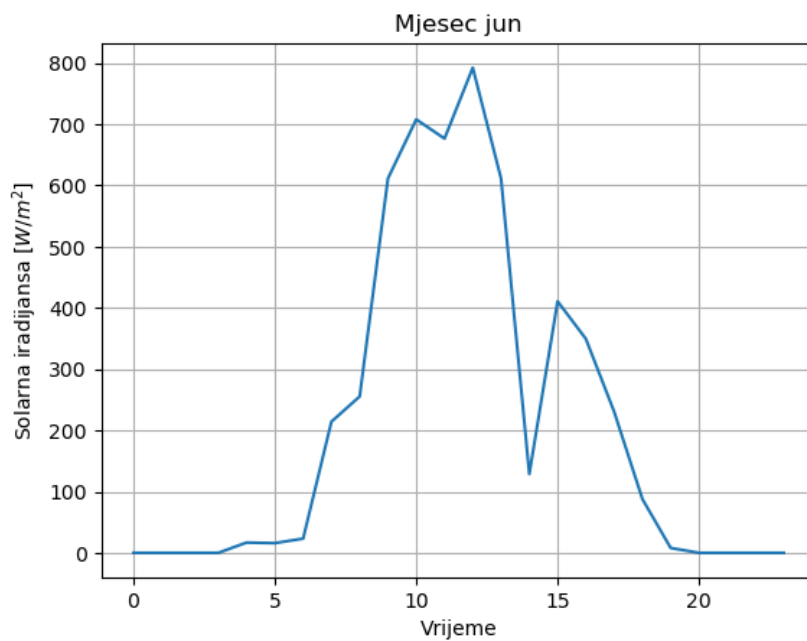
Na Slikama 6.1 i 6.2 prikazani su primjeri stvarnih i sintetičkih podataka u domenu indeksa vedrog neba. Na Slici 6.1 prikazane su stvarne vrijednosti indeksa vedrog neba za mjesec jun i jul, dok je na slici 6.2 prikazan primjer sintetičkih podataka dobijenih TimeGAN mrežom treniranom na sekvencama dužine 25. Za potrebe dugoročnog planiranja korištene su sekvence u domenu iradijance, dobijene množenjem indeksa vedrog neba modelom vedrog neba. Na slikama 6.3 i 6.4 prikazani su primjeri takvih sekvenci za mjesec jun i decembar respektivno. Sekvence su takođe generisane uz pomoć TimeGAN modela treniranog na sekvencama dužine 25.



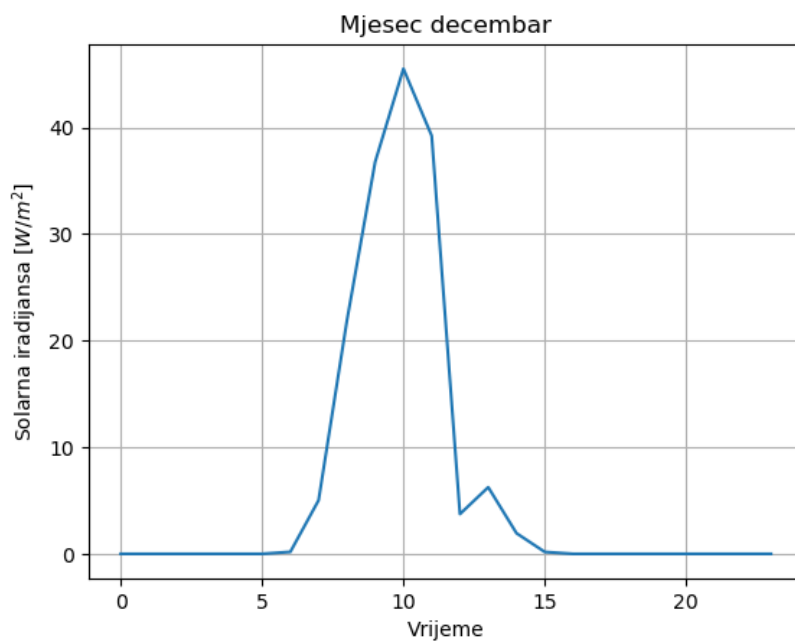
Slika 6.1: Stvarne vrijednosti indeksa vedrog neba



Slika 6.2: Sintetičke vrijednosti indeksa vedrog neba generisane putem TimeGAN mreže



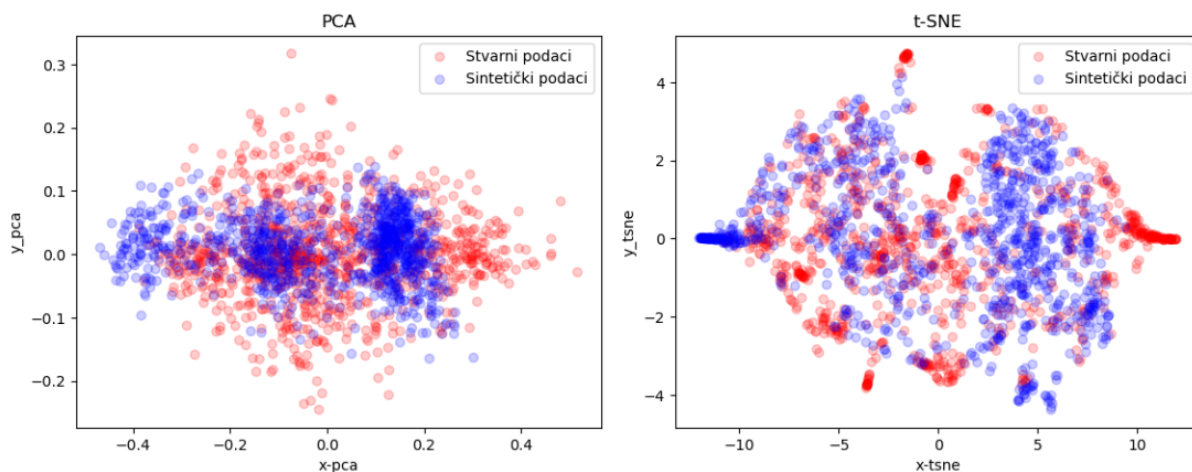
Slika 6.3: TimeGAN sintetička sekvenca, mjesec jun



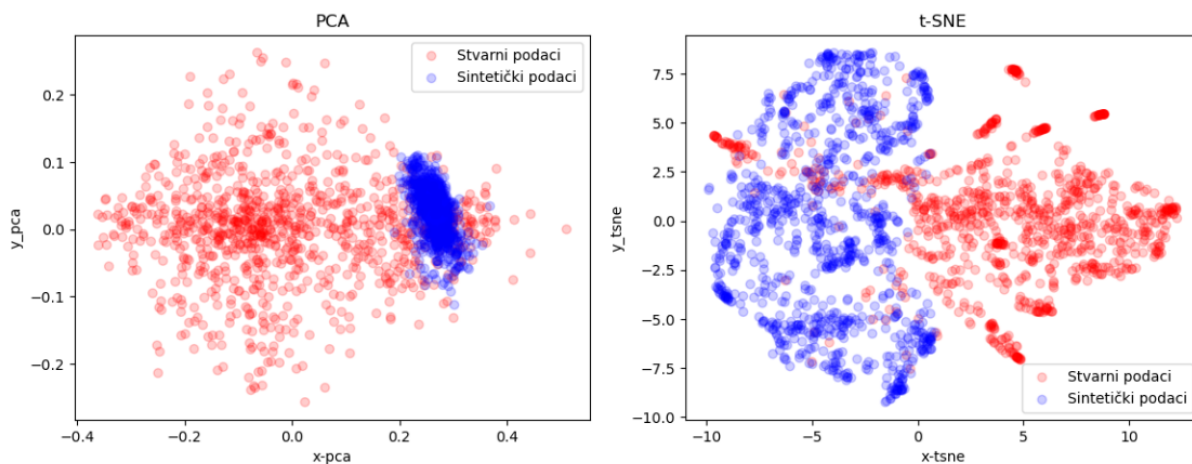
Slika 6.4: TimeGAN sintetička sekvenca, mjesec decembar

6.1 PCA i tSNE dijagrami

Kao što je napomenuto na početku poglavlja, PCA i tSNE dijagrami su prikazani za baze podataka generisane uz pomoć svakog od navedenih modela. Na Slikama 6.5, 6.6, 6.7 i 6.8, prikazani su dijagrami kada su modeli trenirani na sekvencama dužine 7, što je ujedno i dužina generisanih sekvenci. Na dijagramima su prikazane komponente koje odgovaraju generisanim sekvencama i stvarnim sekvencama iz trening skupa. Prva stvar koju je moguće zaključiti posmatranjem dijagrama je da kompleksniji modeli, ExtraMAE i TimeGAN pokazuju mnogo bolje rezultate. Na osnovu dijagrama vidljivo je da postoji dobro preklapanje između podataka generisanih putem ovih modela i originalnih trening podataka na osnovu čega možemo zaključiti da ovi modeli generišu raznovrsne podatke koji potiču iz slične ili iste raspodjele kao originalni. Takođe, sa ovih dijagrama moguće je zaključiti da modeli ne generišu podatke prostim kopiranjem trening skupa stvarnih podataka. Samo na osnovu dijagrama, C-RNN-GAN predstavlja model koji pokazuje najlošije rezultate.

