



Univerzitet u Zenici
Politehnički fakultet
Softversko inženjerstvo
Poslovna inteligencija



Integracija rezultata mašinskog učenja u konfigurabilni model za odabir lokacija održivih industrijskih objekata u funkciji predviđanja ekoloških i ekonomskih faktora

Student: Merjem Bajramović
Broj indeksa: II-77

Profesor: Doc.dr. Esad Kadušić

Zenica, septembar 2024. godine

Sažetak: Ovaj rad istražuje primjenu mašinskog učenja u predviđanju optimalnih lokacija za industrijska postrojenja, s naglaskom na održivost i ekonomsku efikasnost. Koristi modele linearne regresije za predviđanje temperatura i solarne radijacije, te regresije slučajne šume za prognozu emisija CO₂, proizvodnje energije, BDP-a i rasta populacije u evropskim državama. Ovi modeli omogućavaju stvaranje okvira za donošenje odluka, gdje se težinski kriteriji dodjeljuju trenutnim i budućim vrijednostima, što omogućava rangiranje država prema klimatskoj stabilnosti, održivosti i ekonomskom rastu. Fleksibilnost modela omogućava prilagodbu kriterija specifičnim potrebama preduzeća te potencijal za daljnje usavršavanje.

Ključne riječi: *prediktivno modeliranje, linearna regresija, slučajna šuma, ekološki uticaj, emisije CO₂, temperatura*

Abstract: This paper explores the application of machine learning in predicting optimal locations for industrial facilities, with a focus on sustainability and economic efficiency. It employs linear regression models to forecast temperatures and solar radiation, and random forest regression models to predict CO₂ emissions, energy production, GDP, and population growth across European countries. These models provide a decision-making framework where weighted criteria are assigned to both current and future values, enabling the ranking of countries based on climate stability, sustainability, and economic growth. The flexibility of the models allows for criteria adjustment according to specific business needs, with potential for further refinement.

Keywords: *predictive modeling, linear regression, random forest, environmental impact, CO₂ emissions, temperature*

Sadržaj

Uvod	4
Pregled literature.....	5
1. Formiranje modela linearne regresije	7
1.1. Predobrada podataka o vremenskim prilikama	7
1.2. Podjela podataka na trening i testne skupove.....	9
1.3. Bilježenje sezonskih obrazaca i dugoročnih promjena	10
1.4. Testiranje i evaluacija	12
1.4.1. Predikcija temperature	12
1.4.2. Predikcija direktne solarne radijacije	13
1.4.3. Predikcija difuzne solarne radijacije.....	13
1.5. Treniranje na cjelokupnom skupu podataka.....	14
2. Model linearne regresije u kontekstu podataka o vremenskim prilikama	17
2.1. Predviđanje vremenskih trendova.....	18
3. Formiranje modela regresije slučajne šume	21
3.1. Predobrada podataka o emisijama CO2	21
3.2. Podjela podataka na trening i testne skupove.....	23
3.3. Primjena modela slučajne šume	23
3.4. Značaj svojstva (eng. <i>feature</i>) u modelu slučajne šume.....	24
3.5. Metrike performansi modela slučajne šume.....	25
3.6. Treniranje na cjelokupnom skupu podataka.....	28
4. Odluke zasnovane na težinskim kriterijima za optimalno smještanje poslovnih objekata.....	30
4.1. Definisanje kriterija za odabir lokacije	30
4.1.1. Trenutna temperatura i solarna radijacija	30
4.1.2. Emisije CO2 i proizvodnja obnovljive energije	30
4.1.3. BDP i populacija.....	31
4.1.4. Predviđena temperatura i solarna radijacija.....	31
4.1.5. Predviđene emisije CO2, proizvodnja obnovljive energije, BDP i populacija.....	31
4.2. Normalizacija podataka	32
4.3. Dodjeljivanje težina.....	32
4.4. Izračun težinskog rezultata	33
Zaključak.....	39
Popis slika.....	40

Izvori.....	41
-------------	----

Uvod

U doba sve veće ekološke svijesti i rasta troškova energije, preduzeća su pod povećanim pritiskom da usvoje održive prakse, a da pritom zadrže ekonomsku efikasnost [1]. Jedna od ključnih odluka za svaku proizvodnu kompaniju je lokacija njenih postrojenja. Ova odluka ne utiče samo na operativne troškove već i na ekološki uticaj. U ovom radu će biti opisan scenarij planiranja izgradnje postrojenja kompanije u Evropi, uzimajući u obzir odabir lokacije koja balansira održivost i ekonomsku efikasnost.

Idealna lokacija postrojenja trebala bi imati umjerenu klimu kako bi se smanjili troškovi energije za grijanje i hlađenje, te obilne količine sunčevog zračenja kako bi se omogućilo korištenje obnovljivih izvora energije, poput solarnih panela. Kako bi se obezbijedilo smanjenje emisija ugljen-dioksida idealna lokacija treba da ima već snažno prisustvo obnovljivih izvora energije i nizak nivo CO₂. Stabilna ekonomija i pouzdana radna snaga su također jedni od ključnih faktora za osiguranje dugoročne poslovne održivosti.

Ovaj rad ima za cilj identifikovati optimalnu evropsku državu za ovakvo postrojenje korištenjem modela mašinskog učenja i analize klimatskih podataka. Korištenjem linearne regresije (*Linear Regression*) i regresije slučajne šume (Random Forest Regression), predviđaju se buduće vrijednosti ključnih ekoloških i ekonomskih faktora, osiguravajući da odluka uzme u obzir ne samo trenutne uslove, već i projektovane trendove u narednoj deceniji. Primjenom pristupa zasnovanog na podacima, teži se odabiru lokacije koja minimizira ekološki uticaj, a istovremeno maksimizira ekonomsku stabilnost i efikasnost poslovanja.

Ova analiza koristi dva primarna skupa podataka: o emisijama CO₂ s Kaggle-a [2] i o vremenskim prilikama iz Open Power System Data [3] repozitorija. Skup podataka o emisijama CO₂ pruža historijske podatke o emisijama ugljen-dioksida po državama, rastu populacije i BDP-u, omogućujući praćenje trendova emisija ugljen-dioksida u različitim državama. Skup podataka o vremenskim prilikama uključuje klimatske varijable kao što su temperatura i solarna radijacija, koji su ključni faktori u određivanju energetske efikasnosti lokacije postrojenja.

U ovom projektu linearna regresija i regresija slučajne šume se međusobno nadopunjuju. Linearna regresija pruža jasna, interpretabilna predviđanja budućih klimatskih uslova (temperatura i solarna radijacija), koji su generalno vođeni dugoročnim, stabilnim trendovima. S druge strane, model slučajne šume obrađuje složenije i nepredvidive faktore koji utiču na emisije CO₂, gdje više varijabli međusobno djeluje na nelinearan način. Kombinovanjem ovih modela osigurava se obrada i ekoloških i ekonomskih aspekata odluke o lokaciji postrojenja, čime se pruža sveobuhvatna i perspektivna analiza pogodnosti svake države.

Pregled literature

Odabir optimalnih lokacija za industrijska postrojenja predstavlja složen izazov koji uključuje razmatranja održivosti, ekonomske efikasnosti i efikasne upotrebe naprednih alata za predviđanje. U nastavku je obuhvaćen pregled ključnih istraživanja, obavljenih u skladu s ciljevima ovog seminarskog rada, sa dodatkom analize njihovog doprinosa širem polju optimizacije odabira održive lokacije korištenjem mašinskog učenja. Integracijom nalaza iz prethodnih istraživanja, ovaj rad nastoji unaprijediti postojeće metodologije i predstaviti sveobuhvatan okvir za donošenje odluka u kontekstu odabira optimalnih lokacija za izgradnju industrijskih postrojenja.

Manuguerra et al. u svom radu *"A machine learning based method for parametric environmental impact model for electric vehicles"* [4] predlažu inovativan pristup koji integriše analizu životnog ciklusa (LCA - Life Cycle Assessment) s mašinskim učenjem. Njihova metodologija omogućava predviđanje ekoloških uticaja tokom faze dizajna električnih vozila putem regresionih modela. Korištenjem minimalnih ulaznih podataka, ovo istraživanje nudi pojednostavljenu metodu za procjenu ekološkog otiska različitih dizajnerskih karakteristika vozila. Ovo istraživanje daje osnov metodološkog pristupa, koji je paralelan cilju predviđanja održivosti lokacija postrojenja. Istraživanje također koristi modele mašinskog učenja za procjenu dugoročnih ekoloških uticaja, naglašavajući važnost prediktivne analitike u održivom donošenju odluka.

Guo et al. predstavljaju ExioML skup podataka, referentni okvir spreman za mašinsko učenje za analizu održivosti u globalnim ekonomskim aktivnostima. Njihovo istraživanje *"ExioML: Eco-economic dataset for Machine Learning in Global Sectoral Sustainability"* [5] naglašava značaj sveobuhvatnih skupova podataka u procjeni ekoloških uticaja unutar globalnih trgovačkih mreža. Ovaj skup podataka uključuje zadatak regresije emisije stakleničkih gasova, naglašavajući ulogu mašinskog učenja u predviđanju ekoloških ishoda. Fokus na stvaranju i korištenju opsežnih skupova podataka odražava pristup obuhvaćen u ovom seminarskom radu, gdje su kvalitetni skupovi podataka i analiza ključni za određivanje optimalnih lokacija postrojenja. Ovaj seminarski rad nadograđuje Green-ovo istraživanje, fokusirajući se posebno na klimatske i energetske podatke za poticanje procjena održivosti.

Istraživanje *"Model proposition for predicting sustainability classes using multicriteria decision support and artificial intelligence"* [6] koje su proveli Couto I Rangel uvodi model koji kombinuje višekriterijsku podršku u odlučivanju s umjetnom inteligencijom za klasifikaciju održivosti kompanija za distribuciju električne energije. Model generiše pravila odlučivanja na osnovu različitih indikatora, kategorizirajući kompanije u različite klase održivosti. Ovo istraživanje je relevantno jer predstavlja primjenu vještačke inteligencije u procesima donošenja odluka na osnovu višekriterijskog pristupa. Ovaj seminarski rad slično, uključuje višekriterijski okvir, koristeći klimatsku stabilnost i ekonomske faktore rasta za rangiranje potencijalnih lokacija postrojenja.

Bilen et al. [7] predstavljaju aplikaciju pametnog grada koja koristi mašinsko učenje za predviđanje optimalnih poslovnih lokacija analizom ekonomskih i demografskih podataka. Njihov model pomaže u urbanom planiranju i alokaciji resursa, naglašavajući donošenje odluka zasnovano na podacima u urbanom kontekstu. Ovaj pristup je usklađen s metodologijom koju pokriva ovaj

seminarski rad, gdje se mašinsko učenje koristi za unapređenje strateškog postavljanja industrijskih objekata unutar geografskog konteksta. Istraživanje naglašava važnost integracije različitih skupova podataka i prediktivne analitike u optimizaciji odluka o lokaciji, princip koji je ključan za ovaj rad.

Shihab et al. [8] dodatno istražuju upotrebu mašinskog učenja za identifikaciju idealnih lokacija za restorane putem algoritama poput mašina sa potporom vektora, stabala odluke i linearne regresije. Njihovo istraživanje ističe ulogu više faktora u određivanju optimalnih lokacija, što rezonira s fokusom seminarskog rada na klimatske i ekonomske varijable u optimizaciji lokacije postrojenja. Ovo istraživanje podržava pretpostavku da mašinsko učenje može efikasno upravljati složenim procesima donošenja odluka koji uključuju više varijabli, pristup koji ovaj seminarski rad usvaja i proširuje na veći obim u kontekstu industrijskih objekata.

Huo i Chaudhry [9] istražuju upotrebu mašinskog učenja u evaluaciji odluka o globalnoj ekspanziji lokacija unutar kineskog proizvodnog sektora. Njihova studija uključuje faktore poput finansijskog uticaja i nivoa plaća, koristeći neuronske mreže i druge napredne tehnike. Upotreba sveobuhvatnih skupova podataka i sofisticirane analize odražava pristup obuhvaćen u ovom seminarskom radu, gdje se koriste višestruki izvori podataka i modeli mašinskog učenja za predviđanje optimalnih lokacija postrojenja, čime se rad nadograđuje. Na ovaj način, posebno integriranjem klimatskih podataka, se unapređuju prediktivne sposobnosti i daje fokus na održivost u optimizaciji odabira lokacije.

Ovaj seminarski rad integriše metodologije i uvide iz navedenih radova kako bi se kreirao holistički model za optimizaciju lokacija industrijskih pogona. Korištenjem tehnika mašinskog učenja kao što su linearna regresija i regresija slučajnih šuma, ovo istraživanje proširuje primjenu prediktivne analitike u optimizaciji odavira lokacije postrojenja. Ono se diferencira integracijom šireg spektra faktora, uključujući klimatsku stabilnost i ekonomske indikatore, u okvir za donošenje odluka. Na taj način se unapređuje proces strateškog planiranja za industrijska postrojenja, usklađujući se s naprednim metodologijama analize zasnovane na podacima i sistemima za podršku u donošenju odluka zasnovanim na mašinskom učenju koji su pregledani u ovoj literaturi.

1. Formiranje modela linearne regresije

Za predviđanja budućih vrijednosti temperature i solarne radijacije u ovom radu se koristi linearna regresija jer predstavlja interpretabilan i efikasan model koji pretpostavlja linearnu vezu između ulaznih varijabli (historijski podaci vremenskih prilika) i ciljne varijable (buduća temperatura ili solarne radijacije). Ova metoda je idealna za klimatske podatke, gdje se dugoročni trendovi, poput rasta temperature ili promjena u solarnom zračenju, često mogu modelirati linearnim trendovima na osnovu historijskih podataka [4]. Primjenom modela linearne regresije na historijske klimatske podatke, generišu se predviđanja uticaja promjenjivih vremenskih prilika na energetske efikasnost. U kontekstu ovog projekta, linearna regresija omogućava procjenu rasta ili pada temperature i solarne radijacije u svakoj državi, omogućavajući da predviđanje da li će ciljana država ostati energetski efikasna lokacija za postrojenje u budućnosti.

