

Prediktivni modeli potrošnje električne energije

Bubalo, Josip

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://um.nsk.hr/um:nbn:hr:168:334332>

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. 1622

**PREDIKTIVNI MODELI POTROŠNJE ELEKTRIČNE
ENERGIJE**

Josip Bubalo

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1622

**PREDIKTIVNI MODELI POTROŠNJE ELEKTRIČNE
ENERGIJE**

Josip Bubalo

Zagreb, lipanj 2024.

ZAVRŠNI ZADATAK br. 1622

Pristupnik: **Josip Bubalo (0036532007)**
Studij: Elektrotehnika i informacijska tehnologija i Računarstvo
Modul: Računarstvo
Mentor: prof. dr. sc. Hrvoje Pandžić

Zadatak: **Prediktivni modeli potrošnje električne energije**

Opis zadatka:

Istraživanje se usredotočuje na analizu potrošnje u kućanstvima, koristeći stvarne podatke o potrošnji kućanskih uređaja prikupljene u laboratoriju za odziv potrošnje na Fakultetu elektrotehnike i računarstva (DRLab). Posebna pažnja posvećena je analizi potrošnje termostatski upravljivih uređaja, identificiranju odstupanja u podacima i praznina u podacima te razmatranju načina za ispravljanje ili popunjavanje nedostajućih podataka. Glavni cilj istraživanja je razviti prediktivne modele potrošnje tih uređaja kako bi se olakšalo planiranje i upravljanje elektroenergetskim sustavom, istovremeno smanjujući troškove električne energije u kućanstvima.

Rok za predaju rada: 14. lipnja 2024.

SADRŽAJ

1. Uvod	1
2. Analiza podataka	2
2.1. Prikupljanje i spremanje podataka	2
2.2. Otkrivanje i uklanjanje stršćih vrijednosti	2
2.3. Analiza i popunjavanje rupa u podacima	4
2.3.1. Definiranje i klasifikacija rupa u podacima	4
2.3.2. Strategije popunjavanja rupa u podacima	7
2.4. Otkivanje trenda, sezonalnosti, šuma i autokorelacije u podacima . . .	10
3. Prediktivni modeli potrošnje električne energije	13
3.1. Baseline model	13
3.2. Modeli linearne regresije i slučajnih šuma	15
3.3. Točnosti prediktivnih modela strojnog učenja	20
4. Zaključak	22
Literatura	23

1. Uvod

Ovo istraživanje bavi se analizom i izradom prediktivnih modela potrošnje električne energije u kućanstvima, koristeći stvarne podatke prikupljene u laboratoriju za odziv potrošnje na Fakultetu elektrotehnike i računarstva (DRLab). Poseban naglasak stavljen je na analizu podataka i izradu prediktivnih modela potrošnje električne energije za električni grijač vode.

Izradom i korištenjem prediktivnih modela strojnog učenja nastoji se omogućiti lakše planiranje i upravljanje uređajima i istovremeno smanjiti troškove potrošnje električne energije u kućanstvu na temelju napravljenih predikcija. Odgovor na potražnju kao strategija upravljanja električnom energijom potiče na smanjenje i prilagodbu potrošnje električne energije kao odgovor na promjene u cijeni električne energije na tržištu. Korištenjem prediktivnih modela za predviđanje kretanja cijene električne energije na tržištu i potrošnje električne energije uređaja u kućanstvu bi moglo dovesti do značajnih ušteda. Osim samo financijskih benefita ovakav način upravljanja uređajima u kućanstvu povećava energetska učinkovitost, doprinosi stabilnosti elektroenergetske mreže i ima ekološke benefite.

Glavna ideja je prikupiti podatke o potrošnji iz baze i analizom tih podataka otkriti moguća odstupanja i praznine te razviti strategiju otklanjanja tih odstupanja i popunjavanja praznina u podacima kako bi se kasnije ti podatci o potrošnji koristili za izradu prediktivnih modela.

2. Analiza podataka

Za analizu i obradu podataka korišten je PyCharm kao razvojno okruženje Python programskog jezika unutar Jupyter Notebooka. Uz to za razvoj su korištene biblioteke Pandas (1), NumPy (2), SciPy (3), Sklearn (4), MathPlotLib (5), StatsModels i Date-Time

2.1. Prikupljanje i spremanje podataka

Mjerenja potrošnje 25 uređaja iz laboratorija za odziv potrošnje prikupljana su i spremljena u bazi podataka. U bazi podataka spremljeni su podatci od 4. do 12. mjeseca 2022. godine i podatci za cijelu 2023. godinu. Baza podataka sastoji se od jedne tablice s 18 stupaca (id, time, energy, total_active_power, demand_total_active_power, maximum_demand_total_active_power, i1, i2, i3, u12, u23, u13, v1, v2, v3, device, total_power_factor, total_apparent_power), a za analizu i predviđanje relevantan je stupac korišten total_active_power koji predstavlja mjerenja potrošnje uređaja. U ovom radu obrađeni su podatci za električni grijač vode.

Spremljeni podatci mogu sadržavati odstupanja od očekivanih vrijednosti kao rezultat pogrešaka u mjerenju ili spremanju podataka, a u ponekim trenutcima može se dogoditi i da mjerenja nisu zabilježena u bazi podataka. Stoga, potrebno je otkriti moguća odstupanja i rupe u podacima, odnosno vremenske periode u kojima nema spremljenih podataka, a trebalo bi ih biti.

2.2. Otkrivanje i uklanjanje stršćih vrijednosti

Isprva, za otkrivanje stršćih vrijednosti korištena je metoda temeljena na z-scoreu. (6). Z-score je statistička mjera koja pokazuje vezu između vrijednosti za koju se provjerava je li stršćea i prosjeka svih vrijednosti u podacima. Vrijednost z-scorea pokazuje za koliko je standardnih devijacija razmatrana vrijednost udaljena od srednje vrijednosti skupa svih podataka. Ako se neka vrijednost provjerava i njezin z-score je

0, to znači da je ta vrijednost jednaka prosjeku svih podataka. U slučaju da je vrijednost z-scorea jednaka 1.0, provjeravana vrijednost je za jednu standardnu devijaciju udaljenja od prosjeka. Z-score također može biti i negativan ako je vrijednost manja od prosjeka, dok je pozitivan ako je vrijednost veća od prosjeka.

Za izračun z-scorea koristi se formula u kojoj se od vrijednosti koja se provjerava oduzima srednja vrijednost svih podataka te se taj rezultat potom dijeli sa standardnom devijacijom.

$$z = \frac{x - \mu}{\sigma} \quad (2.1)$$

Za detekciju stršćeih vrijednosti u podacima o potrošnji električnog grijača vode, korištena je biblioteka `scipy.stats` i njena metoda za izračun z-scorea. Metoda kao ulazni podatak prima niz te za svaki element iz toga niza računa pripadnu vrijednost z-scorea.

```
z_scores = stats.zscore(df['total_active_power'])
outliers = (z_scores > 4.25) | (z_scores < -0.5)
```

Potrošnja električnog grijača vode većinu vremena iznosi 0 ili je u rasponu između 1400 i 1600 W. Ako je vrijednost snage očitana baš u trenucima paljenja ili gašenja grijača, vrijednost snage može se nalaziti u intervalu $\langle 0, 1400 \rangle$.

Ova metoda rezultira detekcijom većeg broja stršćeih vrijednosti, koje u stvarnosti ne bi svrstali u tu skupinu. Električni grijač vode većinu je vremena isključen te je srednja vrijednost njegove potrošnje vrlo blizu nule. Korištenje ove metode svrstava svaki ciklus u kojem grijač radi kao potencijalnu stršćeu vrijednost, zato što su vrijednosti potrošnje u tim trenucima daleko od prosjeka potrošnje uređaja. Kako potrošnja električnog grijača vode ne prati normalnu distribuciju ovakav način otkrivanja stršćeih vrijednosti nije dobar.

