

# Predykcja cen energii elektrycznej na Rynku Dnia Następnego - TGE

Adam Marek

28 maja 2021

# Spis treści

<b>1</b>	<b>Wprowadzenie</b>	<b>2</b>
1.1	Charakterystyka rynku energii elektrycznej . . . . .	2
1.2	Wyzwania towarzyszące modelowaniu cen energii . . . . .	3
1.3	Uzasadnienie biznesowe . . . . .	5
<b>2</b>	<b>Dane</b>	<b>6</b>
2.1	Źródła danych . . . . .	6
2.1.1	Zmienna celu . . . . .	6
2.1.2	Pozostałe dane . . . . .	6
2.2	Metodologia gromadzenia danych . . . . .	8
<b>3</b>	<b>Eksploracyjna analiza danych</b>	<b>9</b>
3.1	Zmienna celu . . . . .	9
3.2	Charakterystyki . . . . .	14
<b>4</b>	<b>Inżynieria cech</b>	<b>20</b>
<b>5</b>	<b>Model</b>	<b>25</b>
5.1	Model naiwny . . . . .	25
5.2	Model liniowy . . . . .	26
5.3	Model nieliniowy . . . . .	27
<b>6</b>	<b>Wnioski</b>	<b>29</b>
6.1	Porównanie wyników . . . . .	29
6.2	Konkluzje . . . . .	29

# Rozdział 1

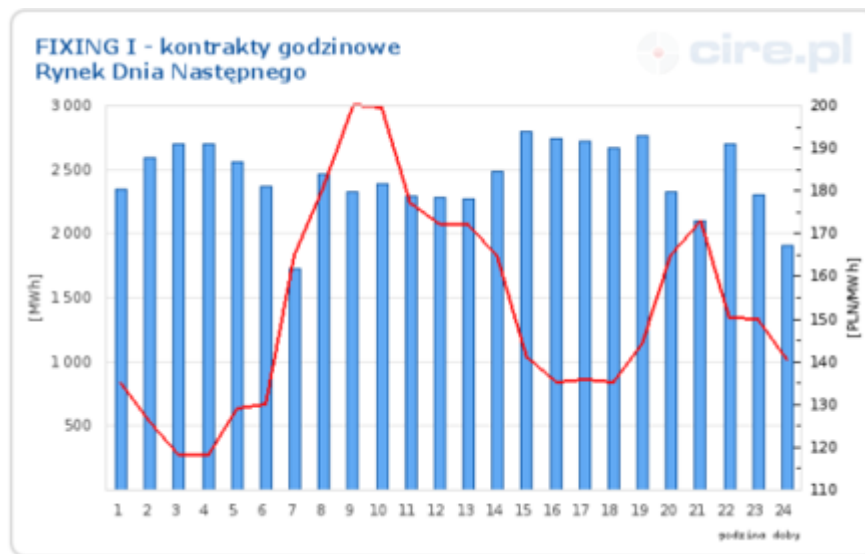
## Wprowadzenie

### 1.1 Charakterystyka rynku energii elektrycznej

Celem projektu jest predykcja cen energii na Towarowej Giełdzie Energii. Głównym obiektem zainteresowań będzie Rynek Dnia Następnego. Jest to rynek spot dla cen energii elektrycznej w Polsce, którego notowania wyznaczają poziom referencyjny dla kontraktów dwustronnych (tzw. kontraktów bilateralnych). Oferowane są tam kontrakty godzinowe oraz blokowe (bazowe, szczytowe i poza szczytowe). Łącznie obecnych jest sześć indeksów:

- TGeBase,
- TGePeak,
- TGeOffpeak,
- TGe9,
- TGe15,
- TGe24.

Ostatni z nich, tj. indeks godzinowy, będzie obiektem modelowania (jest to instrument bazowy dla kontraktów futures). Typowy profil cen na polskim rynku (na datę 20-04-2020) na tle wykresu zapotrzebowania prezentuje poniższy Rysunek.



Rysunek 1.1: Dzienny profil zmienności cen spot - TGE

Podstawowe wzorce, takie jak:

- dolina nocna,
- szczyt poranny,
- dolina popołudniowa,
- szczyt wieczorny

charakteryzują (w różnym stopniu) niemal każdy dzień w roku.

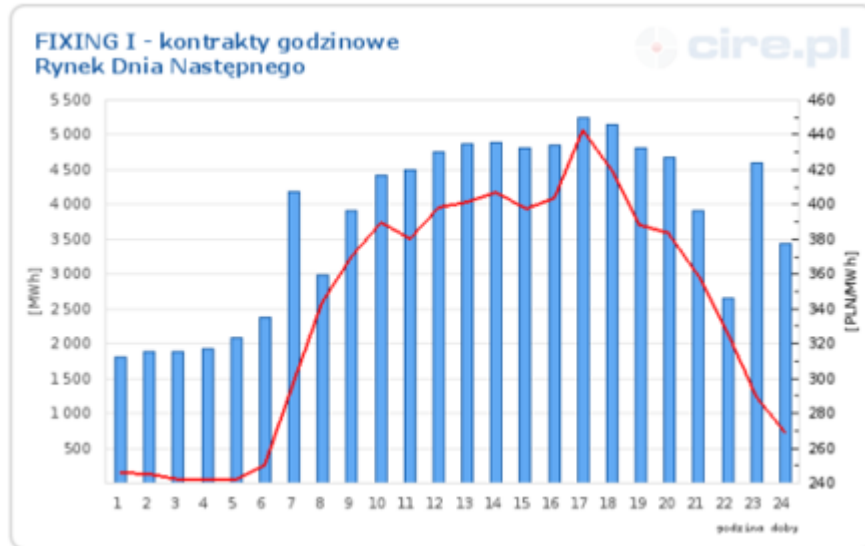
## 1.2 Wyzwania towarzyszące modelowaniu cen energii

Postawione w poprzednim paragrafie zadanie cechuje się dużym stopniem skomplikowania. Podstawowe składniki wpływające na poziom cen to m.in.:

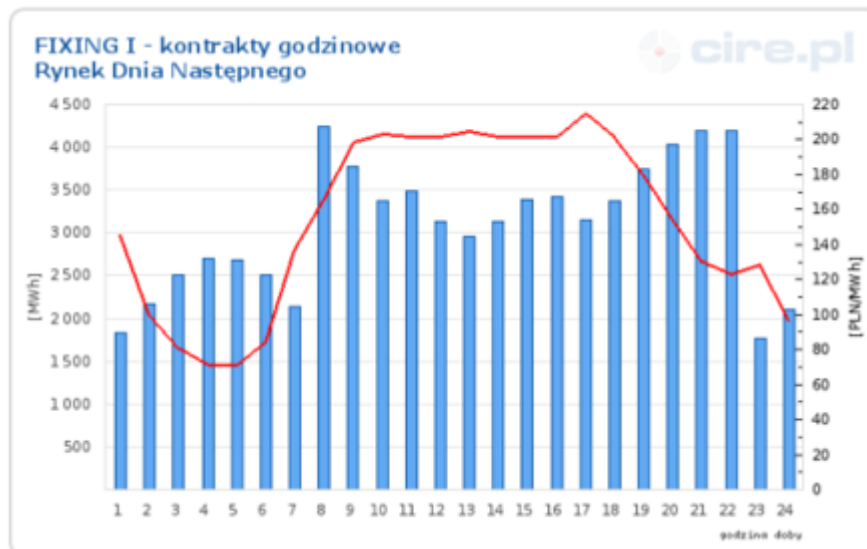
- aktualne zdolności wytwórcze,
- dostępności różnego typu źródeł (OZE, jednostki centralnie dysponowane),
- zdolności przesyłowe,
- import energii elektrycznej,
- warunki atmosferyczne,
- duże wahania zapotrzebowania (a więc i ceny) na energię elektryczną w ciągu dnia (tzw. dzienny profil zapotrzebowania),

- zmiany zapotrzebowania w zależności od dnia tygodnia,
- zmienność średniego zapotrzebowania w przeciągu roku,
- wpływ dni świątecznych.

Wpływ ostatniej kategorii - dni świątecznych - można zobrazować na podstawie poniższych Rysunków. Pierwszy z nich przedstawia stan na dzień 16 grudnia 2020, zaś drugi na dzień 24 grudnia 2020.



Rysunek 1.2: Dzienny profil zmienności cen - 16 grudnia 2020



Rysunek 1.3: Dzienny profil zmienności cen - 24 grudnia 2020

W Wigilię średnia cena energii jest dwukrotnie niższa.

Charakter zmienności wyżej wymienionych czynników również nie jest stały i ulega modyfikacjom na przestrzeni lat (są to przewaźnie szeregi niestacjonarne). Wpływają na to m.in. wzrost konsumpcji, rozwój urządzeń energooszczędnych oraz zmiany legislacyjne zapewniające mechanizmy wsparcia dla OZE. Proces modelowania jest dodatkowo utrudniony poprzez współzależności między wymienionymi czynnikami.