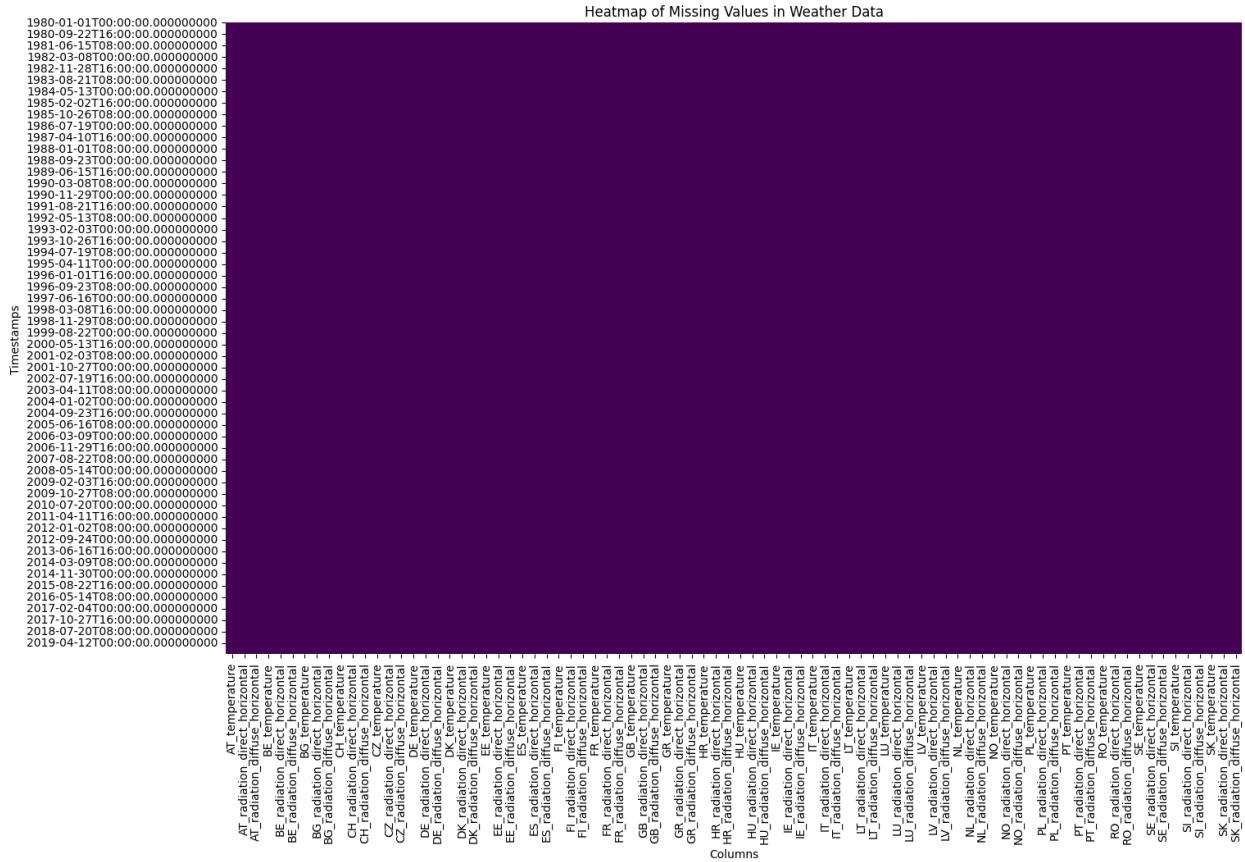
1.1. Predobrada podataka o vremenskim prilikama

Podaci o vremenskim prilikama korišteni u ovoj analizi preuzeti su iz repozitorija Open Power System Data [3], koji pruža sveobuhvatne informacije o vremenskim prilikama za 28 evropskih država. Skup podataka sadrži satne zapise ključnih vremenskih parametara, poput temperature i solarne radijacije po satu u vremenskom periodu od 1980. do 2020. za svaku državu. Na primjer, kolone *AT_temperature* i *AT_radiation_direct_horizontal* odnose se na temperaturu i direktnu solarnu radijaciju u Austriji (Slika 1).

utc_timestamp	AT_temperature	AT_radiation_direct_horizontal	AT_radiation_diffuse_horizontal
1980-01-01T00:00:00Z	-3.64	0	0
1980-01-01T01:00:00Z	-3.803	0	0
1980-01-01T02:00:00Z	-3.969	0	0
1980-01-01T03:00:00Z	-4.076	0	0
1980-01-01T04:00:00Z	-4.248	0	0
1980-01-01T05:00:00Z	-4.527	0	0
1980-01-01T06:00:00Z	-4.84	0.0239	0.4413
1980-01-01T07:00:00Z	-4.703	4.6844	33.5061
1980-01-01T08:00:00Z	-3.835	30.1528	90.8233
1980-01-01T09:00:00Z	-2.804	61.2307	126.7209

Slika 1 Prikaz nekoliko prvih redova skupa podataka o vremenu "Weather Data" za jednu državu (Austrija)

Jedan od prvih koraka u pripremi podataka je osiguranje njihove potpunosti. Nedostajući podaci mogu značajno uticati na tačnost modela mašinskog učenja, pa je identifikacija i rješavanje nedostajućih vrijednosti od ključne važnosti. Nakon učitavanja skupa podataka o vremenskim prilikama provjerava se postojanje bilo kakvih praznih ili nedostajućih vrijednosti koristeći kombinaciju pandas biblioteke u Python-u i vizualizaciju pomoću seaborn-a (Slika 2).



Slika 2 Toplotna mapa (heatmap) vizualizacije nedostajućih vrijednosti (nema nedostajućih vrijednosti)

Ključne vremenske karakteristike relevantne za treniranje modela mašinskog učenja:

- **temperature:** temperatura izražena u stepenima Celzijusa (°C). Odabrana kao primarni faktor za određivanje energetske potrebe postrojenja, posebno za sisteme grijanja i hlađenja.
- **radiation (direct/diffuse):** Direktna/difuzna solarna radijacija mjerena u vatima po kvadratnom metru (W/m^2). Ključna za procjenu potencijala solarne energije, koja bi se mogla koristiti za napajanje postrojenja ili smanjenje njegove ovisnosti o neobnovljivim izvorima energije.

Obzirom na to da su podaci zapisani na satnom nivou, kako bi se smanjila veličina skupa radi efikasnije obrade, izvršena je agregacija u prosječne mjesečne vrijednosti (Slika 3). Ova agregacija pomaže u smanjenju šuma u podacima i čini ih prikladnijim za prediktivno modeliranje.

utc_timestamp	AT_temperature	AT_radiation_direct_horizontal	AT_radiation_diffuse_horizontal
1980-01-31 00:00:00+00:00	-6.103502688	24.11349341	33.06779207
1980-02-29 00:00:00+00:00	-0.42441523	60.92634828	40.65633089
1980-03-31 00:00:00+00:00	1.464170699	54.62848763	76.87203199
1980-04-30 00:00:00+00:00	3.720781944	81.58877583	93.56323097
1980-05-31 00:00:00+00:00	9.083321237	130.6227914	109.5576227
1980-06-30 00:00:00+00:00	14.27126389	118.9179772	126.1006601
1980-07-31 00:00:00+00:00	15.5336828	128.0973038	116.3180677
1980-08-31 00:00:00+00:00	17.2123414	142.7587462	88.03111707
1980-09-30 00:00:00+00:00	13.2235	98.00839333	73.36833611

Slika 3 Prikaz nekoliko prvih redova skupa podataka o vremenu "Weather Data" za jednu državu (Austrija) nakon izvršene agregacije u prosječne mjesečne vrijednosti

1.2. Podjela podataka na trening i testne skupove

Kako bi se osiguralo da model generalizira dobro na neviđene podatke, korišten je dvostepeni proces. Prvo su podaci podijeljeni na trening i testne skupove u omjeru 80:20%, pri čemu je model treniran koristeći samo dio skupa podataka. Nakon što je potvrđeno da model pruža zadovoljavajuće rezultate, on je ponovno treniran na cjelokupnom skupu podataka kako bi se napravile predikcije za buduće vremenske periode. Ovaj pristup omogućava procjenu performansi modela prije nego što se primijeni za dugoročne predikcije, za koje nije moguća direktna validacija.

U početnoj podjeli podataka prvih 80% je korišteno za treniranje, a preostalih 20% za testiranje. Ovaj pristup poštuje vremensku ovisnost podataka, osiguravajući da se model trenira na prošlim podacima i evaluira na budućim, simulirajući stvarne uvjete. Ova podjela omogućava provjeru da li model hvata temeljne trendove u podacima, umjesto da se samo prilagođava prošlim vrijednostima. Za ovaj prvi test, kao fokus su korišteni podaci o vremenskim prilikama Njemačke. Skup podataka podijeljen je na sljedeći način:

- **Trening skup (80%):** Podaci iz ranijih perioda, koji su korišteni za treniranje modela.
- **Testni skup (20%):** Podaci iz novijih perioda, koji su korišteni za evaluaciju sposobnosti modela da predvidi neviđene vremenske obrasce.

Trening skup se koristi za učenje modela o odnosima između ulaznih karakteristika i ciljane varijable. U ovom slučaju, ulazne karakteristike uključuju vremenske karakteristike i nivoe radijacije za Njemačku. Kreirane su ciklične karakteristike kako bi se uhvatile sezonske varijacije u temperaturi i radijaciji:

```
df['month_sin'] = np.sin(2 * np.pi * df['month'] / 12)
df['month_cos'] = np.cos(2 * np.pi * df['month'] / 12)
df['time_trend'] = np.arange(len(df))
```

- `month_sin` i `month_cos` hvataju cikličnu prirodu mjeseci. Sinusne i kosinusne transformacije osiguravaju da je prijelaz iz decembra (mjesec 12) u januar (mjesec 1) gladak, izbjegavajući veliki skok koji bi stvorio jednostavan brojni indeks mjeseca.
- `time_trend` je linearna karakteristika koja predstavlja prolazak vremena. Ovo pomaže modelu da uzme u obzir dugoročne trendove, poput globalnog zagrijavanja, koje može uzrokovati postepeni porast temperature kroz godine.

1.3. Bilježenje sezonskih obrazaca i dugoročnih promjena

Podaci o vremenskim prilikama, posebno temperatura i solarna radijacija, imaju izražene sezonske obrasce. Na primjer, temperature su obično više ljeti, a niže zimi, dok je solarna radijacija jača tokom dužih dana u ljetnim mjesecima, a slabija tokom kraćih dana zimi.

Mjeseci u godini su po svojoj prirodi ciklični; nakon decembra (mjesec 12), podaci prelaze u januar (mjesec 1). Ovo stvara problem za modele ako jednostavno koristimo broj mjeseca kao ulaznu karakteristiku. Na primjer, razlika između decembra (mjesec 12) i januara (mjesec 1) iznosi samo jedan mjesec, ali numerički bi se činilo kao da postoji veliki skok s 12 na 1, što ne odražava stvarni odnos. Kako bi se riješio ovaj problem, koristi se cikličko kodiranje pretvaranjem mjeseca u dvije karakteristike: sinus i kosinus mjeseca. Na taj način, model može razumjeti da su januar i decembar blizu jedan drugome, jer su i sinusne i kosinusne krivulje ciklične i ponavljaju se svakih 12 mjeseci. Formule za ove transformacije su:

$$month_sin = \sin\left(\frac{2\pi \cdot month}{12}\right)$$

$$month_cos = \cos\left(\frac{2\pi \cdot month}{12}\right)$$

Ove transformacije omogućavaju modelu da tretira mjesece kao kontinuirane, ciklične vrijednosti. Na primjer:

- U **januaru** ($month = 1$), $month_sin \approx 0.5$ $month_cos \approx 0.87$.
- U **julu** ($month = 7$), $month_sin \approx -0.5$ $month_cos \approx -0.87$.

Zbog ciklične prirode sinusnih i kosinusnih funkcija, ova kodiranja osiguravaju da model prepoznaje periodičnu prirodu mjeseci, što je ključno za modeliranje sezonskih vremenskih obrazaca.

Pored sezonalnosti, podaci o vremenskim prilikama također pokazuju dugoročne trendove, posebno u kontekstu klimatskih promjena. Na primjer, prosječne temperature mogu postepeno rasti tokom vremena zbog globalnog zagrijavanja. Kako bi se zabilježili ovi dugoročni trendovi, uvodi se karakteristika vremenskog trenda (*time trend*), koja je jednostavno rastući broj koji predstavlja prolazak vremena. Ovo se implementira na sljedeći način:

$$time_trend = 0, 1, 2, \dots, n$$

Gdje je n broj mjeseci u skupu podataka. Vremenski trend omogućava modelu da uzme u obzir postepene promjene kroz vrijeme. Na primjer, ako su temperature stalno rasle tokom posljednje decenije, model može naučiti da veće vrijednosti vremenskog trenda odgovaraju višim temperaturama. Vremenski trend hvata nesezonske, dugoročne obrasce koji se ne mogu objasniti samo mjesečnim ciklusima. Bez ove karakteristike, model bi mogao propustiti ove postepene promjene koje se dešavaju izvan uobičajenih sezonskih obrazaca.

Model je treniran da predvidi nivoe temperature i radijacije koristeći ciklične karakteristike mjeseci, vremenski trend i podatke o radijaciji. Model linearne regresije nastoji uklopiti ravnu liniju kroz podatke, minimizirajući grešku između predviđenih i stvarnih vrijednosti. U ovom slučaju, set za treniranje je korišten za prilagođavanje sljedeće linearne jednačine za temperaturu [10]:

$$y_{temperature} = \beta_0 + \beta_1 \cdot time_trend + \beta_2 \cdot month_sin + \beta_3 \cdot month_cos + \beta_4 \cdot radiation_direct + \beta_5 \cdot radiation_diffuse$$

Model linearne regresije uči ove koeficijente minimiziranjem srednje kvadratne greške (*Mean Squared Error*, MSE), koja mjeri prosječnu kvadratnu razliku između predviđenih i stvarnih vrijednosti:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Gdje je:

y_i stvarna vrijednost

\hat{y}_i predviđena vrijednost iz modela

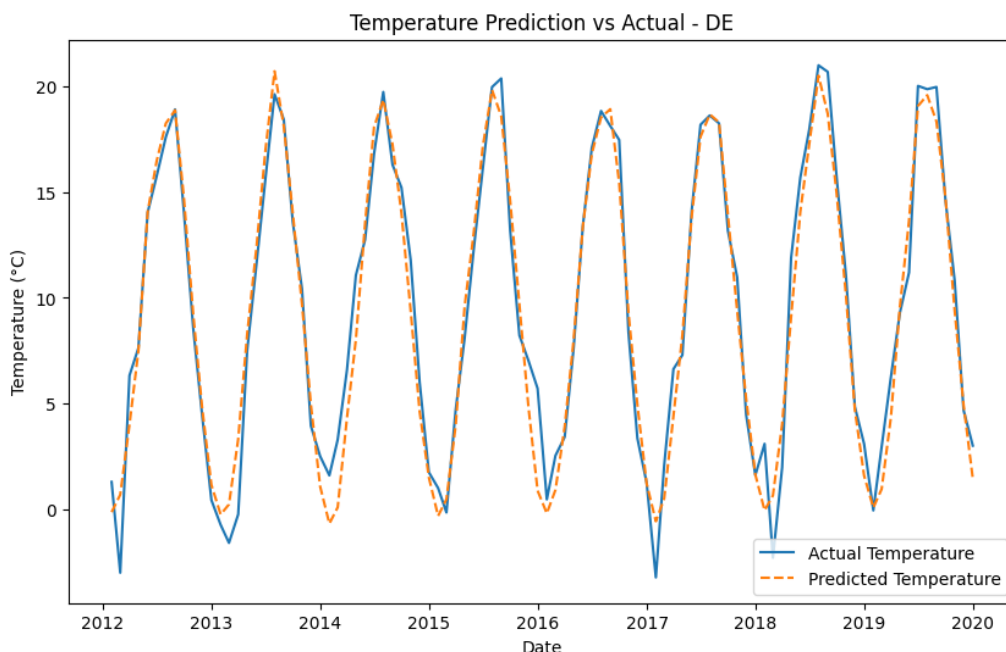
n broj tačaka podataka

1.4. Testiranje i evaluacija

Nakon treniranja modela, njegova se efikasnost evaluira na testnom skupu. Testni skup sadrži podatke koje model ranije nije vidio, što omogućava precizniju procjenu njegove sposobnosti generalizacije. Performanse modela na ovim podacima daju indicaciju o tome kako će on funkcionirati prilikom budućih predikcija. Performanse modela evaluirane su usporedbom MSE s varijansom stvarnih podataka za svaku predviđenu varijablu; temperaturu, direktnu solarnu radijaciju i difuznu solarnu radijaciju. MSE je široko korištena metrika u regresijskoj analizi koja mjeri prosječnu kvadratnu razliku između predviđenih i stvarnih vrijednosti. Niži MSE ukazuje na bolju izvedbu modela jer odražava koliko su predviđanja bliska stvarnim vrijednostima. Kako bi se rezultati bolje interpretirali, MSE je poređen s var stvarnih podataka, što pruža uvid u prirodne fluktuacije podataka. Ako je MSE znatno manji od varijanse, to sugerira da model zahvata većinu varijabilnosti u podacima. Nasuprot tome, ako je MSE blizu varijanse ili veći, to ukazuje na slabe performanse modela.

