

Zastosowanie metod uczenia maszynowego do predykcji generowanej energii w małym systemie PV

Adam Idźkowski, Mateusz Sumorek

Politechnika Białostocka, Wydział Elektryczny, ul. Wiejska 45D 15-351 Białystok

Streszczenie: Artykuł przedstawia problem prognozowania generacji energii elektrycznej w małych systemach fotowoltaicznych (PV). Celem opracowanych długoterminowych prognoz jest możliwość poprawnego zarządzania systemem elektroenergetycznym poprzez podejmowanie odpowiednich działań zachowawczych. Przeanalizowano czynniki atmosferyczne wpływające na pozyskiwanie energii elektrycznej w systemach fotowoltaicznych. Dokonano porównania wybranych modeli prognostycznych z wykorzystaniem uczenia maszynowego, m.in. sieci neuronowych MLP oraz metody wektorów nośnych SVM. Zostały wybrane mierniki pozwalające określić trafność (dokładność) prognoz. Określenie jakości prognoz bazowało na stanach faktycznych pogody, a nie na jej prognozie. Przedstawiono sposób przygotowania danych do utworzenia modeli prognostycznych i zaprezentowano najlepsze modele regresyjne. Do tego celu wykorzystano bibliotekę Scikit-learn umożliwiającą tworzenie skryptów w języku Python. W rozpatrywanym zespole fotowoltaicznym najlepsze rezultaty uzyskano dla modeli MLPRegressor, CatBoostRegressor i SVR. Wykorzystano rzeczywiste dane pomiarowe z systemu paneli ustawionych optymalnie o mocy 3,0 kWp. Dla modelu MLPRegressor osiągnięto największy współczynnik determinacji 0,605 oraz najmniejszy pierwiastek błędu średniokwadratowego 1,79 kWh dla średniej dziennej generacji energii elektrycznej w okresie od kwietnia do września 2022 r. wynoszącej 11,65 kWh.

Słowa kluczowe: elektrownia fotowoltaiczna; uczenie maszynowe; prognozowanie wytwarzanej energii; modele regresji; sieci neuronowe; perceptron wielowarstwowy; regresja metodą wektorów nośnych; regresja grzbietowa

1. Wstęp

Tak nadwyżka, jak i deficyt wytwarzanej energii negatywnie wpływa na funkcjonowanie sieci elektroenergetycznej. Niezbędnym elementem poprawnego funkcjonowania systemu energetycznego jest dostosowanie wygenerowanej energii do aktualnego zapotrzebowania [1]. O ile znajomość zapotrzebowania pozwala na dostosowanie pozyskiwanej energii w elektrowniach konwencjonalnych, to nie jest to do końca możliwe w elektrowniach opartych o źródła odnawialne. W tym celu konieczne jest zrealizowanie zautomatyzowanego prognozowania energii pozyskiwanej przez odnawialne źródła energii – szczególnie gdy udział takich elektrowni jest znaczący w stosunku do pozostałych źródeł energii elektrycznej. Predykcja możliwości wytwarzania energii ze źródeł odnawial-

nych w odstępach minutowych, godzinowych, czy dziennych, umożliwia skuteczniejsze zarządzanie generowaną energią elektryczną oraz zabezpieczenie systemu elektroenergetycznego [2].

Stały rozwój gospodarki, zmiany w życiu codziennym, zmiany społeczne, czy też zmiany środowiskowe, wymuszają opracowywanie różnych profili zużycia energii elektrycznej. Zwiększenie udziału odnawialnych źródeł energii w pozyskiwaniu energii elektrycznej oraz stopniowe przekształcanie sieci elektroenergetycznej w sieć inteligentną powoduje pojawienie się z tym związanych nowych wyzwań [3]. Tak oto pojawia się zapotrzebowanie na metody prognozowania generacji energii elektrycznej przez elektrownie wiatrowe, czy też słoneczne, zależnie od warunków pogodowych [4]. W literaturze występują liczne publikacje z zakresu prognozowania pozyskiwanej energii elektrycznej w systemach fotowoltaicznych. Wiele z nich powstało w ostatnich kilku latach. Opracowania te opisują prognozy ultrakrótkoterminowe (kiluminutowe), krótkoterminowe (godzinne), średnioterminowe (dniowe, tygodniowe) oraz długoterminowe (miesięczne) z wykorzystaniem metod sztucznej inteligencji, najczęściej sztucznych sieci neuronowych.

Sztuczne sieci neuronowe są wykorzystywane w głębokim uczeniu DL (ang. *Deep Learning*), podkategorii uczenia maszynowego ML (ang. *Machine Learning*), aby przeprowa-

Autor korespondujący:

Adam Idźkowski, a.idzkowski@pb.edu.pl

Artykuł recenzowany

nadesłany 04.08.2023 r., przyjęty do druku 18.01.2024 r.



