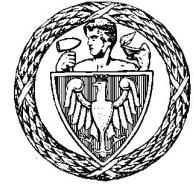


Politechnika Warszawska

WYDZIAŁ GEODEZJI I KARTOGRAFII



Instytut Automatyki i Robotyki

# Praca dyplomowa magisterska

na kierunku Automatyka, Robotyka i Informatyka Przemysłowa

Nazwa pracy

inż. Ivan Kaliankovich

numer albumu 335662

promotor

prof. dr inż. Paweł Wnuk

WARSZAWA 2025



## **Nazwa pracy**

### **Streszczenie**

To jest streszczenie. To jest trochę za krótkie, jako że powinno zająć całą stronę.

  Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

  Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

  Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

  Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

**Słowa kluczowe:** A, B, C



## **Thesis name**

### **Abstract**

This is abstract. This one is a little too short as it should occupy the whole page.

  Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

  Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

  Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

  Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

**Keywords:** X, Y, Z



# Spis treści

<b>1 Wstęp</b>	<b>9</b>
<b>2 Przegląd literatury</b>	<b>12</b>
2.1 Zmienne wejściowe . . . . .	12
2.2 Modele prognozowania . . . . .	14
<b>3 Dane i zmienne</b>	<b>17</b>
3.1 Zmienna zależna . . . . .	17
3.2 Zbiór zmiennych niezależnych . . . . .	19
3.2.1 Dane pogodowe . . . . .	19
3.2.2 Produkcja energii z wybranych źródeł . . . . .	25
3.2.3 Handel energią z państwami sąsiednimi . . . . .	26
3.2.4 Ceny paliw kopalnych i emisji CO <sub>2</sub> . . . . .	27
3.2.5 Straty mocy w systemie elektroenergetycznym . . . . .	28
3.2.6 Zapotrzebowanie na energię i wolumen handlu . . . . .	29
3.2.7 Inne zmienne . . . . .	30
<b>4 Eksploracja danych</b>	<b>34</b>
4.1 Wstępna obróbka danych . . . . .	34
4.2 Analiza korelacji . . . . .	35
4.3 Podział danych . . . . .	39
4.4 Przygotowanie danych . . . . .	40
<b>5 Metodologia</b>	<b>42</b>
5.1 Ocena jakości prognoz . . . . .	42
5.1.1 Mean Absolute Error (MAE) . . . . .	42
5.1.2 Root Mean Squared Error (RMSE) . . . . .	43
5.1.3 Mean Absolute Percentage Error (MAPE) . . . . .	44
5.1.4 Symmetric Mean Absolute Percentage Error (sMAPE) . . . . .	44
5.1.5 Współczynnik determinacji . . . . .	45
5.2 Wybrane metody weryfikacji zbioru danych . . . . .	46
5.2.1 Regresja liniowa . . . . .	46

5.2.2	Regresja grzbietowa (Ridge) . . . . .	47
5.2.3	Prophet . . . . .	48
5.2.4	Wielowarstwowy perceptron (MLP) . . . . .	51
<b>6</b>	<b>Analiza danych</b>	<b>53</b>
6.1	Okres stabilny . . . . .	53
6.2	Okres niestabilny . . . . .	56
<b>7</b>	<b>Podsumowanie wyników i wnioski</b>	<b>57</b>
<b>Bibliografia</b>		<b>59</b>
<b>Wykaz skrótów i symboli</b>		<b>61</b>
<b>Spis rysunków</b>		<b>62</b>
<b>Spis tabel</b>		<b>63</b>
<b>Spis załączników</b>		<b>64</b>

# Rozdział 1

## Wstęp

### Cel pracy

Większość dotychczasowych badań w literaturze naukowej dotycząca prognozowania cen energii elektrycznej, skupia się głównie na rynkach amerykańskich, europejskich oraz azjatyckich, pomijając specyfikę polskiego rynku energii. W związku z tym, celem niniejszej pracy jest opracowanie obszernej bazy danych z polskiego rynku energii, obejmującej zarówno okres umiarkowanej zmienności cen, jak i okres dużej zmienności cen i brak stabilności rynkowej. Baza danych zostanie przygotowana w celu uwzględnienia kluczowych czynników wpływających na ceny energii elektrycznej, charakterystycznych dla polskiego rynku. Następnie, w celu przetestowania użyteczności opracowanej bazy danych, przeprowadzono analizę z wykorzystaniem wybranych modeli uczenia maszynowego: regresji liniowej, regresji grzbietowej, modelu Prophet oraz wielowarstwowej sieci neuronowej.

### Wprowadzenie

Rynek energii elektrycznej to jeden z filarów współczesnej gospodarki, a jego sprawne funkcjonowanie ma kluczowe znaczenie dla stabilności systemów elektroenergetycznych, przedsiębiorstw i codziennego życia konsumentów. W centrum rynku znajduje się Rynek Dnia Następnego (RDN), który działa jako platforma handlu energią na dzień przed jej dostarczeniem. RDN jest miejscem, gdzie producenci energii od elektrowni węglowych po farmy wiatrowe spotykają się z odbiorcami, takimi jak dostawcy energii dla domów mieszkalnych czy duże zakłady przemysłowe, w celu ustalenia ceny i volumenów energii na każdą godzinę kolejnego dnia. Mechanizm działania RDN opiera się na systemie aukcyjnym: uczestnicy składają oferty kupna i sprzedaży, określając, ile energii są w stanie dostarczyć lub kupić oraz po jakiej cenie. System następnie dopasowuje te oferty, ustalając cenę równowagi, która obowiązuje dla danej godziny.

Taki model handlu pozwala na elastyczne reagowanie na zmieniające się warunki - zarówno po stronie podaży, jak i popytu. Na przykład, jeśli prognozy wskazują na silny wiatr, producenci energii wiatrowej mogą zwiększyć podaż, co może obniżyć ceny; z kolei fala upałów może zwiększyć zapotrzebowanie na energię do klimatyzacji, podnosząc ceny. RDN działa w wielu krajach, choć jego

specyfika różni się w zależności od regionu. W Europie, w tym w Polsce, rynek ten jest częścią szerszego systemu integracji rynków energii, który ma na celu zapewnienie płynności i efektywności handlu. W Stanach Zjednoczonych RDN funkcjonuje w ramach regionalnych rynków, takich jak PJM Interconnection czy California ISO (CAISO), gdzie handel energią jest dodatkowo skomplikowany przez różnice regulacyjne między stanami. Niezależnie od regionu, RDN jest kluczowym narzędziem w zarządzaniu systemem elektroenergetycznym, umożliwiając szybkie dostosowanie podaży do popytu w krótkim horyzoncie czasowym.

Jednak handel na RDN to nie tylko szansa na zysk dla wszystkich biorących udział, ale i ogromne ryzyko finansowe, które wynika z nieprzewidywalności cen energii. Na rynku amerykańskim, gdzie mechanizm licytacji między kupującymi (buyers) a sprzedającymi (sellers) jest szczególnie rozwinięty, ryzyko to jest wyjątkowo widoczne. Uczestnicy rynku muszą składać oferty w czasie rzeczywistym, próbując przewidzieć, jak zachowa się cena w danej godzinie. Jeśli producent energii zaoferuje zbyt wysoką cenę, jego energia może nie zostać kupiona, co oznacza utratę przychodów; z kolei kupujący, który zaoferuje zbyt niską cenę, może zostać zmuszony do zakupu energii po znacznie wyższej cenie rynkowej. Przykładem skali tego ryzyka jest kryzys w Teksasie [1] w lutym 2021 roku, znany jako Texas winter storm. W wyniku ekstremalnych mrozów i awarii systemu elektroenergetycznego ceny na rynku ERCOT (Electric Reliability Council of Texas) wzrosły do 9000 USD/MWh - poziomu, który dla wielu uczestników rynku oznaczał straty liczone w dziesiątkach milionów dolarów. W takich warunkach dokładna predykcja cen staje się nie tylko narzędziem do optymalizacji handlu, ale wręcz koniecznością, by uniknąć katastrofalnych strat. Na rynkach krajów rozwiniętych, gdzie Odnawialne Źródła Energii (OZE) odgrywają coraz większą rolę, ceny mogą spadać do wartości ujemnych w godzinach nadprodukcji np. z energii słonecznej, by kilka godzin później gwałtownie wzrosnąć, gdy zapotrzebowanie przewyższa podaż. Ta zmienność sprawia, że handel na RDN przypomina grę o wysoką stawkę, w której każdy ruch musi być dokładnie przemyślany.

Rynek bilansujący stanowi kolejny istotny element systemu elektroenergetycznego, uzupełniając funkcjonowanie RDN. Działa on w czasie rzeczywistym, umożliwiając operatorom systemu elektroenergetycznego (w Polsce jest to Polskie Sieci Elektroenergetyczne, PSE) równoważenie podaży i popytu w sytuacjach, gdy rzeczywiste zużycie energii odbiega od prognoz ustalanych na RDN. Na rynku bilansującym uczestnicy mogą zgłaszać oferty na dostawy energii w bardzo krótkim horyzoncie czasowym, nawet w ciągu kilkunastu minut, co pozwala na szybkie reagowanie na nagłe zmiany, takie jak awarie elektrowni, nieoczekiwane skoki zapotrzebowania czy zmienność produkcji z OZE. Ceny na rynku bilansującym są często bardziej zmienne niż na RDN, co dodatkowo zwiększa ryzyko finansowe dla uczestników, ale jednocześnie podkreśla znaczenie precyzyjnych prognoz cen, które mogą pomóc w lepszym zarządzaniu tymi krótkoterminowymi wahaniami.

W niniejszej pracy magisterskiej skupiam się na analizie RDN w Polsce, który choć różni się od rynku amerykańskiego pod względem skali i regulacji, również zmaga się z podobnymi wyzwaniami. W Europie, w tym w Polsce, RDN jest częścią zintegrowanego systemu handlu energią, który opiera się na współpracy między krajami i dąży do harmonizacji rynków energii w ramach Unii Europejskiej. Mechanizm ustalania cen na RDN w Polsce opiera się na zasadzie jednolitej ceny, gdzie cena

---

rozliczeniowa, jest wyznaczana na podstawie równowagi popytu i podaży dla każdej godziny. Oznacza to, że wszyscy uczestnicy, których oferty zostały zaakceptowane, rozliczają się po tej samej cenie, co zapewnia przejrzystość i efektywność handlu. W Polsce RDN jest prowadzony przez Towarową Giełdę Energii (TGE), która od 2000 roku pełni rolę kluczowej platformy handlu energią. TGE organizuje aukcje na RDN, umożliwiając uczestnikom składanie ofert na każdą godzinę kolejnego dnia. W latach 2016-2024, które obejmują analizowany w pracy okres, ceny na RDN w Polsce wahają się od 200 PLN/MWh aż do ponad 3000 PLN/MWh, co odzwierciedla zarówno lokalne uwarunkowania (np. zależność od węgla, rosnąca rola OZE), jak i globalne trendy. Ważną rolą w takiej zmienności odegrały także niespodziewane czynniki zewnętrzne, zwane czarnymi łącznikami. Z przykładów takich czynników można wymienić wybuch pandemii i wprowadzenie w związku z tym restrykcji na pracę oraz kryzys energetyczny w 2022 roku wywołany konfliktem zbrojnym na Ukrainie. Wszystkie te czynniki sprawiają, że prognozowanie cen energii elektrycznej na RDN w Polsce staje się niezwykle złożonym zadaniem, które wymaga uwzględnienia wielu zmiennych i zastosowania zaawansowanych metod analizy danych.

## Rozdział 2

# Przegląd literatury

Rozdział przedstawia przegląd literatury dotyczącej prognozowania cen energii elektrycznej, ze szczególnym uwzględnieniem zmiennych wejściowych oraz zwykle używanych modeli prognozowania. W analizie odwołano się zarówno do prac profesora Rafała Werona z Politechniki Wrocławskiej, który jest uznany ekspertem w dziedzinie EPF, jak i do badań innych autorów, aby zapewnić kompleksowy kontekst dla przeprowadzonego badania.

### 2.1 Zmienne wejściowe

W przeglądzie literatury dotyczącym prognozowania cen energii elektrycznej na rynkach dnia następnego, artykuł napisany przez Jesus Lago o współautorstwie prof. Werona [2] może stanowić istotny punkt odniesienia pod względem wyboru zmiennych wejściowych. Autorzy przyjęli zestaw dostępnych danych wejściowych dla prognozowania godzinowych cen na rynkach energii elektrycznej Nord Pool, EPEX-BE, EPEX-FR, EPEX-DE oraz PJM. Niezależnie od zastosowanych przez autorów modeli (LEAR czy DNN), dla wszystkich rynków stworzono wektory zawierające po 24 wartości odpowiadające każdej godzinie dla cen z poprzednich trzech dni oraz tygodnia wstępco. Do tych wektorów dodano wektor określający dzień tygodnia w celu uwzględnienia sezonowości dziennej. Dodatkowo pod uwagę wzięto zmienne fundamentalne, różniące się w zależności od analizowanego rynku.

Dla rynku Nord Pool (NP) były to prognozy zapotrzebowania na moc oraz generacji energii wiatrowej.

Dla rynku PJM w Stanach Zjednoczonych uwzględniono dwie serie prognoz zapotrzebowania na moc: prognozę zapotrzebowania dla całego systemu oraz prognozę dla strefy Commonwealth Edison.

Dla rynku EPEX-BE w Belgii wykorzystano prognozę zapotrzebowania na moc we Francji oraz prognozę generacji we Francji, ponieważ wcześniejsze badania wykazują, że właśnie te dwie zmienne są najlepszymi predyktormi cen belgijskich.

Dla rynku EPEX-FR we Francji brano pod uwagę prognozę zapotrzebowania na moc oraz prognozę generacji.

Dla rynku EPEX-DE w Niemczech uwzględniono prognozę zapotrzebowania na moc oraz zagrego-

wane prognozy generacji energii wiatrowej i słonecznej.

Do wszystkich wymienionych zmiennych fundamentalnych dodano wektory godzinowe tych zmiennych na dzień oraz tydzień wstecz.

Kolejnym artykułem wartym uwagi jest artykuł skupiony na postprocesingu [4] w prognozowaniu cen energii elektrycznej metodami probabilistycznymi. Do osiągnięcia celu zostały wykorzystane następujące zmienne:

- Historyczne ceny energii elektrycznej z poprzednich jednego, dwóch, trzech i siedmiu dni
- Prognoza systemowego zapotrzebowania na moc
- Prognoza generacji energii z odnawialnych źródeł będąca sumą prognoz generacji energii wiatrowej i słonecznej
- Ceny uprawnień do emisji dwutlenku węgla
- Ceny gazu ziemnego
- Ceny ropy naftowej Brent
- Cena emisji węgla
- Zmienne czasowe reprezentujące dni tygodnia
- Zmienne czasowe, czyli zmienne binarne reprezentujące dni tygodnia

Warto podkreślić, że autorzy zastosowali transformację hiperbolicznego sinusa do cen energii elektrycznej w celu stabilizacji wariancji przed modelowaniem.

W innym artykule o współautorstwie prof. Werona [13] podkreśla się, że opóźnione ceny oraz zmienne egzogeniczne określające obciążenie systemowe oraz strefowe mają bardzo istotny wpływ na prognozowanie cen.

Artykuł prof. Ziel [16] analizujący niemiecki RDN, potwierdza poprzednie tezy, wskazując na znaczenie cen energii opóźnionych, obciążenia sieci i energii z OZE. Opisujący przez niego model autoregresyjny opisuje każdą z tych zmiennych jako zależną od jej własnych opóźnionych wartości. Wprowadza skomplikowany zestaw cech opóźnionych jak i również opóźnione zależności niektórych z cech:

- Opóźnienia cen energii elektrycznej o 1, 2, ..., 361, 504, 505, 672, 673, 840, 841, 1008, 1009 godzin
- Opóźnienia obciążenia sieci o 1, 2, ..., 361, 504, 505, 672, 673, 840, 841, 1008, 1009 godzin
- Opóźnienia generacji energii z OZE o 1, 2, ..., 361 godzine
- Opóźnienia obciążenia zależnego od ceny
- Opóźnienia ceny zależnej od obciążenia o 1, 2, ..., 361, 504, 505, 672, 673, 840, 841, 1008, 1009 godzin
- Opóźnienia ceny zależnej od generacji energii z OZE o 1, 2, ..., 49 godzin
- Opóźnienia obciążenia zależnego od generacji energii z OZE o 1, 2, ..., 49 godzin

Oprócz opóźnionych wartości zmiennych, model uwzględnia również trend i efekty sezonowe, takie jak efekty czasowe i kalendarzowe, godziny w ciągu dnia, dni tygodnia, święta publiczne (w tym krajowe i regionalne), zmiany czasu letniego.

Innym przykładem jest praca inżynierska napisana przez Mikołaj Kalisz i Adam Mantiuk na Politechnice Warszawskiej [6]. Analizując rynek energii elektrycznej w Polsce w latach 2021-2023, autorzy analizują zmienne średnią temperaturę godzinową w Polsce, handel międzynarodowy, kurs polskiego złotego względem euro oraz dolara oraz poziom inflacji miesiąc do miesiąca i rok do roku oraz rezerwy mocy ponad i poniżej zapotrzebowania. Po analizie korelacji z wielkiego zestawu cech wybierają dziesięć zmiennych, które mają największy wpływ na cenę energii w wybranym okresie. Są to: produkcja energii z fotowoltaiki, rezerwa mocy poniżej zapotrzebowania, cena uprawnienia emisyjnych (CO<sub>2</sub>), inflacja w porównaniu do poprzedniego roku, cena gazu oraz historyczne ceny prądu opóźnione od dwóch do sześciu dni.

Podsumowując, w literaturze dotyczącej prognozowania cen energii elektrycznej na RDN zmienne wejściowe są różnorodne i zależą od specyfiki danego rynku. Wybrane badania podkreślają znaczenie historycznych cen energii, zapotrzebowanie oraz generacji energii z OZE jako kluczowych zmiennych wpływających na prognozy. Dodatkowo, zmienne fundamentalne, takie jak ceny surowców energetycznych czy zmienne czasowe, również odgrywają istotną rolę. Wszystkie z kluczowych cech są wykorzystywane w niniejszej pracy.

## 2.2 Modele prognozowania

Wśród stosowanych metod do predykcji cen energii elektrycznej wyróżnia się różne podejścia. Rafał Weron w swojej ponadczasowej pracy z 2014 roku "Electricity price forecasting: A review of the state-of-the-art with a look into the future" [14] dokonuje przeglądu literatury z ubiegłych wtedy 15 lat, systematyzując szybko rosnącą liczbę publikacji w tej dziedzinie. Weron wyjaśnia mechanizmy kształtowania się cen na rynkach energii elektrycznej, koncentrując się na cenach dnia następnego. Klasyfikuje techniki predykcyjne pod względem horyzontu czasowego i zastosowanej metodologii. Wymienia następujące kategorie modeli:

- multi-agent
- fundamentalne,
- reduced-form,
- statystyczne,
- computational intelligence,
- hybrydowe.

Modele multi-agent symulują zachowanie uczestników rynku energii, takich jak producenci i konsumenci, w celu przewidywania cen. Weron [14] wskazuje, że modele oparte na równowadze Nasha-Cournota czy równowadze funkcji podaży, są przydatne w analizie długoterminowej, ale nie uwzględniają strategiczne obstawianie cen przez uczestników rynku. W związku z tym tego rodzaju metody pasują najbardziej do stabilnych rynków bez dużych wahań cenowych.

Modele fundamentalne opierają się na analizie czynników ekonomicznych i fizycznych, takich jak ceny paliw, emisje CO<sub>2</sub> czy zapotrzebowanie. Weron [14] dzieli je na parameter-rich (uwzględniające

wiele zmiennych) i parsimonious structural (uproszczone). Głównymi wyzwaniami takich modeli są ich złożoność oraz duże ilości danych, które często mogą być niedostępne w czasie rzeczywistym.