1.4.1. Predikcija temperature

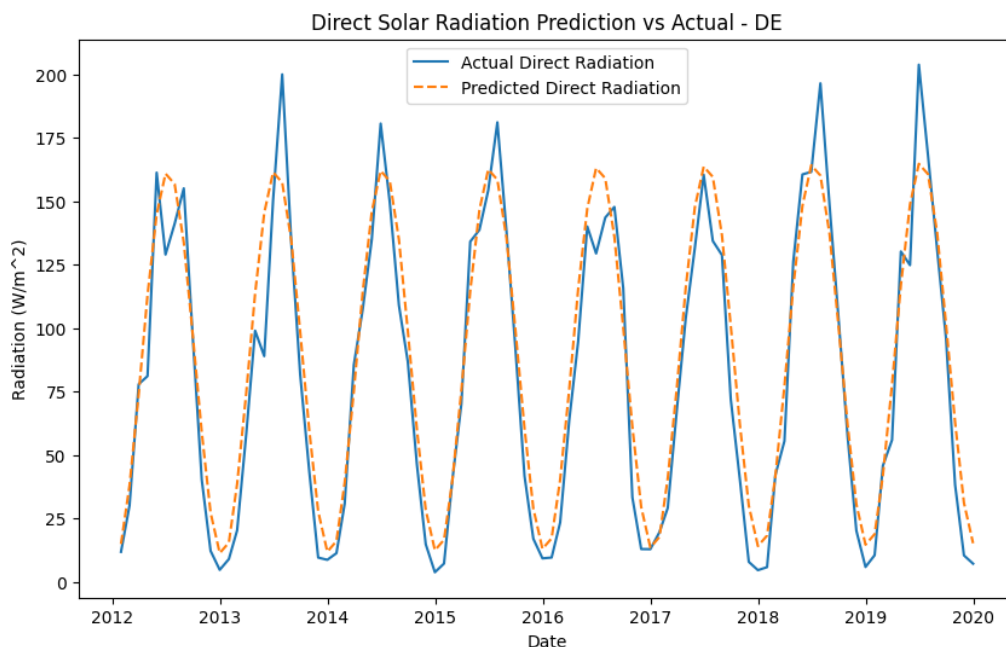
Za predikcije temperature, MSE iznosi 2,37, što je znatno manje od varijanse koja iznosi 48,14. Ovo ukazuje na to da model za temperaturu radi dobro jer zahvata značajan dio varijabilnosti u podacima o temperaturi. Zapravo, MSE predstavlja oko 5% od ukupne varijanse, što sugerira da je greška između predviđenih i stvarnih vrijednosti temperature relativno mala u odnosu na prirodne fluktuacije prisutne u podacima. Stoga, model za temperaturu pruža razumne predikcije, uspješno zahvatajući sezonske cikluse i dugoročne trendove u podacima.



Slika 4 Grafički prikaz rezultata testiranja predikcije temperature

1.4.2. Predikcija direktne solarne radijacije

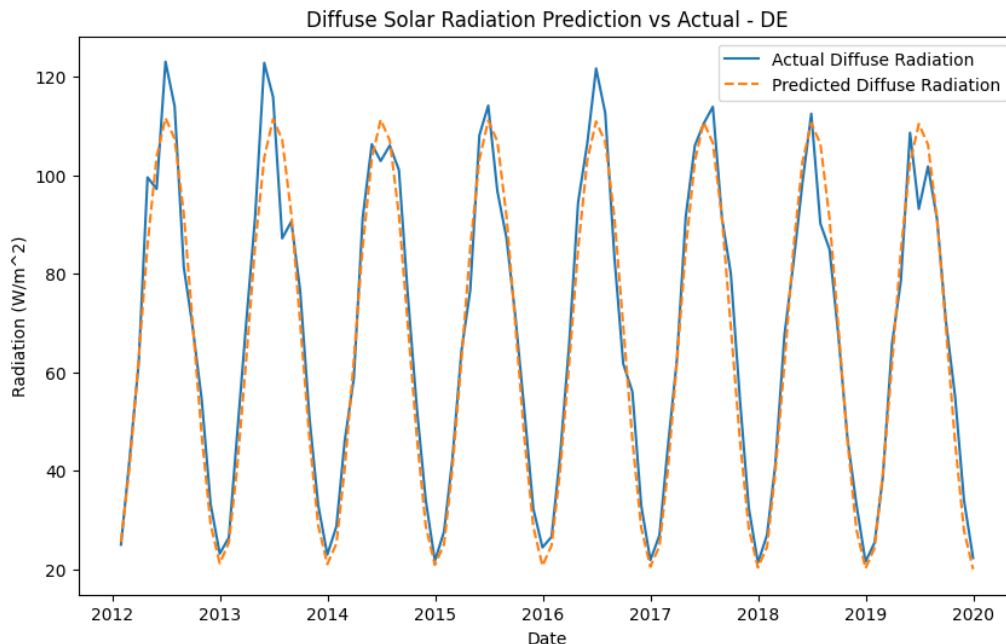
Predikcije direktne solarne radijacije dale su MSE od 291,17 u usporedbi s varijansom od 3455,40. Iako je MSE veći nego kod temperature, i dalje je znatno manji od varijanse (približno 8,4% od ukupne varijanse). Ovo sugerira da model može zahvatiti značajan dio varijabilnosti u direktnoj solarnoj radijaciji, unatoč inherentnim izazovima predikcije ovako promjenljive pojave. Solarnoj radijaciji značajno utječu kratkoročni vremenski uslovi, poput naoblake, što dodaje složenost predikcijama modela. Ipak, performanse modela su zadovoljavajuće jer hvata opći trend direktne solarne radijacije s prihvatljivim nivoom tačnosti.



Slika 5 Grafički prikaz rezultata testiranja predikcije direktne solarne radijacije

1.4.3. Predikcija difuzne solarne radijacije

Za difuznu solarnu radijaciju model je postigao MSE od 42,46, što je relativno nisko u usporedbi s varijansom od 976,31. MSE predstavlja oko 4,3% od ukupne varijanse, što ukazuje na to da model dobro radi u predikciji difuzne solarne radijacije. Ova niža greška u odnosu na varijansu sugerira da model uspješno zahvata sezonske obrasce i dugoročne trendove u difuznoj radijaciji, čineći ga efikasnim prediktorom u ovom kontekstu.



Slika 6 Grafički prikaz rezultata testiranja predikcije difuzne solarne radijacije

Prethodnom evaluacijom je zaključeno da modeli dobro funkcioniraju u predikciji temperature i difuzne solarne radijacije, s MSE vrijednostima koje su značajno niže od odgovarajućih varijanci, što ukazuje na to da modeli zahvataju većinu prirodne varijabilnosti ovih varijabli. Model za direktnu solarnu radijaciju, iako ima veći MSE, i dalje daje zadovoljavajuće rezultate s obzirom na inherentnu varijabilnost solarne radijacije uzrokovanu kratkoročnim vremenskim fluktuacijama. Usporedba između MSE i varijance sugerira da modeli generalno efikasno hvataju temeljne trendove u podacima, posebno za temperaturu i difuznu radijaciju, gdje je greška mali dio prirodne varijabilnosti.

1.5. Treniranje na cjelokupnom skupu podataka

Nakon što je potvrđeno da model funkcionira zadovoljavajuće na podjeli skupa podataka u omjeru 80:20%, izvršeno je treniranje na cjelokupnom skupu podataka kako bi se napravile predikcije za buduće vremenske periode. Ovaj korak je ključan jer u stvarnim scenarijima, budući podaci nisu dostupni za testiranje. Stoga, model mora biti treniran na svim raspoloživim historijskim podacima kako bi pružio najpreciznije predikcije za buduće vremenske periode. Ponovnim treniranjem na punom skupu podataka, model može uhvatiti sve relevantne obrasce i odnose u podacima. Prije treniranja modela, kreirane su ciklične karakteristike kako bi se uzeli u obzir sezonske varijacije i vremenski trend za hvatanje dugoročnih trendova u temperaturi i solarnoj radijaciji. Ciklične karakteristike, kao što je već prethodno objašnjeno, za mjesec pomažu modelu da razumije ponavljajuće sezonske obrasce. Bez ovih karakteristika je moguće da model ne bi pravilno interpretirao cikličnu prirodu vremena, posebno pri prelasku iz decembra u januar. Nakon kreiranja cikličnih karakteristika i vremenskog trenda, treniran je zaseban linearni regresijski model za nivoe temperature i radijacije u svakoj državi.

Priprema ulaznih karakteristika i ciljne varijable za model temperature:

```
X_temp = country_df[['month_sin', 'month_cos', 'time']]
y_temp = country_df[temperature_col]
```

`X_temp` matrica karakteristika za model temperature koja sadrži ulazne karakteristike (također poznate kao prediktori ili nezavisne varijable) za model. U ovom slučaju, uključuje tri kolone:

- `month_sin` i `month_cos` su sinusne i kosinusne transformacije mjeseca, koje hvataju ciklični obrazac godišnjih doba.
- `time` je linearni vremenski trend koji predstavlja protok vremena i modelira dugoročne trendove.

`y_temp` je ciljna varijabla za model temperature, koja predstavlja stvarne vrijednosti temperature za državu koja se modelira. Model će naučiti odnos između ulaznih karakteristika (`X_temp`) i ciljne varijable (`y_temp`) kako bi kasnije mogao predvidjeti temperature na osnovu budućih vrijednosti `month_sin`, `month_cos`, i `time`.

Treniranje modela za temperaturu:

```
model_temp = LinearRegression()
model_temp.fit(X_temp, y_temp)
```

`model_temp = LinearRegression()` predstavlja inicijalizaciju modela linearne regresije. Ovaj algoritam mašinskog učenja nastoji modelirati odnos između ulaznih karakteristika (`X_temp`) i ciljne varijable (`y_temp`) kao linearnu jednačinu u obliku:

$$y = \beta_0 + \beta_1 \cdot month_sin + \beta_2 \cdot month_cos + \beta_3 \cdot time$$

Gdje je:

y ciljana varijabla (predviđena temperatura),

β_0 presjek, koji predstavlja osnovnu vrijednost temperature kada su svi ostali faktori jednaki nuli,

β_1 i β_2 koeficijenti za ciklične karakteristike, koji zahvataju sezonske varijacije i cikluse,

β_3 koeficijent za vremenski trend, koji zahvata dugoročne promjene i trendove u podacima kroz vrijeme.

`model_temp.fit(X_temp, y_temp)` trenira model koristeći ulazne karakteristike (`X_temp`) i ciljnu varijablu (`y_temp`). Model uči koeficijente (težine) β_0 , β_1 , β_2 , β_3 koji minimiziraju MSE između predviđenih i stvarnih vrijednosti temperature u skupu podataka za treniranje.

Na sličan način treniraju se zasebni modeli za direktnu i difuznu solarnu radijaciju koristeći iste ulazne karakteristike (`month_sin`, `month_cos`, i `time`):

```
model_direct_rad = LinearRegression()
model_direct_rad.fit(X_temp, country_df[radiation_cols[0]])

model_diffuse_rad = LinearRegression()
model_diffuse_rad.fit(X_temp, country_df[radiation_cols[1]])
```

`model_direct_rad` i `model_diffuse_rad` su modeli linearne regresije za predviđanje direktne i difuzne solarne radijacije.

`Country_df[radiation_cols[0]]` i `country_df[radiation_cols[1]]` izrazi pristupaju podacima o direktnoj i difuznoj solarnoj radijaciji za državu.

Treniranje modela se vrši na ulaznim karakteristikama (`X_temp`) i ciljnoj varijabli, koju predstavljaju podaci o direktnoj ili difuznoj solarnoj radijaciji. Slično kao kod modela za temperaturu, modeli za radijaciju uče kako se direktna i difuzna radijacija mijenjaju u odnosu na ciklične karakteristike i vrijeme.

2. Model linearne regresije u kontekstu podataka o vremenskim prilikama

Linearna regresija nastoji da pronađe najbolje odgovarajući linearni odnos između ulaznih karakteristika (cikličnih karakteristika i vremenskog trenda) i ciljne varijable (temperature, direktne ili difuzne radijacije). Ovaj postupak se izvodi pronalaženjem optimalnih vrijednosti za koeficijente $\beta_0, \beta_1, \beta_2, \beta_3$ koji minimiziraju razliku između stvarnih i predviđenih vrijednosti ciljne varijable. Model podrazumijeva da je odnos između ulaznih karakteristika i ciljne varijable linearan, što znači da promjene u ulaznim karakteristikama dovode do proporcionalnih promjena u ciljnoj varijabli.

U ovom slučaju, ciklične karakteristike `month_sin` i `month_cos` hvataju sezonske varijacije, omogućavajući modelu da predvidi da će temperatura i radijacija rasti tokom ljeta i opadati zimi. Vremenska karakteristika `time` obuhvata dugoročne trendove, kao što su porast temperatura kroz godine uslijed klimatskih promjena. Učeci koeficijente za ove karakteristike, linearni regresijski model može napraviti prognoze za buduće mjesece koristeći njihove vrijednosti za `month_sin`, `month_cos`, i `time`.