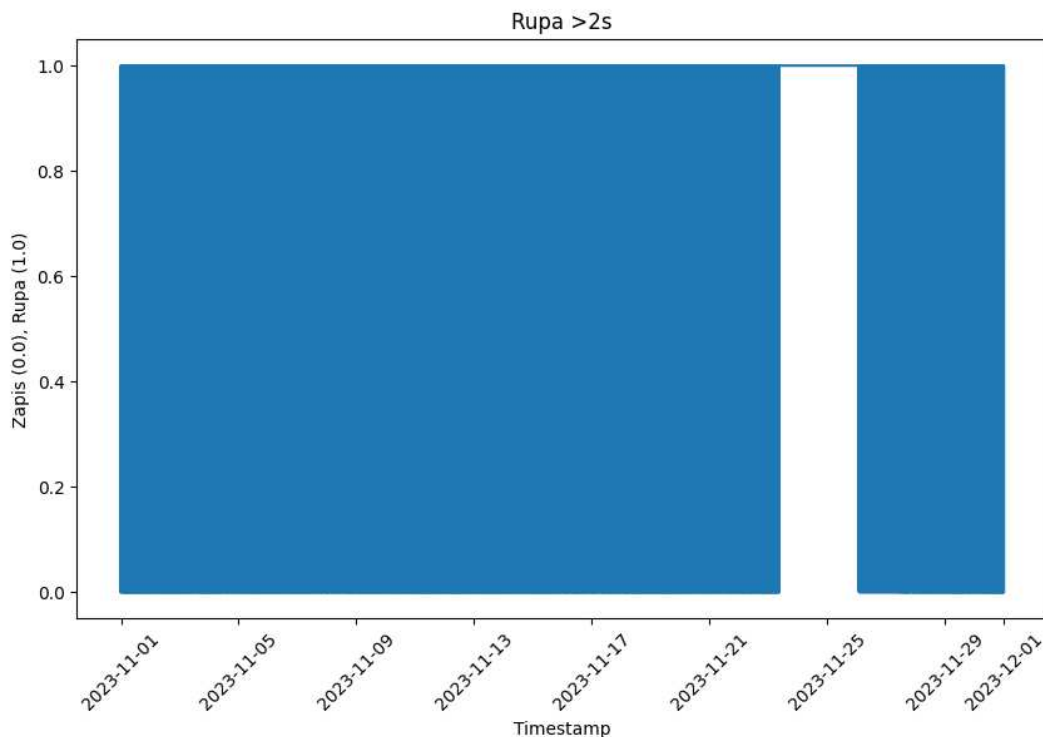
Iz tog razloga korištena je alternativna metoda otkrivanja stršćeih vrijednosti u kojoj se isključivo provjerava je li potrošnja je li potrošnja negativna, što je fizikalno nemoguće te je li potrošnja veća od 1700 W. Ovom metodom nisu detektirane stršćeće vrijednosti te je zaključeno kako su sva mjerenja u ispravnom rasponu.

2.3. Analiza i popunjavanje rupa u podacima

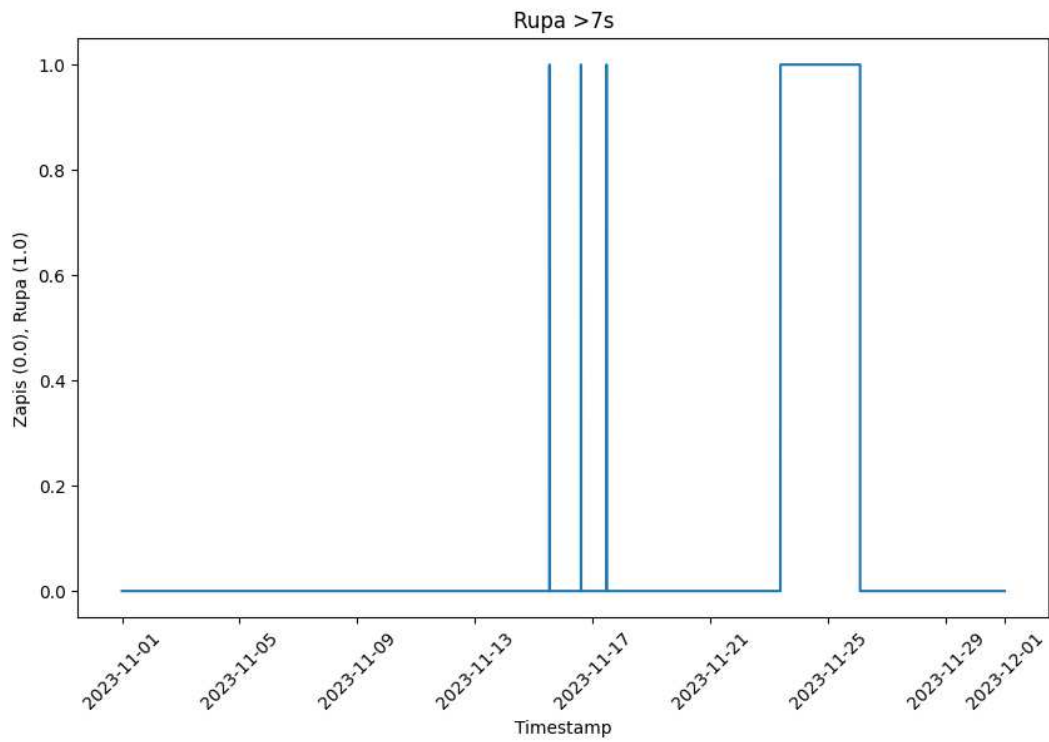
2.3.1. Definiranje i klasifikacija rupa u podacima

Zapisi o potrošnji nisu spremeni u bazu podataka s fiksnom frekvencijom pohrane, zbog čega je nejasno koliki vremenski period bi trebao proći bez pohrane podataka da bi se smatralo da je došlo do pogreške u spremanju. No, većina zapisa u bazi spremljena je između 2 i 3 sekunde od prethodno zaprimljenog zapisa, što sugerira da bi se u slučaju da podatka nema u iduće 2 ili 3 sekunde radilo o grešci. Stoga, najprije je bilo potrebno definirati taj period, a potom i kreirati strategije popunjavanja nedostajajućih zapisa. Kako bi se kreirale čim efikasnije strategije, bilo je potrebno i klasificirati takve zapise po dužini trajanja.

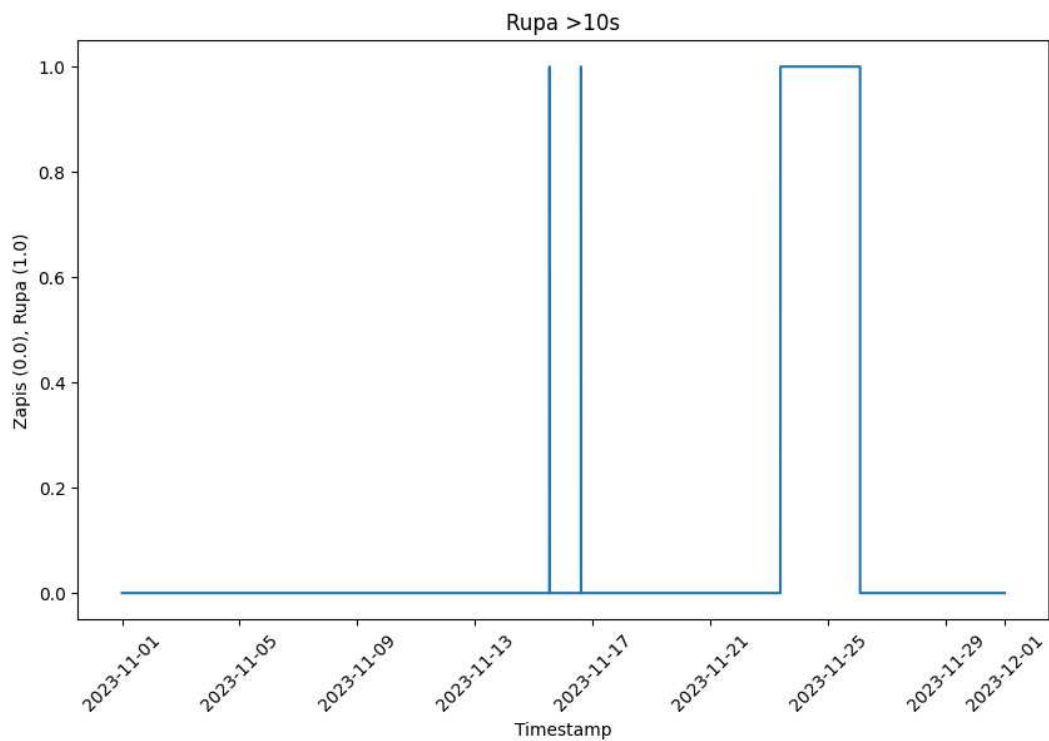
S ciljem određivanja optimalne frekvencije zapisa, pretpostavljeno je nekoliko mogućih frekvencija iz skupa {2,3,5,7,10} sekundi (slike 2.1, 2.2, 2.3, 2.4, 2.5). Za svaku moguću frekvenciju zapisa određen je broj zapisa koji nedostaju, npr. ako je pretpostavljena frekvencija zapisa 10 sekundi, broji se koliko je puta idući zapis u bazi bio tek nakon više od 10 sekundi.



Slika 2.1: Prikaz rupa u podacima većih od 2 sekunde (11. mjesec 2023.)



Slika 2.4: Prikaz rupa u podacima većih od 7 sekundi (11. mjesec 2023.)



Slika 2.5: Prikaz rupa u podacima većih od 10 sekundi (11. mjesec 2023.)

