

DUGOROČNA PROGNOZA POTROŠNJE ELEKTRIČNE ENERGIJE ZASNOVANA NA LINEARNOM MODELU**LONG-TERM LOAD FORECAST BASED ON A LINEAR MODEL**Kristina Polih, *Fakultet tehničkih nauka, Novi Sad***Oblast – Elektrotehničko i računarsko inženjerstvo****Kratak sadržaj** – Rad se bavi problemom dugoročne prognoze potrošnje električne energije. Cilj rada jeste da se upotrebi jednostavan model i pokaže korelacija temperature vazduha sa potrošnjom električne energije. Korišteni su SVM i MLR algoritmi, koji rade prognozu potrošnje na osnovu linearnog modela, dok su za određivanje tačnosti upotrebljeni kvantili, MAPE, MSE, PLF i Winkler funkcija.**Gljučne reči:** Dugoročna prognoza potrošnje električne energije, regresioni algoritmi, kvantili, SVM, MLR**Abstract** – The paper deals with the problem of long-term forecast of electricity consumption. The aim of this paper is to use a simple model and show the correlation of air temperature with electricity consumption. SVM and MLR algorithms were used, which perform consumption forecasting based on a linear model, while quantiles, MAPE, MSE, PLF and Winkler functions were used to determine accuracy.**Keywords:** Long-term load forecast, Regression algorithms, quantiles, MLR, SVM**1. UVOD**

Prognoza potrošnje električne energije, omogućava da se preciznije mogu odrediti buduće akcije koje su potrebne za normalan rad elektroenergetskog sistema. Efikasnije se može rukovati sistemom, kao i kupovinom i prodajom energije na berzi, ako se zna koliko će biti potrebno energije za napajanje svih korisnika.

Prilikom proračuna prognoze potrošnje, mora se voditi računa o parametrima koji nisu statični i koji uvode neizvesnost u same proračune, kao što su atmosferske prilike tj. temperatura vazduha, kiša, brzina vetra, vlažnost vazduha itd.

Dinamika promenljivih parametara čini dugoročnu prognozu zahtevnijom, pa i težom, a samim tim i izazovnijom za istraživanje.

Ovde se posmatra dugoročna prognoza potrošnje na dnevnom nivou za narednih godinu dana. Pretpostavljeno je da predviđanje prognoze potrošnje može biti poboljšano pretragom po sličnim danima (eng. Similar Day search). U radu su upoređeni MLR i SVM algoritmi, nad istim skupom podataka.

Algoritmi su obučavani i testirani na jednostavnom linearnom modelu sa ciljem da se pokaže korelacija između temperature vazduha i potrošnje električne energije.

NAPOMENA:

Ovaj rad proistekao je iz master rada čiji mentor je bio dr Aleksandar Erdeljan, red. prof.

2. DUGOROČNA PROGNOZA POTROŠNJE ELEKTRIČNE ENERGIJE

Prognoza potrošnje električne energije se može podeliti u četiri kategorije: prognoza bliska realnom vremenu – horizont predikcije od nekoliko sati, kratkoročna prognoza potrošnje električne energije (STLF) – obično predstavlja kratak interval od par dana do dve ili tri nedelje, srednjeročna prognoza potrošnje električne energije (MTLF) – često je definisana u periodu od jednog meseca pa do jedne godine i dugoročna prognoza potrošnje električne energije (LTLF) – predstavlja prognozu koja se izvodi od jedne godine pa čak i do trideset godina.

Za različite vrste prognoza potrošnje, mogu da variraju faktori koji će biti prioritetni. Tačnost prognoza takođe varira u odnosu na to koja prognoza je u pitanju.

LTLF omogućava da se uoči kada će biti vrhunac potrošnje u posmatranom periodu, i da se u skladu sa prognoziranom potrošnjom pripremi sistem. Vremenski period za koji se radi dugoročna potrošnja, može biti od jedne godine pa čak do nekoliko decenija. Dugoročne prognoze od nekoliko decenija se koriste za planiranja i proširivanja od strane preduzeća koja se bave proizvodnjom, prenosom, distribucijom električne energije, a najčešće se koriste za uređenje i razvijanje elektroenergetskih sistema. Na osnovu radova koji su napisani na ovu temu, izabrani su MLR i SVM algoritmi, radi njihovog poređenja [1].

