

Izrada strategija trgovanja električnom energijom u Njemačkoj korištenjem metode potpornih vektora

Balaško, Lea

Master's thesis / Diplomski rad

2016

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:217:889471>

Rights / Prava: [In copyright](#)

Download date / Datum preuzimanja: **2022-08-26**



Repository / Repozitorij:

[Repository of Faculty of Science - University of Zagreb](#)



SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO–MATEMATIČKI FAKULTET
MATEMATIČKI ODSJEK

Lea Balaško

**IZRADA STRATEGIJA TRGOVANJA
ELEKTRIČNOM ENERGIJOM U
NJEMAČKOJ KORIŠTENJEM METODE
POTPORNIH VEKTORA**

Diplomski rad

Voditelj rada:
izv. prof. dr. sc. Saša Singer

Zagreb, veljača, 2016.

Ovaj diplomski rad obranjen je dana _____ pred ispitnim povjerenstvom u sastavu:

1. _____, predsjednik
2. _____, član
3. _____, član

Povjerenstvo je rad ocijenilo ocjenom _____.

Potpisi članova povjerenstva:

1. _____
2. _____
3. _____

*Mojoj obitelji i prijateljima bez kojih ovaj rad ne bi bio moguć.
Posebno hvala dr. sc. Marinu Matijašu, koji je ovaj rad zamislio i omogućio mi da ga
provedem do kraja.*

Sadržaj

Sadržaj	iv
Uvod	1
1 Strojno učenje i metoda potpornih vektora	2
1.1 Strojno učenje	2
1.2 Metoda potpornih vektora	6
2 Trgovanje električnom energijom	14
2.1 Općenito o trgovanju električnom energijom	14
3 Analiza podataka	19
3.1 Općenito o podacima	19
3.2 Satna cijena električne energije	20
3.3 Satna temperatura	24
3.4 Satna proizvodnja električne energije iz sunčevih i fotonaponskih elektrana	27
3.5 Satna proizvodnja električne energije iz vjetroelektrana	30
3.6 Satna proizvodnja električne energije iz konvencionalnih elektrana iznad 100 MW instalirane snage	32
3.7 Dnevna proizvodnja električne energije iz hidroelektrana	35
3.8 Satna energija padalina	37
4 Rad algoritma	40
4.1 Priprema podataka za rad	40
4.2 Regresija potpornim vektorima	42
5 Rezultati	44
6 Zaključak	52
Bibliografija	54

7 Prilozi	56
7.1 Kod za unos i obradu podataka	56
7.2 Kod za pronalazak optimalnih parametara	57
7.3 Kod za testiranje modela	58

Uvod

U modernom svijetu, u kojem računala preuzimaju sve veću ulogu u svakodnevnom životu, jača želja da računala postanu "pametna", imaju sposobnost učenja iz iskustva, na način kao što to čine ljudi. I dok je ljudima učenje urođena sposobnost, računala trebaju posebne algoritme da bi pomoću njih "učila". Grana računarstva koja se bavi ovim problemom naziva se strojno učenje. Strojno učenje nastalo je sredinom prošlog stoljeća, te se intenzivno razvija od prelaska na novi milenij. Trenutno je zvijezda strojnog učenja metoda potpornih vektora. Od devedesetih godina prošlog stoljeća, ovaj algoritam pokazuje odlične rezultate u širokom spektru primjena. Misao vodilja ovog rada je pokazati mogućnosti strojnog učenja primjenom metode potpornih vektora na problemu iz stvarnog života. Izabrani problem je trgovanje električnom energijom.

Električna energija je pokretač industrije, gospodarstva i svakodnevnih aktivnosti u kućanstvima. Proces liberalizacije tržišta zahvatio je devedesetih godina prošlog stoljeća i tržište električne energije. Od tada se električnom energijom trguje na burzi. Trgovanje na burzi otvara vrata optimizaciji poslovanja, uz pretpostavku preciznog predviđanja tržišne cijene. Postoji više načina na koje možemo predviđati cijenu – najčešće korištene su statističke metode, te algoritmi strojnog učenja.

Prvo poglavlje ovog rada posvećeno metodi potpornih vektora započinje osnovnim pregledom strojnog učenja, a nastavlja se opisom logike rada metode potpornih vektora. Poglavlje završava matematičkom formulacijom regresije potpornim vektorima, što je verzija metode potpornih vektora primjenjiva u problemu predviđanja cijene električne energije. Drugo poglavlje donosi općeniti pregled trgovanja električnom energijom. Ostala poglavlja pokazuju primjenu metode potpornih vektora na trgovanje električnom energijom. Prvi korak pri rješavanju svakog problema je obrada dostupnih podataka, što je prikazano u trećem poglavlju. Četvrto poglavlje objašnjava pripremu podataka za rad metode potpornih vektora. Algoritam je implementiran u programskom jeziku Python. Korištene funkcije i klase također će biti spomenute u četvrtom poglavlju. Rezultati algoritma dani su u petom poglavlju. Rad završava sažetkom.

Poglavlje 1

Strojno učenje i metoda potpornih vektora

1.1 Strojno učenje

Strojno učenje je grana računarstva usko povezana sa statistikom, vjerojatnošću, rudarenjem podataka, umjetnom inteligencijom, optimizacijom, numeričkom matematikom, te neurobiologijom [12]. Zbog mogućnosti primjene strojnog učenja u rješavanju mnogih problema, ovaj bi se popis mogao poprilično produžiti. Strogu definiciju strojnog učenja je teško izreći. Jedna intuitivna definicija bi bila: "Strojno učenje jest programiranje računala na način da optimiziraju neki kriterij uspješnosti temeljem podatkovnih primjera ili prethodnog iskustva" [7]. U strojnom učenju imamo podatke na temelju kojih pronalazimo algoritam (model) koji uspješno opisuje te podatke.

Primjena strojnog učenja je jako široka. Pogledajmo nekoliko primjera:

- U robotici je bitno da se sustav može prilagođavati okolini. Zato ono mora imati sposobnost "učenja" iz svojih postupaka.
- Prepoznavanje \LaTeX simbola kao kod internet stranice Detexify. Ovdje se "rukom" napiše znak, a stranica ga prepozna i ponudi \LaTeX simbol za taj znak.
- Postoje sustavi koji u pošti automatski razvrstavaju pisma prema mjestu stanovanja primatelja. Kako su mnoge adrese pisane rukom, sustav mora moći prepoznati slova različitih rukopisa. Sa svakim točno razvrstanim pismom, sustav "uči" raspoznavati slova i postaje bolji u svojem poslu.
- Aplikacije koje nude glazbu korisnicima predlažu nove pjesme temeljem korisnikove povijesti preuzimanja. Prijedlog koja bi glazba mogla zanimati korisnika sve je češće rezultat rada algoritma strojnog učenja.

- Pregledom snimaka mozga pacijenata s karcinomom i snimaka mozga zdravih osoba, sustav "uči" razliku među snimkama. Na ovaj način, sustav može brže i točnije pregledati snimke i prepoznati prve znakove bolesti.
- Autonomno vozilo "uči" voziti kroz primjenu kao što to rade i vozači u autoškoli.

Kao što se može primijetiti iz gore opisanih primjena, problem strojnog učenja sastoji se od sljedećih dijelova: dostupni primjeri, ciljna vrijednost, model i novi, zasad nepoznati primjeri. Na temelju dostupnih primjera pokušava se pronaći model koji najbolje određuje ciljne vrijednosti novih, nepoznatih primjera, ili pronalazi nove informacije o dostupnim primjerima.

Podjela strojnog učenja obavlja se prema načinu na koji se doznaje ciljna vrijednost:

- nadzirano učenje
- nenadzirano učenje
- učenje s podrškom.

U nadziranom učenju sustavu se daju primjeri s točno određenim ciljnim vrijednostima temeljem kojih sustav uči. Primjer su:

- klasifikacija – ciljne vrijednosti su kategorije
- regresija – ciljne vrijednosti su vrijednosti funkcije
- predikcija – ciljne vrijednosti su kategorije, ali ne ovise samo o primjeru, već i o kategorijama prethodnih primjera [12].

Kod nenadziranog učenja postoje samo primjeri, bez njihove ciljne vrijednosti. Zadatak algoritma je pronaći neke pravilnosti među primjerima i na temelju tih pravilnosti odrediti ciljne vrijednosti novih primjera. Primjer su:

- grupiranje – ciljne vrijednosti su klasteri, tj. grupe
- otkrivanje iznimki – postoje samo dvije ciljne vrijednosti: iznimka i ne iznimka
- kompresija podataka – cilj je originalne podatke prikazati u nižoj dimenziji uz mogućnost jednoznačnog vraćanja izvornih podataka [12].

Kod učenja s podrškom, sustav za svaki primjer odredi ciljnu vrijednost, te temeljem te vrijednosti prima nagradu ili kaznu. Nakon svakog primjera sustav uči. Primjeri su: igre kao što je šah ili kružić-križić, "učenje" robota [12].

Neki od najpoznatijih algoritama za strojno učenje su: naivni Bayesov klasifikator, neuronske mreže, metoda potpornih vektora, te k -srednjih vrijednosti [12]. Naivni Bayesov klasifikator temelji se na Bayesovom pravilu:

$$p(y|x) = \frac{P(x|y)P(y)}{P(x)},$$

gdje s x označavamo podatke, a s y hipotezu, odnosno model [12]. Algoritam neuronskih mreža inspiriran je radom neurona u ljudskom mozgu. Neuronu u mozgu sami po sebi ne mogu riješiti teške probleme, već samo jednostavne, ali zajedno su snaga koja pokreće čovječanstvo. Na isti način rade i neuronske mreže. Neuronska mreža sastoji se od više umjetnih neurona od kojih jedan sam ne može puno, ali zajedno rješavaju vrlo složene probleme [12]. Algoritam k -srednjih vrijednosti grupira podatke u k grupa. Prvo odabranom metodom bira k primjera zvanih centroidi, oko kojih grupira ostale primjere. Primjer grupira tako da računa udaljenost svakog primjera od svih centroida. Primjer se pridodaje onom centroidu do kojeg je udaljenost najmanja, čime se stvaraju grupe. Nakon toga se za svaku grupu računa novi centroid, tako da se izračuna srednja vrijednost svih primjera u grupi. Postupak grupacije ponavlja se, ili unaprijed određen broj puta, ili dok broj primjera koji mijenjaju grupe ne padne ispod željenog broja. Funkcija kojom se računa udaljenost ovisi o karakteru primjera.

Jedan od najbitnijih teorema u strojnom učenju na kojem se cijela teorija i gradi je PAC (Probably Approximately Correct), u slobodnom prijevodu teorem o vjerojatno aproksimativno točnom.

Teorem 1.1.1 (PAC). *Ukoliko je neka hipoteza izrazito pogrešna – tada će to biti vidljivo s velikom vjerojatnošću već na malom podskupu primjera (preko velike pogreške). Vrijedi i obratno – za bilo koju konzistentnu hipotezu na dovoljno velikom broju primjera malo je vjerojatno da je izrazito pogrešna – tj. vjerojatno je približno točna [12].*

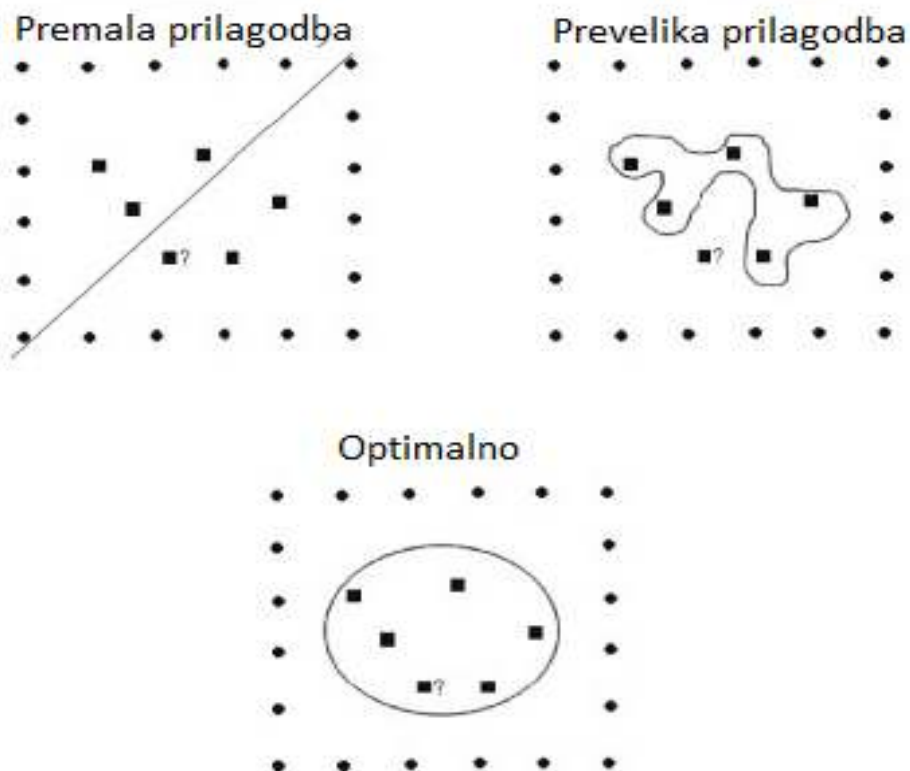
Upravo je ovo logika u pozadini strojnog učenja. Ako model radi dovoljno dobro na većini primjera, onda je vjerojatno dovoljno dobar. Temeljem toga zaključka opravdava se korištenje dobivenog modela.

Najveći izazov u strojnom učenju je odnos između točnosti i generalizacije. S obzirom na to da se model razvija na dostupnim primjerima, tražeći što veću točnost, može se dogoditi da model nauči savršeno na primjerima za učenje, ali radi greške na novim primjerima. S druge strane, ako model ne nauči dovoljno na primjerima za učenje, neće dobro raditi ni na novim primjerima. Prva situacija se naziva prejaka prilagodba (eng. overfitting), a druga premala prilagodba (eng. underfitting). Da bi se ovaj problem dočarao, često se koristi sljedeća priča:

Primjer 1.1.2. *Ivan i Marko su treneri pasa. Ivan godinama trenira samo njemačke ovčare, dok Marko trenira različite vrste pasa i to samo nekoliko dana. Ivan i Marko*

su zamoljeni da svojim znanjem odrede jesu li dvije odabrane životinje psi. Ivanu je pokazana doga, koju on ne smatra psom, jer osobinama ne odgovara onom što on očekuje od psa, to jest što je svojim radom s njemačkim ovčarima naučio da određuje psa. Marku je pokazana mačka za koju on zaključuje da jest pas jer ima četiri noge. Očito, Marko još nije naučio dovoljno osobina koje razlikuju psa od ostalih životinja. Ivan je primjer prejake, a Marko premale prilagodbe.

Ovaj i sljedeći primjer su preuzeti iz [5].



Slika 1.1: Primjer premale, prejake i optimalne prilagodbe

Primjer 1.1.3. Model uči odvojiti točkice i kvadratiće na svim primjerima na slici 1.1 osim primjera označenih upitnikom. Zadatak modela je odrediti grupu za primjer s upitnikom. U prvom slučaju, kada postoji premala prilagodba, model ne nauči točno odvajati primjere. Samim time ne određuje točno ni primjer s upitnikom. U drugom slučaju, model odlično nauči razdvajati primjere za učenje, ali netočno određuje grupu za primjer s upitnikom. Treći slučaj prikazuje optimalan rad modela kada ispravno grupira primjere za učenje, te primjer s upitnikom.

Premale i prevelike prilagodbe nisu poželjne u strojnom učenju. U strojnom učenju želi se postići treći slučaj, optimalna prilagodba, u kojem model na primjerima za učenje nauči dovoljno za razlikovanje grupa, ali ne previše, da ne bi krivo procijenio nove primjere [5].

1.2 Metoda potpornih vektora

Povijesni pregled i primjena

Mnogi algoritmi za strojno učenje pretpostavljaju da je poznata razdioba vjerojatnosti, što u praksi često nije slučaj. Metoda potpornih vektora (eng. Support vector machine, SVM) nema tu pretpostavku pa je primjenjiva u velikom broju situacija iz stvarnog života. Algoritam svoje korijene vuče iz šezdesetih godina prošlog stoljeća, kada je Vladimir Vapnik uveo pojam linearnog klasifikatora i optimizirajućih razdvajajućih ravnina. Taj algoritam se nije previše koristio sve do 1992. godine, kada su Isabelle Guyon, Bernhard Boser i Vladimir Vapnik uveli nelinearni klasifikator koristeći tzv. *kernel* trik [5]. Metoda potpornih vektora je algoritam nadziranog učenja, što znači da postoje primjeri s poznatim ciljnim vrijednostima. Prvotno je algoritam napravljen za kategorizaciju, ali je kasnije razvijena i varijanta za regresiju potpornim vektorima (eng. Support vector regression, SVR). Regresiju potpornim vektorima predložili su Vladimir Vapnik, Harris Drucker, Chris Burges, Linda Kaufman i Alex Smola 1996. godine [5]. Daljnjim proučavanjem pokazalo se da je metoda potpornih vektora zapravo samo jedan slučaj regresije potpornim vektorima. Metoda potpornih vektora je, zbog svojih jedinstvenih svojstava, pronašla široku primjenu. Koristi se u procesiranju slika, prepoznavanju govora, predikciji vremenskih serija, analizi podataka, energetici, sigurnosti, bioinformatici, teoriji kaosa i drugim područjima [5].

Teorija statističkog učenja

Temelj metode potpornih vektora je teorija statističkog učenja koja se primarno bavi vezom između kapaciteta modela za učenje i njegovih performansi [5]. Gradnja teorije statističkog učenja polazi od pretpostavke da postoji n primjera koji se sastoje od parova ulaznih i izlaznih podataka $x_i, y_i, i = 1, 2, \dots, n$, gdje su x_i ulazni podaci, a y_i izlazni, odnosno ciljni podaci za koje se iz pouzdanog izvora zna da su točni. Nastoji se pronaći model koji bi koristio funkcijsko mapiranje $x_i \rightarrow f(x_i, \alpha)$, gdje je α podesivi parametar učenja. Druga pretpostavka je da postoji nepoznata razdioba vjerojatnosti $P(x, y)$, s funkcijom gustoće vjerojatnosti $p(x, y)$, iz koje se ovi podaci uzimaju nezavisno i identički su distribuirani [5]. Tada se očekivana pogreška testiranja ili rizik (R) može izraziti preko parametra α :

$$R(\alpha) = \int_{x,y} \frac{1}{2} |y_i - f(x_i, \alpha)| p(x, y) dx dy. \quad (1.1)$$

U praksi je česta situacija da je funkcija gustoće vjerojatnosti nepoznata, pa je nemoguće izračunati očekivani rizik. Umjesto toga, računa se empirijski rizik (R_{emp}), što je zapravo prosječna pogreška na parovima za učenje [5]

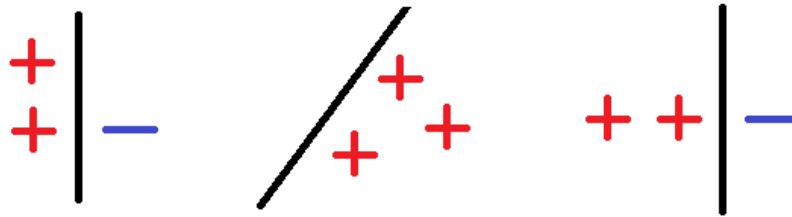
$$R_{emp}(\alpha) = \frac{1}{2n} \sum_{i=1}^n |y_i - f(x_i, \alpha)|. \quad (1.2)$$

R_{emp} je neovisan o funkciji gustoće, te je fiksiran za određeni α i skup primjera za učenje. Izraz na desnoj strani jednadžbe (1.2) naziva se gubitak. Nadalje, uvodi se parametar η , $0 < \eta < 1$, koji će uz odgovarajuće skaliranje, odgovarati gubitku. Kao što je navedeno u [5], Vladimir Vapnik je 1955. godine pokazao da uz vjerojatnost $1 - \eta$ vrijedi sljedeća ograda

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\left(\frac{h(\log(\frac{2n}{h}) + 1) - \log(\frac{\eta}{4})}{n} \right)}, \quad (1.3)$$

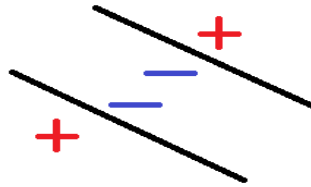
gdje je $h \in \mathbb{N}$ kvantitativna mjera kapaciteta modela za učenje i naziva se Vapnik Chervonenskisova dimenzija, ili kraće, VC dimenzija. Poznavanje mjere h omogućava izračunavanje izraza na desnoj strani jednadžbe (1.3). Naravno, cilj je pronaći model za učenje s minimalnim rizikom. Iz jednadžbe (1.3) vidljivo je da se to može postići odabirom parametra α , koji će rezultirati minimalnom gornjom granicom za rizik [5].

