

**COLLEGIUM DA VINCI**

**Wydział Nauk Stosowanych**

**Kierunek: Big Data - Analytics & Society**

**Studia podyplomowe**

**Andrii Senchyshyn**

**"Analiza i prognozowanie uzysku energii z instalacji PV z wykorzystaniem danych meteorologicznych i metod Big Data"**

**"Analysis and Forecasting of Energy Yield from PV Installations Using Meteorological Data and Big Data Methods»"**

Praca dyplomowa

Promotor  
Krzysztof Kuźmich, PhD

**Poznan 2025**

## 2. Spis treści

3. Abstrakt	3
4. Wstęp	4
5. Tekst zasadniczy	5
5.1. Podstawy teoretyczne	5
5.2. Podstawy narzędziowe	7
5.3. Podstawy metodyczne	8
5.4. Realizacja scenariusza analizy danych	9
5.4.4 Drzewo decyzyjne (Decision Tree Regressor)	15
5.5. Wyniki i dyskusja	24
6. Porównanie międzynarodowe	26
6.1. Dobór lokalizacji do analizy porównawczej	26
6.2. Źródło danych i metodologia	27
6.3. Porównanie prognoz irradacji na 2030 rok	28
7. Analiza praktyczna i inwestycyjna	30
7.1. Koszty instalacji PV w wybranych krajach	30
7.2. Porównanie irradacji a kosztów instalacji	31
7.3. Szacowany czas zwrotu inwestycji (simple payback time)	31
8. Wnioski końcowe	33
9. Bibliografia	35

### **3. Abstrakt**

#### **Abstrakt (PL)**

Praca dotyczy analizy oraz prognozowania uzysku energii z instalacji fotowoltaicznych na podstawie danych meteorologicznych oraz narzędzi analitycznych zaliczanych do kategorii Big Data. W badaniu wykorzystano dane historyczne z lat 2005–2023 pochodzące z serwisu PVGIS (JRC, 2023) (Photovoltaic Geographical Information System), obejmujące takie parametry jak irradiacja globalna  $G(i)$ , temperatura powietrza ( $T_{2m}$ ), prędkość wiatru ( $WS_{10m}$ ) oraz wysokość Słońca nad horyzontem ( $H_{sun}$ ).

Analizie poddano pięć lokalizacji: Zwartowo (Polska), Monachium (Niemcy), Rzym (Włochy) oraz Lizbona (Portugalia), dobrane pod kątem reprezentowania zróżnicowanych warunków klimatycznych Europy. Dla każdej lokalizacji przygotowano zestawy danych, przeprowadzono agregację roczną i zastosowano modele regresji oraz model Random Forest (Breiman, 2001) do predykcji irradiacji w roku 2030.

W dalszej części pracy przeprowadzono uproszczoną analizę inwestycyjną, uwzględniającą koszty instalacji PV oraz ceny energii elektrycznej w wybranych krajach. Na tej podstawie oszacowano wskaźniki efektywności inwestycji oraz prosty czas zwrotu (SPBT). Wyniki pokazują, że mimo umiarkowanego poziomu irradiacji, Polska pozostaje konkurencyjna względem krajów Europy Zachodniej, a najbardziej atrakcyjną lokalizacją inwestycyjną okazała się Lizbona.

#### **Abstract (EN)**

This thesis focuses on the analysis and forecasting of energy yield from photovoltaic (PV) systems based on meteorological data and analytical tools associated with Big Data. The study uses historical data from 2005 to 2023 obtained from the PVGIS (Photovoltaic Geographical Information System), covering parameters such as global irradiance  $G(i)$ , air temperature ( $T_{2m}$ ), wind speed ( $WS_{10m}$ ), and solar altitude ( $H_{sun}$ ).

Five locations were analysed : Zwartowo (Poland), Munich (Germany), Rome (Italy), and Lisbon (Portugal), chosen to represent different climatic zones of Europe. For each site, datasets were processed, annual averages were calculated, and both regression models and a Random Forest model were applied to forecast irradiance for the year 2030.

The second part of the thesis includes a simplified investment analysis, taking into account PV installation costs and electricity prices in the selected countries. Based on this, investment efficiency indicators and simple payback time (SPBT) were estimated. The results show that, despite moderate solar potential, Poland remains competitive compared to Western Europe, while Lisbon stands out as the most favorable investment location.

## **4. Wstęp**

W obliczu zmian klimatycznych oraz rosnącego zapotrzebowania na czystą energię, odnawialne źródła energii, w szczególności energia słoneczna, odgrywają coraz większą rolę w systemach energetycznych państw. Fotowoltaika, jako technologia przekształcająca promieniowanie słoneczne w energię elektryczną, zyskuje na znaczeniu zarówno w skali mikro (gospodarstwa domowe), jak i makro (instalacje przemysłowe) (Fraunhofer ISE, 2023).

W niniejszej pracy skoncentrowano się na analizie danych meteorologicznych i ich wpływie na uzysk energii z instalacji PV. Głównym problemem badawczym jest identyfikacja najważniejszych czynników pogodowych, które decydują o efektywności systemów PV oraz możliwość wykorzystania technik analizy danych do prognozowania przyszłego uzysku energii (JRC, 2023; GSE, 2023).

Wybór tematu został podyktowany jego aktualnością, użytecznością praktyczną oraz możliwością zastosowania nowoczesnych narzędzi analitycznych. Dane źródłowe pozyskano z portalu PVGIS, a analiza zostanie przeprowadzona przy użyciu środowiska Python i bibliotek takich jak pandas, matplotlib, seaborn oraz scikit-learn (Pedregosa et al., 2011; Van Rossum & Drake, 2009).

Struktura pracy obejmuje wprowadzenie teoretyczne (fotowoltaika, parametry meteorologiczne, koncepcja Big Data), opis narzędzi i danych, metody analizy, realizację scenariusza badawczego oraz interpretację wyników. W zakończeniu przedstawione zostaną wnioski i propozycje praktycznego wykorzystania wyników analizy.

W dalszej części pracy, w celu lepszego zrozumienia wpływu warunków klimatycznych na wydajność instalacji PV, zaproponowano również porównanie wyników uzyskanych dla kilku lokalizacji w Europie, reprezentujących różne strefy geograficzne (IEO, 2023; APREN, 2023).

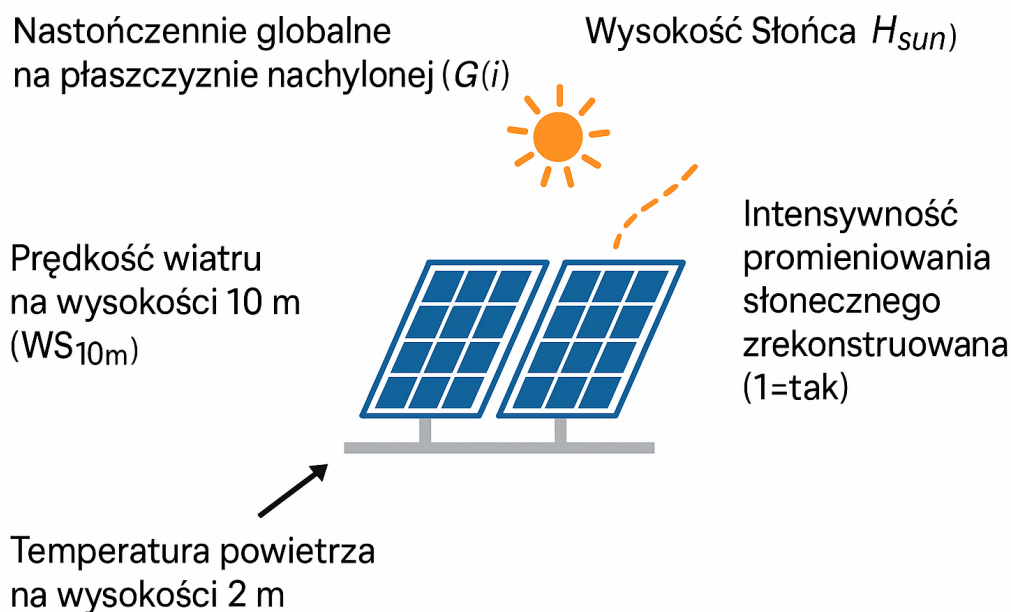
## **5. Tekst zasadniczy**

### **5.1. Podstawy teoretyczne**

Fotowoltaika to dziedzina technologii zajmująca się bezpośrednim przekształcaniem energii promieniowania słonecznego w energię elektryczną przy użyciu materiałów półprzewodnikowych. Proces ten realizowany jest w ogniwach fotowoltaicznych, które stanowią podstawowy element paneli PV. W ostatnich latach obserwuje się dynamiczny rozwój sektora PV, co związane jest nie tylko z postępem technologicznym, lecz także z rosnącym znaczeniem energetyki odnawialnej w kontekście polityki klimatycznej i ekonomicznej (GSE, 2023; APREN, 2023).

Efektywność instalacji PV zależy od wielu czynników, z których kluczowe to: natężenie promieniowania słonecznego (irradiacja), temperatura powietrza i modułów, wilgotność względna, kierunek i prędkość wiatru, ciśnienie atmosferyczne oraz warunki lokalne (zacienienie, kąt nachylenia modułów, ich orientacja) (IEO, 2023; JRC, 2023). Szczególne znaczenie ma irradiacja globalna ( $G(h)$ ) oraz irradiacja bezpośrednia ( $G_b(n)$ ), które bezpośrednio przekładają się na ilość energii docierającej do powierzchni modułów.