Modele reduced-form opisują dynamikę cen za pomocą procesów stochastycznych, takich jak jump-diffusions czy Markov regime-switching. Weron [14] wskazuje, że są one użyteczne w modelowaniu dziennej zmienności cen, ale mogą nie być dokładne w próbie dokładnego liczenia cen godzinowych, gdyż nie uwzględniają wpływu zmiennych fundamentalnych, takich jak sezonowość czy zmiany w podaży i popycie.

Modele statystyczne, takie jak ARIMA i GARCH, są szeroko stosowane w krótkoterminowym EPF. Modele autoregresywne (AR) wykorzystują liniową kombinację przeszłych wartości zmiennej do prognozowania przyszłych wartości. Modele średniej ruchomej (MA) prognozują zmienną na podstawie liniowej kombinacji przeszłych błędów prognoz. Modele ARMA łączy te dwa podejścia, a modele ARIMA dodają różnicowanie, aby uwzględnić niestacjonarność danych. W celu uwzględnienia sezonowości wykorzystuje się również model SARIMA. Zgodnie z [8] modele SARIMA wykazują dobre wyniki w prognozowaniu cen energii elektrycznej, ale ich skuteczność może być ograniczona w przypadku danych o dużej zmienności. Weron [14] wspomina również o modelach ARX i ARMAX, gdzie X odpowiada za zmienne zewnętrzne wyjaśniające. Model GARCH jest szczególnie przydatny w modelowaniu zmienności cen energii elektrycznej, ponieważ uwzględnia heteroskedastyczność i zmienność w czasie.

Modele computational intelligence, oparte na technikach uczenia maszynowego, zyskały na popularności w nowszych badaniach. Wśród popularnych metod wyróżniają się metody rozmyte, metody wektorów nośnych, LSTM oraz CNN. Jednym z przykładów zastosowania takiego modelu jest wspomniana przeze mnie praca inżynierska na Politechnice Warszawskiej [6]. W celu stworzenia modelu do skomplikowanego okresu rynkowego od 2021 do 2023 roku, autorzy używają Perceptronu wielowarstwowego (MLP). W innej pracy Grzegorza Marcjasza [5], proponowany jest dobór hiperparametrów do głębszej sieci neuronowej (DNN). Badanie porównuje DNN do statystycznego modelu LASSO i podkreśla, iż wyniki DNN są lepsze, co może wskazywać na przewagę nowoczesnego uczenia maszynowego nad tradycyjnymi metodami statystycznymi.

Modele hybrydowe łączą elementy różnych kategorii, aby wykorzystać ich zalety. Jinliang Zhang jest przykładem takiego podejścia. W pracy [11] łączy transformację falkową z ARIMA i GARCH. Argumentuje to w sposób, że połączenie WT z modelami ARIMA i GARCH pozwala na skuteczne modelowanie złożonych cech cen energii elektrycznej, takich jak niestacjonarność, nielinowość i wysoka zmienność. W późniejszej pracy Zhang [15] łączy wspomniane transformację falkową oraz ARIMA z metodą najmniejszych kwadratów maszyn wektorów nośnych (LSSVM). Potwierdzając skuteczność takich metod, Weron [14] podaje, że popularnym podejściem hybrydowym jest łączenie modeli statystycznych z sieciami neuronowymi, co pozwala na modelowanie zarówno liniowych, jak i nieliniowych zależności.

W niniejszej pracy stworzony zbiór danych jest analizowany za pomocą regresji liniowej oraz grzbietowej, jak i modelami Prophet i Wielowarstwowy Perceptron (MLP). Według klasyfikacji Werona, regresja liniowa oraz regresja grzbietowa należą do zbiorów modeli statystycznych. Prophet

## *Rozdział 2. Przegląd literatury*

---

również odpowiada kategorii modeli statystycznych, ale jego algorytm jest bardziej złożony od prostej regresji bądź szeroko używanych modeli ARIMA. Z kolei perceptron wielowarstwowy to przykład modelu computational intelligence, który wykorzystuje sieci neuronowe o różnej architekturze do prognozowania zmiennej docelowej.

## Rozdział 3

# Dane i zmienne

Dane odgrywają kluczową rolę w analizie i prognozowaniu cen energii na RDN. Zbiór danych stworzony w pracy obejmuje okres od 1 stycznia 2016 roku do 31 grudnia 2024 roku i zawiera granulację godzinową. Zbiór danych składa się z różnorodnych zmiennych, które można podzielić na kilka kategorii, które są opisane w tym rozdziale. Duża ilość danych pochodzi z Polskie Sieci Elektroenergetyczne (PSE) [9], czyli otwartej platformie, dostarczającej różnego rodzaju danych dostępnych dla analizy. Warto zauważyć, że od 14 czerwca 2024 roku PSE zmieniła sposób raportowania danych i przeszła na nową stronę [10]. W związku z tym PSE obecnie posiada dwie oddzielne strony do raportów historycznych przed i po tym dniu. Wiele z danych zostało pobrane również ze strony [energy.instrat.pl](http://energy.instrat.pl). Jest to strona, która pobiera dane z platformy PSE i udostępnia je w sposób wygodniejszy, dzięki czemu arkusze csv z dowolnego okresu czasowego o dowolnej częstotliwości można pobrać jednym przyciskiem myszy. Dane dotyczące cen paliw kopalnych oraz kursów walut zostały pozyskane z innych źródeł, takich jak publiczne bazy danych rynkowych lub platformy finansowe. Niniejszy rozdział szczegółowo opisuje zmienną zależną i zmienne niezależne podzielone na kategorie, a także prezentuje kluczowe cechy danych za pomocą tabel i wykresów, co wyjaśnia dobór zmiennych i pozwala na lepsze zrozumienie ich specyfiki i wyzwań związanych z modelowaniem.

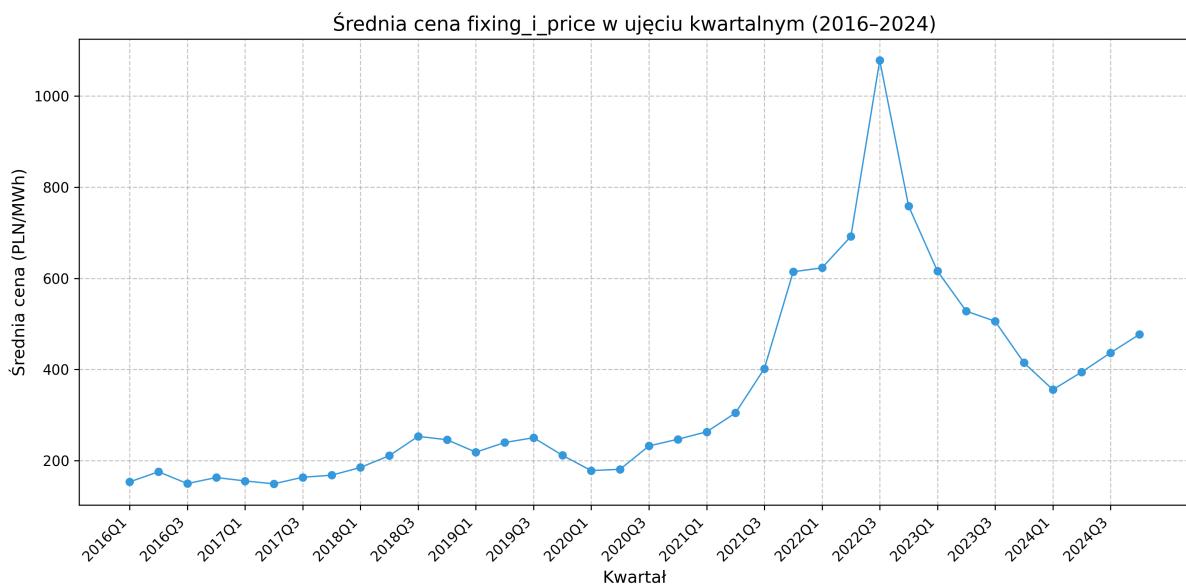
### 3.1 Zmienna zależna

Zmienna zależna w niniejszej pracy to `fixing_i_price`, czyli cena energii elektrycznej na Rynku Dnia Następnego w Polsce, wyrażona w PLN/MWh. Dane dotyczące tej zmiennej zostały pobrane z wymienionej platformy [energy.instrat.pl](http://energy.instrat.pl) w granulacji godzinowej. Zbiór danych obejmuje okres od 1 stycznia 2016 roku do 31 grudnia 2024 roku. Statystyki opisowe zmiennej `fixing_i_price` przedstawiono w Tabeli poniżej.

Statystyka	Wartość
Średnia	344.10 PLN/MWh
Odchylenie std.	268.35 PLN/MWh
Minimum	-360.00 PLN/MWh
25% (Q1)	176.11 PLN/MWh
Mediania	250.00 PLN/MWh
75% (Q3)	434.84 PLN/MWh
Maksimum	3812.45 PLN/MWh

**Tabela 1.** Podstawowe statystyki zmiennej fixing\_i\_price

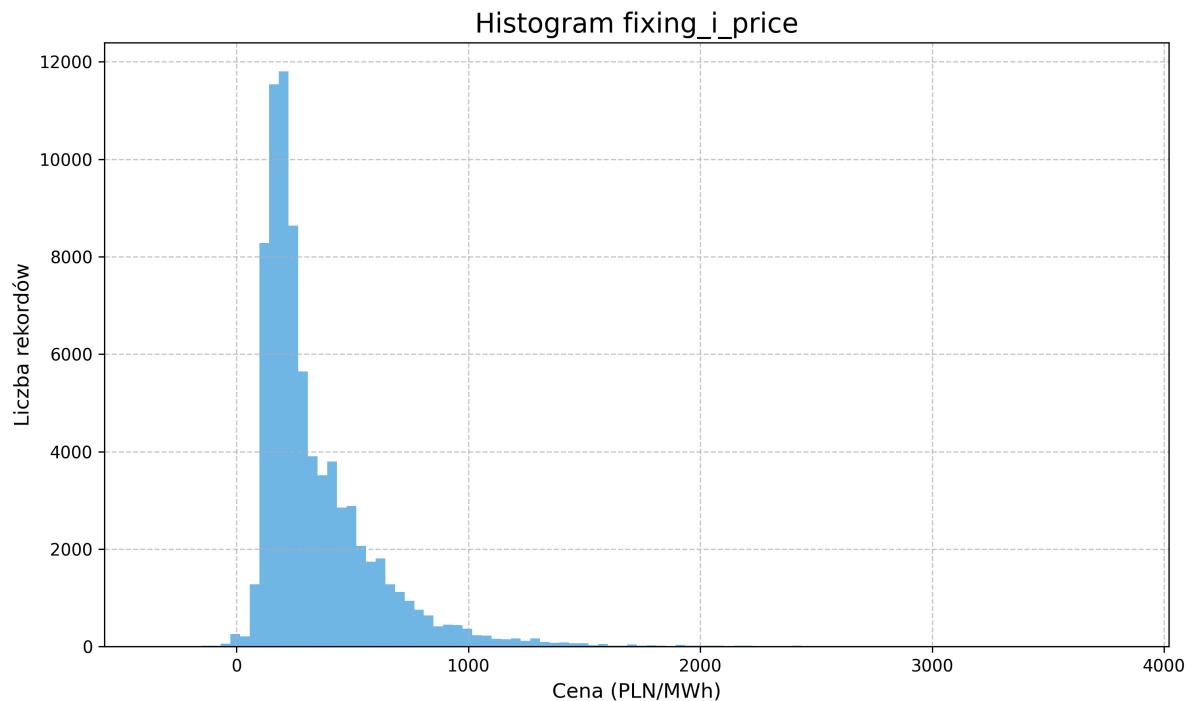
Aby lepiej zrozumieć dynamikę cen energii na RDN, przeanalizowano ich zmienność w całym okresie badania. Rysunek poniżej przedstawia zmienność cen energii w czasie zebranych danych.



**Rysunek 1.** Zmienność cen energii elektrycznej na RDN w latach 2016-2024

Wyraźnie widoczna jest różnica w poziomie cen w różnych okresach: od 2016 do Q4 roku 2020 ceny były stosunkowo stabilne, oscylując w przedziale 100-300 PLN/MWh. Sytuacja zmieniła się w 2020 roku, gdy zaczęły pojawiać się pierwsze skoki cenowe z powodu poważnych obostrzeń z powodu pandemii, a w 2022 roku, w wyniku kryzysu energetycznego wywołanego wojną na Ukrainie i ograniczeniami w dostawach paliw kopalnych, ceny osiągnęły rekordowe poziomy. Pierwszy okres zostanie określony jako okres stabilności cenowej, a drugi jako okres skoków cenowych. Drugi z okresów pokazuje, jak niekorzystne sytuacje gospodarcze mogą wpływać na dynamikę cen energii, co ma istotne implikacje dla modelowania i prognozowania.

Rysunek 2 przedstawia histogram rozkładu zmiennej fixing\_i\_price. Rozkład jest wyraźnie asymetryczny, z długim prawym ogonem, co odzwierciedla występowanie skoków cenowych, takich



Rysunek 2. Histogram rozkładu zmiennej fixing\_i\_price

jak te w 2022 roku. Ujemne ceny, choć rzadkie (ok. 0,4% rekordów), są widoczne w lewej części histogramu, co potwierdza specyficzne cechy danych i rynku.

## 3.2 Zbiór zmiennych niezależnych

W niniejszej pracy wykorzystano różnorodne zmienne niezależne, które można podzielić na kilka kategorii. Obejmują one dane pogodowe, zapotrzebowanie, straty sieciowe, bilans wymiany transgranicznej, dane o produkcji energii przez poszczególne typy generatorów, ceny paliw kopalnych, emisji CO<sub>2</sub> i inne. Wybór tych zmiennych oparty jest na ich potencjalnym wpływie na ceny energii elektrycznej. Poniżej przedstawiono szczegółowy opis każdej z kategorii zmiennych niezależnych, które zostały uwzględnione w analizie.

### 3.2.1 Dane pogodowe

Pierwotnie zbiór danych miał być zestawiony z danych dostępnych za pomocą oficjalnej strony Instytutu Meteorologii i Gospodarki Wodnej, natomiast udostępnione nimi dane historyczne z lat 2016-2024 mają ograniczoną rozdzielcość. Zbierane są przez wiele stacji meteorologicznych rozproszonych po całym kraju, ale tylko w godzinach 6:00, 12:00 oraz 18:00. Aproxymować dane pogodowe w godzinach nocnych jest zadaniem nie do wykonania, szczególnie w przypadku sezonów zimowych, gdzie temperatura w nocy może drastycznie spadać w ciągu godziny. Z tego powodu jako

### *Rozdział 3. Dane i zmienne*

---

źródło danych pogodowych wykorzystano stronę open-meteo.com [7]. Jest to strona, która zbiera dane godzinowe z dokładnością do jednego kilometra. Dostępne są dane pogodowe dla dowolnego okresu czasu i lokalizacji.

Dane pogodowe zostały pobrane dla czterech lokalizacji w Polsce: Warszawy (WAW), Koszalina (KSZ), Krakowa (KRK) i Babimost (BAB). Wybór miast został podkutowany ich zróżnicowaniem geograficznym i klimatycznym, co pozwala uwzględnić regionalne różnice w warunkach pogodowych wpływających na produkcję i zapotrzebowanie na energię.

Warszawa, jako stolica i największe miasto Polski, reprezentuje centralny region kraju o wysokim zapotrzebowaniu na energię, szczególnie w okresach zimowych i letnich. Koszalin, położony na Pomorzu, jest kluczowy ze względu na bliskość farm wiatrowych na Morzu Bałtyckim, co czyni go istotnym punktem dla analizy produkcji energii wiatrowej. Kraków, znajdujący się w południowej Polsce, charakteryzuje się większym udziałem energii słonecznej w miksie energetycznym, a także wysokim zapotrzebowaniem na energię w sezonie grzewczym. Babimost, zlokalizowany w zachodniej Polsce, jest istotny ze względu na swoje położenie na zachodzie Polski w pobliżu granicy z Niemcami.

Parametry pogodowe zostały wybrane z uwzględnieniem ich bezpośredniego wpływu na rynek energii.

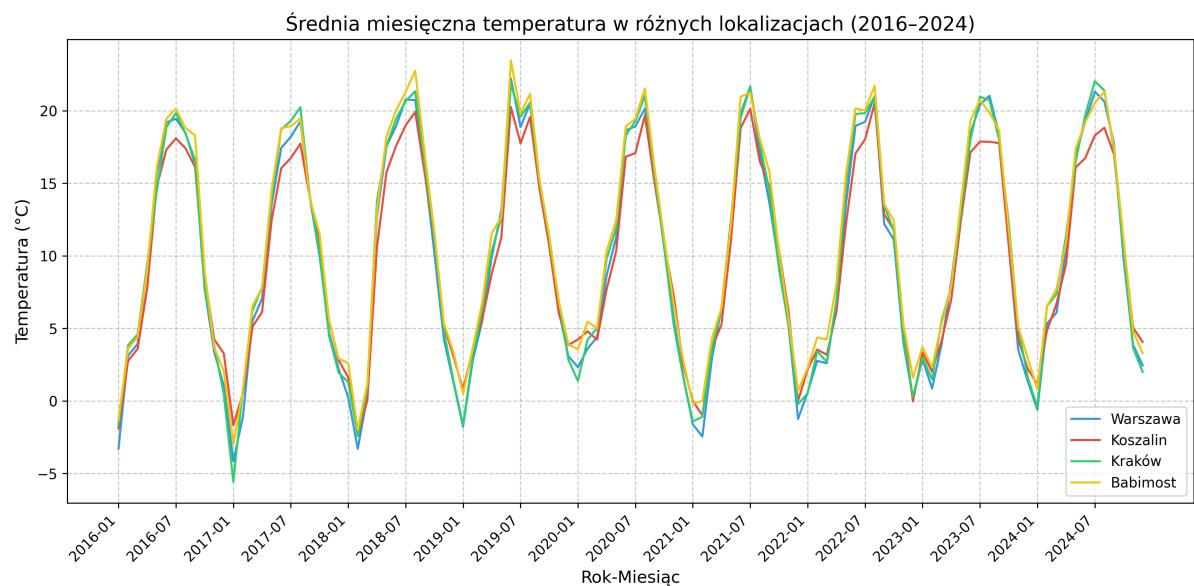
Temperatura jest kluczowym czynnikiem, ponieważ bezpośrednio wpływa na zapotrzebowanie na energię. Niskie temperatury zwiększą zużycie energii na ogrzewanie, natomiast bardzo wysokie temperatury latem mogą podnosić zapotrzebowanie na klimatyzację.

Udostępniona prędkość wiatru została zmierzona na wysokości 100 metrów nad powierzchnią ziemi. Taka wysokość została wybrana, ponieważ jest przeciętna dla turbin wiatrowych w Polsce, co pozwala dokładniej oszacować potencjalną produkcję energii z farm wiatrowych.

