9. Prognozowanie popytu, wizualizacja i inżynieria cech szeregów czasowych w łańcuchach dostaw

Czym jest prognozowanie popytu? Jak skutecznie wizualizować dane o klientach i wyciągać z nich wnioski? Jak przeprowadzić inżynierię cech szeregów czasowych?

Na te i podobne pytania postaramy się udzielić odpowiedzi w poniższym rozdziale.

9.1 Czym jest popyt i prognozowanie popytu?

Ostateczny popyt klienta wprawia w ruch cały łańcuch dostaw (Syntetos i in., 2016). W związku z tym popyt klienta jest kluczowym elementem planowania wszystkich procesów logistycznych w łańcuchu dostaw, a zatem określenie poziomów popytu klienta jest bardzo interesujące dla menedżerów łańcucha dostaw. Prognozowanie popytu jest niezbędnym działaniem do planowania i harmonogramowania działań logistycznych w ramach obserwowanego łańcucha dostaw (Mircetic i in., 2017). Dokładne modele prognozowania popytu bezpośrednio wpływają na obniżenie kosztów logistycznych, ponieważ zapewniają ocenę popytu klientów z wyprzedzeniem czasowym potrzebnym do planowania i działania(Mircetic i in., 2016). Prognozowanie w łańcuchach dostaw wykracza poza operacyjne zadanie ekstrapolacji wymagań popytu na jednym poziomie. Obejmuje ono złożone kwestie, takie jak koordynacja łańcucha dostaw i dzielenie się informacjami między wieloma zainteresowanymi stronami (Syntetos i in., 2016).

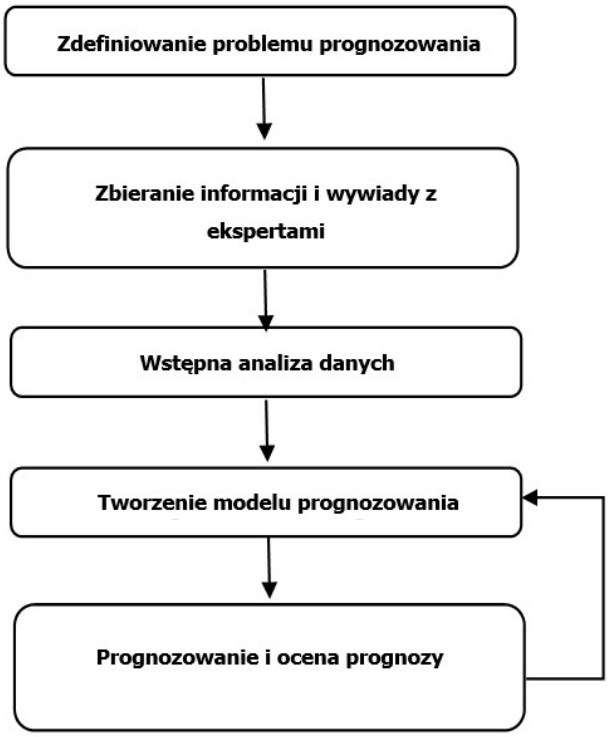
Popyt klientów i towarzyszące mu prognozy mają kluczowe znaczenie dla łańcuchów dostaw (SC), ponieważ zapewniają podstawowe dane wejściowe do planowania i kontroli wszystkich obszarów funkcjonalnych, w tym logistyki, marketingu, produkcji itp. Gdyby ostateczny popyt konsumentów był stały lub znany z dużym wyprzedzeniem, wówczas działanie łańcucha dostaw byłoby prostym (wstecznym) planowaniem. Jednak popyt nie jest znany i dlatego musi być prognozowany. To właśnie niepewność związana z tym popytem sprawia, że zarządzanie łańcuchem dostaw jest bardzo trudne (Syntetos i in., 2016). Na skuteczność

prognozowania popytu wpływa nieodłączna niepewność w szeregach czasowych popytu, które występują w łańcuchach dostaw (Rostami-Tabar, 2013). W związku z tym zajęcie się i zrozumienie tych niepewności jest głównym wyzwaniem dla menedżerów podczas koordynowania i planowania operacji w ramach łańcuchów dostaw (Mircetic, 2018).