Równie istotna jest temperatura powietrza ( $T_{2m}$ ), ponieważ nadmierne nagrzewanie się modułów powoduje spadek ich sprawności (Fraunhofer ISE, 2023). Z kolei wysoka wilgotność względna (RH) może wpływać na efektywność pracy poprzez zwiększenie strat energetycznych i osadzanie się cząstek na powierzchni paneli.



**Rys. 1** Schemat instalacji fotowoltaicznej i powiązanych parametrów meteorologicznych.

Zmienność warunków meteorologicznych w czasie oraz ich wpływ na efektywność instalacji PV sprawia, że analiza danych pogodowych staje się kluczowym elementem optymalizacji działania systemów fotowoltaicznych (JRC, 2023). W tym kontekście, zastosowanie narzędzi analizy danych i metod przewidywania jest nie tylko zasadne, ale wręcz konieczne.

Pojęcie Big Data odnosi się do zbiorów danych, które charakteryzują się dużą objętością, różnorodnością i zmiennością. Dane meteorologiczne, zwłaszcza pozyskiwane z wielu lat w odstępach czasowych godzinowych lub dziennych, spełniają te kryteria (Pedregosa et al., 2011). Ich analiza wymaga zastosowania odpowiednich narzędzi informatycznych oraz znajomości metod statystycznych i uczenia maszynowego.

W kontekście pracy dyplomowej, zastosowanie Big Data obejmuje nie tylko przetwarzanie i wizualizację danych meteorologicznych, ale również budowę modeli

predykcyjnych, które umożliwiają prognozowanie uzysku energii elektrycznej w określonych warunkach (Breiman, 2001). Wykorzystanie takich technik pozwala na lepsze planowanie produkcji energii, prognozowanie kosztów oraz ocenę opłacalności inwestycji w instalacje PV.

Zrozumienie teoretycznych podstaw funkcjonowania instalacji fotowoltaicznych oraz kluczowych parametrów meteorologicznych stanowi fundament dla dalszej analizy danych i budowy modeli predykcyjnych, które zostaną zaprezentowane w kolejnych rozdziałach pracy.

## **5.2. Podstawy narzędziowe**

W celu przeprowadzenia analizy danych meteorologicznych oraz predykcji uzysku energii z instalacji fotowoltaicznych wykorzystano zestaw narzędzi informatycznych, które są powszechnie stosowane w analizie danych i projektach związanych z Big Data. Kluczowym elementem warsztatu pracy było środowisko programistyczne Python, znane z przejrzystości składni oraz bogatego ekosystemu bibliotek wspierających analizę, modelowanie i wizualizację danych (Van Rossum & Drake, 2009).

W analizie wykorzystano między innymi następujące biblioteki:

- pandas – do wczytywania, przekształcania i analizy danych w strukturze DataFrame (Pedregosa et al., 2011),
- matplotlib i seaborn – do wizualizacji danych i generowania wykresów opisujących rozkłady i zależności między zmiennymi,
- scikit-learn – do budowy modeli predykcyjnych, regresji liniowej i oceny skuteczności modeli (Pedregosa et al., 2011),
- numpy – do operacji matematycznych i statystycznych na danych liczbowych,
- statsmodels – w celu przeprowadzania regresji oraz analizy korelacji.

Dodatkowo, dane meteorologiczne pobrano z oficjalnego portalu PVGIS (Photovoltaic Geographical Information System), zarządzanego przez Wspólne Centrum Badawcze Komisji Europejskiej (JRC). Portal PVGIS umożliwia dostęp do danych pogodowych i solarno-energetycznych w ujęciu historycznym oraz modelowanym dla dowolnej lokalizacji geograficznej w Europie i na świecie (JRC, 2023). W ramach pracy wykorzystano dane dla lokalizacji w Polsce (Zwartowo), w zakresie czasowym obejmującym lata 2005–2023.

Dane obejmują m.in.:

- $G(i)$  – irradiacja globalna na płaszczyźnie modułów PV ( $W/m^2$ ),
- $T_{2m}$  – temperatura powietrza ( $^{\circ}C$ ),
- $WS_{10m}$  – prędkość wiatru na 10 m ( $m/s$ ),
- $H_{sun}$  – wysokość słońca nad horyzontem ( $^{\circ}$ ),
- $Int$  – wskaźnik wartości rekonstruowanych.

Plik zawiera dane w rozdzielczości godzinowej i umożliwia analizę czasową na poziomie dni, miesięcy i lat. Format danych (CSV) ułatwia import i obróbkę w środowisku Python.

### **5.3. Podstawy metodyczne**

W celu zapewnienia przejrzystości i rzetelności badania, niniejszy rozdział przedstawia zakres, charakter i metodykę analiz przeprowadzonych w pracy. Analiza danych meteorologicznych oraz prognozowanie uzysku energii z instalacji fotowoltaicznych wymaga określenia ram czasowych, przestrzennych, jak również sprecyzowania przedmiotu badania.

Zakres czasowy obejmuje lata 2005–2023, co pozwala na uwzględnienie zmian warunków atmosferycznych w dłuższej perspektywie oraz analizę sezonowości i długoterminowych trendów. Zakres przestrzenny ogranicza się do jednej lokalizacji –



miejsowości Zwartowo w województwie pomorskim, gdzie znajduje się największa farma fotowoltaiczna w Polsce (IEO, 2023). Taka lokalizacja pozwala na badanie danych rzeczywistych w kontekście realnych zastosowań technologii PV.

Zakres przedmiotowy pracy obejmuje zmienne meteorologiczne ( $G(i)$ ,  $T2m$ ,  $H_{sun}$ ,  $WS10m$ ) oraz ich wpływ na potencjalny uzysk energii z instalacji PV. Celem jest określenie zależności pomiędzy tymi czynnikami a irradacją oraz wskazanie możliwości predykcji wartości  $G(i)$  na podstawie pozostałych parametrów.

Zakres przestrzenny obejmuje przede wszystkim lokalizację Zwartowo w Polsce i danych generowanych dla tej lokalizacji, pobranych z serwisu PVGIS, jednak w dalszej części pracy uwzględniono również trzy dodatkowe lokalizacje: Rzym (Włochy), Monachium (Niemcy) oraz Lizbona (Portugalia), w celu przeprowadzenia analizy porównawczej na poziomie europejskim.

W pracy przyjęto metodę analizy ilościowej, wykorzystującą eksplorację danych (EDA – Exploratory Data Analysis), analizę korelacji oraz regresję. W analizie przewidziano także zastosowanie metod predykcyjnych z zakresu uczenia maszynowego (ML), takich jak regresja liniowa i modele drzewa decyzyjnego (Breiman, 2001). Dane zostały przekształcone i poddane standaryzacji w celu umożliwienia ich obróbki przez modele matematyczne.

Cały proces badawczy realizowany jest z wykorzystaniem języka Python oraz środowiska Jupyter Notebook, co zapewnia transparentność, możliwość wielokrotnego odtworzenia analizy oraz elastyczność w testowaniu różnych podejść (Van Rossum & Drake, 2009; Pedregosa et al., 2011). Kolejny rozdział przedstawi szczegółowy scenariusz pracy analitycznej.

## **5.4. Realizacja scenariusza analizy danych**

W niniejszym rozdziale przedstawiono kolejne etapy analizy danych meteorologicznych oraz budowy modelu predykcyjnego prognozującego irradację ( $G(i)$ ) dla lokalizacji w Zwartowie. Proces rozpoczęto od wczytania danych pobranych z serwisu

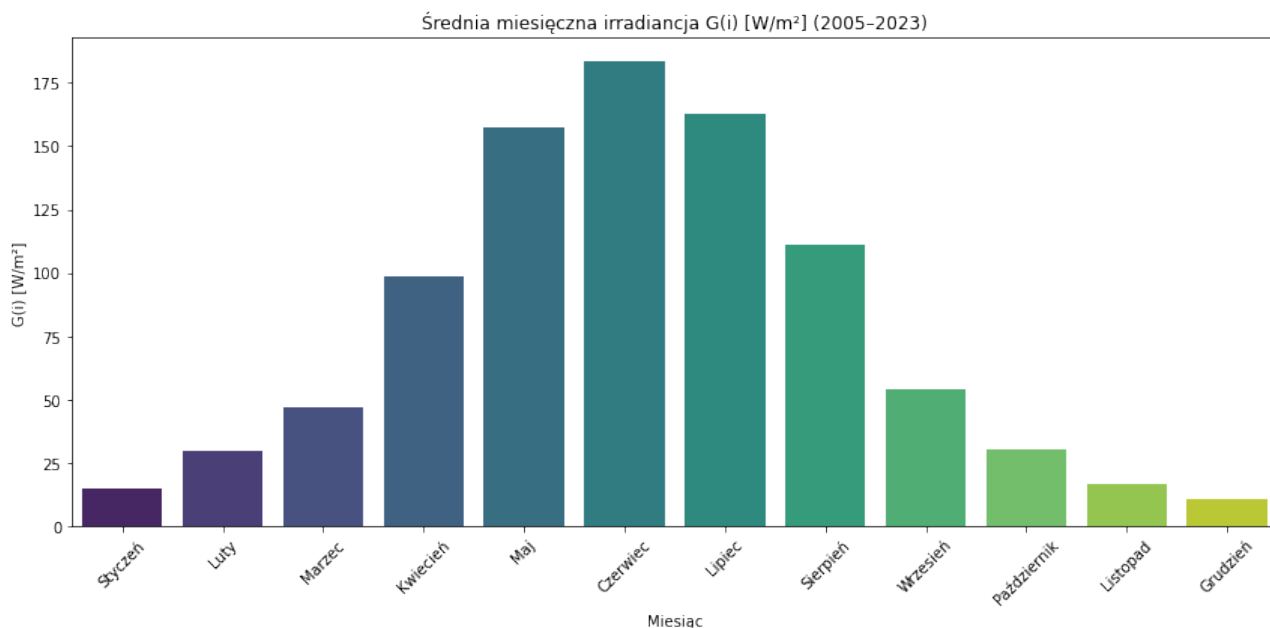
PVGIS w formacie CSV (JRC, 2023). Dane zostały zaimportowane do środowiska Python z wykorzystaniem biblioteki pandas (Van Rossum & Drake, 2009; Pedregosa et al., 2011).