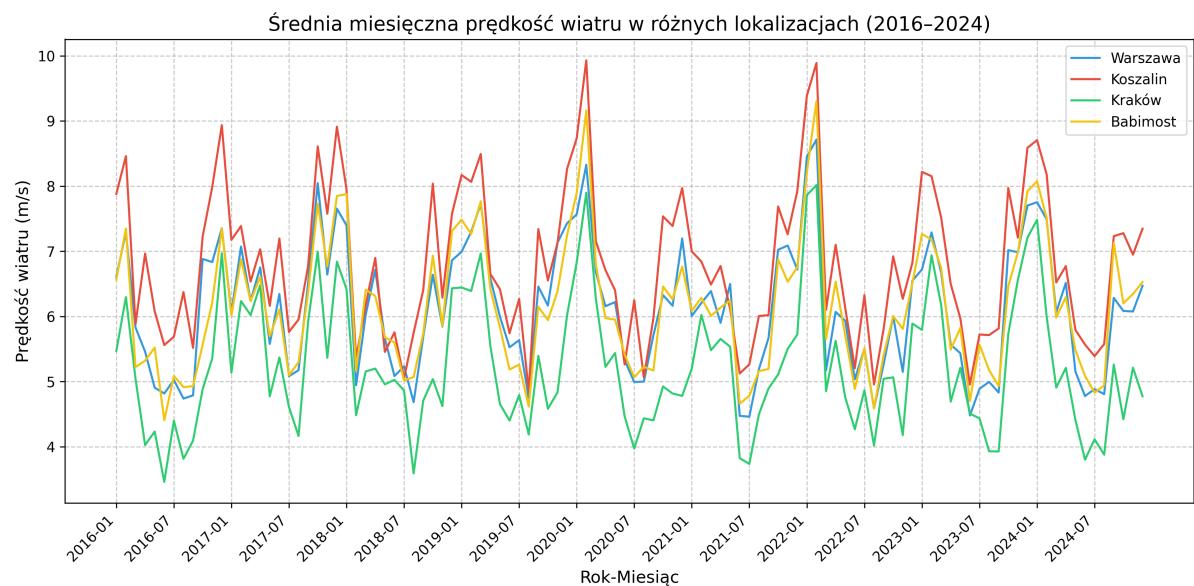
Promieniowanie słoneczne jest istotnym czynnikiem dla produkcji energii z paneli fotowoltaicznych. Zachmurzenie zostało uwzględnione, ponieważ wysoki poziom zachmurzenia zmniejsza efektywność paneli słonecznych. Wybór tych parametrów pozwala na kompleksową analizę wpływu pogody na ceny energii na RDN.

Poniżej przedstawię wykresy dla każdego z parametrów pogodowych, które zostały uwzględnione w analizie. Wykresy przedstawiają zmienność danych pogodowych w czasie. Zachmurzenie jest wyrażone w oktantach (0-8), gdzie 0 oznacza brak zachmurzenia, a 8 oznacza całkowite zachmurzenie.

### 3.2. Zbiór zmiennych niezależnych

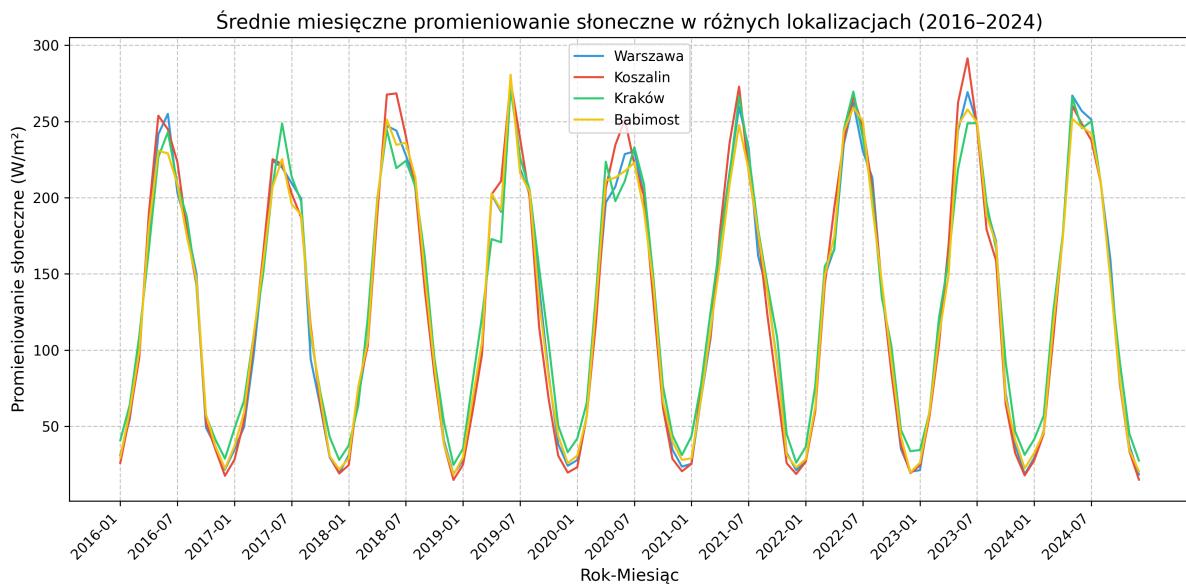


Rysunek 3. Zmienność temperatury w czasie (2016-2024)

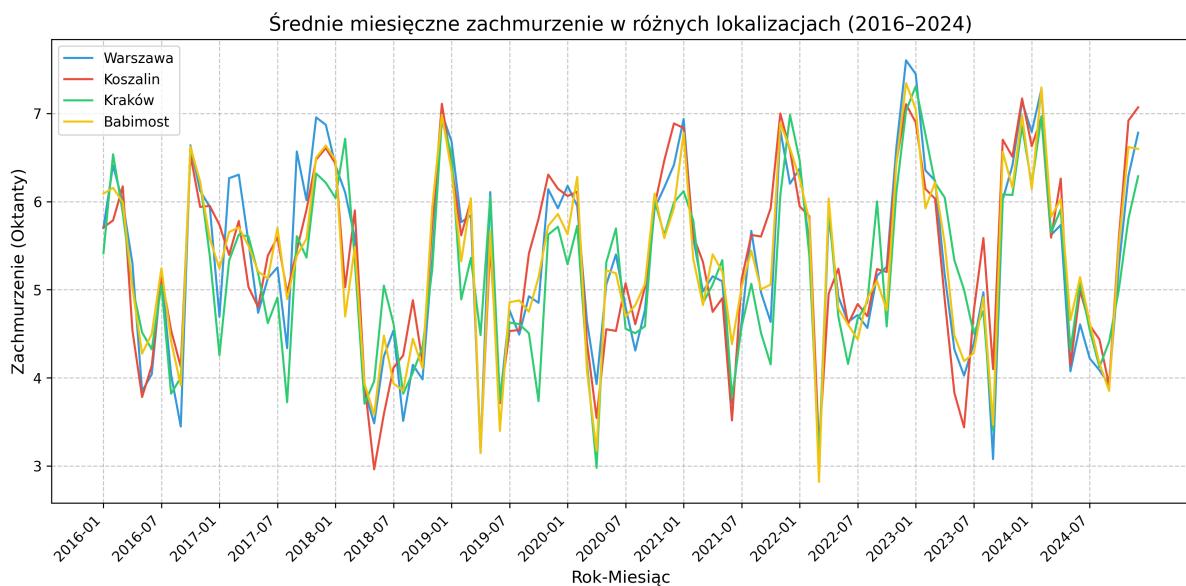


Rysunek 4. Zmienność prędkości wiatru w czasie (2016-2024)

### Rozdział 3. Dane i zmienne



Rysunek 5. Zmienna promieniowania słonecznego w czasie (2016-2024)

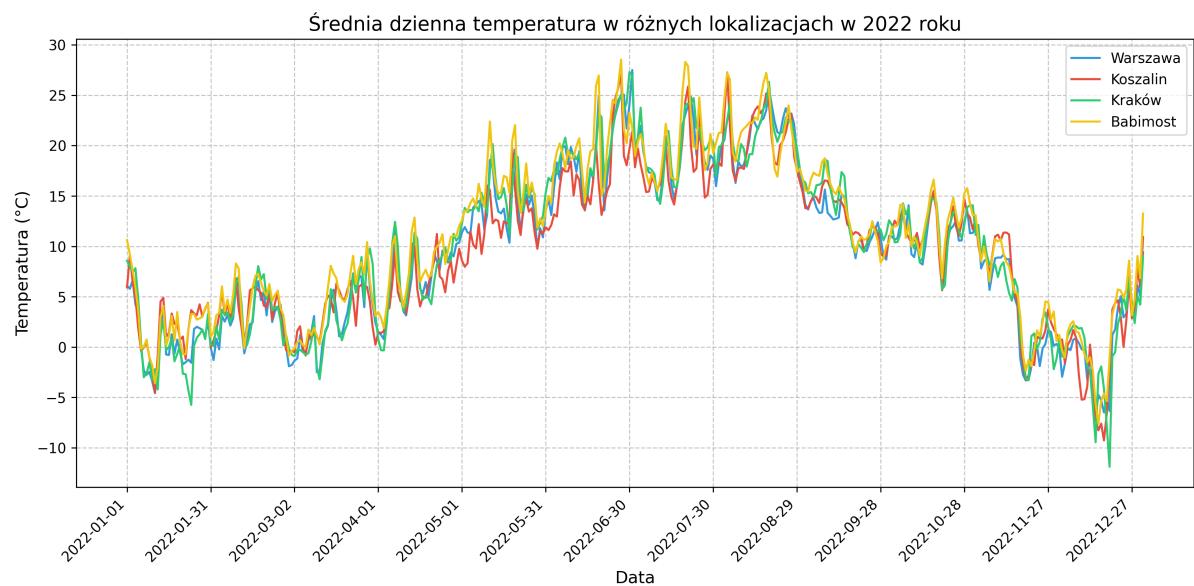


Rysunek 6. Zmienna zachmurzenia w czasie (2016-2024)

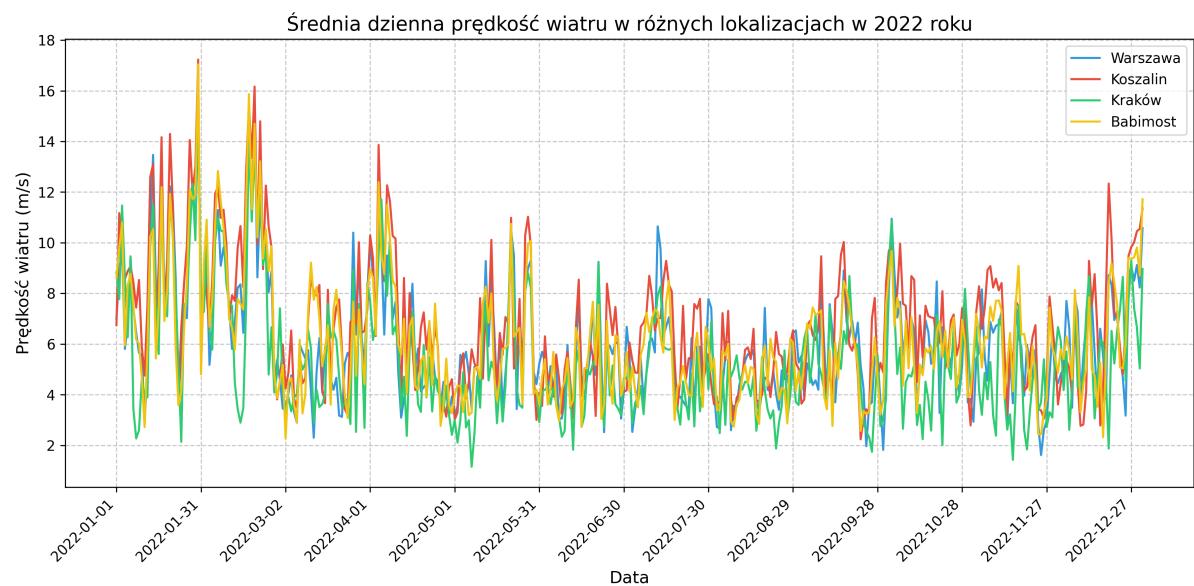
Każdy z wykresów przedstawia zmienność danego parametru pogodowego w wybranych lokalizacjach w przeciągu okresu badawczego. Wyraźnie widać sezonowe wahania parametrów pogodowych, co jest typowe dla klimatu Polski. Temperatura i promieniowanie słoneczne mają wyraźnie większe wartości w sezonach letnich, prędkość wiatru w sezonach zimowych, z kolei zachmurzenie ma bardziej zróżnicowany charakter.

Chciałbym również przybliżyć wykresy i potwierdzić efekt sezonowości na przykładzie 2022 roku.

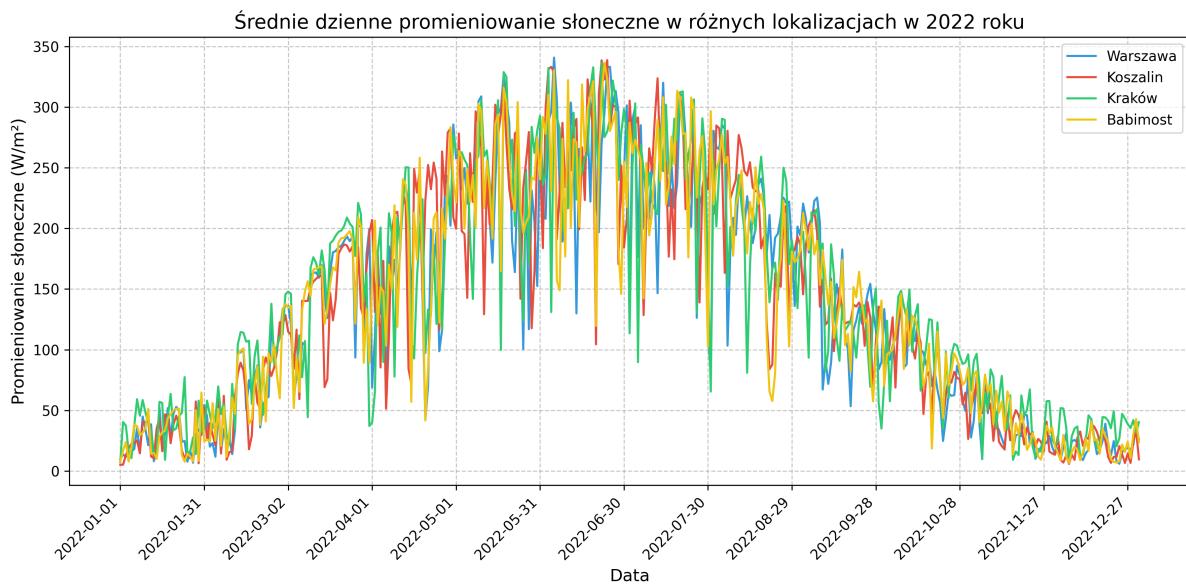
### 3.2. Zbiór zmiennych niezależnych



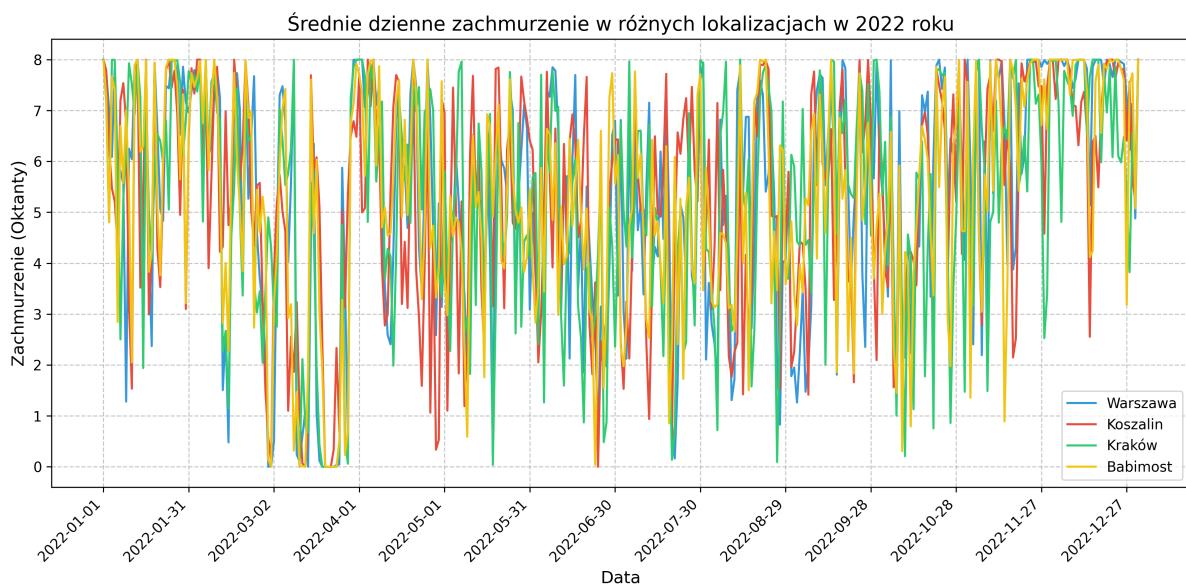
Rysunek 7. Zmiennaść temperatury w czasie (2022)



Rysunek 8. Zmiennaść prędkości wiatru w czasie (2022)



Rysunek 9. Zmienność promieniowania słonecznego w czasie (2022)



Rysunek 10. Zmienność zachmurzenia w czasie (2022)

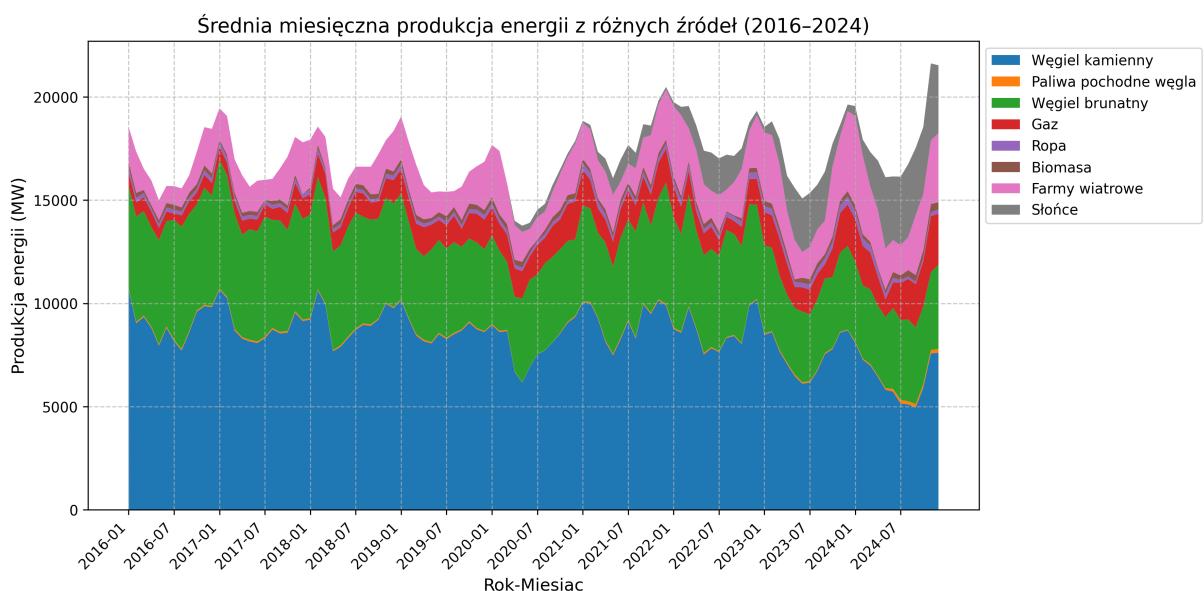
Temperatura jest rzadko uwzględniana w analizach prognozowania cen energii, ponieważ nie jest bezpośrednio związana z produkcją energii. Niemniej jednak, w przypadku Polski, gdzie duża część energii elektrycznej pochodzi z węgla, temperatura może mieć istotny wpływ na zapotrzebowanie, szczególnie w sezonie grzewczym.

### 3.2.2 Produkcja energii z wybranych źródeł

Zmienne dotyczące produkcji energii z różnych źródeł odgrywają kluczową rolę w analizie cen energii na RDN, ponieważ odzwierciedlają strukturę podaży energii w Polsce, która ma bezpośredni wpływ na dynamikę cen. W niniejszej pracy uwzględniono osiem zmiennych opisujących produkcję energii:

- **hard\_coal** - produkcja z węgla kamiennego (MWh),
- **coal\_derived** - produkcja z paliw pochodnych węgla (MWh),
- **lignite** - produkcja z węgla brunatnego (MWh),
- **gas** - produkcja z gazu ziemnego (MWh),
- **oil** - produkcja z ropy naftowej lub jej pochodnych (MWh),
- **biomass** - produkcja z biomasy (MWh),
- **wind** - produkcja z elektrowni wiatrowych lądowych (MWh),
- **solar** - produkcja z paneli fotowoltaicznych (MWh).