Zezwala się na korzystanie z artykułu na warunkach licencji Creative Commons Uznanie autorstwa 3.0

dzać w ich warstwach ukrytych proces uczenia i konstruować modele na podstawie dużych ilości danych [5]. W zależności od publikacji proponowane są różne rodzaje tych sieci, są to między innymi: MLP (ang. *Multi-Layer Perceptron*) [6], RNN (ang. *Recurrent Neural Network*) [7], LSTM (ang. *Long Short-Term Memory*) [8], CNN (ang. *Convolutional Neural Network*) [9], systemy hybrydowe [10]. Stosuje się także metodę wektorów nośnych SVM (ang. *Support Vector Machine*) zastosowaną do regresji [11]. Ostatnio opublikowane badania wykazały, że architektura oparta na dekodzie i enkoderze, zastosowana w modelu sieci typu transformer, nadaje się do modelowania złożonych relacji sekwencyjnych oraz do wieloetapowego prognozowania szeregów czasowych. Transformer jest potężnym narzędziem do wieloetapowego prognozowania szeregów czasowych, co jest trudnym zadaniem do realizacji w przypadku tradycyjnych modeli statystycznych, takich jak ARIMA i GARCH [12].

Na podstawie literatury można zauważyć, iż dominują publikacje dla systemów fotowoltaicznych położonych w takich położeniach geograficznych, gdzie ich zastosowanie jest wyraźnie uzasadnione ekonomicznie. Wynika to z warunków klimatycznych występujących w danych miejscach. W krajach takich jak np. Australia, Indie, Arabia Saudyjska, czy Włochy, liczba dni słonecznych w roku jest wyraźnie większa niż np. w Polsce, co bezpośrednio jest powiązane z irradancją, a finalnie z wytwarzaniem energii elektrycznej. Ponadto w tych państwach możliwe jest zauważenie wyraźniejszej sezonowości i cykliczności pogody w kolejnych latach.

Artykuł skupia się na prognozowaniu generowanej energii elektrycznej w małych systemach fotowoltaicznych metodą regresji. Wykorzystano kilka znanych algorytmów uczenia maszynowego z darmowej biblioteki Scikit-learn, dostępnej dla użytkowników języka programowania Python. Do tego celu zostały wykorzystane zgromadzone dane pomiarowe z lat 2015–2022 z elektrowni hybrydowej położonej na terenie Politechniki Białostockiej (<http://elektrownia.pb.edu.pl/>).

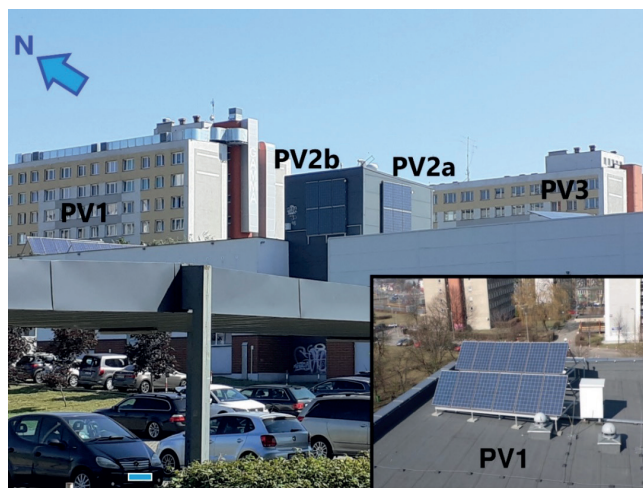
2. Materiały i metody

Elektrownia hybrydowa (rys. 1) usytuowana jest na kampusie Politechniki Białostockiej w Białymstoku. Elementami składowymi elektrowni są dwie turbiny wiatrowe oraz cztery zespoły paneli fotowoltaicznych [13]. Wśród paneli możemy wyróżnić:

- panele ustawione optymalnie, o mocy DC 3,0 kWp (PV1),
- panele na trackerze, o mocy DC 3,0 kWp (PV3),
- panele na elewacji południowo-wschodniej, o mocy DC 1,5 kWp (PV2a),
- panele na elewacji południowo-zachodniej, o mocy DC 1,5 kWp (PV2b).

Prognozy zostały opracowane na podstawie danych pomiarowych historycznych z lat 2015–2021, w okresie od 1 kwietnia do 30 września każdego roku. Dane zgromadzone w roku 2022 zostały wykorzystane do testowania opracowanych modeli prognostycznych. Prognozy zostały wykonane dla różnych wariantów parametrów modeli, najlepsze rezultaty zostały zaprezentowane graficznie. Jakość prognoz została oszacowana na podstawie współczynnika determinacji oraz wybranych mierników (błędów).