Dodatkowo, w celu zapewnienia możliwości potencjalnego wykorzystania predykcji, konieczne jest ich uzyskanie w dzień poprzedzający, do godziny 9:00 (tj. przewidywane wartości cen muszą być dostępne przed fixingiem, na którym są one ustalane na cały następny dzień). Wiąże się to z ograniczeniem danych, na których można bazować estymując prognozy

### **1.3 Uzasadnienie biznesowe**

Ze względu na brak możliwości magazynowania energii na systemową skalę całe zapotrzebowanie musi zostać pokryte. Wolumen niebilansowania jest obiektem obrotu na Rynku Bilansującym, który znany jest z dużej niestabilności, w związku z czym dobrej jakości prognozy na RDN (Rynek Dnia Następnego) niosą ze sobą duże korzyści natury ekonomicznej.

# Rozdział 2

## Dane

### 2.1 Źródła danych

#### 2.1.1 Zmienna celu

Postawiony w poprzednim paragrafie problem można zakwalifikować do klasy nadzorowanych zagadnień uczenia maszynowego. Zmienną celu jest w tym przypadku cena na określoną godzinę (tj. 24 godziny wprzód w stosunku do identyfikatora czasowego każdej obserwacji). Źródłem danych jest oficjalny serwis internetowy Towarowej Giełdy Energii (<https://tge.pl>), na której dane te są publikowane codziennie, a po upływie dostatecznie długiego okresu - archiwizowane.

#### 2.1.2 Pozostałe dane

Głównym założeniem niniejszego projektu jest próba predykcji cen mająca charakter wspomagający traderów w ich decyzjach. Ponieważ fixing na rozważanej giełdzie odbywa się o godzinie 9:00 (lokalnej strefy czasowej), wszystkie dane użyte do uczenia powinny być do godziny 8:00 dostępne, aby można było generować prognozy. Mając to ograniczenie na uwadze wstępny wybór cech obejmował dane datowane na dany dzień (tj. poprzedzający dzień, na który wystawiane są prognozy) oraz sześć dni wstecz (tj. tydzień wstecz w stosunku do dnia prognoz). Taki wstępny wybór ma uzasadnienie w licznych badaniach ekspertów z zakresu rynków energii na całym świecie i jest potwierdzony wynikami eksploracyjnej analizy danych (Rozdział 3). Jako zmienne wyjaśniające przyjęte następujące klasy wielkości:

- wymiana transgraniczna, synchroniczna (1 zmienna przedziałowa),
- wymiana transgraniczna, asynchroniczna (1 zmienna przedziałowa),
- cena pozwoleń do emisji CO<sub>2</sub> wyrażona w lokalnej walucie (1 zmienna przedziałowa),
- przewidywane zdolności wytwórcze jednostek centralnie dysponowanych (1 zmienna przedziałowa),

- przewidywane zdolności wytwórcze pozostałych jednostek (1 zmienna przedziałowa),
- całkowita generacja jednostek centralnie dysponowanych (1 zmienna przedziałowa),
- przewidywana (na następny dzień) całkowita generacja jednostek centralnie dysponowanych (1 zmienna przedziałowa),
- całkowita generacja pozostałych jednostek (1 zmienna przedziałowa),
- przewidywana (na następny dzień) całkowita generacja pozostałych jednostek (1 zmienna przedziałowa),
- generacja jednostek w ramach usługi IRZ - interwencja rezerwy zimnej (1 zmienna przedziałowa),
- generacja jednostek w ramach usługi PI - praca interwencyjna (1 zmienna przedziałowa),
- generacja ze źródeł solarnych (1 zmienna przedziałowa),
- przewidywana (na następny dzień) generacja ze źródeł solarnych (1 zmienna przedziałowa),
- generacja ze źródeł wiatrowych (1 zmienna przedziałowa),
- przewidywana (na następny dzień) generacja ze źródeł wiatrowych (1 zmienna przedziałowa),
- rezerwa operacyjna (2 zmienne),
- zapotrzebowanie na energię elektryczną (1 zmienna przedziałowa),
- przewidywane (na następny dzień) zapotrzebowanie na energię elektryczną (1 zmienna przedziałowa),
- cena energii elektrycznej rynku spot (1 zmienna przedziałowa),
- sumaryczny import energii elektrycznej z różnych państw (6 zmiennych przedziałowa),
- sumaryczny eksport energii elektrycznej z różnych państw (6 zmiennych przedziałowa),
- rzeczywista generacja 138 największych jednostek wytwórczych charakteryzujących się różnym paliwem, sprawnością oraz rolą w systemie elektroenergetycznym (138 zmiennych przedziałowych),
- dane kalendarzowe (3 zmienne: flaga dni świątecznych - zmienna kategorialna, binarna; numer miesiąca oraz dnia roboczego - zmienne kategorialne, nominalne).



Łącznie jest 170 zmiennych na dzień poprzedzający dzień prognozy. Dodatkowo, każda z nich ma odpowiadającą cechę na tydzień wstecz. Sumarycznie zatem dostępnych jest 340 potencjalnych predyktorów do modelu. Wszystkie te wartości spełniają postawione wcześniej kryteria użyteczności.

Większość z nich pochodzi z oficjalnego serwisu polskiego operatora systemu przesyłowego - Polskie Sieci Elektroenergetyczne (<https://pse.pl>). Dodatkowe źródło stanowi serwis europejskiego operatora systemu przesyłowego ENTSOE (<https://transparency.entsoe.eu>)

## 2.2 Metodologia gromadzenia danych

Dane te zostały pobrane z publicznego API, wstępnie oczyszczone i złączone za pomocą skryptu języka Python. Zrzut ich w formacie XLSX został dołączony do projektu. Ostateczny zbiór został podzielony na 3 podzbiory:

- zbiór treningowy - wszystkie obserwacje od początku do dnia 31-07-2020,
- zbiór walidacyjny - obserwacje w okresie od 01-08-2020 do 30-09-2020,
- zbiór testowy - ostatnie dwa miesiące obserwacji.

Podział zbioru został wykonany przy zachowaniu chronologicznego porządku. Dodatkowo, w celu szybszej selekcji najlepiej rokujących klas modeli, ze zbioru testowego wybrany został zredukowany zbiór zawierający obserwację z ostatnich dwóch lat.

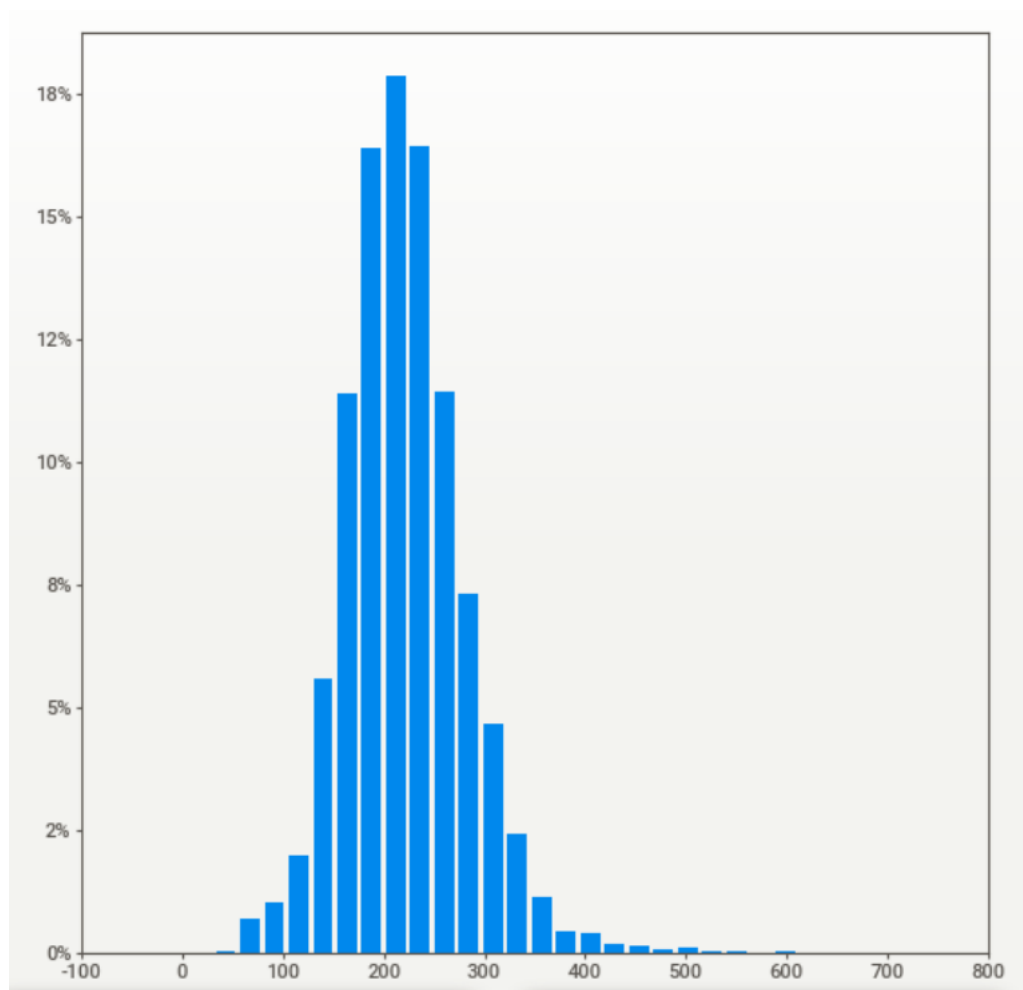
# Rozdział 3

## Eksploracyjna analiza danych

Wartości zmiennej celu są rejestrowane co godzinę. Pierwszy rekord datowany jest na dzień 01-01-2017 na godzinę 00:00 (czasu lokalnego). Ostatni rekord na dzień 31-10-2020 na godzinę 23:00. Sumarycznie jest 33432 obserwacji.