#### 5.4.1 Wstępna analiza danych (EDA)

Pierwszym krokiem była eksploracyjna analiza danych (EDA – Exploratory Data Analysis), która objęła:

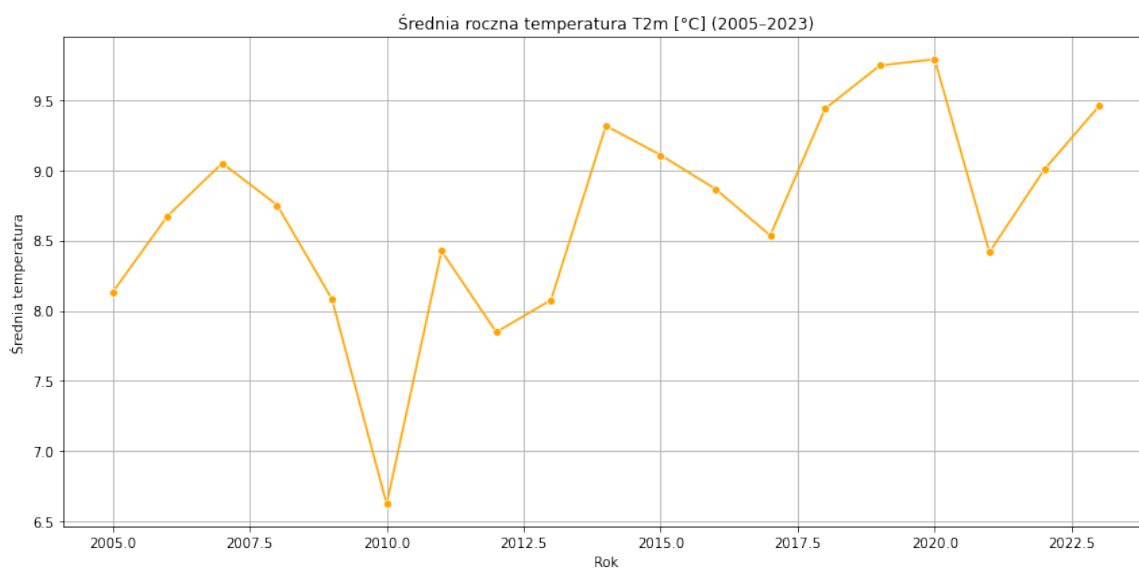
- sprawdzenie kompletności danych i identyfikację braków,
- analizę rozkładów poszczególnych zmiennych ( $T_{2m}$ ,  $H_{sun}$ ,  $WS_{10m}$ ),
- obliczenie podstawowych statystyk opisowych (średnia, mediana, odchylenie standardowe).

Wizualizacje przygotowano przy pomocy bibliotek matplotlib i seaborn (Pedregosa et al., 2011). Na Rys. 2 możemy zobaczyć wizualizację średnich wartości irradacji  $G(i)$  w przekroju miesięcznym, co pozwoliło zaobserwować sezonowość zjawiska i identyfikować miesiące o największym potencjale energetycznym.

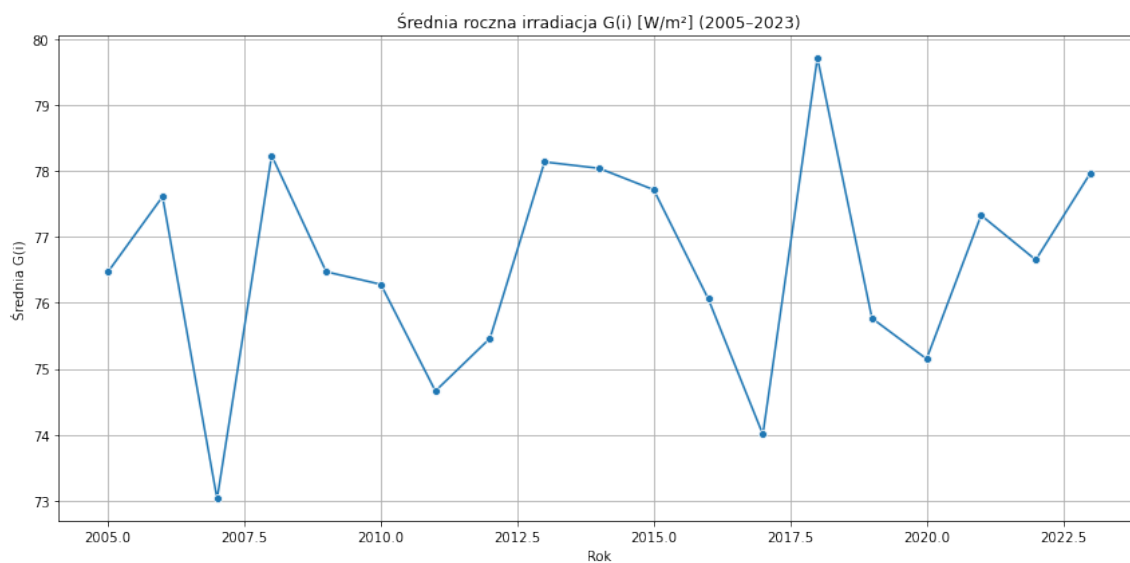


**Rys. 2** Średnia miesięczna irradancja  $G(i)$  [ $W/m^2$ ]

Także było przeanalizowano trendy długoterminowe w danych meteorologicznych – w szczególności średnią roczną temperaturę ( $T_{2m}$ ) Rys. 3 oraz średnią roczną irradiację ( $G(i)$ ) Rys. 4 w okresie 2005–2023. Miało to na celu identyfikację potencjalnych zmian klimatycznych oraz ich wpływu na opłacalność instalacji PV (IEO, 2023; JRC, 2023).

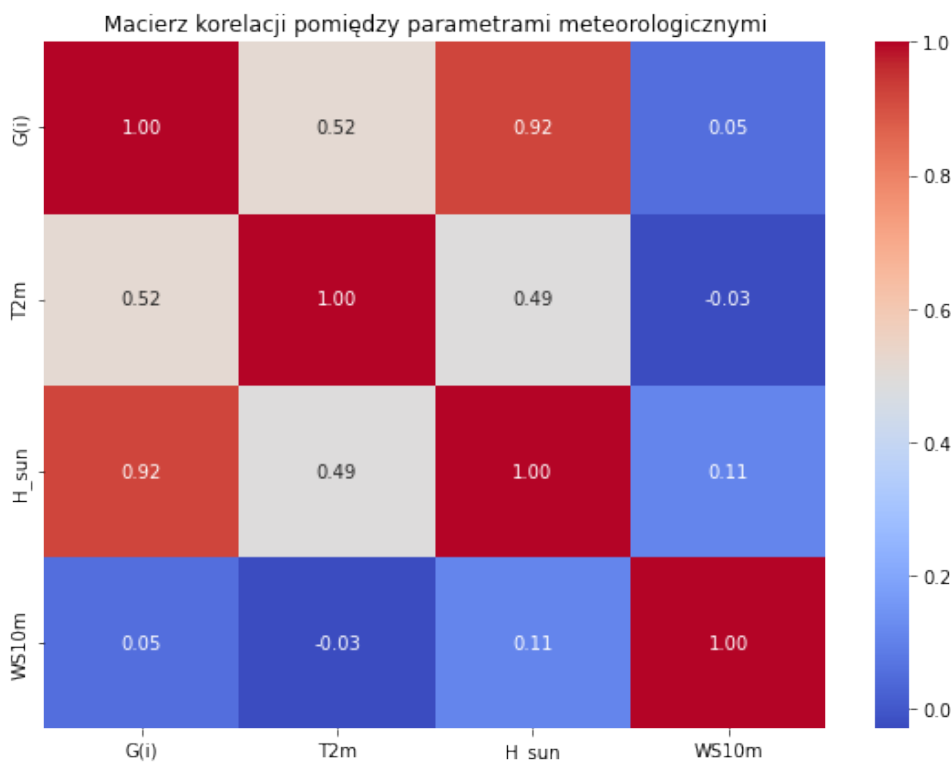


**Rys. 3** Średnia roczna temperatura  $T_{2m}$  [°C] (2005–2023)



**Rys. 4** Średnia roczna irradiacja  $G(i)$  [W/m²] (2005–2023)

Następnie przeprowadzono analizę korelacji między zmiennymi meteorologicznymi a zmienną zależną  $G(i)$  z wykorzystaniem współczynnika korelacji Pearsona (Breiman, 2001). Wyniki zaprezentowano na heatmapie (Rys. 5), wygenerowanej przy użyciu biblioteki seaborn.



**Rys. 5** Macierz korelacji pomiędzy parametrami meteorologicznymi

Na podstawie analizy korelacji (Ryc. 5) można zauważyć, że największy wpływ na wartość irradacji ( $G(i)$ ) ma wysokość Słońca nad horyzontem ( $H_{\text{sun}}$ ), dla której współczynnik korelacji przekracza wartość 0.9. Oznacza to, że im wyżej znajduje się Słońce na niebie, tym większe promieniowanie dociera do powierzchni modułów PV. Dodatnia korelacja występuje również dla temperatury powietrza ( $T2m$ ), choć jej wpływ jest słabszy. Z kolei prędkość wiatru ( $WS10m$ ) wykazuje niewielką lub znikomą korelację z irradacją (JRC, 2023).

