



UNIVERZITET U NIŠU
ELEKTRONSKI FAKULTET



Miloš M. Božić

**KRATKOROČNA PROGNOZA POTROŠNJE
ELEKTRIČNE ENERGIJE ZASNOVANA NA
METODAMA VEŠTAČKE INTELIGENCIJE**

DOKTORSKA DISERTACIJA

Niš, 2014.



UNIVERZITET U NIŠU
ELEKTRONSKI FAKULTET



Miloš M. Božić

**KRATKOROČNA PROGNOZA POTROŠNJE
ELEKTRIČNE ENERGIJE ZASNOVANA NA
METODAMA VEŠTAČKE INTELIGENCIJE**

DOKTORSKA DISERTACIJA

Niš, 2014.



UNIVERSITY OF NIŠ
FACULTY OF ELECTRONIC ENGINEERING



Miloš M. Božić

SHORT-TERM LOAD FORECASTING BASED ON ARTIFICIAL INTELLIGENCE METHODS

DOCTORAL THESIS

Niš, 2014.

INFORMACIJE O MENTORU I ČLANOVIMA KOMISIJE

Mentor: redovni profesor, dr Zoran Stajić, Univerzitet u Nišu, Elektronski fakultet.

Članovi komisije:

- redovni profesor, dr Zoran Stajić, Univerzitet u Nišu, Elektronski fakultet,
- redovni profesor, dr Milena Stanković, Univerzitet u Nišu, Elektronski fakultet,
- redovni profesor, dr Nikola Rajaković, Univerzitet u Beogradu, Elektrotehnički fakultet,
- docent, dr Lidija Korunović, Univerzitet u Nišu, Elektronski fakultet,
- docent, dr Aleksandar Janić, Univerzitet u Nišu, Elektronski fakultet.

Datum odbrane:

ZAHVALNOST

Zahvaljujem se svom mentoru, prof. dr Zoranu Stajiću, na značajnim savetima, uputstvima i sugestijama, koje je nesebično podelio sa mnom pri planiranju i pisanju disertacije.

Najiskrenije se zahvaljujem mom kolegi i velikom prijatelju dr Milošu Stojanoviću, koji mi je pružio veliku i bezrezervnu pomoć u toku istraživanja i realizaciji doktorske disertacije.

Najveću zahvalnost dugujem svojim roditeljima Olgi i Mališi i supruzi Marini, koji su me takođe podržavali i motivisali u toku naučno istraživačkog rada, kao i u izradi disertacije.

U Nišu, 2014. godine

Autor

REZIME

Kratkoročna prognoza potrošnje električne energije zasnovana na metodama veštačke inteligencije

Ova disertacija se bavi problemom kratkoročne prognoze potrošnje električne energije primenom regresionih metoda veštačke inteligencije. Predložena su tri nova modela koji za formiranje nelinearnog regresionog modela koriste *least squares support vector machines* metodu mašinskog učenja.

Pod pretpostavkom da će povećanje broja ulaznih promenljivih, pomoći pri formiranju preciznijeg modela, predložen je model sa prognozom u dve faze. Kod ovog modela se u prvoj fazi generiše prognoza vršnog dnevnog opterećenja koje se onda koristi kao dodatni atribut u drugoj fazi prognoze, gde se prognozira dnevna časovna potrošnja. Sa pretpostavkom dobijanja preciznijeg modela promenom strukture atributa inicijalnog modela, formiran je i model sa izborom atributa, koji za kriterijum izbora koristi količinu zajedničkih informacija. Predložen je i model sa dopunom trening skupa. Ovaj model se na određene intervale vremena retrenira koristeći novo formirani trening skup, dobijenim dopunom inicijalnog trening skupa novim trening vektorima za upravo prognozirani interval i izbacivanjem najstarijih vektora. Model se bazira na pretpostavci da će „osvežavanje“ modela tj. pružanje novih informacija doprineti poboljšanju preciznosti.

Evaluacija predloženih modela obavljena je korišćenjem podataka o električnom opterećenju grada Niša, koji ima oko 260 hiljada stanovnika sa prosečnim godišnjim dnevnim opterećenjem od 182 MW. Kao komparativni modeli korišćeni su dvostruko periodični ARIMA i Holt-Winters model, kao predstavnici grupe klasičnih modela ali i modeli veštačkih neuronski mreža, *least squares support vector machines* i *relevance vector machines* model, kao predstavnici grupe modela veštačke inteligencije. Komparativna analiza modela rađena je sa dobijenim rezultatima prognoze dnevnog časovnog opterećenja za period od godinu dana. Za metriku je korišćena srednja procentualna apsolutna greška i njena simetrična modifikacija, srednja kvadratna greška i apsolutna procentualna greška.

Rezultati su pokazali da najveću preciznost ima predloženi model sa dopunom trening skupa a slede ga dvostruko sezonski ARIMA i model veštačkih neuronskih mreža. Na začelju se nalaze *relevance vector machine model* i dvostruki Holt-Winters model. Pokazano je da se

pobednički model može uspešno koristiti kod rešavanja realnih problema kratkoročne prognoze potrošnje električne energije.

Ključne reči: kratkoročna prognoza potrošnje električne energije, veštačka inteligencija, mašinsko učenje, električno opterećenje, vremenske serije

Naučna oblast: Tehničko-tehnološke nauke

Uža naučna oblast: Elektrotehničko i računarsko inženjerstvo

UDK broj:

ABSTRACT

Short-term load forecasting based on artificial intelligence methods

The topic of this dissertation is a short-term load forecasting using artificial intelligence methods. Three new models with least squares support vector machines for nonlinear regression are proposed.

First proposed model is a model with forecasting in two stages. This model use additional feature, maximum daily load which is not known for day ahead. Forecasting of maximum daily load is obtained in the first stage. This forecasted value is used in second stage, where forecasting of hourly load is done. Model with feature selection, using mutual information for selection criteria, is a second proposed model. This model tries to find an optimal feature set for a given problem. Forecasting model based on an incremental update scheme is a third proposed model. This model is based on the incremental update of the initial training set by adding new instances into it as soon as they become available and throwing out the old ones. Then the model is trained with new training set. By this approach the evolving nature of the load pattern is followed and the model performance is preserved and improved.

For models evaluation, the forecasting of hourly loads for one year is done. Electrical consumption data for the City of Niš, which have about 260000 habitants and average daily demand of 182 MW, is used for testing. Double seasonal ARIMA and Holt-Winters as representatives of classical models and artificial neural networks, least squares support vector machines and relevance vector machines as representatives of artificial models, are used for models evaluation. For a measure of accuracy, mean absolute percentage error, symmetrical mean absolute percentage error, square root mean error and absolute percentage error are used.

Obtained results show that the best model is model with incremental update scheme, followed by double seasonal ARIMA and artificial neural networks models. The worst results are obtained by relevance vector machines and double seasonal Holt-Winters models. It has been shown that the best model could be successfully used with the short-term load forecasting problem.

Keywords: short-term electrical load forecasting, artificial intelligence, machine learning, electrical load, time series

SADRŽAJ

1	UVOD	1
1.1	PROCES PROGNOZE U ELEKTROENERGETSKIM PREDUZEĆIMA	1
1.2	POTREBE ZA PROGNOZOM POTROŠNJE ELEKTRIČNE ENERGIJE	2
1.3	PODELA PROGNOZE POTROŠNJE ELEKTRIČNE ENERGIJE	3
1.4	SADRŽAJ ISTRAŽIVANJA	6
1.5	ORGANIZACIJA DISERTACIJE	8
2	PREGLED OBLASTI	11
2.1	KLASIČNI MODELI	11
2.2	MODELI VEŠTAČKE INTELIGENCIJE	20
2.3	HIBRIDNI MODELI	25
3	NADGLEDANO MAŠINSKO UČENJE	31
3.1	DEFINICIJA MAŠINSKOG UČENJA	31
3.2	OBLAST MAŠINSKOG UČENJA	31
3.3	OSNOVE NADGLEDANOG MAŠINSKOG UČENJA	33
3.4	PROGNOZA VREMENSKIH SERIJA	40
3.4.1	Rekurzivna strategija	42
3.4.2	Direktna strategija	43
3.4.3	Direktno-rekurzivna strategija	43
4	TEORIJSKE OSNOVE	45
4.1	EKSPONENCIJALNO IZRAVNANJE	45
4.2	AUTOREGRESIONI INTEGRISANI PROCES POKRETNIH PROSEKA - ARIMA	48
4.3	VEŠTAČKE NEURONSKE MREŽE	52
4.4	LEAST SQUARES SUPPORT VECTOR MACHINES	60
4.4.1	Osnovne LS-SVM formulacije	61
4.4.2	Kernel funkcije	64
4.4.3	Rešavanje KKT sistema – CG algoritam	67
4.4.4	Tehnike za optimizaciju LS-SVM parametara	69
4.4.5	Dobijanje <i>sparse</i> LS-SVM rešenja	72
4.5	RELEVANCE VECTOR MACHINES	73
4.6	PROCENA ZAJEDNIČKIH INFORMACIJA	76

4.6.1	Definicija zajedničkih informacija	76
4.6.2	Procedura za procenu zajedničkih informacija	80
5	KARAKTERISTIKE POTROŠNJE ELEKTRIČNE ENERGIJE.....	83
6	MODELI ZA KRATKOROČNU PROGNOZU POTROŠNJE ELEKTRIČNE ENERGIJE.....	88
6.1	MODEL KRATKOROČNE PROGNOZE POTROŠNJE ELEKTRIČNE ENERGIJE SA PROGNOZOM U DVE FAZE	90
6.2	MODEL KRATKOROČNE PROGNOZE POTROŠNJE ELEKTRIČNE ENERGIJE SA IZBOROM ATRIBUTA.....	93
6.3	MODEL KRATKOROČNE PROGNOZE POTROŠNJE ELEKTRIČNE ENERGIJE SA DOPUNOM TRENING SKUPA	98
7	REZULTATI EKSPERIMENATA	101
7.1	POSTAVKE EKSPERIMENATA	101
7.2	REZULTATI I DISKUSIJA	105
8	ZAKLJUČAK.....	147

LISTA SLIKA

Slika 1.1 Podela prognoze potrošnje prema horizontu	6
Slika 1.2 Organizaciona šema disertacije	10
Slika 4.1 Struktura tipičnog biološkog neurona [107]	53
Slika 4.2 Struktura veštačkog neurona.....	53
Slika 4.3 Dijagram neurona	54
Slika 4.4 Aktivacione funkcije.....	54
Slika 4.5 Dijagram jednoslojne neuronske mreže.....	55
Slika 4.6 Dijagram višeslojne neuronske mreže sa više izlaza.....	57
Slika 4.7 Preslikavanje $\phi(x)$	64
Slika 4.8 Primarno-dualna interpretacija LS-SVM-a.....	66
Slika 4.9 Shrinking grid-search u dva nivoa.....	70
Slika 4.10 Primer unakrsnih validacija za k=10, 70% skupa za trening i 30% za validaciju modela	72
Slika 4.11 Simbolički prikaz zajedničkih informacija između X i Y	78
Slika 5.1 Srednje dnevno opterećenje i temperatura.....	84
Slika 5.2 Kriva električnog opterećenja u toku sedmice.....	85
Slika 5.3 Dnevno električno opterećenje	86
Slika 5.4 Kriva električnog opterećenja u danima praznika	87
Slika 6.1 Proces prognoze potrošnje električne energije u dve faze.....	93
Slika 6.2 Proces prognoze potrošnje električne energije sa izborom atributa	97
Slika 6.3 Proces prognoze potrošnje električne energije sa dopunom trening skupa	100
Slika 7.1 MAPE greška modela kod prognoze vršnog dnevnog opterećenja.....	107
Slika 7.2 RMSE greška modela kod prognoze vršnog dnevnog opterećenja	108
Slika 7.3 Histogram MAPE greške DS-ARIMA modela	109
Slika 7.4 Histograma MAPE greške DS-HW modela	110
Slika 7.5 Histogram MAPE greške ANN modela	110
Slika 7.6 Histogram MAPE greške RVM modela	111
Slika 7.7 Histogram MAPE greške LSSVM modela.....	111
Slika 7.8 Histogram MAPE greške 2F-LSSVM modela	112
Slika 7.9 Histogram MAPE greške IA-LSSVM modela	112
Slika 7.10 Histogram MAPE greške DTS-LSSVM modela.....	113

Slika 7.11 MAPE greška u januaru.....	114
Slika 7.12 MAPE greška u februaru	115
Slika 7.13 MAPE greška u martu	116
Slika 7.14 MAPE greška u aprilu	117
Slika 7.15 MAPE greška u maju.....	118
Slika 7.16 MAPE greška u junu.....	119
Slika 7.17 MAPE greška u julu.....	120
Slika 7.18 MAPE greška u avgustu	121
Slika 7.19 MAPE greška u septembru	122
Slika 7.20 MAPE greška u oktobru	123
Slika 7.21 MAPE greška u novembru.....	124
Slika 7.22 MAPE greška u decembru	125
Slika 7.23 Raspodela APE greške svakog sata po danima za DS-ARIMA model	128
Slika 7.24 Raspodela APE greške svakog sata po danima za DS-HW model.....	129
Slika 7.25 Raspodela APE greške svakog sata po danima za ANN model	129
Slika 7.26 Raspodela APE greške svakog sata po danima za RVM model.....	130
Slika 7.27 Raspodela APE greške svakog sata po danima za LSSVM model	130
Slika 7.28 Raspodela APE greške svakog sata po danima za 2F-LSSVM model.....	131
Slika 7.29 Raspodela APE greške svakog sata po danima za IA-LSSVM model.....	131
Slika 7.30 Raspodela APE greške svakog sata po danima za DTS-LSSVM model.....	132
Slika 7.31 Časovno opterećenje za period od 19. do 25. januara	134
Slika 7.32 Časovno opterećenje za period od 9. do 15. februara.....	135
Slika 7.33 Časovno opterećenje za period od 2. do 8. marta	136
Slika 7.34 Časovno opterećenje za period od 13. do 19. aprila	137
Slika 7.35 Časovno opterećenje za period od 18. do 24. maja	138
Slika 7.36 Časovno opterećenje za period od 22. do 28. juna	139
Slika 7.37 Časovno opterećenje za period od 6. do 12. jula	140
Slika 7.38 Časovno opterećenje za period od 3. do 9. avgusta.....	141
Slika 7.39 Časovno opterećenje za period od 21. do 27. septembra.....	142
Slika 7.40 Časovno opterećenje za period od 12. do 18. oktobra.....	143
Slika 7.41 Časovno opterećenje za period od 2. do 8. novembra	144
Slika 7.42 Časovno opterećenje za period od 21. do 27. decembra.....	145

LISTA TABELA

Tabela 1.1 Aktivnosti koje koriste prognozu potrošnje električne energije	3
Tabela 1.2 Dostupnost i pouzdanost temperaturnih, ekonomskih i urbanističkih podataka za prognozu potrošnje.....	4
Tabela 1.3 Podela prognoze potrošnje električne energije	5
Tabela 1.4 Veza poslovnih aktivnosti sa različitim tipovima prognoze	6
Tabela 7.1 Prosečna godišnja MAPE, sMAPE i RMSE greška	105
Tabela 7.2 Prosečna godišnja MAPE, sMAPE i RMSE greška za modele sa prognozom u dve faze	107
Tabela 7.3 Prosečna mesečna MAPE greška.....	126
Tabela 7.4 Prosečna MAPE greška po danu u nedelji	127
Tabela 7.5 Srednja APE greška po satu	133

LISTA PROCEDURA

Procedura 4.1 CG algoritam	67
Procedura 4.2 Shrinking grid-search i k-unakrsnih validacija.....	70
Procedura 4.3 Algoritam za postizanje sparse LS-SVM-a	73
Procedura 4.4 RVM regresioni proces.....	76
Procedura 6.1 Prognoza potrošnje električne energije u dve faze	91
Procedura 6.2 Prognoza potrošnje električne energije sa izborom atributa.....	96
Procedura 6.3 Prognoza potrošnje električne energije sa dopunom trening skupa.....	99

SKRAČENICE

SCADA - *Supervisory control and data acquisition*

DSM - *Distribution side management*

VSTLF - *Very short-term load forecasting*

STLF - *Short-term load forecasting*

MTLF - *Mid-term load forecasting*

LTLF - *Long-term load forecasting*

SVM - *Support vector machines*

LS-SVM - *Least squares support vector machines*

kNN - *k nearest neighbours*

ARIMA - *Autoregressive integrated moving average*

AR - *Autoregressive*

ACF - *Autocorrelation function*

PACF - *Partial autocorrelation function*

TFM - *Transfer function model*

MAPE - *Mean absolute percentage error*

RMSE - *Root mean square error*

MRPE - *Mean relative percentage error*

SEP - *Standard error of prediction*

MAE - *Mean absolute error*

APE - *Absolute percentage error*

PSO - *Particle swarm optimization*

ACO - *Ant colony optimization*

BP - *Back-propagation algorithm*

PCA - *Principal components analysis*

ANN - *Artificial neural networks*

MI - *Mutual information*

EMR - *Empirical risk minimization*

SRM - *Structural risk minimization*

VC - *Vapnik - Chervonenkis dimension*

QP - *Quadratic programming*

KKT - *Karush - Kuhn - Tucker conditions*

RBF - *Radial basis function*
MLP - *Multilayer perceptron*
SMO - *Sequential minimal optimization*
MSE - *Mean square error*
LOOCV - *Leave one out cross - validation*
PSO - *Particle swarm optimization*
PDF - *Probability density function*
KL - *Kullback-Leibler divergence*
IID - *Independent and identically distributed*
DS-ARIMA - *Double sesonal ARIMA*
DS-HW - *Double sesonal Holt-Winters*
MIMO - *Multiple inputs-multiple outputs*
DIRMO - *Direct-multiple outputs*
RVM - *Relevance vector machines*

1 UVOD

Ekonomski razvoj jedne zemlje direktno zavisi od dostupnosti električne energije, posebno zato što većina industrije zavisi u potpunosti od njenog korišćenja. Dostupnost stalne, jeftine i održive energije je od ključne važnosti za siguran ekonomski razvoj zemlje. Prognoza potrošnje električne energije je u osnovi nauka ili umetnost predviđanja buduće potrošnje nekog sistema za određeni vremenski period unapred. Prognoza potrošnje električne energije ima ključnu ulogu u obezbeđivanju pouzdanog snabdevanja sistema električnom energijom i smanjenju operativnih troškova.

Ovo poglavlje daje kratak uvod u problem prognoze potrošnje električne energije, ukazuje na značaj i potrebu za njegovom rešavanjem, predstavlja izazove i ograničenja, i prezentuje cilj i sadržaj istraživanja. Na kraju ovog poglavlja dat je pregled disertacije po poglavljima sa organizacionom šemom, radi lakšeg praćenja i razumevanja sadržaja.

1.1 Proces prognoze u elektroenergetskim preduzećima

U većini preduzeća za proizvodnju i distribuciju električne energije u svetu ali i kod nas, proces prognoze potrošnje električne energije je usko povezan sa planiranjem, pa je samim tim zadatak prognoze dodeljen sektoru za planiranje. Ipak, treba naglasiti razliku između ova dva procesa. Planiranje stvara strategiju i odnosi se na to kako i šta jedno preduzeće treba da uradi. Prognoza se odnosi na to šta bi se desilo ukoliko preduzeće proba da sprovede datu strategiju.

Tradicionalni organizacioni problemi koji postoje u preduzećima dovode do toga da je svaki sektor zadužen za prognozu koja mu je potrebna. Tako često nema komunikacije

između sektora po pitanju procesa prognoze i razmene informacija, pa svaki sektor obično radi prognozu potrošnje samo za sopstvene potrebe. Ovo je prvi od dva osnovna problema koji su prisutni i predstavlja primer neefikasnog korišćenja resursa unutar preduzeća. Drugi nedostatak je da kvalitet prognoze može mnogo da varira. Razlog ovoga je to da će sektor prodaje za svoje potrebe da radi prognozu koristeći podatke o izvršenim naplatama, dok će sektor planiranja koristiti podatke sa uređaja za prikupljanje podataka SCADA sistemom (*eng. Supervisory control and data acquisition*). Kako su podaci korišćeni u procesu prognoze u ovim sektorima različiti i kvalitet ostvarenih prognoza potrošnje će se razlikovati što može dovesti do operativnih problema.

1.2 Potrebe za prognozom potrošnje električne energije

Danas je prognoza potrošnje električne energije važan proces u elektroenergetskim preduzećima sa primenom u nekoliko sektora preduzeća kao što su sektor za planiranje, operativu, sektor prodaje itd. Potrebe za kvalitetnom prognozom potrošnje električne energije vezane su za sledeće aktivnosti preduzeća [1]:

- 1) *Kupovina energije*. Prognoza potrošnje električne energije je osnova za proces kupovine energije, bez obzira da li taj proces obavlja samo preduzeće ili taj proces prepušta nekoj spoljnoj, nezavisnoj organizaciji. Preduzeće može da obavlja bilateralnu kupovinu sa obavezom na duži period od npr. 10 godina, može da rezerviše i kupuje na intervalima od jednog meseca do 3 godine unapred ili da prilagođava i kupovinu obavlja na tržištu za dan unapred.
- 2) *Planiranje prenosa i distribucije*. Preduzeća imaju potrebu za održavanjem i blagovremenim unapređenjem sistema kako bi se zadovoljile potrebe rasta potrošnje, obezbedilo stabilno snabdevanje i povećala pouzdanost sistema. Preduzeća moraju da planiraju i kupovinu zemljišta u budućnosti kako bi se stvorila mogućnost za izgradnju novih elektroenergetskih objekata. Planiranje prenosa i distribucije se u velikoj meri oslanja na prostornu prognozu potrošnje koja treba da pruži odgovore na pitanja koliko i gde će porasti potražnja za električnom energijom.
- 3) *Održavanje i operativa*. Na dnevnom nivou, dijagrami opterećenja dobijeni prognozom potrošnje, omogućuju dispečeru da donosi pravovremene odluke o manipulaciji i planira održavanje.

-
- 4) *Upravljanje opterećenjem (eng. Demand side management - DSM)*. Iako većina aktivnosti DSM-a spada u dnevne aktivnosti, poželjno je odvojiti DSM kao poseban proces zbog njegovog značaja kod smart-grid koncepta. Prognoza potrošnje električne energije pomaže donošenju pravih odluka kad su u pitanju regulacija opterećenja i smanjivanje napona. Takođe, pomaže da preduzeća sprovedu dugoročno planiranje bazirano na dobijenim karakteristikama opterećenja krajnjih korisnika tokom procesa prognoze.

U Tabeli 1.1 prikazane su aktivnosti preduzeća koje se baziraju na prognozi potrošnje električne energije sa minimalnim intervalom i maksimalnim horizontom prognoze.

Tabela 1.1 *Aktivnosti koje koriste prognozu potrošnje električne energije*

	Minimalni interval	Maksimalni horizont
Kupovina energije	1 sat	10 i više godina
Planiranje prenosa i distribucije	1 dan	30 godina
Održavanje i manipulacija	15 minuta	2 nedelje
Upravljanje opterećenjem	15 minuta	10 i više godina

1.3 Podela prognoze potrošnje električne energije

Podela prognoze potrošnje električne energije na različite tipove prognoze, ne zavisi samo od potreba elektroenergetskih preduzeća već i od dostupnih promenljivih koje utiču na potrošnju električne energije: meteorološke (ili klimatske na dužem vremenskom periodu), sociološke, ekonomske i urbanističke.

Meteorološke promenljive se odnose na meteorološke elemente kao što su temperatura, vlažnost vazduha, brzina vetra, količina padavina, itd. posmatrane u određenom području u periodu do dve nedelje. Klimatske promenljive podrazumevaju iste meteorološke elemente u određenom području ali posmatrane u dužem vremenskom periodu. Od meteoroloških promenljivih temperatura ima najveći uticaj na potrošnju električne energije i u uskoj je korelaciji sa njom, tako da se često koristi u modelima prognoze potrošnje električne energije. Danas, za potrebe prognoze potrošnje električne energije, prognoza temperature je relativno tačna do dan unapred i može biti neprecizna ali pouzdana do dve nedelje unapred. Neke elektroenergetske kompanije investiraju u sisteme za praćenje meteoroloških

parametara i razvijaju sopstvene modele za njihovu prognozu dok druge koriste usluge specijalizovanih meteoroloških servisa.

Uticaj socioloških faktora tj. ljudskih aktivnosti na potrošnju električne energije se može posmatrati iz nekoliko uglova. Uticaj ljudskih aktivnosti na potrošnju za period od nekoliko sati unapred se ogleda kroz kalendarske promenljive kao što su sat u danu, dan u nedelji i mesec u godini. Uticaj na potrošnju za period od nekoliko meseci ili kvartala se ogleda i kroz različite ekonomske faktore. Podaci o stanju i razvoju ekonomije postaju sve precizniji i pouzdaniji ali trenutno su relativno tačni do jedne godine unapred i neprecizni ali pouzdani do tri godine unapred. Za period od nekoliko godina unapred, i klimatski i ekonomski faktori utiču na potrošnju električne energije, međutim u tom slučaju su nedostupni pa se prognoza potrošnje električne energije može dobiti jedino simulacijom različitih scenarija. Sa druge strane, prognoza potrošnje električne energije za duži vremenski interval a za potrebe širenja mreže, zavisi od povećanja broja stanovnika i urbanističkog razvoja. Ovaj podatak je dostupan kroz urbanistički plan i uglavnom je tačan do jedne godine i neprecizan ali pouzdan do pet godina unapred. Neke zemlje prave urbanistički plan i za do 30 godina unapred, međutim u tom slučaju ostaje nepoznanica šta će se dešavati iz godine u godinu u tom periodu. Prognoza potrošnje električne energije za potrebe širenja mreže korišćenjem simulacije urbanističkog razvoja naselja se naziva prostorna prognoza potrošnje električne energije (*eng. spatial electric load forecasting*) [2].

Tabela 1.2 sumira dostupnost temperaturnih, ekonomskih i urbanističkih podataka za potrebe korišćenja kod prognoze potrošnje električne energije.

Tabela 1.2 Dostupnost i pouzdanost temperaturnih, ekonomskih i urbanističkih podataka za prognozu potrošnje

Tip podataka	Precizno	Neprecizno	Nepouzđano
Temperaturni	1 dan	2 nedelje	> 2 nedelje
Ekonomski	1 mesec	3 godine	> 3 godine
Urbanistički	1 godina	5 godina	> 5 godina

Tabela 1.3 prikazuje jednu podelu prognoze potrošnje električne energije na osnovu dostupnosti ulaznih podataka, intervala i horizonta prognoze: izuzetno kratkoročna (*eng. very short term load forecasting - VSTLF*), kratkoročna (*eng. short term load forecasting - STLF*),

srednjoročna (*eng. medium term load forecasting - MTLF*) i dugoročna prognoza (*eng. long term load forecasting - LTLF*).

Tabela 1.3 Podela prognoze potrošnje električne energije

	Temperaturni	Ekonomski	Urbanistički	Min. Interval	Maks. Horizont
VSTLF	Opciono	Opciono	Opciono	<= 1 sat	1 dan
STLF	Obavezan	Opciono	Opciono	1 dan	2 nedelje
MTLF	Simuliran	Obavezan	Opciono	1 mesec	3 godine
LTLF	Simuliran	Simuliran	Obavezan	1 godina	30 godina

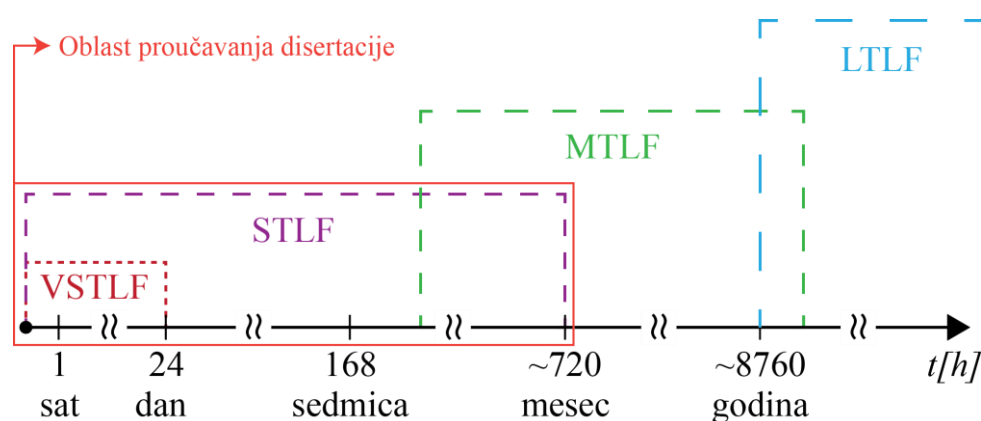
U slučaju VSTLF, nema potrebe za korišćenjem podatka o temperaturi, ekonomskom i urbanističkom razvoju jer je vremenski horizont mali i dovoljni su podaci o prošloj potrošnji, pa se ovakav problem prognoze najčešće svodi na problem prognoze vremenske serije. Kako su ekonomske promene i urbanistički razvoj relativno stabilne u kratkom vremenskom periodu (manje od dve nedelje), ove promenljive su opcione u slučaju STLF. Sa druge strane, temperaturna promenljiva igra veoma važnu ulogu kod STLF. Kako se temperatura ne može precizno prognozirati za period od tri godine, u slučaju MTLF može se jedino koristiti simulirana promena temperature bazirana na promeni temperature u prošlosti. Sa druge strane se ekonomski pokazatelji mogu precizno odrediti za taj period, a kako je i njihov uticaj na potrošnju značajan, kod problema MTLF ekonomska promenljiva je obavezna. Urbanistički podaci koji su vezani za širenje i razvoj naselja ali i industrijskih zona su opciona promenljiva kod VSTLF, STLF i MTLF jer se oni vrlo malo menjaju u periodu od tri godine. Međutim, kad je reč o LTLF, ovaj uticaj je značajan pa su urbanističke promenljive obavezne, dok je temperaturu i ekonomski razvoj teško prognozirati za dug vremenski period pa se mogu koristiti samo njihove simulacije.

Prema podeli potrošnje električne energije datoj u Tabeli 1.3, aktivnosti preduzeća koje koriste prognozu potrošnje električne energije mogu se povezati sa odgovarajućim tipom prognoze, što je prikazano u Tabeli 1.4. Jedna aktivnost preduzeća zbog intervala i horizonta na koji se odnosi, može da budu vezana za više tipova prognoze. Tako se npr. kupovina električne energije može obavljati na 15-to minutnom intervalu, za dan unapred, godinu dana ili 10 godina unapred, pa se može koristiti od VSTLF pa do LTLF.

Tabela 1.4 Veza poslovnih aktivnosti sa različitim tipovima prognoze

	VSTLF	STLF	MTLF	LTLF
Kupovina energije	X	X	X	X
Planiranje prenosa i distribucije		X	X	X
Održavanje i manipulacija	X	X		
Upravljanje opterećenjem	X	X	X	X

Treba napomenuti da različiti autori koriste različite vremenske horizonte kod podele prognoze potrošnje električne energije pa se tako u literaturi mogu naći i malo drugačije podele u smislu vremenskog horizonta koji obuhvataju. Na Slici 1.1 je prikazana podela prognoze potrošnje električne energije koja je prisutna u literaturi, a prema vremenskom horizontu na koji se odnosi.



Slika 1.1 Podela prognoze potrošnje prema horizontu

1.4 Sadržaj istraživanja

Istraživanja disertacije su primarno fokusirana na prognozu potrošnje električne energije primenom metoda veštačke inteligencije (mašinskog učenja) i to se pre svega odnosi na izuzetno kratkoročnu (VSTLF) i kratkoročnu (STLF) prognozu potrošnje, Slika 1.1. Zbog osobina da uvek pronalaze optimalno rešenje, odlične generalizacije i malog broja potrebnih trening parametara, za modelovanje potrošnje je korišćena metoda potpornih vektora (*eng. support vectors machines* - SVM) tj. jedna njena modifikacija, LS-SVM (*eng. least squares support vector machines*).

Izvršena je analiza električne potrošnje kako bi se prepoznale i izdvojile one promenljive koje najviše utiču na potrošnju. U skladu sa tim, formirani su odgovarajući inicijalni modeli.

Kako bi se poboljšali ovi inicijalni modeli, moguće je ići u dva pravca. Prvi je da se iz inicijalnog skupa promenljivih (atributa) koristeći neki od postojećih algoritama, uradi izbor i odstrane one promenljive za koje dati kriterijum ukazuje da su nepotrebne. Međutim, kako je problem potrošnje električne energije dobro poznat i proučavan dugo vremena, u većini slučajeva je dovoljna ekspertska analiza za identifikaciju najkorisnijih promenljivih, pa je pitanje koliko izbor atributa može doprineti poboljšanju preciznosti. Ipak ne treba inicijlno odustati od pristupa sa izborom atributa jer mesta za poboljšanja ima a ujedno se tako može i proveriti da li su ekspertskom analizom izabrani dobri atributi. Tako je u disertaciji predložen model sa izborom atributa koji se bazira na primeni procene zajedničkih informacija za kriterijum izbora.

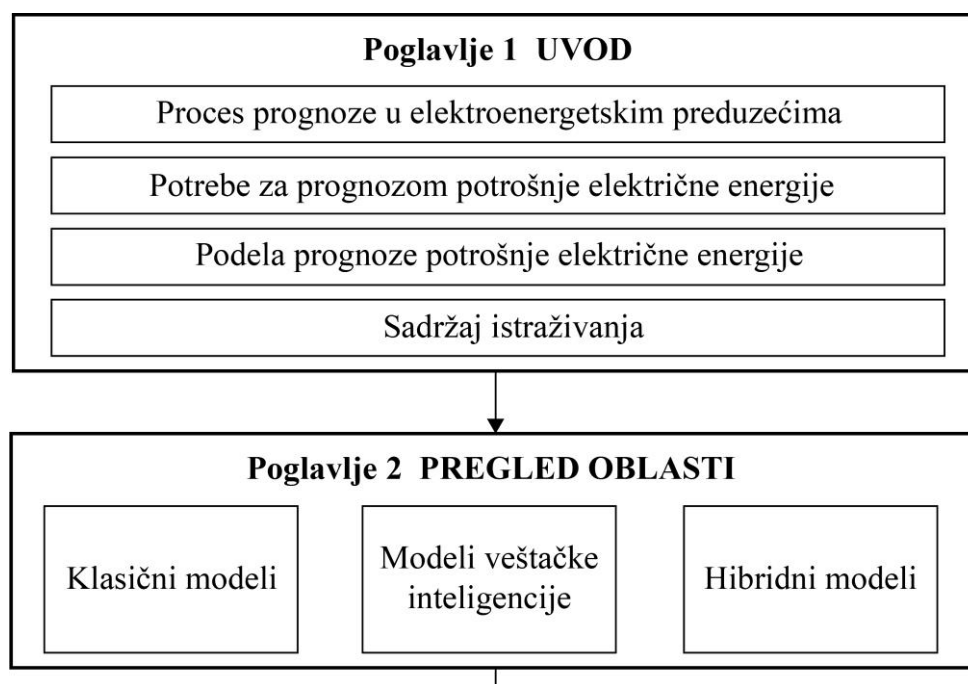
Pružanje dodatne informacije modelu o potrošnji, trebalo bi doprineti poboljšanju prognoze. Ova pretpostavka se uglavnom bazira na praktičnom iskustvu, model kome je dostupna veća količina informacija je precizniji od modela sa nedovoljno informacija. Međutim, ovo jeste tačno pod uslovom da su to korisne informacije za model. Sa tim na umu, inicijalni modeli su dopunjeni "izvedenim" atributom, dnevnom potrošnjom za dan unapred tj. tekući dan za koji se radi prognoza. Dnevna potrošnja može da se odnosi na vršnu, srednju ili minimalnu dnevnu potrošnju. Kako ovaj dodatni atribut nije poznat za dan unapred, prvo je formiran model za prognozu dnevne potrošnje koja se zatim koristi kod modela za prognozu potrošnje po satima za naredni dan. Zatim je urađena komparacija novog modela sa inicijalnim modelom u cilju procene prednosti ovakvog pristupa.

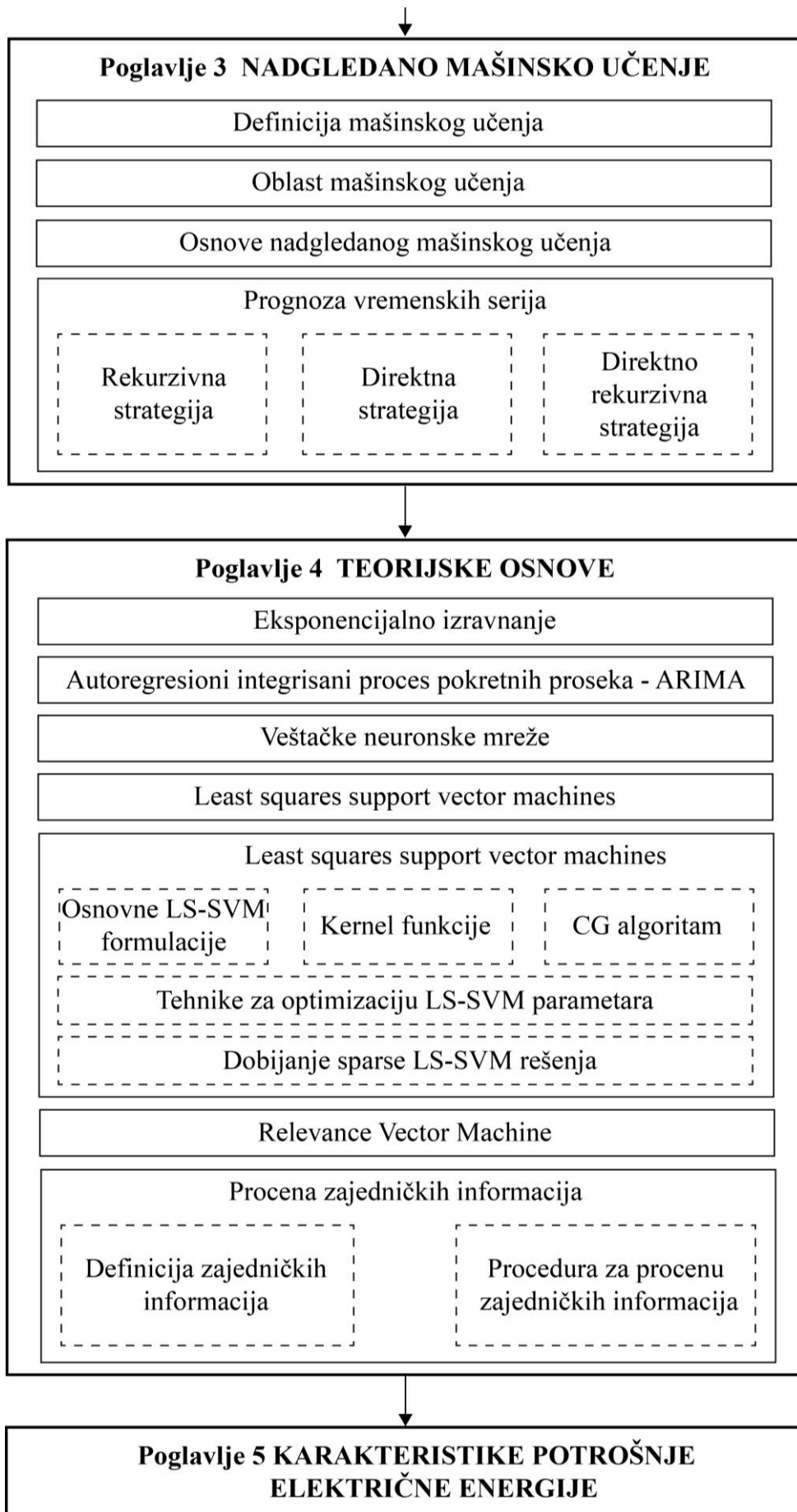
Kako bi se modelu omogućio konstantan priliv svežih informacija predložen je model sa dopunom trening skupa. Kod ovog modela, inicijalni trening skup se dopunjava na određenom intervalu vremena (na mesec dana), tako što se stari trening vektori odbacuju a uključuju novi. Na ovaj način model se konstantno „osvežava“ a vremenski razmak između poslednjeg retrainiranja modela i test vektora je minimalan.

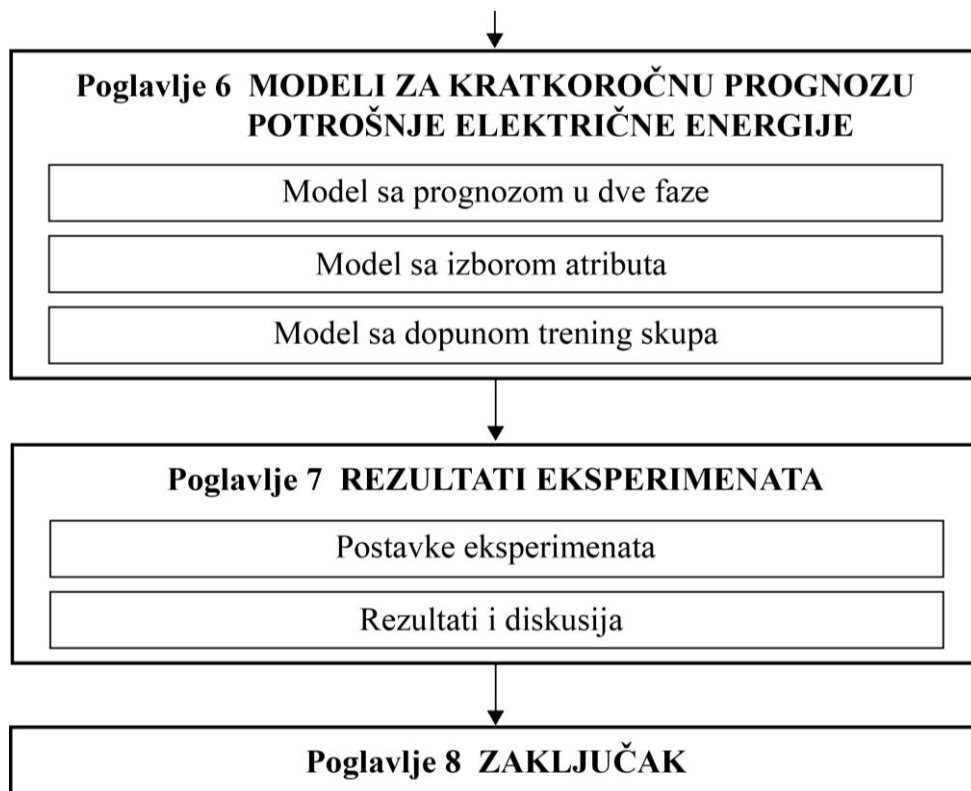
Evaluacija predloženih modela obavljena je na realnim podacima o potrošnji električne energije. Predloženi modeli su poređeni sa inicijalnim LS-SVM modelom ali i sa duplo-periodičnim ARIMA modelom, duplo-periodičnim modelom eksponencijalnog poravnanja, modelom veštačkih neuronskih mreža i RVM modelom (*eng. relevance vector machines*).

1.5 Organizacija disertacije

Kratak pregled disertacije po poglavljima sa organizacionom šemom ima za cilj da sumira sadržaj svakog poglavlja kako bi se čitaocu olakšalo praćenje disertaciju. U Poglavlju 2 je dat prikaz dosadašnjeg rada na temu prognoze potrošnje električne energije, sumirano je šta je do sada urađeno a na kraju su predstavljeni budući pravci u oblasti. U Poglavlju 3 je objašnjen osnovni princip rada regresionih tehnika mašinskog učenja (veštačke inteligencije) i definisana osnovna terminologija neophodna za dalje praćenje sadržaja disertacije. U Poglavlju 4 su date osnovne teorijske postavke svih metoda korišćenjih za formiranje modela za prognozu električne potrošnje ali i definicija zajedničkih informacija koje su korišćenje za meru sličnosti kod izbora atributa. Analiza električne potrošnje, identifikacija promenljivih koje karakterišu električnu potrošnju i ekspertski izbor atributa (promenljivih) za modelovanje električne potrošnje dat je u Poglavlju 5. Osnovne ideje i implementacija predloženih modela za prognozu električne potrošnje prezentovane su u Poglavlju 6. Postavke eksperimenata, rezultati i diskusija dati su u Poglavlju 7, dok je zaključak disertacije dat u Poglavlju 8. Na Slici 1.2 prikazana je detaljna organizaciona šema disertacije.







Slika 1.2 Organizaciona šema disertacije

2 PREGLED OBLASTI

Veliki broj radova na temu prognoze potrošnje električne energije je publikovan u poslednjih 30 godina. U ovom poglavlju je dat pregled literature koja se odnosi na STLF, sa ciljem da se prikažu dosadašnja dostignuća ali i budući izazovi u oblasti. Kako se metode i tehnike korišćene kod problema prognoze potrošnje električne energije mogu podeliti u tri kategorije:

- Klasične (konvencionalne) tehnike,
- Tehnike veštačke inteligencije i
- Hibridne tehnike,

tako je i pregled literature (modeli) grupisan u ove tri kategorije. Pregled literature će pre svega biti fokusiran na doprinos u oblasti, a ne na analizu svakog aspekta predloženog modela, sa osvrtom na specifičnosti i karakteristike koje nam mogu pomoći kod modelovanja kratkoročne prognoze električne potrošnje.

2.1 Klasični modeli

Kada je reč o klasičnim tj. konvencionalnim tehnikama, tu se pre svega misli na metode regresione analize (*eng. regression analysis*), eksponencijalnog poravnanja (*eng. exponential smoothing*), Box-Jenkins tehnike i Kalmanove filtre (*eng. Kalman filters*).

Jedan pristup kratkoročne prognoze potrošnje električne energije baziran na linearnoj regresiji predložen je u [3]. Model prvo prognozira inicijalnu vršnu dnevnu potrošnju koju zatim koristi za prognozu inicijalne časovne potrošnje. Po dobijanju inicijalne časovne

potrošnje vrši se korekcija inicijalne vršne dnevne potrošnje koristeći model linearne regresije sa promenljivama: inicijalnom prognozom vršne potrošnje, maksimumom prognoze inicijalne časovne potrošnje, poslednjom greškom prognoze vršne potrošnje i eksponencijalnom aproksimacijom grešaka prognoze. Na kraju se svaka prognozirana vrednost potrošnje po satu, koja je veća od korigovane vršne dnevne potrošnje koriguje na tu vrednost čime se dobija krajnja korigovana časovna prognoza potrošnje. Predloženi pristup zahteva formiranje četiri modela, po dva u inicijalnim fazama prognoze i dva u korekcionima fazama. Model za prognozu inicijalne vršne potrošnje generiše vrednost vršne potrošnje za dan unapred koristeći vršnu dnevnu potrošnju tekućeg dana i šest prethodnih dana, temperaturu za ove dane, indikator dana u nedelji, prognozu temperature za naredni dan, vreme u godini i podatke o praznicima. U sledećoj fazi, model za prognozu inicijalne časovne potrošnje koristi inicijalnu prognozu vršne potrošnje za dan unapred, potrošnju po časovima za tekući dan, vrednost greške inicijalne vršne potrošnje i skup svih promenljivih kod prognoze vršne potrošnje. Po dobijanju inicijalnih vrednosti potrošnje po satima za dan unapred, maksimalna vrednost potrošnje po satima se koristi kao promenljiva u modelu za korekciju prognoze vršne potrošnje, zajedno sa inicijalnom prognozom vršne potrošnje i greškama ove prognoze. Na kraju se generišu korigovane vrednosti prognoze potrošnje po časovima, koje su iste kao i vrednosti inicijalne prognoze po časovima osim onih vrednosti prognoze koje su veće od korigovanih vrednosti inicijalne vršne potrošnje pa se u tom slučaju koriguju na vrednost inicijalne vršne potrošnje. Prednosti predloženog pristupa se ogledaju kroz modelovanje nekoliko ključnih faktora koji utiču na potrošnju. Modelovanje praznika i posebnih događaja koristeći binarne promenljive koje su definisane za glavne praznike kao što su Nova Godina ali i za sporedne događaje kao što su sportska dešavanja. Binarnim promenljivama su pridruženi odgovarajući koeficijenti kako bi se izvršila korekcija prognoze potrošnje koja je generalno manja u danima praznika ali utiče i na potrošnju u danima koji slede. Uticaj temperature na potrošnju modelovan je koristeći krive hlađenja i grejanja koje se formiraju na osnovu analize uticaja temperature na potrošnju. Tako je kriva hlađenja jednaka nuli sve dok je ispod donjeg temperaturnog praga, zatim kvadratno raste do gornjeg temperaturnog praga, posle koga raste linearno. Kriva grejanja ima istu formu samo što naravno raste kako pada temperatura. Pored ovih prednosti predloženog pristupa, autori ukazuju i na lako određivanje parametara modela koristeći metodu linearne regresije sa najmanjim kvadratima, otpornost modela na velike greške prognozirane temperature i prevenciju od podkoračenja prognoze vršnog opterećenja. Međutim treba napomenuti da iako prezentovani rezultati obećavaju, predloženi pristup je poređen samo sa prethodnim

pristupom autora na jednom test skupu i da je kod testiranja modela korišćena tačna vrednost temperature koja nije dostupna za dan unapred. Iako autori predlažu postupak određivanja buduće temperature dodavanjem šuma trenutnoj temperaturi i daju način za određivanje ovog šuma, ispravnost ovakvog pristupa nije potvrđena. Takođe, formiranje modela zahteva analizu prethodnih podataka koji su jedinstveni za svaku situaciju i određivanje velikog broja parametara modela.

Velike greške koje se javljaju kod prognoze električne potrošnje primenom metode linearne regresije su posledica nelinearne zavisnosti klimatskih parametara i električne potrošnje, što je naročito izraženo u prelaznom periodu kao što je proleće i jesen. Kako bi prevazišli ovaj problem, autori u [4] predlažu jedan pristup koristeći metod linearne regresije sa transformacionom tehnikom. Da bi dobili što precizniju prognozu vršne električne potrošnje, autori imaju za cilj da u model uključe sezonske promene električne potrošnje, godišnji rast potrošnje ali i promene potrošnje u nekoliko poslednjih dana. Na osnovu podataka o potrošnji iz prethodne godine, formira se transformaciona funkcija koristeći metodu najmanjih kvadrata. Model je u tom slučaju definisan sa:

$$P = \alpha_0 + \sum_{j=1}^m \alpha_j f_j(X_j), \quad (2.1)$$

gde je P vršno dnevno opterećenje, α_0 i α_j su parametri modela, X_j je klimatski parametar kao što je vršna temperatura a f_j je transformaciona funkcija. Koristeći ovu funkciju, promenljive modela prognoze se preslikaju a zatim se te preslikane promenljive koriste za nalaženje parametara modela prognoze. Na ovaj način model obuhvata skorije promene potrošnje ali i godišnje promene usled uticaja klimatskih faktora. Međutim ovakav pristup važi samo pod pretpostavkom da je zavisnosti između potrošnje i klimatskih faktora u prethodnoj i tekućoj godini slična, što naravno nije slučaj, pa transformaciona funkcija nije linearna. Zato autori uvode korekcionni parametar, koeficijent pomeraja δ_j pa model sada ima sledeći oblik:

$$P = \alpha_0 + \sum_{j=1}^m \alpha_j f_j(X_j - \delta_j). \quad (2.2)$$

Na ovaj način se transformaciona funkcija koriguje da bude linearna. Autori uvode još jednu korekcionu tehniku, reflektivni metod, kojim se koriguje nagib krive posle transformacije sa translacijom. U procesu testiranja predloženog pristupa autori formiraju tri modela, osnovni model baziran na linearnoj regresiji, model baziran na linearnoj regresiji sa transformacionom funkcijom i model baziran na linearnoj regresiji sa transformacionom funkcijom i sa

translatornom i refleksivnom korekcijom. Testiranje obuhvata interval od četiri godine sa akcentom na prolećni i jesenji period a za evaluaciju rezultata je korišćena srednja procentualna greška i standardna devijacija. Prikazani rezultati ukazuju na prednost uvođenja transformacione funkcije, jer model sa transformacionom funkcijom prevazilazi osnovni model, smanjujući i srednju apsolutnu grešku i standardnu devijaciju u svakoj testiranoj situaciji. Dodatno poboljšanje se postiže primenom modela koji koristi i transformacionu funkciju i korekzione metode translacije i refleksije, ali je to poboljšanje znatno manje nego u prethodnom slučaju.

Veliki uticaj temperature ali i drugih klimatskih faktora na električnu potrošnju inspirisalo je autore u [5] da predlože klimatski prilagodljiv model za prognozu električne potrošnje baziran na metodi linearne regresije. Predloženi pristup se sastoji iz dva koraka. U prvom koraku se radi ukupna prognoza potrošnje električne energije a zatim se u drugom koraku radi prognoza potrošnje po časovima, koristeći dobijenu vrednost iz prvog koraka. Za formiranje (treniranje) oba modela korišćeni su podaci koji su prethodno izabrani tako da odgovaraju sličnim klimatskim uslovima ali i potrošnji za prognozirani dan. Za izbor tj. indentifikaciju ovih dana korišćeno je Euklidsko rastojanje kao mera sličnosti. Model ukupne potrošnje električne energije (potrošnja za period od 24 časa) je dat sa:

$$W_j = \left[k_d + (2f_j - 1)(k_s \Delta T_{ejr} + k_i \Delta T_{einjr}) \right] W_{jr}, \quad (2.3)$$

gde je W_j prognozirana ukupna potrošnja električne energije, W_{jr} - potrošnja električne energije u periodu od 24 časa, k_d - koeficijent tipa dana, k_s - koeficijent temperaturene osetljivosti, k_i - koeficijent temperaturene inercije, ΔT_{ejr} - razlika srednjih temperatura u poslednjih 24 časa i prognoziranog dana j , ΔT_{einjr} - razlika između inercijalnih temperatura pre perioda od 24 časa i prognoziranog dana j i f_j - koeficijent grejanja/hlađenja (ima vrednost 0 ili 1 i definiše da li se radi o krivi grejanja ili hlađenja). Za formiranje trening skupa autori koriste Euklidsko rastojanje kao meru sličnosti između prognoziranog dana i nekog prethodnog dana koristeći sledeće promenljive: ukupna dnevna potrošnja električne energije, dnevna i inercijalna temperatura i trajanje dnevne svetlosti. Na ovaj način autori biraju sedam dana iz baze, čiji se podaci koriste za određivanje koeficijenata modela k_d , k_s i k_i . Prognozirana ukupna potrošnja električne energije se sada koristi u model za prognozu električnog časovnog opterećenja koje je modelovano sa:

$$P_{jh} = \left[p_h + (2f_j - 1)(\Delta p_t A_{jh} + \Delta p_{in} B_j) \right] W_j, \quad (2.4)$$

gde je P_{jh} časovno opterećenje u j -tom danu i h -tom času, W_j - ukupna dnevna potrošnja električne energije dobijena prognozom u prethodnom koraku, p_h - relativna vrednost osnovne komponente opterećenja (klimatski nezavisna komponenta časovnog opterećenja), $\Delta p_t A_{jh}$ - relativna vrednost temperaturno zavisne komponente časovnog opterećenja (predstavlja uticaj grejanja/hlađenja na časovno opterećenje), $\Delta p_{in} B_j$ - relativna vrednost inercijalne komponente časovnog opterećenja u danu j i f_j - binarni koeficijent grejanja/hlađenja. Slično kao i u prethodnom slučaju i ovde se koristi Euklidsko rastojanje kao mera sličnosti za izbor trening skupa sa tom razlikom što autori prvo grupišu dane u četiri grupe, radne dane: utorak, sredu, četvrtak i petak u jednu grupu a subotu, nedelju i ponedeljak u preostale tri. Na taj način, se bira iz baze osam dana koji čine trening, prvo prema podeli na osnovu tipa dana a zatim prema meri sličnosti koristeći Euklidsko rastojanje sa sledećim promenljivama: ukupna dnevna potrošnja električne energije, minimalna, maksimalna i srednja dnevna temperatura i trajanje dnevne svetlosti. Rezultati predstavljeni u radu se odnose na period od 11 godina. Kao meru preciznosti predloženog pristupa, autori koriste srednju apsolutnu procentualnu grešku dnevne potrošnje energije (*eng. mean absolute daily energy error* - MAEE) i srednju apsolutnu procentualnu grešku časovnog opterećenja (*eng. mean absolute hourly loads error* - MALE). Srednje vrednosti MAEE i MALE grešaka na celom test periodu su 1.64% i 2.73% respektivno. Takođe autori prikazuju i kretanje grešaka po mesecima ali i po danima. Pa su tako greške najmanje tokom letnjih meseci (jun - septembar) MAEE 1.21% - 1.47%, tokom zimskih meseci su približne (novembar - mart) MAEE 1.52% - 1.72%, dok su najveće tokom prolećnih i jesenjih meseci (april, maj i oktobar) MAEE 1.82% - 2.58%. Kad su u pitanju dani, najmanje greške su u danima od utorka do petka (MAEE 1.48% - 1.57%) dok su u danima subote, nedelje i ponedeljka nešto veće (MAEE 1.77% - 1.83%). Testiranjem su obuhvaćeni i dani praznika kojih je u test periodu bilo 287 i dobijene su znatno veće greške (MAEE 4.18% u proseku). Međutim treba napomenuti da su greške u oba slučaja računane na godišnjem periodu pa posle usrednjene na celom test intervalu. Bilo bi prikladnije da je za meru preciznosti u prvom koraku kod prognoze dnevne potrošnje električne energije korišćena relativna greška a u drugom koraku srednja apsolutna procentualna greška na dnevnom nivou (MAPE) koje bi posle bile usrednjene na godišnjem ili mesečnom nivou. Takođe sam pristup zavisi i od heurističkog

izbora nekoliko koeficijenta kod određivanja pojedinih parametara modela što može dovesti do formiranja lošeg modela usled pogrešnog izbora.

Kako potrošnja električne energije zavisi od temperature, tipa dana, prošle potrošnje i drugih promenljivih, ona se može predstaviti funkcijom raspodele verovatnoće sa više promenljivih $PDF = f(P, \bar{X})$, gde je P srednja časovna potrošnja a $\bar{X} = (x_1, \dots, x_r)^T$ je vektor promenljivih koje opisuju potrošnju. Vodeći se ovom idejom, autori u [6] predlažu model prognoze potrošnje električne energije baziran na neparametarskoj regresiji, preciznije korišćenjem kernel estimatora za određivanje funkcije raspodele verovatnoće. Koristeći Gausovu raspodelu za kernel funkciju i zamenom u jednačinu neparametarske regresije sa više promenljivih, autori daju sledeću jednačinu:

$$\hat{P}(x) = \frac{\sum_{i=1}^n \left\{ P_i \exp \left(- \sum_{j=1}^r \frac{(x_j - x_{ji})^2}{2h_j^2} \right) \right\}}{\sum_{i=1}^n \left\{ \exp \left(- \sum_{j=1}^r \frac{(x_j - x_{ji})^2}{2h_j^2} \right) \right\}}, \quad (2.5)$$

gde su: h_j - koeficijenti poravnanja, j je j -ota promenljiva vektora \bar{X} , i je i -ti vektor iz test skupa i r je ukupan broj promenljivih. Može se primetiti da ova jednačina daje prognozu potrošnje bez određivanja funkcije raspodele verovatnoće. Kako prethodna jednačina predstavlja samo stacionarni model potrošnje i ne opisuje dinamiku potrošnje usled zavisnosti od temperature, autori uvode novu promenljivu u model kako bi opisali zavisnost potrošnje usled promene klimatskih parametara kao i inertnost samog procesa grejanja/hlađenja u zgradama. Nova promenljiva je definisana sa:

$$T_h(t) = \frac{1}{m} \sum_{i=1}^m T(t-i), \quad (2.6)$$

gde je m dužina prošlog perioda koji se posmatra sa predloženom optimalnom vrednošću $m=8$. Kod evaluacije modela, za nalaženje koeficijenta poravnanja tj. formiranje modela, korišćeni su podaci o potrošnji za tri nedelje u letnjem periodu, razdvojeni u dve grupe po tipu dana (radni dani ili vikend). Po nalaženju tri koeficijenta poravnanja za promenljive: vreme dana, temperaturu i srednju temperatura za poslednjih 8 časa, urađena je prognoza potrošnje sa horizontom prognoze od nedelju dana (168 časova). Autori prikazuju apsolutnu grešku prognoze, daju poređenje predloženog modela sa jednim modelom veštačkih neuronskih mreža ali i predloženog stacionarnog modela (modela bez T_h promenljive). Model

veštačke neuronske mreže ima jedan skriveni sloj sa tri neurona i sledećim promenljivama u ulaznom sloju: čas u danu, temperatura u tom času, prošla potrošnja, srednja temperatura za prethodnih 8 časova i potrošnja i temperatura za poslednja 24 časa. Kao meru preciznosti, autori koriste srednju apsolutnu grešku i srednje kvadratnu grešku. Dobijeni su rezultati za srednje kvadratnu grešku 3.64% za predloženi model, 3.82% za predloženi stacionarni model i 3.57% za model veštačke neuronske mreže. Ovi rezultati definitivno ukazuju na opravdanost predloženog modela ali treba napomenu da su autori koristili tačnu temperaturu kao i da je neuronska mreža trenirana sa malim trening skupom. Takođe horizont prognoze je 168 (cela nedelja) pa bi se sigurno dobili bolji rezultati da je rađena prognoza sa horizontom od 24 tj, iz dana u dan. Sličan pristup baziran na neparametarskom estimatoru za prognozu vršnog opterećenja autori predlažu u [7].

Modelovanje potrošnje električne energije koristeći metodu eksponencijalnog poravnanja (*eng. exponential smoothing*) prikazano je u [8]. Pristup se bazira na činjenici da se potrošnja može predstaviti kao linearna kombinacija poznatih funkcija vremena i komponente šuma: $x(t) = \bar{a}'f(t) + e(t)$, gde su: $\bar{a}(t)$ - vektor kolona sa n konstantnih koeficijenata, $\bar{f}(t)$ - vektor kolona sa n linearno nezavisnih funkcija vremena i $e(t)$ - slučajna promenljiva sa Gausovom raspodelom sa nultim matematičkim očekivanjem i konstantnom standardnom devijacijom (beli šum). Kako se koeficijenti \bar{a} menjaju dovoljno polako da se mogu smatrati konstantnim u vremenu, prognoza potrošnje u trenutku τ se može dobiti ekstrapolacijom prethodne jednačine koristeći poznate funkcije aproksimacije $\bar{f}(t)$ i koeficijente $\bar{a}(T)$: $\hat{x}_\tau(T + \tau) = \bar{a}'(T)\bar{f}(t + \tau)$. Metoda najmanjih kvadrata je korišćena za određivanje koeficijenata sa izabranom konstantom β , $0 < \beta < 1$, koja kontroliše stepen uticaja grešaka. Funkcije aproksimacije su predstavljene Furijeovim redom gde je izbor odgovarajućih harmonika izvršen proračunom autokorelacione funkcije i spektra. Koristeći podatke o potrošnji za period od dve godine, prvo se vrši estimacija parametra modela na osnovu podataka za deset nedelja, pri čemu je izvršen optimalan izbor harmonika i koeficijenta poravnanja, dok je evaluacija modela izvršena na preostalim podacima. Kao meru preciznosti autor koristi standardnu devijaciju koja se kreće u opsegu od 2 do 5 procenta. Predloženi pristup karakteriše jednostavna implementacija, praćenje sedmičnih promena potrošnje kao i uticaj praznika, ali ne uvažava klimatske uticaje na potrošnju. Takođe sama evaluacija je diskutabilna, jer ostaje nejasno zašto autor ne koristi druge greške za meru preciznosti kao npr. srednje kvadratnu grešku.

Box-Jenkins modeli vremenskih serija imaju široku primenu pa su tako prisutni i kod problema prognoze potrošnje električne energije. U [9] autori porede nekoliko Box-Jenkins modela i predlažu pristup za prevazilaženje problema koji karakterišu ove modele, pre svega nemogućnost da modeluju nelinearnu zavisnost potrošnje od temperature. Autori formiraju tri modela: ARIMA model, model prenosne funkcije (*eng. transfer function model* - TFM) i TFM model sa nelinearnom transformacijom. Kod formiranja ARIMA modela, autori koriste autokorelacionu (ACF) i parcijalnu autokorelacionu funkciju (PACF) da bi ispitali stacionarnost procesa i odredili potreban stepen diferenciranja kako bi se formirao model što bliži stacionarnom. Utvrđuju da se radi o periodičnom procesu sa periodom 24 i 168 što odgovara pretpostavci o dnevnoj i sedmičnoj periodičnosti. Parametri modela su određeni principom maksimalne sličnosti (*eng. maximum-likelihood estimation*) a sličan pristup je korišćen i kod formiranja TFM modela. Kako bi modelovali nelinearnu zavisnost potrošnje od temperature, autori ubacuju malu modifikaciju u osnovni TFM model. Oni na osnovu podataka iz prošle godine aproksimiraju temperaturnu krivu polinomom trećeg stepena koju zatim koriste za transformaciju temperature na ulazu u TFM. Za meru preciznosti autori koriste srednju procentualnu apsolutnu grešku (MAPE) a testiranje rade za periode od po 20 dana za zimu, proleće i leto. Predložene modele porede i sa konvencijalnim modelom koji se bazira na usrednjavanju potrošnje po satima za odgovarajući period prošle godine koji se bira na osnovu sezone tj. vremenskog perioda, dana u nedelji i stepenu oblačnosti. Dobijeni rezultati ukazuju na prednost TFM modela sa nelinearnom transformacijom u odnosu na standardni TFM model, sa 3.73% u odnosu na 4.25% i prednost oba modela nad ARIMA modelom koji ima srednju grešku od 4.92%. Takođe svi predloženi modeli su se pokazali bolje od konvencijalnog modela koji ima grešku od 5.75%. Treba napomenuti da su korišćene tačne vrednosti temperature.

