

WOJSKOWA AKADEMIA TECHNICZNA
im. Jarosława Dąbrowskiego
WYDZIAŁ CYBERNETYKI



SPRAWOZDANIE
Metody Eksploracji Danych

Temat laboratorium: **ANALIZA REGRESJI – REGRESJA LINIOWA, PRZEKSZTAŁCANIE FUNKCJI NIELINIOWYCH W RÓWNOWAŻNE LINIOWE, BUDOWA MODELU NA PODSTAWIE DANYCH**

INFORMATYKA

(kierunek studiów)

INŻYNIERIA SYSTEMÓW – ANALIZA DANYCH

(specjalność)

Zespół:

Michał ŚLĘZAK
Szymon OLEŚKIEWICZ

Prowadzący laboratorium:

Dr inż. Romuald Hoffmann, prof.
WAT

Spis treści

Rozdział I. Zadanie 1 - Facebook	4
I.1. Wykorzystane narzędzia i zależności	5
I.2. Co chcemy zbadać i dlaczego ?	10
Wnioski	11
Rozdział II. Analiza zależności przychodów od czasu.....	12
II.1. Modele regresji	12
Rozdział III. Analiza zależności przychodów od liczby użytkowników.....	15
III.1. Modele regresji.....	15
III.2. Model hybrydowy	17
Rozdział IV. Analiza zależności liczby użytkowników od czasu	19
IV.1. Model potęgowy	19
Rozdział V. Prognoza zmiennych na lata 2018-2023.....	20
V.1. Prognoza liczby użytkowników na lata 2018-2023.....	20
V.2. Prognoza przychodów na lata 2018-2023	21
Wnioski.....	21
Bibliografia.....	22
Spis rysunków	23
Spis tabel	23
Załączniki	23

Rozdział I. Zadanie 1 - Facebook

METODY EKSPLORACJI DANYCH

**Laboratorium. Analiza regresji - regresja liniowa
Przekształcanie funkcji nieliniowych w równoważne liniowe
Budowa modelu na podstawie danych (ogólnie dostępnych)**

ZADANIE

Mamy zgromadzone dane dot. dobrze znanego przedsiębiorstwa internetowego Meta w tym odnoszące się do portalu „Facebook”, które zostały zawarte w poniższych tabelach. Dane te dotyczą liczby użytkowników (klientów) w rozliczeniu na kwartały w poszczególnych latach. Pozostałe dane to przychody liczone w milionach dolarów amerykańskich.

Tabela 1. Liczba użytkowników portalu społecznościowego „Facebook”

Kwartał w mln	Liczba użytkowników	Liczba użytkowników	
		Kwartał w mln	Liczba użytkowników
Q3 '08	100	Q1 '09	197
		Q2 '09	242
		Q3 '09	305
		Q4 '09	360
		Q1 '10	431
		Q2 '10	482
		Q3 '10	550
		Q4 '10	608
		Q1 '11	680
		Q2 '11	739
		Q3 '11	800
		Q4 '11	845
		Q1 '12	901
		Q2 '12	955
		Q3 '12	1007
		Q4 '12	1056
		Q1 '13	1110
		Q2 '13	1155
		Q3 '13	1189
		Q4 '13	1228

Tabela 2. Przychody firmy Meta”

Rok	Przychód w mln \$
2007	153
2008	272
2009	777
2010	1974
2011	3711
2012	5089
2013	7872
2014	12466
2015	17928
2016	27638
2017	40653

- 1) Proszę przeanalizować dane i zastanowić się nad tym „Co chcemy zbadać i dlaczego?”, tzn. na jakie pytania chcemy sobie odpowiedzieć przy pomocy jakich modeli.
- 2) W powyższym kontekście, na gruncie regresji, należy zaproponować **model** lub **modele badające wybrane zależności** i wyliczyć m. in. ich parametry strukturalne, odchylenia standardowe, miary dopasowania, czy też przebadać hipotezy, itp.
- 3) Analizę proszę na początek przeprowadzić na podstawie danych, które zostały przedstawione w powyższych tabelach.
- 4) Propowane modele należy uzasadnić oraz wskazać ewentualny sposób ich praktycznego użycia w tym np. do predykcji.
- 5) Na podstawie opracowanych modeli i przeprowadzonych obliczeń **sformułować własne wnioski**.
- 6) Następnie wykorzystując zbiór danych znajdujący się w pliku programu Excel o nazwie „MED-Lab-1-Zadanie-1-Dane.xlsx” należy
 - a) za lata 2018-2020, sprawdzić zachowanie się (wykorzystanie) modeli np. w procesie predykcji za lata 2018-2020.
 - b) za lata 2021-2023, sprawdzić zachowanie się (wykorzystanie) modeli np. w procesie predykcji za lata 2021-2023.
- 7) Wyniki swoich analiz zawierających: postawione pytania „badawcze”, hipotezy, wzory, udokumentowane wyniki obliczeń, wnioski, itp. proszę zatrudnić w postaci **społowodzanie**. Obliczenia można przeprowadzić w dowolnie wybranym narzędziu, np. które do tej pory wykorzystywaliśmy na zajęciach. Do sprawozdania proszę dodać jako załączniki wszystkie pliki z obliczeniami.

I.1. Wykorzystane narzędzia i zależności

I.1.1. Modele

W celu wyznaczenia modeli liniowych i modeli zlinearyzowanych wraz z ich parametrami na podstawie wykładów oraz wiedzy własnej przygotowano program w języku Python wykorzystujący biblioteki Pandas, NumPy, stastmodel oraz scipy oraz użyto narzędzia Gretl do obliczeń oraz wizualizacji ich wyników. Modele oraz ich parametry zostały wyznaczone za pomocą wyżej wspomnianych narzędzi, które oparte są o metodę MNK (Metoda Najmniejszych Kwadratów), gdzie wraz z dopasowaniem danych, wyznaczono wiele parametrów i metryk dla danego stworzonego modelu. Model liniowy jednej zmiennej wyraża się poniższym wzorem.

$$\hat{y} = a_1x + a_0 \quad (1)$$

Należy wyznaczyć takie wartości parametrów a_1 i a_0 , aby najlepiej dopasować przebieg prostej regresji do wartości danych. Aby tego dokonać, należy rozwiązać poniższe zadanie optymalizacyjne.