Korištenje istih karakteristika (`month_sin`, `month_cos`, `time`) u svim modelima omogućava dosljednost i pojednostavljuje proces modeliranja. I ciklične i vremenske karakteristike igraju ključnu ulogu u prepoznavanju sezonskih promjena i dugoročnih trendova u temperaturi i sunčevoj radijaciji:

- **Sezonski obrasci:** Kako temperatura i sunčeva radijacija značajno zavise od sezonskih promjena, ciklične karakteristike omogućavaju modelu da nauči obrasce, poput viših temperatura i veće radijacije tokom ljetnih mjeseci.
- **Dugoročni trendovi:** Vremenska karakteristika pomaže modelu da prepozna postupene promjene tokom vremena, kao što su efekti klimatskih promjena.

Linearni regresijski model pretpostavlja da postoji linearan odnos između ulaznih karakteristika (nezavisnih varijabli) i ciljne varijable (zavisne varijable koja se predviđa). Cilj modela je pronaći jednačinu najbolje odgovarajuće pravolinijske funkcije (ili ravne površine u slučaju više ulaznih karakteristika) koja može predvidjeti ciljnu varijablu na osnovu vrijednosti ulaznih karakteristika. Da bi to postigao, model minimizira zbir kvadratnih razlika između stvarnih zabilježenih vrijednosti i vrijednosti koje model predviđa. Prilagođavanjem koeficijenata ulaznih karakteristika, model nastoji da svoja predviđanja što više približi stvarnim podacima. Ovaj proces osigurava smanjenje ukupne greške između predviđenih i zabilježenih rezultata, što rezultira modelom koji efikasno zahvata linearni odnos između ulaznih i ciljnih varijabli.

Matematički, za svaki model, rješenje uključuje pronalaženje koeficijenata u jednačini:

$$y = \beta_0 + \sum_{i=1}^n \beta_i x_i$$

Gdje je:

y ciljna varijabla (temperatura ili radijacija),

x_i ulazne karakteristike (`month_sin`, `month_cos`, `time`),

β_0 osnovna vrednost ciljne varijable,

β_i koeficijenti koji odgovaraju svakoj ulaznoj karakteristici.

Tokom procesa treniranja modela, ulazne karakteristike `X_temp` uključuju vrijednosti `month_sin`, `month_cos`, i `time` za svaku tačku podataka, dok ciljne varijable (poput `y_temp` za temperaturu) sadrže posmatrane vrijednosti ciljne varijable. Model prilagođava koeficijente β_i da bi minimizovao funkciju troška (obično MSE). Prednosti ovog pristupa uključuju:

- **Zapažanje sezonskih obrazaca:** Korištenjem `month_sin` i `month_cos`, model uspješno uči sezonske obrasce prisutne u podacima.
- **Računanje dugoročnih trendova:** Vremenska karakteristika omogućava modelu da prati dugoročne trendove, kao što su porast ili pad temperatura i radijacije tokom godina.
- **Jednostavnost:** Linearna regresija je jednostavna za implementaciju i interpretaciju, što je čini pogodnom za ovu vrstu analize.

2.1. Predviđanje vremenskih trendova

Model je, nakon treniranja na cjelokupnom skupu podataka, iskorišten za predviđanje vremenskih trendova temperature i solarne radijacije za narednih 10 godina. Ova predviđanja su napravljena na mjesečnoj osnovi, koristeći iste ciklične karakteristike (sinus i kosinus mjeseca) i vremenski trend koji su korišteni tokom procesa treniranja. Time se osigurava da model zahvata i sezonske varijacije, poput promjena u temperaturi i radijaciji tokom godine, kao i dugoročne trendove, poput potencijalnog postepenog porasta temperature zbog klimatskih promjena.

Za početak, kreiran je novi DataFrame koji predstavlja vremensku liniju za narednih 120 mjeseci, što odgovara periodu od 10 godina. Ovo uključuje generisanje budućih datuma i ponovni izračun cikličnih karakteristika i vremenskog trenda kako bi se uskladili s budućim periodom:

```
future_dates = pd.date_range(country_df.index.max() +  
pd.DateOffset(months=1), periods=120, freq='M')
```

```

future_df = pd.DataFrame(index=future_dates)
future_df['month'] = future_df.index.month
future_df['month_sin'] = np.sin(2 * np.pi * future_df['month'] / 12)
future_df['month_cos'] = np.cos(2 * np.pi * future_df['month'] / 12)
future_df['time'] = np.arange(len(country_df), len(country_df) + 120)

```

Nakon pripreme podataka za budući period, korišteni su trenirani modeli za predviđanje temperature i solarne radijacije (direktne i difuzne) za svaku državu. Ovaj proces uključuje unos cikličnih karakteristika i vremenskog trenda u modele kako bi se generisala mjesečna predviđanja za narednu deceniju. Predviđanja su zasnovana na pretpostavci da će se sezonski ciklusi i dugoročni trendovi nastaviti slično u budućnosti. U nastavku je prikazan ključni dio procesa predikcije:

```

future_df[temperature_col] = model_temp.predict(future_df[['month_sin',
'month_cos', 'time']])
future_df[radiation_cols[0]] =
model_direct_rad.predict(future_df[['month_sin', 'month_cos', 'time']])
future_df[radiation_cols[1]] =
model_diffuse_rad.predict(future_df[['month_sin', 'month_cos', 'time']])

```

Nakon generisanja predikcija, rezultati su sumirani izračunavanjem prosječnih nivoa temperature i radijacije za svaku državu tokom cijelog desetogodišnjeg perioda (Slika 7). Ovaj korak prosjekovanja pruža jasniju sliku očekivanih klimatskih uslova u svakoj regiji kroz duži period, pomažući u procjeni održivosti i energetskeg potencijala različitih lokacija za postrojenja. Vrijednosti direktne i difuzne radijacije su kombinovane kako bi se stvorila ukupna mjera predviđenog potencijala solarne energije:

```

avg_temp = future_df[temperature_col].mean()
avg_direct_rad = future_df[radiation_cols[0]].mean()
avg_diffuse_rad = future_df[radiation_cols[1]].mean()

predicted_avg_radiation = avg_direct_rad + avg_diffuse_rad

```

Ovaj pristup omogućava donošenje informisanih odluka o potencijalnim lokacijama za postrojenja. Bilježenjem, kako kratkoročnih sezonskih obrazaca, tako i dugoročnih trendova, model pomaže u identifikaciji regija koje će vjerovatno imati stabilne i povoljne vremenske uslove u narednoj deceniji, čime se osigurava energetska efikasnost i održivost u poslovanju postrojenja.

	average_temperature▼	average_radiation▼	temperature_fluctuation▼	predicted_avg_temp▼	predicted_avg_radiation▼
SK	8.79	158.47	8.63	9.41	162.87
PL	8.56	144.77	8.47	9.21	149.54
LU	9.26	151.12	6.88	9.87	155.14
FI	4.23	113.14	9.11	4.93	116.91
CZ	8.51	153.71	8.21	9.18	158.83
GR	16.77	205.52	6.97	17.18	205.75
DK	8.71	132	6.27	9.38	134.61
LV	6.28	127.39	8.74	6.8	131.91
EE	6	123.81	8.69	6.59	127.58
GB	9.54	136.87	4.68	10.09	138.6
IT	14.23	184.49	7.05	14.7	185.64
IE	9.59	130.21	3.81	10.01	131.08
PT	15.38	202.59	4.39	15.44	204.27
DE	9.24	147.84	7.13	9.94	153.08
BE	10.07	146.23	6.04	10.68	149.93
CH	7.97	163.85	7.22	8.43	168.62
NL	10.23	142.92	5.89	10.9	147.05
HR	11.94	171.69	8.03	12.51	174.41
SE	6.41	121.95	7.62	7.17	124.44
SI	9.84	165.93	7.89	10.43	168.95
FR	11.29	163.87	6.14	11.78	166.85
AT	8.3	160.48	7.87	8.79	164.4
HU	11.16	164.32	8.91	11.79	168.53
NO	4.31	109.48	7.17	4.97	111.05
LT	6.54	131.61	8.82	7.02	135.95
ES	15.3	206.34	6.09	15.53	208.22
BG	11.56	184.44	8.58	12.13	186.52
RO	10.71	170.51	9.22	11.36	175.15

Slika 7 Prosječne temperature i radijacija po državama u posljednjoj deceniji (lijevo) i predviđene vrijednosti za narednu deceniju (desno)

3. Formiranje modela regresije slučajne šume

Emisije CO₂ mogu biti pod utjecajem širokog spektra faktora, poput proizvodnje energije, industrijske aktivnosti i rasta populacije, koji mogu međusobno djelovati na nelinearan način. Zbog toga je za predviđanje emisija CO₂ odabrana regresija slučajne šume. Ovaj model je posebno pogodan za probleme gdje je veza između ulaznih varijabli i ciljne varijable kompleksna i nelinearna, što je često slučaj s emisijama CO₂. Emisije su pod utjecajem mnogih međuzavisnih faktora, uključujući proizvodnju energije (posebno omjer obnovljivih i neobnovljivih izvora), industrijsku aktivnost, rast populacije i ekonomski razvoj.

Slučajna šuma gradi više stabala odluke na osnovu podskupova podataka i zatim agregira rezultate kako bi napravila preciznija predviđanja. Svako stablo odluke je model koji dijeli podatke u manje grupe na osnovu najvažnijih karakteristika (poput proizvodnje energije ili populacije), pokušavajući minimizirati grešku u predviđanju emisija CO₂. Usrednjavanjem rezultata mnogih stabala, model slučajne šume smanjuje rizik od pretjeranog prilagođavanja podacima za treniranje i stvara generalizirani model.

Model slučajne šume u kontekstu emisija CO₂ može uhvatiti složene, nelinearne interakcije između faktora kao što su proizvodnja energije, BDP i populacija. Za razliku od linearne regresije, slučajna šuma ne pretpostavlja linearnu vezu između ulaznih i izlaznih varijabli. Na primjer, kako BDP neke države raste, emisije se ne moraju povećavati konstantno. Umjesto toga, mogu se stabilizirati ili čak smanjiti ako država investira u zelene tehnologije. Model slučajne šume je također manje osjetljiv na izuzetke i šum u podacima, što ga čini idealnim za situacije u kojima postoje značajne varijacije u emisijama zbog naglih promjena u politikama ili promjena u energetskej infrastrukturi. Korištenjem slučajne šume, moguće je kreirati preciznija predviđanja budućih emisija CO₂ za svaku državu, što je ključno za identifikaciju lokacija koje su u skladu s ciljevima kompanije za održivost.

3.1. Predobrada podataka o emisijama CO₂

Skup podataka o emisijama CO₂ korišten za ovu analizu preuzet je s platforme Kaggle [2] i sadrži podatke o potrošnji energije, emisijama i različitim vrstama energije u različitim državama u periodu od 1980. do 2020.. Skup podataka pruža sveobuhvatan pregled energetskeg profila država (Slika 8), uključujući detalje kao što su emisije povezane s ugljem, prirodnim gasom, naftom, obnovljivim izvorima i nuklearnom energijom (*Energy_type*). Pored toga, prate se ključni pokazatelji kao što su energetske intenzitet po glavi stanovnika (*Energy_intensity_per_capita*) i prema BDP-u (*Energy_intensity_by_GDP*), što pruža uvid u to kako je upotreba energije povezana s populacijom i ekonomskim rezultatima.

Column1	Country	Energy_type	Year	Energy_consumption	Energy_production	GDP	Population	Energy_intensity_per_capita	Energy_intensity_by_GDP	CO2_emission
0	World	all_energy_type	1980	292.8997896	296.3372276	27770.91	4298126.522	68.14592081	10.54699996	4946.62713
1	World	coal	1980	78.65613403	80.11419429	27770.91	4298126.522	68.14592081	10.54699996	1409.790188
2	World	natural_gas	1980	53.8652233	54.76104559	27770.91	4298126.522	68.14592081	10.54699996	1081.593377
3	World	petroleum_n_o	1980	132.0640194	133.1111089	27770.91	4298126.522	68.14592081	10.54699996	2455.243565
4	World	nuclear	1980	7.575700462	7.575700462	27770.91	4298126.522	68.14592081	10.54699996	0
5	World	renewables_n	1980	20.70234415	20.77517837	27770.91	4298126.522	68.14592081	10.54699996	0
6	Afghanistan	all_energy_type	1980	0.026583217	0.072561156		13356.5	1.990283134		0
7	Afghanistan	coal	1980	0.002479248	0.002355286		13356.5	1.990283134		0
8	Afghanistan	natural_gas	1980	0.002094	0.06282		13356.5	1.990283134		0
9	Afghanistan	petroleum_n_o	1980	0.014624098	0		13356.5	1.990283134		0
10	Afghanistan	nuclear	1980				13356.5	1.990283134		0
11	Afghanistan	renewables_n	1980	0.00738587	0.00738587		13356.5	1.990283134		0

Slika 8 Prikaz nekoliko prvih redova skupa podataka o emisijama CO2 "Countries CO2 Emission and more..."

Obzirom na to da skup podataka o vremenskim prilikama pokriva specifične evropske države, podaci o emisijama CO2 su ograničeni na 28 evropskih država prisutnih u oba skupa kako bi se očuvala relevantnost. Države izvan ove regije su isključene jer njihovi podaci nisu relevantni za analizu koja se fokusira na energetska održivost u kontekstu vremenskih uslova u Evropi.

Radi pojednostavljenja analize, vrste energije su grupisane u dvije kategorije: obnovljive (uključujući nuklearnu energiju) i neobnovljive (ugalj, prirodni gas i naftu). Ovaj pristup prilagođava podatke za fokus na održivost i dugoročnu energetska efikasnost postrojenja. Pregled strukture podataka (Slika 9):

- **Country:** Država za koju su podaci zabilježeni.
- **Year:** Godina u kojoj su podaci zabilježeni.
- **Energy_type:** Obnovljivi i neobnovljivi izvori energije. Skup podataka uključuje agregirani red za sve vrste energije, koji objedinjuje ukupnu potrošnju i proizvodnju za obnovljive i neobnovljive izvore.
- **Energy_consumption:** Ukupna potrošnja energije za svaku vrstu, mjerenu u kvadrilionima britanskih termalnih jedinica (quad Btu).
- **Energy_production:** Proizvodnja energije za svaku vrstu energije mjerena u quad Btu, omogućavajući poređenje između količine proizvedene i potrošene energije.
- **GDP:** Bruto domaći proizvod države, mjereno u milijardama dolara iz 2015. godine po paritetu kupovne moći (PPP). Pruža ekonomski kontekst u analizi podataka o energiji.
- **Population:** Populacija države, mjerena u hiljadama ljudi, pomaže u normalizaciji podataka u odnosu na veličinu stanovništva.
- **CO2_emission:** Oznaka ukupnih emisija CO2 iz potrošnje energije za svaku vrstu, mjerena u milionima metričkih tona (MMtonnes CO2).