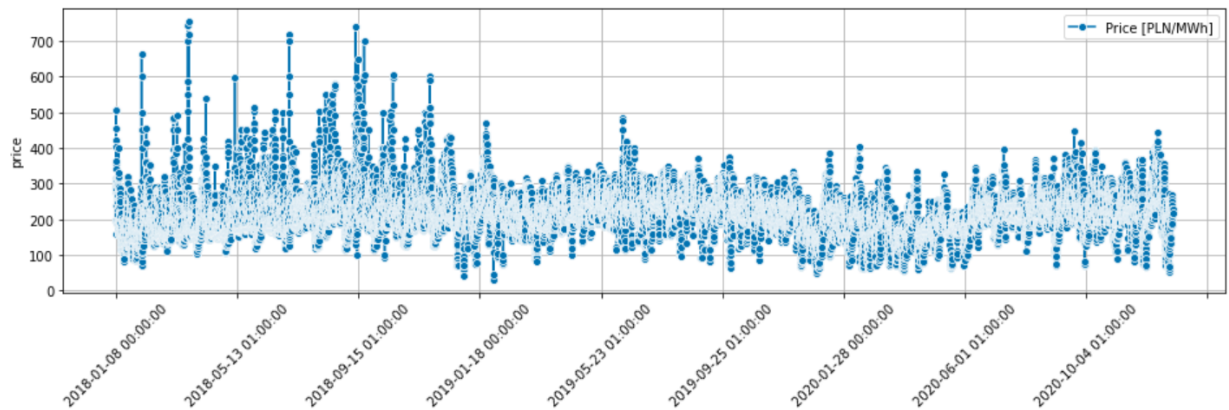
### 3.1 Zmienna celu

Poniższy Rysunek przedstawia rozkład wartości cen.



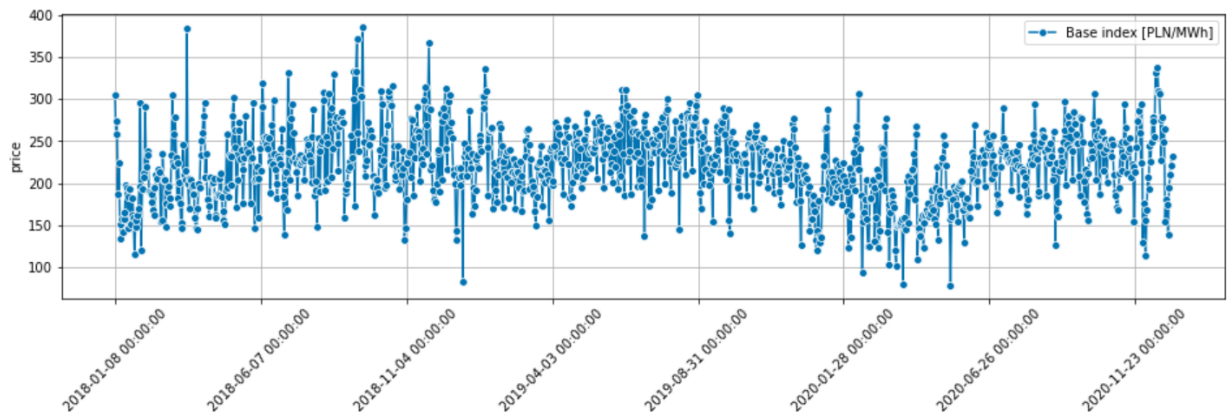
Rysunek 3.1: Histogram wartości cen

Wartości cen posiadają naturalne ograniczenie od dołu (ceny nie mogą być ujemne) oraz brak ograniczenia z góry. Skutkuje to niewielką asymetrią rozkładu. Kolejny Rysunek przedstawia historyczny przebieg zmienności cen z ostatnich trzech lat.



Rysunek 3.2: Zmienność cen na przekroju trzech lat

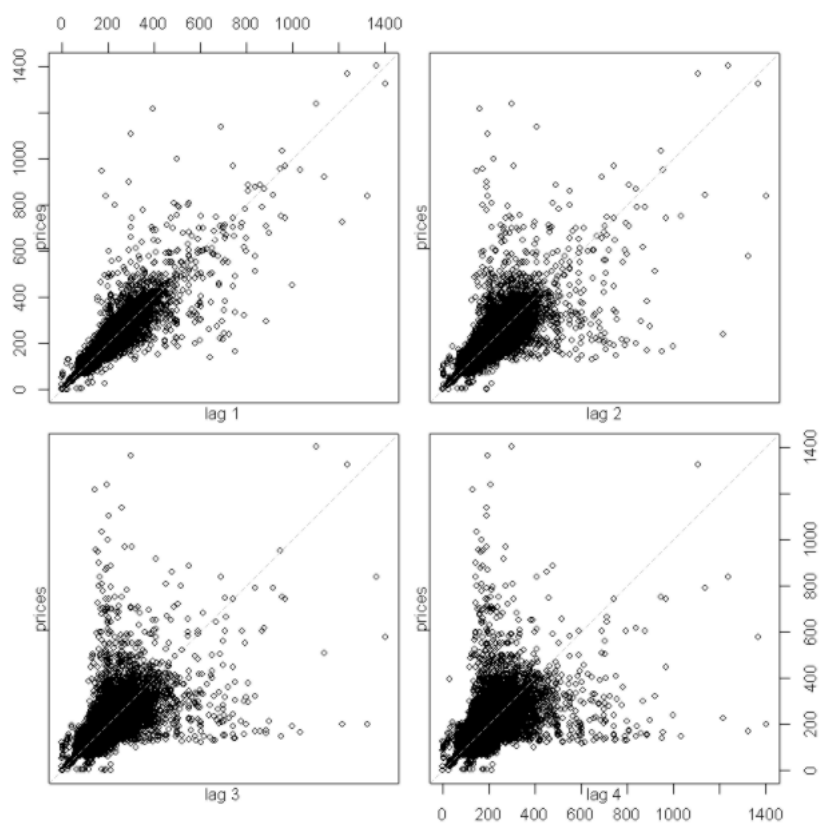
Widoczna jest nierównomierna w czasie wariancja, a także sporo obserwacji odstających (we wczesnych latach), których poprawne przewidywanie jest szczególnie skomplikowane. Dodatkowo, na kolejnym Rysunku pokazano zmienność wskaźnika *base* (będącego średnią arytmetyczną cen z danego dnia).



Rysunek 3.3: Zmienność wskaźnika *base* na przekroju trzech lat

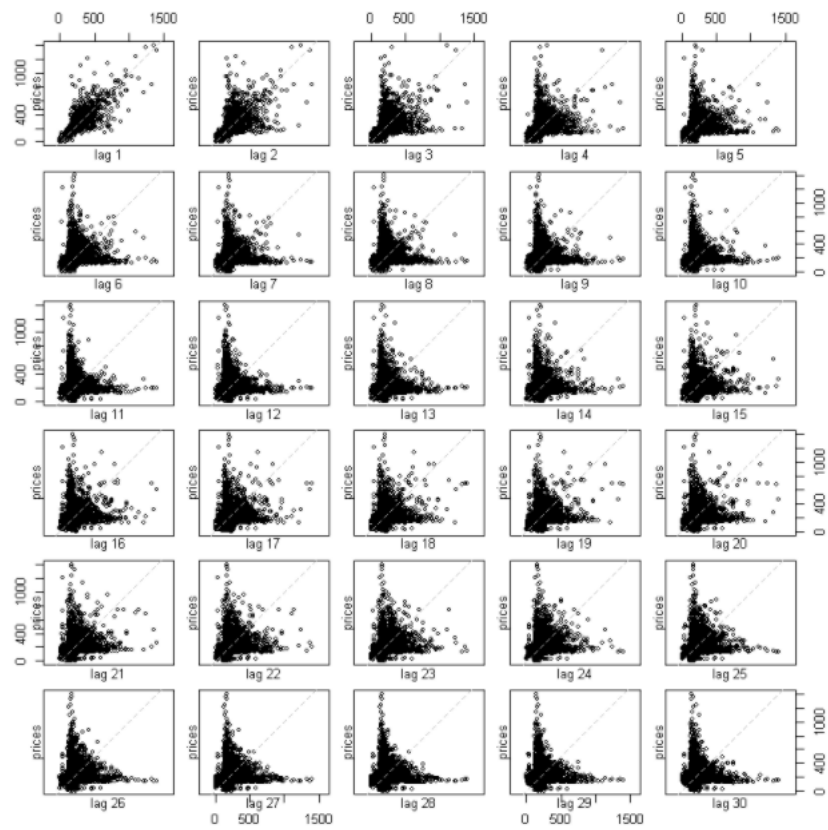
Brak jest tam wyraźnego trendu długoterminowego. Sama wariancja przejawia oznaki większej stabilności (jest to efekt spodziewany ze względu na regularyzujące działanie operacji uśredniania).