Zgodnie z dostępnymi danymi z elektrowni fotowoltaicznej, założono, że prognozy będą wykonywane dla zespołu paneli ustawionych optymalnie, pod kątem nachylenia 38° (PV1), i mają one określać dzienną energię wyrażoną w kWh i będą one dokonane na podstawie: temperatury maksymalnej oraz minimalnej danego dnia, ciśnienia atmosferycznego, średniej



Rys. 1. Ogólny widok elektrowni hybrydowej: PV1 – panele ustawione optymalnie, PV2a oraz PV2b – panele na elewacjach, PV3 – fotowoltaiczny system nadążny

Fig. 1. General view of the hybrid power plant: PV1 – fixed-tilt with panels optimally directed, PV2a and PV2b – panels on the facades, PV3 – solar tracking system

prędkości wiatru, dnia roku (okresowość cyklu słonecznego w ciągu roku). Wybór czynników wykorzystywanych do prognozowania generowanej energii podyktowany był faktem, iż są to powszechnie dostępne informacje w prognozach pogodowych krótko- i średnioterminowych.

Pierwszym etapem jakiegokolwiek analizy jest odpowiednie przygotowanie danych wejściowych. Dane pomiarowe z elektrowni fotowoltaicznej zostały pobrane z serwera elektrownia.pb.edu.pl w postaci plików CSV. Każdy plik był oddzielny dla danego roku, zaś dane w każdym pliku zawierały pomiary w okresie od 1 kwietnia do 30 września. System pomiarowy rejestruje wartości z okresem około 10 sekund, co przekłada się na ponad 2,1 miliona rekordów dla każdego roku w analizowanym okresie [14]. Konieczne zatem było zredukowanie danych przez określenie temperatury minimalnej oraz maksymalnej, maksymalnego ciśnienia atmosferycznego, średniej prędkości wiatru jak i obliczenie ilości pozyskanej energii dla poszczególnych dni z tego okresu. W tym celu opracowany został skrypt w języku Python. Jego zadaniem było wczytanie wybranego pliku CSV oraz wygenerowanie pliku zawierającego pojedynczy rekord danych z danego dnia zawierający:

- datę (przeliczoną na numer dnia z analizowanego okresu, 1, 2, 3, ...),
- energię elektryczną uzyskaną przez zespół PV1,
- temperaturę maksymalną,
- temperaturę minimalną,
- ciśnienie atmosferyczne,
- średnią prędkość wiatru.

W rezultacie z każdego roku uzyskano 183 pozycje – każdą odpowiadającą jednemu dniowi z analizowanego okresu od 1 kwietnia do 30 września, w formacie jak przedstawiono w tab. 1 zgodnie z dotychczas określonymi założeniami, napisano programy odpowiadające za przygotowanie modelu i prognozy, oparte na bazie sieci neuronowych MLP oraz metody wektorów nośnych SVM. Założono tworzenie prognoz długoterminowych. Wykorzystano darmową bibliotekę algorytmów z dziedziny uczenia maszynowego Scikit-learn, napisaną w języku Python [15–17].

W tabeli 2 przedstawiono korelację między poszczególnymi wielkościami a energią wyjściową, otrzymane na podstawie zarejestrowanych danych z okresu lat 2015–2021. Można zauważyć znaczącą korelację energii generowanej z temperaturą maksymalną i ciśnieniem atmosferycznym.

Tab. 1. Fragment danych użytych do opracowania modeli

Tab. 1. A fragment of the data used to develop the models

Numer dnia	Energia	Temperatura maksymalna	Temperatura minimalna	Ciśnienie	Prędkość wiatru
	kWh	°C	°C	hPa	m/s
95	17,806	31,3	14,7	1008,7	1,3
96	17,231	32,4	16,5	1000,9	1,2
97	12,312	33,5	16,8	997,6	2,6
98	17,233	27,4	13,8	997,9	2,9
99	11,907	27,6	16,6	989,7	2,7

Tab. 2. Współczynniki korelacji Pearson'a między poszczególnymi wielkościami wejściowymi a generowaną energią

Tab. 2. Pearson correlations between the input features and produced energy

Wartości wejściowe	Współczynnik korelacji
Numer dnia	-0,11
Temperatura maksymalna	0,40
Temperatura minimalna	0,06
Ciśnienie atmosferyczne	0,32
Średnia prędkość wiatru	-0,08

3. Mierniki jakości prognoz

Prognozy są to estymacje oparte na obserwacjach lub zebranych danych, poparte stosownymi wyliczeniami. Przybliżenie charakteryzuje się tym, iż w pewnym (większym lub mniejszym) stopniu jego wynik będzie odbiegał od rzeczywistości. Po zaistnieniu prognozowanego zdarzenia, możliwe jest skonfrontowanie wartości rzeczywistych z przewidywanymi, a następnie określenie jakości prognozy. W tym celu istnieją różne mierniki służące tym wyliczeniom. Najbardziej stosowne do omawianego zagadnienia to [18–20]:

- MAE – średni błąd bezwzględny,
- MSE – błąd średniokwadratowy,
- $RMSE$ – pierwiastek błędu średniokwadratowego,
- R^2 – współczynnik determinacji.

Średni błąd bezwzględny wyraża średnią wartość błędu bezwzględnego w danym zbiorze próbek. Błąd bezwzględny wyraża różnicę między wartością estymowaną, a rzeczywistą wyrażoną w procentach. MAE można opisać zależnością:

$$MAE(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (1)$$

gdzie: N – liczba próbek, y_i – i -ta wartość rzeczywista, \hat{y}_i – i -ta wartość estymowana.