Country	Year	Energy_type	Energy_consumption	Energy_production	GDP	Population	CO2_emission
Austria	1980	all_energy_types	1.09738269	0.477056917	212.4871	7550.408	0
Austria	1980	non-renewables	0.783267947	0.17778893	637.4613	22651.224	0
Austria	1980	renewables	0.314114744	0.299267987	424.9742	15100.816	0
Austria	1981	all_energy_types	1.086021611	0.470869747	212.2023	7572.555	55.26293027
Austria	1981	non-renewables	0.756221817	0.153903349	636.6069	22717.665	55.26293027
Austria	1981	renewables	0.329799793	0.316966398	424.4046	15145.11	0
Austria	1982	all_energy_types	1.033423966	0.472787608	216.2571	7571.652	51.59746664
Austria	1982	non-renewables	0.705295443	0.155006775	648.7713	22714.956	51.59746664
Austria	1982	renewables	0.328128524	0.317780833	432.5142	15143.304	0
Austria	1983	all_energy_types	1.040713209	0.46138511	222.1881	7561.34	48.86900785

Slika 9 Prikaz nekoliko prvih redova skupa podataka o emisijama CO2 nakon izvršene predobrade

3.2. Podjela podataka na trening i testne skupove

Nakon pripreme skupa podataka o emisijama CO2, izvršena je evaluacija performansi modela slučajne šume koristeći podjelu 80:20% za treniranje i testiranje modela. Cilj ovog testiranja je provjeriti performanse modela na podskupu podataka prije nego što se model trenira na cjelokupnom skupu podataka. Slično pristupu korištenom za vremenske podatke, početni fokus je bio na Njemačkoj.

Skup podataka je filtriran kako bi uključivao samo podatke za Njemačku. Podaci su podijeljeni prema vrsti energije na obnovljive i neobnovljive izvore kako bi se njihovi utjecaji analizirali odvojeno. Za svaku vrstu energije pripremljene su ulazne karakteristike (potrošnja energije) i ciljne varijable (emisije CO2, proizvodnja energije, BDP i populacija). Ova podjela omogućila je dublje razumijevanje kako potrošnja obnovljive i neobnovljive energije utječe na ove faktore.

3.3. Primjena modela slučajne šume

Koristeći regresiju slučajne šume, trenirani su zasebni modeli za obnovljive i neobnovljive izvore energije. Modeli su trenirani za predviđanje nekoliko ciljnih varijabli (emisije CO2, proizvodnja energije, BDP i populacija) na osnovu potrošnje energije kao ulazne karakteristike. Slučajna šuma je metoda učenja koja kombinuje predikcije više stabala odlučivanja kako bi se proizvele tačnije i stabilnije predikcije. Umjesto da se oslanja na jedno stablo odlučivanja, koje može biti sklono pretreniravanju (model previše prilagođen treniranim podacima koji loše generalizira na novim podacima), slučajna šuma gradi više stabala na različitim nasumičnim podskupovima podataka. Svako stablo daje svoju predikciju, a konačni rezultat je prosjek ovih predikcija (u zadacima regresije - Random Forest Regression) ili većinska odluka (u zadacima klasifikacije - Random Forest Classification).

Skup podataka prvo se dijeli na trening i testne skupove. Model se trenira na 80% podataka, a njegova učinkovitost se procjenjuje na preostalih 20% kako bi se spriječilo pretreniravanje. Svako stablo odlučivanja u slučajnoj šumi gradi se koristeći nasumični podskup podataka za treniranje. Ova metoda je poznata kao "*bagging*" [11]. U regresiji, svako stablo predviđa kontinuiranu vrijednost (u ovom slučaju, emisije CO2, proizvodnju energije, BDP ili populaciju na osnovu potrošnje energije).

Svako stablo se gradi rekurzivnim dijeljenjem podataka na različite karakteristike (u ovom slučaju, potrošnja energije je jedina karakteristika) kako bi se minimizirala varijansa ciljne varijable (na primjer, emisije CO₂). Podjele se nastavljaju dok se ne ispuni kriterij zaustavljanja (npr. minimalni broj uzoraka u listu ili maksimalna dubina stabla). `n_estimators=100` označava da model slučajne šume treba izgraditi 100 stabala odlučivanja.

```
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
```

Nakon što su sva stabla trenirana, model kombinuje njihove predikcije uzimanjem prosjeka (budući da se radi o regresijskom zadatku). Ovo pomaže da se predikcije izglate i smanji vjerovatnoća pretreniravanja. Na primjer, za emisije CO₂, svako od 100 stabala će proizvesti predikciju za testni skup, a konačna predikcija će biti prosjek tih predikcija:

$$\hat{y}_{final} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i$$

Gdje je \hat{y}_i predikcija svakog pojedinačnog stabla odlučivanja, a n broj stabala u šumi.

3.4. Značaj svojstva (eng. *feature*) u modelu slučajne šume

U regresiji slučajne šume, važnost svojstva (eng. *feature*) se odnosi na to koliko svako svojstvo doprinosi tačnosti predikcija. Značaj svojstva određuje se na osnovu toga koliko ono smanjuje varijansu (ili grešku) u predviđenim vrijednostima kroz sva stabla odlučivanja u šumi. Za svako stablo odlučivanja, algoritam slučajne šume dijeli podatke na različitim tačkama na osnovu vrijednosti ulazne karakteristike (u ovom slučaju, potrošnja energije). Svaka podjela smanjuje varijansu ciljne varijable (poput emisije CO₂ ili proizvodnje energije), a važnost karakteristike se računa prema tome koliko smanjuje varijansu kroz sva stabla u šumi.

U algoritmu slučajne šume, procjenjuju se sve moguće podjele podataka. Za svako svojstvo se računa koliko ono pomaže u smanjenju nečistoće (u regresiji, to se odnosi na varijansu). Što više svojstvo pomaže u smanjenju varijanse prilikom podjela, veća je njegova važnost. Nakon što se izgradi šuma stabala odlučivanja, važnost svake karakteristike se prosječno izračunava kroz sva stabla, dajući ukupnu ocjenu važnosti. U ovom slučaju, budući da je potrošnja energije jedina ulazna karakteristika, ona će imati punu važnost, odnosno 1.0. To znači da su sve predikcije modela (za emisije CO₂, proizvodnju energije, BDP i populaciju) u potpunosti zasnovane na podacima o potrošnji energije. Nema drugih karakteristika koje bi model mogao procijeniti ili usporediti u smislu važnosti. Navedena linija koda dohvata vrijednosti važnosti karakteristika iz treniranog modela slučajne šume (`rf_model`). Budući da postoji samo jedna karakteristika (potrošnja energije), njena važnost će biti 1.0 (100%):

```
feature_importances = rf_model.feature_importances_
```

Razumijevanje važnosti svojstava pomaže u interpretaciji odluka modela. U ovom kontekstu, model se u potpunosti oslanja na potrošnju energije za predviđanje ciljnih varijabli. Da postoje dodatne karakteristike (poput cijena energije ili tehnoloških napredaka), one bi također mogle doprinijeti predikcijama, a važnost svojstava bi pokazala koliko svaka od tih karakteristika utiče na predikcije modela.

3.5. Metrike performansi modela slučajne šume

Nakon što model napravi predikcije, njegove performanse se procjenjuju korištenjem metrika poput srednje kvadratne greške i koeficijenta determinacije (R^2) [12]. MSE, koji je prethodno detaljnije objašnjen, mjeri koliko su predviđene vrijednosti blizu stvarnim vrijednostima. Niži MSE ukazuje na bolje prilagođavanje modela. Koeficijent determinacije mjeri koliko dobro model objašnjava varijabilnost ciljne varijable. Vrijednost od 1 ukazuje na savršene predikcije, dok 0 znači da model uopšte ne objašnjava varijabilnost.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Gdje je:

y_i stvarna vrijednost ciljne varijable,

\hat{y}_i predviđena vrijednost iz modela,

\bar{y} prosječna vrijednost stvarnih vrijednosti,

n ukupan broj tačaka podataka.

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Predstavlja zbir kvadratnih ostataka ili kvadratne razlike između stvarnih i predviđenih vrijednosti.

$$\sum_{i=1}^n (y_i - \bar{y})^2$$

Predstavlja ukupni zbir kvadrata, koji mjeri ukupnu varijabilnost ciljne varijable.

R^2 daje brz i interpretabilan pokazatelj koliko dobro model odgovara podacima:

- $R^2 = 1$ znači da model savršeno predviđa ciljnu varijablu (sve predikcije tačno odgovaraju stvarnim vrijednostima).
- $R^2 = 0$ znači da model ne objašnjava nikakvu varijabilnost (nije bolji od pukog predviđanja srednje vrijednosti ciljne varijable).
- $R^2 < 0$ ukazuje na to da model radi lošije od jednostavnog predviđanja srednje vrijednosti.

Tabela 1 Rezultati regresije slučajne šume za podatke o obnovljivoj energiji (renewables) - Važnost karakteristike potrošnja energije (Energy_consumption) = 1.0

	MSE	R^2	Varijansa
CO2_emission	0.0	1.0	0.0
Energy_production	0.0028	0.9958	0.6582
GDP	52945.61	0.9902	5407747.1162
Population	521246.3717	0.9998	2895747297.6883

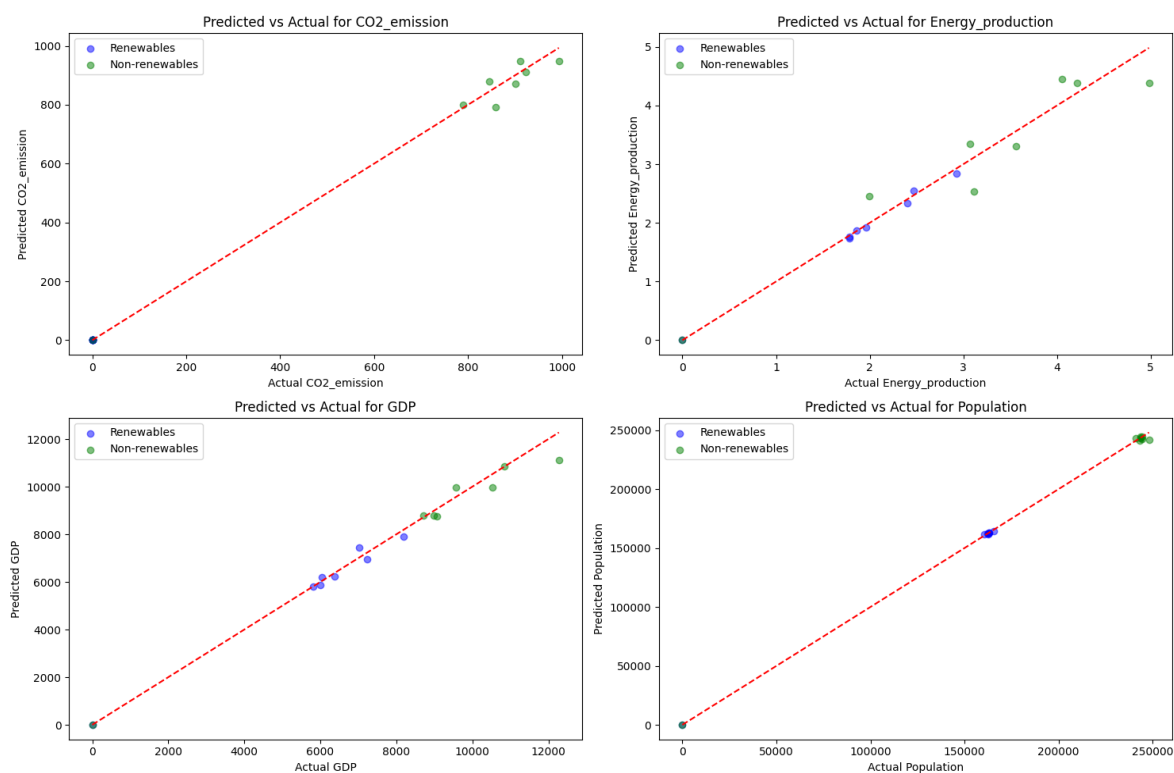
Za obnovljive izvore energije, model je pokazao izuzetno dobre rezultate za sve ciljne varijable (Tabela 1). Vrijednosti R^2 su blizu ili tačno 1.0, što znači da model zadovoljavajuće objašnjava varijabilnost u podacima. R^2 za emisije CO2 iznosi 1.0, što ukazuje na savršene predikcije. MSE za emisije CO2 je 0.0, što pokazuje da model nije napravio nijednu grešku u predikciji za ovu ciljnu varijablu. Obzirom da obnovljivi izvori energije ne proizvode emisije, ovaj rezultat je bio očekivan. Ostale ciljne varijable, kao što su proizvodnja energije ($R^2 = 0.996$) i BDP ($R^2 = 0.990$), također pokazuju visoku prediktivnu snagu uz niske vrijednosti MSE-a. Ovo sugerira da je potrošnja energije iz obnovljivih izvora visoko prediktivna za ove ishode.

Tabela 2 Rezultati regresije slučajne šume za podatke o neobnovljivoj energiji (non-renewables) - Važnost karakteristike potrošnja energije (Energy_consumption) = 1.0

	MSE	R^2	Varijansa
CO2_emission	1278.6616	0.9857	9514.9157
Energy_production	0.1532	0.9268	2.0912
GDP	243600.3815	0.97998	12167431.0115
Population	6411380.2162	0.9990	6515431419.7986

Za obnovljive izvore energije performanse modela su ostale i dalje dobre (Tabela 2). MSE za emisije CO₂ je 1278.66, dok je R^2 0.986, što znači da je model i dalje objašnjava većinu varijanse. Slično tome, R^2 za proizvodnju energije iznosi 0.927, što je nešto niže od modela za obnovljive izvore, ali i dalje snažno. Sveukupno, rezultati pokazuju da je model slučajne šume učinkovit u predviđanju ishoda na osnovu potrošnje energije, posebno za obnovljive izvore, gdje su predikcije izuzetno zadovoljavajuće. Za obnovljive izvore, model također pokazuje dobre performanse, iako veća varijabilnost u podacima o obnovljivim izvorima rezultira nešto višim greškama. Ipak, model je i dalje visoko prediktivan i pouzdan za obje vrste energije. Ovo je također vidljivo iz grafičkog prikaza preformansi (Slika 10), gdje se predviđene vrijednosti zadovoljavajuće podudaraju sa stvarnim vrijednostima, pokazujući sposobnost modela da napravi tačne predikcije za emisije CO₂, proizvodnju energije, BDP i populaciju kod obje vrste energije.

Model regresije slučajne šume je posebno pogodan za ovu analizu jer efikasno zahvata složene, nelinearne odnose između potrošnje energije i ciljnih varijabli poput emisija CO₂, proizvodnje energije, BDP-a i populacije. Za razliku od linearne regresije, koja pretpostavlja linearan odnos između varijabli, slučajna šuma može modelirati složenije obrasce koji postoje u stvarnim podacima. Pored toga, omogućava procjenu važnosti karakteristika, pružajući uvid u to koliko snažno potrošnja energije utiče na ciljne ishode. Još jedna ključna prednost slučajne šume je sposobnost minimiziranja pretreniravanja. Kroz izgradnju više stabala odlučivanja i prosječenje njihovih predikcija, model se bolje generalizira na neviđene podatke, smanjujući rizik pretreniravanja na šum u skupu podataka za treniranje.