$$M(a_0, a_1) = \sum_{i=1}^n [y_i - (a_1x_i + a_0)]^2 \quad (2)$$

Przyrównując pochodne funkcji sumy kwadratów reszt po zmiennych a_0 i a_1 do zera, otrzymamy układ równań, który w postaci macierzowej będzie mieć następującą postać.

$$\begin{bmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{bmatrix} \quad (3)$$

Po lewostronnym przemnożeniu obu stron równania przez macierz odwrotną do stojącej przy macierzy parametrów, otrzymamy wzór na wartość macierzy tych parametrów.

$$\begin{bmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{bmatrix} = \begin{bmatrix} a_1 \\ a_0 \end{bmatrix} \quad (4)$$

Powyższy tok rozumowania można uogólnić dla regresji wielu zmiennych. W tym celu można wykorzystać zapis macierzowy, gdzie po przekształceniach otrzymamy ostatecznie wzór na wartości współczynników dla dowolnej liczby parametrów.

$$A = \begin{bmatrix} a_0 \\ \dots \\ a_k \end{bmatrix} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y \quad (5)$$

W powyższym wzorze macierze X oraz Y zostały wyznaczone na podstawie zbioru danych, takich jak poniżej.

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ \dots \\ y_k \\ \dots \\ y_n \end{bmatrix} \quad (6)$$

W celach realizacji zadania wykorzystano także modele nieliniowe oraz zastosowano ich linearyzację. Wykorzystano następujące, poniższe modele, wraz z pokazaniem sposobu ich linearyzacji.

Niech model będzie **funkcją wykładniczą** z jedną zmienną objaśniającą:

$$\hat{W} = b \cdot a^x \quad (7)$$

Powyższy model sprowadzamy do modelu liniowego logarytmując obustronnie otrzymując:

$$\log \hat{W} = \log b + X \cdot \log a \quad (8)$$

Przyjmując, że $\hat{Y} = \log \hat{W}$, $a_0 = \log b$, $a_1 = \log a$, otrzymujemy następujący równoważny liniowy model:

$$\hat{Y} = a_1 \cdot X + a_0 \quad (9)$$

Rozpatrzmy **model potęgowy** z jedną zmienną objaśniającą:

$$\hat{W} = b \cdot X^{a_1} \quad (10)$$

Powyższy model sprowadzamy do modelu liniowego logarytmując obustronnie otrzymując:

$$\log \hat{W} = \log b + a_1 \cdot \log X \quad (11)$$

Przyjmując, że $\hat{Y} = \log \hat{W}$, $a_0 = \log b$, $X_1 = \log X$ otrzymujemy następujący równoważny liniowy model:

$$\hat{Y} = a_1 \cdot X_1 + a_0 \quad (12)$$

Niech modelem będzie **wielomianem stopnia pierwszego** (model z jedną zmienną objaśniającą):

$$\hat{Y} = a_0 + a_1 \cdot X \quad (13)$$

Powyższy model sprowadzamy do modelu liniowego podstawiając $X = X_1$

W konsekwencji wykonanych podstawięń otrzymujemy równoważny liniowy model:

$$\hat{Y} = a_1 \cdot X_1 + a_0 \quad (14)$$

Dodatkowo, w celu poprawy jakości dopasowania w przypadku gdy jeden model niedoszacowuje a drugi przeszacowuje, zbudowano model złożony jako kombinację modeli potęgowego i wykładniczego.

$$\hat{Y}_{\text{hyb}} = \alpha \cdot \hat{Y}_{\text{pot}} + (1 - \alpha) \cdot \hat{Y}_{\text{wyk}} \quad (15)$$

, gdzie $\alpha \in [0,1]$ jest parametrem wagi określającym udział modelu potęgowego w prognozie. Parametr α został wyznaczony drogą optymalizacji wartości średniego błędu kwadratowego w postaci pierwiastka (RMSE), który ma poniższy wzór.

$$RMSE(Y, \hat{Y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (16)$$

, gdzie y_i oznacza wartość rzeczywistą a \hat{y}_i wartość prognozowaną przez dany model. Dla modelu łączącego potęgowy i wykładniczy, funkcja błędu RMSE wyglądać będzie w takim razie następująco.

$$RMSE_{\text{hyb}}(\alpha) = RMSE(Y, \alpha \cdot \hat{Y}_{\text{pot}} + (1 - \alpha) \cdot \hat{Y}_{\text{wyk}}) \quad (17)$$

Minimalizacja funkcji $RMSE_{\text{hyb}}(\alpha)$ na przedziale $[0,1]$ została przeprowadzona za pomocą funkcji optymalizacyjnej `minimize_scalar()` z modułu `scipy` do znalezienia minimum lokalnego w danym przedziale $[x1, x2]$. Algorytm ten wyszukuje taką wartość $\alpha \in [0,1]$, dla której błąd $RMSE_{\text{hyb}}$ jest najmniejszy.

$$\alpha = \arg \min_{\alpha \in [0,1]} RMSE_{\text{hyb}}(\alpha) \quad (18)$$

I.1.2. Ocena predykcji modelu

W celu oceny predykcji modelu należy rozróżnić dwa typy predykcji: ex ante i post ante.

Predykcja ex ante ma na celu przewidzieć rzeczywistą wartość przed jej poznaniem. W tym celu należy wyznaczyć średni błąd predykcji ex ante dla wskazanej chwili lub wartości τ . Jego wzór podano poniżej.

$$S_{\tau}^P = \sqrt{S_e^2 + x_{\tau} \cdot D^2(A) \cdot x_{\tau}^T} \quad (19)$$

Na podstawie tak wyznaczonej wartości można wyznaczyć wartość względną średniego błędu predykcji w chwili τ jak poniżej.

$$v_{\tau}^P = \frac{S_{\tau}^P}{y_{\tau}^P} \cdot 100\% \quad (20)$$

Predykcja post ante ma na celu zweryfikowanie poprawności przewidywanych przez model wartości do wartości rzeczywistych danych. Istnieje wiele metryk błędu prognozy post ante, poniżej znajduje się omówienie niektórych z nich.

