

# Model predykcyjny produkcji energii elektrycznej z instalacji fotowoltaicznych w Polsce w zależności od lokalizacji i warunków pogodowych

Joanna Wójcicka, Jakub Truszkowski

5 stycznia 2024

## Streszczenie

Praca skupia się na opracowaniu modelu predykcyjnego produkcji energii elektrycznej przez instalacje fotowoltaiczne zlokalizowane na terenie Polski. Celem badania jest zrozumienie wpływu geograficznego położenia oraz warunków pogodowych na efektywność generacji energii z paneli fotowoltaicznych.

Wykorzystując dane z liczników produkcji energii elektrycznej przez instalacje fotowoltaiczne, w połączeniu z danymi pogodowymi uzyskanymi z Photovoltaic Geographical Information System (PVGIS), opracowano narzędzie, które może być efektywnie używane do optymalizacji lokalizacji nowych instalacji fotowoltaicznych oraz do weryfikacji i oceny opłacalności decyzji inwestycyjnych.

## 1 Wstęp

W miarę postępu technologicznego i wzrastającej świadomości ekologicznej, rosnące zainteresowanie odnawialnymi źródłami energii stawia przed nami wyzwanie zrównoważonego i efektywnego wykorzystania dostępnych zasobów. W kontekście energii elektrycznej, przydomowe instalacje fotowoltaiczne stanowią istotny element transformacji energetycznej, umożliwiając jednostkom domowym produkowanie czystej energii przy minimalnym wpływie na środowisko.

Niniejsza praca skupia się na analizie produkcji energii elektrycznej z przydomowych instalacji fotowoltaicznych w Polsce, z wykorzystaniem danych z liczników oraz informacji pogodowych dostarczanych przez Photovoltaic Geographical Information System (PVGIS). Celem badania jest opracowanie modelu predykcyjnego, który pozwoli zoptymalizować lokalizację nowych instalacji fotowoltaicznych, z uwzględnieniem zarówno warunków geograficznych, jak i czynników pogodowych.

Przydomowe instalacje fotowoltaiczne stanowią kluczowy składnik decentralizacji produkcji energii, przyczyniając się do zrównoważonej dystrybucji źródeł zasilania. Skala rozwoju tego sektora w Polsce zaznacza się coraz wyraźniej, co niesie za sobą potrzebę zastosowania narzędzi analitycznych, które pozwolą na efektywne planowanie i zarządzanie tym rodzajem infrastruktury.

Praca składa się z kilku kluczowych etapów, obejmujących analizę danych geograficznych dotyczących lokalizacji instalacji fotowoltaicznych, zebranie informacji meteorologicznych, oraz wykorzystanie algorytmów uczenia maszynowego w celu stworzenia modelu predykcyjnego.

## 2 Kwerenda literaturowa

Przegląd literatury pozwolił nam na identyfikację kilku istotnych prac z zakresu prognozowania produkcji energii z instalacji fotowoltaicznych. Jednakże, jedna praca w szczególności okazała się być szczególnie przydatna dla naszego projektu. Praca ta, zatytułowana "Solar Energy Prediction Model Based on Artificial Neural Networks and Open Data" [1], dostarczała cennych informacji i podejścia do problemu.

Autorzy omówili, jak zmiany klimatyczne wpływają na rosnące zainteresowanie zrównoważonym rozwojem oraz odnawialnymi źródłami energii. Zwrocono uwagę na fakt, że niepewność dotycząca niektórych źródeł energii odnawialnej wynika z ich niskiej niezawodności, szczególnie takich jak zmienne warunki atmosferyczne.

Rysunek 1: Mapa lokalizacji wszystkich 9182 udostępnionych instalacji.



W kontekście energii słonecznej, autorzy zwrócili uwagę na dwie główne formy - ciepłą i fotowoltaiczną. Skoncentrowali się na energii fotowoltaicznej, która jest szeroko stosowana w Europie. Wartości efektywności produkcji energii z paneli fotowoltaicznych były analizowane w kontekście różnych warunków atmosferycznych i lokalizacji.