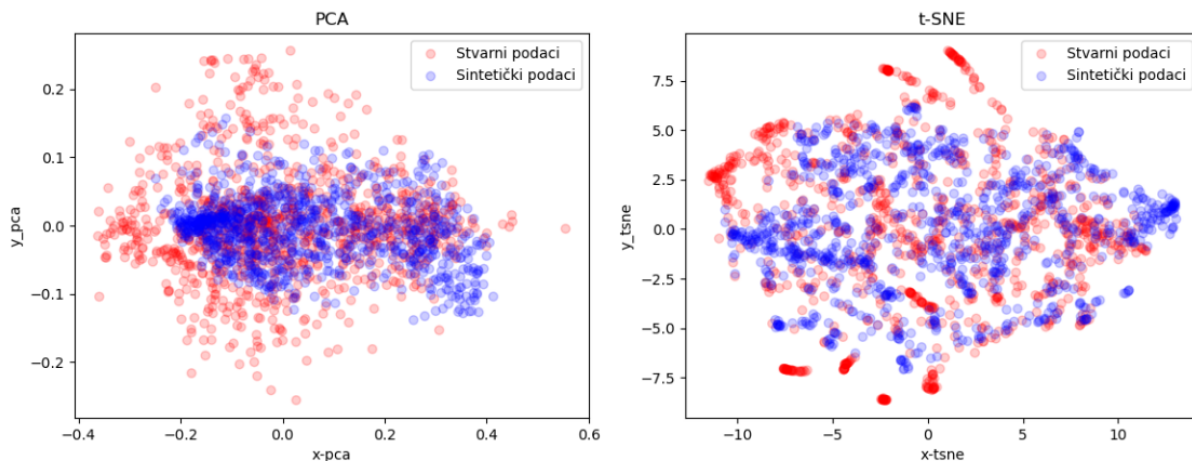


Slika 6.5: PCA i tSNE dijagrami LSTM-GAN podataka dužine 7



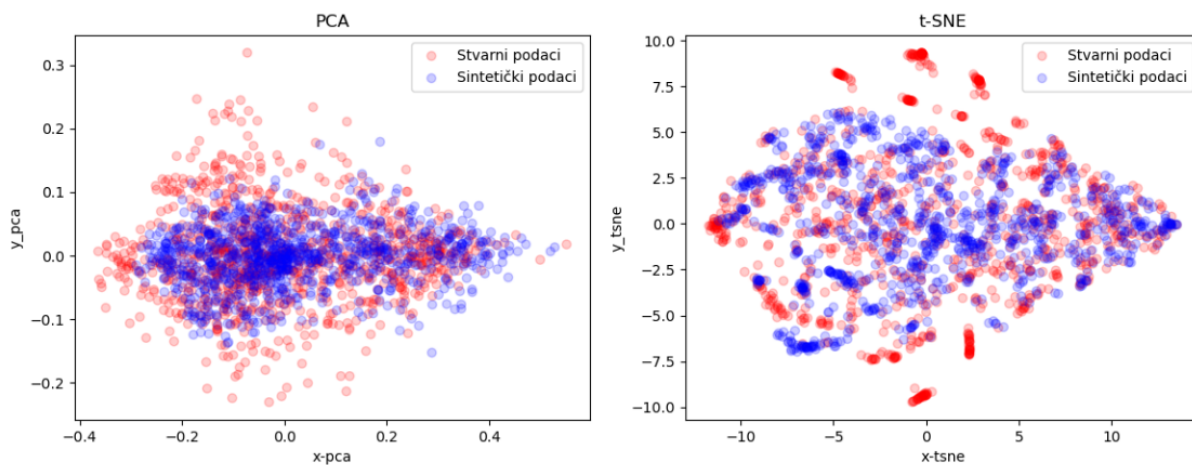
Slika 6.6: PCA i tSNE dijagrami C-RNN-GAN podataka dužine 7

6.1. PCA i tSNE dijagrami



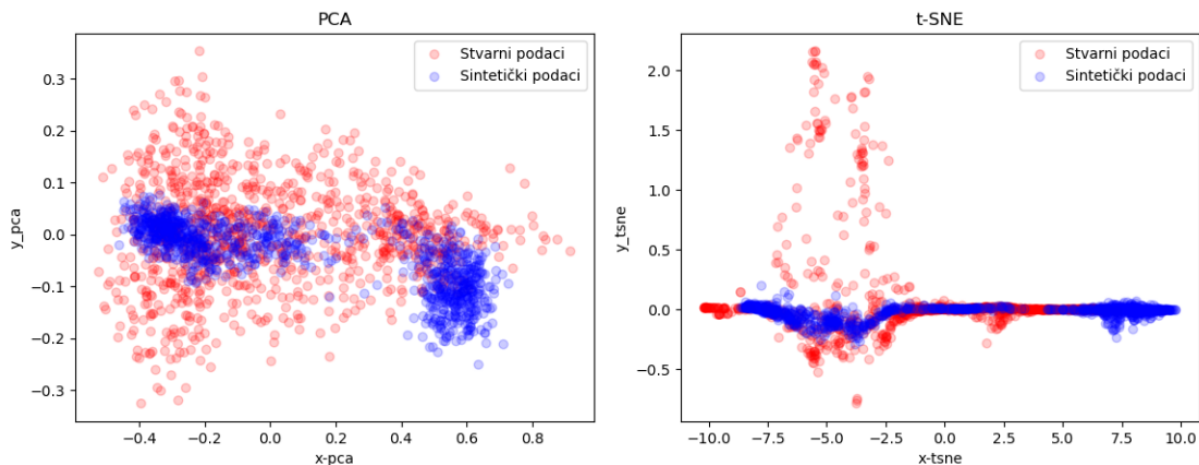
Slika 6.7: PCA i tSNE dijagrami TimeGAN podataka dužine 7

Na Slikama 6.9, 6.10, 6.11 i 6.12, prikazani su dijagrami kada su dužine trening i generisanih sekvenci 25. Ovi dijagrami daju nam slične informacije kao i dijagrami sa sekvencama dužine 7. Ubjedljivo najlošije rezultate pokazuje C-RNN-GAN model, gdje ne postoji nikakvo preklapanje između komponenata stvarnih i sintetičkih podataka. Sa druge strane, ExtraMAE i TimeGAN model pokazuju značajno preklapanje i bolje rezultate u odnosu na ostale modele.

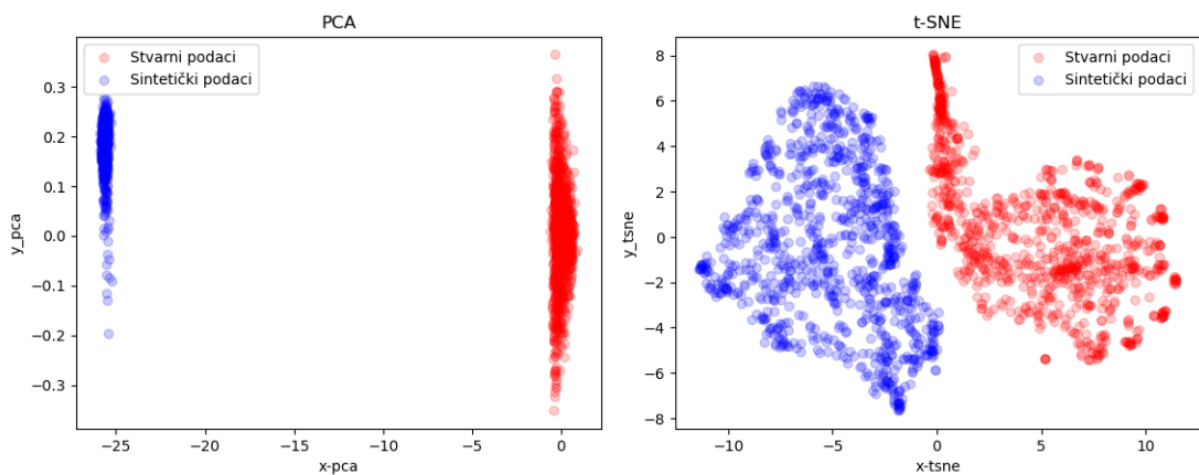


Slika 6.8: PCA i tSNE dijagrami MAE podataka dužine 7

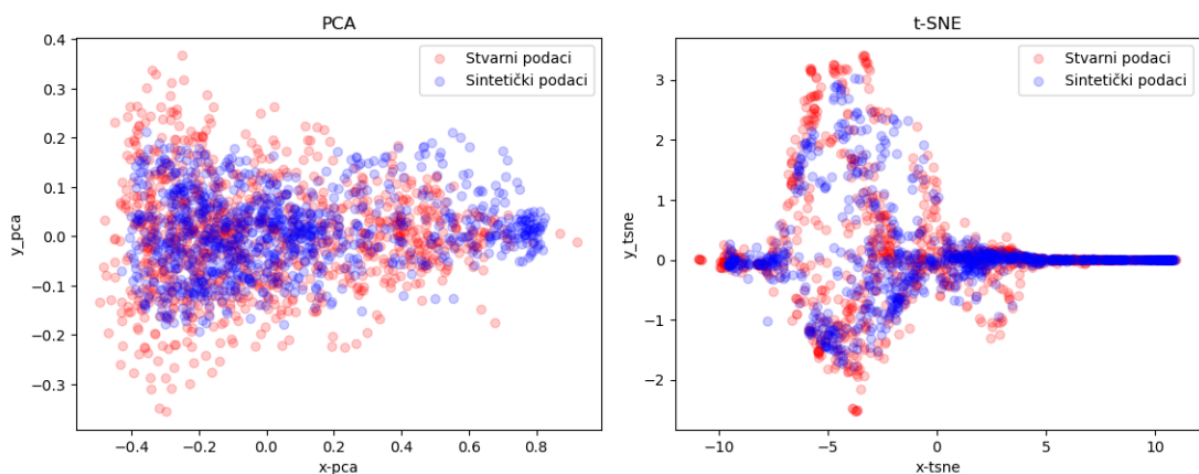
6.1. PCA i tSNE dijagrami



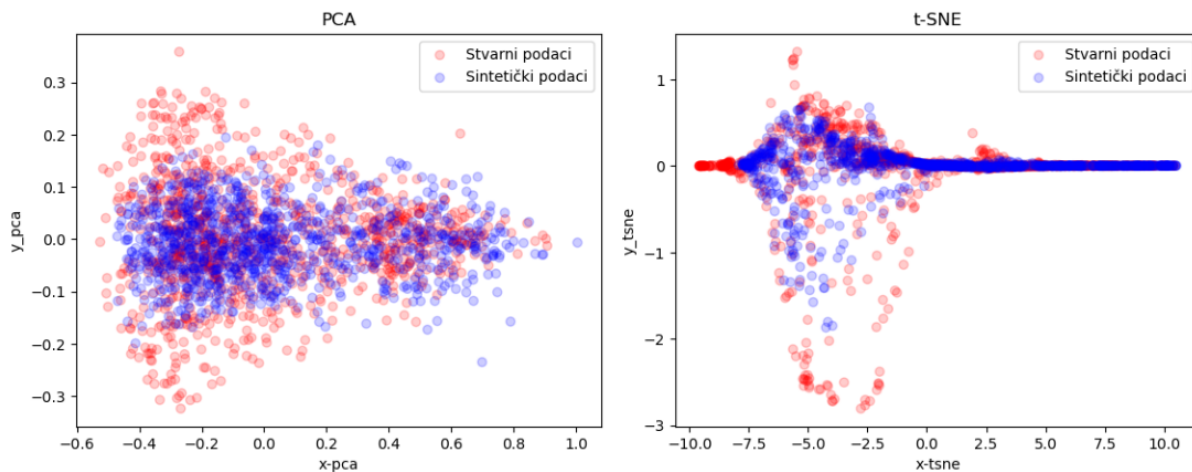
Slika 6.9: PCA i tSNE dijagrami LSTM-GAN podataka dužine 25



Slika 6.10: PCA i tSNE dijagrami C-RNN-GAN podataka dužine 25



Slika 6.11: PCA i tSNE dijagrami TimeGAN podataka dužine 25



Slika 6.12: PCA i tSNE dijagrami MAE podataka dužine 25

6.2 Rezultati prediktora

Kao što je objašnjeno u prethodnoj glavi, sintetički podaci iskorišteni su za obučavanje prediktora solarne iradijance. Ideja je vršenje predikcije indeksa vedrog neba jedan sat unaprijed, na osnovu poznatih nekoliko prethodnih odmjeraka. Testiraju se dva prediktora, AR prediktor i LSTM neuronska mreža. Parametri LSTM prediktora opisani su u Glavi 5.