VC dimenzija definira se preko principa razdvajanja točaka. Poznato je da se n točaka može podijeliti u dvije grupe na 2^n načina. Skup funkcija $f(x)$ razdvaja n točaka, ako za svaki od 2^n načina grupiranja točaka, postoji barem jedna funkcija iz skupa koja savršeno razdvaja grupe. Slika 1.2 pokazuje situaciju u dvodimenzionalnom prostoru s tri točke. U



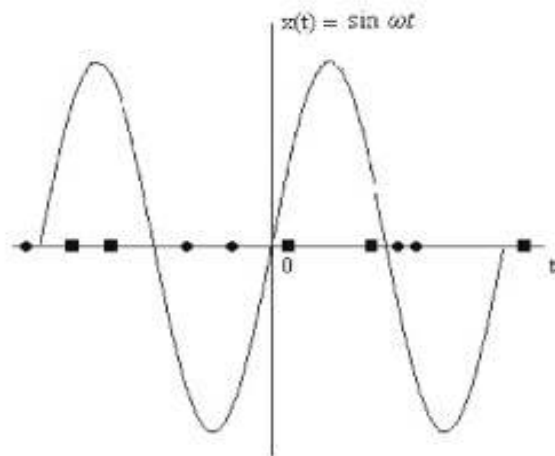
Slika 1.2: Razdvajanje tri točke u dvodimenzionalnom prostoru

ovom slučaju funkcija razdvajanja točaka je usmjereni pravac. Prikazane su tri situacije, ali uz malo razmišljanja vidljivo je da za svaku moguću kombinaciju postoji pravac koji savršeno razdvaja točke. U primjeru s četiri točke u dvodimenzionalnom prostoru, postoji slučaj kada nije moguće razdvojiti dvije grupe pravcem [12] (vidjeti sliku 1.3). VC dimenzija h skupa funkcija $f(x, \alpha)$ je najveći broj primjera za učenje koji mogu biti razdvojeni



Slika 1.3: Primjer kada se grupe ne mogu razdvojiti

funkcijama iz skupa [5]. Za navedeni primjer $f(x, \alpha)$ je set usmjerenih pravaca u \mathbb{R}^2 , te je h jednako $(2 + 1) = 3$. VC dimenzija skupa usmjerenih hiperravnina, što su razdvajajuće funkcije u n -dimenzionalnom prostoru, je $n + 1$. Postoje skupine funkcija za koje je VC dimenzija beskonačna kao što je sinusoida $f(x, \omega) = l(\sin \omega x > 0)$ [5]. Odabirom odgovarajuće vrijednosti parametra ω , sinusoida može razdvojiti proizvoljno mnogo točaka (vidjeti sliku 1.4).

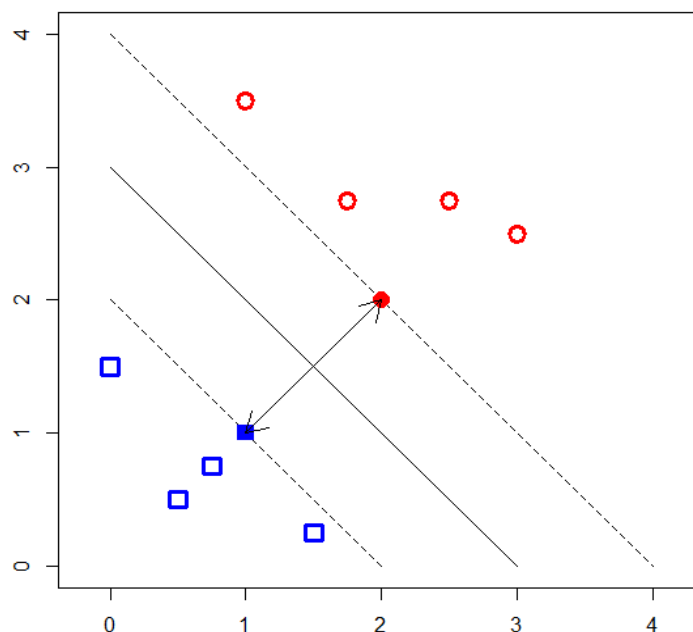


Slika 1.4: Razdvajanje točaka sinusoidom

Cilj je odabrati model optimalne složenosti za zadani skup primjera za učenje. Iz prethodnog izlaganja i jednadžbe (1.3) zaključujemo da ćemo to postići biranjem funkcije s minimalnim $R_{emp}(\alpha)$ iz skupa funkcija s minimalnom VC dimenzijom [5].

Optimizirajuća razdvajajuća ravnina i potporni vektori

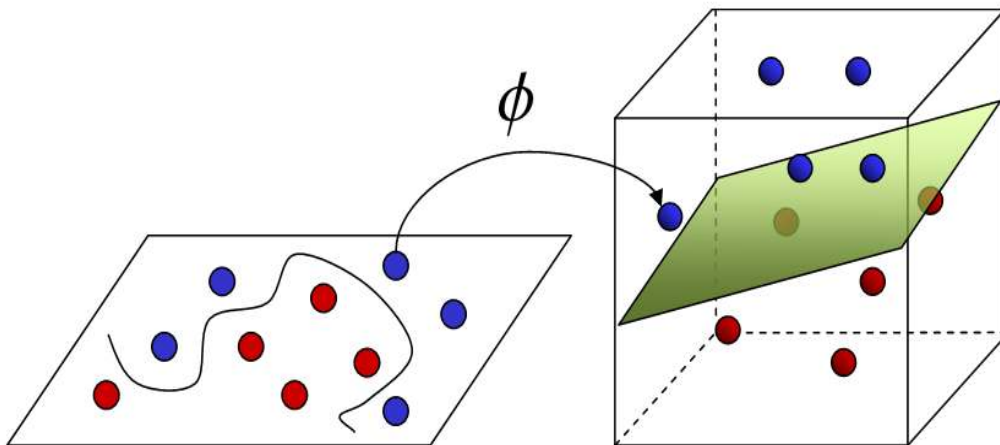
Algoritam nastoji pronaći hiperravninu koja najbolje razdvaja primjere iz različitih kategorija. Najbolje razdvajajuća ravnina je ona koja ima najveću marginu, tj. udaljenost između hiperravnine i primjera koji su joj najbliži, potpornih vektora. Sljedeća slika prikazuje primjere dviju klasa u dvodimenzionalnom prostoru i elemente metode potpornih vektora. Dvije kategorije prikazane su različitim oblicima i bojama, kako bi se što bolje razlikovale.



Slika 1.5: Popunjeni oblici su potporni vektori u primjeru SVM algoritma u dvije dimenzije

Kao što je već rečeno, u dvije dimenzije hiperravnina razdvajanja je zapravo usmjereni pravac. Na slici 1.5 vidi se pravac, kao i margine u obliku strelica, dok su potporni vektori popunjeni oblici. Na ovom primjeru možemo primijetiti da algoritam radi dobro sve dok su kategorije razdvojive. No što kada nisu? To je prvi problem s kojim se susreće metoda potpornih vektora. Drugi problem koji se pojavljuje kada su kategorije razdvojive, je vrijeme potrebno za pronalazak optimalne hiperravnine. Naime, što je razlika među kategorijama manja, povećava se vrijeme potrebno za pronalazak optimalne hiperravnine. Oba problema rješavaju se sljedećim trikom: točke se mapiraju u prostor dimenzije više

od originalne. Slika 1.6, preuzeta iz [8], prikazuje kako problem koji nije razdvojiv u dvije dimenzije, postaje razdvojiv u tri dimenzije. Ovo preslikavanje naziva se *kernel*, a trik preslikavanja u višu dimenziju *kernel* trik. Odabir pravog *kernel*-a je jako bitan. Većinom se koriste linearni, polinomni, radijalni (RBF), Markovljevi, Fourierovi i dvoslojni neuronski *kernel* [11]. Odabir najboljeg *kernel*-a ovisi o problemu, no RBF se eksperimentalno pokazao najboljim za široku paletu problema [5].



Slika 1.6: Primjer preslikavanja iz dvije u tri dimenzije

Regresija potpornim vektorima

Za regresiju potpornim vektorima, umjesto margine između optimalne razdvajajuće hiper-ravnine i potpornih vektora, potrebno je koristiti neku mjeru pogreške aproksimacije. Neka je y predviđena, a f izmjerena vrijednost. Najčešće korištene mjere pogreške su:

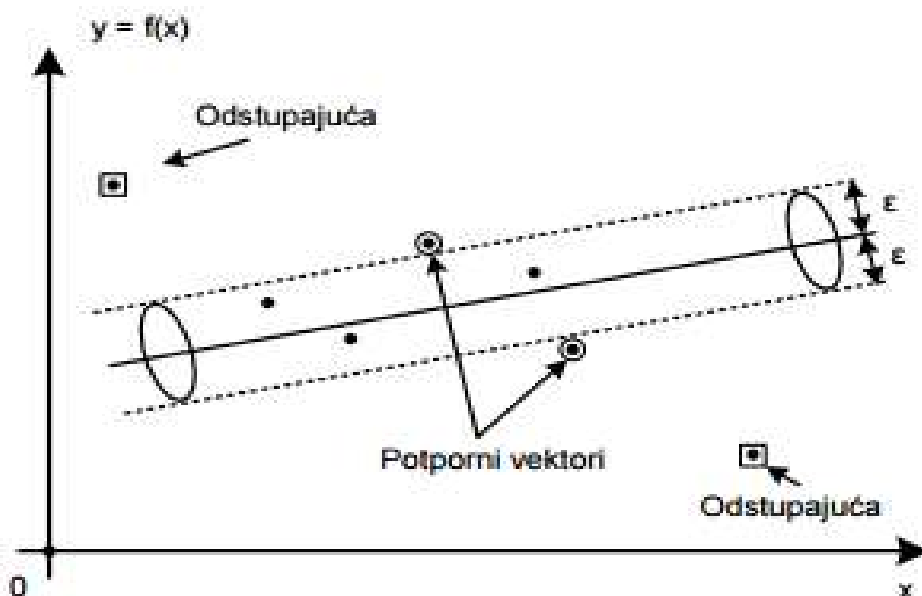
- L_1 norma: $|y - f|$
- L_2 norma: $(y - f)^2$
- Hubertova funkcija gubitka:
$$\begin{cases} \frac{1}{2}(y - f)^2 & \text{za } |y - f| < \epsilon \\ \epsilon|y - f| - \frac{\epsilon^2}{2} & \text{inače.} \end{cases}$$

Definicije mjera preuzete su iz [5].

Izvod regresije potpornim vektorima provodi se korištenjem linearnog modela,

$$\vec{y} = f(\vec{x}) = \vec{w}^T \vec{x} + b. \quad (1.4)$$

Kernel trik omogućava da se dobivena teorija proširi na nelinearni model. Slika 1.7, preuzeta iz [11], prikazuje regresiju potpornim vektorima. Točke koje odstupaju označene su kvadratima, a potporni vektori krugovima. Mjera neosjetljivosti na greške ϵ definira "ε-



Slika 1.7: "ε-cijev" i potporni vektori u regresiji potpornim vektorima

cijev" za koju vrijedi sljedeće:

- ukoliko je predviđena vrijednost unutar "ε-cijevi", greška je 0,
- s druge strane, za predviđene vrijednosti izvan "ε- cijevi", greška je razlika između predviđene vrijednosti i radijusa "ε-cijevi" [5].

Definira se

$$V(u) = \begin{cases} 0, & \text{za } |u| < \epsilon \\ |u| - \epsilon, & \text{inače.} \end{cases} \quad (1.5)$$

Procjena \vec{w} u linearnoj regresiji provodi se minimizacijom funkcije:

$$\min_{\vec{w}} \sum_{i=1}^n V(y_i - f(x_i)) + \frac{\lambda}{2} \|\vec{w}\|^2. \quad (1.6)$$

Regulacijski parametar λ procjenjuje se unakrsnom validacijom. U regresiji potpornim vektorima pokušava se istovremeno minimizirati empirijski rizik i $\|\vec{w}\|^2$ [5].

Sljedeći korak je uvođenje Langrangeovih multiplikatora (α_i, α_i^*) tako da α_i i α_i^* odgovaraju točkama (u parovima) iznad i ispod "ε-cijevi". Iz jednadžba (1.5) i (1.6) optimizacijski problem za regresiju potpornim vektorima može biti zapisan sljedećim formulama:

$$\min_{(\alpha_i, \alpha_i^*)} \epsilon \sum_{i=1}^n (\alpha_i^* + \alpha_i) - \sum_{i=1}^n y_i (\alpha_i^* - \alpha_i) + \frac{1}{2} \sum_{i,j=1}^n (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) (x_i^\top x_j), \quad (1.7)$$

$$\begin{aligned} \sum_{i=1}^n \alpha_i^* &= \sum_{i=1}^n \alpha_i, \\ \alpha_i^* &\leq \frac{1}{\lambda}, \\ 0 &\leq \alpha_i. \end{aligned} \quad (1.8)$$

Rješenjem jednadžbe (1.7), uz uvjete dane u (1.8), određuju se Langrangeovi multiplikatori [5]. Izrazi za varijable dane u formuliranju linearnog modela za regresiju (1.4) izraženi su preko Langrangeovih multiplikatora α_i i α_i^* :

$$\begin{aligned} \vec{w} &= \sum_{i=1}^n (\alpha_i^* - \alpha_i) \vec{x}_i, \\ b &= \frac{1}{n} \sum_{i=1}^n (y_i - \vec{x}_i^\top \vec{w}). \end{aligned} \quad (1.9)$$

Iz (1.9) slijedi da pronalazak Langrangeovih multiplikatora dovodi do optimalnih rješenja za \vec{w} i b . Potporni vektori su rješenja koja zadovoljavaju Karush-Kuhn-Tuckerove uvjete [5].

$$\alpha_i \alpha_i^* = 0, \quad i = 1, 2, \dots, n. \quad (1.10)$$

Jednadžba (1.10) govori da su potporni vektori, zapravo, oni podaci u kojima je točno jedan Langrangeov multiplikator nula.

U praksi se puno češće koristi nelinearni model zbog boljih rezultata. *Kernel* trik, koji je spomenut ranije, upotrebljava se i u ovom slučaju, pa se nelinearni model preslikava u

višu dimenziju, gdje postaje linearni i time se omogućava korištenje formula za linearni model. Optimizacijski problem iz (1.7) sada postaje

$$\min_{(\alpha_i, \alpha_i^*)} \epsilon \sum_{i=1}^n (\alpha_i^* + \alpha_i) - \sum_{i=1}^n y_i (\alpha_i^* - \alpha_i) + \frac{1}{2} \sum_{i,j=1}^n (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) K(x_i, x_j). \quad (1.11)$$

S $K(x_i, x_j)$ označena je *kernel* funkcija [5]. Ponavljajući isti postupak kao za linearni model, to jest pronalaženjem rješenja problema (1.11), uz uvjete iz (1.8), dobivaju se Lagrangeovi multiplikatori. Sada je regresijska funkcija:

$$\vec{y} = f(\vec{x}) = \vec{w}^\top \vec{x} + b = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\vec{x}_i, \vec{x}) + b, \quad (1.12)$$

uz optimalni b :

$$b = \frac{1}{2} \sum_{i=1}^n (\alpha_i - \alpha_i^*) [K(\vec{x}_i, \vec{x}_r) + K(\vec{x}_i, \vec{x}_s)]. \quad (1.13)$$

Jednadžbe (1.12) i (1.13) preuzete su iz [5].

Koraci u radu s regresijom potpornim vektorima

Sada kada smo pokazali teoriju regresije potpornim vektorima, rezimirajmo znanje i prikazimo korake koje trebamo poduzeti prilikom korištenja regresije potpornim vektorima.

1. Skaliranje – u jednadžbama se koristili unutarnji produkt vektora x_i što ukazuje na važnost da skaliramo sve podatke. U praksi su se dovoljnima pokazala linearna skaliranja na intervale $[-1, 1]$ ili $[0, 1]$.
2. Izbor *kernel* funkcije – vidjeli smo u teorijskom prikazu da je *kernel* trik neprocjenjiv za metodu potpornih vektora. Zbog toga je izbor *kernel* funkcije jako bitan. Kao što je već rečeno, RBF *kernel* se pokazao najboljim u praksi, pa se on najčešće preporučuje.
3. Odrediti parametre *kernel* funkcije – naravno, nakon odabira *kernel* funkcije potrebno je pronaći optimalne parametre funkcije. U praksi se to često određuje korištenjem specijaliziranih klasa programskih biblioteka ili ručnim namještanjem.
4. Treniranje – sljedeći korak "učenje" optimalnog modela na podacima za trening.
5. Testiranje – zadnji korak je, naravno, testiranje modela na podacima za testiranje [5].

Poglavlje 2

Trgovanje električnom energijom

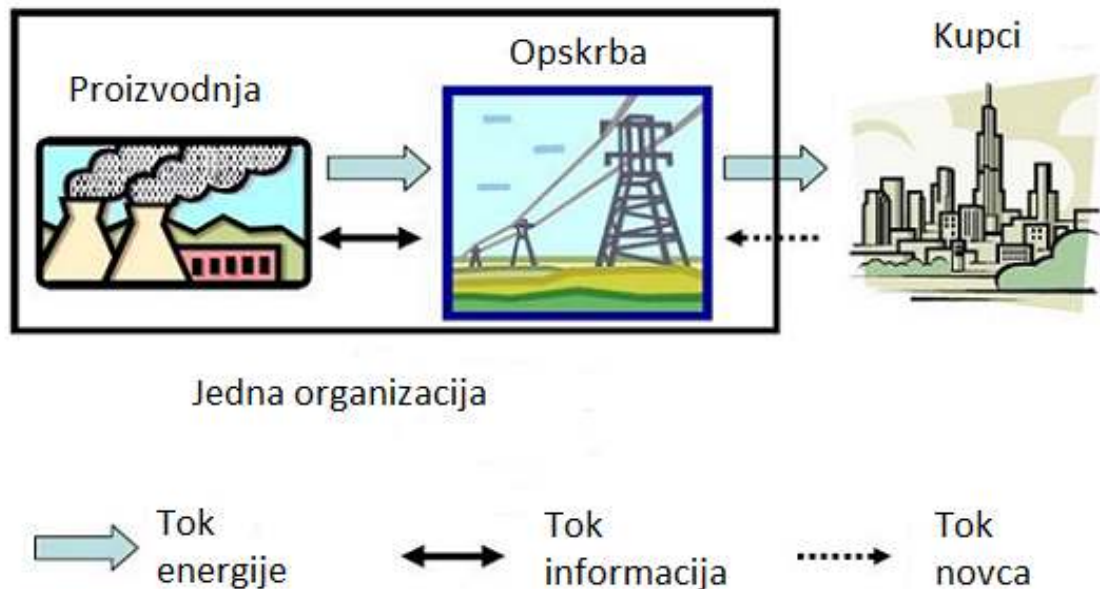
2.1 Općenito o trgovanju električnom energijom

Električnom energijom se trguje kao i svakom drugom robom, uz nekoliko specifičnosti. Do devedesetih godina prošlog stoljeća, prevladavajući model organizacije elektro–energetike bio je vertikalno integrirani sustav. U spomenutom sustavu, proizvodnju i opskrbu provodi jedna tvrtka koja isporučuje proizvedenu energiju krajnjem korisniku i od njega naplaćuje uslugu. Slika 2.1 preuzeta iz [6] pokazuje vertikalno integrirani sustav.

Devedesetih godina prošlog stoljeća krenulo se u liberalizaciju tržišta električnom energijom. Osnovna ideja liberalizacije je odvojiti proizvodnju od distribucije, barem na papiru. Tada jedna tvrtka ne bi mogla proizvoditi i distribuirati električnu energiju kao u vertikalno integriranom sustavu, već bi te dvije djelatnosti morale biti razdvojene. Dozvoljava se da se djelatnosti zadrže u jednoj grupaciji pomoću sistema tvrtki kćeri i holdinga. Liberalizacijom tržišta razvila su se dva prevladavajuća načina trgovanja: bilateralnim ugovorima i trgovanjem na organiziranom tržištu električne energije. Na organiziranom tržištu električne energije postoje dva načina trgovanja: preko brokera i na burzi [6].

Bilateralni ugovori su ugovori između dvije strane, one koja prodaje električnu energiju i one koja je kupuje. Kod bilateralnih ugovora, cijena se određuje pregovaranjem i ne postoji srednja vrijednost koja bi se, u određenom vremenskom terminu, koristila u svim ugovorima. Ovakvi ugovori se, uglavnom, ugovaraju na dulji vremenski period. Prednost je fiksna cijena tijekom duljeg vremenskog perioda, a rizik je nepredvidljivost proizvodnje i potražnje električne energije [6]. Pogledajmo sljedeći primjer:

Primjer 2.1.1. *Tvrtka koja upravlja nuklearnom elektranom sklopila je višegodišnji ugovor s distribucijskom tvrtkom. Prednosti za obje tvrtke su ugovorene količine energije po fiksnoj cijeni. No, ukoliko se dogodi neočekivan kvar na elektrani, čime se smanjuju proizvodni kapaciteti, prva tvrtka mora kupovati električnu energiju da bi ispoštovala ugovor. U suprotnom, mora platiti kaznu definiranu ugovorom. Cijena po kojoj kupuje električnu*



Slika 2.1: Prikaz vertikalno integriranog sustava u kojem jedna tvrtka provodi proizvodnju i opskrbu

energiju može biti viša od one iz ugovora, što čini probleme prvoj tvrtki. Slična situacija može zadesiti drugu tvrtku ako ostane bez velikog broja potrošača, što rezultira viškom električne energije. Pošto ne može skladištiti električnu energiju, tvrtka je mora prodati pa čak i po cijeni koja je manja od ugovorene cijene.

U ovom primjeru pokazali smo dvije situacije: kada prodavatelj ima manjak, a proizvođač višak električne energije i kada prodavatelj ima višak, a proizvođač manjak električne energije.