Jedan pristup zasnovan na kombinaciji regresionog i ARIMA modela predstavljen je u [10]. Prognoza potrošnje električne energije po časovima za period od sedam dana dobijena je primenom ARIMA metode na prethodno normalizovanu vremensku seriju iz koje su odstranjeni pikovi i trend. Regresiona metoda je korišćena za prognozu vršnog opterećenja sa klimatskim promenljivama za atribute.

Potrošnja električne energije kao vremenska serija se može modelovati sezonskim AR modelom čiji se parametri određuju primenom Box-Jenkins metode, što se ustvari svodi na rekurzivnu metodu najmanjih kvadrata. Kako ovakav pristup zahteva nekoliko složenih operacija u svakoj iteraciji, u [11] je predstavljen modifikovani iterativni metod najmanjih

kvadrata (*eng. iteratively reweighted least squares estimation*) za efikasno određivanje parametara modela. Predložena metoda je primenjena na prognozu potrošnje prvih 168 sati u februaru a model je formiran na osnovu 744 vrednosti potrošnje u predhodnom mesecu (januar). Analizom autokorelacione i parcijalne autokorelacione funkcije određen je odgovarajući model $ARMA(2,0,0) \times (2,0,0)_{24}$. Formirano je nekoliko modela baziranih na različitim težinskim funkcijama i prema prikazanim rezultatima srednje i maksimalne apsolutne greške i standardne devijacije, svi oni prevazilaze osnovni Box-Jenkins model.

Poređenje pet metoda primenjenih na problemu kratkoročne prognoze potrošnje električne energije od kojih četiri možemo svrstati u kategoriju klasičnih a jednu u kategoriju metoda veštačke inteligencije, dato je u [12]. Ovde je urađena prognoza električne potrošnje po satima svakom metodom za dva izabrana dana, jedan u letnjem i jedan u zimskom periodu godine. Primljeni su metodi: višepromenljiva linearna regresija, ARIMA, metod generalnog eksponencijalnog poravnanja (*eng. general exponential smoothing*), metod prostora stanja (*eng. state space method*) i ekspertni sistem baziran na skupu znanja (*eng. knowledge-based expert system*). Dobijeni rezultati ukazuju na različito ponašanje modela u letnjem odnosno zimskom periodu godine. Najveću preciznost u letnjem ima ARIMA model dok su u zimskom to ekspertni sistem baziran na skupu znanja i model prostora stanja. Najmanju preciznost u oba slučaja je pokazao model linearne regresije i pored pokušaja da se dobiju precizniji rezultati formiranjem zasebnih modela za različite periode dana. Takođe treba napomenuti da su autori usvojili pretpostavku da je zavisnost između električne potrošnje i ostalih promenljivih linearna, da su klimatske promenljive (temperatura) poznate i da je sedmična periodičnost zanemarljiva.

Upravo su dnevna i sedmična periodičnost koje karakterišu vremensku seriju električne potrošnje, navele autore u [13] da predlože modifikaciju standardne metode eksponencijalnog poravnanja kako bi se ova karakteristika uključila u model. Inspirisani radom [14] u kome je predstavljen dvostruko periodični ARIMA model, autori predlažu dvostruko periodični model eksponencijalnog poravnanja. Preciznije radi se o multiplikativnom dvostruko periodičnom Holt-Winters metodu, koji uključuje u model još jedan periodični parametar. Podaci o polučasovnoj električnoj potrošnji Engleske i Velsa, za period od 12 sedmica su korišćeni za evaluaciju metoda, 8 sedmica sa 2688 vektora za formiranje modela i 4 sedmice sa 1344 vektora za testiranje modela. Formirano je četiri modela: multiplikativni dvostruko periodični ARIMA model, $ARIMA(2,0,0) \times (2,0,2)_{48} \times (2,0,2)_{336}$ (sa dnevnim periodičnim koeficijentom 48 i sedmičnim 336 jer se radi u polučasovnoj potrošnji) standardni multiplikativni Holt-

Winters metod koji uvažava samo dnevnu periodičnost, metod koji uvažava samo sedmičnu periodičnost i novi dvostruko periodični multiplikativni Holt-Winters metod. Rezultati pokazuju da dvostruko periodičan Holt-Winters model prevazilazi standardne Holt-Winters modele ali i multiplikativni dvostruko periodični ARIMA model. Ovo važi u slučaju kada Holt-Winters modeli sadrže AR(1) model za rezidualne a ukoliko to nije slučaj, ARIMA model ima bolju preciznost.

Vremenski promenljiv model prostora stanja kod modelovanja kratkoročne prognoze potrošnje električne energije u kombinaciji sa Kalmanovim filtrom za rekurzivno određivanje optimalnog skupa parametara modela u svakom koraku, predstavljen je u [15]. Model koji predstavlja jednu modifikaciju standardnog ARIMA modela koja se ogleda u kombinaciji više promenljivih: procene prognoze dispečera, temperature i prošle potrošnje, dat je u [16]. Za razliku od standardnog ARIMA modela, ovaj model za inicijalnu vrednost uzima procenu dispečera, ARMA model koji se prilagođava prirodi procesa (Gausov ili ne) potrošnje električne energije dat je u [17], dok je poređenje standardnih AR i ARIMA modela sa AR(1) modelom sa visokofrekventnim filtrom prikazano u [18]. Klimatski prilagodljiv model prostora stanja sa detekcijom loših vrednosti prikazan je u [19], dok je u [20] dat periodičan AR model sa tehnikom za uklanjanje trenda vremenske serije.

2.2 Modeli veštačke inteligencije

U ovu grupu modela spadaju modeli za prognozu potrošnje električne energije zasnovani na fazi logici, ekspertskim sistemima, veštačkim neuronskim mrežama, metodama potpunih vektora itd.

Model prognoze potrošnje električne energije u [21] je baziran na linearnoj regresiji sa metodom fazi logike za određivanje parametara modela. Parametri modela kao fazi promenljive su predstavljeni trougaonom funkcijom pripadnosti. Autori predlažu dva pristupa, prvi gde se izlazna potrošnja tretira kao ne-fazi promenljiva a parametri modela kao fazi promenljive i drugi gde se izlazna potrošnja tretira kao fazi promenljiva. Ulazne promenljive se tretiraju kao ne-fazi promenljive. Formiraju se modeli posebno za letnji i zimski period. Razlika između ova dva modela je to što letnji model pored temperature koristi i vlažnost vazduha kao klimatsku promenljivu a zimski koristi koeficijent hlađenja vetra umesto vlažnosti vazduha. Takođe formiraju se posebni modeli za radne i dane vikenda u okviru svake sezone. Tako je model za letnji period dat sa:

$$\begin{aligned} \underline{Y}_j(t) = & A_0 + \underline{A}_1 T_j(t) + \underline{A}_2 T_j^2(t) + \underline{A}_3 T_j^3(t) + \underline{A}_4 T_j(t-1) + \underline{A}_5 T_j(t-2) + \underline{A}_6 T_j(t-3) \\ & + \underline{A}_7 H_j(t) + \underline{A}_8 H_j(t-1) + \underline{A}_9 H_j(t-2), \end{aligned} \quad (2.7)$$

gde je $\underline{Y}_j(t)$ prognoza potrošnje u trenutku t , $T_j(t)$ temperatura u trenutku t , $H_j(t)$ vlažnost vazduha u trenutku t , A_0, A_1, \dots, A_9 fazi parametri modela i $t=1, \dots, 24$ sat u danu. Da se primetiti da model koristi temperaturu i vlažnost vazduha u trenutku za koji se radi prognoza potrošnje (t), pa kako ove vrednosti nisu unapred poznate treba koristiti njihove prognozirane vrednosti što opet znači da je potrebno imati 24 vrednosti prognoze temperature ali i vlažnosti vazduha u ovom slučaju. Autori ne spominju da li su koristili tačne ili prognozirane vrednosti za klimatske promenljive niti ukazuju na ovaj konkretan nedostatak predloženog modela.

U [22] su predložena dva modela bazirana na fazi logici za prognozu vršne i ukupne dnevne potrošnje električne energije. Modeli su slični pa tako kao ulazne promenljive koriste vršnu odnosno totalnu potrošnju za prethoni dan i maksimalnu i minimalnu dnevnu temperaturu za tekući dan (dan za koji se radi prognoza). Za funkciju pripadnosti autori koriste trougaonu funkciju, mada probaju i trapezoidnu ali sa manje uspeha. Predloženi model autori porede sa jednostavnim ANN modelom sa istim promenljivama na ulazu kao u slučaju fazi modela i jednim skrivenim slojem sa 7 neurona. Koristeći MAPE grešku za evaluaciju, ANN model je pokazao znatno bolje rezultate iako je jednostavan.

Sličan prethodnom fazi modelu, u [23] je predstavljen fazi model za prognozu časovne potrošnje električne energije sa dve ulazne promenljive, "vreme" tj. trenutak u toku dana i temperatura u toku dana. Obe promenljive su predstavljene trougaonom funkcijom pripadnosti stim što je vreme podeljno na 8 intervala a temperatura na 4. Takođe i izlazna promenljiva je predstavljena sa trougaonom funkcijom pripadnosti podeljenom na 8 intervala. Za testiranje su izabrana 3 dana bez napomene o kojim danima se radi, ali je rečeno da pripadaju različitim periodima u godini. Predložen model je poređen sa konvencijalnim modelom ali se ne spominje o kom konkretno modelu se radi a za evaluaciju rezultata je korišćena relativna greška.

Prognoza potrošnje električne energije primenom ekspertskih sistema prikazana je u [24]. Autori daju dva modela, jedan za prognozu potrošnje od jednog do šest sati unapred i drugi za 24 sata unapred. Kako bi prepoznali ključne promenljive koje oblikuju električnu potrošnju, autori rade korelacionu analizu između potrošnje i klimatskih promenljivih (temperatura, vlažnost vazduha, brzina i pravac vetra itd.) i prepoznaju jaku korelaciju sa temperaturom i to negativnom u zimskoj a pozitivnom u letnjoj sezoni. Takođe primećuju

jaki sezonski uticaj na potrošnju, sa dva pika (ujtru i uveče) u zimskoj sezoni i jednim u letnjoj (kasno popodne). Prepoznaju uticaj tipa dana (dan u nedelji) na potrošnju kao i doba dana (sat u danu). Na osnovu izvršene analize, autori definišu niz pravila kojima se formira struktura sistema. Predložen model testiraju na skupu od 4 meseca (februar, maj, juli i oktobar) i koriste srednju apsolutnu procentualnu grešku kao meru preciznosti. Dobljene rezultate ipak treba uzeti sa rezervom jer je korišćena tačna temperatura umesto prognozirane.

Jedan od prvih pokušaja rešavanja problema prognoze potrošnje električne energije primenom metoda veštačkih neuronskih mreža (ANN) predstavljen je u [25]. Autori predlažu tri modela za prognozu vršne dnevne potrošnje, ukupne dnevne potrošnje i potrošnje po časovima. Tako je struktura ANN za prvi model sačinjena od tri sloja sa tri neurona u ulaznom sloju (vršna, srednja i minimalna dnevna temperatura), pet neurona u srednjem skrivenom sloju i jednim neuronom u izlaznom sloju, a ista struktura ANN je i kod modela za ukupnu dnevnu potrošnju. Nešto drugačija struktura ANN je predložena u slučaju modela za časovnu potrošnju gde u ulaznom sloju ima šest neurona (sat u danu, potrošnja i temperatura u prethodna dva sata i prognozirana temperatura u datom satu), u srednjem skrivenom sloju deset neurona i jedan u izlaznom sloju. Da se primetiti da prvi i drugi model koriste srednju, minimalnu i vršnu dnevnu temperaturu za dan unapred, koje nisu poznate pa se moraju koristiti procenjene vrednosti, što autori ne spominju.

Na pitanje da li je i kada potrebno koristiti nelinearni model kod prognoze potrošnje električne energije, pokušali su da odgovore autori u [26]. Oni predlažu da se prvo ispita da li se radi o linearnom ili nelinearnom procesu (vremenskoj seriji). Pa se tako formiraju linearna autokorelaciona funkcija data sa:

$$r(X(t), X(t-k)) = \frac{\text{Cov}(X(t), X(t-k))}{\sqrt{\text{Var}(X(t))}\sqrt{\text{Var}(X(t-k))}}, \quad (2.8)$$

i nelinearna autokorelaciona funkcija koja se bazira na merenju količine zajedničkih informacija između dve promenljive. Onda se ove dve funkcije direktno porede i ako dođe do poklapanja pokazuje se linearna priroda procesa tj. ukoliko ne dođe do poklapanja, nelinearna priroda. Određivanjem prirode procesa (vremenske serije) određuje se da li je potrebno koristiti linearan ili nelinearna model. Autori testiraju predloženi pristup na podacima o električnoj potrošnji Češke, i usled poklapanja autokorelacionih funkcija potvrđuju da se radi o nestacionarnom linearnom procesu i predlažu linearni ARIMA model koji porede sa modelom veštačkih neuronskih mreža kao tipičnim predstavnikom nelinearnih modela

veštačke inteligencije. ANN struktura je takva da ulazni sloj ima 15 neurona (potrošnja sa regresorom $r=1, \dots, 6, 23, 24, 25, 48, 72, 96, 120, 144, 240$) a srednji skriveni sloj 10 neurona. Prikazani su rezultati prognoze sa horizontom od 1, 12, 24 i 36 sata a za meru preciznosti korišćena su srednja kvadratna greška (MSE), srednja apsolutna procentualna greška (MAPE) i vršna apsolutna procentualna greška (maxAPE). Autori konstantuju da dobijeni rezultati potvrđuju da je u konkretnom slučaju dovoljan ARIMA model, međutim ako se rezultati pažljivije pogledaju, vidi se da to nije slučaj jer se ANN model pokazao boljim, naročito u slučaju kada se uključi temperatura u model.

Interesantan pristup u rešavanju problema prognoze potrošnje električne energije baziran na ANN metodi sa aktivnim učenjem, prikazan je u [27]. Autori predlažu klasterizaciju podataka i formiranje više modela čiji broj bi naravno odgovarao broju formiranih klastera. Struktura modela ostaje ista, Elmanova neuronska mreža sa jednim skrivenim slojem sa 10 neurona osim u slučaju modela za posebne dane kada model uključuje i korekcionni koeficijent. Prvo se uradi korelaciona analiza potrošnje između dana u sedmici i na osnovu dobijenih korelacionih koeficijenata odredi se broj potrebnih klastera. Zatim se izvrši klasterizacija čime se formiraju grupe podataka za formiranje modela. Za testiranje su korišćeni podaci o potrošnji Turske za period od 2 godine. U konkretnom slučaju autori prepoznaju 5 klastera, prvi koji čine podaci o časovnoj potrošnji ponedeljka od 0 do 12 časova, drugi koji čini potrošnja svih ostalih radnih dana uključujući i preostale sate ponedeljka, zatim dva klastera za subotu i nedelju posebno i poslednji koji čine posebni dani (praznici). Predloženi model sa srednjom godišnjom MAPE greškom od 1.6% nadmašuje komparativni ARMA model. Međutim ostaje nejasno koji metod je korišćen za klasterizaciju i na koj način, pošto autori ne pominju niti predlažu neki u radu osim da su prostom analizom izvršili razdvajanje podataka.

Jedna od metoda iz grupe veštačkih neuronskih mreža, samoorganizujuće mape (*eng. self-organizing maps - SOM*), primenjena je kod prognoze potrošnje električne energije u [28]. SOM spada u kategoriju metoda nenadgledanog mašinskog učenja pa se tako koristi uglavnom kod problema klasifikacije i klasterizacije. SOM struktura se bazira na neuronima organizovanim u dvodimenzijalnu mrežu sa relacionim funkcijama između suseda. Tako autori predlažu šestougaoonu strukturu sa Gausovom relacionom funkcijom i veličinom mape ne većom od 10×10 , zbog dobrog odnosa preciznosti i vremena izvršenja. Kako bi optimizovali svoj model, autori rade analizu kako izbor atributa ali i veličina i struktura trening skupa utiču na preciznost. Predloženi model testiraju na skupu o potrošnji električne

energije Španije za period od 4 godine a za meru preciznosti koriste MAPE grešku. Tako predloženi model ima blagu prednost u odnosu na standardni ANN model sa MAPE greškom od 2.18% prema 2.59% i solidnu prednost u odnosu na grešku od 3.18% ARIMA modela. Takođe, pored MAPE greške dat je i pregled kretanja greške u zavisnosti od godine, doba godine i sedmice. Pa su tako meseci januar, mart i decembar sa nešto većom greškom a kad su u pitanju dani u nedelji to je ponedeljak.

Modeli prognoze potrošnje električne energije za 1 i 24 sata unapred primenom *echo state network* (ESN) neuronskih mreža predstavljeni su u [29]. Radi se o rekurentnoj neuronskoj mreži sa retko povezanim skrivenim slojem, obično sa povezanošću od 1% do 3%. Model prognoze za sat unapred ima samo dva atributa: trenutnu časovnu potrošnju (potrošnja prethodnog sata u odnosu na sat za koji se radi prognoza) i prognozu temperature u narednom satu (u satu za koji se radi prognoza), dok model prognoze za 24 sata čine 24 modela za prognozu za sat unapred koji se rekurzivno pozivaju a svaki ima samo tri atributa: trenutnu časovnu potrošnju, prognozu temperature za 24 sata unapred (temperature sata za koji se radi prognoza) i tip dana (dan u sedmici). Ovako mali broj atributa modela je posledica arhitekture same ESN metode. Kako predloženi modeli koriste vrednosti prognozirane temperature, autori predlažu i model za prognozu vrednosti temperature za 1 i 24 sata unapred vodeći se sličnim pristupom kao i kod prognoze potrošnje. Međutim, kao i što se moglo očekivati, predloženi model za prognozu temperature ne daje očekivane rezultate. Predloženi modeli sa MAPE greškom od 1.048% i 2.117% za prognozu potrošnje za 1 i 24 sata unapred, su nešto lošiji u odnosu na hibridne modele date u [30] sa greškom od 0.99% i 2.04%.

Pored spomenutih radova koji se bave prognozom potrošnje električne energije primenom metoda ANN, treba spomenuti i radove [31]–[36].

Metode potpornih vektora postaju popularne od 2001. kada nalaze primenu i kod problema prognoze potrošnje električne energije. Prvi zapaženiji rad na tu temu je [37] koji predstavlja porednički model na EUNIT takmičenju u prognozi vršne dnevne potrošnje. Kako bi identifikovali najvažnije attribute za model, autori rade analizu električne potrošnje stavljajući akcenat na tri ključna uticaja na potrošnju: sezonski (godišnje doba, dan u sedmici), klimatski (temperatura, vlažnosti vazduha itd.) i uticaj praznika (Nova Godina, Božić itd.). Na osnovu analize, autori prepoznaju nekoliko ključnih atributa za formiranje modela: dan u sedmici, temperatura i prošla potrošnja (vršna dnevna potrošnja za prethodnih 7 dana). Formirano je više modela sa različitom strukturom atributa i trening skupa: sa svim

atributima, bez prošle potrošnje, bez temperature, bez prošle potrošnje i temperature. Modeli su trenirani sa trening skupom za celu zimu i sa trening skupom za januar-februar. Model sa atributima: dan u sedmici i prošla potrošnja za 7 dana bez temperature, treniran trening skupom za celu zimu, sa MAPE greškom od 1.95% je najbolji model koji je ujedno i pobednički model takmičenja. Ovaj model je nadmašio slične modele koji imaju i temperaturu za atribut pa čak i onaj koji koristi tačnu vrednost temperature (MAPE 2.7%). Ovo je interesantna činjenica jer je prognoza rađena za 31 dan unapred (mesec januar), tj. za prilično dug horizont za koji je prognoza temperature neprecizna, pa je ovakav rezultat gde model bez uvažavanja temperature nadmašuje model sa tačnom temperaturom, ohrabrujući za modele bez temperature.

Modifikacija metode potpornih vektora, metoda potpornih vektora sa najmanjim kvadratima (*eng. least-squares support vector machines - LS-SVMs*), primenjena na rešavanje problema prognoze potrošnje električne energije data je u [38], [39]. Za razliku od standardne SVM metode, kod LS-SVM kvadratni optimizacioni problem je linearizovan što skraćuje potrebno procesorsko vreme izvršenja i zahteva manje procesorskih resursa. Autori formiraju dva modela za prognozu časovne potrošnje za dan unapred, prvi sa horizontom prognoze od jednog sata i drugi sa 24 časovnim horizontom prognoze. Strukturu modela čine sledeće promenljive: prošla 164 časovna potrošnja (regresor $r=164$), temperatura i kalendarski podaci (mesec, dan u nedelji i sat u danu). Iako je ogroman skup podataka bio dostupan za trening modela, 26000 ulaznih vektora, autori za formiranje modela koriste samo od 1% do 4% tog skupa što je bilo sasvim dovoljno za formiranje dobrog modela. Tako je 400 odnosno 800 potpornih vektora bilo dovoljno za formiranje prvog odnosno drugog modela, da bi njihova srednje kvadratna greška bila ne veća od 2%.

2.3 Hibridni modeli

U ovu kategoriju spadaju modeli koji su nastali kombinovanjem dva ili više modela iz prve dve grupe ili kombinacijom nekog od modela iz prve dve grupe sa nekim heurističkim optimizacionim metodom za određivanje parametara modela, metodom za izbor atributa ili nekom metodom za izbor trening skupa. Najveći broj modela, naročito iz novijeg perioda može se svrstati u ovu kategoriju.

Veštačka neuronska mreža u kombinaciji sa fazi logikom za određivanje parametara sistema predstavlja interesantan pristup dat u [40]. Model ima strukturu sa samo 6 atributa,

koja je ista i u slučaju ANN modela koji se koristi za poređenje: minimalnu i maksimalnu dnevnu temperaturu, vršnu potrošnju prethodnog dana, period dana, sezonu i dan u sedmici. Ulazni sloj sa 6 neurona, dva skrivena sloja sa po 40 neurona i jedan neuron u izlaznom sloju, čine strukturu ANN modela. Pored ANN modela, predloženi model je poreden i sa ARIMA(1,1,1)(1,0,1)⁴⁸ modelom. Period testiranja obuhvata 2 dana tj. horizont prognoze 96, pošto se radi o podacima sa polu časovnim vrednostima potrošnje. Za meru preciznosti korišćena je srednje kvadratna greška koja daje sledeći poredak: predloženi fazi-neuro model, ANN model pa ARIMA model. Ipak, manjak testiranja nas navodi da rezultate uzmemo sa određenom rezervom. Takođe, treba napomenuti nedostatak predloženog modela koji se ogleda u velikom broju parametara koje treba odrediti a koji raste sa porastom broja promenljivih (ulaza): broj i tip funkcija pripadnosti, koeficijente osetljivosti i greške itd. Treba napomenuti i da je struktura ANN modela ista kao i kod predloženog fazi-neuro modela, što za posledicu ima loš rezultat, koji bi bio bolji da je uzet nešto drugačiji skup atributa.

Kako vremensku seriju potrošnje električne energije karakteriše trend, sezonska komponenta i lokalne fluktuacije, autori u [30] predlažu model sa diskretnom talasnom transformacijom za višeslojnu dekompoziciju vremenske serije (signala). Zatim se prognoza radi sa tako transformisanom vremenskom serijom pomoću veštačkih neuronskih mreža. Posle formiranja ANN modela, za fino podešavanje parametra modela koristi se evolucionini algoritam. Dvostruka korelaciona analiza služi za izbor optimalnog skupa atributa kod definisanja strukture ANN modela. Kako bi potvrdili opravdanost predloženog modela, autori porede svoj model sa modelima predloženim u [41] i [42]. U prvom radu se radi o modelu sa SOM metodom za klasterizaciju trening skupa na osnovu čega se formira 24 SVM modela za potrošnju u svakom satu. Podaci za testiranje se odnose na električnu potrošnju grada Njujork u Americi a predloženi model u ovom slučaju sa MAPE greškom od 1.68% i MAE 87.01 MW je bolji od modela u [41] koji ima MAPE 1.82% i MAE 106.97 MW. U drugom radu se radi o ANN modelu za horizont prognoze od jednog do šest sati unapred koji koristi Euklidsko rastojanje za izbor sličnih dana (*eng. similar day approach*). Podaci o električnoj potrošnji Australijske države Viktorija korišćeni su za testiranje, pa su tako za horizont prognoze od jednog sata za mesec decembar i jul dobijene MAPE vrednosti 1.17% odnosno 0.73% ili za 6 sati unapred MAPE vrednosti 3.48% i 1.57% što je lošiji rezultat u odnosu na predloženi model u [30] sa 0.94% i 0.61% za sat i 2.56% i 1.41% za šest sati unapred u istim mesecima.

Dekompozicija signala (*eng. decomposition*) tj. vremenske serije je popularna tehnika koja se koristi kako bi se odstranio šum. Prethodno spomenuta, diskretna talasna transformacija je najčešća tehnika koja se koristi kod vremenske serije električne potrošnje. Slično pristupu u [30], talasna transformacija je korišćena i u [43], [44].

Model višeslojne neuronske mreže sa prostiranjem unapred (*eng. feedforward multilayer neural network*) u kombinaciji sa meta-heurističkim optimizacionim algoritmom roja čestica (*eng. particle swarm optimization - PSO*) za učenje predstavljen je u [45]. Strukturu ulaznih vektora sačinjavaju 5 atributa: vrednosti potrošnje u prethodna dva dana, minimalna dnevna temperatura (prognozirana vrednost temperature za dan za koji se radi prognoza), dan u sedmici i dan u mesecu. Arhitekturu predložene neuronske mreže čine tri sloja, ulazni sa 5 neurona, skriveni sa deset neurona i izlazni sa jednim neuronom. Populacija čestica kod PSO algoritma je stavljena da bude 100 a brzina konvergencije 0.1 kako bi se izbegao lokalni minimum ali i oscilacije vrednosti parametara modela. Predloženi model je poređen sa ANN modelom iste arhitekture ali sa algoritmom propagacije greške unazad (*eng. Back-propagation*) za učenje, i ARMA(1,1) modelom. Na test skupu od 30 dana na kome je prognozirana dnevna potrošnja, predloženi model sa MAPE od 2.52% je najbolji, pa sledi BP model sa 2.82% i na kraju ARMA sa 3.93%.

Problem optimizacije parametra modela kod metode potopnih vektora navelo je autore u [46]–[48] da metod potpornih vektora (SVM) za regresiju kombinuju sa nekim optimizacionim metodom iz grupe takozvanih metoda inteligencije roja (*eng. swarm intelligence*). Tako je u [46] za nalaženje optimalnih parametra modela korišćen metod roja mrava (*eng. ant swarm optimization*) koji je poređen sa tri modela: sa modelom u [49] gde je korišćen jedan evolucionim optimizacionim algoritam (GA) sa SVM, sa modelom u [50] koji koristi optimizacioni algoritam roja čestica (PSO) u kombinaciji sa SVM i sa modelom u [51] koji je baziran na ANN metodu. Dobijeni rezultati koji se odnose na električnu potrošnju Tajvana stavljaju na prvo mesto predloženi model, pa model sa PSO koji prati GA model i na poslednjem mestu je ANN model. Sličan pristup predložen je u [47] gde je SVM regresioni metod kombinovan sa metodom veštačke kolonije pčela (*eng. artificial bee colony algorithm*) za optimizaciju parametara dok je u [48] kombinovan sa genetskim algoritmom (GA). U [46]–[48] kao glavnu prednost predloženih modela, autori navode prevazilaženje problema prevremene konvergencije, sporo nalaženje globalnog optimuma i konvergencije ka lokalnom minimumu. Međutim iako predložene modele porede sa nekoliko drugih modela, nigde ne daju direktno poređenje sa standardnim SVM modelom koji koristi pretragu po mreži (*eng.*

grid-search) sa unakrsnim validacijama (*eng. cross-validation*) za određivanje parametara modela, pa se ne može proceniti da li zaista i koliko predloženi model donosi poboljšanja.

Diferencijalni evolucionni algoritam (*eng. differential evolution algorithm*) je korišćen za optimizaciju parametara regresionog SVM modela u [52]. Za razliku od radova [46]–[48], ovde autori porede predloženi model sa standardnim SVM modelom ali i sa ANN modelom sa propagacijom greške unazad i regresionim modelom. MAPE greška od 1.1% za predloženi model u odnosu na 4.8% za standardan SVM regresioni model, 1.5% za ANN model i 2.58% za regresioni model, ukazuje na opravdanost korišćena predložene metodologije. Treba napomenuti da je ovde rađena prognoza godišnje električne potrošnje Pekinga i da je test skup veoma mali, pa je poželjno predloženi model ispitati na znatno većem skupu.

Pored spomenutih, modeli za prognozu električne potrošnje sa meta-heurističkim algoritmima za određivanje parametra regresionog metoda predstavljani su i u [53]–[58]. Tako je u [55] algoritam kolonije mrava (ACO) korišćen za izbor optimalnog skupa atributa u kombinaciji sa SVM kao regresionim metodom. Metod potpornih vektora sa genetskim algoritmom za optimizaciju parametra modela dat je u [57]. Sličan model predložen je u [56], gde je SVM metod kombinovan sa algoritmom simulacije žarenja (*eng. simulated annealing*). Kombinovani hibridni model baziran na ARIMA metodi, sezonskoj eksponencijalnoj metodi i SVM metodi, sa PSO algoritmom za nalaženje parametara modela predstavljen je u [58].

Uporedno poređenje nekoliko modela: dvostruko periodični ARIMA model ($(ARIMA(2,0,1) \times (2,0,1)_{48} \times (1,0,2)_{336})$), dvostruko periodični eksponencijalni Holt-Winters model, ANN model i regresioni model u kombinaciji sa metodom analize glavnih kompeneti (*eng. principal components analysis - PCA*), dato je u [59]. Spomenuti modeli više pripadaju u prehodne dve kategorije ali su stavljeni ovde jer predstavljaju modifikacije osnovnih modela. Modeli su testirani na dva skupa podataka, na podacima polu-časovne električne potrošnje Engleske i Velsa i časovne potrošnje grada Rio de Ženeiro u Brazilu. Iako su računane i ostale greške, kao glavna mera preciznosti korišćena je MAPE greška. Svi modeli se na oba test skupa ponašaju slično iako ovi test skupovi imaju malo sličnosti i svaki ima određene karakteristike. Tako je dvostruko periodični eksponencijalni model pokazao najbolji rezultat a prati ga regresioni model sa PCA metodom. Sledi ARIMA model a ANN model se iznenađujuće pokazao najlošije. Razlog lošeg rezultata ANN modela može da bude činjenica da ne postoje izrazite nelinearnosti testiranih vremenskih serija kao i mala količina šuma. Valjanost predloženog dvostruko periodičnog eksponencijalnog modela potvrđena je u [60] gde je urađena evaluacija na skupovima električne potrošnje 10 Evropskih zemalja.

Predloženi model je opet postigao najbolji rezultat dok ga prate regresioni model sa PCA i dvostruko sezonski ARIMA model.

Kratkoročna prognoza za 10 do 30 minuta unapred spada u grupu VSTLF problema, gde je od interesa vrednost potrošnje u svakom minutu. U [61] je formirano nekoliko modela i dvostruko periodični eksponencijalni model je pokazao najbolje rezultate za prognozu do 30 minuta unapred. Međutim kada se horizont prognoze produži na period veći od 30 minuta, predloženi model za VSTLF nije više pogodan i autori preporučuju da se u tom slučaju koriste modeli koji uvažavaju klimatske promenljive.

Dvostruko periodični modeli uvažavaju dnevnu i sedmičnu periodičnost koje karakterišu potrošnju električne energije. Međutim kod potrošnje električne energije prisutan je i uticaj godišnjeg ciklusa. On se može uključiti u dvostruko periodični model tako što će umesto dnevne periodičnosti biti ubačena godišnja ali se onda gubi uticaj dnevnog ciklusa na potrošnju. Kako bi se uvažio efekat godišnje periodičnosti električne potrošnje u [62] je predstavljen trostruko periodičan ARMA i eksponencijalni Holt-Winters model. Korišćeni podaci su polu-časovne vrednosti električne potrošnje Britanije i Francuske za period od 6 godina, pa je period od 5 godina sa 87648 vrednosti korišćen za formiranje modela a ostatak od 17520 vrednosti za testiranje. Prikazani rezultati nedvosmisleno ukazuju na prednost predloženih trostruko periodičnih modela u odnosu na dvostruko periodične sa dnevnom i sedmičnom periodičnošću ali i sa sedmičnom i godišnjom periodičnošću. Takođe blagu prednost ima trostruko periodični ARMA model u odnosu na eksponencijalni Holt-Winters model.

Hibridni model koji kombinuje metod potpornih vektora, veštačke neuronske mreže i dvostruke Kalmanove filtre predložen je u [63]. Predloženi model prvo koristi regresioni metod potpornih vektora da bi odredio strukturu i inicijalne parametre neuronske mreže. Zatim se Kalmanovi filtri koriste za nalaženje optimalnih parametra neuronske mreže. Na kraju se optimalna troslojna neuronska mreža koristi za prognozu električne potrošnje. Podaci za testiranje koji su korišćeni odnose se na električnu potrošnju Tajvana a predloženi model je poređen sa neuronskom mrežom sa opadajućim gradijentom (*eng. gradient descend*) i modelom sličnim predloženom ali bez regresione metode potpornih vektora. Posebno su formirani modeli za radne, dane vikenda i praznike sa horizontom prognoze od 24, 72 i 128 sata (1, 3 i 7 dana). Predloženi model sa MAPE greškom od 0.6%, 1.03% i 0.99% u radnim danima za horizont prognoze od 24, 72 i 168 sati prednjači u odnosu na sličan model bez SVM-a koji ima 0.79%, 1.84% i 2.03%, odnosno ANN modela sa opadajućim gradijentom sa

0.84%, 2.96% i 5.97%. Isti poredak uspešnosti modela je i u slučaju prognoze za dane vikenda i dane praznika.

Spomenuti radovi iz grupe hibridnih modela su najzapaženiji radovi u oblasti koji su pažnju na sebe privukli jedinstvenim pristupom u rešavanju problema prognoze potrošnje električne energije sa manje ili više uspeha. Međutim oni čine samo mali broj radova koji je publikovan na temu prognoze potrošnje električne energije. Nešto manje zapaženiji radovi koji takođe zaslužuju pažnju zbog svog doprinosa u rešavanju ovog problema su [64]–[82].

3 NADGLEDANO MAŠINSKO UČENJE

Nadgledano mašinsko učenje je grana veštačke inteligencije a sam termin se odnosi na formiranje i proučavanje sistema koji mogu da uče iz skupova podataka. U ovom poglavlju data je definicija i osnove nadgledanog mašinskog učenja. Nakon toga sledi oblast problema koju ono pokriva i definicije pojmova koji se odnose na algoritme nadgledanog mašinskog učenja.

3.1 Definicija mašinskog učenja

Mašinsko učenje se bavi razvojem i proučavanjem metoda koje omogućavaju “programiranje” računara da uče. Arthur Samuel je 1959. godine definisao mašinsko učenje kao: “Oblast koja proučava i daje računarima mogućnost da uče bez da budu eksplicitno programirani”. Tom M. Mitchell je dao često citiranu, formalnu definiciju: “Za računarski program se može reći da uči iz iskustva E u odnosu na neke vrste zadataka T i merilo performansi P , ako se njegove performanse na zadacima iz T , merene sa P , poboljšavaju sa iskustvom E ” [83].

3.2 Oblast mašinskog učenja

Od pojave prvog elektronskog programibilnog računara (ENIAC, 1945.) gotovo da nema oblasti u kojoj se računari ne koriste za rešavanje problema. Za većinu tih problema je moguće napraviti odgovarajući program mada se u praksi javljaju i mnogi problemi za koje je to isuviše složeno ili u nekim slučajevima čak i nemoguće učiniti. Ovi problemi se mogu svrstati u jednu od četiri kategorije [84].

U prvu kategoriju spadaju problemi za čije rešavanje ne postoje eksperti. Na primer, u savremenim automatizovanim proizvodnim postrojenjima potrebno je predvideti otkaz ili kvar mašine pre nego što do njega dođe, očitavanjem informacija koje pristižu sa senzora. Ukoliko su mašine nove, ne postoje eksperti koji mogu pružiti neophodno ekspertsko znanje programerima kako bi razvili računarski sistem za automatsku predikciju kvarova. Međutim, sistem zasnovan na mašinskom učenju može sam da analizira sačuvane podatke o ispravnom i neispravnom radu mašine, i da na osnovu njih formira predikcioni model.

Drugu kategoriju čine problemi za čije rešavanje postoje eksperti, ali nisu u mogućnosti da efektivno prenesu i implementiraju svoje znanje. To je slučaj u mnogim perceptivnim zadacima, kao što su: prepoznavanje govora, prepoznavanje rukopisa i tumačenje govornog jezika. Većina ljudi je veoma uspešna u obavljanju ovih zadataka, ali niko od njih ne može opisati detaljne korake kako se ovi procesi odvijaju. Međutim, mogu se obezbediti karakteristični skupovi pravilno klasifikovanih primera, tako da algoritmi mašinskog učenja mogu da “nauče” da klasifikuju nove, nepoznate primere.

U treću kategoriju spadaju problemi gde se posmatrane pojave dinamički menjaju u vremenu. Na primer, u finansijskoj ekonomiji od interesa je da se predvidi buduće ponašanje berze, kursa neke valute ili potražnje određene grupe proizvoda. Pošto se ponašanje ovakvih sistema menja u toku vremena, čak i u slučajevima kada je moguće napraviti program koji bi vršio predikciju, taj program bi povremeno morao da se ažurira ili čak pravi ponovo. Uz pomoć algoritama mašinskog učenja, moguća je stalna promena i podešavanje skupa parametara predikcionog modela, kako bi se na adekvatan način propratile promene u ponašanju posmatranog sistema.

Četvrtu kategoriju čine aplikacije koje je potrebno individualno prilagoditi svakom korisniku ponaosob. Uzmimo na primer program koji treba da filtrira neželjenu elektronsku poštu, takozvani „spam“ filter. Različitim korisnicima će biti potrebni različiti tipovi filtera. Ne može se očekivati da svaki korisnik sam formira skup svojih pravila za „spam“ filter i da ih povremeno ažurira. Sistem zasnovan na mašinskom učenju može sam da “nauči” koju poštu korisnik svrstava kao nepoželjnu, a koju zadržava, i na taj način automatski ažurira „spam“ filter sistem.

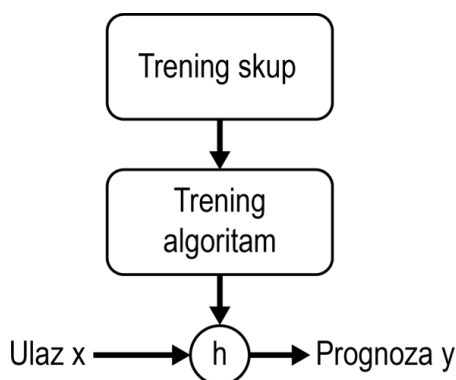
Mašinsko učenje razmatra iste oblasti istraživanja kao statistika i „data-mining“, ali sa drugačijeg aspekta. Statistika se fokusira na razumevanje procesa koji generišu podatke, često sa ciljem testiranja različitih hipoteza o tim procesima. Data-mining tehnike nastoje da

pronađu šablone u podacima koji su razumljivi ljudima. Nasuprot tome, mašinsko učenje se prvenstveno bavi tačnošću i efikasnošću rezultujućeg sistema. Na primer, razmotrimo različite aspekte sa kojih se mogu posmatrati podaci o govoru. Pristup mašinskog učenja se fokusira na formiranje preciznog i efikasnog sistema za prepoznavanje govora. Statistički pristup se fokusira na testiranje hipoteza o mehanizmima koji su u osnovi prepoznavanja govora, dok se kod data-mining pristupa traže obrasci u podacima o govoru.

3.3 Osnove nadgledanog mašinskog učenja

Ovde je definisana osnovna terminologija i notacija nadgledanog mašinskog učenja sa ciljem lakšeg praćenja i razumevanja daljeg sadržaja disertacije.

Označimo sa $x_j^{(i)}$ ulazne promenljive, odnosno attribute (*eng. inputs, features*), a sa $y^{(i)}$ izlazne, odnosno ciljne (*eng. output, target*) promenljive koje pokušavamo da prognoziramo. Par $(x_j^{(i)}, y^{(i)})$, $j = 1, \dots, n$ označava jedan trening primer (*eng. training example*), a skup koji će biti korišćen za obučavanje $S = \{(x_j^{(i)}, y^{(i)})\}, i = 1, \dots, N$ naziva se trening skup. U ovako definisanom trening skupu imamo N trening vektora $x_j^{(i)}$, svaki od njih se sastoji od n atributa, i svakom od njih je pridružena poznata skalarna vrednost ciljne promenljive $y^{(i)}$. Takođe, za oznaku prostora ulaznih vrednosti se koristi $X \in R^n$, a za oznaku prostora ciljnih vrednosti $Y \in R$.



Slika. 3.1 Funkcija hipoteze $h(x)$

Zadatak algoritama nadgledanog mašinskog učenja je da na osnovu zadatog trening skupa S “nauče” funkciju prognoze (*eng. hypothesized function*) $h: X \rightarrow Y$, tako da $h(x)$ budu “dobre” aproksimacije za odgovarajuće vrednosti ciljnih promenljivih y , Slika 3.1.

Formalno gledano, $h \in H$ predstavlja jednu od hipoteza o funkciji koja treba da bude naučena na osnovu zadanog trening skupa S , gde H predstavlja konačan (u nekim slučajevima i beskonačan) skup hipoteza.

U disertaciji se umesto pojma naučiti funkciju prognoze $h(x)$, koristi manje formalan inženjerski termin formirati (trenirati) model. Predikcioni model se formira pomoću trening algoritma, na osnovu izabranog trening skupa, koji se sastoji od vektora čiji su elementi atributi. U slučaju kada su promenljive koje je potrebno prognozirati kontinualne, tada se problem definiše kao regresioni. Ukoliko prognozirane vrednosti mogu da sadrže samo ograničen skup diskretnih vrednosti, tada se problem definiše kao klasifikacioni.

U kontekstu prognoze vremenskih serija, trening i test skupovi su formirani na osnovu date vremenske serije $y = \{y_1, y_2, \dots, y_{n-1}, y_n, y_{n+1}, \dots, y_i, \dots, y_N\}, i = 1, \dots, N$, gde N predstavlja ukupan broj njenih elemenata, a r veličinu regresora (broj prethodnih elemenata na osnovu kojih se vrši prognoza narednog). Svi trening i test vektori se sastoje od r elemenata vremenske serije y , i može ih biti najviše $(N-r)$. Svaki element ovih vektora predstavlja jedan atribut, koji je definisan vrednošću vremenske serije y u datom trenutku.

U nastavku su na primerima linearne i polinomne regresije neformalno definisani osnovni pojmovi koji se odnose na algoritme nadgledanog mašinskog učenja.

Uzmimo na primer da je potrebno prognozirati potrošnju električne energije na osnovu vremenskog trenutka u toku dana (sata u danu), koristeći samo ovaj jedan atribut. Naravno, u realnom problemu ovog tipa skup atributa bi bio znatno veći, ali je ovde problem uprošćen zbog vizualizacije podataka. Da bismo primenili nadgledano mašinsko učenje, moramo prvo odlučiti na koji način ćemo da predstavimo hipotezu h u algoritmu. Uzmimo na primer najjednostavniji slučaj, da smo se opredelili za linearnu funkciju:

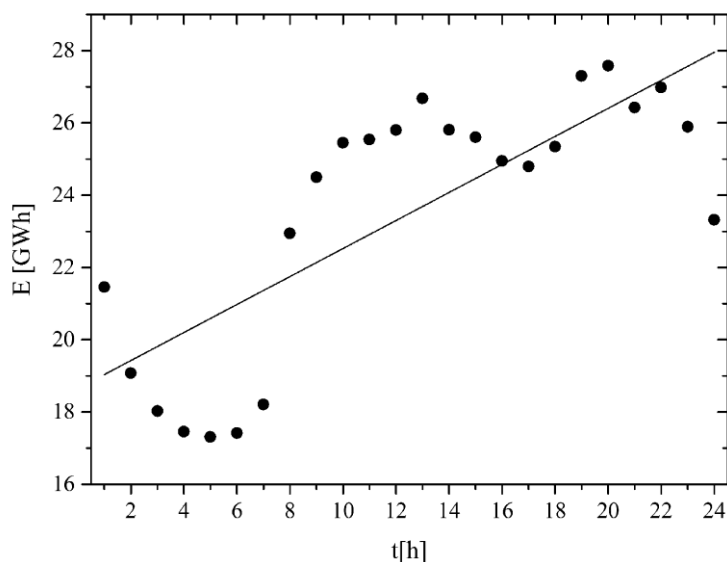
$$h_{\theta}(x) = \theta_0 + \theta_1 x \quad (3.1)$$

gde su θ_0 i θ_1 parametri koji parametrizuju prostor linearnih funkcija mapiranjem iz X u Y . Jedan od načina za optimizaciju parametara θ_i , je da se izaberu tako da $h(x)$ bude što bliže y , makar za sve primere u trening skupu. Definišimo funkciju koja će za sve vrednosti parametara θ_i meriti koliko blizu je $h(x^{(i)})$ do odgovarajućeg $y^{(i)}$. Ova funkcija se naziva optimizaciona ili kriterijumska funkcija (*eng. cost function, loss function*) i ona definiše optimizacioni problem koji treba rešiti:

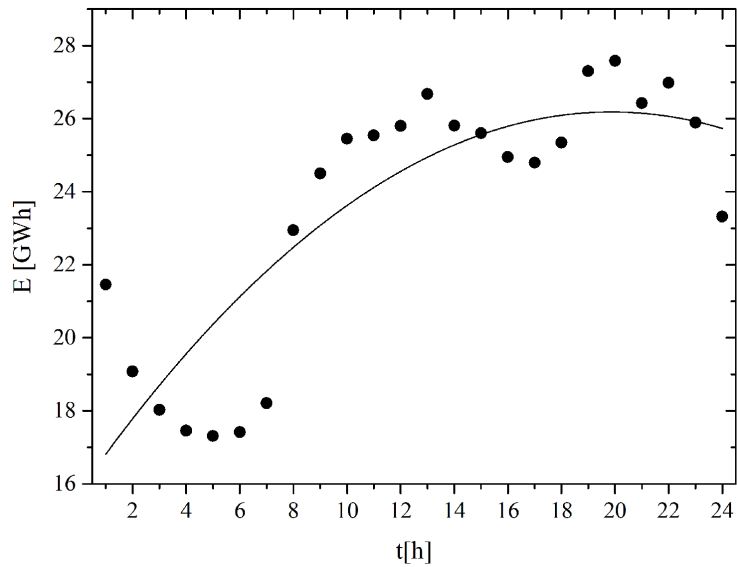
$$J(\theta) = \frac{1}{2} \sum_{i=1}^N (h_{\theta}(x^{(i)}) - y^{(i)})^2 \quad (3.2)$$

Dakle, potrebno je odrediti parametre θ_i koji minimizuju $J(\theta)$. Jedan od načina za to je korišćenje metode gradijentnog spusta (*eng. gradient descent*). Problem definisan izrazom (3.2) ima fiksni i konačan skup parametara θ_i koji se aproksimiraju prema podacima trening skupa. Za ovakav tip algoritma učenja kažemo da je parametarski, jer nakon određivanja parametara θ_i nema potrebe za čuvanjem trening skupa u memoriji radi naredne prognoze. Nasuprot tome, kod neparametarskih algoritama potrebno je čuvati ceo trening skup (nekada samo jedan njegov deo) radi naredne prognoze. Možemo primetiti da je u (3.2) kriterijumska funkcija konveksna (definiše konveksan optimizacioni problem). Takođe, ovako definisana kriterijumska funkcija, koja minimizuje greške na trening skupu predstavlja primer empirijske minimizacije rizika (*eng. empirical risk minimization, EMR*)[85]. Pored EMR, statistička teorija učenja definiše i strukturnu minimizaciju rizika (*eng. structural risk minimization, SRM*), o kojoj se više detalja može pronaći u [86], [87].

Pretpostavimo da smo odredili parametre θ_i i da na Slici 3.2 imamo rezultat aproksimacije linearne funkcije $y = \theta_0 + \theta_1 x$ na trening skupu.



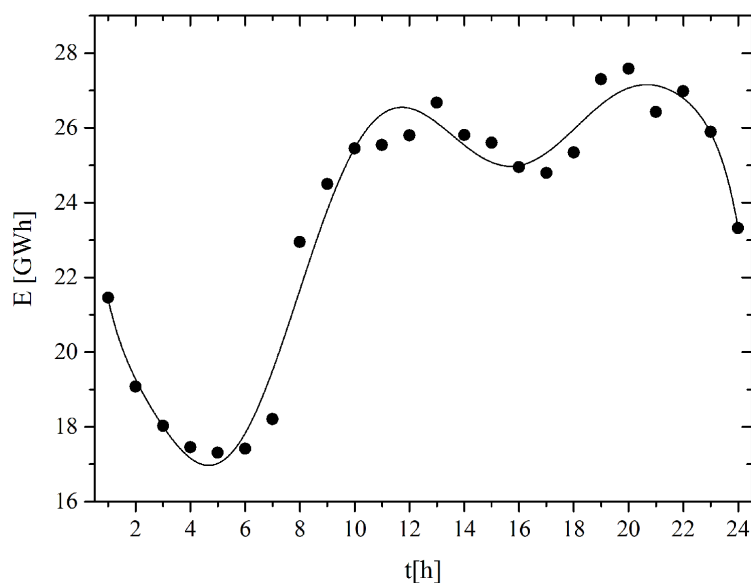
Slika. 3.2 Aproksimacija linearne funkcije na trening skupu



Slika. 3.3 Aproksimacija kvadratne funkcije na trening skupu

Sa Slike 3.2 se može uočiti da se većina tačaka ne nalazi na liniji, tako da linearna funkcija ne predstavlja baš najbolji izbor za funkciju $h(x)$. Ukoliko uvedemo dodatni atribut u trening skup, možemo aproksimirati kvadratnu funkciju $y = \theta_0 + \theta_1 x + \theta_2 x^2$, Slika 3.3. Primećujemo da kvadratna funkcija mnogo bolje aproksimira elemente trening skupa. Na prvi pogled može se učiniti da će aproksimacija biti bolja ukoliko dodamo još atributa, međutim to ne mora da znači. Na Slici 3.4 vidimo rezultat aproksimirane krive polinomom devetog

stepena $y = \sum_{i=1}^9 \theta_i x^i$.



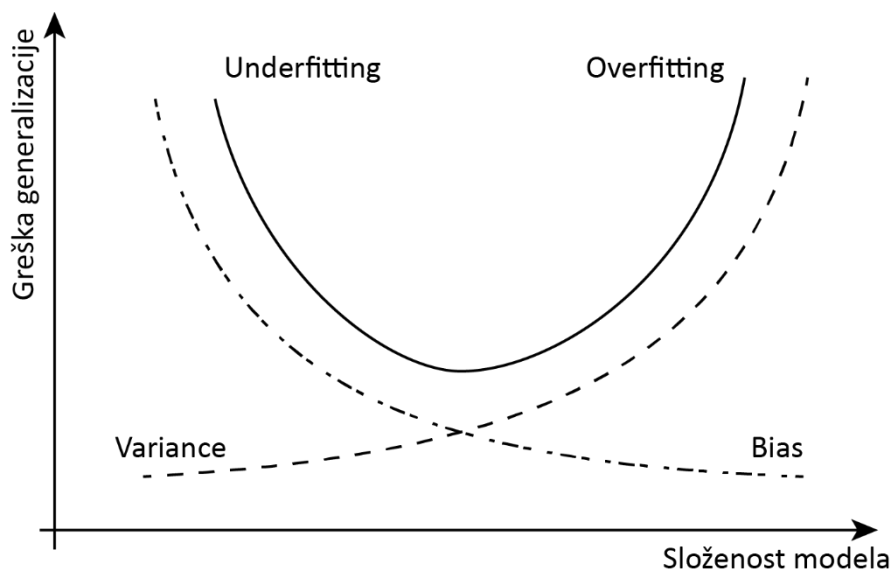
Slika. 3.4 Aproksimacija polinom devetog stepena na trening skupu

Sa Slike 3.4 se može uočiti da bez obzira što aproksimirana kriva prolazi kroz sve elemente trening skupa, ona neće rezultovati dobrom prognozom potrošnje električne energije. Linearna funkcija, prikazana na Slici 3.2 je tipičan primer pod-podešavanja (*eng. underfitting*), kod koga model nije dobro obučen na osnovu trening skupa, odnosno previše je jednostavan. S druge strane, polinom devetog stepena, prikazan na Slici 3.4 je tipičan primer nad-podešavanja (*eng. overfitting*), kod koga je model pre-treniran na trening skupu, odnosno previše je složen. Bez obzira što bi polinom devetog stepena davao dobru prognozu potrošnje na primerima koji se nalaze u trening skupu, nije za očekivati da bi dobili dobru prognozu na nepoznatim primerima iz test skupa. Ovde se polazi od pretpostavke da su elementi trening i test skupa međusobno nezavisne slučajne promenljive koje pripadaju istoj raspodeli (*eng. independent and identically distributed random variables, IID*), odnosno da potiču od strane istog procesa. Drugim rečima, predikcioni model obučen na osnovu trening skupa ne generalizuje dobro primere izvan njega. Ovim se neformalno definiše pojam generalizacije modela, odnosno greške generalizacije (*eng. generalization error*) koja predstavlja očekivanu vrednost greške na primerima koji ne moraju biti u trening skupu.

Linearni model i polinom devetog stepena imaju veliku grešku generalizacije. Međutim, razlog zbog koga se ona javlja je različit kod ova dva modela. Ukoliko je veza između x i y nelinearna, čak i ako aproksimiramo linearni model na trening skupu koji sadrži veoma mnogo primera, on neće moći da ga predstavi dovoljno precizno. Ovim se neformalno definiše očekivana vrednost greške generalizacije modela (odstupanje od prave vrednosti parametara koji se procenjuju) - bias (*eng. bias*). Linearni model iz prethodnog primera ima veliku očekivanu grešku generalizacije modela (tj. bias), odnosno može biti sklon pod-podešavanju (*underfitting*).

Pored bias-a, postoji i druga komponenta koja utiče na grešku generalizacije, varijansa (*eng. variance*) modela. U slučaju aproksimacije polinomom devetog stepena iz prethodnog primera, postoji rizik da treniramo model na osnovu relativno malog (ograničenog) skupa primera koji se nalaze u trening skupu, ali koji ne reprezentuju na odgovarajući način veze koje se mogu javiti između x i y . U našem primeru do ovoga bi moglo doći ukoliko se u trening skupu nalaze vrednosti potrošnje koje znatno odskaču u odnosu na prosek u odgovarajućem vremenskom trenutku (satu u danu). Aproksimiranjem i ovakvih primera iz trening skupa može doći do porasta greške generalizacije. Ovim se neformalno definiše varijansa modela, odnosno model može biti sklon nad-podešavanju (*eng. overfitting*).

U praksi je neophodno napraviti balans između bias-a i varijanse modela (*eng. bias-variance tradeoff*), Slika 3.5.



Slika. 3.5 Balans između bias-a i varijanse

Ukoliko je model previše jednostavan i ima isuviše malo parametara, može se desiti da ima veliki bias, a malu varijansu. Naprotiv, previše složen model, sa previše parametara može imati veliku varijansu, a mali bias. U prethodnom primeru aproksimacija kvadratnom funkcijom daje dobar balans između bias-a i varijanse.

U prethodnom primeru pokazano je da ni linearna funkcija ni polinom devetog stepena ne predstavljaju dobar izbor za predikcioni model. Nameće se pitanje, kako sprečiti nad-podešavanje drugog modela. Ovo je moguće učiniti uvođenjem regularizacionog člana u kriterijumsku funkciju (*eng. cost function*):

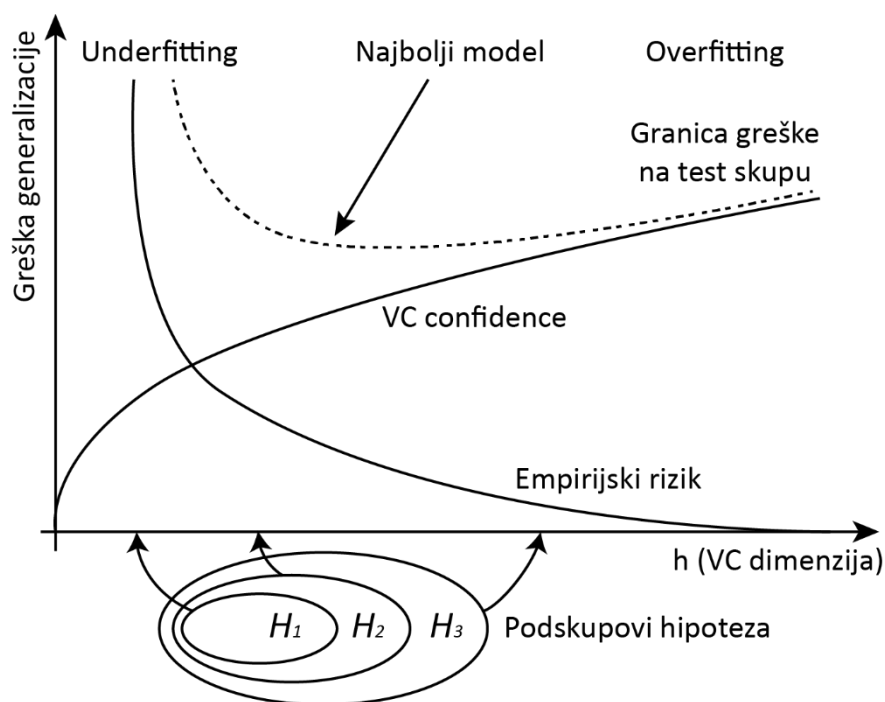
$$J(\theta) = \frac{1}{2} \sum_{i=1}^N (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^n \theta_j^2 \quad (3.3)$$

gde $\lambda > 0$ predstavlja regularizacionu promenljivu (*eng. regularization term*) [88], čija vrednost nije unapred poznata za dati skup podataka, već se može odrediti na osnovu trening skupa primenom tehnika za izbor modela (npr. grid-search i cross-validation). Regularizacija ograničava prostor hipoteza H . U konkretnom primeru potrebno je ograničiti prostor polinomnih funkcija 9-tog stepena. Restrikcija prostora hipoteza H sprečava nad-podešavanje ograničavajući formu potencijalnih funkcija, i na taj način sprečava izbor onih funkcija koje imaju vrednost empirijskog rizika veoma blisku nuli. Uzmimo za primer ekstreman slučaj da je $\lambda = 10^{10}$. Tada bi svi parametri θ_i , osim θ_0 bili jednaki nuli (zbog uslova minimizacije

$J(\theta)$) i model bi se sveo na $h_\theta(x) = \theta_0$. Ukoliko je $\lambda = 0$, model bi se sveo na

$$h_\theta(x) = \sum_{i=1}^9 \theta_i x^i.$$

Kao što je ranije spomenuto, pored empirijske minimizacije rizika, koja minimizuje greške na trening skupu, mnogi algoritmi nadgledanog mašinskog učenja se zasnivaju na naprednijem principu strukturne minimizacije rizika. Strukturna minimizacija rizika definiše princip izbora funkcije hipoteze prilikom treninga modela sa skupovima konačne veličine [89]. Ona pravi balans između složenosti prostora hipoteza (VC dimenzija aproksimacionih funkcija) i kvaliteta aproksimacije na trening skupu (empirijske greške), Slika 3.6.



Slika. 3.6 Strukturna minimizacija rizika

Procedura strukturne minimizacije rizika može se simbolički predstaviti kroz sledeće faze:

1. Koristeći prethodno znanje o problemu, izabrati klasu funkcija (hipoteza), kao što su na primer polinomi n -tog stepena.
2. Podeliti klasu funkcija u hijerarhijski ugnježdene podskupove, u poretku povećanja složenosti. Na primer, polinomi rastućeg stepena.
3. Izvršiti empirijsku minimizaciju rizika na svakom podskupu. Na primer, optimizacija parametara θ_i kod polinomne regresije metodom gradijentnog spusta.

4. Izaberi onu funkciju, čija je suma empirijskog rizika i VC člana iz izraza (3.4) minimalna.

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\frac{h(\log(2N/h) + 1) - \log(\eta/4)}{N}} \quad (3.4)$$

Izraz (3.4) definiše gornju granicu greške generalizacije, odnosno maksimalnu očekivanu grešku na test skupu, koja se sastoji od empirijskog rizika procenjenog na trening skupu i VC člana (*eng. VC confidence*). Podrazumeva se da su elementi trening i test skupa međusobno nezavisne slučajne promenljive koje pripadaju istoj raspodeli. Izraz (3.4) važi sa verovatnoćom $(\eta - 1)$, gde je $0 \leq \eta \leq 1$, $N > h$ je broj trening primera, a h predstavlja *Vapnik-Chervonenkis (VC)* dimenziju, odnosno kapacitet trening algoritma. VC dimenzija predstavlja meru kapaciteta klasifikacionog algoritma, koja se definiše kao kardinalnost najvećeg skupa tačaka koju taj algoritam može da razdvoji (*eng. shatter*).

Više informacija o statističkoj teoriji učenja (*eng. statistical learning theory*), koja formalno definiše prethodno opisane pojmove, se može pronaći u [85]–[87]. U Poglavlju 4 detaljno je analiziran jedan od algoritama nadgledanog mašinskog učenja, LS-SVM, koji je korišćen za formiranje predikcionih modela.

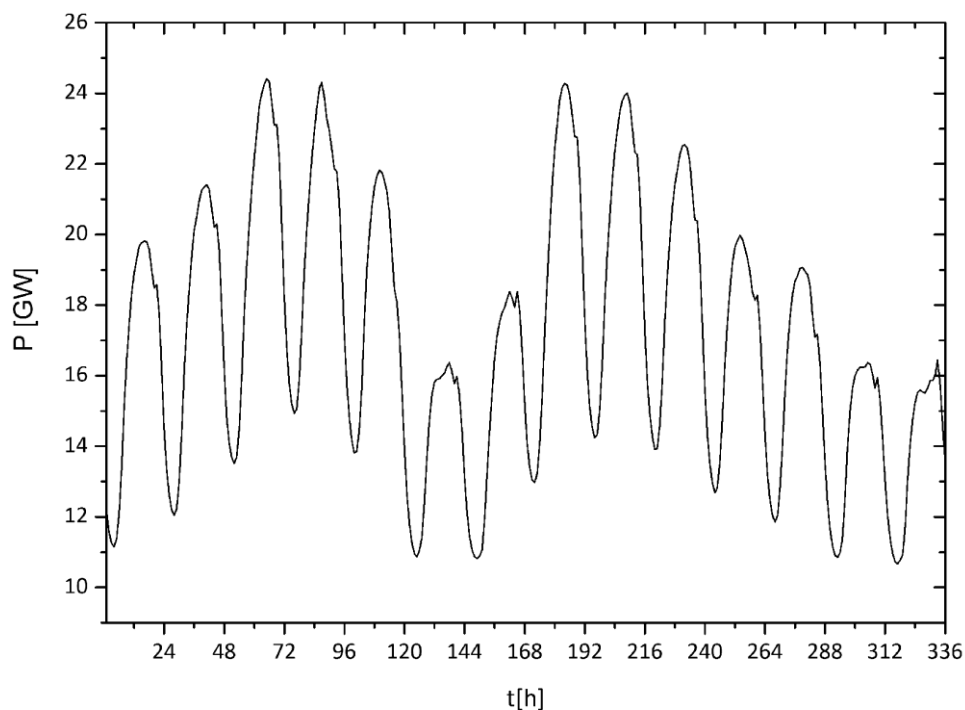
3.4 Prognoza vremenskih serija

Iako pojam vremenskih serija nije direktno vezan za nadgledano mašinsko učenje, već se razne metode nadgledanog mašinskog učenja koriste kao predikcione metode kod problema prognoze vremenskih serija, ovde su dati osnovni pojmovi vremenskih serija i kratak osvrt na strategije predikcije kod vremenskih serija. Takođe treba napomenuti da je sam pojam vremenska serija od interesa zato što upravo potrošnja električne energije predstavlja jednu vremensku seriju.

