

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

| | |
|--|----|
| Wstęp | 4 |
| Rozdział I. Podstawy teoretyczne..... | 5 |
| I.1. Regresja liniowa | 5 |
| I.2. Linearyzacja modeli nieliniowych..... | 5 |
| I.3. Miary jakości modelu | 6 |
| Rozdział II. Opis problemu | 8 |
| II.1. Treść zadania | 8 |
| II.2. Opis problemu badawczego..... | 10 |
| Rozdział III. Postawione hipotezy..... | 11 |
| III.1. Hipoteza I..... | 11 |
| III.2. Hipoteza II..... | 11 |
| III.3. Hipoteza III | 12 |
| III.4. Hipoteza IV | 12 |
| Rozdział IV. Implementacja modeli..... | 14 |
| IV.1. Przygotowanie danych | 14 |
| IV.2. Modele dla Hipotezy I..... | 14 |
| IV.3. Model dla Hipotezy II | 15 |
| IV.4. Modele dla Hipotezy III | 16 |
| IV.5. Modele dla Hipotezy IV | 18 |
| Rozdział V. Wyniki eksperymentów..... | 19 |
| V.1. Wyniki dla Hipotezy I | 19 |
| V.2. Wyniki dla Hipotezy II..... | 23 |
| V.3. Wyniki dla Hipotezy III | 25 |
| V.4. Wyniki dla Hipotezy IV | 30 |
| Podsumowanie..... | 34 |
| Bibliografia | 35 |
| Spis tabel | 36 |
| Spis rysunków | 36 |
| Załączniki | 37 |

Wstęp

Celem niniejszego sprawozdania jest prezentacja wyników analizy regresji przeprowadzonej na danych firmy Meta (dawniej Facebook). W pracy zbadano zależności między liczbą użytkowników portalu społecznościowego Facebook a przychodami generowanymi przez firmę Meta w latach 2007-2023.

Analiza została przeprowadzona w oparciu o cztery główne hipotezy badawcze dotyczące charakteru wzrostu liczby użytkowników oraz przychodów firmy. Wykorzystano różne modele regresji, w tym modele liniowe oraz nieliniowe po przeprowadzeniu linearyzacji: wykładniczy, potęgowy, logistyczny i wielomianowy.

Dane treningowe obejmują lata 2007-2017 (dla przychodów) oraz 2009-2017 (dla liczby użytkowników w rozbiciu kwartalnym). Dane testowe dotyczą lat 2018-2023, co pozwoliło na weryfikację zdolności predykcyjnych opracowanych modeli. Sprawozdanie zawiera szczegółową analizę dopasowania modeli do danych treningowych oraz ocenę jakości prognoz dla okresu testowego z podziałem na lata 2018-2020 oraz 2021-2023.

Rozdział I. Podstawy teoretyczne

I.1. Regresja liniowa

Regresja liniowa jest podstawową metodą statystyczną służącą do modelowania zależności między zmienną objaśnianą Y a jedną lub wieloma zmiennymi objaśniającymi X. Model regresji liniowej dla jednej zmiennej objaśniającej można zapisać jako:

$$\hat{Y} = a_1 \cdot X + a_0 \quad (1)$$

Gdzie:

- \hat{Y} – wartość teoretyczna zmiennej objaśnianej
- X – zmienna objaśniająca
- a_1 – współczynnik regresji (nachylenie prostej)
- a_0 – wyraz wolny (przecięcie z osią Y)

Parametry strukturalne modelu a_0 i a_1 wyznacza się metodą najmniejszych kwadratów, minimalizując sumę kwadratów odchyleń wartości rzeczywistych od teoretycznych.

I.2. Linearyzacja modeli nieliniowych

Wiele rzeczywistych zjawisk nie może być adekwatnie opisanych prostym modelem liniowym. W takich przypadkach stosuje się modele nieliniowe, które następnie przekształca się do postaci liniowej.

I.2.1. Model wykładniczy

Postać nieliniowa:

$$\hat{W} = b \cdot a^X \quad (2)$$

Po obustronnym zlogarytmizowaniu:

$$\log \hat{W} = \log b + X \cdot \log a \quad (3)$$

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

$$\hat{Y} = a_1 \cdot X + a_0 \quad (4)$$

I.2.2. Model potęgowy

Postać nieliniowa:

$$\hat{W} = b \cdot X^a \quad (5)$$

Po obustronnym zlogarytmizowaniu:

$$\log \hat{W} = \log b + a \cdot X \quad (6)$$

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

$$\hat{Y} = a_1 \cdot X_1 + a_0 \quad (7)$$

I.2.3. Model logistyczny

Postać nieliniowa:

$$\hat{W} = \frac{k}{1 + e^{a_1 \cdot X + a_0}} \quad (8)$$

Gdzie k jest parametrem nasycenia (górną asymptotą).

Po przekształceniu:

$$\hat{Y} = \log \left(\frac{k - \hat{W}}{\hat{W}} \right) = a_1 \cdot X + a_0 \quad (9)$$

I.2.4. Model wielomianowy

Dla wielomianu stopnia n :

$$\hat{Y} = a_0 + a_1 \cdot X + a_2 \cdot X^2 + \dots + a_n \cdot X^n \quad (10)$$

Model ten linearyzuje się przez podstawienie $X^2 = X_2$, $X^3 = X_3$, itd.

I.3. Miary jakości modelu

Do oceny jakości dopasowania modelu oraz jego zdolności predykcyjnych wykorzystano często używane miary jakości.

I.3.1. Współczynnik determinacji R^2

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

Gdzie:

- n – liczba obserwacji
- y_i – wartość rzeczywista

- \hat{y}_l – wartość prognozowana przez model
- \bar{y} – wartość średnia

Miara dopasowania modelu do danych uczących. Wartość współczynnik przyjmuje z przedziału $[0; 1]$ jeśli w modelu występuje wyraz wolny.

I.3.2. Średni błąd bezwzględny (MAE – *Mean Absolute Error*)

$$MAE = \frac{1}{n} \cdot \Sigma |y_i - \hat{y}_i| \quad (12)$$

Gdzie:

- n – liczba obserwacji
- y_i – wartość rzeczywista
- \hat{y}_i – wartość prognozowana przez model

MAE wyraża średnią wielkość błędu predykcji w jednostkach zmiennej objaśnianej.

Rozdział II. Opis problemu

II.1. Treść zadania

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 liczony w milionach dolarów amerykańskich.

Tab. 1. Liczba użytkowników portalu społecznościowego „Facebook”

| Kwartał | Liczba użytkowników w mln | Kwartał | Liczba użytkowników w mln | Kwartał | Liczba użytkowników w mln |
|---------|---------------------------|---------|---------------------------|---------|---------------------------|
| Q3 '08 | 100 | Q1 '09 | 197 | Q1 '14 | 1276 |
| | | Q2 '09 | 242 | Q2 '14 | 1317 |
| | | Q3 '09 | 305 | Q3 '14 | 1350 |
| | | Q4 '09 | 360 | Q4 '14 | 1393 |
| | | Q1 '10 | 431 | Q1 '15 | 1441 |
| | | Q2 '10 | 482 | Q2 '15 | 1490 |
| | | Q3 '10 | 550 | Q3 '15 | 1545 |
| | | Q4 '10 | 608 | Q4 '15 | 1591 |
| | | Q1 '11 | 680 | Q1 '16 | 1654 |
| | | Q2 '11 | 739 | Q2 '16 | 1712 |
| | | Q3 '11 | 800 | Q3 '16 | 1788 |
| | | Q4 '11 | 845 | Q4 '16 | 1860 |
| | | Q1 '12 | 901 | Q1 '17 | 1936 |
| | | Q2 '12 | 955 | Q2 '17 | 2006 |
| | | Q3 '12 | 1007 | Q3 '17 | 2072 |
| | | Q4 '12 | 1059 | Q4 '17 | 2129 |
| | | Q1 '13 | 1110 | | |
| | | Q2 '13 | 1155 | | |
| | | Q3 '13 | 1189 | | |
| | | Q4 '13 | 1228 | | |

Tab. 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 |

W analizie proszę:

- 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) Proponowane 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 zawrzeć w postaci sprawozdania. 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.