Burza je organizirano i centralizirano mjesto trgovanja. Kao i na svakoj drugoj burzi, ovdje prodavatelji i kupci ne komuniciraju međusobno, već daju svoje ponude anonimno, na centralno mjesto. Zbog posebnosti električne energije kao robe (ne može se skladištiti, u svakom trenutku količina proizvedene i potrošene energije mora biti podjednaka), većinom se trguje dan unaprijed na razini svakog sata [6]. Burza prima ponude dan unaprijed za određeni sat i na temelju tih ponuda određuje cijenu i količinu električne energije kojom će se trgovati. Prodavatelji električne energije nude količinu energije i cijenu po kojoj su spremni prodati navedenu količinu. Kupci iznose količine energije i cijenu po kojoj su spremni kupiti. Od svih primljenih ponuda za prodaju, formira se krivulja prodaje, a od svih primljenih ponuda za kupnju, krivulja kupnje. Cijena i količina kojom će se trgovati

određuje se presjekom tih dviju krivulja [9]. Pogledajmo sljedeći primjer koji je izrađen temeljem primjera iz [9]:

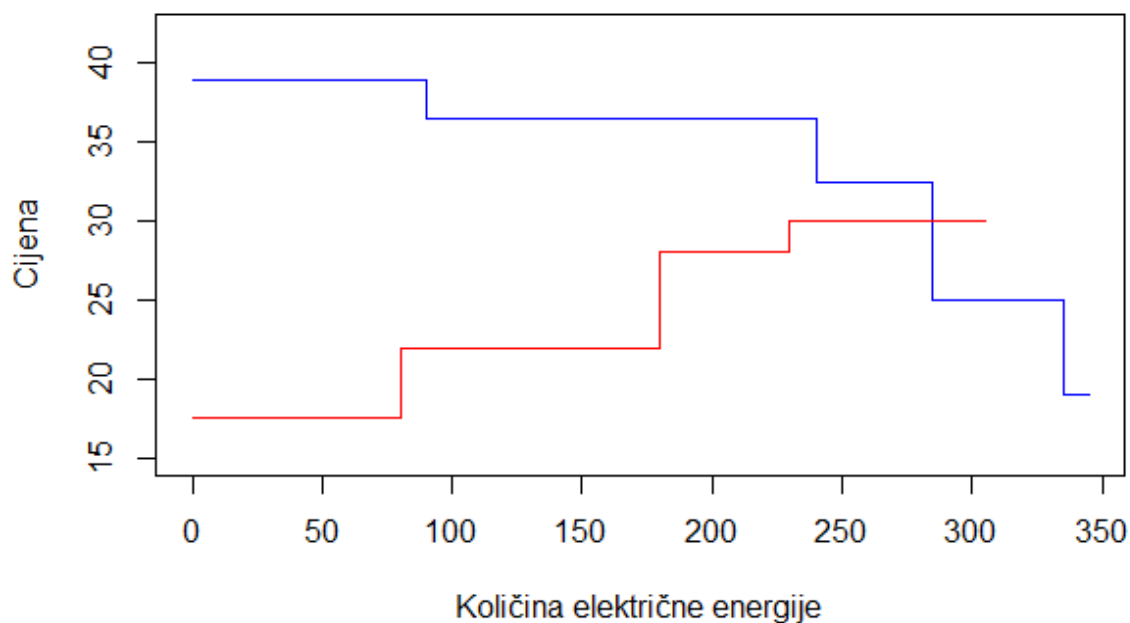
Primjer 2.1.2. *Sljedeća tablica pokazuje ponude za prodaju i kupnju dan unaprijed za dan kasnije, 15:00 sati.*

Ime tvrtke	Količina (MWh)	Cijena (€/MWh)
Prodavatelj 1	100	22
Prodavatelj 2	50	28
Prodavatelj 3	75	30
Prodavatelj 4	80	17.5
Kupac 1	45	32.5
Kupac 2	150	36.5
Kupac 3	90	39
Kupac 4	10	19
Kupac 5	50	25

Iz svih ponuda za prodaju formira se linija ponude (označena crvenom linijom na slici 2.2). Plavom linijom je označena linija potražnje, formirana iz svih ponuda za kupnju. Tržišna cijena nalazi se u sjecištu tih dviju linija, u ovom slučaju to je 30 €/MWh. Ukupno se trguje s 285 MWh električne energije. Prihvaćaju se sve ponude za prodaju s cijenom ispod tržišne. Cijena trgovanja bit će upravo tržišna cijena. Sukladno tome, prihvaćaju se i sve ponude za kupnju iznad tržišne cijene.

Često se bilateralno trgovanje i trgovanje dan unaprijed koriste zajedno, tako da se nadopunjavaju. Bilateralnim ugovorima ugovara se ispostava električne energije u zadanoj količini. U stvarnoj situaciji, rijetko se događa da stvarna potreba za električnom energijom savršeno odgovara ugovorenoj. Mnogo je češće stvarna potražnja ili veća ili manja od ugovorene, ali se ugovorena količina ne može mijenjati i mora se ispoštovati. S obzirom na to da se električna energija ne može skladištiti, postoji potreba za izjednačavanjem. U ovakvoj situaciji prodavatelj i kupac električne energije okreću se trgovanju dan unaprijed [10]. Postoje četiri moguća scenarija:

- Prodavatelj nije u mogućnosti isporučiti ugovorenu količinu, te mu je kupnja energije na burzi jedini izbor.
- Prodavatelj ima više energije od ugovorene količine koju prodaje na burzi.
- Kupac ima veće potrebe za energijom od ugovorene, koju onda zadovoljava kupnjom na burzi.
- Kupac ima manje potrebe od isporučene energije, pa višak prodaje na burzi.



Slika 2.2: Primjer formiranja cijene na sjecištu linije ponude i linije potražnje

Trgovanje dan unaprijed omogućava održanje ugovorenih količina. Osim u suradnji s bilateralnim ugovorima, tržište dan unaprijed je važno samo po sebi. Velika količina energije i novca svaki se dan razmjeni preko burze. Najvažnija europska tržišta na kojima se trguje električnom energijom su [6]:

- EPEX Spot – burza za trgovanje na njemačkom, austrijskom, švicarskom, francuskom, britanskom, danskom i nizozemskom tržištu; centralna europska burza za trgovanje električnom energijom
- Nord Pool – burza za trgovanje na tržištu nordijskih i baltičkih zemalja
- APX – burza trgovanja na britanskom, nizozemskom i belgijskom tržištu
- OMEL – trgovanje na španjolskom i portugalskom tržištu
- GME (Gestore Mercati Energetici) – burza na kojoj se trguje na talijanskom tržištu

- HUPX – burza trgovanja na mađarskom tržištu
- BSP - Southpool – burza trgovanja na slovenskom tržištu
- CROPEX – burza trgovanja na hrvatskom tržištu; trenutno u fazi potpisivanja ugovora s tržišnim sudionicima za trgovanje dan unaprijed.

Izrada strategije trgovanja ovisi o ulozi na burzi – jeste li prodavatelj, kupac ili broker. Također, veliku razliku rade mogućnosti i ograničenja koja imate. Na primjer, manager iz solarne elektrane ne može ponuditi električnu energiju noću, jer se ona tada ne proizvodi, što je ograničenje njegovih mogućnosti. S druge strane, manager iz termoelektrane ima više mogućnosti regulacije ponude, pošto termoelektrane mogu regulirati proizvodnju. Slična ograničenja i mogućnosti imaju i kupci, ovisno o razlozima potreba za električnom energijom. Zanimljivi primjer trgovanja je planiranje proizvodnje prema cijeni električne energije, rašireniji u zapadnim zemljama, dok u Hrvatskoj još nije korišten. Manager tvrtke može regulirati vrijeme proizvodnje, tako da iskoristi manje cijene električne energije. Naravno, postoje i ograničenja ovisna o tome koje je minimalno vrijeme rada koje je isplativo prema specifikacijama strojeva i radnim navikama radnika.

Početni korak kod svake strategije je pronaći model koji će što točnije predviđati cijenu električne energije. Algoritmi strojnog učenja sve se češće koriste za izradu modela. Regresija potpornim vektorima pokazala je odlične rezultate u procjenjivanju cijene električne energije u trgovanju dan unaprijed.

Poglavlje 3

Analiza podataka

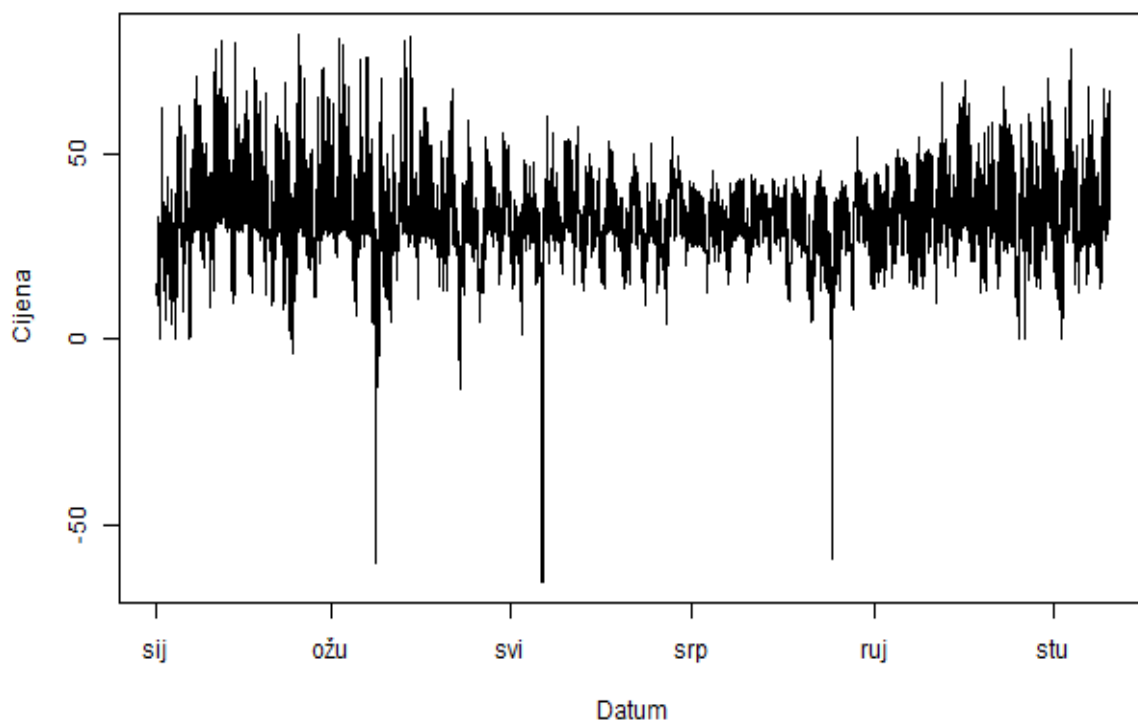
3.1 Općenito o podacima

U prvom poglavlju opisano je kako regresija potpornih vektora predviđa ciljnu vrijednost na temelju naučenog modela i unesenih podataka. Ciljna vrijednost, u ovom projektu, je satna cijena električne energije na njemačkom tržištu. Postavlja se pitanje koji to podaci najviše utječu na formiranje cijene, tj. koji su podaci potrebni računalu da ispravno odredi model i na temelju tog modela što točnije predvidi cijenu. Kako smo opisali u drugom poglavlju, cijena se formira u odnosu na proizvodnju i potrošnju. Električna energija proizvodi se u elektranama koje mogu biti pogonjene na nuklearno gorivo, fosilna goriva, snagu vode, sunca i vjetra. Podaci o proizvodnji podijeljeni su u četiri skupine: dnevna količina proizvodnje iz hidroelektrana, satna proizvodnja iz sunčevih i fotonaponskih elektrana, satna proizvodnja iz vjetroelektrana i satna proizvodnja iz konvencionalnih elektrana (termoelektrane). Još je jedan bitan podatak koji ukazuje na proizvodnju, a to je satna energija padalina. Potrošnja je vidljiva iz podataka o satnoj temperaturi. Iako je možda intuitivno tražiti više podataka o potrošnji, npr. neki indikator gospodarske proizvodnje ili gustoću naseljenosti, eksperimentalno se pokazalo da to ima manji utjecaj na cijenu od podataka o proizvodnji, te ponekada može dovesti do lošijih rezultata modela. Zbog toga, u ovom radu ne koristimo ni jednu od tih vrijednosti. Za potrebe ovog rada, podaci su dobiveni od tvrtke Petrol d.o.o., koja iste koristi za vlastito poslovanje. Tvrtka dio potrebnih podataka preuzima iz javnih izvora kao što su ENTSOE (European network of Transmission System Operators for Electricity), Weather Underground, Eurostat i EMCWF, dok su ostali podaci produkt poslovne suradnje sa specijaliziranim tvrtkama. Podaci su pojedinačno opisani na sljedećim stranicama. Za potrebe treniranja modela sve značajke, s iznimkom satne energije padalina, imaju stvarne vrijednosti, dok se za potrebe testiranja koriste predviđene vrijednosti. Za satnu energiju padalina dostupni su samo predviđeni podaci, što je detaljno objašnjeno u potpoglavlju posvećenom ovoj značajki. Kada ra-

dimo predviđanje za sat, dan ili više dana unaprijed, nemamo stvarne vrijednosti za ciljne vremenske periode, već se koriste očekivane vrijednosti na temelju kojih model predviđa cijenu električne energije. Kako želimo odrediti cijenu električne energije na njemačkom tržištu, sve vrijednosti se odnose na Njemačku.

3.2 Satna cijena električne energije

Kao što smo već rekli, cilj ovog rada je predviđati cijenu električne energije koristeći metodu potpornih vektora. U radu su korišteni podaci za cijenu električne energije od početka 2014. godine do 19. studenog iste godine i to za svaki sat. Svi podaci o cijeni izraženi su u eurima po megawatt satu. Na slici 3.1 prikazani su svi dostupni podaci o cijeni električne energije.

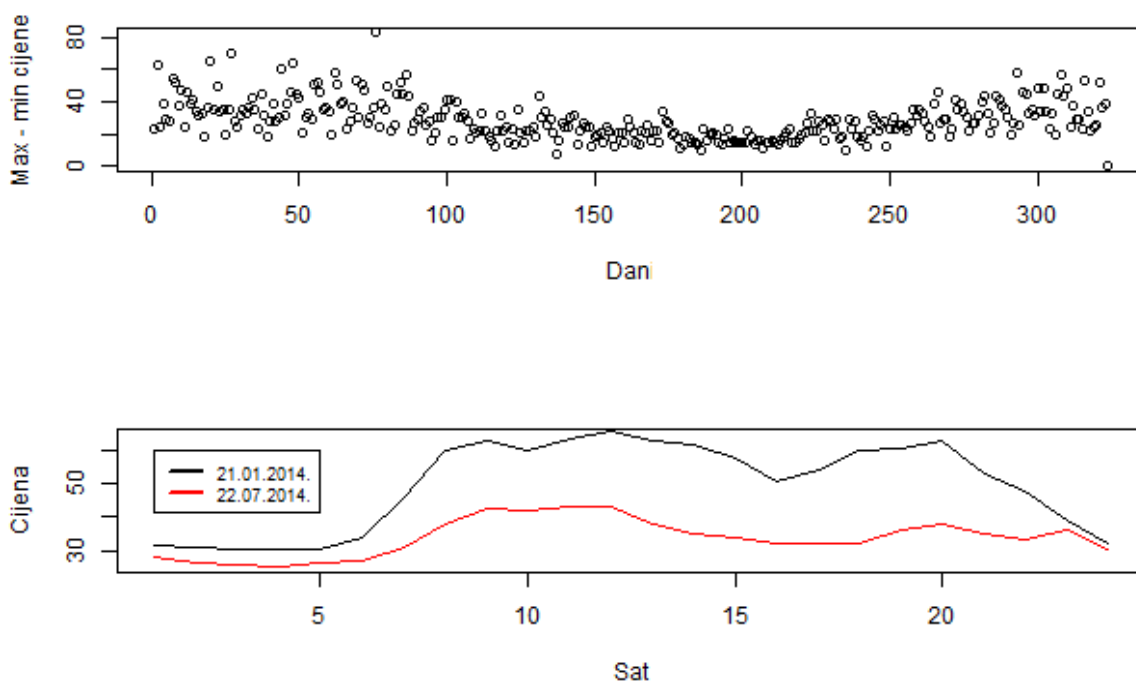


Slika 3.1: Prikaz ponašanja cijene električne energije na njemačkom tržištu za 2014. godinu

Četiri se stvari mogu zapaziti na grafu:

- Cijene ponekad postignu negativne vrijednosti. Razlozi zbog kojih se to događa će biti objašnjeni malo kasnije.
- Cijena jako varira. Na grafu su jasno vidljivi svakodnevni skokovi u cijeni. Mogu se primijetiti i tri izrazita skoka kada je cijena poprimila negativnu vrijednost manju od -50 €/MWh .
- Cijena električne energije u zimskim mjesecima više varira, nego u ljetnim mjesecima.
- U zimskim mjesecima cijene postiže više vrijednosti, nego u ljetnim.

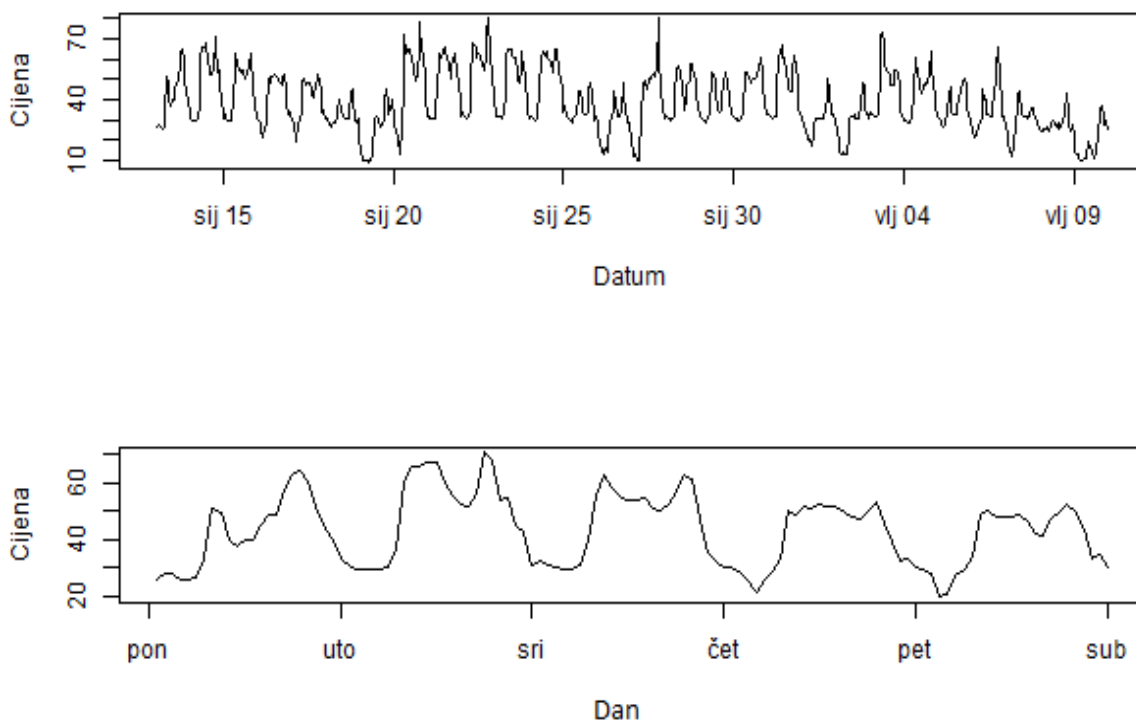
Prvo obrađujemo varijacije u cijeni. U tu svrhu napravljen je graf koji prikazuje razliku maksimalne i minimalne cijene u danu (slika 3.2, gornji graf).



Slika 3.2: Razlika maksimalne i minimalne cijene po danu za 2014. godinu i usporedba ponašanja cijene za jedan zimski i jedan ljetni dan

Ekstremne vrijednosti cijene, ispod -50 €/MWh, izbačene su za potrebe ovog grafa. Vidljivo je da razlika maksimalne i minimalne cijene u danu postiže ekstremnije vrijednosti u zimskim, nego u ljetnim mjesecima. Zbog toga se može zaključiti da su u ljetnim mjesecima cijene stabilnije. Isto se može primijetiti i na drugom grafu, koji prikazuje ponašanje cijene tijekom jednog zimskog i jednog ljetnog dana (slika 3.2, donji graf). Uzeti su podaci za dane 21. 1. i 22. 7. koji su posebno odabrani, jer se radi o sličnim danima: radni dan, utorak, bez neposredne blizine praznika. Na grafu je jasno vidljivo da cijene kreću od približno istih vrijednosti, no za zimski dan postižu veće vrijednosti tijekom dana, nego za ljetni dan. Još jedna stvar je vidljiva na ovom grafu. Cijene tijekom noći su približno jednake, dok se tijekom dana stvara veća razlika, to jest razlika cijena za isti sat odabranih dana je manja noću, a povećava se tijekom dana.

Cijena električne energije ne ovisi samo o periodu godine. Sličan je fenomen primjetan na tjednoj i dnevnoj razini (vidjeti sliku 3.3).

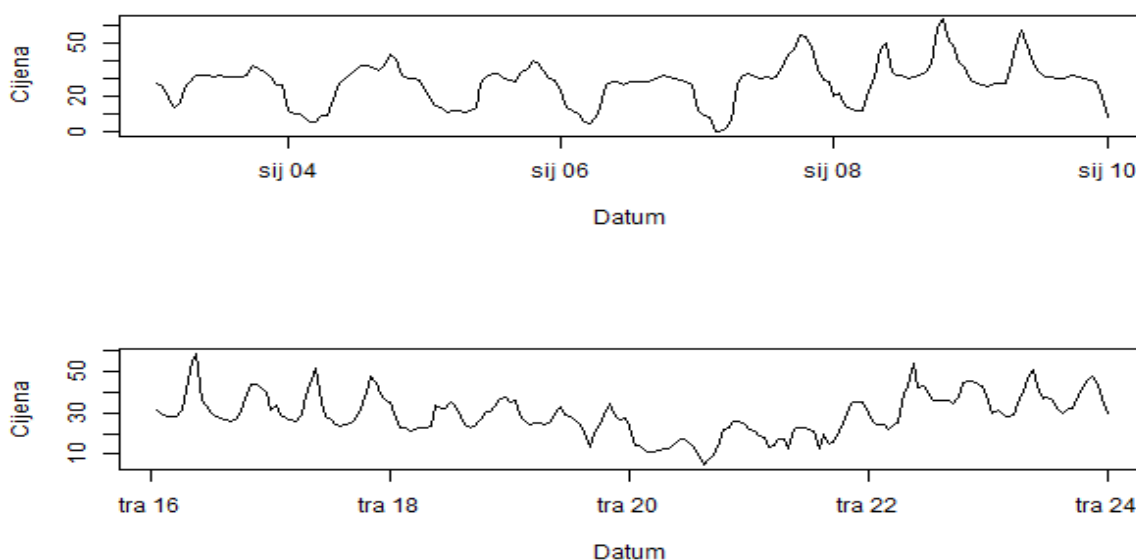


Slika 3.3: Prikaz pada cijene električne energije vikendom i ponašanja u danu

Prvi graf pokazuje kretanje cijene električne energije za period od 13. 1. do 9. 2. 2014. Postoji očita pravilnost pada cijene električne energije tijekom vikenda. Ovaj fenomen je i očekivan, jer se vikendom troši manje električne energije zbog smanjenje industrijske proizvodnje, te cijena pada zbog manje potražnje.

