

Određivanje obrazaca ponašanja potrošnje pametnih kućanskih uređaja

Cvetković, Matea

Undergraduate thesis / Završni rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Electrical Engineering and Computing / Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:168:354345>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2025-03-20**



Repository / Repozitorij:

[FER Repository - University of Zagreb Faculty of Electrical Engineering and Computing repository](#)



SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1623

**ODREĐIVANJE OBRAZACA PONAŠANJA POTROŠNJE
PAMETNIH KUĆANSKIH UREĐAJA**

Matea Cvetković

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1623

**ODREĐIVANJE OBRAZACA PONAŠANJA POTROŠNJE
PAMETNIH KUĆANSKIH UREĐAJA**

Matea Cvetković

Zagreb, lipanj 2024.

ZAVRŠNI ZADATAK br. 1623

Pristupnica: **Matea Cvetković (0036540266)**
Studij: Elektrotehnika i informacijska tehnologija i Računarstvo
Modul: Računarstvo
Mentor: prof. dr. sc. Hrvoje Pandžić

Zadatak: **Određivanje obrazaca ponašanja potrošnje pametnih kućanskih uređaja**

Opis zadatka:

U svrhu razvoja modela za odziv potrošnje, analiziraju se podatci o potrošnji kućanskih uređaja prikupljeni u laboratoriju za odziv potrošnje na Fakultetu elektrotehnike i računarstva (DRLab). Fokus analize usmjeren je na podatke o potrošnji termostatski upravljivih uređaja, s posebnim naglaskom na identifikaciju odstupanja, praznina u podacima i slično. Cilj istraživanja je odrediti karakteristične tjedne obrasce potrošnje takvih uređaja kako bi se olakšala njihova daljnja primjena u kompleksnim modelima za optimalno upravljanje uređajima.

Rok za predaju rada: 14. lipnja 2024.

Sadržaj

1. Uvod	4
2. Odziv potrošnje	6
3. Strojno učenje	7
3.1. Priprema i analiza podataka	7
3.1.1. Priprema podataka	7
3.1.2. Analiza vremenskih serija	7
3.2. Ulazni podaci	9
3.3. Grupiranje	9
3.3.1. Algoritam K-sredina	10
3.3.2. Dinamičko savijanje vremena	11
4. Podaci	12
4.1. Priprema podataka	12
4.1.1. Određivanje intervala s nedostajućim podacima	12
4.1.2. Popunjavanje intervala s nedostajućim podacima	14
4.1.3. Određivanje stršecih vrijednosti	16
4.1.4. Poduzorkovanje	17
4.2. Analiza vremenske serije	17
4.2.1. Stacionarnost	19
4.2.2. Sezonalnost	20
4.2.3. Autokorelacija	21
4.2.4. Korelacija između značajki	21
5. Reprezentativni tjedni i dani potrošnje	24

5.1. Reprezentativni tjedni	24
5.1.1. Grupiranje euklidskom distancom	24
5.1.2. Grupiranje u 2 grupe dinamičkim savijanjem vremena	26
5.1.3. Grupiranje u 3 grupe dinamičkim savijanjem vremena	30
5.2. Reprezentativni dani	32
6. Zaključak	36
Literatura	37
Sažetak	38
Abstract	39

1. Uvod

S povećanjem potrošnje električne energije, kao i novim, sve zastupljenijim obnovljivim izvorima, mijenja se tradicionalan način funkcioniranja elektroenergetskog sustava. Upravljanje uređajima u kućanstvu jedna je od važnijih komponenti za održavanje stabilnosti modernog sustava. S obzirom na to da opterećenje mreže diktira cijenu električne energije, sudjelovanjem u sustavu odziva potrošnje, korisnici mogu iskoristiti niže cijene energije u razdobljima nižeg opterećenja, a istovremeno pružati uslugu elektroenergetskom sustavu. Krajnji korisnici, pomičući vlastitu potrošnju u stvarnom vremenu u ovisnosti o trenutnom opterećenju mreže, odnosno cijeni, pomažu u ostvarivanju jednakosti proizvodnje i potrošnje u svakom vremenskom trenutku. Da bi se korisnike potaknulo na sudjelovanje u ovakvom sustavu, potrebno je pružati uslugu odziva potrošnje tako da kućanstvo ostvari čim veću financijsku dobit, uz minimalno žrtvovanje vlastitih želja i navika. Iz ovog je razloga potrebno razviti inteligentan sustav koji može istovremeno pratiti stanje u mreži i po potrebi upravljati skupinom uređaja. Iako zbog prevelikog utjecaja na uobičajeno ponašanje korisnika nisu svi uređaji u kućanstvu pogodni za automatsko upravljanje (poput paljenja i gašenja svjetla i sl.), velik dio njih, poput grijanja, klime, električnih grijalica vode [1] i drugih, jest. Za upravljanje takvim uređajima potrebno je omogućiti predviđanje potrošnje kućanstva, kao i cijena na tržištima energije i rezervi. Uz to, treba razviti optimizacijske modele koji će odrediti optimalan način rada pojedinog uređaja u kućanstvu s obzirom na spomenuta predviđanja.

Predviđanje potrošnje podrazumijeva mjerenje, pohranu i obradu podataka o potrošnji uređaja unutar kućanstva, a modeli predviđanja cijena obradu podataka o povijesnim cijenama na tržištu, meteorološkim uvjetima i geopolitičkoj situaciji. Pri obradi i predviđanju koriste se statističke metode i algoritmi strojnog učenja.

U ovom radu obrađeni su podaci prikupljeni u Laboratoriju za odziv potrošnje (De-

mand Response Lab - DRLab) na Fakultetu elektrotehnike i računarstva (FER-u) te su, korištenjem algoritama nenadziranog strojnog učenja, određeni reprezentativni obrasci tjedne i dnevne potrošnje električne grijalice pogodni za precizniju implementaciju automatskog upravljanja uređajem.

U drugom poglavlju rada predstavljen je pojam odziva potrošnje te je opisan DRLab. U trećem poglavlju opisan je postupak analize vremenskih serija te osnovni koncepti strojnog učenja. Detaljan prikaz prikupljenih podataka, opis pripreme podataka te njihova analiza dani su u četvrtom poglavlju, dok su u petom prikazane primjene algoritama strojnog učenja i dobiveni rezultati.

2. Odziv potrošnje

Osnovni je preduvjet kvalitetnog elektroenergetskog sustava jednakost proizvodnje i potrošnje električne energije. Porastom udjela obnovljivih proizvođača energije kojima je teško upravljati dolazi do potrebe za fleksibilnim upravljanjem potrošnjom sa strane potrošača kako bi se ta jednakost održala. Odziv potrošnje [2] predstavlja sposobnost potrošača da svoje potrošačke zahtjeve prilagodi trenutačnim potrebama i mogućnostima elektroenergetskog sustava. Iako se odziv potrošnje već primjenjuje kod industrijskih korisnika, značajno je manje prisutan u kućanstvima. Kućanski uređaji sami po sebi nisu veliki potrošači, ali uzme li se u obzir ukupna potrošnja svih uređaja u jednom kućanstvu, i zatim više takvih kućanstava, njihov značaj raste. Kućanstva uključena u odziv potrošnje pomicanjem svoje potrošnje u vremenu pružat će uslugu elektroenergetskom sustavu, a istovremeno će imati smanjene financijske troškove.

U svrhu razvoja sustava odziva potrošnje kućanstva, na Fakultetu elektrotehnike i računarstva dizajniran je Laboratorij za odziv potrošnje, DRLab (Demand Response Lab). Ovaj laboratorij predstavlja jedno aktivno kućanstvo, a sastoji se od velikog broja pametnih električnih kućanskih uređaja, čiji se podaci o potrošnji prikupljaju i obrađuju. U laboratoriju se nalaze sljedeći uređaji: hladnjak, električna grijalica vode, električna pećnica, aparat za kavu, mikrovalna pećnica, perilica za rublje, sušilica rublja, perilica posuđa, napa, podno grijanje, rolete, drobilica za led, štednjak, klima, rasvjeta te utičnice.

Za integraciju modela odziva potrošnje u stvarna kućanstva potrebno je prikupiti podatke i razviti kvalitetan model predviđanja potrošnje kućanstava kao i tržišnih cijena, te razviti alat koji automatizirano upravlja kućanstvom na način optimalan i za potrošača i za sustav. Razvojem ovakvih alata mogućnost sudjelovanja u odzivu potrošnje postat će dostupna širem broju kućanstava.

3. Strojno učenje

Strojno učenje je grana umjetne inteligencije koja se bavi razvojem statističkih algoritama koji na temelju poznatih podataka sakupljaju znanje te ga primjenjuju na nove podatke. Strojno učenje dijeli se na nadzirano i nenadzirano: kod nadziranog učenja podaci dolaze u parovima i potrebno je pronaći neku funkciju zavisnost među njima, dok se kod nenadziranog učenja u skupu podataka traže pravilnosti.

Postoji nekoliko koraka u primjeni algoritama strojnog učenja, od kojih su najbitniji početna priprema i obrada podataka, odabir značajki, smanjenje dimenzionalnosti te odabir, učenje i vrednovanje modela [3].

3.1. Priprema i analiza podataka

3.1.1. Priprema podataka

Prije predaje podataka algoritmu strojnog učenja, podatke je potrebno pripremiti. Podaci prikupljeni u stvarnom svijetu nisu savršeni već sadrže razne pogreške: stršeće vrijednosti koje jako odstupaju od ostalih te nedostajuće podatke odnosno rupe u podacima. Da bi se podaci mogli koristiti, potrebno je izolirati i izbaciti stršeće vrijednosti te na neki način popuniti intervale u kojima nedostaju podaci.

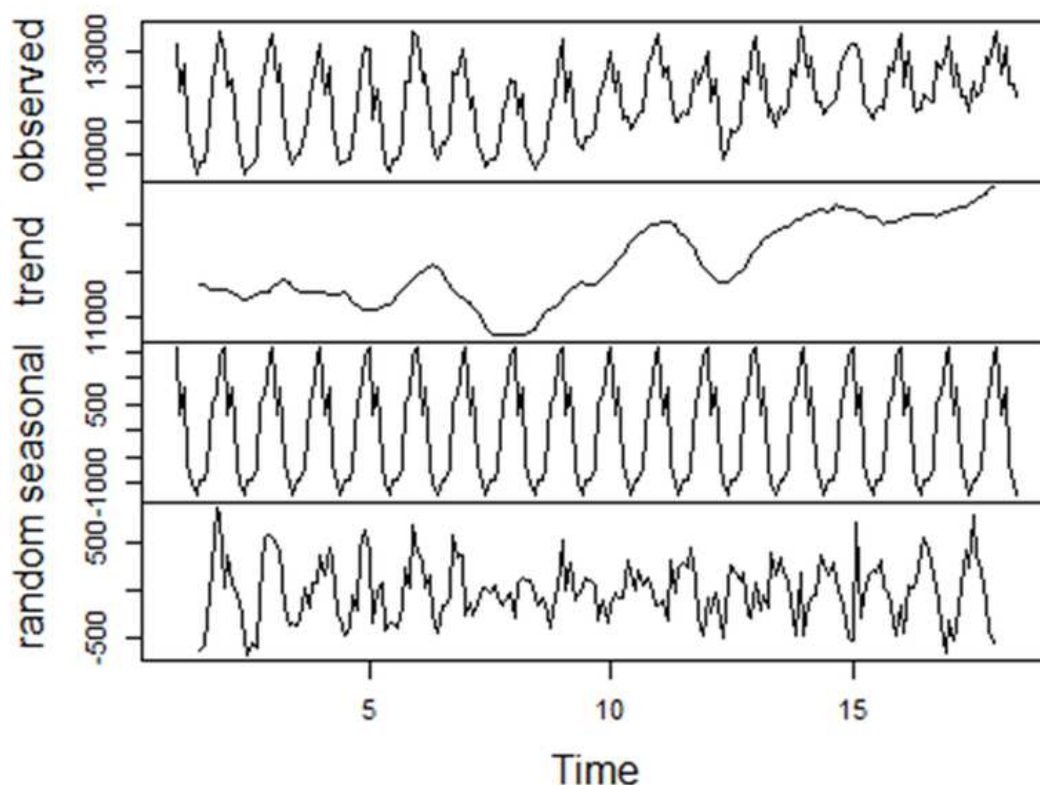
3.1.2. Analiza vremenskih serija

Nakon pripreme podataka, slijedi analiza. U slučaju potrošnje, riječ je o podacima u vremenu, odnosno vremenskoj seriji. Vremenska serija je skup poredanih vrijednosti uzorkovanih s konstantnim periodom. Postoji nekoliko glavnih svojstava vremenskih serija koja se detaljno promatraju u analizi: stacionarnost, sezonalnost i autokorelacija. Vremenske serije mogu se dekompozicijom rastaviti na tri komponente: trend, sezo-

nalnu komponentu i šum (slika 3.1.).

Stacionarnost je svojstvo vremenske serije da su njezina srednja vrijednost i varijanca konstantne u vremenu. Stacionarnost se testira raznim statističkim metodama, od kojih je najpoznatija prošireni Dickey-Fuller (ADF) test [4]. Osim ADF testom, stacionarnost serije vidi se i pri njezinoj dekompoziciji, u komponenti trenda. Trend je komponenta koja prikazuje dugoročnu promjenu srednje vrijednosti serije u vremenu (slika 3.1.). Ako takve promjene nema, ili je vrlo malena, serija je stacionarna.

Sezonalnost je svojstvo vremenske serije da ima uzorak koji se ponavlja u pravilnim intervalima. Ova je karakteristika u dekompoziciji serije prikazana sezonalnom komponentom (slika 3.1.). Sezonalna komponenta prikazuje, kao što je već rečeno, uzorak u vremenu. Sezonalnost i stacionarnost ponekad je teško uočiti zbog šuma. Šum je komponenta vremenske serije koja ne podliježe pravilnosti, već je nasumična i otežava uočavanje pravilnosti u podacima.



Slika 3.1. Dekompozicija vremenske serije [5]

Korelacija između dvije varijable opisuje međusobnu ovisnost njihovih vrijednosti.[6]

Linearna korelacija izražava se Pearsonovim koeficijentom, koji poprima vrijednosti između -1 i 1, opisujući pritom negativnu ili pozitivnu korelaciju. Ako je Pearsonov koeficijent 0, varijable nisu korelirane, a poprimi li vrijednosti između 0 i 1 ili 0 i -1, iskazuje djelomičnu koreliranost.

