

NAUKI O ZARZĄDZANIU

Magdalena ROGALSKA*

Zdzisław HEJDUCKI**

ANALIZA PORÓWNAWCZA PROGNOZOWANIA PRODUKCJI BUDOWLANEJ Z ZASTOSOWANIEM METOD REGRESJI KROKOWEJ, SIECI NEURONOWYCH I ARIMA. CZĘŚĆ II

W pracy analizowano możliwość prognozowania produkcji budowlano montażowej województwa dolnośląskiego metodami regresji, sieci neuronowych i ARIMA (Autoregressive Integrated Moving Average - autoregresyjny zintegrowany proces średniej ruchomej). Do prognozowania w metodzie regresji użyto danych pogodowych dziennych województwa dolnośląskiego oraz danych ekonomicznych dotyczących ilości zatrudnionych pracowników w sektorze budownictwo i średnich zarobków pracowników w tym sektorze. Przeprowadzono analizę wyników obliczając błędy ME, MAE, MPE, MAPE oraz współczynniki Theila I , I_1^2 , I_2^2 , I_3^2 . Sformułowano wnioski z obliczeń.

Słowa kluczowe: prognoza, produkcja budowlano montażowa, regresja krokowa, sieci neuronowe, ARIMA

WSTĘP

Niniejszy artykuł jest kontynuacją tematyki przedstawionej w pracy [5]. Podjęto kolejną próbę prognozowania [1, 3] produkcji budowlano montażowej w województwie dolnośląskim. Stosowano wcześniej utworzoną bazę danych pogodowych wzbogaconą o dane ekonomiczne w postaci średniego wynagrodzenia w sektorze budowlanym oraz ilości zatrudnionych pracowników w budownictwie w latach 2000-2009 w województwie dolnośląskim. Dane ekonomiczne uzyskano w Dolnośląskim Oddziale Głównego Urzędu Statystycznego we Wrocławiu. Celem badań było uzyskanie wyniku lepszego

* dr inż. Magdalena ROGALSKA – Wydział Budownictwa i Architektury Politechniki Lubelskiej

** dr hab. inż. Zdzisław HEJDUCKI - Wydział Budownictwa Lądowego i Wodnego Politechniki Wrocławskiej

od prognozy modelu ARIMA (*Autoregressive Integrated Moving Average - autoregresyjny zintegrowany proces średniej ruchomej*). W modelu ARIMA zakłada się, że można oszacować współczynniki modelu, które opisują kolejne elementy szeregu na podstawie opóźnionych w czasie poprzednich elementów (proces autoregresyjny) oraz że pozostają one pod wpływem realizacji składnika losowego w okresach przeszłych (proces średniej ruchomej). Zatem każda obserwacja składa się ze składnika losowego oraz kombinacji liniowej składników losowych z przeszłości, a wartość szeregu czasowego jest sumą składnika losowego oraz kombinacji liniowej poprzednich obserwacji. W części I pracy najlepszy wynik (prognozę z najmniejszym błędem) uzyskano właśnie, wykorzystując ten model. W pracy autorzy chcieli udowodnić, że wprowadzając do obliczeń odpowiednią ilość i jakość danych, można uzyskać lepszą prognozę niż w modelu ARIMA. Nowe, wprowadzone dane ekonomiczne będą czynnikiem modelującym obliczenia.

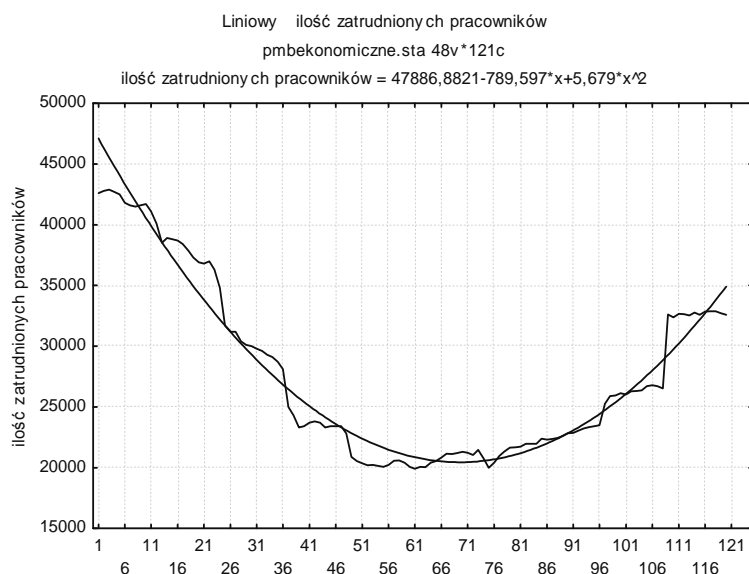
1. DANE

Do analizy przyjęto trzy rodzaje danych: dane dotyczące produkcji budowlano-montażowej, dane pogodowe oraz ekonomiczne (wynagrodzenie średnie w sektorze budowlanym rys. 1 i ilość zatrudnionych pracowników rys. 2). Dane zbierano dla województwa dolnośląskiego z lat 2000-2009. W części I pracy [5] nie uwzględniano danych ekonomicznych. Autorzy mając nadzieję na lepszą prognozę i bardziej zbliżoną linię trendu zastosowali do obliczeń dodatkowe dane. Oznaczenia danych zestawiono w tabeli 1. Nowe dane v20 i v22 przedstawiono odpowiednio na rysunkach 1 i 2.