Ista logika vrijedi i za promjene u danu. Cijena pada tijekom noći, jer je tada potražnja manja. Iako se i proizvodnja smanjuje, što se biti prikazano kasnije, to se ne događa proporcionalno u odnosu na manju potražnju te cijena pada. Prikaz kretanja cijene od 13. 1. do 17. 1. (slika 3.3, donji graf) to potvrđuje. Sada se postavlja pitanje zašto se proizvodnja ne smanjuje proporcionalno potrošnji? Vrijeme potrebno za gašenje i ponovno paljenje postrojenja može biti preveliko da bi se događalo na dnevnoj razini. Osim vremena potrebnog za gašenje i paljenje postrojenja, problem može biti i cijena potrebna za taj postupak. Ovdje se krije uzrok pojavljivanja prije spomenutih negativnih vrijednosti. Proizvođačima električne energije isplativije je plaćati kupcima kupovinu električne energije, nego gasiti i ponovno paliti pogon. Na grafovima na slikama 3.2 i 3.3 primjetna je još jedna karakteristika cijene električne energije. Ako se поближе pogleda kretanje cijene električne energije tijekom dana, primjećuje se sljedeći obrazac: cijena polako raste do podneva, zatim pada i ponovno raste u kasno poslijepodne, gdje najčešće dostiže cijenu postignutu u podne, ili je čak i prestiže. Otprilike oko 20 sati navečer, cijena ponovno počinje padati.

Postoje i posebni dani koji utječu na cijenu električne energije, a to su blagdani.

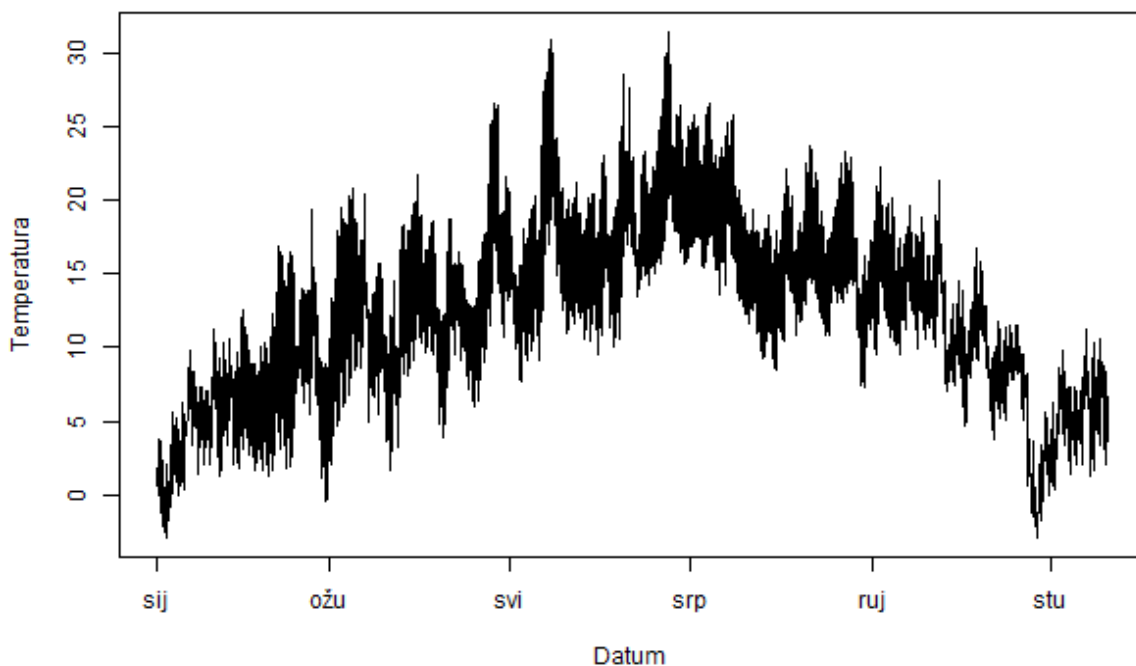


Slika 3.4: Prikaz ponašanja cijene električne energije oko blagdana

Grafovi na slici 3.4 prikazuju kretanje cijena na i oko blagdana. Gornji graf prikazuje situaciju oko blagdana Sveta tri kralja, dok donji graf donosi pregled situacije oko Uskrsa, 20. travnja 2014. Možemo primijetiti da cijena pada što je i očekivano jer potrošnja pada, to jest vrijedi ista logika kao za vikende.

3.3 Satna temperatura

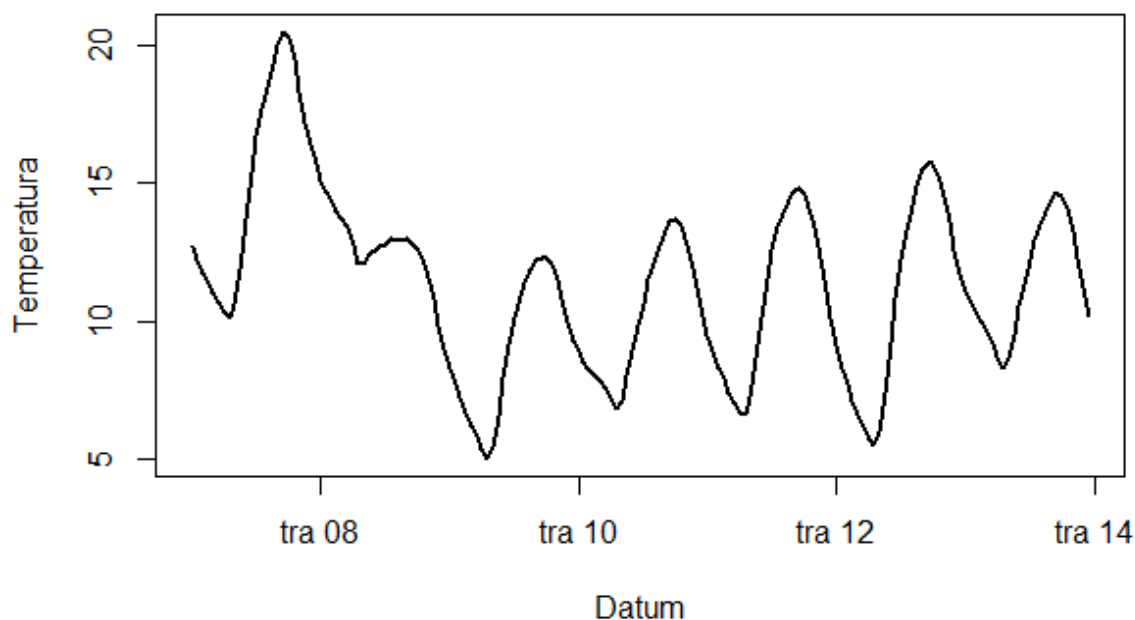
Temperatura utječe na svakodnevni život i navike ljudi, pa time i na cijenu električne energije. Kada su temperature niže ili više od ugodnih za život, koriste se aparati za grijanje, odnosno hlađenje. U ovom radu su korišteni podaci za temperaturu u Njemačkoj za 2014. godinu od 27. 1. do 18. 11. Kao i podaci za cijenu električne energije, podaci za temperaturu su dani na satnoj razini. Sve vrijednosti su u °C.



Slika 3.5: Prikaz ponašanja temperature zraka u Njemačkoj 2014. godine

Graf na slici 3.5 donosi pregled svih podataka o temperaturi korištenih u ovom radu. Uspoređujući graf na slici 3.1 s ovim grafom, odmah se ističe sklonost variranju, koja je

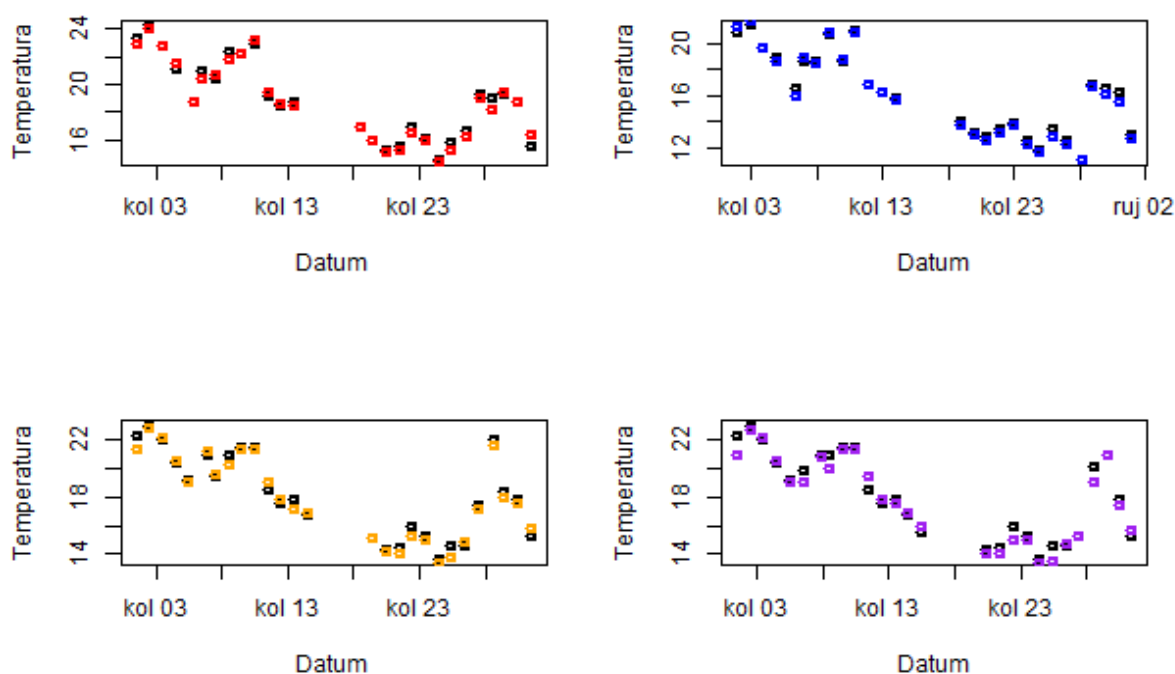
primjetna i kod cijene električne energije i kod temperature. Ovo je dobar pokazatelj o utjecaju temperature na cijenu električne energije. Prije je već komentirana povezanost cijene električne energije s razdobljem u godini. Temperatura pokazuje slične karakteristike, vrijednost je mnogo manja zimi i veća ljeti. Slučaj s cijenom električne energije je obrnut, veće vrijednosti se postižu zimi, a manje ljeti. Ovakav obrnuto proporcionalan odnos je još jedan pokazatelj utjecaja temperature na cijenu električne energije. Pogled na graf temperature u periodu od 7. 4. do 14. 4. 2014. otkriva i treću poveznicu između cijene i temperature (slika 3.6). Radi se o padu temperature u noćnim satima što možemo povezati s padom cijene u noćnim satima komentiranom u prošlom potpoglavlju.



Slika 3.6: Pad temperature u noćnim satima

Kao što je već ranije rečeno, za treniranje modela koriste se stvarne vrijednosti, a za rad modela predviđene vrijednosti. Koliko će model dobro procjenjivati cijenu električne energije ovisi i o točnosti predviđanja temperature. U praksi se koriste podaci o predviđanjima temperature, a i drugih značajki, dobiveni od drugih izvora te se ne može previše utjecati na točnost tih predviđanja. Najviše kako se može utjecati na te podatke je izborom najvjerodostojnijeg izvora podataka, odnosno izvora s čijim podacima se postiže optimalan rad

algoritma. Previđanja se rade u određene sate, koji nisu uvijek isti, i to nekoliko sati unaprijed. Period koliko sati unaprijed će se raditi predviđanje također varira. Rezultat toga je da postoji više predviđanja za određeni trenutak u vremenu, i to napravljenih u različito vrijeme. Na primjer, za 4. sat dana 5. 10. 2014. dostupna su predviđanja napravljena 21. 9., 22. 9., 23. 9., 24. 9., 25. 9., 26. 9., 27. 9., 28. 9., 29. 9., 1. 10. te 2. 10., sva napravljena u 11 sati. Naravno, ova predviđanja se međusobno razlikuju. U praksi se uobičajno uzima najnovije predviđanje. Slika 3.7 prikazuje stvarnu i predviđenu temperaturu za mjesec kolovoz, za izabrano vrijeme predviđanja.



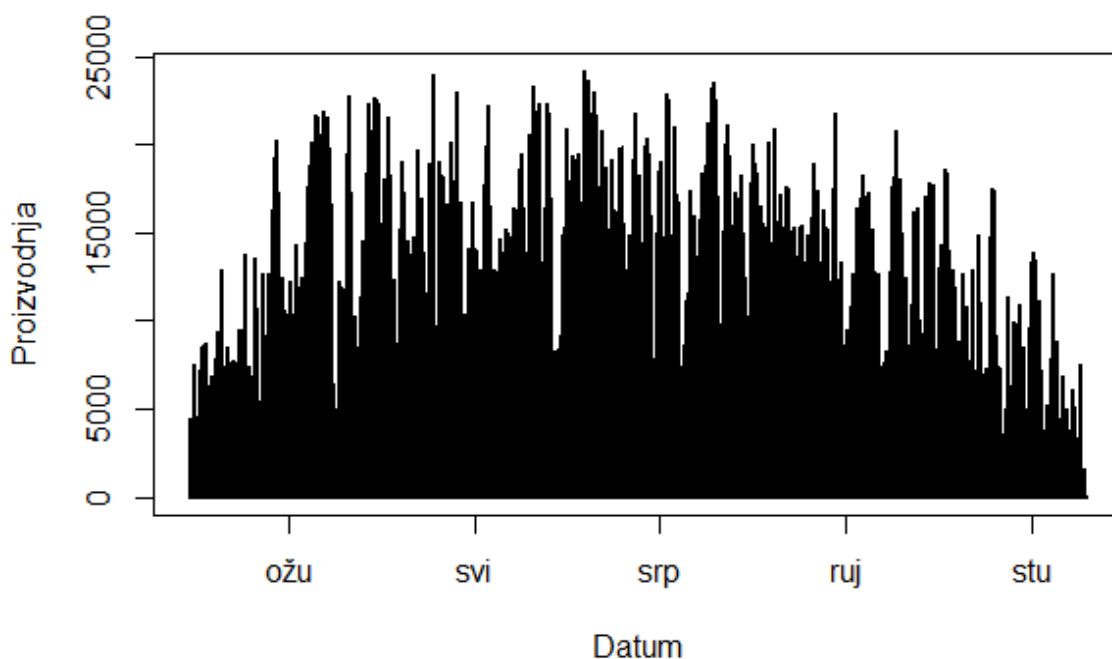
Slika 3.7: Odnos stvarne i predviđene temperature zraka za kolovoz 2014. godine

Napravljena su četiri grafa. Na sva četiri grafa crno su označene točke stvarnih temperatura. Crvenom bojom označene su točke predviđanja napravljenih sat vremena unaprijed, kao što je vidljivo na grafu u gornjem lijevom kutu slike. Graf u gornjem desnom kutu prikazuje predviđanja napravljena 12 sati unaprijed. Ta su predviđanja označena plavom bojom, dok crna opet označava stvarne vrijednosti. Narančasta boja rezervirana je za predviđanja napravljena 24 sata unaprijed. Ti podaci su prikazani na donjem lijevom grafu. Posljednji graf, ljubičastom bojom, prikazuje predviđanja napravljena dva dana unaprijed,

odnosno 48 sati unaprijed. Možemo primijetiti da predviđanja dosta dobro prate stvarne vrijednosti. Ostaje pitanje: Hoće li netočnost, koja ipak postoji u predviđanjima, koliko god ona bila mala, utjecati na točnost algoritma za predviđanje cijene električne energije? Najbolji način da to vidimo je testiranje algoritma.

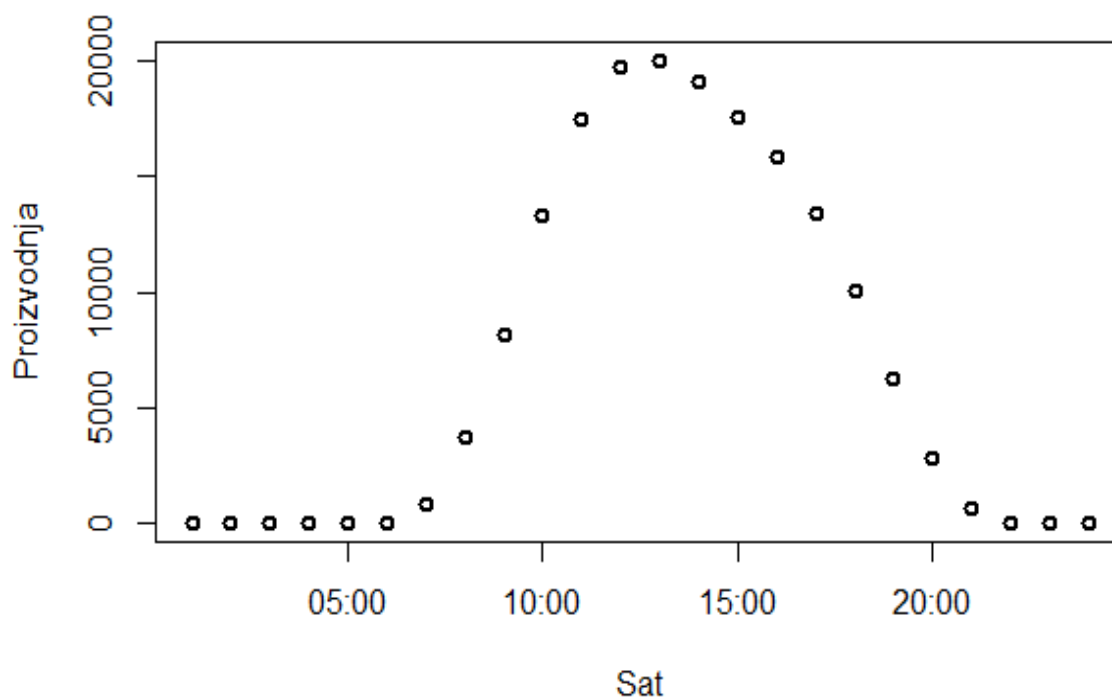
3.4 Satna proizvodnja električne energije iz sunčevih i fotonaponskih elektrana

Prvi podaci o proizvodnji o kojima će se govoriti su podaci o proizvodnji električne energije iz sunčevih i fotonaponskih elektrana. Interval podataka je od 27. 1. do 19. 11. 2014. godine. Većina podataka o proizvodnji, pa i ovi podaci, izražena je u MWh/h što ukazuje da su podaci dostupni u satnom formatu. Pregled cijelog raspona podataka vidljiv je na grafu na slici 3.8.



Slika 3.8: Ponašanje proizvodnje električne energije iz sunčevih i fotonaponskih elektrana

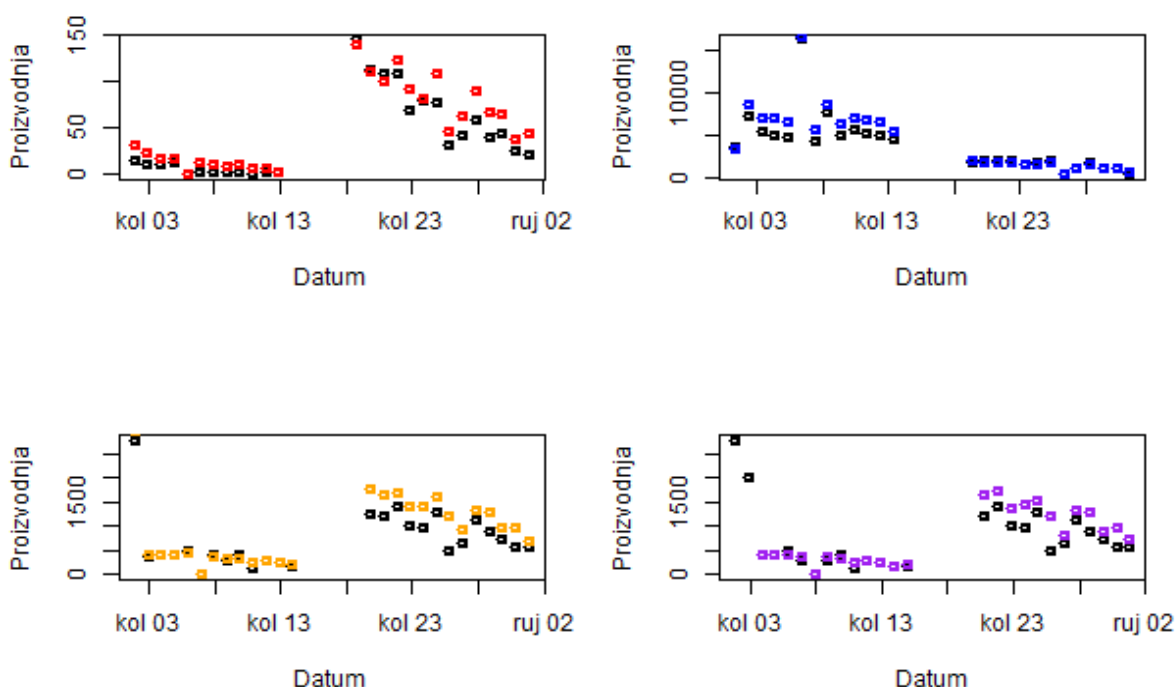
Vrijednosti proizvodnje iz sunčevih i fotonaponskih elektrana nikada ne padaju ispod nule, to jest nula je donja granica što nije bio slučaj s cijenom i temperaturom. S obzirom na to da se radi o proizvodnji, postojanje donje granice je bilo i očekivano, te će biti primjetno i kod proizvodnje iz ostalih izvora. Već viđena varijacija vrijednosti primijetiva je i na ovom grafu. Ponavljanje ovog fenomena u više svojstava djeluje pozitivno na rad algoritma regresije potpornim vektorima. Primjetno je i smanjenje vrijednosti proizvodnje električne energije iz sunčevih i fotonaponskih elektrana tijekom zimskih mjeseci i povećavanje tijekom ljetnih. Isto je vidljivo i kod podataka o cijeni električne energije, što ukazuje na još jednu poveznicu cijene i proizvodnje iz sunčevih i fotonaponskih elektrana. Iako je na grafu vidljivo da se vrijednost nula postiže svakodnevno, zbog velike količine podataka, nemoguće je vidjeti kretanje proizvodnje tijekom dana. Na slici 3.9 prikazani su podaci o proizvodnji električne energije iz sunčevih i fotonaponskih elektrana na dan 31. 8. 2014.