Błąd średniokwadratowy wyrażany jest przez kwadrat różnicy między wartością rzeczywistą, a wartością oczekiwaną (prognozowaną). Błąd średniokwadratowy przyjmuje zawsze wartości dodatnie. Charakteryzuje się tym, iż duże błędy mają

większy wpływ (wagę) na rezultat. MSE można opisać zależnością

$$MSE(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2. \quad (2)$$

$RMSE$ jest to pierwiastek z wartości MSE . Dzięki pierwiastkowi wartości przedstawiane przez błąd MSE są bardziej intuicyjne w interpretacji, ponieważ jego wartość jest wyrażona w jednostce badanej wielkości. Ponadto przy dużych wartościach (ze względu na podniesienie do kwadratu) w takiej postaci jest bardziej reprezentatywny. $RMSE$ można wyrazić za pomocą wyrażenia:

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}. \quad (3)$$

Współczynnik determinacji R^2 jest miarą, która wskazuje na dopasowanie modelu do danych uczących. Mierzy on zmienność (wariancję) w wektorze docelowym (zawierającym wyniki prognoz). Współczynnik R^2 przyjmuje wartości z zakresu od 0 do 1, gdzie wartości zbliżone do jedności mówią o lepszym dopasowaniu. Współczynnik wyraża się zależnością:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (4)$$

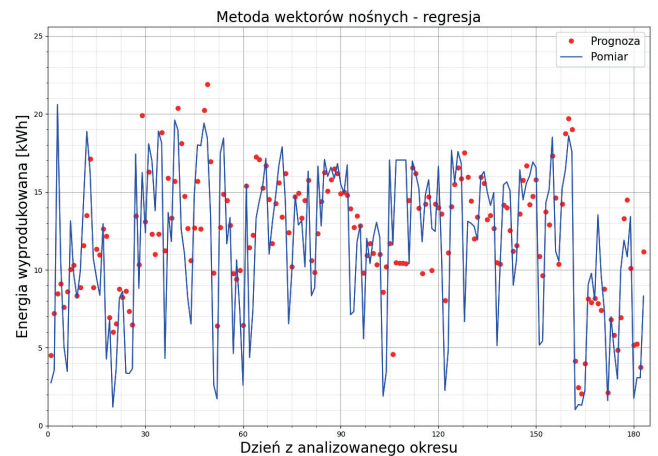
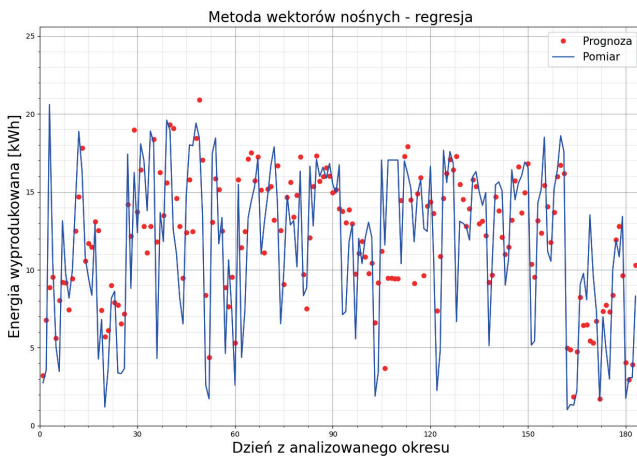
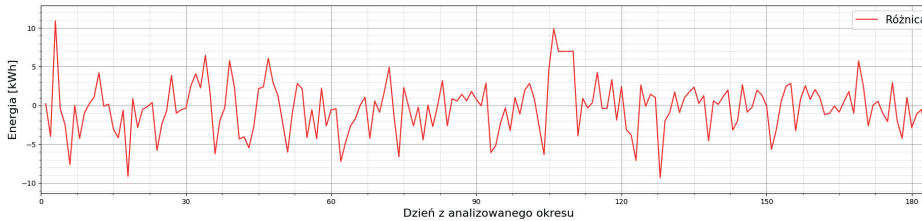
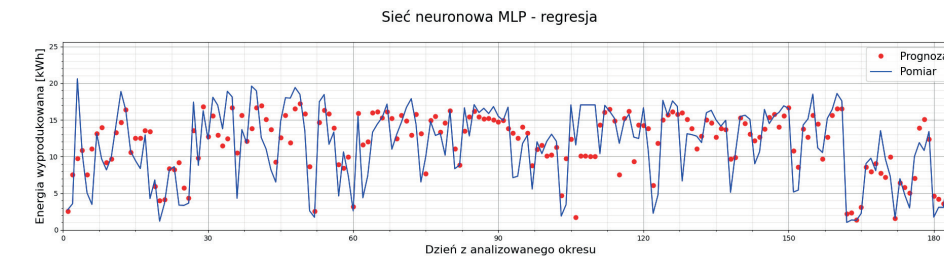
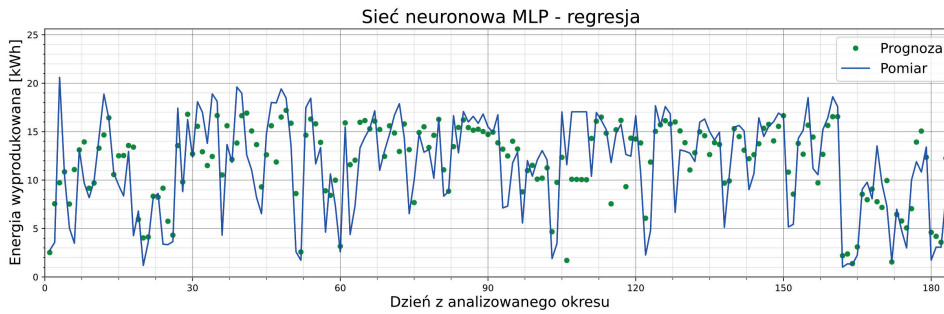
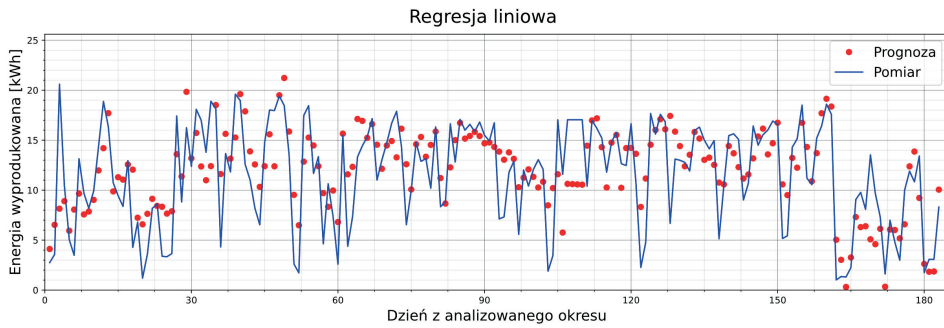
gdzie: \bar{y} – wartość średnia wektora wartości rzeczywistych.