Jednym z ważniejszych rezultatów pracy było zaproponowanie narzędzia opartego na sztucznych sieciach neuronowych, które umożliwia łatwe szacowanie produkcji energii z paneli fotowoltaicznych w danym obszarze. Autorzy podkreślili, że ich podejście koncentruje się na produkcji energii, a nie tylko na promieniowaniu słonecznym. Podejście to zastosowaliśmy również w naszym projekcie.

W związku z powyższym, praca ta stanowiła cenną inspirację i wzór dla naszego projektu, szczególnie w kontekście wykorzystania sztucznych sieci neuronowych, czy w doborze parametrów uczenia.

Przydatnym źródłem informacji była także praca "Deep Neural Networks for Wind and Solar Energy Prediction" [2], której autorzy zdecydowali się korzystać z głębokich sieci neuronowych w celu przewidywania produkcji energii z farm wiatrowych oraz słonecznych.

## 3 Dane

### 3.1 Dane lokalizacyjne instalacji

Zbiór danych, który wykorzystujemy do uczenia modelu, pochodzi z odczytów z mierników produkcji energii elektrycznej instalacji fotowoltaicznych, udostępniony nam w sposób zachowujący anonimowość właściciela przez firmę [PvMonitor](#).

Załączona mapa 1 zawiera lokalizacje znajdujące się nie tylko w Polsce, ale także na terytorium innych państw oraz akwenu Morza Bałtyckiego. W dalszej pracy upewniliśmy się, by punkty te zostały skutecznie odfiltrowane.

### 3.2 Dane pomiarowe instalacji

Uzyskane dane dotyczyły roku 2020 i obejmowały 3028 punktów i dla każdej instalacji zawierały wartość całkowitą wyprodukowanej energii(p), przyrost względem poprzedniego odczytu(dp) oraz efektywność instalacji( $\epsilon$ ), obliczaną ze wzoru:

$$\epsilon = \frac{dp}{pow} \quad (1)$$

Operujemy tu już tylko na 3028 punktach pomiarowych, co stanowi mniej niż jedną trzecią z początkowych 9182. Różnica ta spowodowana jest znaczącym wzrostem liczby zainstalowanych przez firmę mierników po roku 2020, dla których nie mamy dostępu do miesięcznych pomiarów.

### 3.3 Agregacja danych

Uzyskany model predykcyjny przewiduje miesięczne wartości produkcji energii, istniała więc konieczność agregacji danych, by odczyty nie dotyczyły okresów trzymiutowych, a miesięczne. Przy użyciu prostego skryptu, zsumowaliśmy wartości produkcji energii, by dotyczyły okresów miesięcznych. Agregacja ta spowodowała zauważalną zmianę liczby rekordów ze 145.068.684 na zaledwie 26.345. Przy agregacji dokonaliśmy od razu pominięcia instalacji nie posiadających wszystkich odczytów dla któregoś miesiąca. Zdecydowaliśmy w takich sytuacjach pomijać też odczyty z pozostałych miesięcy dla tych instalacji, co uzasadniamy ryzykiem wadliwości miernika, bądź błędu przy instalacji lub przesyłaniu danych.

### 3.4 Filtracja danych

Instalacje przydomowe odznaczają się znaczną zmiennością efektywności, która wynika z różnorodnych czynników. Wśród głównych przyczyn zmienności należy wymienić zróżnicowane kąty nachylenia dachów na budynkach, co bezpośrednio wpływa na optymalne wykorzystanie promieni słonecznych. Dodatkowo, fizyczne przeszkody na dachu, takie jak kominy, drzewa czy inne struktury, mogą wprowadzać cienie i blokować bezpośredni dostęp do światła słonecznego, co negatywnie wpływa na efektywność instalacji.

Warto również zaznaczyć, że błędy koncepcyjne lub instalacyjne, takie jak nieprawidłowe rozmieszczenie paneli fotowoltaicznych czy nieoptymalne ustawienie urządzeń, mogą dodatkowo obniżać efektywność całego systemu. Biorąc powyższe aspekty pod uwagę, postawiliśmy sobie za nadrzędny cel dobranie odpowiednich warunków filtracyjnych.