Slika 10 Grafički prikaz odnosa predviđenih i stvarnih vrijednosti za obnovljive i neobnovljive izvore energije

3.6. Treniranje na cjelokupnom skupu podataka

Nakon što je potvrđeno da se model slučajne šume dobro izvodi na testnim podacima, pristupa se korištenju cjelokupnog skupa podataka za generisanje predikcija za narednih 10 godina za različite države. Cilj je prognozirati BDP, populaciju, emisije CO2 i proizvodnju energije za obnovljive i neobnovljive izvore energije. Nakon učitavanja cjelokupnog skupa podataka (*energy_cleaned.csv*) korišten je rječnik za mapiranje kako bi se nazivi država konvertovali u standardizirane kodove. Ovo mapiranje je potrebno za osiguranje dosljedne reference između različitih skupova podataka i za buduće predikcije, obzirom da su države u skupu podataka o vremenskim prilikama predstavljene kroz standardizirane kodove.

Izvršeno je filtriranje podataka za svaku državu kako bi se izdvojili energetske podaci specifični za tu državu, a zatim su trenirani modeli za predviđanje budućeg BDP-a (*GDP*) i populacije (*Population*) na osnovu potrošnje energije za sve vrste energije (obnovljive i neobnovljive). Da bi se ovo postiglo, model *RandomForestRegressor* je treniran na ukupnoj potrošnji energije te države i korišten je za predviđanje BDP-a i populacije u narednih 10 godina:

```
X_all_energy = all_energy_data[features]
y_all_energy = all_energy_data[['GDP', 'Population']]

rf_all_energy_model = RandomForestRegressor(n_estimators=100,
random_state=42)
rf_all_energy_model.fit(X_all_energy, y_all_energy)

future_predictions_gdp_pop =
rf_all_energy_model.predict(future_data_gdp_pop)
```

Predikcije za obnovljive i neobnovljive izvore energije obrađene su odvojeno. Za svaku vrstu energije izdvojeni su podaci vezani za potrošnju energije i korišteni za predviđanje emisija CO2 (*CO2_emission*) i proizvodnje energije (*Energy_production*). Za svaku vrstu energije treniran je poseban *RandomForestRegressor* model kako bi se uzeli u obzir različiti načini na koje obnovljivi i neobnovljivi izvori doprinose proizvodnji energije i emisijama:

```
X_energy = energy_data[features]
y_energy = energy_data[['CO2_emission', 'Energy_production']]

rf_energy_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_energy_model.fit(X_energy, y_energy)

future_predictions_energy = rf_energy_model.predict(future_data_energy)
```

Nakon generisanja predikcija za svaku državu, sačuvane su prosječne vrijednosti za proizvodnju energije i emisije CO2 iz obnovljivih i neobnovljivih izvora. Predikcije za BDP i populaciju su pohranjene u strukturiranom formatu, omogućavajući analizu dugoročne održivosti energetske upotrebe u svakoj državi:

```
future_predictions['avg_production_renewables'].append(avg_production)
future_predictions['avg_production_non_renewables'].append(avg_production)
future_predictions['avg_CO2_emissions'].append(avg_emissions)

# Append GDP and Population (they will be the same for both energy types)
future_predictions['GDP'].append(future_gdp)
future_predictions['Population'].append(future_population)
```

Ova metoda omogućava projekciju budućeg BDP-a, populacije, emisija CO2 i proizvodnje energije za više evropskih država koristeći regresiju slučajne šume. Razdvajanjem predikcija za obnovljive i neobnovljive izvore energije omogućena je procjena potencijalnih ekoloških i ekonomskih utjecaja oba tipa energije u narednoj deceniji, pružajući detaljniji uvid u energetske strategije svake države. Ove predikcije su ključne za određivanje država koje su najpogodnije za dugoročno održive lokacije za postrojenja.

4. Odluke zasnovane na težinskim kriterijima za optimalno smještanje poslovnih objekata

Cilj izrade prethodno opisana dva modela mašinskog učenja je rangirati države prema njihovoj prikladnosti za smještaj različitih vrsta poslovnih operacija, poput fabrika, industrijskih postrojenja, IT kompanija ili serverskih objekata. Evaluacija se temelji na kombinaciji trenutnih i budućih predikcija ključnih faktora [6] kao što su klimatska stabilnost, proizvodnja energije, emisije CO₂, ekonomski uslovi i dostupnost radne snage. Uzimajući u obzir kako trenutne podatke, tako i predviđanja za budućnost, ovaj pristup pruža sveobuhvatnu i dugoročnu procjenu potencijala svake države kao idealne lokacije za održive i efikasne poslovne operacije.

Analiza uključuje višestruke kriterije, poput trenutnih prosječnih temperatura, nivoa solarne radijacije, proizvodnje obnovljive i neobnovljive energije, BDP-a i populacije, kako bi se procijenili trenutni uslovi. Predikcije, dobivene na osnovu modela treniranih na podacima o vremenu, energiji i emisijama za narednih deset godina, se koriste kako bi se predvidjeli trendovi u ovim faktorima. Rezultirajući rangovi pružaju uvide u to koje će države zadržati ili poboljšati povoljne uslove za poslovanje, čineći ih najoptimalnijim izborom za buduće industrijske ili korporativne investicije.

4.1. Definisanje kriterija za odabir lokacije

Model razvijen za rangiranje država prema optimalnoj lokaciji za smještaj poslovnih objekata temelji se na kombinaciji trenutnih i predikovanih podataka. Kriteriji obuhvataju različite faktore koji su ključni za održivost države kao domaćina industrijskih operacija, s naglaskom na klimatsku stabilnost, održivost i ekonomske performanse. U nastavku su predstavljeni glavni kriteriji korišteni u evaluaciji:

4.1.1. Trenutna temperatura i solarna radijacija

Klimatska stabilnost igra ključnu ulogu u energetskej efikasnosti. Države s umjerenim temperaturama zahtijevaju manje energije za grijanje i hlađenje [1], što je posebno važno za industrije poput proizvodnje ili podatkovnih centara, gdje je kontrola temperature značajan dio operativnih troškova. Solarna radijacija je bitna za procjenu potencijala za proizvodnju obnovljive energije. Korištenje solarne energije može značajno smanjiti ovisnost o neobnovljivim izvorima, čime se smanjuju troškovi i podržavaju ciljevi održivosti. Države s umjerenom klimom i visokim nivoima solarne radijacije su povoljnije jer mogu smanjiti potrošnju energije i operativne troškove.

4.1.2. Emisije CO₂ i proizvodnja obnovljive energije

Povećani naglasak na održivosti zahtijeva od poslovanja da procijeni svoj ekološki otisak. Države s niskim emisijama CO₂ pokazuju posvećenost smanjenju stakleničkih gasova, usklađujući se s globalnim ekološkim ciljevima i osiguravajući poštivanje međunarodnih regulativa. Države s visokim nivoima proizvodnje obnovljive energije nude mogućnost integracije ekološki prihvatljivih izvora energije. Lokacije s niskim emisijama CO₂ i visokom proizvodnjom obnovljive energije su poželjne za dugoročne operacije zbog održivosti i smanjenja troškova.

4.1.3. BDP i populacija

Ekonomska stabilnost i dostupnost radne snage su ključni faktori za poslovni uspjeh. BDP je pokazatelj ekonomske snage države, odražavajući infrastrukturu, kupovnu moć i sposobnost podržavanja velikih poslovnih operacija. Populacija je pokazatelj dostupnosti radne snage, osiguravajući da preduzeća mogu naći kvalificiranu radnu snagu. Države sa stabilnim ili rastućim ekonomijama i velikom populacijom nude neophodne resurse, uključujući kvalificiranu radnu snagu i stabilne ekonomske uvjete za dugoročno poslovanje.

4.1.4. Predviđena temperatura i solarna radijacija

Dugoročne klimatske prognoze su važne za strateška ulaganja. Predviđeni temperaturni trendovi ukazuju na potencijalne promjene u obrascima potrošnje energije, dok je predviđena solarna radijacija ključna za dugoročne procjene potencijala proizvodnje solarne energije. Razmatranje budućih klimatskih trendova pomaže u izbjegavanju lokacija koje bi mogle biti pogođene ekstremnim vremenskim uvjetima. Države sa stabilnim projekcijama solarne radijacije nude dugoročnu mogućnost korištenja obnovljive energije.

4.1.5. Predviđene emisije CO₂, proizvodnja obnovljive energije, BDP i populacija

Iako su trenutni ekonomski i ekološki faktori važni, buduće projekcije osiguravaju dugoročnu održivost. Države koje pokazuju pozitivne trendove u smanjenju emisija CO₂ i povećanju proizvodnje obnovljive energije demonstriraju proaktivan pristup održivosti. Pored toga, predviđeni rast BDP-a i stabilnost populacije osiguravaju da će država i dalje biti pouzdana lokacija za poslovanje. Države s pozitivnim dugoročnim projekcijama u pogledu održivosti i ekonomske snage pružaju stabilno okruženje za poslovanje, osiguravajući kontinuiranu efikasnost i rast.

Ovi kriteriji su odabrani kako bi osigurali sveobuhvatnu evaluaciju trenutnih i budućih uvjeta svake države. Stabilna klima i potencijal za obnovljive izvore energije ključni su za smanjenje operativnih troškova i podršku ciljevima održivosti. Niske emisije CO₂ i visoka proizvodnja obnovljive energije pružaju uvid u ekološku održivost energetske mreže države. Ekonomska stabilnost i dostupnost radne snage su ključni faktori za dugoročnu održivost bilo koje industrijske ili poslovne operacije. Težinskom dodjelom ovih kriterija, model osigurava da odabrana lokacija nije samo idealna za trenutne potrebe, već i otporna i prilagodljiva budućim izazovima.

4.2. Normalizacija podataka

Normalizacija podataka je ključni korak u preprocesiranju prije dodjeljivanja težina različitim kriterijima. Primarni cilj normalizacije je dovesti sve varijable na uporedivu skalu, osiguravajući da nijedan pojedinačni kriterij ne dominira konačnim rangiranjem zbog razlika u jedinicama ili skalama. Na primjer, emisije CO₂ se mogu mjeriti u milionima tona, dok se BDP mjeri u milijardama dolara, a populacija u milionima ljudi. Bez normalizacije, veće numeričke vrijednosti, poput BDP-a, neproporcionalno bi utjecale na konačni rezultat, zasjenjujući jednako važne faktore manje razmjerne, poput emisija CO₂ ili proizvodnje obnovljive energije.

MinMaxScaler transformiše vrijednosti svake karakteristike na skalu između 0 i 1, pri čemu zadržava relativne razlike između vrijednosti, dok ih standardizuje na dosljedan raspon. Ova metoda osigurava da svi kriteriji, bilo da se odnose na ekonomsku stabilnost ili ekološku održivost, ravnopravno doprinose konačnom rangiranju:

```
scaler = MinMaxScaler()
data[current_criteria_cols] =
scaler.fit_transform(data[current_criteria_cols])
data[predicted_criteria_cols] =
scaler.fit_transform(data[predicted_criteria_cols])
```

Primjenom ove transformacije na trenutne i predviđene kriterije, model osigurava uravnotežen i pravedan sistem dodjeljivanja težina. Ovaj korak je ključan za postizanje tačnog i reprezentativnog rangiranja država prema njihovom potencijalu za smještaj poslovnih operacija, osiguravajući da nijedan kriterij ne bude nenamjerno previše ili premalo naglašen zbog razlika u skali.

4.3. Dodjeljivanje težina

Kako bi se uravnotežili različiti kriteriji za rangiranje država, svakom kriteriju se dodjeljuju težine na osnovu njegove percipirane važnosti. Ove težine odražavaju značaj svakog faktora u određivanju optimalne lokacije za poslovne operacije, s naglaskom na održivost, energetska efikasnost i ekonomsku stabilnost:

```
weights = {
    'average_temperature': 0.10, 'average_radiation': 0.10,
    'avg_CO2_emissions': -0.15, 'avg_production_renewables': 0.15,
    'GDP': 0.075, 'Population': 0.075,
    'predicted_avg_temp': 0.10, 'predicted_avg_radiation': 0.10,
    'predicted_avg_CO2_emissions': -0.15,
    'predicted_avg_production_renewables': 0.15,
    'predicted_GDP': 0.075, 'predicted_Population': 0.075
}
```

Održivost je ključna komponenta modernih poslovnih operacija, zbog čega su veće težine dodijeljene faktorima poput proizvodnje obnovljive energije (0.15) i emisija CO2 (-0.15). Ovime se naglašava važnost čistih izvora energije i smanjenja ugljikovog otiska, što je u skladu s globalnim ciljevima održivosti. Temperatura i solarna radijacija, kako trenutne tako i predviđene, dobile su težinu od 0.10. Ovi faktori su presudni za procjenu dugoročne energetske efikasnosti i održivosti lokacije. Ekonomski faktori poput BDP-a i populacije, koji osiguravaju kvalificiranu radnu snagu i stabilno ekonomsko okruženje, dobili su nešto niže težine (0.075), čime se priznaje njihova važnost bez zasjenjivanja ekoloških prioriteta.

U ovom modelu, težine su dodijeljene tako da njihov zbir iznosi 1, što pruža nekoliko prednosti u pogledu jasnoće i ravnoteže:

- **Interpretabilnost:** Svaka težina direktno predstavlja postotak doprinosa određenog kriterija ukupnom rezultatu. Na primjer, težina od 0.10 pokazuje da određeni faktor doprinosi 10% konačnoj odluci, čineći ulogu svakog kriterija u evaluacijskom procesu jasnom i intuitivnom.
- **Relativna važnost svakog kriterija:** Kada zbir težina iznosi 1, postaje lakše uporediti koliko jedan faktor utiče u odnosu na drugi. Na primjer, ako emisije CO2 imaju težinu -0.15, a BDP 0.075, to pokazuje da su emisije CO2 dvostruko važnije (iako negativno, zbog negativne težine) od BDP-a u procesu donošenja odluka. Ova proporcionalna reprezentacija omogućava smislene usporedbe između različitih kriterija.
- **Skaliranje ulaznih podataka:** Korištenje metoda poput MinMaxScaler osigurava da su sve vrijednosti ujednačene (u rasponu od 0 do 1), omogućavajući težinama da odraze uticaj kriterija bez obzira na njihove originalne skale. Iako bi težine tehnički mogle biti bilo koje vrijednosti, osiguravanje da njihov zbir bude 1 garantuje da je doprinos svakog kriterija uravnotežen i usklađen s ciljevima modela.