Vremenska serija predstavlja sekvencu vrednosti, dobijenih merenjima koja se obično izvode u uzastopnim vremenskim trenucima koji su međusobno ekvidistantni. Pa se tako i potrošnja električne energije može posmatrati kao vremenska serija koju sačinjavaju izmerene vrednosti u uzastopnim vremenskim trenucima, vrednosti merene na 15 minuta, 30 minuta, jednog sata ili dana.

U opštem slučaju, vremenska serija može ispoljavati nelinearnost, nestacionarnost, periodičnost, prisustvo haotične komponente i prisustvo šuma. Vremenska serija je linearna

ukoliko se njene buduće vrednosti mogu predstaviti preko linearne funkcije njenih prethodnih vrednosti. Ukoliko se statistička svojstva (srednja vrednost i varijansa) vremenske serije ne menjaju tokom vremena ona se smatra stacionarnom. Vremenska serija koja sadrži komponente koje se periodično ponavljaju, smatra se periodičnom. Specijalni slučaj ovoga su takozvane sezonske vremenske serije koje imaju ponovljivo i predvidljivo pomeranje oko trenda u određenom intervalu. Pored toga vremenska serija može sadržati haotičnu komponentu i šum, koji može biti prisutan samo u njenim pojedinim delovima ili u celoj vremenskoj seriji. Ove osnovne karakteristike vremenskih serija mogu se naći i kod vremenske serije potrošnje električne energije kao što se vidi na Slici 3.7.



Slika. 3.7 Osnovne karakteristike vremenskih serija

Kada je reč o strategiji predikcije kod vremenskih serija, zadatak strategije predikcije vremenskih serija je da definiše kako se vrši prognoza narednih H vrednosti $[y_{N+1}, \dots, y_{N+H}]$ vremenske serije na osnovu njenih prethodnih N vrednosti $[y_1, \dots, y_N]$, gde $H > 1$ označava veličinu horizonta prognoze, a N predstavlja ukupan broj elemenata vremenske serije. Ovde su spomenute tri osnovne strategije (pristupa) kod prognoze vremenskih serija, rekurzivna, direktna i direkno-rekurzivna a postoje još MIMO (eng. *multiple inputs-multiple outputs*) [90] i DIRMO (eng. *Direct-multiple outputs*) [91] koje nisu zastupljene kod prognoze potrošnje električne energije.

3.4.1 Rekurzivna strategija

Ovo je najstarija i najintuitivnija strategija prognoze (*eng. recursive prediction strategy*) [92]–[97]. U literaturi je takođe poznata i pod nazivima: iterativna (*eng. iterative*), rekurentna (*eng. recurrent*) i multi-stage strategija, s obzirom na to da koristi već prognozirane vrednosti umesto poznatih, koje u datim trenucima nisu na raspolaganju za prognozu budućih vrednosti. Kod rekurzivne strategije prvo se vrši prognoza za jedan korak unapred:

$$\hat{y}_{t+1} = f(y_t, y_{t-1}, \dots, y_{t-r+1}, \theta). \quad (3.5)$$

Broj prethodnih vrednosti vremenske serije na osnovu kojih se vrši prognoza narednih vrednosti definisan je veličinom regresora r (*eng. regressor size, window size*), f označava predikcioni model, dok θ definiše skup parametara modela koje je potrebno optimizovati. Za prognozu narednog koraka, koristi se isti model, sa istim skupom parametara:

$$\hat{y}_{t+2} = f(\hat{y}_{t+1}, y_t, y_{t-1}, \dots, y_{t-r+2}, \theta), \quad (3.6)$$

i na kraju za prognozu H -tog koraka, gde H označava veličinu horizonta prognoze, \hat{y}_{t+H} se dobija istim modelom:

$$\hat{y}_{t+H} = f(\hat{y}_{t+H-1}, y_t, y_{t+H-2}, \dots, y_{t-r+H}, \theta). \quad (3.7)$$

U (3.6) je korišćena prognozirana vrednost za \hat{y}_{t+1} , umesto njene tačne vrednosti, koja u datom trenutku nije na raspolaganju. Za prognozu svih H koraka, vrednosti od \hat{y}_{t+2} do \hat{y}_{t+H} se prognoziraju rekurzivno, i sa svakim korakom, broj prognoziranih vrednosti u regresoru raste. U trenutku kada veličina horizonta prognoze H postane jednaka ili veća od veličine regresora r , u ulazima modela se nalaze samo prognozirane vrednosti iz prethodnih koraka.

Prednost rekurzivne strategije se ogleda u tome što je potrebno formiranje samo jednog predikcionog modela, tj. potrebno je optimizovati samo jedan skup parametara za prognozu svih koraka unutar horizonta H . Nedostaci se ogledaju u akumulaciji i propagaciji grešaka kroz korake prognoze, što je naročito izraženo sa porastom veličine horizonta prognoze. U slučaju kada bi predikcioni model bio idealan, ne bi dolazilo do propagacije i akumulacije grešaka kroz korake prognoze, i ona bi onda bila podjednako precizna u svim koracima (ukoliko vremenska serija ne sadrži šum). Naravno, u praksi je nemoguće formirati idealan predikcioni model, tako da akumulacija i propagacija grešaka u zavisnosti od preciznosti modela i količine šuma u podacima, utiču na rezultate prognoze. Ovo je naročito izraženo u

slučajevima kada veličina horizonta prognoze H premaši veličinu regresora r , odnosno kada regresor čine samo prognozirane vrednosti. Uprkos ovim nedostacima, rekurzivna strategija se uspešno koristi kod prognoze vremenskih serija u mnogim praktičnim problemima, u kombinaciji sa različitim metodama nadgledanog mašinskog učenja.

3.4.2 Direktna strategija

Kod direktne strategije, prognoze se u svakom koraku vrše nezavisno jedna od drugih, svaka sa zasebnim modelom i zasebnim skupom trening parametara [92]–[96], [98]:

$$\hat{y}_{t+h} = f_h(y_t, y_{t-1}, \dots, y_{t-r+1}, \theta_h), 1 \leq h \leq H. \quad (3.8)$$

Formira se H različitih modela, svaki sa zasebnim skupom parametara θ_h , od kojih ni jedan ne sadrži prognozirane vrednosti u regresoru. Kod ove strategije ne dolazi do propagacije i akumulacije grešaka kroz korake prognoze, odnosno prisutne su samo greške koje su posledice samih modela, i šuma koji sadrže podaci. Preciznosti prognoze u prethodnim ili narednim koracima nemaju uticaj na preciznost prognoze u tekućem koraku. Međutim, nezavisan trening H modela, ima za pretpostavku međusobnu nezavisnost narednih H koraka prognoze. To utiče na preciznost prognoze, pošto ignoriše postojanje stohastičkih zavisnosti koje postoje među njima. Direktna strategija povećava složenost izračunavanja u odnosu na rekurzivnu strategiju, i u određenim situacijama postiže bolje rezultate prognoze. Najčešće se koristi u kombinaciji sa izborom atributa, gde se za svaki od H modela na osnovu odgovarajućeg algoritma pretrage i kriterijuma izbora, bira podskup atributa inicijalnog trening skupa sa kojim se može povećati kvalitet prognoze i smanjiti vreme potrebno za trening modela. Ukoliko se koristi bez izbora atributa, svaki od H modela je treniran sa istim skupom atributa. Različite metode nadgledanog mašinskog učenja su korišćene u kombinaciji sa direktnom strategijom, npr. veštačke neuronske mreže [94], kNN aproksimator [92] i stabla odlučivanja [99].

3.4.3 Direktno-rekurzivna strategija

Ukoliko se kombinuju rekurzivna i direktna strategija dobija se direktno-rekurzivna strategija [100]. Kod ove strategije se formira novi model u svakom koraku prognoze ali se modelu pridodaje i prognoza iz prethodnih koraka:

$$\begin{aligned}
\hat{y}_{t+1} &= f_1(y_t, y_{t-1}, \dots, y_{t-r+1}, \theta_1) \\
\hat{y}_{t+2} &= f_2(\hat{y}_{t+1}, y_t, y_{t-1}, \dots, y_{t-r+1}, \theta_2) \\
&\vdots \\
\hat{y}_{t+H} &= f_H(\hat{y}_{t+H-1}, \hat{y}_{t+H-2}, \dots, y_t, y_{t-1}, y_{t-r+1}, \theta_H).
\end{aligned}
\tag{3.9}$$

U svakom koraku prognoze se formira novi predikcioni model što je odlika direktne strategije. Sa druge strane, prognozirana vrednost iz prethodnog koraka se pridružuje modelu umesto tačne što je odlika rekurzivne strategije. Veličina regresora raste kroz korake prognoze kao što je prikazano u (3.9). Što je model precizniji, pridodate vrednosti prognoze iz prethodnih koraka će doprineti još većoj preciznosti modela. Ukoliko veličina regresora nije konstantna već raste u svakom koraku prognoze, tada složenost modela raste linearno. Takođe ova strategija se može koristiti u kombinaciji sa izborom atributa kada se na osnovu izbora pridodate prognoze potvrđuje opravdanost uključivanja prognoze iz prethodnog koraka u model. Ukoliko to nije slučaj, tada bi uključivanje prognoze iz prethodnog koraka najverovatnije negativno uticala na preciznost modela u narednim koracima. U [100] je ova strategija uspešno korišćena u kombinaciji sa kNN aproksimatorom i izborom atributa, gde je prevazišla rekurzivnu i direktnu strategiju na podacima električne potrošnje Poljske.

4 TEORIJSKE OSNOVE

U ovom poglavlju su predstavljene metode koje su korišćene u disertaciji za formiranje STLF modela. Detaljno je prezentovana metoda potpornih vektora sa najmanjim kvadratima (LS-SVM) koja predstavlja osnovu svih predloženih STLF modele. Takođe je data definicija zajedničkih informacija (*eng. mutual information, MI*) koja je korišćena kod predloženog modela sa izborom atributa. Pored matematičkih formulacija ove dve metode koje čine osnovu predloženih modela, predstavljene se dve statističke metode za prognozu vremenskih serija i njihove modifikacije za primenu kod STLF: metoda eksponencijalnog izravnjanja i autoregresionih integrisanih procesa pokretnih proseka - ARIMA. Pored ove dve statističke metode za kvalitativno poređenje korišćene su i najpoznatija i najrasprostranjenija metoda nadgledanog mašinskog učenja, metoda veštačkih neuronskih mreža, i „relevance vector machines“ metoda, čije su formulacije takođe date u ovom poglavlju.

4.1 Eksponencijalno izravnjanje

Vremensku seriju X_t možemo prikazati modelom u kome posmatrana pojava iskazuje slučajne fluktuacije oko prosečnog nivoa μ_t :

$$X_t = \mu_t + \varepsilon_t, \quad (4.1)$$

gde je $E(\varepsilon_t) = 0$ i $Var(\varepsilon_t) = \sigma^2$. U bilo kom vremenskom intervalu nivo serije je konstantan, ali postoji mogućnost spore promene tokom vremena. Za procenu nivoa može se koristiti

ponderisana sredina koja veći ponder (težinu) daje novijim podacima. Kada se koriste eksponencijalno opadajući ponderi, za procenu nivoa μ_t dobija se:

$$\begin{aligned}\mu_t &= \alpha X_t + \alpha(1-\alpha)X_{t-1} + \alpha(1-\alpha)^2 X_{t-2} + \dots = \alpha \sum_{j=0}^{\infty} (1-\alpha)^j X_{t-j} = \\ &= \alpha \left[1 + (1-\alpha)B + (1-\alpha)^2 B^2 + \dots + (1-\alpha)^j B^j + \dots \right] X_t.\end{aligned}\quad (4.2)$$

Kako izraz u srednjoj zagradi predstavlja zbir članova opadajuće geometrijske progresije on iznosi $[1 - (1-\alpha)B]^{-1}$, pa se izraz (4.2) može napisati u obliku:

$$[1 - (1-\alpha)B]^{-1} \mu_t = \alpha X_t, \quad (4.3)$$

odnosno u obliku:

$$\mu_t = \alpha X_t + (1-\alpha)\mu_{t-1}, \quad (4.4)$$

izražavajući time trenutnu vrednost nivoa μ_t kao linearnu kombinaciju, odnosno kao ponderisani prosek tekuće vrednosti vremenske serije X_t i prethodne vrednosti nivoa μ_{t-1} . Ponder α naziva se konstanta izravnjanja a postupak rekurzivnog određivanja trenutne vrednosti nivoa koja je definisana sa (4.4) naziva se jednostavno ekponencijalno izravnjanje (*eng. simple exponential smothing*). Postoji i alternativni prikaz jednostavnog ekponencijalnog poravnjanja u obliku korekcije greške (*eng. error correction form*) prema kome se trenutna vrednost nivoa dobija modifikacijom prethodne vrednosti nivoa delom tekuće greške $e_t = X_t - \mu_{t-1}$ koja je nastala korišćenjem ocene nivoa iz prethodnog perioda kao prognoze u tekućem periodu: $\mu_t = \mu_{t-1} - \alpha(X_t - \mu_{t-1}) = \mu_{t-1} + \alpha e_t$.

Kod implementacije modela potrebno je izabrati konstantu izravnjanja i početnu vrednost μ_0 . Na osnovu izraza (4.4) očigledno je, da što je manja vrednost konstante izravnjanja to će ocena nivoa sporije reagovati na promenu lokalnog nivoa vremenske serije a veća vrednost konstante izravnjanja uzrokuje da izravnata vrednost reaguje brzo i to ne samo na stvarne promene nivoa nego i na slučajne fluktuacije. U literaturi se obično pretpostavlja da konstanta izravnjanja uzima vrednosti iz skupa $\alpha \in (0,1)$, a najčešće u intervalu od 0.01 do 0.3. Međutim tu treba biti oprezan, i preporučuje se da se konstanta ocenjuje na osnovu karakteristike vremenske serije. Kada je reč o određivanju početne vrednosti μ_0 , može se primeniti nekoliko pristupa:

- za početnu vrednost se uzme vrednost prve opservacije same vremenske serije ili nula,
- za početnu vrednost se uzme srednja vrednost vremenske serije,
- vremenska serija se podeli na dva dela, pa se srednja vrednost prvog dela uzme za μ_0 a drugi deo serije se koristi za određivanje optimalne vrednosti konstante izravnjanja.

Može se primetiti da model dat sa (4.4) ne uvažava sezonsku tj. periodičnu komponentu u vremenskoj seriji. Ukoliko se u model uključi i sezonska komponenta I_t , čija je dužina perioda s , tada aditivni model ima oblik dat sa (4.5) a multiplikativni sa (4.6). U ovim modelima je dodata jednačina kojom se vrši revizija sezonske komponente korišćenjem nove konstante izravnjanja δ .

$$\begin{aligned}\mu_t &= \alpha(X_t - I_{t-s}) + (1-\alpha)\mu_{t-1} \\ I_t &= \delta(X_t - \mu_t) + (1-\delta)I_{t-s} \\ \hat{X}_t(h) &= \mu_t + I_{t-s+h}\end{aligned}\tag{4.5}$$

$$\begin{aligned}\mu_t &= \alpha(X_t / I_{t-s}) + (1-\alpha)\mu_{t-1} \\ I_t &= \delta(X_t / \mu_t) + (1-\delta)I_{t-s} \\ \hat{X}_t(h) &= \mu_t \cdot I_{t-s+h}\end{aligned}\tag{4.6}$$

Spomenuti modeli eksponencijalnog izravnjanja se odnose na vremensku seriju oblika (4.1), odnosno za konstantan proces sa slučajnim fluktuacijama sa ili bez sezonske komponente. Ukoliko je vremenska serija sa trendom, tada je za seriju primereniji model linearnog trenda:

$$X_t = \mu_t + T_t + \varepsilon_t.\tag{4.7}$$

Kako primena jednostavnog eksponencijalnog izravnjanja na seriju sa linearnim trendom dovodi do toga da izravnata statistika μ_t kasni iza pravog signala za navedeni iznos pristrasnosti, jedan od pristupa koji eliminiše spomenutu pristrasnost svodi se na ponovnu primenu postupka eksponencijalnog izravnjanja na izlazni rezultat [101]. Ovaj postupak se naziva dvostruko ekponencijalno izravnjanje (*eng. double exponential smoothing*) ili Brownov postupak izravnjanja. Brownov metod je proistekao iz korišćenja metoda ponderisanih najmanjih kvadrata (ponder je $\omega = 1 - \alpha$) i predstavlja optimalan metod izravnjanja za procese

sa linearnim trendom. Rekurzivni oblik Brownovog metoda za nesezonsku vremensku seriju dat je sa (4.8).

$$\begin{aligned}\mu_t &= \alpha X_t + (1-\alpha)\mu_{t-1} \\ T_t &= \alpha(\mu_t - \mu_{t-1}) + (1-\alpha)T_{t-1} \\ \hat{X}_t(h) &= \mu_t + \left(\frac{1-\alpha}{\alpha}\right)T_t + hT_t\end{aligned}\tag{4.8}$$

Korišćenje iste konstante izravnjanja za reviziju nivoa i trenda serije je prednost ali i mana Brownovog metoda. Za razliku od Brownove metode, Holt-Wintersov metod [102], [103] ostavlja mogućnost da konstante izravnjanja kod nivoa i trenda serije budu jednake, ali se ovo ograničenje ne nameće. Holt-Wintersov model za nesezonske serije dat je sa (4.9) dok je multiplikativni sezonski model dat sa (4.10).

$$\begin{aligned}\mu_t &= \alpha X_t + (1-\alpha)(\mu_{t-1} + T_{t-1}) \\ T_t &= \gamma(\mu_t - \mu_{t-1}) + (1-\gamma)T_{t-1} \\ \hat{X}_t(h) &= \mu_t + hT_t\end{aligned}\tag{4.9}$$

$$\begin{aligned}\mu_t &= \alpha(X_t / I_{t-s}) + (1-\alpha)(\mu_{t-1} + T_{t-1}) \\ T_t &= \gamma(\mu_t - \mu_{t-1}) + (1-\gamma)T_{t-1} \\ I_t &= \delta(X_t / \mu_t) + (1-\delta)I_{t-s} \\ \hat{X}_t(h) &= (\mu_t + hT_t) \cdot I_{t-s+h}\end{aligned}\tag{4.10}$$

U pogledu izbora konstanti izravnjanja kod Holt-Wintersovog modela, situacija je slična kao kod jednostavnog modela eksponencijalnog izravnjanja. U slučaju apriornog izbora sugerišu se vrednosti od 0.01 do 0.3 za sve tri konstante (nivo, trend i sezonu). Međutim, kao što je spomenuto, ovakav izbor konstanti ne uvažava specifičnost vremenske serije pa se preporučuje izbor konstanti nekom optimizacionom metodom koristeći srednje kvadratnu grešku kao optimizacioni kriterijum. Za metode određivanja početnih vrednosti videti u [104].

4.2 Autoregresioni integrisani proces pokretnih proseka - ARIMA

Prema Waldovoj teoremi razlaganja [105] svaki stacionarni proces se može predstaviti kao zbir dva međusobno nekorelisana procesa, jednog čisto determinističkog¹ i jednog čisto

¹ Pod čisto determinističkim procesom se podrazumeva proces kod koga se postiže perfektno predviđanje budućih vrednosti na osnovu sopstvenih prošlih vrednosti.

nedeterminističkog. Prema tome se čisto nedeterministički proces može izraziti u sledećoj formi:

$$X_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \psi_0 = 1, \quad (4.11)$$

gde je $\{\varepsilon_t : t = 0, \pm 1, \pm 2, \dots\}$ niz nekoreliranih slučajnih promenljivih. Tako je prema Woldovoj teoremi reprezentacija procesa data u formi (4.11) dovoljno opšta da obuhvati sve slabo stacionarne stohastičke procese. Proces definisan izrazom (4.11) naziva se linearni proces. Uvođenjem lag² operatora (*eng. lag*) za koji važi $BX_t = X_{t-1}$ tj. $B^m X_t = X_{t-m}$, linearni process se može napisati u obliku:

$$X_t - \mu = (1 - \psi_1 B + \psi_2 B^2 + \dots) \varepsilon_t = \psi(B) \varepsilon_t, \quad (4.12)$$

gde je sa $\psi(B)$ označen polinom po lag operatoru, pri čemu je $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$. Linearni proces se definiše na osnovu stacionarnog stohastičkog procesa $\{\varepsilon_t\}$ koji se naziva procesom belog šuma ili potpuno slučajan proces.

Izraz (4.11) je ponderisani prosek slučajnih poremećaja ili šokova i predstavlja MA reprezentaciju linearnog procesa. Pored ove postoji i autoregresiona reprezentacija (AR), prema kojoj se tekuća vrednost procesa regresira na svoje prošle vrednosti, a uključen je i slučajni poremećaj ε_t :

$$X_t = \pi_1 X_{t-1} + \pi_2 X_{t-2} + \dots + \varepsilon_t, \quad (4.13)$$

što se korišćenjem lag operatora može napisati u obliku:

$$\pi(B) X_t = \varepsilon_t, \quad (4.14)$$

gde je $\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$, i $1 + \sum_{j=1}^{\infty} |\pi_j| < \infty$ pri tom koeficijente u AR reprezentaciji nazivamo π -ponderi. Ako se linearni proces dat izrazom (4.11) može napisati u formi AR reprezentacije (4.14), tada za proces kažemo da je invertibilan.

Treća klasa modela opisuje tzv. autoregresione procese pokretnih proseka, odnosno reč je o mešovitom modelu koji glasi:

² Ili operatora docnje kako je poznat u literaturi kod nas.

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}. \quad (4.15)$$

U slučaju da AR i MA modeli za konkretnu vremensku seriju sadrže veliki broj koeficijenata tada je mešoviti model njihova realna alternativa. Po pravilu se na osnovu njega sa manjim brojem koeficijenata u odnosu na AR i MA modele postiže približno ista aproksimacija posmatranog procesa.

Ako se u AR reprezentaciju linearnog procesa (4.11) stavi da je $\pi_1 = \phi_1, \pi_2 = \phi_2, \dots, \pi_p = \phi_p$ i $\pi_j = 0$ za $j > p$ dobija se autoregresioni proces reda p . Autoregresioni procesi impliciraju regresiju na sopstvene vrednosti pa otuda i prefiks auto u imenu ovog procesa. Za proces X_t se kaže da je autoregresioni proces reda p ako je opisan jednačinom

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t. \quad (4.16)$$

Znači da je tekuća vrednost procesa linearna kombinacija p sopstvenih prošlih vrednosti plus slučajni poremećaj ε_t . Kako je $\sum_{j=1}^{\infty} \pi_j = \sum_{j=1}^p |\phi_j| < \infty$, proces je uvek invertibilan. Autoregresioni proces (4.16) može se napisati i u sažetijem obliku korišćenjem lag operatora:

$$\phi(B) X_t = \varepsilon_t, \quad (4.17)$$

gde je $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$. Autoregresioni proces reda p se označava kao AR(p).

Za proces X_t se kaže da predstavlja proces pokretnih proseka reda q ako je

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \quad (4.18)$$

koji se korišćenjem lag operatora može napisati u obliku

$$X_t = \theta(B) \varepsilon_t, \quad (4.19)$$

gde je $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$. Proces pokretnih proseka je koristan u modeliranju pojava kod kojih događaji uzrokuju trenutne efekte a koji traju kratak period vremena. Ako u linearnom procesu (4.11) ψ -ponderi uzmu sledeće vrednosti:

$\psi_1 = -\theta_1, \psi_2 = -\theta_2, \dots, \psi_q = -\theta_q, \psi_j = 0$, za $j > q$ dobija se process pokretnih proseka (4.18). Varijansa ovog procesa je $\sigma^2(1 + \theta_1^2 + \dots + \theta_q^2) < \infty$ pa se može zaključiti da će proces pokretnih proseka uvek biti stacionaran. Ovaj proces se označava sa $MA(q)$.

Ako se polinom $\psi(B)$ predstavi kao količnik dva konačna polinoma po lag operatoru B :

$$\psi(B) = \frac{\theta(B)}{\phi(B)} = \frac{1 - \theta_1 B - \dots - \theta_q B^q}{1 - \phi_1 B - \dots - \phi_p B^p}, \quad (4.20)$$

tada se $ARMA(p, q)$ proces definiše isto kao (4.11), tj.

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad (4.21)$$

koji se preko lag operatora može napisati u obliku

$$(1 - \phi_1 B - \dots - \phi_p B^p) X_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t, \quad (4.22)$$

ili u obliku $\phi(B) X_t = \theta(B) \varepsilon_t$.

Predstavljeni modeli do sada su se bavili klasom stacionarnih stohastičkih procesa čije su osnovne karakteristike konstantnost sredine (nivoa serije), konstantnost varijanse i zavisnost autokovarijansi samo od vremenskog intervala. Međutim većina vremenskih serija odstupa od klase stacionarnih serija, već se radi o stohastičkim procesima i vremeskim serijama sa vremenski zavisnim nivoom i/ili varijansom, pa ove procese nazivamo nestacionarnim stohastičkim procesima i serijama. Ovi procesi tj. serije se mogu modelovati na dva načina. Prvi, klasičan pristup je da se sredina nestacionarnog procesa predstavi determinističkom funkcijom vremena. Drugi pristup zasnovan je na postupku diferenciranja vremenske serije kada se dobija stacionarna serija što potvrđuje da prvobitna serija sadrži stohastički trend.

Drugi način modeliranja nestacionarne serije u sredini X_t zasnovana je na pretpostavci da će njena diferencija reda d biti stacionarna. Klasa procesa sa ovim svojstvom naziva se diferencno stacionarni procesi. Tako proces dat izrazom (4.23)

$$X_t = X_{t-1} + \beta_1 + \varepsilon_t, \quad (4.23)$$

predstavlja jedan primer klase nestacionarnih procesa poznate pod nazivom integrisani procesi, kod kojih se višestrukim diferenciranjem postiže stacionarnost, pri tome diferencirana serija može biti autokorelisana. Tako se kod procesa datog izrazom (4.23) prvim diferenciranjem serije dobija stacionaran proces, odnosno proces belog šuma. Ako se u

opštem slučaju diferencirana serija $(1-B)^d X_t$ modelira stacionarnim ARMA(p,q) procesom, tada se dobija:

$$\phi_p(B)(1-B)^d X_t = \theta_0 + \theta_q(B)\varepsilon_t, \quad (4.24)$$

gde su $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ i $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ AR i MA polinomi po lag operatoru, reda p i q .

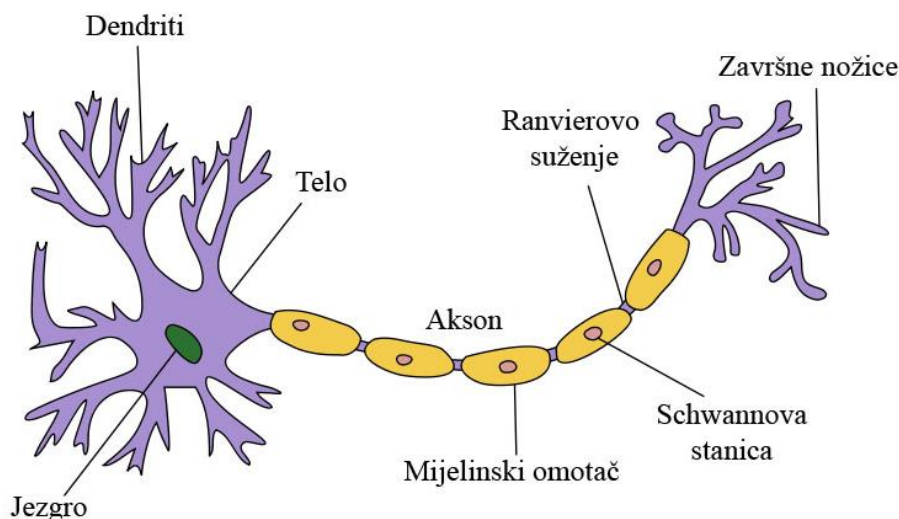
Za razliku od ARMA(p,q) modela u izrazu (4.24) uvedena je konstanta θ_0 koja zavisi od reda diferenciranja. Ako je $d=0$ proces je stacionaran, a konstanta je u relaciji sa sredinom procesa $\theta_0 = \mu(1 - \phi_1 - \dots - \phi_p)$, gde je $\mu = E(X_t)$. Za vrednosti $d>0$, konstanta θ_0 ukazuje na prisustvo determinističkog trenda u seriji. Konstanta se izostavlja iz izraza (4.24) kada nema mesta pretpostavci da će diferencirana serija sadržati deterministički trend.

Homogeno nestacionarni proces dat izrazom (4.24) naziva se autoregresioni integrisani proces pokretnih proseka (*eng. autoregressive integrated moving average*) reda p , d i q ili prema skraćenoj notaciji ARIMA(p,d,q) proces. Takođe, za proces X_t se kaže da je integrisan reda d u oznaci $I(d)$.

4.3 Veštačke neuronske mreže

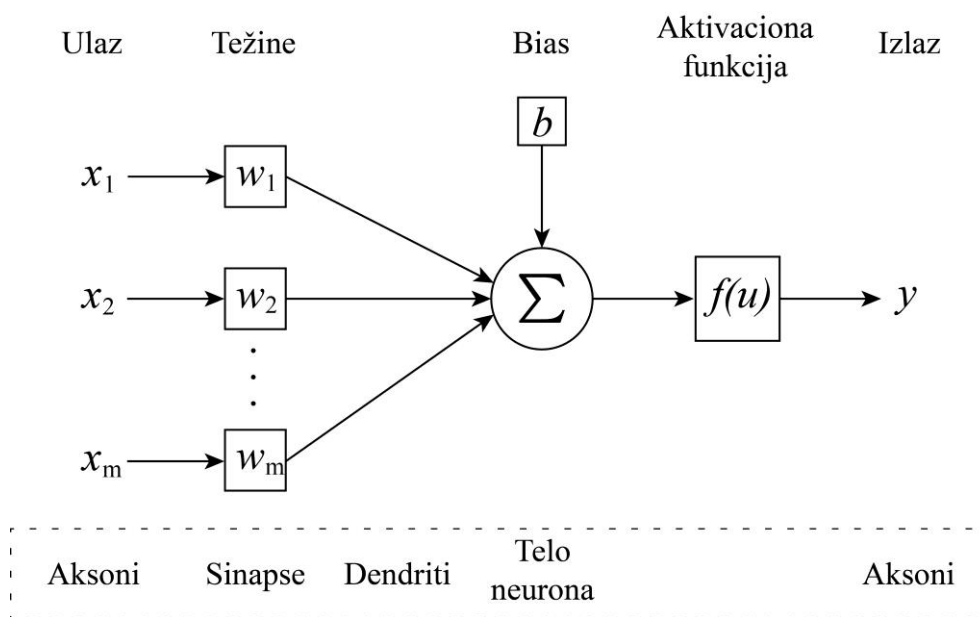
Pojava teorije veštačkih neuronskih mreža se nesumljivo može pripisati Mek Kulohu i Pitu koji su 1943. sprovedli eksperiment modeliranja neuro-sistema pomoću mreže jednostavnih logičkih operacija [106]. Njihova ideja je bila da se pronade jednostavan parametarski nelinearni model za biološki neuron.

Struktura tipičnog biološkog neurona prikazana je na Slici 4.1. Biološki neuron prima nadražaje preko dendrita kroz sinapse. Dendriti su kratki, mnogobrojni produžeci citoplazme koji polaze sa tela nervne ćelije i granaju se, čime se povećava površina kojom primaju signale. Imaju funkciju primanja nadražaja koji dolaze od drugih neurona i njihovom provođenju ka telu nervne ćelije. Kada je ovaj nadražaj dovoljno jak i prelazi određeni prag, neuron se aktivira i šalje nadražaj kroz akson. Akson ili neurit je cilindrični nastavak koji prenosi nadražaj od tela nervne ćelije, tj. akson nadražaj odvodi od tela neurona ka sledećem neuronu. Proces učenja se tako svodi na podešavanje veza između neurona (sinapsi) pa će određeni nadražaj da aktivira samo određene neurone.



Slika 4.1 Struktura tipičnog biološkog neurona [107]

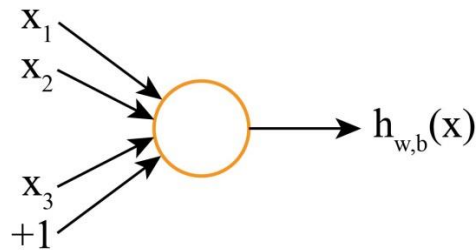
Perceptron, osnovni neuron predstavio je 1957. Rozenblat koji je pokušavao da razume proces učenja i memorisanja kod ljudi nakon čega je predstavio i prvu mašinu koja može da uči da prepoznaje optičke obrasce, početkom 1960. Matematički model perceptrona odnosno veštačkog neurona je formiran na osnovu arhitekture biološkog neurona. Aksoni i sinapse biološkog neurona su modelovani ulazima i težinama. Jačina veze između ulaza i neurona je određena odgovarajućom težinom. Ponderisani ulazi se sabiraju i propuštaju kroz nelinearnu aktivacionu funkciju, koja kontroliše amplitudu izlaza neurona. Struktura modela veštačkog neurona prikazana je na Slici 4.2.



Odgovarajuća struktura kod biološkog neurona

Slika 4.2 Struktura veštačkog neurona

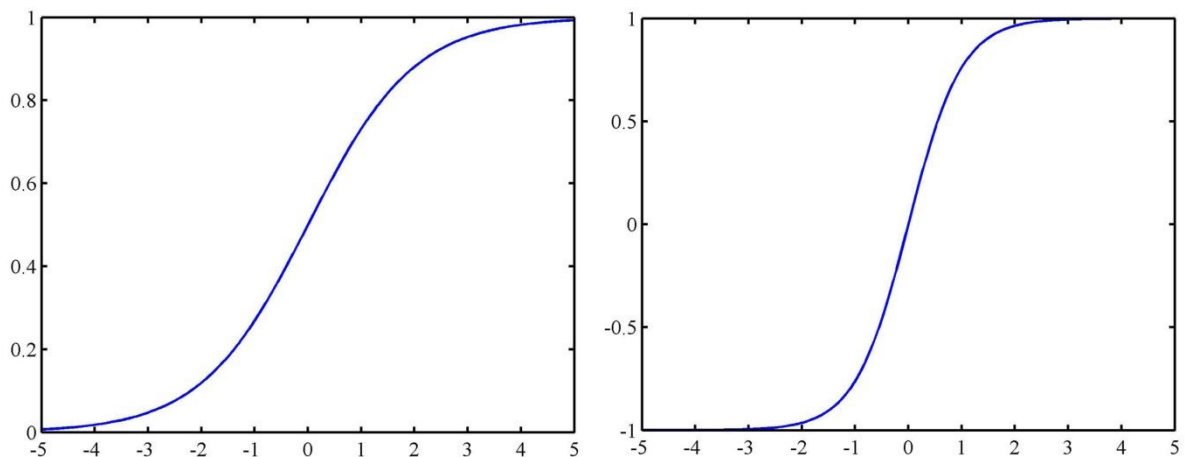
Najjednostavniji oblik veštačke neuronske mreže jeste meža koju čini samo jedan neuron, koja se može predstaviti dijagramom kao na Slici 4.3, pri čemu je usvojeno da neuron ima tri ulaza ($m=3$) zbog lakšeg praćenja.



Slika 4.3 Dijagram neurona

Ovakva neuronska mreža koju čini jedan neuron ima tri ulaza x_1, x_2, x_3 i izlaz $h_{w,b}(x) = f(W^T x) = f(\sum_{i=1}^3 W_i x_i + b)$, gde se $f: R \rightarrow R$ naziva aktivaciona funkcija. Ako se usvoji da je $f(\cdot)$ sigmoidalna logistička funkcija onda ona ima oblik $f(z) = \frac{1}{1 + \exp(-z)}$, a dati neuron vrši mapiranje ulaza na izlaz kao u slučaju logističke regresije. Međutim, pored logističke funkcije mogu se koristiti i funkcija tipa tangensa hiperboličkog

$$f(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \text{ ili funkcija tipa znaka } f(z) = \text{sgn}(z) = \begin{cases} 1, & z > 0 \\ 0, & z = 0 \\ -1, & z < 0 \end{cases}$$

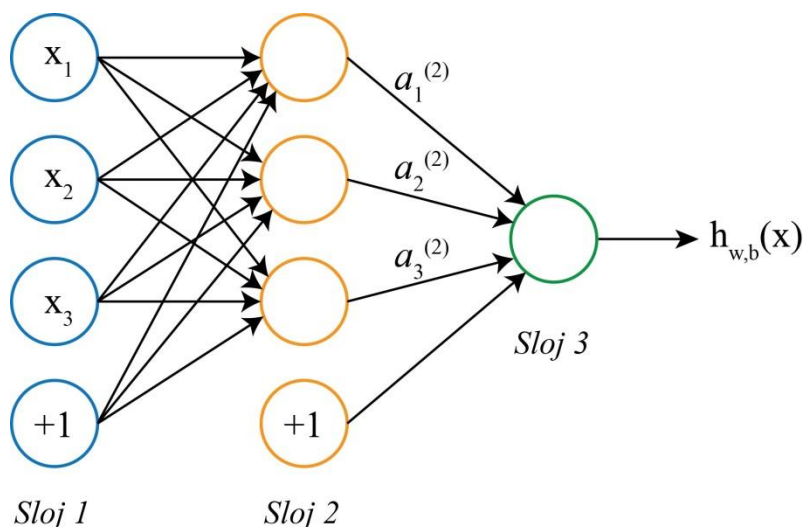


a) Logistička funkcija

b) Tangens hiperbolička funkcija

Slika 4.4 Aktivacione funkcije

Ukoliko se međusobno poveže više neurona tako da izlaz prethodnog bude ulaz narednog, mogu se dobiti različite strukture neuronske mreže sa više slojeva. Na primer, uzmimo u razmatranje jednoslojnu neuronsku mrežu prikazanu na Slici 4.5.



Slika 4.5 Dijagram jednoslojne neuronske mreže

Sloj 1 koji je najdalji levo se naziva ulazni sloj a najdalji sloj desno (sloj 3), izlazni sloj. Srednji sloj (sloj 2) se naziva skriveni sloj. Tako se za konkretnu mrežu kaže da ima 3 sloja sa 3 ulazna čvora, 3 čvora u skrivenom sloju i jednim u izlaznom. Ako sa n_l označimo broj slojeva u mreži, tada će u našem slučaju biti $n_l = 3$. Kako je u opštem slučaju L_l oznaka za sloj l , tako je L_1 ulazni sloj a L_{n_l} izlazni sloj. Parametri mreže su $(W, b) = (W^{(1)}, b^{(1)}, W^{(2)}, b^{(2)})$ gde $W_{ij}^{(l)}$ označava parametar (težinu) dodeljenu vezi između čvora j u sloju l i čvora i u sloju $l+1$. Takođe, $b_i^{(l)}$ je bias koji se odnosi na čvor i u sloju $l+1$. Da se primetiti da bias jedinica nema ulaze i uvek daje na izlazu +1. Broj čvorova u sloju l je s_l , ne računajući bias jedinice.

Označimo izlaz čvora i u sloju l sa $a_i^{(l)}$. Za $l=1$ važi $a_i^{(1)} = x_i$ što predstavlja i -ti ulaz. Parametri mreže W, b definišu funkciju hipoteze $h_{w,b}(x)$. Tako je matematički model predstavljene mreže dat sa:

$$a_1^{(2)} = f(W_{11}^{(1)}x_1 + W_{12}^{(1)}x_2 + W_{13}^{(1)}x_3 + b_1^{(1)}), \quad (4.25)$$

$$a_2^{(2)} = f(W_{21}^{(1)}x_1 + W_{22}^{(1)}x_2 + W_{23}^{(1)}x_3 + b_2^{(1)}), \quad (4.26)$$

$$a_3^{(2)} = f(W_{31}^{(1)}x_1 + W_{32}^{(1)}x_2 + W_{33}^{(1)}x_3 + b_3^{(1)}), \quad (4.27)$$

$$h_{w,b}(x) = a_1^{(3)} = f(W_{11}^{(2)}a_1^{(2)} + W_{12}^{(2)}a_2^{(2)} + W_{13}^{(2)}a_3^{(2)} + b_1^{(2)}). \quad (4.28)$$

Ako ponderisanu sumu ulaza čvora i u sloju l označimo sa $z_i^{(l)}$ (npr. $z_i^{(2)} = \sum_{j=1}^n W_{ij}^{(1)}x_j + b_i^{(1)}$), tada je $a_i^{(l)} = f(z_i^{(l)})$. Jednačine (4.25 - 4.28) se tako mogu kompaktnije napisati u vektorskom obliku:

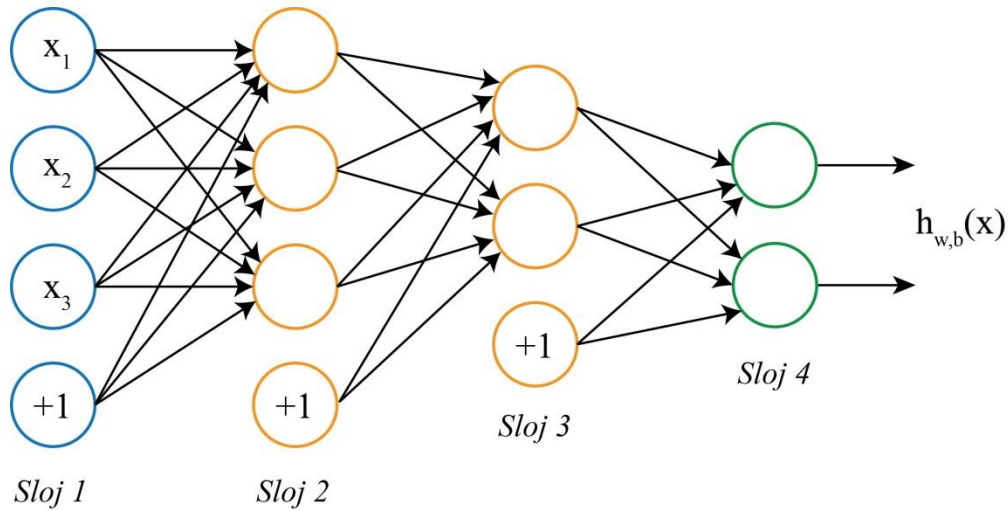
$$\begin{aligned} z^{(2)} &= W^{(1)}x + b^{(1)} \\ a^{(2)} &= f(z^{(2)}) \\ z^{(3)} &= W^{(2)}a^{(2)} + b^{(2)} \\ h_{w,b}(x) &= a^{(3)} = f(z^{(3)}) \end{aligned}$$

Kako je $a^{(1)} = x$ što predstavlja ulaze, tako za poznate aktivacije $a^{(l)}$ u sloju l se mogu odrediti aktivacije $a^{(l+1)}$ sloja $l+1$:

$$z^{(l+1)} = W^{(l)}a^{(l)} + b^{(l)}, \quad (4.29)$$

$$a^{(l+1)} = f(z^{(l+1)}). \quad (4.30)$$

Mreža data na Slici 4.5 je jednoslojna neuronska mreža sa jednim izlazom. Međutim, neuronske mreže mogu da imaju i više skrivenih slojeva i više izlaza, kao što je slučaj sa mrežom na Slici 4.6 koja ima dva skrivena sloja i dva izlaza. Kod ovakvih mreža sa n_l slojeva, gde je prvi sloj ulazni a n_l -ti izlazni i svaki sloj l je povezan sa sledećim $l+1$ slojem, izlaz mreže se dobija sukcesivnim proračunom aktivacija u sloju L_2 , L_3 i tako dalje do sloja L_{n_l} koristeći jednačine (4.29 i 4.30). Ovakve mreže se nazivaju višeslojne neuronske mreže sa prostiranjem unapred (*eng. feedforward multilayer neural networks*) i one nemaju nikakve zatvorene petlje niti povratne sprege.



Slika 4.6 Dijagram višeslojne neuronske mreže sa više izlaza

Funkcija hipoteze $h_{w,b}(x)$ definisana je parametrima mreže W, b koje treba odrediti. Proces određivanja ovih parametara naziva se obučavanje ili treniranje neuronske mreže. Podimo od trening skupa $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$ koji ima m trening primera. Za jedan

trening primer (x, y) može se napisati kriterijumska funkcija $J(W, b; x, y) = \frac{1}{2} \|h_{w,b}(x) - y\|^2$.

Ako se uvažavaju svi trening primeri, tada kriterijumska funkcija ima oblik

$$\begin{aligned}
 J(W, b) &= \left[\frac{1}{m} \sum_{i=1}^m J(W, b; x^{(i)}, y^{(i)}) \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \\
 &= \left[\frac{1}{m} \sum_{i=1}^m \frac{1}{2} \|h_{w,b}(x^{(i)}) - y^{(i)}\|^2 \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2
 \end{aligned} \tag{4.31}$$

Prvi član u jednačini (4.31) je usrednjena ukupna kvadratna greška a drugi je regularizacioni član (*eng. regularization or weight decay term*) koji ima za cilj korekcije težina i pomoć u prevenciji prekoračenja (*eng. overfitting*). Regularizacioni parametar λ kontroliše relativni uticaj ova dva člana.

Kriterijumska funkcija (4.31) se koristi i kod klasifikacionih i regresionih problema. Kod klasifikacije je $y=0$ ili 1 što predstavlja dve klase (jer su izlazi sigmoidalne aktivacione funkcije $[0,1]$ a ako se koristi tangens hiperbolička funkcija onda su izlazi -1 i $+1$). U slučaju regresije, izlazi se skaliraju da budu u opsegu $[0,1]$ ili ako se koristi tangens hiperbolička funkcija, onda u opsegu $[-1,1]$.

Cilj je minimizacija kriterijumske funkcije $J(W, b)$ koja je funkcija po parametrima W i b . Obučavanje mreže počinje inicijalizacijom parametara $W_{ij}^{(l)}$ i $b_i^{(l)}$ na vrlo male vrednosti bliske nuli ($N(0, \varepsilon^2)$, $\varepsilon \approx 0.01$) a zatim se primeni neki optimizacioni algoritam kao što je metoda gradijentnog spusta (*eng. gradient descent*). Ovde treba biti pažljiv jer funkcija $J(W, b)$ nije konveksna pa se tako metoda gradijentnog spusta može zaustaviti u lokalni minimum i da ne dođe do globalnog optimuma. Međutim u praksi i tako jednostavan optimizacioni algoritam radi prilično dobro. Takođe inicijalizaciju početnih vrednosti parametara $W_{ij}^{(l)}$ i $b_i^{(l)}$ treba uraditi nasumičnim dodeljivanjem vrednosti a ne stavljanjem svih vrednosti parametara na nulu ili neku drugu vrednost. Ukoliko bi se to uradilo i svi parametri imali istu početnu vrednost, tada bi težine $W_{ij}^{(l)}$ bile iste za svako i , odnosno $a_1^{(2)} = a_2^{(2)} = a_3^{(2)} = \dots$ za svaki ulaz x .

Jedna iteracija metode gradijentnog spusta procesa određivanja parametara W, b data je sa:

$$W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b), \quad (4.32)$$

$$b_i^{(l)} = b_i^{(l)} - \alpha \frac{\partial}{\partial b_i^{(l)}} J(W, b), \quad (4.33)$$

gde je α koeficijent obučavanja. Ovde je ključna stvar izračunavanje parcijalnih diferencijala

$\frac{\partial}{\partial W_{ij}^{(l)}} J(W, b)$ i $\frac{\partial}{\partial b_i^{(l)}} J(W, b)$, što rešava algoritam propagacije greške unazad (*eng. backpropagation algorithm*).

Ovaj algoritam učenja ustvari omogućava određivanje

$\frac{\partial}{\partial W_{ij}^{(l)}} J(W, b; x, y)$ i $\frac{\partial}{\partial b_i^{(l)}} J(W, b; x, y)$ tj. parcijalne diferencijale kriterijumske funkcije

$J(W, b; x, y)$ za jedan primer (x, y) . Kada se odrede ovi diferencijali za jedan primer tada se

uvažavanjem jednačine (4.31) dobijaju i diferencijali:

$$\frac{\partial}{\partial W_{ij}^{(l)}} J(W, b) = \left[\frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b; x^{(i)}, y^{(i)}) \right] + \lambda W_{ij}^{(l)}, \quad (4.34)$$

$$\frac{\partial}{\partial b_i^{(l)}} J(W, b) = \frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial b_i^{(l)}} J(W, b; x^{(i)}, y^{(i)}). \quad (4.35)$$

Za dati skup primera (x, y) prvo se radi “propagacija unapred” kako bi se izračunale sve aktivacije kroz mrežu, uključujući i izlaznu vrednost funkcije hipoteze $h_{W, b}(x)$. Zatim za svaki čvor i u sloju l , računa se “greška“ $\delta_i^{(l)}$ koja predstavlja meru koliko određeni čvor ima uticaja na ukupnu grešku izlaza. Algoritam propagacije greške je tako dat kroz sledeće korake:

1. Kroz propagaciju unapred, izračunati aktivacije za slojeve L_2, L_3 sve do izlaznog sloja L_{n_l} .
2. Za svaki izlazni čvor i u izlaznom sloju n_l , izračunati:

$$\delta_i^{(n_l)} = \frac{\partial}{\partial z_i^{(n_l)}} \frac{1}{2} \|y - h_{W, b}(x)\|^2 = -(y_i - a_i^{(n_l)}) \cdot f'(z_i^{(n_l)}).$$

3. Za $l = n_l - 1, n_l - 2, n_l - 3, \dots, 2$, za svaki čvor i u sloju l odrediti:

$$\delta_i^{(l)} = \left(\sum_{j=1}^{s_{l+1}} W_{ji}^{(l)} \delta_j^{(l+1)} \right) f'(z_i^{(l)}).$$

4. Izračunati parcijalne diferencijale:

$$\frac{\partial}{\partial W_{ij}^{(l)}} J(W, b; x, y) = a_j^{(l)} \delta_i^{(l+1)}$$

$$\frac{\partial}{\partial b_i^{(l)}} J(W, b; x, y) = \delta_i^{(l+1)}.$$

Kada su parcijalni diferencijali određeni kroz proces propagacije greške unazad, možemo se vratiti metodi gradijentnog spusta za određivanje parametra mreže $W_{ij}^{(l)}$ i $b_i^{(l)}$ što je i bio cilj. Jedna iteracija metode gradijentnog spusta se može predstaviti sledećim pseudo-kodom:

-
1. Staviti $\Delta W^{(l)} = 0$ i $\Delta b^{(l)} = 0$ za svako l .
 2. Za svako $i=1$ do m ,
 - a. odrediti $\nabla_{W^{(l)}} J(W, b; x, y)$ i $\nabla_{b^{(l)}} J(W, b; x, y)$ koristeći algoritam propagacije greške unazad,
 - b. izračunati $\Delta W^{(l)} = \Delta W^{(l)} + \nabla_{W^{(l)}} J(W, b; x, y)$,
 - c. izračunati $\Delta b^{(l)} = \Delta b^{(l)} + \nabla_{b^{(l)}} J(W, b; x, y)$.
 3. Ažurirati parametre:

$$W^{(l)} = W^{(l)} - \alpha \left[\left(\frac{1}{m} \Delta W^{(l)} \right) + \lambda W^{(l)} \right],$$

$$b^{(l)} = b^{(l)} - \alpha \left[\frac{1}{m} \Delta b^{(l)} \right].$$

Treba napomenuti da u prethodnom pseudo kodu, $\Delta W^{(l)}$ je matrica dimenzije iste kao i $W^{(l)}$, a $\Delta b^{(l)}$ je vektor, iste dimenzije kao i $b^{(l)}$.

Da se primetiti da se obučavanje mreže svodi na iterativni proces optimizacije kriterijumske funkcije $J(W, b)$. Ovde je zbog lakšeg praćenja za optimizaciju korišćen jednostavan metod gradijentnog spusta. Međutim u praksi se koriste napredniji optimizacioni algoritmi kao što je L-BFGS (*eng. limited-memory Broyden–Fletcher–Goldfarb–Shanno*) ili „conjugate gradient“ koji su znatnije brži od metode gradijentnog spusta.

4.4 Least squares support vector machines

Kako se predloženi modeli u disertaciji za STLF baziraju na LS-SVM metodi, ovde je detaljno prezentovan ovaj metod za nelinearnu regresiju kao i njegove specifičnosti u odnosu na “standardni” SVM. Analizirane su kernel funkcije i njihova uloga u metodama potpornih vektora. Razmatran je jedan od načina za rešavanje optimizacionog problema koji formira LS-SVM model, „conjugate gradient“ (CG) algoritam. Prikazana su i dva pristupa za optimizaciju LS-SVM parametara: „grid-search“ algoritam u kombinaciji sa „k-fold cross-validation“ pristupom. Na kraju poglavlja naveden je i osnovni algoritam za dobijanje sparse LS-SVM rešenja.

4.4.1 Osnovne LS-SVM formulacije

Jedno od svojstava “standardnog” SVM-a predloženog u [87] od strane Vapnik-a je to da se problemi nelinearne klasifikacije i regresije mogu rešavati metodama konveksnog kvadratnog programiranja (*eng. convex quadratic programming, QP*). Pored toga što SVM uvek pronalazi optimalno rešenje QP problema, u njemu figurišu samo primeri iz trening skupa koji mu najviše doprinose, takozvani potporni vektori (*eng. support vectors*), odnosno formira se sparsno rešenje (*eng. sparse solution*). Čak i veličina QP problema ne zavisi od dimenzije ulaznog prostora, već samo od broja trening primera. Optimizacioni problem koji definiše SVM se zasniva na strukturnoj minimizaciji rizika, tako da se minimizuje gornja granica greške generalizacije, što takođe doprinosi kvalitetu rešenja. Postavlja se pitanje za koliko se može uprostiti formulacija SVM-a bez gubitaka u kvalitetu rešenja?

U [108] predložena je jedna modifikacija SVM-a koja vodi do rešavanja sistema linearnih jednačina, koji je jednostavnije rešiti u odnosu na QP problem.

Neka je dat trening skup $\{x_k, y_k\}, k=1, \dots, N$ gde N predstavlja ukupan broj trening primera, sa ulazima $x_k \in R^n$ i izlazima $y_k \in R$. Može se formirati regresioni model u primarnom težinskom prostoru (*eng. primal weight space*) korišćenjem nelinearnog preslikavanja $\varphi(\cdot): R^n \rightarrow R^{n_h}$ koje preslikava ulazni prostor atributa (*eng. feature space*) u neki višedimenzionalni prostor, i u njemu definiše linearnu regresiju.

Regresioni model u primarnom prostoru je definisan sa:

$$y(x) = \omega^T \varphi(x) + b, \quad (4.36)$$

gde ω predstavlja vektor parametara (*eng. weight vector*), a b definiše pomeraj (*eng. bias term*).

Optimizacioni problem koji je potrebno rešiti formira se u primarnom prostoru:

$$\min_{\omega, b, e} J_p(\omega, e) = \frac{1}{2} \omega^T \omega + \frac{1}{2} \gamma \sum_{k=1}^N e_k^2, \quad (4.37)$$

uz ograničenja:

$$y_k = \omega^T \varphi(x_k) + b + e_k, k = 1, \dots, N, \quad (4.38)$$

gde su e_k dozvoljene greške prilikom formiranja regresionog modela, a γ je parametar koji im dodeljuje relativne težine.

U odnosu na standardnu formulaciju SVM-a ovde imamo dve modifikacije. Prvo, u (4.38) ograničenja nejednakosti (*eng. inequality constraints*) su zamenjena sa ograničenjima jednakosti (*eng. equality constraints*). Drugo u (4.37), usvojena je kvadratna funkcija greške (*eng. squared loss function*). Kao što će biti pokazano, ove dve modifikacije će značajno pojednostaviti rešavanje problema.

U cilju rešavanja optimizacionog problema formulisanog sa (4.37) i (4.38) potrebno je konstruisati dualni problem korišćenjem Lagranževove funkcije:

$$L(\omega, b, e; \alpha) = \frac{1}{2} \omega^T \omega + \gamma \frac{1}{2} \sum_{k=1}^N e_k^2 - \sum_{k=1}^N \alpha_k \{ \omega^T \varphi(x_k) + b + e_k - y_k \}, \quad (4.39)$$

gde su α_k Lagranževovi multiplikatori, koji sada mogu biti pozitivni ili negativni zbog ograničenja jednakosti definisanih u (4.38)). Uslovi optimalnosti definisani su sa (4.40 - 4.43):

$$\frac{\partial L}{\partial \omega} = 0 \rightarrow \omega = \sum_{k=1}^N \alpha_k \varphi(x_k), \quad (4.40)$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{k=1}^N \alpha_k = 0, \quad (4.41)$$

$$\frac{\partial L}{\partial e_k} = 0 \rightarrow \alpha_k = \gamma e_k, k = 1, \dots, N, \quad (4.42)$$

$$\frac{\partial L}{\partial \alpha_k} = 0 \rightarrow \omega^T \varphi(x_k) + b + e_k - y_k = 0, k = 1, \dots, N, \quad (4.43)$$

Iz uslova optimalnosti sledi linearni sistem jednačina:

$$\begin{bmatrix} I & 0 & 0 & -\varphi^T \\ 0 & 0 & 0 & -1_v^T \\ 0 & 0 & \gamma I & -I \\ \varphi & 1_v & I & 0 \end{bmatrix} \begin{bmatrix} \omega \\ b \\ e \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ y \end{bmatrix}, \quad (4.44)$$

gde je $\varphi = [\varphi(x_1), \dots, \varphi(x_N)]^T$, $y = [y_1, \dots, y_N]$, $e = [e_1, \dots, e_N]$, $\alpha = [\alpha_1, \dots, \alpha_N]$, $1_v = [1, \dots, 1]$, a I je jedinična matrica reda $N \times N$.

Nakon eliminacije ω i e iz (4.40 - 4.43) dobija se linearni *Karush-Kuhn-Tucker* (KKT) sistem jednačina:

$$\begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & K(x_1, x_1) + 1/\gamma & \cdots & K(x_1, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & K(x_N, x_1) & \cdots & K(x_N, x_N) + 1/\gamma \end{bmatrix} \begin{bmatrix} b \\ \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ \vdots \\ y_N \end{bmatrix}, \quad (4.45)$$

odnosno u formi blok-matrice:

$$\begin{bmatrix} 0 & \mathbf{1}_v^T \\ \mathbf{1}_v & \Omega + I\gamma^{-1} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}, \quad (4.46)$$

gde su $y = [y_1, \dots, y_N]$, $\mathbf{1}_v = [1, \dots, 1]$, $\alpha = [\alpha_1, \dots, \alpha_N]$ predstavlja Lagranžeove multiplikatore, I je jedinična matrica, a $\Omega_{kl} = \varphi(x_k)^T \varphi(x_l) = K(x_k, x_l)$, $k, l = 1, \dots, N$ označava matricu jezgra (*eng. kernel matrix*). Nakon rešavanja sistema (4.46) direktnom metodom eliminacije, dobijaju se rešenja za b i α :

$$b = \frac{\mathbf{1}_v^T (\Omega + I\gamma^{-1})^{-1} y}{\mathbf{1}_v^T (\Omega + I\gamma^{-1})^{-1} \mathbf{1}_v}, \quad (4.47)$$

$$\alpha = (\Omega + I\gamma^{-1})^{-1} (y - b\mathbf{1}_v). \quad (4.48)$$

Sada se regresioni model u dualnoj formi može predstaviti kao:

$$y(x) = \sum_{k=1}^N \alpha_k K(x, x_k) + b, \quad (4.49)$$

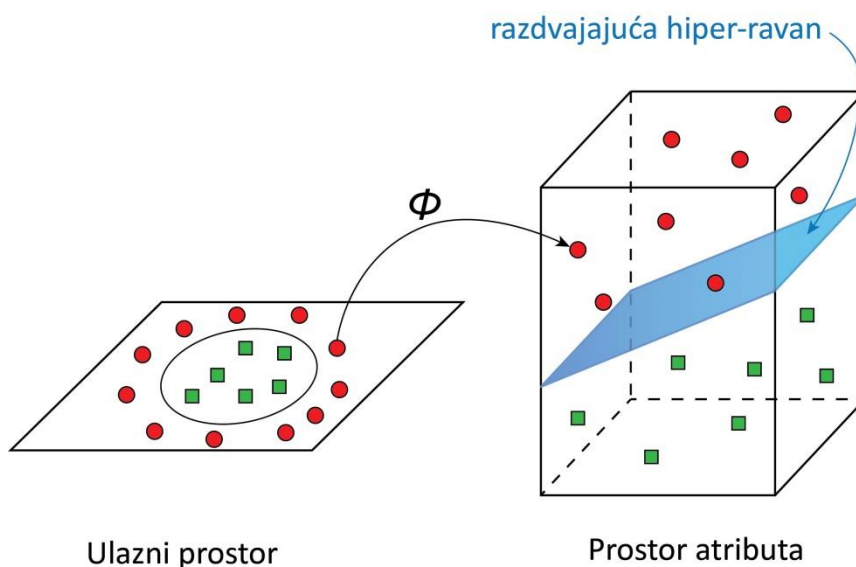
gde su α_k i b rešenja sistema (4.46). Kao i u slučaju SVM-a rešenja KKT sistema garantuju globalni optimum (s tim što je u ovom slučaju KKT sistem linearan). Rešavanje KKT sistema je fundamentalna stavka kod LS-SVM-a, kao i u nelinearnim optimizacionim problemima. U sekciji 4.4.3 je diskutovana jedna od metoda koja se često koristi u svrhu rešavanja ovog problema – „*conjugate gradient*“ algoritam. Biblioteka funkcija koja implementira LS-SVM može se preuzeti sa [109].

Nedostatak ove reformulacije SVM-a se ogleda u tome što kod LS-SVM-a svi trening primeri učestvuju u formiranju rešenja u (4.49), odnosno javlja se takozvani nedostatak sparsnosti rešenja (*eng. lack of sparseness*). To je direktna posledica izraza (4.42). U slučaju LS-SVM-a svaki trening vektor predstavlja potporni vektor (*eng. support vector*), s tim što su neki od njih značajniji od drugih, što je određeno sa $|\alpha_k|$. Tehnike za postizanje sparsnog rešenja “skraćivanjem” (*eng. sparseness by pruning*) ukratko su pomenute u sekciji 4.4.5.

Skalarni proizvod u (4.49) $K(x, x_k) = \varphi(x)^T \varphi(x_k)$ predstavlja funkciju jezgra (eng. *kernel function*). Kernel funkcije koje zadovoljavaju *Mercer-ove* uslove (eng. *Mercer's conditions*) omogućavaju izračunavanje skalarnog proizvoda u višedimenzionalnom prostoru atributa (eng. *higher dimensional feature space*) korišćenem atributa iz ulaznog prostora, bez potrebe za eksplicitnim korišćenjem preslikavanja $\varphi(x)$. U literaturi ova “prečica” je poznata pod nazivom *kernel trik* (eng. *kernel trick*), i detaljnije je diskutovana u narednoj sekciji.

4.4.2 Kernel funkcije

Najznačajniji napredak u teoriji SVM-a je načinjen 1995 godine od strane *Vapnik-a*, kada je linearni SVM proširen kernel funkcijama, i time postao primenljiv i na grupu nelinearnih problema [86]. U tom cilju, ulazni atributi su preslikani u višedimenzionalni prostor atributa, koji čak može biti i beskonačnodimenzionalan. Zatim se u tom prostoru formira linearna regresija (odnosno linearna razdvajajuća hiper-ravan u slučaju klasifikacije, Slika 4.7).