Potrošnja zavisi od mnogobrojnih faktora, koji mogu da variraju iz dana u dan, pa samim tim i prilikom proračuna prognoze potrošnje se moraju uzeti u obzir svi važni elementi koji utiču na potrošnju. [6] Jedan od bitnijih faktora su vremenske prilike (temperatura i vlažnost vazduha, vetar, osunčanost i sl.), od čega zavisi da li će korisnici uključiti grejalice, klime, itd. Ako je jako visoka ili niska temperatura, veliki deo stanovništva će uključiti belu tehniku za grejanje ili rashlađivanje prostora ili stanova. Stoga, predviđanje potrošnje električne energije zavisi od vremenske prognoze i naravno temperature.

3. DETERMINISTIČKI ALGORITMI

Deterministički pristup je korišten u cilju izbora optimalnih parametara, npr. koliko je dana potrebno da se iskoristi za obučavanje modela da bi se dobili dovoljno tačni rezultati, itd. Zajednički parametri za sve varijacije modela su: broj dana nad kojim se radi pretraga po sličnom danu, granica temperature i broj dana koji su ulaz u model. Prilikom odabira optimalnih parametara, ulazni modeli se razlikuju, po broju dana nad kojima se radi pretraga po sličnom danu, po graničnim vrednostima temperature, koja je parametar za pretragu po sličnim

danima i po broju dana koji su ulaz za obučavanje modela.

3.1. Višestruka linearna regresija (MLR)

MLR pokušava da uspostavi odnos između dve ili više nezavisnih promenljivih i jedne zavisne promenljive. Iako odnos između zavisne i nezavisnih promenljivih može biti nelinearan u ovom radu je MLR bazirana nad sledećim pretpostavkama [8]: 1. linearna zavisnost između zavisne i nezavisne promenljive, 2. nezavisne promenljive nisu u međusobnoj korelaciji, 3. varijacija ostataka je konstantna, 4. nezavisnost posmatranja, 5. multivarijantna normalnost.

Prema tome, vrednost predviđanja y je

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

gde su: β_0 – presek tj. vrednost predviđanja kada su svi ulazi nula, $\beta_1, \beta_2, \dots, \beta_n$ – koeficijenti regresije, X_1, X_2, \dots, X_n – ulazi kao nezavisne promenljive, ε – greška predviđanja.

3.2. Support Vector Machine (SVM)

SVM su nadzirani obučavajući modeli koji su povezani sa određenim algoritmima za potrebe analiziranja podatka potrebnih za klasifikaciju i regresione analize. Problem linearnog regresionog SVM algoritma jeste da pronađe funkciju koja radi aproksimaciju preslikavanja ulaznih podataka u realne brojeve, na osnovu obučavajućih uzoraka [8].

3.3. Algoritmi za procenu tačnosti

Najvažniji deo, prilikom upotrebe svakog algoritma je određivanje koliko je tačan. U zavisnosti od toga koja se metrika koristi za procenu tačnosti, mogu se dobiti bolji i lošiji rezultati [10].

U radu će biti korišteni: Mean Absolute Percentage Error (MAPE) i Mean Square Error (MSE). Iako izgleda da je MAPE veoma jednostavan za upotrebu, postoji mnogi nedostaci u praktičnoj primeni, takođe postoje i mnoga istraživanja o nedostacima i o obmanjujućim rezultatima MAPE algoritma [11].

4. PROBABILISTIČKI ALGORITMI

Algoritam A se ponaša kao matematička funkcija preslikavanja tj. mapiranja i to na takav način da ako primenimo A na isti ulaz X, nekoliko puta, uvek ćemo dobiti isti izlaz Y. Matematički zapis takve funkcije bi bio: $A : X \rightarrow Y$, gde A predstavlja deterministički algoritam sa ulazima X i izlazima Y. U odnosu na deterministički algoritam, probabilistički algoritam A je takav da njegovo ponašanje, u najvećem broju slučajeva, kontrolišu slučajni događaji. Prilikom proračuna izlaza Y, koristeći ulaz X, rezultat algoritma zavisi od ograničenog broja nasumičnih eksperimenata.

Konkretno, primenom algoritma A na isti ulaz X, više puta, mogu se dobiti potpuno različiti rezultati [16-17]. Algoritmi su u prvoj iteraciji korišteni kao deterministički, radi utvrđivanja optimalnog izbora parametara. Nakon čega su modifikovani po pravilima koja nalažu probabilistički algoritmi, tako da koriste slučajan parametar koji algoritmima obezbeđuje nasumičnost.