Važno je napomenuti da se ove težine mogu prilagoditi specifičnim ciljevima i potrebama preduzeća. Kompanije fokusirane na ekonomsku efikasnost mogu povećati težinu BDP-a ili populacije, dok one s naglaskom na održivost mogu dati prednost obnovljivoj energiji i smanjenju emisija CO2. Trenutna dodjela težina je primjer koji se može prilagoditi strateškim ciljevima i operativnim prioritetima preduzeća. Ova fleksibilnost osigurava da se model može prilagoditi potrebama bilo koje organizacije, bilo da daje prednost održivosti, ekonomskom učinku ili dostupnosti radne snage u svom procesu donošenja odluka.

4.4. Izračun težinskog rezultata

Kako bi se države rangirale prema njihovoj prikladnosti za poslovne operacije, za svaku od njih je izračunat težinski zbir kombiniranjem trenutnih i predviđenih budućih vrijednosti. Ovaj sveobuhvatni pristup omogućava procjenu ne samo trenutnih performansi države, već i njenog potencijala u narednoj deceniji, uzimajući u obzir faktore poput klimatskih uvjeta, ekonomske stabilnosti i održivosti. Izračun uključuje dodjeljivanje težine svakom kriteriju, kao što je definirano

u modelu, te množenje svakog kriterija s pripadajućom težinom. Za svaku državu, trenutne i buduće vrijednosti ovih kriterija kombinirane su korištenjem sljedeće formule:

```
data['Score'] = (
    data['average_temperature'] * weights['average_temperature'] +
    data['average_radiation'] * weights['average_radiation'] +
    data['avg_CO2_emissions'] * weights['avg_CO2_emissions'] +
    data['avg_production_renewables'] *
weights['avg_production_renewables'] +
    data['GDP'] * weights['GDP'] +
    data['Population'] * weights['Population'] +
    data['predicted_avg_temp'] * weights['predicted_avg_temp'] +
    data['predicted_avg_radiation'] * weights['predicted_avg_radiation'] +
    data['predicted_avg_CO2_emissions'] *
weights['predicted_avg_CO2_emissions'] +
    data['predicted_avg_production_renewables'] *
weights['predicted_avg_production_renewables'] +
    data['predicted_GDP'] * weights['predicted_GDP'] +
    data['predicted_Population'] * weights['predicted_Population']
)
```

Ovaj izračun osigurava da se trenutni faktori, poput BDP-a i proizvodnje obnovljive energije, te buduće projekcije, kao što su rast BDP-a i emisije CO₂, uzimaju u obzir pri konačnom rangiranju. Države koje postižu dobre rezultate u sadašnjosti i imaju povoljne prognoze za budućnost, rangiraju se više. Težinski izračun omogućava sveobuhvatnu procjenu dugoročnog potencijala države, pružajući podatkovno utemeljenu osnovu za odabir najprikladnije lokacije. Sabiranjem težinskih kriterija dobija se konačni rezultat koji odražava trenutne performanse države i njen budući potencijal. Ovaj pristup omogućava da se države rangiraju prema trenutnim prednostima i dugoročnoj održivosti, čime se olakšava donošenje strateških odluka o lokaciji poslovnih postrojenja.

Primjer rangiranja država prema prethodno zadanim težinama (Slika 11):

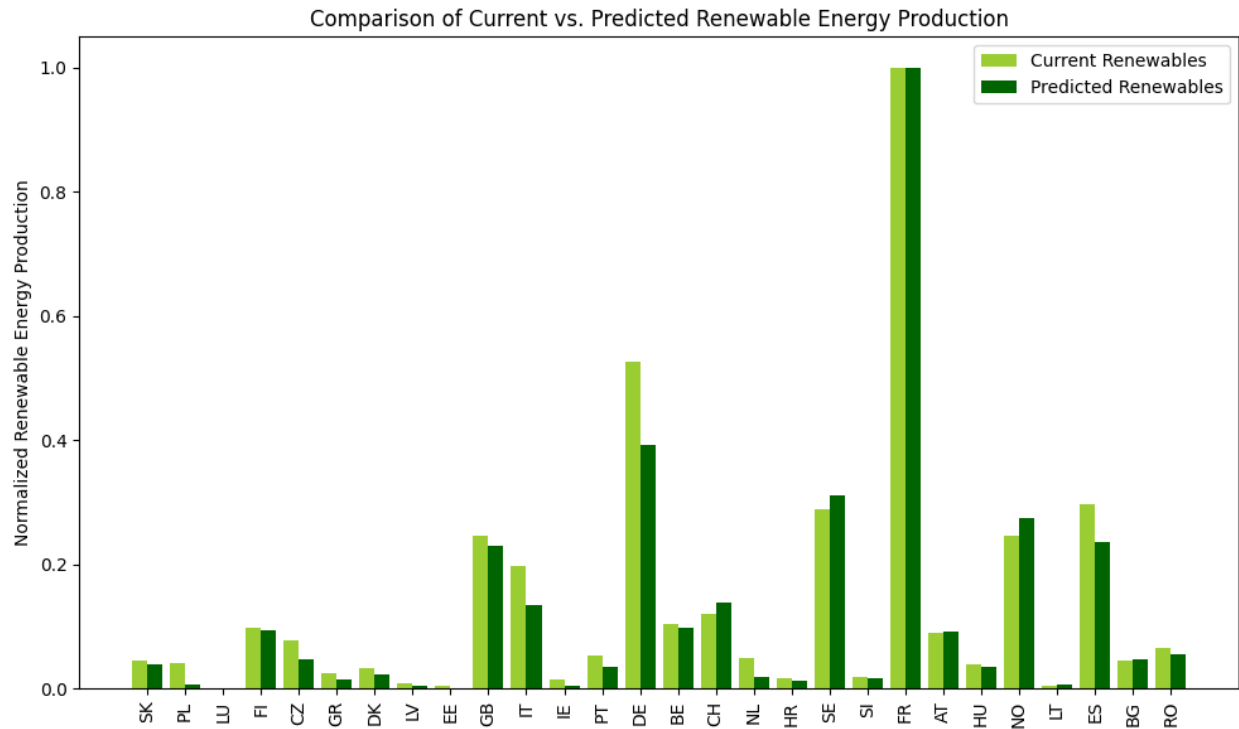
- **Francuska (FR)** se nalazi na prvom mjestu, prvenstveno zahvaljujući visokoj proizvodnji obnovljive energije (Slika 12), niskim emisijama CO₂ u odnosu na populaciju (Slika 14) i snažnoj ekonomskoj stabilnosti i populacijom (Slika 17). Dodatno, buduće projekcije BDP-a i proizvodnje obnovljive energije ukazuju na nastavak rasta i održivosti.
- **Španija (ES)** je na drugom mjestu zahvaljujući snažnoj solarnoj radijaciji (Slika 16), dobroj proizvodnji obnovljive energije (Slika 12) i relativno niskim emisijama CO₂ u odnosu na populaciju (Slika 14). Budući rast u obnovljivoj energiji i stabilni ekonomski uvjeti čine Španiju atraktivnom lokacijom za dugoročne investicije.

- **Italija (IT)** postiže dobar rang zahvaljujući stabilnoj solarnoj radijaciji (Slika 16), proizvodnji obnovljive energije (Slika 12) i budućoj stabilnosti BDP-a (Slika 17) i proizvodnje obnovljive energije. Iako su emisije CO₂ nešto veće, Italija zadržava snažan ukupni rezultat.
- **Finska (FI)** je nisko rangirana zbog niske prosječne radijacije (Slika 16) i umjerene proizvodnje obnovljive energije (Slika 12). Iako ima relativno niske emisije CO₂ (Slika 13), njen budući potencijal za rast u obnovljivoj energiji i ekonomskim pokazateljima je slabiji u odnosu na države s višim rangom.
- **Estonija (EE)** također ima nizak rang, prvenstveno zbog niske trenutne i predviđene proizvodnje obnovljive energije (Slika 12) te slabih ekonomskih pokazatelja (Slika 16). Njena proizvodnja obnovljive energije je minimalna, što ograničava mogućnosti održivosti.
- **Latvija (LV)** je rangirana nisko zbog relativno niske proizvodnje obnovljive energije (Slika 12) i solarne radijacije (Slika 16). Njene buduće projekcije u obnovljivoj energiji i ekonomskom rastu su skromne (Slika 17), što objašnjava niži rang.

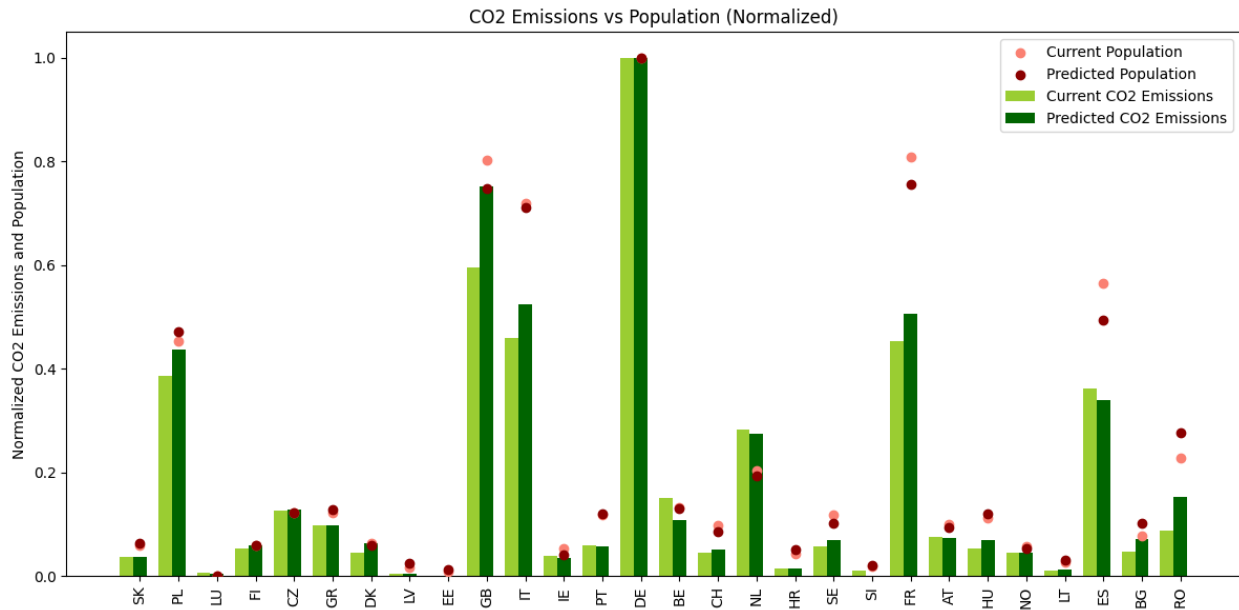
Rank	Country_Code	Score	Ranking	Country_Code	Score
1	FR	0.596867	15	BE	0.201168
2	ES	0.485195	16	SK	0.192716
3	IT	0.407537	17	CZ	0.173208
4	GR	0.402075	18	LU	0.167656
5	PT	0.390343	19	SE	0.163648
6	DE	0.301518	20	NL	0.154463
7	BG	0.284244	21	IE	0.134265
8	RO	0.266003	22	PL	0.133126
9	HR	0.261267	23	DK	0.128839
10	HU	0.245956	24	LT	0.088561
11	CH	0.229784	25	NO	0.083741
12	GB	0.225208	26	LV	0.075608
13	SI	0.214698	27	EE	0.06146
14	AT	0.204842	28	FI	0.037426

Slika 11 Primjer rezultata rangiranja država

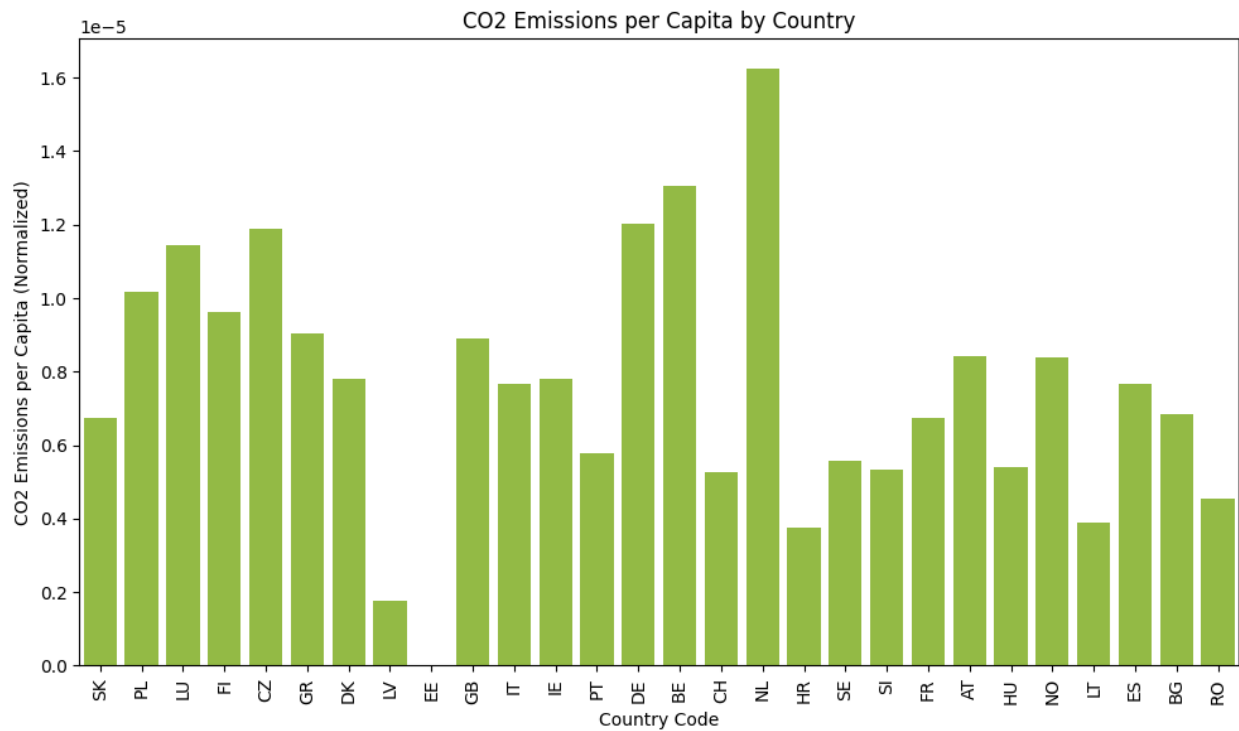
Države koje se dobro pozicioniraju u trenutnim pokazateljima održivosti (poput proizvodnje obnovljive energije i niskih emisija) i u budućim ekonomskim i ekološkim projekcijama dominiraju u vrhu rang liste. Ovaj pristup osigurava da preduzeća ne biraju samo optimalne lokacije na temelju današnjih podataka, već donose i strateške odluke koje će ostati relevantne i u narednoj deceniji.