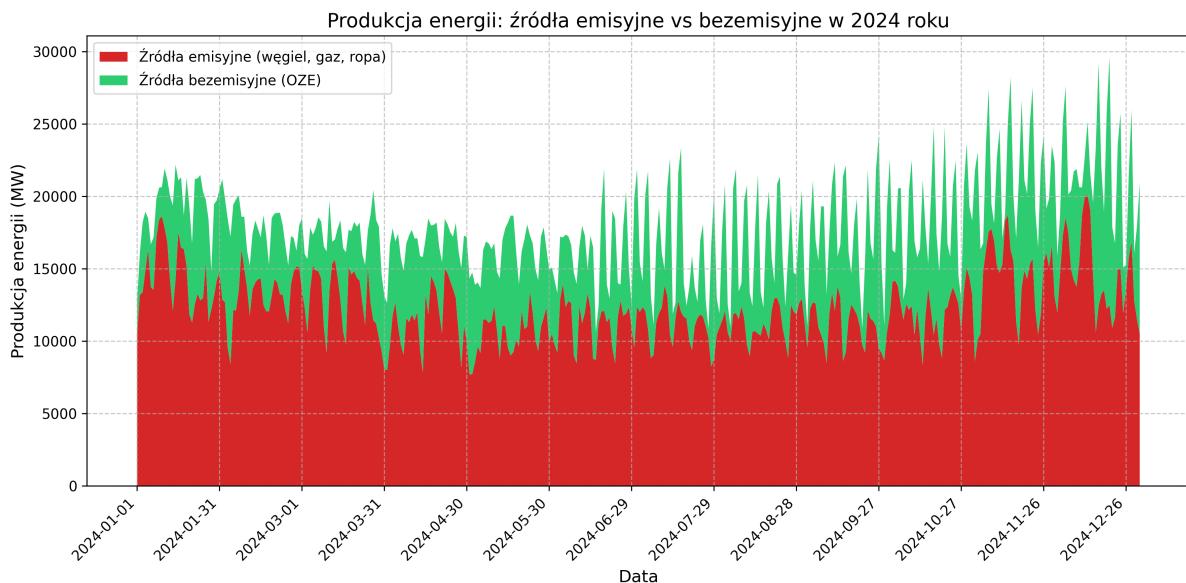
Dane te pochodzą z Polskich Sieci Elektroenergetycznych (PSE) i zostały dopasowane do godzinowego formatu danych RDN, co pozwoliło na ich integrację z pozostałymi zmiennymi.



Rysunek 11. Zmienna produkcji energii z różnych źródeł w czasie (2016-2024)

Wykres powyżej przedstawia średnią miesięczną produkcję energii z różnych źródeł w Polsce w latach 2016-2024, co jest istotne w kontekście prognozowania cen energii na RDN. Wykres obszarowy ukazuje dominację węgla kamiennego i węgla brunatnego, które do tej pory odpowiadają za większość produkcji energii elektrycznej. Biomasa oraz paliwa pochodne węgla stanowią znikomą część podaży i mogą być pominięte dla redukcji ilości zmiennych. Produkcja gazu i ropy stale posiada niewielką, ale istotną część produkcji energii. Warto zauważyć, że produkcja z OZE, szczególnie ze słońca znacząco rośnie w ostatnich latach, co może mieć istotny wpływ na ceny energii i umiejętności jej prognozowania. Do poprzednio omówionych zmiennych dodano również zmienną

non\_emissive\_sources\_percentage, która reprezentuje procentowy udział energii produkowanej z OZE w całkowitej produkcji energii elektrycznej w Polsce.



**Rysunek 12.** Porównanie produkcji energii z emisyjnych i bezemisyjnych źródeł w 2024 roku

W najbardziej korzystne dla gospodarki momenty udział OZE w produkcji energii może przekraczać udział źródeł emisyjnych, co może zwykle prowadzi do spadku cen. Warto również zauważyć, że produkcja z węgla kamiennego i brunatnego jest bardziej stabilna i przewidywalna niż produkcja z OZE, co wpłynie na dokładność prognoz. W związku z tym, zmienne dotyczące produkcji energii z różnych źródeł są istotnym elementem modelowania cen energii.

### 3.2.3 Handel energią z państwami sąsiednimi

Zmienne dotyczące wymiany energii z innymi krajami są kolejnym istotnym elementem analizy cen energii na Rynku Dnia Następnego, ponieważ pozwalają na uwzględnienie wpływu handlu międzynarodowego na ceny energii w Polsce. W niniejszej pracy uwzględnione zostały bilanse handlu energią z Państwami sąsiednimi, z którymi Polska ma połączenia transgraniczne. Są to Państwa takie jak:

- Niemcy (DE),
- Czechy (CZ),
- Słowacja (SK),
- Litwa (LT),
- Szwecja (SE),
- Ukraina (UA).

Dane dotyczące wymiany energii z tymi krajami pochodzą z Polskich Sieci Elektroenergetycznych i są dostępne w granulacji godzinowej. Warto zauważyć, że Polska jest jednym z kluczowych graczy

### 3.2. Zbiór zmiennych niezależnych

---

na rynku energii w Europie Środkowo-Wschodniej, co sprawia, że wymiana energii zwykle występuje z każdym krajem w każdej godzinie czasu rzeczywistego. W okresach wysokiego zapotrzebowania na energię lub w sytuacjach poważnych awarii, Polska sięga po energię z innych krajów, co prowadzi do wzrostu cen. Z drugiej strony, w okresach niskiego zapotrzebowania, Polska może eksportować, czyli sprzedawać nadwyżki energii, w momentach spadków cen. Prawdopodobnie, taki aktywny handel może również wynikać z różnic cenowych pomiędzy krajami. Poniżej przedstawiam wykresy dotyczące wymiany energii z sąsiadami w latach 2016-2024.

Rok	Niemcy (MW)	Czechy (MW)	Litwa (MW)	Słowacja (MW)	Szwecja (MW)	Ukraina (MW)
2016	994.26	-761.81	68.21	-475.47	294.36	108.99
2017	836.24	-612.73	119.92	-499.21	340.96	102.45
2018	803.00	-373.78	102.53	-366.14	310.83	160.96
2019	1149.22	-312.31	216.38	-367.32	329.80	157.17
2020	1275.75	-264.90	201.93	-348.95	428.81	167.98
2021	957.79	-901.92	110.04	-519.59	364.69	93.55
2022	918.62	-1000.89	86.53	-684.25	429.17	126.95
2023	851.27	-526.46	58.40	-382.18	431.60	5.76
2024	1005.26	-544.15	-112.90	-389.65	296.29	-68.03

**Tabela 2.** Średni bilans wymiany energii z sąsiadami w latach 2016-2024

Wartości te są średniogodzinnymi bilansami wymiany energii z sąsiadami w latach 2016-2024. Wartości dodatnie oznaczają eksport energii, a wartości ujemne import energii. Polska ma dodatni bilans wymiany energii z Niemcami, co oznacza, że regularnie eksportuje energię do Niemiec. Największe ujemne bilanse Polska posiada z Czechami, prawdopodobnie spowodowane jest to funkcjonującymi w Czechami elektrowniami atomowymi. Przepływy transgraniczne są warte uwagi i uwględnienia w zbiorze.

#### 3.2.4 Ceny paliw kopalnych i emisji CO<sub>2</sub>

Zmienne dotyczące cen paliw kopalnych oraz emisji CO<sub>2</sub> odgrywają kluczową rolę w analizie cen energii na RDN, ponieważ koszty paliw i emisji mają bezpośredni wpływ na ceny energii elektrycznej w Polsce, gdzie dominującym źródłem energii jest węgiel. W niniejszej pracy uwzględniono cztery zmienne, które zostały opisane w Tabeli 3.

Nazwa zmiennej	Opis	Źródło danych	Granulacja
gas_price	Cena gazu ziemnego (PLN/MWh)	energy.instrat.pl	Dzienna
coal_price	Cena węgla kamiennego (PLN/GJ)	energy.instrat.pl	miesięczna
co2_price	Cena uprawnień do emisji CO <sub>2</sub> (PLN/tCO <sub>2</sub> )	energy.instrat.pl	tygodniowa
brent_price	Cena ropy Brent Europe (PLN/bar)	www.eia.gov	Dzienna

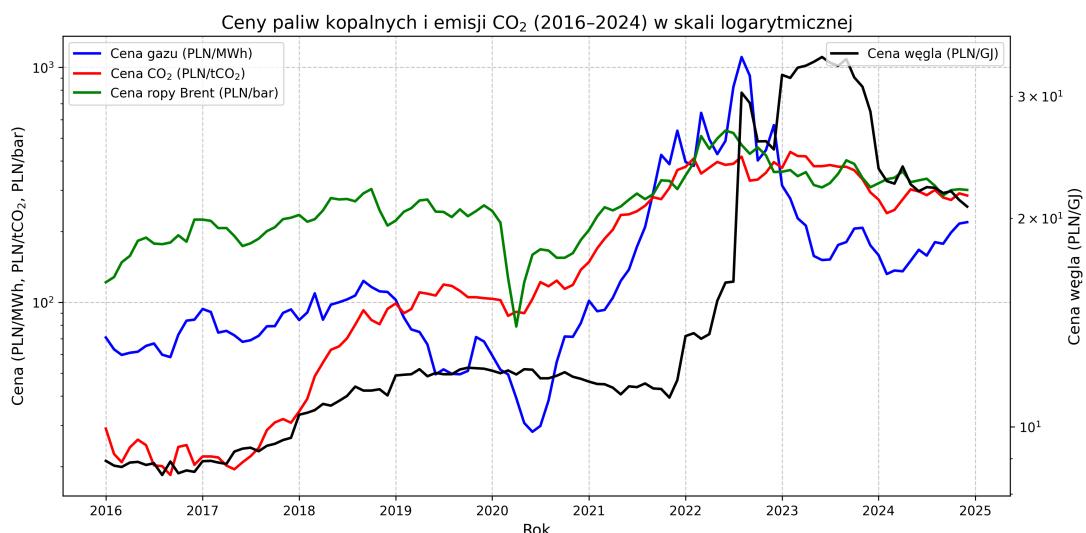
**Tabela 3.** Opis zmiennych dotyczących cen paliw kopalnych i emisji CO<sub>2</sub>.

### Rozdział 3. Dane i zmienne

Dane te zostały dopasowane do godzinowego formatu danych RDN, w taki sposób, że w dniach pracy giełdy i znanych cen paliw, cena stała jest przypisana do godzin, a w dniach wolnych od pracy lub bez znanych cen, wartości są interpolowane.

Ceny paliw kopalnych i emisji CO<sub>2</sub> są istotne w kontekście prognozowania cen energii, ponieważ wpływają na koszty produkcji energii w elektrowniach konwencjonalnych, które dominują w polskim miksie energetycznym. Na przykład, wzrost ceny węgla (coal\_price) lub uprawnień do emisji CO<sub>2</sub> (co2\_price) zwiększa koszty produkcji energii w elektrowniach węglowych. Cena gazu (gas\_price) jest kluczowa dla elektrowni gazowych, które pełnią rolę bilansującą w systemie. Cena ropy Brent (brent\_price) ma mniejszy bezpośredni wpływ na produkcję energii w Polsce, ale jest istotna w kontekście globalnych trendów cen paliw i skorelowana z innymi cenami surowców energetycznych.

Aby zilustrować dynamikę tych zmiennych, na rysunku poniżej przedstawiono zmiany cen paliw kopalnych i emisji CO<sub>2</sub> w latach 2016-2024 w ujęciu miesięcznym. Dla lepszego przedstawienia danych użyłem skali logarytmicznej.



**Rysunek 13.** Ceny paliw kopalnych i emisji CO<sub>2</sub> w latach 2016-2024 (w ujęciu miesięcznym). Źródło: Opracowanie własne na podstawie danych rynkowych.

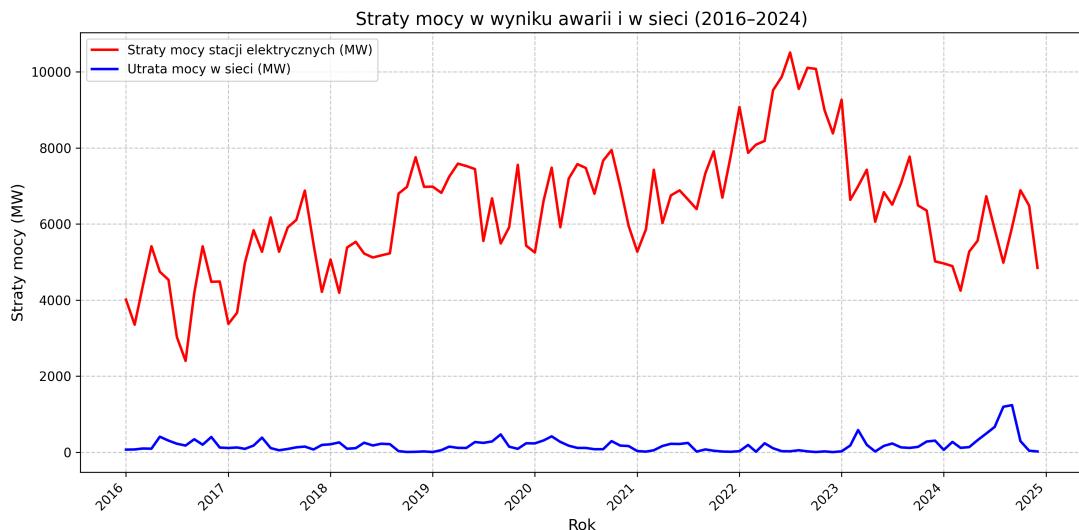
Ceny paliw kopalnych w okresie wysokiej zmienności szybko rosną, co jest jedną z przyczyn rosnących cen energii. Największy wzrost ma cena gazu, która od rozpoczęcia konfliktu zbrojnego wzrosła ponad 10-krotnie w ciągu roku.

#### 3.2.5 Straty mocy w systemie elektroenergetycznym

Zmienne dotyczące strat mocy w systemie elektroenergetycznym odgrywają istotną rolę w analizie cen energii na Rynku Dnia Następnego (RDN), ponieważ wpływają na dostępność energii w systemie oraz koszty jej przesyłu i dystrybucji. W niniejszej pracy uwzględniono następujące dane zbierane przez systemy PSE: power\_loss (utrata mocy w wyniku awarii w MW) oraz network\_loss (utrata mocy w sieci w MW). Udostępnione dane mają granulację godzinową.

Straty mocy w systemie elektroenergetycznym są kluczowe w kontekście prognozowania cen energii, ponieważ zmniejszają ilość energii dostępnej dla odbiorców, co może prowadzić do wzrostu cen na RDN oraz Rynek Bilansujący (RB). Utrata mocy w wyniku awarii

Aby zilustrować dynamikę tych zmiennych, na rysunku 14 przedstawiono zmiany strat mocy w wyniku awarii (power\_loss) oraz strat mocy w sieci (network\_loss) w latach 2016-2024 w ujęciu miesięcznym.



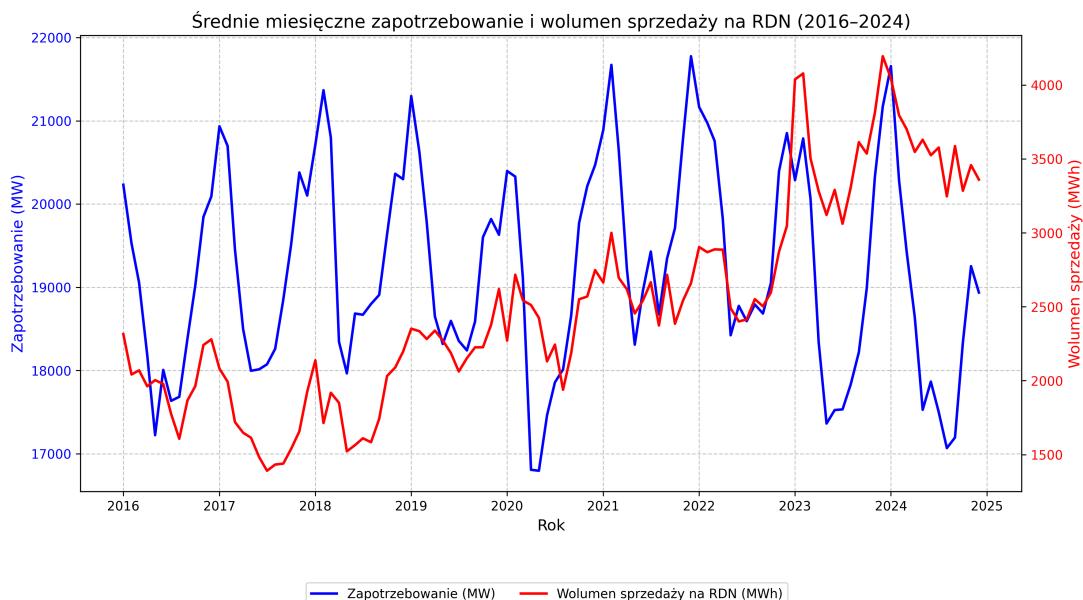
**Rysunek 14.** Straty mocy w wyniku awarii i w sieci w latach 2016-2024 (w ujęciu miesięcznym). Źródło: Opracowanie własne na podstawie danych z PSE.

Straty sieciowe na wykresie po 06.2024 są większe od poprzednich. Wynika to ze zmiany metodyki obliczania strat sieciowych przez PSE. Niemniej jednak straty mocy są znaczące, ale wynikają prawdopodobnie z tego, że nie wszystkie jednostki wytwórcze pracują w danym momencie lub z ograniczonej ich eksploatacji.

### 3.2.6 Zapotrzebowanie na energię i wolumen handlu

Kolejnymi z ważnych zmiennych są zapotrzebowanie na energię i wolumen handlu. Przedstawiają one popyt na rynku, przez co można je określić podstawowymi czynnikami kształtującym ceny. W niniejszej pracy uwzględniono następujące zmienne: load (zapotrzebowanie na energię w MWh) oraz trade\_volume (wolumen handlu w MWh). Obie zmienne pochodzą z Polskich Sieci Elektroenergetycznych (PSE) i są dostępne w granulacji godzinowej.

Poniżej na wspólnym wykresie przedstawię różnice pomiędzy zapotrzebowaniem a wolumenem handlu.



**Rysunek 15.** Zapotrzebowanie na energię i wolumen handlu w latach 2016-2024 (w ujęciu miesięcznym). Źródło: Opracowanie własne na podstawie danych z PSE.

Rozbieżność między zapotrzebowaniem a wolumenem sprzedaży na RDN ma swój powód. Wolumen sprzedaży na RDN zwykle nie przekracza 5000 MWh. Porównanie do zapotrzebowania na poziomie ponad 15 000 MWh wskazuje, że RDN pokrywa jedynie część całkowitego zapotrzebowania. Pozostała część jest zaspokajana przez: (1) kontrakty bilateralne (OTC), (2) rynek bilansujący, na którym PSE obsługuje zakup oraz sprzedaż energii w czasie rzeczywistym, (3) import energii z sąsiednich krajów, jak pokazano w podrozdziale 3.2.3 oraz innego rodzaju transakcje.

Wyraźnie widoczne są duże szczyty zapotrzebowania zapotrzebowania w okresie zimowym. Wynika to z zwiększonego zapotrzebowania na energię elektryczną w okresie grzewczym, co jest typowe dla klimatu Polski.

### 3.2.7 Inne zmienne

W swoim zbiorze danych uwzględnione zostały inne zmienne, które mogą mieć wpływ na ceny energii na RDN.