Radi poređenja rezultata oba prediktora prvo su obučena na stvarnim trening podacima. Za predikciju su i u ovom slučaju korištene dvije dužine sekvenci. Prva dužina je 6, a druga dužina je 24 odmjerka. Prediktori koji se obučavaju na sintetičkim podacima obučeni su tako da unutar sekvence dužine 7, prediktor na osnovu prvih 6 tačaka estimira sedmu, nakon čega se koristi iduća sekvenca, bez preklapanja. Ovakav princip korišten je kako bi se osigurala vremenska zavisnost odmjeraka. Ista logika je primijenjena na sekvence dužine 25, gdje se na osnovu 24 odmjerka vrši predikcija naredne vrijednosti.

U Tabeli 6.1 prikazani su rezultati prediktora kada je dužina sekvence 7. Oba prediktora obučena su na stvarnim podacima i na sintetičkim podacima generisanim od strane svakog od modela. Potom je svaka verzija prediktora testirana na predikciji podataka iz testnog skupa i izračunata je srednja apsolutna greška i srednjekvadratna greška. LSTM mreža, kao kompleksniji prediktor, daje bolje rezultate predikcije za skoro sve testirane baze podataka. Ono što je moguće zaključiti je da i u ovom slučaju podaci generisani TimeGAN i MAE modelima pokazuju bolje performanse. C-RNN-GAN daje najlošije rezultate, što je bilo moguće pretpostaviti na osnovu PCA i tSNE dijagrama prikazanih u prethodnom odjeljku. Za ovo postoji nekoliko potencijalnih objašnjenja. C-RNN-GAN je primarno korišten za generisanje muzike i obučavan na sekvencama značajno veće dužine nego što su sekvence solarne iradijance. Sem toga, osnovni problem C-RNN-GAN-a tokom obučavanja je velika razlika koja se javlja između funkcija cijene generatora i diskriminatora, pri čemu je ažuriranje generatora u tom slučaju minimalno. Prilikom obučavanja mreže primijenjen je mehanizam zamrzavanja diskriminatora kada razlika postane prevelika, ali ovo nije dovelo do značajnih poboljšanja modela.

Tabela 6.1: Rezultati prediktora obučeni na podacima dužine sekvence 7

	AR		LSTM	
	MAE	MSE	MAE	MSE
Stvarni podaci	0,2021	0,0837	0,2053	0,0676
LSTM-GAN	0,2698	0,1273	0,2392	0,0885
C-RNN-GAN	0,5210	0,3829	0,2466	0,0885
MAE	0,2200	0,1004	0,2152	0,0921
TimeGAN	0,2344	0,1085	0,1956	0,0747

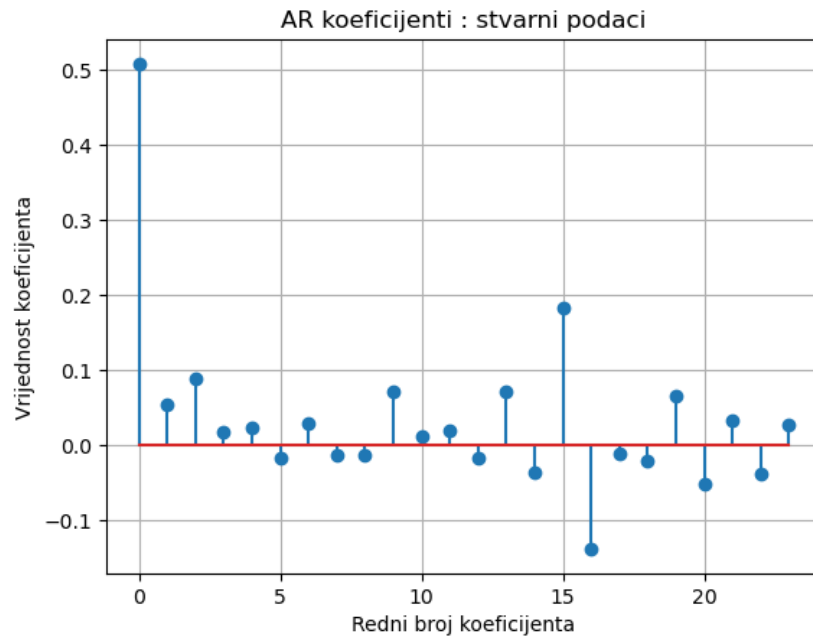
U Tabeli 6.3 prikazani su rezultati prediktora kada je dužina sekvence 25. Ukoliko se ovi rezultati uporede sa onima dobijenim kada je dužina sekvence 7, moguće je zaključiti da podaci sa većom dužinom sekvence generalno pokazuju bolje rezultate. Podaci generisani TimeGAN modelom, sa dužinom sekvence 25 daju najbolje rezultate, pri čemu su oni prilično slični rezultatima dobijenim kada su prediktori trenirani na stvarnim podacima. Ukoliko posmatramo dva modela sa najboljim performansama, MAE i TimeGAN, može se primijetiti da MAE daje bolje rezultate pri korištenju AR prediktora, dok su za LSTM prediktor povoljniji TimeGAN podaci.

Tabela 6.2: Rezultati prediktora obučeni na podacima dužine sekvence 25

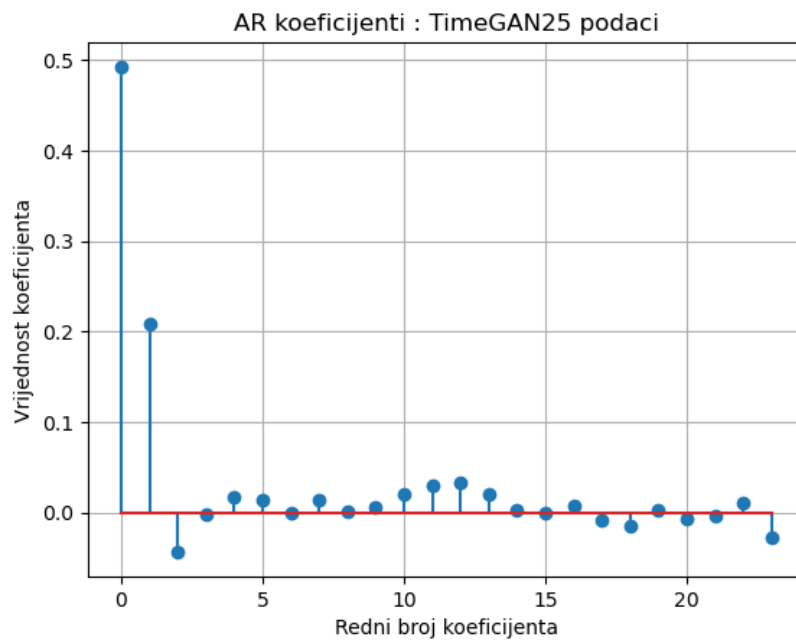
	AR		LSTM	
	MAE	MSE	MAE	MSE
Stvarni podaci	0,1916	0,0731	0,1523	0,0476
LSTM-GAN	0,2020	0,0773	0,2503	0,1257
C-RNN-GAN	0,2207	0,0979	0,3187	0,1647
MAE	0,1911	0,0811	0,1800	0,0757
TimeGAN	0,2033	0,0853	0,1632	0,0594

Ukoliko koeficijente AR prediktora obučenog na stvarnim podacima i AR prediktora obučenog na sintetičkim podacima sa najboljim vrijednostima metrika, odnosno TimeGAN podacima dužine 25 prikazane na Slikama 6.13 i 6.14 možemo primijetiti da u oba slučaja prvi koeficijent ima najveću vrijednost. Ovo implicira da najveći uticaj na predikciju budućeg odmjerka ima odmjerak koji mu neposredno prethodi. Ovo nam potvrđuju i dijagrami parcijalne autokorelacije, PACF (eng. *Partial Autocorrelation Function*) za stvarne i TimeGAN podatke, prikazani na Slikama 6.15 i 6.16. Na osnovu njih možemo vidjeti da je najveća korelacija prisutna između dva susjedna odmjerka.

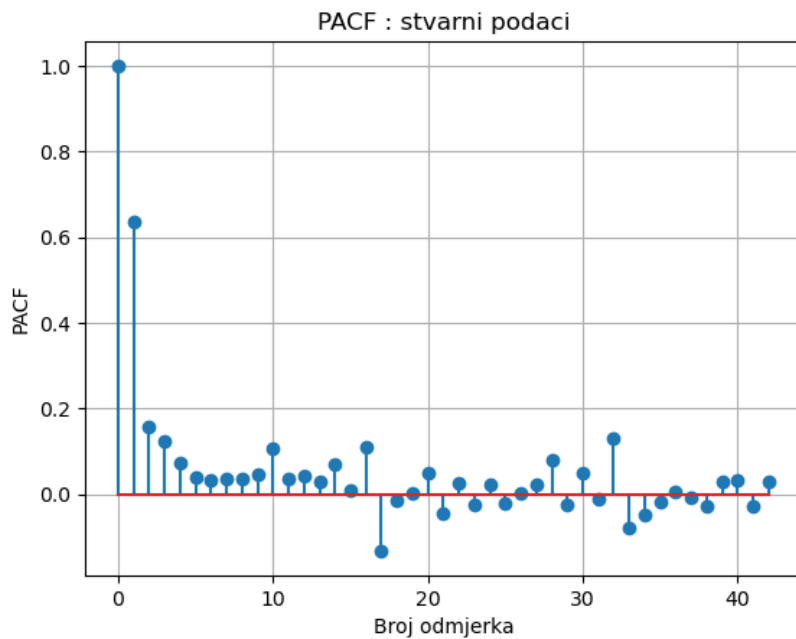
Na Slikama 6.17, 6.18, 6.19 i 6.20 prikazane su stvarne vrijednosti testnog skupa u poređenju sa vrijednostima dobijenim predikcijom, za dva skupa sintetičkih podataka sa najboljim performansama, MAE i TimeGAN podataka dužine sekvenci 25, kao i uvećanih prvih 200 odmjeraka za svaki od slučaja.