Tabela 1. Zestawienie danych niezależnych v1 – v21 oraz zmienna zależna v22

Numer zmiennej v _i	Nazwa zmiennej j	Uwagi	Nazwa zmiennej v _i	Nazwa zmiennej	Uwagi
v1	lp	Liczba porządkowa lp=1,...,120	v12	ww	Wilgotność względna
v2	lp ²	v2=v1 ²	v13	r	Współczynnik mieszania
v3	lp ³	v3=v1 ³	v14	kw	Kierunek wiatru
v4	lp ⁴	v4=v1 ⁴	v15	ukw	Uogólniony kierunek wiatru
v5	t	Okres t=1,...,12	v16	pt	Potencjalna temperatura
v6	t ²	v6=v5 ²	v17		v17=v16 ²
v7	t ³	v7=v5 ³	v18	ept	Ekwiwalentna potencjalna temperatura
v8	t ⁴	v8=v5 ⁴	v19	wpt	Wirtualna potencjalna temperatura
v9	p	Ciśnienie atmosferyczne	v20	zp	Ilość zatrudnionych pracowników w sektorze budownictwo
v10	t _{min}	Temperatura minimalna	v21	wp	Średnie miesięczne wynagrodzenie w sektorze budownictwo
v11	t _{max}	Temperatura maksymalna	v22	pbm	Produkcja budowlano-montażowa

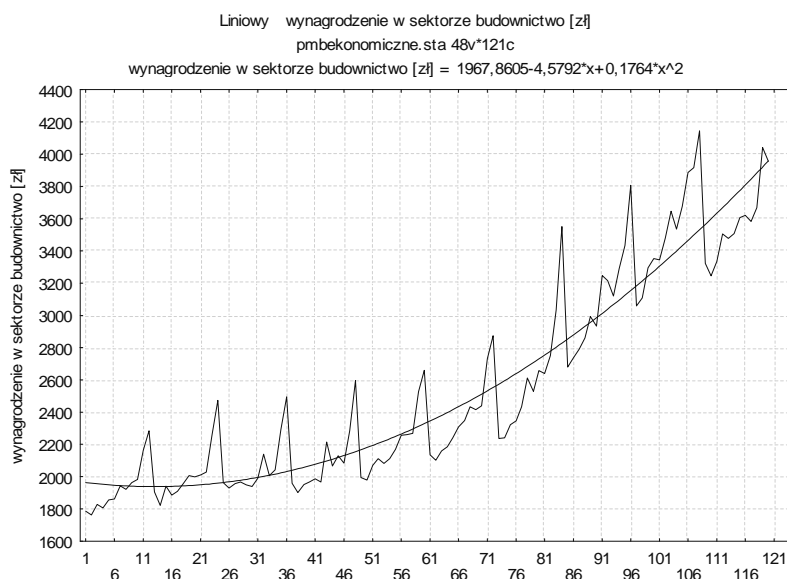
Źródło: Opracowanie własne



Rys. 1. Wykres zależności ilości zatrudnionych pracowników w sektorze budowlanym w województwie dolnośląskim w badanych okresach od stycznia 2000 do grudnia 2009. Na osi X oznaczono okresy odpowiadające kolejnym miesiącom

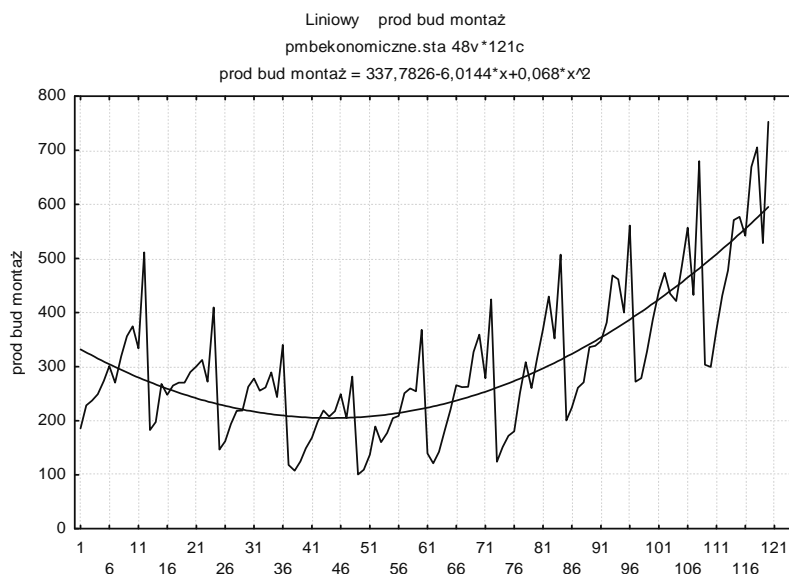
Źródło: Opracowanie własne

Wykres zależności średniego wynagrodzenia w sektorze budowlanym w województwie dolnośląskim w okresach od stycznia 2000 do grudnia 2009 jest wizualnie podobny do wykresu produkcji budowlano-montażowej (rys. 3.). Posiada wahania sezonowe oraz podobną linię trendu.