4. Wyniki prognoz

Przeprowadzone liczne próby skutkowały opracowaniem wielu wariantów prognoz przedstawionych w tabeli 3, dla których zestawiono najlepsze wyniki przy zastosowaniu poszczególnych modeli prognostycznych dla zespołu paneli ustawionego w pozycji optymalnej (PV1).

Prognozy zostały wykonane dla różnych parametrów modelu. Zmieniany był stopień wielomianu m w zakresie od 1 do 5. Najlepsze rezultaty uzyskano przy $m = 2$, czyli dla przypadku, kiedy zależność między zmienną objaśniającą a objaśnianą w modelu regresji jest nieliniowa (funkcja kwadratowa).

W przypadku opracowania prognoz z wykorzystaniem sieci neuronowych, niemożliwe było zastosowanie procedury jak w przypadku regresji w celu doboru najlepszych parametrów



Rys. 4. Regresja metodą wektorów nośnych SVR z liniową funkcją jądrową w modelu – uzysk i prognoza energii
 Fig. 4. Support vector regression SVR used with the linear kernel function in the model – energy yield and forecast

Rys. 5. Regresja metodą wektorów nośnych SVR z nieliniową funkcją jądrową w modelu (wielomian 35 stopnia) – uzysk i prognoza energii
 Fig. 5. Support vector regression SVR with the polynomial of degree 35 as a kernel function in the model – energy yield and forecast

Rys. 2. Energia uzyskana w każdym dniu analizowanego okresu oraz prognozy dla elektrowni PV1 z wykorzystaniem modelu regresji liniowej dla $m = 1$ oraz z wykorzystaniem modelu MLPRegressor
 Fig. 2. The daily energy yield in the analyzed period and the forecast for the PV1 power plant using the linear regression model for $m = 1$ and the MLPRegressor model

Rys. 3. Energia wygenerowana w każdym dniu analizowanego okresu wraz z prognozą dla elektrowni PV1 z wykorzystaniem modelu MLPRegressor; różnica między prognozą i pomiarem
 Fig. 3. The daily energy yield in the analyzed period and the forecast for the PV1 power plant using the MLPRegressor model; the difference between forecast and measurement

Tab. 3. Zestawienie najlepszych prognoz dla poszczególnych metod

Tab. 3. A summary of the best forecasts for individual methods

Model	MAE [kWh]	MAE [%]	MSE [kWh ²]	RMSE [kWh]	R ²
Linear regression ¹	2,71	44,22	12,13	3,48	0,538
Polynomial regression ²	2,65	42,05	11,96	3,46	0,544
Polynomial regression ³	2,66	38,51	12,24	3,50	0,533
XGBRegressor	2,90	53,08	13,19	3,63	0,497
RandomForestRegressor	2,85	48,32	13,11	3,62	0,500
CatBoostRegressor ⁴	2,64	41,38	11,79	3,43	0,551
MLPRegressor ⁵	2,43	37,48	3,22	1,79	0,605
Kernel Ridge Regression	2,80	48,79	12,55	3,54	0,521
Support Vector Regression ⁶	2,64	39,94	11,77	3,43	0,551
Support Vector Regression ⁷	2,62	41,11	11,59	3,40	0,558