Autokorelacija je mjera korelacije između trenutne vrijednosti varijable i njezinih prošlih vrijednosti. Kada su ovakve korelacije prisutne, ukazuju na to da prošle vrijednosti varijable utječu na trenutnu vrijednost. Autokorelacija narušava pretpostavku brojnih algoritama strojnog učenja o međusobnoj neovisnosti vrijednosti varijable u vremenu.

3.2. Ulazni podaci

Ulaz algoritama strojnog učenja su primjeri (engl. *sample*). Primjer je jedan podatak, a se sastoji od nekoliko značajki (engl. *feature*). Značajke su atributi primjera koji se u konkretnom problemu smatraju bitnima. Primjer može imati mnogo značajki, međusobno zavisnih ili nezavisnih. Visoko korelirane značajke mogu se izbaciti iz ulaznog skupa zato što usporavaju algoritam, a same po sebi ne nose nikakve dodatne informacije.

Skup svih primjera koji se predaju algoritmu je ulazni prostor. Ulazni prostor je n -dimenzijski vektorski prostor, u kojem svaki vektor odgovara jednom primjeru. Dimenzija n tog prostora odgovara broju značajki, što u realnom svijetu može biti i do nekoliko stotina ili tisuća značajki. Kako bi se osigurala optimalnost algoritama i kvaliteta rezultata, kod velikog broja značajki potrebno je smanjiti dimenzionalnost ulaznog prostora, odnosno broj značajki, bez da se izgube vrijedne informacije o podacima. Tako se iz ulaznog prostora brišu redundantne značajke, značajke koje su međusobno visoko korelirane, te one koje su beznačajne.

3.3. Grupiranje

Jedan od zadataka nenadziranog strojnog učenja jest grupiranje (engl. *clustering*), što je postupak podjele primjera u grupe ovisno o njihovoj međusobnoj sličnosti. Postoje particijsko grupiranje, pri kojem su grupe međusobno disjunktne, te hijerarhijsko grupi-

ranje, koje dopušta da grupe budu podijeljene na podgrupe. Razlikujemo i tvrdo i meko grupiranje, gdje prvo dopušta da primjer pripada isključivo jednoj grupi, a potomje dopušta pripadnost primjera većem broju grupa.

3.3.1. Algoritam K-sredina

Jedan od algoritama koji se koriste za grupiranje je algoritam K-sredina (engl. *K-means*). Algoritam K-sredina je primjer partijskog algoritma koji dijeli primjere u tvrde, disjunktne grupe. U ovom algoritmu svaka je grupa predstavljena svojim centroidom (srednjom vrijednošću), a svaki je primjer smješten u grupu čijem je centroidu najbliži. Udaljenost primjera od centroida računa se kao euklidska udaljenost. Za neki n-dimenzionalan prostor, euklidska udaljenost dvije točke $T_1(p_1, p_2, \dots, p_n)$ i $T_2(q_1, q_2, \dots, q_n)$ iskazana je kao:

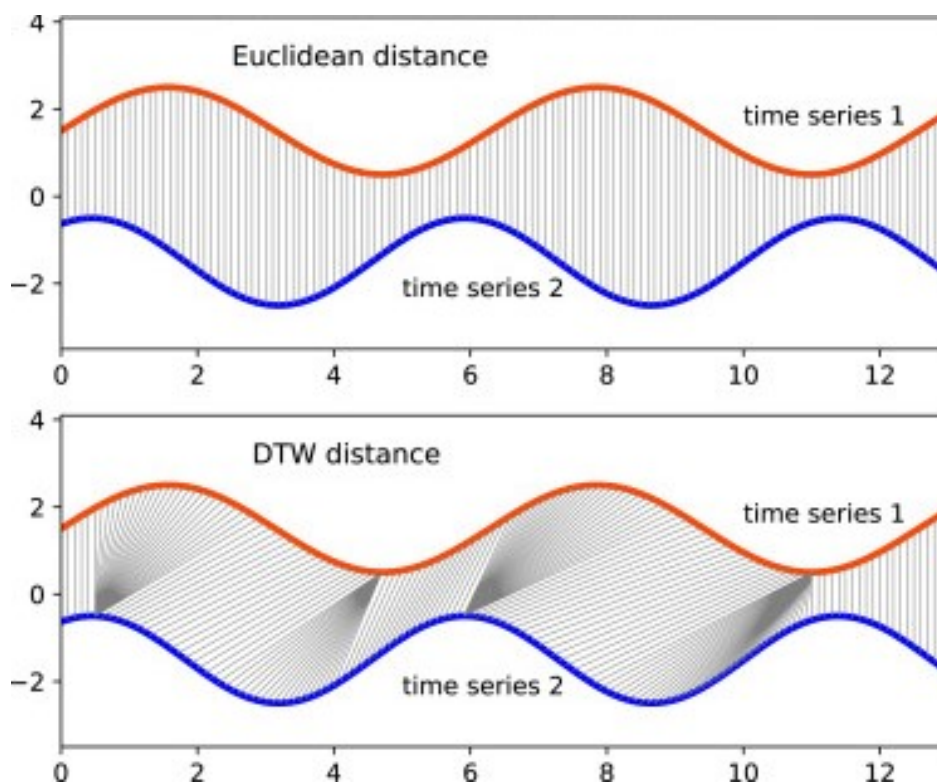
$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}. \quad (3.1)$$

Broj grupa u koji se primjeri dijele zadaje se unaprijed. U početnom koraku algoritam slučajno odabire K centroida, kojima pridjeljuje odgovarajuće primjere. To rezultira promjenom vrijednosti centroida, prema kojoj se zatim primjeri ponovno svrstavaju u grupe. Postupak se izvršava određeni broj iteracija. S obzirom na to da se početni centriodi biraju nasumično, algoritam ne daje jednake grupe pri svakom pokretanju. Zato ga se pokreće više puta za isti broj grupa i iste primjere, te se odabire najtočnije (i najsmislenije) grupiranje.

S obzirom na to da često nije unaprijed poznat broj grupa na koje se primjeri trebaju podijeliti, potrebne su metode za odabir optimalnog broja grupa. Algoritam K-sredina, osim što svrstava primjere u grupe i računa predstavnike grupa, računa i pogrešku grupiranja J , koja je definirana kao

$$J = \sum_{k=1}^K \sum_{i=1}^N b_k^{(i)} \|x^{(i)} - \mu_k\|^2 \quad (3.2)$$

gdje je K broj grupa, N broj primjera, $b_k^{(i)}$ binarna varijabla koja indicira pripada li primjer i grupi k (1) ili ne (0), te $\|x^{(i)} - \mu_k\|^2$ kvadrat euklidske udaljenosti između primjera $x^{(i)}$ i centroida μ . Najjednostavnija metoda odabira broja grupa je metoda koljena. Metoda



Slika 3.2. Usporedba euklidske udaljenosti i dinamičkog savijanja vremena [7]

koljena zahtijeva da se algoritam K-sredina pokrene više puta za različit broj grupa, te se izradi graf ovisnosti pogreške J o broju grupa. Optimalan broj grupa je koljeno krivulje, točka u kojoj pogreška prestaje naglo padati - za ovaj broj grupa pogreška se značajno smanjila, a daljnjim povećanjem broja grupa neznatno će se smanjivati i za njime nema potrebe. Dodatno, optimalan broj grupa određen je i brojem primjera koji pripadaju svakoj grupi, jer kod grupa s malo primjera dolazi do pretjeranog prilagođavanja.

3.3.2. Dinamičko savijanje vremena

Algoritam K-sredina, osim euklidske udaljenosti, kao metodu za izračun može koristiti i dinamičko savijanje vremena (engl. *Dynamic Time Warping, DTW*). DTW je metoda izračuna udaljenosti pogodna za vremenske serije, jer jednoj točki u seriji može pridijeliti više njih iz druge serije. Tako se omogućuje uočavanje uzoraka u vremenu, čak i ako su različite brzine ili amplitude (slika 3.2.). Zbog tog svojstva, DTW za vremenske serije daje bolje rezultate od euklidske udaljenosti.

4. Podaci

Podaci za rad prikupljeni su u Laboratoriju za odziv potrošnje na Fakultetu elektrotehnike i računarstva za razdoblje od 1.4.2022. do 31.12.2023. Laboratorij sadrži 18 uređaja, a za svaki se uređaj u bazu zapisuju sljedeći podaci: id i vrijeme zapisa (mjerenja), naziv uređaja, energija, radna snaga, maksimalna radna snaga, struje u sve tri faze, linijski naponi, fazni naponi, faktor snage te prividna snaga. Frekvencija zapisa u bazu je između 2 i 3 sekunde. Uređaj na kojem je provedena dalje opisana analiza je električna grijalica vode. Za analizu potrošnje najkorisnija značajka ovog uređaja je snaga, koja je u bazi pohranjena pod imenom `total_active_power`.

Kod korišten za pripremu, analizu i grupiranje podataka nalazi se u repozitoriju na referenci [8].

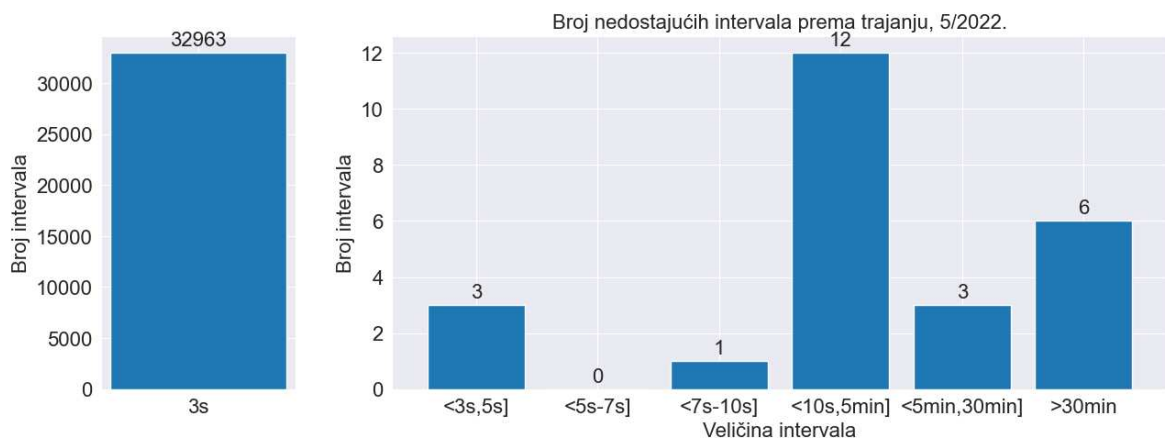
4.1. Priprema podataka

4.1.1. Određivanje intervala s nedostajućim podacima

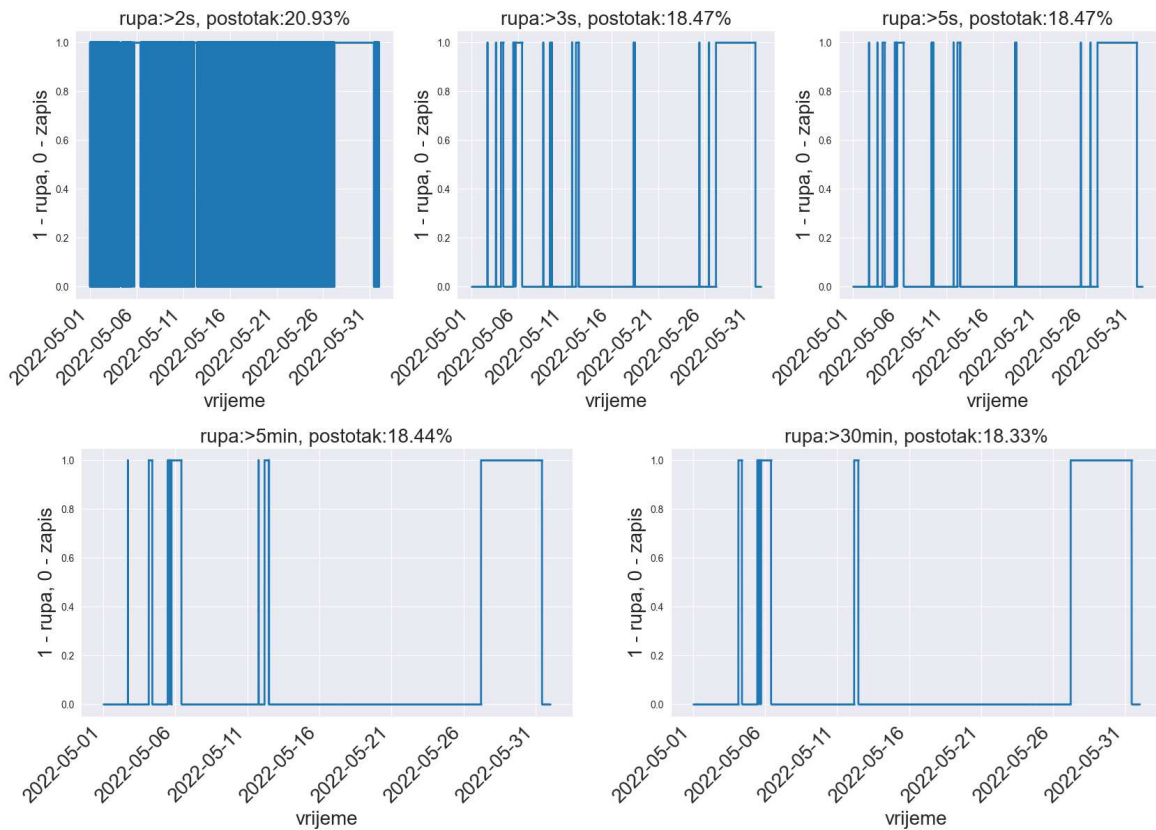
Prvi korak u pripremi podataka za grupiranje je identificiranje intervala u kojima nedostaju podaci te njihovo popunjavanje. Za svaki mjesec napravljena je tablica takvih intervala, s podacima o njihovom početku, kraju i trajanju, te je izračunata prosječna mjesečna frekvencija zapisa podataka. Dobiveni intervali podijeljeni su u kategorije prema trajanju rupa u podacima, te je prema tome izračunata frekvencija pojedine kategorije (slika 4.1.).