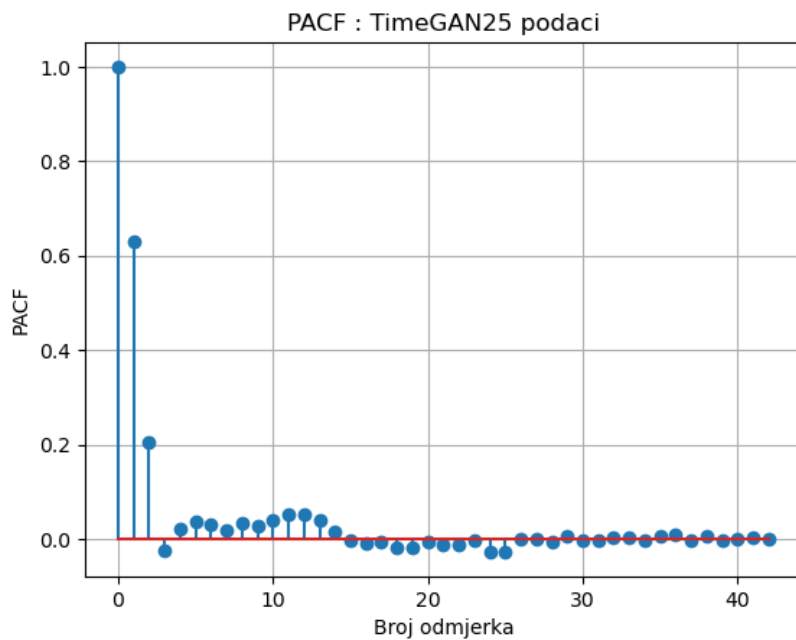
Slika 6.13: Koeficijenti AR modela obučenog na stvarnim podacima



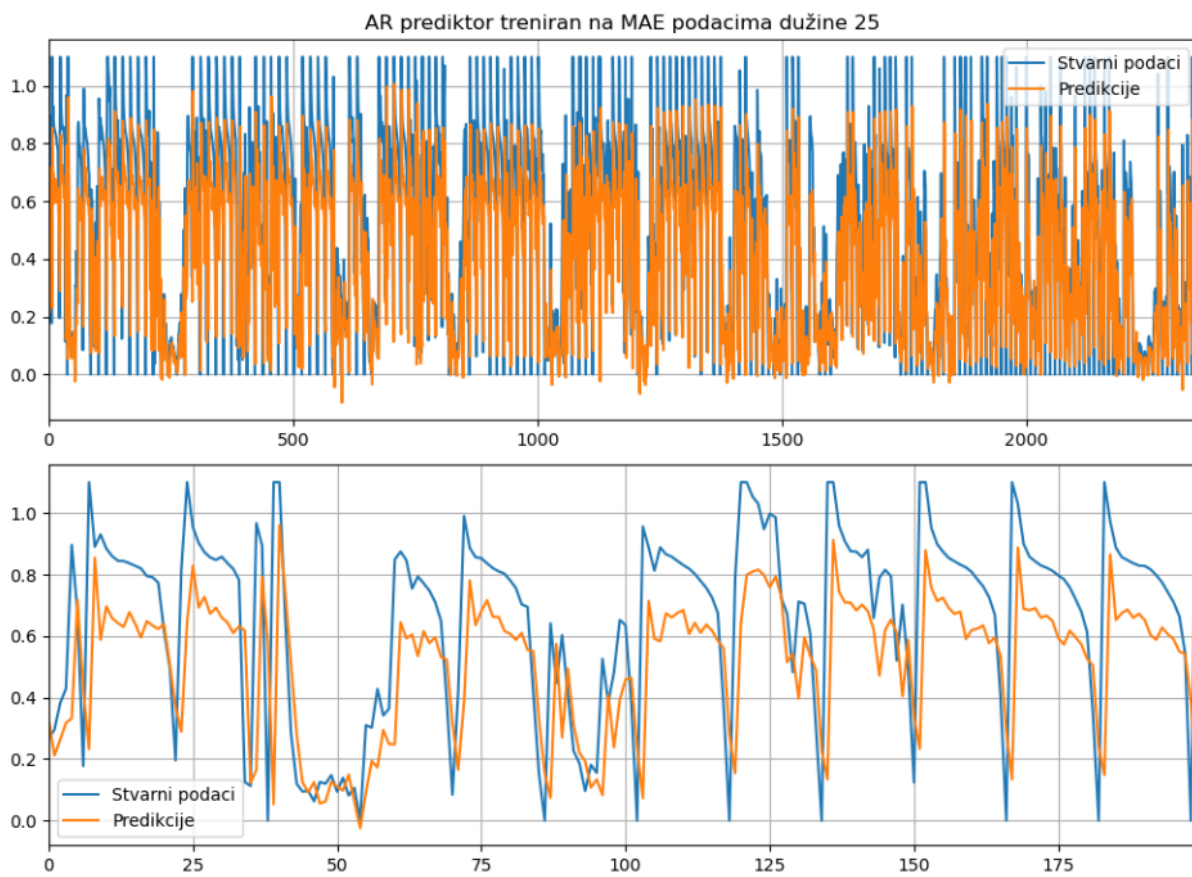
Slika 6.14: Koeficijenti AR modela obučenog na TimeGAN25 podacima