Średni błąd predykcji ME pokazuje czy model systematycznie przeszacowuje (dodatni ME) lub nie doszacowuje (ujemny ME) wartości rzeczywistych danych. Wartości powinny być jak najbardziej bliskie zera. Opisuje go poniższy wzór.

$$ME = \frac{1}{m} \sum_{\tau=1}^m (y_{\tau} - y_{\tau}^p) \quad (21)$$

Średni błąd absolutny MAE mierzy przeciętną wielkość błędu bez uwzględniania jego kierunku, jego wartość jest zawsze dodatnia. Pokazuje jak bardzo przewidywane wartości są średnio odległe od danych rzeczywistych. Określony jest on poniższym wzorem.

$$MAE = \frac{1}{m} \sum_{\tau=1}^m |y_{\tau} - y_{\tau}^p| \quad (22)$$

Pierwiastek średniego błędu kwadratowego RMSE mocniej wyróżnia większe błędy przez podniesienie do kwadratu. Opisuje go poniższy wzór.

$$RMSE = \sqrt{\frac{1}{m} \sum_{\tau=1}^m (y_{\tau} - y_{\tau}^p)^2} \quad (23)$$

Jak już wcześniej wspomniano, do obliczeń wykorzystano narzędzia w postaci języka programowania Python oraz narzędzia Gretl, które, m.in. obliczają niektóre metryki.

Odchylenie standardowe modelu S_e uzyskamy pierwiastkując wariancję S_e^2 .

$$S_e^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - k - 1} \quad (24)$$

Na podstawie wartości odchylenia standardowego S_e modelu jednej zmiennej można wyznaczyć odchylenia parametrów a_0 i a_1 , tak jak poniżej.

$$S_{a_0} = S_e \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x}_i)^2}} \quad (25)$$

$$S_{a_1} = \frac{S_e}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_i)^2}} \quad (26)$$

W celu wyznaczenia odchylenia standardowego każdego parametru modelu należy wyznaczyć jego kowariancję jak poniżej.

$$\text{Cov}(A) = D^2(A) = S_e^2(X^T \cdot X)^{-1} \quad (27)$$

W celu zbadania jak dobrze wyznaczona prosta regresji oddaje przebieg zmiennej można wykorzystać współczynnik determinacji, który obliczany jest na podstawie poniższego wzoru. Współczynnik ten przyjmuje wartości od 0 do 1. Im bliżej 1, tym nasz model jest bardziej dopasowany.

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (28)$$

Następną miarą oceny dopasowania modelu do danych rzeczywistych jest współczynnik zmienności losowej, którego wzór zamieszczono poniżej. Współczynnik ten porównuje się z wartością krytyczną $W_e^* \in [0\%, 25\%]$, która gdy $W_e \leq W_e^*$ to model jest dobrze dopasowany.

$$W_e = \frac{S_e}{\bar{y}} \cdot 100\% \quad (29)$$

W stworzonym modelu warto też sprawdzić poziom istotności parametrów. W tym celu sprawdzamy czy współczynniki a_0 i a_1 istotnie różnią się od zera. W tym celu można wykonać poniższy test.

Hipotezy:

$$H_0: a_i = 0$$

$$H_1: a_i \neq 0$$

Statystyka testu:

$$I_j = \frac{|a_j|}{S(a_j)}$$

(30)

Wartość krytyczna:

$$t_k = t_{\alpha, n-k-1} : P(Y \geq t_{\alpha, n-k-1})$$

(31)

Jeśli $I_j > t_k$ to odrzucamy H_0 i przyjmujemy H_1 , w przeciwnym wypadku mamy brak podstaw do odrzucenia H_0 . Wartości testu t wyznaczono za pomocą narzędzia Gretl a wartość krytyczną na podstawie tablic statystycznych.

I.2. Co chcemy zbadać i dlaczego ?

Podstawowym celem analiz, była potrzeba prognozy rozwoju (na lata 2018-2024) portalu ‘Facebook’ (Meta), pod względem liczby użytkowników oraz przychodów. W tym celu wymagane jest znalezienie zależności pomiędzy posiadanymi wskaźnikami (liczba użytkowników (kwartalna) i przychody (roczne)), wykorzystując również zmienną czasową (podane dane występują w postaci szeregów czasowych). Postawiono zatem następujące hipotezy:

Hipoteza 1: Zależność przychodu od czasu

H_1 : Przychody portalu Facebook wykazują istotną statystycznie zależność od czasu, przy czym w latach 2018-2024 obserwuje się trend wzrostowy przychodów zgodny z modelem liniowym lub nieliniowym (np. wykładniczym, wielomianowym).

H_0 (hipoteza zerowa): Nie istnieje istotna statystycznie zależność między przychodami a czasem, co oznacza brak systematycznego trendu w analizowanym okresie.

Hipoteza 2: Zależność przychodu od liczby użytkowników

H_2 : Przychody portalu Facebook są istotnie statystycznie skorelowane z liczbą użytkowników, przy czym wzrost liczby użytkowników prowadzi do proporcjonalnego (lub bardziej niż proporcjonalnego) wzrostu przychodów.

H_0 (hipoteza zerowa): Nie istnieje istotna statystycznie zależność między przychodami a liczbą użytkowników, co oznacza że zmiany w liczbie użytkowników nie wpływają na poziom przychodów.

Hipoteza 3: Zależność liczby użytkowników od czasu

H₃: Liczba użytkowników portalu Facebook wykazuje istotną statystycznie zależność od czasu, przy czym w latach 2018-2024 obserwuje się trend wzrostowy liczby użytkowników, który może być opisany modelem liniowym, logistycznym lub innym modelem wzrostu.

H₀ (hipoteza zerowa): Nie istnieje istotna statystycznie zależność między liczbą użytkowników a czasem, co oznacza stabilizację lub brak systematycznych zmian w bazie użytkowników.

Wnioski

...