Slika 3.9: Ponašanje proizvodnje električne energije iz sunčevih i fotonaponskih elektrana nalikuju kretanju sunca na obzoru

Na grafu je odmah vidljivo da proizvodnja prati kretanje sunca na obzoru. Noću nema proizvodnje, ujutro se povećava, dostiže maksimum u 12 sati, te zatim pada do 20 sati kada ponovno prestaje proizvodnja. Prisjećanjem na graf kretanja cijene tijekom dana (vidjeti sliku 3.2), primjetna je određena sličnost. Savršenu sličnost narušava pad cijene u poslijepodnevним satima i rast oko 20 sati.

Kao što smo već komentirali, bitna stavka je i odnos predviđenih i stvarnih vrijednosti. Sukladno slici 3.7 napravljena je slika 3.10 koja sadrži prikaz odnosa stvarnih i predviđenih vrijednosti.

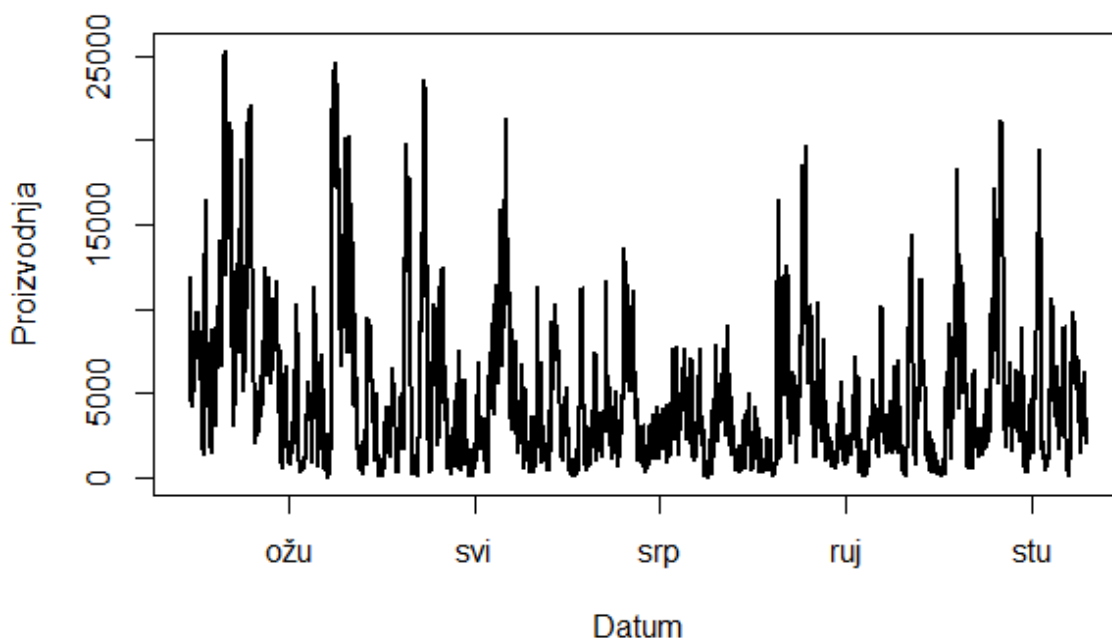


Slika 3.10: Odnos stvarne i predviđene proizvodnja iz sunčevih i fotonaponskih elektrana za kolovoz 2014.

Boje ponovno predstavljaju koliko sati unaprijed je predviđanje napravljeno (crvena – jedan sat, plava – dvanaest sati, narančasta – dvadeset četiri sata i ljubičasta – četrdeset osam sati). Opet je zamjetna određena podudarnost podataka, iako ne baš toliko dobra, kao što je bio slučaj s temperaturom.

3.5 Satna proizvodnja električne energije iz vjetroelektrana

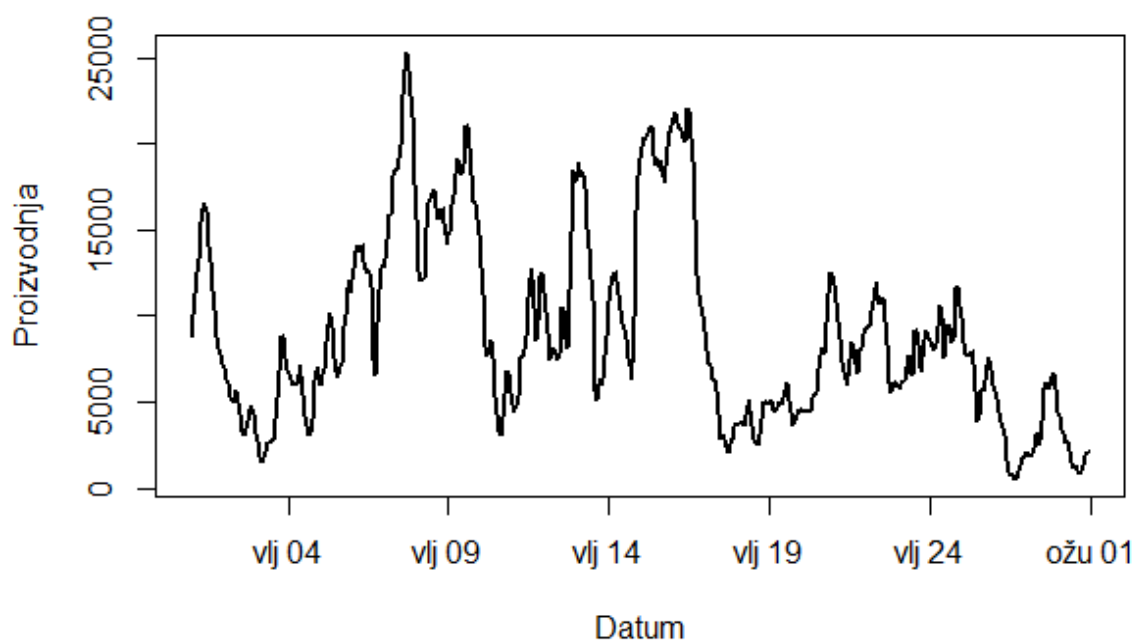
Povećanjem iskorištavanja obnovljivih izvora energije, povećava se i proizvodnja iz vjetroelektrana, a samim tim i povezanost cijene s proizvodnjom iz ovog obnovljivog izvora. Svi podaci o proizvodnji iz vjetroelektrana prikazani su na grafu na slici 3.11. Radi se o rasponu od 27. 1. do 19. 11. 2014. godine u intervalima od svakog sata. Kao što je već rečeno, mjerna jedinica je MWh/h.



Slika 3.11: Ponašanje proizvodnje električne energije iz vjetroelektrana za 2014. godinu

Odmah se primjećuju izraziti skokovi u proizvodnji. Skokovi su manjeg intenziteta u ljetnim mjesecima u odnosu na zimske mjesece, što je ponašanje primijećeno i kod cijene električne energije (slika 3.1). Usporedba grafa cijene na slici 3.1 s proizvodnjom iz vjetroelektrana (graf 3.11) pokazuje postojanje veza između stabilnosti cijena u ljetnim mjesecima i stabilnosti proizvodnje u spomenutim mjesecima. Ovo je dobar pokazatelj da će proizvodnja iz vjetroelektrana biti bitan faktor u formiranju cijene. Kod cijene je sličan

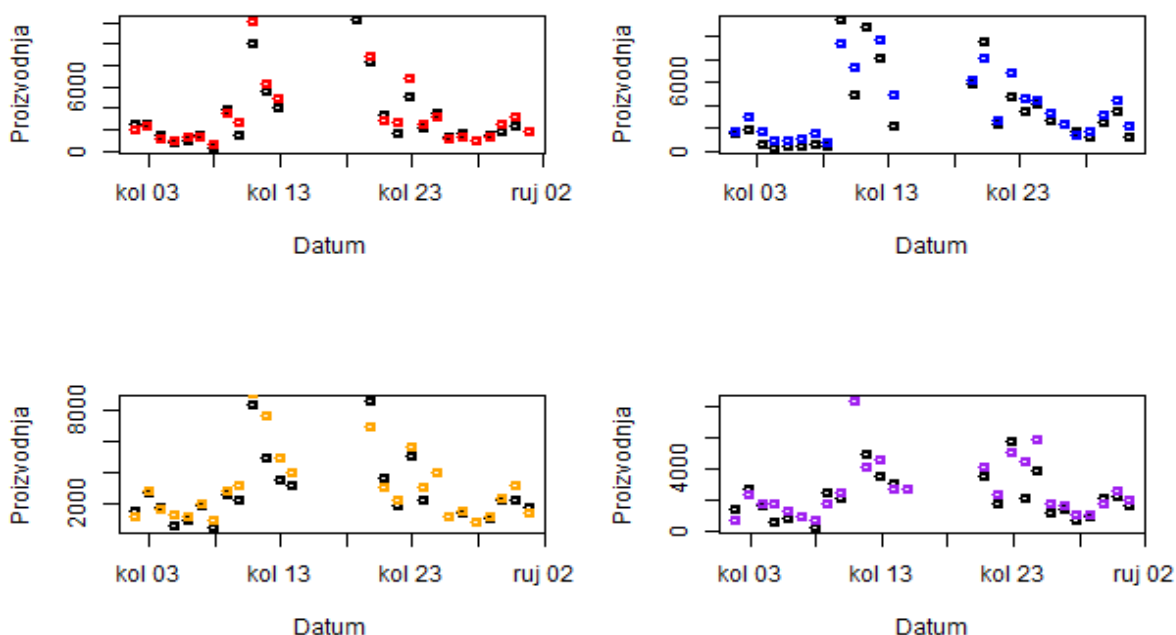
fenomen prisutan i na tjednoj i na dnevnoj razini. Kod proizvodnje iz vjetroelektrana, to nije slučaj kao što je vidljivo na sljedećem grafu (slika 3.12). Prikazana je proizvodnja električne energije iz vjetroelektrana za mjesec veljaču.



Slika 3.12: Ne postoji fenomen smanjenja proizvodnje iz vjetroelektrana na tjednoj i dnevnoj razini

Ukoliko se pokaže da je uistinu proizvodnja iz vjetroelektrana bitan faktor, povećava se pritisak na točnost predviđanja. Odnos stvarnih i predviđenih podataka iskazan je na grafovima na slici 3.13.

Vidljivo je da predviđanja prate stvarne vrijednosti, ali ovdje je primjetno više odstupanja obzirom na grafove stvarnih i predviđenih vrijednosti temperature i proizvodnje iz sunčevih i fotonaponskih elektrana (slike 3.7 i 3.10).

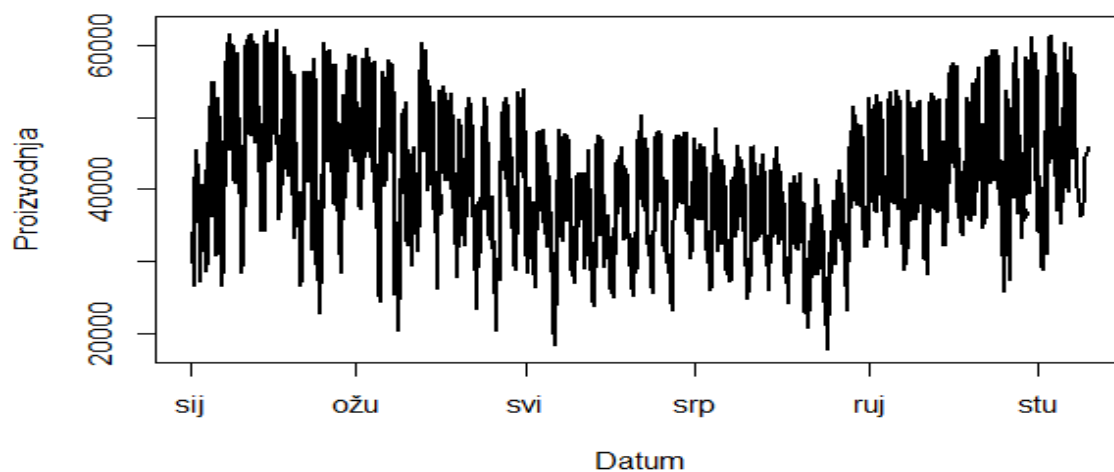


Slika 3.13: Odnos stvarne i očekivane proizvodnje iz vjetroelektrana za kolovoz

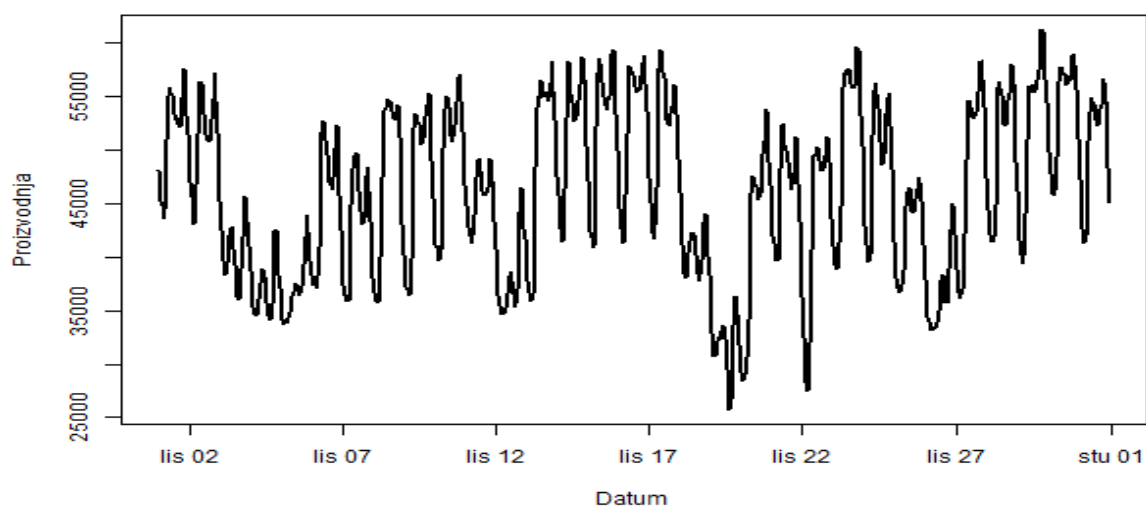
3.6 Satna proizvodnja električne energije iz konvencionalnih elektrana iznad 100 MW instalirane snage

Povećanje korištenja obnovljivih izvora energije nije smanjilo važnost konvencionalnim izvorima, kao što su ugljen, nafta, plin i nuklearna energija. Proizvodnja iz ovih izvora još uvijek predstavlja važan dio ukupne proizvodnje električne energije. U ovom radu korišteni su podaci u periodu od 1. 1. do 19. 11. 2014. godine, koji su prikazani na sljedećem grafu (slika 3.14).

Već dobro poznato ponašanje velikog variranja vrijednosti uočljivo je i na ovom grafu. U ovom slučaju su ekstremi u nižim vrijednostima veći, dok su skokovi u većim vrijednostima ujednačeniji. Na godišnjoj razini, vrijednosti su stabilnije u ljetnim mjesecima nego u zimskim, gdje pronalazimo poveznicu s cijenom. Graf nagovještava da bi i na tjednoj i na dnevnoj razini mogle postojati zanimljive pravilnosti. U svrhu pronalaska tih pravilnosti napravljen je graf proizvodnje za mjesec listopad (slika 3.15).

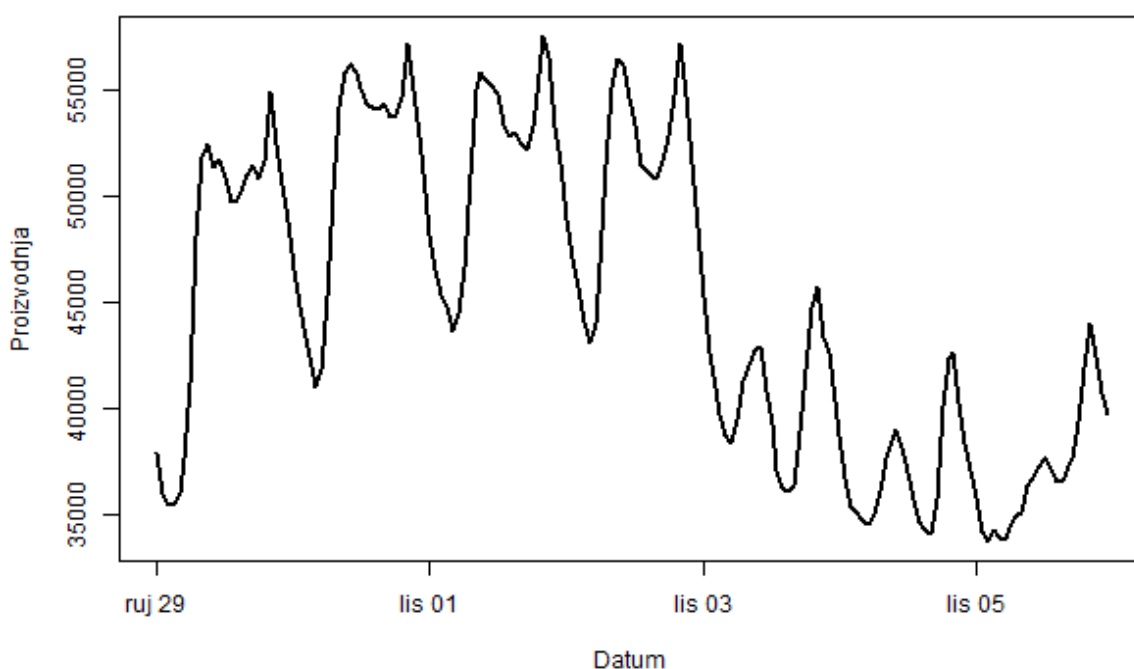


Slika 3.14: Ponašanje proizvodnje električne energije iz konvencionalnih elektrana za 2014. godinu



Slika 3.15: Pad proizvodnje iz konvencionalnih izvora na tjednoj razini

Svakog vikenda događa se pad proizvodnje električne energije iz konvencionalnih izvora. Razlog za to je mogućnost reguliranja proizvodnje iz konvencionalnih izvora, što nije slučaj s proizvodnjom iz sunčevih elektrana i vjetroelektrana. Kontrolori u konvencionalnim elektranama prate potrošnju u sustavu i reguliraju proizvodnju sukladno tome. S obzirom na to da je potrošnja manja vikendom, što je objašnjeno u poglavlju o satnoj cijeni električne energije, smanjuje se i proizvodnja iz konvencionalnih elektrana. Ovaj proces regulacije proizvodnje iz konvencionalnih izvora događa se i na dnevnoj razini, te rezultira situacijom vidljivom na grafu 3.16.

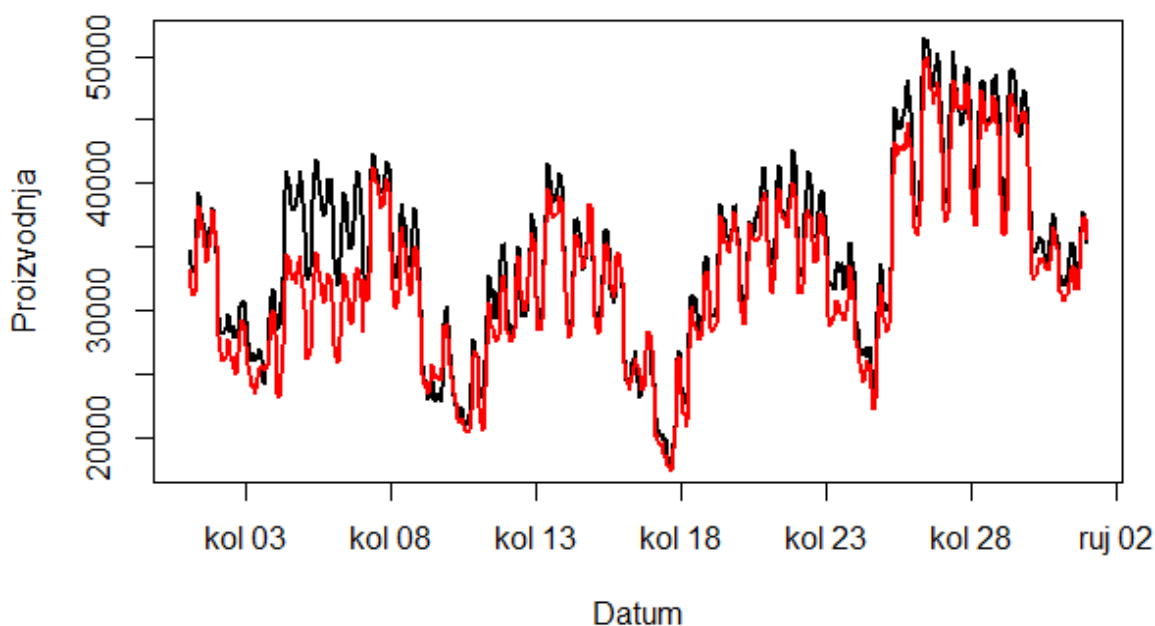


Slika 3.16: Pad proizvodnje električne energije iz konvencionalnih elektrana na dnevnoj razini

Prikazane su vrijednosti za tjedan od 29. rujna do 6. listopada. Primjetno je da proizvodnja iz konvencionalnih izvora prati cijenu i na dnevnoj razini, što znači da su tijekom noći vrijednosti male, rastu tijekom jutra, kada postižu lokalni maksimum, padaju tijekom popodneva, te ponovno rastu navečer kada postižu globalni maksimum.

Kod prethodno analiziranih podataka, postojalo je više predviđanja za isti trenutak u vremenu. Ovdje to nije slučaj, dostupno je točno jedno predviđanje za svaki vremenski

trenutak. Kao i u prošlim potpoglavljima, prikazat ćemo odnos stvarnih i predviđenih vrijednosti za mjesec kolovoz (slika 3.17).