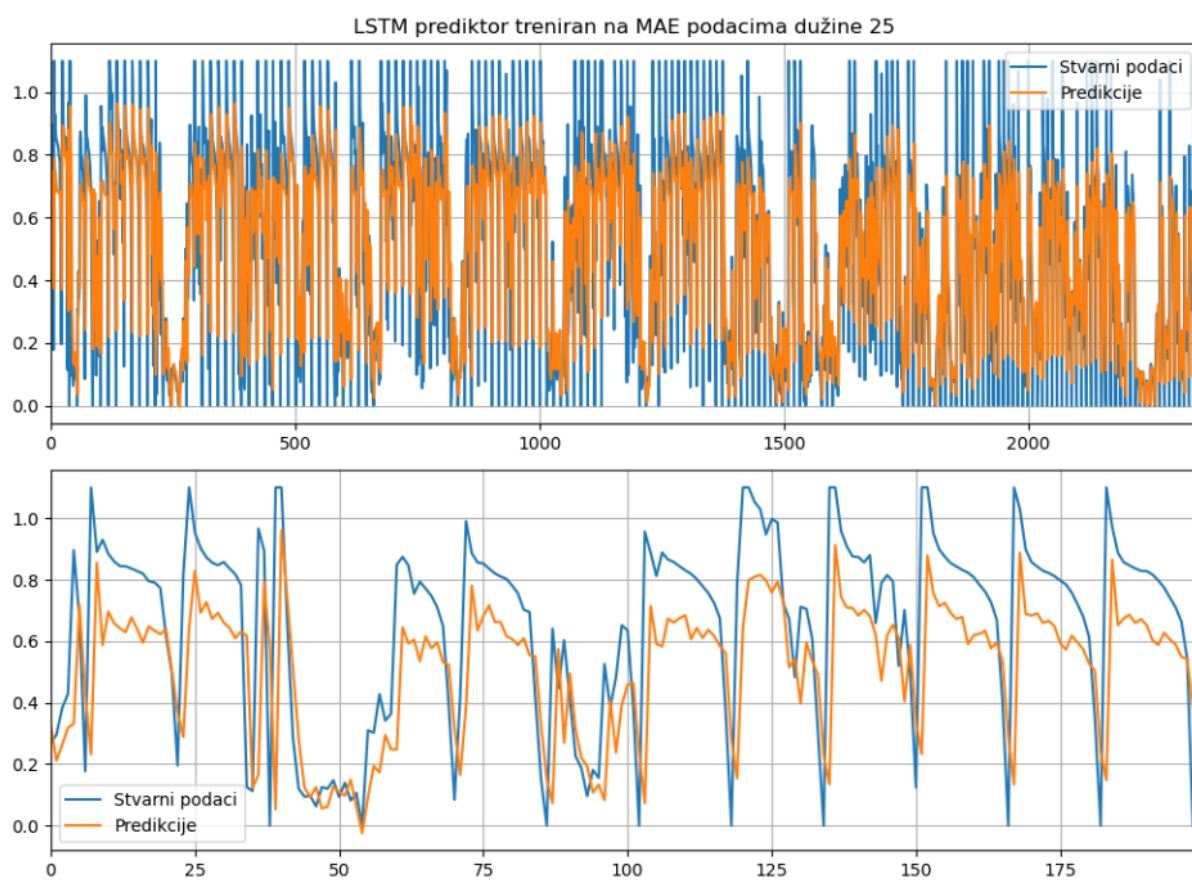
Slika 6.15: PACF stvarnih podataka



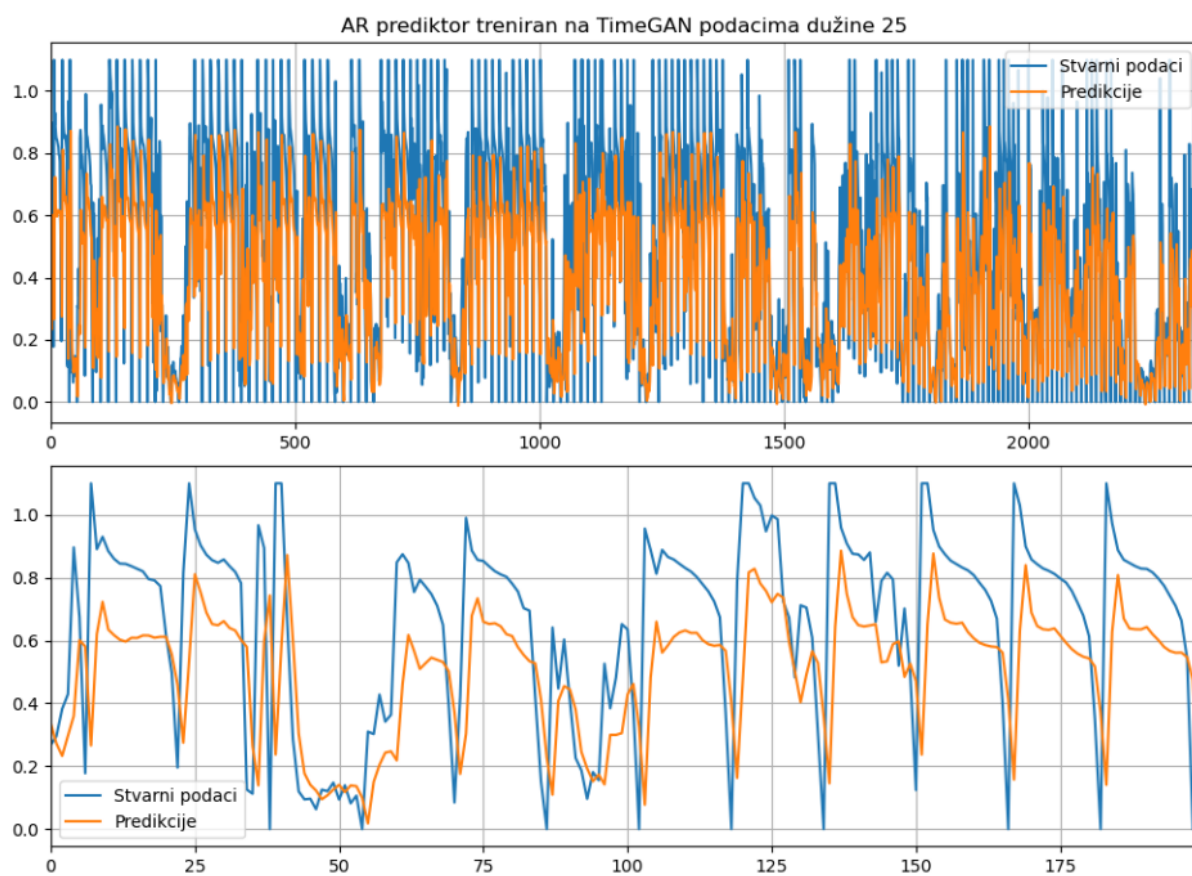
Slika 6.16: PACF TimeGAN25 podataka



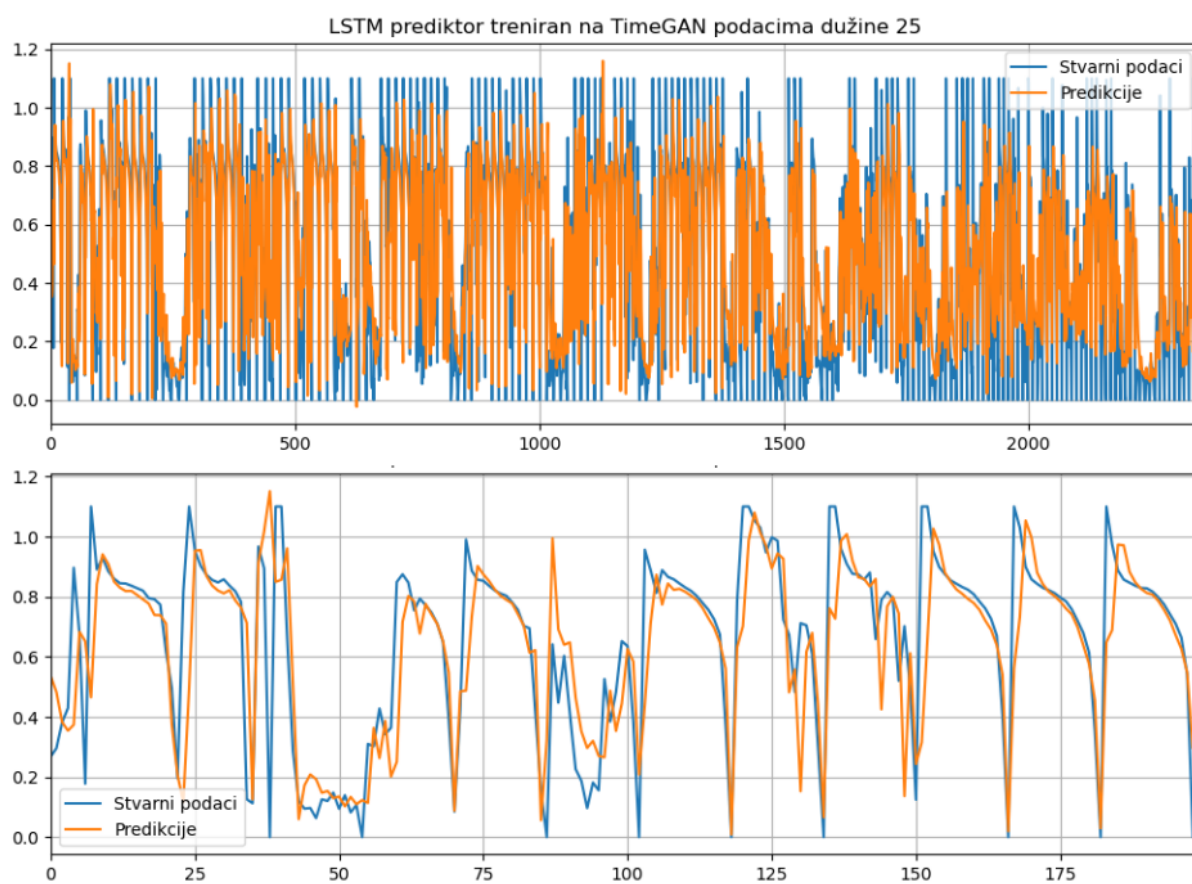
Slika 6.17: Predikcije dobijene AR prediktorom obučenom na MAE podacima dužine 25



Slika 6.18: Predikcije dobijene LSTM prediktorom obučenom na MAE podacima dužine 25



Slika 6.19: Predikcije dobijene AR prediktorom obučenom na TimeGAN podacima dužine 25



Slika 6.20: Predikcije dobijene LSTM prediktorom obučenom na TimeGAN podacima dužine 25

6.3 Rezultati kalkulatora

Podaci koji su pokazali najbolje rezultate na prethodnim testovima izabrani su za validaciju uz pomoć kalkulatora snage PV sistema. U ovom slučaju to su podaci generisani pomoću TimeGAN mreže, sa dužinom sekvence 25. Podaci su iz domena indeksa vedrog neba ponovo preračunati u vremenski domen, nakon čega je izvršeno razdvajanje direktne i difuzne komponente.

Simulira se vremenski period od 6 mjeseci, kolika je dužina testnog skupa. Prema tome, iz skupa sintetičkih podataka odabrani su podaci koji odgovaraju upravo danima u ovom vremenskom periodu. Radi preciznijeg poređenja, sintetički podaci za mjeseci od interesa generisani su za 10 različitih godina. Za energiju koja je izračunata za ove godine određena je srednja vrijednost i standardna devijacija. Sem podataka o indeksu vedrog neba iz sintetičkog skupa podataka izdvojeni su podaci o temperaturi koji takođe predstavljaju ulazne podatke kalkulatora.

Prvo se uz pomoć kalkulatora računa energija koju PV sistem generiše za 6 mjeseci ukoliko su ulazni podaci stvarni, izmjereni podaci o solarnoj iradijansi. Dobijena energija pod ovakvim uslovima iznosi $4,2305 \cdot 10^5$ Wh. Ukoliko su u isti sistem uneseni podaci generisani pomoću TimeGAN modela za 10 godina, dobijena srednja vrijednost energije iznosi $4,4981 \cdot 10^5$ Wh sa standardnom devijacijom $0,1503 \cdot 10^5$ Wh. Na osnovu dobijenog rezultata moguće je zaključiti da se korištenjem sintetičkih podataka prilično realistično može simulirati rad fotonaponskog sistema na određenoj lokaciji.

Tabela 6.3: Generisana energija fotonaponskog sistema

Podaci	Stvarni podaci	TimeGAN25 podaci
Generisana energija [MWh]	423,05	449,81 \pm 15,03

Tabela 6.4: Generisana energija fotonaponskog sistema u MWh po mjesecu

	Jul	Avgust	Septembar	Oktobar	Novembar	Decembar
Stvarni podaci	118,72	110,01	87,39	51,37	32,08	23,48
TimeGAN25	119,55 \pm 8,27	108,24 \pm 8,23	85,98 \pm 5,68	60,94 \pm 6,40	40,76 \pm 7,27	33,78 \pm 9,13

Sem ukupnog rezultata za testnih 6 mjeseci u Tabeli 6.4 prikazane su i vrijednosti generisane energije fotonaponskog sistema za svaki od mjeseci u testnom periodu. Sem zaključka da se rad elektrane može realistično simulirati korištenjem sintetičkih podataka, na osnovu vrijednosti standardnih devijacija možemo zaključiti da model daje različite scenarije i da postoji varijacija između pojedinih godina.

Glava 7

Zaključak

Potreba za sve većom količinom električne energije, kao i ekološka kriza izazvana velikom upotrebom fosilnih goriva podstakli su rapidan rast udjela obnovljivih izvora energije u elektroenergetskom sistemu, pri čemu se kao najznačajnija ističe solarna energija. Zajedno sa porastom upotrebe solarne energije, opada i cijena opreme, samim tim i cijena električne energije dobijene iz ovog izvora. Zahvaljujući tome, očekuje se i daljnja ekspanzija ovih tehnologija. Ova nagla ekspanzija stvara i određene probleme, od čega većina potiče od stohastičke prirode obnovljivih izvora. Snaga fotonaponskih sistema direktno zavisi od trenutne jačine Sunčevog zračenja što otežava planiranje rada elektroenergetskog sistema sa visokim udjelom ovakvih proizvođača. S tim u vezi, nastaje sve veće interesovanje za mehanizme koji bi omogućili predikciju snage Sunčevog zračenja određeni vremenski interval unaprijed, kako bi se olakšalo planiranje rada. Sem toga, postoji potreba za dugoročnom simulacijom rada fotonaponskog sistema na određenoj lokaciji, prije same njegove instalacije kako bi se olakšala njegova integracija u elektroenergetski sistem. S obzirom na to da su podaci o Sunčevom zračenju za određenu lokaciju teško dostupni i da mjerenje ovih podataka predstavlja jako dug proces, javlja se potreba za sintetičkim podacima o solarnoj iradijansi koji bi bili primjenjivi kada su u pitanju navedene upotrebe.