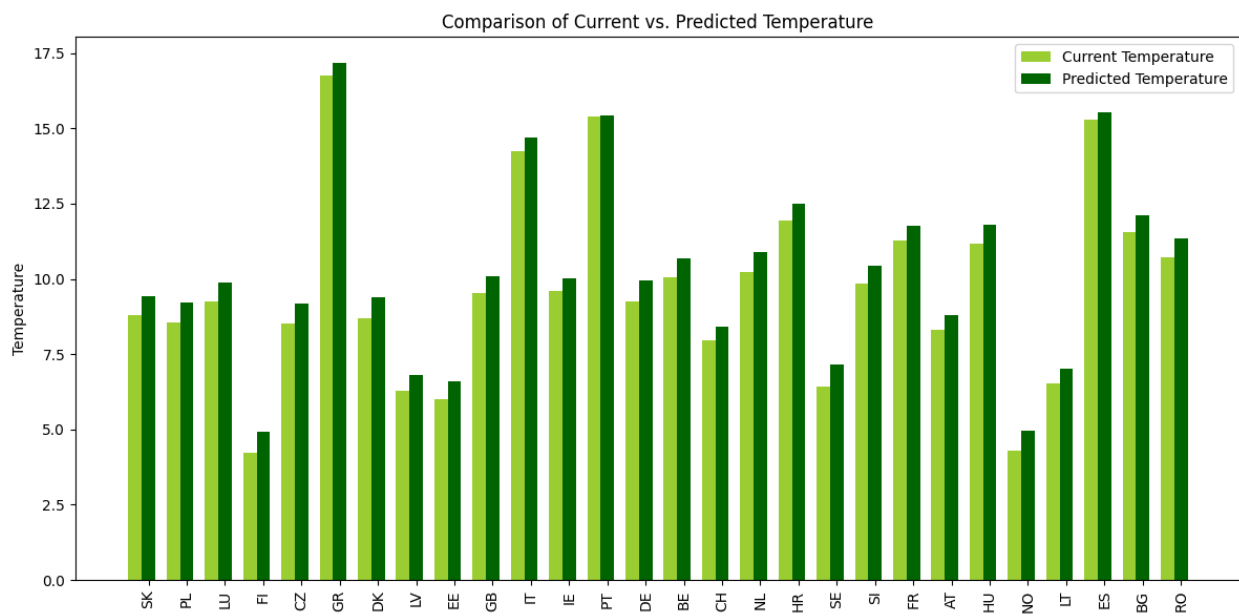
Slika 12 Usporedba trenutne i predviđene proizvodnje obnovljive energije (prosjeak posljednjih 10 u poređenju sa budućih 10 godina)



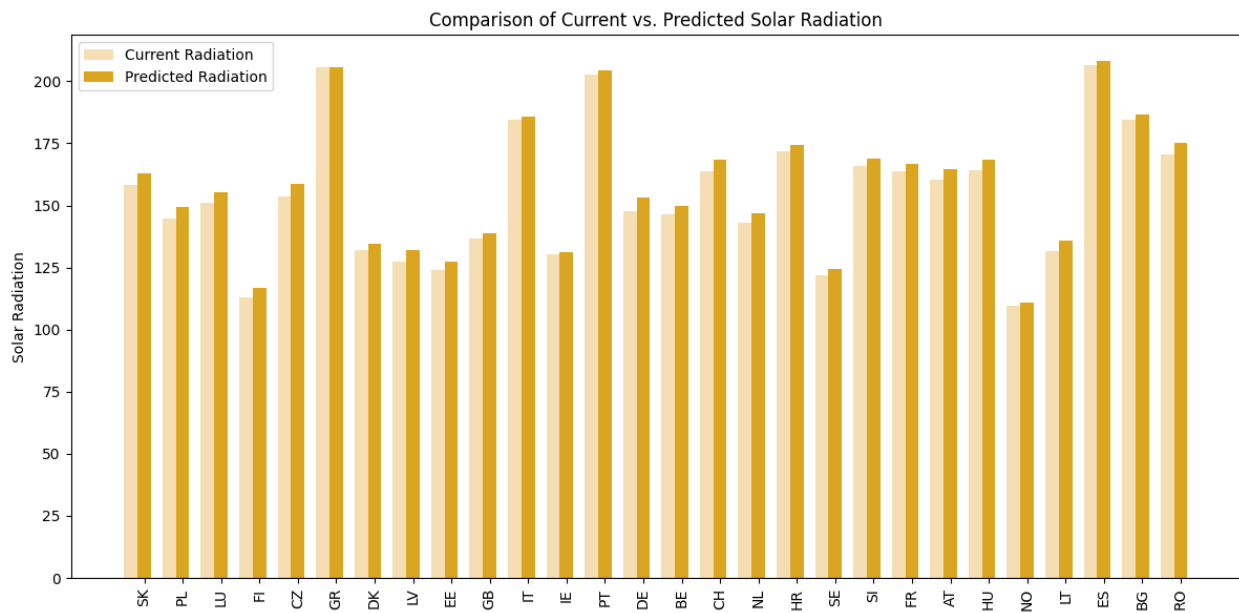
Slika 13 Usporedba trenutnih i predviđenih emisija CO2 (prosjeak posljednjih 10 u poređenju sa budućih 10 godina) u odnosu na populaciju



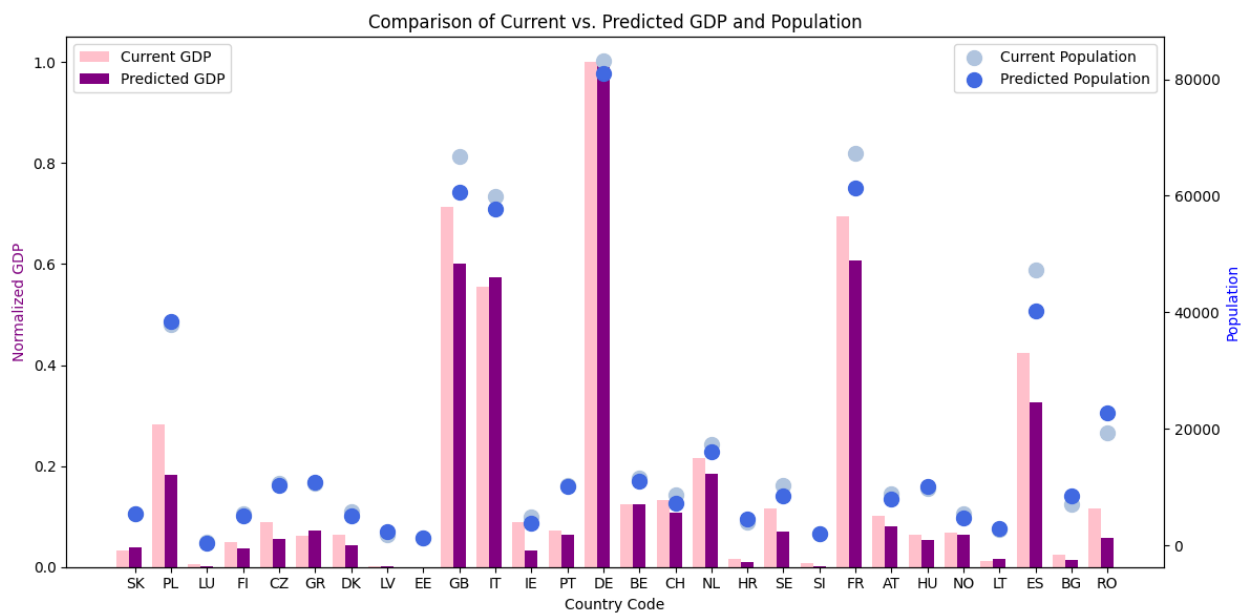
Slika 14 Emisije CO2 po glavi



Slika 15 Usporedba trenutne i predviđene temperature (prosjeak posljednjih 10 u poređenju sa budućih 10 godina)



Slika 16 Usporedba trenutne i predviđene solarne radijacije (prosjek posljednjih 10 u poređenju sa budućih 10 godina)



Slika 17 Usporedba trenutnog (posljednji zapis u skupu podataka) i predviđenog (za 10 godina) BDP-a u odnosu na populaciju

Zaključak

Proces dodjele težinskih vrijednosti pri donošenju odluka opisan u ovom radu omogućava kompanijama da donose odluke zasnovane na podacima prilikom odabira optimalnih lokacija za fabrike, serverske objekte ili druge poslovne operacije. Kombinacijom trenutnih faktora (poput klimatskih uvjeta, emisija CO₂ i ekonomske stabilnosti) i budućih projekcija (predviđeni temperaturni trendovi i potencijal obnovljive energije), ovaj pristup pruža sveobuhvatan pregled prikladnosti svake države. Fleksibilnost modela omogućava prilagođavanje težina dodijeljenih svakom kriteriju, što kompanijama omogućava da evaluaciju prilagode svojim specifičnim potrebama ili zahtjevima industrije. Na primjer, kompanije fokusirane na održivost mogu dodijeliti veće težine proizvodnji obnovljive energije i smanjenju emisija, dok one koje prioritet daju ekonomskom rastu mogu naglasiti BDP i dostupnost radne snage.

Ova metoda se može unaprijediti uključivanjem dodatnih podataka u modele koji se koriste za predviđanja vremena i emisija. Dodavanje informacija o atmosferskim faktorima kao što su vlažnost i padavine, moglo bi poboljšati tačnost predviđanja solarne radijacije i temperature. Slično, podaci o cijenama električne energije i drugi ekonomski pokazatelji mogli bi omogućiti detaljnije razumijevanje operativnih troškova u svakoj državi, omogućavajući nijansiraniju procjenu financijske održivosti potencijalnih lokacija. Kontinuiranim ažuriranjem modela relevantnijim podacima, ovaj okvir za donošenje odluka može evoluirati i pružati preciznije uvide za dugoročno poslovno planiranje.

Popis slika

Slika 1 Prikaz nekoliko prvih redova skupa podataka o vremenu "Weather Data" za jednu državu (Austrija).....	7
Slika 2 Toplotna mapa (heatmap) vizualizacije nedostajućih vrijednosti (nema nedostajućih vrijednosti)	8
Slika 3 Prikaz nekoliko prvih redova skupa podataka o vremenu "Weather Data" za jednu državu (Austrija) nakon izvršene agregacije u prosječne mjesečne vrijednosti.....	9
Slika 4 Grafički prikaz rezultata testiranja predikcije temperature	12
Slika 5 Grafički prikaz rezultata testiranja predikcije direktne solarne radijacije	13
Slika 6 Grafički prikaz rezultata testiranja predikcije difuzne solarne radijacije	14
Slika 7 Prosječne temperature i radijacija po državama u posljednjoj deceniji (lijevo) i predviđene vrijednosti za narednu deceniju (desno)	20
Slika 8 Prikaz nekoliko prvih redova skupa podataka o emisijama CO2 "Countries CO2 Emission and more..."	22
Slika 9 Prikaz nekoliko prvih redova skupa podataka o emisijama CO2 nakon izvršene predobrade	23
Slika 10 Grafički prikaz odnosa predviđenih i stvarnih vrijednosti za obnovljive i neobnovljive izvore energije.....	27
Slika 11 Primjer rezultata rangiranja država	35
Slika 12 Usporedba trenutne i predviđene proizvodnje obnovljive energije (prosjeak posljednjih 10 u poređenju sa budućih 10 godina)	36
Slika 13 Usporedba trenutnih i predviđenih emisija CO2 (prosjeak posljednjih 10 u poređenju sa budućih 10 godina) u odnosu na populaciju	36
Slika 14 Emisije CO2 po glavi	37
Slika 15 Usporedba trenutne i predviđene temperature (prosjeak posljednjih 10 u poređenju sa budućih 10 godina)	37
Slika 16 Usporedba trenutne i predviđene solarne radijacije (prosjeak posljednjih 10 u poređenju sa budućih 10 godina).....	38
Slika 17 Usporedba trenutnog (posljednji zapis u skupu podataka) i predviđenog (za 10 godina) BDP-a u odnosu na populaciju	38

Izvori

- [1] International Energy Agency. *"Heating."* International Energy Agency, 2020, www.iea.org/reports/heating. Pristupljeno 12 Sep. 2024.
- [2] Vanous, Benjamin. *"CO2 Emission by Country's Growth and Population."* Kaggle, www.kaggle.com/datasets/lobosi/c02-emission-by-countrys-growth-and-population. Pristupljeno 12 Sep 2024.
- [3] Open Power System Data. *"Weather Data."* Open Power System Data, www.data.open-power-system-data.org/weather_data/. Pristupljeno 12 Sep. 2024.
- [4] Manuguerra, Luca, et al. *"A Machine Learning Based Method for Parametric Environmental Impact Model for Electric Vehicles."* *Journal of Cleaner Production*, vol. 454, 2024, p. 142308. ScienceDirect, <https://doi.org/10.1016/j.jclepro.2024.142308>.
- [5] Guo, Yanming, et al. *"ExioML: Eco-economic Dataset for Machine Learning in Global Sectoral Sustainability."* *arXiv*, 2406.09046v2, 6 July 2024, <https://doi.org/10.48550/arXiv.2406.09046>.
- [6] Couto, A. B. G., and L. A. D. Rangel. *"Model Proposition for Predicting Sustainability Classes Using Multicriteria Decision Support and Artificial Intelligence."* *Gestão & Produção*, vol. 29, 2022, p. e6922. <https://doi.org/10.1590/1806-9649-2022v29e6922>.
- [7] Bilen, T., M. Erel-Özçevik, Y. Yaslan, and S. F. Oktug. *"A Smart City Application: Business Location Estimator Using Machine Learning Techniques."* 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), 2018, pp. 1314-1321, <https://doi.org/10.1109/HPCC/SmartCity/DSS.2018.00219>.
- [8] Shihab, I. F., M. M. Oishi, S. Islam, K. Banik, and H. Arif. *"A Machine Learning Approach to Suggest Ideal Geographical Location for New Restaurant Establishment."* 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), 2018, pp. 1-5, <https://doi.org/10.1109/R10-HTC.2018.8629845>.
- [9] Huo, Da, and Hassan Rauf Chaudhry. *"Using Machine Learning for Evaluating Global Expansion Location Decisions: An Analysis of Chinese Manufacturing Sector."* *Technological Forecasting and Social Change*, vol. 163, 2021, 120436, <https://doi.org/10.1016/j.techfore.2020.120436>.
- [10] Pino-Mejías, Rafael, et al. *"Comparison of Linear Regression and Artificial Neural Networks Models to Predict Heating and Cooling Energy Demand, Energy Consumption and CO 2 Emissions."* *Energy*, vol. 118, 1 Jan. 2017, pp. 24–36., <https://doi.org/10.1016/j.energy.2016.12.022>.

- [11] Dašić, Darin. *"Postupci Ansambla Stabala Odluke."* Završni rad br. 4848, Sveučilište u Zagrebu, Fakultet Elektrotehnike i Računarstva, Jun 2017.
- [12] Kumari, Surbhi, and Sunil Kumar Singh. *"Machine Learning-Based Time Series Models for Effective CO2 Emission Prediction in India."* Environmental Science and Pollution Research, vol. 30, 2023, pp. 116601–116616. Springer, <https://doi.org/10.1007/s11356-022-21723-8>.