4.1. Algoritmi za procenu tačnosti

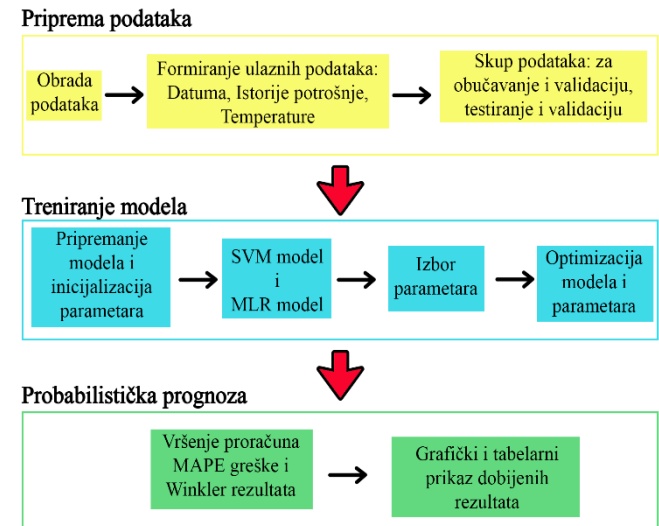
Funkcije gubitaka su veoma važne i u teoriji i u praksi predviđanja. Prilikom određivanja tačnosti probabilistič-

kih algoritama često se koriste funkcije gubitaka u kombinaciji sa kvantilima [17]. Kvantili predstavljaju presečne tačke koje odvajaju opsege raspodele verovatnoće na neprekidne intervale sa jednakim verovatnoćama. Za probablističke rezultate su korišteni: Pinball Loss Function (PLF) i Winkler Function. Pošto Winkler i PLF računaju vrednosti za jedan i-ti vremenski trenutak, funkcije bodovanja su korištene da bi se napravio prosek vrednosti: Quantile Score (QS) i Winkler Score (WS).

5. IMPLEMENTACIJA

U radu su korišteni linearni regresioni SVM algoritam i MLR algoritam. Iskorišteni su implementirani algoritmi koji se nalaze u Accord.NET biblioteci. SVM algoritam koristi pojednostavljenu verziju Chih-Jen Lin i Jorge Moré's TRON, Newton-ove metode za rešavanje velikih optimizaciono-ograničenih problema [1].

Upotreba algoritama i implementacija se grubo može podeliti na tri etape: priprema podataka, treniranje modela i prikazivanje rezultata probablističke prognoze. Proces probablističke prognoze je prikazan na slici 5.1.



Slika 5.1. Tok implementacije probablističke prognoze potrošnje

Korišteni podaci su u periodu od 2007. do 2015. godine, na dnevnom nivou, na nivou Srbije. Potrošnja električne energije je izražena u MW jedinicama, a pribavljena je sa sajta Entsoe za potreban period od 2007 do 2015. Izvor podataka koji je objavio ENTSO-E su OPS-ovi članovi [19]. Potrošnja je bila predstavljena u satima za svaki dan, pa je za potrebe ovog rada urađen prosek na dnevnom nivou. Korištena je prosečna temperatura na nivou Srbije za posmatrani period, za svaki dan. Temperatura je pribavljena sa POWER Data Access sajta [20].

Skup podataka pomoću kojeg je određen optimalan deterministički pristup, koji je osnova za probablistički pristup i sa kojim su rađena dalja testiranja je organizovan u formatu Datum, Potrošnja i Temperatura.

Varijacije determinističkih algoritama koji su korišteni u radu se razlikuju po broju dana koje algoritam koristi za obučavanje i prognozu potrošnje. Sve varijacije algoritama koriste iste parametre i iste podatke za obučavanje. Parametri koji se koriste su prethodna potrošnja električne energije i srednja temperatura vazduha, kako bi se dokazala korelacija između potrošnje i temperature vazduha. Nakon odabira najtačnije varijacije, algoritmi su pobolj-

šani tako da koriste Similar Day pretragu (pretraga po sličnim danima), odnosno da pomoćni algoritam pretražuje određeni broj dana, kako bi našao vrednosti koje upadaju u postavljeni okvir.

Pretraga po sličnim danima je testirana na MLR algoritmu. U cilju poboljšanja tačnosti algoritma prilikom pretrage sličnih dana, kada se posmatra parametar temperatura, napravljen je pomoćni algoritam. Pomoćni algoritam beleži koje su vrednosti iskorištene u prethodnim iteracijama, jer bi se u suprotnom uvek dobili isti rezultati, kada se algoritam pokrene više puta.