Rys. 2. Wykres zależności średniego wynagrodzenia w sektorze budowlanym w województwie dolnośląskim w badanych okresach od stycznia 2000 do grudnia 2009. Na osi X oznaczono okresy odpowiadające kolejnym miesiącom

Źródło: Opracowanie własne



Rys. 3. Wykres zależności produkcji budowlano montażowej w województwie dolnośląskim w badanych okresach od stycznia 2000 do grudnia 2009. Na osi X oznaczono okresy odpowiadające kolejnym miesiącom

Źródło: Opracowanie własne

2. PROGNOZOWANIE METODĄ REGRESJI KROKOWEJ

Obliczenia wykonano w programie STATISTICA firmy Statsoft [2, 4, 8]. Metoda regresji krokowej wstecznej polega na poszukiwaniu zależności funkcyjnych pomiędzy danymi statystycznymi. Poszukiwana wartość (w naszym przypadku wartość produkcji budowlano montażowej) zwana jest zmienną zależną, natomiast dane, które służą do jej wyznaczenia to zmienne niezależne. W wyniku obliczeń metodą regresji krokowej otrzymuje się między innymi parametr R , który jest jednym z czynników dobroci dopasowania. Parametr R (opisujący, jak mocno dane przewidywane przez model korelują z danymi eksperymentalnymi), R^2 oraz skorygowane R^2 podawane są w oknie wyników programu STATISTICA. Parametr R^2 odzwierciedla zasób zmienności opisany przez model. R^2 obniżony jest skorygowanym R^2 ze względu na dodatkowe stopnie swobody wprowadzane przez kolejne zmienne (dla modelu z jedną zmienną R^2 jest równe skorygowanemu R^2). Wprowadzanie zbyt wielu zmiennych w odniesieniu do liczby przypadków prowadzi do nadmiernego dopasowania (*overfitting*) i można uzyskać nieprawdziwy model (o R^2 równym nawet 1).

W wyniku prowadzonych obliczeń regresyjnych z wykorzystaniem bazy danych zestawionych w tabeli 2. uzyskano 6 modeli regresji. W modelach 1 i 2 zmienne niezależne to dane ekonomiczne v_{20} i v_{21} oraz dane szeregu czasowego t i lp^2 . W modelach 3,4 i 5 wykorzystano dane pogodowe i ekonomiczne. Modele różnią się między sobą danymi szeregu czasowego: w modelu 3 nie występują dane szeregu, w modelu 4 zastosowano daną okres t , a w modelu 5 daną okres podniesiony do kwadratu t^2 . Model 6 to najlepszy uzyskany model przy zastosowaniu danych pogodowych oraz szeregu czasowego lp^2 i t^3 opisany w I części pracy. Najlepszy wynik określony przez wartość maksymalną R^2 oraz potwierdzony późniejszymi obliczeniami p.4. pracy otrzymano w odniesieniu do modelu 4. Model ten przyjęto jako najlepszy

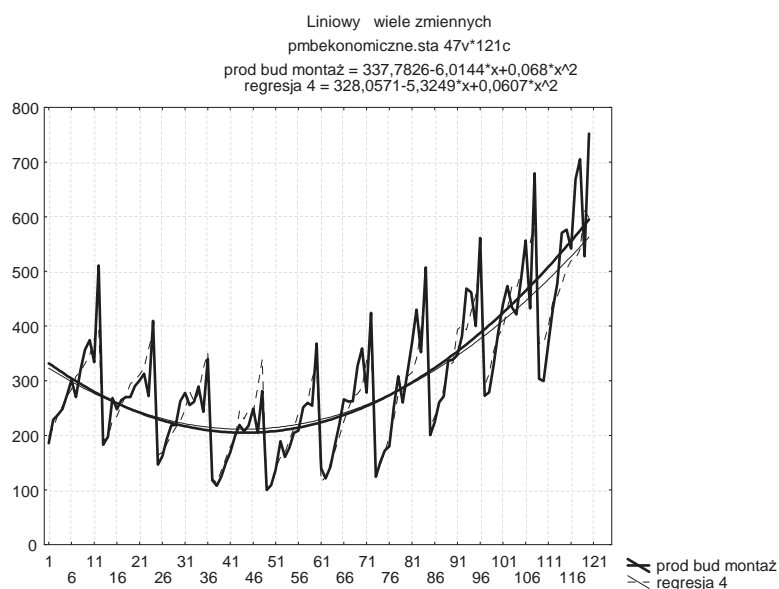
w obliczeniach metodą regresji i właśnie on będzie porównywany do wyników otrzymanych innymi metodami.

Wizualizacja obliczeń modelu 4 wraz ze zmienną prognozowaną produkcją budowlano-montażową przedstawiona jest na rysunku 4.