Następny Rysunek ilustruje autokorelacje cen dla opóźnienia godzinowego.



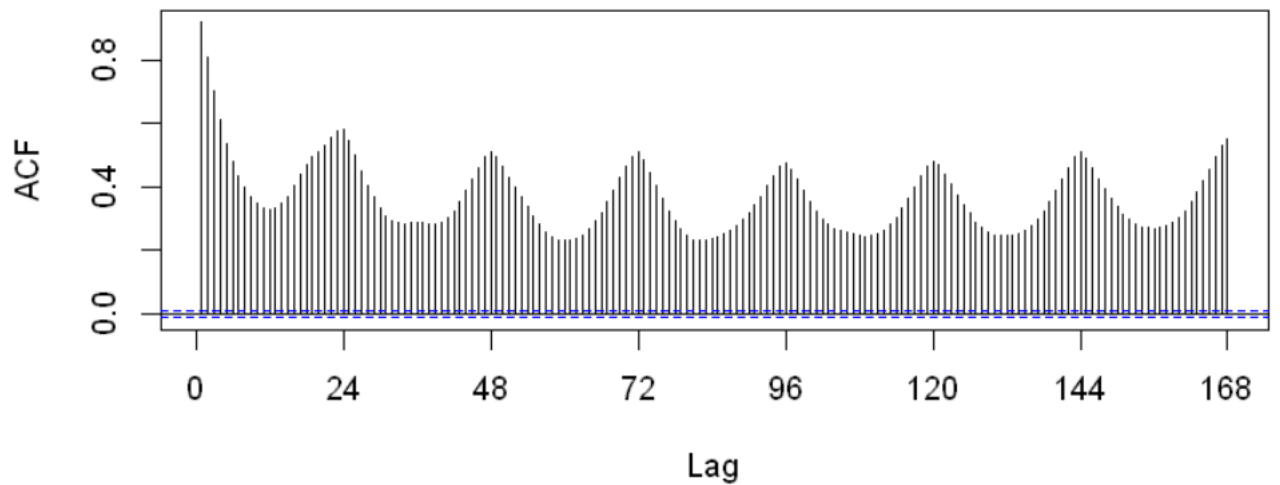
Rysunek 3.4: Wykresy autokorelacji dla czterech godzinowych opóźnień

Wyrazny jest wpływ tej samej godziny dnia poprzedniego. Każde kolejne opóźnienie charakteryzuje się słabszym stopniem skorelowania. Obserwację tę potwierdza kolejny Rysunek, który przedstawia analogiczną charakterystykę dla 24 godzinnych opóźnień.

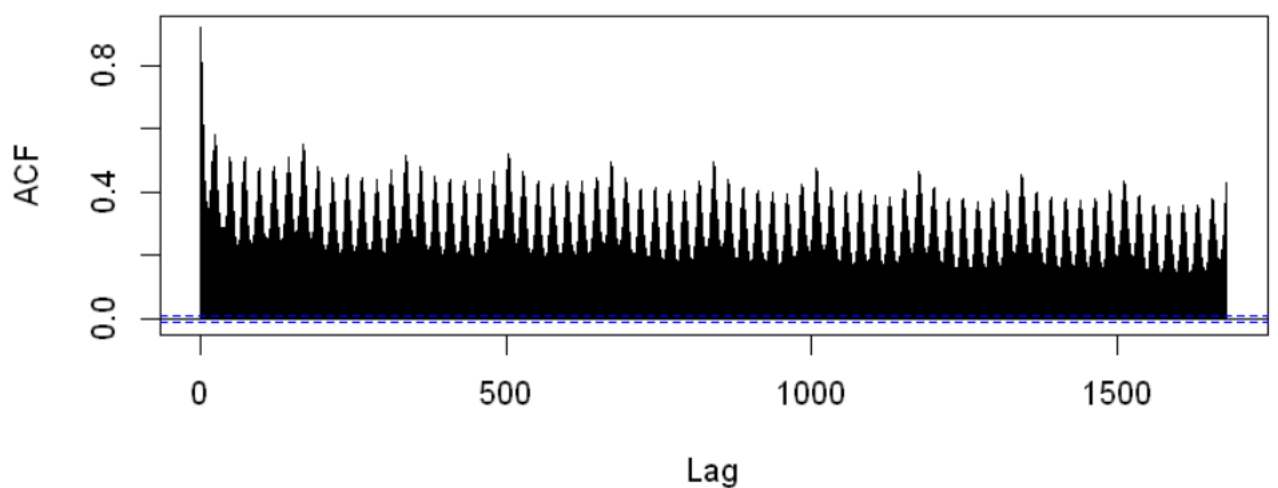


Rysunek 3.5: Wykresy autokorelacji dla dwudziestu czterech godzinowych opóźnień

Następnie ukazane zostały autokorelacje dla opóźnienia dziennego dla uśrednionych wartości cen (indeks *base*) w ramach jednego dnia. Pierwszy Rysunek ogranicza się do tygodnia (168 dziennych opóźnień), kolejny zaś nie ma ograniczenia.



Rysunek 3.6: Wykresy autokorelacji indeksu *base* dla dziennych opóźnień - ograniczone do tygodniowej historii



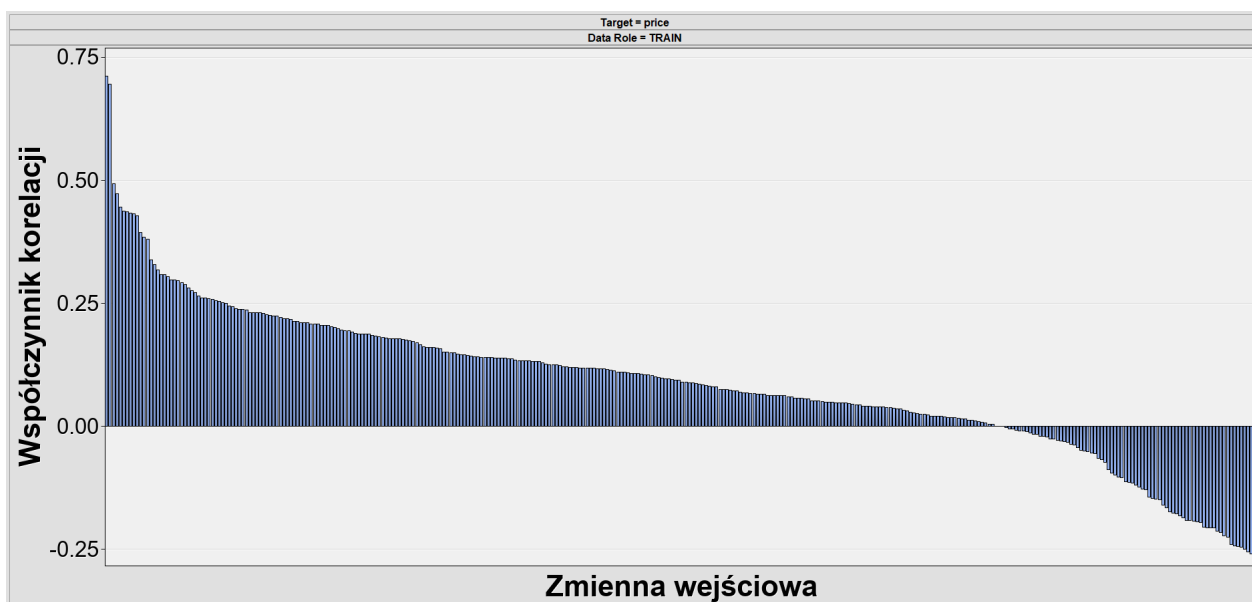
Rysunek 3.7: Wykresy autokorelacji indeksu *base* dla dziennych opóźnień

Widoczne piki wskazują największą korelację dla tego samego dnia tygodnia oraz dnia poprzedzającego, co usprawiedliwia przyjęcie tych danych do procesu uczenia.

## 3.2 Charakterystyki

Spośród dostępnych 340 cech nie sposób oczekiwać, aby każda wpływała na zmienną celu w jednakowy sposób. Istotnie - wpływ (mierzony w tym przypadku współczynnikiem

korelacji Pearsona) zmiennych przedziałowych zmienia się w sposób ciągły, co ilustruje poniższy Rysunek.