Slika 4.7 Preslikavanje $\phi(x)$

Međutim, nema potrebe za eksplicitnom upotrebom preslikavanja $\phi(x)$. Ovo je moguće zbog sledeće osobine: za svaku simetričnu kontinualnu funkciju $K(x, z)$ koja zadovoljava *Mercer-ove* uslove [110], postoji Hilbertov prostor H , preslikavanje $\phi(\cdot): R^n \rightarrow H$ i pozitivni brojevi λ_i , takvi da važi:

$$K(x, z) = \sum_{i=1}^{n_H} \lambda_i \phi_i(x) \phi_i(z), \quad (4.50)$$

gde $x, z \in \mathbb{R}^n$ a n_H predstavlja dimenziju od H . Mercer-ovi uslovi zahtevaju da bude ispunjena i relacija:

$$\int K(x, z) g(x) g(z) dx dz \geq 0, \quad (4.51)$$

za svaku kvadratnu integrabilnu funkciju $g(x)$. Izraz (4.50) se može napisati kao

$$K(x, z) = \sum_{i=1}^{n_H} \sqrt{\lambda_i} \phi_i(x) \sqrt{\lambda_i} \phi_i(z), \text{ gde su } \varphi_i(x) = \sqrt{\lambda_i} \phi_i(x) \text{ i } \varphi_i(z) = \sqrt{\lambda_i} \phi_i(z). \text{ Nakon toga}$$

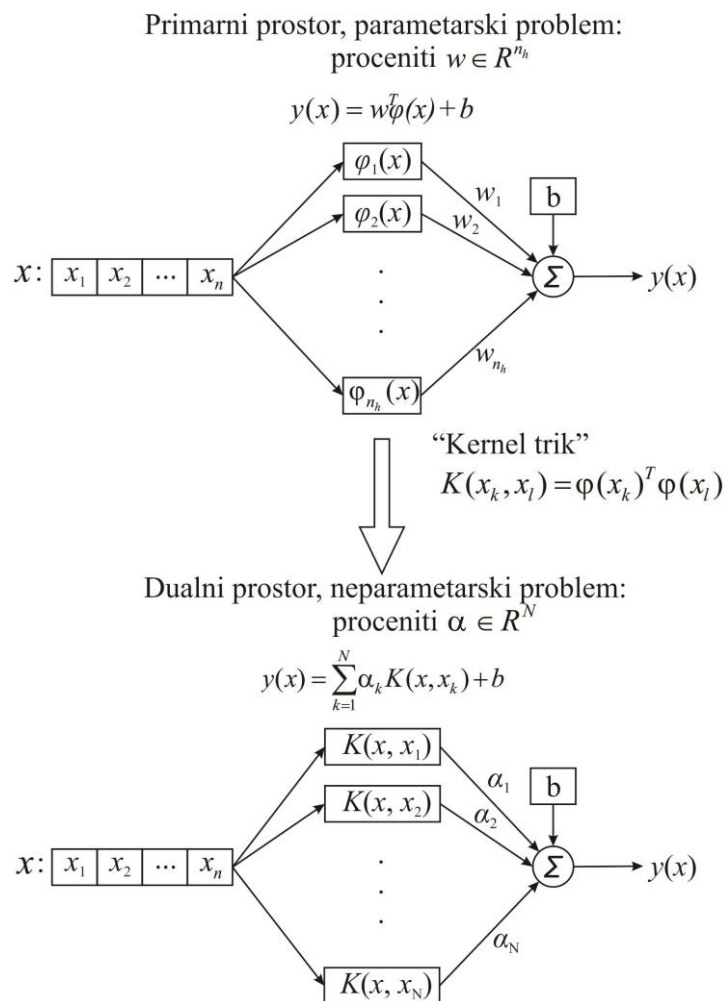
kernel funkcija se može predstaviti preko skalarnog proizvoda:

$$K(x, z) = \varphi_i(x)^T \varphi_i(z). \quad (4.52)$$

Kernel funkcija mora da bude simetrično pozitivno definisana da bi važila relacija (4.52), što ima za posledicu da je i sam kernel K separabilan. Primena relacije (4.52) je u literaturi poznata kao kernel trik. On nam omogućava da radimo u višedimenzionalnim prostorima bez potrebe da vršimo eksplicitna izračunavanja u njima. U slučaju LS-SVM-a (kao i kod SVM-a) početna formulacija problema je u primarnom težinskom prostoru sa višedimenzionalnim prostorom atributa (dobijenim nakon primene preslikavanja $\varphi(x)$). Međutim problem se ne rešava u primarnom težinskom prostoru, već u dualnom prostoru Lagranževih multiplikatora, nakon primene kernel trika. Na ovaj način omogućen je rad u višedimenzionalnom prostoru atributa bez eksplicitnog obavljanja izračunavanja u njemu (sva izračunavanja obavljaju se sa vektorima dimenzije n , iz originalnog nepreslikanog prostora).

U primarnom prostoru problem je određen parametrima ω (problem je parametarski sa fiksnom dimenzijom vektora $\omega \in \mathbb{R}^{n_h}$, gde je n_h dimenzija “preslikanog” prostora). U dualnom prostoru, problem postaje neparametarski (umesto određivanja vektora ω određuje se vektor α) i složenost rešenja $\alpha \in \mathbb{R}^N$ raste samo sa porastom broja trening ulaza N . Ovo ima za posledicu da na veličinu KKT sistema ne utiče dimenzija ulaznog prostora n , već samo broj trening ulaza N , Slika 4.8.

U slučaju izrazito velikog N (npr. $N \approx 10\,000$), moguće je regresioni problem definisan sa (4.36 - 4.38) rešavati u primarnom prostoru. Više detalja o ovome može se pronaći u [111], [112].



Slika 4.8 Primarno-dualna interpretacija LS-SVM-a

Postoji više mogućih izbora za kernel $K(\cdot, \cdot)$, i neki od njih su:

Linearan:
$$k(x, x_k) = x_k^T x, \tag{4.53}$$

Gausov (RBF):
$$k(x, x_k) = e^{-\frac{\|x-x_k\|^2}{\sigma^2}}, \tag{4.54}$$

Polinomni:
$$k(x, x_k) = (\tau + x_k^T x)^d, \tag{4.55}$$

MLP kernel:
$$k(x, x_k) = \tanh(k_1 x_k^T x + k_2). \tag{4.56}$$

Mercerovi uslovi su ispunjeni za sve vrednosti parametara σ i sve pozitivne vrednosti parametra τ u slučaju RBF (*eng. radial basis function*) i polinomnog kernela respektivno, ali ne i za sve kombinacije k_1 i k_2 u slučaju MLP (*eng. multi layer perceptron*) kernela. U problemima nelinearne regresije najčešće je u upotrebi RBF kernel kao mera sličnosti dva

vektora, tako da je on korišćen u eksperimentima opisanim u sekciji sa rezultatima, iako u opštem slučaju optimalni izbor kernela za dati problem nije unapred poznat. U slučaju izbora RBF kernela, primećujemo da samo α i b slede kao rešenja KKT sistema (4.46), ali ne i parametar koji određuje širinu kernela σ . U prethodnoj sekciji, u izrazu (4.37) javlja se i parametar γ koji takođe nije unapred poznat. Da bi došli do optimalnog rešenja za α i b , neophodno je prvo odrediti optimalnu kombinaciju parametara (γ, σ) . Može se primetiti da je potrebno odrediti samo dva parametra (γ, σ) za razliku od SVM-a gde je potrebno odrediti tri $(\gamma, \sigma, \varepsilon)$. Postoji više načina za određivanje ovih parametara i u sekciji 4.4.4 analizirani su najčešće korišćeni algoritmi: „*grid-search*“ i „*k-fold cross-validation*“.

4.4.3 Rešavanje KKT sistema – CG algoritam

Kao što je spomenuto u sekciji 4.4.1 prilikom formiranja LS-SVM modela, neophodno je rešavanje linearnog KKT sistema. Iako je sistem moguće rešiti metodama direktne eliminacije, veličina matrice iz izraza (4.46) raste sa porastom broja ulaza N . Rešavanje sistema direktnom eliminacijom ograničeno je na trening skupove koji sadrže relativno mali broj ulaza (npr. do $N=2000$). Za veće trening skupove preporučuje se upotreba iterativnih metoda, kao što je „*conjugate gradient*“ algoritam [113].

Za sistem linearnih jednačina $Ax = b$, gde je kvadratna matrica A simetrična i pozitivno definisana, CG algoritam iterativnim postupkom nalazi numeričko rešenje sistema, x_* . Prva iteracija može da startuje sa vektorom $x_0=0$, koji je inicijalna aproksimacija za x_* . Algoritam se prekida kada r_{k+1} dostigne unapred definisanu toleranciju, npr. $\varepsilon = 10^{-6}$. Za praćenje rada i numeričku proveru konvergencije algoritma poželjno je u svakoj iteraciji pratiti da kvadratna funkcija gubitka $f(x) = \frac{1}{2}x^T Ax - x^T \beta$, koja takođe minimizuje x_* monotono opada u svakoj iteraciji. CG je prikazan pseudo-kodom u Proceduri 4.1:

Procedura 4.1 CG algoritam

$$r_0 = b - Ax_0$$

$$p_0 = r_0$$

$$k = 0$$

repeat

$$\alpha_k = \frac{r_k^T r_k}{p_k^T A r_k}$$

$$x_{k+1} = x_k + \alpha_k p_k$$

$$r_{k+1} = r_k - \alpha_k A p_k$$

If $r_{k+1} < \varepsilon$

$$\beta_k = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$$

$$p_{k+1} = r_{k+1} - \beta_k p_k$$

$$k=k+1$$

end if

end repeat

Da bi se na rešavanje sistema (4.46) mogao primeniti CG algoritam, potrebno je prvo sistem transformisati tako da bude pozitivno definisan. Iako je $Z = \Omega + I\gamma^{-1}$ pozitivno definisana matrica, rezultujuća matrica reda $(n+1) \times (n+1)$ u (4.46) to ne mora da bude. Sistem (4.45) se može transformisati u pozitivno definisan sistem, na način opisan u [114], i rezultujući sistem postaje:

$$\begin{bmatrix} Z & 0 \\ 0 & 1_v^T Z^{-1} 1_v \end{bmatrix} \begin{bmatrix} \alpha + b Z^{-1} 1_v \\ b \end{bmatrix} = \begin{bmatrix} 1_v^T Z^{-1} y \\ 0 \end{bmatrix}, \quad (4.57)$$

gde je $S = 1_v^T Z^{-1} 1_v > 0$. Nakon toga se sistem (4.57) reda $(n+1) \times (n+1)$ razlaže na dva sistema (4.58) i (4.59), oba reda $n \times n$, koji se nezavisno rešavaju po η i v primenom CG algoritma:

$$Z\eta = 1_v, \quad (4.58)$$

$$Zv = y. \quad (4.59)$$

Nakon toga, rešenja sistema (4.57) su definisana sa (4.60) i (4.61):

$$b = \eta^T y / s, \quad (4.60)$$

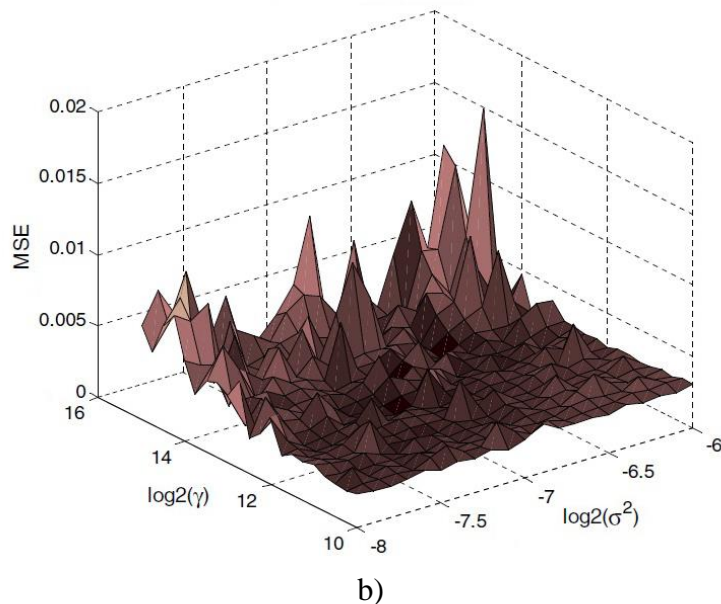
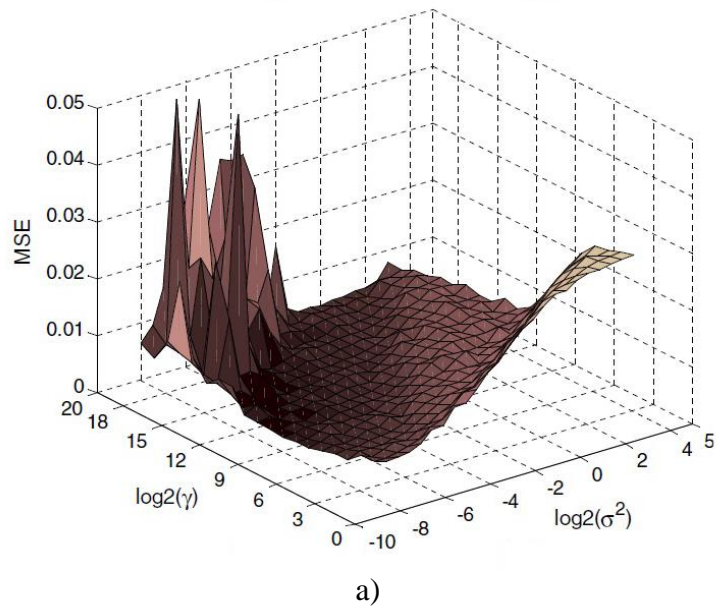
$$\alpha = v - b\eta. \quad (4.61)$$

Treba napomenuti da je ovde prikazan samo osnovni CG algoritam primenjen na rešavanje LS-SVM KKT sistema. Postoji više modifikacija polaznog sistema kao i samog CG algoritma koje značajno poboljšavaju performanse numeričkih izračunavanja. Neke od njih mogu se naći u [115] i [116]. Takođe u upotrebi je i modifikacija „*Platt-ovog sequential minimal optimization*“ (SMO) algoritma [117], koja je pokazala značajno bolje performanse u rešavanju LS-SVM KKT sistema koga karakterišu veliki trening skupovi.

4.4.4 Tehnike za optimizaciju LS-SVM parametara

U sekciji 4.4.2 i 4.4.3 pokazano je da rešavanje KKT sistema daje vrednosti za vektor α i parametar b . Takođe, pokazano je da optimalne vrednosti parametara γ i širine kernela σ (u slučaju izbora RBF) nisu unapred određene za dati skup podataka. Štaviše, da bi uopšte moglo da se započne rešavanje KKT sistema, moraju biti definisane bar inicijalne vrednosti ovih parametara. Optimalni izbor ovih parametara je od ključnog uticaja na kvalitet rešenja iz razloga što oni direktno kontrolišu balans između nad-podešavanja (*overfitting*) i pod-podešavanja (*underfitting*) modela (*eng. bias-variance tradeoff*). Jako male vrednosti parametra σ mogu uticati na pojavu nad-podešavanja (*eng. low bias-higher variance*), dok njegove jako velike vrednosti mogu prouzrokovati pod-podešavanje (*eng. high bias-lower variance*). Takođe jako male vrednosti parametra γ mogu uticati na pojavu nad-podešavanja, odnosno njegove veoma velike vrednosti mogu prouzrokovati pod-podešavanje. U literaturi je ovaj problem poznat kao izbor modela (*eng. model selection*) [118].

Postoji više načina za određivanje optimalne kombinacije ovih parametara. Najjednostavniji od njih je da se inicijalni skup podataka podeli na tri disjunktne podskupa: trening skup, validacioni skup i test skup. Zatim se na validacionom skupu ispituju kombinacije parametara (γ, σ) po unapred formiranoj mreži (*eng. grid*) (koraci za γ i σ u mreži ne moraju biti ekvidistantni), i na kraju se odabere ona kombinacija čije vrednosti daju najbolje rezultate na validacionom skupu. Nedostatak ove tehnike je taj što izbor parametara zavisi od izbora validacionog skupa. Bolja alternativa je „*shrinking grid-search*“, Slika 4.9 a) i b), u kombinaciji sa k -unakrsnih validacija [119]. Kombinacija ovih prisupa prikazana je u Proceduri 4.2.



Slika 4.9 Shrinking grid-search u dva nivoa

Procedura 4.2 Shrinking grid-search i k -unakrsnih validacija

1. Podeliti inicijalni skup podataka u odnosu: 2/3 za trening i validacioni podskup i 1/3 za test podskup.
2. Prva iteracija, $i=0$. Startovati k -fold cross-validation proceduru (npr. $k=10$) nad trening-validacionim podskupom za svaku kombinaciju parametara (γ, σ) iz inicijalnog skupa kandidata (Σ_0, Γ_0) u prvoj iteraciji (npr. $\Sigma_0 = \{5, 10, 15, 20, 25, \dots, 500\}$ i $\Gamma_0 = \{0.01, 0.05, 0.1, 0.5, 1, \dots, 10\}$).

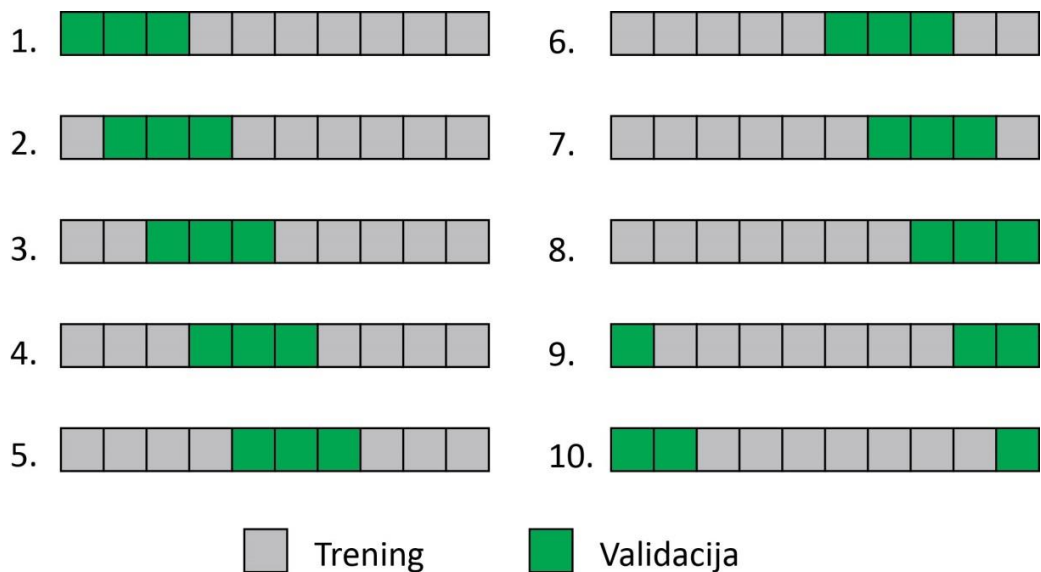
-
3. Izabrati optimalnu kombinaciju parametara (γ, σ) iz skupova Σ_i, Γ_i na osnovu procenjenih *k-fold cross-validation* performansi.
 4. Ako je $i=i_{MAX}$ (npr. $i_{MAX}=2$) ići na korak 5. U suprotnom $i=i+1$, zatim konstruisati lokalnu mrežu (*grid*) “finije rezolucije” oko optimalnih parametara (γ, σ) iz prethodne iteracije i ići na korak 3.
 5. Trenirati LS-SVM regresioni model koristeći ceo trening-validacioni skup sa prethodno određenom optimalnom kombinacijom parametara (γ, σ) .
 6. Proceniti uspešnost treniranog modela na test skupu.

U slučaju „*k-fold cross-validation*“ algoritma, trening-validacioni skup se na slučajan način deli na k disjunktnih podskupova jednake veličine. Od k podskupova, jedan se usvaja za validaciju tekućeg modela, dok se preostalih $k-1$ podskupova koriste za treniranje modela, Slika 4.10. Proces unakrsnih validacija se zatim ponavlja k puta (*eng. k-folds*), uz uslov da je svaki od k podskupova upotrebljen tačno jednom za validaciju modela. Zatim se rezultati iz svih k validacija usrednjavaju (ili na drugi način kombinuju) da bi se dobila ukupna mera uspešnosti modela. Kod regresije se za meru uspešnosti najčešće koristi srednja kvadratna greška (*eng. mean squared error, MSE*) iz svih validacija, mada je moguće koristiti i druge mere, kao što je npr. srednja apsolutna greška (*eng. mean absolute error, MAE*). Prednost ove metode je u tome što su svi podaci iz trening-validacionog skupa korišćeni i za treniranje i za validaciju, a svaki podskup od njih k je korišćen za validaciju tačno jednom. Najčešće je u upotrebi $k=10$ unakrsnih validacija, iako optimalna vrednost parametra k nije unapred definisan parametar. U eksperimentima prezentovanim u sekciji sa rezultatima za određivanje optimalne kombinacije (γ, σ) LS-SVM parametara korišćen je prethodno opisan pristup sa $k=10$ unakrsnih validacija u kombinaciji sa „*shrinking grid-search*“ algoritmom u dva nivoa. Granice i korak grida u prvom nivou su definisani u opsegu $\gamma \in \{2^0, 2^1, 2^2, \dots, 2^{20}\}$ i $\sigma \in \{2^{-10}, 2^{-9}, 2^{-8}, \dots, 2^{10}\}$, u skladu sa preporukom iz [120].

„*Leave-one-out cross-validation*“ (LOOCV), kao što joj naziv sugeriše, koristi po jedan primer (vektor) iz trening-validacionog skupa za validaciju modela, dok sve ostale primere koristi za treniranje modela. Ova procedura se ponavlja tako da se svaki primer iz trening skupa koristi tačno jednom za validaciju modela. LOOCV se svodi na k -unakrsnih validacija gde je $k = N$. LOOCV je zahtevnija što se tiče vremena izvršavanja u odnosu na k -unakrsnih

validacija jer zahteva mnogo više iteracija treniranja modela (broj iteracija je jednak broju trening primera N).

U cilju redukcije vremena „*grid-search*“ algoritma, u skorije vreme koriste se sve više meta-heurističke metode za optimizaciju parametara SVM-a i LS-SVM-a bazirane na evolucionim algoritmima (*eng. evolutionary algorithms*) i algoritmima inteligencije roja (*eng. swarm intelligence*) kao što su npr. „*genetic algorithm*“, „*particle swarm optimization*“, „*ant colony optimization*“, „*artificial bee colony algorithm*“ itd.



Slika 4.10 Primer unakrsnih validacija za $k=10$, 70% skupa za trening i 30% za validaciju modela

4.4.5 Dobijanje *sparse* LS-SVM rešenja

Kao što je spomenuto, jedan od nedostataka LS-SVM-a je taj što svi trening vektori učestvuju u formiranju rešenja, odnosno rešenje nije sparse. Ovo je direktna posledica uslova optimalnosti (4.42). Postoji više načina kojima se može postići sparse rešenje, ukoliko je potrebna. Najjednostavniji način za to je odbacivanje primera iz trening skupa koji najmanje utiču na rešenje.

Iz linearnog sistema (4.46) određuju se vrednosti α_k . Logika odbacivanja je sledeća: vektori kojima odgovaraju male vrednosti $|\alpha_k|$ su manje relevantni za formiranje modela po analogiji sa SVM-om gde vektori kojima odgovaraju vrednosti za α_k jednake nuli ne učestvuju u formiranju modela. Algoritam odbacivanja funkcioniše u nekoliko koraka, pri čemu u svakom koraku biva odbačen relativno mali broj vektora, npr. 5% onih koji

odgovaraju najmanjim vrednostima $|\alpha_k|$. Nakon svakog koraka linearni sistem (4.46) se rešava ponovo. Osnovni metod odbacivanja prikazan je u Proceduri 4.3:

Procedura 4.3 Algoritam za postizanje sparse LS-SVM-a

1. Trenirati LS-SVM model na osnovu svih N vektora u trening skupu,
2. ukloniti mali deo vektora, (npr. 5% od celog trening skupa), onih koji odgovaraju po apsolutnoj vrednosti najmanjim vrednostima α_k u sortiranom α_k nizu,
3. re-trenirati LS-SVM model, ovaj put sa redukovanim trening skupom,
4. ukoliko performanse modela ne opadnu ispod unapred zadatog kriterijuma, ići na korak 2. U suprotnom proveriti da li ponovna procena parametara γ i σ (u slučaju RBF kernela) može popraviti performance.

U ovoj sekciji prikazan je samo osnovni algoritam odbacivanja za postizanje sparsnosti (eng. *sparse*) LS-SVM rešenja. Postoje poboljšane verzije ovog algoritma, koje se mogu pronaći u [121], [122].

4.5 Relevance Vector Machines

Tipping u [123] predlaže „*relevance vector machines*“ (RVM) metod koji se kao i SVM bazira na istim idejama ali preinačenim kroz kontekst Bajesove teoreme. Za dati trening skup $\{\mathbf{x}_n, t_n\}_{n=1}^N$, generalizovan linearni regresioni model koji opisuje zavisnost između ulaznog vektora \mathbf{x} i skalarne ciljne vrednosti t dat je sa:

$$t_n = y(\mathbf{x}_n, \mathbf{w}) + \varepsilon_n, \mathbf{t} = \mathbf{y} + \boldsymbol{\varepsilon}, \quad (4.62)$$

pri čemu je y funkcija definisana sa

$$y(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^M \omega_i \psi_i(x) = \mathbf{w} \boldsymbol{\phi}(x), \quad (4.63)$$

a ε je šum sa Gausovom raspodelom, očekivanjem 0 i disperzijom σ^2 , dok je M u (4.63) broj baznih funkcija. Shodno tome, t_n će biti slučajna promenljiva sa raspodelom $p(t_n | x) = N(t_n | y(x_n); \sigma^2)$. Funkcija verodostojnosti za celokupan skup podataka biće određena sa

$$p(\mathbf{t} | \mathbf{w}, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{\|\mathbf{t} - \Phi\mathbf{w}\|^2}{2\sigma^2}\right\}, \quad (4.64)$$

gde je $\mathbf{t} = [t_1, \dots, t_N]^T$, $\mathbf{w} = [w_1, \dots, w_M]^T$, i Φ je $N \times M$ matrica sa $\Phi_{nm} = \phi_m(x_n)$.

Svakoj težini w_i pridružuje se jedan hiperparametar α_i . Zadatak hiperparametra jeste da odredi kada bazna funkcija više nije relevantna za algoritam. Bazna funkcija se izbacuje kada više nije statistički značajna, čime se dobija bolja sparsnost (*eng. sparse*) i brži algoritam. Shodno tome raspodela parametara težine određuje se kao

$$p(\mathbf{w} | \alpha) = \prod_{i=1}^N N(w_i | 0; \alpha_i^{-1}), \quad (4.65)$$

dok se za raspodelu hiperparametara uzima najčešće gama raspodela:

$$p(\alpha) = \prod_{i=1}^N \text{Gamma}(\alpha_i | a, b), \quad (4.66)$$

$$p(\sigma^2) = \prod_{i=1}^N \text{Gamma}(\sigma^2 | c, d), \quad (4.67)$$

$$\text{Gamma}(\alpha | x, y) = \Gamma(x)^{-1} y^x \alpha^{x-1} \exp^{-y\alpha},$$

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$$

Parametri a, b, c, d zadaju se u zavisnosti od problema ali u najvećem broju slučajeva zahteva se da α i σ^2 imaju približno uniformne raspodele. Stoga se za a, b, c, d uzimaju vrednosti bliske nuli ($\approx 10^{-4}$).

Sada kada je problem priora rešen ostaje pitanje kako proceniti posterior za sve nepoznate $p(\mathbf{w}, \alpha, \sigma^2 | \mathbf{t})$, ako su poznati ulazni podaci. Za rešavanje ovog problema može poslužiti Bajesova teorema. Dakle, za posterior imamo:

$$p(\mathbf{w}, \alpha, \sigma^2 | \mathbf{t}) = \frac{p(\mathbf{t} | \mathbf{w}, \alpha, \sigma^2) p(\mathbf{w}, \alpha, \sigma^2)}{p(\mathbf{t})}, \quad (4.68)$$

što se ne može odrediti analitički, pa se tako posterior može razložiti na sledeći način:

$$p(\mathbf{w}, \alpha, \sigma^2 | \mathbf{t}) = p(\mathbf{w} | \alpha, \sigma^2, \mathbf{t}) p(\alpha, \sigma^2 | \mathbf{t}). \quad (4.69)$$

Prvi činilac iz (4.69) (*eng. weight posterior term*) se računa kao konvolucija Gausijana:

$$\begin{aligned}
p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}, \sigma^2) &= \frac{p(\mathbf{t} | \mathbf{w}, \sigma^2) p(\mathbf{w} | \boldsymbol{\alpha})}{p(\mathbf{t} | \boldsymbol{\alpha}, \sigma^2)} \\
&= (2\pi)^{-(N+1)/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{w} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{w} - \boldsymbol{\mu}) \right\},
\end{aligned} \tag{4.70}$$

gde su:

$$\boldsymbol{\Sigma} = (\sigma^{-2} \boldsymbol{\Phi}^T \boldsymbol{\Phi} + \mathbf{A})^{-1}, \tag{4.71}$$

$$\boldsymbol{\mu} = \sigma^{-2} \boldsymbol{\Sigma} \boldsymbol{\Phi}^T \mathbf{t}. \tag{4.72}$$

Matrica \mathbf{A} je dijagonalna matrica hiperparametara α , $\mathbf{A} = \text{diag}(\alpha_0, \dots, \alpha_N)$.

Drugi činilac u (4.69) (*eng. hyperparameter posterior term*) se ne računa direktno. Aposteriorne funkcije gustina hiperparametara se aproksimiraju delta funkcijom u okolini modusa. Dakle posterior se zamenjuje sa $\delta(\alpha_{MAP}, \sigma_{MAP}^2)$ pri čemu je

$$(\alpha_{MAP}, \sigma_{MAP}^2) = \arg \max_{\alpha, \sigma^2} p(\boldsymbol{\alpha}, \sigma^2 | \mathbf{t}). \tag{4.73}$$

Tako je problem obučavanja RVM-a sveden na maksimizaciju $p(\alpha, \sigma^2 | \mathbf{t})$. Prema Bajesovoj teoremi ovaj izraz se može napisati kao:

$$p(\boldsymbol{\alpha}, \sigma^2 | \mathbf{t}) = p(\mathbf{t} | \boldsymbol{\alpha}, \sigma^2) p(\boldsymbol{\alpha}) p(\sigma^2). \tag{4.74}$$

Kako je uvedena pretpostavka o uniformnoj raspodeli hiperparametara, potrebno je maksimizovati samo prvi činilac iz (4.74).

$$\begin{aligned}
p(\mathbf{t} | \boldsymbol{\alpha}, \sigma^2) &= \int p(\mathbf{t} | \mathbf{w}, \sigma^2) p(\mathbf{w} | \boldsymbol{\alpha}) d\mathbf{w} \\
&= (2\pi)^{-N/2} |\sigma^2 \mathbf{I} + \boldsymbol{\Phi} \mathbf{A}^{-1} \boldsymbol{\Phi}^T|^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{t}^T (\sigma^2 \mathbf{I} + \boldsymbol{\Phi} \mathbf{A}^{-1} \boldsymbol{\Phi}^T)^{-1} \mathbf{t} \right\}.
\end{aligned} \tag{4.75}$$

Diferenciranjem $\log p(\mathbf{t} | \boldsymbol{\alpha}, \sigma^2)$ po α i σ^2 , i izjednačavanjem dobijenih izraza sa nulom dobijamo iterativne jednačine:

$$\alpha_i^{Novo} = \frac{\gamma_i}{\mu_i^2}, \tag{4.76}$$

$$(\sigma^2)^{Novo} = \frac{\|\mathbf{t} - \boldsymbol{\Phi} \boldsymbol{\mu}\|^2}{N - \sum_{i=1}^M \gamma_i}, \tag{4.77}$$

gde je γ_i definisano sa:

$$\gamma_i = 1 - \alpha_i \Sigma_{ii}. \quad (4.78)$$

Proces predikcije dat je kroz sledeći algoritam:

Procedura 4.4 RVM regresioni proces

1. Inicijalizovati sve $\{\alpha_i\}$ i σ^2 .
2. Izračunati $\boldsymbol{\mu}$ i $\boldsymbol{\Sigma}$ (*weight posterior*).
3. Izračunati svako $\{\gamma_i\}$ a zatim ponovo izračunati $\{\alpha_i\}$ (i σ^2 ako je potrebno).
4. Vratiti se na korak 2 sve dok uslov konvergencije ne bude ispunjen.
5. “Obrisati“ težine (bazne funkcije) koje odgovaraju optimalnim vrednostima $\alpha_i = \infty$.
6. Za nove podatke odrediti predikcionu raspodelu sa dobijenim vrednostima $\boldsymbol{\alpha}_{MP}$ i σ_{MP}^2

$$p(t_* | \mathbf{t}) = \int p(t_* | \mathbf{w}, \sigma_{MP}^2) p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) d\mathbf{w}, \quad (4.79)$$

dok $y(\mathbf{x}_*; \boldsymbol{\mu})$ predstavlja očekivanje.

4.6 Procena zajedničkih informacija

Ovde je data definicija zajedničkih informacija na osnovu diferencijalne entropije. Analizirane su najvažnije osobine zajedničkih informacija koje se odnose na izbor atributa i vektora. Ukratko su analizirana ograničenja procene zajedničkih informacija pomoću histogram i kernel estimatora. Zatim je prezentovan i analiziran pristup za procenu zajedničkih informacija na osnovu algoritma k najbližih suseda koji prevazilaze ova ograničenja, kao i strategija izbora parametra k . Na kraju sekcije diskutovani su najbitniji detalji vezani za implementaciju k NN-MI estimatora.

4.6.1 Definicija zajedničkih informacija

Definicija zajedničkih informacija je izvedena na osnovu diferencijalne entropije iz teorije informacija [124]. Može da detektuje i nelinearne zavisnosti između slučajnih promenljivih bez prethodnog poznavanja strukture njihovih odnosa, i u tome se ogleda njegova prednost u odnosu na linearne estimatore, kao što je npr. korelacioni koeficijent. Pored toga, MI može biti definisan i za grupe promenljivih. MI između dve slučajne

promenljive X i Y određuje količinu informacija koju one dele. Formalno definisano, količina zajedničkih informacija koju promenljive dele, određuje za koliko poznavanje jedne od promenljivih smanjuje neizvesnost, odnosno neodređenost one druge. Označimo sa X i Y kontinualne slučajne promenljive sa zajedničkom gustinom verovatnoće $\mu_{X,Y}$ i marginalnim gustinama verovatnoće:

$$\mu_X(x) = \int_{-\infty}^{+\infty} \mu_{X,Y}(x, y) dy, \quad (4.80)$$

$$\mu_Y(y) = \int_{-\infty}^{+\infty} \mu_{X,Y}(x, y) dx. \quad (4.81)$$

Neizvesnost promenljive Y je definisana diferencijalnom entropijom:

$$H(Y) = - \int_{-\infty}^{+\infty} \mu_Y(y) \log \mu_Y(y) dy, \quad (4.82)$$

gde \log označava prirodni logaritam, tako da se količina informacija meri u “prirodnim” jedincama (*eng. natural units*).

Ukoliko je promenljiva X poznata, rezultujuća neizvesnost promenljive Y je definisana uslovnom entropijom definisanom sa:

$$H(Y|X) = - \int_{-\infty}^{+\infty} \mu_X(x) \int_{-\infty}^{+\infty} \mu_Y(y|X=x) \log \mu_Y(y|X=x) dx dy. \quad (4.83)$$

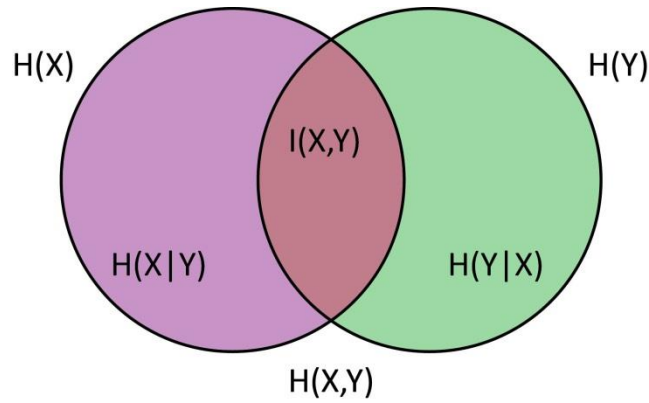
Zajednička neizvesnost para promenljivih (X, Y) je određena zajedničkom entropijom:

$$H(X, Y) = - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mu_{X,Y}(x, y) \log \mu_{X,Y}(x, y) dx dy. \quad (4.84)$$

Količina zajedničkih informacija između promenljivih X i Y je definisana sa

$$I(X, Y) = H(Y) - H(Y|X) \quad (4.85)$$

i simbolički prikazna na Slici 4.11:



Slika 4.11 Simbolički prikaz zajedničkih informacija između X i Y

Izraz (4.85) se može interpretirati na sledeći način: “smanjenje neizvesnosti promenljive Y kada je promenljiva X poznata”. Bitno je naglasiti da količina zajedničkih informacija predstavlja redukciju neizvesnosti, a ne ekvivalenciju sa neizvešnošću. MI se takođe može predstaviti i u ekvivalentnim formama:

$$I(X, Y) = H(X) + H(Y) - H(Y, X), \quad (4.86)$$

$$I(X, Y) = H(X) - H(X|Y), \quad (4.87)$$

$$I(X, Y) = H(X, Y) - H(X|Y) - H(Y|X). \quad (4.88)$$

U nastavku su nabrojane neke od osobina koje važe za MI, a od značaja su prilikom izbora atributa i vektora:

1. Simetrija, odnosno $I(X, Y) = I(Y, X)$,
2. MI je uvek veće ili jednako od nule, $I(X, Y) > 0$, odnosno poznavanjem promenljive Y ne može se umanjiti neodređenošću promenljive X ,
3. MI je jednako nuli $I(X, Y) = 0$, ako i samo ako su promenljive X i Y nezavisne,

4. Relacija

$$d(X, Y) = H(X, Y) - I(X, Y) = H(X) + H(Y) - 2I(X, Y) = H(X|Y) + H(Y|X),$$

definiše metriku, i u literaturi je poznata kao varijacija informacija,

5. MI između dve slučajne promenljive X i Y predstavlja *Kullback-Leibler divergence* (KL) između njihove zajedničke gustine verovatnoće i proizvoda njihovih marginalnih gustina verovatnoće: $I(X, Y) = KL(\mu_{X,Y}(x, y) \| \mu_X(x)\mu_Y(y))$,
6. Invarijantnost u odnosu na reparametrizaciju marginalnih promenljivih X i Y . Ako su $X'=F(X)$ i $Y'=G(Y)$ homeomorfizmi, onda važi $I(X', Y') = I(X, Y)$ i
7. Svojstvo hijerarhijske dekompozicije $I(X, Y, Z) = I((X, Y), Z) + I(X, Y)$.

Iz (4.80 - 4.84)) MI se može izračunati kao:

$$I(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mu_{X,Y}(x, y) \log \frac{\mu_{X,Y}(x, y)}{\mu_X(x)\mu_Y(y)} dx dy. \quad (4.89)$$

Kao što se može zaključiti iz (4.89) za procenu zajedničkih informacija potrebno je poznavati zajedničku gustinu verovatnoće (*eng. joint probability density function-PDF*) za par promenljivih (X, Y) . Međutim u praktičnim problemima zajedničke gustine verovatnoća nisu unapred poznate, što ima za posledicu da se količina zajedničkih informacija ne može direktno izračunati primenom relacije (4.89), već se mora proceniti na osnovu skupa podataka.

Procena zajedničkih informacija je veoma aktuelan problem, i u tu svrhu se najčešće koriste histogram i kernel estimatori koji na osnovu skupa podataka vrše procenu nepoznatih funkcija gustine verovatnoće. Za procenu nepoznatih funkcija gustine verovatnoće, najintuitivniji i najkorišćeniji je estimator koji se zasniva na histogramu (*eng. histogram based estimator-HB*) [125]. Osnovna ideja kod histogram estimatora je ta da se ulazni prostor kontinualnih promenljivih podeli u više diskretnih particija. Nakon toga entropija i količina zajedničkih informacija se na osnovu ovih particija mogu proceniti zamenom integrala sa konačnim sumama. Za kontinualnu procenu entropije i zajedničkih informacija u [126] predložen je estimator koji se zasniva na kernelima (*eng. kernel based estimator-KB*). On na osnovu izabrane kernel funkcije i njenih parametara vrši procenu nepoznatih funkcija gustine verovatnoće direktno na osnovu skupa podataka, bez potrebe za particionisanjem ulaznog prostora u diskretne podskupove kao kod histogram estimatora.

Međutim, iako veoma aktuelne, ove metode često ne daju dobre rezultate kada rade sa višedimenzionalnim promenljivima i malim brojem primera. Razlog za to je što broj primera potrebnih za preciznu procenu zajedničkih gustina verovatnoće raste eksponencijalno sa porastom dimenzije prostora [127]. S obzirom na to da je broj raspoloživih primera na

osnovu kojih se vrši ova procena u praksi uvek ograničen, i ukoliko su primeri višedimenzionalni vektori, može se dogoditi da većina particija kod histogram estimatora bude prazna, što može dovesti do vrlo nepreciznih procena. Takođe, procena zajedničkih gustina verovatnoće biće neprecizna ukoliko broj primera na osnovu kojih se ona vrši nije znatno veći u odnosu na broj formiranih particija. Iako kernel estimator prevazilazi neka od ovih ograničenja, i kod njega može doći do problema u proceni, pošto se on u suštini svodi na “glatki” histogram. Na primer, vreme potrebno za procenu zajedničkih informacija raste značajno sa porastom broja primera, procene u višedimenzionalnim prostorima sa malim brojem primera još uvek nisu dovoljno precizne, itd. Parametarski estimator zasnovan na algoritmu k najbližih komšija predložen je u [127]. Njegova prednost se ogleda u tome što on ne zahteva eksplicitne procene funkcija gustine verovatnoće, i time zaobilazi jednu od najproblematičnijih stavki u proceni zajedničkih informacija. Na taj način, estimator zasnovan na algoritmu k najbližih komšija, može precizno da vrši procenu zajedničkih informacija između višedimenzionalnih promenljivih direktno na osnovu skupova podataka.

4.6.2 Procedura za procenu zajedničkih informacija

Razmotrimo skup od N ulazno-izlaznih parova $z^i = (x^i, y^i)$, $i = 1, \dots, N$ koji čine međusobno nezavisne i sa identičnim raspodelama realizacije slučajne promenljive $Z=(X,Y)$, gde x i y mogu biti skalari ili vektori. Za svaki par tačaka z i z' iz prostora Z koristi se maksimalna norma za poređenje ulazno-izlaznih parova, definisana sa:

$$\|z - z'\| = \max \{ \|x - x'\|, \|y - y'\| \}, \quad (4.90)$$

dok se u prostorima X i Y mogu koristiti bilo koje norme (koje čak ne moraju biti iste).

Osnovna ideja algoritma predloženog u [127] je da se $I(X,Y)$ proceni preko rastojanja (procenjenih maksimalnom normom) od z^i do njegovih k najbližih komšija, usrednjenih preko celog z^i . Označimo sa $z^{k(i)} = (x^{k(i)}, y^{k(i)})$ k -tog najbližeg komšiju od z^i . Važno je naglasiti da $x^{k(i)}$ i $y^{k(i)}$ predstavljaju ulazno-izlazni par od $z^{k(i)}$ respektivno, i stoga ne neophodno k -tog najbližeg komšiju od x^i i y^i . Definišimo $d_X^i = \|x^i - x^{k(i)}\|$, $d_Y^i = \|y^i - y^{k(i)}\|$, $d_Z^i = \|z^i - z^{k(i)}\|$. Očigledno je da važi, $d^i = \max(d_X^i, d_Y^i)$. Zatim, treba odrediti n_X^i , odnosno broj tačaka x^j čije je rastojanje od x^i strogo manje od d^i , i slično n_Y^i , odnosno broj tačaka y^j čije je rastojanje od y^i strogo manje od d^i . Zatim, $I(X,Y)$ se može proceniti kao:

$$I(X, Y) = \psi(k) - \frac{1}{N} \sum_{i=1}^N [\psi(n_X^i + 1) + \psi(n_Y^i + 1)] + \psi(N), \quad (4.91)$$

gde je ψ di-gama funkcija:

$$\psi(t) = \frac{\Gamma'(t)}{\Gamma(t)} = \frac{d}{dt} \ln \Gamma(t), \quad (4.92)$$

a $\Gamma(t)$ gama funkcija:

$$\Gamma(t) = \int_0^{\infty} u^{t-1} e^{-u} du. \quad (4.93)$$

Funkcija ψ se može izračunati na osnovu rekurentne relacije $\psi(x+1) = \psi(x) + 1/x$ i početnog uslova $\psi(1) = C$, gde $C = -0.5772156$ predstavlja *Euler-Mascheroni-jevu* konstantu.

Još jedna alternativa je da se n_X^i odredi na osnovu broja tačaka x^j koje zadovoljavaju nejednakost $\|x^i - x^j\| \leq \frac{d_X^i}{2}$, i slično n_Y^i na osnovu broja tačaka y^j koje zadovoljavaju nejednakost $\|y^i - y^j\| \leq \frac{d_Y^i}{2}$. Zatim se, $I(X, Y)$ može proceniti kao:

$$I(X, Y) = \psi(k) - \frac{1}{k} - \frac{1}{N} \sum_{i=1}^N [\psi(n_X^i) + \psi(n_Y^i)] + \psi(N). \quad (4.94)$$

Za više promenljivih X_1, X_2, \dots, X_M MI estimator je definisan:

$$I(X_1, X_2, \dots, X_M) = \psi(k) - \frac{1}{N} \sum_{i=1}^N [\psi(n_{X_1}^i + 1) + \psi(n_{X_2}^i + 1) + \dots + \psi(n_{X_M}^i + 1)] + (M-1)\psi(N), \quad (4.95)$$

odnosno:

$$I(X_1, X_2, \dots, X_M) = \psi(k) - \frac{(M-1)}{k} - \frac{1}{N} \sum_{i=1}^N [\psi(n_{X_1}^i) + \psi(n_{X_2}^i) + \dots + \psi(n_{X_M}^i)] + (M-1)\psi(N). \quad (4.96)$$

Na osnovu osobine hijerarhijske dekompozicije, sledi:

$$I(X, Y, Z) = I((X, Y), Z) + I(X, Y). \quad (4.97)$$

Iteracijom izraza (4.97) može se izvršiti dekompozicija $I(X_1, X_2, \dots, X_M)$ za svako $M > 2$ i za bilo koji podskup skupa (X_1, X_2, \dots, X_M) na MI unutar elemenata u podskupu i na MI

između podskupova. Ova osobina je naročito značajna kod izbora atributa, jer se na taj način može izračunati MI između svih podskupova ulaznih atributa i ciljnih promenljivih. Druga alternativa kod izbora atributa je da se podskupovi atributa tretiraju kao jedna višedimenzionalna promenljiva, pa se procena MI vrši na osnovu (4.91) ili (4.94) s tim što su tada $x^i \in X$ vektori dimenzije M . U [128] je pokazano da oba estimatora (4.91) i (4.94) daju vrlo bliske rezultate u proceni zajedničkih informacija. Implementacija drugog pristupa, iz (4.94) se može naći u [129].

Iz (4.91) i (4.94) može se primetiti da parametar k , koji određuje broj suseda na osnovu kojih se vrši procena zajedničkih informacija nije unapred poznat. Zbog toga procena ovih estimatora zavisi od unapred izabrane vrednosti parametra k . U [127] je pokazano da se statistička greška estimatora povećava ukoliko k opada. To implicira da treba koristiti vrednosti $k > 1$ u cilju smanjenja statističke greške. Međutim, velike vrednosti parametra k treba izbegavati, jer tada povećanje sistematske greške može da premaši smanjenje statističke. Kao što je predloženo u [128] u eksperimentima prikazanim u disertaciji korišćena je vrednost $k=6$.

Kada se implementira direktno, algoritam najviše procesorskog vremena troši na lociranje k najbližih suseda. U tom slučaju potrebne su dve ugnježdene petlje kojima će se iterirati kroz sve elemente, i na taj način locirati susede. Složenost algoritma je u tom slučaju $O(N^2)$. Dok je ovo prihvatljivo kod manjih skupova podataka, npr. $N < 2000$, za veće skupove potrebni su brži algoritmi za lociranje suseda. Kod većih skupova podataka npr. $N > 2000$ moguće je koristiti $k-d$ tree algoritam [130]. Složenost algoritma u tom slučaju je $O(N \log N)$. Međutim, ovaj algoritam nije pogodan za korišćenje u slučaju višedimenzionalnih promenljivih, kada mu se složenost svodi na $O(N^2)$, tj. pogodan je za korišćenje samo ako je $N \gg 2^n$, gde je n dimenzija promenljivih, a N njihov broj. U tom slučaju potrebno je implementirati neki od algoritama za aproksimativno određivanje suseda. Neki od njih su: *best bin first* [131] - modifikacija osnovnog $k-n$ tree algoritma, *locality sensitive hashing* [132] i ϵ -*approximate nearest neighbor search* [133]. U praksi se mogu javiti slučajevi kada se u test skupovima naiđe na više tačaka koje imaju identične koordinate. U tom slučaju može se desiti da MI estimator zasnovan na k NN algoritmu daje neprecizne procene. Najjednostavniji način za rešavanje ovog problema je dodavanje Gausovog šuma vrlo male amplitude podacima (npr. $\approx 10^{-10}$), kao što je preporučeno u [127].

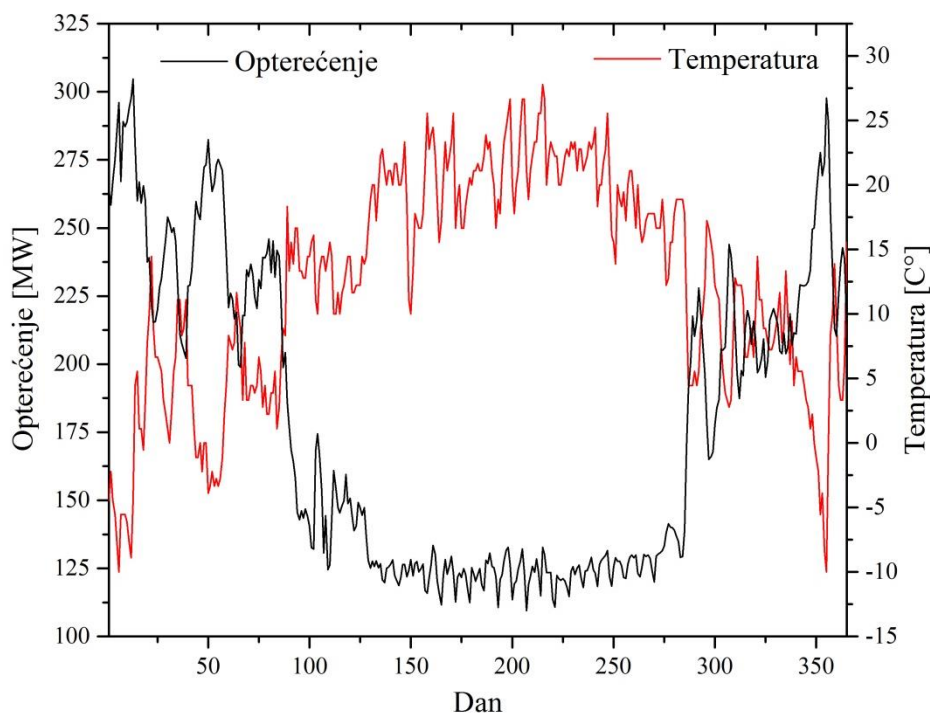
5 KARAKTERISTIKE POTROŠNJE ELEKTRIČNE ENERGIJE

U ovom poglavlju je analiziran uticaj promenljivih na krivu električne potrošnje sa ciljem da se ustanovi na koji način i u kolikoj meri ove promenljive oblikuju krivu potrošnje i tako izaberu odgovarajuće promenljive za attribute modela. U analizi su korišćeni podaci o električnom opterećenju grada Niša.

Kriva potrošnje električne energije oblikovana je uticajem velikog broja promenljivih. Zbog toga je za formiranje preciznog modela za kratkoročnu prognozu potrošnje električne energije značajno prepoznati one promenljive koje imaju najveći uticaj na potrošnju. Ove promenljive možemo smestiti u dve osnovne grupe: meteorološke i sociološke. Meteorološke promenljive se odnose na meteorološke elemente kao što su temperatura, vlažnost vazduha, brzina vetra, oblačnost, količina padavina, itd., posmatrane u određenom području u kratkom vremenskom periodu, najčešće do dve nedelje. Kada je reč o sociološkim promenljivama tu se misli na promenljive koje su definisane uticajem ljudskih aktivnosti na potrošnju električne energije. Ove sociološke promenljive se kod problema kratkoročne prognoze potrošnje svode na kalendarske promenljive: sat u danu, dan u nedelji, mesec u godini, period godine, indikator praznika.

Od meteoroloških promenljivih, temperatura ima najveći uticaj na potrošnju električne energije i u uskoj je korelaciji sa njom, tako da se često koristi kod modela prognoze potrošnje električne energije. Danas, se za potrebe prognoze potrošnje električne energije, smatra da je prognoza temperature relativno precizna do dan unapred. Na Slici 5.1 prikazano je srednje dnevno električno opterećenje i temperatura za period od godinu dana. Vidi se da opterećenje i temperatura imaju jak negativan korelacioni odnos. Drugim rečima, sa padom

temperature dolazi do porasta opterećenja i obrnuto, sa porastom temperature dolazi do pada opterećenja.

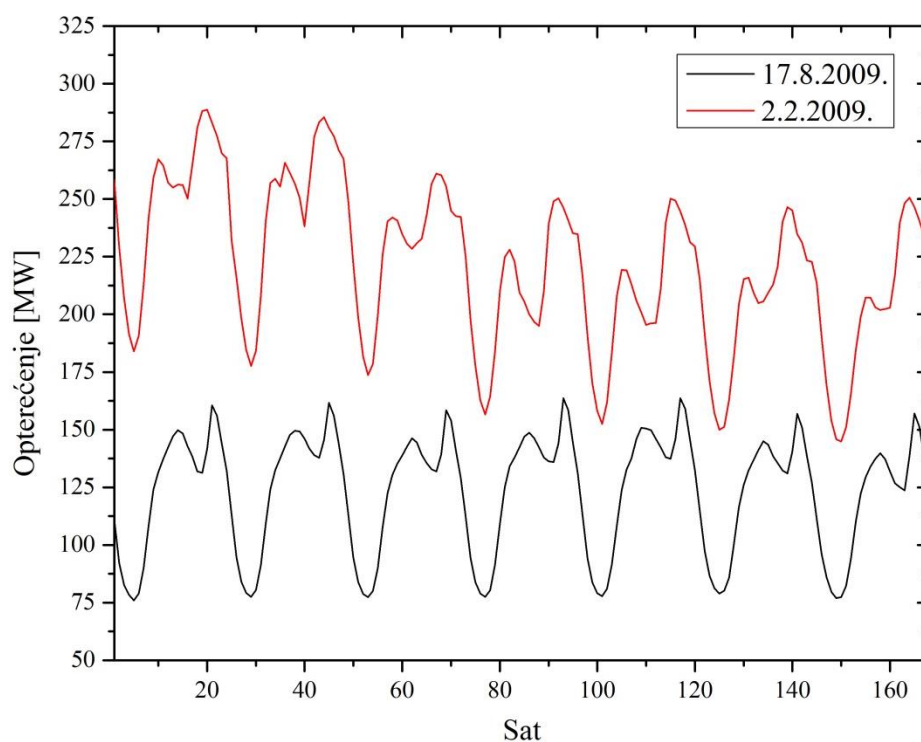


Slika 5.1 Srednje dnevno opterećenje i temperatura

Tako u ovom slučaju korelacioni koeficijent između opterećenja i temperature za zimski period (za period od oktobra do marta) iznosi -0.9 a u letnjem periodu (od aprila do septembra) je -0.59. Ovako visok korelacioni koeficijent naročito u zimskom periodu, potvrđuje značajan uticaj koji temperatura ima na krivu potrošnje električne energije. U skladu sa tim, temperatura je poželjan atribut kod formiranja modela prognoze električne potrošnje. Međutim, činjenica koja se ne može ignorisati jeste da je vrednost temperature za dan unapred nepoznata, tako da model može jedino da koristi prognoziranu vrednost temperature. Kako je prognoza temperature problem mnogo složeniji od problema prognoze potrošnje električne energije, iluzorno je formirati sopstveni model za prognozu temperature, već je najbolje koristiti usko specijalizovane servise koji nude tu uslugu. Naravno temperatura je atribut koji je poželjno uključiti u model ali ne i nužno, pa je tako danas sve više metoda koji izbegavaju korišćenje temperature ili bilo kog drugog meteorološkog atributa. Treba napomenuti da korelacija između temperature i električne potrošnje varira od skupa do skupa podataka pa čak i od godine do godine, pa se tako može desiti da je ovaj odnos loš pa se samim tim uključivanje temperature može loše odraziti na preciznost modela

[37]. Zbog toga, uvek kod procesa izbora atributa treba izvršiti analizu uticaja temperature na električnu potrošnju pre uključivanja ovog atributa u model.

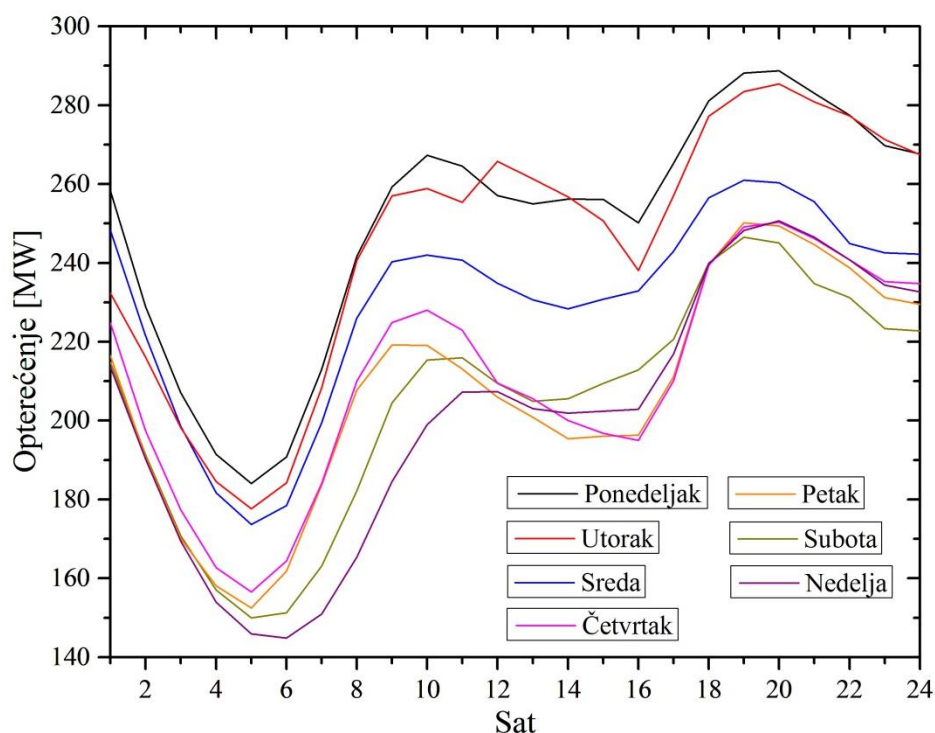
Uticao ponašanja ljudi na potrošnju električne energije se ogleda pre svega kroz dnevne i sedmične aktivnosti ljudi ali i kroz aktivnosti tokom dana praznika koji se izdvajaju u odnosu na obične dane. Sa Slike 5.1 se vidi da je opterećenje u letnjem periodu godine znatno manje i stabilnije nego u zimskom periodu gde su varijacije opterećenja izraženije. Na Slici 5.2 prikazane su krive električnog opterećenja u toku sedmice za zimski i letnji period godine, počevši od 2.2.2009. za zimski, odnosno 17.8.2009. za letnji period godine. Može se potvrditi još jednom da je u zimskom periodu električno opterećenje približno dvostruko veće sa izraženim varijacijama i promenom trenda od dana do dana, što nije slučaj u letnjem periodu godine. Letnji period karakteriše stabilnije opterećenje, sa manjim varijacijama i skoro bez trenda. Promena trenda u toku sedmice se ogleda kroz povećanu potrošnju u danima posle vikenda (ponedeljak i utorak), kada su ljudi nešto aktivniji u odnosu na drugu polovinu sedmice. Dane vikenda karakteriše smanjena potrošnja, što je direktna posledica neradnih dana kada su ljudi manje aktivni a industrijska proizvodnja stoji.



Slika 5.2 Kriva električnog opterećenja u toku sedmice

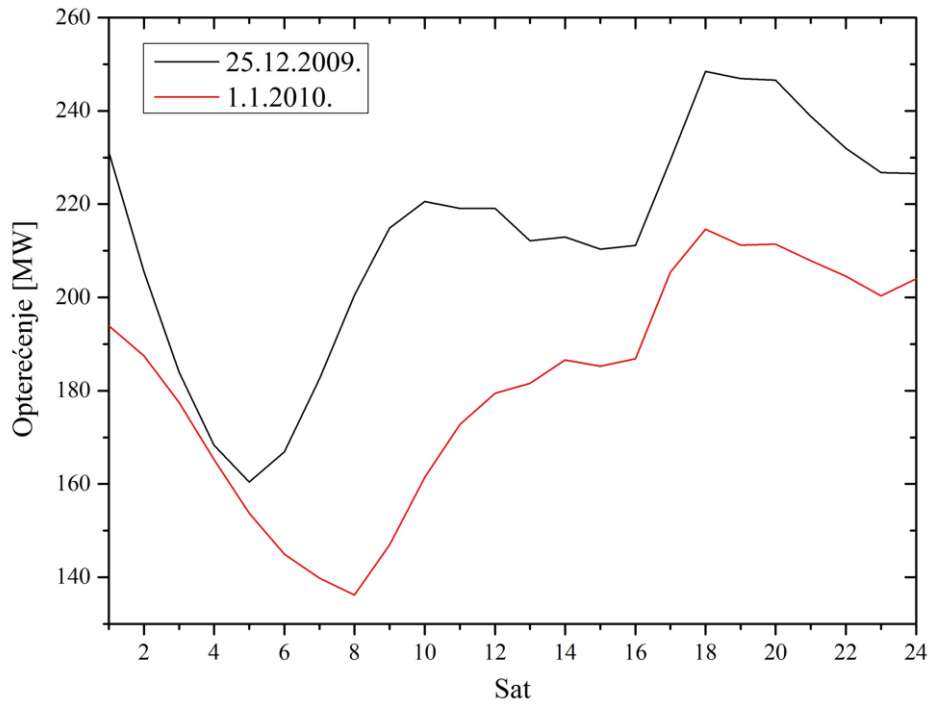
Na Slici 5.3 su prikazane krive dnevnog električnog opterećenja za jednu sedmicu u zimskom periodu godine. Kriva dnevnog električnog opterećenja je oblikovana uticajem

dnevnih ljudskih aktivnosti. Kriva ima dva vršna opterećenja, prvo je između 9 i 10 časova a drugo je između 19 i 20 časova, i to su periodi u toku dana kada ljudske aktivnosti dostižu vrhunac. Najniže opterećenje je oko 5 časova kada su ljudi najmanje aktivni, posle čega sledi početak dnevnih aktivnosti posle buđenja. Kao i sa krive opterećenja u toku sedmice, i ovde se može primetiti povećano opterećenje tokom ponedeljka i utorka i opadanje opterećenja u danima koji slede, sa najmanjim opterećenjem u danima vikenda.