II.2. Opis problemu badawczego

Problem badawczy koncentruje się na zrozumieniu dynamiki wzrostu firmy Meta poprzez analizę dwóch kluczowych wskaźników:

- **Liczba użytkowników** – podstawowa metryka odzwierciedlająca zasięg i popularność platformy. Zrozumienie charakteru wzrostu bazy użytkowników (liniowy, wykładniczy, logistyczny) ma kluczowe znaczenie dla przewidywania przyszłego rozwoju platformy.
- **Przychody firmy** – wskaźnik efektywności biznesowej, pokazujący zdolność monetyzacji bazy użytkowników. Analiza relacji między liczbą użytkowników a przychodami pozwala ocenić, czy firma zwiększa efektywność generowania przychodów na użytkownika.

Kluczowe pytania badawcze obejmują:

- Jaki model najlepiej opisuje wzrost liczby użytkowników w czasie?
- Czy istnieje punkt nasycenia dla liczby użytkowników?
- Jaki model najlepiej przewiduje przyszłe przychody firmy?
- Jaka jest relacja między liczbą użytkowników a generowanymi przychodami?

Rozdział III. Postawione hipotezy

III.1. Hipoteza I

III.1.1. Treść hipotezy

Wzrost liczby użytkowników portalu Facebook w czasie jest wykładniczy lub potęgowy.

III.1.2. Pytania badawcze

1. Czy model wykładniczy lub potęgowy prowadzi do lepszej prognozy w latach 2018-2023 niż model liniowy?

III.1.3. Uzasadnienie

Wzrost platformy społecznościowej charakteryzuje się szybkim wzrostem bazy użytkowników spowodowanej polecaniem platformy przez użytkowników swoim znajomym. Taki wzrost jest nieliniowy i może być najlepiej opisywany przez model wykładniczy lub potęgowy.

III.1.4. Dane wykorzystane w analizie

- Zmienna objaśniana: liczba użytkowników (*Users*) w milionach
- Zmienna objaśniająca: numer kwartału (*NoQ*)
- Dane treningowe: Q1 2009 - Q4 2017 (36 kwartałów)
- Dane testowe: Q1 2018 - Q4 2023 (24 kwartały)

III.2. Hipoteza II

III.2.1. Treść hipotezy

Wzrost liczby użytkowników portalu Facebook nie jest nieograniczony i dąży do określonego punktu nasycenia, co sprawia, że model logistyczny będzie charakteryzował się najlepszym dopasowaniem.

III.2.2. Pytania badawcze

1. Jaki jest teoretyczny pułap liczby użytkowników portalu i kiedy tempo przyrostu nowych klientów zacznie/zaczęło trwale spadać?
2. Czy model logistyczny prowadzi do lepszej prognozy w latach 2018-2023 niż model liniowy?

III.2.3. Uzasadnienie

Facebook jako globalna platforma jest naturalnie ograniczony liczbą ludności z dostępem do internetu, konkurencją oraz efektem nasycenia rynku. Model logistyczny, charakteryzujący się punktem przegięcia oraz górną asymptotą, może lepiej odzwierciedlać te ograniczenia niż modele zakładające nieograniczony wzrost.

III.2.4. Dane wykorzystane w analizie

- Zmienna objaśniana: liczba użytkowników (*Users*) w milionach
- Zmienna objaśniająca: numer kwartału (*NoQ*)
- Dane treningowe: Q1 2009 - Q4 2017 (36 kwartałów)
- Dane testowe: Q1 2018 - Q4 2023 (24 kwartały)

III.3. Hipoteza III

III.3.1. Treść hipotezy

Przychody firmy Meta rosną w tempie wykładniczym względem upływu lat.

III.3.2. Pytania badawcze

1. Czy model oparty o funkcję wykładniczą daje lepsze wyniki prognozy dla lat 2018-2023 niż model oparty o funkcję potęgową lub wielomianową?
2. Czy przychody firmy rosną szybciej niż baza jej użytkowników?

III.3.3. Uzasadnienie

Na podstawie wyników finansowych firmy z lat 2007 – 2017 widać bardzo szybki wzrost przychodów, który najlepiej może tłumaczyć model wykładniczy charakteryzujący się nagłym przyśpieszeniem wzrostu.

III.3.4. Dane wykorzystane w analizie

- Zmienna objaśniana: przychody firmy Meta (*Revenue*) w milionach dolarów
- Zmienna objaśniająca: numer roku (*NoY*)
- Dane treningowe: 2007-2017 (11 lat)
- Dane testowe: 2018-2023 (6 lat)

III.4. Hipoteza IV

III.4.1. Treść hipotezy

Istnieje silna, nieliniowa zależność między liczbą użytkowników a generowanym przychodem, co sugeruje rosnącą efektywność monetyzacji pojedynczego użytkownika.

III.4.2. Pytania badawcze

1. Ile wzrośnie przychód, jeśli baza użytkowników wzrośnie o 5%?
2. Który model najlepiej opisuje relację przychodu z liczbą użytkowników?

III.4.3. Uzasadnienie

Jeśli efektywność monetyzacji użytkowników rośnie w czasie, przychody powinny rosnąć szybciej niż liniowo względem liczby użytkowników. Nieliniowa zależność może wynikać z efektów skali, lepszej personalizacji reklam oraz rosnącej wartości danych generowanych przez użytkowników.

III.4.4. Dane wykorzystane w analizie:

- Zmienna objaśniana: przychody firmy Meta (*Revenue*) w milionach dolarów
- Zmienna wyjaśniająca: średnia roczna liczba użytkowników (*Users*) w milionach
- Dane treningowe: 2009-2017 (9 lat)
- Dane testowe: 2018-2023 (6 lat)

Rozdział IV. Implementacja modeli

IV.1. Przygotowanie danych

Implementację przeprowadzono w języku Python z wykorzystaniem następujących bibliotek:

- **pandas** – zarządzanie danymi tabelarycznymi
- **numpy** – operacje matematyczne
- **statsmodels** – modelowanie statystyczne i regresja
- **matplotlib** – wizualizacje wyników
- **sklearn** – metryki oceny modeli
- **scipy** – optymalizacja parametrów

Dane zostały zorganizowane w strukturze *DataFrame* biblioteki *pandas*. Dla zarządzania danymi kwartalnymi zaimplementowano klasę *Quarter*, umożliwiającą porównywanie i sortowanie okresów kwartalnych.

Podział danych:

- Dane użytkowników: 36 kwartałów treningowych (2009-2017), 24 kwartały testowe (2018-2023)
- Dane przychodów: 11 lat treningowych (2007-2017), 6 lat testowych (2018-2023)

IV.2. Modele dla Hipotezy I

W ramach weryfikacji Hipotezy I zaimplementowano cztery modele.

IV.2.1. Model liniowy (referencyjny)

Model liniowy służy jako punkt odniesienia dla rozpatrywanych modeli nieliniowych.

Model wyrażony jest jako równanie (1):

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

, gdzie:

- \hat{Y} – liczba użytkowników w milionach (*Users*),
- X – numer kwartału (*NoQ*).

IV.2.2. Model wykładniczy

Zlogarytmizowana postać modelu (5) :

$$\log \hat{Y} = a_1 \cdot X + a_0$$

, gdzie:

- \hat{Y} – liczba użytkowników w milionach (*Users*),
- X – numer kwartału (*NoQ*).

IV.2.3. Model potęgowy

Zlogarytmizowana postać modelu (2):

$$\log \hat{Y} = a_1 \cdot \log X + a_0$$

, gdzie:

- \hat{Y} – liczba użytkowników w milionach (*Users*),
- X – numer kwartału (*NoQ*).

IV.2.4. Model hybrydowy

Po opracowaniu modelu wykładowicznego i modelu potęgowego zauważono, że poprawny model znajduje się pomiędzy wynikami powyższych modeli, z tego powodu opracowano model hybrydowy, będący kombinacją liniową tych modeli.

$$\hat{Y}_{hybrydowy} = \alpha \cdot \hat{Y}_{wykładowiczny} + (1 - \alpha) \cdot \hat{Y}_{potęgowy} \quad (13)$$

, gdzie parametr α , wybrany zostaje poprzez rozwiązywanie zadania optymalizacyjnego:

$$\alpha^* = \arg \min_{\alpha \in [0;1]} MAE(\alpha) \quad (14)$$

gdzie MAE jest funkcją zwracającą średni błąd bezwzględny prognozy modelu hybrydowego dla danych z lat 2018-2023, opisana równaniem (12).

IV.3. Model dla Hipotezy II

Do weryfikacji Hipotezy II opracowano model logistyczny, wyniki modelu porównano z rezultatami uzyskanymi podczas weryfikacji Hipotezy I.