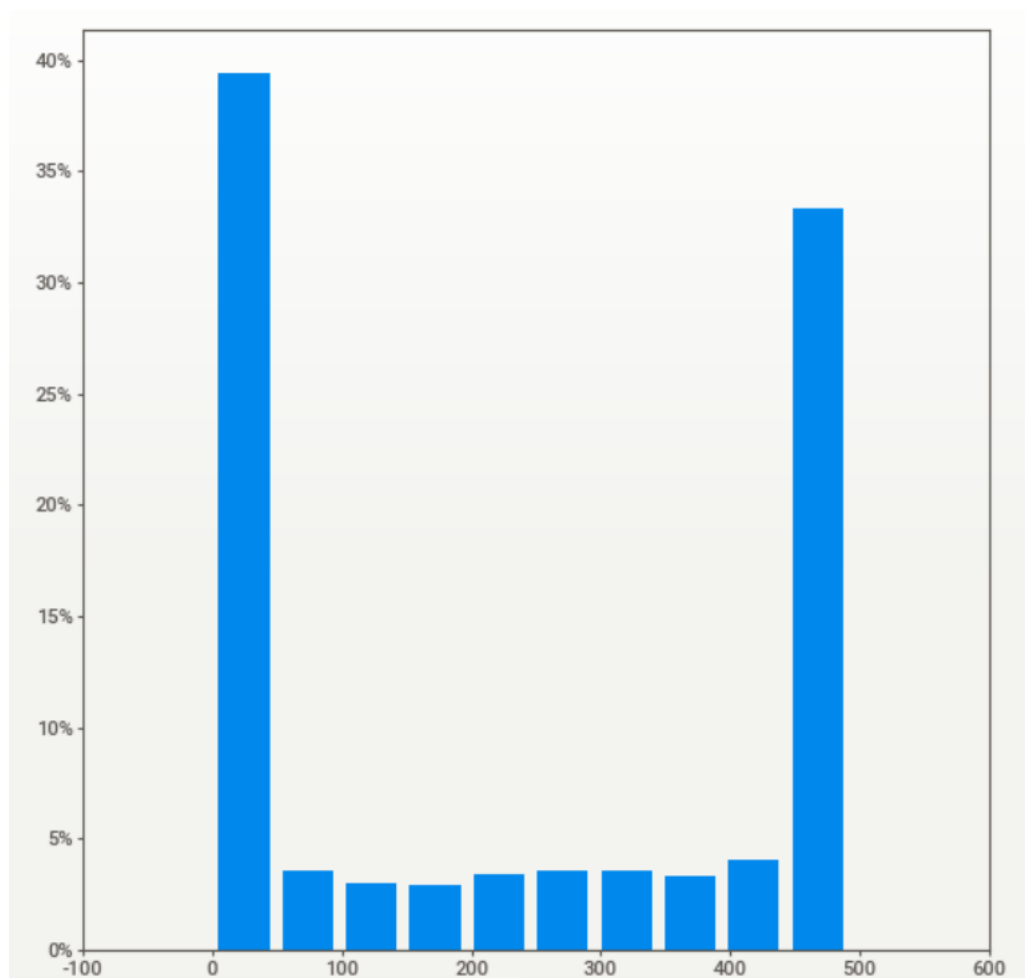


Rysunek 3.8: Wpływ cech na zmienną celu

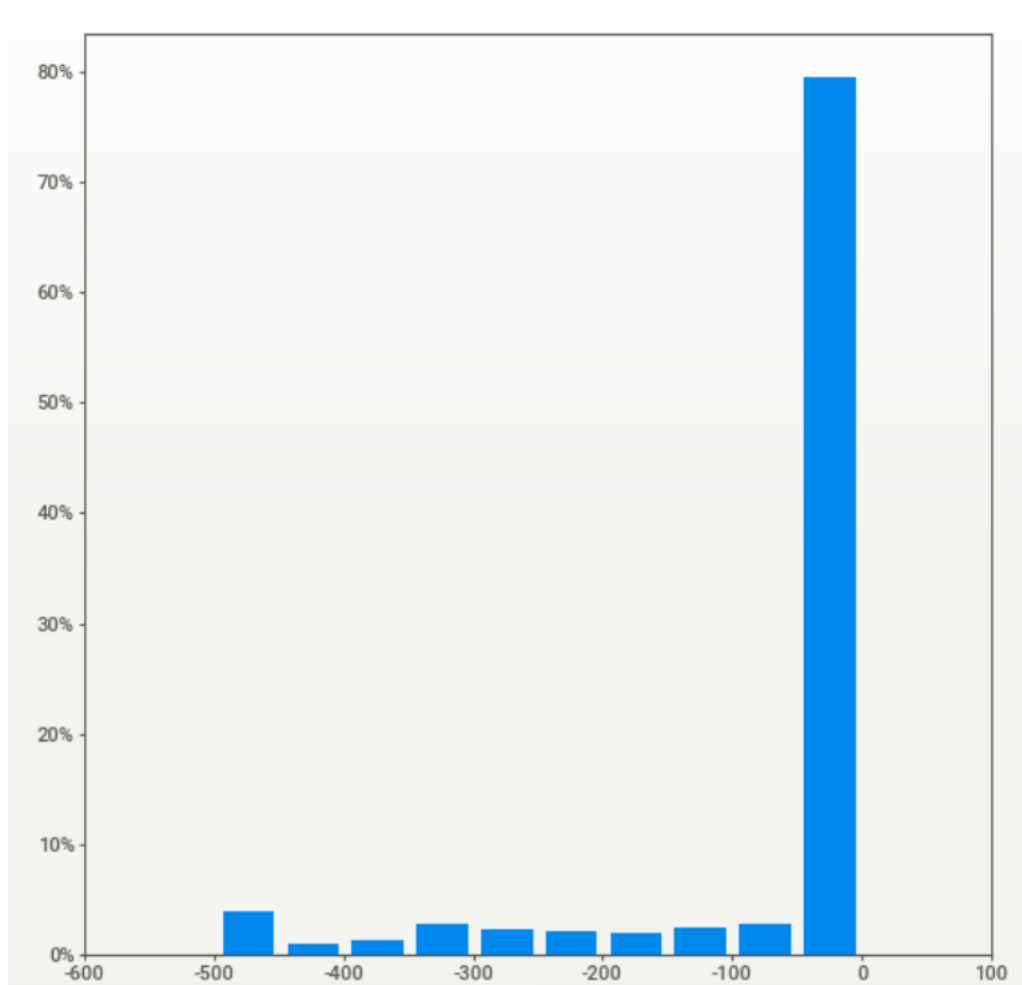
Najbardziej znaczącymi cechami są te o charakterze agregacyjnym (sumaryczna moc zainstalowana, sumaryczne zapotrzebowanie), a także poziomy cen z przeszłości i dane kalendarzowe. Zdecydowana większość zmiennych, których wpływ zmienia się w sposób ciągły to generacje na poziomie jednostki wytwórczej. Dokładniejsza analiza pokazuje, że te najmniej istotne zmienne (o najmniejszym module współczynnika korelacji) odpowiadają generacjom bloków małych, o niskiej sprawności, których rola w systemie elektroenergetycznym jest bardziej regulacyjna, aniżeli podstawowa. Problem ten został zaadresowany w ramach inżynierii cech (rozdział 4.1).

Dodatkowo, znalezione zostały silnie asymetryczne rozkłady zmiennych związanych z wymianą transgraniczną. Poniższe dwa Rysunki przedstawiają import i eksport z Litwą.



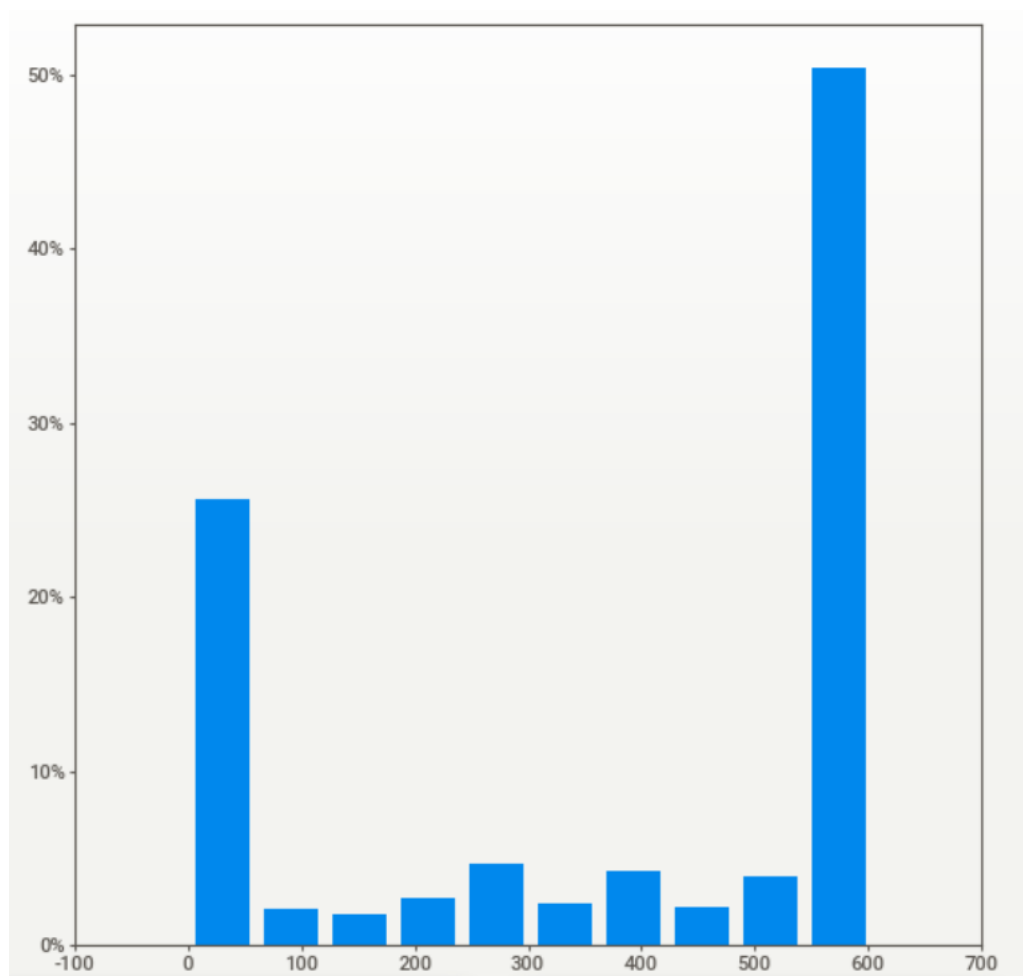


Rysunek 3.9: Wielkość importu (Litwa)