### 5.4.2 Przygotowanie danych do modelowania

Dane zostały przekształcone do odpowiedniego formatu:

- agregacja do wartości dziennych lub miesięcznych,
- standaryzacja zmiennych niezależnych,
- podział zbioru danych na część treningową i testową (80/20).

### 5.4.3 Regresja liniowa

Model regresji liniowej został wykorzystany jako podstawowy punkt odniesienia do prognozowania wartości irradacji  $G(i)$ . Wykorzystano implementację z biblioteki scikit-learn (Pedregosa et al., 2011). Poniżej zaprezentowano przykładowy fragment kodu oraz wykres rzeczywistych vs przewidywanych wartości  $G(i)$  (Rys. 6).

```
from sklearn.linear_model import LinearRegression

from sklearn.metrics import mean_squared_error, r2_score

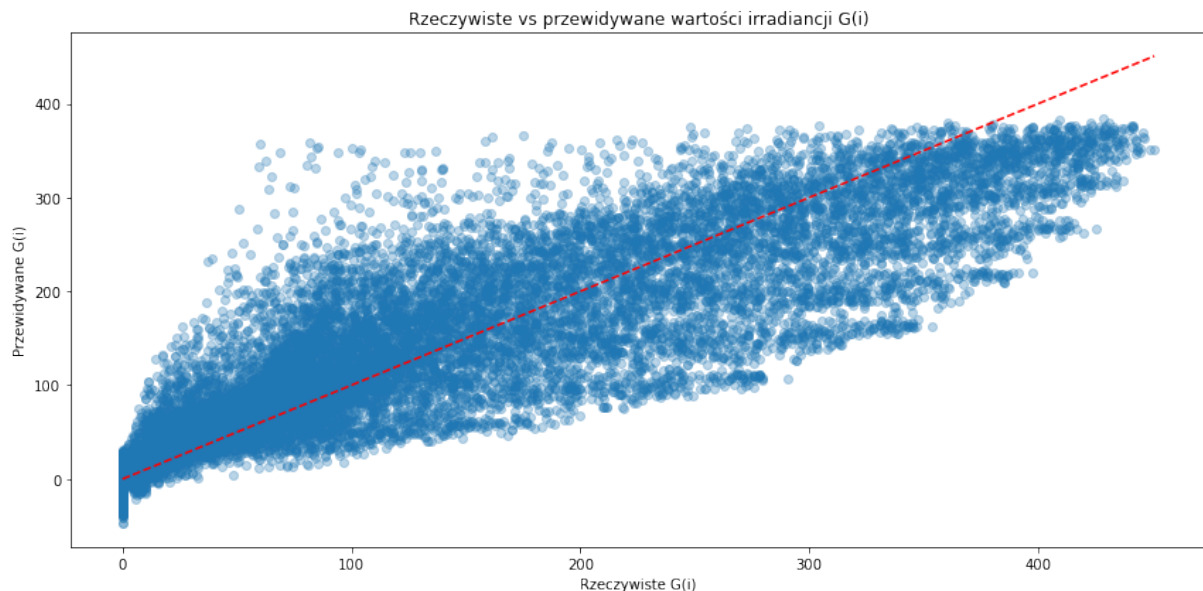
model = LinearRegression()

model.fit(X_train, y_train)

y_pred = model.predict(X_test)

r2 = r2_score(y_test, y_pred)

rmse = mean_squared_error(y_test, y_pred, squared=False)
```



**Rys. 6** Wykres rzeczywistych vs przewidywanych wartości  $G(i)$  – regresja liniowa.

Uzyskane wyniki:

- $R^2 = 0,853$
- $RMSE = 43,42$

Choć model dobrze uchwycił ogólną zależność między zmiennymi, analiza wykresu rzeczywistych vs przewidywanych wartości wykazała istotne rozrzuty. W szczególności:

- Model nie doszacowuje wartości wysokie i przeszacowuje niskie.
- Punkty są znacznie rozproszone wokół linii idealnej ( $y = x$ ), co wskazuje na niewystarczającą precyzję w ekstremalnych przypadkach.

Model regresji liniowej, mimo swojej prostoty, okazał się niewystarczający do pełnego odwzorowania nieliniowej zależności obecnej w danych.

#### 5.4.4 Drzewo decyzyjne (Decision Tree Regressor)

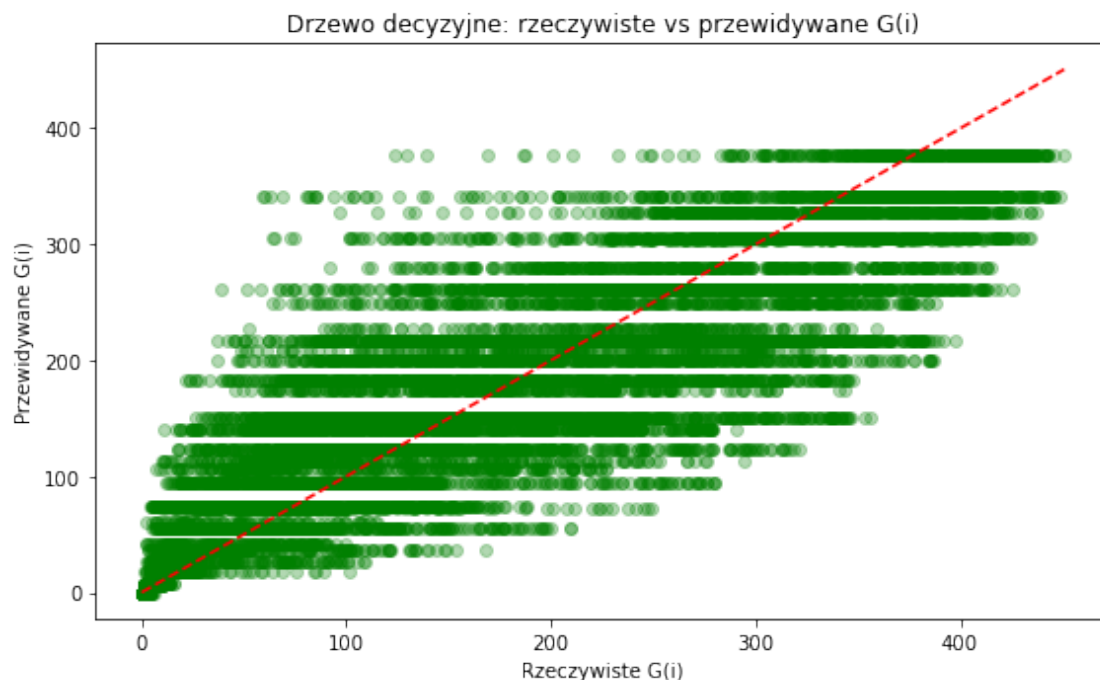
W kolejnym etapie zastosowano model drzewa decyzyjnego, który lepiej radzi sobie z nieliniowymi zależnościami. Zastosowano klasę `DecisionTreeRegressor` ze zbioru `scikit-learn` (Pedregosa et al., 2011). Poniżej zamieszczono fragment kodu oraz wykres predykcji (Ryc. 7).

```
from sklearn.tree import DecisionTreeRegressor

tree = DecisionTreeRegressor(random_state=42)

tree.fit(X_train, y_train)

y_tree_pred = tree.predict(X_test)
```



**Rys. 7** Przewidywane vs rzeczywiste wartości  $G(i)$  – model drzewa decyzyjnego.

Wyniki były nieco lepsze niż w przypadku regresji liniowej:

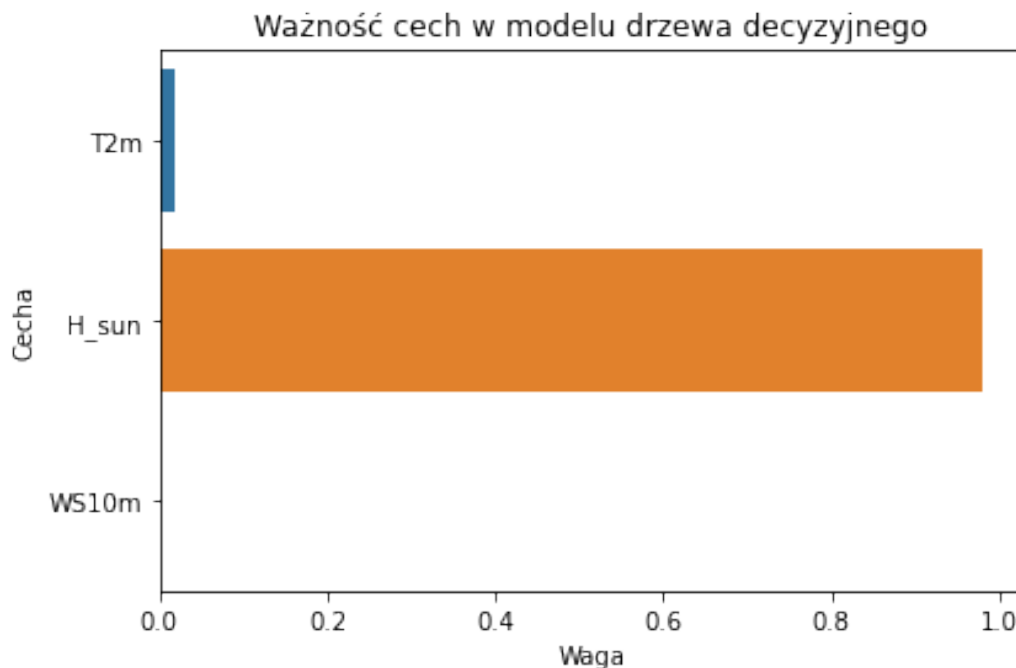
- $R^2 = 0,861$
- $RMSE = 42,12$

Na wykresie rzeczywistych vs przewidywanych wartości widać poprawę w odwzorowaniu danych, jednak:

- Model generuje prognozy schodkowe, co jest typowe dla pojedynczego drzewa (Breiman, 2001).
- Nadal występują zauważalne odchylenia przy bardzo wysokich i niskich wartościach  $G(i)$ .