Veličina rupe / sekunde	Postotak rupa / %
>2	15.701
>3	6.454
>5	6.453
>7	6.453
>10	6.452

Slika 2.6: Prikaz postotka rupa u odnosu na zapise (11. mjesec 2023.)

Nakon što je napravljena analiza nad svim dostupnim podacima uočeno je da je gotovo svaki put najveća razlika u broju podataka koji nedostaju između 2 i 3 sekunde dok su za preostale moguće frekvencije razlike vrlo male (slika 2.6). Donesen je zaključak da je očekivana frekvencija zapisa mjerenja u bazu 3 sekunde (zapis mora doći unutar 3 sekunde) te se u slučaju pohrane zapisa nakon više od 3 sekunde od prethodnog može govoriti o grešci u pohrani podataka i smatrati da imamo 'rupu' u podacima koju je potrebno popuniti.

Nadalje, promatranjem količine vremenskih perioda u kojima nedostaju zapisi, 'rupe' su klasificirane u tri različite skupine:

1. Male rupe koje su u intervalu od 3 sekunde do 30 sekundi (<3s, 30s]
2. Srednje rupe koje su u intervalu od 30 sekundi do 1 sata (<30s, 1h]
3. Velike rupe koje su veće od jednog sata

Za svaku od tih skupina rupa razvijena je posebna strategija popunjavanja.

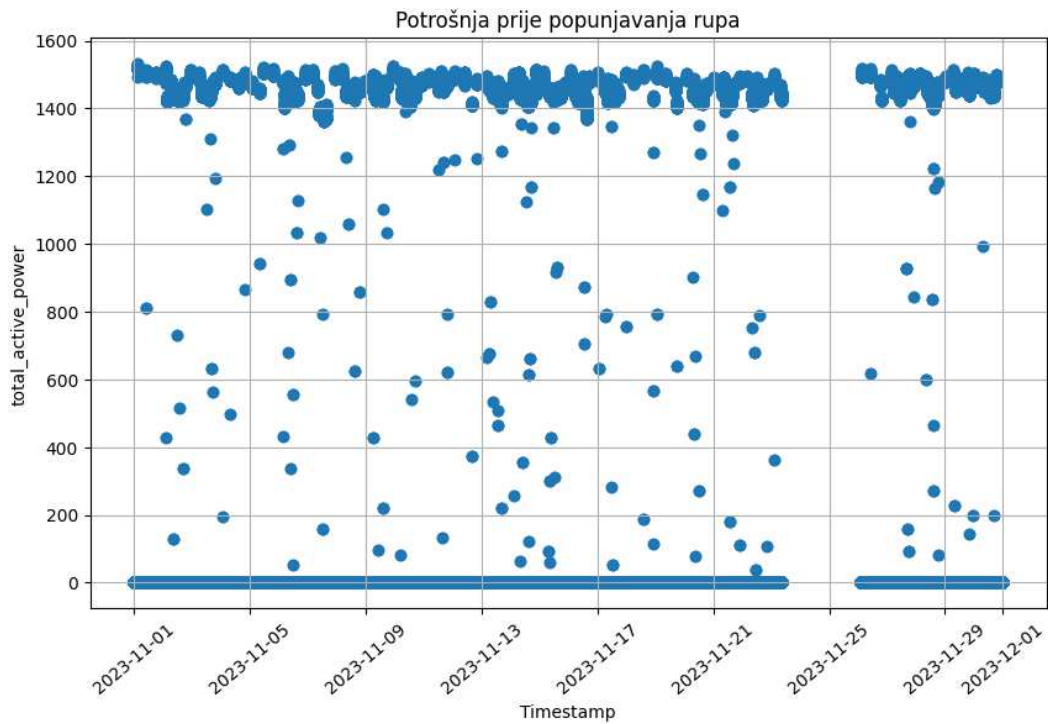
2.3.2. Strategije popunjavanja rupa u podacima

U ovisnosti kategoriji kojoj pripadaju, postoje različite strategije popunjavanja rupa. Definirane su strategije popunjavanja podataka koji nedostaju: jedna za male rupe, a druga za srednje i velike rupe (slike 2.7 i 2.8).

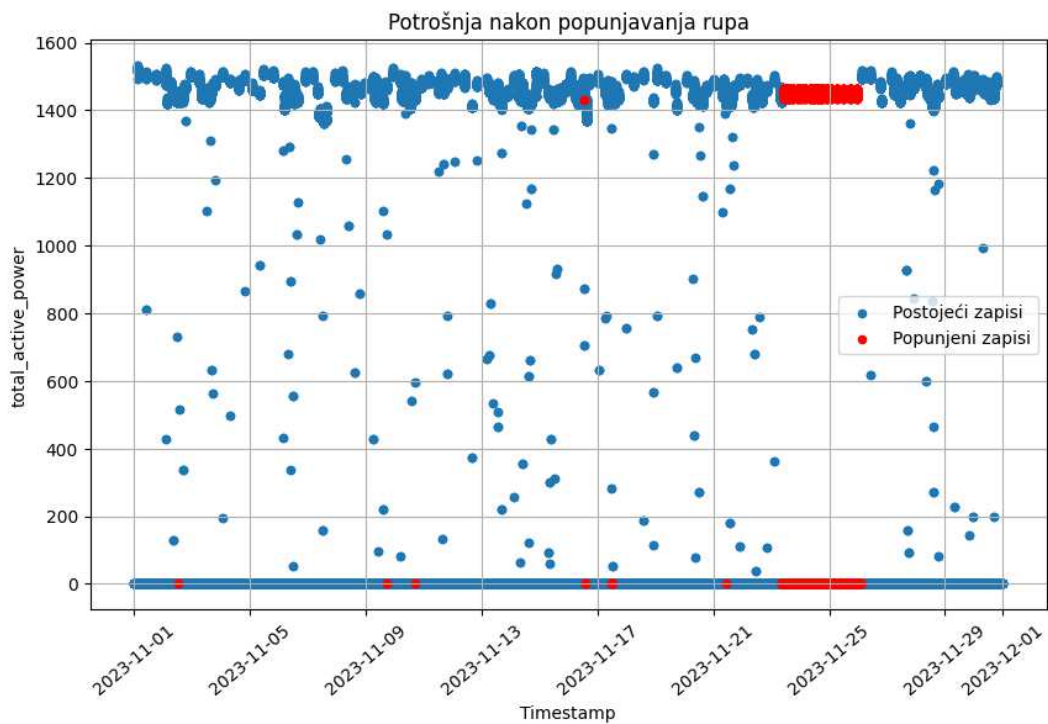
Za male rupe razvijena je strategija kojom se rupe popunjavaju interpolacijom, za što se koristila metoda `interpolate()` iz Pandas biblioteke. Interpolacija označava matematičku metodu konstrukcije novih točaka unutar raspona poznatih točaka podataka. Metoda `interpolate()` služi za popunjavanje NaN vrijednosti nekog skupa podataka korištenjem raznih interpolacijskih tehnika. Za popunjavanje malih rupa u podacima električnog grijača vode korištena je polinomijalna interpolacija (slika 2.9).

```
mask = (df['total_active_power'].isna()) & (df['time_diff'] <= 30)
df.loc[mask, 'total_active_power'] = df['total_active_power']
.interpolate(method='polynomial', limit_area='inside', order=2)
```