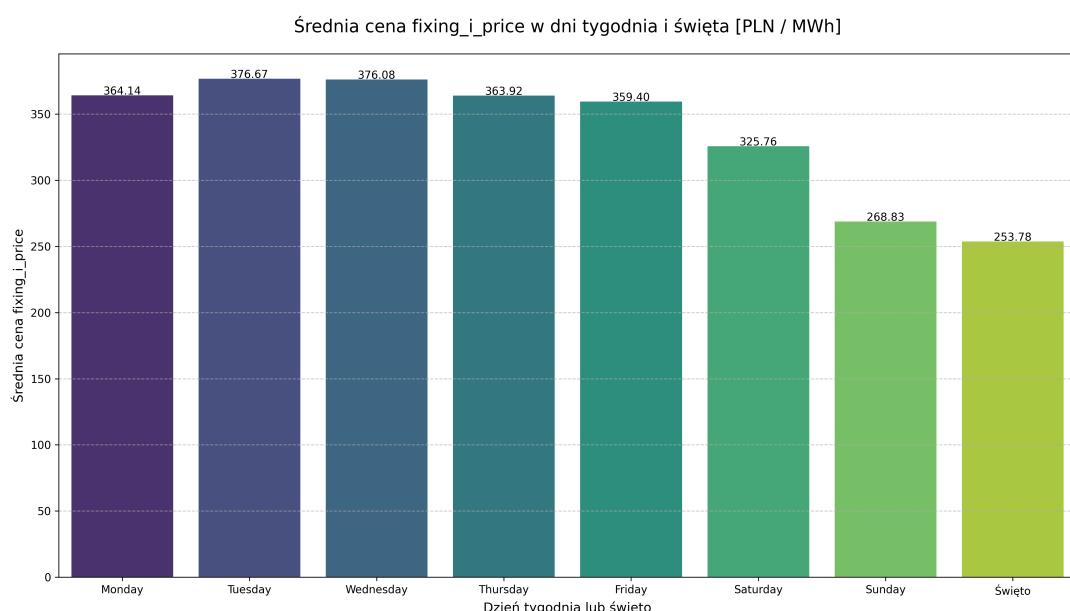
#### Zmienne sezonowe

Zmienne sezonowe, takie jak `day_of_week`, `month` i `is_holiday`, zostały wprowadzone do zbioru danych w celu uchwycenia cykliczności i wzorców sezonowych w cenach energii na Rynku Dnia Następnego.

Zmienna `month` reprezentuje miesiąc roku (1-12, gdzie 1 to styczeń, a 12 to grudzień). Na temat cykliczności miesięcznej wspomniałem wcześniej w podrozdziałach 3.2.4 i 3.2.6. Niemniej jednak, wprowadzenie zmiennych `day_of_week` i `month` do modelu pozwala lepiej uchwycić cykliczność

i sezonowość w cenach energii, na poszczególne godziny dnia o czym świadczy poniższy wykres korelacji.

Zmienna `day_of_week` reprezentuje dzień tygodnia w zakresie od 1 do 7. Zmienna pozwala uwzględnić modelowi cykliczność tygodniową w cenach energii. Z wykresu poniżej wynika, że zapotrzebowanie na energię jest zazwyczaj wyższe w dni robocze, gdy działa przemysł i biura, a niższe w weekendy i święta, gdy aktywność gospodarcza jest mniejsza. Ta cykliczność przekłada się na ceny energii na RDN: w dni robocze ceny są zazwyczaj wyższe, szczególnie w godzinach szczytu. Wprowadzone zmienne sezonowe pozwalają modelowi lepiej uchwycić te wzorce, co jest istotne w dokładnym modelowaniu.



Rysunek 16. Korelacja dni tygodnia z cenami energii na RDN.

Dodatkowo wprowadzono zmienną `peak_hour`, która jest zmienną binarną określającą godziny szczytu, w których występuje zwiększone zapotrzebowanie na energię elektryczną. Przyjmuje wartość 1, jeśli spełniony jest jeden z dwóch warunków, w przeciwnym razie 0:

- W dni wolne lub dni weekendowe, zapotrzebowanie przekracza 18 000 MW, a godzina należy do przedziałów porannych (7:00-9:00) lub wieczornych (16:00-18:00).
- W dni robocze, zapotrzebowanie przekracza 23 000 MW i posiada takie same przedziały godzinowe jak w dni wolne.

Definicja ta odzwierciedla wyższe zapotrzebowanie na energię w godzinach szczytu, z różnymi progami dla dni roboczych i wolnych, co jest zgodne z charakterystyką polskiego rynku energii.

### Ceny historyczne

Zmienne historyczne `fixing_i_price_lag24`, `fixing_i_price_lag48`, `fixing_i_price_lag72`, `fixing_i_price_lag96`, `fixing_i_price_lag120`, `fixing_i_price_lag144` oraz `fixing_i_price_lag168`, zostały wprowadzone do zbioru danych w celu uchwycenia autokorelacji w cenach energii na RDN. Zmienne te zostały wygenerowane na podstawie kolumny `fixing_i_price`, która reprezentuje cenę energii na RDN w danej godzinie (w PLN/MWh), poprzez przesunięcie wartości o odpowiednią liczbę godzin.

Oprócz ceny opóźnionej, zostały wprowadzone średnie kroczące z ostatnich 24 i 48 godzin. W wyniku tego do datasetu zostały dodane zmienne `fixing_i_price_mean24` i `fixing_i_price_mean48`, które reprezentują średnie ceny energii na RDN w ostatnich 24 i 48 godzinach. Wartości te zostały obliczone na podstawie danych z kolumny `fixing_i_price` z przesunięciem o jedną godzinę, żeby nie w średniej kroczącej nie brać pod uwagę obecnej wartości.

Wprowadzenie zmiennych historycznych do modelu pozwala lepiej uchwycić autokorelację i cykliczność w cenach energii, co jest kluczowe dla poprawy dokładności prognoz. Zgodnie z literaturą, ceny historyczne są często jednymi z najważniejszych predyktorów w modelach prognozowania cen energii.

### Kurs wymiany PLN/USD

Zmienna `pln_usd` reprezentuje kurs wymiany złotego polskiego względem dolara amerykańskiego (USD/PLN) w danej godzinie. Dane te zostały pozyskane z oficjalnej strony Narodowego Banku Polskiego (NBP) w granulacji dziennej. W celu dopasowania danych do godzinowego formatu RDN, wartości kursu zostały przypisane do wszystkich godzin w danym dniu. W przypadku dni wolnych od pracy lub braku dostępnych danych, wartości kursu zostały interpolowane.

Rok	Średni kurs PLN/USD
2016	3.94
2017	3.78
2018	3.61
2019	3.84
2020	3.90
2021	3.86
2022	4.46
2023	4.20
2024	3.98

**Tabela 4.** Średni kurs wymiany PLN/USD w latach 2016-2024. Źródło: Opracowanie własne na podstawie danych NBP.

Kurs wymiany PLN/USD jest istotnym czynnikiem w kontekście prognozowania cen energii na RDN, ponieważ Polska importuje znaczną część paliw kopalnych, takich jak gaz ziemny i ropa

### *3.2. Zbiór zmiennych niezależnych*

---

naftowa, które są wyceniane w dolarach amerykańskich. Wzrost kursu PLN/USD, czyli osłabienie złotego względem dolara zwiększa koszty importu surowców. Kurs polskiego złotego odzwierciedla też zmiany w gospodarce krajowej i globalnej. Próbuje stabilizować kurs, NBP zmienia politykę monetarną kraju, co również wpływa na koszt produkcji energii i przekłada się na jej cenę.

## Rozdział 4

# Eksploracja danych

W niniejszym rozdziale przedstawiono proces przygotowania omówionego zestawu danych. Proces eksploracji danych ubejmuje obsługę braków wartości, dostosowanie granulacji zmiennych, analizę korelacji, podział danych na okresy o różnej zmienności, podział na zbiory treningowy, walidacyjny i testowy oraz preprocessing danych przed modelowaniem.

### 4.1 Wstępna obróbka danych

#### Obsługa braków wartości

Pierwszym krokiem eksploracji jest analiza i obsługa brakujących wartości w zbiorze danych. Braki wartości w zmiennych objaśniających mogą wynikać z przyczyny, że w trakcie niektórych godzin pomiary nie były zbierane, lub z powodu błędów w danych, które mogły prowadzić do ich usunięcia. Na szczęście pierwotny zbiór danych zawierał niewiele braków i nie były one zbyt ciągłe w kontekście całego zbioru. W związku z tym rekordy zawierające brakujące wartości w miejscu dowolnej zmiennej zostały usunięte. Z oczekiwanych 78912 rekordów godzinowych pozostało 78451. Jest to zaledwie 0.55% braków, co jest nie powinno stanowić problemu w dalszej analizie.

Niektóre zmienne, takie jak kursy walut czy ceny paliw kopalnych miały braki w okresach zamkniętego rynku, czyli weeekendy i święta. W takich przypadkach brakujące wartości zostały przeniesione z poprzedniego dnia handlu.

#### Obsługa granulacji

Dane użyte w pracy charakteryzowały się różną granulacją czasową. PSE udostępnia dane dotyczące sieci energii elektrycznej, w tym handlu w rozdzielczości godzinowej. Zmienne pogodowe również mają granulację dzienną. Natomiast zmienne makroekonomiczne, takie jak ceny paliw, są dostępne w częstotliwości dniowej lub tygodniowej.

Aby ujednolicić granulację do poziomu godzinowego, zastosowano dwie techniki. W przypadku zmiennych o granulacji dziennej, takich jak dane pogodowe, założono, że wartości w ciągu doby nie ulegają zmianie. Wartości dzienne przypisano więc każdej godzinie danego dnia, co pozwoliło

na zachowanie prostoty przy jednoczesnym dostosowaniu danych do godzinowej rozdzielczości ceny energii.

Dla zmiennych o rzadszych granulacjach, na przykład dla ceny węgla (miesięczna) oraz emisji CO<sub>2</sub> (tygodniowa), przeprowadzono uzupełnianie danych metodą interpolacji liniowej między sąsiednimi wartościami. Poniższy przykład dla granulacji tygodniowej.

$$\text{cena w dniu } d = \text{cena w tygodniu } t + \left( \frac{\text{cena w tygodniu } t+1 - \text{cena w tygodniu } t}{6} \right) \times d$$

, gdzie  $d \in [t, t+1]$

Następnie, podobnie jak w przypadku danych dziennych, wartości te przypisano każdej godzinie w danej dobie. Podejście to umożyliwiło ujednolicenie wszystkich zmiennych do godzinowej rozdzielczości danych docelowych, co jest niezbędnym krokiem dla poddania zbioru danych analizie.

## 4.2 Analiza korelacji

Aby zbadać zależności między zmiennymi objaśniającymi a zmienną docelową, przeprowadzono analizę korelacji, wykorzystując dwa współczynniki: Pearsona i Spearmana. Wybór odpowiedniego współczynnika dla każdej zmiennej oparto na charakterze jej relacji z `fixing_i_price`, co pozwoliło na bardziej precyzyjne oszacowanie siły i rodzaju zależności.

Współczynnik korelacji Pearsona ( $r$ ) mierzy liniową zależność między dwiema zmiennymi. Jest on zdefiniowany wzorem:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}},$$

gdzie  $x_i$  i  $y_i$  to wartości zmiennych i danej zmiennej objaśniającej),  $\bar{x}$  i  $\bar{y}$  to ich średnie, a  $n$  to liczba obserwacji. Współczynnik Pearsona przyjmuje wartości w przedziale  $[-1, 1]$ , gdzie  $r = 1$  oznacza doskonałą dodatnią zależność liniową,  $r = -1$  doskonałą ujemną zależność liniową, a  $r = 0$  brak liniowej zależności.

Pearson jest odpowiedni dla zmiennych, których relacja jest liniowa. Na przykład wzrost zapotrzebowania zwykle większa ceny energii liniowo.

Współczynnik korelacji Spearmana ( $\rho$ ) mierzy monotoniczną zależność między zmiennymi, co czyni go bardziej odpowiednim dla relacji nieliniowych. Spearman opiera się na rangach wartości zmiennych, a jego wzór to:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)},$$

gdzie  $d_i$  to różnica między rangami wartości  $x_i$  i  $y_i$ . Współczynnik Spearmana również przyjmuje wartości w przedziale  $[-1, 1]$ , ale nie zakłada liniowości relacji - wystarczy, że wzrost jednej zmiennej odpowiada wzrostowi (lub spadkowi) drugiej w sposób monotoniczny.

#### *Rozdział 4. Eksploracja danych*

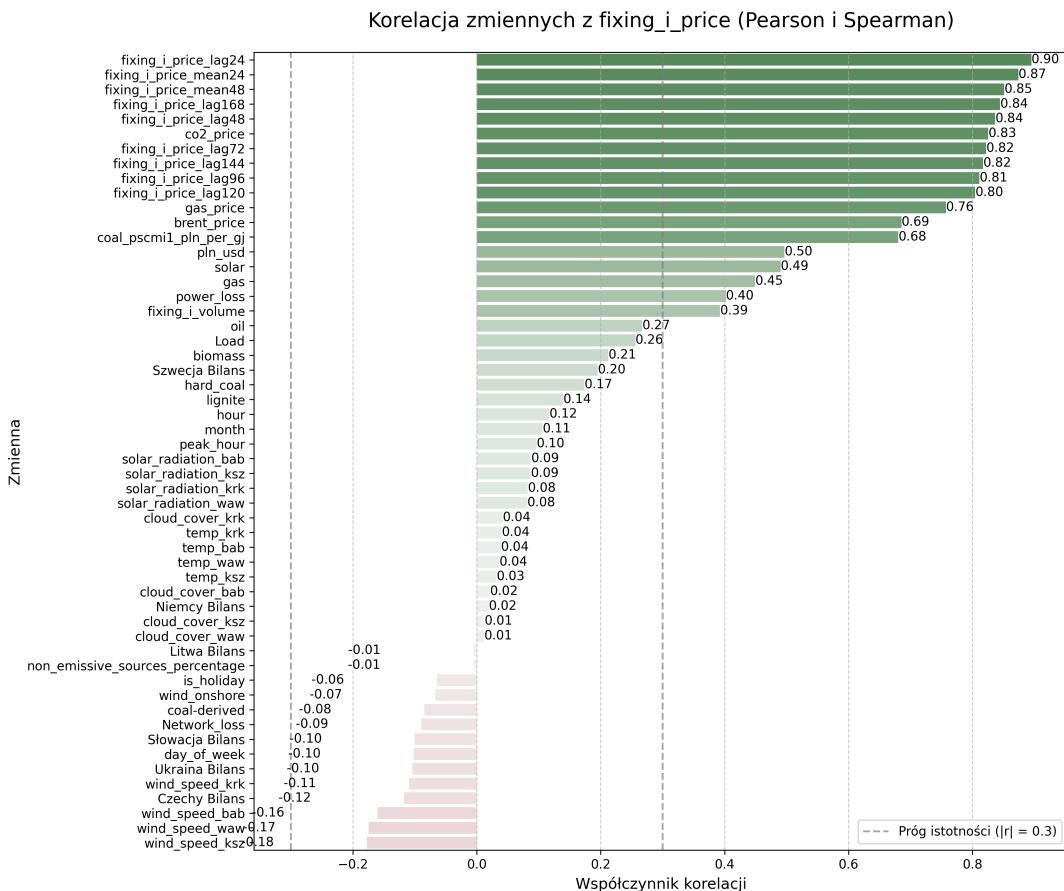
---

Spearman jest szczególnie użyteczny dla zmiennych o nieliniowej relacji z `fixing_i_price`. Na przykład wzrost cen gazu w 2022 roku prowadził do nieproporcjonalnego wzrostu cen energii, co lepiej oddaje Spearman niż Pearson.

W celu wyboru odpowiedniego współczynnika korelacji obliczono zarówno korelację Pearsona, jak i Spearmana dla wszystkich zmiennych objaśniających względem `fixing_i_price`. Następnie obliczono bezwzględną różnicę między tymi współczynnikami ( $|\rho - r|$ ). Zmienne, dla których różnica była większa niż 0,1, uznano za posiadające nieliniową relację z `fixing_i_price`, stosując dla nich korelację Spearmana. W pozostałych przypadkach wybrano korelację Pearsona, zakładając liniową zależność.

Analiza wykazała, że zmienne takie jak `co2_price` (różnica 0,114), `coal_pscmi1_pln_per_gj` (0,161), `pln_usd` (0,112), `solar` (0,172), `gas` (0,153), `biomass` (0,174), `coal-derived` (0,142) oraz zmienne związane z promieniowaniem słonecznym (różnice 0,106-0,115) mają nieliniową relację z `fixing_i_price`. Nieliniowość wynika z charakteru tych zmiennych - na przykład wysoka produkcja energii z fotowoltaiki w miesiącach letnich obniża ceny energii w sposób nieproporcjonalny, co prowadzi do bardzo niskich, a czasem ujemnych cen. Podobnie gwałtowny wzrost cen CO2 w okresie niespokojnym (2020-2023) zwiększa koszty produkcji energii w elektrowniach węglowych, ale efekt ten był wzmacniany przez inne czynniki, takie jak spekulacje rynkowe.

W wyniku tego powstał następujący wykres korelacji poniżej



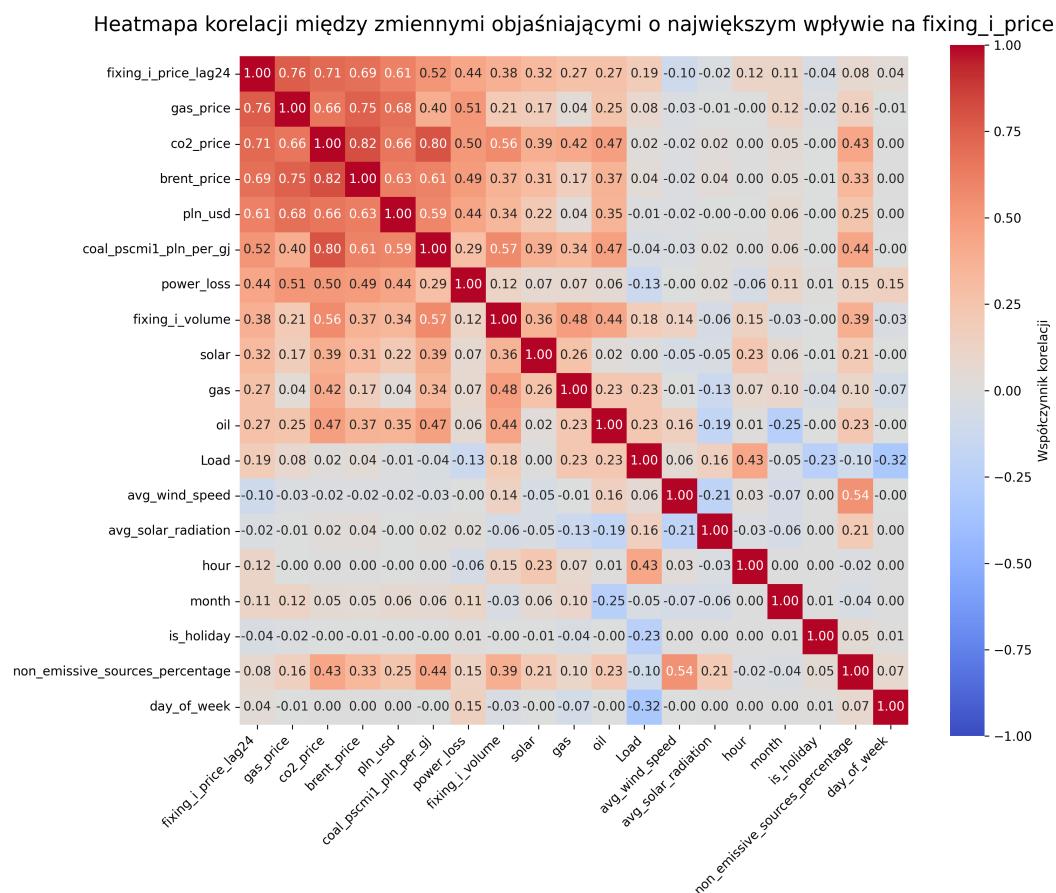
Rysunek 17. Wykres korelacji zmiennych objaśniających względem zmiennej docelowej fixing\_i\_price.