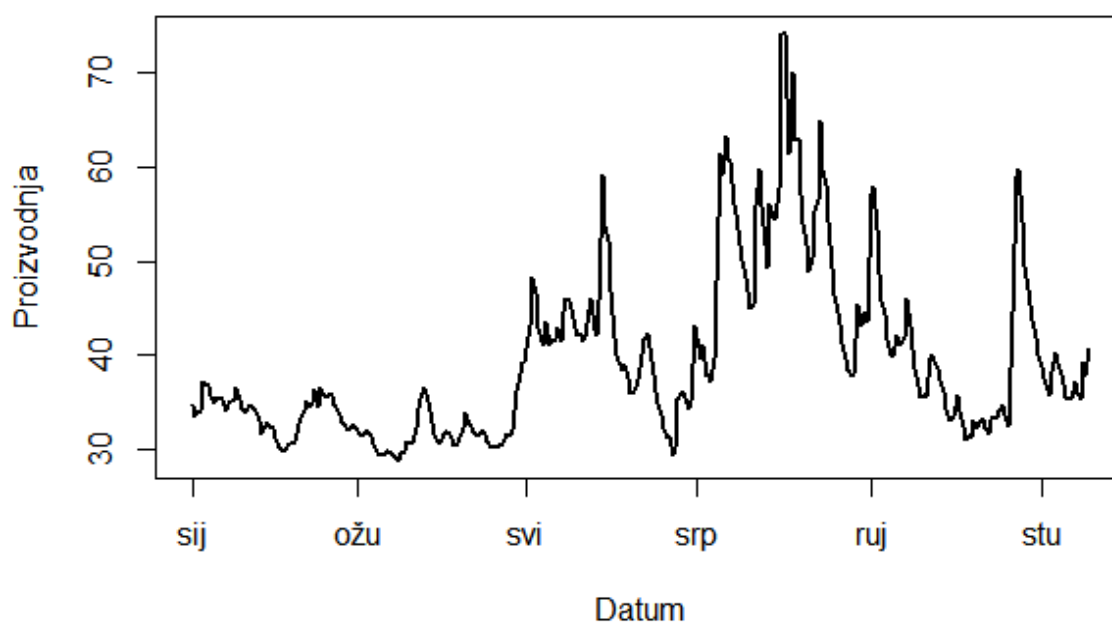
Slika 3.17: Odnos stvarne i predviđene proizvodnje iz konvencionalnih izvora

Crnom bojom prikazana je stvarna vrijednost, a crvenom predviđena. Izuzev odstupanja početkom mjeseca, predviđanja dobro prate stvarne vrijednosti – uspješno predviđaju pad proizvodnje vikendom, te kretanje proizvodnje na dnevnoj razini.

3.7 Dnevna proizvodnja električne energije iz hidroelektrana

Dok su ostali podaci dostupni za svaki sat, podaci o proizvodnji električne energije iz hidroelektrana dostupni su samo za pojedini dan. U radu sa stvarnim podacima, česta je situacija da podaci nisu dostupni u formatu kojem bismo to željeli. Prije treniranja modela, morat ćemo odlučiti na koji način ćemo ove podatke koristiti, ali o tome više u četvrtom poglavlju. Sada ćemo prikazati podatke u formatu u kojem su dostupni, to jest prikazat ćemo

ih po danima. Dostupni su podaci od prvog siječnja do 18. studenog 2014. godine. S obzirom na to da se radi o proizvodnji za cijeli dan mjerna jedinica je GWh.



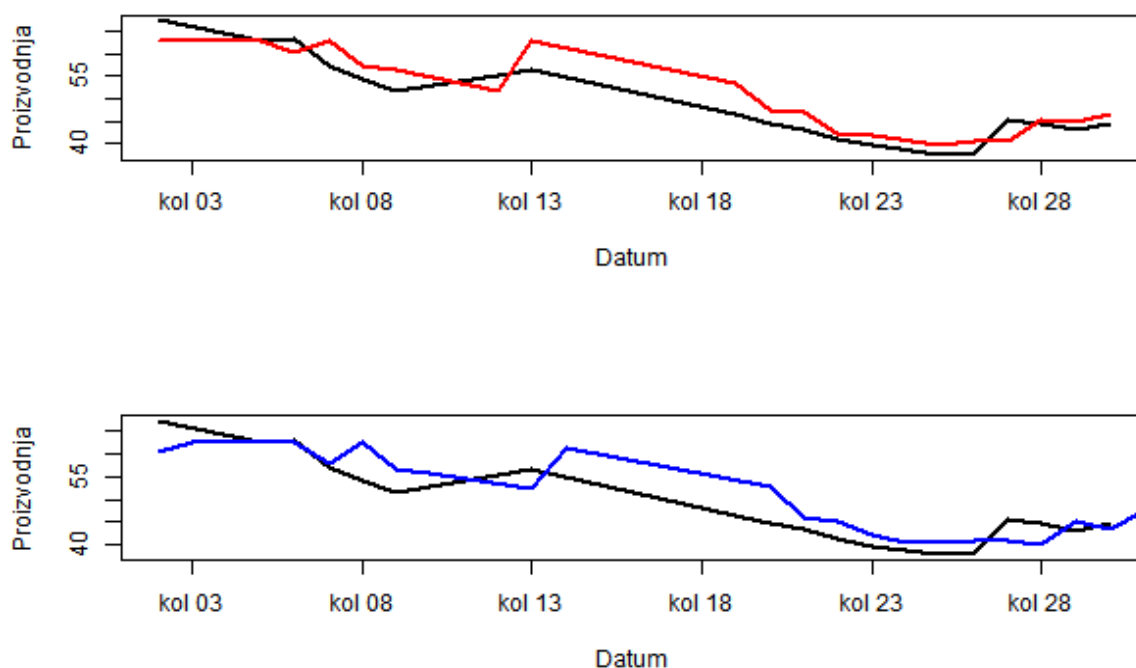
Slika 3.18: Ponašanje proizvodnje električne energije iz hidroelektrana u 2014. godini

Uočljivo je nekoliko stvari:

- Prva četiri mjeseca u godini postižu se manje vrijednosti u odnosu na ostale mjesece.
- Vrijednosti jako variraju, a skokovi su svugdje pristupni.
- Ne postoji nikakva vidljiva veza s cijenom električne energije na godišnjoj, tjednoj ili dnevnoj razini.

Slično kao za prije opisane podatke, za jedan dan postoji više predviđanja koja su napravljena na različite dane. Graf predviđenih i stvarnih vrijednosti dan je na slici 3.19.

Crnom linijom označene su stvarne vrijednosti, crvenom i plavom predviđanja napravljena dan, odnosno dva unaprijed. Predviđanja ne prate stvarne vrijednosti onoliko dobro

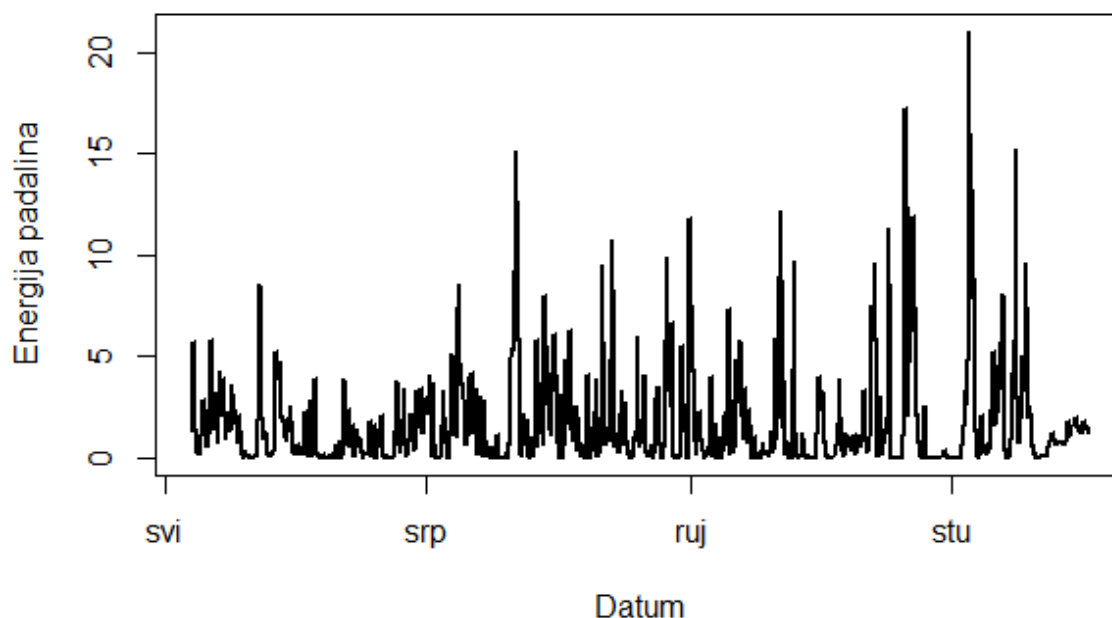


Slika 3.19: Odnos predviđene i stvarne vrijednosti za kolovoz

koliko smo naučeni iz razmatranja prethodnih podataka. Vidljivo je da predviđanja proizvodnje iz hidroelektrana nije lako točno izvesti i zbog toga mogu predstavljati problem kod predviđanja cijene električne energije.

3.8 Satna energija padalina

Energija padalina je vrijednost koja govori koliko se električne energije potencijalno može dobiti proizvodnjom u hidroelektranama od količine padalina koja je pala na nekom području u nekom trenutku. Podaci koji se koriste u ovom radu odnose se na cijelo područje Njemačke u periodu od 7. 5. do 2. 12. 2014. godine za svaki sat. Mjerna jedinica je GWh, jer se radi o potencijalnoj količini električne energije. S obzirom na to da se radi o potencijalu padaline, vrijednosti se ne mogu egzaktno mjeriti. Umjesto toga koriste se podaci o količini oborina, te se iz tih podataka procjenjuje potencijal padalina. Graf na slici 3.20 prikazuje sve podatke.

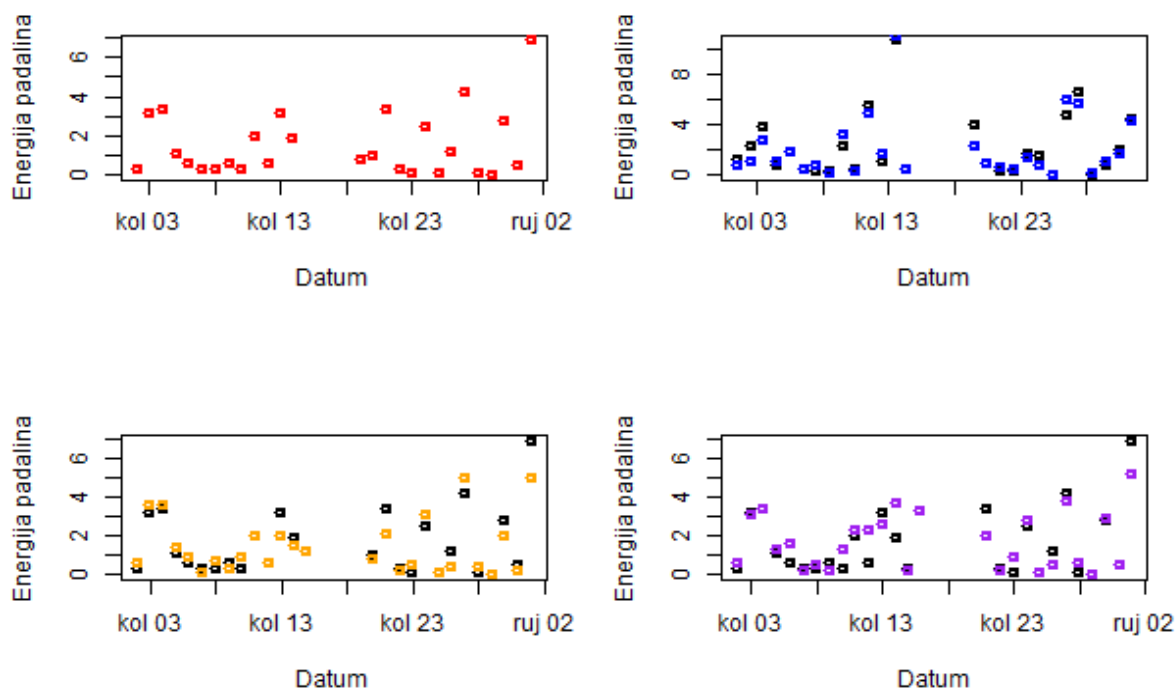


Slika 3.20: Ponašanje energije padalina za period 7. 5. do 2. 12. podsjeća na ponašanje proizvodnje iz hidroelektrana

Pogled na graf 3.18 – proizvodnja električne energije iz hidroelektrana, otkriva jasnu povezanost dviju značajki. Povećanje energije padalina rezultira povećanjem proizvodnje iz hidroelektrane i obrnuto, smanjenje energije padalina rezultira smanjenjem proizvodnje. Na grafu je primjetno i da energija padalina jako varira, kao i sve druge značajke do sada. Ovakva povezanost značajki međusobno, i s ciljnom vrijednosti, je poželjna za optimalan rad algoritma regresije potpornim vektorima.

Objašnjeno je kako se podatak o energiji padalina dobiva od podatka o količini padalina. Kada radimo dan ili više unaprijed, količina padalina nije poznata. Zato se koriste modeli koji procjenjuju količinu padalina. Primjer iz svakodnevnog života su vremenske prognoze. Ti procijenjeni podaci se onda unose u algoritam koji iz njih predviđa energiju padalina. Na slici 3.21 su četiri grafa. Na svima su crnom bojom označene točke koje odgovaraju energiji padalina predviđenoj iz podatka o stvarnoj količini padalina. Predviđanja koja su gore opisanim postupkom napravljena sat vremena unaprijed, prikazana su crvenom bojom. Preostale boje, plava, narančasta i ljubičasta odgovaraju predviđanjima na-

pravljenima dvanaest, dvadeset četiri i četrdeset osam sati unaprijed.



Slika 3.21: Predviđanja energije padalina za kolovoz

Savršenu podudarnost ima predviđanje napravljeno sat vremena unaprijed. Razlog tome leži u činjenici da su predviđanja količine padalina za jedan sat unaprijed izrazito točna. Predviđanja napravljena više sati unaprijed pokazuju odstupanja, što je i očekivano. Naime, predviđanje količine padalina gubi na točnosti što se povećava razmak između trenutka predviđanja i trenutka za koji se predviđanje izvodi.

Poglavlje 4

Rad algoritma

4.1 Priprema podataka za rad

U obradi podataka primjetno je da postoje određene inkonzistencije u podacima, poput dostupnog vremenskog perioda i formata (satne ili dnevne vrijednosti). Zbog toga je prije rada algoritma potrebno pripremiti podatke kako bi bili primjereni ulazu koji algoritam očekuje. Obavljene su sljedeće radnje:

- U podacima postoje dvije anomalije: 30. 3. nedostaju podaci za dvadeset četvrti sat, a 26. 10. postoji podatak za dvadeset peti sat. Pogledom na spomenute datume, može se primijetiti da se radi o danima kada se vrijeme mijenjalo sa zimskog računanja na ljetno i obrnuto. U matematičkom izvodu regresije potpornim vektorima mnoge jednadžbe koriste skalarni produkt vektora $x_i^T x_i$. Ovo čini regresiju potpornim vektorima osjetljivom na odstupanja tipa jedan sat manje ili više. Problem nedostajućeg sata je riješen tako da su svi sati iza dva sata ujutro pomaknuti za jedan sat unaprijed što je rezultiralo time da nedostaje zapis za drugi sat što je i ispravno s obzirom na to da se promjena sata izvodi tako da se sat pomiče s dva sata na tri sata. Podatak za drugi sat je dodan kao aritmetička sredina podataka za prvi i treći sat. Sličan postupak je obavljen i za 26. 10. Svi podaci za sate iza dva sata su pomaknuti za jedan sat unatrag što je rezultiralo s dva zapisa za drugi sat 26. 10. Ovakav zapis odgovara pomicanju sata s ljetnog na zimskog računanje, kada se s tri sata kazaljka vraća na dva sata. Jedinstveni podatak za dva sata je uzet kao aritmetička sredina dvaju podataka koji su postojali.
- Kod obrade cijene električne energije, primijećen je obrazac kretanja u ovisnosti o vikendima i praznicima. Algoritmu je potrebno pružiti podatak o tome kakav je određen dan. U tu svrhu dodana su tri nova svojstva: radni dan, vikend i praznik. Sva tri svojstva imaju dvije moguće vrijednosti: 1 ako je dan određenog svojstva, 0

ako nije. Na primjer 6. 1. 2014. je bio praznik Sveta tri kralja pa on ima 0 na svojstvu radni dan, 0 na svojstvu vikend i 1 na svojstvu praznik svojstvu. 19. 1. 2014. je bila subota pa ima 0 na svojstvu radni dan, 1 na svojstvu vikend i 0 na svojstvu praznik. Naravno, 22. 1. 2014. , običan radni dan, ima 1 na svojstvu radni dan, 0 na svojstvu vikend i svojstvu praznik.

- Nisu svi podaci dostupni u istom vremenskog razdoblju. Praznine u podacima nisu dobre za rad algoritma, pa se mora odlučiti što učiniti s njima. Jedna opcija je dopuniti nedostajuće podatke, na primjer aritmetičkom sredinom postojećih podataka oko nedostajućeg. Druga opcija je odbaciti vremenske periode kojima neki podatak nedostaje. Odabir koju opciju koristiti uvelike ovisi o ukupnoj količini podataka, te mjestima na kojima nedostaju podaci. Podaci mogu nedostajati tako da ih do određenog datuma nema, zatim su dostupni svi te onda opet nedostaju neki. Na primjer, u podacima korištenim u ovom radu, vrijednosti o cijeni električne energije su dostupne od 1. 1. do 19. 11. 2014. , a za temperaturu od 27. 1. do 17. 11. 2014. U ovom slučaju, može se koristiti najveći interval u kojem su podaci dostupni za sve značajke. Naravno, taj interval može sadržavati jako malo podataka, te se tada mora pristupiti dopunjavanju podataka. U ovom radu korišten je interval od 7. 5. do 11. 11. 2014. Donju granicu određuje dostupnost podataka o energiji padalina, a gornju proizvodnja iz konvencionalnih elektrana. To je rezultiralo s 4536 sati za koje su dostupni svi podaci. Ovo je sasvim dovoljno za optimalan rad regresije potpornim vektorima. Ukoliko najveći interval ne sadrži dovoljno podataka, isti se moraju nadopuniti. Kako će se nadopuniti ovisi o podacima, na primjer ako prate neku razdiobu, ta razdioba se može koristiti za nadopunu. Druga moguća situacija je da ne postoji interval u kojem su dostupni podaci za sve značajke. Tada je nadopunjavanje obvezno i mora mu se pristupiti izuzetno oprezno zbog utjecaja koji će imati na rad algoritma.
- Podaci o proizvodnji električne energije iz hidroelektrana su dostupni u dnevnom formatu. Ostali podaci su dostupni u satnom, što je i format na koji se cilja u ovom radu, s obzirom na to kako se određuje cijena električne energije na burzi koja posluje po modelu trgovanja dan unaprijed. Iz tog razloga, dnevne vrijednosti podijeljene su s dvadeset četiri sata čime su pretvorene u satni format. U nekom drugom slučaju, na primjer da se određuje prosječna dnevna cijena, bilo bi bolje satne vrijednosti pretvoriti u dnevne.

Ispravno postupanje s podacima ovisi najviše o njima i cilju rada algoritma. Ne postoji univerzalno rješenje kako raditi s podacima.

Rezultat svih ovih promjena je oblik podataka koje će algoritam koristiti. Oblik podataka prikazan je u sljedećoj tablici.

Ime svojstva	Format	Mjerna jedinica
Sat u danu	prirodni broj	-
Radni dan	{0, 1}	-
Vikend	{0, 1}	-
Praznik	{0, 1}	-
Temperatura	decimalan broj	°C
Proizvodnja električne energije iz sunčevih i fotonaponskih elektrana	decimalan broj	MWh
Proizvodnja električne energije iz vjetroelektrana	decimalan broj	MWh
Proizvodnja električne energije iz konvencionalnih elektrana	decimalan broj	MWh
Proizvodnja električne energije iz hidroelektrana	decimalan broj	GWh
Energija padalina	decimalan broj	GWh
Cijena električne energije	decimalan broj	€/MWh

4.2 Regresija potpornim vektorima

Svi podaci su zapisani u matrici, u kojoj jedan red odgovara zapisu podataka za jedan sat u gore opisanom formatu. Iz Excel datoteke se učitavaju korištenjem programskog jezika Python. Python ima odličnu biblioteku Sklearn za strojno učenje, koja sadrži funkcije za pretprocesiranje podataka, funkcije za optimizaciju modela, te implementaciju algoritama strojnog učenja.

Kao što je rečeno na kraju prvog poglavlja, prvi korak u radu s regresijom potpornim vektorima je skaliranje podataka. U ovom radu korištena je standardizacija (Z normalizacija) koja normalizira podatke oko 0 sa standardnom devijacijom 1.

$$Z = \frac{x - \mu}{\sigma} \text{ gdje je } \mu = 0, \sigma = 1. \quad (4.1)$$

U tu svrhu je korištena klasa `StandardScaler()` čiji je zadatak skalirati podatke na standardnu normalizaciju [2].

Sljedeći korak koji se često provodi u strojnom učenju je selekcija značajki. Neki problemi imaju jako puno značajki, na primjer, problem određivanja spola pisca iz članka gdje se svaka riječ spaja s brojem pojavljivanja u članku, to jest svaka riječ predstavlja značajku. S obzirom na to da veći broj značajki rezultira duljim radom algoritma, provodi se selekcija. Računa se korelacija između svake značajke i ciljne vrijednosti i biraju one s većom korelacijom, a odbacuju one koje nemaju veliki utjecaj na točnost. Koliko značajki odbaciti, a koliko ostaviti ovisi o podacima i najbolje se procjenjuje testiranjem. S obzirom

na to da u ovom radu postoji deset značajki, selekcija nije potrebna sa stajališta brzine, ali može pridonijeti točnosti. Korištena je Python-ova funkcija `SelectKBest()` za $k = 5$, to jest korišteno je pet svojstava koja imaju najveću koreliranost s ciljnom vrijednošću [4] [1].