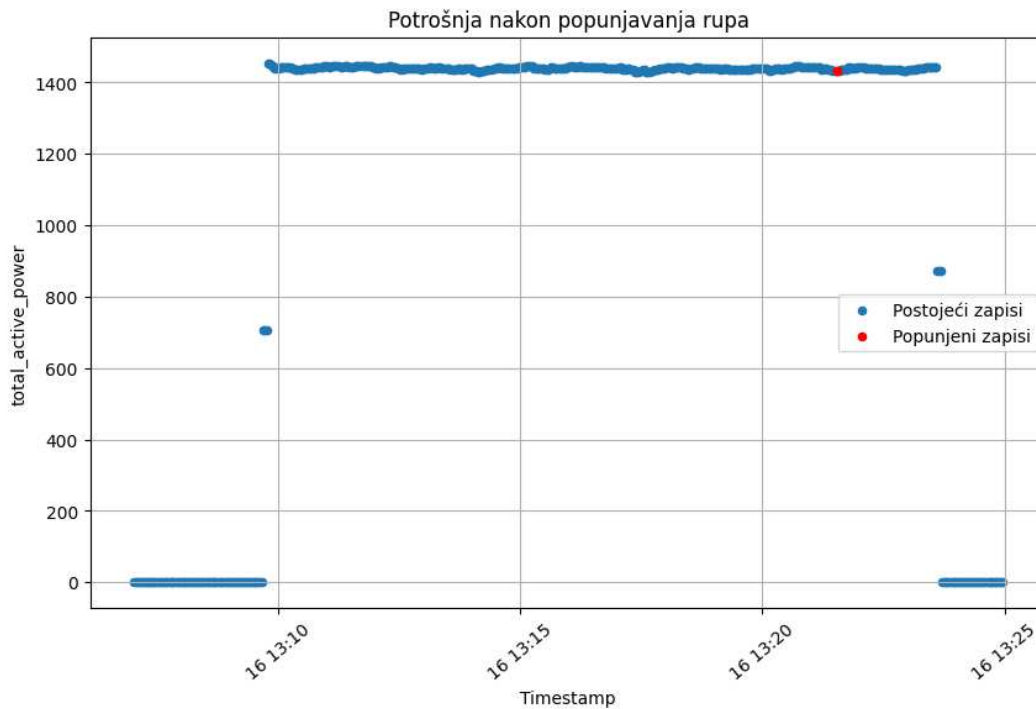
Zbog svog trajanja i zbog prosječnog trajanja rada električnog grijača (9 minuta 45 sekundi), za srednje i velike grupe trebalo je osmisliti drukčiju strategiju popunjavanja, zato što bi spomenuta strategija rezultirala fizikalno neispravnim podacima. Pri popunjavanju srednjih i velikih rupa uzima se prosječno trajanje rada, prosječnu potrošnju za vrijeme rada i prosječno vrijeme između rada električnog grijača vode te na temelju tih vrijednosti popunjava rupe. Kako se prosječne vrijednosti potrošnje razlikuju u ovisnosti o danu i dobu dana, pri popunjavanju je uzeto u obzir o kojem se vremenskom trenutku radi. Za vrijeme radnog dana između 7 i 17 sati električni grijač vode radi češće nego vikendima ili neradnim satima zbog puno češćeg korištenja tople vode. Algoritam provjerava kad je električni grijač vode zadnji put bio uključen te razlikuje dva slučaja. U prvom slučaju rupa se pojavila kad je električni grijač vode bio isključen te se provjerava koliko dugo je već isključen, ako je duže od prosječnog vremena između ciklusa rada onda se uključuje odmah, dok ako je kraće od tog vremena ostaje isključen dok ne dosegne prosječno vrijeme između dva ciklusa rada. Drugi slučaj je da je električni grijač vode radio kad se pojavila rupa, te tada nastavlja s radom dok ne dosegne prosječno vrijeme rada. Ciklus rada se popunjava prosječnom vrijednosti potrošnje grijača s dodanim nasumičnim pomakom od 5%.



Slika 2.7: Prikaz potrošnje prije popunjavanja rupa (11. mjesec 2023.)



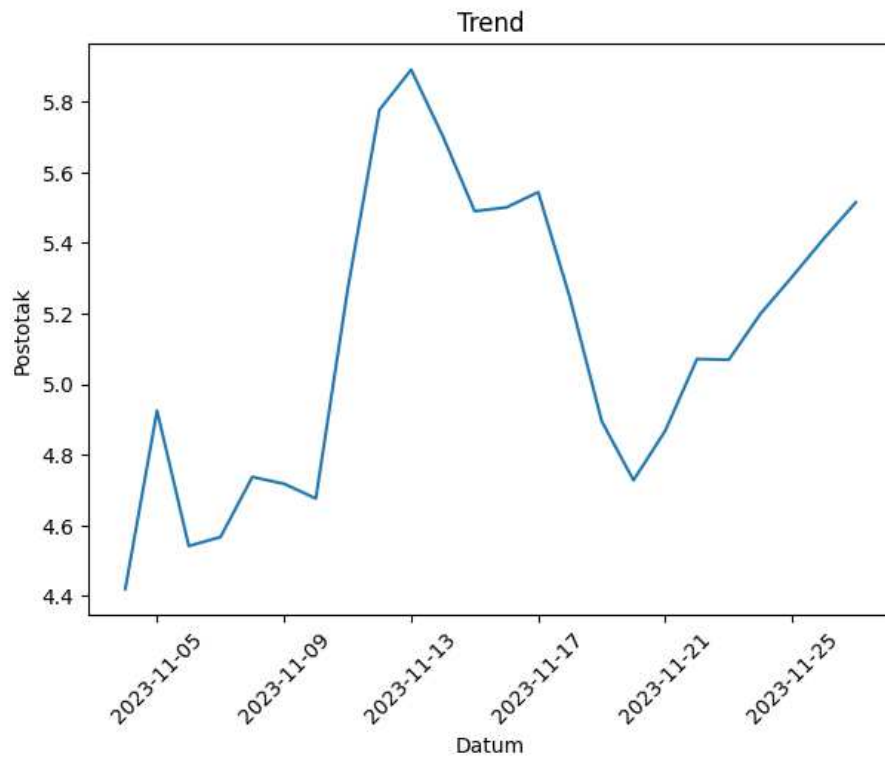
Slika 2.8: Prikaz potrošnje nakon popunjavanja rupa (11. mjesec 2023.)



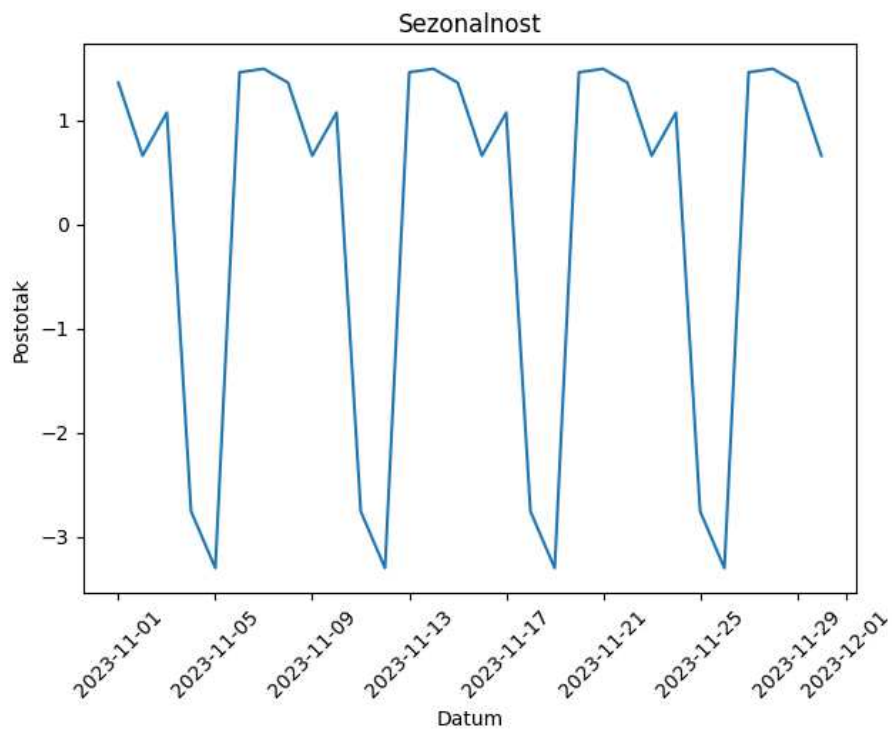
Slika 2.9: Prikaz popunjavanja kratke rupe (11. mjesec 2023.)