Zgodnie z literaturą, największy wpływ na ceny energii mają zmienne, które powstały w wyniku przetworzenia zmiennej objaśnianej, czyli ceny opóźnione i średnie. Koljenymi zmiennymi wykazującymi silną korelację są ceny na surowce, kurs polskiego złotego oraz zapotrzebowanie i volumen. Sprzeczna z logiką może być pozytywna korelacja ceny z zmienną solar, która wskazuje na produkcję energii z paneli fotowoltaicznych. Wynika to prawdopodobnie z faktu, w momentach, gdy świeci słońce i produkcja energii z OZE jest wysoka, zapotrzebowanie również jest zwiększone i to powoduje wzrost cen. Z kolei wartości reprezentujące parametry pogodowe nie wykazują istotnej korelacji.

Produkcja energii z odnawialnych źródeł i gazu również odgrywa rolę. Zmienna solar (Spearman: 0,490) wskazuje, że wysoka produkcja energii z fotowoltaiki obniża ceny energii, szczególnie w miesiącach letnich, gdzie nadpodaż energii z OZE może prowadzić do bardzo niskich cen. gas (Spearman: 0,449) pokazuje, że produkcja energii z gazu ma nieliniowy wpływ, zależny od cen gazu i dostępności innych źródeł energii. Ponadto zmienne takie jak fixing\_i\_volume (Spearman: 0,442), power\_loss (Spearman: 0,441), oil (Spearman: 0,349) oraz Load (Pearson: 0,256) mają umiarkowany wpływ, odzwierciedlając znaczenie volumenu obrotu, strat w sieci, produkcji z oleju oraz zapotrzebowania na energię.

## Zbiór danych skrócony

Na podstawie analizy korelacji stworzony został skrócony zbiór danych, który z pierwotnych 54 regresorów zostawia najbardziej istotne. Tymi zmiennymi zostały wszystkie przekraczające próg istotności na poziomie 0.2, zmienne sezonowe określające dzień tygodnia, miesiąc, godzinę oraz święta, średnia arytmetyczna ze zmiennych objaśniających prędkość wiatru w Polsce, gdyż wykazują one najmocniejszą odwrotną korelację oraz zmienna objaśniająca generację OZE w procentach, ponieważ jest to ważna zmienna z punktu widzenia literatury. W wyniku tego powstał zbiór danych z 27 zmiennymi objaśniającymi. Taki zbiór danych został określony jako zbiór danych skrócony i będzie użyty w dalszej części pracy do analizy w celu sprawdzenia istotności zbioru danych o największej ilości parametrów. Poniżej przedstawiono mapę cieplną dla zmiennych objaśniających w zbiorze danych skróconym bez uwzględnienia zmiennych opóźnionych i średnich.



Rysunek 18. Mapa cieplna korelacji zmiennych objaśniających w skróconym zbiorze danych.

Z powyższej mapy cieplnej wynika, że zmienne o największej korelacji z ceną energii w dużym stopniu są również skorelowane pomiędzy sobą.

## 4.3 Podział danych

### Podział na okresy spokojny i niespokojny

Dane zostały podzielone na dwa okresy w celu uwzględnienia różnych warunków rynkowych i ich wpływu na ceny energii. Okres spokojny (2016-2019) charakteryzuje się stabilnymi cenami energii, wynikającymi z braku znaczących szoków podażowych, łatwiej przewidywalnych cen paliw oraz łagodnego wzrostu cen CO<sub>2</sub> w ramach polityki klimatycznej UE. W tym okresie nie występowały większe kryzysy geopolityczne ani pandemie, co pozwoliło na utrzymanie cen w stosunkowo wąskim zakresie.

Okres niespokojny (2020-2023) został zdominowany przez szereg wydarzeń, które drastycznie wpłynęły na rynek energii. Pandemia COVID-19 w latach 2020-2021 początkowo obniżyła zapotrzebowanie na energię, ale ożywienie gospodarcze w 2021 roku spowodowało gwałtowny wzrost cen. Kryzys energetyczny w latach 2021-2022, związany z ograniczoną podażą gazu, rekordowymi cenami CO<sub>2</sub> oraz wysokimi cenami węgla, doprowadził do ekstremalnych skoków cen energii. Wybuch wojny na Ukrainie w 2022 roku dodatkowo zaostrzył sytuację, powodując przerwanie dostaw gazu z Rosji, sankcje i spekulacje rynkowe, co przełożyło się na rekordowe ceny energii. W tym okresie pojawiły się również ujemne ceny, wynikające z nadpodaży energii z OZE i ograniczonej elastyczności systemu elektroenergetycznego.

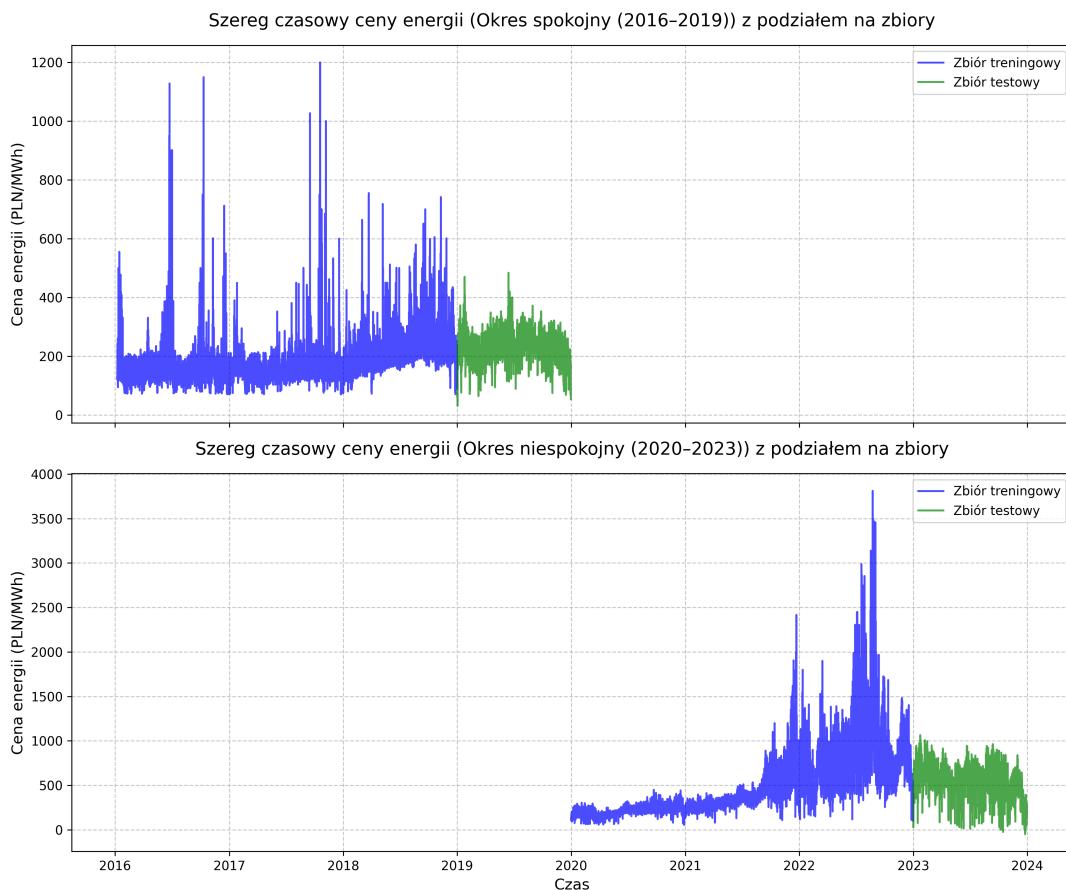
Statystyki opisowe dla obu okresów przedstawiono w tabeli 5. Okres spokojny charakteryzuje się niższą średnią ceną, mniejszą zmiennością i dużo mniejszym odchyleniem standardowym, co odzwierciedla stosunkowo stabilne warunki rynkowe. W okresie niespokojnym średnia cena wzrosła, a współczynnik zmienności i odchylenie znaczco rosną. W latach 2020-2023 pojawiły się również ujemne ceny oraz rekordowe maksima. Podział na te dwa okresy i poddanie ich osobnej analizie pozwoli lepiej ocenić skuteczność wybranych cech i modeli w różnych warunkach rynkowych.

**Tabela 5.** Porównanie statystyk opisowych cen energii w okresach spokojnym (2016-2019) i niespokojnym (2020-2023).

Miara	Okres spokojny	Okres niespokojny
Średnia (PLN/MWh)	193,51	478,06
Mediana (PLN/MWh)	182,00	412,00
Odchylenie standardowe (PLN/MWh)	70,67	321,39
Współczynnik zmienności (%)	36,52	67,23
Kwartyl Q1 (25%) (PLN/MWh)	143,66	246,41
Kwartyl Q3 (75%) (PLN/MWh)	229,35	609,00
Minimum (PLN/MWh)	31,00	-50,00
Maksimum (PLN/MWh)	1199,53	3812,45
Procent dni z ceną powyżej 500 PLN/MWh (%)	0,41	36,94

### Podział danych na zbiory treningowe i testowe

Dane zostały podzielone na zbiory treningowe i testowe w obrębie każdego z dwóch okresów, aby uwzględnić strukturę szeregów czasowych. Podział został przeprowadzony sekwencyjnie w proporcji 75/25, co zapewnia dużą ilość danych do treningu (3 lata) oraz odpowiednią ilość danych do testowania (1 rok). W przypadku takiego podziału w zbiorach testowych można przetestować wszystkie okresy sezonowe.



Rysunek 19. Podział szeregów czasowych cen energii na zbiory treningowe i testowe

## 4.4 Przygotowanie danych

Przed przystąpieniem do modelowania dane zostały poddane szeregu kroków preprocessingu, aby zapewnić ich odpowiednią jakość i format dla wybranych modeli.

Pierwszym krokiem jest kodowanie zmiennych cyklicznych. Zmienne sezonowe mają charakter cykliczny (np. po godzinie 23 następuje 0, po grudniu następuje styczeń). Aby uwzględnić tę cykliczność, zastosowano kodowanie za pomocą funkcji sinusoidalnych:

- Dla `day_of_week`:  $\sin\left(\frac{2\pi \cdot \text{day\_of\_week}}{7}\right)$ ,

- Dla month:  $\sin\left(\frac{2\pi \cdot \text{month}}{12}\right)$ ,
- Dla hour:  $\sin\left(\frac{2\pi \cdot \text{hour}}{24}\right)$ .

Oryginalne zmienne zostały usunięte, a ich zakodowane wersje dodano do zbioru danych. Kodowanie sinusoidalne pozwala modelom lepiej uchwycić cykliczność danych, w przeciwieństwie do kodowania typu one-hot, które zwiększyłoby wymiarowość danych i nie uwzględniało cykliczności.

Następnie, wszystkie zmienne numeryczne niebinarne zostały poddane standaryzacji StandardScaler z biblioteki sklearn. Standaryzacja polega na przekształceniu zmiennych, aby miały średnią 0 i odchylenie standardowe 1. Wartości zmiennych zostały przekształcone według wzoru:

$$z = \frac{x - \mu}{\sigma},$$

gdzie  $z$  to wartość po standaryzacji,  $x$  to wartość przed standaryzacją,  $\mu$  to średnia zmiennej, a  $\sigma$  to odchylenie standardowe.

Standaryzacja jest kluczowym krokiem w preprocessingu danych, ponieważ większość algorytmów uczenia maszynowego zakłada, że dane mają podobną skalę. W przeciwnym razie algorytmy mogą być wrażliwe na różnice w skali zmiennych, co prowadzi do nieoptimalnych wyników. Proces standaryzacji został przeprowadzony osobno dla każdego z okresów spokojnego i niespokojnego osobno, aby uwzględnić różnice w rozkładach danych między tymi okresami. W obrębie każdego okresu parametry standaryzacji (średnia i odchylenie standardowe) obliczono na zbiorze treningowym i zastosowano zarówno do danych treningowych, jak i testowych, aby uniknąć wycieku informacji.

Wartości odstające (np. ekstremalnie wysokie ceny w okresie niespokojnym, takie jak 3812,45 PLN/MWh) nie zostały zmodyfikowane, ponieważ odzwierciedlają rzeczywiste zjawiska rynkowe (np. kryzys energetyczny w 2022 roku). Ich wpływ na modele będzie monitorowany podczas analizy wyników.

## Rozdział 5

# Metodologia

W tym rozdziale przedstawiono metodologię przeprowadzonych badań. Rozdział składa się z dwóch sekcji. W pierwszej przedstawiono metodykę oceny jakości prognoz, a w drugiej omówiono metodykę prognozowania cen energii elektrycznej.

### 5.1 Ocena jakości prognoz

Ocena jakości modeli prognozowania cen energii elektrycznej jest kluczowym etapem analizy, ponieważ pozwala na porównanie skuteczności różnych podejść. W niniejszej pracy zastosowano następujące popularne metryki oceny: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (sMAPE) oraz  $R^2$ . Wszystkie z tych metryk są omawiane w literaturze porównania skuteczności modeli [3]. W pracy prof. Werona [14] podano, że nie ma standardu obliczenia metryk EPF i wspomina o innych metrykach stosowanych przez innych autorów, między innymi wymieniono - Ważony Średni Błąd Bezwzględny (WMAE), średni błąd dniowy (MDE) i tygodniowy (MWE). Niemniej jednak, w tej pracy skupiono się na tych najszerzej stosowanych metrykach. Każda z tych metryk ma swoje zalety i ograniczenia, których omówienie jest przedstawione poniżej, wraz z ich matematycznymi definicjami i przykładami zastosowania w EPF.

#### 5.1.1 Mean Absolute Error (MAE)

Mean Absolute Error jest jedną z najprostszych i najczęściej stosowanych metryk w prognozowaniu szeregów czasowych, w tym w EPF. MAE mierzy średnią wartość bezwzględnych błędów prognoz, co pozwala na ocenę dokładności modelu bez uwzględniania kierunku błędu (nad-lub niedoszacowania).

Matematyczna definicja MAE jest następująca:

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|$$

gdzie:

- $y_t$  to rzeczywista cena energii w godzinie  $t$ ,
- $\hat{y}_t$  to przewidywana cena energii w godzinie  $t$ ,
- $n$  to liczba obserwacji w zbiorze testowym.

MAE jest wyrażane w tej samej jednostce co prognozowane wartości (w omawianym przypadku jest to PLN/MWh), co łatwo interpretować. Na przykład, jeśli MAE wynosi 10 PLN/MWh, oznacza to, że średni błąd prognozy wynosi 10 PLN na każdą megawatogodzinę.

#### Zalety MAE:

- Prosta interpretacja i obliczenia.
- Równomierne traktowanie wszystkich błędów, niezależnie od ich kierunku.

#### Ograniczenia MAE:

- Nie uwzględnia kwadratu błędów, przez co nie penalizuje większych odchyleń w sposób szczególny, co może być problematyczne w EPF, gdzie duże skoki cen (np. w godzinach szczytu) są istotne.

### 5.1.2 Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) jest kolejną popularną metryką w EPF, która uwzględnia kwadrat błędów, co powoduje większy wpływ dużych odchyleń między wartościami rzeczywistymi a przewidywanymi. RMSE jest szczególnie użyteczne w sytuacjach, gdzie duże błędy prognoz mogą mieć poważne konsekwencje ekonomiczne.

Definicja RMSE jest następująca:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$$

gdzie:

- $y_t$ ,  $\hat{y}_t$  i  $n$  mają takie same znaczenie jak w MAE.

RMSE jest również wyrażane w jednostkach oryginalnych danych, co ułatwia interpretację. Na przykład, RMSE równe 15 PLN/MWh oznacza, że typowy błąd prognozy w sensie średniego kwadratu wynosi 15 PLN na megawatogodzinę.

#### Zalety RMSE:

- Większa wrażliwość na duże błędy, co jest istotne w EPF, gdzie skoki cen mogą być kosztowne.

#### Ograniczenia RMSE:

- Wrażliwość na wartości odstające - pojedyncze duże błędy mogą znaczco zawyżyć wartość RMSE.
- Mniej intuicyjne w interpretacji niż MAE, ponieważ kwadrat błędów zmienia skalę.

### 5.1.3 Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) jest metryką wyrażającą błąd prognozy jako procent rzeczywistej wartości, co czyni ją szczególnie użyteczną w porównaniach między różnymi zbiorami danych lub rynkami o różnych poziomach cen.

Definicja MAPE jest następująca:

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100$$

gdzie:

- $y_t$ ,  $\hat{y}_t$  i  $n$  mają takie same znaczenie jak wcześniej.

MAPE jest wyrażane w procentach, co ułatwia interpretację. Na przykład, MAPE równe 5% oznacza, że średni błąd prognozy wynosi 5% rzeczywistej ceny. W kontekście RDN, jeśli cena energii wynosi 200 PLN/MWh, a MAPE wynosi 5%, średni błąd wynosi 10 PLN/MWh.

**Zalety MAPE:**

- Intuicyjna interpretacja w procentach, nie trzeba zastanawiać się nad jednostkami bądź kursami walutowymi.

**Ograniczenia MAPE:**

- Problemy z wartościami bliskimi zera - jeśli  $y_t$  jest bardzo małe, co jest możliwe w godzinach nocnych, dzielenie przez  $y_t$  prowadzi do bardzo dużych wartości procentowych, a nawet do matematycznego błędu dzielenia przez zero.
- Asymetria - MAPE bardziej penalizuje niedoszacowania niż przeszacowania, co może prowadzić do nieobiektywnej oceny.

### 5.1.4 Symmetric Mean Absolute Percentage Error (sMAPE)

Symmetric Mean Absolute Percentage Error (sMAPE) jest zmodyfikowaną wersją MAPE, która rozwiązuje problem asymetrii i dzielenia przez zero. sMAPE uwzględnia zarówno rzeczywiste, jak i przewidywane wartości w mianowniku, co czyni ją bardziej stabilną w sytuacjach, gdy ceny energii są niskie.

Definicja sMAPE jest następująca:

$$\text{sMAPE} = \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{(|y_t| + |\hat{y}_t|) / 2} \times 100$$

gdzie:

- $y_t$ ,  $\hat{y}_t$  i  $n$  mają takie same znaczenie jak wcześniej.

Podobnie jak MAPE, sMAPE jest wyrażane w procentach. Na przykład, sMAPE równe 4% oznacza, że średni błąd symetryczny wynosi 4% średniej wartości rzeczywistej i przewidywanej ceny.

**Zalety sMAPE:**

- Rozwiązuje problem dzielenia przez zero, co jest istotne w EPF, gdzie ceny mogą być bliskie zera.
- Symetria - traktuje nad-i niedoszacowania w bardziej zrównoważony sposób niż MAPE.

**Ograniczenia sMAPE:**

- Nadal może być wrażliwe na skrajne wartości, choć w mniejszym stopniu niż MAPE.
- Interpretacja jest mniej intuicyjna niż w przypadku MAE czy RMSE, ponieważ uwzględnia zarówno  $y_t$ , jak i  $\hat{y}_t$  w mianowniku.

### 5.1.5 Współczynnik determinacji

Współczynnik determinacji, oznaczany jako  $R^2$ , jest metryką powszechnie stosowaną w analizie regresji i prognozowaniu.  $R^2$  mierzy, jak dobrze model wyjaśnia zmienność danych rzeczywistych, czyli jaki procent wariancji zmiennej zależnej jest wyjaśniony przez model progностyczny. Jest to metryka szczególnie użyteczna w ocenie modeli liniowych, ale znajduje zastosowanie również w bardziej złożonych modelach, w celu ogólnej oceny ich dopasowania do danych.