Rysunek 2 przedstawia efektywność instalacji w zależności od miesiąca przed filtracją. Warto zauważyć, że konieczność wykonania filtracji danych przejawia się obecnością wielu odczytów w bliskiej zeru odległości. Ostateczny rezultat powinien ukazać odczyty układające się w trend zbliżony do funkcji wielomianowej, osiągającej dwa maksima na przełomie marca i kwietnia oraz sierpnia i września. Występowanie tych maksimów jest ściśle powiązane z typowym nachyleniem paneli fotowoltaicznych w Polsce, które wynosi 45 stopni. Takie ustawienie sprzyja idealnemu kątowi padania promieni słonecznych na panele wiosną i jesienią, co skutkuje najlepszymi wynikami w tych porach roku.

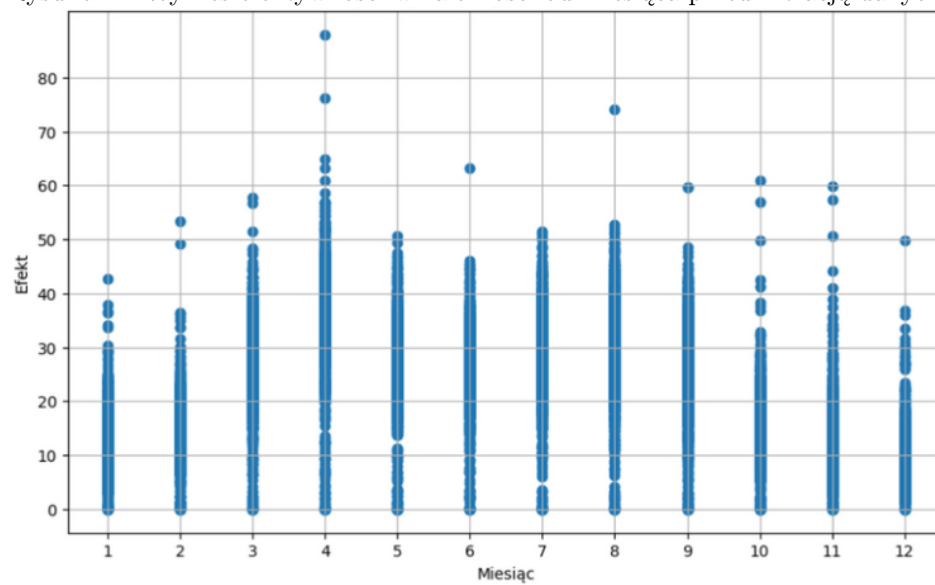
Biorąc to wszystko pod uwagę, zdecydowaliśmy się na filtrację danych odstających, które identyfikujemy jako:

- Instalacje, dla których brakuje ważnych parametrów takich jak moc
- Instalacje, które przynajmniej raz w miesiącu nie osiągnęły efektywności  $\geq 40$
- Instalacje, które w miesiącach zimowych (styczeń, luty, marzec, październik, listopad, grudzień) nie osiągnęły średniej miesięcznej efektywności rzędu 5
- Instalacje, które w miesiącach letnich (kwiecień - wrzesień) nie osiągnęły średniej miesięcznej efektywności rzędu 10

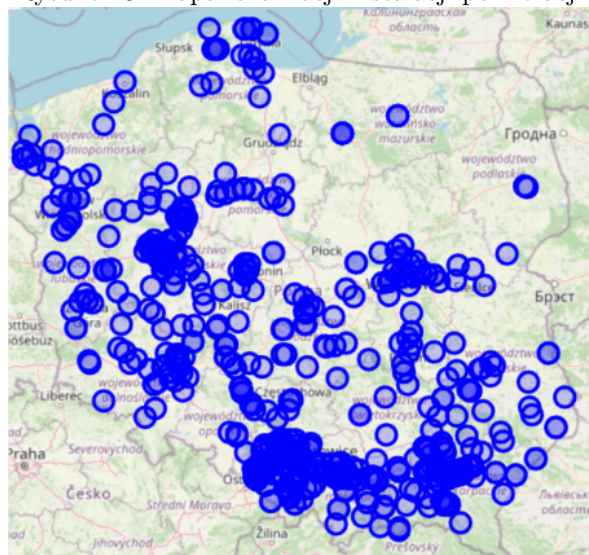
Ponadto zdecydowaliśmy się też zrezygnować z instalacji, dla których nie posiadamy wszystkich odczytów z okresu 12 miesięcy.