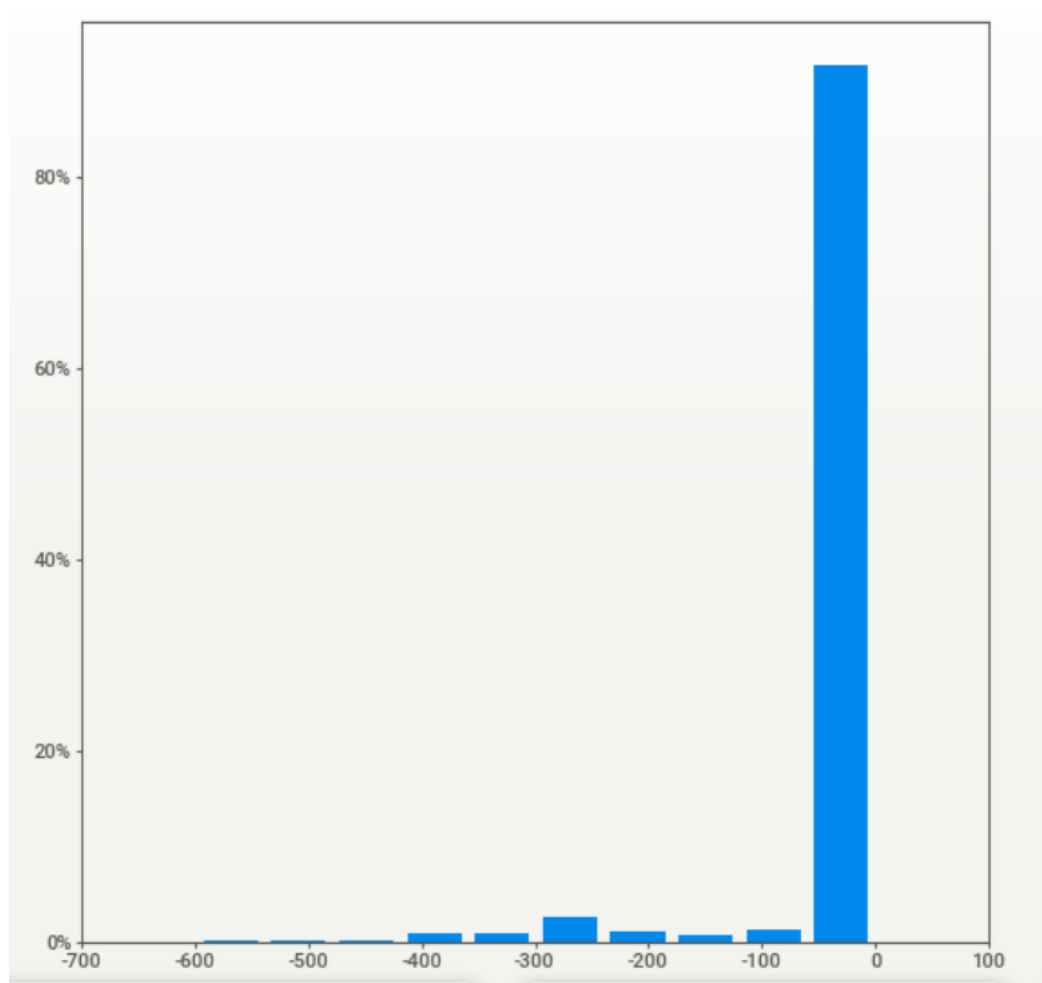


Rysunek 3.10: Wielkość eksportu (Litwa)

Kolejne dwa Rysunki ilustrują podobny problem związany z wymianą transgraniczną ze Słowacją.



Rysunek 3.11: Wielkość importu (Słowacja)



Rysunek 3.12: Wielkość eksportu (Słowacja)

Kwestia silnie skośnych wykresów również została zaadresowana na etapie inżynierii cech (Rozdział 4.1).

# Rozdział 4

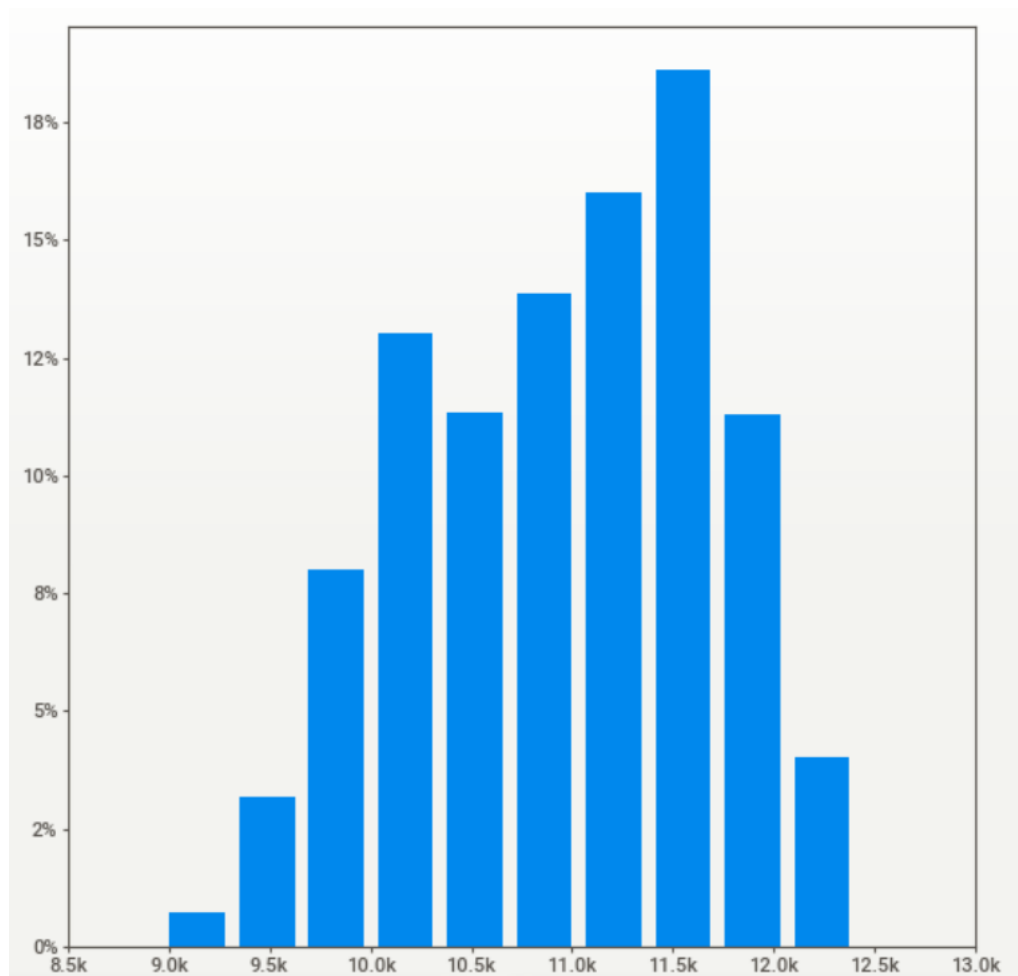
## Inżynieria cech

Na etapie eksploracyjnej analizy danych (rozdział 3.2) nakreślony został problem niskiej korelacji cech skojarzonych z generacją dla dużej klasy bloków energetycznych. Z tego względu, w ramach inżynierii cech, zostały wygenerowane charakterystyki agregujące - w grupy zebrane zostały bloki energetyczne operujące na tym samym typie paliwa o zbliżonych mocach nominalnych (moce bloków energetycznych zazwyczaj przyjmują konkretne, ściśle określone wartości, jednak ze względu na remonty ich moce rzeczywiste mogą się nieco różnić od wartości znamionowych). Powstały następujące kategorie zmiennych:

- Bloki na węgiel kamienny o mocy do 200 MW,
- Bloki na węgiel kamienny o mocy do 300 MW,
- Bloki na węgiel kamienny o mocy do 500 MW,
- Bloki na węgiel kamienny o mocy do 1000 MW,
- Bloki na węgiel brunatny o mocy do 200 MW,
- Bloki na węgiel brunatny o mocy do 300 MW,
- Bloki na węgiel brunatny o mocy do 500 MW,
- Bloki na węgiel brunatny o mocy do 1000 MW,
- Bloki na biomasę o mocy do 200 MW,
- Bloki na biomasę o mocy do 300 MW,
- Bloki na biomasę o mocy do 500 MW,
- Bloki na biomasę o mocy do 1000 MW,
- Bloki gazowe o mocy do 200 MW,
- Bloki gazowe o mocy do 300 MW,

- Bloki gazowe o mocy do 500 MW,
- Bloki gazowe o mocy do 1000 MW,

Każda z powyższej kategorii miała odpowiadające dwie zmienne: sumaryczną generację oraz sumaryczną moc dysponowaną w ramach danej klasy bloków. Przykładowy rozkład mocy dostępnej dla bloków klasy 200 MW działających w oparciu o węgiel kamienny przedstawia poniższy Rysunek.