Definicja  $R^2$  jest następująca:

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2}$$

gdzie:

- $y_t$ ,  $\hat{y}_t$  i  $n$  mają takie same znaczenie jak wcześniej,
- $n$  to liczba obserwacji w zbiorze testowym.

Licznik w wyrażeniu  $\sum_{t=1}^n (y_t - \hat{y}_t)^2$  to suma kwadratów reszt, czyli całkowity błąd modelu, natomiast mianownik  $\sum_{t=1}^n (y_t - \bar{y})^2$  to całkowita suma kwadratów, czyli całkowita wariancja danych względem ich średniej.  $R^2$  przyjmuje wartości w przedziale od 0 do 1, gdzie:

- $R^2 = 1$  oznacza, że model idealnie przewiduje wszystkie wartości (błąd wynosi 0),
- $R^2 = 0$  oznacza, że model nie wyjaśnia żadnej zmienności danych i jest równoważny prostemu modelowi średniej ( $\hat{y}_t = \bar{y}$ ).

W kontekście EPF, na przykład na RDN,  $R^2$  równe 0,85 oznaczałoby, że model wyjaśnia 85% zmienności cen energii.

**Zalety  $R^2$ :**

- Intuicyjna interpretacja -  $R^2$  jasno wskazuje, jaki procent zmienności danych jest wyjaśniony przez model.
- Bez jednostek - umożliwia porównanie modeli na różnych zbiorach danych, niezależnie od skali cen (np. PLN/MWh na RDN vs. EUR/MWh na EEX).

**Ograniczenia  $R^2$ :**

- Wrażliwość na przeuczenie -  $R^2$  może być zawyżone w modelach o dużej liczbie parametrów, szczególnie w przypadku małych zbiorów danych, co może prowadzić do mylnego wniosku o dobrym dopasowaniu modelu.

- Brak informacji o kierunku błędów -  $R^2$  nie rozróżnia, czy model nad-czy niedoszacowuje wartości, co w EPF może być istotne z ekonomicznego punktu widzenia.

W niniejszej pracy  $R^2$  zostanie wykorzystane jako dodatkowa metryka oceny, aby uzupełnić analizę opartą na MAE, RMSE, MAPE, sMAPE.

## 5.2 Wybrane metody weryfikacji zbioru danych

Stworzony zbiór danych z cechami objaśniającymi ceny energii elektrycznej należy zweryfikować pod kątem jego skuteczności. W związku z tym zostały wybrane cztery metody prognozowania.

### 5.2.1 Regresja liniowa

#### Opis metody

Regresja liniowa jest jednym z najprostszych i najczęściej stosowanych modeli statystycznych w analizie zbiorów danych. Zakłada liniową zależność między zmienną zależną, a zestawem zmiennych niezależnych (predyktorów). W kontekście EPF regresja liniowa jest często stosowana jako model bazowy, który pozwala na szybkie uzyskanie prognoz i ocenę wpływu poszczególnych zmiennych na ceny energii. Jej zaletą jest prostota interpretacji oraz niski koszt obliczeniowy, co czyni ją odpowiednią do analizy dużych zbiorów danych, co czyni ją odpowiednią do tej pracy.

#### Wzór modelu

Model regresji liniowej można zapisać jako:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon \quad (1)$$

gdzie:

- $y$  - zmienna zależna, cena energii elektrycznej
- $\beta_0$  - wyraz wolny,
- $\beta_1, \beta_2, \dots, \beta_p$  - współczynniki regresji dla zmiennych niezależnych,
- $x_1, x_2, \dots, x_p$  - zmienne niezależne (predyktory, np. zmienne związane z zapotrzebowaniem, cenami paliw czy danymi kalendarzowymi),
- $\epsilon$  - składnik losowy (błąd), zakładany jako  $\epsilon \sim N(0, \sigma^2)$ .

W macierzowej formie model przyjmuje postać:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2)$$

gdzie:

- $\mathbf{y}$  - wektor obserwacji zmiennej zależnej,
- $\mathbf{X}$  - macierz projektowa zawierająca wartości zmiennych niezależnych,
- $\boldsymbol{\beta}$  - wektor współczynników regresji,
- $\boldsymbol{\epsilon}$  - wektor błędów.

### Estymacja parametrów

Parametry modelu  $\beta$  są estymowane za pomocą metody najmniejszych kwadratów (OLS), która minimalizuje sumę kwadratów błędów:

$$\min \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

gdzie  $\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$  to przewidywana wartość dla  $i$ -tej obserwacji. Rozwiążanie analityczne to:

$$\beta = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (4)$$

### Istotne parametry modelu

W niniejszej pracy regresja liniowa została zaimplementowana za pomocą biblioteki scikit-learn w Pythonie. Kluczowe parametry modelu obejmują:

- `fit_intercept=True`: Włączenie wyrazu wolnego ( $\beta_0$ ).
- `normalize=False`: Brak normalizacji zmiennych przed estymacją.
- `solver='auto'`: Automatyczny wybór algorytmu estymacji, domyślnie OLS.

### Zalety i ograniczenia

Regresja liniowa jest łatwa do interpretacji, ponieważ współczynniki  $\beta_j$  wskazują, o ile zmieni się cena energii przy wzroście zmiennej  $x_j$  o jednostkę (przy założeniu stałości pozostałych zmiennych). Jednak model zakłada liniowe zależności między zmiennymi, co może być ograniczeniem w przypadku bardziej złożonych, nieliniowych wzorców w danych cen energii, szczególnie w okresie niespokojnym.

## 5.2.2 Regresja grzbietowa (Ridge)

### Opis metody

Regresja grzbietowa (ang. Ridge Regression) jest rozszerzeniem regresji liniowej, które wprowadza regularyzację L2, aby zapobiec przeuczeniu i poprawić stabilność modelu w przypadku współliniowości między zmiennymi wyjaśniającymi. W prognozowaniu cen energii regresja grzbietowa jest szczególnie użyteczna, gdy zestaw danych zawiera wiele zmiennych, które mogą być skorelowane. Regularizacja pozwala na zmniejszenie wpływu mniej istotnych zmiennych, co poprawia generalizację modelu.

### Wzór modelu

Model regresji grzbietowej opiera się na tej samej zależności liniowej co regresja liniowa:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon \quad (5)$$

Jednak estymacja parametrów uwzględnia dodatkową karę regularizacyjną L2. Funkcja kosztu w regresji grzbietowej to:

$$\min \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p \beta_j^2 \quad (6)$$

gdzie:

- Pierwsza część ( $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ ) to suma kwadratów błędów, jak w OLS.
- Druga część ( $\lambda \sum_{j=1}^p \beta_j^2$ ) to kara L2 na wielkość współczynników  $\beta_j$ .
- $\lambda \geq 0$  - parametr regularizacji, który kontroluje siłę kary (większe  $\lambda$  oznacza silniejszą regularizację).

W macierzowej formie funkcja kosztu to:

$$\min \| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \|_2^2 + \lambda \|\boldsymbol{\beta}\|_2^2 \quad (7)$$

Rozwiążanie analityczne dla parametrów to:

$$\boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y} \quad (8)$$

gdzie  $\mathbf{I}$  to macierz jednostkowa.

### Istotne parametry modelu

Regresja grzbietowa została zaimplementowana w Pythonie za pomocą biblioteki scikit-learn.

Kluczowe parametry modelu to:

- `alpha=1.0`: Domyślna wartość parametru regularizacji  $\lambda$ . W pracy metodą empiryczną spróbuje się różnych wartości `alpha` (np. 0.1, 1.0, 10.0, 100.0) za pomocą walidacji krzyżowej, aby wybrać optymalną.
- `fit_intercept=True`: Włączenie wyrazu wolnego ( $\beta_0$ ).
- `normalize=False`: Brak normalizacji zmiennych przed estymacją (zmienne przeskalowano wcześniej za pomocą StandardScaler).
- `solver='auto'`: Automatyczny wybór algorytmu (domyślnie Cholesky dla małych zbiorów danych lub SAG dla dużych).

### Zalety i ograniczenia

Regresja grzbietowa jest bardziej odporna na współliniowość i przeuczenie od regresji liniowej, co czyni ją odpowiednią do zestawów danych z dużą liczbą zmiennych objaśniających. Jest to szczególnie przydatne, gdyż w pracy uwzględnione zostają parametry temperatury z całej Polski, które zdecydowanie mają korelację. Jednak, podobnie jak regresja liniowa, zakłada liniowe zależności, co może ograniczać jej skuteczność w modelowaniu bardziej złożonych wzorców, szczególnie w niestabilnych okresach rynkowych.

## 5.2.3 Prophet

### Opis metody

Prophet [12] to model prognozowania szeregów czasowych opracowany przez Facebooka, zaprojektowany do analizy danych z wyraźną sezonowością i trendami, które mogą ulegać zmianom w czasie. W kontekście prognozowania cen energii elektrycznej (EPF) Prophet jest szczególnie ciekawy ze względu na zdolność do modelowania cyklicznych wzorców jakie występują na tym rynku oraz uwzględniania efektów specjalnych, takich jak święta. Model jest oparty na addytywnym

podejściu, które rozkłada szereg czasowy na składowe trendu, sezonowości i efektów dodatkowych. Jego intuicyjna parametryzacja i możliwość automatycznego dopasowania do danych czynią go atrakcyjnym narzędziem w analizie dużych zbiorów danych, takich jak te wykorzystane w niniejszej pracy.

### **Wzór modelu**

Prophet modeluje zmienną zależną jako sumę trzech głównych składowych plus składnik losowy:

$$y(t) = g(t) + s(t) + h(t) + r(t) + \epsilon_t \quad (9)$$

gdzie:

- $y(t)$  - wartość prognozowana,
- $g(t)$  - składowa trendu, modelująca długoterminowe zmiany w danych,
- $s(t)$  - składowa sezonowości, modelująca cykliczne wzorce (np. dobowe, tygodniowe),
- $h(t)$  - składowa efektów specjalnych takich jak święta,
- $r(t)$  - składowa zmiennych objaśniających, uwzględniająca wpływ dodatkowych regresorów, takich jak zapotrzebowanie czy dane pogodowe,
- $\epsilon_t$  - składnik losowy (błąd), zakładany jako  $\epsilon_t \sim N(0, \sigma^2)$ .

### **Składowa trendu ( $g(t)$ )**

Trend w modelu Prophet jest modelowany za pomocą nieliniowej funkcji z punktami zmiany (ang. changepoints), które pozwalają na elastyczne dopasowanie do nagłych zmian w danych. Standardowo używa się funkcji liniowej z punktami zmiany:

$$g(t) = (k + \mathbf{a}(t)^T \boldsymbol{\delta})t + (m + \mathbf{a}(t)^T \boldsymbol{\gamma}) \quad (10)$$

gdzie:

- $k$  - współczynnik nachylenia trendu,
- $m$  - wyraz wolny,
- $\mathbf{a}(t)$  - wektor binarny wskazujący punkty zmiany,
- $\boldsymbol{\delta}$  - wektor zmian nachylenia w punktach zmiany,
- $\boldsymbol{\gamma}$  - wektor przesunięć dla ciągłości trendu w punktach zmiany.

### **Składowa sezonowości ( $s(t)$ )**

Sezonowość jest modelowana za pomocą szeregu Fouriera, który aproksymuje cykliczne wzorce:

$$s(t) = \sum_{n=1}^N \left( a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right) \quad (11)$$

gdzie:

- $P$  - okres sezonowości (np. 24 godziny dla sezonowości dobowej, 168 godzin dla tygodniowej),
- $a_n, b_n$  - współczynniki szeregu Fouriera,
- $N$  - liczba składników szeregu (kontrolowana przez parametr `fourier_order`).

### Składowa efektów specjalnych ( $h(t)$ )

Efekty specjalne, takie jak święta, są modelowane jako:

$$h(t) = \mathbf{Z}(t)\boldsymbol{\kappa} \quad (12)$$

gdzie:

- $\mathbf{Z}(t)$  - macierz binarna wskazująca wystąpienie efektów specjalnych (np. 1 dla dni świątecznych, 0 w pozostałych),
- $\boldsymbol{\kappa}$  - wektor efektów dla każdego zdarzenia.

### Składowa zmiennych objaśniających ( $r(t)$ )

Zmienne objaśniające, takie jak zapotrzebowanie, dane pogodowe czy bilanse handlowe, są uwzględniane jako dodatkowa już znana składowa liniowa:

$$r(t) = \beta_1 x_1(t) + \beta_2 x_2(t) + \dots + \beta_p x_p(t) \quad (13)$$

### Estymacja parametrów

Parametry modelu ( $k, m, \delta, \gamma, a_n, b_n, \boldsymbol{\kappa}, \beta_1, \dots, \beta_p$ ) są estymowane za pomocą maksymalizacji funkcji wiarogodności lub metod bayesowskich. Prophet wykorzystuje algorytm L-BFGS do optymalizacji w trybie domyślnym, co zapewnia szybkie dopasowanie modelu. Punkty zmiany są automatycznie wykrywane, a ich liczba i rozmieszczenie są kontrolowane przez parametry modelu, takie jak `n_changepoints` i `changepoint_prior_scale`.

### Istotne parametry modelu

W niniejszej pracy model Prophet został zaimplementowany w Pythonie za pomocą biblioteki `prophet`. Dane wejściowe zostały przygotowane w formacie wymaganym przez Prophet, gdzie kolumna `ds` zawiera znaczniki czasowe (`timestamp`), a kolumna `y` zawiera ceny energii (`fixing_i_price`). Kluczowe parametry modelu to:

- `n_changepoints`: Liczba punktów zmiany trendu, umożliwiająca dopasowanie do potencjalnych zmian w danych (domyślana wartość 25).
- `changepoint_prior_scale`: Siła regularyzacji punktów zmiany (domyślana wartość 0.05).
- `yearly_seasonality=False`: Wyłączenie sezonowości rocznej, ponieważ dane cen energii wykazują głównie sezonowość dobową i tygodniową.
- `weekly_seasonality=True`: Włączenie sezonowości tygodniowej.
- `daily_seasonality=True`: Włączenie sezonowości dobowej.
- `fourier_order`: Liczba składników szeregu Fouriera dla każdej sezonowości (domyślana wartość 10).
- `holidays`: Włączono efekty dni świątecznych na podstawie zmiennej `is_holiday` z danych.
- Dodatkowe regresory: W przypadku pełnego zestawu danych wszystkie zmienne objaśniające, takie jak zapotrzebowanie, dane pogodowe, bilanse handlowe czy ceny paliw, zostały dodane za pomocą funkcji `add_regressor`.

### Zalety i ograniczenia

Prophet jest intuicyjny i dobrze radzi sobie z danymi o wyraźnej sezonowości, co czyni go odpowiednim do modelowania cen energii w stabilnych okresach. Automatyczne wykrywanie punktów zmiany, obsługa efektów specjalnych oraz możliwość włączenia wszystkich zmiennych objaśniających ułatwiają jego stosowanie w praktyce. Jednak model może mieć trudności z modelowaniem bardzo dużych wahań cen, takich jak te obserwowane w okresie niespokojnym (2020-2023), szczególnie jeśli zmienne objaśniające nie w pełni tłumaczą zmienność. Ponadto Prophet zakłada addytywną strukturę szeregu czasowego, co może ograniczać jego zdolność do wychwytywania bardziej złożonych, nieliniowych zależności.

#### 5.2.4 Wielowarstwowy perceptron (MLP)

##### Opis metody

Wielowarstwowy perceptron (MLP) to rodzaj sztucznej sieci neuronowej wykorzystywany w zadaniach uczenia maszynowego. Składa się z warstw neuronów: wejściowej, ukrytych i wyjściowej, które są w pełni połączone.

##### Wzór modelu

MLP przekształca wektor zmiennych wejściowych  $\mathbf{x} = [x_1, x_2, \dots, x_p]^T$  w wartość prognozowaną  $\hat{y}$  (cenę energii, oznaczaną w pracy jako `fixing_i_price`) poprzez sekwencję warstw neuronów. Dla sieci z jedną warstwą ukrytą model można zapisać jako:

$$\hat{y} = f_o(\mathbf{w}_o^T \mathbf{h} + b_o) \quad (14)$$

gdzie:

- $\mathbf{h} = f_h(\mathbf{W}_h \mathbf{x} + \mathbf{b}_h)$  - wektor aktywacji warstwy ukrytej,
- $\mathbf{x} = [x_1, x_2, \dots, x_p]^T$  - wektor zmiennych objaśniających,
- $\mathbf{W}_h$  - macierz wag między warstwą wejściową a ukrytą,
- $\mathbf{b}_h$  - wektor biasów warstwy ukrytej,
- $f_h(\cdot)$  - funkcja aktywacji warstwy ukrytej (np. `tanh`),
- $\mathbf{w}_o$  - wektor wag między warstwą ukrytą a wyjściową,
- $b_o$  - bias warstwy wyjściowej,
- $f_o(\cdot)$  - funkcja aktywacji warstwy wyjściowej (dla regresji zazwyczaj liniowa, tj.  $f_o(z) = z$ ).

Dla sieci z wieloma warstwami ukrytymi proces jest analogiczny, z kolejnymi przekształceniemi dla każdej warstwy:

$$\mathbf{h}_k = f_k(\mathbf{W}_k \mathbf{h}_{k-1} + \mathbf{b}_k), \quad k = 1, 2, \dots, K \quad (15)$$

gdzie  $\mathbf{h}_0 = \mathbf{x}$ ,  $K$  to liczba warstw ukrytych, a  $\mathbf{h}_K$  to wejście do warstwy wyjściowej.

W części analitycznej pracy są przedstawione wyniki z różną ilością warstw ukrytych.

##### Estymacja parametrów

Parametry modelu ( $\mathbf{W}_k, \mathbf{b}_k$  dla każdej warstwy oraz  $\mathbf{w}_o, b_o$ ) są estymowane przez minimalizację

funkcji kosztu, czyli średniego błędu kwadratowego (MSE):

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (16)$$

gdzie  $y_i$  to rzeczywista cena energii, a  $\hat{y}_i$  to przewidywana wartość dla  $i$ -tej obserwacji.

Estymacja parametrów odbywa się za pomocą algorytmu wstecznej propagacji błędu (backpropagation) w połączeniu z optymalizatorem, takim jak Adam. Proces treningu polega na iteracyjnym dostosowywaniu wag i biasów w celu zmniejszenia błędu na zbiorze treningowym, z uwzględnieniem walidacji na oddzielnym zbiorze danych w celu uniknięcia przeuczenia.

### Istotne parametry modelu

W niniejszej pracy model MLP został zaimplementowany w Pythonie za pomocą modułu Keras z biblioteki TensorFlow. Kluczowe parametry modelu, wraz z ich domyślnymi wartościami w bibliotece Keras, obejmują:

- **units:** Liczba neuronów w każdej warstwie ukrytej.
- **activation:** Funkcja aktywacji dla warstw ukrytych (domyślnie `relu` dla warstw gęstych).
- **optimizer:** Algorytm optymalizacji (domyślnie `rmsprop` dla modelu sekwencyjnego).
- **learning\_rate:** Szybkość uczenia dla optymalizatora (domyślnie 0.001 dla optimizera Adam).
- **batch\_size:** Rozmiar partii danych w każdej iteracji treningu (domyślnie 32 w metodzie `fit`).
- **epochs:** Maksymalna liczba epok treningu.

### Zalety i ograniczenia

MLP jest elastycznym modelem zdolnym do wychwytywania nieliniowych zależności w danych, co czyni go odpowiednim do modelowania cen energii w okresach o dużej zmienności. Możliwość dostosowania architektury sieci i hiperparametrów pozwala na optymalizację modelu pod kątem specyfiki danych. Jednak MLP wymaga starannego doboru hiperparametrów i preprocessingu danych, a jego trening jest bardziej kosztowny obliczeniowo niż w przypadku modeli statystycznych. Ponadto model może być podatny na przeuczenie, jeśli liczba warstw lub neuronów jest zbyt duża w stosunku do dostępnych danych.