Mimo tych ograniczeń, drzewo decyzyjne potrafiło lepiej dopasować się do danych niż regresja liniowa, co wskazuje na obecność nieliniowych zależności w zbiorze danych.

Dodatkowo zaprezentowano wykres ważności cech (feature importance), pozwalający zidentyfikować zmienne o największym wpływie na wynik predykcji (Rys. 8).



**Rys. 8** Ważność cech w modelu drzewa decyzyjnego



### 5.4.5 Model Random Forest

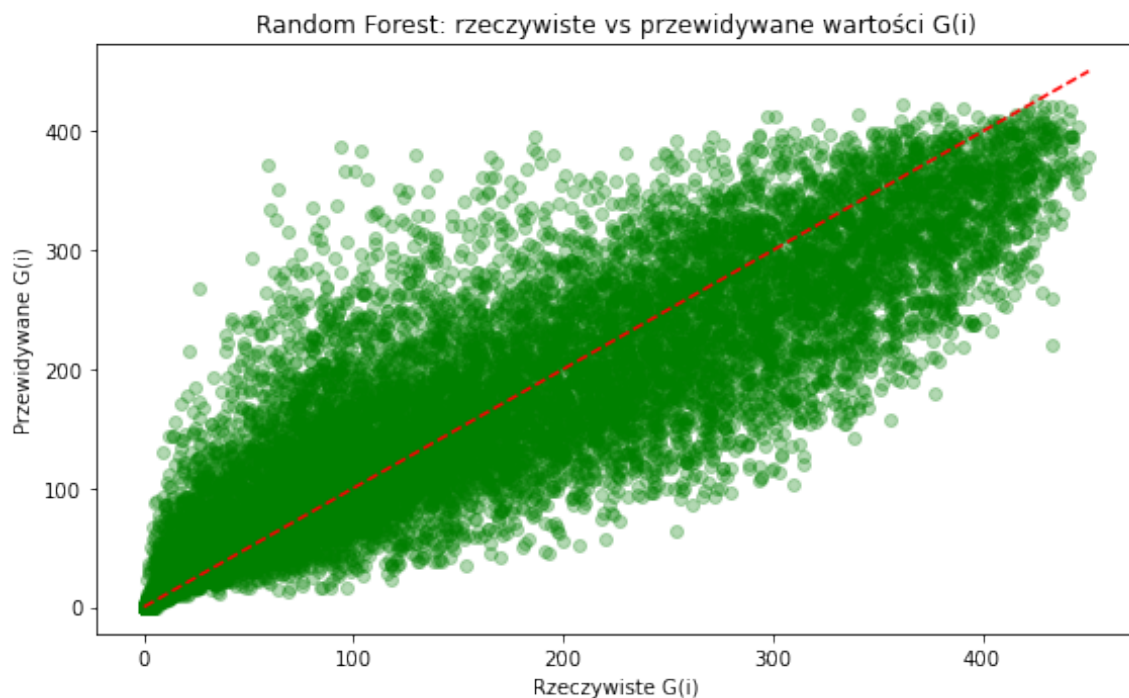
Na koniec przeprowadzono analizę przy użyciu modelu zespołowego Random Forest, który okazał się najbardziej efektywny ze względu na swoje właściwości, które czynią go wyjątkowo skutecznym w przypadku danych o potencjalnie nieliniowych zależnościach oraz większej złożoności (Breiman, 2001; Pedregosa et al., 2011). Przedstawiono kod oraz wykres predykcji dla tego modelu (Rys. 9).

```
from sklearn.ensemble import RandomForestRegressor

rf = RandomForestRegressor(n_estimators=100, random_state=42)

rf.fit(X_train, y_train)

y_rf_pred = rf.predict(X_test)
```



**Rys. 9** Random Forest: rzeczywiste vs przewidywane wartości G(i)

Wyniki uzyskane dla tego modelu:

- $R^2 = 0,875$
- $RMSE = 40,00$

Osiągnięte rezultaty okazały się najlepsze spośród wszystkich przetestowanych modeli regresyjnych. Model ten wyjaśnia 87,5% zmienności w danych i generuje prognozy z najmniejszym średnim błędem (RMSE)

#### **5.4.6 Analiza rozkładu reszt**

W celu oceny dokładności predykcji, zbadano rozkład błędów (reszt) dla modelu Random Forest. Histogram rozkładu przedstawiono na Rys. 10.

```
residuals_rf = y_test - y_rf_pred

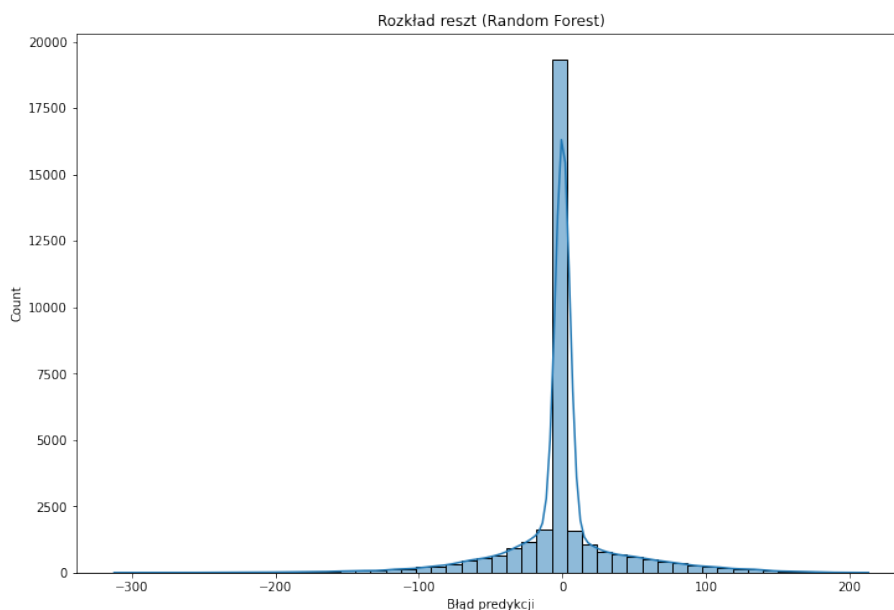
plt.figure(figsize=(12, 8))

sns.histplot(residuals_rf, bins=50, kde=True)

plt.title("Rozkład reszt (Random Forest)")

plt.xlabel("Błąd predykcji")

plt.show()
```



**Rys. 10** Histogram rozkładu reszt modelu Random Forest.

Rozkład błędów jest symetryczny i skoncentrowany wokół zera, co świadczy o braku błędów systematycznych. Większość predykcji mieści się w przedziale  $\pm 10$  W/m<sup>2</sup>, co potwierdza wysoką precyzję modelu. Pojedyncze duże błędy (np. powyżej  $\pm 200$ ) są sporadyczne i typowe dla danych rzeczywistych.

Na podstawie histogramu reszt modelu Random Forest można wyciągnąć następujące wnioski:

#### 1. Symetria wokół zera

Rozkład reszt jest wyraźnie skoncentrowany wokół wartości 0, co świadczy o tym, że model nie wykazuje systematycznego błędu – nie ma tendencji do regularnego przeszacowywania ani nie doszacowywania wartości  $G(i)$ .

#### 2. Wysoki, wąski szczyt

Większość błędów predykcji znajduje się bardzo blisko zera (w przedziale około -10 do +10), co oznacza, że model jest precyzyjny w dużej liczbie przypadków.

### 3. Długie ogony rozkładu

Na wykresie widoczne są również pojedyncze przypadki dużych błędów (np. -300 lub +200). Jest to typowe zachowanie dla modeli typu Random Forest, szczególnie w odniesieniu do punktów nietypowych lub rzadko występujących w zbiorze treningowym.