Tabela 2. Zestawienie modeli regresji krokowej wraz z wartościami parametrów skorygowanych R^2

Lp.	Nazwa i opis modelu	Skorygowane R^2
1	Regresja 1 $f(x_i)=f(lp^2,t,zp,wp)$	0,8864
2	Regresja 2 $f(x_i)=f(t,zp,wp)$	0,8796
3	Regresja 3 $f(x_i)=f(pt,zp,wp)$	0,7886
4	Regresja 4 $f(x_i)=f(t,pt, zp,wp)$	0,8921
5	Regresja 5 $f(x_i)=f(t^2,pt,zp,wp)$	0,8888
6	Regresja 6 $f(x_i)=f(lp^2, t^3,pt)$	0,6829

Źródło: Opracowanie własne



Rys. 4. Wykres zależności produkcji budowlano-montażowej w województwie dolnośląskim w badanych okresach od stycznia 2000 do grudnia 2009 oraz prognozy z zastosowaniem modelu 4. Na osi X oznaczono okresy odpowiadające kolejnym miesiącom

Źródło: Opracowanie własne

3. PROGNOZOWANIE METODĄ SZTUCZNYCH SIĘCI NEURONOWYCH

Istnieje wiele typów i rodzajów sieci neuronowych, które różnią się między sobą strukturą i zasadami działania [6, 7, 9]. Obecnie najpopularniejsza architektura sieci związana jest z koncepcją wielowarstwowego perceptronu MLP (*Multilayer Perceptron*). MLP składa się z wejścia, służącego do wprowadzania wartości zmiennych, warstw ukrytych neuronów, warstwy wyjściowej neuronów oraz wyjścia, które oznaczają wynik obliczeń np. prognozowaną wartość.

Do neuronu dociera pewna liczba sygnałów – wartości wejściowych. Są to wartości danych pierwotnych podawanych do sieci z zewnątrz jako dane do obliczeń albo sygnały pośrednie (pochodzące z wyjść innych neuronów wchodzących w skład sieci). Zwykle wprowadza się dodatkową wartość wejściową $x_0=1$ zwaną BIASEM lub progamiem. Wprowadzenie BIASU umożliwia takie usytuowanie funkcji agregacji, że nie musi ona przechodzić przez początek układu współrzędnych. Każda wartość wejściowa wprowadzana jest do neuronu przez połączenie o pewnej sile (wadze). Każdy neuron posiada również pojedynczą wartość progową, określającą jak silne musi być jego pobudzenie, by doszło do zapłonu. W neuronie obliczana jest ważona suma wejść (tzn. suma wartości sygnałów wejściowych pomnożonych przez odpowiednie współczynniki wagowe), a następnie odejmowana jest od niej wartość progowa. Uzyskana w ten sposób wartość pomocnicza określa pobudzenie neuronu. Sygnał reprezentujący łączne pobudzenie neuronu przekształcany jest z kolei przez ustaloną funkcję aktywacji neuronu (która określana jest również niekiedy jako funkcja przejścia neuronu). Wartość obliczona przez funkcję aktywacji jest ostatecznie wartością wyjściową (sygnałem wyjściowym) neuronu. Obliczenia wykonywano przy użyciu 20 rodzajów sieci MLP i RBF. Najlepsze wyniki uzyskano dla 5 sieci MLP. Wyniki zestawiono w tabeli 3.

Oznaczenia sieci np. MLP 21:10-4-1:1 oznaczają, że sieć jest w postaci wielowarstwowego perceptronu, ma 21 wejść (zmiennie niezależne) i jeden BIAS, 10 neuronów w pierwszej warstwie ukrytej, 4 neurony w drugiej warstwie ukrytej, 1 neuron w warstwie wyjściowej i jedno wyjście (zmienną zależną). Najlepszą jakość uczenia uzyskano dla sieci nr 5. Jednak o dobroci dopasowania decydują błędy ME, MAE, MPE, MAPE oraz współczynniki Theila.

Tabela 3. Zestawienie 5 najlepszych wyników obliczeń przy wykorzystaniu automatycznych sieci neuronowych pakietu STATISTICA z danymi pogodowymi i ekonomicznymi

Podsumowanie aktywnych sieci (produkcja budowlano montażowa z danymi ekonomicznymi)						
Id sieci	Nazwa sieci	Jakość (uczenie)	Jakość (testowanie)	Jakość (walidacja)	Błąd (uczenie)	Błąd (testowanie)
1	MLP 21:10-4-1:1	0,955859	0,953617	0,906951	492,3850	1265,429
2	MLP 21:10-9-1:1	0,942677	0,971425	0,921203	640,7550	848,918
3	MLP 21:10-9-1:1	0,947809	0,940087	0,904281	584,7846	1353,081
4	MLP 21:10-8-1:1	0,952782	0,963637	0,911730	525,9695	1036,614
5	MLP 21:10-4-1:1	0,960861	0,965165	0,905141	437,6101	901,799
Podsumowanie aktywnych sieci (produkcja budowlano montażowa z danymi ekonomicznymi)						
Błąd (walidacja)	Algorytm uczenia	Funkcja błędu	Aktywacja (ukryte)	Aktywacja (wyjściowe)		
807,8278	BFGS 20	SOS	Logistyczna	Liniowa		
722,8808	BFGS 9	SOS	Wykładnicza	Tanh		
843,0924	BFGS 7	SOS	Liniowa	Logistyczna		
763,9928	BFGS 16	SOS	Wykładnicza	Wykładnicza		
818,8423	BFGS 25	SOS	Wykładnicza	Logistyczna		

Źródło: Opracowanie własne

W celach porównawczych wykonano analogiczne obliczenia, nie uwzględniając danych ekonomicznych. Wyniki obliczeń zestawiono w tabeli 4.