Parametry i hiperparametry modeli: 1 – $m = 1$; 2 – $m = 2$; 3 – $m = 3$; 4 – iterations = 1000, loss function='RMSE', l2_leaf_reg=30, depth = 6; 5 – activation = 'tanh', solver = 'lbfgs', hidden_layers_size = 5; 6 – kernel='linear', epsilon = 0,1; 7 – kernel = 'poly', degree = 35, epsilon = 0,1.

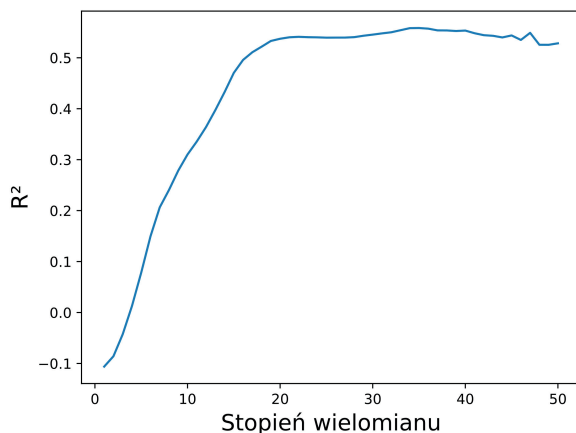
Rys. 6. Zależność współczynnika determinacji R^2 od stopnia wielomianu (regresja SVR)

Fig. 6. Dependence of the coefficient of determination R^2 on the degree of the polynomial (SVR regression)

do wyliczenia modelu. Wynika to ze stochastycznego charakteru procesu nauczania sieci neuronowej. Wprowadzając ten sam zestaw danych uczących, za każdym razem uzyskany zostanie inny model. W związku z tym wybór najlepszych wyników został przeprowadzony metodą eksperymentalną. Przeprowadzone zostały liczne próby z różnymi hiperparametrami modeli. Po przeprowadzeniu tych prób zostały przedstawione najlepsze uzyskane rezultaty wraz z dobranymi parametrami modelu.

Na podstawie przedstawionego zestawienia możliwe jest wytypowanie najlepszej jakościowo prognozy. Stosując identyczne kryteria jak przy wyborze najlepszego wariantu (maksymalizacja wartości R^2) można jednoznacznie wskazać, iż najlepiej sprawdziła się sztuczna sieć neuronowa MLP

w wariancie regresyjnym (MLPRegressor). Jak można zauważyć wyższy współczynnik determinacji R^2 niekoniecznie koreluje się wartością MAE . Niższa jego wartość nie musi wiązać się z lepszą korelacją między prognozą, a stanem rzeczywistym (pomiarami). Patrząc na wartości błędu $RMSE$ w rozpatrywanym zestawieniu, można już taką korelację dostrzec, im mniejszy błąd średniokwadratowy MSE , tym większa wartość współczynnika R^2 .

Poszerzając analizę o wizualną reprezentację prognoz subiektywnie można stwierdzić iż każda z zaprezentowanych prognoz (rys. 2–5) pokrywa się z pomiarami. Niebieską linią zostały przedstawione dane historyczne (rzeczywiste), zaś kolorem czerwonym wartości prognozowane. Wartości osi X to numery dni z analizowanego okresu, tj. dzień o numerze 1 odpowiada dacie 1 kwietnia, a 183 dacie 30 września 2022 r. Możliwe jest dostrzeżenie trendów, natomiast nie zaobserwowano cykliczności.

Rysunki 4 i 5 przedstawiają wyniki najlepszych prognoz metodą regresji SVR. Rysunek 6 przedstawia współczynnik determinacji R^2 w zależności od stopnia wielomianu.

5. Podsumowanie i wnioski końcowe

Biorąc pod uwagę wszystkie zgromadzone dane, zarówno w postaci metryk (tabela 3) jak i wykresów reprezentujących rezultaty, trudno jest bazować na jednym parametrze metryk jako uniwersalnym wskaźniku jakości prognozy. W przypadku sieci neuronowych MLP, mimo zbliżonych wartości współczynnika R^2 do prognoz uzyskanych za pomocą klasycznej regresji liniowej, widoczna jest duża rozbieżność wartości $RMSE$ między tymi metodami. Porównując najlepsze rezultaty między klasyczną regresją, a regresją z wykorzystaniem sieci MLP, trudno jest wskazać realny wpływ na jakość prognoz. Rozszerzając analizę o pozostałe przeprowadzone próby, różnice w wynikach są niewielkie. Mimo

wszystko, sieć MLP umożliwi poprawę wyniku w sytuacji gdzie w metodzie regresyjnej niemożliwe jest już poprawienie wyniku z tym samym zestawem danych, co wynika między innymi ze stochastycznego charakteru procesu uczenia sieci MLP. Błąd *RMSE* jest wyraźnie mniejszy dla modelu regresji MLP. Dla badanych modeli osiągnięto R^2 na poziomie 0,6, a sieć neuronowa typu MLP pozwoliła uzyskać najmniejszą wartość *RMSE*.