### Wniosek końcowy

Model Random Forest cechuje się bardzo dobrym dopasowaniem:

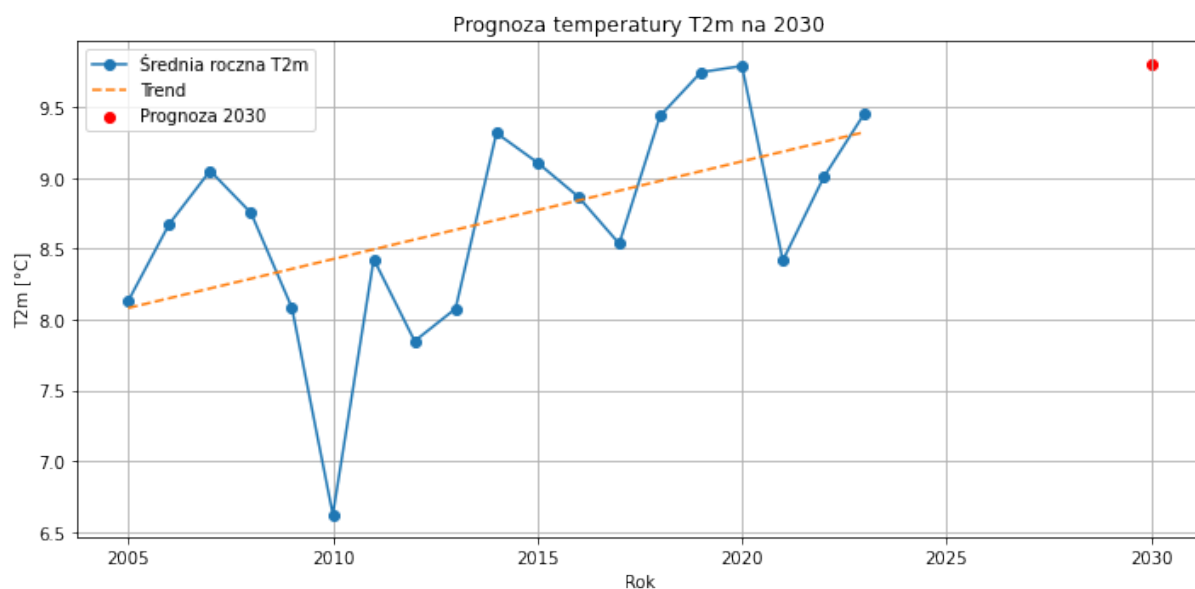
- jest stabilny,
- nie generuje systematycznych błędów,
- zapewnia dużą dokładność w większości przypadków,
- a obecność pojedynczych dużych błędów mieści się w akceptowalnych granicach.

Całość analizy została przeprowadzona w sposób powtarzalny i udokumentowana w środowisku Jupyter Notebook co zapewnia pełną powtarzalność i transparentność procesu badawczego. W kolejnym rozdziale zaprezentowano podsumowanie wyników oraz interpretację uzyskanych rezultatów w kontekście założonych celów badawczych.

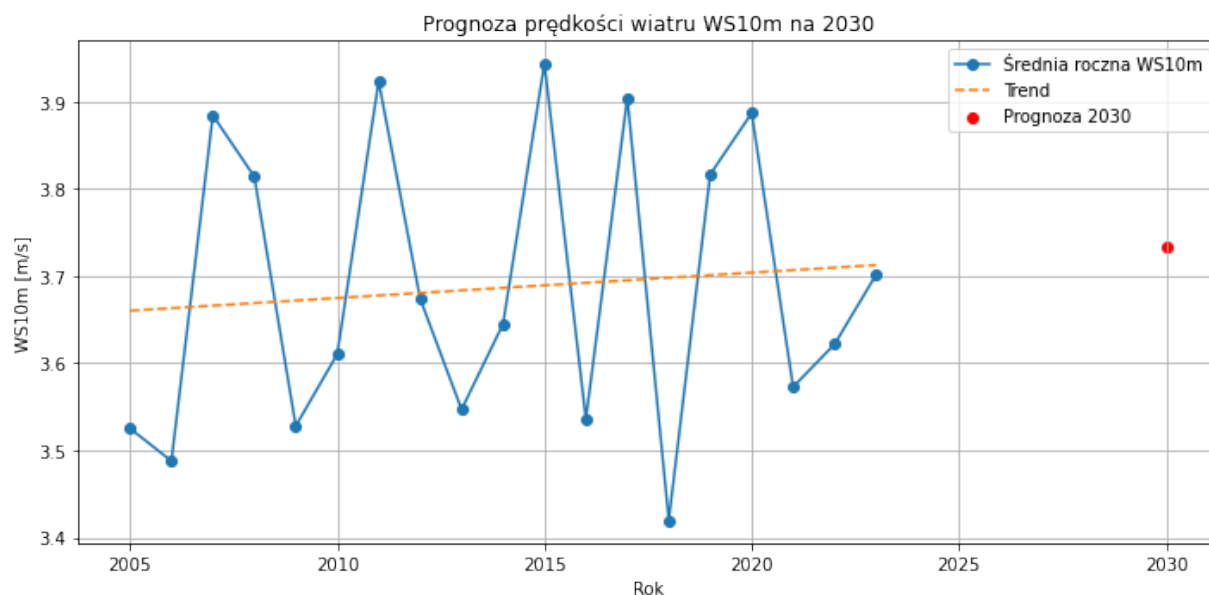
### **5.4.7 Prognoza irradacji na 2030 na podstawie prognozowanej pogody**

W celu zaprezentowania praktycznego zastosowania modelu Random Forest przeprowadzono eksperyment prognostyczny, mający na celu oszacowanie irradacji  $G(i)$  w roku 2030.

W pierwszym kroku, na podstawie danych historycznych z lat 2005–2023, zbudowano liniowe modele trendów dla temperatury (T2m) Rys. 11 i prędkości wiatru (WS10m) Rys. 12. Na podstawie tych modeli wyznaczono przewidywane wartości meteorologiczne na 2030 rok.



**Rys. 11** Prognoza temperatury (T2m)



**Rys. 12** Prognoza prędkości wiatru (WS10m)

Prognozowana temperatura T2m  $\approx 9,82^{\circ}\text{C}$

Prognozowana prędkość wiatru WS10m  $\approx 3,90 \text{ m/s}$

Następnie założono średnią roczną wartość wysokości Słońca (H\_sun) – obliczoną na podstawie całego zbioru danych:

Fragment kodu:

```
h_sun_avg = df['H_sun'].mean()
```

```
print(f'Średnia wartość H_sun : {h_sun_avg:.2f}°»)
```

Średnia wartość H\_sun : 20.78°

Z wykorzystaniem powyższych wartości jako danych wejściowych, przeprowadzono predykcję irradacji G(i) za pomocą wcześniej wytrenowanego modelu Random Forest Regressor (Breiman, 2001):

```
g_pred_2030 = rf.predict(future_input)[0]
```

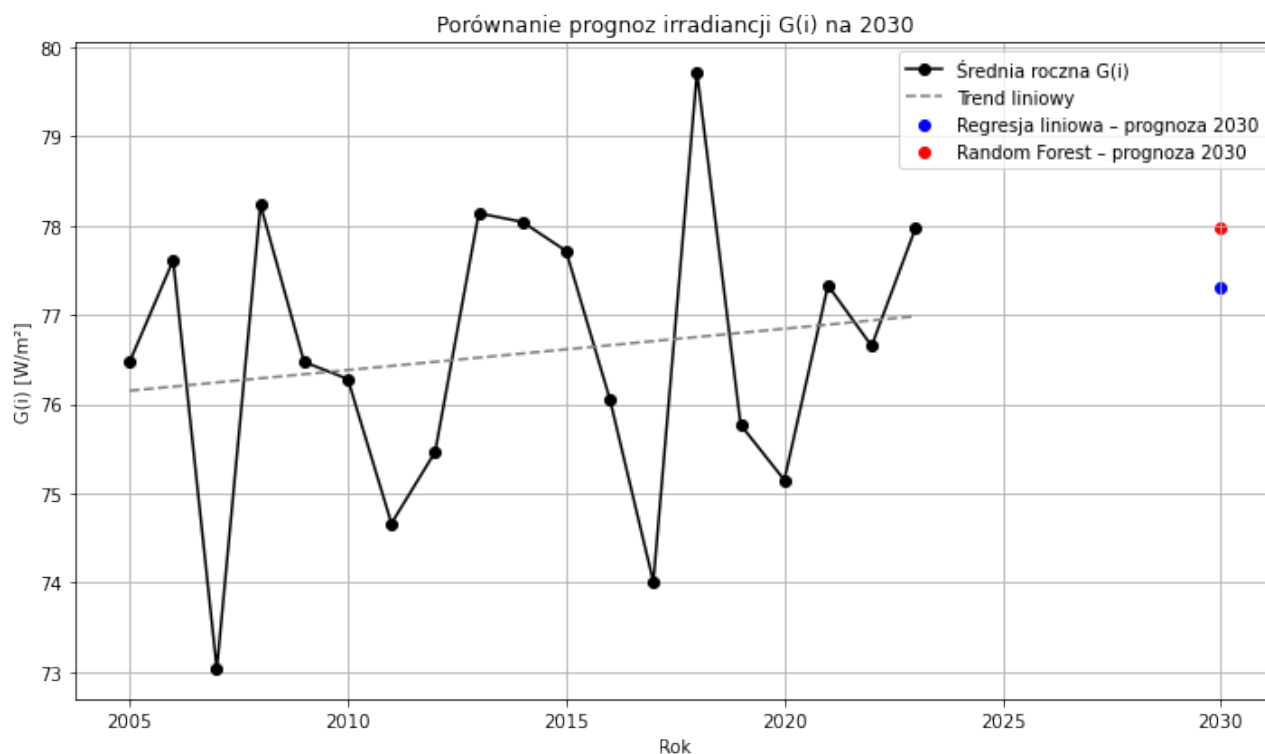
```
print(f'Prognozowana irradacja G(i) w 2030: {g_pred_2030:.2f} W/m²»)
```

Prognozowana irradacja G(i) (średnioroczna) na 2030: 77.97 W/m²

Aby uzyskać dodatkowy punkt odniesienia i potwierdzić kierunek trendu, zastosowano również model regresji liniowej na danych rok vs G(i) w celu oszacowania przewidywanej irradacji na 2030 rok (Pedregosa et al., 2011).

Model	Prognozowana irradacja G ( i ) [W/m²]
Regresja liniowa	77,31
Random Forest	77,97

Wizualne porównanie przedstawiono na poniższym wykresie Rys. 13, który zawiera linię trendu liniowego oraz prognozy obu modeli dla roku 2030.