Tabela 4. Zestawienie 5 najlepszych wyników obliczeń przy wykorzystaniu automatycznych sieci neuronowych pakietu STATISTICA z danymi pogodowymi **bez** danych ekonomicznych

Podsumowanie aktywnych sieci (produkcja budowlano montażowa bez danych ekonomicznych)						
Id sieci	Nazwa sieci	Jakość (uczenie)	Jakość (testowanie)	Jakość (walidacja)	Błąd (uczenie)	Błąd (testowanie)
1	MLP 19:10-4-1:1	0,982920	0,979903	0,977455	295,7542	544,992
2	MLP 19:10-4-1:1	0,982490	0,974892	0,959260	300,8798	663,827
3	MLP 19:10-4-1:1	0,958876	0,978240	0,941745	696,8860	1005,520
4	MLP 19:10-7-1:1	0,981676	0,975608	0,964387	314,1591	632,168
5	MLP 19:10-12-1:1	0,949681	0,971860	0,945222	854,0518	922,120
Podsumowanie aktywnych sieci (produkcja budowlano montażowa bez danych ekonomicznych)						
Błąd (walidacja)	Algorytm uczenia	Funkcja błędu	Aktywacja (ukryte)	Aktywacja (wyjściowe)		
281,7672	BFGS 102	SOS	Tanh	Tanh		
470,5932	BFGS 61	SOS	Wykładnicza	Tanh		
686,3286	BFGS 16	SOS	Tanh	Liniowa		
432,8166	BFGS 89	SOS	Wykładnicza	Wykładnicza		
682,1566	BFGS 16	SOS	Wykładnicza	Wykładnicza		

Źródło: Opracowanie własne

4. ANALIZA WYNIKÓW

Celem analitycznej oceny poprawności predykcji 3 metodami regresji, sieci neuronowych i ARIMA obliczono błędy : ME, MAE, MPE i MAPE dane wzorami (1), (2), (3) i (4). Otrzymane wyniki zestawiono w Tabeli 5 i na rysunku 5.

$$ME = \frac{1}{T-n} \sum_{i=T-n}^T (Y_i - Y_{ip}) \quad (1)$$

$$MAE = \frac{1}{T-n} \sum_{i=T-n}^T |Y_i - Y_{ip}| \quad (2)$$

$$MPE = \frac{1}{T-n} \sum_{i=T-n}^T \frac{Y_i - Y_{ip}}{Y_i} \quad (3)$$

$$MAPE = \frac{1}{T-n} \sum_{i=T-n}^T \frac{|Y_i - Y_{ip}|}{Y_i} \quad (4)$$

gdzie:

ME – średni błąd (mean error)

MAE – średni osiągnięty błąd (mean average error)

MPE – średni błąd procentowy (mean percentage error)

$MAPE$ – średni absolutny procentowy błąd (mean absolute percentage error)

T – suma ilości okresów obliczeniowych i prognozowanych

n – ilość okresów prognozowanych

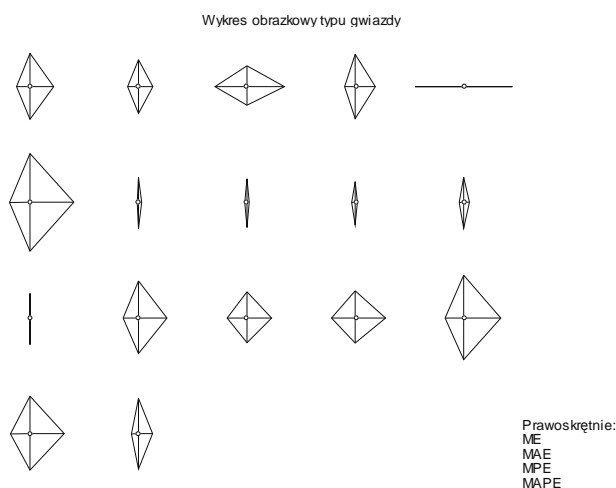
Y_i – wartość rzeczywista zmiennej w okresie i

Y_{ip} – wartość prognozowana zmiennej w okresie i .