Niepewność popytu jest jednym z najważniejszych wyzwań dla współczesnych łańcuchów dostaw. Niedawna pandemia COVID-19 jeszcze bardziej podkreśliła tę kwestię, powodując powszechne zakłócenia, które skomplikowały planowanie i kontrolę łańcucha dostaw (Nikolopoulos i in., 2020;). Prognozowanie popytu w łańcuchach dostaw często wiąże się z przewidywaniem popytu na wiele produktów. Osoby zajmujące się prognozowaniem w łańcuchach dostaw zazwyczaj ekstrapolują dane szeregów czasowych dla każdej jednostki magazynowej indywidualnie. Na przykład sprzedawca detaliczny może wykorzystywać dane z punktów sprzedaży do generowania prognoz na poziomie poszczególnych sklepów (Mircetic i in., 2022).

9.2 Etapy prognozowania popytu w łańcuchach dostaw

Zgodnie z powyższymi stwierdzeniami i wnioskami dotyczącymi znaczenia popytu i prognozowania popytu dla łańcuchów dostaw, ważne jest, aby postępować zgodnie z określonymi krokami podczas opracowywania prognoz w łańcuchach dostaw (rysunek 9.1).



Rysunek 9.1 Podstawowe kroki dla prawidłowego wdrożenia prognoz w firmie

Źródło: Makridakis i in., 1998; Makridakis i in., 1983

Każdy z etapów przedstawionych na rysunku 9.1 ma swoje zalety i przyczynia się do tworzenia wiarygodnych i użytecznych (zorientowanych biznesowo) prognoz. W związku z tym zdefiniowanie problemu jest często najtrudniejszą częścią prognozowania i wymaga zrozumienia, w jaki sposób prognozy będą wykorzystywane, a także roli funkcji prognozowania w obserwowanej firmie. Prognozujący powinien poświęcić dużo czasu na komunikację z wszystkimi osobami zaangażowanymi w gromadzenie danych, utrzymanie bazy danych i wykorzystywanie prognoz do planowania przyszłości. Jednym z głównych czynników obciążających definicję problemu jest to, w jaki sposób ostateczna prognoza będzie wykorzystywana w codziennych operacjach logistycznych (jaka platforma, projekt oprogramowania, interfejs użytkownika itp.).

Na etapie gromadzenia informacji zawsze potrzebne są co najmniej dwa rodzaje informacji: dane statystyczne i zgromadzona wiedza osób, które gromadzą dane i wykorzystują prognozy. W praktyce często trudno jest uzyskać dane historyczne, aby stworzyć dobry model statystyczny. Ponadto istnieje duże nieporozumienie co do tego, czym są dane dotyczące popytu i co można wykorzystać jako ich zamiennik. Istnieje zła praktyka wykorzystywania danych o wysyłkach i dostawach jako zamiennika danych o popycie,

co tylko pogorszy proces podejmowania decyzji w oparciu o prognozy sporządzone na podstawie łatwego rodzaju danych. Dane o sprzedaży są jedynym wiarygodnym zamiennikiem danych o popycie (Syntetos i in., 2016), chociaż to uproszczenie nie jest idealne, szczególnie w łańcuchach dostaw, w których występuje wiele sytuacji braku zapasów.

Na etapie wstępnej analizy danych zaleca się, aby zawsze rozpoczynać analizę danych od reprezentacji graficznych, aby odpowiedzieć na następujące pytania. Czy istnieją spójne wzorce? Czy istnieje znaczący trend? Czy istnieje zauważalna sezonowość? Czy istnieją dowody na cykle koniunkturalne? Jak silne są zależności między zmiennymi? Są to pytania, na które prosta grafika może dostarczyć odpowiedzi i umożliwić dalszą analizę danych poprzez zawężenie zakresu modeli, które należy zastosować do odkrytych cech popytu. Zwykle prosty model, określony w ten sposób, może pokonać bardziej wyrafinowane i skomplikowane (Rostami-Tabar & Mircetic, 2023).