Izrađena je funkcija koja za različite veličine intervala računa broj i postotak nedostajućih podataka u odnosu na podatke koji bi trebali biti zapisani u tom mjesecu, te grafom prikazuje položaj tih intervala u mjesecu (slika 4.2.).

Budući da se vrijeme spremanja zapisa u bazu zaokružuje na cijelu sekundu, a frekvencija zapisa je između 2 i 3 sekunde, vremenski razmak između dvaju uzastopnih zapisa bit će zaokružen na 2 ili 3 sekunde. Stoga se intervalom u kojem nedostaju podaci smatra bilo koji razmak između zapisa čija je razlika veća od 3 sekunde. Primijećeno je da 2022. godina ima značajno veći postotak nedostajućih podataka od 2023.: kolovoz 2022. nema niti jedan zapis, a u ostalim mjesecima postotak nedostajućih zapisa varira između 15 i 76 %. Za 2023. taj se postotak kreće oko 17%.



Slika 4.1. Frekvencija različitih veličina praznih intervala



Slika 4.2. Intervali s nedostajućim podacima, 5. mjesec 2022.

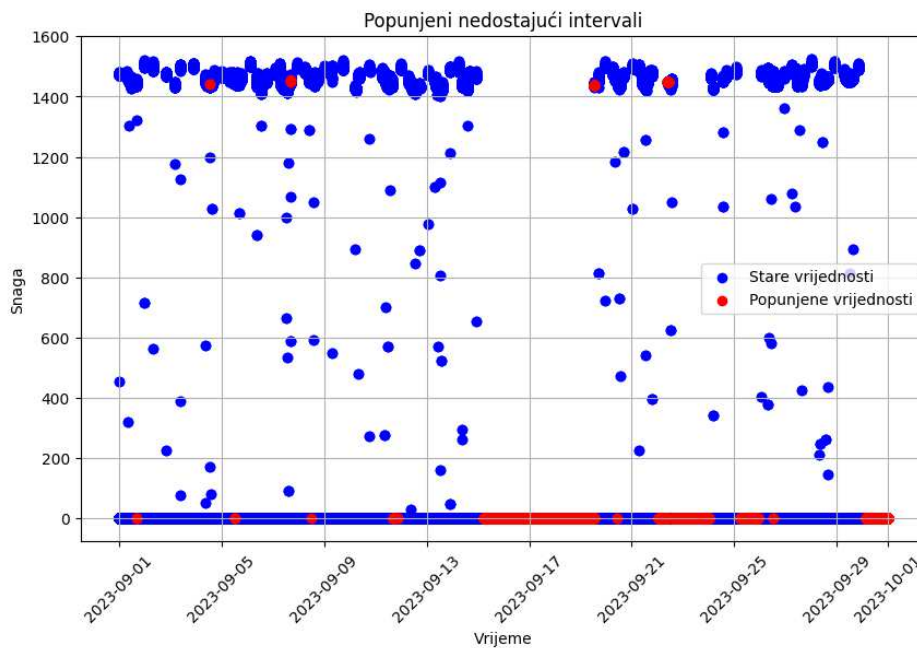
4.1.2. Popunjavanje intervala s nedostajućim podacima

Prazni intervali kraći od 30 sekundi popunjeni su polinomnom (kvadratnom) interpolacijom, dok je za one veće osmišljena vlastita metoda.

Prije popunjavanja određeno je prosječno i maksimalno kontinuirano trajanje rada uređaja, u kojem se on nije ugasio i nema nedostajućih zapisa, na cijelom skupu podataka. Intervali iznad 30 s trajanja popunjeni su na sljedeći način: ako je na početku i kraju intervala vrijednost 0 pa uređaj nije radio niti na početku niti na kraju intervala, čitav se interval popunjava nulama. Ako je s jedne strane 0, a s druge neka druga vrijednost snage, interval se popunjava rubnom vrijednosti snage sve dok trajanje rada ne dođe do prosječnog, prethodno izračunatog trajanja rada, a nakon toga se popunjava nulama. Za situacije u kojima je s obje strane intervala vrijednost veća od nula, provjerava se hoće li se popunjavanjem intervala vrijednostima kao da je uređaj upaljen i radi prijeći maksimalno kontinuirano trajanje rada uređaja za taj mjesec: ako neće, interval se popunjava aritmetičkom sredinom rubova, a ako hoće popunjava se od rubova odgova-

rajućim rubnim vrijednostima dok se ne postigne prosječno vrijeme rada, a ostatak se puni nulama.

Ovakva metoda omogućava realno popunjavanje praznih intervala trajanja do nekoliko minuta, ali kod većih intervala od nekoliko sati ona neće točno predstaviti ponašanje uređaja, jer čitav interval u sredini puni nulama. Na slici 4.3. prikazan je jedan mjesec popunjen ovakvom metodom: veći interval popunjen na ovaj način očigledno odskače od stvarnog ponašanja uređaja.



Slika 4.3. 9.mjesec 2023. popunjen početnom strategijom

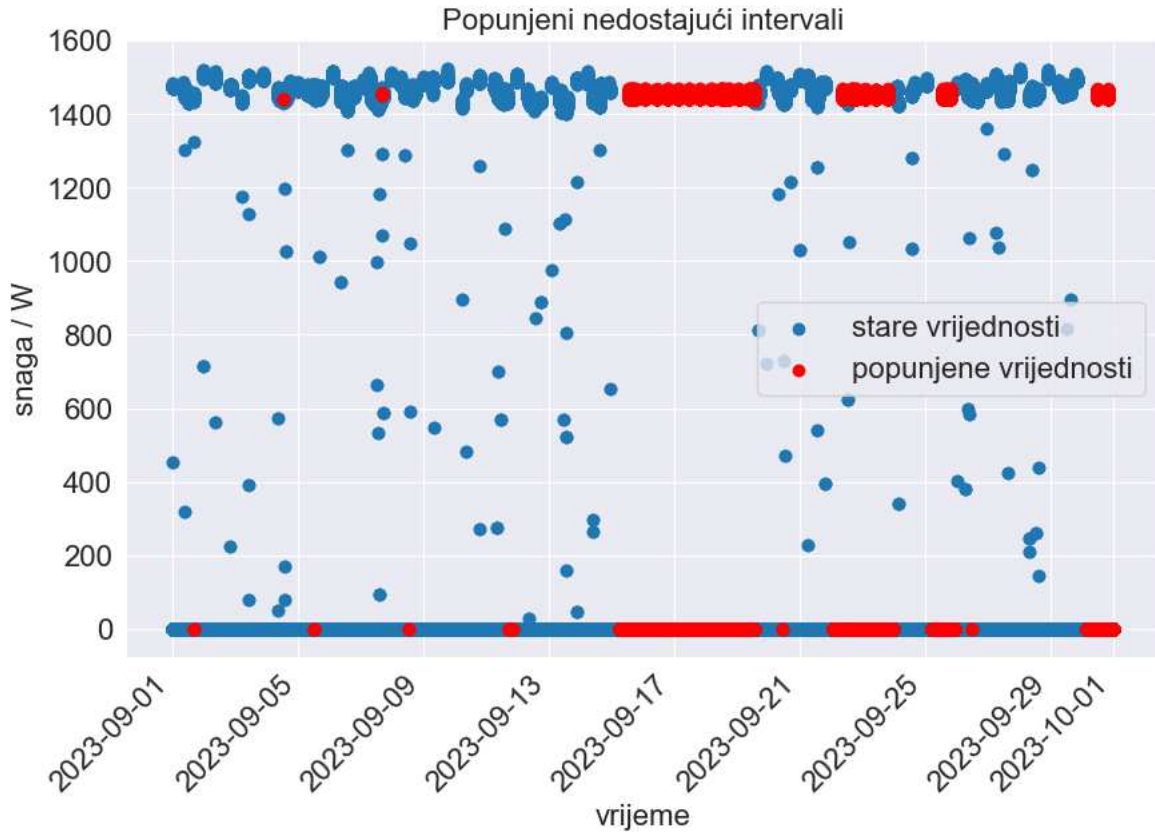
Iz ovog je razloga naknadno razvijena drugačija metoda [9].

Prije popunjavanja određeno je prosječno vrijeme u kojem uređaj ne radi za: radno vrijeme radnih dana (radni dani između 7 i 17 sati), noć radnih dana (radni dani 17 h - 7 h) te vikende, kao i prosječno kontinuirano trajanje rada uređaja te prosječna snaga pri radu \bar{P}_{rad} .

Intervali iznad 30 sekundi popunjeni su na sljedeći način: interval se puni nulama dok se ne dostigne odgovarajuće prosječno vrijeme u kojem uređaj ne radi; potom se popunjava vrijednostima snage dok se ne dostigne prosječno vrijeme rada, a zatim se postupak ponavlja. Da bi vrijednosti snage kojima se popunjavaju trenutci u kojima uređaj radi bile realnije, dobivene su pribrajanjem odmaka na prethodno izračunatu prosječnu

radnu snagu. Odmak se dobiva slučajno iz uniformne razdiobe s rubnim vrijednostima $\pm 0.01 \cdot \bar{P}_{\text{rad}}$.

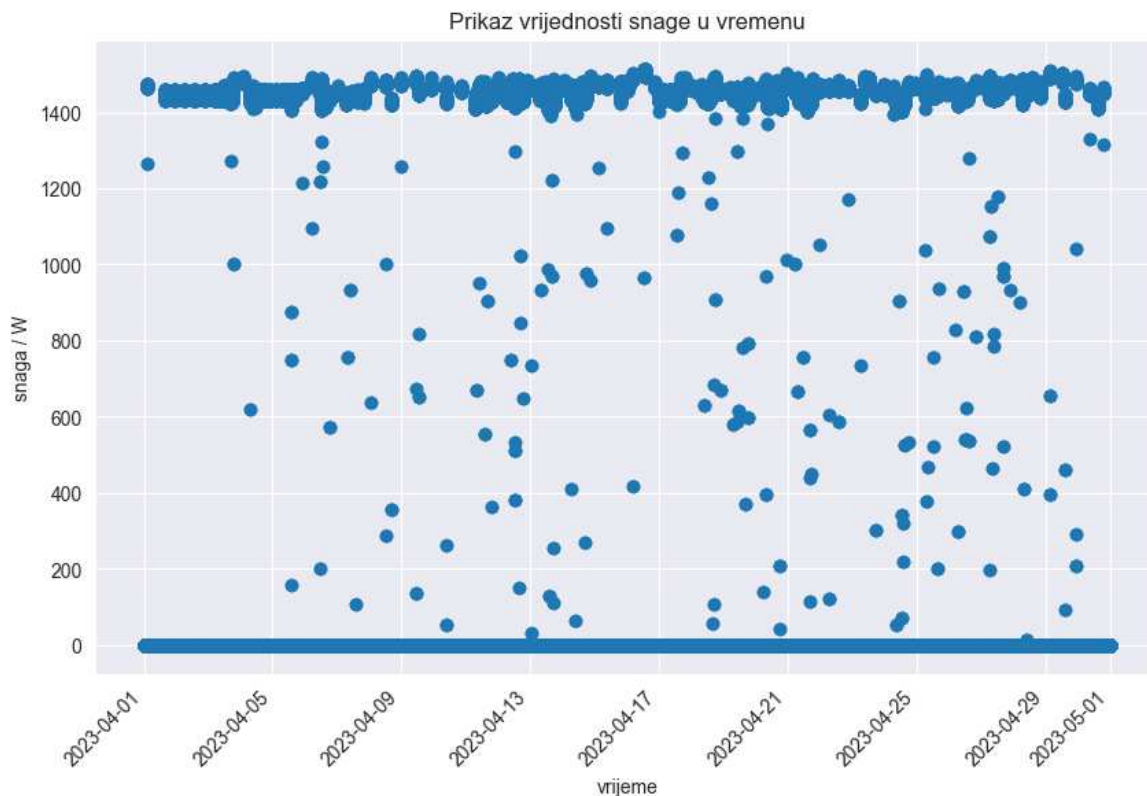
Na slici 4.4. prikazan je 9. mjesec 2023. popunjen ovom metodom.



Slika 4.4. 9.mjesec 2023. popunjen konačnom strategijom

4.1.3. Određivanje stršćih vrijednosti

Za određivanje stršćih vrijednosti korišteni su grafički prikazi snage u vremenu (slika 4.5.), pomoću kojih je određeno da one ne postoje. S grafa je vidljivo da snaga teži u dvije vrijednosti, 0 i drugu pozitivnu vrijednost. Ta druga vrijednost je prosječna vrijednost snage kada uređaj radi i približno iznosi 1447 W, što je blisko nazivnoj snazi uređaja. Stršće vrijednosti su one koje su negativne ili jako prelaze nazivnu snagu. S obzirom na to da se sve vrijednosti snage zapisane u bazu nalaze u intervalu $[0, 1541]$, zaključeno je da ne postoje stršće vrijednosti. Vrijednosti koje su po veličini između nazivne snage i 0 zapisuju se ako do očitavanja snage dođe upravo pri paljenju ili gašenju uređaja te zbog toga nisu krive, iako na grafu odskaču.