2.4. Otkivanje trenda, sezonalnosti, šuma i autokorelacije u podacima

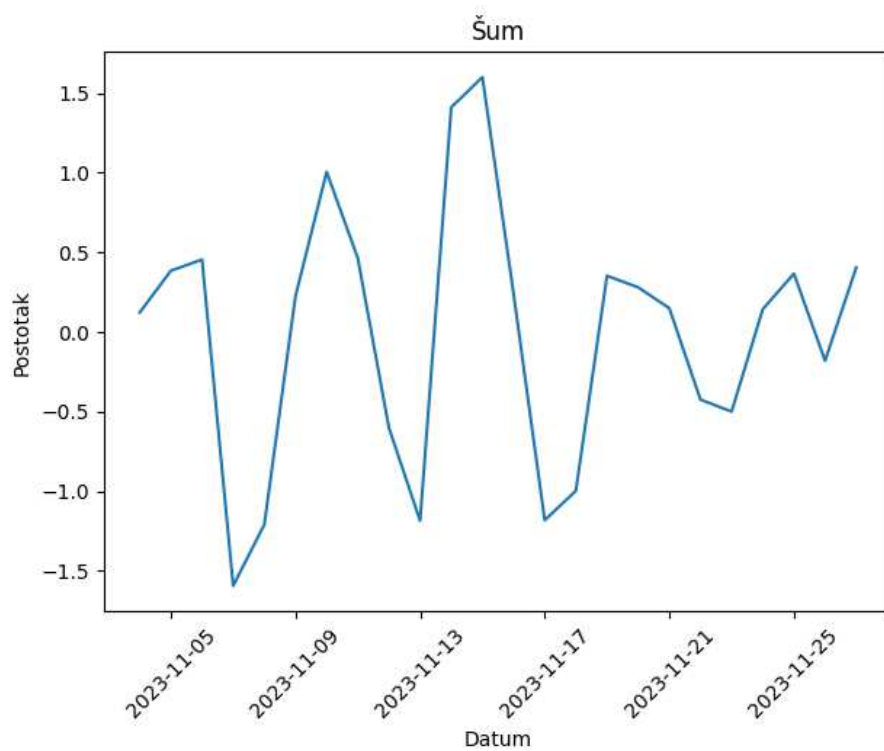
Vrijednosti trenda, sezonalnosti, šuma i autokorelacije u podacima o potrošnji mogli bi pomoći kod predviđanja ponašanja električnog grijača vode i potrošnje električne energije. Za otkrivanje vrijednosti trenda, sezonalnosti i šuma u podacima korištena je metoda `seasonal_decompose()` koja je dio `StatsModels` biblioteke (7). Nakon pokretanja te metode dobiven je rezultat iz kojeg je vidljiv grafovi za trend, sezonalnost i šum (8) (slike 2.10, 2.11, 2.12). Iz dobivenih rezultata uočena je sezonalnost tijekom svakog tjedna unutar jednog mjeseca podataka kao pravilnost koja bi mogla utjecati na predikciju potrošnje. Također porvjerena je i vrijednost autokorelacije između podataka o potrošnji sa odmakom od jednog dana (slika 2.13). Taj podatak koristan je kako bi se uočilo imali smisla koristiti podatke od prijašnjeg dana za predviđanje potrošnje. Graf autokorelacije dobiven je metodom `plot_acf()` koja je također dio `StatsModels` biblioteke.



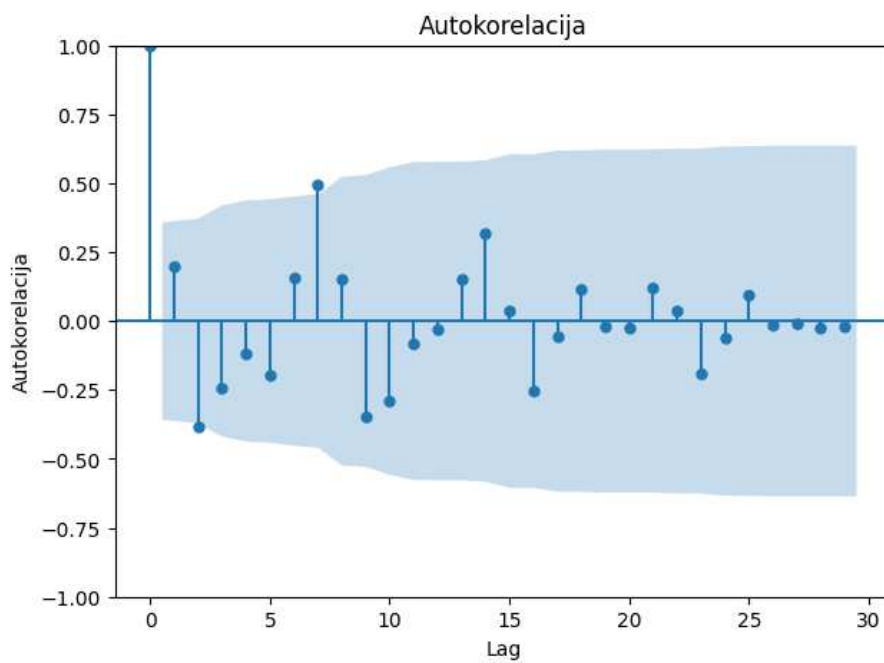
Slika 2.10: Trend u podatcima potrošnje (11. mjesec 2023.)



Slika 2.11: Sezonalnost u podatcima potrošnje (11. mjesec 2023.)



Slika 2.12: Šum u podacima potrošnje (11. mjesec 2023.)



Slika 2.13: Autokorelacija u podacima potrošnje (11. mjesec 2023.)

3. Prediktivni modeli potrošnje električne energije

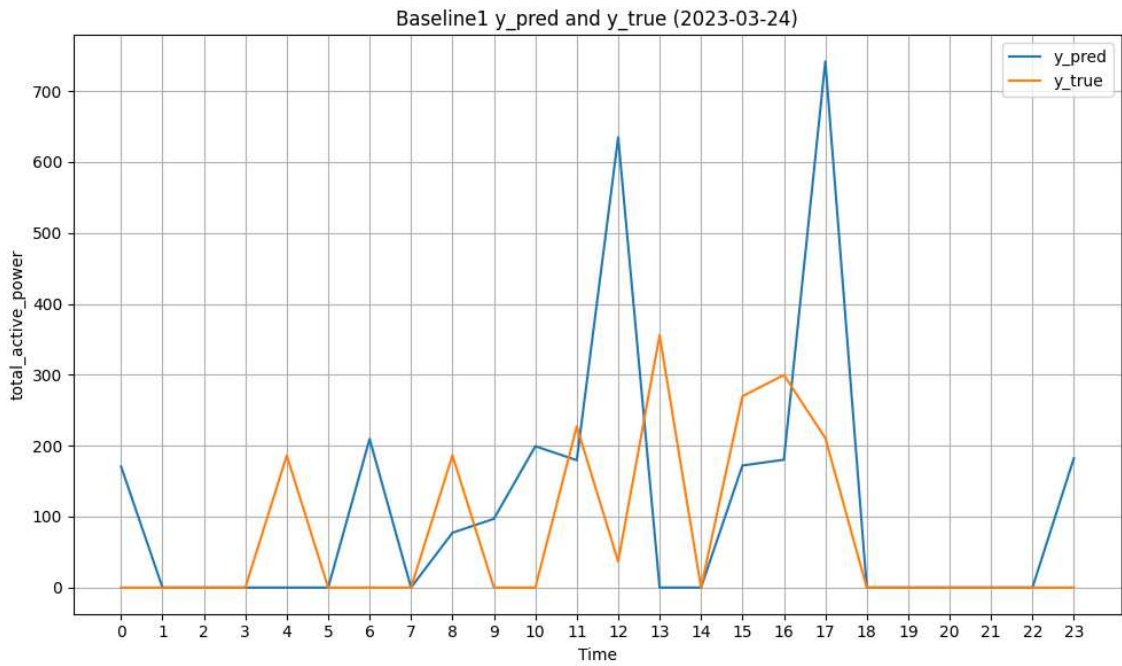
Nakon napravljene analize podataka, potrebno je izmijeniti rezoluciju podataka na jednosatnu prije nego je moguće kreirati i trenirati modele za predviđanje potrošnje energije električnog grijača vode u kućanstvu. To je moguće napraviti metodom `resample()` koja je dio Pandas biblioteke.

```
df_hourly = df.resample('H', on='time').mean()
```

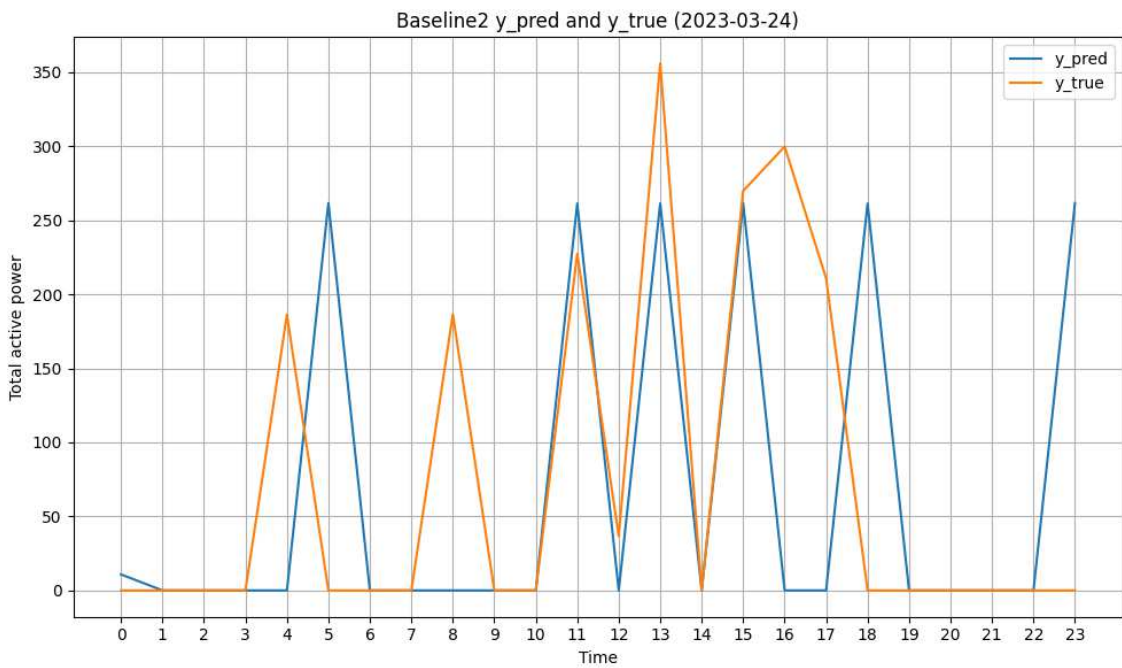
Model bi trebao davati predikciju potrošnje za jedan dan unaprijed na temelju svih prikupljenih podataka osim podataka od dana kad je pokrenut model. Za predviđanje potrošnje kreirani su baseline modeli za usporedbu performansi te modeli temeljeni na linearnoj regresiji i algoritmu slučajnih šuma.