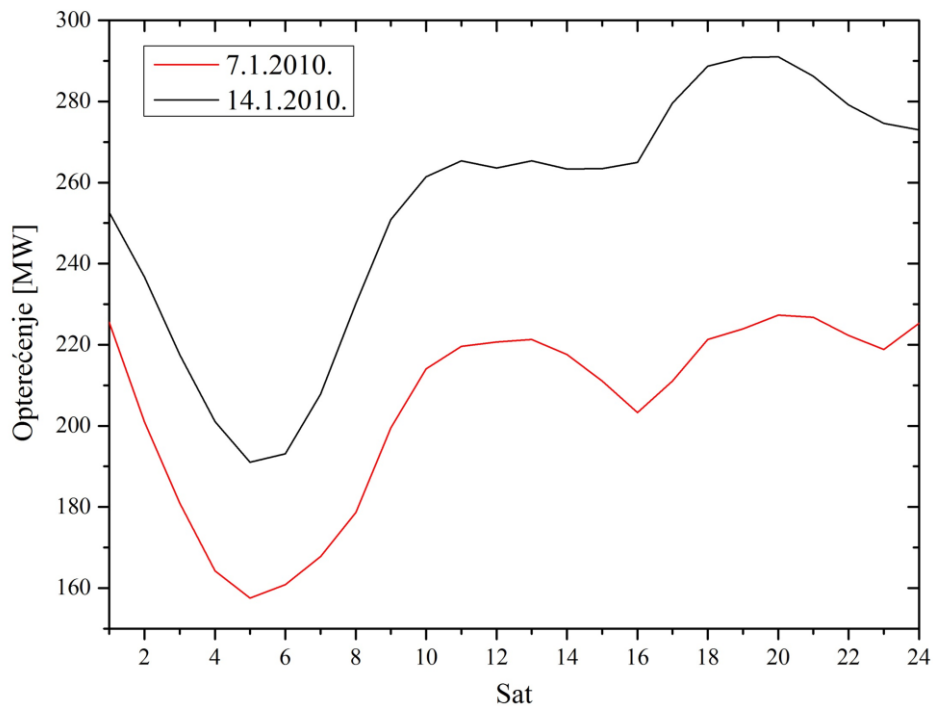


Slika 5.3 Dnevno električno opterećenje

Krive električnog opterećenja u danima praznika prikazane su na Slici 5.4. Prikazana su opterećenja prvog dana Nove Godine (Slika 5.4 a) i na dan Božića (Slika 5.4 b). Kako bi se stekao uvid kako se kreće opterećenje u danima praznika, prikazana su i opterećenja prvog istog susednog dana. Kako se 1.1.2010. pada u petak, dato je i opterećenje za prvi prethodni petak 25.12.2009., i slično kod Božića gde je prikazano opterećenje za prvi sledeći četvrtak 14.1.2010. Prikazane krive opterećenja ukazuju da je opterećenje u danima praznika manje nego uobičajeno što je i očekivano jer su to dani odmora kada ljudi nemaju svakodnevnu obavezu odlaska na posao. Jasno je da su dani praznika posebni, međutim modelovanje ovih dana kod problema prognoze električne potrošnje je problem koji zahteva posebnu pažnju jer je ovih dana u toku godine malo (svega nekoliko), pa model teško može da „uhvati“ promene opterećenja jer ima malo trening primera. U ovoj disertaciji, dani praznika biće tretirani kao obični dani.



a) opterećenje prvog dana Nove Godine



b) opterećenje na Božić

Slika 5.4 Kriva električnog opterećenja u danima praznika

6 MODELI ZA KRATKOROČNU PROGNOZU POTROŠNJE ELEKTRIČNE ENERGIJE

U ovom poglavlju su predstavljeni predloženi modeli za kratkoročnu prognozu potrošnje električne energije. Predložena su tri modela, model sa prognozom u dve faze, model za izborom atributa i model sa dopunjavanjem trening skupa. Date su osnovne ideje na kojima se baziraju predloženi modeli i detaljno prikazani njihovi algoritmi. Kod svih predloženih modela biće korišćen LS-SVM metod za formiranje nelinearnog regresionog modela sa napomenom da predloženi modeli ne zavise od korišćene metode mašinskog učenja za formiranje regresionog modela.

Pretpostavimo da je potrebno formirati inicijalni trening skup na osnovu vremenske serije električne potrošnje definisane u (6.1):

$$p = \{p_1, p_2, \dots, p_{n-1}, p_n, p_{n+1}, \dots, p_i, \dots, p_N\}, i = 1, \dots, N, \quad (6.1)$$

gde p_i predstavlja časovnu električnu potrošnju u i -tom trenutku a N ukupan broj elemenata tj. ukupan broj očitavanja časovne potrošnje. Ukoliko pored atributa vremenske serije uključimo i dodatne attribute kao što su dan u nedelji (d), sat u danu (s), temperaturu (T) itd., onda je inicijalni trening skup $S = \{X, P\}$ formiran na osnovu zadate vremenske serije p i skupa dodatnih atributa, definisan sa:

$$X = \begin{bmatrix} d_n & s_n & \cdots & T_n & p_n & p_{n-1} & \cdots & p_1 \\ d_{n+1} & s_{n+1} & \cdots & T_{n+1} & p_{n+1} & p_n & \cdots & p_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{N-1} & s_{N-1} & \cdots & T_{N-1} & p_{N-1} & p_{N-2} & \cdots & p_{N-n} \end{bmatrix}_{(N-n) \times (n+S_a)} \quad (6.2)$$

$$P = \begin{bmatrix} p_{n+1} \\ p_{n+2} \\ \vdots \\ p_N \end{bmatrix}_{(N-n) \times 1} \quad (6.3)$$

U (6.2) svaka vrsta matrice X predstavlja jedan trening vektor $x_k \in R^{n+S_a}$ gde je S_a broj dodatnih atributa (pored vremenske serije), dok u (6.3) svaki element vektora P predstavlja njemu pridruženu ciljnu vrednost $p_k \in R, k = 1, \dots, (N-n)$. Sa $(N-n)$ je označena veličina inicijalnog trening skupa, odnosno ukupan broj ulazno-izlaznih parova iz X i P , a sa $n+S_a$ ukupan broj atributa, gde n predstavlja veličinu regresora vremenske serije.

Pretpostavimo da želimo da dobijemo prognozu za H koraka unapred, odnosno $p_{N+1}, p_{N+2}, \dots, p_{N+H}$, gde H predstavlja veličinu horizonta prognoze. Za postizanje tog cilja rekurzivna strategija predviđanja, predstavlja najintuitivniji i najkorišćeniji pristup. Njena osnovna karakteristika je da koristi prognozu iz prethodnih koraka, u cilju prognoze narednih koraka, umesto tačnih vrednosti koje u datom trenutku nisu na raspolaganju. U skladu sa tim, trenira se jedan predikcioni model, koji prvo vrši prognozu za jedan korak unapred:

$$p_{N+1} = f(p_N, p_{N-1}, \dots, p_{N-n+1}). \quad (6.4)$$

Zatim se za prognozu narednog koraka, koristi isti model:

$$p_{N+2} = f(p_{N+1}, p_{N1}, \dots, p_{N-n+2}). \quad (6.5)$$

Predviđanje H -tog koraka sledi na osnovu:

$$p_{N+H} = f(p_{N+H-1}, p_{N+H-2}, \dots, p_{N+H-n}). \quad (6.6)$$

Glavni nedostaci rekurzivne strategije prognoze se ogledaju u propagaciji i akumulaciji grešaka kroz korake prognoze. Ovo naročito dolazi do izražaja sa povećanjem veličine horizonta prognoze H . Ukoliko je veličina regresora n veća od H , onda imamo $n-H$ poznatih podataka u regresoru za prognozu H -tog koraka. S druge strane, ukoliko H premaši n , svi podaci u regresoru će biti prognozirane vrednosti, što negativno utiče na kvalitet narednih prognoza. Ipak u praksi se pokazalo da je rekurzivna strategija prognoze dominantna u

pogledu preciznosti i uspešno se primenjuje kod mnogih predikcionih problema. Svi predloženi predikcioni modeli u ovom poglavlju se baziraju na rekurzivnoj strategiji prognoze.

6.1 Model kratkoročne prognoze potrošnje električne energije sa prognozom u dve faze

U Poglavlju 5 je data kratka analiza promenljivih koje oblikuju potrošnju električne energije i koje su zbog svog uticaja na potrošnju izabrane za attribute kod formiranja regresionog modela mašinskog učenja. Već je spomenuto da su atributi koji se koriste za formiranje modela veoma bitni i da se korišćenjem atributa koji su u bliskoj zavisnosti sa potrošnjom mogu formirati precizniji modeli. Sa druge strane, ako se u model uključe atributi čiji uticaj na potrošnju nije značajan, tada se modelu dodaje šum čime se ne doprinosi kvalitetu modela, odnosno preciznost modela se čak može i smanjiti. Pretpostavka je da se model prognoze potrošnje električne energije može unaprediti dodavanjem atributa koji značajno utiču na oblikovanje krive potrošnje. Skup atributa koji se koriste kod formiranja modela za kratkoročnu prognozu potrošnje električne energije je prilično univerzalan a određen pre svega ekspertskom analizom. Skup atributa koji se koriste kod formiranja modela za kratkoročnu prognozu potrošnje električne energije najčešće čine sledeći atributi: atributi vremenske serije čiji broj zavisi od veličine izabranog regresora koji se sa druge strane bira na osnovu horizonta prognoze (kod prognoze časovne potrošnje za dan unapred, najčešće se uzima za regresor $r=24$) i dodatni atributi. U dodatne attribute spadaju atributi koji nisu izvedeni iz vremenske serije potrošnje i tu se pre svega misli na: period u godini (godišnje doba, mesec, nedelja u godini), dan u nedelji, sat u danu, indikator paznika, a od klimatskih najčešće dnevna temperatura (vršna, srednja ili minimalna), a ređe brzina vetra, vlažnost vazduha i oblačnost.

Ako pretpostavimo da će se dodavanjem novog atributa poboljšati preciznost modela, postavlja se pitanje da li je i na koj način moguće uključiti neki novi atribut kako bi se modelu pružile dodatne informacije? Odgovor na ovo pitanje može se naći u nekim drugim oblastima u kojima se koriste metode mašinskog učenja. Tako se kod problema prognoze promene trenda kretanja berze [134], [135], koji u stvari predstavlja problem prognoze vremenske serije, koriste izvedeni atributi iz vremenske serije kao što su npr. logaritamski prinos i eksponencijalni pokretni proseki. Kod problema prepoznavanja slika [136] primena

metoda mašinskog učenja danas je standard a da bi ove metode dale što bolje rezultate koriste se razni izvedeni atributi kao npr. jednostavna detekcija ivica.

Vodeći se osnovnom idejom da se uključivanjem novog atributa u model može formirati precizniji model, i primerima iz drugih oblasti gde je korišćenje izvedenih atributa doprinelo poboljšanju modela, ovde je predložen model [137] u koji će se pored atributa vremenske serije i dodatnih atributa uključiti i jedan izveden atribut. Kako model generiše prognozu časovne potrošnje električne energije za dan unapred, od interesa je u model uključiti neku od sledećih promenljivih: srednja, vršna, minimalna vrednosti ili eksponencijalni pokretni prosek električne potrošnje (opterećenja) za dan unapred tj. dan za koji se radi prognoza. Bitno je naglasiti da bez obzira koji se izvedeni atribut uključi u model, koristi se njegova vrednost za dan unapred. Jasno je da ta vrednost u trenutku prognoze nije dostupna jer se radi o budućem događaju. Međutim ovu vrednost je moguće prognozirati sličnim pristupkom kao i kod prognoze časovne potrošnje i tako je uključiti u model u sledećem koraku prognoze gde je cilj prognoza potrošnje po satima. Kako se kod predloženog pristupa radi prvo prognoza vrednosti izvedenog atributa koja se zatim uključuje u model prognoze potrošnje po satima, predloženi model je nazvan model prognoze potrošnje u dve faze. Ovakav naziv je prikladan i zbog toga što ovakvim pristupom dobijamo prvo vrednost srednje, vršne ili minimale dnevne potrošnje za dan unapred što je takođe vrednost koja je od interesa dispečeru, pa zatim i vrednosti potrošnje po satima. Kod predloženog modela u disertaciji za dodatni tj. izvedeni atribut je uzeto vršno dnevno opterećenje. Proces prognoze potrošnje električne energije u dve faze prikazan je kroz Proceduru 6.1.

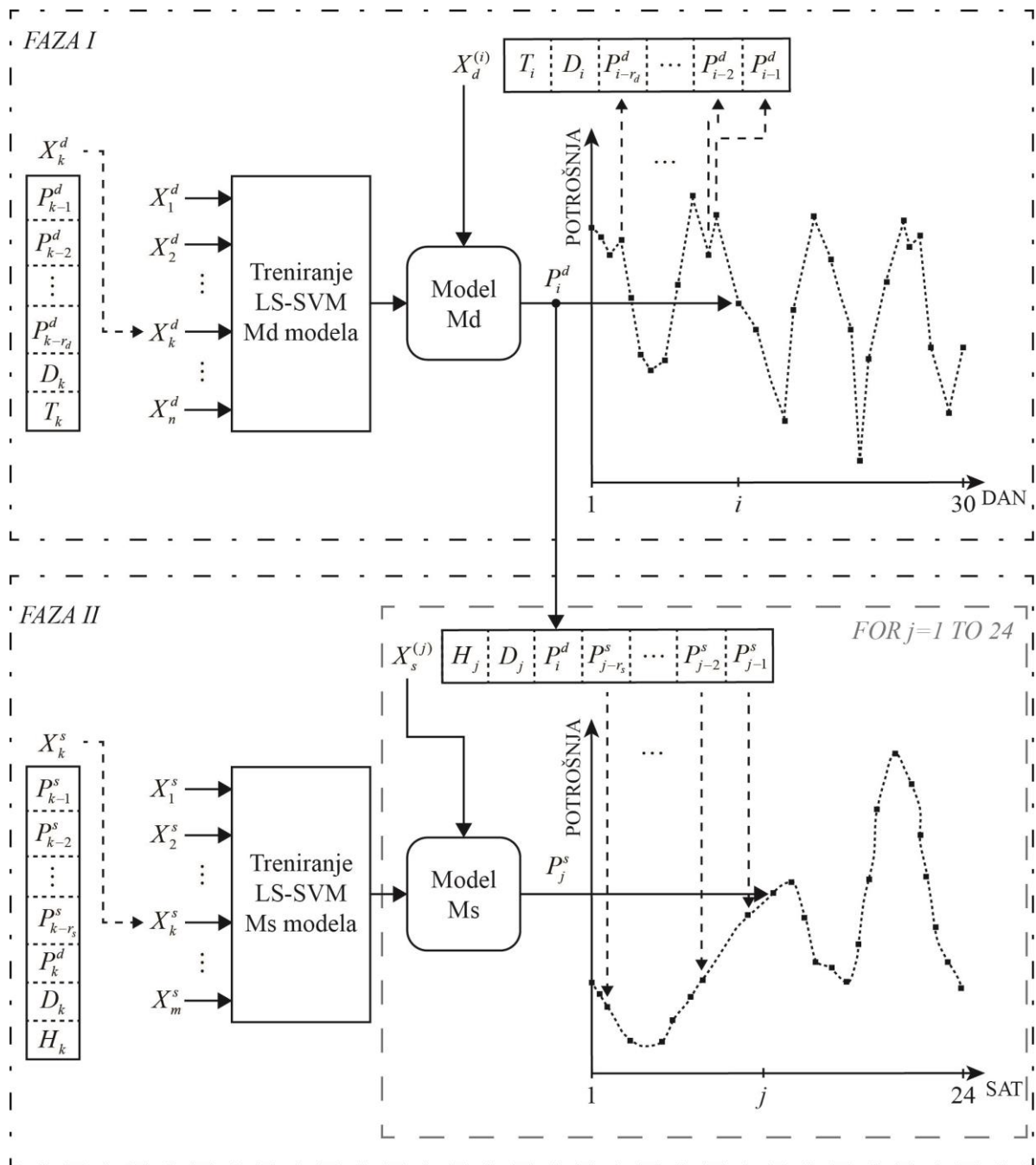
Procedura 6.1 Prognoza potrošnje električne energije u dve faze

1. Na osnovu izabranih struktura modela, formirati trening skupove $S_d = \{X_d, P_d\}$ i $S_s = \{X_s, P_s\}$ od vremenske serije potrošnje električne energije (6.1) i dodatnih atributa (d, s, T, \dots), u skladu sa izabranom veličinom regresora r_d i r_s i izrazima (6.2) i (6.3). Indeksi d i s se odnose na model prognoze izvedenog atributa u prvoj fazi prognoze M_d i model prognoze časovne potrošnje u drugoj fazi M_s , respektivno.
2. Trenirati model M_d koristeći formiran trening skup S_d .
3. Formirati test vektor $X_d^{(i)}$ u skladu sa izabranom strukturom modela M_d i generisati prognozu izvedenog atributa P_i^d .
4. Trenirati model M_s koristeći formiran trening skup S_s .

-
5. Formirati test vektor za prvi korak $X_s^{(j=1)}$ u skladu sa izabranom strukturom modela Ms i ubacivanjem prognozirane vrednosti izvedenog atributa P_i^d iz koraka 3. Generisati prognozu za prvi sat potrošnje $P_{j=1}^s$.
 6. Ažurirati vektor $X_s^{(j)}$ pomeranjem za jedno mesto, i dodavanjem prognoze iz predhodnog koraka, u skladu sa rekurzivnom strategijom predviđanja.
 7. Generisati prognozu potrošnje P_j^s za tekući korak na osnovu novog test vektora $X_s^{(j)}$ i modela Ms.
 8. Ponavljati korake **6** i **7** $H-1$ puta, gde H predstavlja veličinu horizonta prognoze ($H=24$ kod prognoze potrošnje po satu za dan u napred).

Dakle prvo se definišu strukture modela za prognozu izvedenog atributa Md i modela za kratkoročnu prognozu potrošnje električne energije Ms. Radi se o dva međusobno nezavisna modela gde se za svaki model može koristiti neka od regresionih metoda mašinskog učenja. Zatim se formiraju trening skupovi S_d i S_s u skladu sa definisanim strukturama modela a na osnovu vremenske serije potrošnje električne energije, dodatnih atributa (d, s, T, \dots) i izabranim veličinama regresora r_d i r_s . Sa formiranim trening skupom S_d trenirati model Md, formirati test vektor $X_d^{(i)}$ u skladu sa izabranom strukturom modela Md i generisati prognozu izvedenog atributa P_i^d . Prognozom izvedenog atributa završava se proces u prvoj fazi, a ova prognozirana vrednost se koristi kao ulaz kod modela u drugoj fazi.

U drugoj fazi se prvo na osnovu formiranog trening skupa S_s trenira model Ms. Zatim se formira test vektor $X_s^{(j)}$ u skladu sa izabranom strukturom modela Ms i ubacivanjem prognozirane vrednosti izvedenog atributa iz prve faze. Generiše se prognozirana vrednost potrošnje za prvi korak $P_j^s, j=1$. Dalje se ovaj proces ponavlja za preostala $H-1$ koraka a u skladu sa rekurzivnom strategijom prognoze. Formira se novi test vektor pomeranjem atributa vremenske serije i ubacivanjem prognozirane vrednosti iz predhodnog koraka a zatim se generiše prognozirana vrednost potrošnje za tekući korak. U našem slučaju, horizont prognoze je $H=24$ jer se radi prognoza električne potrošnje po satima za dan unapred. Veličine regresora r_d i r_s zavise od izabrane strukture modela za prvu, odnosno drugu fazu prognoze. Za model prognoze srednje, vršne ili minimalne električne potrošnje za dan unapred, obično se za regresor uzima $r_d=7$, a kod modela za prognozu električne potrošnje po satima $r_s=24$. Prikaz opisanog procesa datog kroz Proceduru 6.1 dat je na Slici 6.1.



Slika 6.1 Proces prognoze potrošnje električne energije u dve faze

6.2 Model kratkoročne prognoze potrošnje električne energije sa izborom atributa

Ukoliko se struktura modela izabere tako da sadrži irelevantne ili redundantne attribute, formiran model najverovatnije neće biti optimalan u pogledu očekivane preciznosti. Kada je broj atributa mali, model nema dovoljno informacija o problemu i može se javiti *underfitting*. Sa druge strane, ukoliko je broj atributa prevelik, moguće je da među njima ima redundantnih i

nepotrebnih pa se može javiti *overfitting*. U oba slučaja se ne dobija optimalan model pa se zato pribegava strategiji izbora atributa pre svega sa ciljem povećanja preciznosti modela. Međutim ovde treba napomenuti da primena strategije izbora atributa ne znači da će se nužno poboljšati preciznost inicijalnog modela, naročito u slučaju kada se inicijalni model formira uz pomoć dobre ekspertske analize konkretnog problema. Ipak, ovaj model može poslužiti za procenu da li je izabrana struktura modela dobra, odnosno da li je ekspertska analiza atributa urađena kako treba.

Kada je reč o metodama izbora atributa, razlikujemo dve osnovne metode: filter metode [138] (*eng. Filter method*) i wrapper metode [139] (*eng. Wrapper method*). Filter metode na osnovu raspoloživog skupa podataka biraju optimalan podskup ulaznih atributa. Izbor se vrši na osnovu nekog izabranog kriterijuma koji meri povezanost svakog podskupa ulaznih atributa sa pridruženim ciljnim vrednostima. Kod Wrapper metoda, kriterijum za izbor podskupa ulaznih atributa direktno je određen algoritmom za obučavanje modela. Tako je sama metoda u stvari deo optimizacione funkcije modela.

Pošto su uključene u samu optimizacionu funkciju modela, Wrapper metode su zahtevne sa stanovišta procesorskog vremena izvršenja. Njihova prednost se ogleda u sigurnom nalaženju optimalnog skupa atributa u kombinaciji sa korišćenom regresionom metodom. Sa druge strane, filter metode nisu deo optimizacione funkcije modela pa za pronalazak optimalnog podskupa atributa koriste kriterijum koji se ne poklapa sa optimizacionim kriterijumom korišćenog regresionog metoda. Stoga ovaj metod ne garantuje da će izabrani podskup atributa davati optimalne rezultate u kombinaciji sa izabranim regresionim metodom. Filter metode su generalno brže i preporučuju se u situacijama kada se koristi složen regresioni model.

Opisane metode izbora atributa se odnose na način kako se koristi kriterijum izbora, da li je on vezan za samu regresionu metodu kao kod wrapper metoda gde je izbor atributa nerazdvojni deo regresione metode ili se koristi nezavisno od nje kao kod filter metoda. Međutim bez obzira koji se pristup koristi, izbor atributa se svodi na pretragu u okviru inicijalnog skupa atributa. Kod optimalne strategije izbora treba ispitati svih $2^d - 1$ kombinacija ulaznih atributa, gde je d ukupan broj ulaznih atributa i izabrati podskup koji daje najbolje rezultate u skladu sa kriterijumom izbora. Ovakva strategija izbora atributa naziva se potpuna strategija izbora (*eng. exhaustive search*) i ona garantuje pronalaženje optimalnog podskupa atributa ali praktična primena ove strategije je ograničena jer sa porastom broja atributa vreme izbora raste eksponenijalno.

Situacije kada je broj atributa veliki pa je potpuna strategija izbora neprimenljiva, mogu se prevazići primenom strategije izbora atributa sa pretragom u napred (*eng. Forward selection*) ili pretragom unazad (*eng. Backward selection*). Ove strategije spadaju u grupu “*greedy*“ algoritama [140] kod kojih se u svakoj fazi bira lokalni minimum sa ciljem nalaženja globalnog minimuma. Kod pretrage unapred, polazi se od praznog skupa ulaznih atributa S pa se u svakom koraku pridodaje najbolji raspoloživi atribut tj. onaj čije dodavanje maksimizuje kriterijum izbora između ostalih atributa i ciljnih vrednosti. Atributi se dodaju sve dok veličina skupa ne dostigne broj atributa. Rezultat izbora je onaj podskup atributa skupa S za koji se ima optimalna vrednost kriterijuma izbora. Slično, kod pretrage unazad se počinje od punog skupa ulaznih atributa S a zatim se kroz korake najmanje bitni atributi odbacuju tako da se maksimizuje kriterijum izbora između ostalih atributa i ciljnih vrednosti, sve dok veličina skupa S ne dostigne vrednost 1. Rezultat izbora je onaj podskup atributa skupa S za koji se ima optimalna vrednost kriterijuma izbora. U obe strategije se ispituje ukupno $\frac{d(d+1)}{2}$ različitih kombinacija podskupova ulaznih atributa, te stoga izbor optimalnog podskupa atributa nije zagarantovan. Nedostatak jednosmerne pretrage kod strategija pretrage u napred i pretrage unazad može se nadomestiti primenom strategije pretrage napred-nazad (*eng. forward-backward*). Kod ove strategije pretrage, ponovo se razmatraju atributi koji su prethodno odbačeni i koji su prethodno pridodati. Pretraga se može započeti od bilo koje konfiguracije ulaznog skupa atributa tj. može se početi od praznog, punog ili na slučajan način inicijalizovanog ulaznog skupa. Ovde takođe nema garancije da će biti izabran optimalan podskup atributa, a izbor atributa zavisi od početne inicijalizacije polaznog skupa.

Primarni cilj modela kratkoročne prognoze potrošnje električne energije sa izborom atributa jeste povećanje preciznosti modela a ne smanjenje dimenzije prostora atributa jer je broj atributa kod problema prognoze potrošnje relativno mali. Iako su atributi koji se koriste izabrani dobrom ekspertskom analizom, potencijalno postoji prostor za poboljšanje modela njihovim reizborom a može se i proveriti urađena ekspertska analiza. Kako je broj atributa $d \approx 30$, imamo $\approx 2^{30}$ kombinacija atributa. Ovo nije mali broj kombinacija i zahteva veliko procesorsko vreme ukoliko se koristi potpuna strategija izbora, pa je predložen izbor atributa sa strategijom pretrage napred-nazad kako bi se prevazišao ovaj problem. Kompletan proces prognoze potrošnje električne energije sa izborom atributa koristeći pretragu napred-nazad, dat je kroz Proceduru 6.2.

Procedura 6.2 Prognoza potrošnje električne energije sa izborom atributa

1. Formirati inicijalni trening skup od vremenske serije potrošnje električne energije (6.1) i dodatnih atributa (d, s, T, \dots), u skladu sa izabranom veličinom regresora r i izrazima (6.2) i (6.3).
2. Inicijalizovati skup S sa izabranim atributima i skup F sa neizabranim atributima tj. atributima koji se ne nalaze u S . Skup S može biti inicijalno prazan, pun ili nasumično popunjen.
3. Izračunati $KriterijumIzbora(S, Y)$.

4. Naći:

$$X_s = \arg \max_{X_{i,j}} \left\{ \text{KriterijumIzbora}(\{S, X_j\}, Y) \right\} \cup \left\{ \text{KriterijumIzbora}(S \setminus X_i, Y) \right\}, X_i \in S, X_j \in F$$

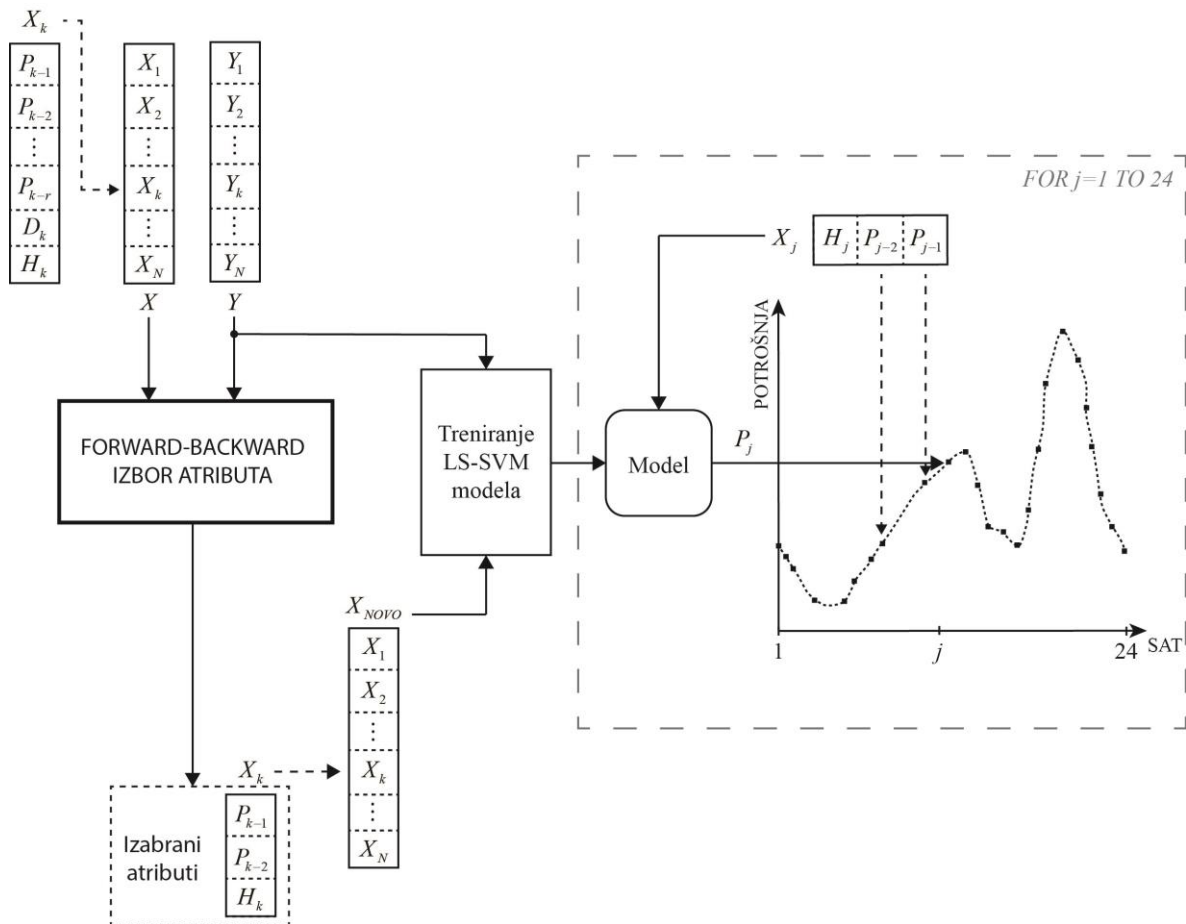
Formirati novi skup izabranih atributa S' , izbacivanjem ili dodavanjem atributa X_s .

5. Izračinati $KriterijumIzbora(S', Y)$.
6. Ukoliko je $KriterijumIzbora(S, Y) > KriterijumIzbora(S', Y)$, preći na korak 7. U suprotnom, ažurirati skup S dodavanjem ili odbacivanjem atributa X_s ($S=S'$) i vratiti se na korak 4.
7. Rezultat izbora atributa je skup S .
8. Na osnovu izabrane kombinacije atributa S i inicijalnog trening skupa, formirati novi trening skup i trenirati model M .
9. Formirati test vektor x_t u skladu sa izabranom kombinacijom atributa S i rekurzivnom strategijom prognoze.
10. Generisati prognozu za k -ti sat potrošnje $p_t(t=k)$, $k=1, \dots, H$ (H – horizont prognoze).
11. Ponoviti korake **9** i **10** H puta kako bi se generisala prognoza za svaki sat.

Proces izbora atributa se u suštini svodi na izbor optimalne strukture modela pa se sam proces započinje definisanjem inicijalnog skupa atributa tj. inicijalne strukture modela. Na osnovu inicijalnog skupa atributa, formira se skup S , koji predstavlja skup izabranih atributa i skup F , koji predstavlja skup odbačenih atributa. Skup S se može inicijalizovati tako da bude prazan, pun ili nasumično popunjen a u skladu sa tim, skup F će sadržati one atribute koji nisu u S . Po inicijalizaciji skupova S i F , računa se $KriterijumIzbora(S, Y)$, gde je Y ciljna promenljiva a za kriterijum izbora je uzeta procena zajedničkih informacija.

Za svaki atribut X_j iz skupa F , računa se $KriterijumIzbora(\{S, X_j\}, Y)$ tj. procena zajedničkih informacija između ciljne vrednosti Y i skupa izabranih atributa iz S sa pridodatim atributom X_j iz F . Takođe za svaki atribut X_i iz skupa S , računa se $KriterijumIzbora(\{S \setminus X_i\}, Y)$ tj. procena zajedničkih informacija između ciljne vrednosti Y i skupa izabranih atributa iz S bez jednog atributa X_i iz S . Uzima se kombinacija sa najvećom vrednošću kriterijuma izbora tj. procene zajedničkih informacija i ona predstavlja novi skup izabranih atributa S' . Skup S' je ustvari skup S kome je u skladu sa prethodnom procedurom, ili dodat novi atribut iz F ili izbačen jedan atribut iz samog skupa S .

Sada se porede kriterijumi izbora između ciljne vrednosti Y i skupova S i S' . Ukoliko je $KriterijumIzbora(S, Y) > KriterijumIzbora(S', Y)$ to znači da novo formirani skup atributa S' nije optimalan u smislu kriterijuma izbora, u našem slučaju procena zajedničkih informacija, pa je optimalni skup atributa skup S a izbor atributa se prekida. Međutim ukoliko to nije slučaj pa je $KriterijumIzbora(S', Y) > KriterijumIzbora(S, Y)$, za novi skup izabranih atributa se stavlja $S' (S=S')$ i izbor se nastavlja prema prethodno spomenutoj proceduri.



Slika 6.2 Proces prognoze potrošnje električne energije sa izborom atributa

Na osnovu optimalnog skupa atributa S , formirati novi trening skup i trenirati model M . Formirati test vektor x_t u skladu sa izabranom kombinacijom atributa S i rekurzivnom strategijom prognoze i generisati prognozu za k -ti sat potrošnje $p_i(t=k)$, $k=1, \dots, H$. Ponoviti postupak H puta kako bi se generisala prognoza za svaki sat. Prikaz opisanog procesa datog kroz Proceduru 6.2 dat je na Slici 6.2.

6.3 Model kratkoročne prognoze potrošnje električne energije sa dopunom trening skupa

Optimalan model baziran na metodama mašinskog učenja, karakterišu optimalan skup atributa, optimalan trening skup i optimalani parametri modela. Optimalan skup atributa se može definisati dobrom ekspertskom analizom problema ili nekom metodom za izbor atributa kao što je slučaj kod modela u Poglavlju 6.2. Optimalni parametri modela u slučaju LS-SVM metode, mogu se naći jednostavnom ali uspešnom *grid-search* pretragom u kombinaciji sa *cross-validation* metodom kao što je opisano Poglavlju 4.4.4. Tako ostaje problem formiranja optimalnog trening skupa čija je struktura trening vektora definisana prethodno formiranim skupom atributa.

Postoje razni metodi za izbor trening skupa koji koriste razne kriterijume za izbor. Ti kriterijumi se baziraju na meri sličnosti između tekućeg test vektora i trening vektora. Za meru sličnosti može da se koristi procena zajedničkih informacija [141], [142] ili korelacioni koeficijent a može i jednostavna podudarnost po nekom parametru, godišnje doba, mesec, dan, sat, dan u nedelji, temperaturni opseg itd. Kod problema prognoze potrošnje električne energije najčešće nije potrebno koristiti posebne metode za izbor trening skupa već je dovoljno izabrati trening skup prema periodu godine. Treniranje modela trening skupom koji se odnosi na veliki vremenski period daje modelu veliku količinu informacija pa je i za očekivati da će preciznost takvog modela biti bolja nego u slučaju treniranja sa malim trening skupom. Međutim kod kratkoročne prognoze potrošnje električne energije radi se o časovnoj potrošnji, što znači da trening skup za period od godinu dana ima oko 8760 trening vektora (365x24). Kako složenost LS-SVM modela raste eksponencijalno sa porastom trening primera, tako raste i vreme potrebno za formiranje modela, pa je model sa trening skupom većim od godinu dana nepraktičan. Imajući ovo u vidu, u disertaciji su svi predloženi modeli trenirani sa trening skupom za preiod od godinu dana.

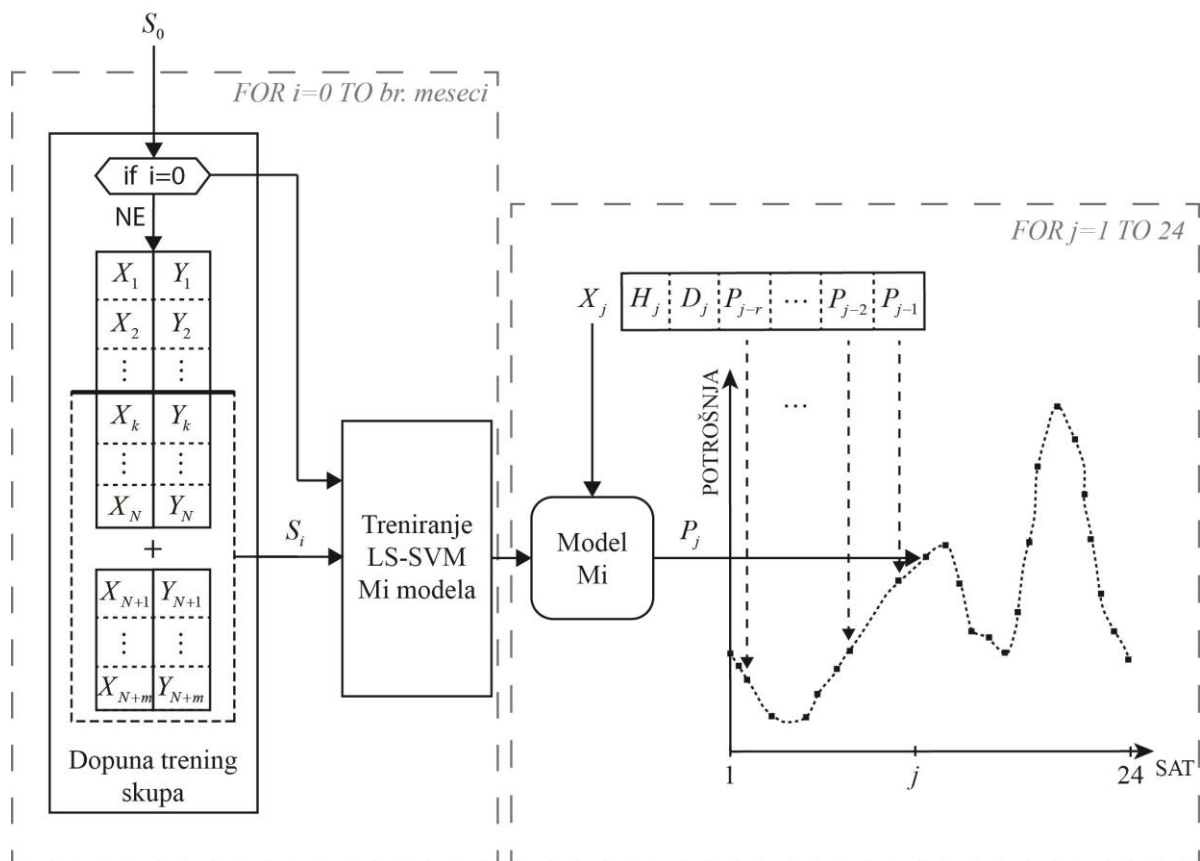
Međutim, kako je potrošnja električne energije nestacionaran proces koji stalno prate neke promene, postavlja se pitanje, za koliki period je validan model treniran trening skupom za prethodnu godinu? Na primer, ako smo trenirali model sa podacima za 2006. godinu a prognozira se 2009., model nema informacije o čitave dve godine, pa je i za očekivati da će preciznost modela biti manja nego u slučaju da se prognoza radi modelom treniranim sa podacima za 2008. Kako bi se prevazišao ovaj problem i modelu omogućio priliv novih informacija, predložen je model sa dopunom trening skupa [143] iz meseca u mesec, tako što se trening vektori za poslednji mesec odbacuju a trening skup se osvežava novim trening vektorima za mesec koji je upravo prošao. Na ovaj način, model se stalno osvežava prilivom novih informacija a stari, manje validni trening vektori se izbacuju. Predloženi model dat je kroz Proceduru 6.3.

Procedura 6.3 Prognoza potrošnje električne energije sa dopunom trening skupa

1. Formirati inicijalni trening skup S_0 od vremenske serije potrošnje električne energije (6.1) i dodatnih atributa (d, s, T, \dots), u skladu sa izabranom veličinom regresora n i izrazima (6.2) i (6.3). $S_i = S_0$.
2. Trenirati model M_i koristeći formiran trening skup S_i .
3. Formirati test vektor za prvi korak $X^{j=1}$ u skladu sa izabranom strukturom modela M_i . Generisati prognozu za prvi sat potrošnje $P_{j=1}$.
4. Ažurirati vektor $X^{(j)}$ pomeranjem za jedno mesto, i dodavanjem prognoze iz predhodnog koraka, u skladu sa rekurzivnom strategijom prognoze.
5. Generisati prognozu potrošnje P_j za tekući korak na osnovu novog test vektora $X^{(j)}$ i modela M_i .
6. Ponavljati korake **5** i **6**, $H-1$ puta, gde H predstavlja veličinu horizonta prognoze i generisati dnevnu časovnu potrošnju.
7. Generisati prognozu svakog dana tekućeg meseca ponavljajući korake od **3** do **6**.
8. Formirati novi trening skup S_i , dopunom trening skupa novim trening vektorima za upravo prognozirani mesec i izbacivanjem trening vektora za prvi mesec.
9. Vratiti se na korak 2.

Proces prognoze potrošnje električne energije dat kroz Proceduru 6.3 i prikazan na Slici 6.3, započinje se formiranjem inicijalnog trening skupa S_0 . Inicijalnim trening skupom se trenira model M_0 a zatim se za svaki dan u mesecu generiše prognoza potrošnje po satima.

Prognoza po satima se započinje formiranjem test vektora za prvi korak prognoze u skladu sa izabranom strukturom modela. Na osnovu ovog test vektora, generiše se prognoza za prvi sat i ažurira se test vektor u skladu sa rekurzivnom strategijom prognoze. Ovaj postupak se ponavlja $H-1$ puta, gde H predstavlja veličinu horizonta prognoze a u našem slučaju to je 24. Po završetku časovne prognoze potrošnje, formira se novi test vektor za prvi sat u sledećem danu i proces se ponavlja za svaki sledeći dan. Kada se dođe do poslednjeg dana u mesecu (u slučaju kad se dopuna trening skupa vrši na mesec dana), trening skup se dopunjuje tako što se trening vektori za poslednjih mesec dana dodaju trening skupu (kako se ovde radi o časovnoj potrošnji broj trening vektora će u tom slučaju biti br. dana u mesecu pomnožen brojem sati u danu) a trening vektori koji se odnose na prvi mesec se izbacuju. Model se retrenira sa novo formiranim trening skupom S_i i proces prognoze po satima za svaki dan se ponavlja. Interval za dopunu trening skupa može biti kraći ali i duži, u skladu sa potrebom da se isprati promena potrošnje.



Slika 6.3 Proces prognoze potrošnje električne energije sa dopunom trening skupa

7 REZULTATI EKSPERIMENATA

Postavke i rezultati eksperimenata prezentovani su u ovom poglavlju. U prvom delu dat je opis korišćenih podataka za eksperimente, prikazani su formirani modeli, date njihove osnovne karakteristike i predstavljene korišćene metrike za evaluaciju modela. U drugom delu su prikazani rezultati eksperimenata komparativne analize predloženih i komparativnih modela, izvršena je analiza i diskutovano značenje tih rezultata u pogledu uspešnosti predloženih modela u disertaciji, koja se ogleda u ostvarenoj preciznosti modela na test skupu.

7.1 Postavke eksperimenata

Podaci³ za validaciju predloženih modela i komparativnu analizu korišćeni u ovoj disertaciji se odnose na časovno električno opterećenje grada Niša u Srbiji koji ima oko 260000 stanovnika sa površinom od 597 km². Grad Niš se nalazi na nadmorskoj visini od 195m sa prosečnom godišnjom temperaturom od 11.2 C° i prosečnim godišnjim dnevnim opterećenjem od 182.2 MW. Korišćeni podaci se odnose na period 2008. i 2009. godine, pri čemu su za testiranje korišćeni podaci za 2009. godinu a za treniranje podaci za 2008. godinu, osim ako nije drugačije napomenuto za konkretni model.

Za rešavanje problema kratkoročne prognoze potrošnje električne energije u disertaciji su predložena tri modela koji koriste LS-SVM metodu mašinskog učenja (veštačke inteligencije) za formiranje regresionog modela. Evaluacija predloženih modela rađena je komparativnom

³ Korišćeni podaci u ovoj disertaciji su vlasništvo privrednog društva „Jugoistok“ d.o.o. Niš i neće biti korišćeni u druge svrhe osim u svrhu naučnog istraživanja.

analizom sa pet modela. Za komparativne modele uzeta su dva statistička modela iz grupe klasičnih metoda i tri modela iz grupe metoda veštačke inteligencije. Formirani su sledeći modeli:

- DS-ARIMA – dvostruko sezonski ARIMA model [59],
- DS-HW – dvotruko sezonski Holt-Winters model [13],
- ANN – model veštačkih neuronskih mreža [26],
- RVM – *relevance vector machine* model [123],
- LSSVM – model potpunih vektora sa najmanjim kvadratima [38], [144],
- 2F-LSSVM – LS-SVM model sa prognozom u dve faze [137],
- IA-LSSVM – LS-SVM model sa izborom atributa [92], [141],
- DTS-LSSVM – LS-SVM model sa dopunom trening skupa [143].

Dvostruko sezonski model autoregresionog integrisanog procesa pokretnih proseka (DS-ARIMA) spada u grupu klasičnih modela. Za razliku od standardnog ARIMA modela, DS-ARIMA model uvažava dve sezonske periodičnosti električne potrošnje, dnevnu ($s_1=24$) i sedmičnu ($s_2=168$). Komparativni DS-ARIMA model ima sledeću strukturu, $ARIMA(3,0,3) \times (3,0,3)_{24} \times (3,0,3)_{168}$ pri čemu je prošli period od 20 sedmica korišćen za formiranje modela. Pored ovog modela, formirani su i $ARIMA(2,0,1) \times (2,0,1)_{24} \times (2,0,1)_{168}$ i $ARIMA(1,1,1) \times (1,1,1)_{24} \times (1,1,1)_{168}$ modeli čija je preciznost za nijansu lošija od $ARIMA(3,0,3)$, zbog čega je on uzet kao referentni. Dvostruko sezonski Holt-Winters (DS-HW) model, za razliku od standardnog aditivnog modela eksponencijalnog izravnjanja, uvažava dve sezonske periodičnosti električne potrošnje, dnevnu i sedmičnu. Za formiranje DS-HW modela korišćeni su podaci za prošli period od godinu dana. Kod formiranje modela veštačkih neuronskih mreža (ANN) korišćeni su podaci za period 2008. godine (model je treniran sa 8760 vektora). Struktura ANN modela je sledeća: ulazni sloj sa 26 čvorova (strukturu vektora čine 26 atributa, 24 atributa vremenske serije, regresor $r=24$ tj. vrednosti opterećenja za poslednja 24 sata i 2 dodatna atributa, dan u nedelji i sat u danu), jedan srednji skriveni sloj sa 24 čvora i jednim izlaznim čvorom. RVM model je treniran sa podacima za 2008. godinu sa ukupno 8760 trening vektora sa skupom atributa kao i kod ANN modela. Slično kao i prethodna dva modela, LSSVM model ima istu strukturu atributa a treniran je sa 200 poslednjih dana 2008. godine tj. 4800 vektora.

Svi predloženi modeli koriste LS-SVM⁴ metod za formiranje regresionog modela, pri čemu 2F-LSSVM model koristi LS-SVM metod i kod formiranja modela za prognozu vršnog dnevnog opterećenja u prvoj fazi i kod modela za prognozu časovnog dnevnog opterećenja u drugoj fazi prognoze. U prvoj fazi, formira se model sa strukturom od 7 atributa vremenske serije (regresor $r_1=7$, vrednosti vršnog dnevnog opterećenja za sedam poslednjih dana) i dva dodatna atributa, dan u nedelji i srednja dnevna temperatura. Ovaj model je treniran sa podacima o vršnom dnevnom opterećenju za period poslednjih godinu dana. Strukturu modela u drugoj fazi kod prognoze časovnog opterećenja, čine 27 atributa: 24 atributa vremenske serije (regresor $r_2=24$), dva dodatna atributa (dan u nedelji i čas u danu) i vrednost vršnog dnevnog opterećenja iz prve faze. Ovde treba napomenuti da je vrednost vršnog dnevnog opterećenja u trenutku prognoze nepoznata, pa se zato koristi prognozirana vrednost iz prve faze. Model u drugoj fazi je treniran sa podacima za prošlu 2008. godinu.

Pored spomenutog 2F-LSSVM modela formirana su još dva slična modela sa prognozom u dve faze:

- 2F-LSSVM-P – model sa prognozom u dve faze sa tačnim vršnim dnevnim opterećenjem i
- 2F-LSSVM-T – model sa prognozom u dve faze sa tačnom srednjom dnevnom temperaturom.

Model 2F-LSSVM-P je u stvari model sa prognozom u dve faze koji u drugoj fazi za ulaz iz prve faze koristi tačnu vrednost vršnog dnevnog opterećenja. 2F-LSSVM-T model koristi kao i 2F-LSSVM model, prognoziranu vrednost vršnog dnevnog opterećenja iz prve faze, ali za razliku od 2F-LSSVM modela kod prognoze u prvoj fazi koristi tačnu vrednost srednje dnevne temperature a ne prognoziranu.

Model za izborom atributa IA-LSSVM, koristi procenu zajedničkih informacija za izbor atributa primenom pretrage napred-nazad. Od $2^{26}-1$ mogućih kombinacija atributa (za inicijalni skup atributa je uzet skup od 26 atributa kao kod ostalih modela veštačke inteligencije datih u disertaciji) koje bi trebalo da se ispitaju primenom potpune strategije pretrage, ovde je ispitano 495 kombinacija i pobednička kombinacija atributa je sat u danu i opterećenje prethodnog sata. Tako strukturu modela čine samo dva atributa a model je treniran kao i u prethodnom slučaju, skupom za prethodnu 2008. godinu. Treba napomenuti da je pored inicijalnog skupa od 26 atributa, pretraga rađena i sa inicijalnim skupom od 50

⁴ LSSVM oznaka se odnosi na konkretno formirani komparativni model, dok LS-SVM označava metodu mašinskog učenja, koja se koristi za formiranje regresionog modela.

atributa (uzet je regresor $r=48$) i da je pretragom napred-nazad sa ispitanih 1101 kombinacija izabran isti skup atributa kao u prethodnom slučaju.

Poslednji predloženi DTS-LSSVM model, sličan je modelu LSSVM jer koristi istu strukturu atributa tj. isti skup atributa i treniran je istim brojem trening vektora, skup trening vektora za period poslednjih godinu dana. Međutim sa jednom bitnom razlikom, model DTS-LSSVM se re-trenira na period od mesec dana sa novim trening skupom sa istim brojem trening vektora pri čemu se odbacuju vektori za najdalji mesec a pridodaju vektori za novi mesec koji je upravo prošao.

Kod svih modela koji koriste LS-SVM metod mašinskog učenja, korišćena je *grid-search* pretraga sa *k fold cross* validacijama ($k=10$) za nalaženje optimalnih parametara modela. Takođe nisu korišćene tehnike pretprocesiranja u cilju postizanja stacionarnosti jer za to nije bilo potrebe, kao što su npr. uklanjanje sezonske komponente (*eng. deseasonalizing*) i eliminacija trenda (*eng. detrending*). Međutim kod svih metoda u disertaciji, osim klasičnih metoda DS-ARIMA i DS-HW, pre treniranja modela urađeno je skaliranje atributa na opseg $[0,1]$, kako bi se sprečila dominacija pojedinih atributa i postigla brža konvergencija i bolja stabilnost prilikom rešavanja optimizacionog problema.

Za komparativnu analizu između modela u disertaciji, korišćene su sledeće metrike:

srednja apsolutna procentualna greška (*eng. mean absolute percentage error - MAPE*), definisana sa (7.1):

$$MAPE[\%] = 100 \frac{1}{n} \sum_{i=1}^n \left| \frac{p_i - \hat{p}_i}{p_i} \right|, \quad (7.1)$$

gde p_i i \hat{p}_i predstavljaju tačnu i prognoziranu vrednost opterećenja u i -tom koraku, a n je horizont prognoze. MAPE spada u grupu takozvanih procentnih grešaka (*eng. percentage errors*) i kao takva je „nezavisna od podataka” (*eng. scale independent*) i najčešće se koristi kod evaluacije prognoze vremenskih serija [145]. MAPE greška ima taj hendikep da ima veću vrednost kod pozitivnog nego negativnog odstupanja. Zbog toga je definisana takozvana „simetrična“ MAPE greška [146] (*eng. symmetric mean absolute percentage error - sMAPE*) data sa (7.2):

$$sMAPE[\%] = 200 \frac{1}{n} \sum_{i=1}^n \frac{|p_i - \hat{p}_i|}{p_i + \hat{p}_i} \quad (7.2)$$

Traba napomenuti da „simetrična“ MAPE greška nije baš simetrična kao sto ime kaže jer za istu vrednost opterećenja p_i vrednost izraza $|p_i - \hat{p}_i| / (p_i + \hat{p}_i)$ ima veću vrednost za manju vrednost prognoze nego u slučaju veće prognoze [147].

Pored ovih procentualnih grešaka, korišćena je i srednje kvadratna greška (*eng. root mean square error* - RMSE) definisana sa (7.3):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - \hat{p}_i)^2} \quad (7.3)$$

Za razliku od prethodno spomenutih, RMSE spada u grupu grešaka „zavisnih od podataka“ (*eng. scale-dependent*) i ona je osetljiva na velika odstupanja [148].

Za praćenje preciznosti modela kroz časove, korišćena je apsolutna procentualna greška (*eng. absolute percentage error* - APE) data sa (7.4):

$$APE_i = 100 \frac{|p_i - \hat{p}_i|}{p_i} \quad (7.4)$$

7.2 Rezultati i diskusija

Ovde su prikazani rezultati eksperimenata u kojima je rađena prognoza časovnog opterećenja grada Niša za 2009. godinu. U Tabeli 7.1 date su prosečne godišnje MAPE, sMAPE i RMSE greške za svaki model.

Tabela 7.1 Prosečna godišnja MAPE, sMAPE i RMSE greška

Model	MAPE [%]	sMAPE [%]	RMSE [MW]
DS-ARIMA	4.12	4.14	9.08
DS-HW	5.62	5.64	12.4
ANN	4.13	4.13	9.06
RVM	4.67	4.64	10.95
LSSVM	4.51	4.42	10.18
2F-LSSVM	4.2	4.2	8.94
IA-LSSVM	4.6	4.58	10.07
DTS-LSSVM	3.87	3.86	8.84

Rezultati predstavljeni u Tabeli 7.1 odnose se na MAPE, sMAPE i RMSE greške računane na dnevnom nivou ($n=24$) usrednjene za period cele godine. Kada je u pitanju MAPE greška, poredak modela je takav da ubedljivo prednjači predloženi DTS-LSSVM model sa dopunom trening skupa a zatim ga prate DS-ARIMA i ANN modeli koji imaju skoro identičan rezultat. Zatim slede, model 2F-LSSVM sa prognozom u dve faze, standardan LSSVM model, model za izborom atributa IA-LSSVM, RVM model i poslednji je DS-HW model. Poredak ostaje isti i u slučaju kada se za meru uspešnosti posmatra sMAPE greška, što je i očekivano jer se radi o „simetričnoj“ verziji MAPE greške. Međutim, poredak modela je nešto drugačiji kada se preciznost modela posmatra kroz RMSE grešku. I u ovom slučaju je DTS-LSSVM najprecizniji model pa sledi 2F-LSSVM model koji je bio na četvrtom mestu iza DS-ARIMA i ANN modela, koji su sada iza njega. Slede ih IA-LSSVM model, koji je sada ispred LSSVM modela, a na začelju su RVM i poslednji DS-HW model. Razlog promene redosleda modela leži u činjenici da je RMSE greška za razliku od MAPE i sMAPE „zavisna od podataka“ i da je naročito osetljiva na odstupanja. Bolji rezultat 2F-LSSVM modela od DS-ARIMA i ANN modela u ovom slučaju, je posledica toga da DS-ARIMA i ANN imaju velika odstupanja u pojedinim koracima prognoze (u našem slučaju prognoza u pojedinim satima u danu) iako su globalno precizniji od 2F-LSSVM modela koji nema takva odstupanja. Ista je situacija i sa IA-LSSVM modelom, koji je ispred LSSVM modela.

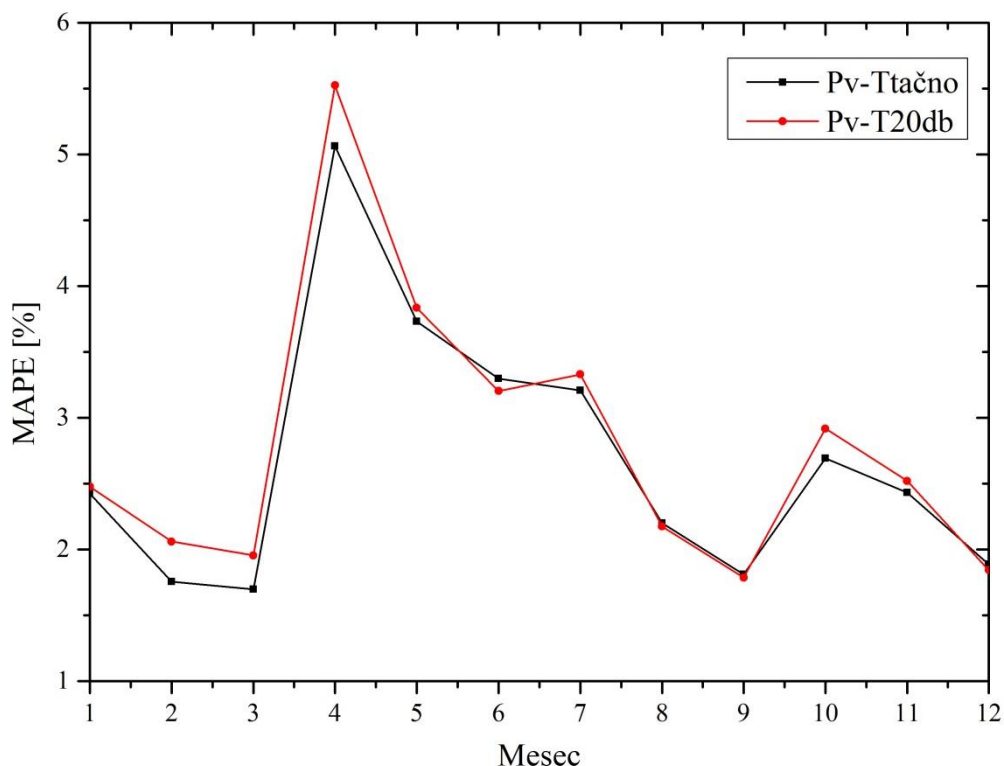
Prosečne godišnje MAPE, sMAPE i RMSE greške za modele sa prognozom u dve faze, date su u Tabeli 7.2. Model 2F-LSSVM-P se odnosi na model sa prognozom u dve faze pri čemu je za rezultat prognoze vršnog dnevnog opterećenja u prvoj fazi, uzeto tačno opterećenje. Kod modela 2F-LSSVM-T se u prvoj fazi prognoza vršnog dnevnog opterećenja radi sa tačnom srednjom dnevnom temperaturom. Međutim, kako ova temperatura nije poznata za dan unapred, a specijalizovani servisi ne čuvaju zapise o prognoziranim vrednostima temperature u prošlosti, kod modela 2F-LSSVM je kod prognoze vršnog dnevnog opterećenja u prvoj fazi korišćena tačna vrednost temperature kojoj je dodat beli šum. Na ovaj način je simulirana prognozirana vrednost dnevne temperature koja nije dostupna. Kolona P_{\max} označava da li model u drugoj fazi koristi tačnu ili prognoziranu vrednost vršne dnevne snage a kolona T_{srednje} da li model u prvoj fazi koristi tačnu ili prognoziranu (veštački zaprljanu kako bi simulirali prognoziranu vrednost) vrednost temperature. Može se primetiti da se u Tabeli 7.2, radi o istom 2F-LSSVM modelu kao i u Tabeli 7.1. Ovaj model je uzet jer je realan model za razliku od modela 2F-LSSVM-P i 2F-

LSSVM-T, koji koriste tačne vrednosti atributa vršnog dnevnog opterećenja, odnosno srednje dnevne temperature, koji nisu poznati u trenutku prognoze.

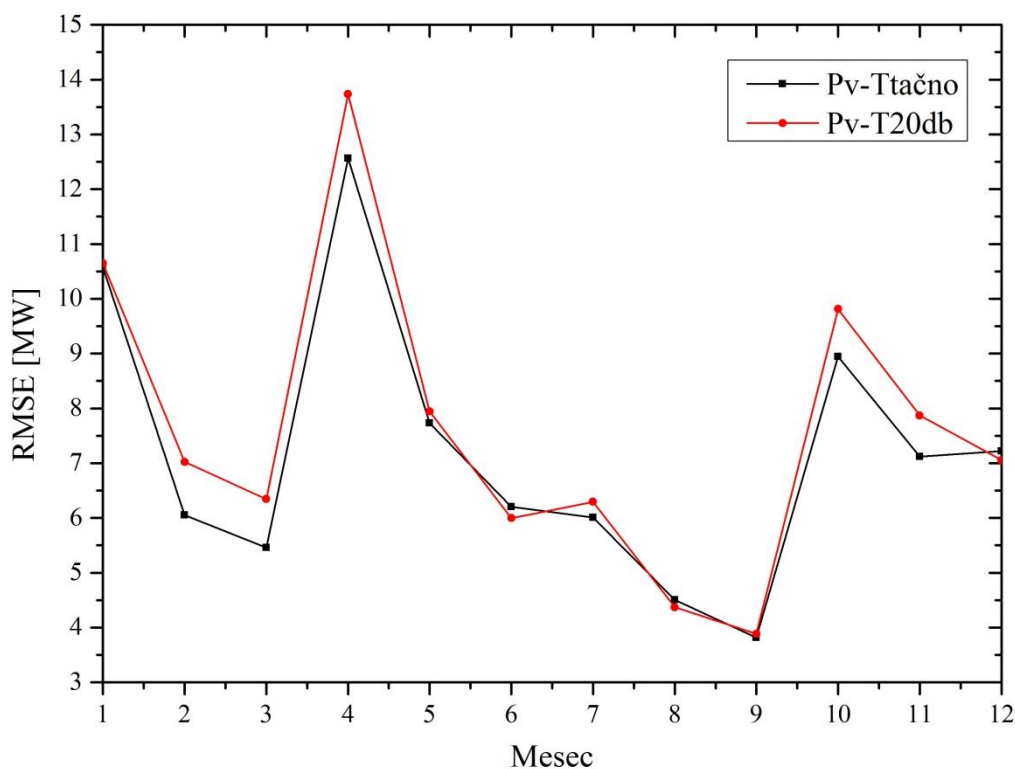
Tabela 7.2 Prosečna godišnja MAPE, sMAPE i RMSE greška za modele sa prognozom u dve faze

Model	P_{\max}	T_{srednje}	MAPE [%]	sMAPE [%]	RMSE [MW]
2F-LSSVM-P	Tačno		3.77	3.77	7.96
2F-LSSVM-T	Prognozirano	Tačno	4.15	4.16	8.84
2F-LSSVM	Prognozirano	Tačno + beli šum 20db	4.2	4.2	8.94

Na Slikama 7.1 i 7.2 prikazana je mesečna MAPE i RMSE greška respektivno, za modele prognoze vršnog dnevnog opterećenja Pv-Ttačno i Pv-T20db. Model Pv-Ttačno se odnosi na model za prognozu vršnog dnevnog opterećenja koji koristi tačnu vrednost srednje dnevne temperature. Model Pv-T20db koristi prognoziranu vrednost srednje dnevne temperature ali kako je ova vrednost nedostupna (specijalizovani servisi ne skladište prognozirane vrednosti temperature), ovaj model koristi veštački zaprljanu vrednost tačne temperature Gausovim šumom kao simulaciju prognozirane vrednosti.



Slika 7.1 MAPE greška modela kod prognoze vršnog dnevnog opterećenja



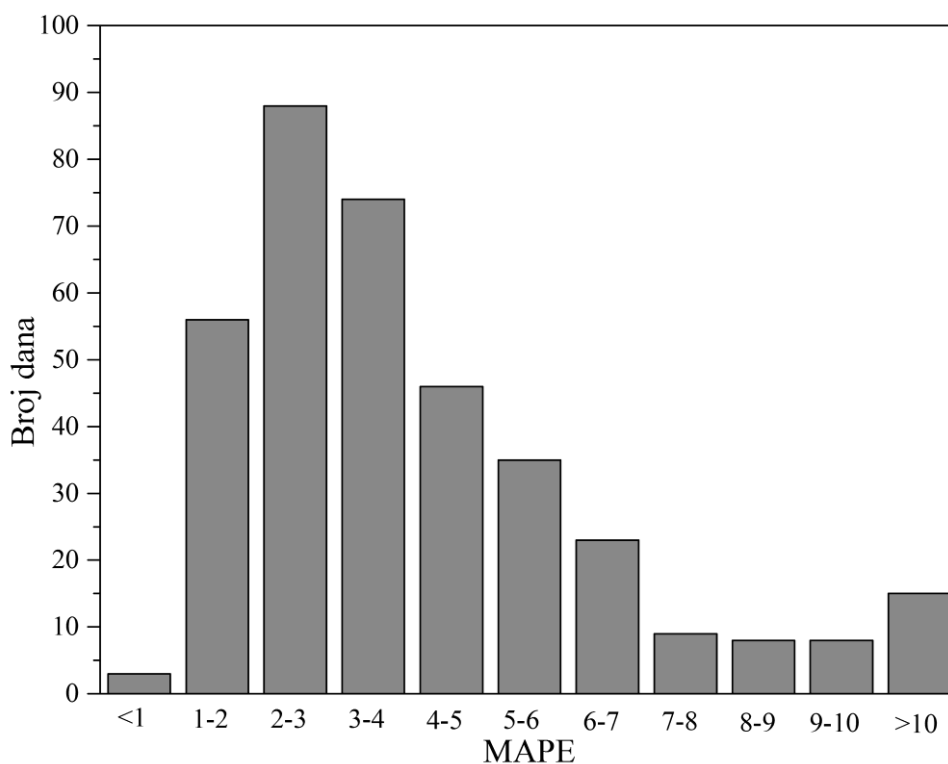
Slika 7.2 RMSE greška modela kod prognoze vršnog dnevnog opterećenja

Kod prognoze vršnog dnevnog opterećenja u prvoj fazi prognoze kod predloženog modela za prognozu časovnog opterećenja u dve faze, Pv-Ttačno model koji koristi tačnu vrednost temperature je očekivano precizniji od modela Pv-T20db koji koristi simuliranu prognozu temperature. Prognoza ovih modela tj. vrednost vršnog dnevnog opterećenja se koristi u drugoj fazi, pa tako model 2F-LSSVM-T koristi prognozu modela Pv-Tačno a model 2F-LSSVM, prognozu modela Pv-T20db. Rezultati iz Tabele 7.2 su očekivani, prednjači model koji koristi tačnu vrednost vršnog dnevnog opterećenja 2F-LSSVM-P, prati ga model 2F-LSSVM-T i na začelju je 2F-LSSVM model.