Slika 4.5. Snaga kroz 4. mjesec 2023. - nema stršćih vrijednosti

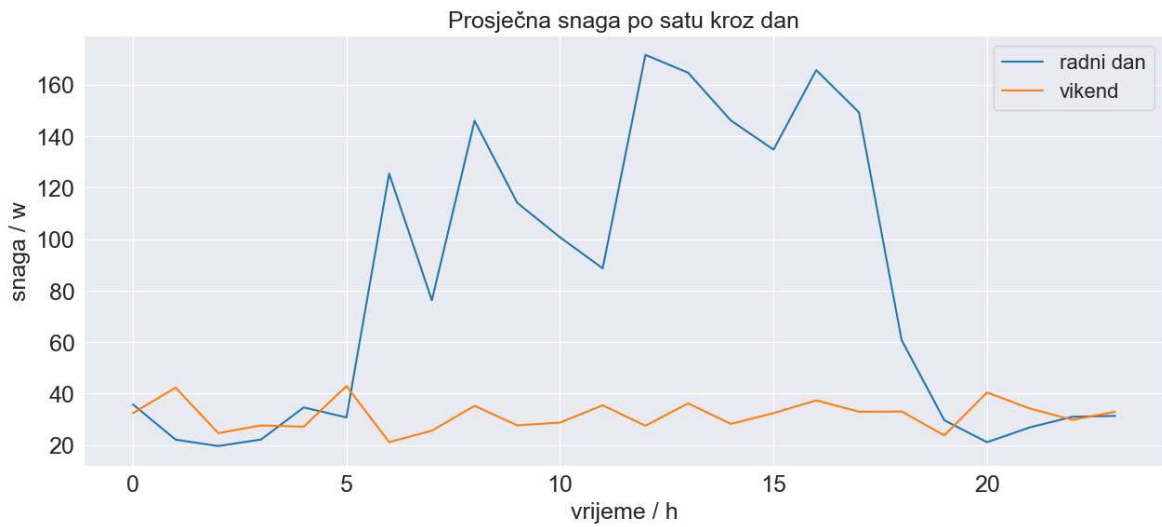
4.1.4. Poduzorkovanje

Premda algoritmi strojnog učenja služe za rad na velikoj količini podataka, zbog povećanja brzine i smanjenja složenosti potrebno je provesti poduzorkovanje. Kod prevelike količine podataka koje se ne koristi dobro može doći do prenaučivosti, prilagođavanja modela šumu. Prevelik broj zapisa dovodi do ogromne dimenzionalnosti, koja ne odgovara metodi K-sredina. Konačno, prevelika preciznost u ovoj će primjeni naglašavati nebitne razlike i otežati uočavanje sličnog uzorka između tjedana. U skladu s budućim prelaskom energetske tržišta sa satne na 15-minutnu razinu, podaci su poduzorkovani na period od 15 minuta. Za vrijednost snage za svaki je period uzeta srednja vrijednost.

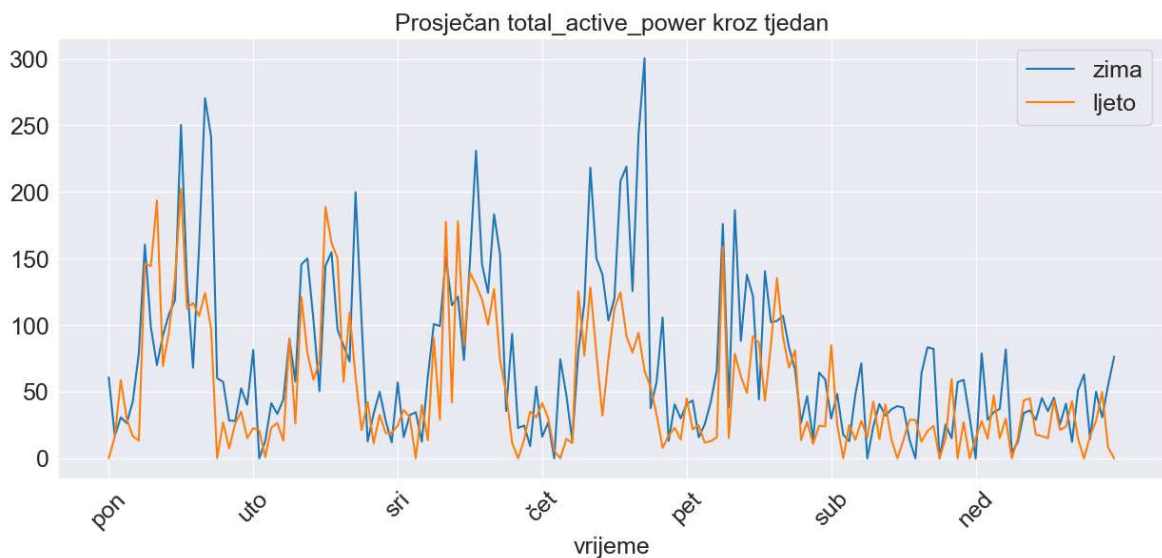
4.2. Analiza vremenske serije

Prije primjene uobičajenih testova za određivanje karakteristika vremenskih serija, korisno je grafički prikazati nekoliko vremenskih perioda kako bi se uočili uzorci u podacima i olakšala njihova interpretacija. Na slici 4.6. prikazana je usporedba snage na prosječan radni dan i vikend, a na slici 4.7. prikazana je prosječna snaga kroz jedan ljetni i

zimski tjedan. Uočljiva je razlika u vrijednostima snage, koje su niže vikendom u odnosu na radni dan, te blago niže ljeti u odnosu na zimu.

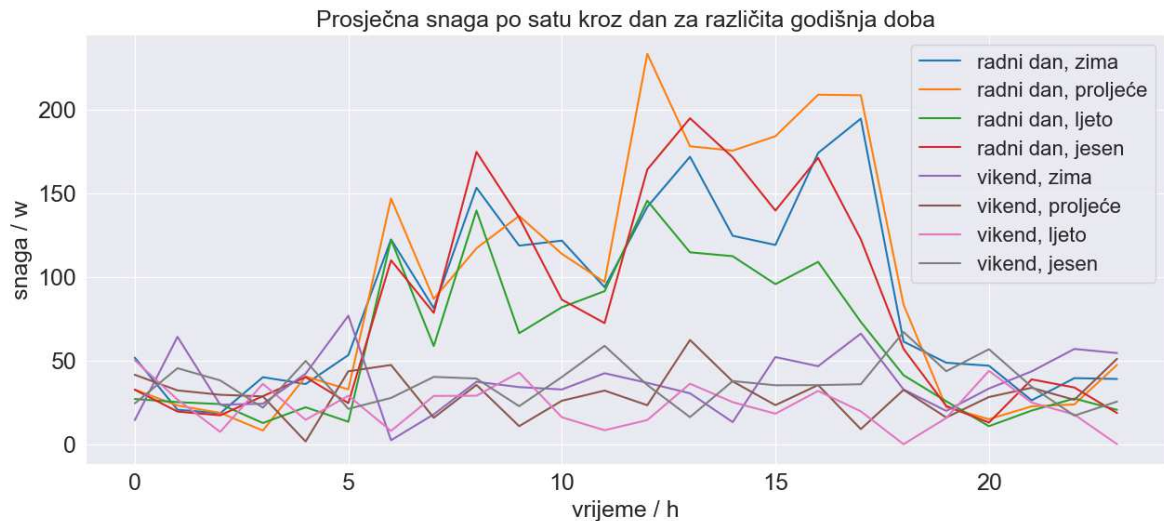


Slika 4.6. Usporedba potrošnje na prosječni rani dan i vikend



Slika 4.7. Usporedba prosječne satne potrošnje kroz ljetni i zimski tjedan

Na slici 4.8. prikazane su prosječne vrijednosti snage po satu za radne dane i vikende različitih godišnjih doba. Dok se vikendom ove snage kreću oko sličnih vrijednosti bez obzira na godišnje doba, radnim danom najniže vrijednosti su one ljetne. Ipak, razlike u potrošnji po godišnjim dobima nisu jako naglašene kao one po radnim i neradnim danima. Iz do sada prikazanog ne može se sa sigurnošću zaključiti da vremenska serija pokazuje godišnju sezonalnost. S druge strane, razlika između potrošnje radnim danom i vikendom svakako postoji.



Slika 4.8. Usporedba prosječne satne potrošnje kroz dane različitih godišnjih doba

Kao što je već spomenuto pri opisu popunjavanja nedostajućih podataka, određena su sljedeća prosječna vremena za cijelu godinu:

prosječno kontinuirano vrijeme rada, $\bar{T}_{\text{rad}} = 9 : 44 \text{ min}$

prosječno kontinuirano vrijeme ugašenosti za radno vrijeme radnih dana,
 $\bar{T}_{\text{ugrad}} = 01 : 49 : 12 \text{ h}$

prosječno kontinuirano vrijeme ugašenosti za neradno vrijeme radnih dana,
 $\bar{T}_{\text{ugnerad}} = 04 : 53 : 34 \text{ h}$

prosječno kontinuirano vrijeme ugašenosti vikendom, $\bar{T}_{\text{ugvik}} = 07 : 11 : 19 \text{ h}$.

Uređaj se tijekom radnog dana prosječno pali 5 puta, tijekom noći radnih dana 2 do 3 puta, a vikendom 3 do 4 puta.

4.2.1. Stacionarnost

Da bi se provjerila stacionarnost vrijednosti snage u vremenu proveden je ADF test. Početna hipoteza ADF testa je da su podaci nestacionarni - da imaju nekakav trend ili velike promjene u vremenu. Test vraća P-vrijednost, što je vjerojatnost koja mjeri jačinu dokaza protiv hipoteze, vrijednost ADF statistike te tri kritične vrijednosti za različite razine značajnosti. Mala P-vrijednost indicira da se hipoteza može odbaciti, a za odbacivanje hipoteze potrebno je još i da je ADF statistika manja od sve tri kritične vrijednosti.

Vrijednosti dobivene za snagu električne grijalice vode iznose:

ADF statistika: -11.153425634856497

P-vrijednost: $2.9053613256300523 \times 10^{-20}$

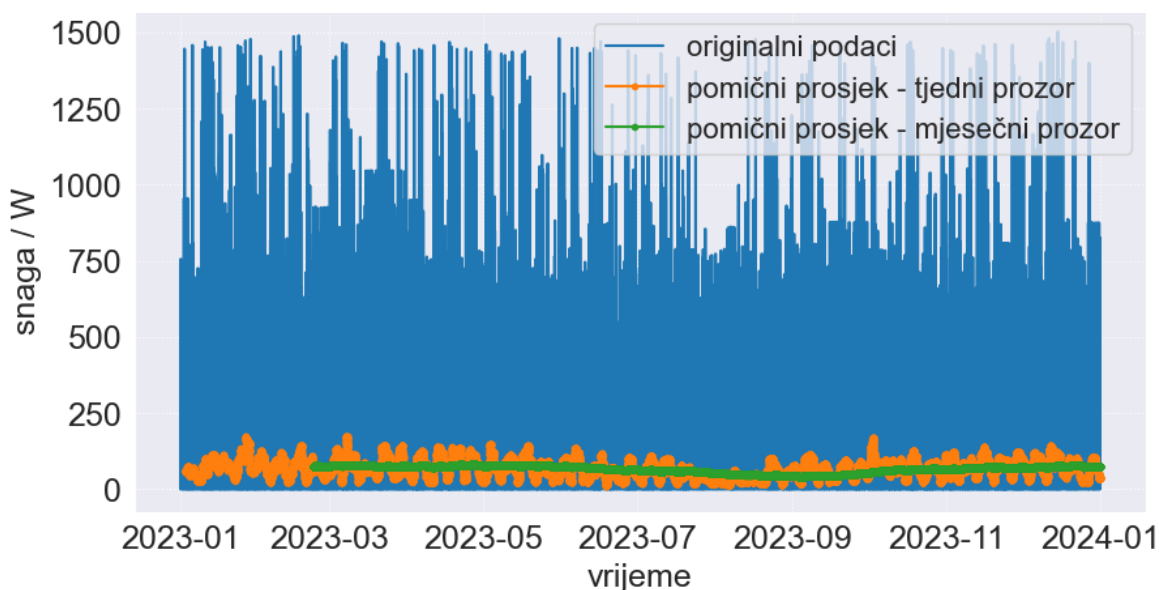
Kritične vrijednosti:

1% : -3.43

5% : -2.86

10% : -2.57

Iz dobivenih podataka zaključuje se da bi vremenska serija mogla biti stacionarna. Na slici 4.9. prikazan je pomični prosjek snage za različite vremenske okvire. Primjetan je lagan pad prosječne vrijednosti u kolovozu, ali taj je pad dovoljno malen da bi se podaci i dalje mogli smatrati stacionarnima.

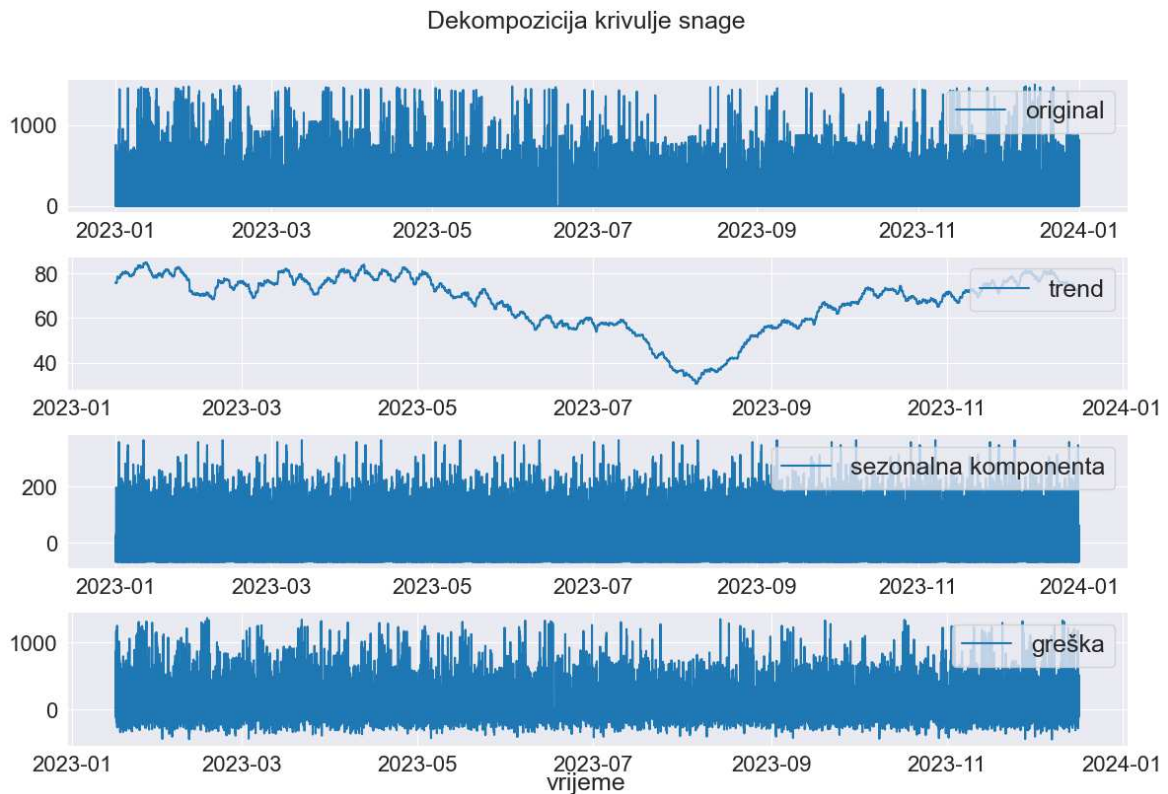


Slika 4.9. Pomičan prosjek kroz 2023.