Tabela 5. Wartości błędów ME, MAE, MPE i MAPE prognoz metodami regresji, sieci neuronowych i ARIMA

Lp.	Nazwa i opis	ME	MAE	MPE	MAPE
1	Regresja 1 $f(x_i)=f(lp^2,t,zp,wp)$	38,64	82,41	3,26	15,51
2	Regresja 2 $f(x_i)=f(t,zp,wp)$	11,58	71,83	- 2,05	14,68
3	Regresja 3 $f(x_i)=f(pt,zp,wp)$	- 12,87	98,22	- 9,24	21,55
4	Regresja 4 $f(x_i)=f(t,pt, zp,wp)$	34,58	78,28	2,64	14,64
5	Regresja 5 $f(x_i)=f(t^2,pt,zp,wp)$	- 99,44	110,68	- 25,58	27,15
6	Regresja 6 $f(x_i)=f(lp^2, t^3,pt)$	102,72	105,18	17,24	17,85
7	MLP 21:10-4-1:1	3,32	59,51	- 2,34	11,56
8	MLP 21:10-9-1:1	- 1,94	58,23	- 3,40	11,79
9	MLP 21:10-9-1:1	- 13,85	58,10	- 5,75	12,29
10	MLP 21:10-8-1:1	4,84	62,05	- 2,59	12,58
11	MLP 21:10-4-1:1	2,57	55,72	- 2,30	11,12
12	MLP 19:10-4-1:1	53,87	87,95	5,55	16,18
13	MLP 19:10-4-1:1	9,67	83,65	- 3,98	17,59
14	MLP 19:10-4-1:1	12,85	90,06	- 3,81	18,93
15	MLP 19:10-7-1:1	78,22	97,51	10,93	17,22
16	MLP 19:10-12-1:1	57,13	94,34	6,10	17,59
17	ARIMA	46,72	71,60	5,51	13,44

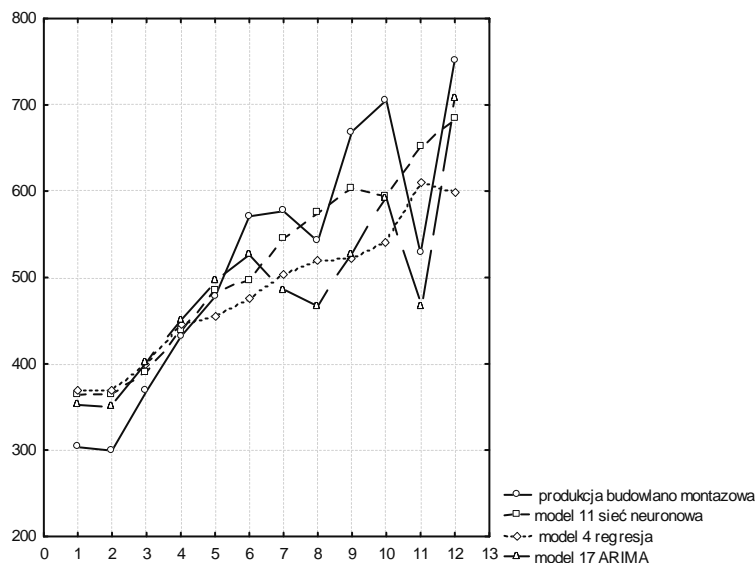
Źródło: Opracowanie własne



Rys. 5. Wartości błędów ME, MAE, MPE i MAPE prognoz metodami regresji, sieci neuronowych i ARIMA

Źródło: Opracowanie własne

Najmniejszą wartość błędów uzyskano dla modelu 11 MLP 21:10-4-1:1. Celem porównania najlepszych wyników z poszczególnych metod (regresji, sieci neuronowych i ARIMA), przedstawiono na rysunku 6 otrzymane prognozy z modeli 4, 11 i 17 wraz z wartościami rzeczywistymi produkcji budowlano montażowej w okresie prognozowanym.



Rys. 6. Prognozy produkcji budowlano montażowej metodami regresji, sieci neuronowych i ARIMA w zestawieniu z danymi rzeczywistymi z roku 2009

Źródło: Opracowanie własne

W procesie oceny prognozy pomocne są również współczynniki rozbieżności Theila, które dostarczają informacji o źródłach błędów prognoz. Podstawowy współczynnik I^2 zdefiniowany jest wzorem (5).

$$I^2 = \frac{\sum_{i=T-n}^T (Y_{ip} - Y_i)^2}{\sum_{i=T-n}^T Y_i^2} \quad (5)$$

Pierwiastek kwadratowy z parametru I^2 informuje, jaki jest przeciętny względny błąd dopasowania V . Interpretację współczynnika Theila ułatwia jego dekompozycja na trzy składowe I_1^2 , I_2^2 i I_3^2 zgodnie z równaniem (6). Składowe opisane równaniami (7-9) pozwalają sprecyzować rząd i charakter błędów aproksymacji. Składnik I_1^2 jest miarą błędu systematycznego, który wynika z obciążenia metody estymacji modelu, inaczej z obciążenia predykcji (niedostateczna zgodność przeciętnych wartości Y_i i Y_p). Parametr I_2^2 mierzy stopień odwzorowania przez model wariancji zmiennej objaśnianej i określa rozmiary błędu związanego z niedostateczną elastycznością predykcji. Innymi słowy, odzwierciedla błędy spowodowane faktem, że zróżnicowanie prognoz nie jest zgodne z faktycznymi wahaniami zmiennej prognozowanej. Składnik I_3^2 charakteryzuje błąd wynikający z braku pełnej zgodności kierunku zmian wartości teoretycznych ze zmianami wartości empirycznych, a więc z nie dość dobrej predykcji punktów zwrotnych. Ogólnie, jeżeli odbiorca prognozy nie poda własnych kryteriów dopuszczalności prognoz, przyjmujemy, że jeżeli względny miernik dokładności predykcji V przyjmuje wartości:

- $V \leq 3\%$, to prognozy są wysoce dokładne;
- $3\% < V \leq 5\%$, to prognozy uznajemy za dobre;

- $5\% < V \leq 10\%$, to prognozy są mało dokładne, ale mogą być przyjęte (dopuszczalne);
- $V > 10\%$, to prognozy są niedokładne, nie powinny być przyjęte (niedopuszczalne).

$$I^2 = I_1^2 + I_2^2 + I_3^2 \quad (6)$$

$$I_1^2 = \frac{(\bar{Y}_{ip} - \bar{Y}_i)^2}{\frac{1}{T-n} \sum_{i=T-n}^T Y_i^2} \quad (7)$$

$$I_2^2 = \frac{(s_{ip} - s_i)^2}{\frac{1}{T-n} \sum_{i=T-n}^T Y_i^2} \quad (8)$$

$$I_3^2 = \frac{2s_{ip}s_i(1 - r_{ipi})}{\frac{1}{T-n} \sum_{i=T-n}^T Y_i^2} \quad (9)$$

gdzie:

I^2, I_1^2, I_2^2, I_3^2 – współczynniki Theila

T – suma ilości okresów obliczeniowych i prognozowanych

n – ilość okresów prognozowanych

Y_i – wartość rzeczywista zmiennej w okresie i

Y_{ip} – wartość prognozowana zmiennej w okresie i

s_i – odchylenie standardowe zmiennej prognozowanej

s_{ip} – odchylenie standardowe prognoz

r_{ipi} – współczynnik korelacji liniowej pomiędzy wielkościami Y_i i Y_{ip}

\bar{Y}_i – średnia arytmetyczna realizacji zmiennej prognozowanej

\bar{Y}_{ip} – średnia arytmetyczna wartości prognoz.

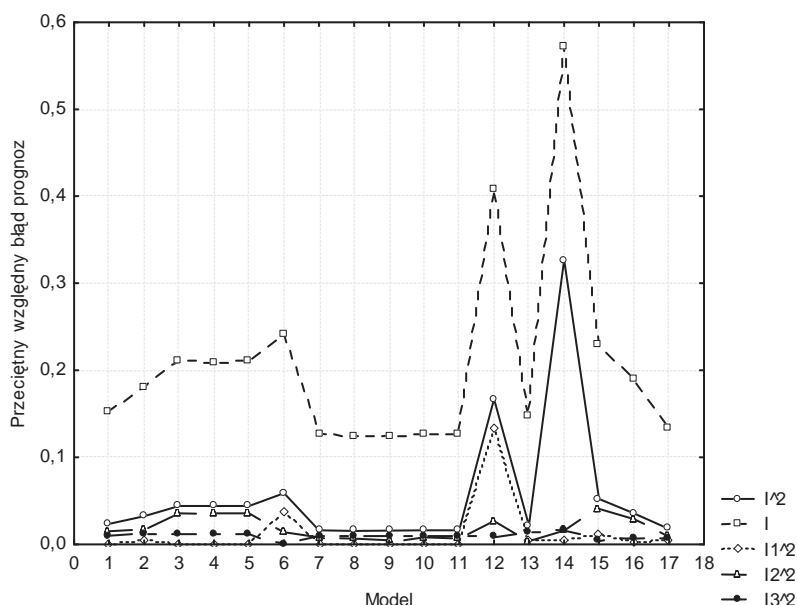
Obliczone zgodnie z wzorami (6-9) współczynniki $I, I^2, I_1^2, I_2^2, I_3^2$ zamieszczono w tabeli 6 i zilustrowano na rysunku 7.

Tabela 6. Wartości współczynników Theila prognoz metodami regresji, sieci neuronowych i ARIMA w modelach 1-17

Lp.	Nazwa i opis	I^2	I	I_1^2	I_2^2	I_3^2
1	Regresja 1 $f(x_i)=f(lp^2,t,zp,wp)$	0,0235	0,153297	0,0005	0,0152	0,0100
2	Regresja 2 $f(x_i)=f(t,zp,wp)$	0,0322	0,179444	0,0052	0,0171	0,0123
3	Regresja 3 $f(x_i)=f(pt,zp,wp)$	0,0441	0,210000	0,0006	0,0356	0,0119
4	Regresja 4 $f(x_i)=f(t,pt, zp,wp)$	0,0440	0,209762	0,0006	0,0356	0,0119
5	Regresja 5 $f(x_i)=f(t^2,pt,zp,wp)$	0,0441	0,210000	0,0006	0,0356	0,0119
6	Regresja 6 $f(x_i)=f(lp^2, t^3,pt)$	0,0588	0,242487	0,0366	0,0139	0,00105
7	MLP 21:10-4-1:1	0,0162	0,127279	0,0000	0,0079	0,0097
8	MLP 21:10-9-1:1	0,0155	0,124499	0,0000	0,0065	0,0104
9	MLP 21:10-9-1:1	0,0156	0,124900	0,0006	0,0058	0,0104
10	MLP 21:10-8-1:1	0,0162	0,127279	0,0000	0,0079	0,0097
11	MLP 21:10-4-1:1	0,0162	0,127279	0,0000	0,0079	0,0097