3.1. Baseline model

Baseline modeli su jednostavni modeli koji se koriste kao početna točka za razvoj složenijih modela. Oni su mjerilo prema kojem se mogu analizirati i ocjenjivati kompleksniji modeli. U ovom radu kreirana su dva baseline modela. Prvi model predviđa potrošnju tako da samo uzima podatke od dana prije nego što je pokrenut model te te podatke ispisuje kao predikciju za potrošnju idućeg dana (slika 3.1). Drugi model radi na sličnom principu, no on, umjesto da uzima podatke od dana prije, uzima podatke od dana iz prošlog tjedana koji odgovara danu za koji se želi predvidjeti potrošnja (slika 3.2).



Slika 3.1: Predikcije jednog dana prvog baseline modela (24.03.2023.)



Slika 3.2: Predikcije jednog dana drugog baseline modela (24.03.2023.)

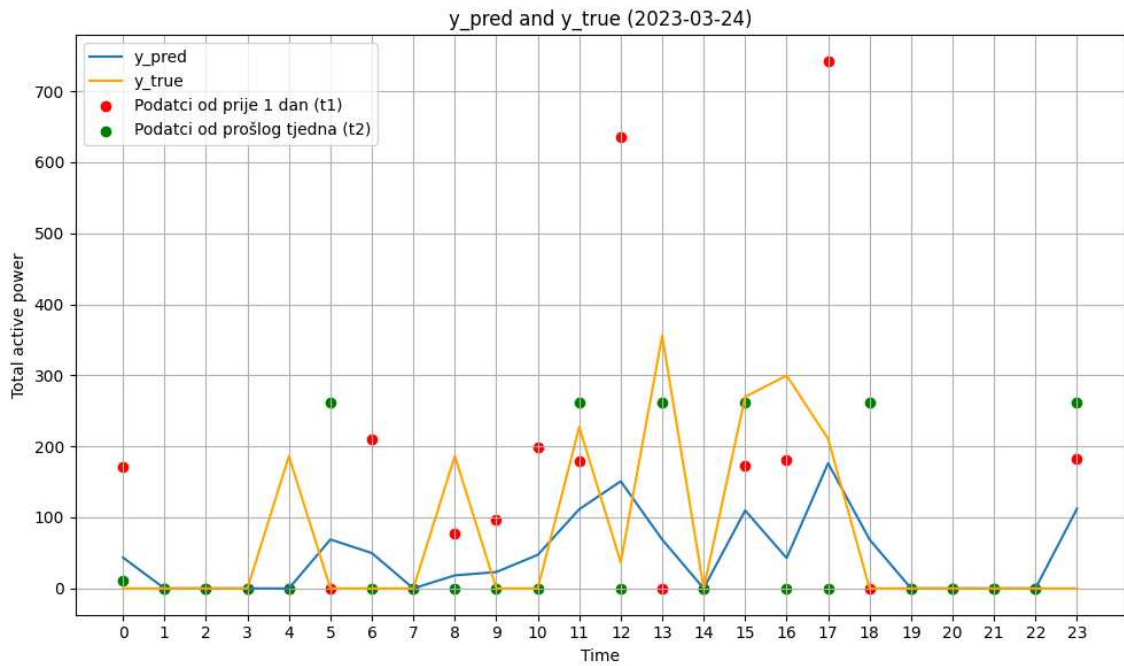
3.2. Modeli linearne regresije i slučajnih šuma

Linearna regresija (9) i algoritam slučajnih šuma (10) prediktivni su modeli strojnog učenja koji uzimaju jednu ili više ulaznih varijabli, odnosno značajki, te u ovisnosti o tim značajkama izračunavaju predikciju. Razlika između navedenih algoritama je u tome što linearna regresija povezuje ulazne varijable s ciljom kroz matematičke funkcije, dok algoritam slučajnih šuma kombinira izlaze većeg broja stabla odluke u jedan izlaz koji predstavlja ciljnu varijablu.

Implementirano je pet modela linearne regresije s pet različitih ulaza i jedan model algoritama slučajnih šuma. Za implementaciju obe vrste modela potrebno je odrediti ulazne značajke tih modela i razdvojiti podatke na skup za treniranje i skup za testiranje. Za razdvajanje ta dva skupa postoji metoda `train_test_split()` koja je također dio Sklearn biblioteke. Podatci su podijeljeni tako da se 60% njih koristi za učenje modela, a 40% za testiranje modela. Za određivanje ulaznih značajki postoji funkcija koja kreira matricu dizajna i funkcija s kojom se kreira matrica očekivanih izlaza. Obe matice se sastoje od 24 reda, za svaki sat u danu jedan red podataka. Modeli su implementirani metodama `LinearRegression()` i `RandomForestRegressor()` koje su dio biblioteke Sklearn.

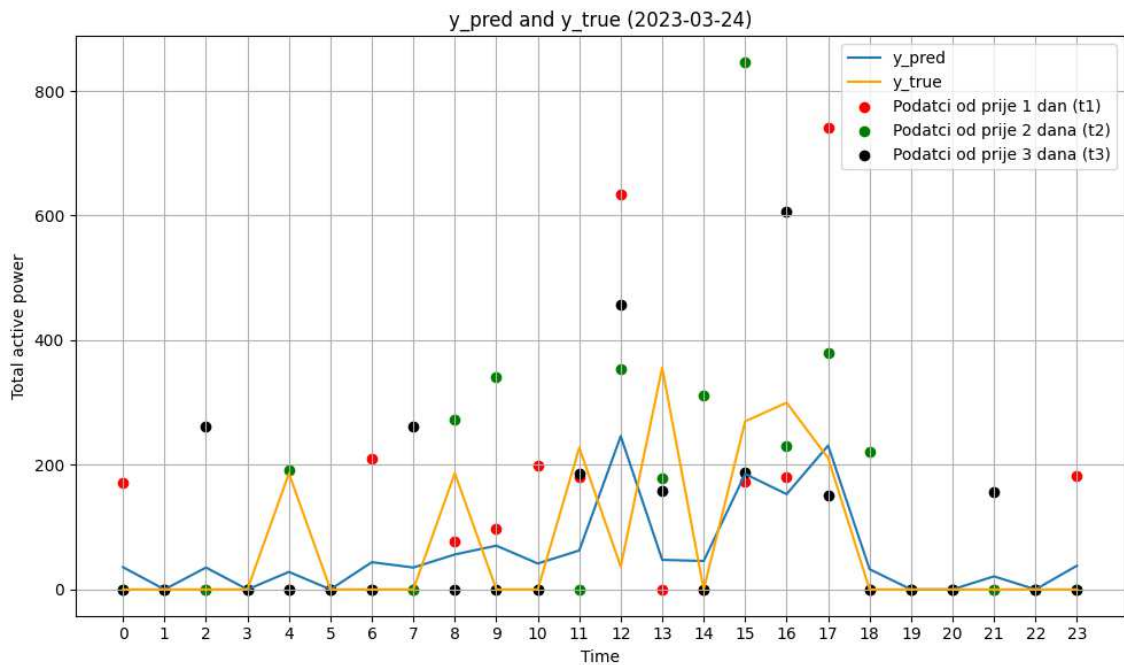
```
train_data, test_data = train_test_split(df, test_size=0.4,
shuffle=False)
```

Prvi model linearne regresije za ulazne značajke koristi podatke o potrošnji električnog bojlera u svakom satu od dana prije nego što je pokrenut model i dana od prošlog tjedna koji odgovara danu za koji se predviđa potrošnja (slika 3.3).