4.2.2. Sezonalnost

Na temelju rezultata ADF testa može se zaključiti da u vremenskoj seriji ne postoji značajan trend. Za razliku od trenda, bilo bi smisleno očekivati sezonalnost na godišnjoj razini zbog razlike u temperaturama po godišnjim dobima. Dekompozicijom serije dobivene su komponente prikazane na slici 4.10. Suprotno očekivanjima, nije uočena očita sezonalnost. Iz ovih se komponenti vidi da električna grijalice vode nije pravilno peri-

odičan uređaj (ne pali se i gasi svaki dan u isto vrijeme), što je i očekivano iz prethodne analize.



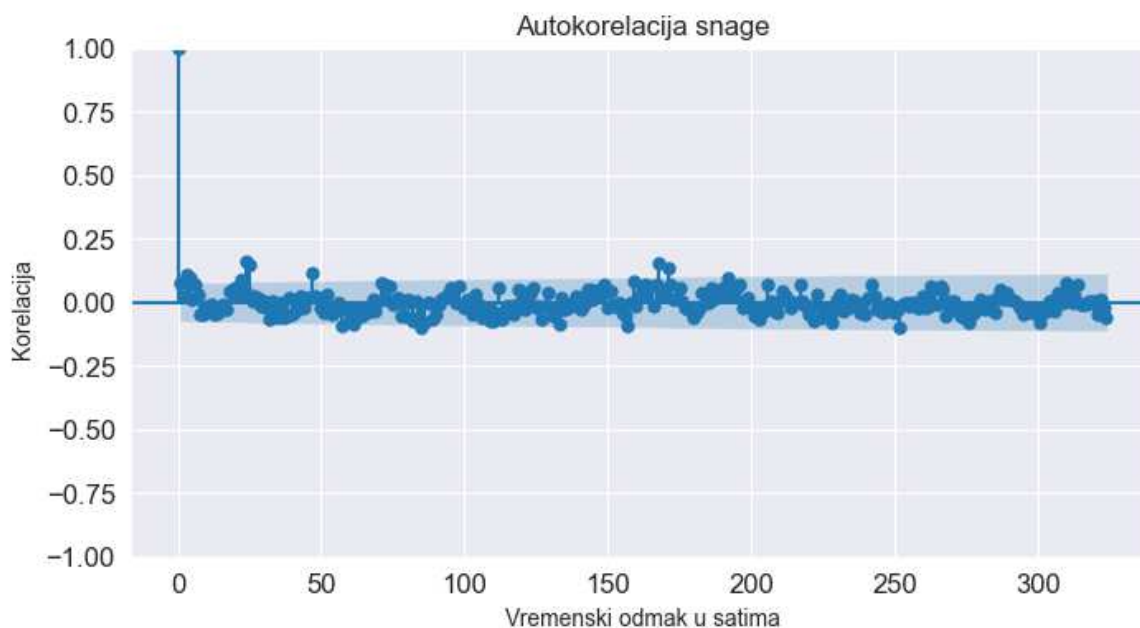
Slika 4.10. Dekompozicija krivulje snage 2023. godine

4.2.3. Autokorelacija

Napravljena je i analiza autokorelacije vrijednosti snage za različite vremenske trenutke/odmake. Graf autokorelacije prikazan je na slici 4.11. Iako su vrijednosti autokorelacije niske, uočljiva je dnevna periodičnost.

4.2.4. Korelacija između značajki

Da bi se odredilo postoje li značajne značajke osim snage, izračunata je korelacija među značajkama (slika 4.12.). Očekivano, korelacija između snage i struje je visoka. Postoji i korelacija između snage i ostalih značajki povezanih s njom, poput faktora snage i prividne snage. Primijećeno je i da su neke od značajki, poput faznog napona, konstante u vremenu te zbog toga nisu relevantne pri grupiranju.



Slika 4.11. Autokorelacija vremenske serije

	energy	total_active_power	demand_total_active_power	maximum_demand_total_active_power	i1	v1	total_power_factor	total_apparent_power
energy	1	-0.00098	0.000302	0.23085	-0.00101	-0.04308	0.000376	-0.00095
total_active_power	-0.00098	1	0.601183	0.019932	0.996773	-0.24593	0.442681	0.999909
demand_total_active_power	0.000302	0.601183	1	0.025024	0.606767	-0.23229	0.020258	0.601397
maximum_demand_total_active_power	0.23085	0.019932	0.025024	1	0.020045	-0.03109	-0.00285	0.01994
i1	-0.00101	0.996773	0.606767	0.020045	1	-0.24802	0.702862	0.996859
v1	-0.04308	-0.24593	-0.23229	-0.03109	-0.24802	1	-0.02929	-0.24596
total_power_factor	0.000376	0.442681	0.020258	-0.00285	0.702862	-0.02929	1	0.447652
total_apparent_power	-0.00095	0.999909	0.601397	0.01994	0.996859	-0.24596	0.447652	1

Slika 4.12. Korelacija značajki

Iz analize je zaključeno da, zbog visoke korelacije pojedinih značajki, nema potrebe za dodavanjem novih značajki iz baze. Vremenska serija ne pokazuje godišnju sezonalnost, pa stoga nisu potrebne ni značajke temeljene na godišnjem dobu. Serija je stacionarna što omogućava korištenje algoritma K-sredina bez ikakvih ograničenja. Traženje repre-

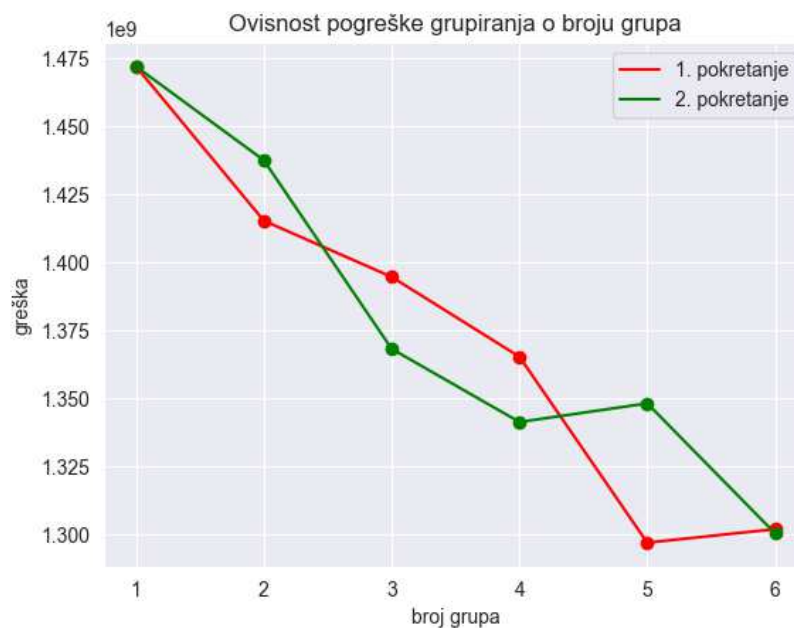
zentativnog tjedna svodi se na grupiranje tjednih krivulja potrošnje uređaja. Značajke koje se koriste pri grupiranju su snage kroz čitav tjedan s periodom od 15 minuta, što je $4 \cdot 24 \cdot 7 = 672$ vrijednosti snage tjedno, tj. 672 značajke po primjeru. Grupiranjem dana s po 96 značajki (što odgovara $4 \cdot 24$), dobiva se reprezentativni dan koji služi kao dobar pokazatelj kvalitete grupiranja tjedana i metode K-sredina s DTW distancom.

5. Reprezentativni tjedni i dani potrošnje

5.1. Reprezentativni tjedni

5.1.1. Grupiranje euclidskom distancom

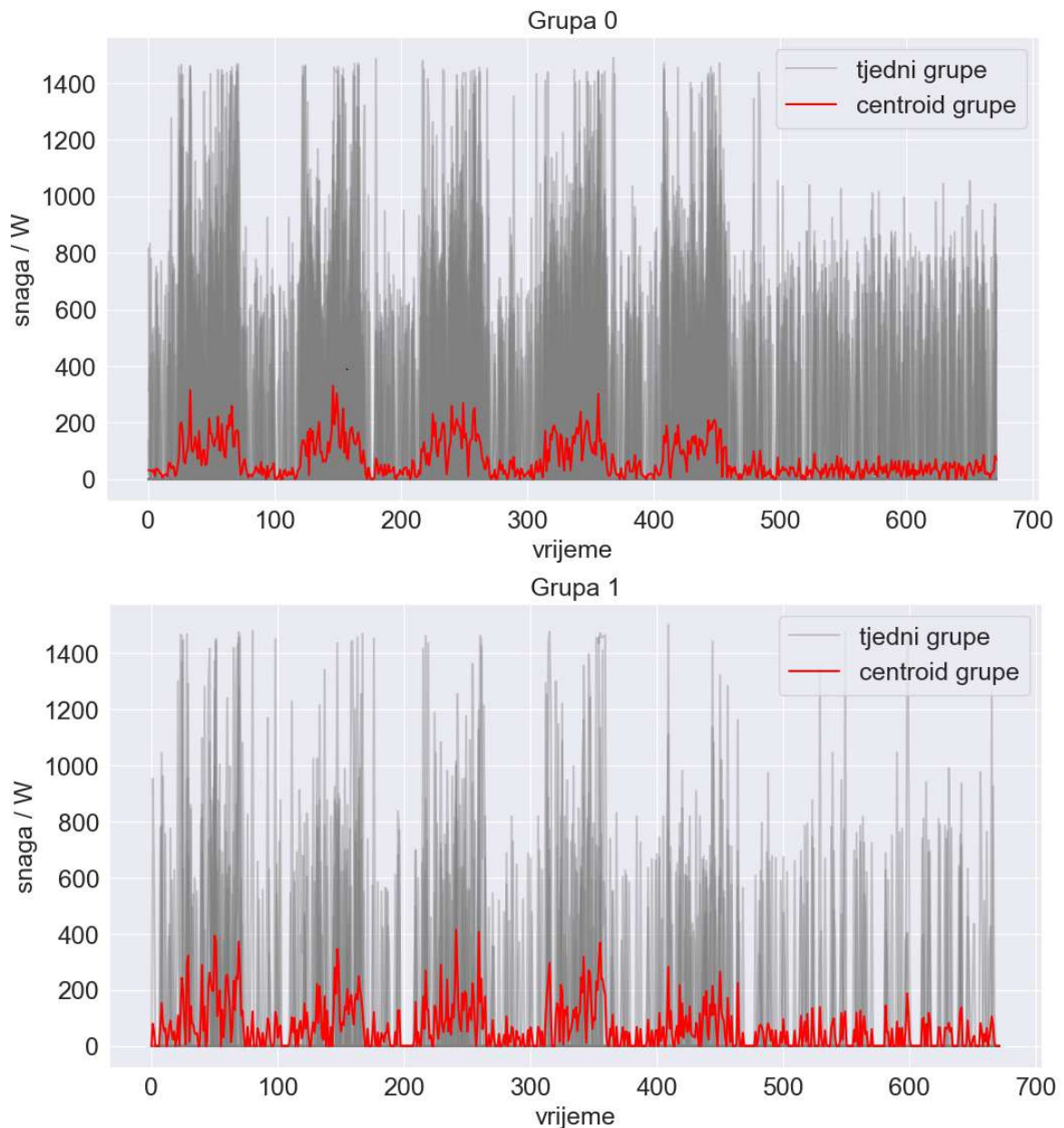
Provedeno je grupiranje tjedana koristeći algoritam K-sredina s euclidskom distancom. Grupiranje je pokrenuto više puta za broj grupa od 1 do 6, te je iscrtan grafički prikaz ovisnosti pogreške grupiranja J o broju grupa (slika 5.1.). Kao dobra rješenja nameću se grupiranja u 2 te u 4 grupe.



Slika 5.1. Algoritam K-sredina s euclidskom udaljenosti: prikaz ovisnosti pogreške J o broju grupa

Kao što je prethodno spomenuto, euclidaska udaljenost ne daje uvijek dobre rezultate za vremenske serije koje su različitih brzina i/ili amplituda. Reprezentativni tjedan pri-

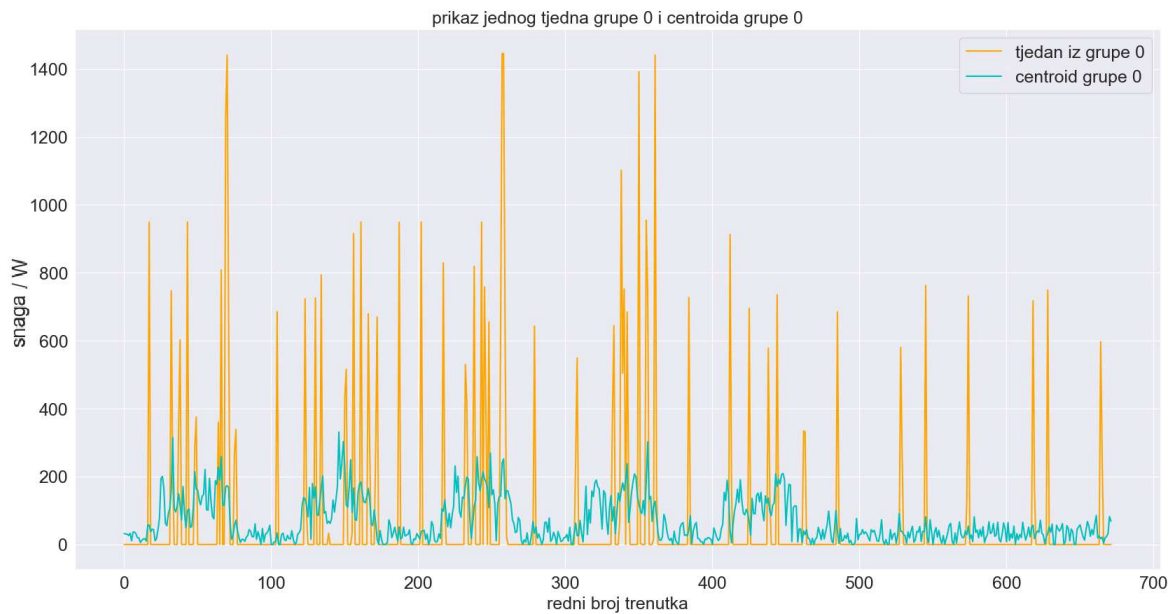
kazan za grupiranje u dvije grupe na slici 5.2. zapravo je aritmetička sredina vrijednosti značajki svih tjedana grupe u svakoj točki. Zbog vremenskog pomaka u radu uređaja, i posljedično različitih vremena u kojima se pojavljuju vršne vrijednosti snage, ovakav izračun nije točan i vrijednosti snage dobivenog centroida su daleko niže od vrijednosti stvarnih snaga grupiranih tjedana (slika 5.3.).