Większe możliwości prognozowania szeregów czasowych stwarzają techniki głębokiego uczenia. Oprócz tradycyjnych, jednoblokowych, modeli ML do prognozowania generowanej energii fotowoltaicznej, obecnie używane są złożone modele oparte na sieciach typu LSTM [21] lub sieciach typu transformer z mechanizmem uwagi (ang. *transformer models with attention*) pozwalającym skupić się na części podzbioru danych [22]. Struktura sieci neuronowych typu transformer oparta jest na architekturze enkoder-dekoder. Takie modele można tworzyć w języku Python z pomocą bibliotek Keras lub TensorFlow.

Na jakość opracowanych modeli duży wpływ miała ilość danych historycznych. Modele powstały na podstawie danych z siedmiu lat. Modele bazowały na podstawowych danych klimatycznych, bez uwzględnienia irradiancji jako wielkości wejściowej. Posiadanie danych z dłuższego okresu niewątpliwie pozytywnie by wpłynęło na uzyskane rezultaty. Ponadto jakość prognoz była określana względem danych pomiarowych z jednego okresu w 2022 r. W związku z tym wszelkie odstępstwa od typowych warunków dla danych okresów (zauważalnie mała powtarzalność pogody w Polsce w poszczególnych dniach czy miesiącach w ostatnich latach), negatywnie wpłynęły na wartości metryk.

Bibliografia

- Sabat M., Baczyński D., Szafranek K. *Analiza szeregów czasowych produkcji energii ze źródeł odnawialnych pod kątem niezależności energetycznej wybranego obszaru*. „Przegląd Elektrotechniczny”, Vol. 93, No. 9, 2017, 11–15, DOI: 10.15199/48.2017.09.02.
- Nawrowski R., Filipiak M., Węgrzyn W., Kurz D., *Analiza możliwości zarządzania i rozdziału energii elektrycznej, wyprodukowanej w prosumenckiej instalacji fotowoltaicznej, w budynku z automatyką budynkową*. „Przegląd Elektrotechniczny”, Vol. 98, No. 11, 2022, 253–258, DOI: 10.15199/48.2022.11.53.
- ElNozahy M.S., Salama M.M.A. *Technical impacts of grid-connected photovoltaic systems on electrical networks—a review*. “Journal of Renewable and Sustainable Energy”, Vol. 5, No. 3, 2013, DOI: 10.1063/1.4808264.
- Moreira M.O., Kaizer B.M., Ohishi T., Bonatto B.D., Zambroni de Souza A.C., Balestrassi P.P., *Multivariate Strategy Using Artificial Neural Networks for Seasonal Photovoltaic Generation Forecasting*. “Energies”, Vol. 16, No. 1, 2023, DOI: 10.3390/en16010369.
- Soori M., Arezoo B., Dastres R., *Artificial intelligence, machine learning and deep learning in advanced robotics, a review*. “Cognitive Robotics”, Vol. 3, 2023, 54–70, DOI: 10.1016/j.cogr.2023.04.001.
- Markovics D., Mayer M.J., *Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction*. “Renewable and Sustainable Energy Reviews”, Vol. 161, 2022, DOI: 10.1016/j.rser.2022.112364.
- Kusuma V., Privadi A., Setya Budi A. L., Budiharto Putri V. L., *Photovoltaic Power Forecasting Using Recurrent Neural Network Based On Bayesian Regularization Algorithm*. [in:] IEEE International Conference in Power Engineering Application (ICPEA), Malaysia, 2021, 109–114, DOI: 10.1109/ICPEA51500.2021.9417833.
- Jailani N.L.M., Dhanasegaran J.K., Alkaws G., Alkah-tani A.A., Phing C.C., Baashar Y., Capretz L.F., Al-Shetwi A.Q., Tiong S.K., *Investigating the Power of LSTM-Based Models in Solar Energy Forecasting*. “Processes”, Vol. 11, No. 5, 2023, DOI: 10.3390/pr11051382.
- Thangavel A., Govindaraj V., *Forecasting Energy Demand Using Conditional Random Field and Convolution Neural Network*. “Elektronika ir Elektrotechnika”, Vol. 28, No. 5, 2022, 12–22, DOI: 10.5755/j02.eie.30740.
- Mukhtar M., Oluwasanmi A., Yimen N., Qinxiu Z., Ukwu-oma C.C., Ezurike B., Bamisile O., *Development and Comparison of Two Novel Hybrid Neural Network Models for Hourly Solar Radiation Prediction*. “Applied Sciences”, Vol. 12, No. 3, 2022, DOI: 10.3390/app12031435.
- Preda S., Oprea S.-V., Băra, A., Belciu A., *PV Forecasting Using Support Vector Machine Learning in a Big Data Analytics Context*. “Symmetry”, Vol. 10, No. 12, 2018, DOI: 10.3390/sym10120748.
- Tehrani S., Juan J., Caro E., *Electricity Spot Price Modeling and Forecasting in European Markets*. “Energies”, Vol. 15, No. 16, 2022, DOI: 10.3390/en15165980.
- Trzasko W., *Analiza wydajności dwuosioowego solarnego układu nadążnego*. „Pomiary Automatyka Robotyka”, Vol. 22, No. 1, 2018, 11–17, DOI: 10.14313/PAR_227/11.
- Idzkowski A., Karasowska K., Walendziuk W., *Temperature Analysis of the Stand-Alone and Building Integrated Photovoltaic Systems Based on Simulation and Measurement Data*. “Energies”, Vol. 13, No. 16, 2020, DOI: 10.3390/en13164274.
- Machine Learning in Python*. <https://scikit-learn.org/stable/index.html>.
- Geron A., *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media Inc., USA, 2019.
- Müller A.C., Guido S., *Introduction to Machine Learning with Python*. O'Reilly Media Inc., USA, 2016.
- Li G., Wei X., Yang H., *Decomposition integration and error correction method for photovoltaic power forecasting*. “Measurement”, Vol. 208, 2023, DOI: 10.1016/j.measurement.2023.112462.
- Sumorek M., Idzkowski A., *Time Series Forecasting for Energy Production in Stand-Alone and Tracking Photovoltaic Systems Based on Historical Measurement Data*. “Energies”, Vol. 16, No. 17, 2023, DOI: 10.3390/en16176367.
- Radicioni M., Lucaferri V., De Lia F., Laudani A., Lo Presti R., Lozito G.M., Riganti Fulginei F., Schioppo R., Tucci M., *Power Forecasting of a Photovoltaic Plant Located in ENEA Casaccia Research Center*. “Energies”, Vol. 14, No. 3, 2021, DOI: 10.3390/en14030707.
- Wang K., Qi X., Liu H., *Photovoltaic power forecasting based LSTM-Convolutional Network*. “Energy”, Vol. 189, 2019, DOI: 10.1016/j.energy.2019.116225.
- Al-Ali EM., Hajji Y., Said Y., Hleili M., Alanzi AM., Laatar AH., Atri M., *Solar Energy Production Forecasting Based on a Hybrid CNN-LSTM-Transformer Model*. “Mathematics”, Vol. 11, No. 3, 2023, DOI: 10.3390/math11030676.