Po uwzględnieniu powyższych filtrów, zostaliśmy z ostatecznym zbiorem 541 instalacji, przedstawionych na mapie 3 i 6.492 odczytów.

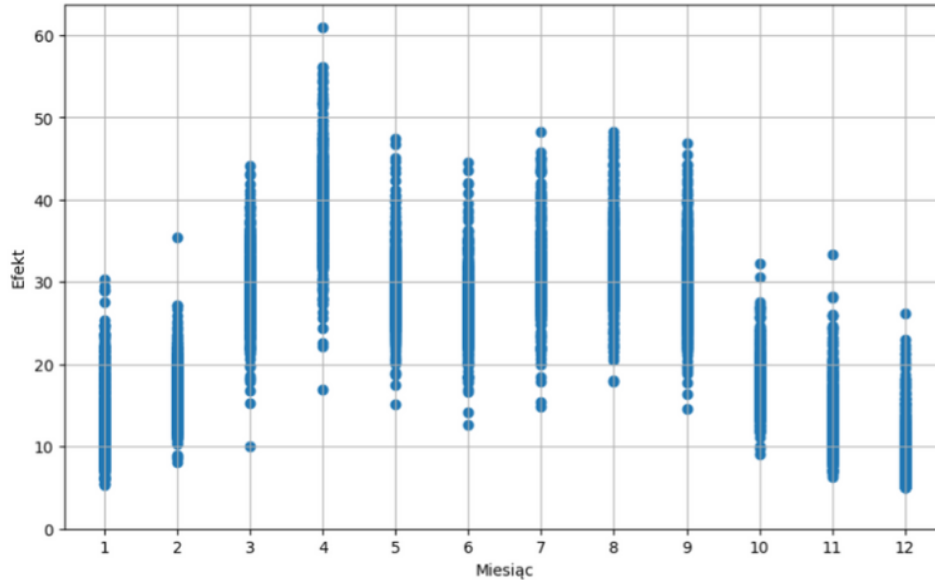
Rysunek 2: Wykres efektywności w zależności od miesiąca przed filtracją danych.



Rysunek 3: Mapa lokalizacji instalacji po filtracji.



Rysunek 4: Wykres efektywności w zależności od miesiąca po filtracji danych.



### 3.5 Wynik filtracji

Wyniki filtracji danych przedstawiają się obiecująco, eliminując znaczną część danych odstających, obserwowanych na wykresie przed filtracją. Po zastosowaniu odpowiednich technik filtrowania, odczyty układają się w bardziej klarowny kształt, co pozwala lepiej dostrzec istniejące tendencje. Zredukowanie odczytów bliskich zeru pozwala na dokładniejsze zobaczenie charakterystyki efektywności produkcji w zależności od miesiąca na rysunku 4.

### 3.6 Dane pogodowe

Ostatnim zbiorem danych niezbędnym do kompleksowego trenowania modelu są informacje pogodowe. Wpływ warunków atmosferycznych na efektywność instalacji fotowoltaicznych jest kluczowy. Aby uzyskać pełniejszy obraz tych zależności, skorzystaliśmy z [Photovoltaic Geographical Information System \(PVGIS\)](#), dostarczającego średnie miesięczne dane dotyczące warunków pogodowych dla każdej lokalizacji. Pogrupowane według miesięcy, te informacje obejmują kluczowe aspekty, takie jak temperatura i nasłonecznienie. Wprowadzenie tych danych do modelu umożliwi uwzględnienie wpływu warunków pogodowych na efektywność instalacji, co pozwoli na bardziej precyzyjne prognozowanie.

## 4 Model Predykcyjny

Aby stworzyć jak najlepszy model predykcyjny efektywności instalacji fotowoltaicznych, poszukiwaliśmy najbardziej efektywnego algorytmu uczenia maszynowego dla naszego problemu. Z racji, że naszym celem było przewidzenie wartości efektywności w zależności od lokalizacji i warunków pogodowych, zdecydowaliśmy się na użycie modelu regresyjnego.

### 4.1 Ocena Modelu

Do oceny skuteczności modelu wybraliśmy 4 metryki, na które składają się:

- Mean Squared Error (MSE): MSE mierzy średnią kwadratową różnicę pomiędzy rzeczywistymi a przewidywanymi wartościami. Im niższa wartość MSE, tym lepiej model radzi sobie z dopasowywaniem danych.
- R-squared ( $R^2$ ): R-squared ilustruje stopień, do jakiego model wyjaśnia zmienność danych. Wartość R-squared między 0 a 1, gdzie 1 oznacza doskonałe dopasowanie modelu do danych.