Rozdział II. Analiza zależności przychodów od czasu

II.1. Modele regresji

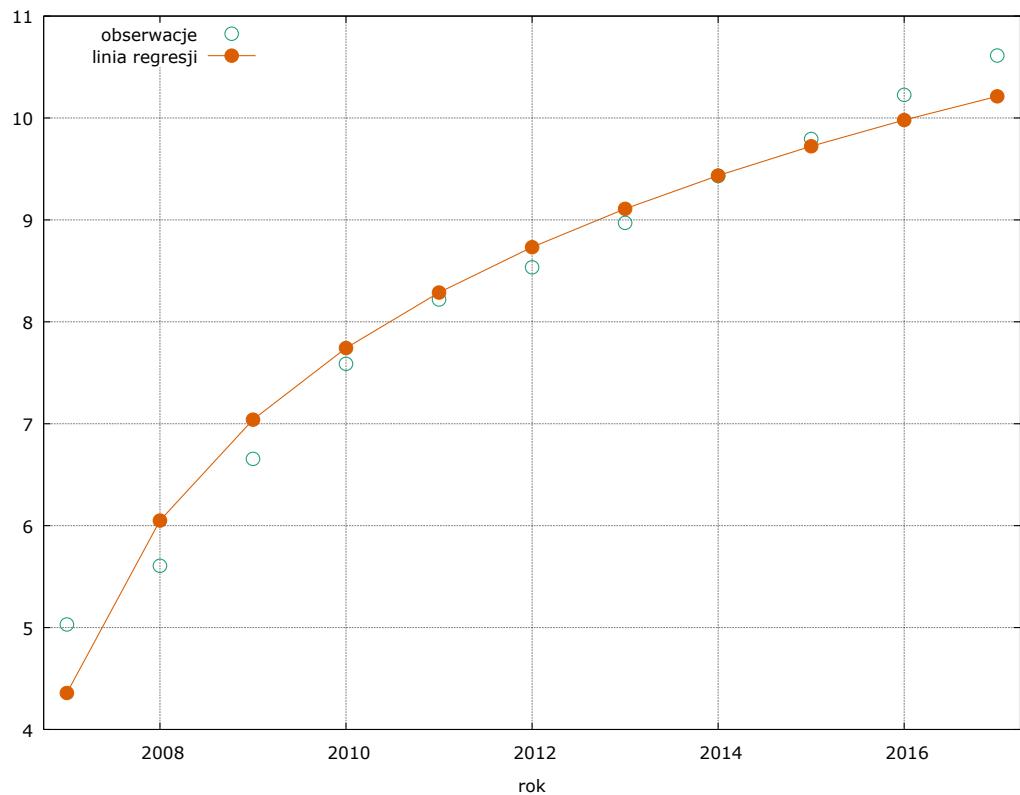
Po przeprowadzeniu wstępnej analizy danych oraz biorąc pod uwagę naturę badanego zjawiska (przychodów) jako potencjalne modele opisujące zależność wybrano model wykładniczy oraz model potęgowy. Aby dobrać współczynniki w modelach dokonano ich linearyzacji. Do estymacji metodą MNK, wykorzystano obserwacje z przedziału 2007-2017.

II.1.1. Model potęgowy

Wykorzystano równanie modelu potęgowego (12). Przyjmując za zmienną objaśnianą przychód, a za zmienną objaśniającą czas (rok).

$$\hat{Y} = a_1 \cdot X_1 + a_0$$

Parametr	Wartość	Błąd stand.	t-Studenta	wartość p
a_0	4,35769	0,25985	16,77	<0,0001
a_1	2,44165	0,14913	16,37	<0,0001
S_e	0,35137			
R^2	0,96752			
W_e	4,26%			

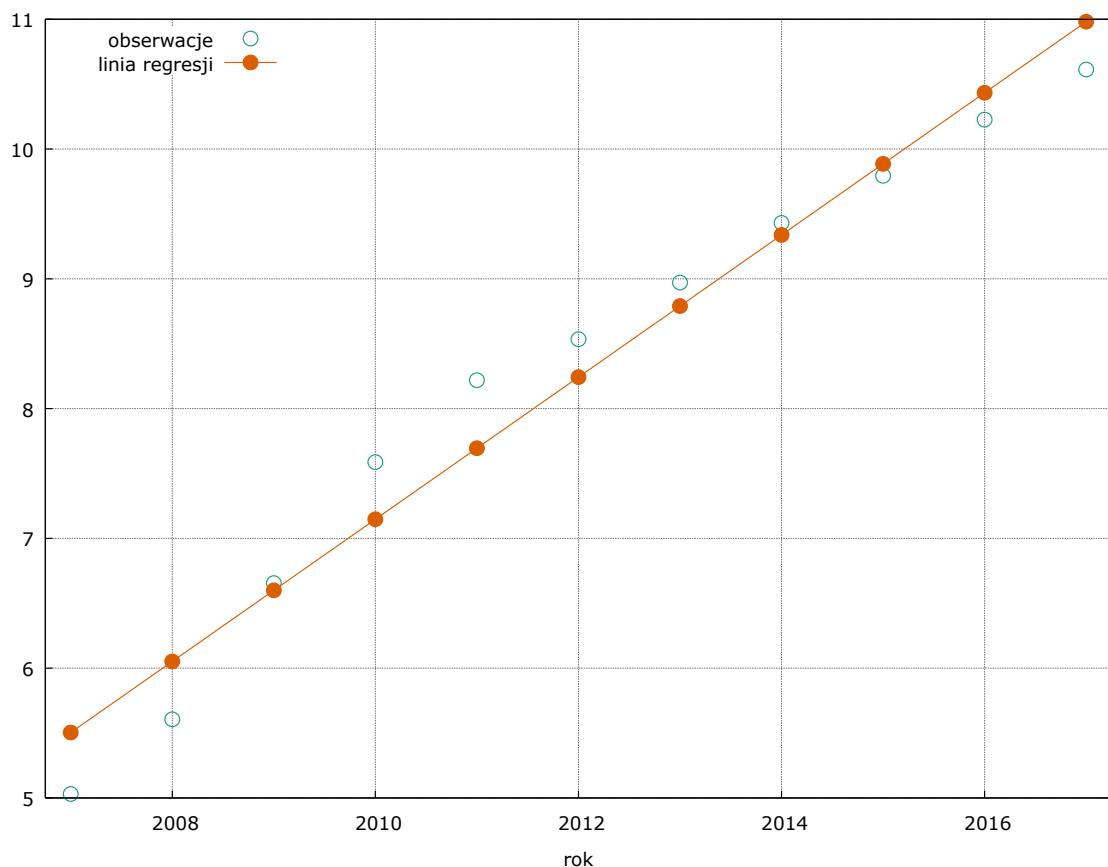


II.1.2. Model wykładniczy

Wykorzystano równanie modelu wykładniczego (-). Przyjmując za zmienną objaśnianą przychód, a za zmienną objaśniającą czas (rok).