U ovom radu razmatrana je mogućnost generisanja realističnih sintetičkih podataka o solarnoj iradijansi korištenjem neuronskih mreža. Testirana su četiri različita modela, prethodno korištena za generisanje različitih tipova vremenskih sekvenci. Modeli su zasnovani na GAN i MAE arhitekturama. Za treniranje modela korišteni su podaci snimljeni na području Banjaluke, meteorološkom stanicom smještenom na Akademiji nauka i umjetnosti Republike Srpske. Svi podaci su svedeni u domen indeksa vedrog neba i razmatran je uticaj dužine sekvence podataka. Generisani podaci su testirani na tri različita načina: vizualizacijom pomoću PCA i tSNE dijagrama, testiranjem na kratkoročnim prediktorima budućih vrijednosti solarne iradijanse i testiranjem pomoću kalkulatora koji na osnovu podataka o solarnoj iradijansi za određenu lokaciju računa koliko energije generiše određeni fotonaponski sistem. Na osnovu dobijenih rezultata moguće je zaključiti da je uz pomoć kompleksnijih modela neuronskih mreža kao što su TimeGAN i maskirani autoenkoder moguće generisati sintetičke podatke koji su prikladni za navedene upotrebe i ostvaruju rezultate slične onima ostvarenim korištenjem stvarnih, snimljenih podataka

o solarnoj iradijansi. Pokazano je da je uz pomoć ovih sintetičkih podataka o solarnoj iradijansi moguće obučiti različite tipove prediktora solarne iradijanse, kao i simulirati rad solarne elektrane na dužem vremenskom periodu. Prikazom pomoću PCA i tSNE dijagrama zaključeno je da su podaci raznovrsni, a u isto vrijeme i realistični u poređenju sa stvarnim podacima, prema tome mogu se koristiti za simulaciju različitih vrsta scenarija koji se mogu desiti u fotonaponskim sistemima.

Ono što ostaje kao potencijalno neriješeno pitanje je generisanje podataka za lokacije koje su teško dostupne, ili za koje postoji mnogo manja količina izmjerenih podataka nego što je korišteno u ovom radu, npr. mjerenja u toku nekoliko mjeseci, ili čak nekoliko tipičnih dana u godini. Jedno potencijalno rješenje koje je moguće razmotriti je pretreniranje generativnih modela na podacima sa različitih lokacija na kojima su mjerenja dostupna, potom fino podešavanje na maloj količini podataka dostupnoj za lokaciju od interesa. Ovo bi potencijalno olakšalo treniranje generativnih modela i generisanje značajne količine podataka na osnovu mjerenja dobijenih na kratkom vremenskom periodu.

Postoji još niz specifičnosti koje je moguće implementirati, kao i niz različitih tipova validacije. Jedno zanimljivo pitanje bi bilo stvaranje univerzalnog modela, koji bi mogao da generiše podatke za proizvoljnu lokaciju. Primjena neuronskih mreža za rad, ne samo sa podacima o solarnoj iradijansi, već sa vremenskim nizovima generalno, je oblast koja još uvijek nije u potpunosti razvijena i za koju još uvijek ne postoje standardne procedure. U ovom radu pokazano je da primjena neuronskih mreža itekako ima potencijal u ovoj oblasti, posebno kada su u pitanju podaci o sunčevom zračenju i da postoji velika količina materijala za dalje istraživanje.

Bibliografija

- [1] Directive 2009/28/ec of the european parliament and of the council of 23 april 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing directives 2001/77/ec and 2003/30/ec. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:140:0016:0062:en:PDF>. Accessed: 2023-07-23.
- [2] Paris agreement. <https://unfccc.int/process-and-meetings/the-paris-agreement>. Accessed: 2023-07-23.
- [3] International energy outlook. <https://www.eia.gov/outlooks/aeo/data/browser/#/?id=17-IE02021®ion=4-0&cases=Reference&start=2010&end=2050&f=A&linechart=Reference-d210719.1-17-IE02021.4-0&sourcekey=0>. Accessed: 2023-07-23.
- [4] Energy management solutions to design, operate, and automate power systems. https://etap.com/images/default-source/product/photovoltaic-solar-panels/etap-photovoltaic-array-analysis-3.jpg?sfvrsn=b8a2bf7f_6. Accessed: 2023-10-12.
- [5] National geographic: Solar energy. <https://education.nationalgeographic.org/resource/solar-energy/>. Accessed: 2023-10-12.
- [6] Solar energy global LCOE and auction values. <https://www.irena.org/Data/View-data-by-topic/Costs/Global-LCOE-and-Auction-values>. Accessed: 2023-07-23.
- [7] 2021. solar energy electricity generation, country rankings. <https://www.irena.org/Data/View-data-by-topic/Capacity-and-Generation/Country-Rankings>. Accessed: 2023-07-23.
- [8] AP Grantham, PJ Pudney, and JW Boland. Generating synthetic sequences of global horizontal irradiation. *Solar Energy*, 162:500–509, 2018.
- [9] T Cebeacauer and M Suri. Typical meteorological year data: Solargis approach. *Energy Procedia*, 69:1958–1969, 2015.
- [10] L Mazonra Aguiar, Brais Pereira, Philippe Lauret, F Díaz, and Mathieu David. Combining solar irradiance measurements, satellite-derived data and a numerical weather

- prediction model to improve intra-day solar forecasting. *Renewable Energy*, 97:599–610, 2016.
- [11] G Peter Zhang. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50:159–175, 2003.
- [12] Ralf C Staudemeyer and Eric Rothstein Morris. Understanding LSTM—a tutorial into long short-term memory recurrent neural networks. *arXiv preprint arXiv:1909.09586*, 2019.
- [13] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [14] Felix A Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with LSTM. *Neural computation*, 12(10):2451–2471, 2000.
- [15] Greg Van Houdt, Carlos Mosquera, and Gonzalo Nápoles. A review on the long short-term memory model. *Artificial Intelligence Review*, 53:5929–5955, 2020.
- [16] Thomas Fischer and Christopher Krauss. Deep learning with long short-term memory networks for financial market predictions. *European journal of operational research*, 270(2):654–669, 2018.
- [17] Alaa Sagheer and Mostafa Kotb. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing*, 323:203–213, 2019.
- [18] Vinay Kumar Reddy Chimmula and Lei Zhang. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, solitons & fractals*, 135:109864, 2020.
- [19] Sujan Ghimire, Ravinesh C Deo, Nawin Raj, and Jianchun Mi. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. *Applied Energy*, 253:113541, 2019.
- [20] Xiangyun Qing and Yugang Niu. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy*, 148:461–468, 2018.
- [21] Enviromonitors: Kipp & Zonen SMP3 pyranometer. <https://www.enviromonitors.co.uk/kipp-and-zonen-smp3-pyranometer>. Accessed: 2023-10-12.
- [22] What is a pyrhelimeter? <https://www.hukseflux.com/applications/solar-energy-pv-system-performance-monitoring/what-is-a-pyrhelimeter>. Accessed: 2023-10-12.
- [23] Čedomir Zeljković. Obnovljivi izvori energije solarna energetika. *Univerzitet u Banjoj Luci/Elektrotehnički fakultet (Banja Luka) i Akademska misao (Beograd)*, 2018.
- [24] Joshua S Stein, Clifford W Hansen, and Matthew J Reno. Global horizontal irradiance clear sky models: implementation and analysis. Technical report, Sandia National Laboratories (SNL), Albuquerque, NM, and Livermore, CA . . . , 2012.

- [25] Philippe Lauret, Rodrigo Alonso-Suárez, Josselin Le Gal La Salle, and Mathieu David. Solar forecasts based on the clear sky index or the clearness index: Which is better? In *Solar*, volume 2, pages 432–444. MDPI, 2022.
- [26] M Daneshyar. Solar radiation statistics for Iran. *Sol. Energy;(United States)*, 21(4), 1978.
- [27] GW Paltridge and D Proctor. Monthly mean solar radiation statistics for Australia. *Solar Energy*, 18(3):235–243, 1976.
- [28] F Kasten. A simple parameterization of the pyrhelimetric formula for determining the linke turbidity factor. *Meteorol. Rdsch.*, 33:124–127, 1980.
- [29] Pierre Ineichen and Richard Perez. A new airmass independent formulation for the Linke turbidity coefficient. *Solar Energy*, 73(3):151–157, 2002.
- [30] Richard E Bird and Carol Riordan. Simple solar spectral model for direct and diffuse irradiance on horizontal and tilted planes at the earth’s surface for cloudless atmospheres. *Journal of Applied Meteorology and Climatology*, 25(1):87–97, 1986.
- [31] Mireille Lefevre, Armel Oumbe, Philippe Blanc, Bella Espinar, Benoît Gschwind, Zhipeng Qu, Lucien Wald, Marion Schroedter-Homscheidt, Carsten Hoyer-Klick, Antti Arola, et al. Mcclear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. *Atmospheric Measurement Techniques*, 6(9):2403–2418, 2013.
- [32] Wenjie Zhang, Yonghong Luo, Ying Zhang, and Dipti Srinivasan. SolarGAN: Multivariate solar data imputation using generative adversarial network. *IEEE Transactions on Sustainable Energy*, 12(1):743–746, 2020.
- [33] RJ Aguiar, Manuel Collares-Pereira, and JP Conde. Simple procedure for generating sequences of daily radiation values using a library of Markov transition matrices. *Solar Energy*, 40(3):269–279, 1988.
- [34] Ricardo Aguiar and MTAG Collares-Pereira. TAG: a time-dependent, autoregressive, Gaussian model for generating synthetic hourly radiation. *Solar energy*, 49(3):167–174, 1992.
- [35] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.
- [36] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.
- [37] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- [38] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang,