U prvom poglavlju je objašnjeno kako je izbor *kernel*-a jako bitan i da je najčešće korišten RBF *kernel* (korišten i u ovom radu), definiran u [5],

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|). \quad (4.2)$$

Sljedeći korak opisan u prvom poglavlju bio je odabir optimalnih parametara. Iz definicije RBF *kernel*-a očito je da γ parametar moramo zadati. γ govori koliko daleko seže utjecaj pojedinog primjera. S obzirom na to da su potporni vektori izrazito bitni u regresiji potpornih vektora, bitno je koliki utjecaj imaju, što regresiju čini jako osjetljivom na parametar γ . Ukoliko je γ prevelik, doći će do prevelike prilagodbe, a ukoliko je premali, do premale prilagodbe. Osim γ , zadaje se i parametar C . C je parametar koji određuje koliko algoritam mora točno raditi na primjerima za trening. Veća vrijednost parametra C će rezultirati boljim radom na trening setu, ali može dovesti do prevelike prilagodbe. Male vrijednosti parametra C mogu dovesti do premale prilagodbe. Odabir optimalnih parametara obavlja se testiranjem. Pomoć pri testiranju može biti Python-ova funkcija `GridSearchCV` koja prima niz vrijednosti za parametre γ i C , te vraća najbolje. S obzirom na to da se ne može testirati cijeli \mathbb{R} , većinom se koriste potencije broja 10. Nakon pronalaska optimalnih parametara funkcijom `GridSearchCV()`, preporučljivo je ručno pretražiti prostor oko vraćenih parametara [3]. U ovom radu korišten je upravo opisani postupak.

Dobiveni parametri korišteni su u regresiji potpornim vektorima. Regresija je provedena klasom `SVR()` u kojoj je implementirana regresija potpornim vektorima koja je opisana u prvom poglavlju. U sljedećem poglavlju bit će prezentirani rezultati rada algoritma.

Poglavlje 5

Rezultati

Prije prezentacije rezultata slijede definicije dviju mjera koje su korištene za evaluaciju rezultata.

$$R^2 = 1 - \frac{\text{Var}(\vec{Y} - \hat{\vec{Y}})}{\text{Var}(\vec{Y})} \quad (5.1)$$

\vec{Y} je vektor točnih vrijednosti, a $\hat{\vec{Y}}$ vektor predviđenih vrijednosti.

$$\text{MAPE} = \frac{100 \%}{n} \left| \frac{\vec{Y} - \hat{\vec{Y}}}{\vec{Y}} \right| \quad (5.2)$$

Obje definicije su preuzete iz [13]. MAPE je postotna mjera pogreške.

Prvi korak je bilo skaliranje podataka na prije opisan način. U jednom slučaju je zatim provedeno traženje optimalnih parametara, a u drugom je prvo provedena selekcija značajki. Funkcija `SelectKBest()` je odabrala sljedećih pet značajki:

- sat u danu
- radni dan
- vikend
- proizvodnja električne energije iz vjetroelektrana
- proizvodnja električne energije iz konvencionalnih elektrana.

Funkcija `GridSearchCV()` korištena je za nalaženje najboljih parametara C i γ takvih da je $C \in \{10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3, 10^4, 10^5, 10^6\}$ i $\gamma \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$. Zanimljivo je da je `GridSearchCV()` odabrao iste parove parametara kao optimalne u slučaju kad je proveden na svim značajkama i u slučaju kad je proveden nakon selekcije. Radi se

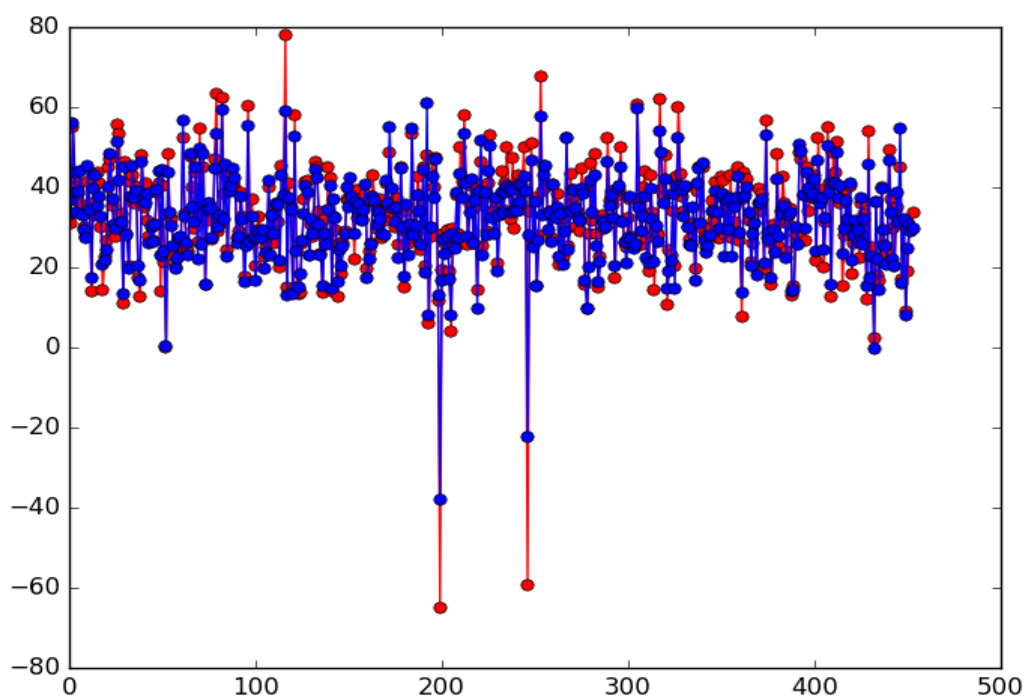
o parametrima:

$$C = 10000,$$
$$\gamma = 9.9999999999999995e^{-03}.$$

U slučaju bez selekcije $R^2 = 0.6735$, a u slučaju sa selekcijom $R^2 = 0.6201$. Ovo su zadovoljavajuće vrijednosti za R^2 , s obzirom na to da male vrijednosti ukazuju na to da model nije dovoljno istreniran, a velike da postoji prevelika prilagodba. Oba slučaja želimo izbjeći. U ovom trenutku nije moguće procijeniti koji model bi mogao dati bolje rezultate.

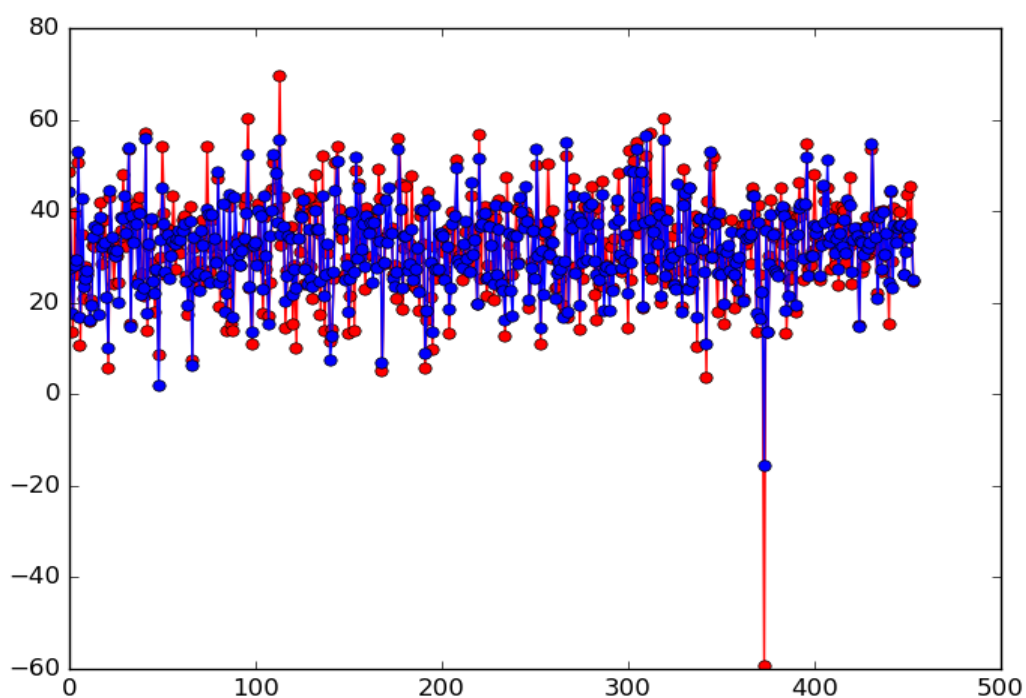
Sljedeći korak je ručno pretraživanje parametara. Korištena je postotna mjera pogreške i izgled grafa. Slučajnim odabirom je odabrano deset posto primjera koji su odvojeni od ostalih. Na preostalim primjerima je model treniran, a na odvojenima testiran. Da bi se osigurala točnost, ovaj korak je proveden deset puta za svaki isprobani par parametara. Za situaciju bez selekcije pronađeni su parametri $C = 13000$ i $\gamma = 9.9999999999999995e^{-03}$, za koje je prosječna postotna mjera pogreške $\text{MAPE} = 15.37\%$.

Na slici 5.1 prikazan je graf koji je dobiven u jednoj od iteracija za optimalne parametre s $\text{MAPE} = 10.04\%$.



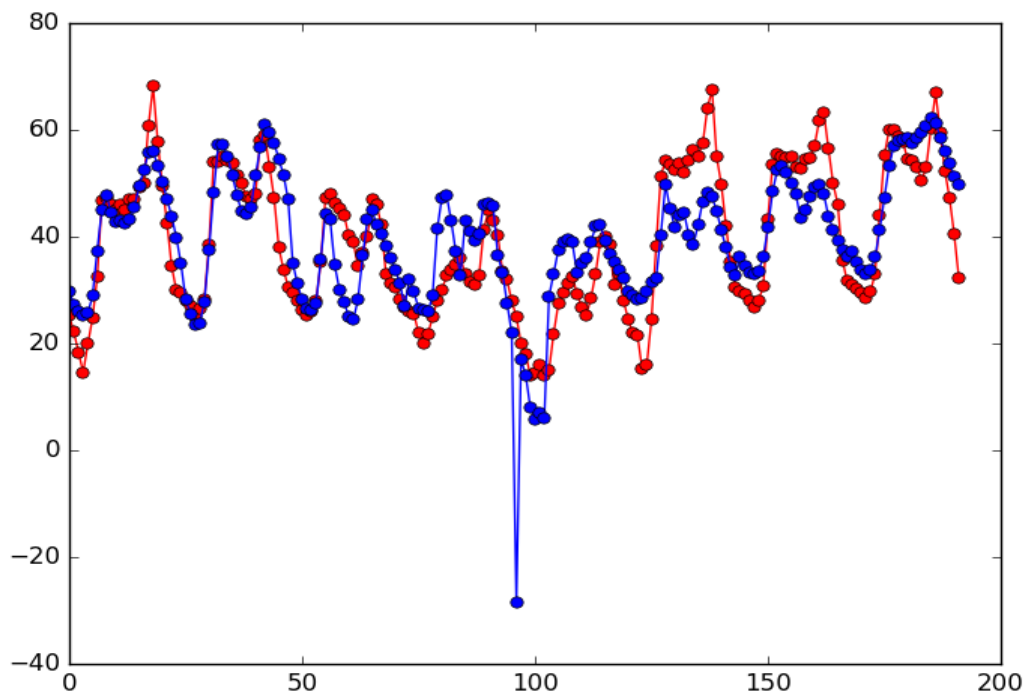
Slika 5.1: Usporedba stvarnih i predviđenih cijena, model bez selekcije

Crvenom bojom su označene stvarne vrijednosti, a plavom predviđene. Vidljivo je da model prati varijacije cijene. Za situaciju kada je selekcijom odabrano pet svojstava, pronađeni su parametri $C = 14000$ i $\gamma = 9.999999999999995e^{-03}$, za koje je prosječna vrijednost postotne mjere pogreške $MAPE = 17.38 \%$. Graf dobiven u ovoj situaciji je prikazan na slici 5.2, $MAPE = 13.94 \%$.



Slika 5.2: Usporedba stvarnih i predviđenih cijena, model sa selekcijom

Posljednji korak je uključivao treniranje modela na optimalnim parametrima na svim podacima za koje postoje stvarne vrijednosti. S obzirom na to da su modeli poprilično slični, ovaj korak je proveden na oba modela. Stvarni podaci su dostupni od 7. 5. do 11. 11. 2014. , uključujući i taj dan. Od modela će se tražiti predviđanje cijene električne energije za tjedan koji slijedi nakon dostupnih podataka za učenje. Radi se o tjednu 12. 11. do 19. 11. 2014. što rezultira sa 192 testna primjera. Za svaki sat u tom tjednu podaci o značajkama dostupni dan unaprijed, znači radi se o predviđanjima, složeni su u format koji model očekuje. Na temelju tih podataka model je predviđao cijene električne energije. Model bez selekcije je dao rezultat $MAPE = 19.08 \%$. Na slici 5.3 vidljiva je usporedba kretanja stvarnih (crvena boja) i predviđenih (plava boja) vrijednosti.



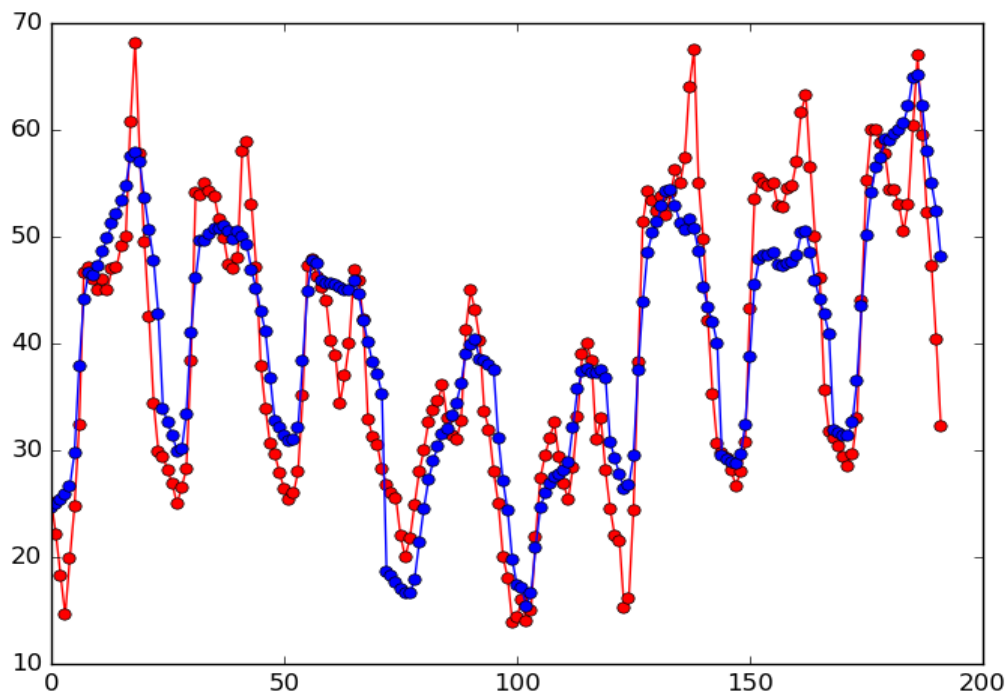
Slika 5.3: Predviđanje cijene, model bez selekcije

Primjećujemo da model radi veliku grešku nešto prije stotog primjera što odgovara danu 15. studenom navečer.

Najbolji rezultat je postignut radom modela sa selekcijom, MAPE = 14.14 % što je prikazano na slici 5.4.

Iz grafa je vidljivo da predviđanja napravljena od strane modela, lijepo prate stvarne vrijednosti. Model regresije potpornim vektorima s parametrima $\gamma = 9.9999999999999995e^{-03}$ i $C = 14000$ naučio je predviđati cijenu električne energije na zadovoljavajućoj razini.

Prilikom obrade značajka, promatrali smo odnos stvarnih i predviđenih vrijednosti značajka. Postavili smo pitanje: hoće li netočnost predviđanja imati utjecaj na rad algoritma? Jedan način kako to možemo provjeriti je da za isti trenutak algoritmu pružimo stvarne podatke i podatke predviđene dan unaprijed i vidimo kako će se algoritam ponašati. Model sa selekcijom, te parametrima $C = 14000$ i $\gamma = 9.9999999999999995e^{-03}$ treniran je na stvarnim podacima do datuma 4. 11. 2014., isključujući taj datum. Zatim su mu dani stvarni podaci o značajkama od datuma 4. 11. 2014. do 11. 11. kako bi predvidio cijenu električne energije. Za isti period, model je predvidio cijenu električne energije na po-

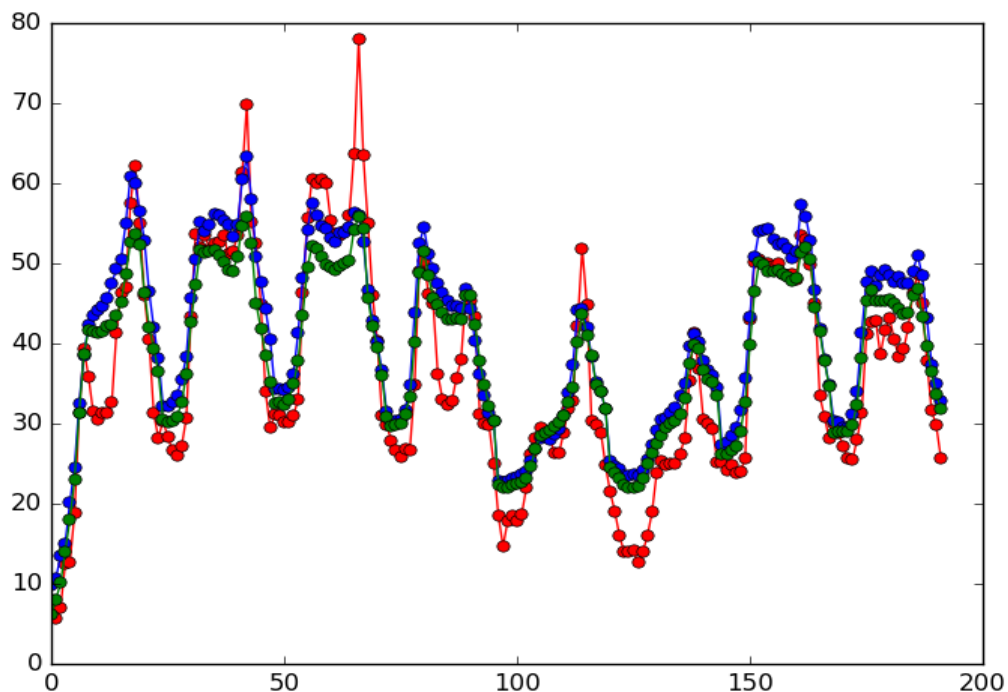


Slika 5.4: Predviđanje cijene, model sa selekcijom

dacima predviđenima dan unaprijed. U prvom slučaju, $MAPE = 16.34 \%$, a u drugom $MAPE = 16.56 \%$. Ovaj rezultat ukazuje da predviđeni podaci o značajkama ne utječu mnogo na rad algoritma.

Na grafu na slici 5.5 crvenom su bojom prikazane stvarne cijene električne energije, plavom cijene predviđene temeljem stvarnih podataka i zelenom cijene predviđene temeljem predviđenih podataka za svaki sat od 4. 11. 2014. do 11. 11. 2014. Vidljivo je da se predviđanja dobivena temeljem stvarnih i predviđenih podataka razlikuju, ali u prosjeku imaju podjednake greške.

Razlog predviđanja cijene električne energije bio je stvaranje strategije trgovanja. Kao što je objašnjeno u drugom poglavlju, izrada strategije ovisi o ulozi pojedinca na tržištu, njegovim mogućnostima i ograničenjima. U svrhu ovog rada, pretpostavit ćemo sljedeću situaciju: broker želi trgovati električnom energijom. Može kupovati i prodavati. Nema ograničenja na količinu kojom trguje, ali na kraju dana ne smije imati neku količinu električne energije u svojem vlasništvu (jer mu ne treba). Također ne smije se dogoditi da proda električnu energiju nekome, a ne isporuči je. Dakle, na kraju dana, ukupna količina



Slika 5.5: Predviđanja dobivena temeljem stvarnih i predviđenih podataka se ne razlikuju mnogo

električne energije koju kupi i proda mora dati nulu. Pretpostavimo da broker ima sljedeću strategiju.

- Kada stvarna cijena bude deset posto manje od predviđene, broker kupuje 50 MW električne energije.
- Kada stvarna cijena bude deset posto veća od predviđene, broker prodaje svu količinu električne energije koju je kupio. Ukoliko u tom trenutku još nije kupio električnu energiju, broker prodaje 50 MW električne energije koju će kasnije kupiti.
- Zadnji sat u danu, broker čini poravnanje. Ukoliko je prodao električnu energiju koju nema, kupuje ju, a ukoliko je kupio električnu energiju koju nije prodao, prodaje ju.

Dodatna pretpostavka u ovom primjeru je da uvijek postoji netko spreman trgovati s brokerom. Pogledajmo kako bi izgledalo ovo trgovanje za dan 12. 11. korištenjem predviđanja sa slike 5.4.