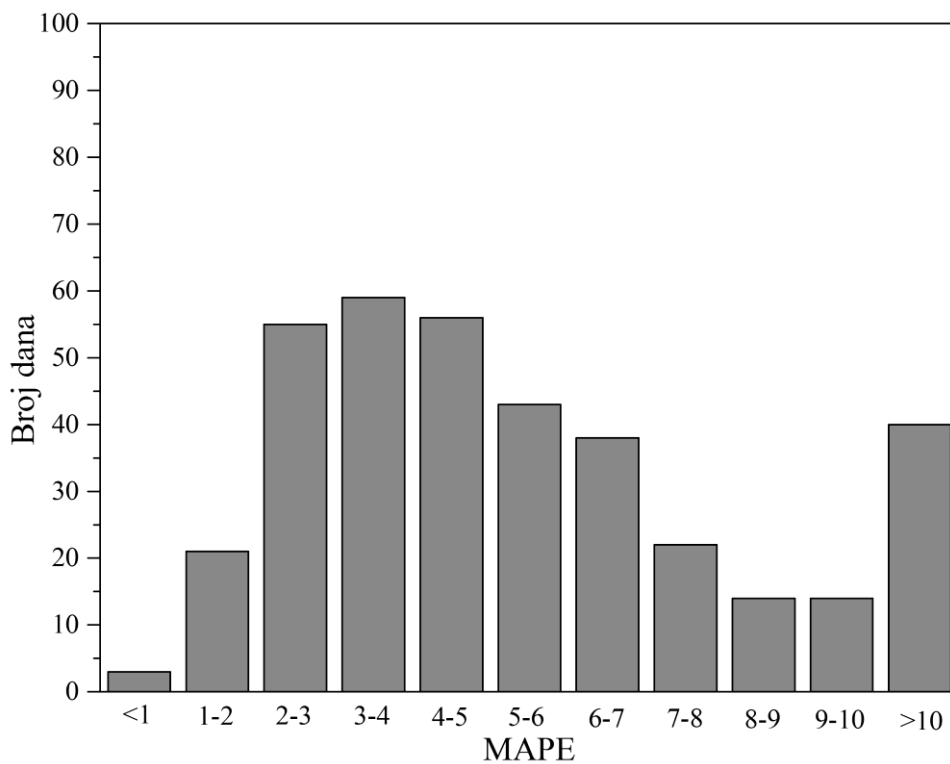
Kod definisanja modela sa prognozom u dve faze, pošlo se od pretpostavke da se preciznost modela može poboljšati, ukoliko se modelu pruži dodatan izvor informacija. Tako je u model uključen dodatni atribut, vršno dnevno opterećenje. Da je ova pretpostavka tačna, potvrđuje rezultat koji je ostvario model 2F-LSSVM-P. Ako se uporede rezultati iz Tabele 7.1 i rezultat modela 2F-LSSVM-P iz Tabele 7.2, može se primetiti da je model 2F-LSSVM-P najbolji, što ukazuje da se kod formiranja modela sa prognozom u dve faze pošlo od ispravne pretpostavke. Međutim kako ovaj model nije realan, jer koristi tačnu vrednost vršnog dnevnog opterećenja za dan unapred koja je nepoznata u trenutku prognoze, ostaje model 2F-LSSVM koji koristi prognoziranu vrednost vršnog dnevnog opterećenja. Naravno,

očekivano je da ovaj model ima manju preciznost od 2F-LSSVM-P i 2F-LSSVM-T modela, ali trebalo je videti, gde se ovaj model nalazi u odnosu na ostale modele iz Tabele 7.1.

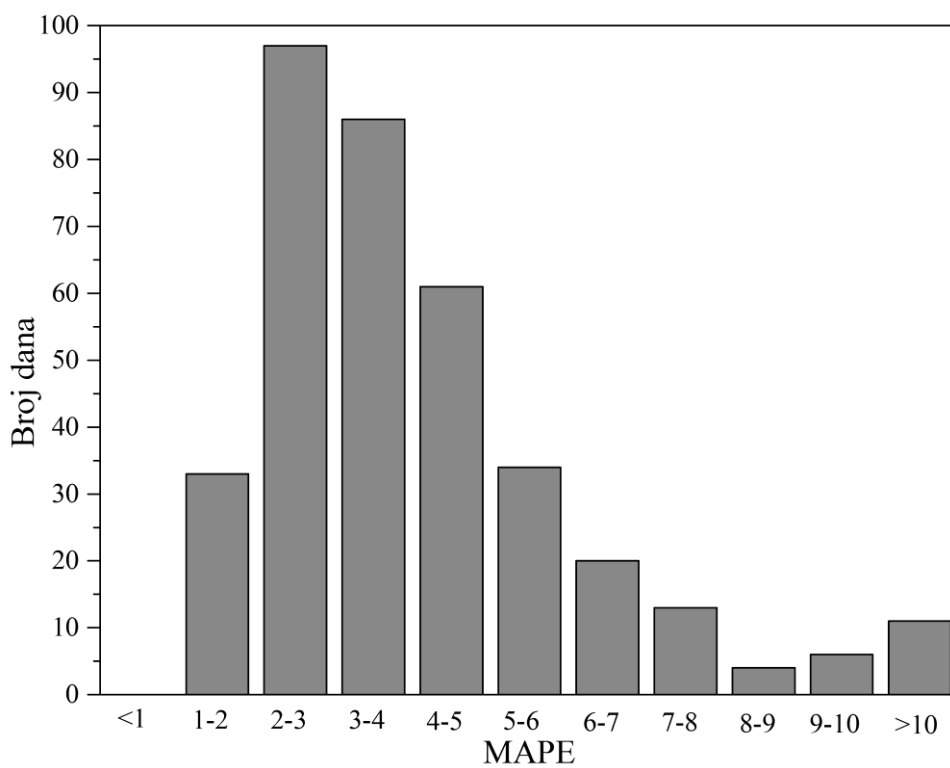
Na Slikama 7.3 - 7.10 prikazani su histogrami MAPE greške svih modela datih u disertaciji. Horizontalna osa predstavlja opsege MAPE greške, a vertikalna osa predstavlja broj dana u godini koji pripadaju datom opsegu. Definisano je 11 opsega, za vrednosti manje od 1, za vrednosti između [1-2], [2-3], [3-4], [4-5], [5-6], [6-7], [7-8], [8-9], [9-10] i vrednosti veće od 10. Tako imamo 11 pravougaonika (kolona) koji odgovaraju definisanim opsezima a čija visina odgovara broju dana u godini koji imaju vrednost MAPE greške u određenom opsegu. Naravno, ukupan zbir dana svih opsega, jednak je broju dana u godini.



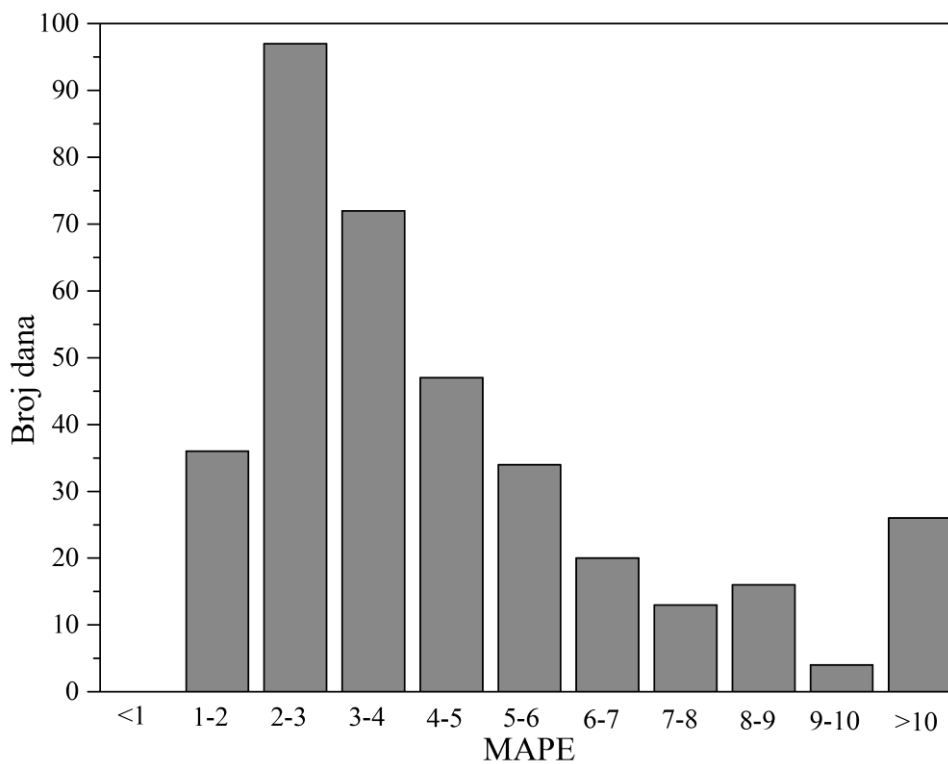
Slika 7.3 Histogram MAPE greške DS-ARIMA modela



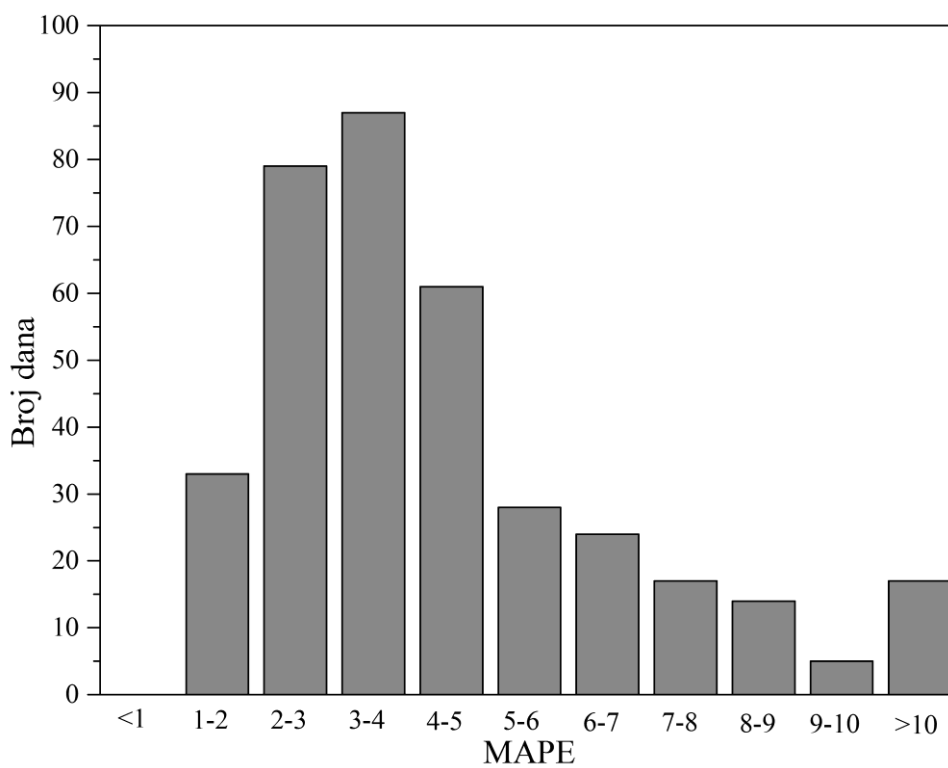
Slika 7.4 Histograma MAPE greške DS-HW modela



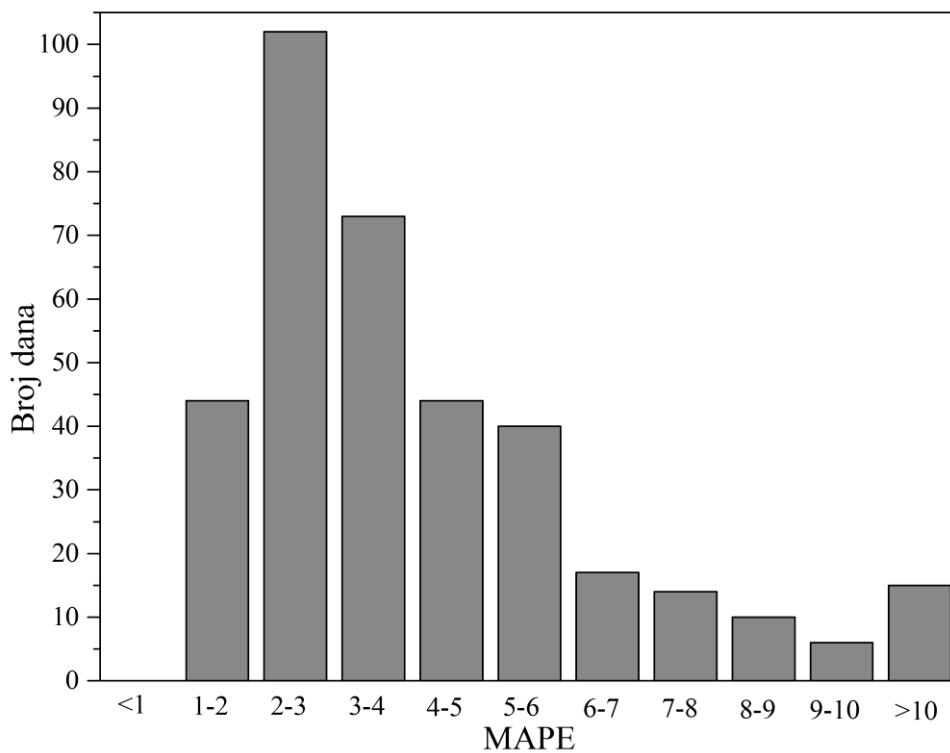
Slika 7.5 Histogram MAPE greške ANN modela



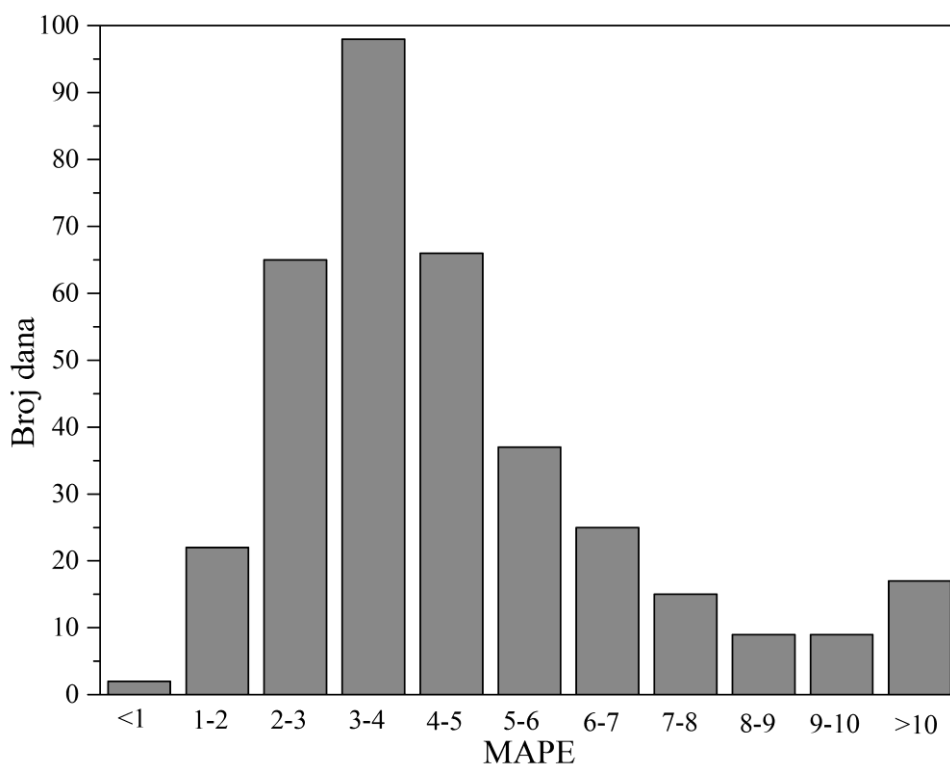
Slika 7.6 Histogram MAPE greške RVM modela



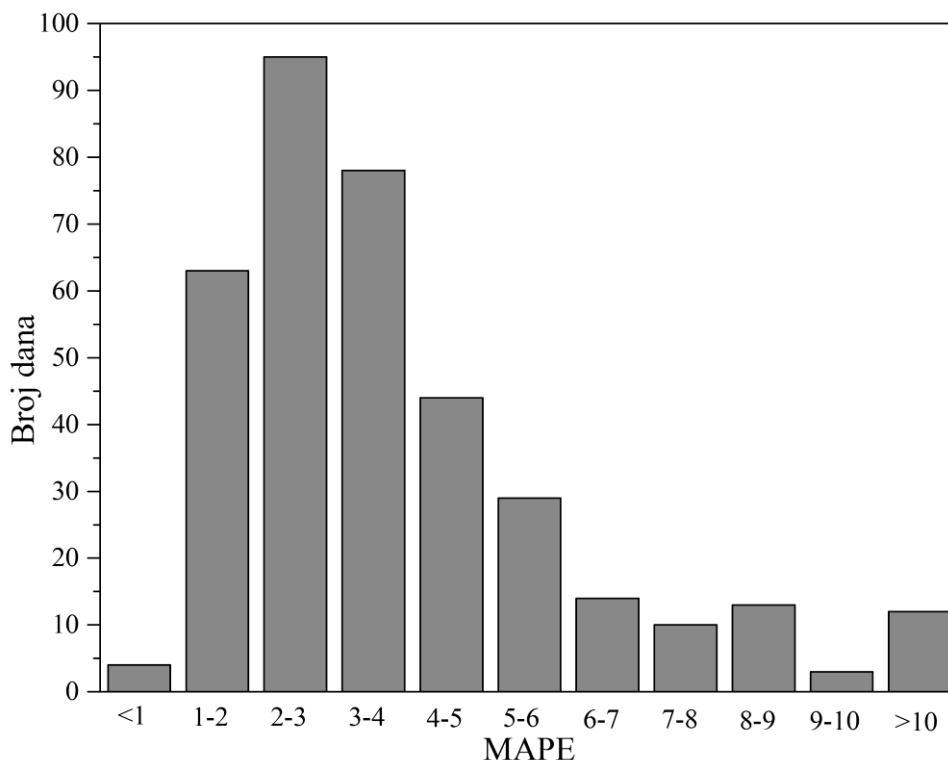
Slika 7.7 Histogram MAPE greške LSSVM modela



Slika 7.8 Histogram MAPE greške 2F-LSSVM modela



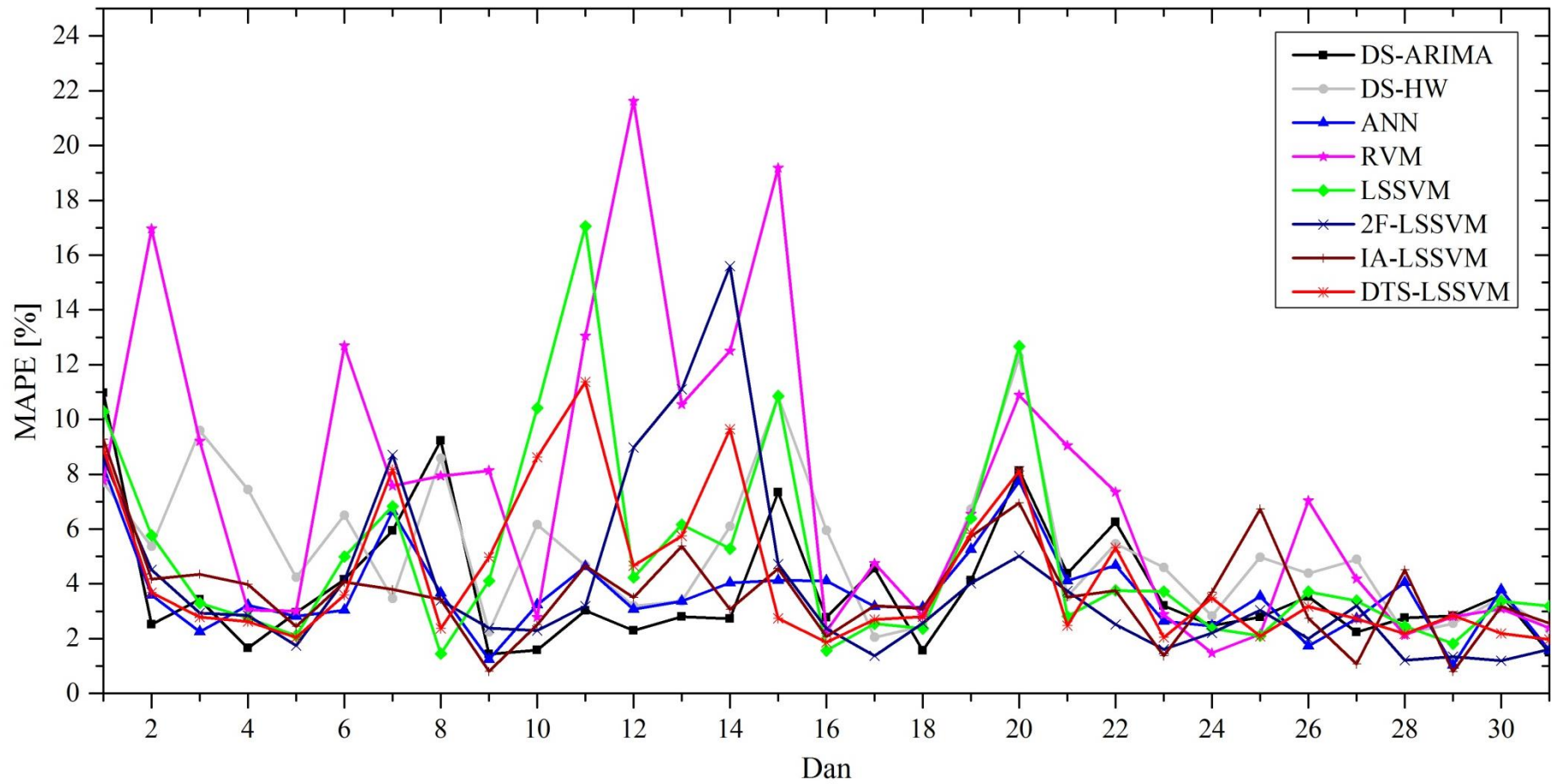
Slika 7.9 Histogram MAPE greške IA-LSSVM modela



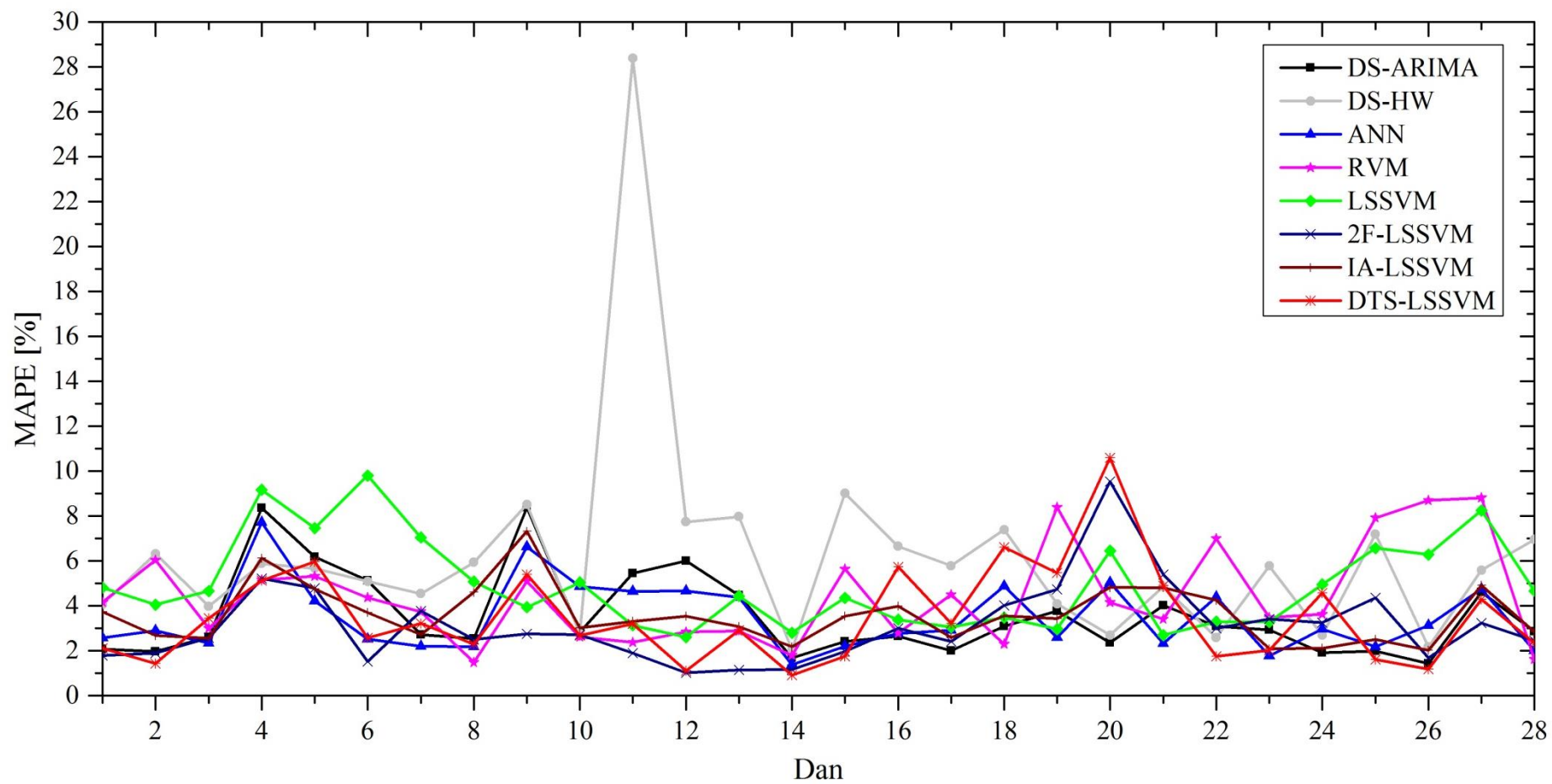
Slika 7.10 Histogram MAPE greške DTS-LSSVM modela

Najveći broj dana u godini ima vrednosti MAPE greške u opsegu od 2 do 3%, kod svih modela osim LSSVM, IA-LSSVM i DS-HW modela kod kojih je ovaj broj najveći u opsegu od 3 do 4%. Raspodela broja dana prema opsegu MAPE greške je slična kod svih modela osim DS-HW modela koji se izdvaja sa velikim brojem dana u opsegu većih vrednosti MAPE greške. Broj dana sa vrednostima MAPE greške većim od 10% nije zanemarljiv i kreće se oko 15 dana kod svih modela, osim RVM i DS-HW modela gde je taj broj i veći. Takođe broj dana sa vrednostima MAPE greške većim od 8% je oko 30 kod većine modela, što je ceo jedan mesec u godini. Ovo ukazuje da treba uraditi dublju analizu o kojim se konkretno danima radi i naći razloge zašto je preciznost modela u tim danima loša.

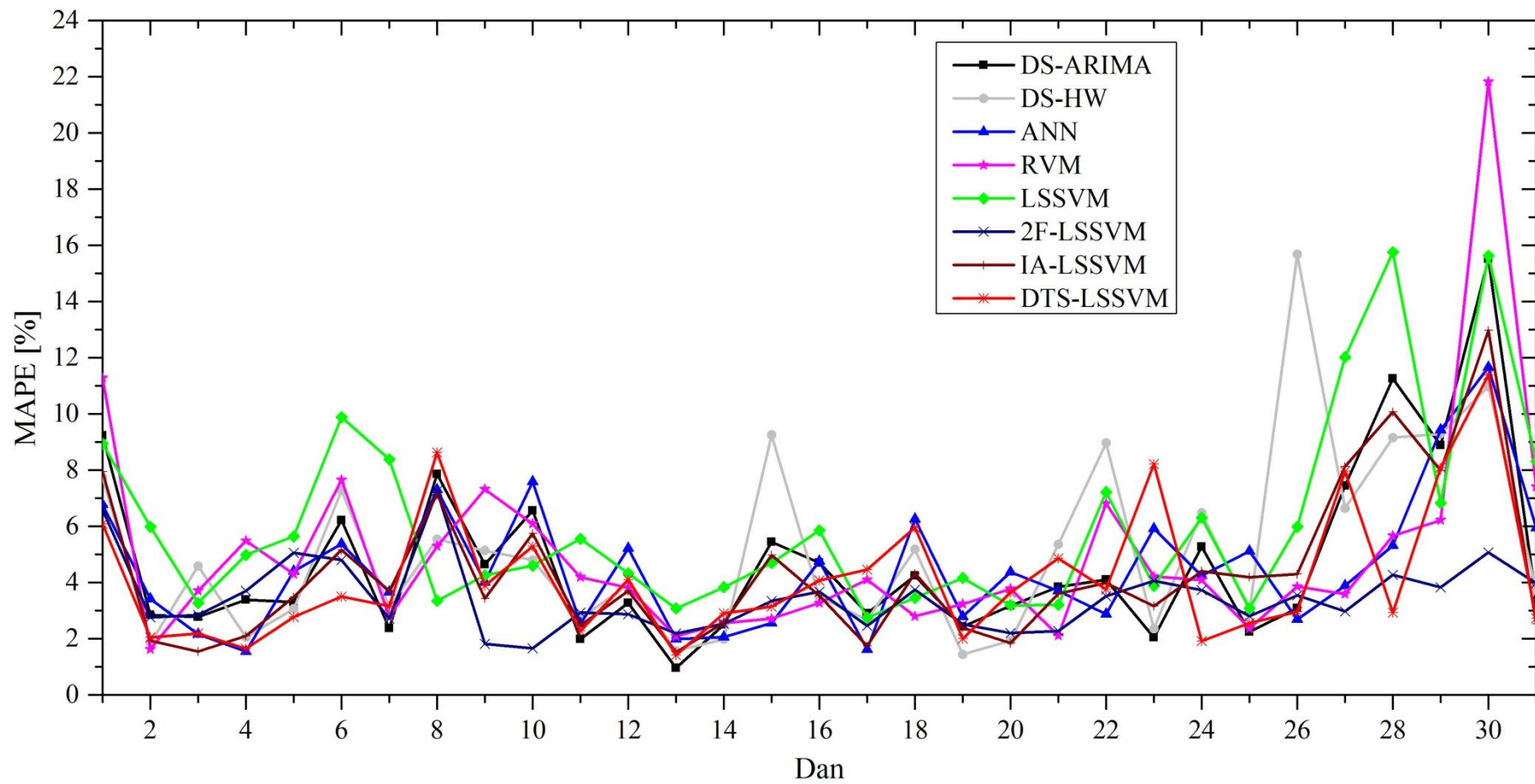
Na Slikama 7.11 - 7.22 prikazane su MAPE greške svakog modela za svaki mesec u godini. Na vertikalnoj osi date su vrednosti MAPE greške a na horizontalnoj osi označeni su dani u mesecu.



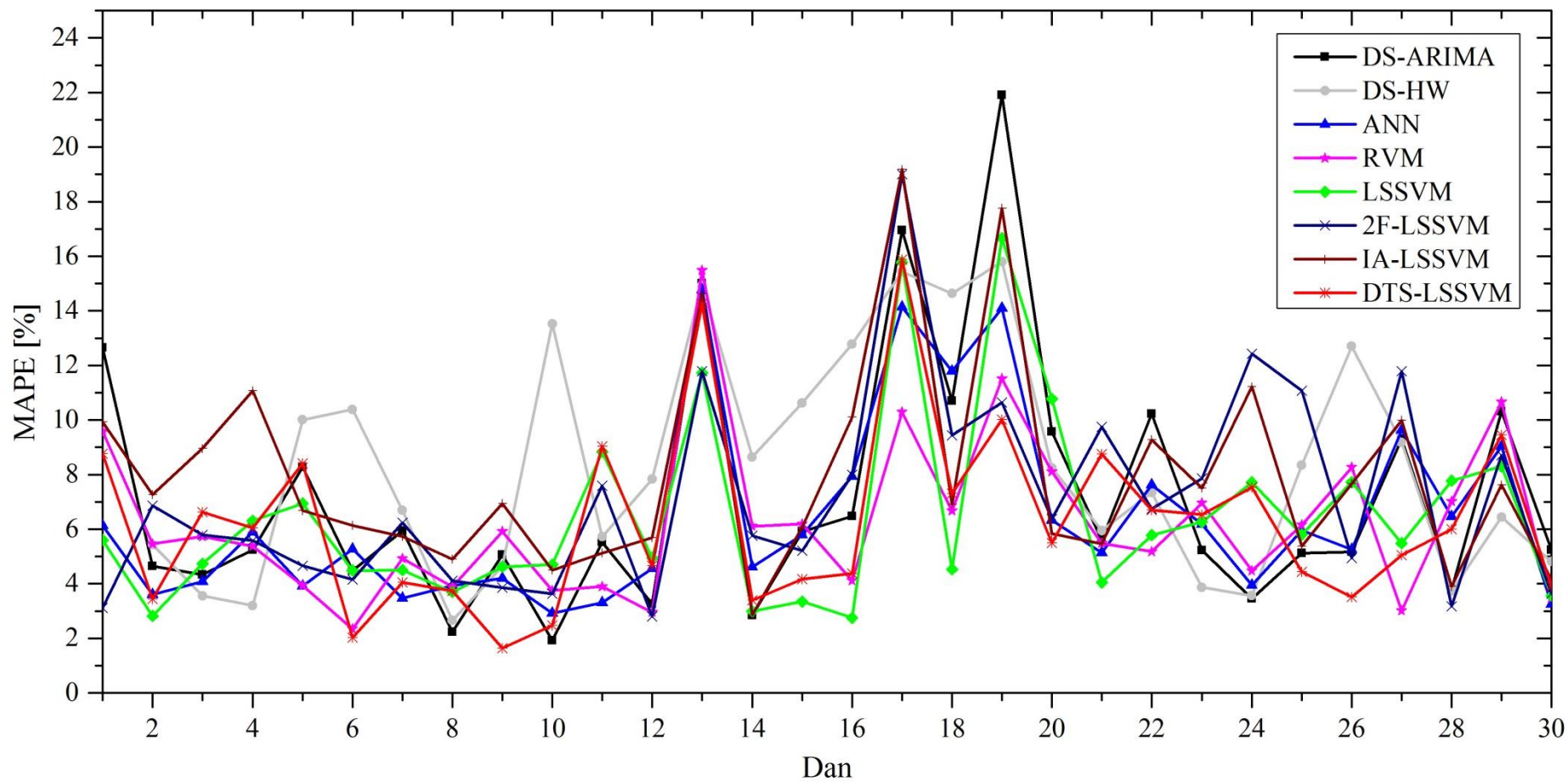
Slika 7.11 MAPE greška u januaru



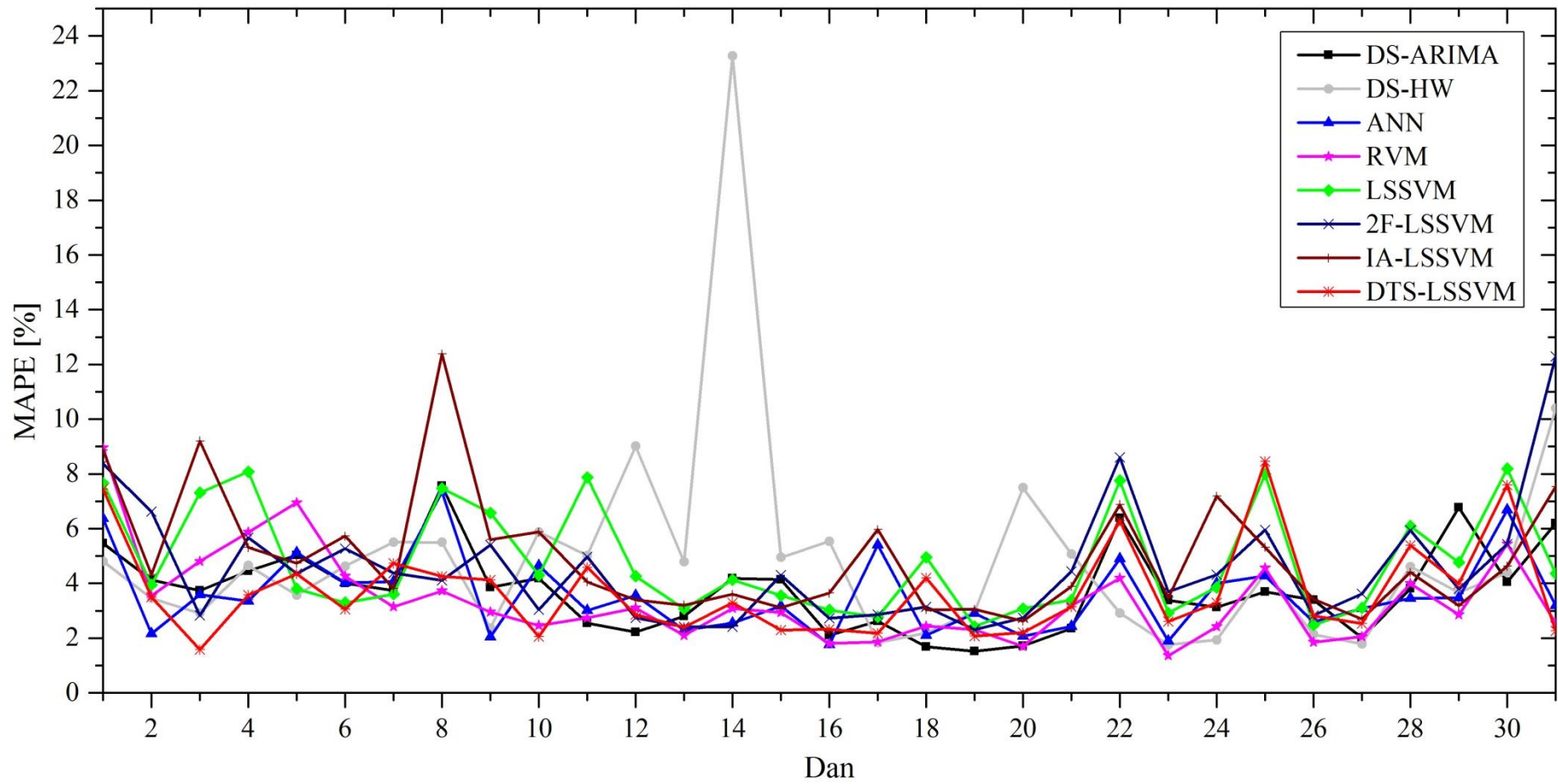
Slika 7.12 MAPE greška u februaru



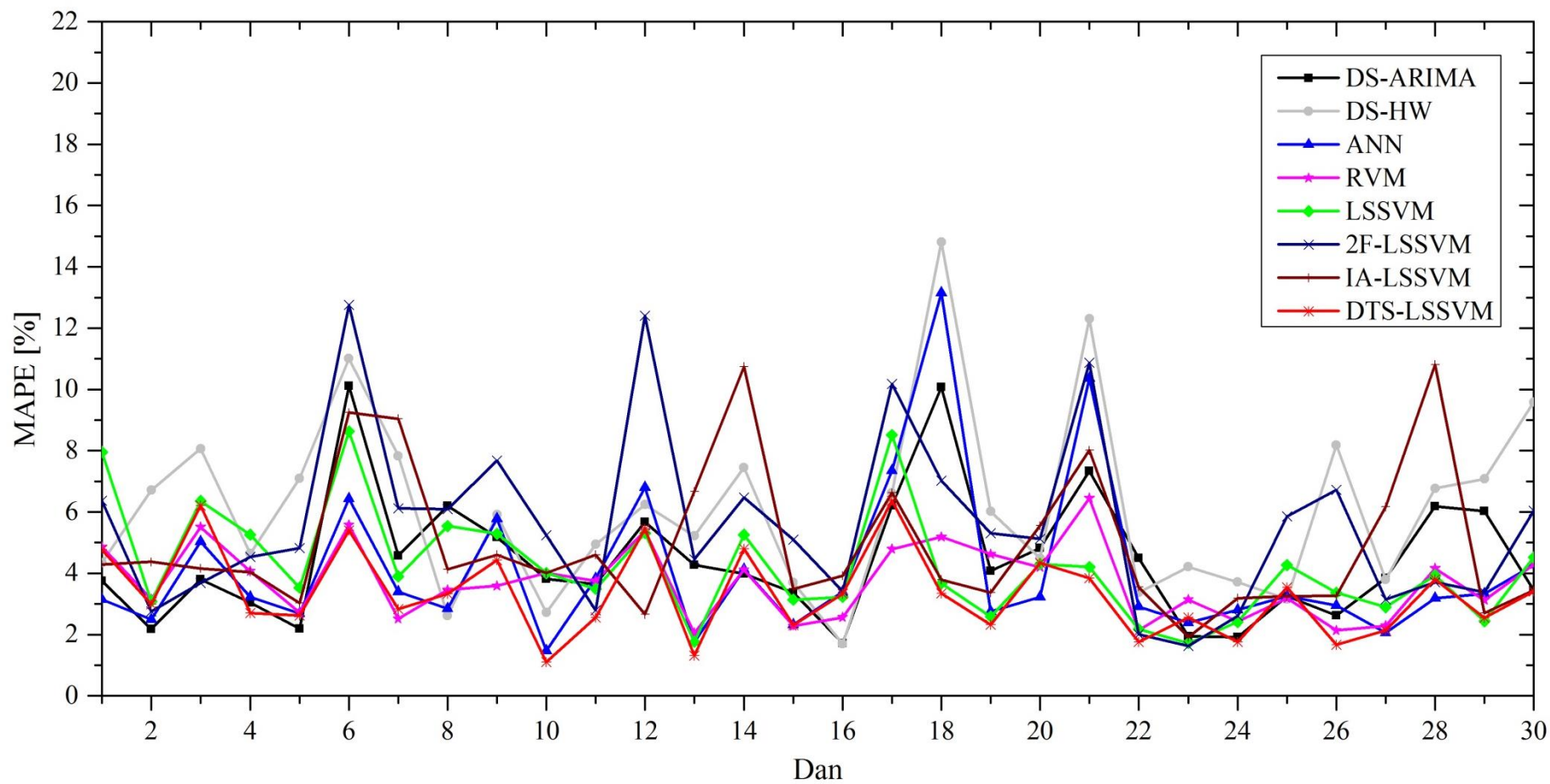
Slika 7.13 MAPE greška u martu



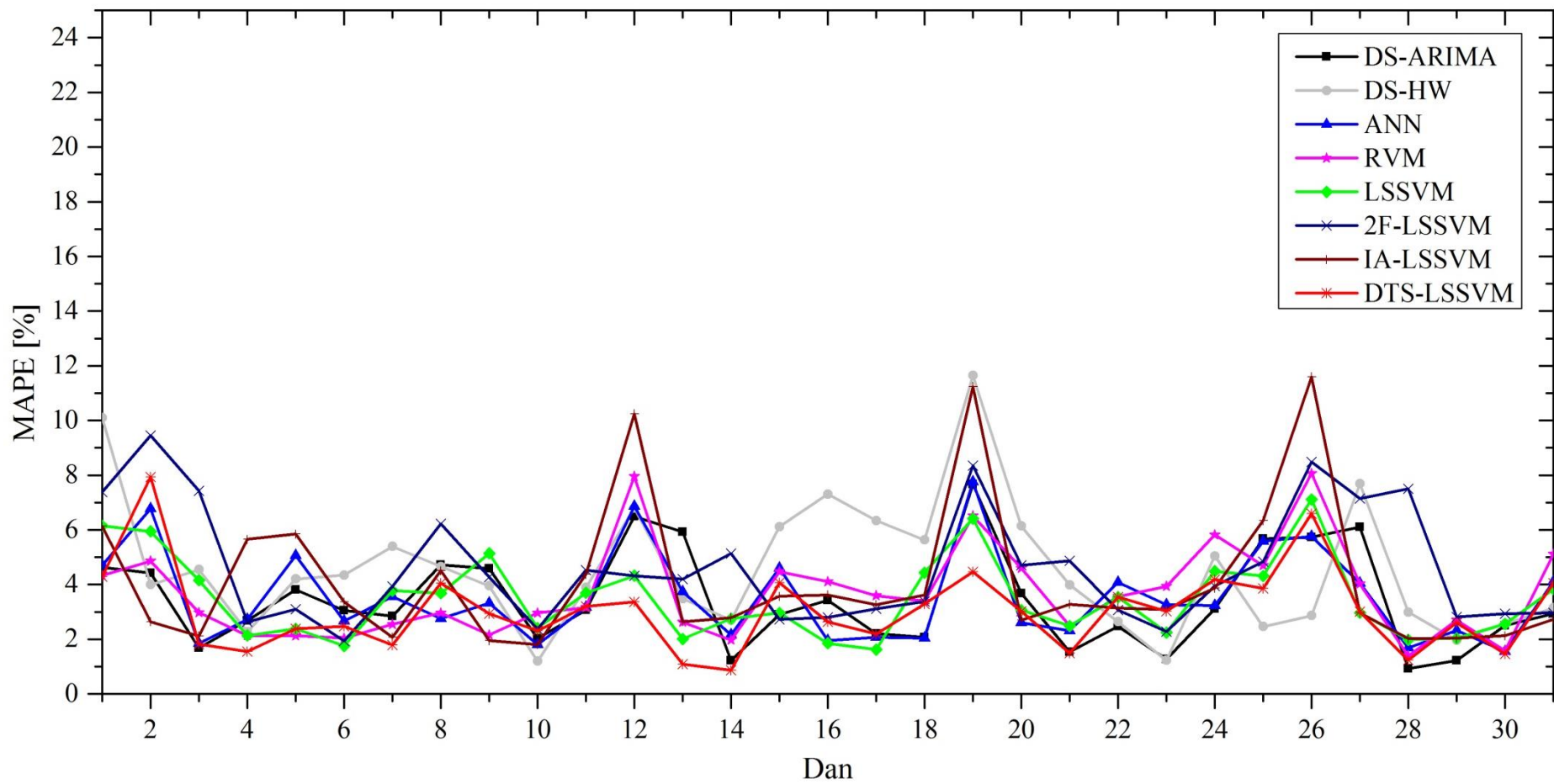
Slika 7.14 MAPE greška u aprilu



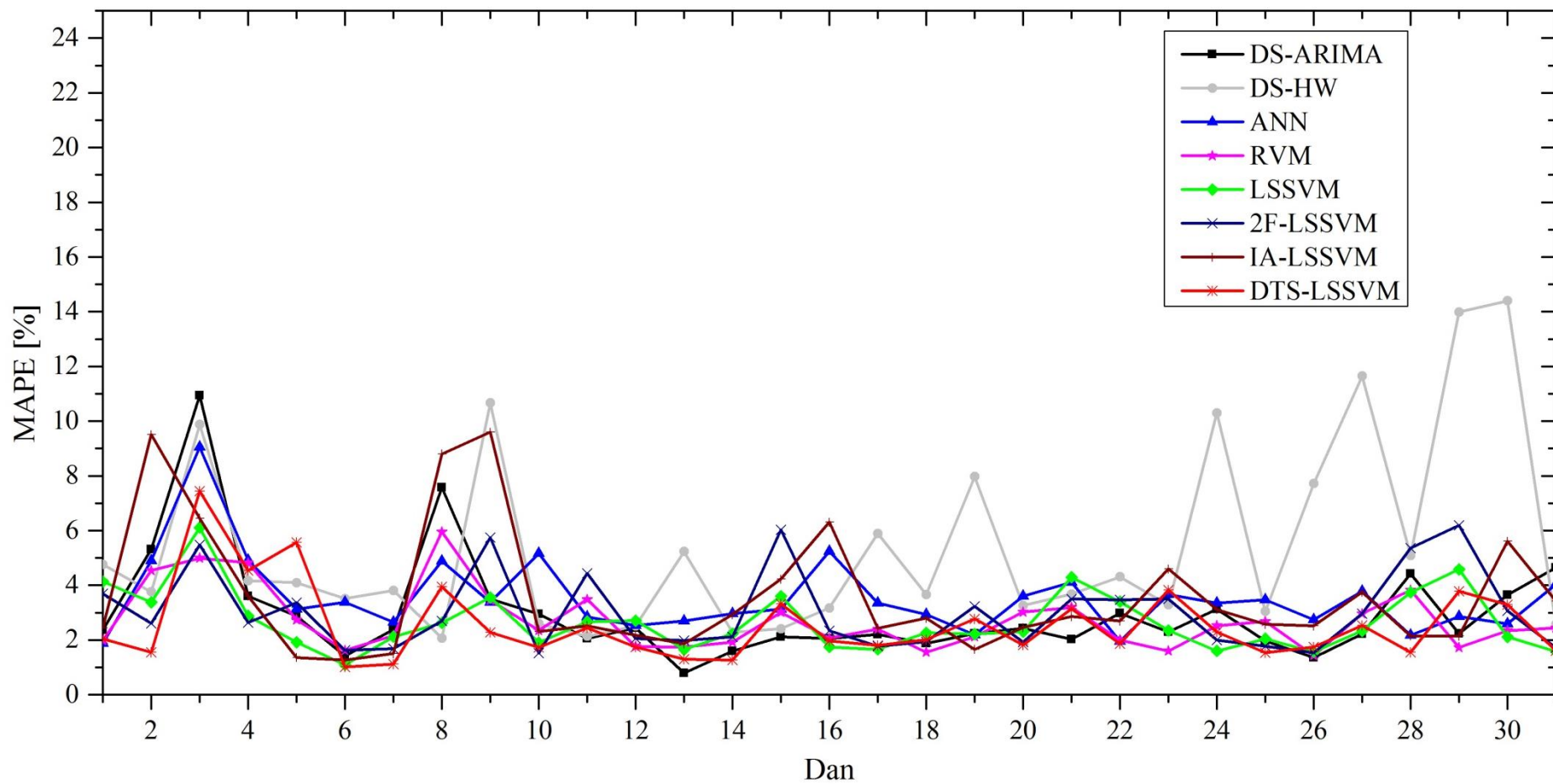
Slika 7.15 MAPE greška u maju



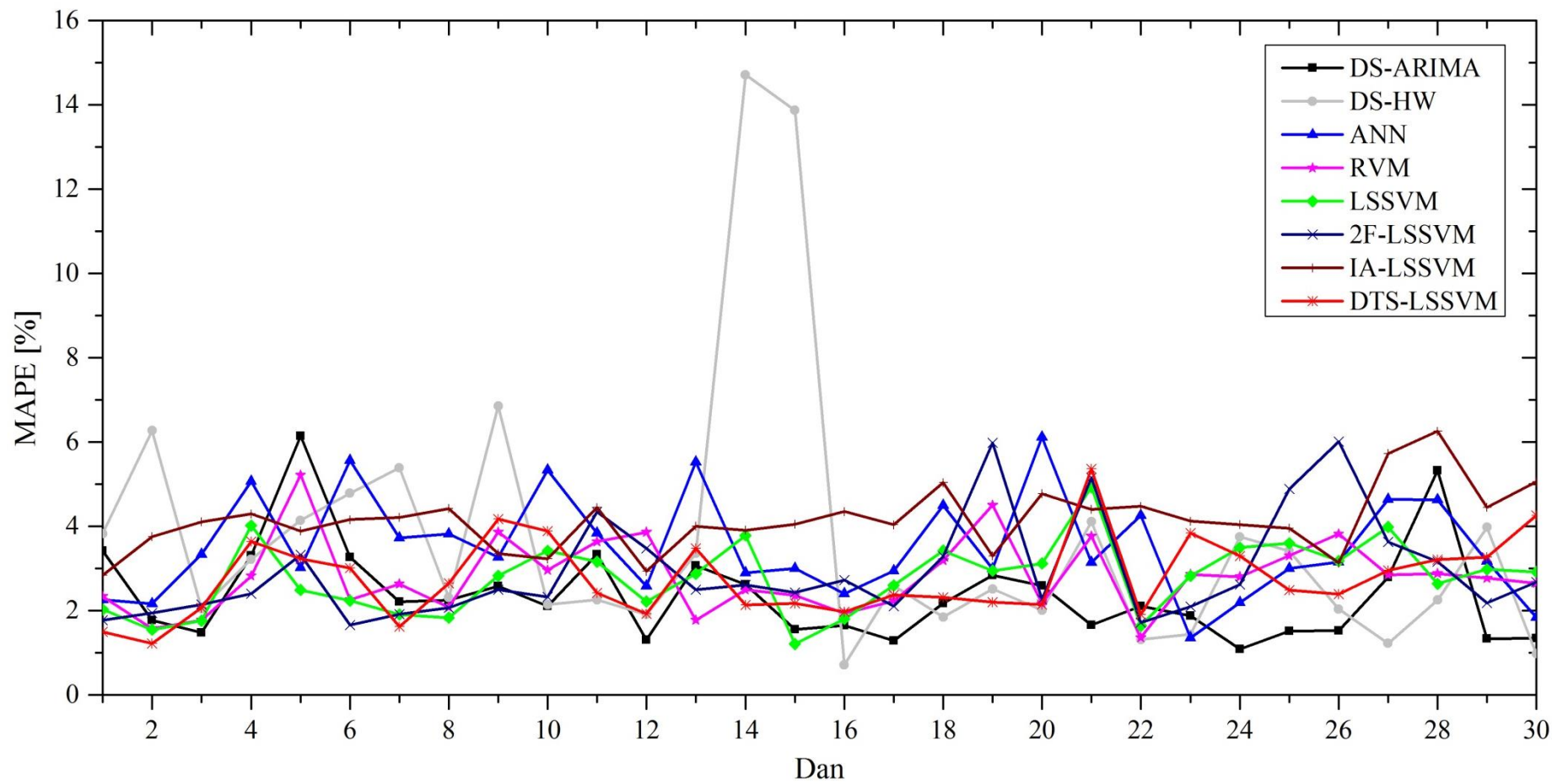
Slika 7.16 MAPE greška u junu



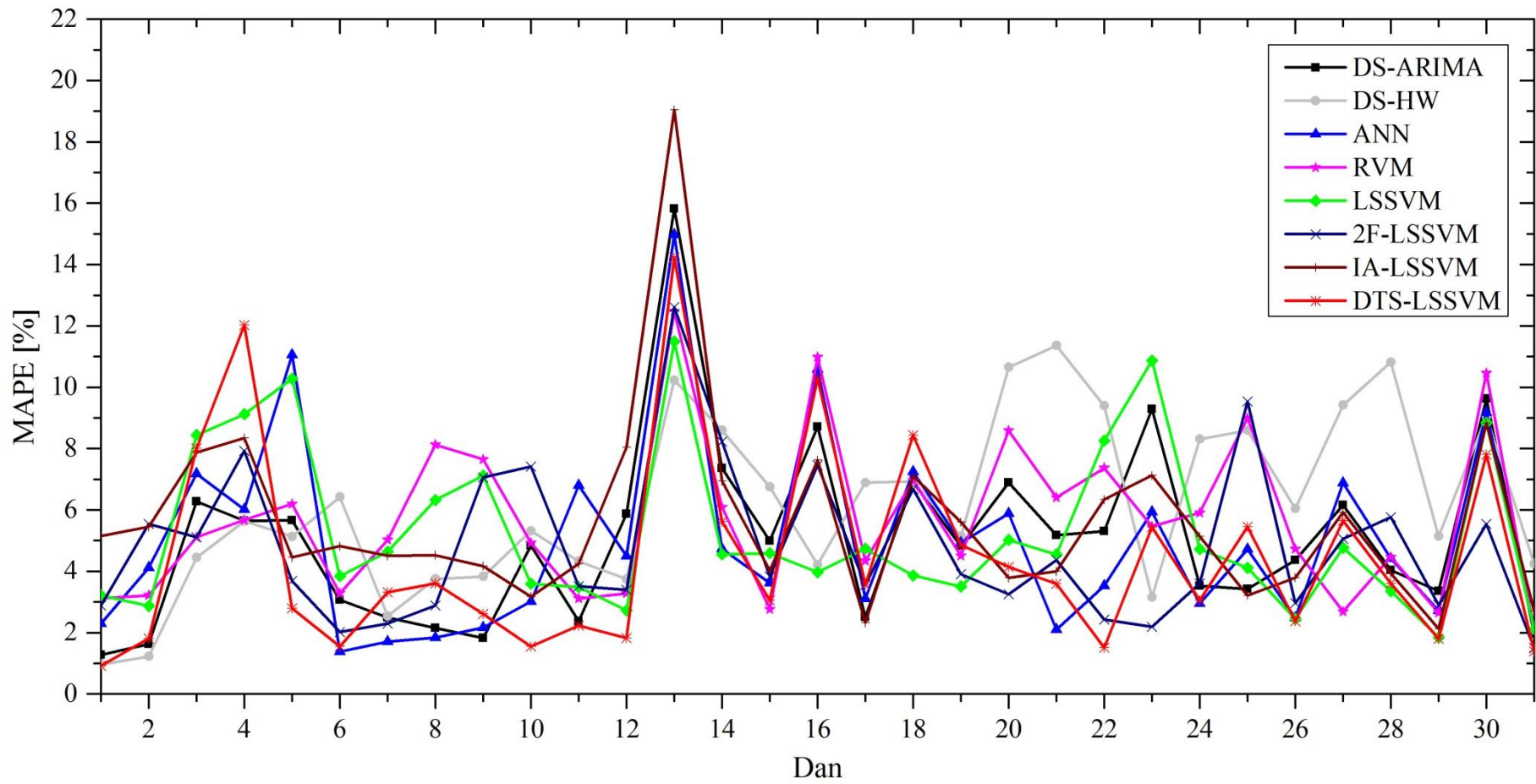
Slika 7.17 MAPE greška u julu



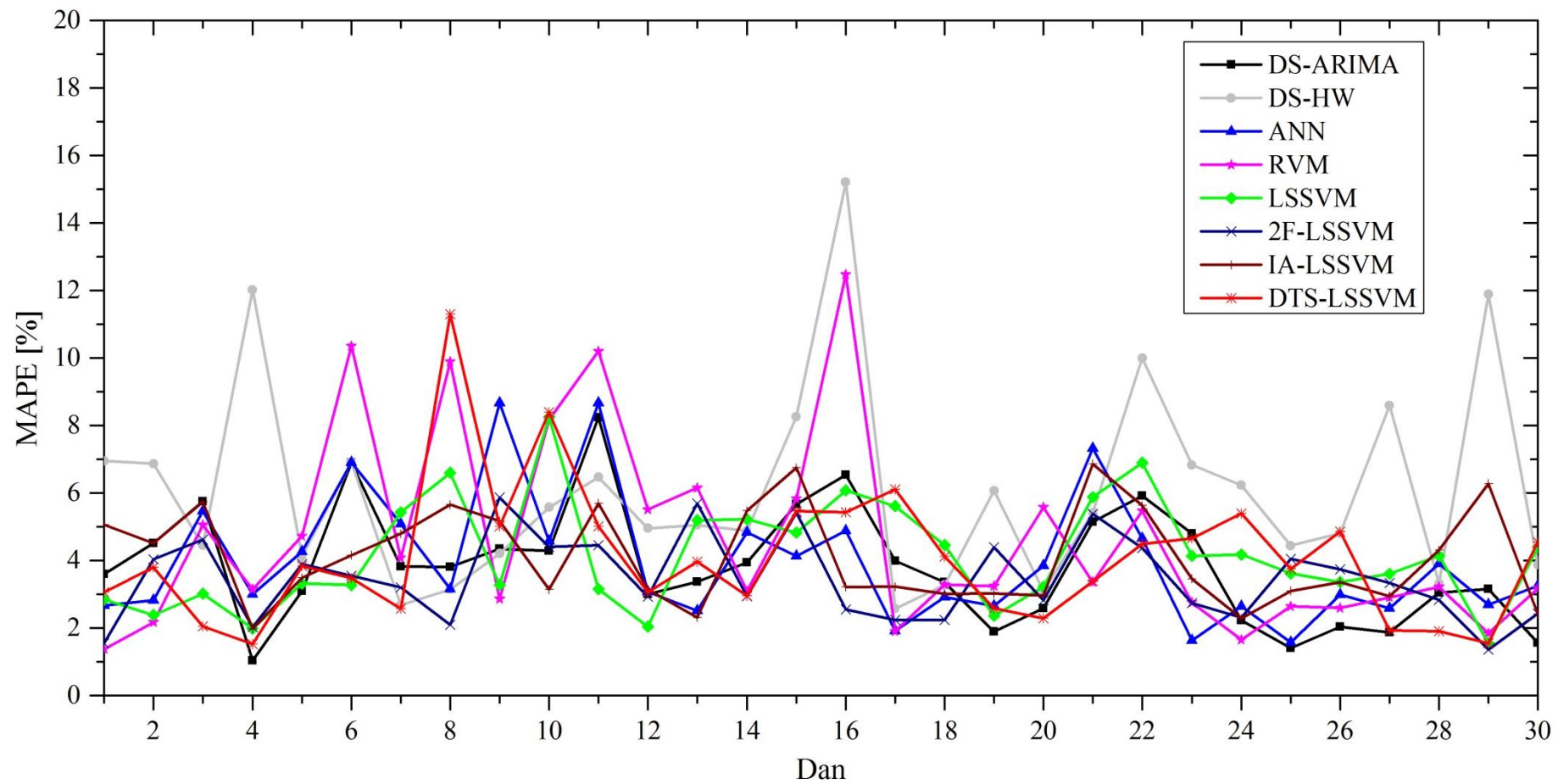
Slika 7.18 MAPE greška u avgustu



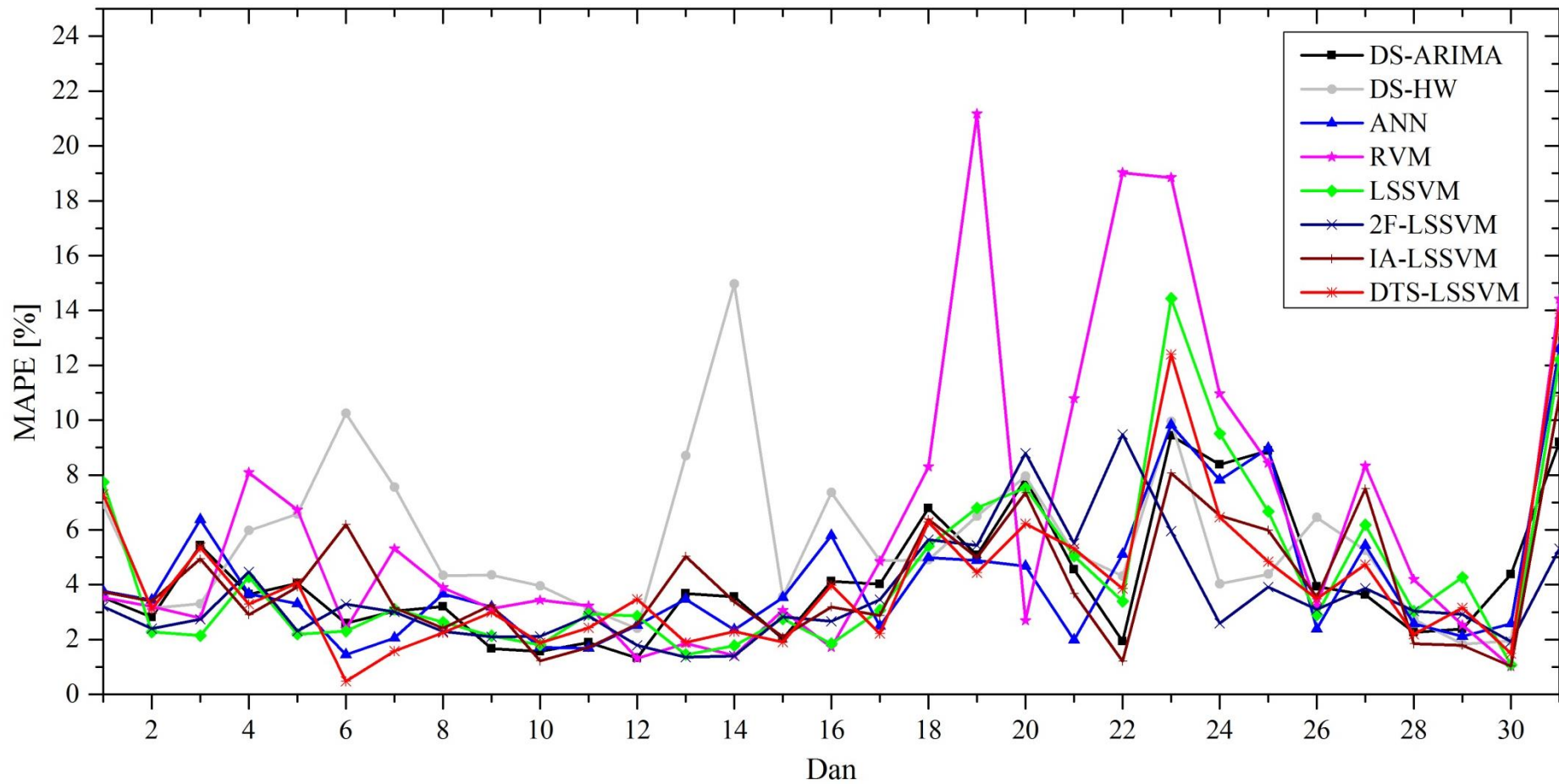
Slika 7.19 MAPE greška u septembru



Slika 7.20 MAPE greška u oktobru



Slika 7.21 MAPE greška u novembru



Slika 7.22 MAPE greška u decembru

U Tabeli 7.3 data je prosečna mesečna MAPE greška svakog modela. Prva kolona označava mesec a ostale kolone predstavljaju vrednost MAPE greške za konkretan model. Poslednja kolona predstavlja prosečnu mesečnu MAPE grešku svih modela.