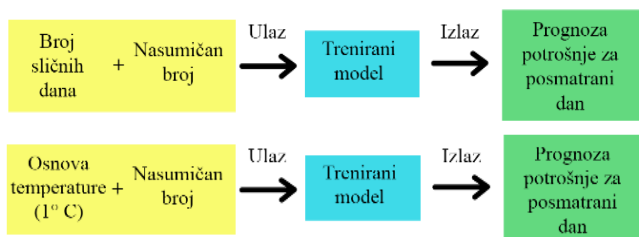
5.1. Predložena rešenja usled nedostatka temperature vazduha

Jedan od velikih problema na koje se nailazi kod dugoročne prognoze potrošnje je nedostatak predviđene temperature vazduha za duži vremenski period, npr. od jedne godine. Problem nedostatka temperature se najčešće prevazilazi, tako što se uradi srednja vrednost temperature ili se radi predviđanje temperature. Srednja vrednost temperature se računa na osnovu par godina podataka.

Predviđanje temperature je rađeno pomoću MLR algoritma, jer su se na prethodnom istraživanju dobili najtačniji rezultati. Opseg podataka koji je korišten za obučavanje modela je od 2007. godine zaključno sa 2011. godinom. Podaci koji su korišteni za predviđanje su u rasponu od 2012. do 2016. godine. Isti skup podataka je korišten za MLR, takođe i za SVM algoritam, dok za prosečenje temperature je potreban samo skup koji je bio korišten za predviđanje tj. podaci od 2012. do 2016. Upoređivanjem dobijenih rezultata greški, može se utvrditi da je najmanja greška dobijena kada je korišteno prosečenje temperature. Greška MLR algoritma je skoro 2.5 puta veća u odnosu na grešku koja je dobijena kada se koristi prosek temperature, dok je greška SVM algoritma najveća.

5.2. Probabilistički pristup

Pošto su opisani algoritmi primarno deterministički, onda je potrebno da se prilikom njihovog izvršavanja ubaci parametar koji je nasumičan, kako bi pomenuti algoritmi bili probabilistički. Jedan od načina na koji je implementiran probabilistički pristup u radu jeste da se dodaje nasumičan broj dana na osnovni broj dana koji se uzima za pronalaženje sličnih dana, što je dalje u tekstu referencirano kao *Prva varijacija*. Osnovni broj dana je pet, dok nasumični broj dana varira između jedan i petnaest, jer se prethodnim istraživanjem dokazalo da se tačnost algoritama pogoršava, kada se koristi preko dvadeset dana za pronalaženje sličnog dana. Na slici 5.2 prikazan je tok ulaza i izlaza podataka iz treniranog modela.



Slika 5.2. Prikaz ulaza i izlaza iz modela, kada je nasumičan parametar broj sličnih dana nad kojima se radi pretraga i kada je nasumičan parametar broj stepeni, respektivno

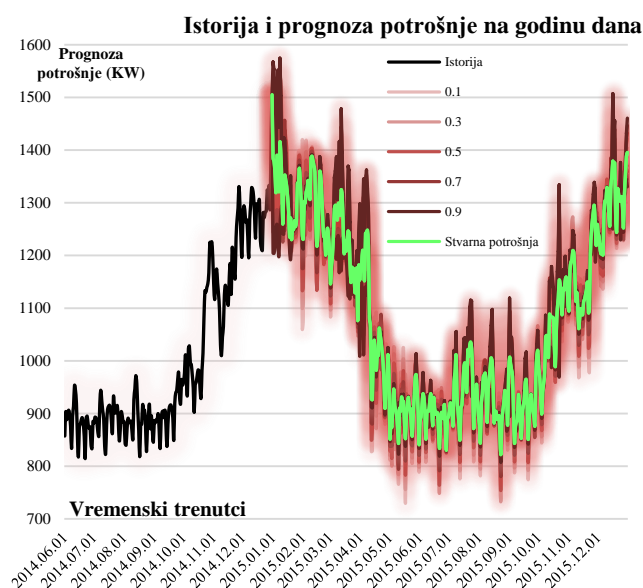
U trenirani model ulazi šest vrednosti od kojih su: tri istorija potrošnje i tri temperatura. Nasumičan broj samo povećava verovatnoću da se pronade dan koji je najbliži posmatranom danu po temperaturi.

Drugi probabilistički pristup koji je implementiran jeste da se dodaje nasumičan broj stepeni na osnovu, koji je dalje u tekstu referenciran kao *Druga varijacija*. Osnova na koju se dodaje nasumičan broj stepeni je jedan stepen, kao što je prikazano na slici 5.2. Nasumičan broj stepeni koji se dodaje na osnovu varira između nula i jednog stepena. U trenirani model ulazi šest vrednosti od kojih su: tri istorija potrošnje i tri temperatura.