- U jedan sat stvarna vrijednost je 25.14 €/MWh, a predviđena 26.37 €/MWh. Stvarna cijena je 95.33 % predviđene cijene, to jest stvarna cijena je 4.67 % manja od predviđene, pa po strategiji, broker ne radi ništa.
- U dva sata stvarna vrijednost je 22.15 €/MWh, a predviđena 26.18 €/MWh. Stvarna cijena je 84.6 % predviđene, to jest, stvarna cijena je 15.4 % manja od predviđene, pa po strategiji, broker kupuje 50 MW električne energije. Ukupna cijena trgovanja je 1107.5 €, što znači da je novčano stanje brokera –1107.5 € i ima 50 MW električne energije koju kasnije može prodati.
- U tri sata stvarna vrijednost je 18.28 €/MWh, a predviđena 26.58 €/MWh. Stvarna cijena je 68.77 % predviđene cijene, što znači da broker kupuje 50 MW električne energije. Ukupna cijena trgovanja je 914 €, čime je novo novčano stanje brokera –2021.5 € i ima 100 MW električne energije.
- U četiri sata stvarna vrijednost je 14.6 €/MWh, a predviđena 27.11 €/MWh. Stvarna cijena je 53.85 % predviđene cijene, te, po strategiji, broker kupuje 50 MW električne energije. S obzirom na to da je ukupna cijena trgovanja 730 €, novo novčano stanje brokera je –2751.5 € i ima 150 MW električne energije.
- U pet sati stvarna vrijednost je 19.92 €/MWh, a predviđena 28.1 €/MWh. Stvarna cijena je 70.88 % predviđene cijene, to jest stvarna cijena je 29.12 % manja od predviđene, pa broker kupuje 50 MW električne energije. Ukupna cijena trgovanja je 996 €, te je novo novčano stanje brokera –3747.5 €. Trenutno broker ima 200 MW električne energije dostupne za trgovanje.
- U šest sati stvarna vrijednost je 24.73 €/MWh, a predviđena 31.07 €/MWh. Stvarna cijena je 79.59 % predviđene cijene, što znači da broker kupuje 50 MW električne energije. Ukupna cijena trgovanja je 1236.5 €, te je novo novčano stanje brokera –4984 €. Trenutno broker ima 250 MW električne energije dostupne za trgovanje.
- U sedam sati stvarna vrijednost je 32.38 €/MWh, a predviđena 39.34 €/MWh. Stvarna cijena je 82.3 % predviđene cijene, te, po strategiji, broker kupuje 50 MW električne energije. S obzirom na to da je ukupna cijena trgovanja 1619 €, novo novčano stanje brokera je –6603 € i ima 300 MW električne energije.
- U osam sati stvarna vrijednost je 46.68 €/MWh, a predviđena 46.01 €/MWh. Stvarna cijena je 101.45 % predviđene cijene, te, po strategiji, broker ne radi ništa. Ostaje staro novčano stanje –6603 € i 300 MW električne energije.
- Od devet sati pa do osamnaest sati, stvarna vrijednost je unutar deset posto od predviđene vrijednosti što, po strategiji, znači da broker ne radi ništa, te se njegovo novčano stanje i količina električne energije za trgovanje ne mijenjaju.

- U devetnaest sati stvarna vrijednost je 68.13 €/MWh, a predviđena 57.67 €/MWh. Stvarna cijena je 118.13 % predviđene cijene, to jest 18.13 % veća od predviđene vrijednosti što znači da broker prodaje svu količinu električne energije koju ima na raspolaganju, znači 300 MW. Ukupna vrijednost trgovanja je 20439 € što znači da je novo novčano stanje brokera 13836 €. Više nema električne energije.
- U dvadeset sati stvarna vrijednost je 57.81 €/MWh, a predviđena 50.83 €/MWh. Stvarna cijena je 113.73 % predviđene cijene, što znači da broker prodaje. S obzirom na to da nema električne energije na raspolaganju, prodaje samo 50 MW. Ukupna vrijednost trgovanja je 2890.5 €, pa je novo novčano stanje brokera 16726.5 €. Sada mu nedostaje 50 MW električne energije koju će morati kupiti do kraja dana.
- U dvadeset jedan sat stvarna vrijednost je 49.51 €/MWh, a predviđena 51.39 €/MWh, to jest stvarna cijena je 96.34 % predviđene cijene, što, po strategiji, znači da broker ne radi ništa.
- U dvadeset dva sata stvarna vrijednost je 42.5 €/MWh, a predviđena 47 €/MWh, to jest stvarna cijena je 90.42 % predviđene cijene, što, po strategiji, znači da broker ne radi ništa.
- U dvadeset tri sata stvarna vrijednost je 34.39 €/MWh, a predviđena 43.22 €/MWh, što znači da je stvarna vrijednost 79.56 % predviđene i broker kupuje 50 MW. Vrijednost trgovanja je 1719.5 € čime je novo novčano stanje brokera 15007 € i nema električne energije za prodaju niti je dužan kupiti ju.
- U dvadeset četiri sata, po strategiji broker treba kupiti ili prodati električnu energiju, ovisno o tome u kojem je stanju. S obzirom na to da broker u ovom primjeru nema kupljene, a ne prodane električne energije ili prodane, a ne kupljene, nema potrebe za poravnavanjem. Na kraju dana brokerova dobit je 15007 €.

Na primjeru ove strategije, vidljivo je da rezultati predviđanja cijene električne energije mogu poslužiti u planiranju i optimizaciji trgovanja električnom energijom.

Poglavlje 6

Zaključak

Za potrebe ovog diplomskog rada, razrađen je primjer predviđanja cijene električne energije u Njemačkoj korištenjem metode potpornih vektora. Motivacija iza ovog rada bila je pokazati mogućnosti strojnog učenja općenito, a posebno metode potpornih vektora u primjeni u stvarnom, svakodnevnom životu. Podaci u radu dobiveni su od tvrtke Petrol d.o.o., koja ih koristi u svakodnevnom poslovanju i temeljem njih vlastitim modelima predviđa cijenu električne energije, pokušavajući time poboljšati dosadašnje poslovanje. Korišteni su podaci o cijeni električne energije, temperaturi zraka, proizvodnji električne energije iz sunčevih i fotonaponskih elektrana, vjetroelektrana, konvencionalnih elektrana iznad 100 MW instalirane snage, hidroelektrana, te energiji padalina.

Analizom dostupnih podataka, primijećeni su određeni obrasci. Vrijednosti svih podataka veoma variraju, te su zabilježeni mnogi skokovi. Kod cijene električne energije vidljiv je obrazac pada vrijednosti u ljetnim mjesecima, vikendima, praznicima, te noću. Slični su obrasci primijećeni i kod temperature koja ima pad vrijednosti u zimskim mjesecima, te proizvodnji iz konvencionalnih izvora koja pokazuje iste obrasce kao cijena. Proizvodnja električne energije iz sunčevih i fotonaponskih elektrana prestaje noću, raste do podneva, te zatim pada. Kod cijene električne energije prisutan je i poseban fenomen, postizanje negativnih vrijednosti što nije očekivano s obzirom na to da negativne vrijednosti cijene nisu pojava u svakodnevnom životu. Proizvodnja iz hidroelektrana pokazuje posljedičnu vezu s energijom padalina. Predviđanja za sve značajke dosta dobro prate stvarne vrijednosti. Netočnost se povećava kako se povećava razlika između trenutka kada je predviđanje napravljeno i trenutka za koje je napravljeno.

Vremenski raspon značajki se razlikuje pa je u stvaranju modela korišten najveći raspon u kojem su podaci za sve značajke dostupni na satnoj razini od 7. 5. do 11. 11. Za potrebe validacije modela korišteni su predviđeni podaci o značajkama za tjedan od 12. 11. do 19. 11.

Izrađena su dva modela korištenjem metode potpornih vektora za regresiju: prvi koriš-

tenjem svih značajki, te drugi korištenjem pet značajki s najvećom korelacijom s ciljnom vrijednošću, u ovom slučaju, cijenom električne energije. Regresija je korištena s obzirom na to da je cijena, po svojoj prirodi, funkcijska vrijednost. Oba modela su pokazala zadovoljavajuće performanse, s rezultatima postotne mjere pogreške 19.08 % za model sa svim značajkama i 14.14 % za model sa selekcijom na tjednu za validaciju. Na temelju rezultata predviđanja modela sa selekcijom, uspješno je izrađena strategija trgovanja koja rezultira zaradom. Sve ovo ukazuje da prikazan pristup može biti korišten u trgovanju električnom energijom.

Prostora za unaprjeđenje modela ima. Za početak, dublja analiza tržišta električnom energijom mogla bi rezultirati otkrivanjem novih važnih podataka koji bi se koristili u izradi modela. Također, detaljnija analiza značajki i njihovih međusobnih veza ukazala bi na odabir koje značajke koristiti prilikom treniranja modela, a koje odbaciti. U ovom radu korišten je *kernel* RBF, ali mogli bi se isprobati i drugi *kernel*-i, te se usporediti dobiveni rezultati. Skaliranje je još jedan korak u kojem se mogu isprobati druge mogućnosti, već spomenuto skaliranje na intervale $[0, 1]$ i $[-1, 1]$, ali i druga skaliranja. Ne može se unaprijed znati koliko je značajki najbolje odabrati. U ovom radu su napravljena dva modela, na svim značajkama i na njih pola, ali svaki drugi broj značajki također može biti isproban, te rezultati takvog modela uspoređeni s rezultatima ostalih modela. Naposljetku, ostaje problem optimalnih parametara. S obzirom na to da je nemoguće pretražiti cijeli \mathbb{R} , ovdje uvijek ostaje mogućnost unaprjeđenja. Proširenje skupa iz kojih se odabiru optimalni parametri, ima potencijal unaprjeđenja modela. Također, ranije je spomenuto da je velika vrijednost R^2 mjere rezultat prevelike prilagodbe modela. Nije lako odrediti što je velika vrijednost R^2 . U ovom radu su modeli, koji su služili kao početna točka za ručno pretraživanje parametara, imali $R^2 = 0.6735$ i $R^2 = 0.6201$. Ostaje pitanje bi li model s manjom vrijednošću R^2 , na primjer s vrijednošću iz intervala $[0.5, 0.6]$, dao bolje rezultate. Ovo je još jedna mogućnost unaprjeđenja modela.

Bibliografija

- [1] *F regression*, http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.f_regression.html.
- [2] *Preprocessing data*, <http://scikit-learn.org/stable/modules/preprocessing.html>.
- [3] *RBF SVM parameters*, http://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html.
- [4] *SelectKBest*, http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html.
- [5] U. Abhisek, *Intelligent Systems and Signal Processing in Power Engineering*, Springer, 2007.
- [6] Nastavnici asistenti i studenti energetike FSB-a, *Energetska tržišta*, http://powerlab.fsb.hr/enerpedia/index.php?title=ENERGETSKA_TR%C5%BDI%C5%A0TA, Enerpedia.
- [7] B. Dalbelo Bašić i J. Šnajder, *Uvod u strojno učenje*, https://www.fer.unizg.hr/_download/repository/SU-1-Uvod%5B1%5D.pdf, Bilješke sa predavanja kolegija Strojno učenje, FER.
- [8] P. Joshi, *Support Vector Machines*, <http://prateekvjoshi.com/2012/08/24/support-vector-machines/>.
- [9] S. Krajčar i Ž. Tomšić, *Ekonomija u energetici*, https://www.fer.unizg.hr/_download/repository/5_-_organizacija_trzista_biljeske.pdf, Bilješke sa predavanja kolegija Organizacija tržišta, FER.
- [10] N. Mandić, *Analiza pojave negativne cijene na spot tržištu električne energije europske burze EEX*, *Energija* **58** (2009), br. 2, 136–165.

- [11] M. Matijaš, *Predviđanje potrošnje električne energije regresijom zasnovanom na potpornim vektorima*, https://www.fer.unizg.hr/_download/repository/KDI_mmatijas.pdf.
- [12] T. Šmuc, *Strojno učenje uvod*, <https://web.math.pmf.unizg.hr/nastava/su/materijali/>, Bilješke sa predavanja kolegija Strojno učenje, PMF.
- [13] E. Sotiroulos, *Modeling of German Electricity Load for Pricing of Forward Contracts*, Disertacija, Swiss Federal Institute of Tehnology, Zurich, 2012.

Poglavlje 7

Prilozi

7.1 Kod za unos i obradu podataka

```
import pandas as pd
import numpy as np
from sklearn import svm, grid_search,
cross_validation, preprocessing
from sklearn.feature_selection import SelectKBest, f_regression

#unos podataka
df = pd.read_excel('Podaci.xlsx')
data = df.as_matrix()

#odvajanje cijene u poseban stupac i odvajanje znacajki
# u posebnu matricu
Y = data[:,10]
X = data[:, [0,1,2,3,4,5,6,7,8,9] ]

#standardno skaliranje
scalerX = preprocessing.StandardScaler()
scalerY = preprocessing.StandardScaler()
X = scalerX.fit_transform(X)
Y = scalerY.fit_transform(Y)

#selekcija znacajki
selector = SelectKBest(f_regression, k=5)
selector.fit(X,Y)
```

```
X = selector.transform(X)
```

7.2 Kod za pronalazak optimalnih parametara

```
#parametri za testirati
C_range = 10.0 ** np.arange(-2, 6)
gamma_range = 10.0 ** np.arange(-3, 3)
param_grid = {'C':C_range, 'gamma': gamma_range}

#svr
svr = svm.SVR()

#odabir optimalnih parametara funkcijom GridSearchCV
grid = grid_search.GridSearchCV(svr, param_grid, scoring = 'r2')

grid.fit(X,Y)

grid.best_estimator_
grid.best_score_

#rucno trazenje optimalnih parametara

#definicija MAPE
def mean_absolute_percentage_error(y_true, y_pred):
    sum = 0
    for i in range(0, len(y_true)):
        sum = sum + np.abs((y_pred[i]-y_true[i])/y_true[i])
    result = sum / len(y_true)
    return result

#razdvajanje podataka na set za treniranje i set za testiranje
X_train, X_test, Y_train, Y_test =
cross_validation.train_test_split(X, Y, test_size=0.1)

#stvaranje modela sa zadanim parametrima
svr = svm.SVR(C=14000.0, cache_size=200, coef0=0.0, degree=3,
epsilon=0.1, gamma=9.9999999999999995e-03, kernel='rbf',
```

```
max_iter=-1, shrinking=True, tol=0.001, verbose=False)

#treniranje modela
svr = svr.fit(X_train, Y_train)

#model predvida
Y_pred = svr.predict(X_test)

#validacija modela
Y_test = scalerY.inverse_transform(Y_test)
Y_pred = scalerY.inverse_transform(Y_pred)
mean_absolute_percentage_error(Y_test, Y_pred)
```

7.3 Kod za testiranje modela

```
#unos podataka za predviđanje
dfp = pd.read_excel('Predviđanje.xlsx')
dataPred = dfp.as_matrix()

#odvajanje cijene u poseban stupac i odvajanje značajki
# u posebnu matricu
YPred = dataPred[:,10]
XPred = dataPred[:, [0,1,2,3,4,5,6,7,8,9] ]

#skaliranje prije definiranim skaliranjem
XPred = scalerX.transform(XPred)
YPred = scalerY.transform(YPred)

#selekcija prije definiranim selektorom
XPred = selector.transform(XPred)

#prije definiran model predvida
YSVR = svr.predict(XPred)

#validacija modela
YSVR = scalerY.inverse_transform(YSVR)
```



```
YPred = scalerY.inverse_transform(YPred)
mean_absolute_percentage_error(YPred, YSVR)
```

Sažetak

”Strojno učenje jest programiranje računala na način da optimiziraju neki kriterij uspješnosti temeljem podatkovnih primjera ili prethodnog iskustva” [7]. Najvažniji teorem strojnog učenja je PAC teorem koji kaže: ukoliko model radi dovoljno dobro na većini primjera, onda je vjerojatno dovoljno dobar. Model strojnog učenja trenira se na primjerima za trening, na kojima se pokušava postići optimalna prilagodba. Žele se izbjeći premala prilagodba (kada model ne radi točno ni na primjerima za treniranje) i prevelika prilagodba (kada model savršeno radi na primjerima za testiranje, ali na novim primjerima griješi).

Metoda potpornih vektora, koju je šezdesetih godina prošlog stoljeća stvorio Vladimir Vapnik, nije se previše koristila sve do 1992. godine, kada su Isabelle Guyon, Bernhard Boser i Vladimir Vapnik uveli nelinearni klasifikator koristeći tzv. *kernel* trik. Verziju za regresiju predložili su 1997. godine Vladimir Vapnik, Harris Drucker, Chris Burges, Linda Kaufman i Alex Smola. Metoda potpornih vektora temelji se na teoriji statističkog učenja koja nastoji minimizirati rizik i kompleksnost modela. U metodi potpornih vektora spomenuto se pokušava postići pronalaskom optimalne razdvajajuće hiperravnine (one s najmanjom udaljenošću između hiperravnine i njoj najbližih primjera, potpornih vektora). U situaciji kada problem nije razdvojiv ili je vrijeme traženja optimalne hiperravnine predugo, točke se mapiraju u prostor više dimenzije od originalne što se naziva *kernel* trik, a preslikavanje *kernel*.

Burza je organizirano i centralizirano mjesto trgovanja gdje prodavatelji i kupci anonimno predaju svoje ponude za prodaju, odnosno kupnju. Iz svih primljenih ponuda za prodaju i kupnju, formiraju se krivulje ponude i potražnje. Tržišna cijena i količina kojom se trguje određuju se presjekom tih dviju krivulja. Strategija trgovanja električnom energijom započinje pronalaskom modela koji što preciznije predviđa cijene električne energije i ovisi o ulozi na tržištu.

U ovom radu obrađen je primjer izrade strategije trgovanja električnom energijom u Njemačkoj korištenjem metode potpornih vektora. Pritom korišteni podaci sastoje se od sljedećih značajki: satna temperatura, satna proizvodnja električne energije iz sunčevih i fotonaponskih elektrana, satna proizvodnja iz vjetroelektrana, satna proizvodnja iz konvencionalnih elektrana iznad 100 MW instalirane snage, dnevna proizvodnja električne energije iz hidroelektrana i satna energija padalina, te ciljne vrijednosti: satna cijena električne

energije. Svi podaci pokazuju varijacije s više ili manje izraženim skokovima. Obrasci na godišnjoj, tjednoj i dnevnoj razini primijećeni su kod cijene električne energije, temperature, proizvodnje iz sunčevih i fotonaponskih elektrana, vjetroelektrana, te konvencionalnih elektrana. Također je primijećena veza između proizvodnje električne energije iz hidroelektrana i energije padalina.

Priprema podataka i rad algoritma proveden je funkcijama i klasama programskog jezika Python. Podaci od 7. 5. do 11. 11. 2014. godine korišteni su za treniranje modela, dok je model validiran na podacima od 12. 11. do 19. 11. 2014. Najbolji rezultat postignut je modelom sa selekcijom i parametrima $C = 1400$ i $\gamma = 9.9999999999999995e^{-03}$. Kod tog modela, postotna mjera pogreške na testnom tjednu je $MAPE = 14.14 \%$. Na temelju predviđenih podataka modela, moguće je stvoriti strategiju koja će poboljšati trgovanje. Primjer jedne strategije pokazan je u radu na predviđanjima modela za dan 12. 11. 2014.

Summary

Machine learning is the theory of programming computers so that they optimize certain success criteria regarding data examples or previous experience [7]. The most important theorem for machine learning is the PAC theorem that says: if a model does well enough on most examples, it is probably good enough. Machine learning model is trained on training examples, in which it is tried, to achieve optimal adjustment. An attempt is made to avoid underfitting, (when the model does poorly on training examples), and overfitting, (when the model does perfectly on training examples, but makes mistakes on new examples).

Support Vectors Machine was created by Vladimir Vapnik in 1960's, but was not used until 1992, when Isabelle Guyon, Bernhard Boser and Vladimir Vapnik introduced nonlinear classifier using the *kernel* trick. Regression version was proposed in 1997, by Vladimir Vapnik, Harris Drucker, Chris Burges, Linda Kaufman and Alex Smola. Statistical learning theory seeks to minimize the risk and complexity of the model. Support vectors machine is based on statistical learning theory. In support vectors machine, the minimization of the risk and complexity of the model is aimed by finding optimal separating hyper plane. Optimal hyper plane is the one with smallest distance between hyper plane and support vectors. In situations when data is not separable or time to discover optimal hyper plane is too long, *kernel* trick is used: points are mapped to higher dimensional space than original.

The marketplace is an organized and centralized place of trading, where sellers and buyers present their offers. Lines of supply and demand are created considering all received offers. Market price and quantity for trading, are determined upon intersection of the supply and demand line. Trading strategy begins by seeking a model with the most precision in predicting market price of electric energy, and it differs regarding trader role on the market.

This thesis elaborates the example of the strategy development of electricity trading in Germany by using the support vectors machine. Data used are: temperature, production of electric energy from solar power plants, production of electric energy from wind power plants, production of electric energy from conventional power plants over 100 MW output power, production of electric energy from hydro power plants, energy of precipitation, and price of electric energy measured in hourly time intervals. All data have variations with

extreme jumps between values. Annually, weekly and daily patterns are shown in the price of electric energy and temperature. Similar patterns are seen in the production of electric energy from solar power plants, wind power plants, and from conventional power plants. There is a distinct connection between the production of electric energy from hydro power plants, and the energy of precipitation.

Functions and classes of programming language Python were used for data preparation and algorithm implementation. Data from 7. 5. to 11. 11. 2014 were used for training of model, and data from 12. 11. to 19. 11. 2014 were used for model validation. Model with selection, $C = 1400$ and $\gamma = 9.999999999999995e^{-03}$ gave the best results with mean absolute percentage error 14.14 %. It is possible to create a strategy of electricity trading based on data produces by described model. Example of one such strategy is given in this thesis for 12. 11. 2014.

Životopis

Moje ime je Lea Balaško. Rođena sam 14. 9. 1991. godine u Zaboku. Najstarije sam od troje djece Štefice i Branka Balaška. Osnovnoškolsko obrazovanje primila sam u Osnovnoj školi Sveti Križ Začretje. Završetkom osnovne škole 2006. godine, upisala sam matematičko–prirodoslovni smjer Gimnazije Antuna Gustava Matoša u Zaboku. Maturirala sam 2010. godine, nakon čega sam upisala Preddiplomski sveučilišni studij Matematike na Prirodoslovno–matematičkom fakultetu Sveučilišta u Zagrebu. Preddiplomski studij završila sam 2013. godine, te nastavila obrazovanje na istom fakultetu upisavši Diplomski sveučilišni studij Računarstvo i matematika.