- Mean Absolute Error (MAE): MAE to średnia wartość bezwzględna różnic między rzeczywistymi a przewidywanymi wartościami. Jest miarą bezwzględnego odchylenia przewidywań modelu.
- Explained Variance Score: Explained Variance Score oblicza, jak dobrze model wyjaśnia wariancję danych. Wartość ta mieści się w zakresie od 0 do 1, gdzie 1 oznacza doskonałe dopasowanie.

Poprzez analizę tych miar otrzymujemy kompleksową ocenę zdolności modelu do generalizacji na nowe dane oraz jego skuteczności w przewidywaniu docelowych wartości. Przyjęcie wielu miar pozwala nam uzyskać pełniejszy obraz efektywności modelu w różnych aspektach predykcyjnych.

## 4.2 Przygotowanie Danych

Przed przystąpieniem do trenowania modelu, dokonaliśmy odpowiedniego przygotowania danych. Zbiór danych zawierał informacje o miesiącu, mocy instalacji, długości i szerokości geograficznej, nasłonecznieniu oraz temperaturze. Przeprowadziliśmy standaryzację danych przy użyciu Standard Scalera [3].

## 4.3 Wybór Modelu

W procesie selekcji modelu regresyjnego rozważaliśmy różne opcje dostępne w bibliotece scikit-learn [4]. Ze względu na zdolność do modelowania nieliniowych zależności i elastyczność konfiguracji, ostatecznie wybraliśmy **MLPRegressor** (Multi-layer Perceptron Regressor) - model sztucznej sieci neuronowej. Pozwolił nam on uzyskać lepsze wyniki niż inne głębokie sieci neuronowe w wielu przetestowanych konfiguracjach.

W celu znalezienia najlepszej liczby warstw ukrytych i ich liczby neuronów, przeprowadziliśmy iteracyjne testowanie różnych ustawień. Ostatecznie, najlepszą konfiguracją okazała się taka z dwoma warstwami ukrytymi o rozmiarze 250 i 260 neuronów, uczona w tysiącu iteracji.

```
model = MLPRegressor(hidden_layer_sizes=(250, 260), max_iter=1000,
random_state=42)
```

Podczas procesu oceny wydajności modelu, skorzystaliśmy z walidacji krzyżowej w celu dokładnej oceny jego zdolności generalizacji na zbiorze testowym. Poniżej znajdują się wyniki tej oceny:

### 4.3.1 Wyniki na zbiorze testowym z użyciem walidacji krzyżowej

- Mean Squared Error: *17.02*
- R-squared: *0.83*
- Mean Absolute Error: *3.04*
- Explained Variance Score: *0.83*