Dodavanjem nasumičnog broja stepeni na osnovu, povećavamo verovatnoću da se pronade dan koji ima najviše sličnosti sa posmatranim danom.

6. ANALIZA REZULTATA

U nastavku će biti prikazani rezultati kada se koristi MLR algoritam za prognozu potrošnje u kombinaciji sa probabilističkim varijacijama algoritama i algoritmima za prognozu temperature. Korišteni algoritmi za prognozu temperature su MLR, SVM i Average (prosečenje). Procena tačnosti algoritama je rađena preko MAPE algoritma. Evaluacijom rezultata se može utvrditi, da prilikom upotrebe SVM algoritma za prognozu temperature u kombinaciji sa prvom varijacijom, dobije se najmanja greška prognoze potrošnje koja varira između 5.13% i 5.37%.



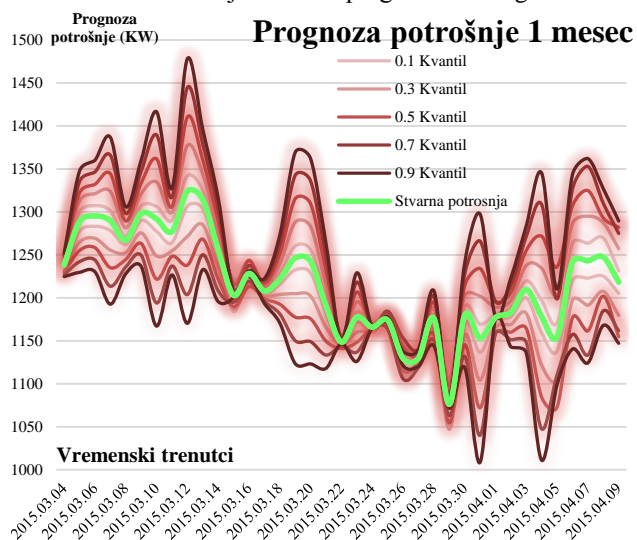
Slika 0.1. Prikaz istorije i prognoze potrošnje za jednu godinu, primarni MLR algoritam, MLR algoritam za prognozu temperature

Kada se koristi druga varijacija, najmanja greška se dobije ukoliko se koristi MLR algoritam za prognozu temperature. Greške tada variraju između 5.61% i 5.70%.

Na slici 6.1. prikazana je prognoza potrošnje, koristeći kvantile i Pinball Loss vrednosti. Primarni algoritam koji je upotrebljen je MLR, algoritam koji je korišten za prognozu temperature je takođe MLR.

Pošto je jako puno vrednosti prikazano na grafiku i kvantili ne mogu da se uoče, a samim tim i analiziraju, na sledećem grafiku su prikazane vrednosti prognoze u periodu od jednog meseca. Na grafiku 6.2., mogu se uočiti svi percentili od 0,1 do 0,9. Interval predviđanja je

procena intervala u kojem će završiti budući rezultati, ali sa određenom verovatnoćom. Svaki kvantil nam govori koliko je vrednost veća u odnosu na stvarnu potrošnju. Svi ostali rezultati su jako slični po grafičkom izgledu.



Slika 0.2. Izdvojeni prikaz prognoze potrošnje za jedan mesec, primarni MLR algoritam, MLR algoritam za prognozu temperaturu

Tačnost prognoze potrošnje, kada je korišten MLR algoritam je raznolika. Sa MAPE algoritmom greška je varirala između 5.05% i 5.71%, pri čemu je prosečna greška iznosila 5.45%. Kada je korišten Winkler Score za procenu tačnosti, greška je u odnosu na SVM algoritam bila veća za skoro 2×10^3 . Prilikom upotrebe Quantile Score za procenu tačnosti, greška je bila značajno manja u odnosu na SVM algoritam.

Tačnosti prognoze potrošnje, prilikom upotrebe SVM algoritma su takođe varirale. Kada je korišten MAPE algoritam za procenu tačnosti, greška je varirala između 5.22% i 7.91%, a u proseku je iznosila 6.12%. Greška koja je dobijena, kada je korišten Winkler Score za procenu tačnosti, u odnosu na MLR algoritam je bila manja za skoro 2×10^3 . Kada je korišten Quantile Score za određivanje tačnosti algoritma, dobijena greška je veća u odnosu na MLR algoritam.