$$\hat{Y} = a_1 \cdot X_1 + a_0$$

Parametr	Wartość	Błąd stand.	t-Studenta	wartość p
a_0	4,95626	0,23698	20,91	<0,0001
a_1	0,547731	0,034941	15,68	<0,0001
S_e	0,36645			
R^2	0,96467			
W_e	4,45%			



Rozdział III. Analiza zależności przychodów od liczby użytkowników

III.1. Modele regresji

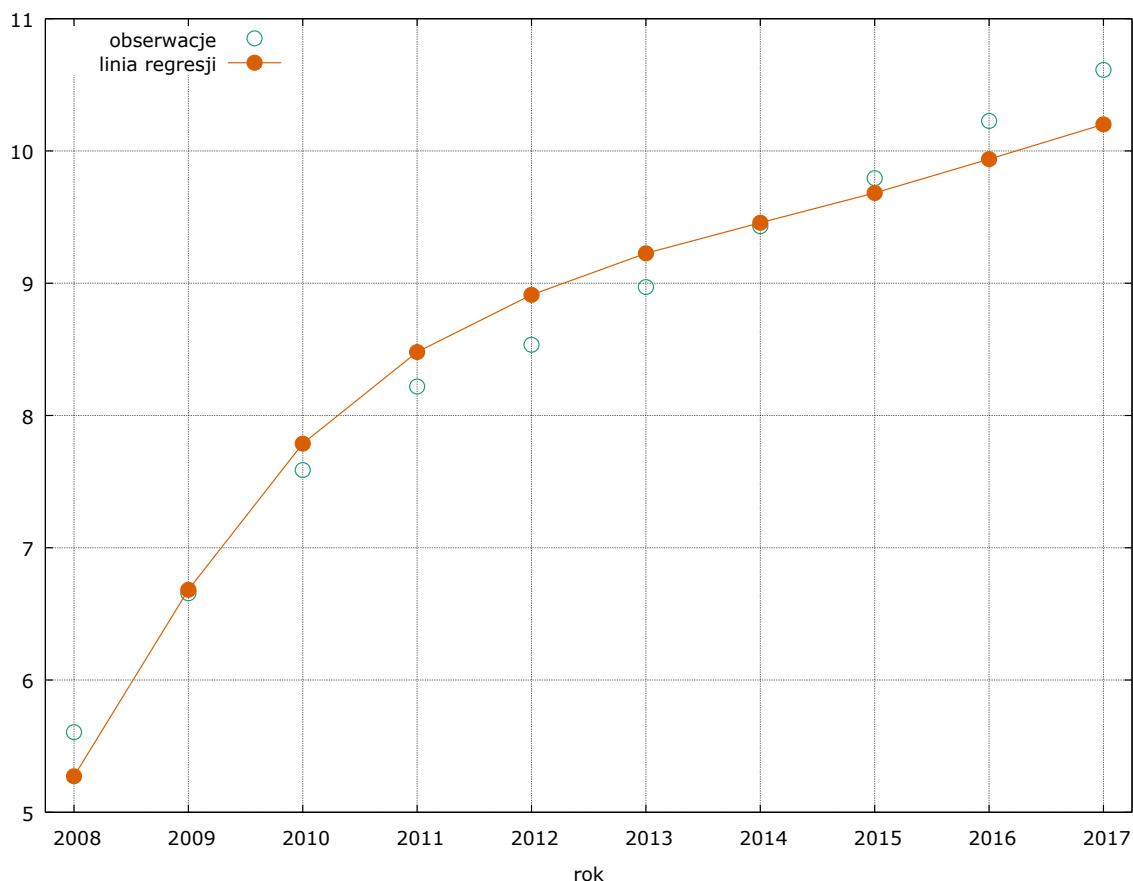
Aby polepszyć wyniki modeli zależności przychodów od czasu wykonano analizę zależności przychodów od liczby użytkowników. Ponownie poddano analizie model potęgowy oraz model wykładniczy, dokonując linearyzacji. Wykorzystano obserwacje z przedziału 2008-2017.

III.1.1. Model potęgowy

Wykorzystano równanie modelu potęgowego (12). Przyjmując za zmienną objaśnianą przychód, a za zmienną objaśniającą liczbę użytkowników.

$$\hat{Y} = a_1 \cdot X_1 + a_0$$

Parametr	Wartość	Błąd stand.	t-Studenta	wartość p
a_0	-3,2192	0,74255	-4,335	0,0025
a_1	1,76160	0,11014	15,99	<0,0001
S_e	0,29435			
R^2	0,96968			
W_e	3,43%			