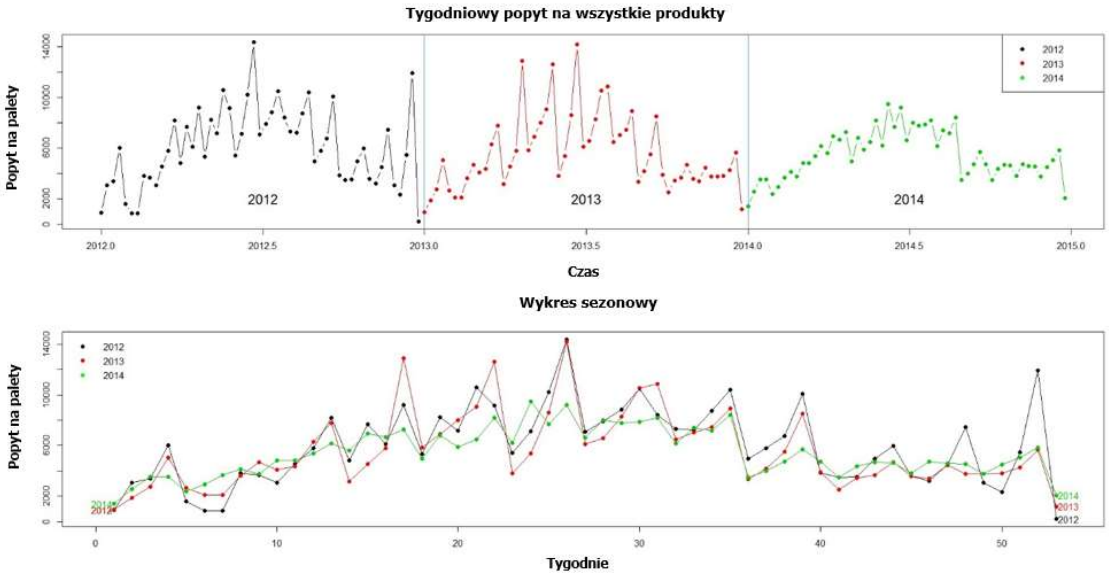
Wybór i tworzenie modeli prognostycznych jest najważniejszym krokiem podczas budowy modelu prognostycznego. Wybór modelu zależy od kilku czynników, z których najważniejsze to dostępność danych historycznych oraz korelacja między zmiennymi zależnymi i niezależnymi. Podczas wyboru modelu często porównuje się dwa lub trzy potencjalne modele. Każdy model jest sztuczną konstrukcją opartą na zestawie założeń (jawnych i ukrytych) i generalnie obejmuje jeden lub więcej parametrów, które muszą zostać utworzone przy użyciu znanych danych historycznych. W poniższym rozdziale przedstawiony zostanie proces rozwoju i zastosowania modelu ARIMA, jako jednego z najlepszych i najpopularniejszych modeli.

Ocena modelu prognostycznego jest krokiem, który mierzy użyteczność stworzonego modelu. Po wybraniu modelu prognostycznego i oszacowaniu jego parametrów, model jest wykorzystywany do tworzenia prognoz. Dokładność modelu jest oceniana za pomocą różnych statystyk podobnie jak trafność prognoz, ale ważne jest również testowanie prognoz za pomocą miar implikacji biznesowych (tj. wskaźników użyteczności).