Tabela 7.3 Prosečna mesečna MAPE greška

Mesec	DS-ARIMA	DS-HW	ANN	RVM	LSSVM	2F-LSSVM	IA-LSSVM	DTS-LSSVM	MAPE/mesec
Januar	3.83	5.17	3.64	7.35	4.96	4.01	3.71	4.31	4.62
Februar	3.56	6.16	3.47	4.40	4.92	3.12	3.60	3.52	4.09
Mart	4.75	5.19	4.55	5.04	6.08	3.48	4.39	4.21	4.71
April	7.21	8.34	6.44	6.23	6.44	7.16	7.95	6.26	7.00
Maj	3.77	4.95	3.60	3.39	4.84	4.57	4.97	3.72	4.23
Jun	4.46	6.14	4.09	3.72	4.22	5.61	4.96	3.32	4.57
Juli	3.44	4.56	3.55	3.71	3.49	4.61	4.11	3.06	3.82
Avgust	3.02	5.41	3.54	2.72	2.66	3.03	3.60	2.54	3.32
Septembar	2.39	3.71	3.52	2.83	2.78	2.93	4.16	2.77	3.14
Oktober	5.08	6.20	4.97	5.70	5.27	4.81	5.62	4.45	5.26
Novembar	3.83	6.07	3.98	4.62	4.15	3.45	4.11	4.08	4.29
Decembar	4.22	5.82	4.21	6.25	4.38	3.61	4.09	4.17	4.59

Tabela 7.3 pruža uvid u kretanje MAPE greške svakog modela kroz određene periode godine, koji su ovde definisani kao period za svaki mesec u godini. Ponašanje modela u određenom periodu (za konkretan mesec), prikazano je na Slikama 7.11 - 7.22. Uvidom u poslednju kolonu Tabele 7.3, primećuje se da je preciznost svih modela najmanja u aprilu i oktobru. Sa druge strane, septembar, avgust i juli su meseci kada je preciznost modela najveća. Moglo se pretpostaviti da je razlog ovakvog ponašanja modela velika fluktuacija temperature u aprilu i oktobru, i stabilno kretanje temperature u septembru, avgustu i julu. Međutim da to nije razlog, leži u činjenici da je prosečna mesečna promena dnevne temperature u aprilu 13.64 C°, u oktobru 10.6 C° a u septembru 13.88 C°, avgustu 14.33 C° i julu 13.87 C°. Avgust koji ima najveću fluktuaciju temperature ujedno je i mesec sa drugim najboljim rezultatom. Sa Slike 7.14 može se primetiti da su u aprilu 13., 17., 19. i 29. dani sa izraženijom greškom u odnosu na ostale dane u mesecu, koji najviše doprinose lošem rezultatu modela u ovom mesecu. Naći razlog zašto je u ovim danim greška posebno velika

nije lako. Ovi dani nisu ništa drugačiji od ostalih dana u mesecu kada je u pitanju opterećenje ili veća fluktuacija temperature.

Preciznost datih modela je veća u letnjem (maj – septembar) nego u zimskom periodu godine (novembar - mart). April i oktobar su prelazni meseci koji se posebno izdvajaju i koji se kod modelovanja mogu svrstati u posebnu grupu. Ovo su meseci kada se završava odnosno počinje sezona grejanja. Uglavnom niske jutarnje temperature i visoke dnevne temperature dovode do potrebe dogrevanja korišćenjem električne energije jer nema potrebe za aktivnim grejanjem tokom celog dana. Veće greške ostalih zimskih meseci u odnosu na letnje mesece, leži u činjenici da veliki broj potrošača koristi električnu energiju za aktivno grejanje. Najmanje greške u julu, avgustu i septembru se mogu objasniti time da je su ovo meseci godišnjih odmora kada je i ljudska aktivnost smanjena. Nešto veća greška u junu u odnosu na ostale letnje mesece je posledica početka letnjih vrućina i uključivanja klima uređaja koja su nasumična i nepredvidiva.

Tabela 7.4 predstavlja prosečnu MAPE grešku po danu u nedelji za svaki model. Prva kolona označava dan u nedelji, sledećih osam kolona se odnose na prosečnu MAPE grešku za konkretan model a poslednja kolona prikazuje prosečnu MAPE grešku svih modela za konkretan dan u nedelji.

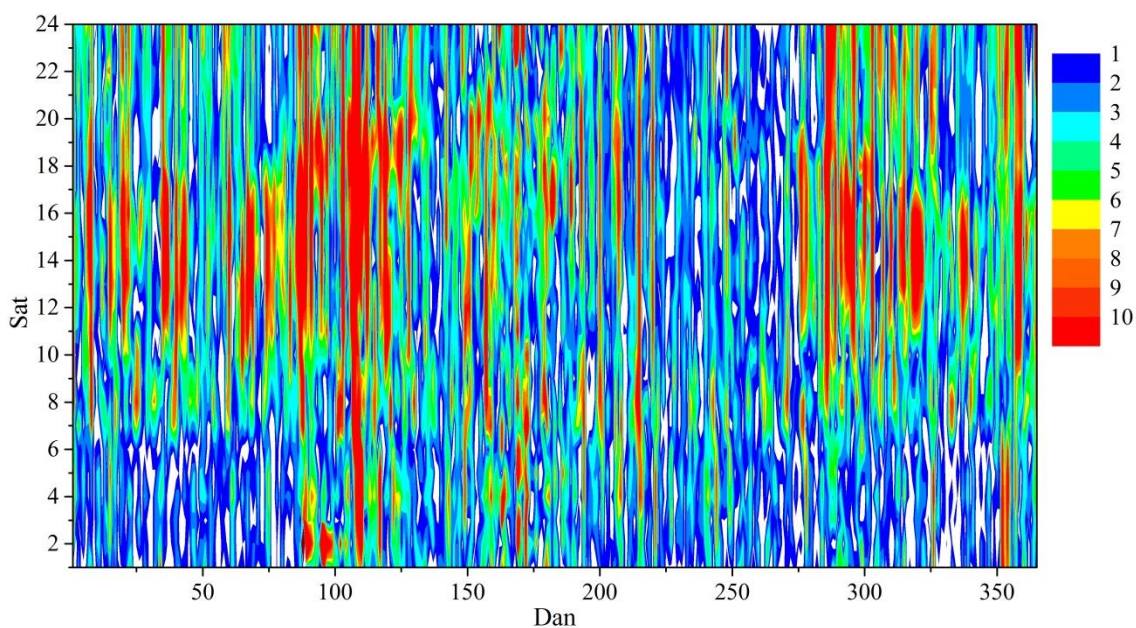
Tabela 7.4 Prosečna MAPE greška po danu u nedelji

Dan	DS-ARIMA	DS-HW	ANN	RVM	LSSVM	2F-LSSVM	IA-LSSVM	DTS-LSSVM	MAPE/dan
Ponedeljak	4.68	6.09	4.34	4.96	4.57	4.04	4.40	3.96	4.63
Utorak	3.47	5.10	3.96	4.61	4.19	3.99	3.76	3.75	4.10
Sreda	3.92	5.84	4.01	4.51	4.00	4.12	4.10	4.04	4.32
Četvrtak	3.99	5.76	4.05	4.78	4.19	3.77	4.09	3.54	4.27
Petak	4.22	4.91	4.26	5.15	5.11	4.74	4.60	4.03	4.63
Subota	3.92	4.90	3.52	4.03	4.63	4.28	4.57	3.57	4.18
Nedelja	4.72	6.84	4.78	4.65	4.91	4.50	6.73	4.20	5.17

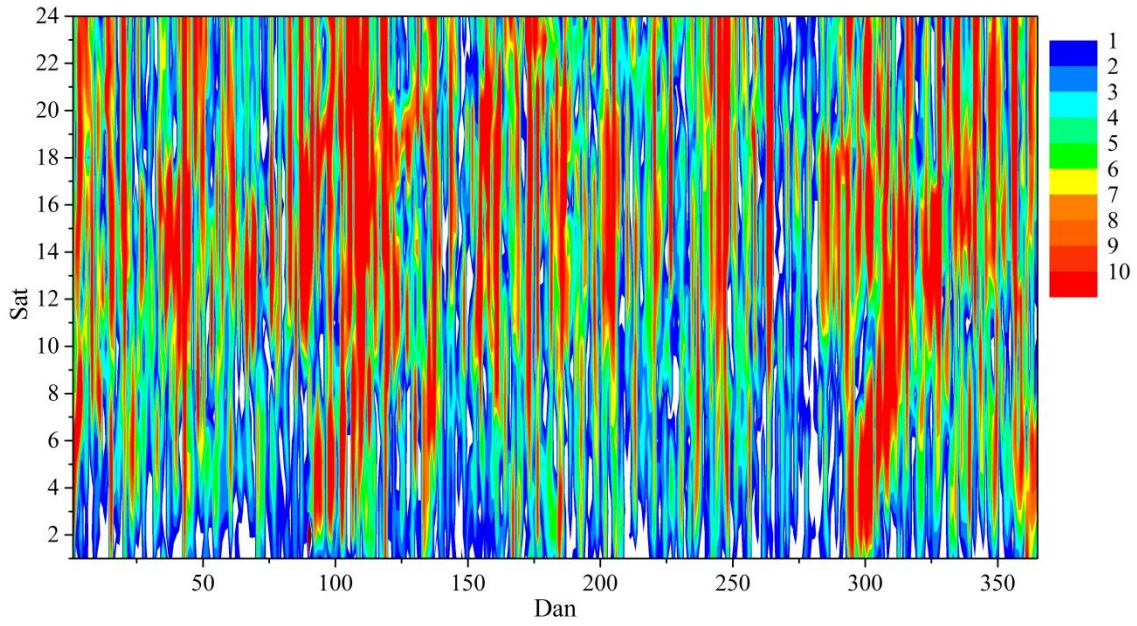
Kada je u pitanju tip dana tj. dan u nedelji, varijacije MAPE greške nisu izražene kao u slučaju različitih meseci u toku godine. Najveća greška je prisutna u nedelji koja je poslednji dan vikenda pa su i ljudske aktivnosti nešto drugačije. Tako se izdvajaju i ponedeljak i petak. Ponedeljak je karakterističan kao prvi radni dan kada su ljudske aktivnosti fokusirane na

početak radne sedmice a petak je poslednji dan radne sedmice i početak priprema ljudi za vikend. Da se primetiti da postoje varijacije u kretanju greške od dana do dana u skladu sa promenom i opterećenja u ovim danima što je sve uslovljeno sedmičnim ljudskim aktivnostima.

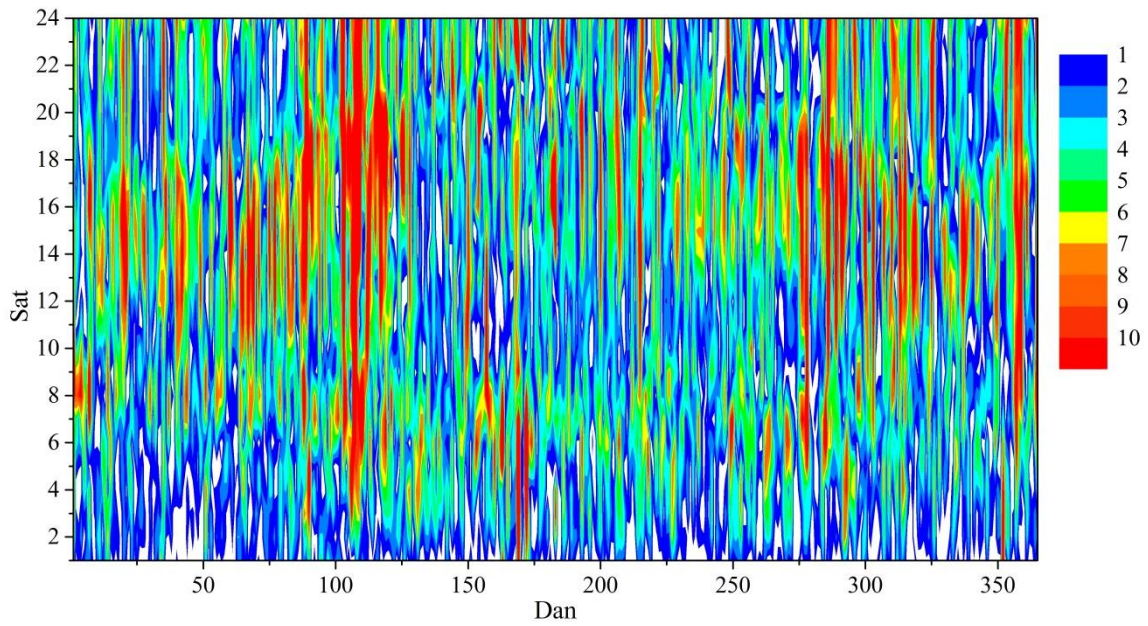
Na Slikama 7.23 - 7.30 prikazane su raspodele APE greške svakog sata po danima za svaki model. Horizontalna osa predstavlja sat u danu a vertikalna dan u godini. Legenda označava koje vrednosti APE greške odgovaraju kojoj boji. Hladnije boje (bela, plava) odgovaraju malim vrednostima APE greške dok toplije boje (žuta, crvena), odgovaraju velikim vrednostima APE greške. Tako bele oblasti predstavljaju APE greške ispod 1%, nijanse plave oblasti na slici predstavljaju male APE greške, između 1% i 4%. Toplije boje, zelena-žuta, predstavljaju oblasti sa greškom između 4% i 7% a nijanse crvene od 7% do 10%. Crvena boja označava oblasti sa APE greškom iznad 10%.



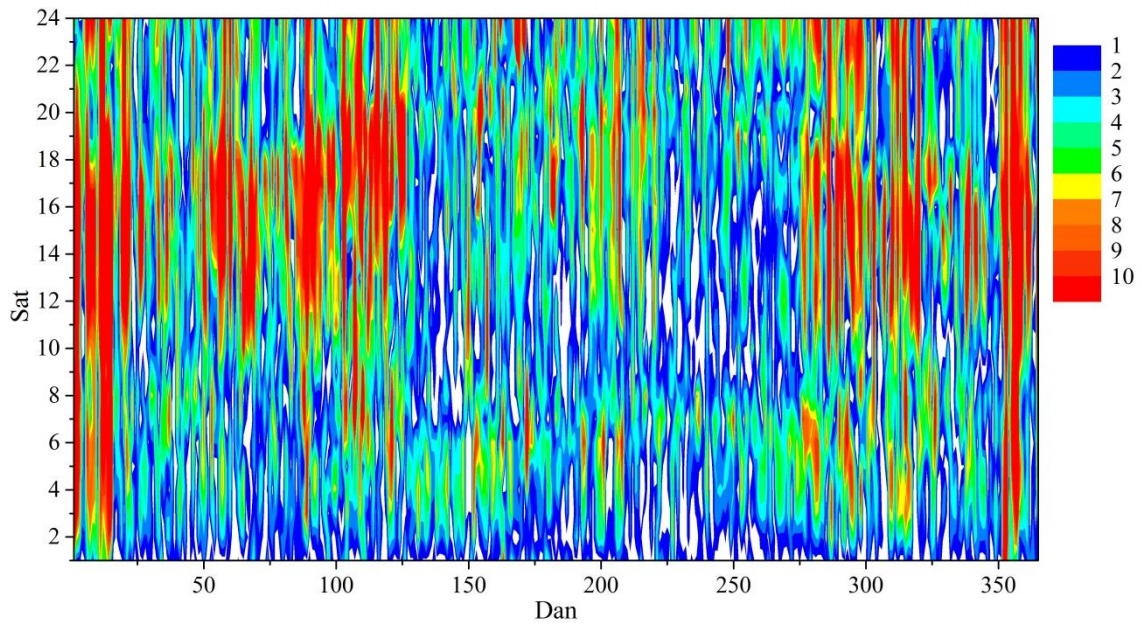
Slika 7.23 Raspodela APE greške svakog sata po danima za DS-ARIMA model



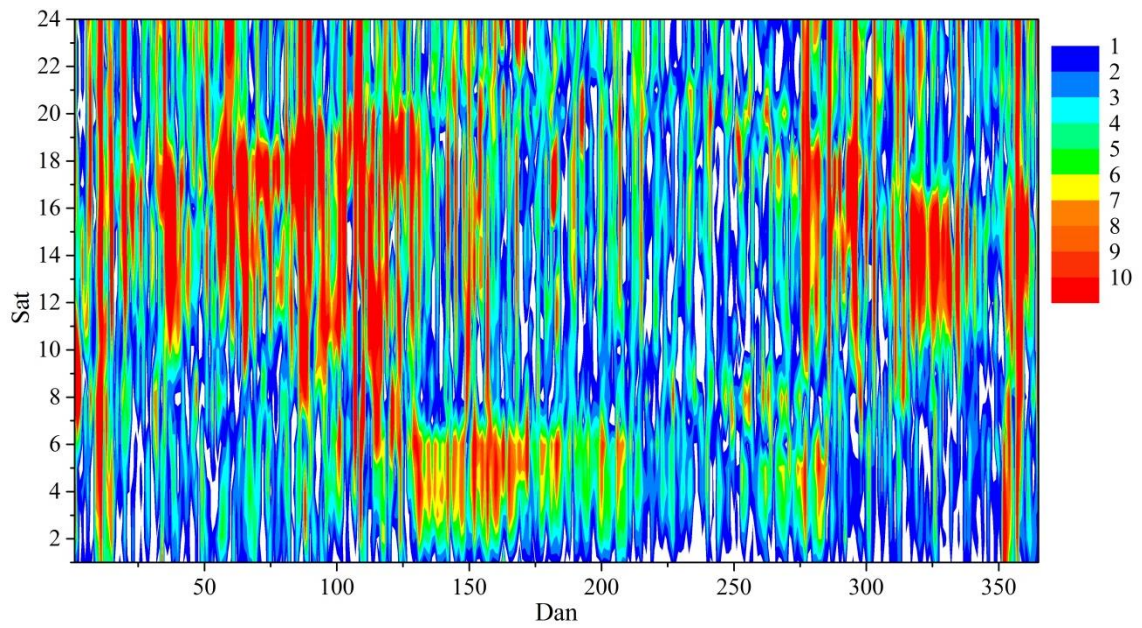
Slika 7.24 Raspodela APE greške svakog sata po danima za DS-HW model



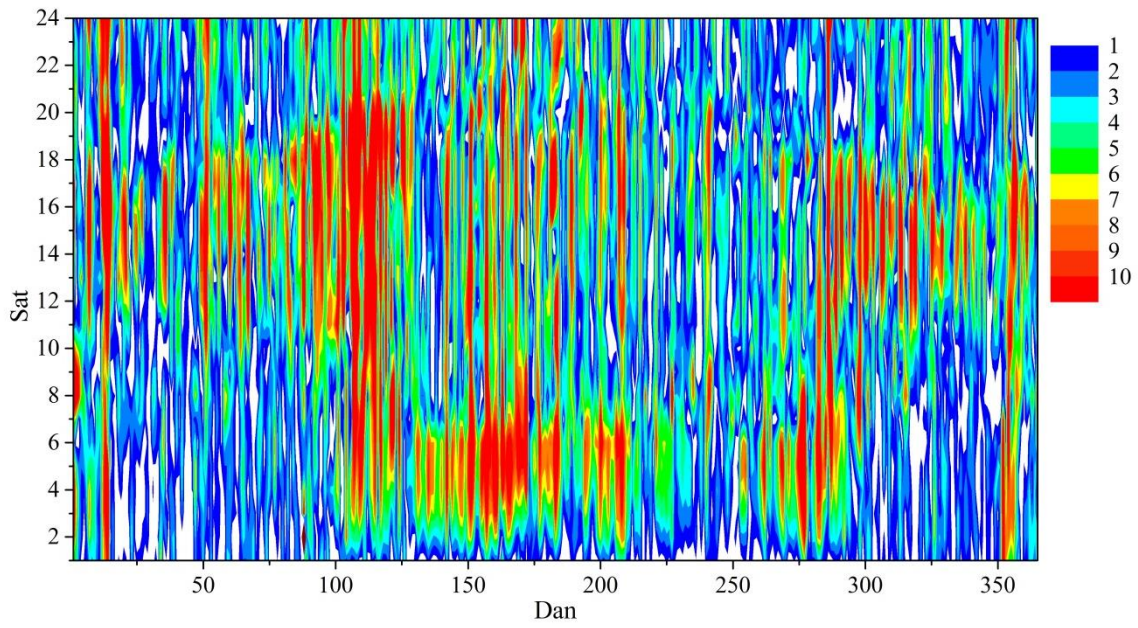
Slika 7.25 Raspodela APE greške svakog sata po danima za ANN model



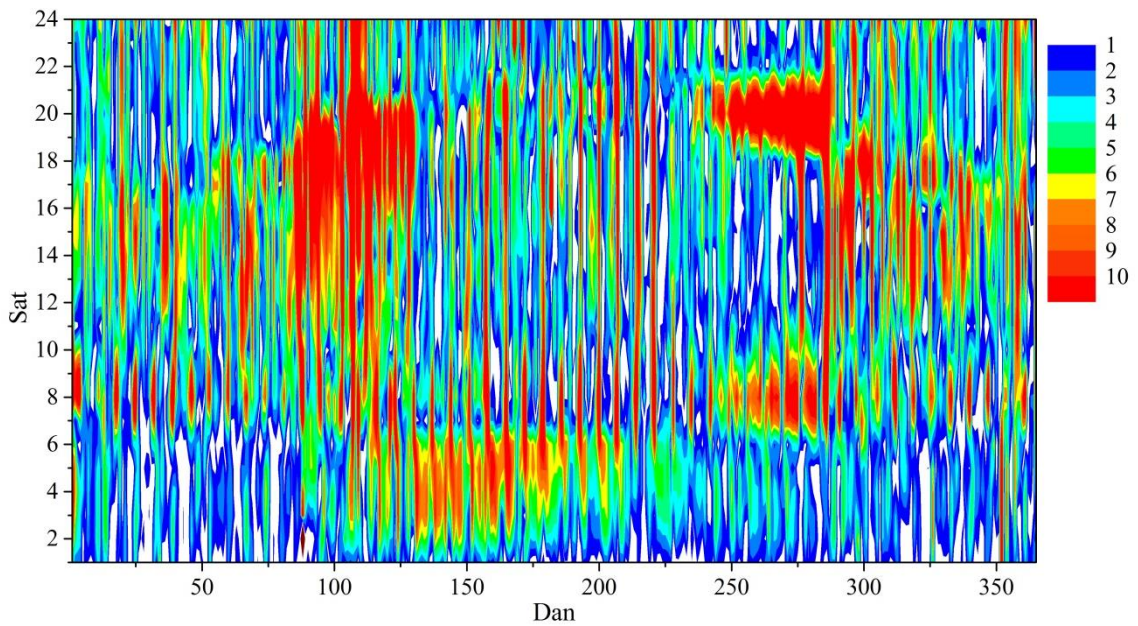
Slika 7.26 Raspodela APE greške svakog sata po danima za RVM model



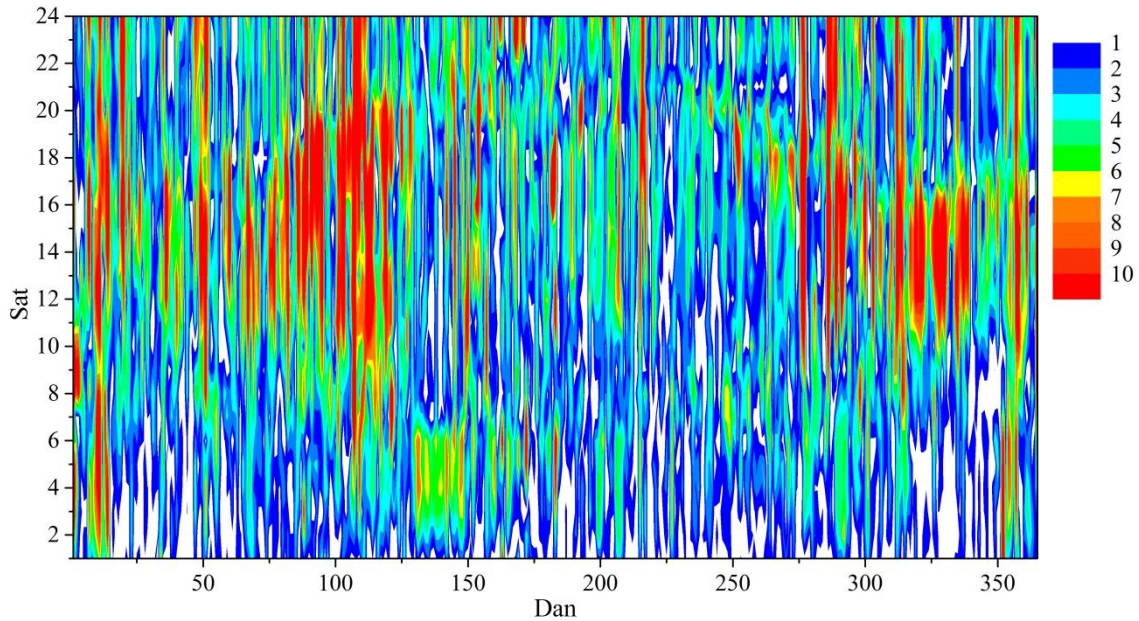
Slika 7.27 Raspodela APE greške svakog sata po danima za LSSVM model



Slika 7.28 Raspodela APE greške svakog sata po danima za 2F-LSSVM model



Slika 7.29 Raspodela APE greške svakog sata po danima za IA-LSSVM model



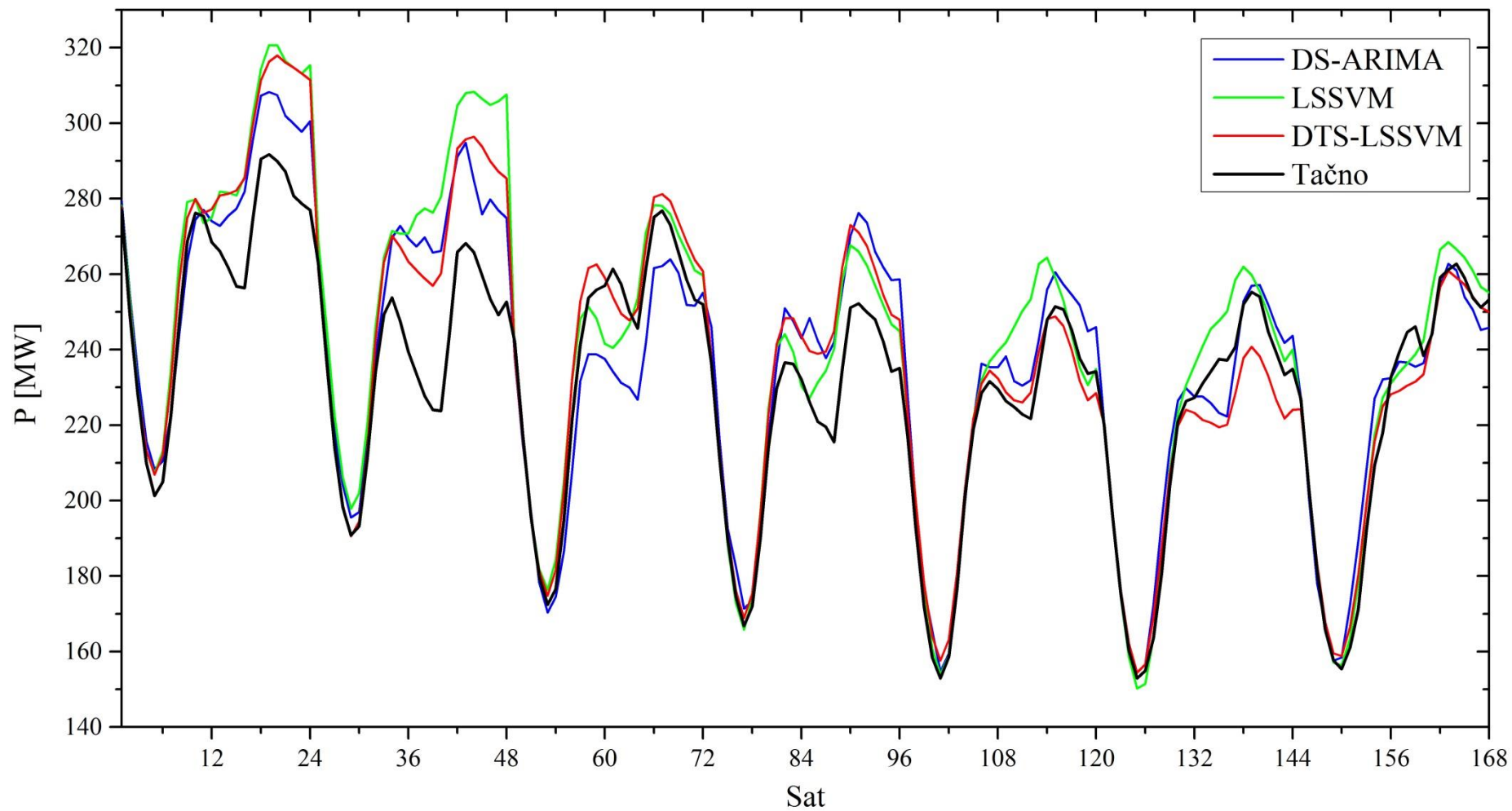
Slika 7.30 Raspodela APE greške svakog sata po danima za DTS-LSSVM model

U Tabeli 7.5 data je prosečna APE greška po satu za svaki model. Prva kolona označava sat u danu, u sledećih osam kolona date su vrednosti prosečne APE greške za konkretan model, a poslednja kolona predstavlja prosečnu vrednost ovih grešaka svih modela po satu u danu. Slike 7.23 - 7.30 i Tabela 7.5 pružaju kompletan uvid u ponašanje modela kroz svaki sat svakog dana. Tako možemo primetiti da se u letnjem periodu godine (od 125. do 275. dana) kod modela IA-LSSVM, 2F-LSSVM, LSSVM, RVM i ANN velike greške javljaju u periodu između 3 i 6 časova. Kod istih modela u zimskom periodu, najveće greške se javljaju između 13 i 20 časova. Dani između 90. i 125. su dani sa najvećim greškama kod svih modela osim RVM modela gde su to dani na početku i kraju godine godine, do 25. i od 350. dana do kraja godine. Ovo se poklapa sa rezultatima u Tabeli 7.3 gde se vidi da se najveće greške javljaju u aprilu. Da je DTS-LSSVM najprecizniji model pokazuje Slika 7.30, na kojoj preovlađuju hladne boje tj. oblasti sa malom greškom a malo je oblasti sa toplim bojama. Malo više oblasti sa toplijim bojama i većom greškom imaju DS-ARIMA i ANN model, Slike 7.23 i 7.25. Ovakav poredak se poklapa sa rezultatima u Tabeli 7.1 gde je na poslednjem mestu model DS-HW što potvrđuje i Slika 7.24 na kojoj preovlađuju toplije boje. Da se najveće greške javljaju između 13 i 20 časova potvrđuju i rezultati u Tabeli 7.5, što je i očekivano jer je to period dana kada su ljudi najaktivniji pa je i potrošnja električne energije povećana u odnosu na ostatak dana. Najmanje greške su u noćnim satima, između 1 i 6 časova je vezano za smanjenje ljudskih aktivnosti i znatno manju potrošnju.

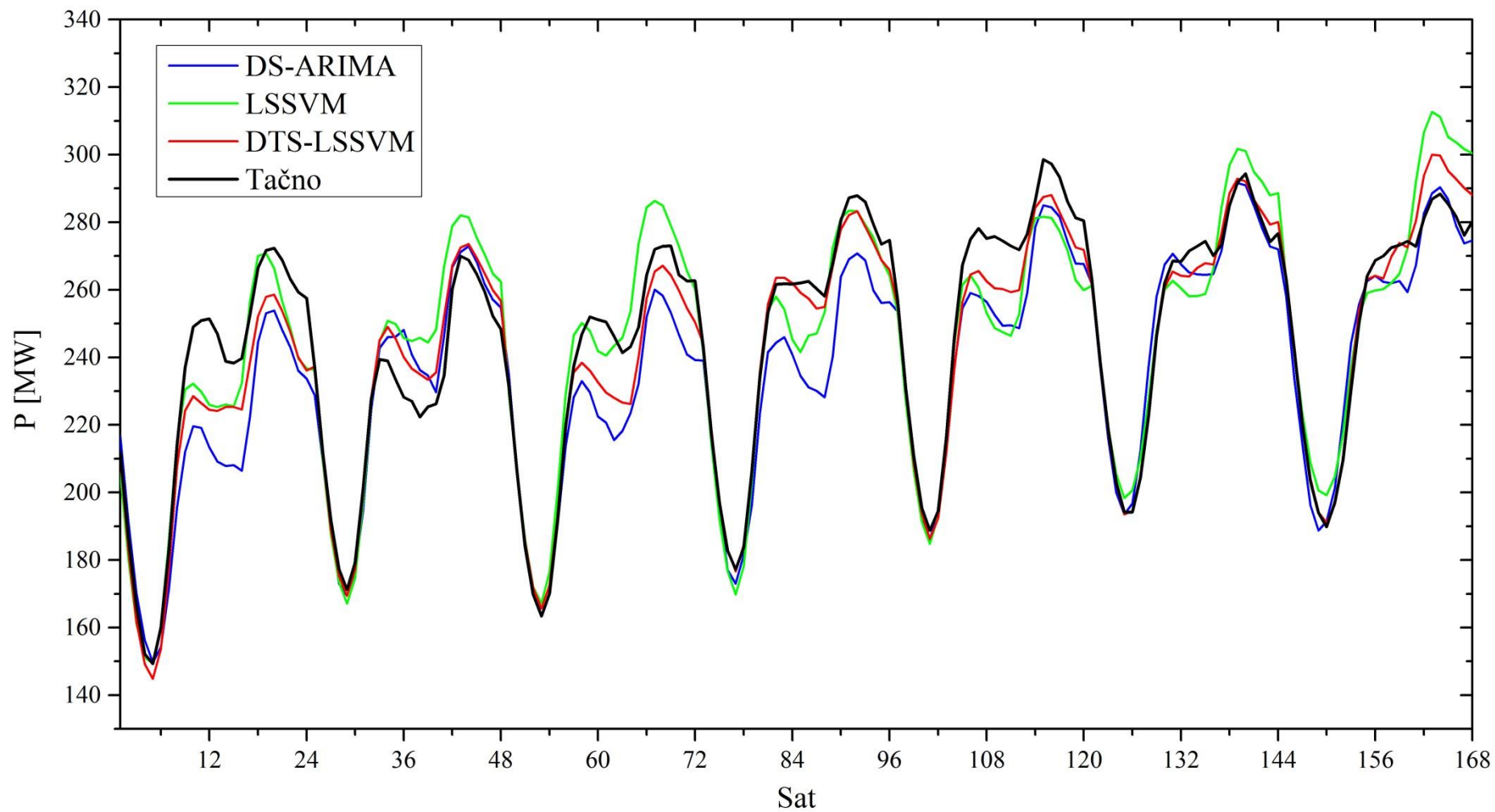
Tabela 7.5 Srednja APE greška po satu

Sat	DS-ARIMA	DS-HW	ANN	RVM	LSSVM	2F-LSSVM	IA-LSSVM	DTS-LSSVM	APE/sat
1	2.34	1.99	1.39	1.49	1.49	1.61	1.47	1.27	1.63
2	3.08	2.76	2.15	2.27	2.55	2.68	2.58	1.85	2.49
3	2.52	3.52	2.69	3.17	3.25	3.53	3.31	2.22	3.03
4	3.29	4.07	3.16	3.72	3.87	4.25	3.71	2.67	3.59
5	2.99	4.41	3.50	3.99	4.09	4.69	3.78	2.85	3.79
6	2.58	4.63	4.07	4.33	4.25	4.82	3.91	3.02	3.95
7	3.67	4.77	4.50	4.20	3.62	4.19	4.63	3.03	4.08
8	4.31	4.82	4.50	3.89	3.76	4.12	5.93	3.27	4.33
9	3.83	4.91	3.74	3.31	3.57	3.67	4.96	3.17	3.90
10	4.03	5.24	3.56	3.46	3.68	3.60	4.01	3.40	3.87
11	4.22	5.65	3.86	4.12	4.47	4.14	3.88	4.07	4.30
12	4.64	6.31	4.41	4.97	5.16	4.64	4.38	4.74	4.91
13	5.27	6.79	4.81	5.59	5.58	5.01	4.83	5.15	5.38
14	5.41	7.17	5.28	6.16	5.86	5.24	5.15	5.36	5.70
15	5.49	7.45	5.59	6.74	6.14	5.70	5.40	5.57	6.01
16	5.77	7.45	5.70	7.02	6.48	5.99	5.72	5.63	6.22
17	5.36	7.18	5.56	7.06	6.72	5.82	6.28	5.29	6.16
18	4.90	7.02	5.12	6.51	6.32	5.33	6.27	4.98	5.81
19	4.76	6.91	4.58	5.64	5.26	4.24	6.52	4.57	5.31
20	4.28	6.66	4.39	5.03	4.81	3.79	6.69	4.39	5.01
21	3.76	6.35	3.94	4.36	4.24	3.38	4.93	3.83	4.35
22	3.99	6.18	4.05	4.60	4.22	3.38	3.85	3.91	4.27
23	4.34	6.39	4.29	5.22	4.49	3.62	4.12	4.32	4.60
24	4.31	6.33	4.32	5.24	4.48	3.69	4.25	4.32	4.62

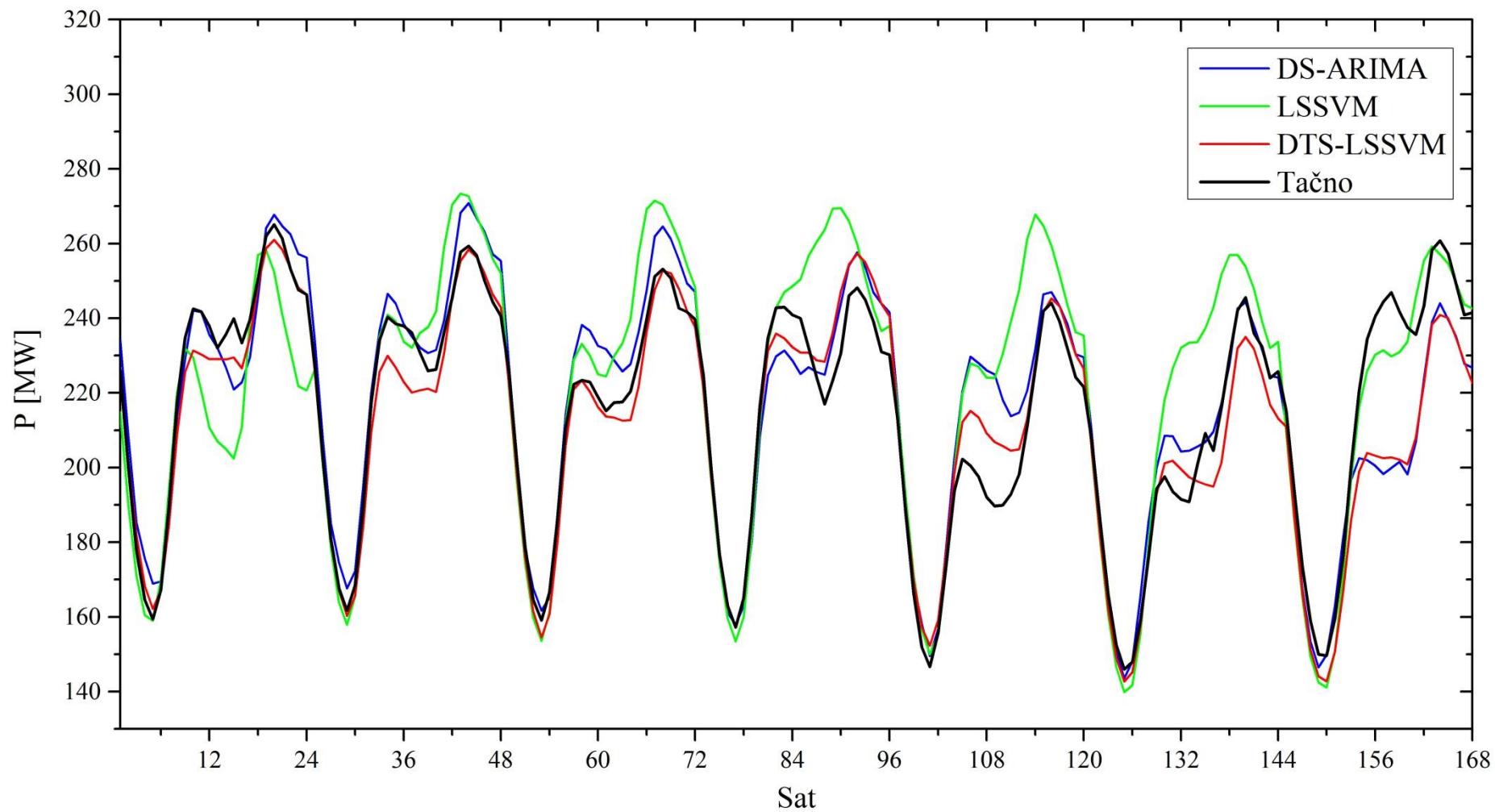
Na Slikama 7.31 - 7.42 prikazane su krive časovnog opterećenja za period od nedelju dana svakog meseca u godini. Horizontalna osa predstavlja opterećenje a vertikalna sat u sedmici. Prikazane su krive tačnog i krive opterećenja DS-ARIMA, LSSVM i DTS-LSSVM modela.



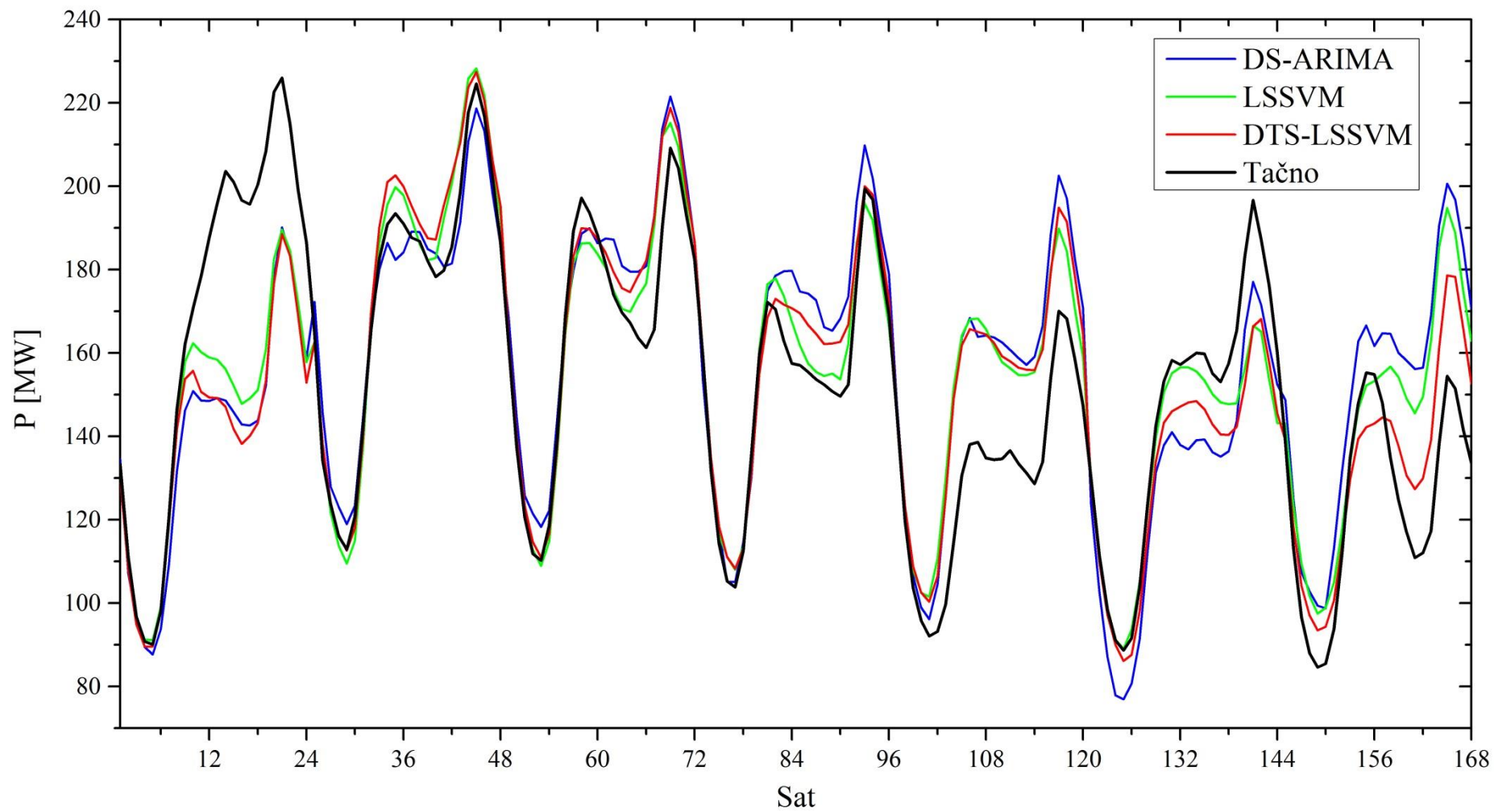
Slika 7.31 Časovno opterećenje za period od 19. do 25. januara



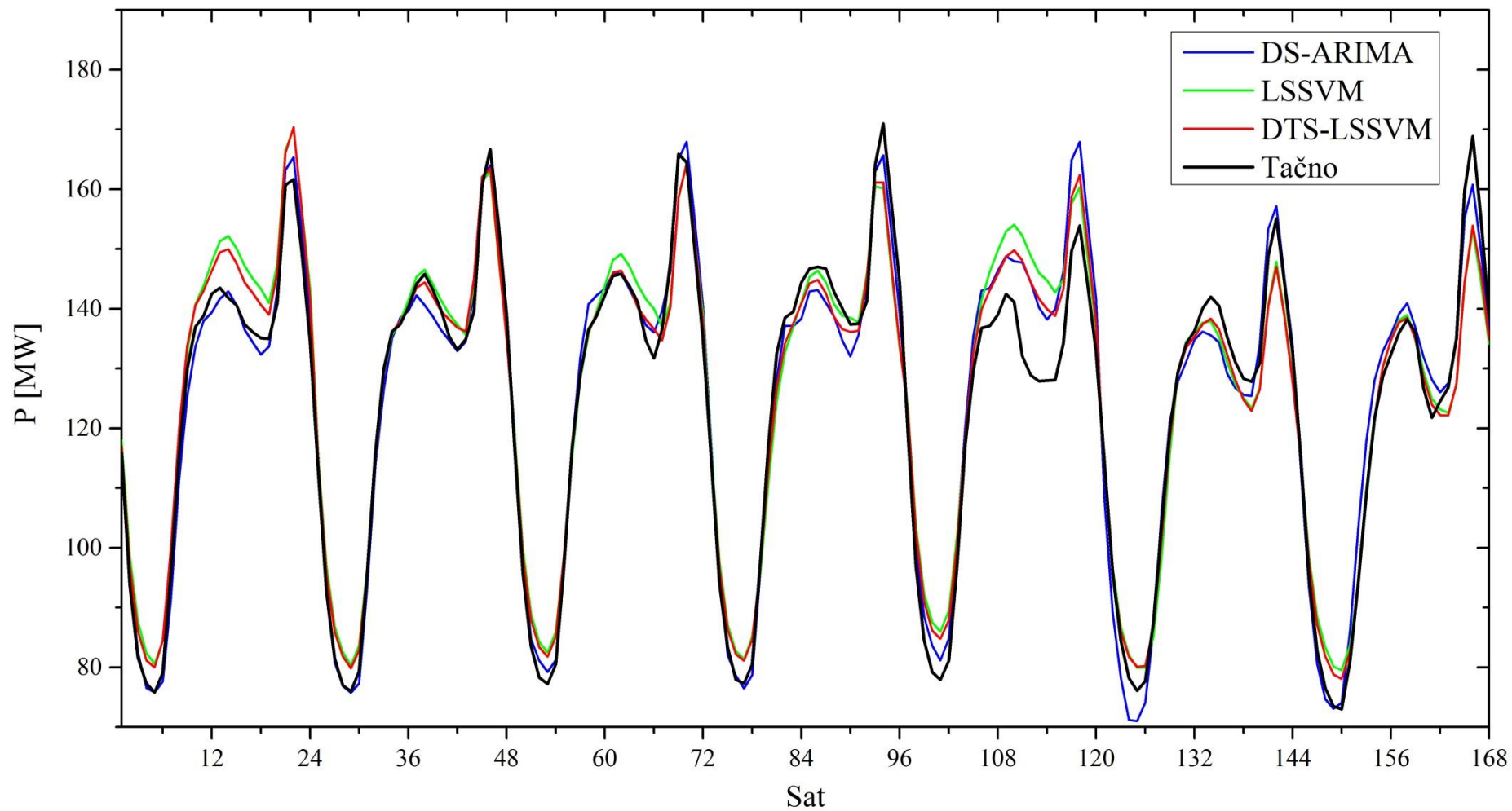
Slika 7.32 Časovno opterećenje za period od 9. do 15. februara



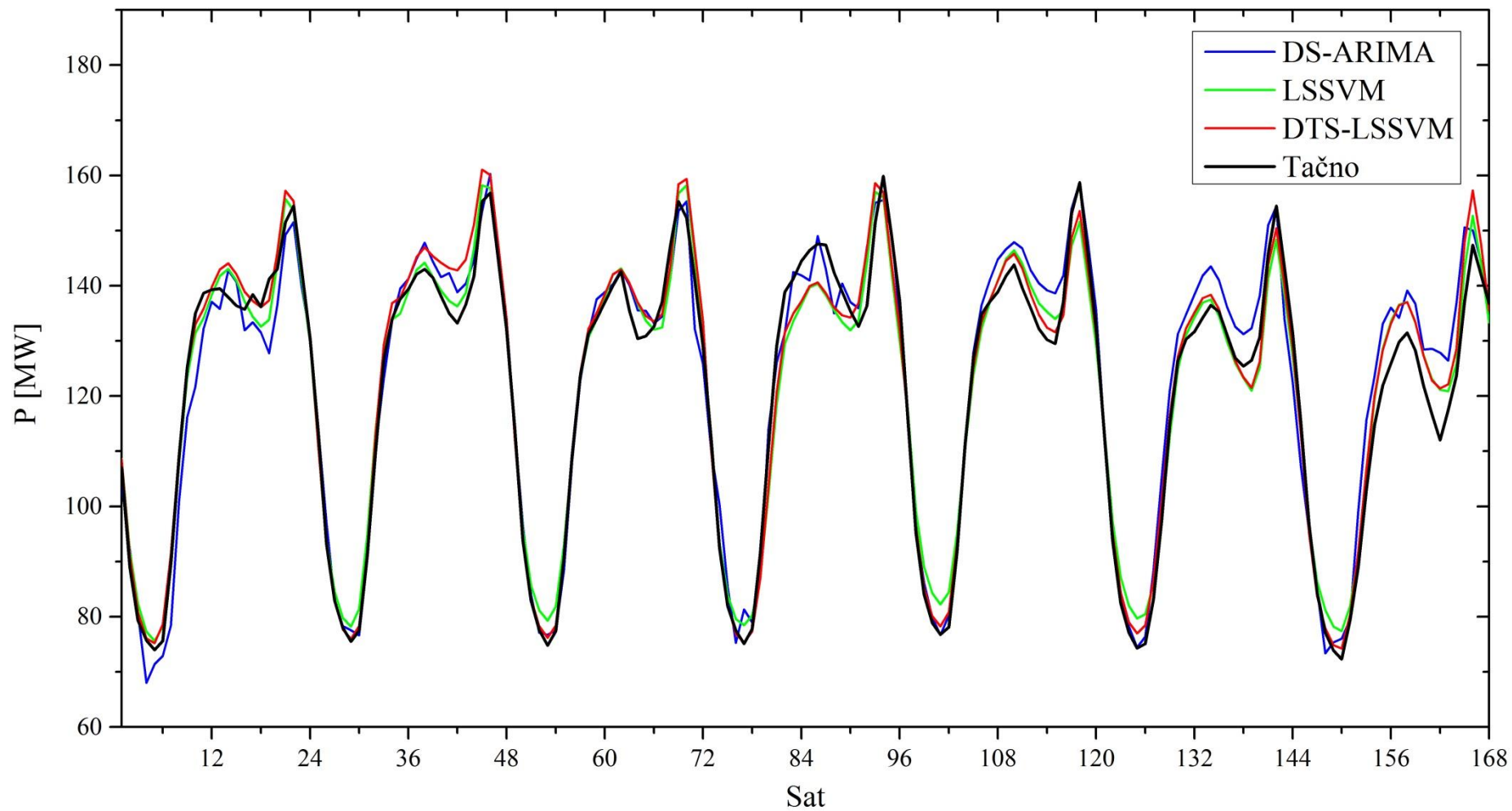
Slika 7.33 Časovno opterećenje za period od 2. do 8. marta



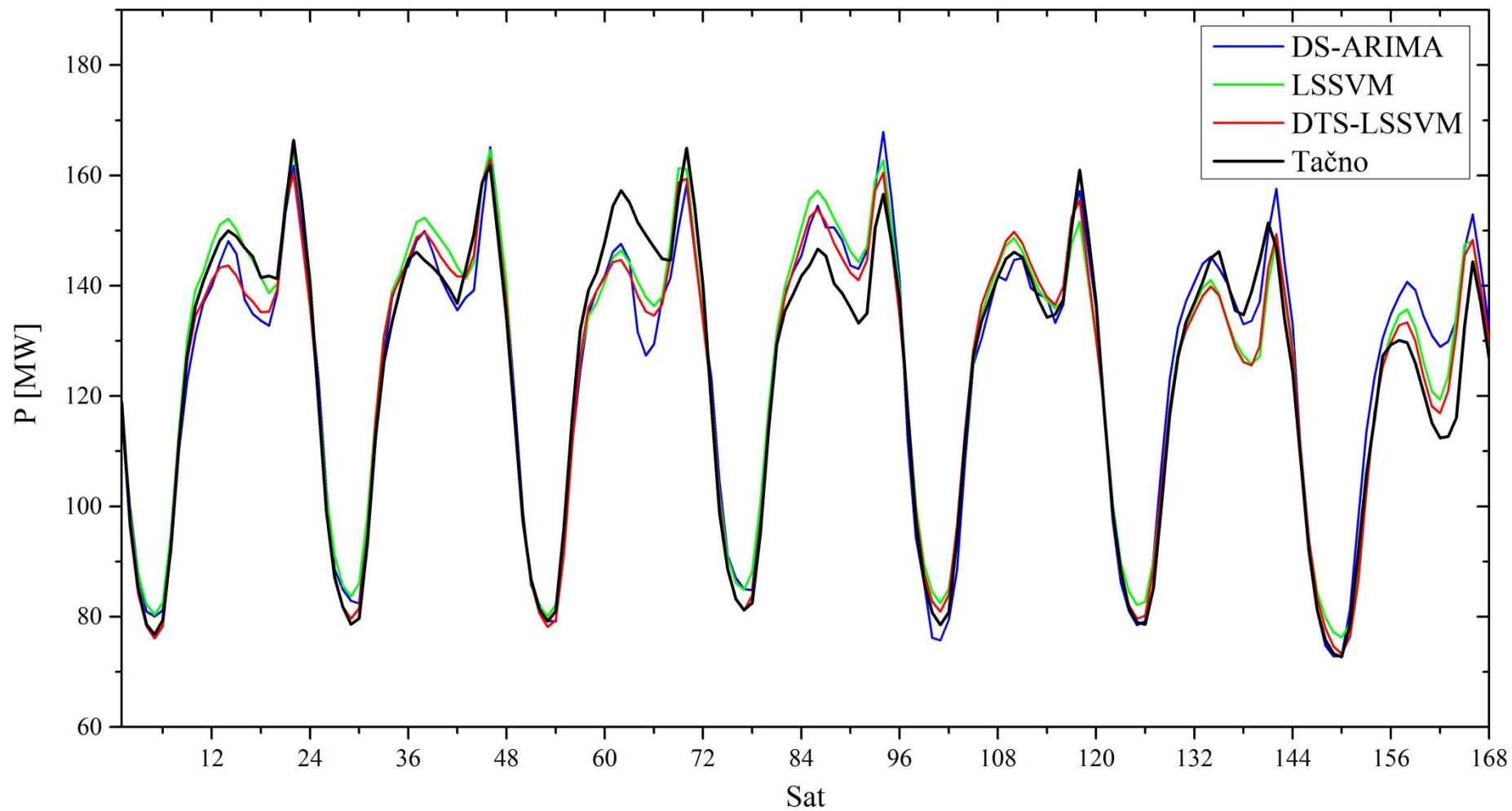
Slika 7.34 Časovno opterećenje za period od 13. do 19. aprila



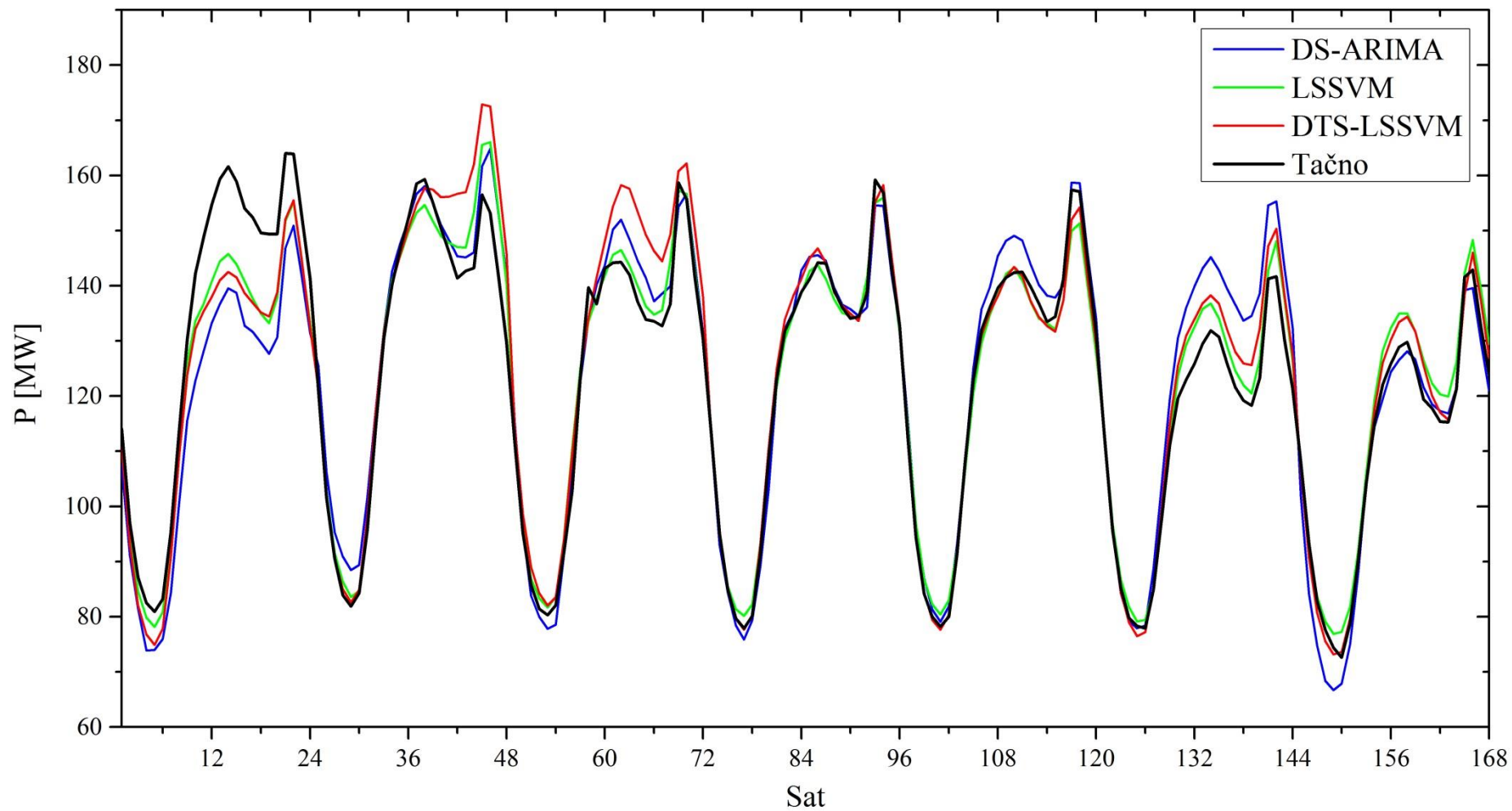
Slika 7.35 Časovno opterećenje za period od 18. do 24. maja



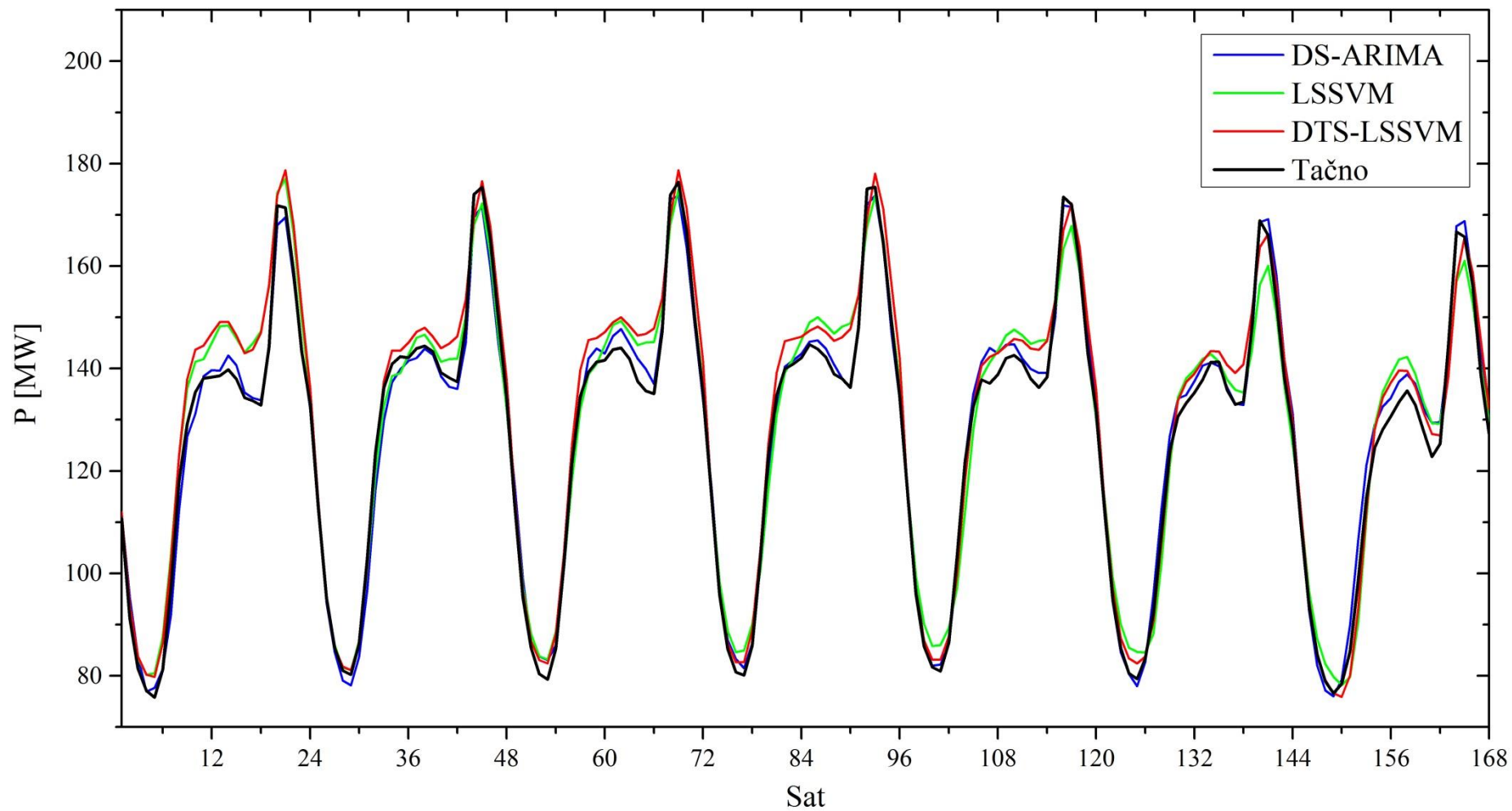
Slika 7.36 Časovno opterećenje za period od 22. do 28. juna