- et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.
- [39] Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. Seqgan: Sequence generative adversarial nets with policy gradient. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.
- [40] Olof Mogren. C-RNN-GAN: Continuous recurrent neural networks with adversarial training. *arXiv preprint arXiv:1611.09904*, 2016.
- [41] Shuntaro Takahashi, Yu Chen, and Kumiko Tanaka-Ishii. Modeling financial time-series with generative adversarial networks. *Physica A: Statistical Mechanics and its Applications*, 527:121261, 2019.
- [42] Cristóbal Esteban, Stephanie L Hyland, and Gunnar Rätsch. Real-valued (medical) time series generation with recurrent conditional gans. *arXiv preprint arXiv:1706.02633*, 2017.
- [43] Zhe Wang and Tianzhen Hong. Generating realistic building electrical load profiles through the Generative Adversarial Network (GAN). *Energy and Buildings*, 224:110299, 2020.
- [44] Jinsung Yoon, Daniel Jarrett, and Mihaela Van der Schaar. Time-series generative adversarial networks. *Advances in neural information processing systems*, 32, 2019.
- [45] Hao Ni, Lukasz Szpruch, Magnus Wiese, Shujian Liao, and Baoren Xiao. Conditional sig-wasserstein gans for time series generation. *arXiv preprint arXiv:2006.05421*, 2020.
- [46] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, Pierre-Antoine Manzagol, and Léon Bottou. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, 11(12), 2010.
- [47] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [48] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022.
- [49] Christoph Feichtenhofer, Yanghao Li, Kaiming He, et al. Masked autoencoders as spatiotemporal learners. *Advances in neural information processing systems*, 35:35946–35958, 2022.
- [50] Mengyue Zha, SiuTim Wong, Mengqi Liu, Tong Zhang, and Kani Chen. Time series generation with masked autoencoder. *arXiv preprint arXiv:2201.07006*, 2022.

- [51] Âzeddine Frimane and Jamie M Bright. Validation of synthetic solar irradiance data. In *Synthetic Solar Irradiance: Modeling Solar Data*, pages 4–1. AIP Publishing LLC Melville, New York, 2021.
- [52] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11), 2008.
- [53] Geoffrey E Hinton and Sam Roweis. Stochastic neighbor embedding. *Advances in neural information processing systems*, 15, 2002.
- [54] Karl Pearson. LIII. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572, 1901.
- [55] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.
- [56] Vladimir Risojević. Multimedijalni sistemi. *Univerzitet u Banjoj Luci/Elektrotehnički fakultet (Banja Luka)*, 2018.
- [57] Michael E Wall, Andreas Rechtsteiner, and Luis M Rocha. Singular value decomposition and principal component analysis. In *A practical approach to microarray data analysis*, pages 91–109. Springer, 2003.
- [58] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [59] Aron P Dobos. PVWatts version 5 manual. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2014.
- [60] Richard Perez, Pierre Ineichen, Robert Seals, Joseph Michalsky, and Ronald Stewart. Modeling daylight availability and irradiance components from direct and global irradiance. *Solar energy*, 44(5):271–289, 1990.
- [61] John Boland, Barbara Ridley, and Bruce Brown. Models of diffuse solar radiation. *Renewable Energy*, 33(4):575–584, 2008.

Biografija autora

Romanela Lajić rođena je 27.7.1997. godine u Banjoj Luci. Završila je osnovnu školu "Jovan Dučić" u Lamovitoj 2012. godine i proglašena učenikom generacije. Gimnaziju u Banjoj Luci završila je 2016. godine. Tokom osnovnog i srednjeg obrazovanja učestvovala je i nagrađivana na raznim takmičenjima i manifestacijama. 2016. godine upisala je prvi ciklus studija na Elektrotehničkom fakultetu Univerziteta u Banjoj Luci, na smjeru elektroenergetika i automatika. Završni rad prvog ciklusa studija, pod naslovom "Digitalno upravljanje elektromotornim pogonom sa BLDC motorom", odbranila je 2020. godine. Tokom studiranja više puta je nagrađivana kao jedan od najboljih studenata na manifestacijama povodom dana fakulteta. Od 2020. godine zaposlena je kao asistent na Elektrotehničkom fakultetu Univerziteta u Banjoj Luci, na katedri za opštu elektrotehniku. Autor je više naučnih radova. Sem maternjeg, srpskog jezika, tečno govori engleski i njemački jezik.

УНИВЕРЗИТЕТ У БАЊОЈ ЛУЦИ
ПОДАЦИ О АУТОРУ ОДБРАЊЕНОГ МАСТЕР/МАГИСТАРСКОГ РАДА

Име и презиме аутора мастер/магистарског рада **Романела Лајић**

Датум, мјесто и држава рођења аутора **27.7.1997. год. Бања Лука, Босна и
Херцеговина**

Назив завршеног факултета/Академије аутора и година дипломирања

Електротехнички факултет Универзитета у Бањој Луци, 2020.

Датум одбране завршног/дипломског рада аутора **28.9.2020.**

Наслов завршног/дипломског рада аутора **„Дигитално управљање електромоторним
погоном са BLDC мотором“**

Академско звање коју је аутор стекао одбраном завршног/дипломског рада

Дипломирани инжињер електротехнике-240 ECTS бодова

Академско звање које је аутор стекао одбраном мастер/магистарског рада **Мастер
електротехнике-300 ECTS -Електроенергетски и индустријски системи**

Назив факултета/Академије на коме је мастер/магистарски рад одбрањен

Електротехнички факултет Универзитета у Бањој Луци

Наслов мастер/магистарског рада и датум одбране **„Примјена неуронских мрежа за
генерисање синтетичких података о соларној ирадијанси“, 9.2.2024. год.**

Научна област мастер/магистарског рада према CERIF шифрарнику **T 121**

Имена ментора и чланова комисије за одбрану мастер/магистарског рада

проф. др Зденка Бабић, предсједник
проф. др Владимир Рисојевић, ментор
проф. др Чедомир Зељковић, члан

У Бањој Луци, дана 31. 1. 2024.

Декан



ИЗЈАВА О АУТОРСТВУ

**Изјављујем да је
мастер/магистарски рад**

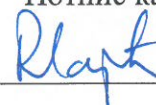
Наслов рада Примјена неуронских мрежа за генерисање синтетичких података о соларној ирадијанси

Наслов рада на енглеском језику Generating synthetic solar irradiance data using neural networks

- резултат сопственог истраживачког рада,
- да мастер/магистарски рад, у цјелини или у дијеловима, није био предложен за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио интелектуалну својину других лица.

У Бањој Луци 31.1.2024. године

Потпис кандидата



**Изјава којом се овлашћује Електротехнички факултет/ Академија умјетности
Универзитета у Бањој Луци да мастер/магистарски рад учини јавно доступним**

Овлашћујем Електротехнички факултет/ Академију умјетности Универзитета у Бањој Луци да мој мастер/магистарски рад, под насловом

Примјена неуронских мрежа за генерисање синтетичких података о соларној ирадијанси

који је моје ауторско дјело, учини јавно доступним.

Мастер/магистарски рад са свим прилозима предао/ла сам у електронском формату, погодном за трајно архивирање.

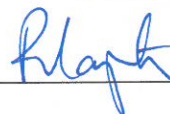
Мој мастер/магистарски рад, похрањен у д и г и т а л н и р е п о з и т о р и ј у м Универзитета у Бањој Луци, могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (*Creative Commons*), за коју сам се одлучио/ла.

1. Ауторство
2. Ауторство - некомерцијално
3. Ауторство - некомерцијално - без прераде
4. Ауторство - некомерцијално - дијелити под истим условима
5. Ауторство - без прераде
6. Ауторство - дијелити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, кратак опис лиценци дат је на полеђини листа).

У Бањој Луци 31.1.2024. године

Потпис кандидата



Изјава о идентичности штампане и електронске верзије
мастер/магистарског рада

Име и презиме аутора Романела Лајић

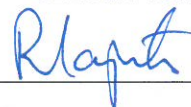
Наслов рада Примјена неуронских мрежа за генерисање синтетичких података о соларној ирадијанси

Ментор проф. др. Владимир Рисојевић

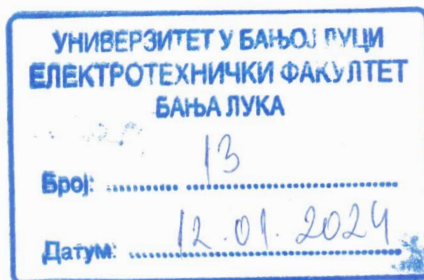
Изјављујем да је штампана верзија мог мастер/магистарског рада идентична електронској верзији коју сам предао/ла за дигитални репозиторијум Универзитета у Бањој Луци.

У Бањој Луци 31.1.2024. год.