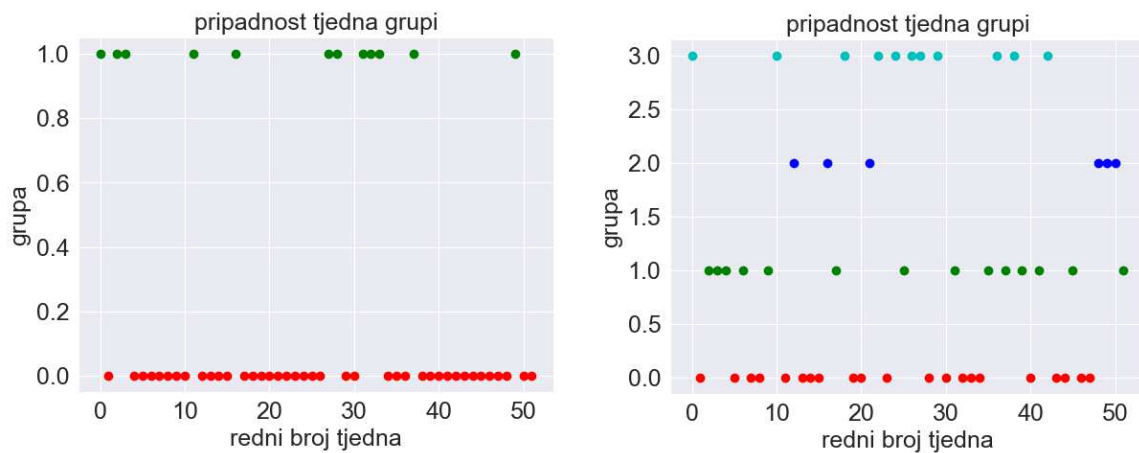
Slika 5.2. Algoritam K-sredina s euklidskom udaljenošću: grupirani tjedni i centriodi grupa

Rezultati dobiveni za podjelu na 4 grupe imaju identičan problem. Na slici 5.4. prikazana je pripadnost tjedana kroz godinu grupama, za 2 i 4 grupe. Sukladno očekivanjima, grupiranje nije pokazalo nikakvu sezonalnost na godišnjoj razini - tjedni se nisu prirodno

grupirali po godišnjim dobima.



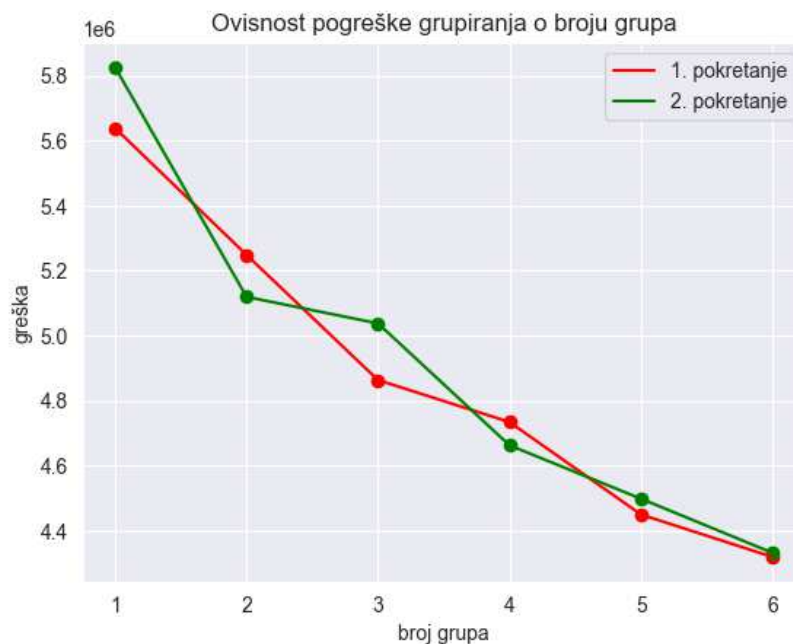
Slika 5.3. Algoritam K-sredina s euklidskom udaljenošću: prikaz jednog tjedna iz grupe 1 i pripadnog centroida



Slika 5.4. Algoritam K-sredina s euklidskom udaljenošću: raspodjela tjedana po grupama

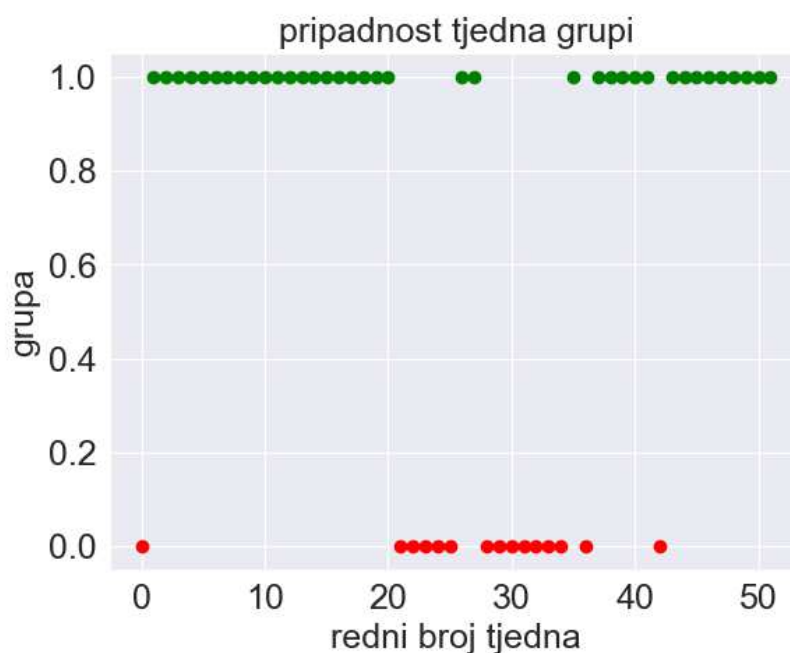
5.1.2. Grupiranje u 2 grupe dinamičkim savijanjem vremena

Algoritam K-sredina s dinamičkim savijanjem vremena donio je bolje rezultate. Algoritam je pokrenut više puta za broj grupa od 1 do 6. Na slici 5.5. prikazan je graf ovisnosti pogreške grupiranja o broju grupa. Koljeno krivulje u nekim se slučajevima nalazi u točki krivulje za koju je broj grupa 2, a u nekim u točki s 3 grupe.



Slika 5.5. Algoritam K-sredina s DTW-om: prikaz ovisnosti pogreške J o broju grupa

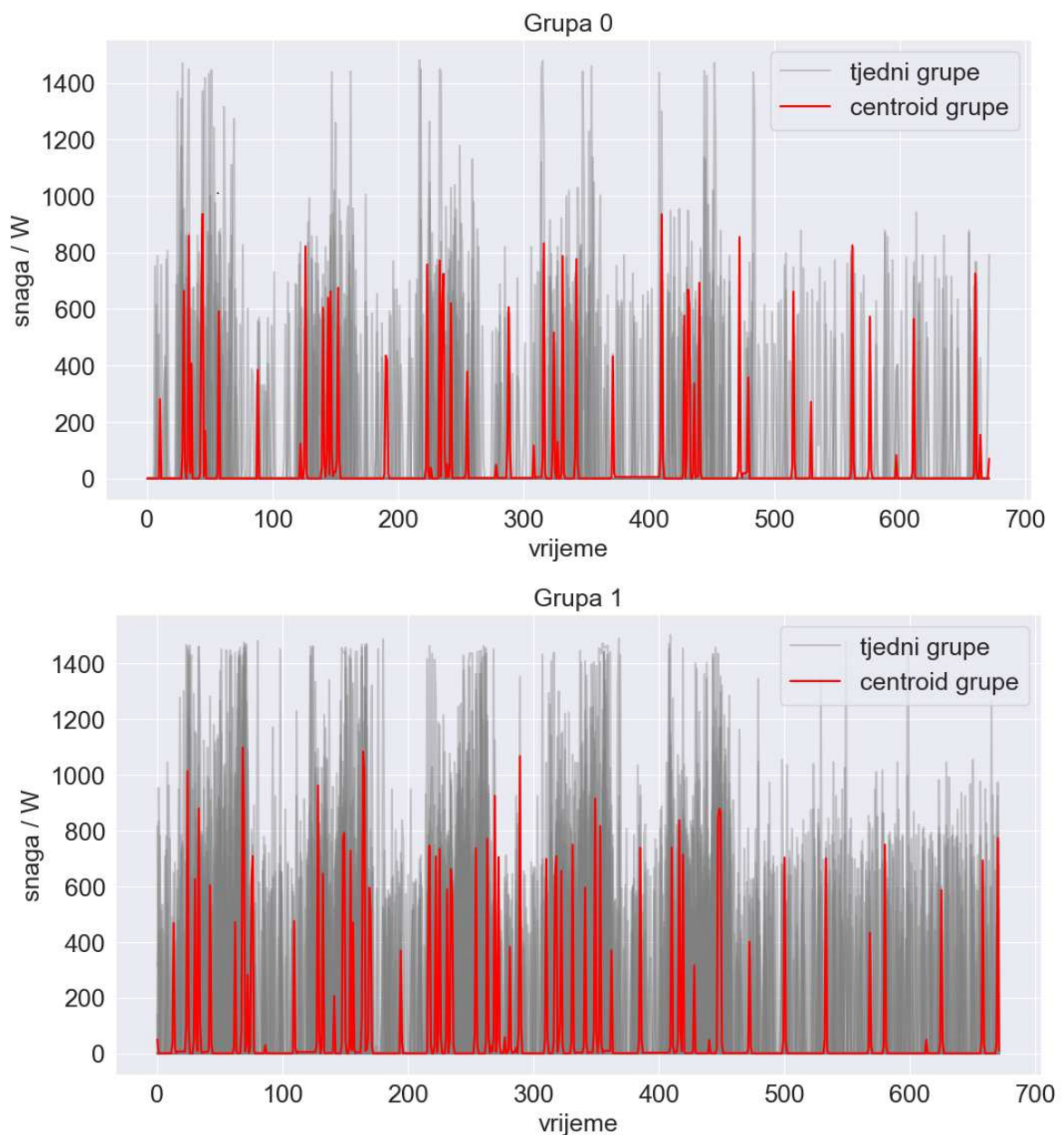
Dvije grupe u koje su tjedni grupirani razdvajaju radne tjedne kroz godinu te tjedne u kojima su godišnji odmori, što je prikazano na slici 5.6.



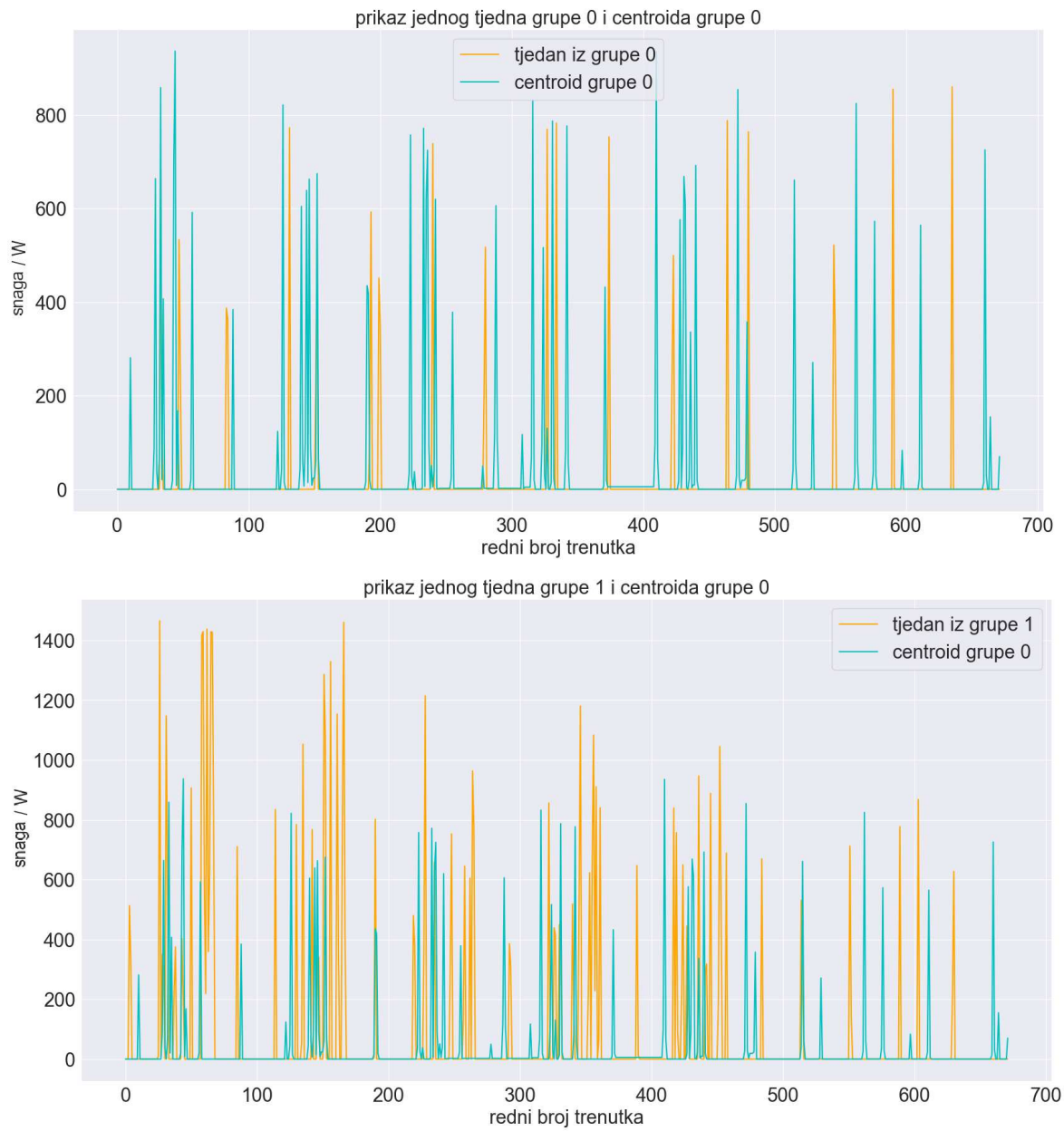
Slika 5.6. Algoritam K-sredina s DTW-om: raspodjela tjedana po grupama