Rysunek 4.1: Rozkład mocy dyspozycyjnej bloków na węgiel kamienny klasy 200 MW

Kolejny problem zidentyfikowany podczas eksploracyjnej analizy danych dotyczył asymetrycznych wykresów związanych z wymianą transgraniczną. Eksport (lub import) jest na jednym lub dwóch poziomach przez większość czasu (patrz Rysunki ...). Zmienne te zostały skategoryzowane do dwóch klas w przypadku rozkładu jednomodalnego oraz trzech klas w przypadku rozkładu dwumodalnego (owe klasy rozdzielały skrajne wartości zmiennej ciągłej na równe przedziały).

Ostatecznie, do modelowania zostały wyselekcjonowane następujące cechy:

- cena na RDN (zmienna przedziałowa),
- cena pozwoleń do emisji CO<sub>2</sub> wyrażona w walucie EUR (zmienna przedziałowa),
- przewidywana rezerwa mocy (zmienna przedziałowa),
- obciążenie systemu elektroenergetycznego (zmienna przedziałowa),
- sumaryczna generacja jednostek centralnie dysponowanych (zmienna przedziałowa),
- prognoza generacji ze źródeł wiatrowych na następny dzień (zmienna przedziałowa),
- sumaryczna generacja energii elektrycznej bloków na węgiel kamienny klasy 200 MW (zmienna przedziałowa),
- sumaryczna moc dostępna bloków na węgiel kamienny klasy 200 MW (zmienna przedziałowa),
- sumaryczna generacja energii elektrycznej bloków na węgiel kamienny klasy 300 MW (zmienna przedziałowa),
- sumaryczna moc dostępna bloków na węgiel kamienny klasy 300 MW (zmienna przedziałowa),
- sumaryczna generacja energii elektrycznej bloków na węgiel kamienny klasy 500 MW (zmienna przedziałowa),
- sumaryczna moc dostępna bloków na węgiel kamienny klasy 500 MW (zmienna przedziałowa),
- sumaryczna generacja energii elektrycznej bloków na węgiel kamienny klasy 1000 MW (zmienna przedziałowa),
- sumaryczna moc dostępna bloków na węgiel kamienny klasy 1000 MW (zmienna przedziałowa),
- sumaryczna generacja energii elektrycznej bloków na węgiel brunatny klasy 200 MW (zmienna przedziałowa),
- sumaryczna moc dostępna bloków na węgiel brunatny klasy 200 MW (zmienna przedziałowa),
- sumaryczna generacja energii elektrycznej bloków na węgiel brunatny klasy 300 MW (zmienna przedziałowa),
- sumaryczna moc dostępna bloków na węgiel brunatny klasy 300 MW (zmienna przedziałowa),
- sumaryczna generacja energii elektrycznej bloków na węgiel brunatny klasy 500 MW (zmienna przedziałowa),

- sumaryczna moc dostępna bloków na węgiel brunatny klasy 500 MW (zmienna przedziałowa),
- sumaryczna generacja energii elektrycznej bloków na węgiel brunatny klasy 1000 MW (zmienna przedziałowa),
- sumaryczna moc dostępna bloków na węgiel brunatny klasy 1000 MW (zmienna przedziałowa),
- sumaryczna generacja energii elektrycznej bloków na biomasę klasy 200 MW (zmienna przedziałowa),
- sumaryczna moc dostępna bloków na biomasę klasy 200 MW (zmienna przedziałowa),
- sumaryczna generacja energii elektrycznej bloków na biomasę klasy 300 MW (zmienna przedziałowa),
- sumaryczna moc dostępna bloków na biomasę klasy 300 MW (zmienna przedziałowa),
- sumaryczna generacja energii elektrycznej bloków na biomasę klasy 500 MW (zmienna przedziałowa),
- sumaryczna moc dostępna bloków na biomasę klasy 500 MW (zmienna przedziałowa),
- sumaryczna generacja energii elektrycznej bloków na biomasę klasy 1000 MW (zmienna przedziałowa),
- sumaryczna moc dostępna bloków na biomasę klasy 1000 MW (zmienna przedziałowa),
- sumaryczna generacja energii elektrycznej bloków gazowych klasy 200 MW (zmienna przedziałowa),
- sumaryczna moc dostępna bloków gazowych klasy 200 MW (zmienna przedziałowa),
- sumaryczna generacja energii elektrycznej bloków gazowych klasy 300 MW (zmienna przedziałowa),
- sumaryczna moc dostępna bloków gazowych klasy 300 MW (zmienna przedziałowa),
- sumaryczna generacja energii elektrycznej bloków gazowych klasy 500 MW (zmienna przedziałowa),
- sumaryczna moc dostępna bloków gazowych klasy 500 MW (zmienna przedziałowa),
- sumaryczna generacja energii elektrycznej bloków gazowych klasy 1000 MW (zmienna przedziałowa),



- sumaryczna moc dostępna bloków gazowych klasy 1000 MW (zmienna przedziałowa),
- import energii elektrycznej z Ukrainy (zmienna nominalna),
- eksport energii elektrycznej z Ukrainy (zmienna nominalna),
- import energii elektrycznej ze Słowacji (zmienna nominalna),
- eksport energii elektrycznej ze Słowacji (zmienna nominalna),
- import energii elektrycznej z Niemiec (zmienna nominalna),
- eksport energii elektrycznej z Niemiec (zmienna nominalna),
- import energii elektrycznej z Litwy (zmienna nominalna),
- eksport energii elektrycznej z Litwy (zmienna nominalna),
- import energii elektrycznej z Czech (zmienna nominalna),
- eksport energii elektrycznej z Czech (zmienna nominalna),
- numer dnia tygodnia (zmienna nominalna),
- numer miesiąca (zmienna nominalna).

Każda z powyższych występowała w dwóch wersjach - w dzień poprzedzający dzień prognozy oraz na tydzień wstecz.

# Rozdział 5

## Model

We wszystkich porównywanych modelach zmienną celu była  $\ell_2$ -norma residuów. Wybór taki podyktowany był celem penalizacji zbyt odległych prognoz od wartości rzeczywistych (w sensie wartości bezwzględnej). Wybrane zostały trzy klasy modeli:

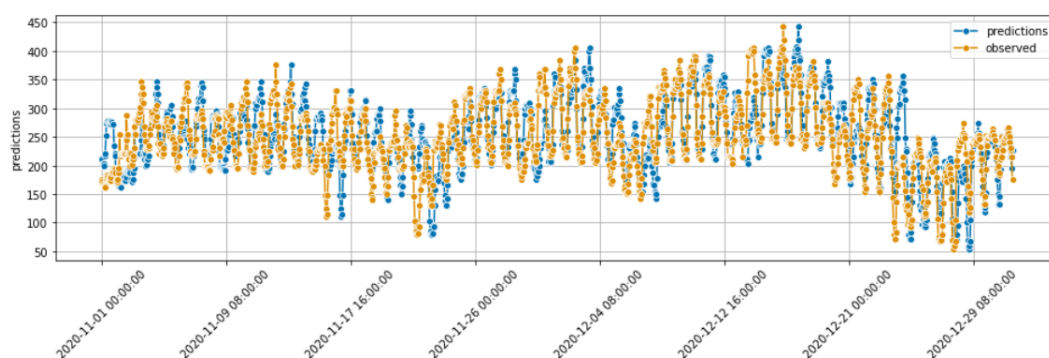
- Model naiwny - przewidujący na dany dzień wartość ceny z dnia poprzedniego (dla odpowiadających sobie godzin) ,
- Model liniowy - będący wynikiem regularyzowanej regresji liniowej (z członem autoregresyjnym oraz dodatkowymi cechami),
- Model nieliniowy - rekurencyjna sieć neuronowa o architekturze GRU.

### 5.1 Model naiwny

Model naiwny został wprowadzony w celach referencyjnych - błędy popełniane przez ten typ modeli stanowi ograniczenie górne, tj. próg dla potencjalnych, nowych modeli. Jest to analog klasyfikatora losowego (lub trywialnego).

Działanie modelu opiera się na wspomnianej sezonowości dziennej - predykcja ceny dla konkretnej godziny jest tożsama z ceną analogicznej godziny dnia poprzedniego. Jest to szczególny przypadek modeli autoregresyjnych (AR(1)). Niemniej, model ten nie był tranowany w klasycznym sensie (współczynniki nie były 'dopasowane' do zbioru treningowego) ze względu na charakter tego modelu - miał on służyć jako domyślny wybór.

Predykcje dla zbioru testowego przedstawia poniższy Rysunek.



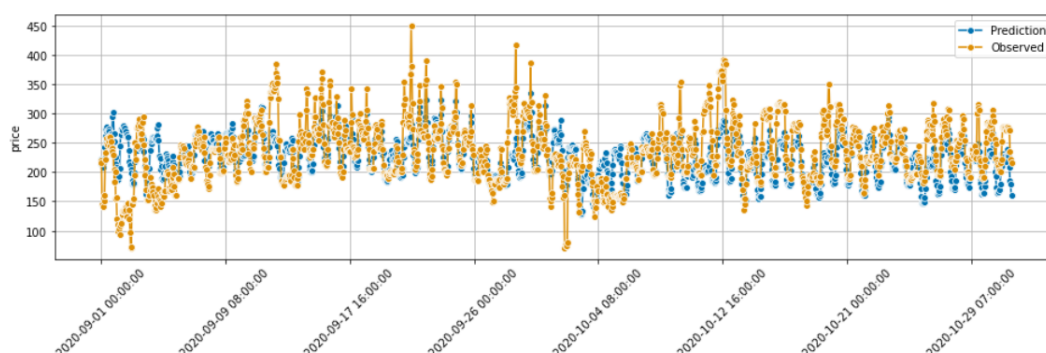
Rysunek 5.1: Zestawienie predykcji modelu naiwnego z wartościami obserwowanymi

Dla tego modelu wartość metryki RMSE (ang. *Root mean squared error* - pierwiastek błędu średniokwadratowego) wynosiła  $49.8 \text{ PLN/MWh}$ . Interpretacja tego wyniku jest taka, że model średnio myli się o wspomnianą wartość. Będzie ona stanowiła punkt wyjścia do porównywania pozostałych modeli.

## 5.2 Model liniowy

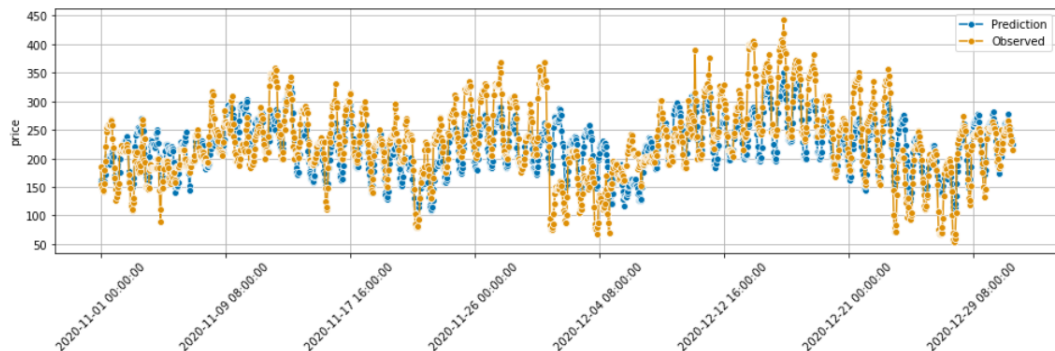
Modele liniowe charakteryzują się dużym stopniem interpretowalności oraz intuicyjności swoich wyników. Dodatkowo, znana jest z ekonometrii uogólniona zależność między popytem a podażą, która ma liniowy charakter. W związku z tym kolejny model jest reprezentantem tej grupy.

Klasa modeli regresji liniowej została wybrana jako dogodna do tworzenia modelu ze względu na dużą prostotę dodawania dodatkowych zmiennych (w przeciwieństwie na przykład do modeli klasy SARIMA). Dodatkowo, ze względu na dużą liczbę zmiennych konieczne było dodanie regularyzacji. Wybrana została w tym celu metoda elastycznej siatki (ang. *Elastic net*) będąca liniową kombinacją regularyzacji  $L1$  (ang. *LASSO*) oraz  $L2$  (ang. *Ridge*). Hiperparametr został dobrany za pomocą zbioru walidacyjnego. Poniższy Rysunek przedstawia wyniki dla zbioru walidacyjnego.



Rysunek 5.2: Zestawienie predykcji modelu liniowego z wartościami obserwowanymi dla zbioru walidacyjnego

Metryka RMSE przyjęła w tym przypadku wartość  $41.5\text{ PLN}/\text{MWh}$ . Kolejny Rysunek ilustruje przewidywania dla zbioru testowego.

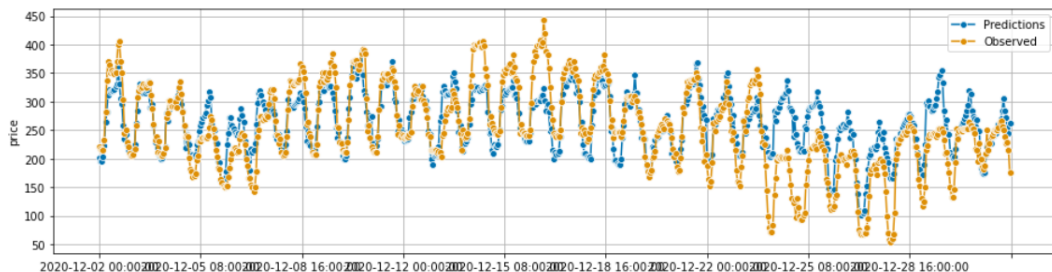


Rysunek 5.3: Zestawienie predykcji modelu liniowego z wartościami obserwowanymi dla zbioru testowego

W tym przypadku RMSE wynosiło  $44.5\text{ PLN}/\text{MWh}$ .

### 5.3 Model nieliniowy

Szeregi czasowe są szczególnym przypadkiem danych sekwencyjnych. W ostatnim czasie silnie rozwinęły się metody analizy innego typu danych sekwencyjnych - ciągów tekstowych. Obszar ten (*NLP* - ang. *Natural language processing*) jest z powodzeniem rozwijany przy użyciu sieci rekurencyjnych, których zdolność do głębokiej pamięci historycznej znacząco ułatwia rozpoznawanie tekstu. W związku z pewnym podobieństwem, wybrana została klasa rekurencyjnych sieci neuronowych jako dodatkowy model. W ramach wspomnianych danych tekstowych najpopularniejszą okazała się warstwa *LSTM* (warstwa o długiej pamięci krótkotrwałym - ang. *Long short-term memory*). Jednak, ze względu na duży stopień skomplikowania, wybrana została jej uproszczona wersja. Warstwa o architekturze GRU stanowi w tym przypadku dobry kompromis między jakością wyników a złożonością obliczeniową. Sama architektura tej warstwy została zoptymalizowana przy użyciu zbioru walidacyjnego. Poniższy Rysunek przedstawia wyniki dla zbioru testowego.



Rysunek 5.4: Zestawienie predykcji modelu nieliniowego z wartościami obserwowanymi dla zbioru testowego

Metryka RMSE przyjęła w tym przypadku wartość  $45.5\text{ PLN}/\text{MWh}$ .

# Rozdział 6

## Wnioski

### 6.1 Porównanie wyników

Poniższa tabela przedstawia zestawienie wyników uzyskanych przez stworzone modele, a także ocenę jakości predykcji komercyjnych (w sensie analizowanej metryki RMSE) dla wspólnego zbioru testowego.

Model	RMSE
Naiwny	49.8 PLN/MWh
Liniowy	44.5 PLN/MWh
Nieliniowy	45.5 PLN/MWh
Komercyjny	39.7 PLN/MWh

Zarówno model liniowy, jak i nieliniowy charakteryzują się lepszymi jakościowo wynikami niż model trywialny. Niemniej, wciąż są istotnie gorsze w stosunku do stosowanego komercyjnie modelu. Dodatkowo, różnica między nimi nie jest duża. Co więcej, może ona zależeć w silny sposób od wybranego horyzontu predykcji. Ze względu na lepszą wyjaśnialność i stabilność wyników, rekomendowany byłby model liniowy w użyciu produkcyjnym.

### 6.2 Konkluzje

Poniższe punkty przedstawiają kluczowe wnioski dotyczące wytrenowanych modeli:

- Oba modele są podobne do siebie w kontekście stosowanej metryki oceny modelu,
- Model referencyjny daje istotnie gorsze wyniki od modeli wytrenowanych sugerując iż wnoszą one wartość dodaną dla analityków rynku,
- Przewidywania modelu komercyjnego nadal pozostają istotnie lepsze - istnieje potencjalne pole do usprawnień,
- Ze względu na prostotę modeli liniowego (interpretowalność, intuicyjność, wyjaśnialność, stabilność) jest on rekomendowanym wyborem w świetle aktualnych wyników.

Następny zestaw punktów stanowi ogólne refleksje związane z doświadczeniami podczas modelowania cen na polskim rynku energii elektrycznej.

- Polski rynek energii jest ‘płytki’ - duzi gracze (koncerny energetyczne takie jak PGE) mogą istotnie wpływać na rynek,
- Rynek jest dość młody i dynamicznie się zmienia - regulacje prawne towarzyszące transformacji energetycznej mają silny wpływ na kształtowanie się cen,
- Ze względu na krótką historię do modelowania pozostaje relatywnie niewielki wolumen danych (szczególnie dla modeli uczenia głębokiego),
- Krótka historia danych do modelowania szczególnie uwidacznia się w przypadku świąt / dni wolnych od pracy, dla których stosownym byłby oddzielny model ze względu na ich specyfikę,
- Najistotniejszą rolę w przypadku budowania modelu cen na polskim rynku odgrywa dobra inżynieria cech (wiedza i doświadczenie tradera).