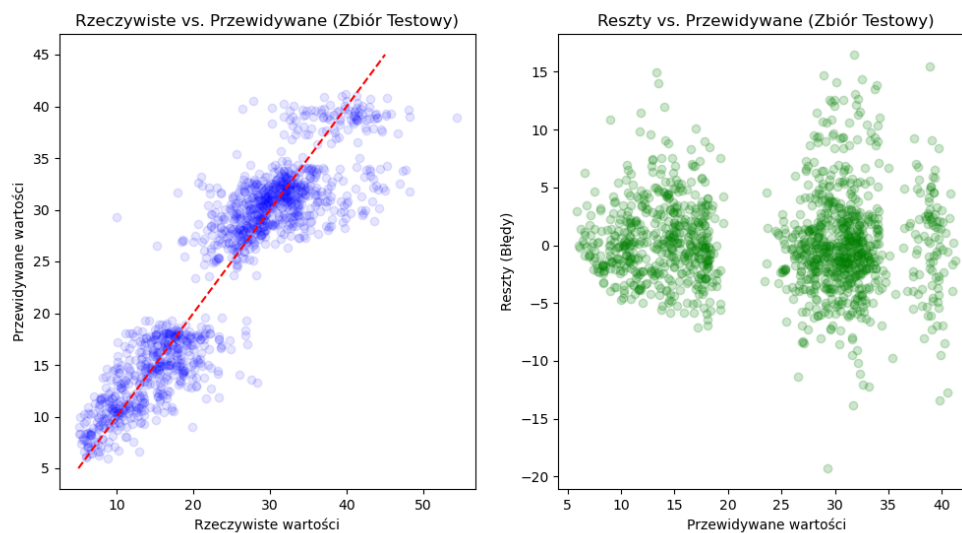
### 4.3.2 Wyniki na zbiorze treningowym

- Mean Squared Error: *15.61*
- R-squared: *0.84*
- Mean Absolute Error: *2.91*
- Explained Variance Score: *0.84*

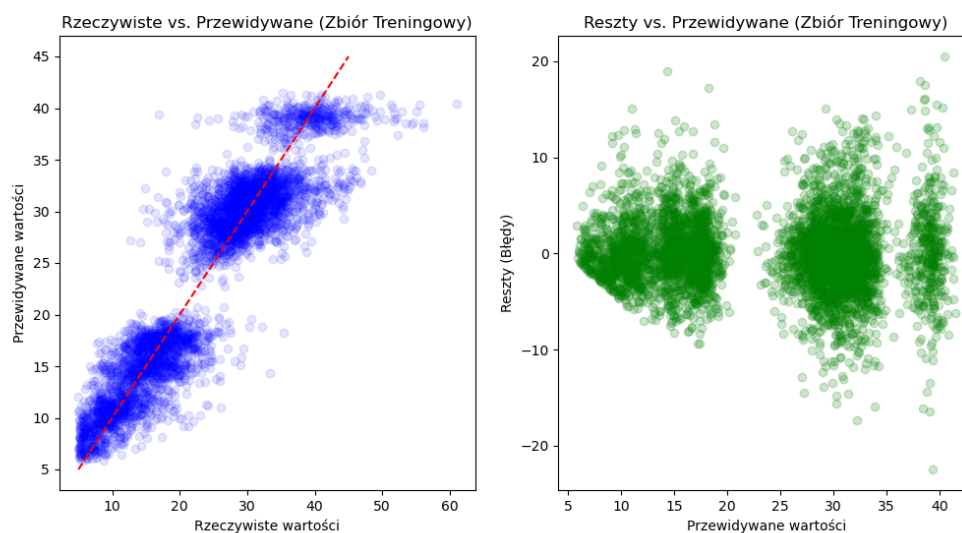
Wizualizacja wyników przedstawiona na rysunku 5 oraz 6 pokazuje porównanie prognoz modelu na zbiorze testowym i treningowym. Otrzymane wyniki potwierdzają skuteczność wybranego modelu regresyjnego w kontekście predykcji wartości efektywności. Średni błąd bezwzględny na poziomie 3,04 jest jak najbardziej akceptowalny w kontekście przewidywania efektywności, która przyjmuje wartości z przedziału 0-100.



Rysunek 5: Wykresy wartości przewidywanych w zależności do wartości rzeczywistych oraz błędów w zależności do przewidywanych wartości dla zbioru testowego.



Rysunek 6: Wykresy wartości przewidywanych w zależności do wartości rzeczywistych oraz błędów w zależności do przewidywanych wartości dla zbioru treningowego.



## 4.4 Wnioski i konkluzje dotyczące działania modelu

Analizując wykresy 5 oraz 6, można zauważyć, że model najczęściej popełnia błędy w przewidywaniu efektywności dla najlepiej działających instalacji. To zjawisko może być związane z istnieniem grupy właścicieli instalacji, którzy osiągają znacznie lepsze wyniki niż pozostali właściciele. Warto podkreślić, że różnice te mogą być spowodowane różnymi czynnikami, takimi jak specyfika techniczna czy ekonomiczna danej instalacji.

Jednym z możliwych powodów tej sytuacji jest fakt, że instalacje przydomowe charakteryzują się znaczną zmiennością efektywności, co zostało szczegółowo omówione w sekcji 3.4. W analizie danych uwzględniliśmy tę zmienność, eliminując na przykład prężnie działające, małe farmy fotowoltaiczne, które mogły wprowadzać dodatkową trudność w prognozowaniu efektywności. Mimo jednak podejmowanych działań filtracyjnych, model wciąż wykazuje ograniczenia w przypadku instalacji, które wyróżniają się swoją efektywnością na tle innych.