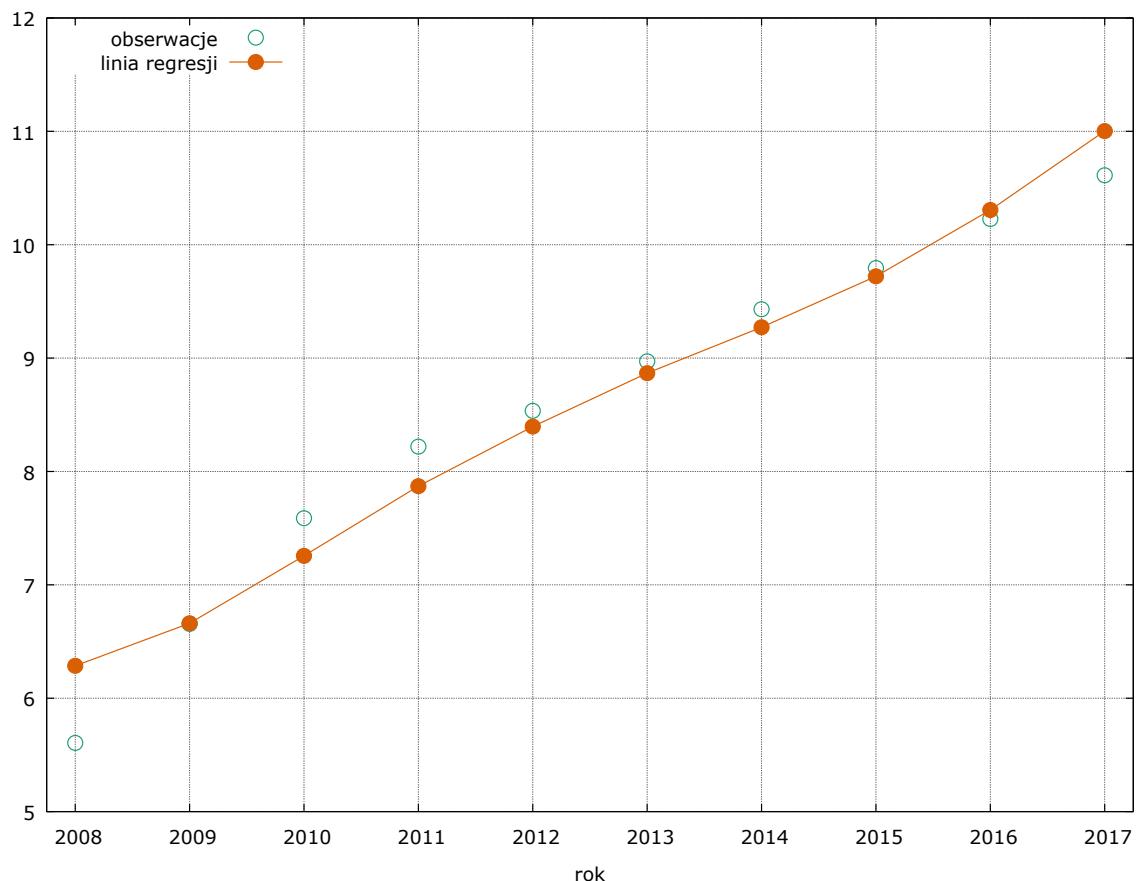


III.1.2. Model wykładniczy

Wykorzystano równanie modelu wykładniczego (-). Przyjmując za zmienną objaśnianą przychód, a za zmienną objaśniającą liczbę użytkowników.

$$\hat{Y} = a_1 \cdot X_1 + a_0$$

Parametr	Wartość	Błąd stand.	t-Studenta	wartość p
a_0	5,97978	0,214765	27,84	<0,0001
a_1	0,00246808	0,000177954	13,87	<0,0001
S_e	0,33778			
R^2	0,96007			
W_e	3,94%			



III.2. Model hybrydowy

Zauważono, iż model potęgowy wydaje się przeszacowywać zmienną objaśnianiom, natomiast model wykładniczy wydaje się ją niedoszacowywać, zauważono zatem, że oczekiwana wartość znajduje się gdzieś pomiędzy prognozami obu modeli. Zdecydowano o połączeniu modeli w nowy model hybrydowy, będący ważoną kombinacją prognoz obydwu modeli. Znalezienie odpowiednich wag umożliwiło dobranie optymalnych proporcji udziału modeli potęgowego i wykładniczego, które minimalizują błąd.

$$\hat{Y}_{\text{hyb}} = \alpha \cdot (a_1 \cdot X_1 + a_0) + (1 - \alpha) \cdot (b_1 \cdot X_1 + b_0)$$

OLS Regression Results							
Dep. Variable:	y	R-squared:	0.995				
Model:	OLS	Adj. R-squared:	0.994				
Method:	Least Squares	F-statistic:	1621.				
Date:	Mon, 08 Dec 2025	Prob (F-statistic):	1.59e-10				
Time:	18:53:34	Log-Likelihood:	8.2586				
No. Observations:	10	AIC:	-12.52				
Df Residuals:	8	BIC:	-11.91				
Df Model:	1						
Covariance Type:	nonrobust						
	coef	std err	t	P> t	[0.025	0.975]	
const	3.5456	0.130	27.245	0.000	3.246	3.846	
x1	2.8672	0.071	40.264	0.000	2.703	3.031	
Omnibus:		0.268	Durbin-Watson:	0.923			
Prob(Omnibus):		0.874	Jarque-Bera (JB):	0.410			
Skew:		0.078	Prob(JB):	0.815			
Kurtosis:		2.020	Cond. No.	8.13			
OLS Regression Results							
Dep. Variable:	y	R-squared:	0.962				
Model:	OLS	Adj. R-squared:	0.957				
Method:	Least Squares	F-statistic:	199.9				
Date:	Mon, 08 Dec 2025	Prob (F-statistic):	6.09e-07				
Time:	18:53:34	Log-Likelihood:	-2.0356				
No. Observations:	10	AIC:	8.071				
Df Residuals:	8	BIC:	8.676				
Df Model:	1						
Covariance Type:	nonrobust						
	coef	std err	t	P> t	[0.025	0.975]	
const	5.2088	0.259	20.077	0.000	4.611	5.807	
x1	0.5162	0.037	14.138	0.000	0.432	0.600	
Omnibus:		1.505	Durbin-Watson:	0.646			
Prob(Omnibus):		0.471	Jarque-Bera (JB):	0.605			
Skew:		-0.593	Prob(JB):	0.739			
Kurtosis:		2.794	Cond. No.	17.9			

Parametr	Wartość	Błąd stand.	t-Studenta	wartość p
α	0.971?	n/d	n/d	n/d
a_0	3.5456	0.130	27.245	<0.0001
a_1	2.8672	0.071	40.264	<0.0001
b_0	5.2088	0.259	20.077	<0.0001
b_1	0.5162	0.037	14.138	<0.0001
S_e	2362.52406 8969097			
R^2	0.96856799 12858128			
W_e	19%			