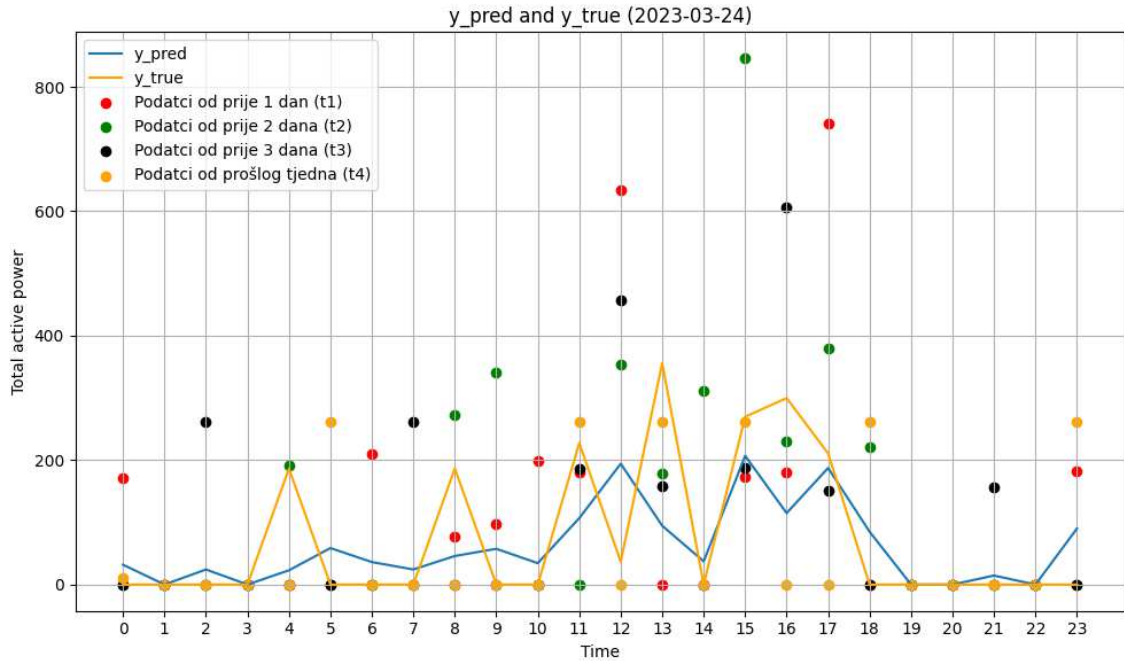
Slika 3.3: Predikcije jednog dana prvog modela linearne regresije (24.03.2023.)

Drugi model linearne regresije za ulazne značajke koristi podatke o potrošnji električnog bojlera od jednog do tri dana prije nego što je pokrenut model (slika 3.4).



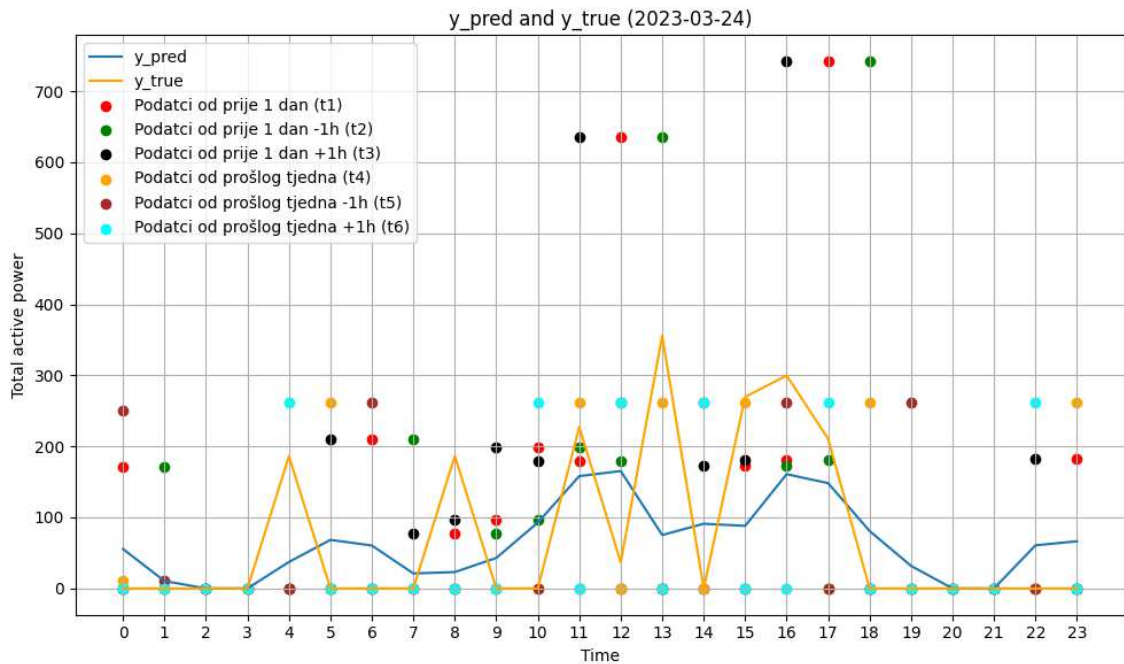
Slika 3.4: Predikcije jednog dana drugog modela linearne regresije (24.03.2023.)

Treći model linearne regresije za ulazne značajke koristi podatke o potrošnji električnog bojlera od jednog do tri dana prije nego što je pokrenut model te dana od prošlog tjedna koji odgovara danu za koji se predviđa potrošnja (slika 3.5).



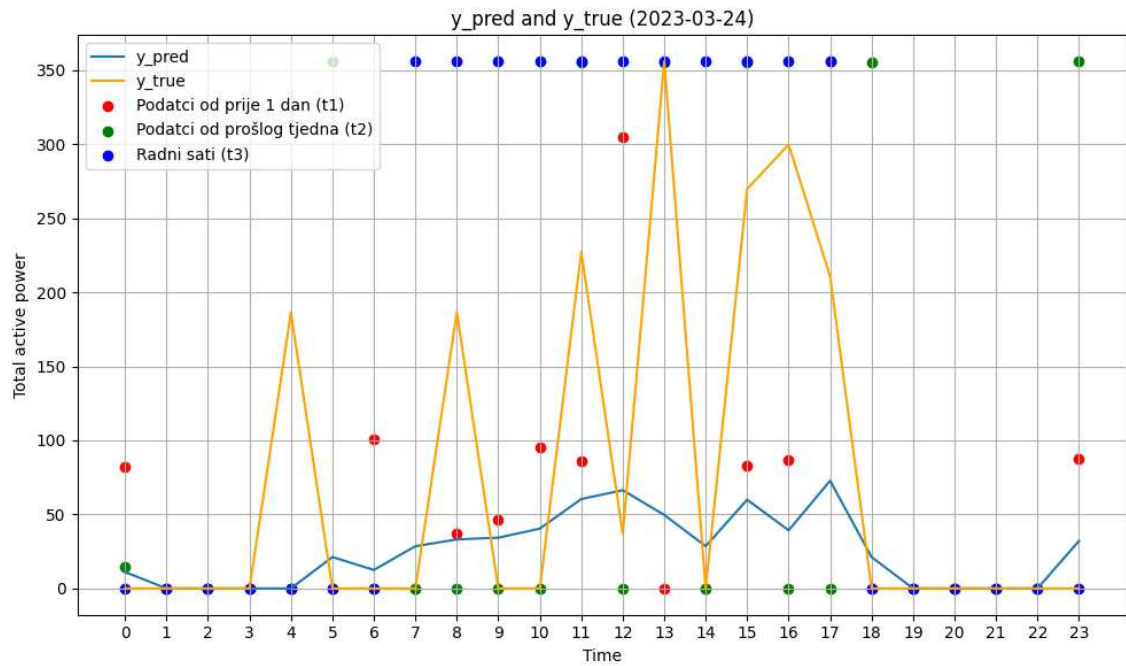
Slika 3.5: Predikcije jednog dana trećeg modela linearne regresije (24.03.2023.)

Četvrti model linearne regresije koristi jednake ulazne značajke kao i prvi model, ali s dodanim pomacima od jednog sata unaprijed i unazad, tako da ukupno ima šest značajki (slika 3.6).