Lp.	Nazwa i opis	I^2	I	I_1^2	I_2^2	I_3^2
12	MLP 19:10-4-1:1	0,1670	0,408656	0,1347	0,0268	0,0083
13	MLP 19:10-4-1:1	0,0215	0,146629	0,0057	0,0030	0,0142
14	MLP 19:10-4-1:1	0,3269	0,571752	0,0039	0,0156	0,0158
15	MLP 19:10-7-1:1	0,0523	0,228692	0,0114	0,0407	0,0038
16	MLP 19:10-12-1:1	0,0357	0,188944	0,0030	0,0290	0,0067
17	ARIMA	0,0177	0,133041	0,0038	0,0085	0,0067

Źródło: Opracowanie własne



Rys. 7. Wartości współczynnika Theila wraz ze składowymi, na osi x oznaczenie numeru modelu, na osi y względny miernik dokładności predykcji I

Źródło: Opracowanie własne

WNIOSKI

Jak wynika z przeprowadzonych obliczeń, nie uzyskano prognozy dokładnej. W prognozie najlepszej (model 8) uzyskano przeciętny względny błąd prognozy wynoszący 12,4499%, zatem jeśli nie założono dokładności predykcji, inaczej należałoby uznać, że prognoza jest zbyt mało dokładna, by można było uznać ją za prawidłową. Są jednak działy nauki, gdzie za prognozę dopuszczalną uznaje się modele z przeciętnym względnym błędem prognozy nawet do 30%. Model 8 to automatyczna sieć neuronowa MLP 21:10-9-1:1. Sukcesem przeprowadzonych obliczeń jest to, że uzyskano wynik lepszy niż obliczony metodą ARIMA (autoregresyjny zintegrowany proces średniej ruchomej) wynoszący 13,3041%. Po wprowadzeniu do obliczeń danych ekonomicznych uzyskano lepszą prognozę w metodzie regresji wielorakiej (model 1 z danymi ekonomicznymi posiada błąd prognozy 15,3297% , natomiast tylko z danymi pogodowymi 24,2487%) oraz w metodzie automatycznych sieci neuronowych, gdzie uzyskano polepszenie wyników przeciętnego względnego błędu prognozy z 18,8944% do 12,4499%. Wydaje się zatem trafne poszerzenie bazy danych wyjściowych o dane ekonomiczne dotyczące ilości zatrudnionych pracowników oraz ich zarobków. Należy poszukiwać kolejnych danych mogących być predyktorami produkcji budowlano montażowej. Celem dalszych prac jest uzyskanie przeciętnego względnego błędu prognozy o wartości mniejszej niż 10%.

LITERATURA

- [1] Cieślak M., *Prognozowanie gospodarcze : metody i zastosowanie*, Wydawnictwo Naukowe PWN, Warszawa 2001.
- [2] Kot S., Jakubowski J., Sokołowski A., *Statystyka*. Difin, Warszawa 2007.
- [3] Martyniak Z., *Wstęp do inwentyki*. Wydawnictwo Uczelniane, Kraków 1997, s. 32.
- [4] *Podręcznik internetowy STATISTICA*, [online] [dostęp: 2010]. Dostępny w Internecie: <http://www.statsoft.pl/textbook/sttimser.htm>
- [5] Rogalska M., Hejducki Z., *Analiza porównawcza prognozowania produkcji budowlanej z zastosowaniem metod regresji krokowej, sieci neuronowych i ARIMA*, [w:] „Zeszyty Naukowe WSOWL”, nr 3/2010.
- [6] Skorupka D., *Risk Management in Building Projects*, [in:] AACE International Transaction, (CSC.1.91– CSC.1.96), The Association for the Advancement of Cost Engineering, USA, Orlando 2003.
- [7] Skorupka D., *Neural Networks in Risk Management of a Project*, 2004 AACE International Transaction, (CSC.1.51– CSC.1.57), The Association for the Advancement of Cost Engineering, USA, Washington 2004.
- [8] Stanisław A., *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny*, T 1. StatSoft Polska Sp. z o.o., Kraków 2006.
- [9] Witkowska D., *Sztuczne sieci neuronowe i metody statystyczne*, C.H.BECK, Warszawa 2002

COMPARATIVE ANALYSIS OF CONSTRUCTION PRODUCTION FORECASTING WITH STEPWISE REGRESSION, NEURAL NETWORKS AND ARIMA METHODS

Summary

The article analyses the possibilities of forecasting Lower Silesia's construction production with stepwise regression, neural networks and ARIMA (Autoregressive Integrated Moving Average) methods. For the forecasting with the stepwise regression method the authors use the daily weather data of Lower Silesia and economic data: the number of employees in the construction sector and the average earnings of workers in this sector.

The analysis of results by calculating errors ME, MAE, MPE, MAPE and Theil coefficients I , I_1^2 , I_2^2 , I_3^2 was performed. The authors of the article also propose the scope for further research.

Key words: *forecasting, construction and assembly production, stepwise regression, neural networks, ARIMA*

Artykuł recenzował: płk dr hab. inż. Dariusz SKORUPKA, prof. nadzw. WSOWL