# Rozdział 6

## Analiza danych

W niniejszym rozdziale przedstawiono wyniki modelowania cen energii za pomocą czterech modeli: regresji liniowej, regresji Ridge, Propheta oraz MLP. Analiza została podzielona na dwa podrozdziały, odpowiadające okresom stabilnemu i niestabilnemu. Zbiór danych został poddany testom w celu oceny skuteczności. W związku z tym dla każdego modelu przeprowadzono strojenie hiperparametrów.

### 6.1 Okres stabilny

#### Regresja liniowa i Ridge

Przeanalizowano modele regresji liniowej oraz regresji Ridge na danych z okresu stabilnego. Modele trenowane na danych z lat 2016–2018, a testowano na danych z 2019 roku.

Wyniki dla pełnego zbioru o 56 parametrach wejściowych przedstawiono w tabeli 6. Regresja Ridge osiągnęła lepsze wyniki niż regresja liniowa, co jest widoczne w niższych wartościach MAE (16.07 vs. 16.41), RMSE (20.62 vs. 21.19) oraz MAPE (7.96% vs. 8.17%). Wynika to z regularizacji L2 zastosowanej w modelu Ridge.

**Tabela 6.** Wyniki regresji liniowej i Ridge dla pełnego zbioru danych w okresie stabilnym (2019).

Model	MAE	RMSE	MAPE (%)	sMAPE (%)	$R^2$
Regresja liniowa	16.41	21.19	8.17	7.60	0.8178
Regresja Ridge	16.07	20.62	7.96	7.46	0.8275

Najlepsza wartość hiperparametru  $\alpha$  w regresji Ridge wyniosła 500.0. Wysoka wartość  $\alpha = 500.0$  sugeruje, że w pełnym zbiorze danych występuje istotna współliniowość między zmiennymi objaśniającymi. Może to być spowodowane wysoką korelacją między zmiennymi, np. zmiennymi pogodowymi.

Następnie przeprowadzono analizę skróconego zbioru danych opisanego w 4.2. Wyniki dla skróconego zbioru danych przedstawiono w tabeli 7, wraz z różnicami w metrykach względem pełnego zbioru danych. Wyniki dla skróconego zbioru były nieco gorsze: dla regresji Ridge MAPE wyniosło 8.15% w porównaniu do 7.96% dla pełnego zbioru, a  $R^2$  spadło z 0.8275 do 0.8144. Różnice

## Rozdział 6. Analiza danych

---

w metrykach wskazują, że dodatkowe zmienne w pełnym zbiorze danych wnoszą istotną информацию, mimo niższego poziomu korelacji ze zmienną objaśnianą. Jednak różnice nie są bardzo duże (np. MAPE różni się o 0.19% dla regresji Ridge), co może sugerować, że skrócony zbiór danych nadal zawiera najważniejsze zmienne objaśniające.

**Tabela 7.** Wyniki regresji liniowej i Ridge dla skróconego zbioru danych w okresie stabilnym (2019) wraz z różnicami względem pełnego zbioru.

Model	MAE	RMSE	MAPE (%)	sMAPE (%)	$R^2$	Różnica MAPE (%)
Regresja liniowa	17.57	22.20	8.54	8.02	0.8001	+0.37
Regresja Ridge	16.73	21.39	8.15	7.69	0.8144	+0.19

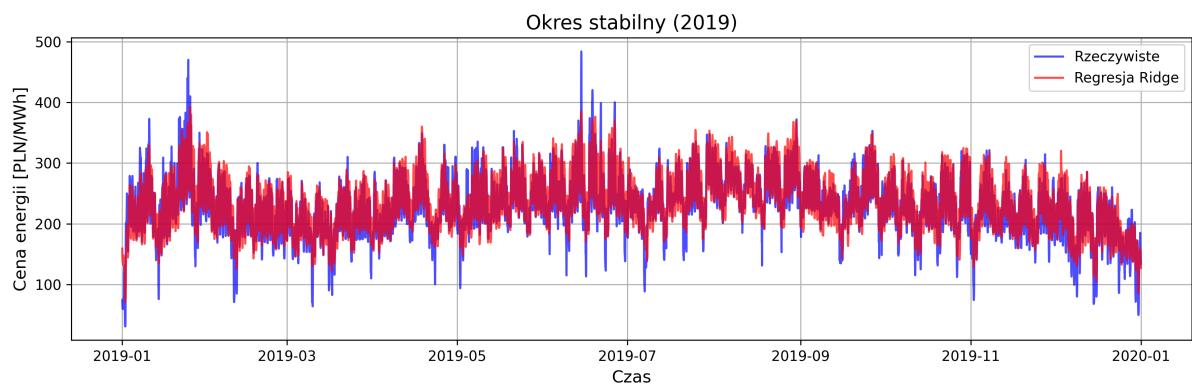
W celu potencjalnego polepszenia wyników zastosowano logarytmizację zmiennej wyjściowej (`fixing_i_price`), co miało na celu zmniejszenie skośności rozkładu cen i poprawę dopasowania modelu. Wyniki z logarytmizacją przedstawiono w tabeli 8.

**Tabela 8.** Wyniki regresji liniowej i Ridge z logarytmizacją dla okresu stabilnego (2019).

Model i zbiór danych	MAE	RMSE	MAPE (%)	sMAPE (%)	$R^2$
Regresja liniowa (pełny)	21.63	28.36	9.85	9.16	0.6736
Regresja Ridge (pełny)	19.99	26.06	9.21	8.65	0.7244
Regresja liniowa (skrócony)	26.36	33.07	11.95	10.94	0.5563
Regresja Ridge (skrócony)	23.29	29.38	10.69	9.89	0.6498

Logarytmizacja nie przyniosła spodziewanych korzyści i pogorszyła wyniki we wszystkich metrykach. Największe pogorszenie zaobserwowano dla skróconego zbioru danych, gdzie MAPE dla regresji liniowej wzrosło do 11.95%, a  $R^2$  spadło do 0.5563. Przyczyną jest najprawdopodobniej fakt, że rozkład cen w okresie stabilnym nie jest mocno skośny ze średnią wynoszącą 193,51 PLN oraz medianą 182 PLN, więc logarytmizacja wprowadziła niepotrzebne nieliniowości, które utrudniły dopasowanie modeli liniowych. Dodatkowo, odwrócenie transformacji logarytmicznej może amplifikować błędy predykcji, co wpłynęło na wzrost RMSE i MAE.

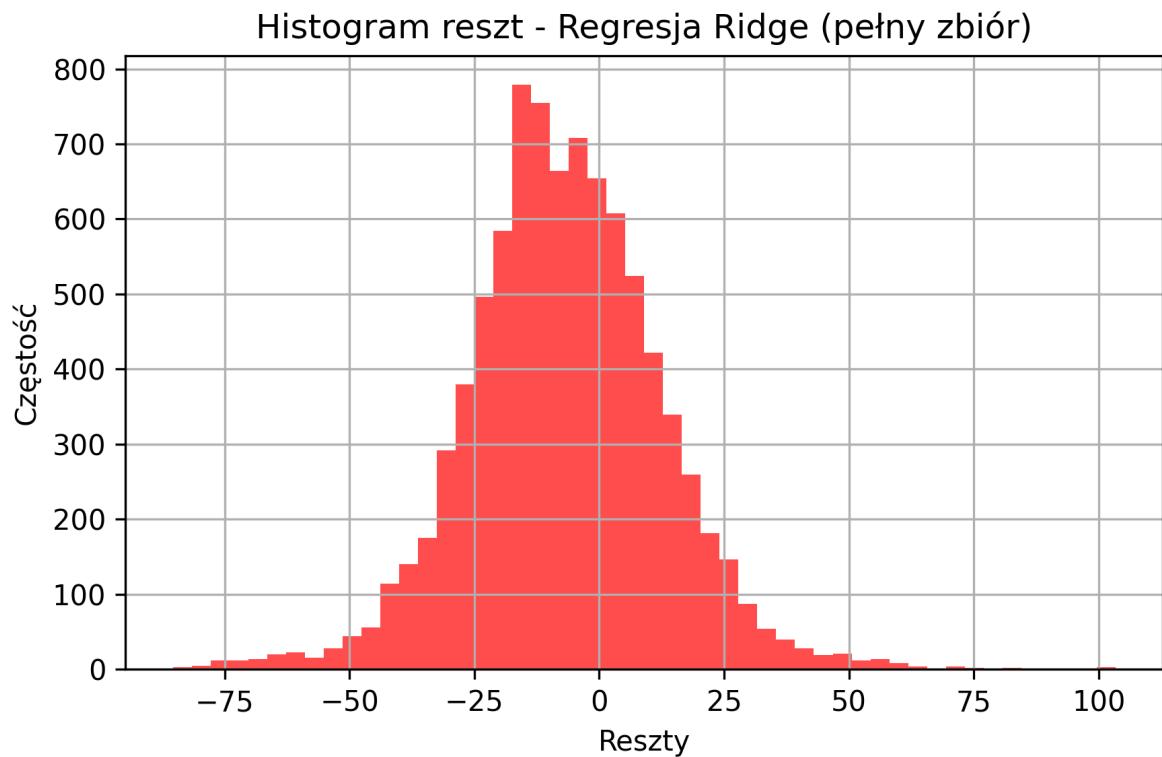
Wyraźne jest, że regresja Ridge w porównaniu do regresji liniowej lepiej radzi sobie z danymi o dużej liczbie zmiennych objaśniających, co potwierdzają wyniki dla pełnego zbioru danych. Poniżej przedstawiono wykresy rzeczywistych i przewidywanych wartości.



**Rysunek 20.** Porównanie rzeczywistych i przewidywanych wartości cen energii dla regresji Ridge w okresie stabilnym.

Widoczne na wykresie jest to, że model nie nadaje za zbyt wysokimi bądź zbyt niskimi wartościami. Może to sugerować, że model nie jest wystarczająco elastyczny, aby uchwycić te zmiany. Wartości przewidywane przez model są bardziej liniowe od rzeczywistych cen energii, co jest typowe dla modeli regresyjnych.

W związku z tym, przygotowany został histogram reszt dla regresji Ridge, aby ocenić, czy reszty są rozkładem normalnym.



**Rysunek 21.** Histogram reszt dla regresji Ridge w okresie stabilnym.

Histogram posiada jeden ze szczytów w okolicy zera i większość reszt skupia się w okolicy zera. Rozkład reszt jest zbliżony do normalnego, ale jest asymetryczny w lewą stronę, co może sugerować, że model jest skłonny do niedoszacowywania wartości. Potwierdza się teza z wykresu rzeczywistych i przewidywanych wartości, która sugeruje, że znajdują się obserwacje z dużymi resztami zarówno dodatnimi, jak i ujemnymi.

## **6.2 Okres niestabilny**

Ridge : W pracy wartość  $\lambda$  została wybrana na podstawie walidacji krzyżowej (ang. cross-validation), aby zoptymalizować równowagę między dopasowaniem modelu do danych a jego zdolnością do generalizacji.

## **Rozdział 7**

### **Podsumowanie wyników i wnioski**



# Bibliografia

- [1] Busby, J. W. i in., „Cascading risks: Understanding the 2021 winter blackout in Texas”, *Energy Research & Social Science*, s. 102106, 2021, ISSN: 2214-6296. DOI: <https://doi.org/10.1016/j.erss.2021.102106>. adr.: <https://www.sciencedirect.com/science/article/pii/S2214629621001997>.
- [2] Lago, J., Marcjasz, G., De Schutter, B. i Weron, R., „Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark”, *Applied Energy*, t. 293, s. 116983, 2021, ISSN: 0306-2619. DOI: <https://doi.org/10.1016/j.apenergy.2021.116983>. adr.: <https://www.sciencedirect.com/science/article/pii/S0306261921004529>.
- [3] Laitkos, V., Vontzos, G., Paraschoudis, P., Tsampasis, E., Bargiolas, D. i Tsoukalas, L. H., „The State of the Art Electricity Load and Price Forecasting for the Modern Wholesale Electricity Market”, *Energies*, t. 17, nr. 22, 2024, ISSN: 1996-1073. DOI: [10.3390/en17225797](https://doi.org/10.3390/en17225797). adr.: <https://www.mdpi.com/1996-1073/17/22/5797>.
- [4] Lipiecki, A., Uniejewski, B. i Weron, R., „Postprocessing of point predictions for probabilistic forecasting of day-ahead electricity prices: The benefits of using isotonic distributional regression”, *Energy Economics*, t. 139, s. 107934, 2024, ISSN: 0140-9883. DOI: <https://doi.org/10.1016/j.eneco.2024.107934>. adr.: <https://www.sciencedirect.com/science/article/pii/S014098832400642X>.
- [5] Marcjasz, G., „Forecasting Electricity Prices Using Deep Neural Networks: A Robust Hyper-Parameter Selection Scheme”, *Energies*, t. 13, nr. 18, 2020, ISSN: 1996-1073. DOI: [10.3390/en13184605](https://doi.org/10.3390/en13184605). adr.: <https://www.mdpi.com/1996-1073/13/18/4605>.
- [6] Mikołaj Kalisz, A. M., „Uczenie ciągłe w modelowaniu cen energii elektrycznej”, prac. mag., Politechnika Warszawska, Mechatroniki, 2025.
- [7] Openmeteo, *Prognozy pogody*, <https://open-meteo.com>, Accessed: 2025-04-09, 2025.
- [8] Poggi, A., Di Persio, L. i Ehrhardt, M., „Electricity Price Forecasting via Statistical and Deep Learning Approaches: The German Case”, *AppliedMath*, t. 3, nr. 2, s. 316–342, 2023, ISSN: 2673-9909. DOI: [10.3390/appliedmath3020018](https://doi.org/10.3390/appliedmath3020018). adr.: <https://www.mdpi.com/2673-9909/3/2/18>.
- [9] PSE, *Dane systemowe*, <https://www.pse.pl/dane-systemowe>, Accessed: 2025-04-09, 2024.

## Bibliografia

---

- [10] PSE, *Raporty historyczne*, <https://raporty.pse.pl>, Accessed: 2025-04-09, 2025.
- [11] Tan, Z., Zhang, J., Wang, J. i Xu, J., „Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models”, *Applied Energy*, t. 87, nr. 11, s. 3606–3610, 2010, ISSN: 0306-2619. DOI: <https://doi.org/10.1016/j.apenergy.2010.05.012>. adr.: <https://www.sciencedirect.com/science/article/pii/S0306261910001807>.
- [12] Taylor, S. i Letham, B., „Forecasting at scale”, wrz. 2017. DOI: 10.7287/peerj.preprints.3190v2.
- [13] Uniejewski, B., Nowotarski, J. i Weron, R., „Automated Variable Selection and Shrinkage for Day-Ahead Electricity Price Forecasting”, *Energies*, t. 9, nr. 8, 2016, ISSN: 1996-1073. DOI: 10.3390/en9080621. adr.: <https://www.mdpi.com/1996-1073/9/8/621>.
- [14] Weron, R., „Electricity price forecasting: A review of the state-of-the-art with a look into the future”, *International Journal of Forecasting*, t. 30, nr. 4, s. 1030–1081, 2014, ISSN: 0169-2070. DOI: <https://doi.org/10.1016/j.ijforecast.2014.08.008>. adr.: <https://www.sciencedirect.com/science/article/pii/S0169207014001083>.
- [15] Zhang, J., Tan, Z. i Yang, S., „Day-ahead electricity price forecasting by a new hybrid method”, *Computers & Industrial Engineering*, t. 63, nr. 3, s. 695–701, 2012, Energy Management and Economics, ISSN: 0360-8352. DOI: <https://doi.org/10.1016/j.cie.2012.03.016>. adr.: <https://www.sciencedirect.com/science/article/pii/S036083521200085X>.
- [16] Ziel, F., Steinert, R. i Husmann, S., „Efficient modeling and forecasting of electricity spot prices”, *Energy Economics*, t. 47, s. 98–111, 2015, ISSN: 0140-9883. DOI: <https://doi.org/10.1016/j.eneco.2014.10.012>. adr.: <https://www.sciencedirect.com/science/article/pii/S0140988314002576>.

# Wykaz skrótów i symboli

**MLP** Wielowarstwowy Perceptron 15

**OZE** Odnawialne Źródła Energii 10

**PSE** Polskie Sieci Elektroenergetyczne 17

**RB** Rynek Bilansujący 29

**RDN** Rynek Dnia Następnego 9

# Spis rysunków

1	Zmienna cen energii elektrycznej na RDN w latach 2016-2024 . . . . .	18
2	Histogram rozkładu zmiennej fixing_i_price . . . . .	19
3	Zmienna temperatury w czasie (2016-2024) . . . . .	21
4	Zmienna prędkości wiatru w czasie (2016-2024) . . . . .	21
5	Zmienna promieniowania słonecznego w czasie (2016-2024) . . . . .	22
6	Zmienna zachmurzenia w czasie (2016-2024) . . . . .	22
7	Zmienna temperatury w czasie (2022) . . . . .	23
8	Zmienna prędkości wiatru w czasie (2022) . . . . .	23
9	Zmienna promieniowania słonecznego w czasie (2022) . . . . .	24
10	Zmienna zachmurzenia w czasie (2022) . . . . .	24
11	Zmienna produkcji energii z różnych źródeł w czasie (2016-2024) . . . . .	25
12	Porównanie produkcji energii z emisyjnych i bezemisyjnych źródeł w 2024 roku . . . . .	26
13	Ceny paliw kopalnych i emisji CO <sub>2</sub> w latach 2016-2024 (w ujęciu miesięcznym). Źródło: Opracowanie własne na podstawie danych rynkowych. . . . .	28
14	Straty mocy w wyniku awarii i w sieci w latach 2016-2024 (w ujęciu miesięcznym). Źródło: Opracowanie własne na podstawie danych z PSE. . . . .	29
15	Zapotrzebowanie na energię i wolumen handlu w latach 2016-2024 (w ujęciu miesięcznym). Źródło: Opracowanie własne na podstawie danych z PSE. . . . .	30
16	Korelacja dni tygodnia z cenami energii na RDN. . . . .	31
17	Wykres korelacji zmiennych objaśniających względem zmiennej docelowej fixing_i_price. . . . .	37
18	Mapa cieplna korelacji zmiennych objaśniających w skróconym zbiorze danych. . . . .	38
19	Podział szeregów czasowych cen energii na zbiory treningowe i testowe . . . . .	40
20	Porównanie rzeczywistych i przewidywanych wartości cen energii dla regresji Ridge w okresie stabilnym. . . . .	55
21	Histogram reszt dla regresji Ridge w okresie stabilnym. . . . .	55

# **Spis tabel**

1	Podstawowe statystyki zmiennej fixing_i_price . . . . .	18
2	Średni bilans wymiany energii z sąsiadami w latach 2016-2024 . . . . .	27
3	Opis zmiennych dotyczących cen paliw kopalnych i emisji CO <sub>2</sub> . . . . .	27
4	Średni kurs wymiany PLN/USD w latach 2016-2024. Źródło: Opracowanie własne na podstawie danych NBP. . . . .	32
5	Porównanie statystyk opisowych cen energii w okresach spokojnym (2016-2019) i niespokojnym (2020-2023). . . . .	39
6	Wyniki regresji liniowej i Ridge dla pełnego zbioru danych w okresie stabilnym (2019). . . . .	53
7	Wyniki regresji liniowej i Ridge dla skróconego zbioru danych w okresie stabilnym (2019) wraz z różnicami względem pełnego zbioru. . . . .	54
8	Wyniki regresji liniowej i Ridge z logarytmizacją dla okresu stabilnego (2019). . . . .	54

## **Spis załączników**