Przynajmniej w przewidywaniu efektywności dla tych specyficznych przypadków może być brak odpowiednich danych treningowych lub subtelne niuanse związane z danymi, które są trudne do uchwycenia przez model. W trakcie analizy nie udało się znaleźć sposobu na poprawienie tego stanu rzeczy bez negatywnego wpływu na ogólną skuteczność modelu. Jest to obszar wymagający dalszych badań i optymalizacji modelu w celu lepszego radzenia sobie z różnorodnością efektywności instalacji fotowoltaicznych.

## 5 Interfejs graficzny

Ta aplikacja została zaprojektowana jako narzędzie prezentujące działanie modelu przewidującego produkcję energii elektrycznej. Aplikacja posiada prosty interfejs graficzny, który pozwala użytkownikowi wybrać lokalizację w jeden z dwóch sposobów:

- Wprowadzenie Adresu: Użytkownik może wprowadzić konkretny adres, na podstawie którego aplikacja określi geograficzne współrzędne lokalizacji. Jest to przydatne w przypadku, gdy użytkownik chce sprawdzić potencjalną produkcję energii w określonym miejscu.
- Zczytanie Lokalizacji przez Adres IP: Aplikacja ma również możliwość automatycznego zlokalizowania użytkownika na podstawie jego adresu IP. Jest to szybki i wygodny sposób na określenie bieżącej lokalizacji użytkownika bez konieczności ręcznego wpisywania danych.

Po określeniu lokalizacji, użytkownik musi podać moc posiadanej instalacji fotowoltaicznej. Na podstawie tych danych oraz danych pogodowych z wybranej lokalizacji, aplikacja wykorzystuje wytrenowany model do przewidzenia ilości energii elektrycznej, którą można wygenerować. Model ten bierze pod uwagę nasłonecznienie oraz temperaturę w danych miesiącach, które mają wpływ na wydajność paneli słonecznych.

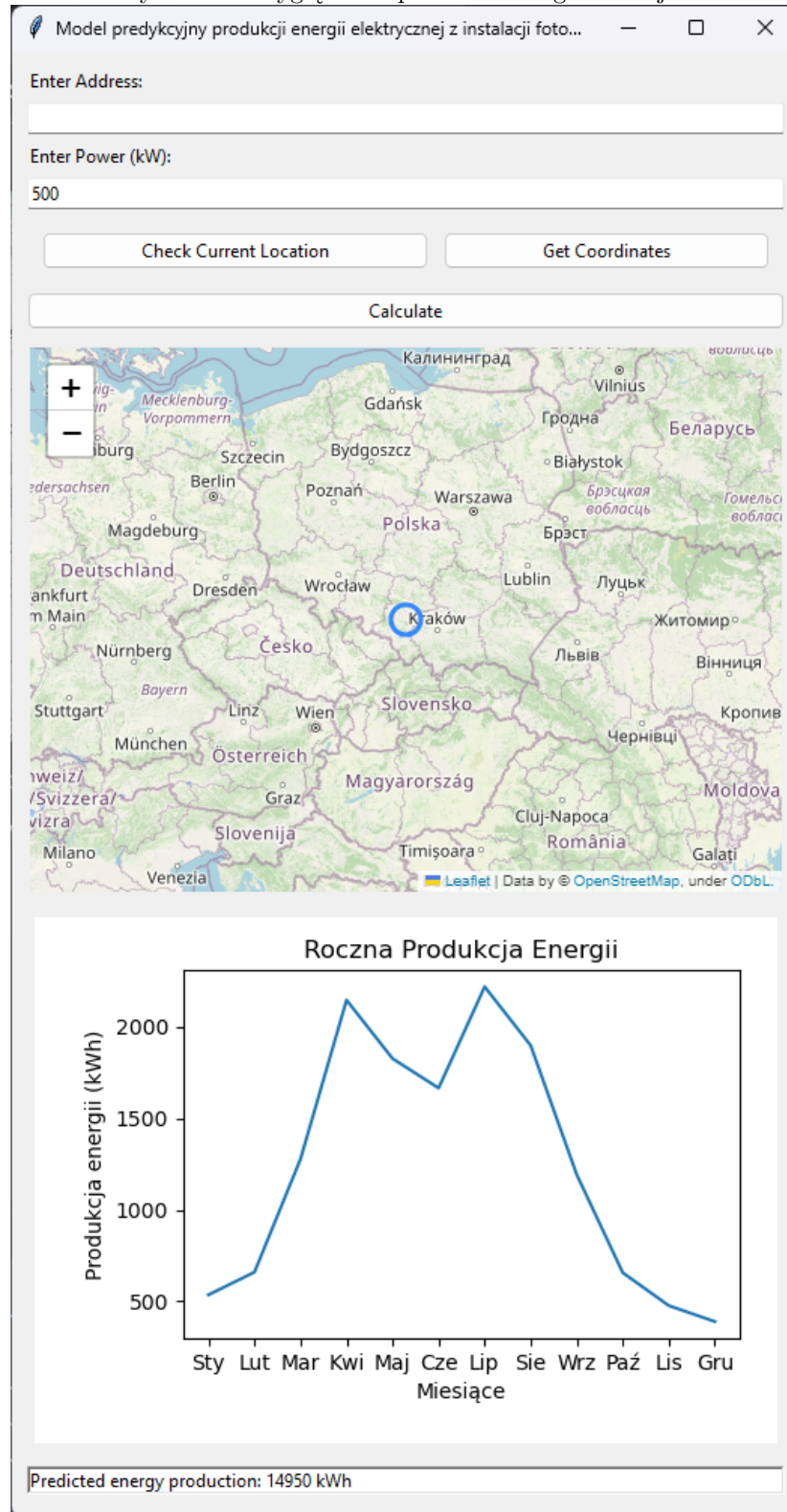
Zaimplementowany interfejs graficzny jest przyjazny dla użytkownika i umożliwia łatwe i szybkie uzyskanie szacowanych wartości produkcji energii. Pozwala to na lepszą ocenę potencjału instalacji fotowoltaicznych w różnych lokalizacjach, co może być pomocne zarówno dla osób prywatnych, jak i przedsiębiorstw planujących inwestycje w odnawialne źródła energii.