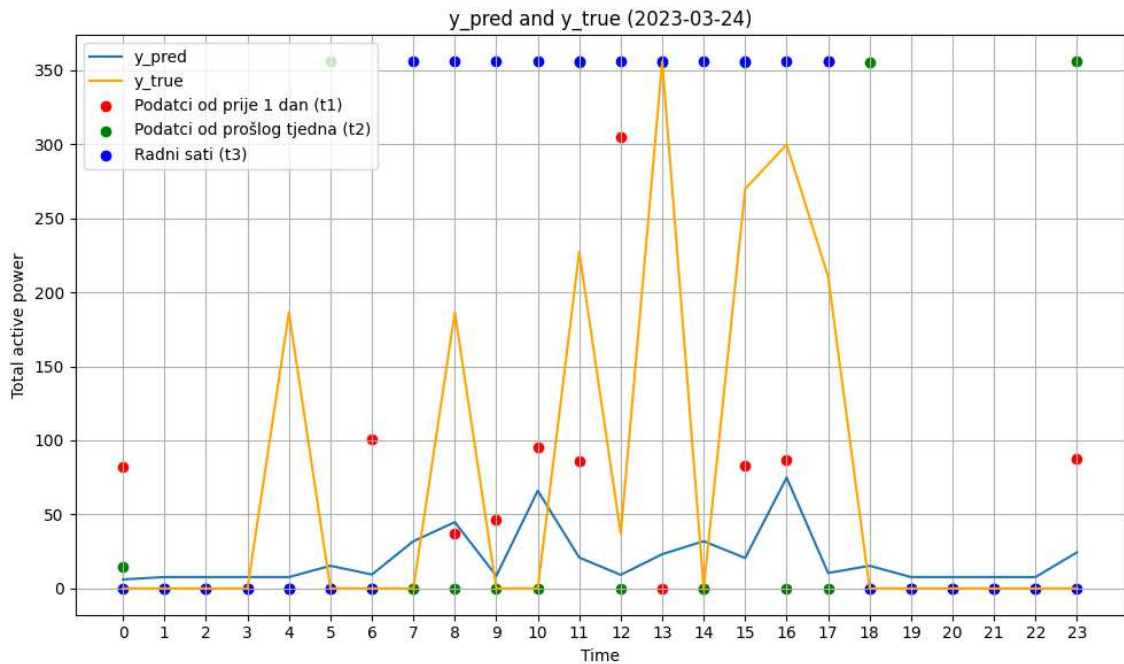
Slika 3.6: Predikcije jednog dana četvrtog modela linearne regresije (24.03.2023.)

Peti model linearne regresije za ulazne značajke koristi podatke o potrošnji električnog bojlera svakog sata od dana prije nego što je pokrenut model i dana od prošlog tjedna koji odgovara danu za koji se predviđa potrošnja. Ti podatci su normalizirani korištenjem `MinMaxScaler()` metode koja je dio `Sklearn` biblioteke. Dodana je još jedna značajka koja označava radno vrijeme, ta značajka ima vrijednost u 1 kada je period između 7 i 17 sati za vrijeme radnog dana, a inače je 0 (slika 3.7).



Slika 3.7: Predikcije jednog dana petog modela linearne regresije (24.03.2023.)

Model algoritma slučajne šume ima jednake ulazne značajke kao i peti model linearne regresije (slika 3.8).



Slika 3.8: Predikcije jednog dana modela slučajne šume (24.03.2023.)

3.3. Točnosti prediktivnih modela strojnog učenja

Kao pokazatelji točnosti modela izračunati su: prosječna kvadratna pogreška (engl. *Mean Squared Error*, MSE), korijen prosječne kvadratne pogreške (engl. *Root Mean Squared Error*, RMSE), prosječna apsolutna pogreška (*Mean Absolute Error*, MAE) i koeficijent determinacije (*Coefficient of determination*, R^2).

Prosječna kvadratna pogreška računa se kao prosjek kvadriranih razlika predviđenih vrijednosti i stvarnih vrijednosti. Programski je implementirana metodom `mean_squared_error()` koja kao parametre prima predviđene vrijednosti i stvarne vrijednosti te je dio Sklearn biblioteke.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (3.1)$$

Korijen prosječne kvadratne pogreške programski je implementiran metodom `root_mean_squared_error()` koja kao parametre prima predviđene vrijednosti i stvarne vrijednosti te je dio Sklearn biblioteke.

$$RMSE = \sqrt{MSE} \quad (3.2)$$

Prosječna apsolutna pogreška računa se kao prosjek apsolutne vrijednosti razlika predviđenih vrijednosti i stvarnih vrijednosti. Programski je implementirana metodom `mean_absolute_error()` koja kao parametre prima predviđene vrijednosti i stvarne vrijednosti te je dio Sklearn biblioteke.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (3.3)$$

Koeficijent determinacije je statistička mjera koja pokazuje koliko dobro graf dobiven regresijom prati stvarne vrijednosti. Programski je implementiran metodom `r2_score()` koja kao parametre prima predviđene vrijednosti i stvarne vrijednosti te je dio Sklearn biblioteke.

$$R^2 = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2} \quad (3.4)$$

Model	MSE	RMSE	MAE	R^2
Baseline1	24676.1219	157.0864	95.8028	-0.7885
Baseline2	21522.4830	146.7054	85.4919	-0.5599
Linearna regresija 1	14505.0204	120.4368	71.6779	-0.0549
Linearna regresija 2	15202.0084	123.2964	75.2325	-0.1056
Linearna regresija 3	13964.5419	118.1716	74.0338	-0.0156
Linearna regresija 4	13161.9619	114.7255	75.1400	0.04269
Linearna regresija 5	12170.2878	110.3190	81.6786	0.1148
Random forest	13377.5001	115.6611	80.1070	0.02701

Slika 3.9: Tablica točnosti svih modela

Najmanju prosječnu kvadratnu pogrešku ima peti model linearne regresije, najmanju apsolutnu pogrešku ima prvi model linearne regresije dok koeficijent determinacije najbliže nuli ima treći model linearne regresije (slika 3.9). Općenito kompleksniji modeli linearne regresije i algoritma slučajnih šuma imaju puno bolje rezultate točnosti od baseline modela.

4. Zaključak

Izrada ovoga rada bila je podijeljena u dva glavna djela: analiza i obrada podataka te izrada prediktivnih modela potrošnje električne energije. Obradeni su podatci o potrošnji električnog grijača vode no uz manje modifikacije i algoritam popunjavanja rupa u podacima i prediktivni modeli mogli bi se prilagoditi za sve ostale uređaje unutar laboratorija. Nakon obrade svih podataka izrađeni su baseline modeli, modeli linearne regresije i model algoritma slučajnih šuma. Usporedbom točnosti svih tih modela zaključeno je raditi predikcije bazirane na kompleksnijim algoritmima poput modela linearne regresije i algoritma slučajnih šuma jer daju puno bolje rezultate nego jednostavni baseline modeli.okus daljnjeg istraživanja bit će predikcije s kompleksnijim algoritmima, kao što su XGBoost te poboljšanjem algoritma slučajnih šuma kako bi se potencijalno moglo omogućiti lakše upravljanje uređajima u kućanstvu i smanjenje potrošnje.

LITERATURA

- [4] *Pandas documentation*, 2024. URL: <https://pandas.pydata.org/docs/>.
- [2] *NumPy documentation*, 2024. URL: <https://numpy.org/doc/1.26/>.
- [3] *SciPy documentation*, 2024. URL: <https://docs.scipy.org/doc/scipy/>.
- [4] *Sklearn documentation*, 2024. URL: <https://scikit-learn.org/stable/>.
- [5] *MathPlotLib documentation*, 2024. URL: <https://matplotlib.org/stable/users/index>.
- [6] Scott Nevil. *Z-Score: Meaning and Formula*. Investopedia, 2024. URL: <https://www.investopedia.com/terms/z/zscore.asp>.
- [7] *Statsmodels.tsa.seasonal.seasonal_decompose()*, 2024. URL: https://www.statsmodels.org/stable/generated/statsmodels.tsa.seasonal.seasonal_decompose.html.
- [8] Data Science Wizards. *Introduction to Time Series Analysis*. Medium, 2023. URL: <https://medium.com/@datasciencewizards/introduction-to-time-series-analysis-i-82d614d462b0>.
- [9] Jan Šnajder. *P03 Linearna regresija*. Fakultet elektrotehike i računarstva, 2023. URL: https://www.fer.unizg.hr/_download/repository/SU1-2022-P03-LinearnaRegresija.pdf.
- [10] IBM. *What is random forest?*, 2023. URL: <https://www.ibm.com/topics/random-forest>.

Prediktivni modeli potrošnje električne energije

Sažetak

Tema ovog rada je izrada prediktivnih modela strojnog učenja za potrošnju električne energije u kućanstvu. Nakon prikupljanja podataka nad njima je napravljena analiza u kojoj su otklonjene stršeće vrijednosti i popunjene rupe. S kompletiranim podacima izrađeno je tri vrste modela predviđanja, osnovni baseline modeli, modeli linearne regresije i model slučajne šume. Na kraju, napravljena je usporedba uspješnosti predviđanja te tri vrste modela.

Ključne riječi: Strojno učenje, analiza podataka, predviđanje, Linearna regresija, Random forest

Prediction models for electric energy consumption

Abstract

The topic of this thesis is the development of predictive machine learning models for household electricity consumption. After collecting the data, an analysis was made, in which outliers were removed and holes were filled. Three types of prediction models were created with the completed data: basic baseline models, linear regression models and random forest models. Finally, a comparison of the prediction performance of those three types of models was made.

Keywords: Machine learning, data analysis, prediction, Linear regression, Random forest