9.3 Prognozowanie popytu w przemyśle spożywczym

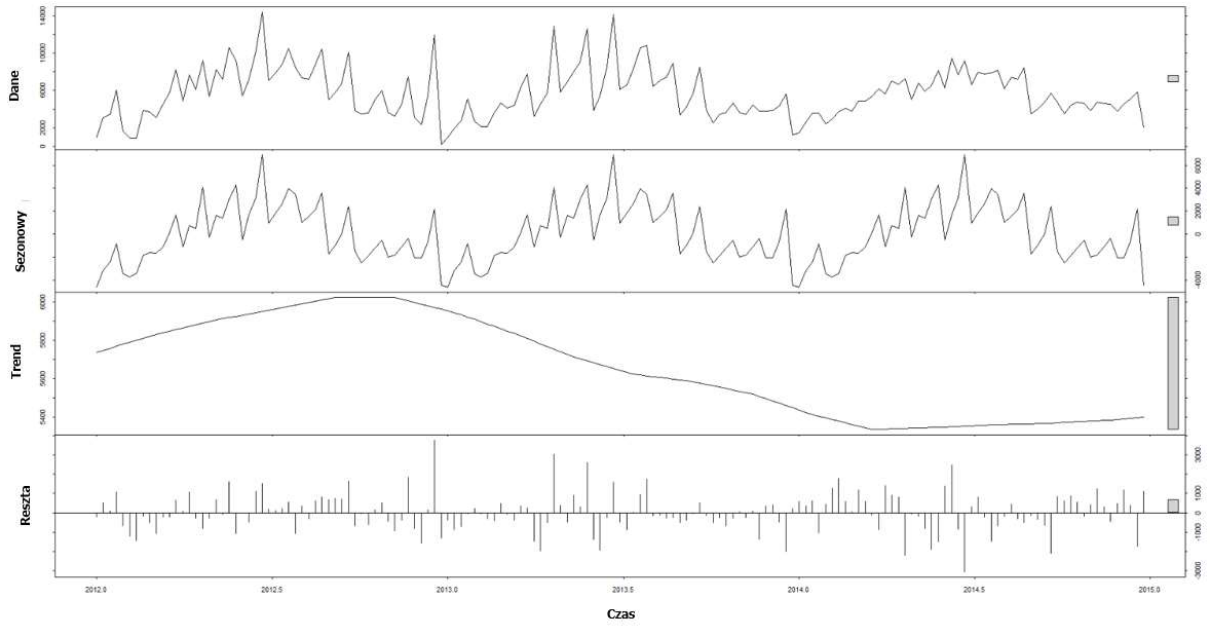
Dane dotyczące konsumpcji wszystkich produktów obserwowanego przedsiębiorstwa z branży spożywczej przedstawiono jako tygodniowy popyt obejmujący okres od stycznia 2012 r. do grudnia 2014 r. (Wykres na rysunku 9.2). Oś x reprezentuje czas, podczas gdy

wartości popytu są przedstawione na osi y. Dane te są przedstawione w odstępach tygodniowych, ponieważ odpowiada to okresowi, w którym prowadzona jest dostawa do końcowych punktów sprzedaży. W związku z tym kierownictwo firmy koncentruje się na prognozowaniu tygodniowej konsumpcji na rynku.



Rysunek 9.2 Dane dotyczące popytu dla obserwowanej firmy spożywczej

Rysunek 9.2 pokazuje kilka ważnych cech i właściwości danych. Po pierwsze, dane mają silny charakter sezonowy, a szczyt sprzedaży przypada na połowę roku (miesiące letnie). Po drugie, wykres sezonowy (dolny wykres na rysunku 9.2) pokazuje znaczącą zmianę wzorca trendu spadkowego w 2014 roku! Jest to bardzo istotna charakterystyka dla wyboru właściwego modelu prognozowania i dla kierowników wyższego szczebla w firmie, ponieważ ujawnia znaczny spadek konsumpcji i utratę rynku. Aby dokładniej zbadać zaobserwowane cechy, zastosowano metodologię dekompozycji trendu i sezonowości (STL) (rysunek 9.3). Dekompozycja STL dzieli pierwotne wzorce popytu na trzy komponenty: sezonowy, trend i komponent pozostały



Rysunek 9.3 Dekompozycja STL danych dotyczących popytu

STL ujawnił, że obserwowane szeregi czasowe mają charakter addytywny co oznacza, że wahania wokół krzywej trendu-cyklu nie zwiększają się znacząco w czasie. W rezultacie transformacje Box-Cox nie były konieczne dla surowych szeregów czasowych. Dekompozycja ujawniła, że składnik sezonowy jest dominujący w obserwowanej serii, wykazując duże wahania w ciągu jednego roku. Biorąc pod uwagę ograniczoną liczbę lat obserwacji, trudno jest zidentyfikować cykl koniunkturalny. Dekompozycja wykazała również, że trend w szeregu jest minimalny, z malejącym wzorcem rozpoczynającym się od połowy 2013 roku.