W aplikacji zaimplementowano funkcję generowania wykresu, który przedstawia przewidywaną roczną produkcję energii elektrycznej z instalacji fotowoltaicznych. Wykres ten umożliwia użytkownikom wizualizację miesięcznych estymacji produkcji energii, co ułatwia zrozumienie sezonowych zmian i efektywności paneli słonecznych w różnych warunkach pogodowych. Jest to istotne narzędzie dla użytkowników do analizy potencjału produkcyjnego instalacji fotowoltaicznych, wspierając decyzje związane z inwestycjami w odnawialne źródła energii.

Podsumowując, aplikacja ta stanowi praktyczne narzędzie do oceny możliwości produkcyjnych instalacji fotowoltaicznych, bazując na danych lokalizacyjnych i pogodowych, co jest kluczowe w kontekście rosnącego zainteresowania odnawialnymi źródłami energii.



Rysunek 7: Wygląd zaimplementowanego interfejsu



## Literatura

- <sup>1</sup> Alejandro Maté Juan Carlos Trujillo Jose Manuel Barrera, Alejandro Reina Reina. Solar energy prediction model based on artificial neural networks and open data. 2020. URL [https://www.researchgate.net/publication/343867043\\_Solar\\_Energy\\_Prediction\\_Model\\_Based\\_on\\_Artificial\\_Neural\\_Networks\\_and\\_Open\\_Data#fullTextFileContent](https://www.researchgate.net/publication/343867043_Solar_Energy_Prediction_Model_Based_on_Artificial_Neural_Networks_and_Open_Data#fullTextFileContent).
- <sup>2</sup> Adil Omari José R. Dorronsoro David Díaz-Vico, Alberto Torres-Barrán. Deep neural networks for wind and solar energy prediction. 2017. URL <https://link.springer.com/article/10.1007/s11063-017-9613-7>.
- <sup>3</sup> Scikit-learn preprocessing standard scaler documentation. URL <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>.
- <sup>4</sup> Scikit-learn neural network mlpregressor documentation. URL [https://scikit-learn.org/stable/modules/generated/sklearn.neural\\_network.MLPRegressor.html](https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html).