Rozdział IV. Analiza zależności liczby użytkowników od czasu

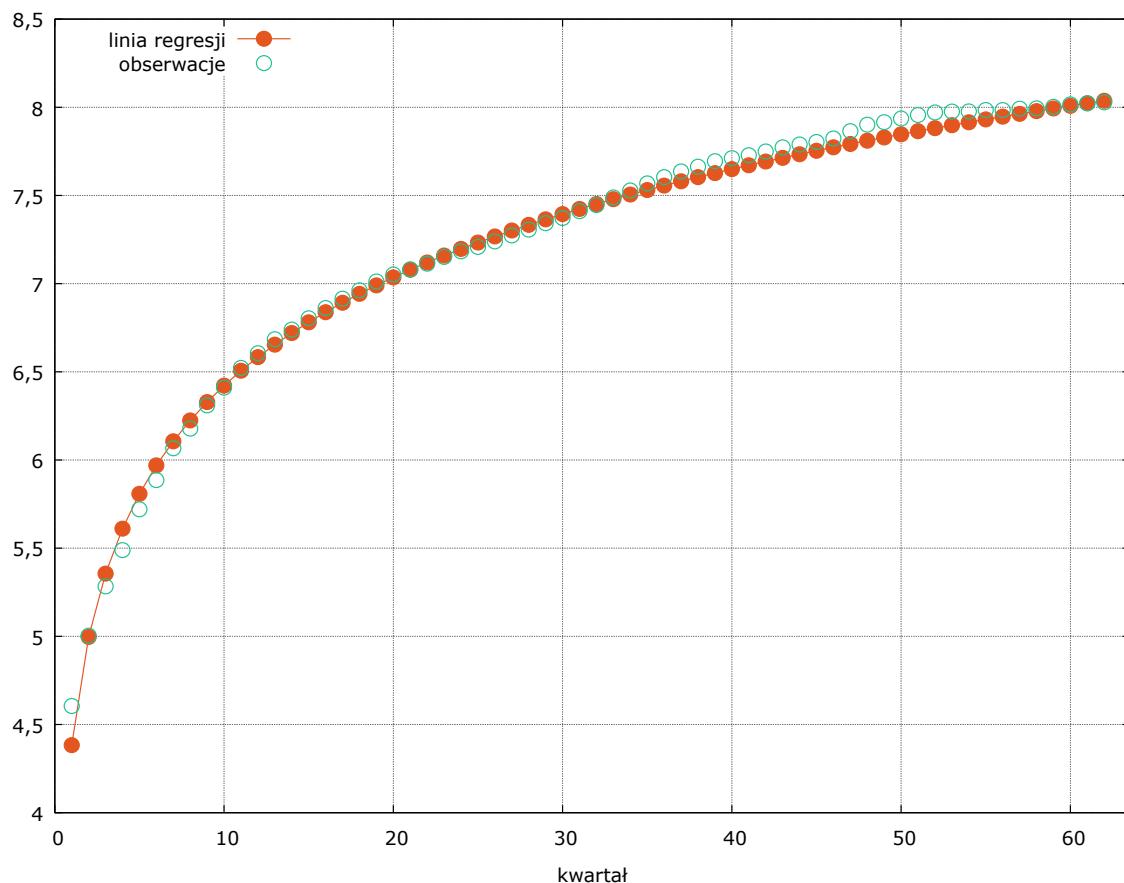
IV.1. Model potęgowy

Na podstawie wstępnej analizy danych oraz uwzględniając naturę badanego zjawiska (przyrost liczby użytkowników), najlepszym modelem opisującym zależność liczby użytkowników od czasu będzie model potęgowy, ponieważ uwzględnia on stopniowe hamowanie wzrostu.

Wykorzystano równanie modelu potęgowego (12). Przyjmując za zmienną objaśnianą liczbę użytkowników, a za zmienną objaśniającą czas (kwartał - liczone od trzeciego kwartału roku 2008). Wykorzystano obserwacje z przedziału Q3'08 – Q4 '17.

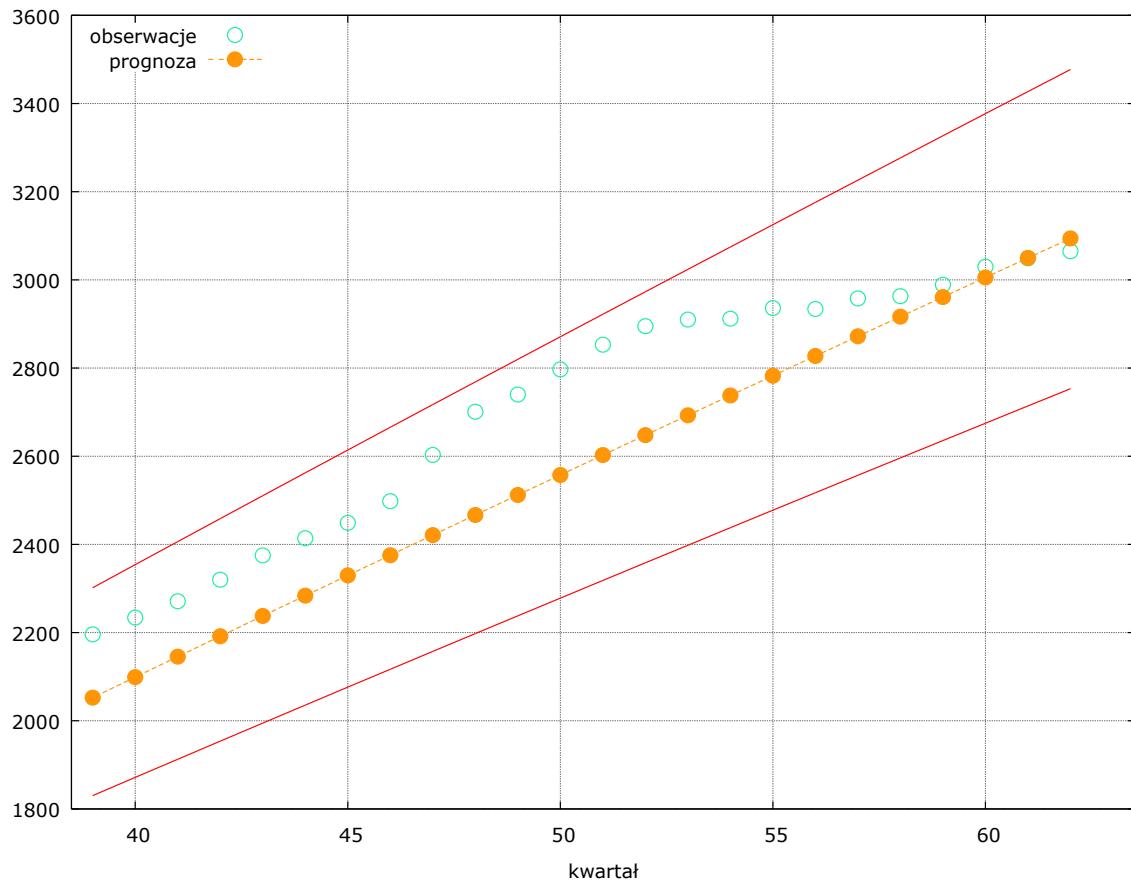
$$\hat{Y} = a_1 \cdot X_1 + a_0$$

Parametr	Wartość	Błąd stand.	t-Studenta	wartość p
a_0	4,38310	0,0295135	148,5	<0,0001
a_1	0,885391	0,0103837	85,27	<0,0001
S_e	0,054920			
R^2	0,995073			
W_e	0,81%			