**Rys. 13** Porównanie prognoz irradancji  $G(i)$  na 2030 rok

Wyniki porównania:

Na podstawie przeprowadzonej analizy trendu historycznego średniej rocznej irradancji  $G(i)$  w latach 2005–2023 dokonano prognozy wartości na rok 2030 za pomocą dwóch podejść:

- regresji liniowej (model prosty),
- modelu Random Forest (bardziej złożony, uwzględniający nieliniowe zależności).

Obie prognozy zostały przedstawione na wspólnym wykresie. Model regresji liniowej oszacował wartość  $G(i)$  na poziomie  $77,31 \text{ W/m}^2$ , natomiast Random Forest –  $77,97 \text{ W/m}^2$ . Różnica pomiędzy przewidywaniami jest niewielka (ok.  $0,66 \text{ W/m}^2$ ), co potwierdza spójność uzyskanych modeli.

#### Wniosek:

Pomimo różnic w strukturze modeli, obie metody wskazują na kontynuację obecnego trendu i umiarkowany wzrost irradacji w perspektywie do 2030 roku. Model Random Forest, dzięki swojej zdolności do modelowania nieliniowych zależności i niższemu błędowi predykcji (RMSE), może być traktowany jako bardziej wiarygodne źródło predykcji. Przeprowadzone porównanie stanowi również potwierdzenie praktycznej użyteczności opracowanego modelu uczenia maszynowego do prognozowania warunków meteorologicznych wpływających na produkcję energii z instalacji PV.

Całość analizy została przeprowadzona w sposób powtarzalny i udokumentowana w środowisku Jupyter Notebook. W kolejnym rozdziale zaprezentowano podsumowanie wyników oraz interpretację uzyskanych rezultatów w kontekście założonych celów badawczych.

## 5.5. Wyniki i dyskusja

W ramach przeprowadzonej analizy utworzono trzy modele predycyjne mające na celu oszacowanie wartości irradacji ( $G(i)$ ) na podstawie parametrów meteorologicznych: temperatury powietrza ( $T2m$ ), wysokości Słońca ( $H_{\text{sun}}$ ) oraz prędkości wiatru ( $WS10m$ ). Porównano skuteczność regresji liniowej, drzewa decyzyjnego oraz lasu losowego (Random Forest Regressor).



Regresja liniowa osiągnęła współczynnik determinacji  $R^2 = 0,853$  oraz błąd średniokwadratowy  $RMSE \approx 43,42 \text{ W/m}^2$ . Choć model wykazał dobrą zgodność ogólną, zauważalne były pewne odchylenia od linii idealnej, szczególnie przy niższych i wyższych wartościach  $G(i)$  (Pedregosa et al., 2011).

Drzewo decyzyjne poprawiło jakość predykcji, osiągając  $R^2 = 0,861$  oraz  $RMSE \approx 42,12 \text{ W/m}^2$ . Wykresy wskazały jednak typowy dla drzew efekt „schodkowych” predykcji, który może obniżać precyzję w niektórych przypadkach.

Model Random Forest Regressor, jako model zespołowy, uzyskał najlepsze rezultaty:  $R^2 = 0,875$  i  $RMSE = 40,00 \text{ W/m}^2$ . Jego przewidywania cechowały się wyższą dokładnością i większą stabilnością (Breiman, 2001). Analiza rozkładu reszt wykazała, że błędy predykcji były symetrycznie rozłożone wokół zera, a większość przypadków mieściła się w przedziale od  $-10$  do  $+10 \text{ W/m}^2$ . Obecność pojedynczych dużych błędów (do  $\pm 300 \text{ W/m}^2$ ) była ograniczona i typowa dla danych rzeczywistych.

Poniższa tabela prezentuje zestawienie metryk dla wszystkich modeli:

Model	$R^2$	RMSE [ $\text{W/m}^2$ ]
Regresja liniowa	0,853	43,42
Drzewo decyzyjne	0,861	42,12
Random Forest Regressor	0,875	40,00

Wizualizacje wspierające analizę zawierały:

- wykresy rzeczywistych vs przewidywanych wartości irradacji,
- wykresy ważności cech (feature importance),
- histogram rozkładu reszt dla modelu Random Forest.

Ostatecznie, model Random Forest Regressor został uznany za najbardziej efektywny i może być rekomendowany jako podstawowy model predykcyjny dla prognozowania uzysku energii z instalacji PV w warunkach polskiego klimatu.

Na podstawie dotychczasowych analiz stwierdzono, że model Random Forest Regressor osiągnął najlepsze wyniki i wykazał najwyższą stabilność predykcyjną. W związku z tym, w dalszej części pracy, to właśnie ten model zostanie zastosowany do analiz dla pozostałych krajów europejskich.

W ramach praktycznej weryfikacji przydatności modelu Random Forest przeprowadzono także prognozę irradacji  $G(i)$  na rok 2030 na podstawie przewidywanych warunków pogodowych ( $T2m$ ,  $WS10m$ ). Wynik uzyskany z modelu Random Forest porównano z prognozą opartą na prostym modelu regresji liniowej trendu historycznego irradacji. Obie prognozy były zbliżone — odpowiednio  $77,97 \text{ W/m}^2$  (Random Forest) oraz  $77,31 \text{ W/m}^2$  (regresja liniowa). Niewielka różnica między wynikami potwierdza spójność podejść oraz dodatkowo wzmacnia wiarygodność modelu zespołowego jako narzędzia predykcyjnego.

W kolejnym rozdziale przedstawione zostaną porównania międzynarodowe oraz wizualizacja wyników z wykorzystaniem narzędzi typu Business Intelligence.

## **6. Porównanie międzynarodowe**

### **6.1. Dobór lokalizacji do analizy porównawczej**

W celu rozszerzenia analizy przeprowadzonej dla lokalizacji w Zwartowie, postanowiono porównać potencjał irradacji ( $G(i)$ ) w wybranych miastach reprezentujących różne regiony klimatyczne Europy. Celem tego porównania jest ocena konkurencyjności Polski na tle innych krajów pod względem uzysku energii z instalacji fotowoltaicznych.

Do analizy wybrano następujące lokalizacje:

Lokalizacja	Kraj	Współrzędne geograficzne	Region klimatyczny
Zwartowo	Polska	54.698° N, 17.813° E	Europa Środkowa
Monachium	Niemcy	48.1351° N, 11.5820° E	Europa Centralna
Rzym	Włochy	41.9028° N, 12.4964° E	Europa Południowa
Lizbona	Portugalia	38.7169° N, -9.1399° E	Europa Zachodnia

Lokalizacje dobrano w taki sposób, aby reprezentowały zróżnicowane warunki klimatyczne, od chłodniejszych stref nadbałtyckich po ciepłe południowe regiony basenu Morza Śródziemnego.

## 6.2. Źródło danych i metodologia

Dane meteorologiczne dla każdej z lokalizacji zostały pobrane z bazy PVGIS - Photovoltaic Geographical Information System, zarządzanej przez Wspólne Centrum Badawcze Komisji Europejskiej (JRC) analogicznie jak w przypadku analizy dla Zwartowa. Zakres czasowy obejmował lata 2005–2023. Dla zapewnienia porównywalności wyników, dla wszystkich lokalizacji zastosowano identyczną procedurę:

- wczytanie i wstępne oczyszczenie danych w Pythonie,
- agregacja do wartości średniorocznych dla parametrów: temperatura powietrza (T2m), wysokość Słońca (H\_sun), prędkość wiatru (WS10m), irradiaacja G(i),
- budowa liniowych modeli trendu dla T2m i WS10m,
- predykcja wartości na rok 2030 i przygotowanie zestawu wejściowego dla modelu Random Forest (wytrenowanego wcześniej dla Zwartowa),
- oszacowanie irradiaacji G(i) w 2030 r. w każdej lokalizacji i porównanie wyników.

Dzięki takiemu podejściu możliwe było bezpośrednie zestawienie prognoz irradacji dla regionów o różnych warunkach klimatycznych.

Wyniki prognoz zostały przedstawione w kolejnym podrozdziale (6.3) w formie tabelarycznej i graficznej.

Dla porównania z regionami o znacząco niższym nasłonecznieniu, przeprowadzono również analizę trendu irradacji w Uppsali (Szwecja). Na podstawie danych z lat 2005–2023, liniowy model regresji oszacował wartość irradacji  $G(i)$  na rok 2030 na poziomie 72,26 W/m<sup>2</sup>, co jest wyraźnie niższe niż w Polsce (Zwartowo – 77,87 W/m<sup>2</sup>) oraz pozostałych krajach południowych. Pokazuje to, że Polska znajduje się w korzystnym położeniu, jeśli chodzi o potencjał produkcji energii słonecznej w skali europejskiej.

### **6.3. Porównanie prognoz irradacji na 2030 rok**

Na podstawie przeprowadzonych analiz i prognoz dla wybranych lokalizacji europejskich, oszacowano średnioroczne wartości irradacji  $G(i)$  na rok 2030 z wykorzystaniem wytrenowanego wcześniej modelu Random Forest. Dane meteorologiczne dla każdej lokalizacji zostały pobrane z bazy PVGIS (JRC, 2023). Dla każdej lokalizacji przygotowano zestaw predykcyjny obejmujący:

- prognozowaną średnią roczną temperaturę powietrza ( $T_{2m}$ ),
- prognozowaną średnią roczną prędkość wiatru ( $WS_{10m}$ ),
- średnią roczną wysokość Słońca ( $H_{sun}$ ) – przyjętą jako wartość historyczna dla danej lokalizacji.