Centroidi grupa određeni na ovaj način bolje predstavljaju svoje grupe, s realnijim vrijednostima snaga i učestalosti paljenja. Centroidi grupa prikazani su zajedno s tjednima koje predstavljaju na slici 5.7. Razlika između grupa, odnosno njihovih predstavnika,

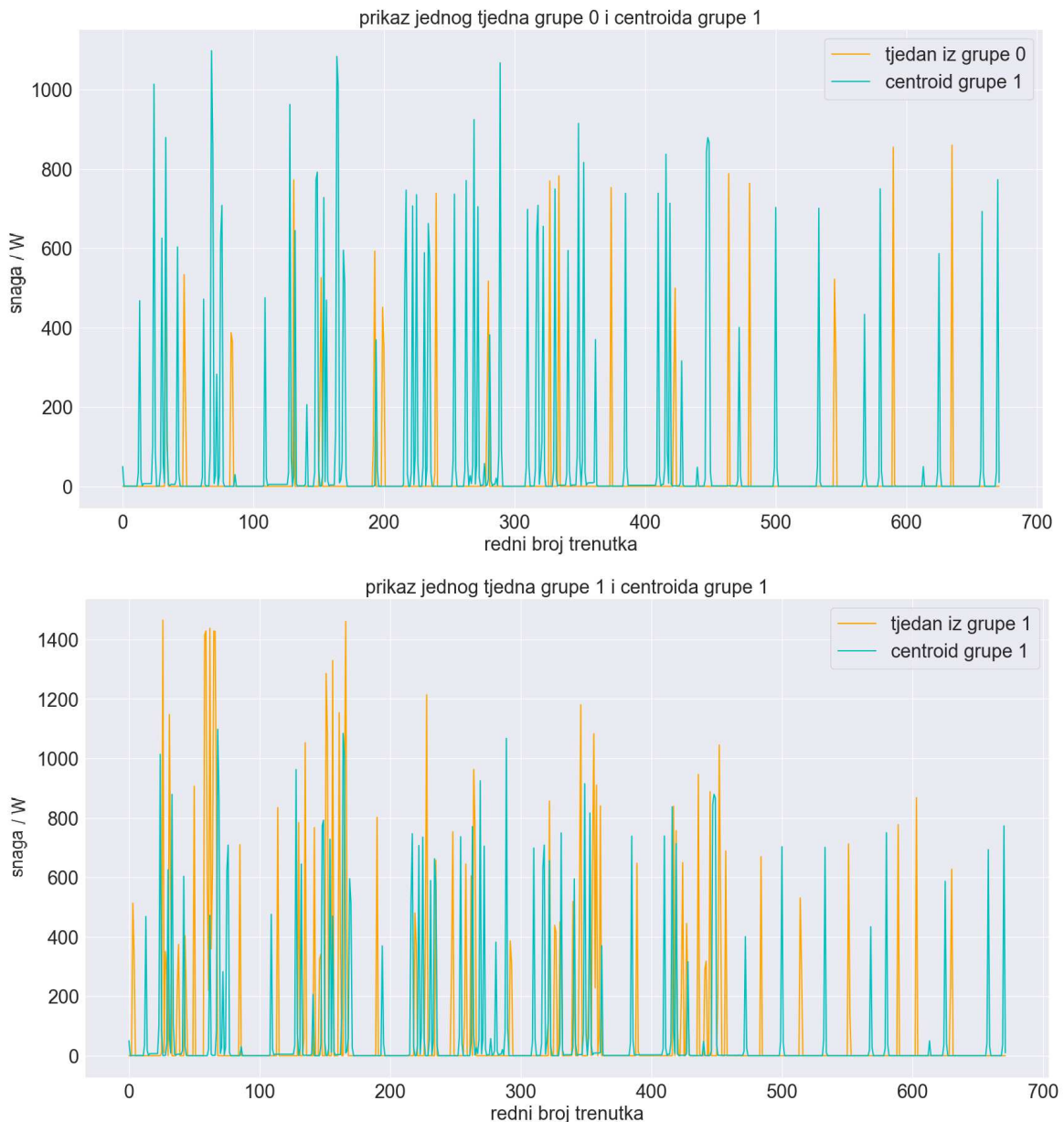
leži u učestalosti paljenja uređaja: u grupi 1 broj i gustoća vrhova krivulje su značajno veći (slika 5.7., 300. - 400. trenutak), što sugerira da ta grupa predstavlja radne tjedne. Primjetna je i sličnost krivulja vikendima (slika 5.7., 500. - 700. trenutak), kad uređaj radi rjeđe. Slike 5.8. i 5.9. prikazuju usporedbu po jednog tjedna svake grupe s centroidima: centroid grupe 1 vidljivo je gušći od tjedna iz grupe 0. Isto tako, vrhovi centroida grupe 0 prerijetki su da bi predstavljali tjedne iz grupe 1. Ovakvo grupiranje podržava rezultate dobivene analizom potrošnje, kojom je uočena niža potrošnja u ljetnim tjednima, koji su ujedno i neradni.



Slika 5.7. Algoritam K-sredina s DTW-om: grupirani tjedni i centroidi grupa



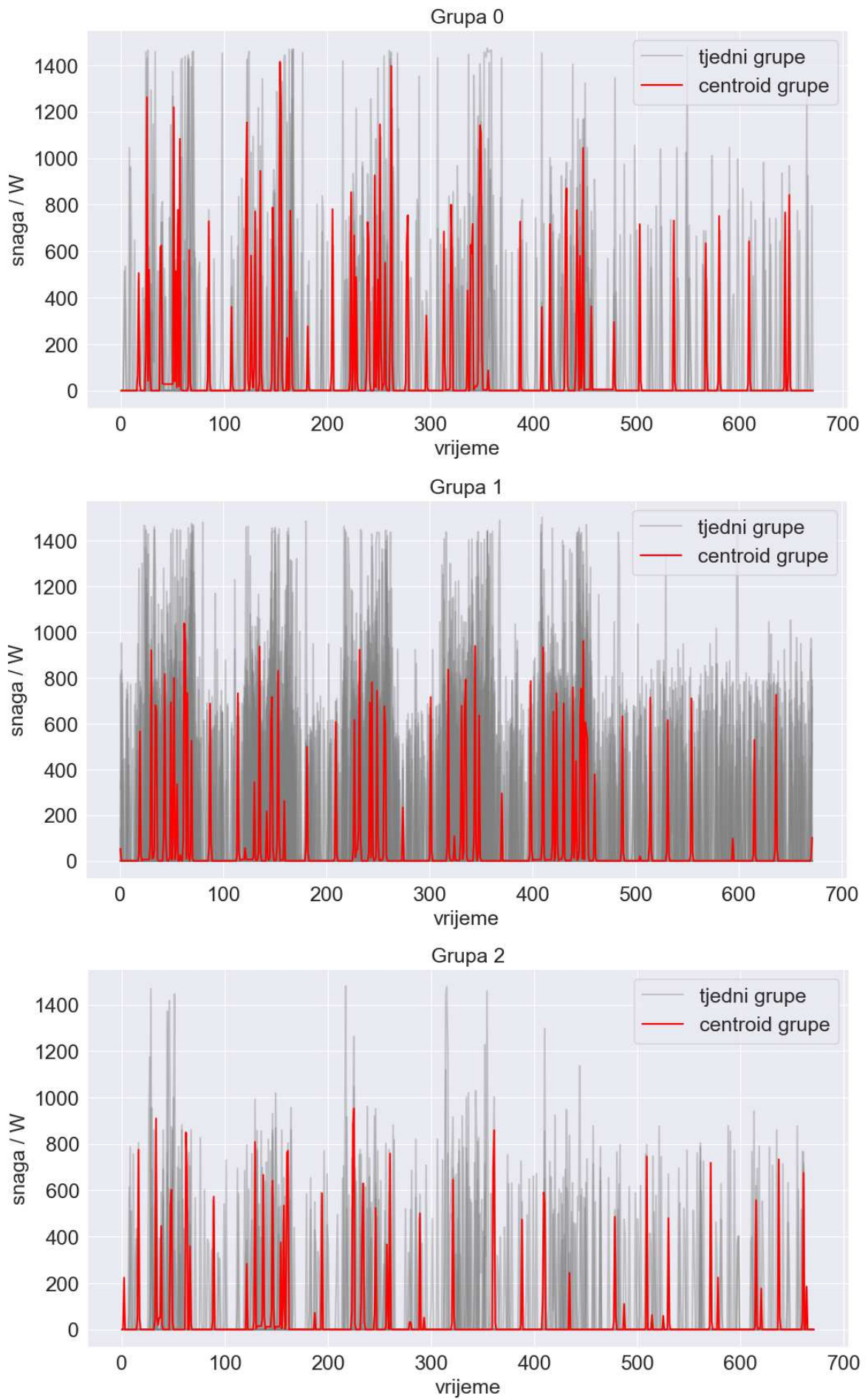
Slika 5.8. Algoritam K-sredina s DTW-om: centroid grupe 0 i tjedni iz svake od grupa



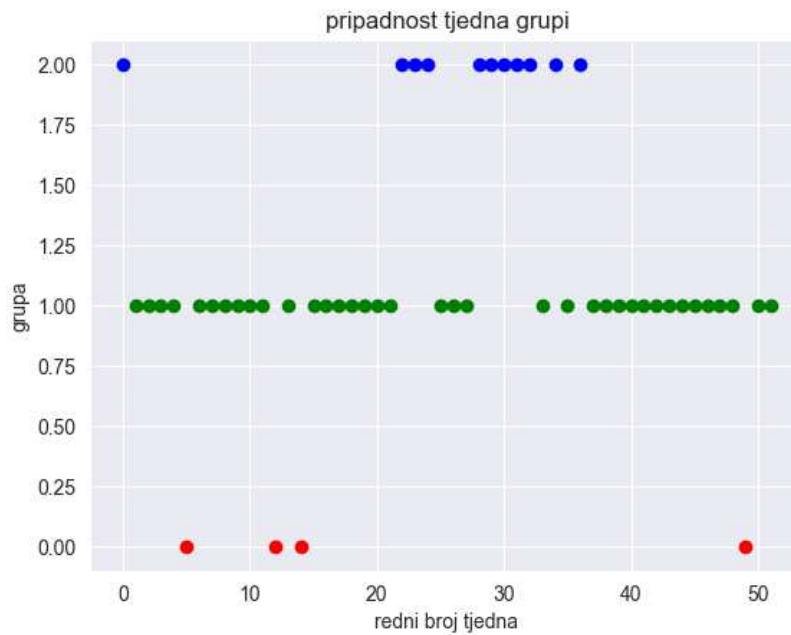
Slika 5.9. Algoritam K-sredina s DTW-om: centroid grupe 1 i tjedni iz svake od grupa

5.1.3. Grupiranje u 3 grupe dinamičkim savijanjem vremena

U slučaju grupiranja u 3 grupe (slika 5.10.) greška J se u nekim slučajevima smanjuje, ali iz podjele tjedana po grupama (slika 5.11.) vidi se da dolazi do pretjerane prilagodivosti: grupi 0 pripadaju svega 4 tjedna. Ovakvo grupiranje, s velikom većinom tjedana koji pripadaju dvjema grupama, je besmisleno.



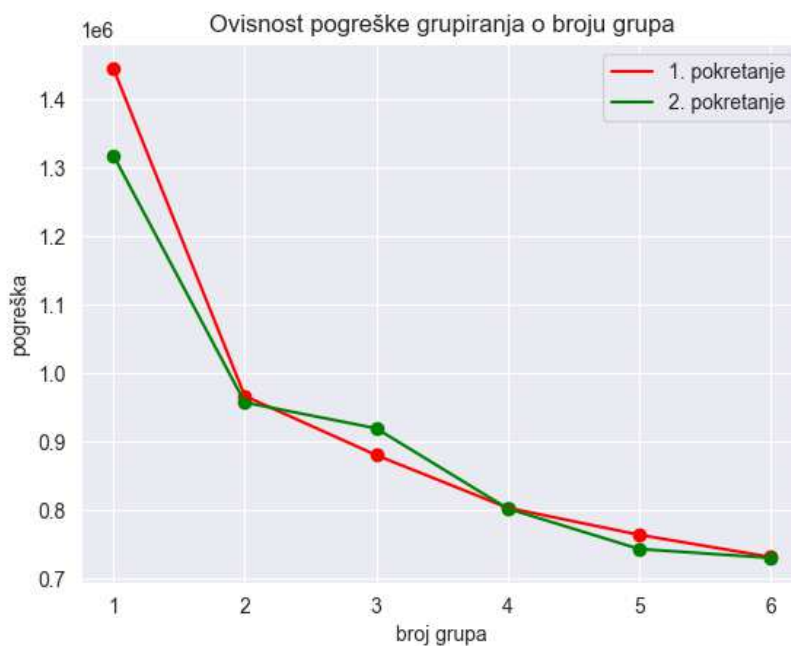
Slika 5.10. Algoritam K-sredina s DTW-om: tjedni grupirani u 3 grupe i centriodi grupa 31



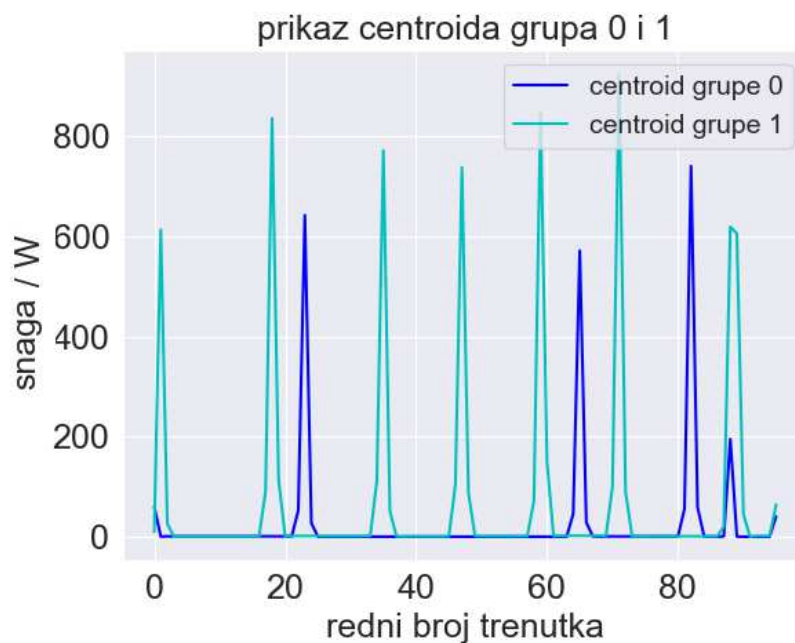
Slika 5.11. Algoritam K-sredina s DTW-om: raspodjela tjedana po grupama za 3 grupe

5.2. Reprezentativni dani

Algoritmom K-sredina s dinamičkim savijanjem vremena provedeno je i grupiranje primjera dana, te je dobiven graf pogreške na slici 5.12. S obzirom na koljeno krivulje, optimalno grupiranje je u dvije grupe, čiji su centroidi prikazani na slici 5.13. Grupu 0 odlikuju rjeđi vrhovi krivulje snage, koji su i niže vrijednosti u odnosu na grupu 1.