Потпис кандидата



УНИВЕРЗИТЕТ У БАЊОЈ ЛУЦИ
ЕЛЕКТРОТЕХНИЧКИ ФАКУЛТЕТ
Патре 5
78000 Бања Лука



Др Зденка Бабић, редовни професор
Универзитет у Бањој Луци, Електротехнички факултет

Др Владимир Рисојевић, ванредни професор
Универзитет у Бањој Луци, Електротехнички факултет

Др Чедомир Зељковић, ванредни професор
Универзитет у Бањој Луци, Електротехнички факултет

НАУЧНО – НАСТАВНОМ ВИЈЕЋУ ЕЛЕКТРОТЕХНИЧКОГ ФАКУЛТЕТА УНИВЕРЗИТЕТА У БАЊОЈ ЛУЦИ

Одлуком Научно–наставног вијећа Електротехничког факултета Универзитета у Бањој Луци број 20/3.437-4/23 од 24.06.2023. године, именовани смо за чланове Комисије за завршни рад II циклуса студија кандидата Романела Лајић, дипл. инж. ел., под називом „Примјена неуронских мрежа за генерисање синтетичких података о соларној ирадијанси“. Након прегледа приложеног рада, подносимо сљедећи

ИЗВЈЕШТАЈ

1. БИОГРАФСКИ ПОДАЦИ КАНДИДАТА

Романела Лајић је рођена 27.07.1997. године. Гимназију у Бањој Луци је завршила 2016. године. Исте године је уписала Електротехнички факултет Универзитета у Бањој Луци, студијски програм Електроенергетика и аутоматика. Дипломирала је 2020. године са темом завршног рада “Дигитално управљање електромоторним погоном са BLDC мотором” и просјечном оцјеном у току студија 9,33. Други циклус студија на Електротехничком факултету, студијски програм, Електротенергетски и индустријски системи, је уписала 2020. године и положила све испите предвиђене планом и програмом. Тренутно је запослена као асистент на Катедри за Општу електротехнику Електротехничког факултета. До сада је објавила четири рада:

1. Romanela Lajić, Petar Matić, “Digital Position Control System With a BLDC Motor Using Field Oriented Control”, In *Proceedings of the 2023 22nd International Symposium INFOTEH-JAHORINA (INFOTEH)*, IEEE, ISSN: 2767-9470 (Online), 15-17 March 2023, pp. 1-6
2. Romanela Lajić, Petar Matić, “Speed Control of BLDC Motor With Ripple Effect Reduction Using Recurrent Wavelet Neural Networks”, *International Journal of Electrical Engineering and Computing*, 6(2), pp. 57-64, 2022

- Romanela Lajić, Čedomir Zeljković, Vladimir Risojević, "Generating Synthetic Hourly Solar Radiation Sequences Using Generative Adversarial Networks", In *Proceedings of the 30th International Electrotechnical and Computer Science Conference ERK 2021*, ISSN: 2591-0442 (online), Portoroz, Slovenia, 20-21. September 2021.
- Romanela Lajić, "Digitalno upravljanje elektromotornim pogonom sa BLDC motorom", In *Proceedings of the 2021 20th International Symposium INFOTEH-JAHORINA (INFOTEH)*, IEEE, 17-19 March 2021.

2. ОСНОВНИ ПОДАЦИ О РАДУ

Завршни рад другог циклуса студија кандидата Романеле Лајић, дипл. инж. електротехнике, под називом "Примјена неуронских мрежа за генерисање синтетичких података о соларној ирадијанси" има обим од 74 странице. Садржи насловну страну на српском и енглеском језику, информације о ментору и раду на српском и енглеском језику, садржај, посвету, списак табела (укупно 4 табеле), списак слика (укупно 32 слике) и списак скраћеница. Рад је организован у седам поглавља:

- Увод
- Потреба за синтетичким подацима о соларној ирадијанси
- Мјерење и репрезентација података о соларној ирадијанси
- Алгоритми за генерисање синтетичких података
- Методологија
- Експериментални резултати
- Закључак

На крају рада дати су преглед кориштене литературе (61 референца) и биографија кандидата.

У првој глави завршног рада је уочен растући значај соларне енергије у модерном свијету који се огледа у повећању глобалне инсталисане снаге соларних произвођача те у смањењу просјечне цијене kWh електричне енергије добијене из фотонапонских панела. Са порастом употребе соларне енергије порасла је и потреба за механизмима за оптимизацију рада постојећих соларних постројења као и алатима за планирање нових постројења. Имајући у виду да је прикупљање података о сунчевом зрачењу у дужем временском периоду тешко, као алтернатива кориштењу измјерених података у планирању и оптимизацији предложено је генерисање кориштење секвенци соларне ирадијансе генерисане кориштењем неуронских мрежа. У овој глави је изложена и структура самог рада.

У другој глави је изложена мотивација за генерисање и кориштење синтетичких података о соларној ирадијанси за дугорочно планирање и краткорочну предикцију рада соларне електране. У наставку је укратко описана типична метеоролошка година, као класичан алат за моделовање соларног капацитета неког подручја и уочени су њени недостаци за чије отклањање је предложено кориштење синтетичких података. Такође су изложени често кориштени модели за краткорочну предикцију сунчевог зрачења што је основни предуслов за планирање рада соларне електране. Наглашено је да је основна препрека при кориштењу наведених модела потреба за њиховим обучавањем на историјским подацима који нису увијек доступни.

У трећој глави су описани инструменти и процедуре за мјерење података о соларној ирадијанси. Такође су описани и начини репрезентације података погодни за примјену у сврху обучавања неуронских мрежа. Конкретно, дискутована је могућност кориштења индекса ведрине и индекса ведрога неба за нормализацију података и уклањање детерминистичке компоненте из измјерених вриједности ирадијансе чиме се модели за генерисање синтетичких података могу фокусирати искључиво на стохастичку компоненту података.

У четвртој глави су описани модели који су у литератури кориштени за генерисање различитих типова временских низова. Најприје су изложени Марковљеви ланци, као класичан метод за генерисање временских низова, а посебна пажња је посвећена техникама заснованим на кориштењу неуронских мрежа, као што су генеративне супротстављајуће мреже и маскирани аутоенкодерси.

У петој глави је изложена кориштена методологија за припрему података за обучавање мрежа, сам поступак обучавања модела и методи за валидацију синтетичких података. Пошто још увијек не постоји стандардна процедура према којој се провјерава валидност синтетичких података, у овом раду су генерисани подаци најприје квалитативно упоређени са стварним, а затим је анализирана њихова употребљивост са становишта практичних примјена, дугорочног планирања рада електране и краткорочне предикције соларне ирадијансе.

Добијени експериментални резултати су приказани у шестој глави. Дати су резултати добијени кориштењем четири различита модела, обучавана на временским секвенцама различитих дужина. Квалитативна оцјена синтетичких података изведена је визуелизацијом синтетичких и стварних података трансформисаних кориштењем PCA и tSNE трансформација, а оцјена употребљивости обучавањем AR и LSTM предиктора на синтетичким и стварним подацима и поређењем резултујуће средњеквадратне и средње апсолутне грешке предикције на секвенцама стварних података. Коначно, употребљивост синтетичких података за дугорочно планирање рада соларне електране анализирана је кориштењем калкулатора снаге фотонапонског система.

У седмој глави су, на основу експерименталних резултата, изведени закључци. Изложене су главне предности и недостаци имплементираних генератора синтетичких секвенци, те потенцијални правци за даље истраживање у предметној области.

3. АНАЛИЗА И НАЈВАЖНИЈИ ДОПРИНОСИ РАДА

Разматрајући завршни рад другог циклуса студија кандидата Романеле Лајић, Комисија је закључила да својим садржајем, постигнутим резултатима и закључцима задовољава критеријуме који се постављају пред завршни рад другог циклуса студија. Рад у цјелини има добро систематизовану структуру и план излагања. Наслов рада је јасно формулисан, разумљив, прецизно описује предмет истраживања и у потпуности указује на садржај рада.

Свеобухватном теоријском анализом као и конкретним експерименталним радом, кандидат Романела Лајић је показала зрелост и способност да савлада и систематизује знања из једне истраживачке области.

Имајући у виду значај проблема планирања рада соларне електране и предикције сунчевог зрачења, тешкоће у прикупљању историјских података потребних за обучавање модела, те актуелност истраживања генеративних модела у области машинског учења, овај завршни рад обухвата актуелна истраживања и представља допринос стању у области.

Комисија констатује да је рад написан у складу са образложењем у пријави теме, као и да су остварени сви резултати који су били и планирани у образложењу пријаве теме:

А. Обучавање модела за генерисање синтетичких података о соларној ирадијанси

Анализом литературе која се бави генерисањем синтетичких временских низова установљено је да генеративни модели засновани на неуронским мрежама имају потенцијал за генерисање синтетичких секвенци соларне ирадијансе. За обучавање су кориштени измјерени подаци о соларној ирадијанси и метеоролошким параметрима на подручју Бање Луке. Подаци за обучавање су претпроцесирани, недостајући подаци су интерполирани и, рачунањем индекса ведрог неба, уклоњена је детерминистичка компонента из секвенци што значи да модели треба да науче да генеришу само стохастичку компоненту соларне ирадијансе. У раду су обучена четири модела за генерисање синтетичких секвенци соларне ирадијансе: LSTM-GAN, C-RNN-GAN, TimeGAN и ExtraMAE. Кориштењем обучених модела генерисане су синтетичке секвенце које ће бити кориштене за евалуацију квалитета генерисаних података и самих модела.

Б. Квалитативна евалуација генерисаних синтетичких података

Квалитет генерисаних синтетичких података је оцијењен њиховим поређењем са стварним подацима. Извршена је редукција димензионалности података кориштењем PCA и tSNE метода, након чега су подаци визуелизовани у дводимензионалном простору. Поређењем визуелизација синтетичких и стварних података, установљено је да су секвенце генерисане TimeGAN и ExtraMAE моделима много сличније стварним него секвенце генерисане LSTM-GAN и C-RNN-GAN моделима.

В. Анализа употребљивости генерисаних података у примјенама

Да би се испитала употребљивост генерисаних података у примјенама које захтијевају велике количине историјских података о соларној ирадијанси, одабрани су проблеми краткорочне предикције соларне ирадијансе и дугорочне предикције производње соларне електране. У оба проблема су поређени резултати добијени кориштењем синтетичких и стварних секвенци. За краткорочну предикцију кориштени су AR модел и LSTM неуронска мрежа, а као мјере перформанси су кориштене средња апсолутна и средњеквадратна грешка предикције. Експериментални резултати показују да перформансе предиктора обученог на синтетичким подацима не заостају много за перформансама предиктора обученог на стварним подацима. У другом тесту су стварни и генерисани подаци искориштени за рачунање енергије коју генерише фотонапонски систем. Добијени резултати показују да се добијене вриједности у великој мјери слажу те да постоји изражена варијабилност у синтетичким подацима што их чини погодним за анализу различитих сценарија у раду електране.

4. ЗАКЉУЧАК И ПРИЈЕДЛОГ

Комисија сматра да завршни рад II циклуса студија под називом „Примјена неуронских мрежа за генерисање синтетичких података о соларној ирадијанси“, кандидата Романеле Лајић, дипл. инж. електротехнике, садржи све потребне елементе и резултате којима су остварени постављени циљеви истраживања, те предлаже Научно-наставном вијећу Електротехничког факултета Универзитета у Бањој Луци да усвоји извјештај Комисије и одобри заказивање јавне усмене одбране.

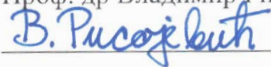
Бања Лука, 11.01.2024. године

Комисија:

Проф. др Зденка Бабић, предсједник



Проф. др Владимир Рисојевић, ментор



Проф. др Чедомир Зељковић, члан