IV.3.1. Model logistyczny

Model logistyczny przedstawia równanie (8):

$$\hat{W} = \frac{k}{1 + e^{a_1 \cdot X + a_0}}$$

, gdzie:

- X – numer kwartału (*NoQ*),
- k – parametr nasycenia (maksymalna teoretyczna liczba użytkowników),

Po przekształceniach otrzymany równoważny model liniowy (9):

$$\hat{Y} = \log \left(\frac{k - \hat{W}}{\hat{W}} \right) = a_1 \cdot X + a_0$$

, gdzie:

- \hat{Y} – liczba użytkowników w milionach (*Users*).

Parametr k został wyznaczony przez rozwiązywanie zadania optymalizacyjnego:

$$k^* = \arg \min_{k \in [3000, 8000]} (MAE_{train}(k) + MAE_{test}(k)) \quad (15)$$

, gdzie:

- MAE_{train} – funkcja zwracająca średni błąd bezwzględny prognozy modelu dla danych z lat 2009-2017 (dane treningowe),
- MAE_{test} – funkcja zwracająca średni błąd bezwzględny prognozy modelu dla danych z lat 2018-2023 (dane testowe).

Punkt przegięcia funkcji logistycznej, w którym tempo wzrostu osiąga maksimum i od którego zaczyna spadać, występuje dla $\hat{W} = \frac{k}{2}$.

IV.4. Modele dla Hipotezy III

Do weryfikacji Hipotezy III opracowano sześć modeli nieliniowych.

IV.4.1. Model wykładniczy

Zlogarytmizowana postać modelu (5) :

$$\log \hat{Y} = a_1 \cdot X + a_0$$

, gdzie:

- \hat{Y} – przychody firmy *Meta* w milionach dolarów (*Revenue*),
- X – numer roku (*No Y*).

IV.4.2. Model potęgowy

Zlogarytmizowana postać modelu (2):

$$\log \hat{Y} = a_1 \cdot \log X + a_0$$

, gdzie:

- \hat{Y} – przychody firmy *Meta* w milionach dolarów (*Revenue*),
- X – numer roku (*No Y*).

IV.4.3. Model hybrydowy

Ponownie zaobserwowano zależność, w której model poprawnie opisujący zależność przychodów od czasu leży pomiędzy modelem wykładniczym a potęgowym, stąd opracowano model hybrydowy będący kombinacją liniową obu modeli (13):

$$\hat{Y}_{hybrydowy} = \alpha \cdot \hat{Y}_{wykładniczy} + (1 - \alpha) \cdot \hat{Y}_{potęgowy}$$

, gdzie parametr α , wybrany zostaje poprzez rozwiązywanie zadania optymalizacyjnego:

$$\alpha^* = \arg \min_{\alpha \in [0;1]} (MAE_{train}(\alpha) + MAE_{test}(\alpha)) \quad (16)$$

, gdzie:

- MAE_{train} – funkcja zwracająca średni błąd bezwzględny prognozy modelu dla danych z lat 2007-2017 (dane treningowe),

- MAE_{test} – funkcja zwracająca średni błąd bezwzględny prognozy modelu dla danych z lat 2018-2023 (dane testowe).

IV.4.4. Model wielomianowy 3. stopnia

Model wielomianowy opracowany na podstawie równania (10):

$$\hat{Y} = a_2 \cdot X_2 + a_3 \cdot X_3$$

, gdzie

- \hat{Y} – przychody firmy *Meta* w milionach dolarów (*Revenue*),
- $X_2 = X^2$,
- $X_3 = X^3$,
- X – numer roku (*NoY*).

Po eksperymentach ustalono, że wyraz wolny oraz wyraz X_1 nie poprawiają istotnie dopasowania modelu, stąd zastosowano model bez tych składowych.

IV.4.5. Model logistyczny

Dla sprawdzenia hipotezy o możliwym nasyceniu przychodów opracowano model logistyczny (8):

$$\hat{W} = \frac{k}{1 + e^{a_1 \cdot X + a_0}}$$

, gdzie:

- X – numer roku (*NoY*),
- k – parametr nasycenia (maksymalny teoretyczny przychód),

Po przekształceniach otrzymany równoważny model liniowy (9):

$$\hat{Y} = \log\left(\frac{k - \hat{W}}{\hat{W}}\right) = a_1 \cdot X + a_0$$

, gdzie:

- \hat{Y} – przychody firmy *Meta* w milionach dolarów (*Revenue*).

Parametr k został wyznaczony przez rozwiązywanie zadania optymalizacyjnego:

$$k^* = \arg \min_{k \in [140000, 600000]} (MAE_{train}(k) + MAE_{test}(k)) \quad (17)$$

, gdzie:

- MAE_{train} – funkcja zwracająca średni błąd bezwzględny prognozy modelu dla danych z lat 2008-2017 (dane treningowe),
- MAE_{test} – funkcja zwracająca średni błąd bezwzględny prognozy modelu dla danych z lat 2018-2023 (dane testowe).

Punkt przegięcia funkcji logistycznej, w którym tempo wzrostu osiąga maksimum i od którego zaczyna spadać, występuje dla $\hat{W} = \frac{k}{2}$.

IV.5. Modele dla Hipotezy IV

Analiza relacji użytkownicy-przychody wymagała przekształcenia danych kwartalnych do postaci rocznej. Zastosowano agregację przez średnią arytmetyczną liczby użytkowników w kwartałach danego roku, co zachowuje informację o możliwych fluktuacjach w bazie użytkowników na przestrzeni roku.

Do weryfikacji Hipotezy IV opracowano cztery modele nieliniowe.

IV.5.1. Model wykładniczy

Zlogarytmizowana postać modelu (5) :

$$\log \hat{Y} = a_1 \cdot X + a_0$$

, gdzie:

- \hat{Y} – przychody firmy *Meta* w milionach dolarów (*Revenue*),
- X – liczba użytkowników w milionach (*Users*).

IV.5.2. Model potęgowy

Zlogarytmizowana postać modelu (2):

$$\log \hat{Y} = a_1 \cdot \log X + a_0$$

, gdzie:

- \hat{Y} – przychody firmy *Meta* w milionach dolarów (*Revenue*),
- X – liczba użytkowników w milionach (*Users*).

IV.5.3. Model hybrydowy

Ponownie zaobserwowano zależność, w której model poprawnie opisujący zależność przychodów od czasu leży pomiędzy modelem wykładniczym a potęgowym, stąd opracowano model hybrydowy będący kombinacją liniową obu modeli (13):

$$\hat{Y}_{hybrydowy} = \alpha \cdot \hat{Y}_{wykładniczy} + (1 - \alpha) \cdot \hat{Y}_{potęgowy}$$

, gdzie parametr α , wybrany zostaje poprzez rozwiązywanie zadania optymalizacyjnego (14):

$$\alpha^* = \arg \min_{\alpha \in [0;1]} MAE(\alpha)$$

IV.5.4. Model wielomianowy 3. stopnia

Model wielomianowy opracowany na podstawie równania (10):

$$\hat{Y} = a_3 \cdot X_3 + a_0$$

, gdzie

- \hat{Y} – przychody firmy *Meta* w milionach dolarów (*Revenue*),
- $X_3 = X^3$,
- X – liczba użytkowników w milionach (*Users*).

Po eksperymentach ustalono, wyrazy X_1 oraz X_2 nie poprawiają istotnie dopasowania modelu, stąd zastosowano model bez tych składowych.