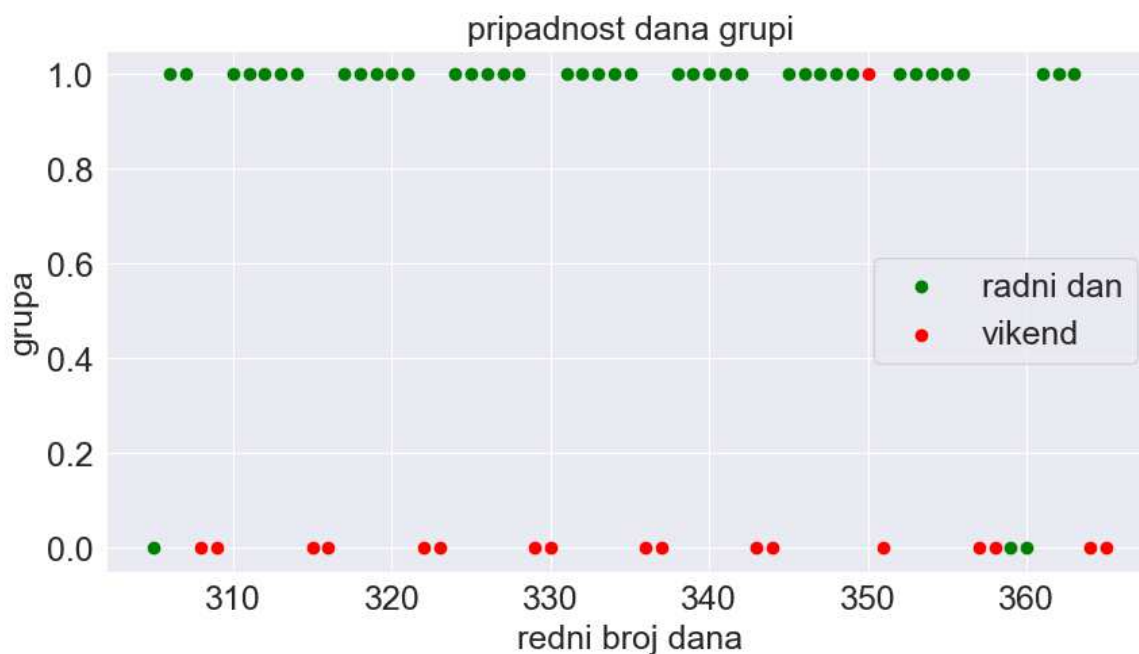


Slika 5.12. Grupiranje dana: pogreška J u ovisnosti o broju grupa



Slika 5.13. Grupiranje dana: centriodi svake od dviju grupa u koje su podijeljeni dani

Slično kao kod grupiranja tjedana, dobivene grupe dijele dane na radne i neradne. Na slici 5.14. jasno se vidi da grupi neradnih dana, uz vikende, prirodno pripadaju i praznici poput 1.11. (305. dan), 23.-26.12. (357. - 360. dan) te 30. i 31.12. (364. i 365. dan).



Slika 5.14. Grupiranje dana: raspodjela dijela dana 2023. po grupama

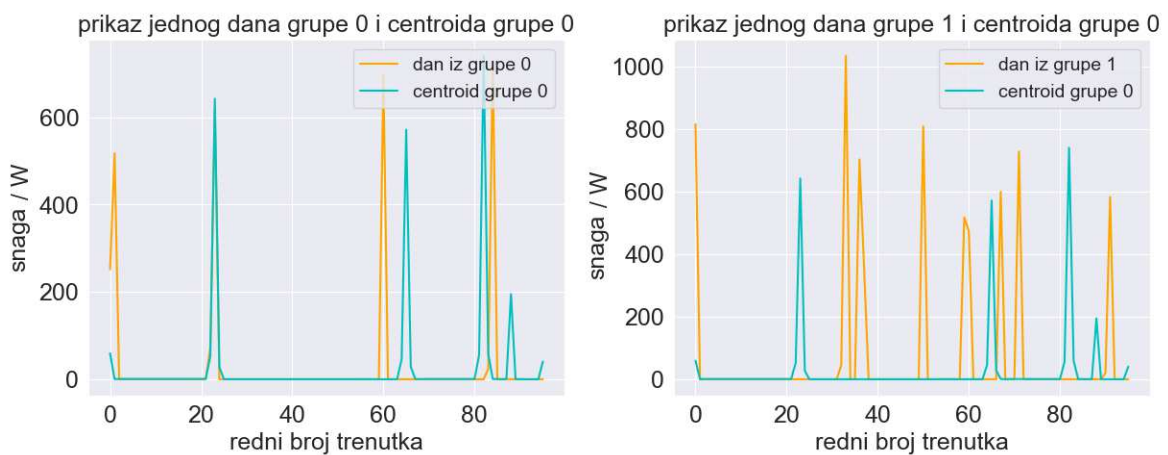
Na slikama 5.15. i 5.16. prikazani su dani iz svake od grupa uz centriode grupa: jasno

se vidi razlika između tjedana koji pripadaju različitim grupama, kao i sličnost pripadnim im centroidima, što potvrđuje smislenost ovakvog grupiranja.

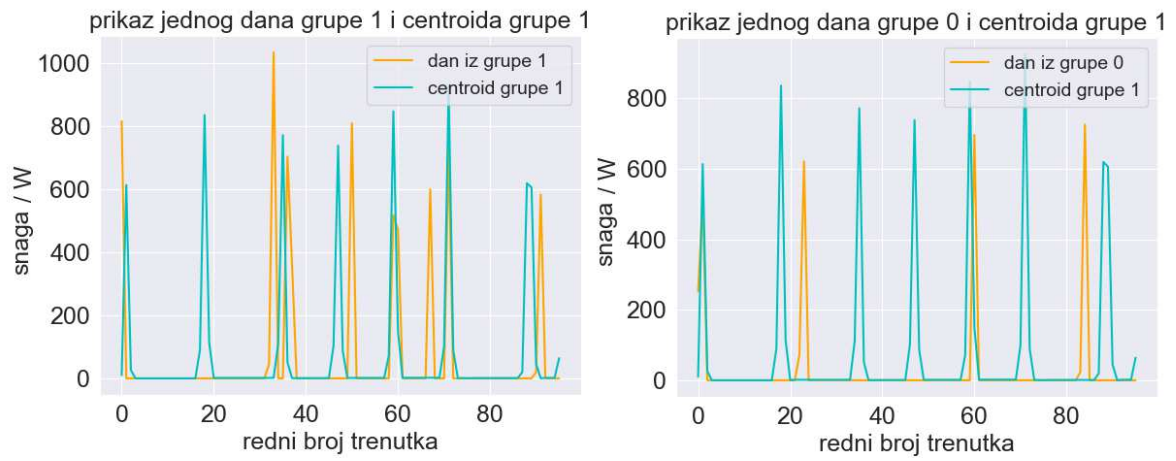
Centroidi dobiveni ovakvim grupiranjem u suglasnosti su s prethodnom analizom vremenske serije: kao što je uočeno u analizi, uređaj se radnim danom pali oko 5 puta tijekom radnog vremena, 2 - 3 puta tijekom noći radnog dana (slika 5.13., centroid grupe 1), te 3 do 4 puta vikendom (slika 5.13., centroid grupe 0).

Slika 5.15. s prikazom centroida 0. grupe i njezinim tjednom odlično prikazuje način na koji radi dinamičko savijanje vremena: iako vrhovi grafa nisu u istim vremenskim točkama, algoritam je prepoznao slično ponašanje različite brzine.

Rezultati grupiranja dana korištenjem dinamičkog savijanja vremena potvrđuju njegovu kvalitetu i pouzdanost. Sukladno rezultatima analize, dobivena je značajna razlika u krivulji snage vikendom i radnim danom.



Slika 5.15. Grupiranje dana: prikaz centroida grupe 0 s po jednim tjednom iz svake grupe



Slika 5.16. Grupiranje dana: prikaz centroida grupe 1 s po jednim tjednom iz svake grupe

6. Zaključak

Provedena je analiza potrošnje električne grijalice vode iz Laboratorija za odziv potrošnje na FER-u. Glavni cilj rada bio je odrediti reprezentativne obrasce tjedne potrošnje kako bi se olakšala njihova daljnja primjena u modelima za optimalno upravljanje uređajima pomoću nenadziranog strojnog učenja. U radu su uspoređeni rezultati dobiveni implementacijom algoritma K-sredina s euklidskom udaljenošću te dinamičkim savijanjem vremena.

Pri pripremi podataka pokazalo se da sustav mjerenja i pohrane podataka iz laboratorija mjeri izrazito točno (nema stršećih vrijednosti), ali i da pohrana podataka ne funkcionira savršeno, o čemu svjedoči povelik broj nedostajućih podataka. Analiza podataka pokazala je da postoji određena dnevna periodičnost u podacima, ali i da sezonalnost na godišnjoj razini ne postoji te da su podaci o potrošnji stacionarni. Uočena je visoka korelacija između pojedinih značajki poput snage i struje, što je i očekivano jer proizlaze jedna iz druge.

Iako algoritam K-sredina s euklidskom udaljenošću ima svoje primjene, verzija koja koristi dinamičko savijanje vremena pokazala se uvjerljivo boljom u određivanju obrasca potrošnje, odnosno grupiranju krivulja vremenske serije. Tjedni su njome podijeljeni u dvije grupe - radne i neradne - ovisno o frekvenciji rada uređaja. Za svaku je grupu određen pripadni reprezentativni tjedan. Isti algoritam, primijenjen na dane, dane dijeli u grupe radnih i neradnih, od kojih svaka ima svog predstavnika.

Literatura

- [1] M. A. Fotouhi Ghazvini, J. Soares, O. Abrishambaf, R. Castro, i Z. Vale, “Demand response implementation in smart households”, *Energy and Buildings*, sv. 143, str. 129–148, 2017. <https://doi.org/https://doi.org/10.1016/j.enbuild.2017.03.020>
- [2] M. Albadi i E. El-Saadany, “A summary of demand response in electricity markets”, *Electric Power Systems Research*, sv. 78, br. 11, 2008. <https://doi.org/https://doi.org/10.1016/j.epsr.2008.04.002>
- [3] J. Šnajder, “Strojno učenje 1 - predavanja”, UNIZG FER, ak. god. 2022./2023.
- [4] G. Kirchgässner i J. Wolters, *Introduction to Modern Time Series Analysis*. Springer, 2007.
- [5] Similarity analysis of the patterns of the monthly electric energy demand time series - scientific figure on researchgate. [Mrežno]. Adresa: https://www.researchgate.net/figure/Additive-decomposition-of-time-series-PL_fig4_331018191
- [6] (2013. – 2024.) Hrvatska enciklopedija, mrežno izdanje. [Mrežno]. Adresa: <https://www.enciklopedija.hr/clanak/korelacija>
- [7] L. Ge i S. Chen, “Exact dynamic time warping calculation for weak sparse time series”, *Applied Soft Computing*, sv. 96, 2020. <https://doi.org/https://doi.org/10.1016/j.asoc.2020.106631>
- [8] M. Cvetković, “Završni rad”, 2023./2024. [Mrežno]. Adresa: <https://github.com/mateacvetkovic/Zavrsni-rad.git>
- [9] J. Bubalo, “Prediktivni modeli potrošnje električne energije”, 2024.

Sažetak

Određivanje obrazaca ponašanja potrošnje pametnih kućanskih uređaja

Matea Cvetković

Povećanje udjela obnovljivih izvora u proizvodnji električne energije za sobom povlači potrebu za sustavom koji omogućava upravljanje ravnotežom sustava sa strane potrošača. Kako bi se podaci potrošnje pojedinih uređaja jednostavnije koristili u modelima za odziv potrošnje, provedena je analiza podataka prikupljenih u Laboratoriju za odziv potrošnje Fakulteta elektrotehnike i računarstva. Korištenjem nenadziranog strojnog učenja: algoritma K sredina s dinamičkim savijanjem vremena, određeni su karakteristični tjedni obrasci potrošnje uređaja.

Ključne riječi: vremenske serije, grupiranje, dinamičko savijanje vremena, analiza vremenskih serija, algoritam K-sredina, strojno učenje

Determining Consumption Patterns of Smart Household Devices

Abstract

Određivanje obrazaca ponašanja potrošnje pametnih kućanskih uređaja

Matea Cvetković

Increasing number of renewable energy resources in power system requires the deployment of a system that shifts the responsibility of system balancing to consumers side. In order to simplify the use of consumption data from individual devices in demand response models, an analysis of data collected in the Demand Response Laboratory at the Faculty of Electrical Engineering and Computing was conducted. Characteristic weekly consumption patterns were identified using unsupervised machine learning: K-means algorithm with dynamic time warping.

Keywords: time series, clustering, dynamic time warping (DTW), time series analysis, K-means algorithm, machine learning