Rozdział V. Prognoza zmiennych na lata 2018-2023

V.1. Prognoza liczby użytkowników na lata 2018-2023

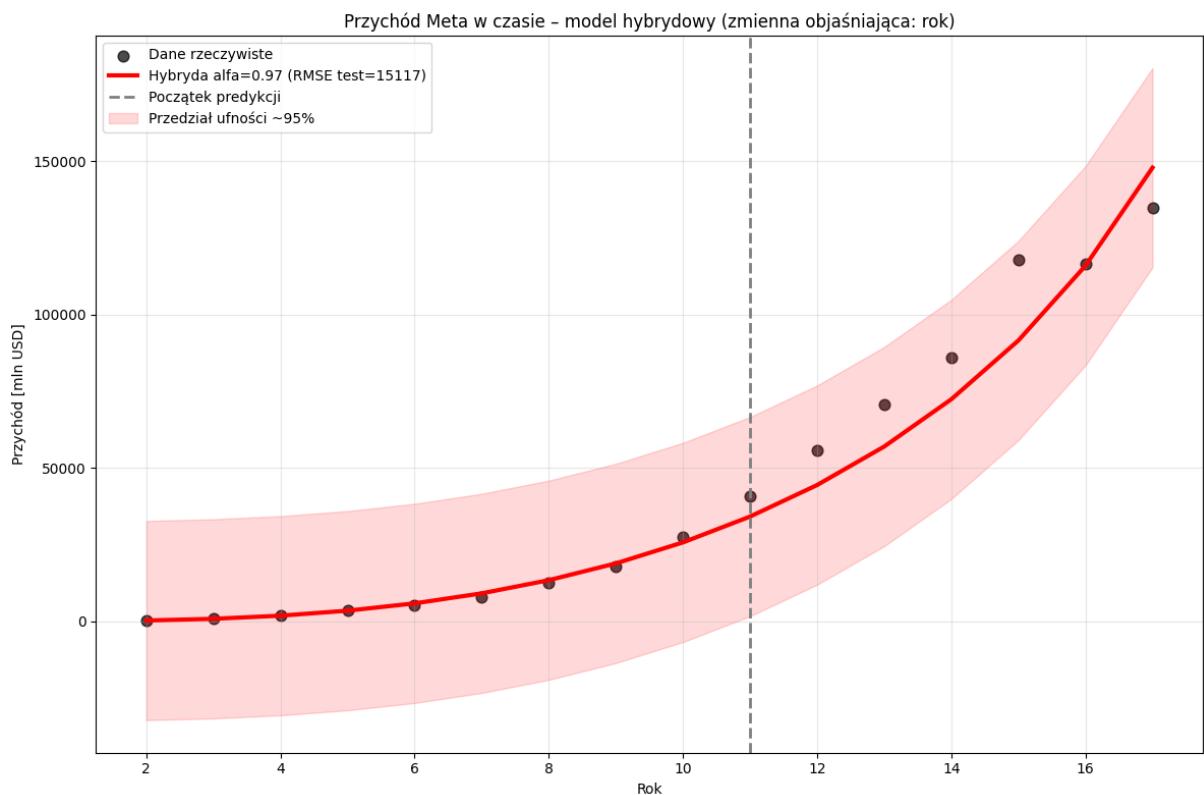


Dla 95% przedziału ufności, $t(36, 0,025) = 2,028$

kwartał	obserwacja	prognoza	błąd standardowy	95% przedział
39	7,69439	7,62679	0,0565131	(7,51217, 7,74140)
40	7,71155	7,64920	0,0565597	(7,53450, 7,76391)
41	7,72798	7,67107	0,0566064	(7,55626, 7,78587)
42	7,74932	7,69240	0,0566530	(7,57750, 7,80730)
43	7,77275	7,71324	0,0566995	(7,59824, 7,82823)
44	7,78904	7,73359	0,0567460	(7,61850, 7,84868)
45	7,80344	7,75349	0,0567923	(7,63831, 7,86867)
46	7,82325	7,77295	0,0568385	(7,65767, 7,88822)
47	7,86442	7,79199	0,0568846	(7,67662, 7,90736)
48	7,90138	7,81063	0,0569305	(7,69517, 7,92609)
49	7,91571	7,82889	0,0569763	(7,71333, 7,94444)
50	7,93630	7,84677	0,0570218	(7,73113, 7,96242)

51	7,95613	7,86431	0,0570672	(7,74857, 7,98004)
52	7,97074	7,88150	0,0571124	(7,76567, 7,99733)
53	7,97591	7,89836	0,0571573	(7,78244, 8,01428)
54	7,97660	7,91491	0,0572021	(7,79890, 8,03093)
55	7,98480	7,93116	0,0572467	(7,81506, 8,04726)
56	7,98412	7,94711	0,0572910	(7,83092, 8,06331)
57	7,99227	7,96278	0,0573351	(7,84650, 8,07907)
58	7,99396	7,97818	0,0573790	(7,86181, 8,09455)
59	8,00269	7,99332	0,0574226	(7,87686, 8,10978)
60	8,01632	8,00820	0,0574660	(7,89165, 8,12475)
61	8,02257	8,02283	0,0575092	(7,90620, 8,13947)
62	8,02780	8,03723	0,0575522	(7,92051, 8,15395)

V.2. Prognoza przychodów na lata 2018-2023



Wnioski

...

Bibliografia

<https://obliczeniastatystyczne.pl/wspolczynnik-zmiennosci/>

Spis rysunków

Nie znaleziono żadnych pozycji spisu treści.

Spis tabel

Nie znaleziono żadnych pozycji spisu treści.

Załączniki