Wyniki predykcji prezentuje poniższa tabela:

Lokalizacja	Kraj	T2m [°C]	WS10m [m/s]	H_sun [°]	Prognoza G(i) [W/m <sup>2</sup> ]
Zwartowo	Polska	9.4	3.43	18.51	77.87
Monachium	Niemcy	10.26	2.45	19.42	112.04
Rzym	Włochy	17.4	2.32	24.14	142.81
Lizbona	Portugalia	17.96	2.53	25.32	163.06

Na powyższym zestawieniu widać wyraźny wzrost irradacji w kierunku południowym. W warunkach polskich prognozowana irradacja osiąga wartość 77.87 W/m<sup>2</sup>, co jest zauważalnie niższe niż w przypadku Niemiec (112.04 W/m<sup>2</sup>), Włoch (142.81 W/m<sup>2</sup>) czy Portugalii (163.06 W/m<sup>2</sup>).

Wnioski z porównania:

- Polska, choć nie osiąga irradacji południowych krajów Europy, znajduje się w korzystniejszej pozycji niż kraje skandynawskie (np. Szwecja), gdzie wartość trendu liniowego dla irradacji na 2030 r. wyniosła jedynie ok. 72.26 W/m<sup>2</sup>.

- Monachium – mimo położenia bardziej na północ niż Rzym – osiąga wyraźnie wyższe wartości irradacji niż Polska, co może świadczyć o mniejszym zachmurzeniu i lepszym profilu nasłonecznienia.

- Lizbona prezentuje się jako najbardziej korzystna lokalizacja pod względem potencjału PV, z prognozowaną irradacją przekraczającą 160 W/m<sup>2</sup>.

Może to mieć istotne znaczenie dla firm planujących ekspansję fotowoltaiczną w Europie – Polska, nadal oferuje atrakcyjny poziom nasłonecznienia przy niższych kosztach gruntów i instalacji niż w Europie Zachodniej.

W kolejnym rozdziale dokonano interpretacji uzyskanych wyników w kontekście praktycznego potencjału inwestycyjnego oraz konkurencyjności Polski względem krajów Europy Zachodniej i Południowej.

## 7. Analiza praktyczna i inwestycyjna

### 7.1. Koszty instalacji PV w wybranych krajach

W celu dokonania realistycznego porównania potencjału inwestycyjnego systemów fotowoltaicznych, niezbędne jest uwzględnienie nie tylko poziomu irradacji, ale również aktualnych kosztów instalacji PV w różnych krajach.

Dane kosztowe (średnie ceny instalacji PV na poziomie gospodarstwa domowego, €/kWp) (Fraunhofer ISE, 2023; IEO, 2023; GSE, 2023):

Kraj	Średni koszt instalacji PV [€/kWp]	Źródło danych
Polska	1100	IEO, 2023
Niemcy	1450	Fraunhofer, 2023
Włochy	1600	GSE, 2023
Portugalia	1500	APREN, 2023

Ceny te obejmują koszt projektu, modułów PV, falowników, montażu oraz podstawowej konfiguracji technicznej dla systemów o mocy 3–10 kWp. Dane mogą się różnić regionalnie, ale stanowią dobrą bazę porównawczą dla analizowanej skali.

## 7.2. Porównanie irradacji a kosztów instalacji

Po zestawieniu danych o irradacji z kosztami instalacji PV, możliwe jest przeprowadzenie wstępnej analizy efektywności potencjalnych inwestycji.

Poniższa tabela prezentuje wskaźnik opłacalności w uproszczonej formie: irradacja  $G(i)$  podzielona przez koszt instalacji ( $W/m^2$  na 1 €/kWp), co może wskazywać na relatywną efektywność inwestycji.

Lokalizacja	$G(i)$ prognoza 2030 ( $W/m^2$ )	Koszt instalacji (€/kWp)	Wskaźnik efektywności ( $G(i)/€/kWp$ )
Zwartowo	77.87	1100	0.071
Monachium	112.04	1450	0.077
Rzym	142.81	1600	0.089
Lizbona	163.06	1500	0.109

Jak wynika z powyższych danych, mimo niższego poziomu irradacji, Polska może konkurować kosztowo z Niemcami. Jednak w ujęciu absolutnym, najwyższą opłacalność inwestycji wykazuje Lizbona, dzięki wysokiej irradacji oraz umiarkowanemu kosztowi instalacji.

## 7.3. Szacowany czas zwrotu inwestycji (simple payback time)

Aby przeprowadzić uproszczoną analizę zwrotu z inwestycji, przyjmujemy:

- koszt instalacji PV dla gospodarstwa domowego o mocy 5 kWp (średni przypadek),

- Sprawność systemu (po uwzględnieniu strat): 0,85,
- Cena energii elektrycznej (dla gospodarstw domowych):
  - Polska: 0.20 €/kWh
  - Niemcy: 0.30 €/kWh
  - Włochy: 0.25 €/kWh
  - Portugalia: 0.24 €/kWh
- prosty model „simple payback time”:

(SPBT) = Koszt instalacji/(Roczna produkcja energii×Cena energii):

- Produkcja roczna energii:

$$E=G(i) \times 5 \times 0,85$$

Obliczenia:

Lokalizacja	G(i) [W/m <sup>2</sup> ]	Produkcja roczna [kWh]	Koszt [€]	Cena energii [€/kWh]	SPBT (lata)
Zwartowo	77.87	$5 \times 77.87 \times 0.85 = 331.45$	5500	0.20	8.3
Monachium	112.04	$5 \times 112.04 \times 0.85 = 476.17$	7250	0.30	5.1
Rzym	142.81	$5 \times 142.81 \times 0.85 = 607.0$	8000	0.25	5.3
Lizbona	163.06	$5 \times 163.06 \times 0.85 = 693.0$	7500	0.24	4.5



Wnioski:

- Najkrótszy czas zwrotu odnotowano dla Lizbony, co czyni ją szczególnie atrakcyjną lokalizacją inwestycyjną.
- Polska (Zwartowo) ma najdłuższy czas zwrotu, ale nadal pozostaje w akceptowalnych granicach (poniżej 10 lat).
- Monachium i Rzym prezentują zbliżoną efektywność inwestycyjną, mimo różnic klimatycznych

## 8. Wnioski końcowe

Na podstawie przeprowadzonych analiz można sformułować następujące wnioski dotyczące potencjału wykorzystania energii słonecznej w Polsce na tle innych krajów europejskich:

1. Polska wciąż ma umiarkowany, ale stabilny potencjał irradycyjny

Prognozowana wartość irradycji  $G(i)$  dla lokalizacji w Zwartowie w 2030 roku wyniosła  $77,87 \text{ W/m}^2$ , co plasuje Polskę powyżej krajów skandynawskich (np. Szwecja:  $\sim 72,26 \text{ W/m}^2$ ), ale wyraźnie poniżej krajów południowej Europy (Portugalia:  $163,06 \text{ W/m}^2$ , Włochy:  $142,81 \text{ W/m}^2$ ).

To wskazuje, że Polska nie jest liderem w zakresie potencjału słonecznego, ale może być konkurencyjna przy korzystnych warunkach inwestycyjnych.

2. Koszty instalacji PV w Polsce należą do najniższych

Z analiz wynika, że Polska posiada jeden z najniższych kosztów instalacji PV (ok. 1100 €/kWp), co czyni ją atrakcyjną z punktu widzenia kosztów wejścia na rynek. W porównaniu do Niemiec (1450 €/kWp) czy Włoch (1600 €/kWp), inwestycja w Polsce wymaga mniejszego kapitału początkowego.

### 3. Czas zwrotu inwestycji w Polsce jest nadal akceptowalny

Chociaż czas zwrotu inwestycji w Polsce (ok. 8,3 roku) jest dłuższy niż w Portugalii (4,5 roku) czy Niemczech (5,1 roku), pozostaje on w akceptowalnych granicach dla konsumentów indywidualnych i inwestorów. Uwzględniając niskie koszty instalacji oraz stabilizujący się rynek energetyczny, inwestycje PV w Polsce są racjonalne ekonomicznie.

### 4. Potencjał wzrostu i rola polityki wspierającej

Dzięki wsparciu programów takich jak „Mój Prąd” czy ulg podatkowych, Polska ma szansę zwiększyć udział energii słonecznej w miksie energetycznym. Rozsądne decyzje inwestycyjne w PV mogą stanowić realną alternatywę dla tradycyjnych źródeł energii, zmniejszając zależność od importu oraz emisję CO<sub>2</sub>.

### Rekomendacje

- Dla inwestorów prywatnych: Polska stanowi rozsądną lokalizację inwestycyjną z akceptowalnym czasem zwrotu i niskimi kosztami startowymi.
- Dla polityki energetycznej: Należy kontynuować i wzmacniać instrumenty wsparcia dla rozwoju OZE, szczególnie mikroinstalacji PV.
- Dla kolejnych badań: Można rozważyć rozszerzenie analizy o modelowanie sezonowe, efekty zachmurzenia oraz wpływ orientacji instalacji PV.

## 9. Bibliografia

- Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Eurostat. (2023). Electricity prices for household consumers – bi-annual data. <https://ec.europa.eu/eurostat>
- Fraunhofer ISE. (2023). Photovoltaics Report. <https://www.ise.fraunhofer.de>
- GSE – Gestore dei Servizi Energetici. (2023). Rapporto Statistico sul Solare Fotovoltaico. <https://www.gse.it>
- IEO – Instytut Energetyki Odnawialnej. (2023). Rynek Fotowoltaiki w Polsce 2023. <https://ieo.pl>
- JRC – Joint Research Centre. (2023). Photovoltaic Geographical Information System (PVGIS). European Commission. <https://ec.europa.eu/jrc/en/pvgis>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
- Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. CreateSpace.
- APREN – Associação de Energias Renováveis. (2023). Relatório Anual. <https://www.apren.pt>