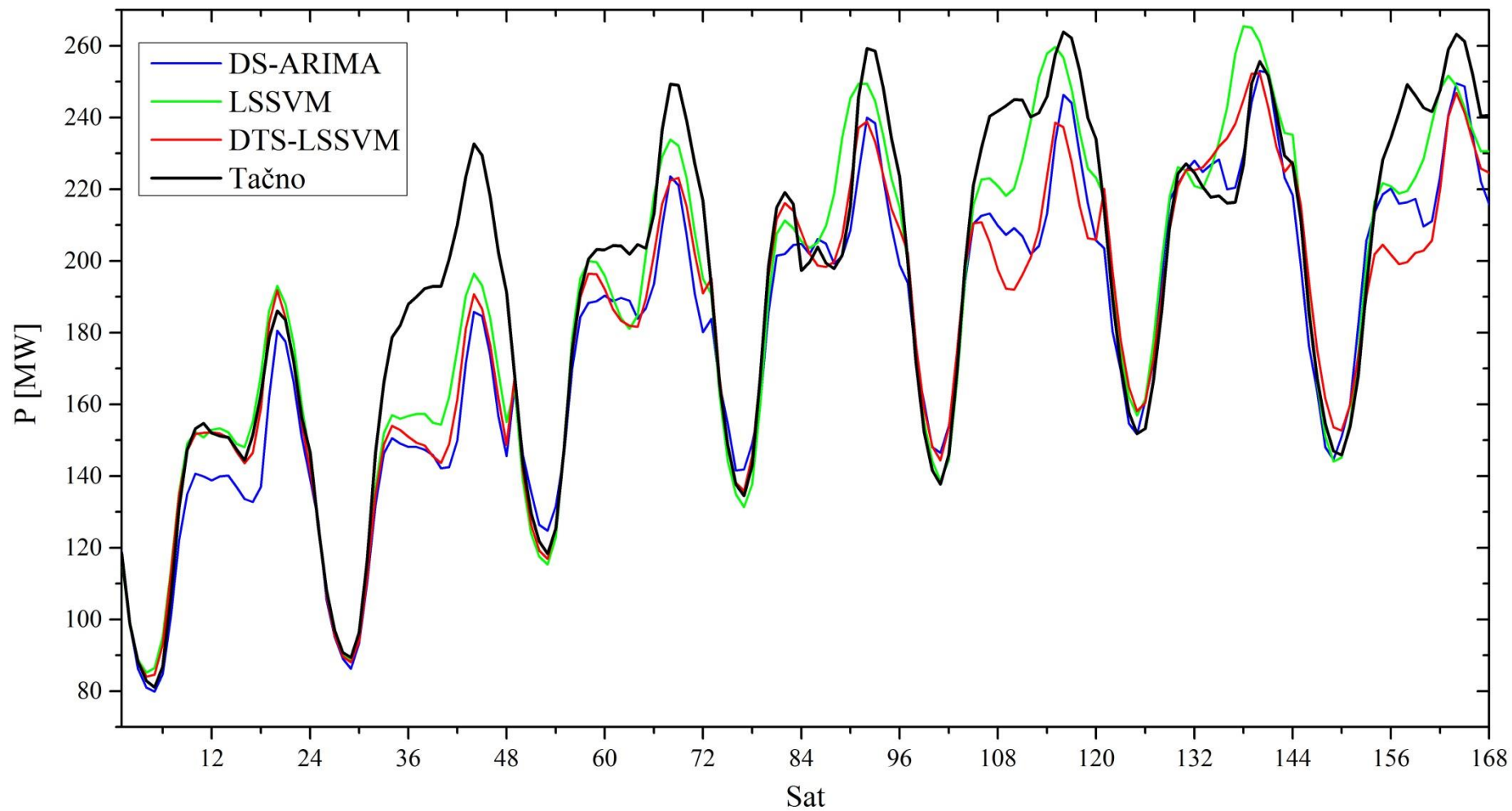
Slika 7.37 Časovno opterećenje za period od 6. do 12. jula



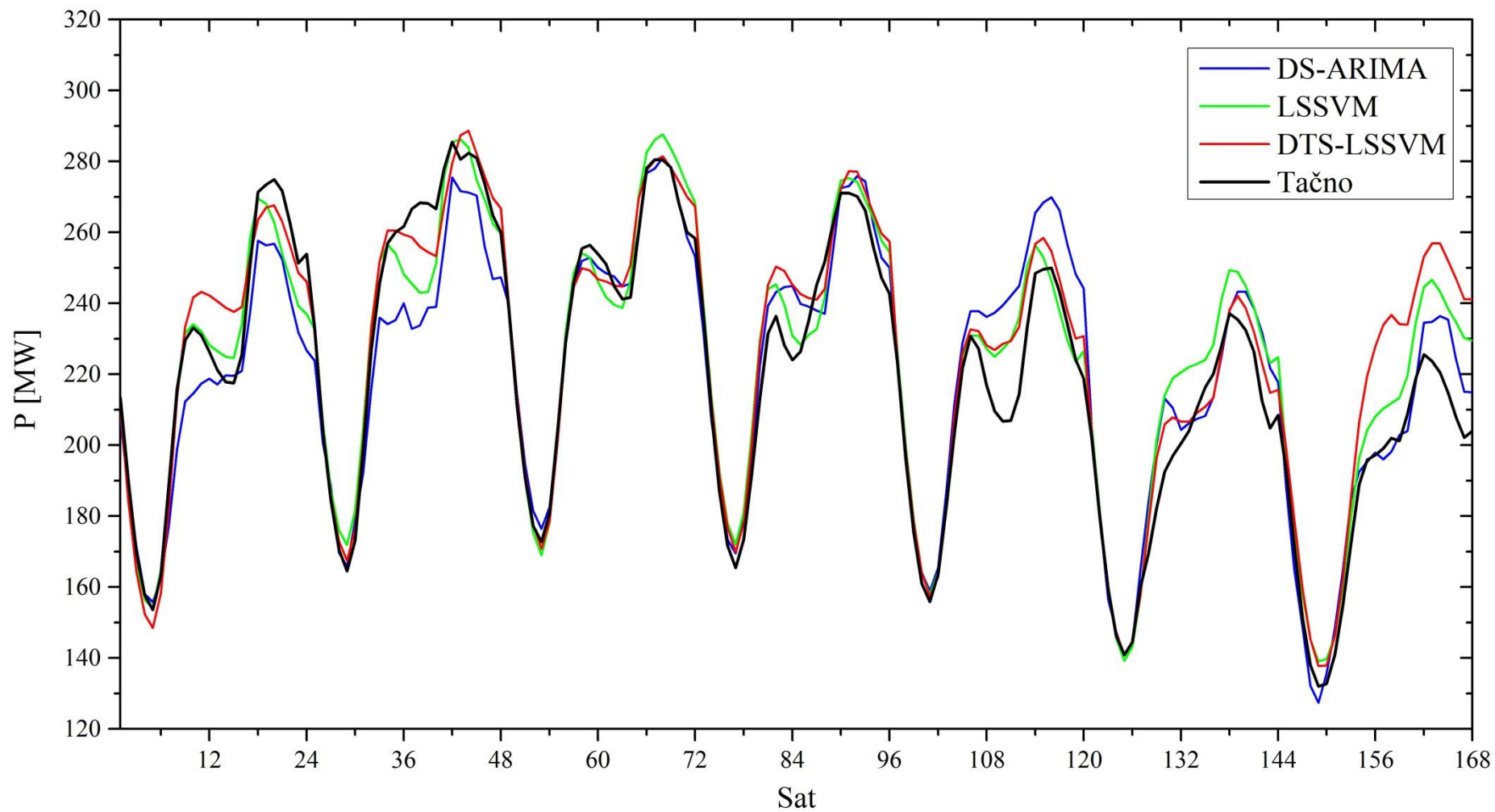
Slika 7.38 Časovno opterećenje za period od 3. do 9. avgusta



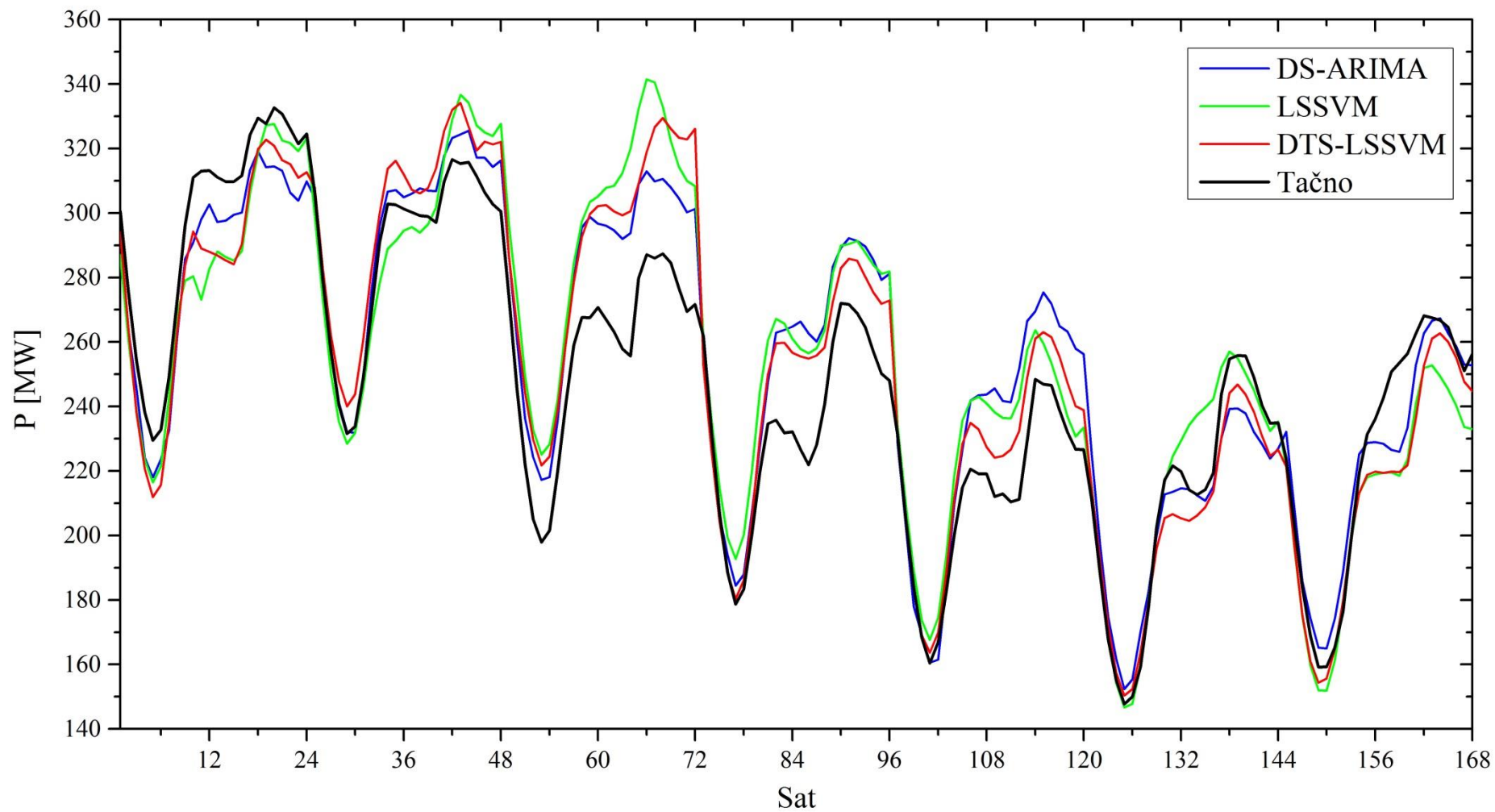
Slika 7.39 Časovno opterećenje za period od 21. do 27. septembra



Slika 7.40 Časovno opterećenje za period od 12. do 18. oktobra



Slika 7.41 Časovno opterećenje za period od 2. do 8. novembra



Slika 7.42 Časovno opterećenje za period od 21. do 27. decembra

Modeli DS-ARIMA, LSSVM i DTS-LSSVM čije su krive časovnog opterećenja za period od sedam dana, za svaki mesec u godini prikazane na Slikama 7.31 - 7.42, izabrani su iz više razloga. DTS-LSSVM model je izabran jer je ukupno gledano najprecizniji model, LSSVM model predstavlja osnovnu varijantu modela na kojoj se baziraju predloženi modeli i DS-ARIMA model je izabran jer predstavlja model iz grupe klasičnih modela a ujedno to je i prvi sledeći model po preciznosti posle DTS-LSSVM modela. Takođe, kako su rezultati DS-ARIMA modela skoro identični rezultatima ANN modela, nema potrebe prikazati krive i jednog i drugog.

Slike 7.31 - 7.42 se odnose na period od sedam dana, gde je prvi dan uvek ponedeljak a svaka kriva ima 168 vrednosti. Sa Slike 7.34 na kojoj su date krive opterećenja za jednu sedmicu u aprilu, može se videti kako izgledaju prognozirane krive za dva dana sa velikim MAPE greškama. Radi se o 13. (0 – 24. sat) i 17. aprilu (96. – 120. sat) sa MAPE greškama 15% i 16.9% za DS-ARIMA model, 11.7% i 15.7% za LSSVM, odnosno 14.3% i 18.8% za DTS-LSSVM model. Prema rezultatima u Tabeli 7.3, već je konstantovano da je april mesec u kome su svi modeli pokazali najlošije rezultate. Sa druge strane, u septembru su svi modeli pokazali najveću preciznost a prognozirane krive opterećenja za ovaj mesec prikazane su na Slici 7.39. MAPE greške za prvi dan tj. prva 24 sata su 1.66% za DS-ARIMA, 4.93% za LSSVM i 5.36% za DTS-LSSVM. Može se primetiti da u ovom slučaju DTS-LSSVM model ne daje optimalnu krivu, već je to kriva koju generiše DS-ARIMA model. Naravno ovo nije jedini dan kada DTS-LSSVM model nije najprecizniji, čak se može desiti da u nekim slučajevima za konkretan mesec neki drugi model bude bolji od globalnog pobedničkog modela. U našem slučaju, prema Tabeli 7.3, DS-ARIMA model je precizniji u januaru i novembru od globalnog pobedničkog DTS-LSSVM modela.

8 ZAKLJUČAK

Problem kratkoročne prognoze potrošnje električne energije, intenzivno se izučava od 80-tih godina prošlog veka do danas. Iako se radi o dobro poznatom problemu, potrebe za što optimalnim korišćenjem električne energije neprestano stavljaju nove izazove pred stručnjake, sa zadatkom ostvarivanja bolje preciznosti prognoze potrošnje električne energije. Ovo je i glavni motiv zbog koga se u disertaciji proučava problem kratkoročne prognoze potrošnje električne energije, sa ciljem da se pronađu novi uspešniji modeli za rešavanje ovog problema.

Pregled oblasti dat u Poglavlju 2, pruža uvid u to šta je urađeno na temu kratkoročne prognoze potrošnje električne energije, koje su metode korišćene i sa koliko uspeha. Tako, dominacija metoda veštačke inteligencije i rezultati koje one postižu u većini slučajeva, ukazuju da se primenom ovih metoda na inovativan način, mogu prevazići trenutna dostignuća u oblasti. Krećući se u tom smeru, zbog lakšeg shvatanja koncepta na koji se baziraju ove metode, u Poglavlju 3 su dati osnovni pojmovi nadgledanog mašinskog učenja i predstavljena je rekurzivna strategija prognoze vremenskih serija koja je korišćena kod predloženih modela u disertaciji. Matematička postavka modela koji su često korišćeni kod problema prognoze potrošnje, zbog čega su izabrani za komparativne modele kod evaluacije predloženih modela u disertaciji, data je u Poglavlju 4. Definisane strukture modela kroz izbor atributa, bilo ekspertski ili nekom metodom, važan je proces kod formiranja modela mašinskog učenja. Osnovne karakteristike krive električnog opterećenja prikazane su u Poglavlju 5 kako bi se izabrao optimalan skup atributa. Predloženi modeli koji se baziraju na LS-SVM metodi za formiranje nelinearnog regresionog modela sa strukturom koju čine prethodno izabrani atributi, predstavljeni su u Poglavlju 6. Validacija predloženih modela rađena je kroz eksperimente čije su osnovne postavke sa rezultatima date u Poglavlju 7.

U disertaciji su predložena tri modela koji za formiranje nelinearnog regresionog modela koriste LS-SVM metod mašinskog učenja sa strukturom atributa koji su izabrani ekspertskom analizom. Prvi predloženi 2F-LSSVM model je model sa prognozom u dve faze. Ovaj model se bazira na činjenici da se LSSVM model sa određenom strukturom atributa može poboljšati dodavanjem dodatnog atributa postojećem skupu atributa, ukoliko je dodatni atribut tako izabran da pruža nove korisne informacije modelu. Tako je za dodatni atribut uzeto vršno dnevno opterećenje i formiran je model sa novom strukturom atributa. Međutim, ovaj dodatni atribut koji je priključen nije poznat u trenutku prognoze za dan unapred pa je jedino moguće uključiti njegovu prognoziranu vrednost. Zbog toga je u prvoj fazi, formiran model za prognozu vršnog dnevnog opterećenja koja se zatim koristi u drugoj fazi kod formiranja modela prognoze časovnog dnevnog opterećenja.

Drugi predloženi model je IA-LSSVM model, koji predstavlja model sa izborom atributa koristeći pretragu napred-nazad sa količinom zajedničkih informacija kao kriterijumom izbora. Model polazi od inicijalnog skupa atributa dobijenog ekspertskom analizom problema i pretragom bira optimalan skup na osnovu zadatog kriterijuma, u ovom slučaju to je količina zajedničkih informacija. Ovde postoje dva problema koja mogu da predstavljaju prepreku uspešnosti ovog modela. Prvi je da je inicijalan skup atributa obično dobro izabran jer je problem prognoze električne potrošnje dobro poznat i kvalitetnom ekspertskom analizom moguće je izabrati skoro optimalan skup atributa. Drugi je da je broj kombinacija prilikom pretrage, korišćenjem potpune pretrage izuzetno veliki, pa je praktična primena nemoguća. Kod pretrage napred-nazad sa druge strane nije zagarantovano nalaženje optimalnog skupa atributa. Iako ove prepreke postoje, ne znači da izbor atributa neće doprineti poboljšanju modela. Naročito u situacijama kada ne postoji dobra ekspertstka analiza problema, iz inicijalno generisanog skupa atributa ovaj model može izabrati ključne attribute koji su optimalni za formiranje modela.

Treći predloženi DTS-LSSVM model, predstavlja model sa dopunom trening skupa. Ovaj model se bazira na pretpostavci da se preciznost osnovnog modela može poboljšati ukoliko se model stalno „osvežava“ i ima priliv novih informacija. Ovakav pristup se lako ostvaruje ukoliko se model retrenira na određenim vremenskim intervalima koristeći novi trening skup formiran od starog trening skupa, izbacivanjem trening vektora za vremenski period jednak intervalu retreniranja koji su vremenski najdalji trenutku prognoze. Umesto ovih trening vektora, ubacuju se novi, „sveži“, koji se odnose na prethodni interval. Na ovaj

način, model uvek odgovara poslednjem scenariju prognoze jer se vremenski ne udaljava od njega sa svakim sledećim korakom prognoze.

Evaluacija predloženih modela rađena je kroz komparativnu analizu sa izabranim modelima: DS-ARIMA i DS-HW modeli iz grupe klasičnih modela i ANN, RVM i LSSVM modeli iz grupe modela veštačke inteligencije. Iako su svrstani u grupu klasičnih modela, DS-ARIMA i DS-HW modeli predstavljaju modifikacije standardnih ARIMA, odnosno modela eksponencijalnog izravnjanja, koji uvažavaju dvostuku periodičnost električnog opterećenja, pa su samim tim u prednosti u odnosu na njihove standardne definicije. Vrlo često korišćen model, ne samo kod prognoze električne potrošnje već i šire, je ANN model koji je zbog velike primene i izabran da bude jedan od komparativnih modela. Kako se svi predloženi modeli baziraju na metodi LS-SVM, za komparativni model je uzet i standardni LSSVM model da bi se uradilo direktno poređenje osnovnog sa predloženim modelima. Poređenje je rađeno i sa RVM modelom, koji predstavlja najnoviju modifikaciju SVM metoda i u literaturi je označen kao metod od koga se očekuju bolji rezultati kod problema gde se primenjuje i SVM metod.

Rezultati prognoze časovnog opterećenja za period od godinu dana, pokazali su da je sveobuhvatno gledano pobednički model, DTS-LSSVM model. Prati ga DS-ARIMA, ANN, 2F-LSSVM, LSSVM, IA-LSSVM, RVM i na začelju je DS-HW model. Iako je DTS-LSSVM najbolji model, iznenađujuć je rezultat koji je postigao DS-ARIMA model. Ovaj model pokazao se veoma uspešnim a u dva meseca je čak bio bolji od pobedničkog modela. Njemu sličan rezultat, ostvario je ANN model, koji je takođe iznenadio jer se radi o najjednostavnijoj konfiguraciji neuronske mreže, sa samo jednim skrivenim slojem i svega 24 neurona u njemu. Negativno iznenađenje je rezultat koji je ostvario RVM model od koga se očekivalo više prema natpisima u literaturi. Standardni LSSVM model je malo podbacio, međutim razlog za takav rezultat leži u činjenici da ovaj model nije treniran kao i većina sa podacima za celu godinu, na račun postizanja veće brzine.

Što se tiče uspeha koji su postigli predloženi modeli, on je u skladu sa pretpostavkama od kojih se krenulo pri formiranju istih. Na začelju predloženih modela je IA-LSSVM model, što je očekivano jer je inicijalni skup atributa izabran dobrom ekspertskom analizom i blizak je optimalnom skupu atributa za konkretan set podataka. Ovaj model koristio je samo dva atributa a ipak je postigao solidan rezultat. Međutim, iako se ovaj model nalazi na poslednjem mestu od predloženih modela, postoje dani kada on nadmašuje pobednički DTS-LSSVM model. 2F-LSSVM je model koji je uzet kao realan za razliku od dva modela sa prognozom u

dve faze koji koriste tačne vrednosti vršnog dnevnog opterećenja, odnosno srednje dnevne temperature. Kada koristi tačno vršno dnevno opterećenje, ovaj model izbija na prvo mesto, ispred DTS-LSSVM modela. Ipak, ovakav scenario je nerealan jer ta vrednost nije dostupna u trenutku prognoze. Rezultat koji je postigao 2F-LSSVM model se može dodatno poboljšati ukoliko se poveća preciznost prognoze u prvoj fazi, na čemu treba raditi u budućnosti. Pretpostavka od koje se pošlo kod formiranja DTS-LSSVM modela se definitivno ispostavila tačnom jer je ovaj model postigao najbolji rezultat. Međutim, ovaj model verovatno može postići još bolji rezultat ukoliko bi se dopuna trening skupa radila bez odbacivanja najstarijih vektora. Razlog zašto to nije rađeno u disertaciji, leži u potrebnom velikom vremenu formiranja modela koje raste eksponencijalno sa porastom broja vektora, pa sam model u jednom trenutku postaje nepraktičan. Od svih predloženih modela, pobednički model je definitivno najjednostavniji i najlakše primenljiv u praksi. Kako su za testiranje korišćeni realni podaci i simuliran realan scenario prognoze, DTS-LSSVM je svakako model koji se može uspešno koristiti u elektroenergetskim preduzećima.

Sprovedena istraživanja u disertaciji su dala odgovore na mnoga pitanja vezana za kratkoročnu prognozu potrošnje električne energije. Ipak, kroz ova istraživanja otvoreno je i mnogo novih pitanja koja čekaju odgovor, a koja ujedno treba da predstavljaju dalje pravce istraživanja. Tako treba ispitati mogućnost kombinovanja IA-LSSVM modela sa pobedničkim DTS-LSSVM modelom kako bi se prevazišle situacije kada DTS-LSSVM model podbaci a IA-LSSVM model poboljša rezultat. Ovak pristup se svodi na formiranje hibridnih modela, pa u tom pravcu i treba nastaviti istraživanja kombinovanjem više različitih modela npr. DTS-LSSVM, DS-ARIMA i ANN modela. Drugi pravac istraživanja je posledica neujednačene preciznosti modela u različitim periodima u toku godine. Tako treba ispitati mogućnost formiranja po jednog optimalnog modela za svaki period godine. Ovakav pristup nije ispitan u disertaciji zbog ograničenog skupa podataka koji je bio dostupan za period od dve godine, pri čemu su za treniranje modela bili dostupni samo podaci za godinu dana a podaci za preostalih godinu dana su korišćeni za testiranje.

Predloženi modeli u disertaciji su testirani samo na jednom skupu podataka. Poželjno je testiranja izvršiti na više različitih skupova podataka a razlog zašto to nije urađeno jeste ograničen obim disertacije. Sa druge strane, period testiranja se odnosi na period od godinu dana što je veliki period testiranja kada je u pitanju kratkoročna prognoza potrošnje električne energije ali poželjan, jer daje kompletnu sliku o ponašanju modela tokom različitih godišnjih doba, meseci, dana. Takođe treba spomenuti da se podaci za testiranje odnose na mali

potrošački konzum što dodatno komplikuje problem. U literaturi se za testiranje najčešće koriste podaci vezani za velike potrošačke konzume koji se obično odnose na električnu potrošnju na teritoriji čitave zemlje. I pored toga, dobijeni rezultati u disertaciji su u rangu najboljih rezultata u literaturi.

LITERATURA

- [1] T. Hong, "Short Term Electric Load Forecasting," North Carolina State University, 2010.
- [2] H. Lee Willis, *Spatial Electric Load Forecasting*. CRC Press, 2002.
- [3] A. D. Papalexopoulos and T. C. Hesterberg, "A regression-based approach to short-term system load forecasting," *IEEE Trans. Power Syst.*, vol. 5, no. 4, pp. 1535–1550, 1990.
- [4] T. Haida and S. Muto, "REGRESSION BASED PEAK LOAD FORECASTING," *IEEE Trans. Power Syst.*, vol. 9, no. 4, pp. 1788–1794, 1994.
- [5] S. Ružić, A. Vučković, and N. Nikolić, "Weather Sensitive Method for Short Term Load Forecasting in Electric Power Utility of Serbia," *IEEE Trans. Power Syst.*, vol. 18, no. 4, pp. 1581–1586, 2003.
- [6] W. Charytoniuk, M. S. Chen, and P. Van Olinda, "Nonparametric regression based short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 13, no. 3, pp. 725–730, 1998.
- [7] W. Charytoniuk, M.-S. Chen, P. Kotas, and P. Van Olinda, "Demand forecasting in power distribution systems using nonparametric probability density estimation," *Power Syst. IEEE Trans.*, vol. 14, no. 4, pp. 1200–1206, 1999.
- [8] W. R. Christiaanse, "SHORT-TERM LOAD FORECASTING USING GENERAL EXPONENTIAL SMOOTHING W.," *IEEE Trans. POWER Appar. Syst.*, vol. PAS-90, no. 2, pp. 900–911, 1971.
- [9] M. T. Hagan and S. M. Behr, "The Time Series Approach to Short Term Load Forecasting," *IEEE Trans. Power Syst.*, vol. PWRS-2, no. 3, pp. 785–791, 1987.
- [10] B. Krough, E. S. De Llinas, and D. Lesser, "Design and Implementation of an On-Line Load Forecasting Algorithm," *Power Eng. Rev.*, vol. PER-2, no. 9, p. 47, 1982.
- [11] G. A. N. Mbamalu and M. E. El-Hawary, "Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation," *IEEE Trans. Power Syst.*, vol. 8, no. 1, pp. 343–348, 1993.
- [12] I. Rahman and S. Moghram, "Analysis and evaluation of five short-term load forecasting techniques," *IEEE Trans. Power Syst.*, vol. 4, no. 4, pp. 1484–1491, 1989.
- [13] J. W. Taylor, "Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing," *J. Oper. Res. Soc.*, vol. 54, no. 0, pp. 799–805, 2003.
- [14] W. Laing and D. Smith, "A comparison of time series forecasting methods for predicting the CEGB demand," in *Proceedings of the Ninth Power Systems Computation Conference*, 1987.

-
- [15] H. M. Al-Hamadi and S. a. Soliman, "Short-term electric load forecasting based on Kalman filtering algorithm with moving window weather and load model," *Electr. Power Syst. Res.*, vol. 68, no. 1, pp. 47–59, Jan. 2004.
- [16] N. Amjady, "Short-term hourly load forecasting using time-series modeling with peak load estimation capability," *IEEE Trans. Power Syst.*, vol. 16, no. 4, pp. 798–805, 2001.
- [17] S. Huang, S. Member, and K. Shih, "Short-Term Load Forecasting Via ARMA Model Identification Including Non-Gaussian," *Power Syst. IEEE Trans.*, vol. 18, no. 2, pp. 673–679, 2003.
- [18] S. Saab, E. Badr, and G. Nasr, "Univariate modeling and forecasting of energy consumption: the case of electricity in Lebanon," *Energy*, vol. 26, no. 1, pp. 1–14, Jan. 2001.
- [19] R. Campo and P. Ruiz, "Adaptive Weather-Sensitive Short Term Load Forecast," *IEEE Trans. Power Syst.*, vol. PWRS-2, no. 3, pp. 592–598, 1987.
- [20] D. G. Infield and D. C. Hill, "Optimal smoothing for trend removal in short term electricity demand forecasting," *IEEE Trans. Power Syst.*, vol. 13, no. 3, pp. 1115–1120, 1998.
- [21] a. M. Al-Kandari, S. a. Soliman, and M. E. El-Hawary, "Fuzzy short-term electric load forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 26, no. 2, pp. 111–122, Feb. 2004.
- [22] D. K. Ranaweera, N. F. Hubele, and G. G. Karady, "Fuzzy logic for short term load forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 18, no. 4, pp. 215–222, May 1996.
- [23] S. Chenthur Pandian, K. Duraiswamy, C. Christober Asir Rajan, and N. Kanagaraj, "Fuzzy approach for short term load forecasting," *Electr. Power Syst. Res.*, vol. 76, no. 6–7, pp. 541–548, Apr. 2006.
- [24] S. Rahman and Rahul Bhatnagar, "An Expert System Based Algorithm For Short Term Load Forecasting," *IEEE Trans. Power Syst.*, vol. 3, no. 2, pp. 392–399, 1988.
- [25] D. C. Park, M. a. El-Sharkawi, R. J. Marks, L. E. Atlas, and M. J. Damborg, "Electric load forecasting using an artificial neural network," *IEEE Trans. Power Syst.*, vol. 6, no. 2, pp. 442–449, May 1991.
- [26] G. A. Darbellay and M. Slama, "Forecasting the short-term demand for electricity Do neural networks stand a better chance?," *Int. J. Forecast.*, vol. 16, pp. 71–83, 2000.
- [27] A. K. Topalli, I. Erkmén, and I. Topalli, "Intelligent short-term load forecasting in Turkey," *Int. J. Electr. Power Energy Syst.*, vol. 28, no. 7, pp. 437–447, Sep. 2006.
- [28] M. López, S. Valero, C. Senabre, J. Aparicio, and a. Gabaldon, "Application of SOM neural networks to short-term load forecasting: The Spanish electricity market case study," *Electr. Power Syst. Res.*, vol. 91, pp. 18–27, Oct. 2012.
-

-
- [29] A. Deihimi and H. Showkati, "Application of echo state networks in short-term electric load forecasting," *Energy*, vol. 39, no. 1, pp. 327–340, Mar. 2012.
- [30] N. Amjady and F. Keynia, "Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm," *Energy*, vol. 34, no. 1, pp. 46–57, Jan. 2009.
- [31] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks : The state of the art," *Int. J. Forecast.*, vol. 14, pp. 35–62, 1998.
- [32] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: a review and evaluation," *Power Systems, IEEE Transactions on*, vol. 16, no. 1, pp. 44–55, 2001.
- [33] W. Charytoniuk and M.-S. Chen, "Very short-term load forecasting using artificial neural networks," *Power Systems, IEEE Transactions on*, vol. 15, no. 1, pp. 263–268, 2000.
- [34] M. Djukanovic, S. Ruzic, B. Babic, D. J. Sobajic, and Y.-H. Pao, "A neural-net based short term load forecasting using moving window procedure," *Int. J. Electr. Power Energy Syst.*, vol. 17, no. 6, pp. 391–397, Dec. 1995.
- [35] J. P. S. Catalão, S. J. P. S. Mariano, V. M. F. Mendes, and L. a. F. M. Ferreira, "Short-term electricity prices forecasting in a competitive market: A neural network approach," *Electr. Power Syst. Res.*, vol. 77, no. 10, pp. 1297–1304, Aug. 2007.
- [36] P. Lauret, E. Fock, R. N. Randrianarivony, and J.-F. Manicom-Ramsamy, "Bayesian neural network approach to short time load forecasting," *Energy Convers. Manag.*, vol. 49, no. 5, pp. 1156–1166, May 2008.
- [37] B.-J. Chen, M.-W. Chang, and C.-J. Lin, "Load forecasting using support vector Machines: a study on EUNITE competition 2001," *Power Systems, IEEE Transactions on*, vol. 19, no. 4, pp. 1821–1830, 2004.
- [38] M. Espinoza, J. K. Suykens, and B. Moor, "Load Forecasting Using Fixed-Size Least Squares Support Vector Machines," in *Computational Intelligence and Bioinspired Systems SE - 125*, vol. 3512, J. Cabestany, A. Prieto, and F. Sandoval, Eds. Springer Berlin Heidelberg, 2005, pp. 1018–1026.
- [39] M. Espinoza, J. K. Suykens, and B. Moor, "Fixed-size Least Squares Support Vector Machines: A Large Scale Application in Electrical Load Forecasting," *Comput. Manag. Sci.*, vol. 3, no. 2, pp. 113–129, 2006.
- [40] A. Abraham and B. Nath, "A neuro-fuzzy approach for modelling electricity demand in Victoria," *Appl. Soft Comput.*, vol. 1, no. 2, pp. 127–138, Aug. 2001.
-

-
- [41] S. Fan and L. Chen, "Short-term load forecasting based on an adaptive hybrid method," *Power Systems, IEEE Transactions on*, vol. 21, no. 1, pp. 392–401, 2006.
- [42] P. Mandal, T. Senjyu, and T. Funabashi, "Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market," *Energy Convers. Manag.*, vol. 47, no. 15–16, pp. 2128–2142, Sep. 2006.
- [43] H. T. Nguyen and I. T. Nabney, "Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models," *Energy*, vol. 35, no. 9, pp. 3674–3685, Sep. 2010.
- [44] T. Nengling, J. Stenzel, and W. Hongxiao, "Techniques of applying wavelet transform into combined model for short-term load forecasting," *Electr. Power Syst. Res.*, vol. 76, no. 6–7, pp. 525–533, Apr. 2006.
- [45] M. El-Telbany and F. El-Karmi, "Short-term forecasting of Jordanian electricity demand using particle swarm optimization," *Electr. Power Syst. Res.*, vol. 78, no. 3, pp. 425–433, Mar. 2008.
- [46] W.-C. Hong, "Application of chaotic ant swarm optimization in electric load forecasting," *Energy Policy*, vol. 38, no. 10, pp. 5830–5839, Oct. 2010.
- [47] W.-C. Hong, "Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm," *Energy*, vol. 36, no. 9, pp. 5568–5578, Sep. 2011.
- [48] W.-C. Hong, Y. Dong, W. Y. Zhang, L.-Y. Chen, and B. K. Panigrahi, "Cyclic electric load forecasting by seasonal SVR with chaotic genetic algorithm," *Int. J. Electr. Power Energy Syst.*, vol. 44, no. 1, pp. 604–614, Jan. 2013.
- [49] W.-C. Hong, "Hybrid evolutionary algorithms in a SVR-based electric load forecasting model," *Int. J. Electr. Power Energy Syst.*, vol. 31, no. 7–8, pp. 409–417, Sep. 2009.
- [50] W.-C. Hong, "Chaotic particle swarm optimization algorithm in a support vector regression electric load forecasting model," *Energy Convers. Manag.*, vol. 50, no. 1, pp. 105–117, Jan. 2009.
- [51] C.-C. Hsu and C.-Y. Chen, "Regional load forecasting in Taiwan—applications of artificial neural networks," *Energy Convers. Manag.*, vol. 44, no. 12, pp. 1941–1949, Jul. 2003.
- [52] J. Wang, L. Li, D. Niu, and Z. Tan, "An annual load forecasting model based on support vector regression with differential evolution algorithm," *Appl. Energy*, vol. 94, pp. 65–70, Jun. 2012.
- [53] C.-M. Huang, "A Particle Swarm Optimization to Identifying the ARMAX Model for Short-Term Load Forecasting," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 1126–1133, May 2005.
- [54] F.-Y. Ju and W.-C. Hong, "Application of seasonal SVR with chaotic gravitational search algorithm in electricity forecasting," *Appl. Math. Model.*, Jun. 2013.
-

-
- [55] D. Niu, Y. Wang, and D. D. Wu, "Power load forecasting using support vector machine and ant colony optimization," *Expert Syst. Appl.*, vol. 37, no. 3, pp. 2531–2539, Mar. 2010.
- [56] P.-F. Pai and W.-C. Hong, "Support vector machines with simulated annealing algorithms in electricity load forecasting," *Energy Convers. Manag.*, vol. 46, no. 17, pp. 2669–2688, Oct. 2005.
- [57] P.-F. Pai and W.-C. Hong, "Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms," *Electr. Power Syst. Res.*, vol. 74, no. 3, pp. 417–425, Jun. 2005.
- [58] J. Wang, S. Zhu, W. Zhang, and H. Lu, "Combined modeling for electric load forecasting with adaptive particle swarm optimization," *Energy*, vol. 35, no. 4, pp. 1671–1678, Apr. 2010.
- [59] J. W. Taylor, L. M. de Menezes, and P. E. McSharry, "A comparison of univariate methods for forecasting electricity demand up to a day ahead," *Int. J. Forecast.*, vol. 22, no. 1, pp. 1–16, Jan. 2006.
- [60] J. W. Taylor and P. E. Mcsharry, "Short-Term Load Forecasting Methods : An Evaluation Based on European Data," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 2213–2219, 2008.
- [61] J. W. Taylor, "An evaluation of methods for very short-term load forecasting using minute-by-minute British data," *Int. J. Forecast.*, vol. 24, no. 4, pp. 645–658, Oct. 2008.
- [62] J. W. Taylor, "Triple seasonal methods for short-term electricity demand forecasting," *Eur. J. Oper. Res.*, vol. 204, no. 1, pp. 139–152, Jul. 2010.
- [63] C.-N. Ko and C.-M. Lee, "Short-term load forecasting using SVR (support vector regression)-based radial basis function neural network with dual extended Kalman filter," *Energy*, vol. 49, pp. 413–422, Jan. 2013.
- [64] H. Hahn, S. Meyer-Nieberg, and S. Pickl, "Electric load forecasting methods: Tools for decision making," *Eur. J. Oper. Res.*, vol. 199, no. 3, pp. 902–907, Dec. 2009.
- [65] W. K. Wong, M. Xia, and W. C. Chu, "Adaptive neural network model for time-series forecasting," *Eur. J. Oper. Res.*, vol. 207, no. 2, pp. 807–816, Dec. 2010.
- [66] L. J. Soares and M. C. Medeiros, "Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data," *Int. J. Forecast.*, vol. 24, no. 4, pp. 630–644, Oct. 2008.
- [67] C. Kang, X. Cheng, Q. Xia, Y. Huang, and F. Gao, "Novel approach considering load-relative factors in short-term load forecasting," *Electr. Power Syst. Res.*, vol. 70, no. 2, pp. 99–107, Jul. 2004.

-
- [68] C. Wang, G. Grozev, and S. Seo, "Decomposition and statistical analysis for regional electricity demand forecasting," *Energy*, vol. 41, no. 1, pp. 313–325, May 2012.
- [69] E. Almeshaiei and H. Soltan, "A methodology for Electric Power Load Forecasting," *Alexandria Eng. J.*, vol. 50, no. 2, pp. 137–144, Jun. 2011.
- [70] M. S. Kim, "Modeling special-day effects for forecasting intraday electricity demand," *Eur. J. Oper. Res.*, vol. 230, no. 1, pp. 170–180, Oct. 2013.
- [71] H. T. Pao, "Forecasting energy consumption in Taiwan using hybrid nonlinear models," *Energy*, vol. 34, no. 10, pp. 1438–1446, Oct. 2009.
- [72] J. R. Cancelo, A. Espasa, and R. Grafe, "Forecasting the electricity load from one day to one week ahead for the Spanish system operator," *Int. J. Forecast.*, vol. 24, no. 4, pp. 588–602, Oct. 2008.
- [73] P. Zhou, B. Ang, and K. Poh, "A trigonometric grey prediction approach to forecasting electricity demand," *Energy*, vol. 31, no. 14, pp. 2839–2847, Nov. 2006.
- [74] S. S. Pappas, L. Ekonomou, D. C. Karamousantas, G. E. Chatzarakis, S. K. Katsikas, and P. Liatsis, "Electricity demand loads modeling using AutoRegressive Moving Average (ARMA) models," *Energy*, vol. 33, no. 9, pp. 1353–1360, Sep. 2008.
- [75] C.-M. Lee and C.-N. Ko, "Short-term load forecasting using lifting scheme and ARIMA models," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5902–5911, May 2011.
- [76] G.-D. Li, C.-H. Wang, S. Masuda, and M. Nagai, "A research on short term load forecasting problem applying improved grey dynamic model," *Int. J. Electr. Power Energy Syst.*, vol. 33, no. 4, pp. 809–816, May 2011.
- [77] J. W. Taylor and R. D. Snyder, "Forecasting intraday time series with multiple seasonal cycles using parsimonious seasonal exponential smoothing," *Omega*, vol. 40, no. 6, pp. 748–757, Dec. 2012.
- [78] S. Zhu, J. Wang, W. Zhao, and J. Wang, "A seasonal hybrid procedure for electricity demand forecasting in China," *Appl. Energy*, vol. 88, no. 11, pp. 3807–3815, Nov. 2011.
- [79] M. Matijaš, J. a. K. Suykens, and S. Krajcar, "Load forecasting using a multivariate meta-learning system," *Expert Syst. Appl.*, vol. 40, no. 11, pp. 4427–4437, Sep. 2013.
- [80] J. Wu, J. Wang, H. Lu, Y. Dong, and X. Lu, "Short term load forecasting technique based on the seasonal exponential adjustment method and the regression model," *Energy Convers. Manag.*, vol. 70, pp. 1–9, Jun. 2013.
- [81] K. Song, Y. Baek, D. H. Hong, and G. Jang, "Short-Term Load Forecasting for the Holidays Using," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 96–101, 2005.
-

-
- [82] M. Amina, V. S. Kodogiannis, I. Petrounias, and D. Tomtsis, "A hybrid intelligent approach for the prediction of electricity consumption," *Int. J. Electr. Power Energy Syst.*, vol. 43, no. 1, pp. 99–108, Dec. 2012.
- [83] Tom M. Mitchell, *Machine Learning*. McGraw-Hill Science, 1997, p. 432.
- [84] T. Dietterich, "Machine Learning," *ACM Comput. Surv.*, vol. 28, no. 4es, Dec. 1996.
- [85] O. Bousquet, S. Boucheron, and G. Lugosi, "Introduction to Statistical Learning Theory," in *Advanced Lectures on Machine Learning SE - 8*, vol. 3176, O. Bousquet, U. Luxburg, and G. Rätsch, Eds. Springer Berlin Heidelberg, 2004, pp. 169–207.
- [86] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1995.
- [87] V. N. Vapnik, "An overview of statistical learning theory," *Neural Networks, IEEE Trans.*, vol. 10, no. 5, pp. 988–999, Sep. 1999.
- [88] T. Evgeniou, T. Poggio, M. Pontil, and A. Verri, "Regularization and statistical learning theory for data analysis," *Comput. Stat. Data Anal.*, vol. 38, no. 4, pp. 421–432, Feb. 2002.
- [89] M. Sewell, "Structural Risk Minimization," 2008.
- [90] G. Bontempi, "Long Term Time Series Prediction with Multi-Input Multi-Output Local Learning," in *Proceedings of the 2nd European Symposium on Time Series Prediction*, 2008, no. 1.
- [91] S. Ben Taieb, G. Bontempi, A. Sorjamaa, and A. Lendasse, "Long-term prediction of time series by combining direct and MIMO strategies," *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*. pp. 3054–3061, 2009.
- [92] A. Sorjamaa, J. Hao, N. Reyhani, Y. Ji, and A. Lendasse, "Methodology for long-term prediction of time series," *Neurocomputing*, vol. 70, no. 16–18, pp. 2861–2869, Oct. 2007.
- [93] C. Hamzaçebi, D. Akay, and F. Kutay, "Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3839–3844, Mar. 2009.
- [94] G. P. Zhang, *Neural networks in business forecasting*. IGI Global, 2004.
- [95] H. Cheng, P.-N. Tan, J. Gao, and J. Scripps, "Multistep-Ahead Time Series Prediction," in *Advances in Knowledge Discovery and Data Mining SE - 89*, vol. 3918, W.-K. Ng, M. Kitsuregawa, J. Li, and K. Chang, Eds. Springer Berlin Heidelberg, 2006, pp. 765–774.
- [96] G. C. Tiao and R. S. Tsay, "Some advances in non-linear and adaptive modelling in time-series," *J. Forecast.*, vol. 13, no. 2, pp. 109–131, 1994.
-

-
- [97] L. J. Herrera, H. Pomares, I. Rojas, a. Guillén, a. Prieto, and O. Valenzuela, “Recursive prediction for long term time series forecasting using advanced models,” *Neurocomputing*, vol. 70, no. 16–18, pp. 2870–2880, Oct. 2007.
- [98] Y. Ji, J. Hao, N. Reyhani, and A. Lendasse, “Direct and Recursive Prediction of Time Series Using Mutual Information Selection,” in *Computational Intelligence and Bioinspired Systems SE - 124*, vol. 3512, J. Cabestany, A. Prieto, and F. Sandoval, Eds. Springer Berlin Heidelberg, 2005, pp. 1010–1017.
- [99] V. T. Tran, B.-S. Yang, and A. C. C. Tan, “Multi-step ahead direct prediction for the machine condition prognosis using regression trees and neuro-fuzzy systems,” *Expert Syst. Appl.*, vol. 36, no. 5, pp. 9378–9387, Jul. 2009.
- [100] A. Sorjamaa and A. Lendasse, “Time Series Prediction using DirRec Strategy,” in *European Symposium on Artificial Neural Networks*, 2006, no. April.
- [101] Robert Goodell Brown, *Smoothing, Forecasting and Prediction of Discrete Time Series*. 2004, p. 468.
- [102] C. C. Holt, “Forecasting seasonals and trends by exponentially weighted moving averages,” *Int. J. Forecast.*, vol. 20, no. 1, pp. 5–10, Jan. 2004.
- [103] P. R. Winters, “Forecasting Sales by Exponentially Weighted Moving Averages,” *Manage. Sci.*, vol. 6, no. 3, pp. 324–342, 1960.
- [104] L. A. J. Douglas C. Montgomery, *Forecasting and time series analysis*. McGraw-Hill Science, 1976, p. 304.
- [105] E. J. Hannan, *Multiple Time Series*. John Wiley & Sons, Inc., 1970.
- [106] W. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, 1943.
- [107] Quasar Jarosz, “Neuron Hand-tuned,” *Wikipedia*, 2009. [Online]. Available: http://en.wikipedia.org/wiki/File:Neuron_Hand-tuned.svg.
- [108] Johan A K Suykens, Tony Van Gestel, Jos De Brabanter, Bart De Moor, and Joos Vandewalle, *Least Squares Support Vector Machines*. 2002, p. 308.
- [109] J. V. J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, “<http://www.esat.kuleuven.be/sista/lssvmlab>.” [Online]. Available: <http://www.esat.kuleuven.be/sista/lssvmlab>.
- [110] C. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998.
-

-
- [111] L. Hoegaerts, J. A. K. Suykens, J. Vandewalle, and B. De Moor, "Primal space sparse kernel partial least squares regression for large scale problems," in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, 2004, vol. 1, p. -563.
- [112] M. Espinoza, J. A. K. Suykens, and B. De Moor, "Least squares support vector machines and primal space estimation," in *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, 2003, vol. 4, pp. 3451–3456 vol.4.
- [113] J. R. Shewchuk, "An Introduction to the Conjugate Gradient Method Without the Agonizing Pain," Carnegie Mellon University, Pittsburgh, PA, USA, 1994.
- [114] J. A. K. Suykens and L. Lukas, "Least Squares Support Vector Machine Classifiers: a Large Scale Algorithm," in *European Conference on Circuit Theory and Design*, 1999, pp. 839–842.
- [115] X. Xia, K. Li, and M. Fei, "A Fast Training Algorithm For Least-Squares Support Vector Machines," in *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2007.
- [116] W. Chu, C. J. Ong, and S. S. Keerthi, "An improved conjugate gradient scheme to the solution of least squares SVM," *IEEE Trans. neural networks*, vol. 16, no. 2, pp. 498–501, Mar. 2005.
- [117] S. S. Keerthi and S. K. Shevade, "SMO algorithm for least-squares SVM formulations," *Neural Comput.*, vol. 15, no. 2, pp. 487–507, Mar. 2003.
- [118] M. M. Adankon and M. Cheriet, "Model selection for the LS-SVM. Application to handwriting recognition," *Pattern Recognit.*, vol. 42, no. 12, pp. 3264–3270, Dec. 2009.
- [119] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Stat. Surv.*, vol. 4, pp. 40–79, 2010.
- [120] S. Crone and S. Pietsch, "A naive support vector regression benchmark for the NN3 forecasting competition," in *Proceedings of International Joint Conference on Neural Networks*, 2007, pp. 12–17.
- [121] Kristiaan Pelckmans, Johan A.K. Suykens, and Bart De Moor, "Sparse LS-SVMs using additive regularization with a penalized validation criterion," in *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2004, pp. 435–440.
- [122] J. López, K. De Brabanter, J.R. Dorronsoro, and J.A.K. Suykens, "Sparse LS-SVMs with L0–norm minimization," in *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2011, pp. 189–194.
- [123] M. E. Tipping, S. A. Solla, T. K. Leen, and K.-R. Müller, "The Relevance Vector Machine," in *Advances in Neural Information Processing Systems 12*, 2000, pp. 652–658.
-

-
- [124] T. Cover and J. Thomas, *Elements of information theory*. John Wiley & Sons, 2006.
- [125] R. Moddemeijer, “A statistic to estimate the variance of the histogram-based mutual information estimator based on dependent pairs of observations,” *Signal Processing*, vol. 75, no. 1, pp. 51–63, Jan. 1999.
- [126] Y. Moon, B. Rajagopalan, and U. Lall, “Estimation of mutual information using kernel density estimators,” *Phys. Rev. E*, vol. 52, no. 3, pp. 2318–2321, Sep. 1995.
- [127] A. Kraskov, H. Stögbauer, and P. Grassberger, “Estimating mutual information,” *Phys. Rev. E*, vol. 69, no. 6, 2004.
- [128] H. Stögbauer, A. Kraskov, S. Astakhov, and P. Grassberger, “Least-dependent-component analysis based on mutual information,” *Phys. Rev. E*, vol. 70, no. 6, 2004.
- [129] “<http://www.klab.caltech.edu/~kraskov/MILCA/>” .
- [130] S. Dhanabal and S. Chandramathi, “A Review of various k-Nearest Neighbor Query Processing Techniques,” *Int. J. Comput. Appl.*, vol. 31, no. 7, 2011.
- [131] D. Marshall, “Nearest neighbour searching in high dimensional metric space,” 2006.
- [132] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, “Locality-sensitive hashing scheme based on p-stable distributions,” in *Proceedings of the twentieth annual symposium on Computational geometry*, 2004.
- [133] J. Kleinberg, “Two algorithms for nearest-neighbor search in high dimensions,” in *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, 1997, pp. 599–608.
- [134] S. Lahmiri, “A Comparison of PNN and SVM for Stock Market Trend Prediction using Economic and Technical Information,” in *International Journal of Computer Applications*, 2011, vol. 29, no. 3, pp. 24–30.
- [135] P. Ou and H. Wang, “Prediction of Stock Market Index Movement by Ten Data Mining Techniques,” *Mod. Appl. Sci.*, vol. 3, no. 12, pp. 28–42, Nov. 2009.
- [136] K.-S. Fu and A. Rosenfeld, “Pattern Recognition and Image Processing,” *Comput. IEEE Trans.*, vol. C-25, no. 12, pp. 1336–1346, 1976.
- [137] M. Božić, M. Stojanović, Z. Stajić, and D. Tasić, “A New Two-Stage Approach to Short Term Electrical Load Forecasting,” *Energies*, vol. 6, no. 4, pp. 2130–2148, Apr. 2013.
- [138] L. Yu and H. Liu, “Feature selection for high-dimensional data: A fast correlation-based filter solution,” in *In Proceedings of The Twentieth International Conference on Machine Learning (ICML-03)*, 2003, pp. 856–863.
-

-
- [139] R. Kohavi and H. John, “Wrappers for feature subset selection,” *Artif. Intell.*, vol. 97, no. 97, pp. 273–324, 2011.
- [140] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, vol. 818. MIT press, 2009, p. 1292.
- [141] M. Božić, M. Stojanović, Z. Stajić, and N. Floranović, “Mutual Information-Based Inputs Selection for Electric Load Time Series Forecasting,” *Entropy*, vol. 15, no. 3, pp. 926–942, Feb. 2013.
- [142] M. B. Stojanović, M. M. Božić, M. M. Stanković, and Z. P. Stajić, “A methodology for training set instance selection using mutual information in time series prediction,” *Neurocomputing*, vol. 141, pp. 236–245, Oct. 2014.
- [143] M. Stojanović, M. Božić, Z. Stajić, and M. Milošević, “LS-SVM model for electrical load prediction based on incremental training set update,” *Electr. Rev.*, vol. 89, no. 4, pp. 195–199, 2013.
- [144] Miloš Božić, Z. Radonjić, D. Tasić, I. Radović, and Z. Stajić, “Short-term load forecasting with least square support vector machines,” in *ECOS 2011*, 2011, pp. 1010–1019.
- [145] R. J. Hyndman and A. B. Koehler, “Another look at measures of forecast accuracy,” *Int. J. Forecast.*, vol. 22, no. 4, pp. 679–688, Oct. 2006.
- [146] Spyros Makridakis, “Accuracy measures: theoretical and practical concerns,” *Int. J. Forecast.*, vol. 9, no. 4, pp. 527–529, 1993.
- [147] P. Goodwin and R. Lawton, “On the asymmetry of the symmetric MAPE,” *Int. J. Forecast.*, vol. 15, no. 4, pp. 405–408, Oct. 1999.
- [148] J. S. Armstrong and F. Collopy, “Error measures for generalizing about forecasting methods : Empirical comparisons,” *Int. J. Forecast.*, vol. 08, no. 1, pp. 69–80, 1992.

BIOGRAFIJA

Miloš Božić, diplomirani inženjer elektrotehnike za industrijsku energetiku, rođen je 31.12.1982. godine u Nišu, Republika Srbija. Osnovnu školu i gimnaziju „Bora Stanković“ završio je u Nišu sa odličnim uspehom 2001. godine. Iste godine upisuje Elektronski fakultet u Nišu, na kome je diplomirao 2008. godine na smeru za Industrijsku energetiku sa prosečnom ocenom 8.72 u toku studija i ocenom 10 na diplomskom ispitu.

Neposredno posle završenih studija, postaje stipendista Ministarstva za nauku i tehnološkog razvoja Republike Srbije i upisuje doktorske studije na Elektronskom fakultetu u Nišu iz naučne oblasti Elektroenergetika. Angažovan je na fakultetu kao mladi istraživač do kraja 2010. godine kada počinje da radi na istom kao saradnik za naučno-istraživački rad sve do danas.

U svom naučno-istraživačkom radu u periodu od 2008. do danas, kao autor ili koautor, publikovao je više naučnih i stručnih radova u vrhunskim domaćim i međunarodnim časopisima iz oblasti elektroenergetike i veštačke inteligencije. Učestvovao je u realizaciji tri projekta Ministarstva nauke Republike Srbije, a trenutno učestvuje u realizaciji dva projekta Ministarstva prosvete i nauke Republike Srbije u oblasti integralnih, interdisciplinarnih istraživanja. Autor je i nekoliko tehničkih rešenja.

Od 2010. godine bavi se naučno-istraživačkim radom u oblasti prognoze potrošnje električne energije primenom metoda mašinskog učenja. Takođe radi na razvoju modela za prognozu vremenskih serija sa izborom atributa i trening skupa. Primena metoda dubinskog učenja (*eng. deep learning*) kod prognoze vremenskih serija su njegovi budući pravci istraživanja.

SPISAK RADOVA

Radovi u međunarodnim časopisima (M21)

1. M. Stojanović, **M. Božić**, M. Stanković, Z. Stajić, *A methodology for training set instance selection using mutual information in time series prediction*, Neurocomputing, vol. 141, pp. 236–245, 2014.

Radovi u međunarodnim časopisima (M22)

1. **M. Božić**, M. Stojanović, Z. Stajić, D. Tasić, *A New Two-stage Approach to Short Term Electrical Load Forecasting*, Energies, Special Issue on Hybrid Advanced Techniques for Forecasting in Energy Sector, vol. 6, No. 4, pp. 2130-2148, 2013.
2. **M. Božić**, M. Stojanović, Z. Stajić, N. Floranović, *Mutual Information-Based Inputs Selection for Electric Load Time Series Forecasting*, Entropy, vol. 15, pp. 926–942, 2013.
3. M. Stojanović, **M. Božić**, M. Stanković, Z. Stajić, *Adaptive least squares support vector machines method for short-term load forecasting based on mutual information for inputs selection*, International review of electrical engineering (IREE), vol. 7, No.1, 2012.

Radovi u međunarodnim časopisima (M23)

4. **M. Božić**, M. Stojanović, Z. Stajić, and Đ. Vukić, *Power Transformer Fault Diagnosis based on Dissolved Gas Analysis with Logistic Regression*, Przegľad Elektrotechniczny, vol. 89, No. 6, pp. 83–87, 2013.
5. M. Stojanović, **M. Božić**, Z. Stajić, and M. Milošević, *LS-SVM model for electrical load prediction based on incremental training set update*, Przegľad Elektrotechniczny, vol. 89, No. 4, pp. 195–199, 2013.

Radovi u međunarodnim časopisima verifikovanim posebnom odlukom (M24)

6. M. Stojanović, **M. Božić**, M. Stanković, *Mid-Term Load Forecasting Using Recursive Time Series Prediction Strategy With Support Vector Machines*, Facta Universitatis, Ser.: Elec. Energ., vol. 23, No. 3, pp. 287-298, 2010.

Radovi u časopisima nacionalnog značaja ili radovi u međunarodnim časopisima koji nisu na SCI listi (M52)

7. **M. Božić**, M. Stojanović, *Application of SVM Methods for Mid-Term Load Forecasting*, Serbian Journal of Electrical Engineering, Vol. 8, No. 1, 73-83, 2011.
8. **M. Božić**, M. Stojanović, Z. Stajić, *Srednjoročno predviđanje električnog opterećenja primenom metoda podržavajućih vektora*, Elektroprivreda, Broj 1, 48-55, 2011.
9. **M. Božić**, M. Stojanović, Z. Stajić, *Short-term Electric Load Forecasting Using Least Squares Support Vector Machines*, Facta Universitatis, Ser.: Automatic Control and Robotics Vol. 9, No 1, pp. 141 – 150, 2010.

Saopštenja sa međunarodnih skupova štampana u celini (M33)

10. **M. Božić**, Z. Radonjić, D. Tasić, I. Radović, Z. Stajić, *Short-term load forecasting with least square support vector machines*, ECOS 2011, Novi Sad, Serbia, 2011.
11. M. Stojanović, **M. Božić**, M. Stanković, Z. Stajić, *Least squares support vector machines for short-term electrical load prediction*, ICT Innovations 2011, Skoplje, Macedonia, 2011.
12. M. Stojanović, **M. Božić**, M. Stanković, *Application of Support Vector Machines On Medium Term Load Forecasting*, ICEST 2010, Vol. 1, pp. 293-296, Ohrid, Macedonia, 2010.

Saopštenja sa skupova nacionalnog značaja štampana u celini (M63)

13. **M. Božić**, M. Stojanović, *Primena SVM metoda za predviđanje maksimalnog dnevnog opterećenja*, ETRAN 2010, Donji Milanovac, 2010.
14. **M. Božić**, M. Stojanović, Z. Stajić, *Kratkoročno predviđanje električnog opterećenja primenom metoda podržavajućih vektora*, Infoteh 2011, Jahorina, 2011.

IZJAVE AUTORA



IZJAVA O AUTORSTVU

Izjavljujem da je doktorska disertacija, pod naslovom

Kratkoročna prognoza potrošnje električne energije zasnovana na metodama veštačke inteligencije

- rezultat sopstvenog istraživačkog rada,
- da predložena disertacija, ni u celini, ni u delovima, nije bila predložena za dobijanje bilo koje diplome, prema studijskim programima drugih visokoškolskih ustanova,
- da su rezultati korektno navedeni i
- da nisam kršio autorska prava, niti zloupotrebio intelektualnu svojinu drugih lica.

U Nišu, 14.10.2014.

Autor disertacije: Miloš Božić

Potpis doktoranda:



**IZJAVA O ISTOVETNOSTI ŠTAMPANE I ELEKTRONSKE VERZIJE
DOKTORSKE DISERTACIJE**

Ime i prezime autora:

Miloš Božić

Studijski program:

Elektroenergetika

Naslov rada:

Kratkoročna prognoza potrošnje električne energije zasnovana na metodama veštačke inteligencije

Mentor:

Prof. dr Zoran Stajić

Izjavljujem da je štampana verzija moje doktorske disertacije istovetna elektronskoj verziji, koju sam predao za unošenje u **Digitalni repozitorijum Univerziteta u Nišu**.

Dozvoljavam da se objave moji lični podaci, koji su u vezi sa dobijanjem akademskog zvanja doktora nauka, kao što su ime i prezime, godina i mesto rođenja i datum odbrane rada, i to u katalogu Biblioteke, Digitalnom repozitorijumu Univerziteta u Nišu, kao i u publikacijama Univerziteta u Nišu.

U Nišu, 14.10.2014.

Autor disertacije: Miloš Božić

Potpis doktoranda:



IZJAVA O KORIŠĆENJU

Ovlašćujem Univerzitetsku biblioteku „Nikola Tesla“ da, u Digitalni repozitorijum Univerziteta u Nišu, unese moju doktorsku disertaciju, pod naslovom:

Kratkoročna prognoza potrošnje električne energije zasnovana na metodama veštačke inteligencije

koja je moje autorsko delo.

Disertaciju sa svim priložima predao sam u elektronskom formatu, pogodnom za trajno arhiviranje.

Moju doktorsku disertaciju, unetu u Digitalni repozitorijum Univerziteta u Nišu, mogu koristiti svi koji poštuju odredbe sadržane u odabranom tipu licence Kreativne zajednice (Creative Commons), za koju sam se odlučio.

1. Autorstvo
2. Autorstvo, nekomercijalno
3. Autorstvo, nekomercijalno – bez prerade
4. Autorstvo, nekomercijalno – deliti pod istim uslovima
5. Autorstvo, bez prerade
6. Autorstvo, deliti pod istim uslovima

U Nišu, 14.10.2014.

Autor disertacije: Miloš Božić

Potpis doktoranda:
