

mgr Wiktor Kolat¹, mgr inż. Wiktoria Galecka²

Uniwersytet Ekonomiczny w Katowicach

Wiktor Kowalski³

Uniwersytet Ekonomiczny we Wrocławiu

PORÓWNANIE REGRESJI LINIOWEJ (DYNAMICZNEJ AR(2)-X) I SIECI NEURONOWYCH W PROGNOZOWANIU CEN NIERUCHOMOŚCI MIESZKANIOWYCH W POLSCE – ANALIZA NA DANYCH KWARTALNYCH 2014-2024

COMPARISON OF LINEAR REGRESSION (DYNAMIC AR(2)-X) AND NEURAL NETWORKS IN FORECASTING RESIDENTIAL REAL ESTATE PRICES IN POLAND – ANALYSIS ON QUARTERLY DATA 2014-2024

Streszczenie

Celem rozdziału jest porównanie skuteczności prognozowania cen nieruchomości mieszkaniowych w Polsce za pomocą modelu regresji dynamicznej typu AR(2)-X (OLS/MNK) oraz sztucznych sieci neuronowych (ANN). Analiza oparta jest na danych

¹ <https://orcid.org/0009-0004-8718-5905>, e-mail: wiktor.kolat@ue.katowice.pl

² <https://orcid.org/0000-0001-9975-4897>

³ <https://orcid.org/0009-0009-1449-0366>

Kolat, W., Galecka, W., Kowalski, W. (2026). Porównanie regresji liniowej (dynamicznej AR(2)-X) i sieci neuronowych w prognozowaniu cen nieruchomości mieszkaniowych w Polsce – analiza na danych kwartalnych 2014-2024. W: A. Siedlecka, M. Pyra (red.), *AI 360°: innowacje i zastosowania sztucznej inteligencji w gospodarce, społeczeństwie i nauce* (s. 135-160). Biała Podlaska: Akademia Bialska im. Jana Pawła II. https://doi.org/10.29316/9788368103366_8

GUS z lat 2014-2024 dotyczących rynku wtórnego w 16 miastach wojewódzkich. Przedstawiono konstrukcję modeli, ich walidację oraz ocenę jakości prognoz. Model regresji dynamicznej AR(2)-X charakteryzował się wysoką dokładnością, osiągając MAPE 2,05% i MAE 206,71 zł/m². Model sieci neuronowej MLP, choć przewyższał modele naiwne, uzyskał wyższe błędy predykcyjne: MAPE 2,24% i MAE 220,76 zł/m².

Słowa kluczowe: rynek nieruchomości, regresja liniowa, sztuczne sieci neuronowe, prognozowanie cen, analiza porównawcza

Abstract

The aim of this chapter is to compare the effectiveness of forecasting residential property prices in Poland using a dynamic regression model of the AR(2)-X type (estimated by OLS) and artificial neural networks (ANN). The analysis is based on Statistics Poland (GUS) data for 2014-2024 covering the secondary market in 16 regional capital cities. The paper presents the construction of the models, their validation, and an assessment of forecast accuracy. The AR(2)-X dynamic regression model exhibited high accuracy, achieving a MAPE of 2.05% and an MAE of PLN 206.71 per m². The MLP neural network model, although outperforming naïve benchmarks, produced higher prediction errors: a MAPE of 2.24% and an MAE of PLN 220.76 per m².

Keywords: real estate market, linear regression, artificial neural networks, price forecasting, comparative analysis

Wstęp

Rynek nieruchomości to złożony segment gospodarki, którego struktura i mechanizmy działania ulegają ciągłym przemianom (Żróbek i in., 2024).

Współczesny rynek nieruchomości mieszkaniowych charakteryzuje się cechami, które odróżniają go od innych segmentów gospodarki, są to m.in. (Bryx, 2006; Gołąbeska, 2011; Grzywińska-Rapca i in., 2024): niska elastyczność podaży, zmienna sytuacja popytowa, trwałość i unikatowość nieruchomości, długi cykl inwestycyjny, wysoka kapitałochłonność, podatność na regulacje prawne, znaczenie emocji i preferencji kupujących, cykliczność rynku, lokalność rynku oraz rola inwestorów na rynku (Bryx, 2006; Gołąbeska, 2011). W Polsce występują dwa główne segmenty rynku nieruchomości

mieszkaniowych (Kucharska-Stasiak, 2005; Gołąbeska, 2011; Grzywińska-Rapca i in., 2024):

- pierwotny – obejmuje nowe nieruchomości budowane przez deweloperów, które nie były wcześniej zamieszkane;
- wtórny – obejmuje nieruchomości już użytkowane, które są sprzedawane przez ich dotychczasowych właścicieli.

Poza czynnikami fizycznymi, które wpływają na wartość nieruchomości, a co za tym idzie na cenę (np.: zużycie budynku, lokalizacja, kształt działki, infrastruktura techniczna etc.), na jej wartość mają wpływ również czynniki makroekonomiczne (np. poziom bezrobocia, wysokość stóp procentowych, stan rozwoju gospodarczego, etap cyklu koniunkturalnego, etc.) (Mach, 2014; Zydrón i Walkowiak, 2013).

W kontekście złożoności rynku nieruchomości mieszkaniowych oraz wielości czynników wpływających na kształtowanie się cen, powstaje potrzeba opracowania skutecznych metod prognozowania.

Celem niniejszego badania jest porównanie skuteczności prognozowania średnich zagregowanych cen mieszkań w Polsce dla 16 miast wojewódzkich z wykorzystaniem dwóch podejść: modelu regresji dynamicznej AR(2)-X estymowanego MNK dla danych czasowych oraz metod z obszaru uczenia maszynowego, w szczególności sieci neuronowych. Analizę oparto na danych GUS dotyczących średnich kwartalnych cen transakcyjnych za 1 m² na rynku wtórnym dla 16 miast wojewódzkich w Polsce w latach 2014-2024.

Badanie ma na celu nie tylko porównanie dokładności prognoz obu podejść, ale także identyfikację ich względnych zalet i ograniczeń w kontekście praktycznego zastosowania na polskim rynku nieruchomości.

Podmiotami, dla których badania mogą okazać się przydatne są fundusze inwestycyjne, firmy deweloperskie czy banki, które mogą wykorzystywać dane makroekonomiczne do prognozy przyszłych cen nieruchomości w celu ustalenia strategii inwestycyjnej i rozwojowej nie dla konkretnego rynku, ale na zagregowanym poziomie całego kraju. Wyniki badań mogą zostać zastosowane również przez osoby fizyczne, które nie mają dostępu do specjalistycznych danych, których pozyskanie często wiąże się z poniesieniem dużych kosztów.

Regresja liniowa a sztuczne sieci neuronowe (ANN)

W rozdziale przez podejście regresyjne rozumie się modele liniowe estymowane metodą najmniejszych kwadratów, w tym specyfikacje dynamiczne dla danych czasowych uwzględniające opóźnienia zmiennej zależnej (ARX), co pozwala uchwycić inercję cen.

Regresja liniowa jest jedną z najstarszych technik statystycznych powszechnie stosowaną w metrologii (Puchalski, Warszawa, 2022; Warszawa, Puchalski, 2020), inżynierii (Vakhshouri, Nejadi, 2015), ekonomii, finansach (Welc, 2009), medycynie (Gorriz i in., 2026; Zapf i in., 2024; Chicco i in., 2020) oraz informatyce (Gorriz i in., 2026; Silhavy i in., 2017; Flores Hinostroza i in., 2025). Stanowi również podstawę dla wielu zaawansowanych algorytmów uczenia maszynowego. Metoda stosowana jest do szacowania liniowych zależności między zmienną zależną (objaśnianą), a jedną lub więcej zmiennymi niezależnymi (objaśniającymi) (Gorriz i in., 2026; Puchalski, Warszawa, 2022; Ampuła, 2014).

Model regresji liniowej często jest formalnie definiowany jako oczekiwana wartość zmiennej zależnej Y pod warunkiem danej wartości zmiennej niezależnej X , z dodanym składnikiem losowym (ϵ), który odpowiada za niezdeklarowane zakłócenia, błędy pomiarowe oraz czynniki nieujęte w modelu (Gorriz i in., 2026; Ampuła, 2014; Puchalski, Warszawa, 2022; Warszawa, Puchalski, 2020). W badaniach prognostycznych dla szeregów czasowych często stosuje się rozszerzenia regresji liniowej o składniki dynamiczne, w szczególności opóźnienia zmiennej zależnej (modele ARX/ARDL), co pozwala uchwycić inercję procesu i poprawić własności prognostyczne (Enders, 2015).

Standardowa metoda dopasowywania linii regresji, znana jako metoda najmniejszych kwadratów (ang. *Ordinary Least Squares* – OLS), polega na minimalizacji sumy kwadratów reszt, czyli różnic między wartościami obserwowanymi a przewidywanymi przez model (Ampuła, 2014; Puchalski, Warszawa, 2022; Silhavy i in., 2017; Warszawa, Puchalski, 2020). OLS jest powszechnie stosowana (Ampuła, 2014; Gorriz i in., 2026) i stanowi fundament dla wielu zaawansowanych technik uczenia maszynowego, takich jak regresja grzbietowa (ang. *ridge*) czy tzw. regresja Lasso (Gorriz i in., 2026). W kontekście danych czasowych dodatkowego znaczenia nabierają założenia dotyczące braku autokorelacji składnika losowego oraz ryzyka regresji pozornej przy

niestacjonarnych szeregach, dlatego w analizie uwzględnia się weryfikację stacjonarności oraz diagnostykę reszt (Enders, 2015; Stock, Watson, 2020).

W kontekście danych czasowych kluczowe znaczenie mają założenia dotyczące braku autokorelacji składnika losowego oraz ryzyka regresji pozornej przy niestacjonarnych szeregach, dlatego w analizie ekonometrycznej szeregów czasowych standardowo stosuje się testy stacjonarności ADF i KPSS zarówno dla zmiennej objaśnianej, jak i potencjalnych regresorów (Ghouse i in., 2018; Sarkodie, Owusu, 2020). Komplementarne zastosowanie obu testów pozwala na jednoznaczną klasyfikację procesów $I(0)$ vs $I(1)$ oraz identyfikację konieczności różnicowania zmiennych, eliminując niejednoznaczności pojedynczych testów jednostkowych (Sarkodie, Owusu, 2020). Taka procedura prewencyjnie ogranicza ryzyko regresji pozornej w modelach dynamicznych typu ARDL/ARX, stanowiąc fundament teoretyczny przed estymacją równań progностycznych (Ghouse i in., 2018).

Zalety regresji liniowej to jej prostota implementacji i interpretacji (Ampuła, 2014). Jest to szczególnie wartościowe narzędzie do tworzenia modeli progностycznych, nawet przy niskich rozmiarach prób (Fernandes i in., 2023). Metoda ta jest wykorzystywana do przewidywania przyszłych zdarzeń lub do sterowania zmienną zależną poprzez manipulowanie zmiennymi niezależnymi (Ampuła, 2014). Można zatem uznać, że regresja umożliwia analizę zależności statystycznych i identyfikację współzmienności kluczowych czynników, natomiast wnioskowanie przyczynowe wymaga spełnienia dodatkowych założeń identyfikacyjnych.

Mimo swojej użyteczności, regresja liniowa obarczona jest istotnymi ograniczeniami i powinna być stosowana z pełną świadomością założeń oraz warunków poprawności (Ampuła, 2014; Hazimeh, Mazumder, 2020; Zapf i in., 2024). Wymaga spełnienia kilku kluczowych założeń, takich jak liniowość, niezależność, homoskedastyczność i brak doskonałej współliniowości (Ampuła, 2014; Aguinis i in., 2005; Poole, O'Farrell, 1971; Zapf i in., 2024). Niespełnienie tych założeń, zwłaszcza homoskedastyczności, może sprawić, że analizy oparte na teście F stają się nieprecyzyjne (obciążone) i statystycznie nieważne. Jest to widoczne w przypadkach z heteroskedastycznością, gdzie założenia testu F nie zostają spełnione (Baum, Lewbel, 2019; Gorris i in., 2026). W kontekście uczenia maszynowego, chociaż testy istotności statystycznej i przedziały predykcji mogą być przeprowadzane na resztach modeli uczenia

maszynowego, to jest rzadko spotykane w literaturze ML (ang. *machine learning* – uczenie maszynowe), skupia się ona na empirycznych miarach wydajności (Gorritz i in., 2026). Ponadto, dla danych niezgodnych z rozkładem Gaussa, tradycyjne metody mogą być niewystarczające (Gorritz i in., 2026). W modelach z wieloma współzmiennymi i heteroskedastycznością, pojawiają się złożoności w wnioskowaniu (Cattaneo i in., 2018). Autorzy (Warsza, Puchalski, 2020) zauważyli, że w pewnych specyficznych sytuacjach, gdy mamy do czynienia ze skorelowanymi danymi pomiarowymi, i jeśli ich macierz kowariancji nie jest dodatnio ani ujemnie określona w sensie matematycznym, mianownik może stać się równy zeru ($\Delta = 0$). Autorzy zwracają również uwagę, że w literaturze dotyczącej regresji liniowej brakuje informacji na temat tej osobliwości oraz wskazówek, jak jej uniknąć lub jak postępować, gdy ona wystąpi (Warsza, Puchalski, 2020).

Sztuczne sieci neuronowe (ang. *Artificial Neural Networks* – ANN) to nieliniowe narzędzia do modelowania danych statystycznych, zdolne do opisywania złożonych interakcji między danymi wejściowymi i wyjściowymi oraz odkrywania wzorców (Mondal i in., 2025; Qamar, Zardari, 2023). Zalicza się je do najważniejszych i najbardziej rozpowszechnionych klas algorytmów uczenia maszynowego (Haykin, 2009; Schmidhuber, 2015). Ich koncepcja została zainspirowana biologicznymi strukturami mózgu, w szczególności mechanizmem przetwarzania informacji przez neurony (Haykin, 2009; Schmidhuber, 2015). Podstawowym elementem ANN jest sztuczny neuron, który przyjmuje sygnały wejściowe, przekształca je za pomocą wag i funkcji aktywacji, a następnie przekazuje wynik dalej do kolejnych warstw (Heaton, 2017).

Formalnie, aktywacja pojedynczego neuronu opisana jest równaniem (Heaton, 2017):

$$Y = f\left(\sum_{i=1}^n w_i x_i + b\right)$$

gdzie:

- x_i – sygnały wejściowe,
- w_i – wagi,
- b – bias (składnik przesunięcia),
- $f(\cdot)$ – nieliniowa funkcja aktywacji (np. sigmoid, ReLU, tanh),
- Y – sygnał wyjściowy neuronu

Sieci neuronowe występują w różnych architekturach – od klasycznych, wielowarstwowych perceptronów (MLP), poprzez specjalistyczne sieci konwolucyjne (CNN) i rekurencyjne (RNN), aż po nowoczesne modele oparte na architekturze transformerów (LeCun, Bengio, Hinton, 2015).

Do głównych zalet sztucznych sieci neuronowych należą (Hornik, 1991; LeCun i in., 2015; Mondal i in., 2025; Pires i in., 2023; Schmidhuber, 2015):

- uniwersalność i wszechstronność: wykazują zdolność do adaptacyjnego uczenia się na podstawie dostarczonych danych treningowych lub wcześniejszych doświadczeń;
- skalowalność i zdolność do przetwarzania dużych zbiorów danych: mogą uczyć się i podejmować decyzje w oparciu o masowe i złożone dane wejściowe, takie jak obrazy, głos i tekst. Architektury głębokich sieci neuronowych zapewniają kompozycyjne modele, w których obiekt jest przedstawiany jako warstwowa kompozycja prymitywów, gdzie cechy z niższych warstw mogą być łączone z tymi z innych warstw;
- automatyczne wykrywanie cech: w przeciwieństwie do regresji liniowej, sztuczne sieci ANN są zdolne do automatycznego odkrywania znaczących cech bezpośrednio z surowych danych, co eliminuje potrzebę ręcznej inżynierii cech.

Pomimo ogromnego potencjału, sztuczne sieci neuronowe nie są wolne od ograniczeń. Przede wszystkim ANN wymagają dużych zbiorów danych (Pires i in., 2023). Sieci neuronowe są często określane mianem „czarnych skrzynek”, ponieważ ich działanie może być trudne do odczytania i zrozumienia (Pires i in., 2023; Lipton, 2018). Kolejne ograniczenie stanowi ryzyko tzw. przeuczenia się (*overfitting*) – mają one dużą zdolność do dopasowania się do szumu w danych, co może wpływać na obniżanie ich skuteczności predykcyjnej (Goodfellow i in., 2016; Mondal i in., 2025; Pires i in., 2023). Jednocześnie wymagają dużych zasobów obliczeniowych, starannego przygotowania danych oraz świadomości ograniczeń związanych z interpretacją wyników.

Metodologia

W badaniu zastosowano regresję dynamiczną typu AR(2)-X (ARX) estymowaną metodą najmniejszych kwadratów (OLS), traktowaną jako benchmark prognostyczny w porównaniu z podejściami z obszaru uczenia maszynowego.

Zmienną objaśnianą (Y) stanowiły średnie arytmetyczne kwartalne ceny (transakcyjne) za 1 m² (zagregowana dla 16 miast wojewódzkich bez Torunia i Gorzowa Wielkopolskiego) w latach 2014Q1-2024Q4. Zmiennymi objaśniającymi (X), które pierwotnie zostały wybrane do modelu, w ujęciu kwartalnym, były: przychody przedsiębiorstw budowlanych (mln zł) (1), inwestycje na budynki i budowle (mln zł) (2), mieszkania na realizację których wydano pozwolenia lub dokonano zgłoszenia z projektem budowlanym (tys.) (3), mieszkania, których budowę rozpoczęto (tys.) (4), mieszkania oddane do użytkowania (tys.) (5), stopa oprocentowania referencyjna (stan na koniec kwartału) (%) (6), PKB wygenerowane przez budownictwo (mld zł) (7), przeciętna liczba osób w budownictwie (tys.) (8), stopa bezrobocia (%) (9), przeciętne miesięczne wynagrodzenie nominalne brutto (zł) (10) oraz inflacja (%) (11). Po zbadaniu poziomu zmienności rozumianego jako iloraz odchylenia standardowego i średniej zmienna 8 została odrzucona, ponieważ nie spełniała kryterium poziomu zmienności powyżej 10%.

Ze względu na możliwość występowania niestacjonarności szeregów czasowych przeprowadzono testy ADF oraz KPSS dla zmiennej objaśnianej i zmiennych objaśniających (w poziomach oraz dla pierwszych różnic). Wyniki wskazywały, że dla większości analizowanych zmiennych nie można przyjąć stacjonarności w poziomach, co uzasadniało dodatkową kontrolę ryzyka regresji pozornej. Zastosowano regresję dynamiczną z dwoma opóźnieniami zmiennej objaśnianej ($cena_{-1}$; $cena_{-2}$) oraz współczesnymi zmiennymi objaśniającymi. Liczbę opóźnień ograniczono do dwóch ze względu na kwartalną częstotliwość danych oraz relatywnie krótką próbę, aby zachować stabilność estymacji i uniknąć nadmiernej utraty obserwacji. Dobór regresorów przeprowadzono w sposób obiektywny poprzez pełne przeszukanie przestrzeni modeli w rozważanej klasie: przetestowano wszystkie kombinacje podzbiorów X1-X7 oraz X9-X11, przy czym w każdej specyfikacji uwzględniano $cena_{-1}$; $cena_{-2}$. Kryterium wyboru stanowiło kryterium informacyjne BIC minimalizowane w próbie estymacyjnej, natomiast jakość prognoz oceniano dodatkowo poza próbą na danych z 2024 r. Dodatkowo, aby ograniczyć ryzyko regresji pozornej wynikające z niestacjonarności części szeregów, przetestowano stacjonarność reszt wybranej specyfikacji (ADF i KPSS), uzyskując wyniki zgodne ze stacjonarnością reszt. Najlepszą specyfikację (o najniższej wartości BIC) stanowi model:

$$y = -273 + 5,35X_3 - 7,28X_5 + 3,72X_7 + 0,11X_{10} + 1,33cena_{-1} - 0,38cena_{-2}$$

W celu prognozowania cen transakcyjnych zastosowano również model sztucznej sieci neuronowej typu MLP (Multi-Layer Perceptron), który stanowi nieliniowe uogólnienie modelu regresyjnego. Model ten pozwala na uchwycenie bardziej złożonych zależności pomiędzy zmiennymi objaśniającymi, a zmienną objaśnianą, w tym nieliniowych efektów reakcji rynku mieszkaniowego na egzogeniczne zmienne makroekonomiczne (Bishop, 1995; Hornik i in., 1989). Zastosowanie sieci jest uzasadnione dynamicznym oraz wieloczynnikowym charakterem rynku oraz możliwością adaptacyjną modelu do zmieniających się warunków gospodarczych (Gupta, Miller, 2012; Lim i in., 2016).

Zastosowana w modelu zmienna objaśniana została wystandaryzowana w postaci przyrostów logarytmicznych $\Delta \log(Y_t) = \log(Y_t) - \log(Y_{t-1})$, co ogranicza heteroskedastyczność oraz umożliwia interpretację w kategoriach procentowych zmian cen (Stock i Watson, 2020; Wooldridge, 2009). Prognozy uzyskane w skali $\Delta \log(Y)$ zostały skonwertowane do poziomu cen transakcyjnych zgodnie z relacją $\hat{Y}_t = Y_{t-1} \cdot e^{\Delta \log(\hat{Y}_t)}$ – co umożliwia porównanie predykcji modelu z rzeczywistą ceną 1 m² na rynku wtórnym (Hamilton, 1994). Do modelu wprowadzono autoregresyjną zmienną pierwszego rzędu (Y_{t-1}), która pozwala modelowi na wykorzystanie informacji o przeszłym poziomie cen. Ponadto, uwzględniono sezonowość cyklu rocznego poprzez wprowadzenie trzech dychotomicznych zmiennych (Q_2, Q_3, Q_4), przy czym pierwszy kwartał stanowi kategorię odniesienia (Harvey, 1981).

Z pełnego zbioru 11 egzogenicznych zmiennych makroekonomicznych zastosowano wieloetapową selekcję, łącząc filtrację statystyczną oraz ocenę informacyjności modeli wstępnych:

1. Standaryzacja częstotliwości danych – wszystkie szeregi przekształcono do częstotliwości kwartalnej (Hamilton, 1994).
2. Utworzono transformacje⁴ (Enders, 2015; Stock, Watson, 2020):
 - $\text{Lag}(X_{t-1})$ – uchwycenie opóźnionych reakcji rynku;
 - $\Delta \log(X_t)$ – zatem $\log(X_t) - \log(X_{t-1})$ – stabilizacja wariancji i interpretacja w kategoriach tempa wzrostu;

⁴ Transformacje lag oraz $\Delta \log$ konstruowano wyłącznie na obserwacjach z okresów wcześniejszych niż prognozowany ($t-1$), co eliminuje ryzyko *look-ahead bias* oraz odzwierciedla rzeczywisty proces podejmowania decyzji w czasie.

3. Filtracja statystyczna (Elisseef i Guyon, 2003; Greene, 2020; Mukaka, 2012):
 - usunięto szeregi o niskiej zmienności;
 - obliczono korelację Pearsona i Spearmana z ceną $\Delta \log(Y)$. Zachowano zmienne dla których $|r| \geq 0,2$;
 - sprawdzono współliniowość poprzez wskaźnik VIF (*Variance Inflation Factor*). W parach o $VIF > 5$ zachowywano zmienną, która wykazywała większą stabilność wpływu w czasie.
4. Ocena informacyjności (Breiman, 2001; Lundberg, Lee, 2017):
 - Gini importance – średnią redukcję nieczystości;
 - istotność permutacji – spadek trafności po losowym permutowaniu zmiennej;
 - wartości SHAP – marginalny wkład zmiennej w predykcję.

Wyniki analiz zostały uporządkowane w formie rankingu i poddane uśrednieniu. Następnie wyłoniono zestaw pięciu zmiennych o najwyższej wartości predykcyjnej, przy jednoczesnym zachowaniu stabilności kierunku wpływu. Ostatecznie do zbioru zmiennych egzogenicznych włączono te, które spełniły oba kryteria w najbardziej zrównoważony sposób:

1. PKB wygenerowane przez budownictwo (mld zł) – lag;
2. mieszkania oddane do użytkowania (tys.) – $\Delta \log$;
3. stopa oprocentowania referencyjna (stan na koniec okresu) – $\Delta \log$;
4. przeciętne miesięczne wynagrodzenie nominalne brutto (zł) – $\Delta \log$;
5. inflacja (%) – $\Delta \log$.

Po wytypowaniu zmiennych objaśniających przystąpiono do doboru struktury sieci i parametrów uczenia. Z uwagi na to, że zmienna objaśniana została przetransformowana do postaci przyrostu logarytmicznego $\Delta \log(Y_t)$ model zwraca prognozę tempa zmian między kolejnymi kwartałami; w przypadku małej próbki obserwacji (40 okresów) istnieje ryzyko predyktowania zbyt gwałtownych zmian pomiędzy okresami, prowadzących do nienaturalnych skoków cen (Enders, 2015; Hamilton, 1994). W celu zapewnienia realizmu dynamiki prognostycznej zastosowano korektę maksymalnej zmiany na poziomie $|\Delta \log(\hat{Y}_t)| \leq 3\%$ na kształt praktyki winsoryzacji; wartość ta została wyznaczona empirycznie, jako maksymalne tempo wzrostu kwartalnego obserwowane w danych historycznych na zbiorze uczącym – zabieg ten stabilizuje

model oraz ogranicza propagację pojedynczych anomalii (Seheult i in., 1989; Wilcox, 2012). Sieć uczona była algorytmem wstecznej propagacji błędu z optymalizatorem Adam z domyślnymi parametrami ($\beta_1 = 0,9$; $\beta_2 = 0,999$) przez 100 epok, co jest szczególnie korzystne w przypadku danych o zmiennym tempie trendu (Kingma i Ba, 2014). Walidacji modelu dokonano za pomocą *expanding-window cross-validation*. Początkowe okno treningowe obejmowało cztery pierwsze obserwacje, a w każdej kolejnej iteracji było rozszerzane o jedną obserwację; dokładność prognozy oceniano dla horyzontu jednookresowego ($t+1$), by uniknąć efektu *look-ahead* (Hyndman, Athanasopoulos, 2021). Jako funkcję aktywacji zastosowano ReLU, która umożliwia modelowi aproksymację nieliniowych zależności przy zachowaniu stabilności obliczeń. Funkcja ReLU jest standardem w modelach MLP ze względu na brak problemu zanikania gradientu charakterystycznego dla funkcji sigmoidalnych (Glorot i in., 2011). W celu doboru optymalnego modelu wykorzystano procedurę *Grid search* na przestrzeni hiperparametrów, która obejmowała (Bergstra, Bengio, 2012):

$$\left\{ \begin{array}{l} \text{warstwy ukryte} \in \{(3,); (4,); (4,2); (6,2)\} \\ \lambda_{L1} \in \{0,01; 0,03; 0,1\} \\ \text{funkcja straty} \in \{"MSE"; "Huber"\} \\ \text{learning rate} \in \{0,001; 0,003\} \\ \text{batch size} \in \{4, 6, 8\} \end{array} \right.$$

Zatem łącznie wzięto pod uwagę 144 modele, a jako kryterium wyboru zastosowano minimalizację błędu walidacyjnego MAPE. W trakcie badania zaobserwowano, że większe architektury ((4,2); 6,2)) prowadziły do przeuczenia objawiającego się niskim błędem na próbie uczącej, lecz wyraźnie pogorszonym błędem walidacyjnym oraz nierealistyczną dynamiką prognoz. Ostatecznie wybrano architekturę: *MLP*($9 \rightarrow 4 \rightarrow 1$), *ReLU*, *hidden* = (4,), $\lambda_{L1} = 0,03$, *loss* = *Huber*, $\eta = 0,001$, *batch* = 6.

Parametry te zostały wyznaczone jako optymalne w rozumieniu kompromisu pomiędzy zdolnością aproksymacyjną modelu, a jego stabilnością prognostyczną. Funkcja straty Huber okazała się w tym kontekście bardziej odpowiednia niż klasyczny błąd średniokwadratowy (MSE), ponieważ ogranicza wpływ obserwacji odstających na gradient podczas uczenia (Heaton, 2017). Jest to szczególnie istotne w przypadku rynku nieruchomości, gdzie na wartość

wskaźników makroekonomicznych mogą okresowo oddziaływać incydentalne wstrząsy podażowo-popytowe lub jednorazowe zmiany regulacyjne (Siemińska, 2011). Regularyzacja L1 dodatkowo pełniła funkcję *miękkiej selekcji wag*, zapobiegając nadmiernej czułości modelu na zmienne o słabym wpływie. Wybrany model charakteryzuje się 9 wejściami, odpowiadającymi: autokorelacji pierwszego rzędu Y_{t-1} , trzem zmiennymi sezonowymi Q_2, Q_3, Q_4 oraz pięciu zmiennymi egzogenicznymi opisującym sytuację makroekonomiczną rynku budowlanego. Taka konstrukcja pozwala modelowi jednocześnie uwzględniać zarówno krótkookresową dynamikę inercyjną cen, cykliczność sezonową, jak i wpływ czynników fundamentalnych.

Wyniki

Dane wykorzystane w analizie zostały zestawione w Tabeli 1. Obejmują one okres od I kwartału 2014 r. do IV kwartału 2024 r. Do estymacji modelu wykorzystano dane z pierwszej dekady tego przedziału, tj. od I kwartału 2014 r. do IV kwartału 2023 r. Ostatnie cztery obserwacje (2024 r.) posłużyły do weryfikacji jakości prognoz oraz obliczenia błędów typu *ex ante* i *ex post*. Parametry modelu należy interpretować jako warunkowe zależności predykcyjne w ramach przyjętej specyfikacji dynamicznej, a nie jako efekty przyczynowe.

Dla zmiennych w modelu oznacza to:

- wzrost liczby mieszkań, na realizację których wydano pozwolenia lub dokonano zgłoszenia z projektem budowlanym (X3) o 1 tys. jednostek jest związany ze wzrostem prognozowanej średniej ceny 1 m² o ok. 5,35 zł;
- wzrost liczby mieszkań oddanych do użytkowania (X5) o 1 tys. jednostek jest związany ze spadkiem prognozowanej średniej ceny 1 m² o ok. 7,28 zł;
- wzrost PKB wytworzonego przez budownictwo (X7) o 1 jednostkę (zgodnie z jednostką w danych) jest związany ze wzrostem prognozowanej średniej ceny 1 m² o ok. 3,72 zł;
- wzrost przeciętnego nominalnego miesięcznego wynagrodzenia brutto (X10) o 1 zł jest związany ze wzrostem prognozowanej średniej ceny 1 m² o ok. 0,11 zł;

- współczynniki przy $cena_{-1}$; $cena_{-2}$ potwierdzają istotną inercję procesu cenowego: wartość bieżąca jest w znacznym stopniu determinowana przez poziom cen w dwóch poprzednich kwartałach.

Stworzony model został wykorzystany do predykcji średnich cen nieruchomości na rynku wtórnym dla 16 miast wojewódzkich w Polsce w czterech kwartałach 2024 r. (Tabela 2).

Tabela 1. Dane wykorzystane do modelu

Czas	Cena za m ²	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁
I 2014	4305,36	16279,00	7118,80	34,80	32,20	35,60	2,50	19391,00	408,60	13,50	3895,31	0,60
II 2014	4353,91	40834,70	17627,80	41,90	40,00	30,90	2,50	29962,00	410,30	12,00	3739,97	0,30
III 2014	4395,25	69060,80	30039,60	43,80	42,30	33,70	2,00	35884,30	410,30	11,50	3781,14	-0,30
IV 2014	4377,21	101908,80	48985,50	36,50	33,60	43,00	2,00	41269,20	416,90	11,40	3942,67	-0,70
I 2015	4382,60	17772,40	8217,80	38,60	31,80	31,70	1,50	21853,80	383,70	11,50	4054,89	-1,50
II 2015	4381,84	41964,50	20034,30	48,60	48,50	32,30	1,50	33102,20	382,90	10,20	3854,88	-0,90
III 2015	4431,32	70499,30	34023,10	51,20	47,70	37,40	1,50	39213,40	389,30	9,70	3895,33	-0,70
IV 2015	4415,74	105919,70	53633,00	50,40	40,30	46,40	1,50	47474,80	394,50	9,70	4066,95	-0,60
I 2016	4425,82	16125,00	7241,80	42,70	34,30	37,40	1,50	20931,70	382,50	9,90	4181,49	-0,90
II 2016	4460,69	40284,10	16962,00	55,60	52,00	36,30	1,50	32111,10	384,00	8,70	4019,08	-0,90
III 2016	4498,84	65243,30	27823,60	56,20	46,80	38,30	1,50	34801,90	383,00	8,30	4055,04	-0,80
IV 2016	4540,22	94447,10	42659,30	57,00	40,80	51,30	1,50	44083,60	391,20	8,20	4218,92	0,20
I 2017	4535,49	16718,10	6455,60	60,70	44,20	40,50	1,50	19629,90	382,60	8,00	4353,55	2,00
II 2017	4673,72	41408,60	15469,90	70,60	61,80	37,80	1,50	32295,00	382,20	7,00	4220,69	1,80
III 2017	4772,84	69313,50	26189,10	59,60	54,80	45,90	1,50	37831,30	389,50	6,80	4255,59	1,90
IV 2017	4830,63	102769,70	43758,90	60,00	45,10	54,10	1,50	44048,60	401,40	6,60	4516,69	2,20
I 2018	4914,05	20019,10	7239,20	66,60	48,00	44,60	1,50	24624,80	399,90	6,60	4622,84	1,50
II 2018	5064,94	48485,70	18191,50	65,60	65,70	38,20	1,50	39032,50	405,10	5,80	4521,08	1,70
III 2018	5237,41	80641,30	31388,20	61,80	60,30	47,00	1,50	43881,00	413,90	5,70	4580,20	2,00
IV 2018	5362,68	117197,40	51831,60	63,60	47,90	55,30	1,50	50902,80	419,60	5,80	4863,74	1,40
I 2019	5472,61	19841,70	9955,00	57,30	53,90	47,40	1,50	27625,80	422,10	5,90	4950,94	1,20
II 2019	5676,53	50835,40	23321,70	72,30	61,50	47,10	1,50	42539,90	423,60	5,30	4839,24	2,40

Czas	Cena za m ²	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁
III 2019	5828,08	84533,80	37945,40	71,80	63,10	51,20	1,50	47373,60	424,40	5,10	4931,59	2,80
IV 2019	6026,11	121185,70	61162,10	67,40	58,80	61,70	1,50	52346,60	432,60	5,20	5198,58	2,80
I 2020	6212,91	22334,70	11569,70	59,40	52,50	49,60	1,00	30195,40	426,10	5,40	5331,47	4,50
II 2020	6445,40	51898,70	24348,00	62,90	47,40	47,50	0,10	43278,60	429,90	6,10	5024,48	3,20
III 2020	6498,41	82396,80	37645,90	70,10	67,40	59,20	0,10	44192,70	421,50	6,10	5168,93	3,00
IV 2020	6651,32	119187,90	57507,80	83,70	56,50	64,60	0,10	50551,60	427,90	6,80	5457,98	2,80
I 2021	6732,27	23863,80	10925,60	84,80	63,70	53,10	0,10	24376,90	419,60	6,90	5681,56	2,70
II 2021	6985,95	55900,00	24208,70	87,10	80,90	52,40	0,10	43689,70	425,40	6,40	5504,52	4,50
III 2021	7187,19	90962,80	40093,50	83,50	71,80	58,70	0,10	44717,30	422,30	6,10	5657,30	5,40
IV 2021	7456,13	144971,30	63667,80	85,80	61,10	70,60	1,75	49548,50	430,30	5,80	5995,09	7,70
I 2022	7734,93	28377,30	12112,40	78,00	53,30	54,70	3,50	29134,20	423,70	5,80	6235,22	9,70
II 2022	8070,09	70158,90	29181,10	93,40	66,40	54,50	6,00	48015,30	425,40	5,20	6156,25	13,90
III 2022	8248,52	113042,60	48352,40	66,20	45,10	57,80	6,75	48077,80	435,90	5,10	6480,67	16,30
IV 2022	8209,61	165226,90	76320,40	60,70	35,50	71,50	6,75	55878,40	449,80	5,20	6733,49	17,30
I 2023	8167,47	34387,70	14346,90	52,10	38,60	55,10	6,75	33783,10	417,20	5,40	7124,26	17,00
II 2023	8349,56	78337,70	33261,90	60,00	47,00	56,80	6,75	57409,40	423,60	5,10	7005,76	13,10
III 2023	8672,71	123855,30	55179,80	62,30	53,30	49,00	6,00	54555,60	422,70	5,00	7194,95	9,70
IV 2023	8965,75	179730,90	86175,10	67,30	50,20	60,40	5,75	66893,20	429,80	5,10	7540,36	6,40
I 2024	9419,38	32351,50	13239,20	70,00	60,10	48,40	5,75	33927,50	415,3	5,30	8147,38	2,80
II 2024	9918,43	75222,30	31092,10	74,20	62,30	47,20	5,75	57768,40	415,1	4,90	8038,41	2,50
III 2024	10221,11	120528,50	50451,70	76,30	59,10	49,40	5,75	56933,40	425,8	5,00	8161,62	4,50
IV 2024	10222,81	176824,10	79131,20	71,20	52,30	55,10	5,75	68195,70	430,2	5,10	8477,21	4,80

Źródło: opracowanie własne na podstawie: Kwartalne wskaźniki cen towarów i usług konsumpcyjnych od 1995 r. (GUS), Kwartalne wskaźniki makroekonomiczne (GUS), Baza cen nieruchomości mieszkaniowych (NBP).

Tabela 2. Prognoza średnich cen nieruchomości na rynku wtórnym dla 16 miast wojewódzkich w Polsce w 2024 r. wraz z błędami *ex ante*

Kwartał	Cena za m ²	Prognoza	Błąd <i>ex ante</i>	95% przedział predykcji
I 2024	9419,38	9322,98	56,66	9207,42-9438,54
II 2024	9918,43	9792,87	94,11	9600,92-9984,82
III 2024	10221,10	10284,67	122,22	10035,41-10533,93
IV 2024	10222,80	10764,15	143,21	10472,08-11056,22

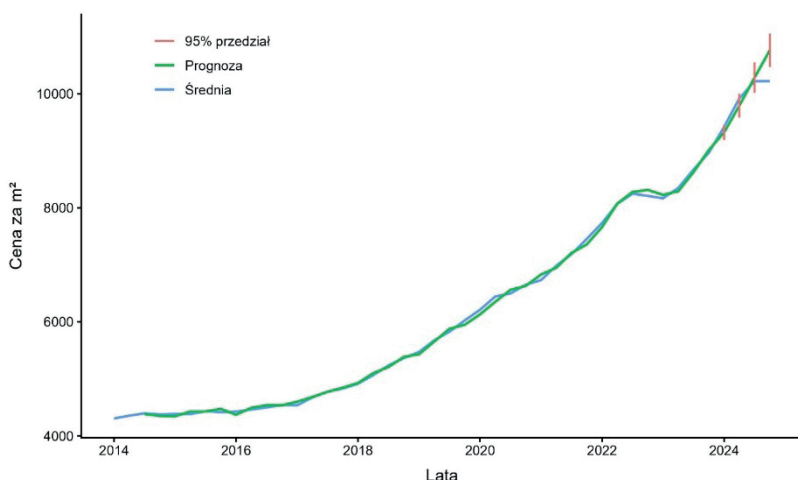
Źródło: opracowanie własne z wykorzystaniem programu R.

W analizowanym przypadku wartości błędu typu *ex ante* dla czterech prognozowanych kwartałów roku 2024 wykazują tendencję rosnącą (od 56,66 do 143,21), co odzwierciedla wzrost niepewności prognoz wraz z wydłużaniem horyzontu od ostatniego okresu estymacji (IV kwartał 2023 r.). Jest to zjawisko typowe dla modeli szeregów czasowych z komponentem autoregresyjnym: w prognozie dynamicznej kolejne wartości prognozowane wykorzystują wcześniejsze prognozy jako zmienne opóźnione, co prowadzi do kumulacji niepewności w horyzoncie wielookresowym. Wartości *ex ante* interpretowane są jako odchylenie standardowe błędu prognozy, wyznaczone na podstawie oszacowanej wariancji składnika losowego oraz parametrów części autoregresyjnej AR(2). Na tej podstawie skonstruowano 95% przedziały predykcji dla prognoz na 2024 r., które – przy przyjętych założeniach modelu – powinny w długim okresie obejmować rzeczywistą wartość zmiennej w ok. 95% przypadków.

W ocenie wiarygodności prognoz istotne znaczenie mają błędy *ex post*, które umożliwiają porównanie jakości predykcyjnej modelu regresji dynamicznej (AR(2)-X) z alternatywnymi podejściami, w tym modelami opartymi na sieciach neuronowych. Obliczenie miar ME, MAE, MPE i MAPE stanowi standardowy element weryfikacji jakości prognoz (Rogalska, 2018). Miary wyznaczono na podstawie prognoz poza próbą dla czterech kwartałów 2024 r.

Średni błąd prognozy (ME) wyniósł -95,730 zł/m². Ujemna wartość ME oznacza, że model wykazywał przeciętną tendencję do przeszacowywania poziomu cen – prognozy były średnio wyższe od wartości rzeczywistych o ok. 95,73 zł/m². Średni błąd bezwzględny (MAE) osiągnął poziom 206,710 zł/m², co wskazuje, że przeciętne odchylenie prognoz od obserwacji wynosiło ok. 206,71 zł/m². Średni błąd procentowy (MPE) wyniósł -0,907%, co również potwierdza niewielkie przeszacowanie prognoz w ujęciu względnym. Z kolei

średni absolutny błąd procentowy (MAPE) równy 2,052% oznacza, że przeciętnie prognozy różniły się od wartości rzeczywistych o ok. 2,05% (Rysunek 1).



Rysunek 1. Rzeczywiste i prognozowane średnie ceny transakcyjne 1 m² na rynku wtórnym w Polsce w latach 2014-2024

Źródło: opracowanie własne z wykorzystaniem programu R.

Podsumowując, uzyskane miary wskazują na relatywnie wysoką dokładność prognoz krótkookresowych w analizowanym horyzoncie (2024 r.). Należy jednak podkreślić, że ocena oparta jest na czterech obserwacjach poza próbą, co ogranicza możliwość uogólnienia wniosków; dlatego miary błędu należy interpretować jako ocenę jakości predykcji w krótkim, jednorocznym oknie walidacyjnym.

Dla sieci neuronowych wykorzystano te same dane wejściowe co dla modelu regresji liniowej, a prognozy cen transakcyjnych na rynku wtórnym w 16 miastach wojewódzkich dla czterech kwartałów 2024 r. przedstawiono poniżej (Tabela 3).

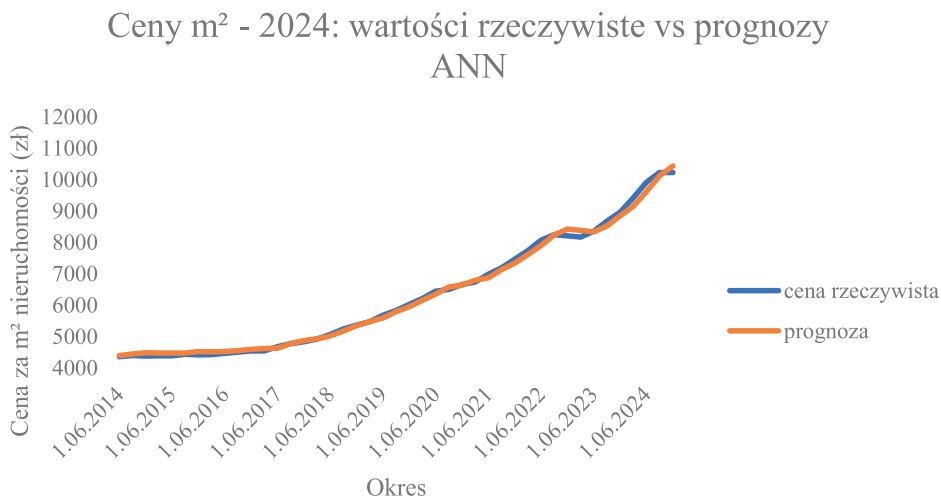
Tabela 3. Prognoza średnich cen nieruchomości wg modelu MLP oraz błędy *ex post*

Kwartał	Cena rzeczywista	Prognoza	Error	%APE
31.03.2024	9419,38	9149,51	269,87	2,87
30.06.2024	9918,43	9612,44	305,99	3,09
30.09.2024	10221,11	10121,72	99,39	0,97
31.12.2024	10222,81	10430,60	-207,79	2,03

Źródło: opracowanie własne z wykorzystaniem biblioteki scikit-learn oraz tensorflow.

W analizowanym przypadku wartości błędów prognozy w poszczególnych kwartałach roku 2024 mieszczą się w przedziale od 0,97% do 3,09%, co świadczy o stosunkowo stabilnej dokładności modelu w horyzoncie krótkookresowym. Najwyższe wartości błędu (%APE) wystąpiły w I i II kwartale 2024 r., czyli w okresie przyspieszenia dynamiki cen mieszkań. Jest to zgodne z obserwacją, że nieliniowe modele uczone na krótkim szeregu mogą reagować z opóźnieniem na nagłe zmiany trendu (Heaton, 2017). Od III kwartału dokładność prognozy ulega poprawie, a w IV kwartale model zachowuje poprawną trajektorię cen, choć obserwuje się niewielkie przeszacowanie poziomu cen, czego odzwierciedleniem jest ujemny błąd ME w ostatnim okresie.

Do analizy zastosowano standardowe miary: ME, MAE, MPE oraz MAPE. Średni błąd prognozy (ME) wyniósł 116,87, co oznacza niewielką tendencję modelu do niedoszacowania cen prognozowane wartości były średnio niższe od rzeczywistych o 116,87 zł/m². Średni błąd bezwzględny (MAE) osiągnął poziom 220,76zł/m², co wskazuje, że przeciętne odchylenie prognoz od cen rzeczywistych wyniosło 220,76 zł/m². Średni błąd procentowy (MPE) równy 1,22 potwierdza występowanie niewielkiego systematycznego niedoszacowania, rzędu 1,22%. Z kolei średni absolutny błąd procentowy (MAPE) wyniósł 2,24%, co oznacza, że przeciętnie prognozy odbiegały od wartości rzeczywistych o 2,24%. Zgodnie z klasyfikacją stosowaną w badaniach prognostycznych, wartości $1\% < \text{MAPE} \leq 3\%$ interpretowane są jako prognozy wysokiej dokładności (Rogalska, 2018) (Rysunek 2).



Rysunek 2. Rzeczywista i prognozowana zmiana cen nieruchomości – sieci neuronowe
Źródło: opracowanie własne z wykorzystaniem programu Excel.

Osiągnięte błędy *ex post* mieszczą się w zakresie uznawanym za akceptowalny dla prognoz rynku nieruchomości, co potwierdza użyteczność modelu w zastosowaniach praktycznych.

Wnioski

Przeprowadzone badanie miało na celu porównanie skuteczności prognozowania cen nieruchomości mieszkaniowych w Polsce z wykorzystaniem modelu regresji dynamicznej typu AR(2)-X estymowanego metodą najmniejszych kwadratów oraz sztucznych sieci neuronowych. Analiza oparta została na danych GUS dotyczących średnich kwartalnych cen transakcyjnych za 1 m² na rynku wtórnym w 16 miastach wojewódzkich w latach 2014-2024.

Wyniki wskazują na nieznaczną przewagę modelu regresji dynamicznej typu AR(2)-X w zakresie dokładności prognoz krótkookresowych w analizowanym oknie walidacyjnym. Dla prognoz poza próbą (2024Q1-2024Q4) uzyskano MAPE = 2,05% oraz MAE = 206,71 zł/m². Jednocześnie odnotowano ujemny ME = -95,73 zł/m² oraz MPE = -0,91%, co oznacza niewielką przeciętną tendencję modelu do przeszacowywania poziomu cen (prognozy średnio nieco przewyższały wartości obserwowane). Istotną zaletą podejścia regresyjnego jest możliwość jawnego wyznaczania 95% przedziałów predykcji

dla prognoz *ex ante*, co pozwala oceniać niepewność prognoz w sposób spójny z założeniami modelu.

Model sieci neuronowej (MLP), pomimo zastosowania zaawansowanych technik regularyzacji i walidacji kroczącej, osiągnął gorsze rezultaty: średni absolutny błąd procentowy (MAPE) osiągnął poziom 2,24%, natomiast średni błąd bezwzględny (MAE) wyniósł 220,76. Różnice między metodami są umiarkowane, jednak w analizowanym przypadku wskazują na lepszą kalibrację progностyczną modelu regresji dynamicznej.

Gorsze wyniki sieci neuronowej można przypisać kilku czynnikom, w tym ograniczonej wielkości próby, obejmującej 40 obserwacji kwartalnych. Ponadto, względna prostota zależności w badanym zbiorze danych mogła nie wymagać zaawansowanego modelowania nieliniowego, co uwidoczniło przewagę prostszych modeli liniowych w kontekście dokładności progностycznej.

Otrzymane wyniki wskazują, że w przypadku niewielkich i stabilnych zbiorów danych tradycyjne modele liniowe mogą osiągać porównywalną lub nawet wyższą dokładność progностyczną niż sieci neuronowe. Jednocześnie większe możliwości uogólniania charakterystyczne dla ANN mogą ujawniać się dopiero przy większej liczbie obserwacji oraz bardziej złożonej strukturze relacji. W takich sytuacjach uzasadnione może być podejście hybrydowe, które łączy stabilność modeli regresyjnych z elastycznością nieliniowego modelowania sieciowego.

Ograniczenia i przyszłe badania

Przeprowadzone badanie charakteryzuje się następującymi ograniczeniami, które mogą wpływać na interpretację i generalizację wyników. Po pierwsze, ograniczony zakres czasowy analizy, obejmujący dane z lat 2014-2024, może nie odzwierciedlać pełnego spektrum długoterminowych cykli koniunkturalnych i strukturalnych na rynku nieruchomości. Dłuższy horyzont czasowy pozwoliłby na lepsze uchwycenie zmienności i dynamiki tego złożonego segmentu gospodarki w Polsce.

Po drugie, agregacja danych do poziomu 16 miast wojewódzkich może ograniczać dokładność analizy, ponieważ ceny i czynniki wpływające na nie mogą znacznie różnić się w zależności od konkretnej lokalizacji.

Po trzecie, model sztucznych sieci neuronowych (MLP) został zbudowany na podstawie ograniczonej wielkości próby, obejmującej 40 obserwacji

kwartalnych. Jest to często niewystarczające dla pełnego wykorzystania potencjału nieliniowych modeli uczenia maszynowego, które uczą się najlepiej, gdy mają dostęp do tzw. dużych zbiorów danych. Mimo zastosowania technik przeciwdziałających przeuczeniu, takich jak standaryzacja cech, regularyzacja L1, wczesne zatrzymanie treningu i redukcja współczynnika uczenia, mała próba mogła ograniczyć zdolność ANN do modelowania złożonych zależności i optymalizacji hiperparametrów.

Dodatkowo, należy pamiętać, że sieci neuronowe są często określane mianem „czarnych skrzynek”, ponieważ ich działanie i proces podejmowania decyzji mogą być trudne do odczytania i zrozumienia. Jest to istotne ograniczenie, szczególnie w kontekście podejmowania decyzji, gdzie cenna jest możliwość analizy przyczynowo-skutkowej.

W przypadku modelu regresji dynamicznej AR(2)-X, dla czterech prognozowanych kwartałów 2024 roku, wartości błędu typu *ex ante* wykazywały tendencję rosnącą, od 56,66 do 143,21. Ilustruje to stopniowy wzrost niepewności prognozy wraz z oddalaniem się od ostatniego okresu estymacji.

Literatura

1. Aguinis, H., Beaty, J. C., Boik, R. J., Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: a 30-year review. *Journal of Applied Psychology*, 90(1), 94-107. <https://doi.org/10.1037/0021-9010.90.1.94>.
2. Ampuła, D. (2014). Regresja jako metoda procesu predykcji. *Problemy Techniki Uzbrojenia*, 130(2), 67-78.
3. Baum, C. F., Lewbel, A. (2019). Advice on using heteroskedasticity-based identification. *The Stata Journal: Promoting Communications on Statistics and Stata*, 19(4), 757-767. <https://doi.org/10.1177/1536867X19893614>
4. Bergstra, J., Bengio, Y. (2012). Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, 13(1), 281-305.
5. Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press. <https://doi.org/10.1093/oso/9780198538493.001.0001>
6. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>

7. Bryx, M. (2006). Zarządzanie nieruchomością jako faza procesu inwestycyjnego. *Studia i Materiały Towarzystwa Naukowego Nieruchomości*, 14(1), 27-34.
8. Cattaneo, M. D., Jansson, M., Newey, W. K. (2018). Inference in linear regression models with many covariates and heteroscedasticity. *Journal of the American Statistical Association*, 113(523), 1350-1361. <https://doi.org/10.1080/01621459.2017.1328360>
9. Chicco, D., Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(6), 1-13. <https://doi.org/10.1186/s12864-019-6413-7>
10. Elisseeff, A., Guyon, I. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157-1182.
11. Enders, W. (2015). *Applied econometric time series* (4. edycja). Wiley.
12. Fernandes, A.C.P., Fonseca, A., Pacheco, F., Sanchez Fernandez, L. F. (2023). Water quality predictions through linear regression – a brute force algorithm approach. *MethodsX*, 10, s102153. <https://doi.org/10.1016/j.mex.2023.102153>
13. Flores Hinostroza, E. M., Mendoza, D. J., Navarro Cejas, M., Palacios Trujillo, E. P. (2025). Linear regression model to predict the use of artificial intelligence in experimental science students. *International Electronic Journal of Mathematics Education*, 20(1), em0807. <https://doi.org/10.29333/iejme/15736>.
14. Ghouse, G., Khan, S.A., Rehman, A.U. (2018). ARDL model as a remedy for spurious regression problems, performance and prospectus. *Munich Personal RePEc Archive*, 83973.
15. Glorot, X., Bordes, A., Bengio, Y. (2011). Deep sparse rectifier neural networks. W: *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (s. 315-323). PMLR.
16. Gołabeska, E. (2011). Efektywność inwestowania na rynku nieruchomości przy uwzględnieniu ryzyka. *Studia i Prace Kolegium Zarządzania i Finansów / Szkoła Główna Handlowa*, 110, 9-22.
17. Gorriz, J. M., Ramirez, J., Segovia, F., Jimenez-Mesa, C., Martinez-Murcia, F. J., Suckling, J. (2026). Statistical agnostic regression: a machine learning method to validate regression models. *Journal of Advanced Research*, 80, 503-533. <https://doi.org/10.1016/j.jare.2025.04.026>

18. Grzywińska-Rapca, M., Andruszkiewicz, K., Grzybowska-Brzezińska, M., Oktaba, P. (2024). *Determinanty preferencji klientów a perspektywa rozwoju rynku nieruchomości mieszkaniowych w Polsce*. Wydawnictwo Ostrołęckiego Towarzystwa Naukowego im. Adama Chętnika.
19. Greene, W. (2020). *Econometric analysis* (8th edition, global edition). Pearson.
20. Gupta, R., Miller, S. M. (2012). The time-series properties of house prices: a case study of the Southern California market. *The Journal of Real Estate Finance and Economics*, 44(3), 339-361. <https://doi.org/10.1007/s11146-010-9234-7>
21. Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press. <https://doi.org/10.1515/9780691218632>
22. Harvey, A. C. (1981). *The econometric analysis of time series*. Wiley.
23. Haykin, S. (2009). *Neural Networks and Learning Machines*. 3. edycja. Prentice Hall.
24. Hazimeh, H., Mazumder, R. (2020). Fast best subset selection: coordinate descent and local combinatorial optimization algorithms. *Operations Research*, 68(5), 1517-1537. <https://doi.org/10.1287/opre.2019.1919>
25. Heaton, J. (2017). Ian Goodfellow, Yoshua Bengio, and Aaron Courville: deep learning. *Genetic Programming and Evolvable Machines*, 19(1-2), 305-307. <https://doi.org/10.1007/s10710-017-9314-z>
26. Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2), 251-257. [https://doi.org/10.1016/0893-6080\(91\)90009-T](https://doi.org/10.1016/0893-6080(91)90009-T)
27. Hornik, K., Stinchcombe, M., White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359-366. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)
28. Hyndman, R. J., Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3. edycja). Otexts, Online Open-Access Textbooks.
29. Kingma, D. P., Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv*, 1412.6980. <https://doi.org/10.48550/ARXIV.1412.6980>
30. Kucharska-Stasiak, E. (2005). *Podstawy funkcjonowania rynku nieruchomości – ujęcie teoretyczne*. Wydawnictwo Uniwersytetu Łódzkiego.
31. LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>

32. Lim, W. T., Wang, L., Wang, Y., Chang, Q. (2016). Housing price prediction using neural networks. W: *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* (s. 518-522). IEEE. <https://doi.org/10.1109/FSKD.2016.7603227>
33. Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM (CACM)*, 61(10), 36-43. <https://doi.org/10.1145/3233231>
34. Lundberg, S., Lee, S.-I. (2017). A unified approach to interpreting model predictions. *arXiv*, 1705.07874. <https://doi.org/10.48550/ARXIV.1705.07874>
35. Mach, Ł. (2014). Czynniki kształtujące wartość nieruchomości mieszkaniowych w kontekście uwarunkowań makro-, mikro- oraz ultraotoczenia. *Ekonometria*, 46(4), 52-61. <https://doi.org/10.15611/ekt.2014.4.05>
36. Mondal, S., Maity, R., Nag, A. (2025). An efficient artificial neural network-based optimization techniques for the early prediction of coronary heart disease: Comprehensive analysis. *Scientific Reports*, 15(1), 4827. <https://doi.org/10.1038/s41598-025-85765-x>
37. Mukaka, M. M. (2012). Statistics corner: a guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal: The Journal of Medical Association of Malawi*, 24(3), 69-71.
38. Pires, P. B., Santos, J. D., Pereira, I. V. (2023). Artificial Neural Networks: history and state of the art. W: *Encyclopedia of Information Science and Technology, Sixth Edition*. IGI Global. <https://doi.org/10.4018/978-1-6684-7366-5.ch037>
39. Poole, M. A., O'Farrell, P. N. (1971). The assumptions of the linear regression model. *Transactions of the Institute of British Geographers*, 52, 145-158.
40. Puchalski, J., Warsza, Z. (2022). Regression and uncertainty of a straight-line for measurements of x and y variables with all correlations. *Pomiary Automatyka Robotyka*, 26(2), 47-58. https://doi.org/10.14313/PAR_244/47
41. Qamar, R., Zardari, B. A. (2023). Artificial Neural Networks: an overview. *Mesopotamian Journal of Computer Science*, 2023, 130-139. <https://doi.org/10.58496/MJCSC/2023/015>
42. Rogalska, M. (2018). Wycena nieruchomości z wykorzystaniem metody wieloczynnikowych modeli statystycznych. W: A. Sobotka, E. Radziszewska-Zielina (red.), *Inżynieria przedsięwzięć budowlanych: problemy, modele, metody* (s. 257-290). Komitet Inżynierii Lądowej i Wodnej PAN.

43. Sarkodie, S. A., Owusu, P. A. (2020). How to apply the novel dynamic ARDL simulations (dynardl) and Kernel-based regularized least squares (krls). *MethodsX*, 7, 101160. <https://doi.org/10.1016/j.mex.2020.101160>
44. Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neural Networks*, 61, 85-117. <https://doi.org/10.1016/j.neunet.2014.09.003>
45. Seheult, A. H., Green, P. J., Rousseeuw, P. J., Leroy, A. M. (1989). Robust regression and outlier detection. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 152(1), 133. <https://doi.org/10.2307/2982847>
46. Siemińska, E. (2011). *Inwestowanie na rynku nieruchomości*. Poltext.
47. Silhavy, R., Silhavy, P., Prokopova, Z. (2017). Analysis and selection of a regression model for the Use Case Points method using a stepwise approach. *The Journal of Systems and Software*, 125, 1-14. <https://doi.org/10.1016/j.jss.2016.11.029>
48. Stock, J. H., Watson, M. W. (2020). *Introduction to econometrics* (4th edition, global edition). Pearson.
49. Vakhshouri, B., Nejadi, S. (2015). Prediction of compressive strength in Light-Weight Self-Compacting Concrete by ANFIS Analytical Model. *Archives of Civil Engineering*, 61(2), 53-72. <https://doi.org/10.1515/ace-2015-0014>.
50. Warsza, Z., Puchalski, J. (2020a). Uncertainty of measurement in the linear regression method. Part 1. Straight-line and its uncertainty bands for uncorrelated measurement data. *Pomiary Automatyka Robotyka*, 24(3), 79-91. https://doi.org/10.14313/PAR_237/79
51. Warsza, Z., Puchalski, J. (2020b). Uncertainty of measurement in the linear regression method part 2. Uncertainty bands of the regression straight-line for the correlated data of y variable. *Pomiary Automatyka Robotyka*, 24(4), 61-72. https://doi.org/10.14313/PAR_238/61
52. Warsza, Z. L., Puchalski, J. (2020c). Estimation of uncertainties in indirect multivariable measurements: Part 1. Case of correlated quantities. W: R. Szweczyk, C. Zieliński, M. Kaliczyńska (red.), *Automation 2020: Towards Industry of the Future* (T. 1140, s. 309-325). Springer International Publishing. https://doi.org/10.1007/978-3-030-40971-5_29
53. Welc, J. (2009). Regresja liniowa jako narzędzie szacowania fundamentalnych współczynników beta na przykładzie spółek giełdowych z sektora maszyn przemysłowych. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, nr 65, 213-223.

54. Wilcox, R. R. (red.). (2012). *Introduction to robust estimation and hypothesis testing* (3rd edition). Academic Press. <https://doi.org/10.1016/B978-0-12-386983-8.00001-9>
55. Wooldridge, J. M. (2009). *Introductory econometrics: A modern approach* (4th edition). South-Western.
56. Zapf, A., Wiessner, C., König, I. R. (2024). Regression analyses and their particularities in observational studies. *Deutsches Ärzteblatt International*, 121, 128-134. <https://doi.org/10.3238/arztebl.m2023.0278>
57. Żróbek, S., Renigier-Biłozor, M., Żróbek, R. (2024). From administrative price to market value of real estate. The evolution of the valuation system in Poland. *Real Estate Management and Valuation*, 3, 127-142. <https://doi.org/10.2478/remav-2024-0030>
58. Zydrón, A., Walkowiak, R. (2013). Analiza atrybutów wpływających na wartość nieruchomości niezabudowanych przeznaczonych na cele budowlane w gminie Mosina. *Rocznik Ochrona Środowiska*, 15, 2911-2924.