The Use of Machine Learning Algorithms to Forecast Energy Production in a Small PV System

Abstract: The article presents data analysis for predicting energy production in photovoltaic (PV) power plant systems. The purpose of long-term forecasts is to determine the effectiveness of preventive actions and manage the power system effectively. Climate variables affecting the production of electricity in photovoltaic systems were analyzed. Forecasting methods using machine learning techniques such as Multi-Layer Perceptron (MLP) neural networks and Support Vector Machine (SVM) were compared. In addition, metrics were selected to determine the quality of forecasts. Determining the quality of forecasts was based on the actual varying conditions, not on the weather forecast data. The way of data preparation to create forecasting models were presented and the models with the best metrics were selected. For this purpose, the Scikit-learn library was used to create scripts in Python. The best results were obtained for regression models: MLPRegressor, CatBoostRegressor and Support Vector Regression. Actual measurement data from a system of optimally-positioned panels with a power of 3.0 kWp were used. For the MLPRegressor model, the highest coefficient of determination 0.605 was achieved with the smallest root-mean-square error of 1.79 KWh.

Keywords: photovoltaic power station; machine learning; energy forecasting; regression models; neural networks; Multi-Layer Perceptron; Support Vector Regression; Kernel Ridge Regression

dr hab. inż. Adam Idźkowski

a.idzkowski@pb.edu.pl
ORCID: 0000-0002-1671-8979

Absolwent Wydziału Elektrycznego Politechniki Białostockiej 2001, doktorat 2013, habilitacja 2019. Obecnie profesor Politechniki Białostockiej. Autor i współautor ponad 90 publikacji na temat technik i aplikacji pomiarowych. Zainteresowania naukowe: modelowanie statystyczne, sensory i systemy pomiarowe, uczenie maszynowe.



mgr inż. Mateusz Sumorek

mateusz.sumorek@pb.edu.pl
ORCID: 0009-0003-2440-9556

Ukończył studia magisterskie na Wydziale Elektrycznym Politechniki Białostockiej (kierunek Elektronika i Telekomunikacja, specjalność Aparatura Elektroniczna i Telekomunikacja). Obecnie jest uczestnikiem Szkoły Doktorskiej Politechniki Białostockiej w dyscyplinie automatyka, elektronika, elektrotechnika i technologie kosmiczne. Przedmiotem jego badań jest lokalizacja w pomieszczeniach zamkniętych z wykorzystaniem technik radiowych.