Te zidentyfikowane cechy stanowiły ważne dane wejściowe podczas procesu projektowania odpowiedniego modelu prognostycznego. W tym celu wybrano model S-ARIMA (ang. Seasonal Autoregressive Integrated Moving Average). Model S-ARIMA jest bardzo skuteczny w prognozowaniu, ponieważ łączy w sobie zarówno komponenty autoregresyjne, jak i średnie ruchome, wraz z różnicowaniem w celu uczynienia danych stacjonarnymi. Model ten jest szczególnie skuteczny w wychwytywaniu i modelowaniu wzorców sezonowych w danych szeregów czasowych, co czyni go idealnym dla branż o cyklicznych wzorcach popytu, takich jak przemysł spożywczy. Dodatkowo, zdolność S-ARIMA do radzenia sobie ze złożonymi strukturami sezonowymi i trendami pozwala na dokładniejsze i bardziej wiarygodne prognozy, które są kluczowe dla skutecznego planowania łańcucha dostaw

i zarządzania zapasami. Strukturalna postać modelu S-ARIMA została przedstawiona w ównaniu (9.1).

$$\Phi(B^m)\phi(B)(1-B^m)^D(1-B)^d y_t = c + \Theta(B^m)\theta(B)\varepsilon_t, \quad (9.1)$$

gdzie $\Phi(z)$ i $\Theta(z)$ są wielomianami odpowiednio rzędu *P* i *Q*, z których każdy nie zawiera pierwiastków wewnątrz okręgu jednostkowego. *B* jest operatorem przesunięcia wstecznego używanym do opisu procesu różniczkowania, tj. $By_t=y_{t-1}$. Jeśli $c\neq 0$, istnieje implikowany wielomian rzędu *d+D* w funkcji prognozy. Ponieważ model S-ARIMA jest modelem wysoce sparametryzowanym, kluczową kwestią podczas korzystania z modelu S-ARIMA jest wybór odpowiedniego rzędu modelu, co wiąże się z określeniem wartości of *p, q, P, Q, D*, i *d*. Jeżeli *d* oraz *D* są nieznane, rzędy *p, q, P*, i *Q* można wybrać za pomocą kryterium informacyjnego, takiego jak Kryterium Informacyjne Akaike (AIC) lub Bayesowskie Kryterium Informacyjne (BIC). Wzory na AIC i BIC są następujące:

$$AIC = -2\log(L) + 2(k);$$

$$BIC = N\log\left(\frac{SSE}{N}\right) + (k+2)\log(N) \quad (9.2)$$

gdzie $k=p+q+P+Q+1$, jeśli uwzględniono stałą, a 0 w przeciwnym razie, *L* jest zmaksymalizowanym prawdopodobieństwem modelu dopasowanego do zróżnicowanych danych, *SSE* jest sumą błędów podniesionych do kwadratu, *N* jest liczbą obserwacji wykorzystanych do estymacji, a *k* jest liczbą predyktorów w modelu

W celu określenia optymalnego zestawu parametrów Hyndman i Khandakar (2007) zaproponowali test pierwiastka jednostkowego Canova-Hansena i KPSS w następujących krokach:

- Użyj testu Canova-Hansena do określenia *D* w ramach ARIMA,
- Wybierz *d*, stosując kolejny test pierwiastka jednostkowego KPSS na danych zróżnicowanych sezonowo (jeśli $D = 1$) lub na danych oryginalnych (jeśli $D = 0$),
- Wybierz optymalne wartości dla *p, q, P* i *Q* poprzez minimalizację AIC.

9.4 Opracowanie modelu prognostycznego S-ARIMA

Opracowanie modelu prognostycznego S-ARIMA

Zgodnie z powyższą procedurą przetestowano kilka ustawień parametrów, a ich wydajność przedstawiono w tabeli 9.1. Do testowania wydajności różnych modeli zastosowano kilka miar: średni bezwzględny błąd procentowy (MAE), średni błąd kwadratowy (RMSE), średni bezwzględny błąd skalowany (MASE), AIC i BIC (równania 2 i 3).

$$MAPE = \frac{100}{L}\sum_{l=1}^{L}\left|\frac{e_l}{y_l}\right|;$$

;

$$RMSE = \sqrt{\frac{1}{L}\sum_{l=1}^{L} e_l^2}; \tag{9.3}$$

,

gdzie e_l są wartościami rezydualnymi, ale liczonymi dla L okresów następnych po N okresach stanowiących próbkę, z której danych otrzymano szacunek modelu: $e_l = y_l - y_l^*$, Tak liczone reszty nazywa się błędami prognoz y_l^*, czyli wartości teoretycznych z predykcji poza próbkę. Do miar trafności prognoz zaliczamy także średni absolutny błąd skalowany:

.

$$MASE = \frac{\frac{1}{L}\sum_{l=1}^{L}|e_l|}{z}; \tag{9.4}$$

Gdzie z stanowi współczynnik skalowania danych sezonowych o długości cyklu S

$$z = \frac{1}{N-m}\sum_{t=m+1}^{N}|y_t - y_{t-S}|. \tag{9.5}$$

Im niższa wartość MASE tym trafniej prognozuje model. Najpopularniejszymi miarami dla menedżerów w łańcuchach dostaw są RMSE i MAPE, ponieważ zapewniają one „poczucie" tego, jak dobrze model działa w liczbach rzeczywistych (RMSE) i procentach (MAPE).

Tabela 9.1 Wydajność modeli S-ARIMA z różnymi ustawieniami parametrów

Modele[a]	RMSE	MAPE	MASE	AIC	BIC
S-ARIMA $(5,0,1)(1,0,0)_{52}$[b]	1023	14,18 %	0,60	86,62	110,50
S-ARIMA $(4,0,0)(1,0,0)_{52}$[c]	1041	14,53 %	0,62	86,73	105,31
S-ARIMA $(4,0,0)(0,1,1)_{52}$[d]	1882	22,12 %	1,05	41,74	53,560
S-ARIMA $(4,0,1)(1,0,0)_{52}$[e]	1050	14,18 %	0,60	88,23	109,47
S-ARIMA $(0,0,1)(0,1,0)_{52}$[f]	1797	21,42 %	1,02	36,52	40,460

[a] Błędy modelu są obliczane na zestawie danych testowych.

[b] Szczegóły dotyczące modelu S-ARIMA $(5,0,1)(1,0,0)_{52}$ są podane poniżej.

[c] $(1-0{,}39B+0{,}06B+0{,}04B-0{,}46B)(1-0{,}65B^{52})y_t = 8{,}52.$

[d] $(1-0{,}29B+0{,}19B-0{,}05B-0{,}19B)(1-B^{52})y_t = (1+0{,}12B^{52})e_t.$

[e] $(1-0{,}29B+0{,}01B+0{,}05B-0{,}48B)(1-0{,}65B^{52})y_t = 8{,}52+(1+0{,}13B)e_t.$

[f] $(1-B^{52})y_t = (1+0{,}3B)e_t.$

Tabela 9.1 pokazuje, że model S-ARIMA $(5,0,1)(1,0,0)_{52}$, gdzie cykl sezonowy trwa $S = 52$ tygodnie, przewyższył konkurencyjne modele, osiągając najniższe błędy RMSE, MAPE i MASE. Postać modelu S-ARIMA $(5,0,1)(1,0,0)_{52}$ przedstawiono w równaniu (9.4), gdzie $\phi_1 = -0{,}5421$, $\phi_2 = 0{,}2962$, $\phi_3 = -0{,}099$, $\phi_4 = 0{,}3974$, $\phi_5 = 0{,}4994$, $\Phi_1 = 0{,}9558$, $c = 8{,}523$, i $\Theta_1 = 0{,}6345$.

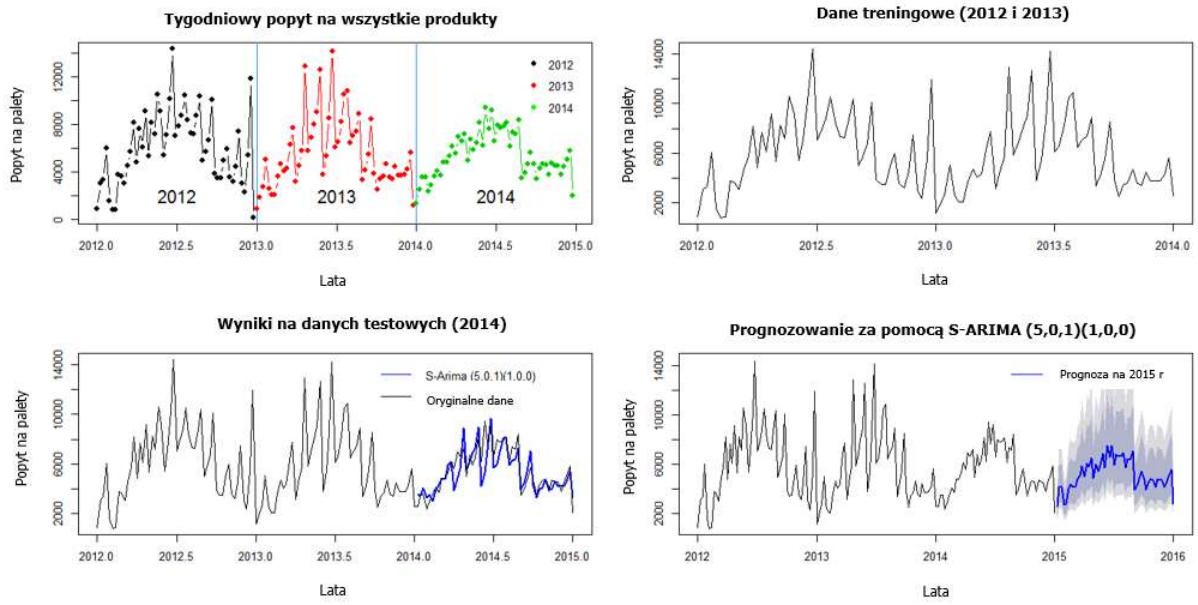
$$\left(1-\phi_1 B-\phi_2 B-\phi_3 B-\phi_4 B-\phi_5 B\right)\left(1-\Phi_1 B^{52}\right)y_t = c+\left(1+\theta_1 B\right)e_t\,, \tag{9.6}$$

Podobne wyniki zaobserwowano w przypadku modeli S-ARIMA $(4,0,0)(1,0,0)_{52}$ i S-ARIMA $(4,0,1)(1,0,0)_{52}$. Podczas wykonywania prognoz model S-ARIMA $(5,0,1)(1,0,0)_{52}$ generuje średnio błąd 1023 sztuk, co przekłada się na 14,18%. Wyniki te podkreślają znaczenie starannego doboru warunków do uwzględnienia w modelu S-ARIMA, ponieważ nie ma znaczących różnic między wynikami modelu z różnymi parametrami. Ocena wykazała, że modele zawierające składniki autoregresyjne (*p, P*) i średniej ruchomej (*q, Q*) były bardziej skuteczne w prognozowaniu spożycia napojów niż te zawierające sezonowe lub niesezonowe różnicowanie. Ponadto przegląd porównawczy ujawnił pewne nieoczekiwane ustalenia, takie jak modele uwzględniające sezonowe różnicowanie S-ARIMA $(4,0,0)(0,1,1)_{52}$

i S-ARIMA $(0,0,1)(0,1,0)_{52}$, które wypadły gorzej niż prosty średni naiwny model prognostyczny, o czym świadczą ich błędy MASE przekraczające jeden.

9.5 Prognozy dotyczące przyszłego popytu.

Na rysunku 9.4 przedstawiono wydajność metody S-ARIMA $(5,0,1)(1,0,0)_{52}$. Lewy górny panel przedstawia początkowe dane popytu, pokolorowane w celu łatwiejszego rozróżnienia różnych lat marketingowych. Prawy górny panel to dane wejściowe do modelu S-ARIMA $(5,0,1)(1,0,0)_{52}$, reprezentujące dane treningowe, na podstawie których parametry w modelu S-ARIMA są określane zgodnie z procedurą opisaną w podrozdziale 9.3. Jest to bardzo ważne, aby zrozumieć, jak trudne jest zadanie modelu prognostycznego, ponieważ w tym przypadku dysponuje on dwoma latami danych jako danymi wejściowymi i musi prognozować przyszły popyt z rocznym wyprzedzeniem! Jest to dość powszechny scenariusz w łańcuchach dostaw i logistyce, ponieważ istnieje niepisana zasada, że firmy przechowują historię swoich danych przez trzy lata, po czym je odrzucają.



Rysunek 9.4 Trening, test i prognozowany popyt modelu S-ARIMA $(5,0,1)(1,0,0)_{52}$

Lewy dolny panel na rysunku 9.4 przedstawia wydajność obserwowanego modelu. Wydajność modelu jest już przedstawiona w tabeli 9.1 za pomocą różnych miar statystycznych, ale dla menedżerów zwykle trudno jest uzyskać opinię, jak dobry lub zły jest model. W tym celu panel graficznie przedstawia wydajność modelu. Można argumentować, że model dość dobrze podąża za danymi testowymi i w większości okresów wykazuje

doskonałą wydajność. W celu wygenerowania przyszłych prognoz, model S-ARIMA $(5,0,1)(1,0,0)_{52}$ został ponownie dopasowany poprzez dodanie danych z 2014 roku do danych treningowych. Następnie model wygenerował 52-tygodniowe prognozy na rok 2015. Prognozy na rok 2015 przedstawiono w lewym dolnym panelu. Prognozom towarzyszą 80% i 95% przedziały predykcyjne, które pokazują, że możliwe przyszłe prognozy różnią się od średnich przewidywanych wartości. Model przewiduje dalszy spadek popytu, który rozpoczął się w 2014 roku. Przyczyny tego spadku mogą być różne i powinny być dalej badane przez menedżerów na strategicznym poziomie firmy.

BIBLIOGRAFIA DO ROZDZIAŁU 9.

- Hyndman, R., & Khandakar, Y. (2007). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software, 27*(3). http://www.jstatsoft.org/v27/i03
- Makridakis, S., Wheelwright, S., & Hyndman, R. J. (1998). *Forecasting: Methods and applications* (3rd ed.). New York: John Wiley & Sons.
- Makridakis, S., Wheelwright, S., & McGee, V. (1983). *Forecasting: Methods and applications* (2nd ed.). New York: John Wiley & Sons.
- Mircetic, D. (2018). *Boosting the performance of top-down methodology for forecasting in supply chains via a new approach for determining disaggregating proportions*. University of Novi Sad.
- Mircetic, D., Nikolicic, S., Maslaric, M., Ralevic, N., & Debelic, B. (2016). Development of S-ARIMA model for forecasting demand in a beverage supply chain. *Open Engineering, 6*(1).
- Mircetic, D., Rostami-Tabar, B., Nikolicic, S., & Maslaric, M. (2022). Forecasting hierarchical time series in supply chains: An empirical investigation. *International Journal of Production Research, 60*(8), 2514-2533.
- Mircetic, D., Nikolicic, S., Stojanovic, D., & Maslaric, M. (2017). Modified top-down approach for hierarchical forecasting in a beverage supply chain. *Transportation Research Procedia, 22*, 193–202.
- Nikolopoulos, K., Punia, S., Schäfers, A., Tsinopoulos, C., & Vasilakis, C. (2020). Forecasting and planning during a pandemic: COVID-19 growth rates, supply chain disruptions, and governmental decisions. *European Journal of Operational Research, 290*(1), 99–115.
- Rostami-Tabar, B. (2013). *ARIMA demand forecasting by aggregation*. Université Sciences et Technologies Bordeaux I.
- Rostami-Tabar, B., & Mircetic, D. (2023). Exploring the association between time series features and forecasting by temporal aggregation using machine learning. *Neurocomputing, 548*, 126376.
- Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research, 252*(1), 1-26.