7. ZAKLJUČAK

Električna energija je postala neophodna u svakodnevnom životu. Shodno tome, planiranje potražnje i potrošnje električne energije postaje sve popularnije, a moglo bi se reći skoro pa neophodno.

Upoređivanjem MLR i SVM algoritma, može se zaključiti, da je MLR algoritam bolja opcija za dugoročnu prognozu potrošnje za posmatrani skup podataka. Ako bi se upotrebili hibridni algoritmi i selekcija ili odabir podataka, verovatno bi se dobila bolja tačnost algoritama.

8. LITERATURA

[1] J. H. Pujar, "Fuzzy Ideology based Long Term Load Forecasting", Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering, 2010

[2] D. Ali, M. Yohana, P. M. Ijasini, M. B. Garkida, "Application of fuzzy – Neuro to model weather parameter variability impacts on electrical load based on long-term forecasting", Alexandria Engineering Journal, Volume 57, Issue 1, 2018, Pages 223-233, ISSN 1110-0168, 2017

[3] D. Ali, M. Yohana, M.I. Puwu, B.M. Garkida, "Long-term load forecast modelling using a fuzzy logic approach", Natural Science and Engineering, Volume 18, Issue 2, 2016, Pages 123-127

[4] Xu Liwen, Li Chengdong, Xie Xiuying, Zhang Guiqing, "Long-Short-Term Memory Network Based Hybrid Model for Short-Term Electrical Load Forecasting", Information 9, no. 7: 165, 2018

[5] Tao Hong, Jason Wilson, Jingrui Xie, "Long Term Probabilistic Load Forecasting and Normalization With Hourly Information", Ieee Transactions On Smart Grid, Vol. 5, 456-462, 2014

[6] N. Ayub, N. Javaid, S. Mujeeb, M. Zahid, W. Z. Khan, and M. U. Khattak, "Electricity Load Forecasting in Smart Grids Using Support Vector Machine", COMSATS University Islamabad, Islamabad 44000, 2019

[7] Yasin, Z. M., Salim, N. A., Aziz, N. F. A., Ali, Y. M., & Mohamad, H, "Long-term load forecasting using grey wolf optimizer-least-squares support vector machine", IAES International Journal of Artificial Intelligence, 2020

[8] Berry, W. D. "Understanding Regression Assumptions, Sage University Paper series on Quantitative Applications in the Social Sciences", Newbury Park, 1993

[9] M. Pontil, A. Verri, "Properties of Support Vector Machines", INFM, Via Dodecaneso, Genova, 1998

[10] B. Dalvi, A. Mishra, W. W. Cohen, "Hierarchical semi-supervised classification with incomplete class hierarchies", Carnegie Mellon Uni., Uni. of Massachusetts, 193-202, 2016

[11] J. V. Mynsbrugge, "Bidding Strategies Using Price Based Unit Commitment in a Deregulated Power Market", K.U.Leuven, 2010

[12] Tofallis, "A Better Measure of Relative Prediction Accuracy for Model Selection and Model Estimation", 2015

[13] Hyndman, Rob J., A. B. Koehler, "Another look at measures of forecast accuracy" IJOF, 2006

[14] Kim, Sungil, H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts", 2016

[15] Makridakis, Spyros, "Accuracy measures: theoretical and practical concerns", IJOF, 9(4):527-529, 1993

[16] Delfs H., Knebl H., "Probabilistic Algorithms. In: Introduction to Cryptography. Information Security and Cryptography", Springer, Berlin, Heidelberg, 2007

[17] Tilmann Gneiting, "Quantiles as optimal point forecasts, International Journal of Forecasting", Volume 27, Issue 2, 2011, Pages 197-207

[18] Chih-Jen Lin, Jorge J. More, "Newton's Method For Large Bound-Constrained Optimization Problems", SIAM J. OPTIM, vol. 9, no. 4, pp. 1100-1127, 1999 Society For Industrial And Applied Mathematics

[19] <https://www.entsoe.eu/data/power-stats/> (pristupljeno septembar 2021. godine)

<https://power.larc.nasa.gov/data-access-viewer/> (pristupljeno septembar 2021. godine)

Kratka biografija:



Kristina Polih rođena je u Novom Sadu 1996. god. Master rad na Fakultetu tehničkih nauka iz oblasti Elektrotehničko i računarsko inženjersvo – Primenjeno softversko inženjersvo, odbranila je master rad 2021.godine.
kontakt: christina.polih@yahoo.com