Rozdział V. Wyniki eksperymentów

V.1. Wyniki dla Hipotezy I

V.1.1. Model liniowy

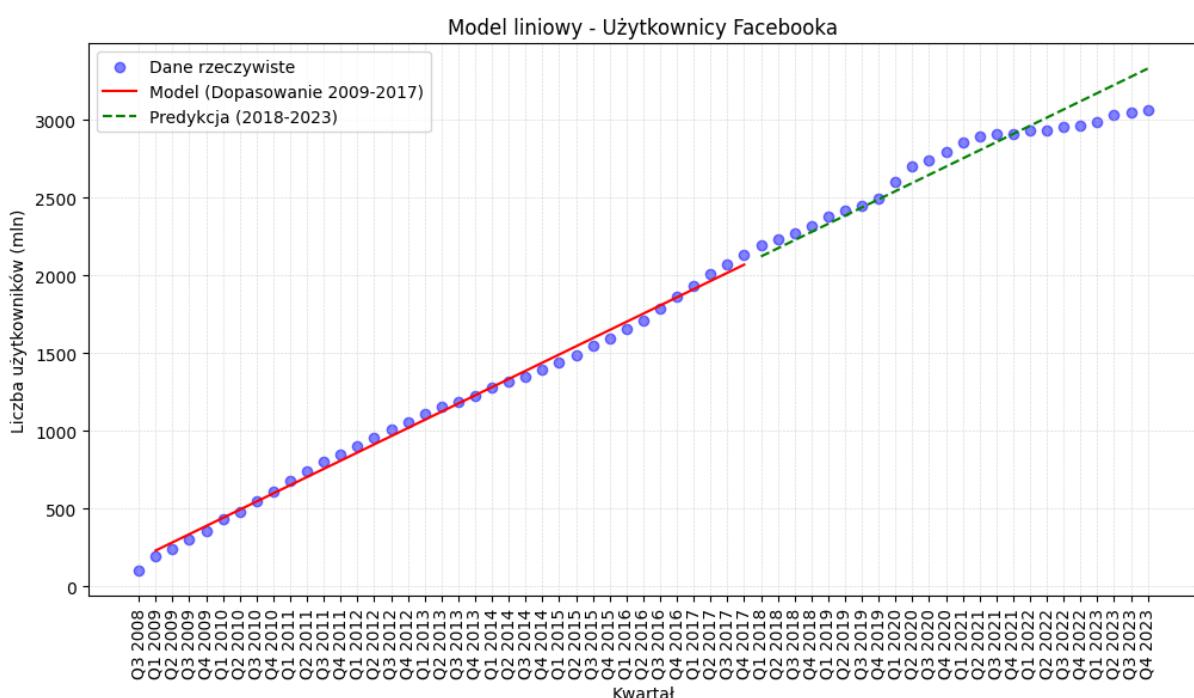
Parametry modelu:

- $a_0 = 125.8 \pm 13.3$,
- $a_1 = 52.5 \pm 0.6$,
- $R^2 = 99.6\%$.

Średni błąd bezwzględny prognozy:

- dla lat 2018-2020: 55.95 mln użytkowników,
- dla lat 2021-2023: 124.17 mln użytkowników.

Wyniki modelu na tle rzeczywistych danych przedstawia Rys. 1.



Rys. 1. Wizualizacja modelu liniowego na tle rzeczywistych danych liczby użytkowników

Model liniowy wykazuje bardzo dobre dopasowanie do danych treningowych ($R^2 = 99.6\%$), jednak jakość prognozy znaczco spada dla okresu 2021-2023, co sugeruje że liniowy charakter wzrostu nie utrzymuje się w dłuższej perspektywie.

V.1.2. Model wykładniczy

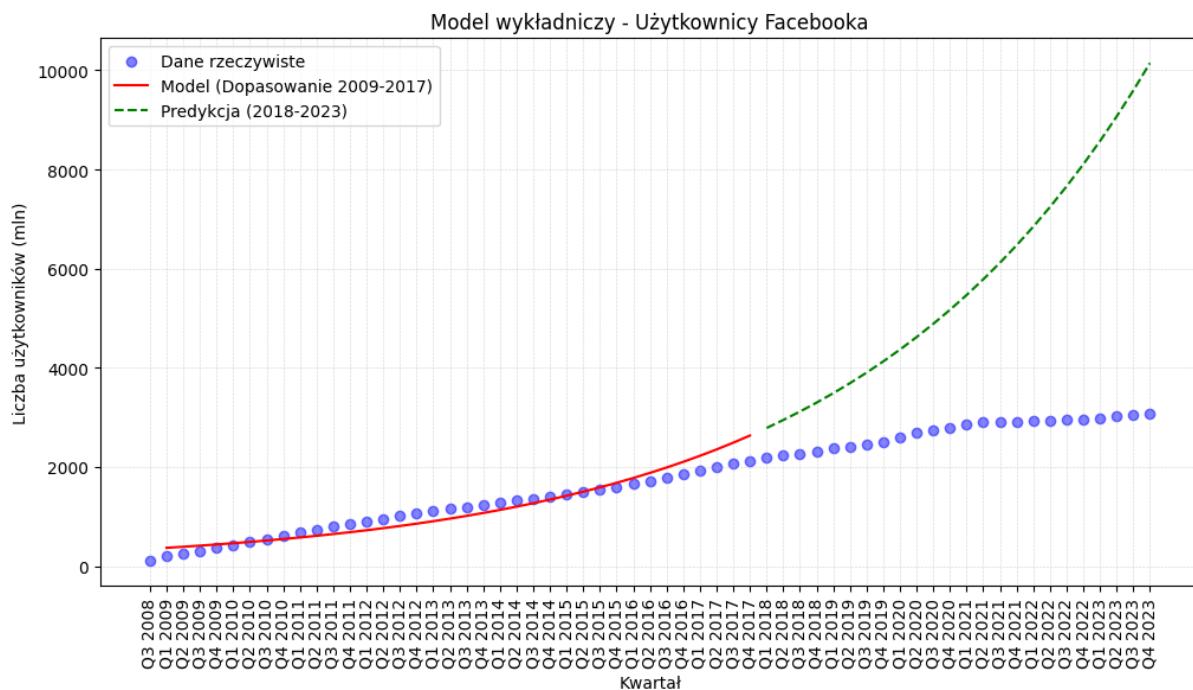
Parametry modelu (zlinearyzowanego):

- $a_0 = 5.796 \pm 0.073$,
- $a_1 = 0.056 \pm 0.003$,
- $R^2 = 89.6\%$.

Średni błąd bezwzględny prognozy:

- dla lat 2018-2020: 1399.2 mln użytkowników,
- dla lat 2021-2023: 4630.9 mln użytkowników.

Wyniki modelu na tle rzeczywistych danych przedstawia Rys. 2.



Rys. 2. Wizualizacja modelu wykładniczego na tle rzeczywistych danych liczby użytkowników

Model wykładniczy wykazuje gorsze dopasowanie do danych treningowych ($R^2 = 89.6\%$), niż model liniowy oraz wykazuje o wiele gorszą jakość prognozy dla okresu 2018-2023. Model wykładniczy nie opisuje zjawiska lepiej niż model liniowy.

V.1.3. Model potęgowy

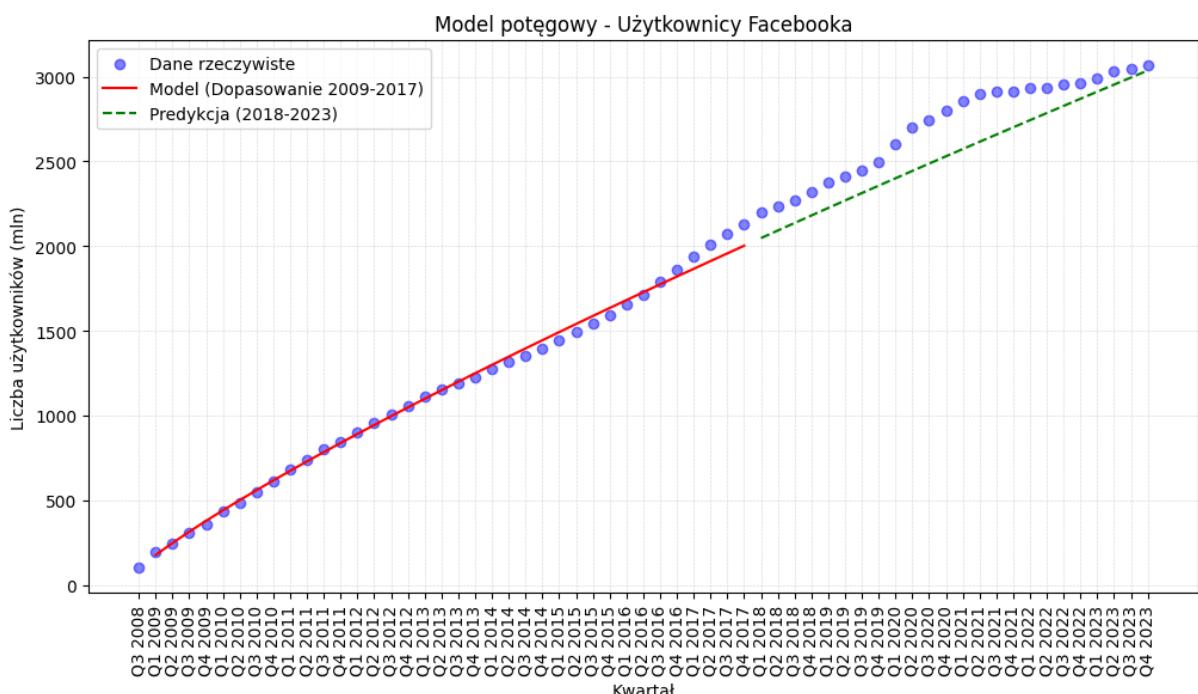
Parametry modelu (zlinearyzowanego):

- $a_0 = 4.593 \pm 0.022$,
- $a_1 = 0.833 \pm 0.008$,
- $R^2 = 99.7\%$.

Średni błąd bezwzględny prognozy:

- dla lat 2018-2020: 176.9 mln użytkowników,
- dla lat 2021-2023: 152.5 mln użytkowników.

Wyniki modelu na tle rzeczywistych danych przedstawia Rys. 3.



Rys. 3. Wizualizacja modelu potęgowego na tle rzeczywistych danych liczby użytkowników

Model liniowy wykazuje bardzo dobre dopasowanie do danych treningowych ($R^2 = 99.7\%$), porównywalne z modelem liniowym, jednak jakość prognozy w okresie 2018-2023 spada względem modelu liniowego. Model potęgowy nie opisuje zjawiska znaczająco lepiej niż model liniowy.

V.1.4. Model hybrydowy

Model hybrydowy wykorzystuje modele potęgowy i wykładniczy opisane w poprzednich punktach wykorzystując je w kombinacji liniowej z parametrem α (Równanie (13)).

Wyznaczona wartość parametru $\alpha = 0.07$.

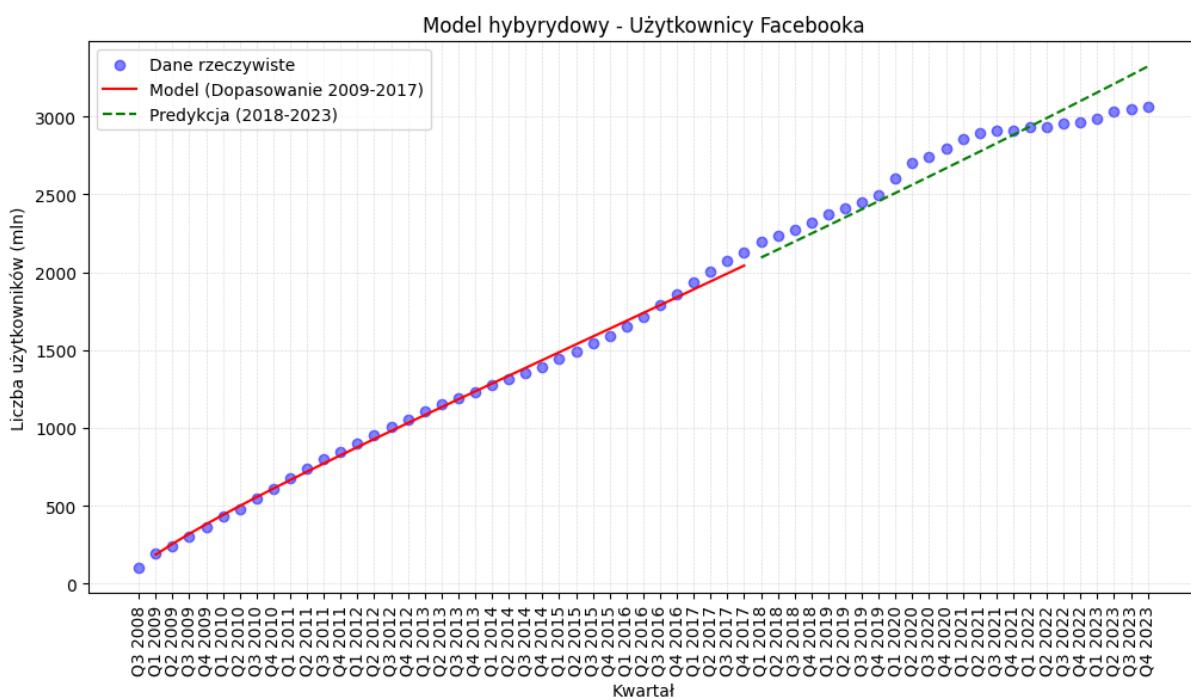
Niska wartość parametru oznacza niski udział modelu wykładniczego w modelu hybrydowym, co potwierdzają znacznie gorsze wyniki modelu wykładniczego od modelu potęgowego.

Parametr $R^2 = 99.6\%$.

Średni błąd bezwzględny prognozy:

- dla lat 2018-2020: 87.2 mln użytkowników,
- dla lat 2021-2023: 121.9 mln użytkowników.

Wyniki modelu na tle rzeczywistych danych przedstawia Rys. 4.



Rys. 4. Wizualizacja modelu hybrydowego na tle rzeczywistych danych liczby użytkowników

Model hybrydowy bardzo dobrze dopasowuje się do danych treningowych ($R^2 = 99.6\%$), porównywalnie z modelem liniowym. Jakość prognozy dla lat 2018-2020 jest lekko gorsza od prognozy modelu liniowego, stąd nie ma znaczących powodów, aby opisywać zjawisko modelem hybrydowym, zamiast tego według dokonanych doświadczeń zjawisko najlepiej opisuje model liniowy.

V.1.5. Wnioski

Analizowane modele nieliniowe nie opisują zjawiska znaczco lepiej od modelu liniowego, zarówno pod kątem dopasowania do danych treningowych jak i jakości prognozy dla okresu 2018-2023.

Wszystkie modele wykazują znaczy wzrost błędu MAE dla okresu 2021-2023, co sugeruje zmianę charakteru zjawiska w tym okresie i popiera hipotezę o nieliniowym charakterze przyrostu liczby użytkowników w czasie, jednak poddane analizie modele nie tłumaczą zachowania zjawiska.

V.2. Wyniki dla Hipotezy II

V.2.1. Model logistyczny

W modelu logistycznym występuje parametr k , będący wartością nasycenia, którą wyznaczono poprzez rozwiązywanie zadania optymalizacyjnego (Równanie (15)), otrzymując:

$$k = 3277 \text{ mln.}$$

Parametry modelu (zlinearyzowanego):

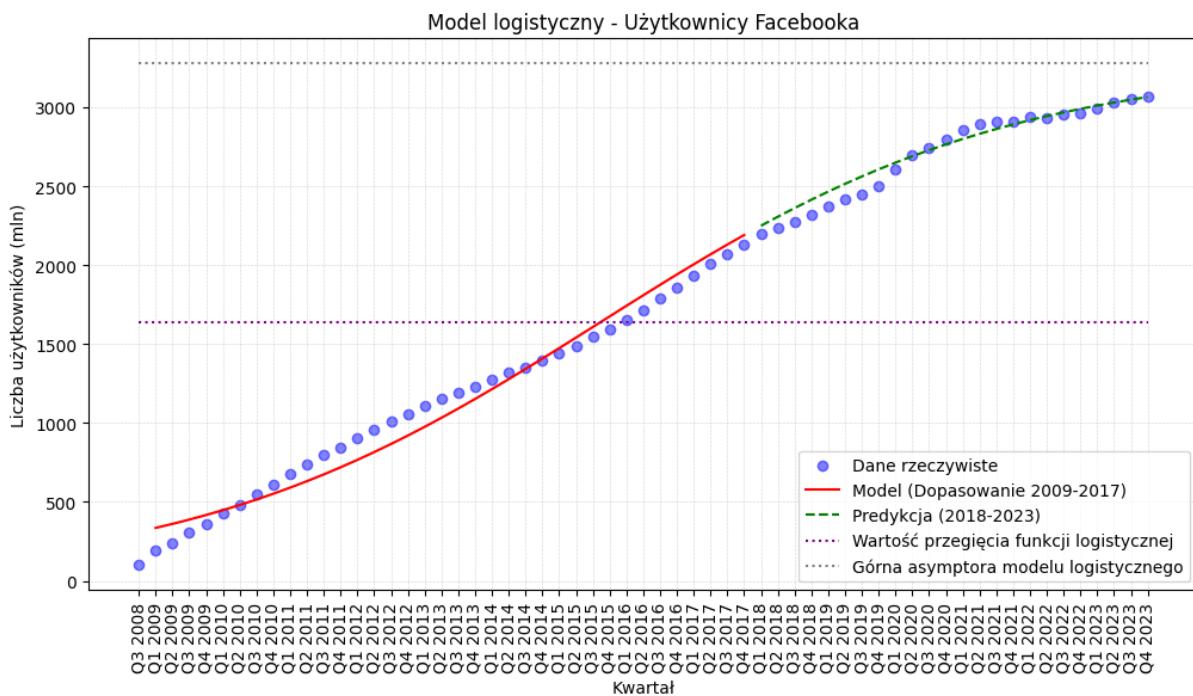
- $a_0 = 2.335 \pm 0.066$,
- $a_1 = -0.0821 \pm 0.003$,
- $R^2 = 95.7\%$.

Średni błąd bezwzględny prognozy:

- dla lat 2018-2020: 69.7 mln użytkowników,
- dla lat 2021-2023: 22.6 mln użytkowników.

Na podstawie wartości parametru k , wyznaczono punkt przegięcia modelu, który znajduje się pomiędzy trzecim, a czwartym kwartałem 2015 roku. Od tego momentu tempo wzrostu liczby użytkowników według modelu zaczyna stale zwalniać.

Wyniki modelu na tle rzeczywistych danych przedstawia Rys. 5.



Rys. 5. Wizualizacja modelu logistycznego na tle rzeczywistych danych liczby użytkowników

Model logistyczny wykazuje bardzo dobre dopasowanie do danych treningowych ($R^2 = 95.7\%$), które jest nieco gorsze, lecz wciąż porównywalne do modelu liniowego. Natomiast znaczająco poprawia się jakość prognozy dla okresu 2021-2023 (poprawa o ponad 80%).

V.2.2. Wnioski

Model logistyczny znacznie lepiej prognozuje liczbę użytkowników w okresie 2021-2023, nie tracąc znaczco na dopasowaniu do danych treningowych i jakości prognozy w okresie 2018-2020, co sugeruje, że zjawisko jest dobrze opisywane przez model logistyczny.

Opracowany model logistyczny pozwolił na określenie maksymalnej teoretycznej liczby użytkowników (punktu nasycenia rynku), który został wyznaczony na liczbę 3277 mln. Dodatkowo określono moment od którego tempo wzrostu liczby użytkowników zaczęło stale spadać, punkt ten znajduje się pomiędzy 3. a 4. Kwartałem 2015 roku.

V.3. Wyniki dla Hipotezy III

V.3.1. Model wykładniczy

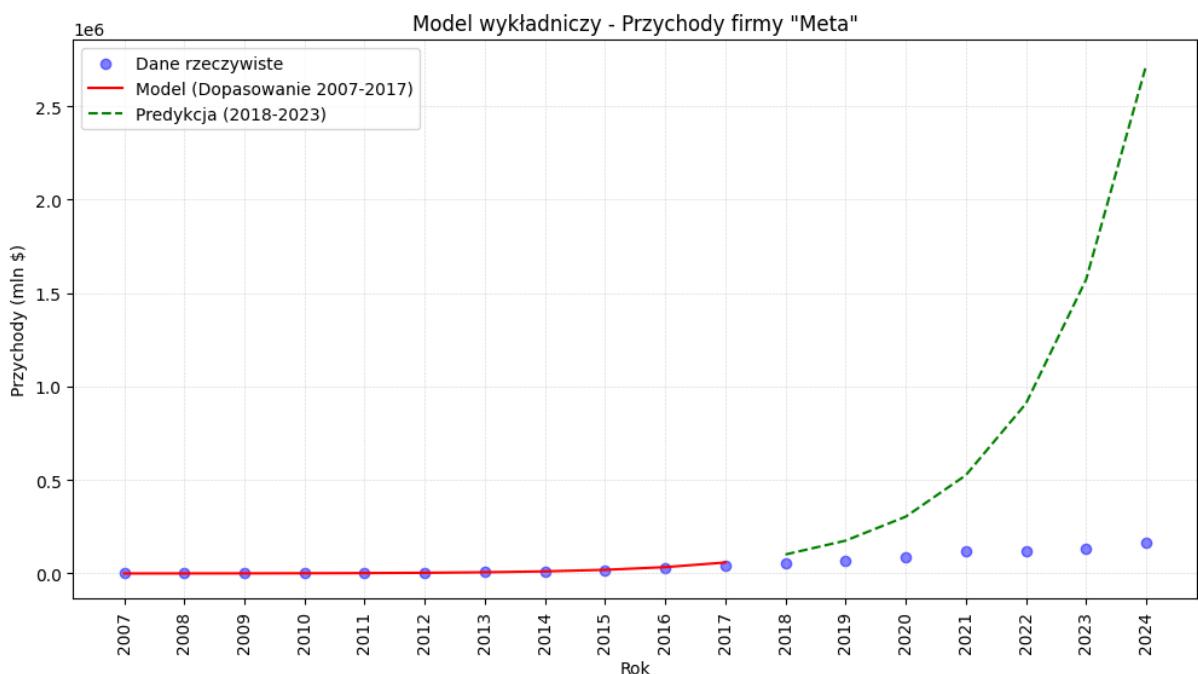
Parametry modelu (zlinearyzowanego):

- $a_0 = 4.96 \pm 0.24$,
- $a_1 = 0.548 \pm 0.035$,
- $R^2 = 96.5\%$.

Średni błąd bezwzględny prognozy:

- dla lat 2018-2020: 122926.2 mln dolarów,
- dla lat 2021-2023: 1297570.9 mln dolarów.

Wyniki modelu na tle rzeczywistych danych przedstawia Rys. 6.



Rys. 6. Wizualizacja modelu wykładniczego na tle rzeczywistych danych przychodów firmy *Meta*

Model wykładniczy wykazuje bardzo dobре dopasowanie do danych treningowych ($R^2 = 96.5\%$), jednak jakość prognozy dla danych testowych znaczco spada, co sugeruje że pomimo szybkiego wzrostu przychodów nie rosą one wykładniczo bez ograniczenia.

V.3.2. Model potęgowy

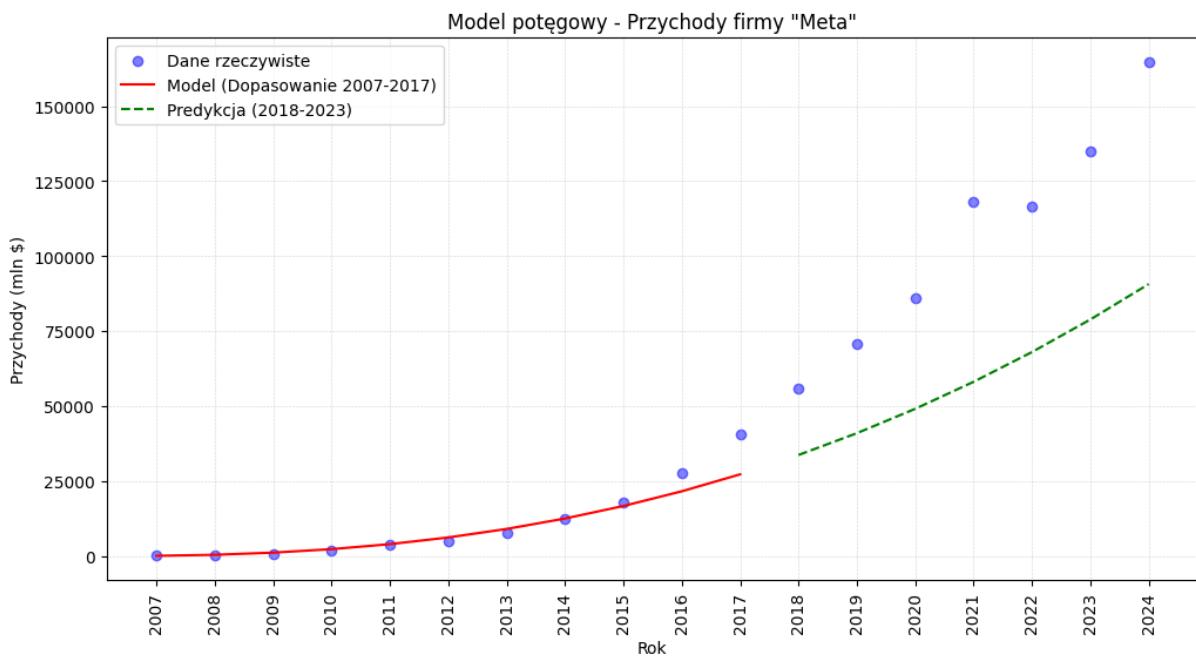
Parametry modelu (zlinearyzowanego):

- $a_0 = 4.36 \pm 0.26$,
- $a_1 = 2.44 \pm 0.15$,
- $R^2 = 96.8\%$.

Średni błąd bezwzględny prognozy:

- dla lat 2018-2020: 29586.9 mln dolarów,
- dla lat 2021-2023: 59578.0 mln dolarów.

Wyniki modelu na tle rzeczywistych danych przedstawia Rys. 7.



Rys. 7. Wizualizacja modelu potęgowego na tle rzeczywistych danych przychodów firmy Meta

Model potęgowy wykazuje bardzo dobre dopasowanie do danych treningowych ($R^2 = 96.8\%$), porównywalne do modelu wykładniczego. Jakość prognozy dla danych testowych jest widocznie lepsza od modelu wykładniczego, lecz na Rys. 7 widać znaczące niedoszacowanie wartości przychodów, co sugeruje że zjawisko nie rośnie potęgowo.

V.3.3. Model hybrydowy

Model hybrydowy wykorzystuje modele potęgowy i wykładniczy opisane w poprzednich punktach wykorzystując je w kombinacji liniowej z parametrem α (Równanie (13)).

Wyznaczona wartość parametru $\alpha = 0.18$.

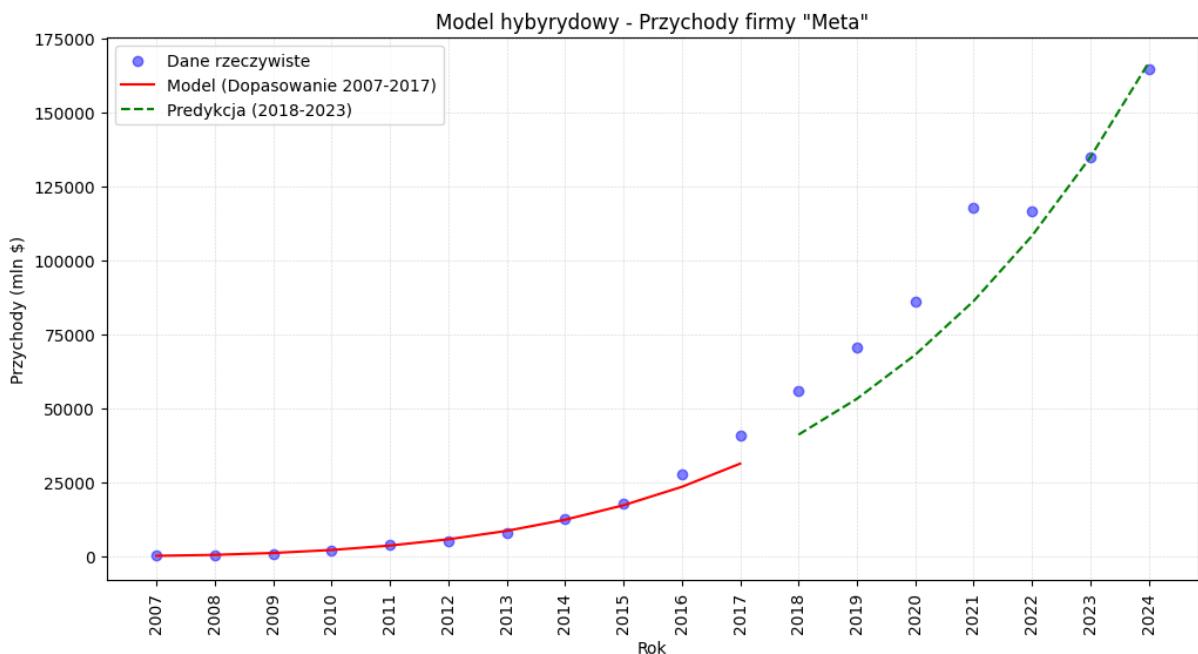
Niska wartość parametru oznacza niski udział modelu wykładniczego w modelu hybrydowym, co potwierdzają znacznie gorsze wyniki modelu wykładniczego od modelu potęgowego w prognozie.

Parametr $R^2 = 93.8\%$.

Średni błąd bezwzględny prognozy:

- dla lat 2018-2020: 16717.3 mln dolarów,
- dla lat 2021-2023: 10598.0 mln dolarów.

Wyniki modelu na tle rzeczywistych danych przedstawia Rys. 8.



Rys. 8. Wizualizacja modelu hybrydowego na tle rzeczywistych danych przychodów firmy *Meta*

Model hybrydowy znajdujący się pomiędzy modelem wykładniczym a potęgowym, o wiele lepiej prognozuje przychody w okresie 2021-2023, lecz widoczne jest spore niedoszacowanie w okresie 2018-2020. Model nieznacznie gorzej dopasowuje dane treningowe od modeli wchodzących w jego skład ($R^2 = 93.8\%$).

V.3.4. Model wielomianowy 3. stopnia

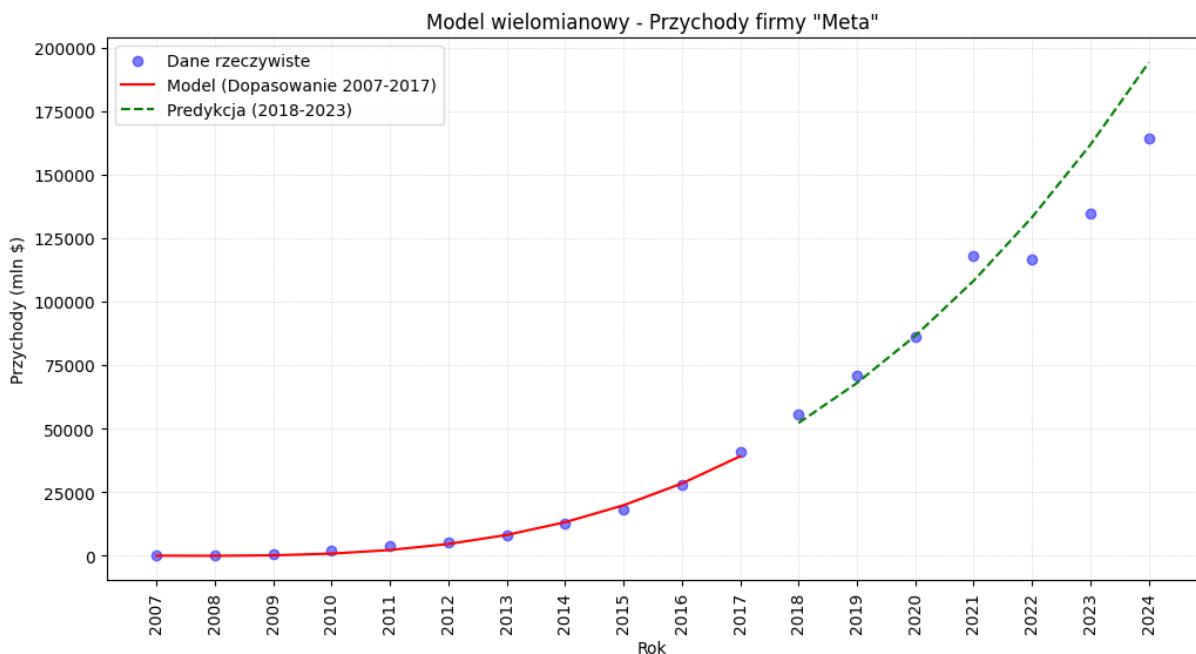
Parametry modelu (zlinearyzowanego):

- $a_2 = -111.1 \pm 34.7$,
- $a_3 = 39.48 \pm 3.59$,
- $R^2 = 99.6\%$.

Średni błąd bezwzględny prognozy:

- dla lat 2018-2020: 2315.4 mln dolarów,
- dla lat 2021-2023: 20752.6 mln dolarów.

Wyniki modelu na tle rzeczywistych danych przedstawia Rys. 9.



Rys. 9. Wizualizacja modelu wielomianowego 3. stopnia na tle rzeczywistych danych przychodów firmy *Meta*

Model wielomianowy 3. stopnia o wiele lepiej dopasowuje się do danych treningowych ($R^2 = 99.6\%$), co wiąże się z o wiele lepszą prognozą w okresie 2018-2020, lecz ulega znacznemu pogorszeniu prognoza w okresie 2021-2023 w porównaniu do modelu hybrydowego.

V.3.5. Model logistyczny

W modelu logistycznym występuje parametr k , będący wartością nasycenia, którą wyznaczono poprzez rozwiązywanie zadania optymalizacyjnego (Równanie(17)), otrzymując:

$$k = 140250 \text{ mln dolarów.}$$

Parametry modelu (zlinearyzowanego):

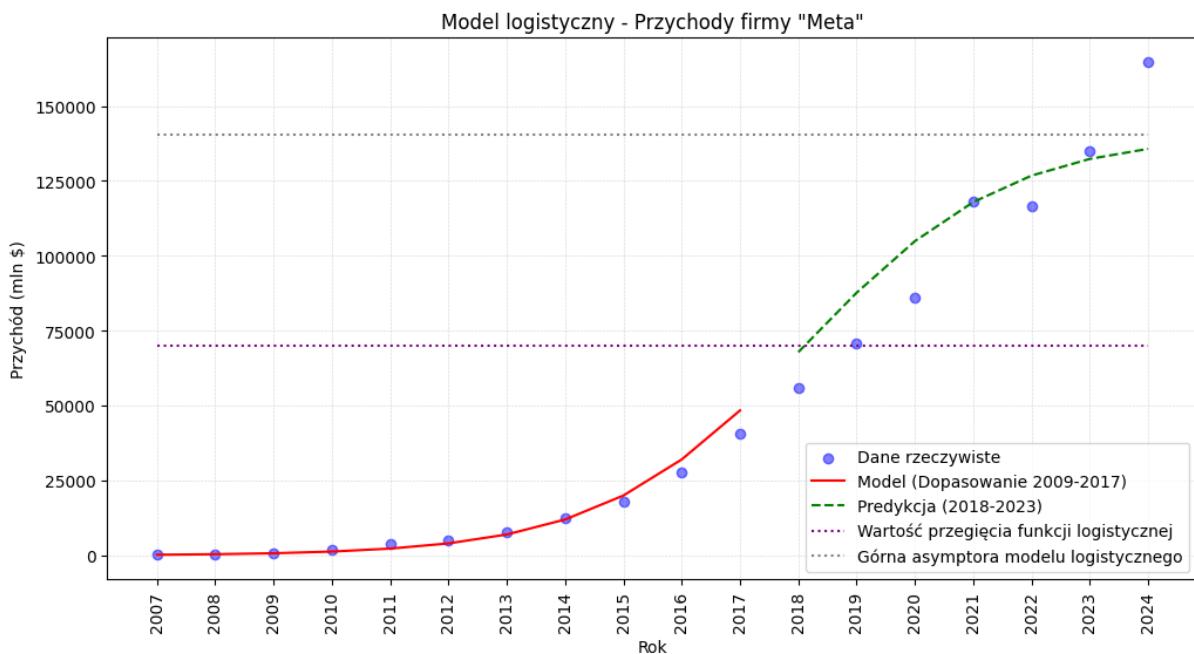
- $a_0 = 6.98 \pm 0.22$,
- $a_1 = -0.576 \pm 0.031$,
- $R^2 = 97.4\%$.

Średni błąd bezwzględny prognozy:

- dla lat 2018-2020: 15983.1 mln dolarów,
- dla lat 2021-2023: 10381.7 mln dolarów.

Na podstawie wartości parametru k , wyznaczono punkt przegięcia modelu, który znajduje się w pierwszym kwartale 2018 roku. Od tego momentu tempo wzrostu przychodów według modelu zaczyna stale zwalniać.

Wyniki modelu na tle rzeczywistych danych przedstawia Rys. 10.



Rys. 10. Wizualizacja modelu logistycznego na tle rzeczywistych danych przychodów firmy *Meta*

Model logistyczny bardzo dobrze dopasowuje się do danych treningowych ($R^2 = 97.4\%$), jakość prognozy dla okresu 2018-2023 jest nieznacznie lepsza od modelu hybrydowego. Model logistyczny pozwolił na określenie górnej granicy przychodów, która znajduje się pod ostatnim punktem danych rzeczywistych (z 2024 roku), co sugeruje, że charakterystyka zjawiska zmieniła się, bądź model logistyczny nie jest najlepszym modelem opisu przyrostu przychodów.

V.3.6. Wnioski

Wszystkie analizowane modele nie potrafią dobrze opisać przyrostu przychodów firmy *Meta*, najlepszym modelem pod kątem prognozy w okresie 2018-2020 okazał się model wielomianowy 3. stopnia, ale pod kątem całego zakresu danych treningowych najlepiej poradził sobie model logistyczny, który jednak nie tłumaczy wszystkich obserwacji.

Wyniki modelu logistycznego pokazują, że zjawisko wzrostu przychodów firmy *Meta* mogło zmienić swój charakter w ostatnich latach, co można uzasadnić np. rozwojem technologii generatywnych modeli sztucznej inteligencji, pozwalających na jeszcze lepszą monetyzację użytkowników, co przyczyniło się do osiągnięcia przychodu powyżej górnej asymptoty modelu logistycznego.

Porównując parametr a_1 modelu logistycznego dla przyrostu liczby użytkowników w czasie, a parametr a_1 modelu logistycznego dla przyrostu przychodów w czasie, można zauważać, że przychody firmy wzrastają znacznie szybciej od bazy jej użytkowników.

V.4. Wyniki dla Hipotezy IV

V.4.1. Model wykładniczy

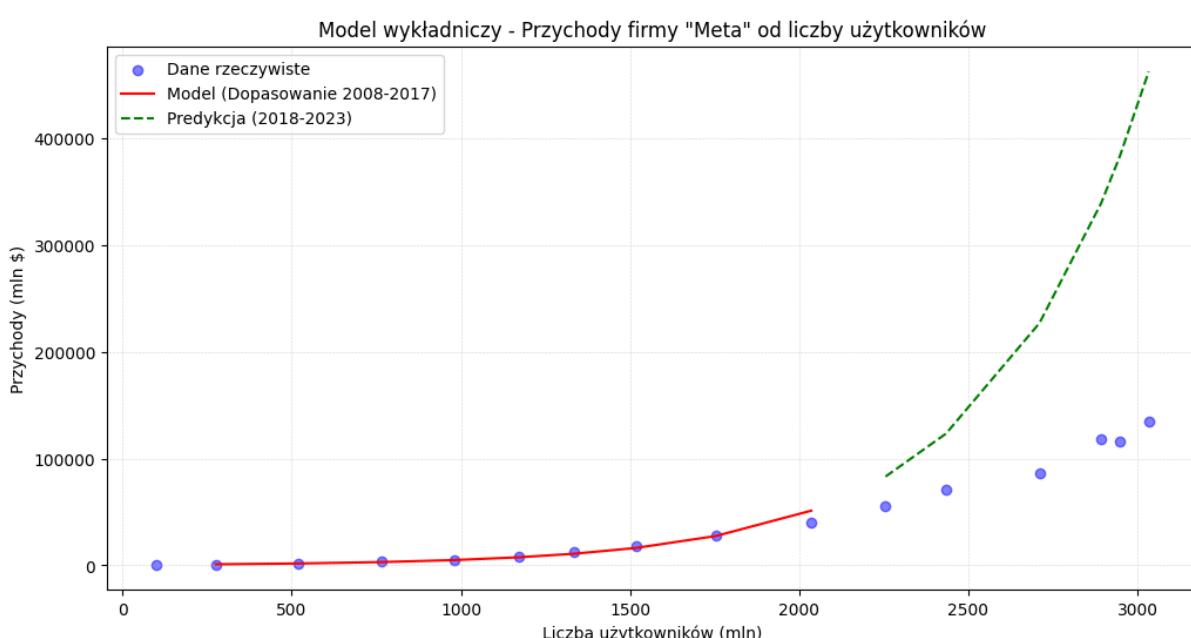
Parametry modelu (zlinearyzowanego):

- $a_0 = 6.36 \pm 0.14$,
- $a_1 = 0.022 \pm 0.001$,
- $R^2 = 98.4\%$.

Średni błąd bezwzględny prognozy:

- dla lat 2018-2020: 73754.2 mln dolarów,
- dla lat 2021-2023: 272006.9 mln dolarów.

Wyniki modelu na tle rzeczywistych danych przedstawia Rys. 11.



Rys. 11. Wizualizacja modelu wykładniczego na tle rzeczywistych danych przychodów firmy *Meta* w zależności od liczby użytkowników

Model wykładniczy mimo bardzo dobrego dopasowania do danych treningowych ($R^2 = 98.4\%$), ma bardzo słabej jakości prognozy w pełnym zakresie danych testowych. Model wykładniczy nie jest najlepszym opisem relacji między przychodami a liczbą użytkowników.

V.4.2. Model potęgowy

Parametry modelu (zlinearyzowanego):

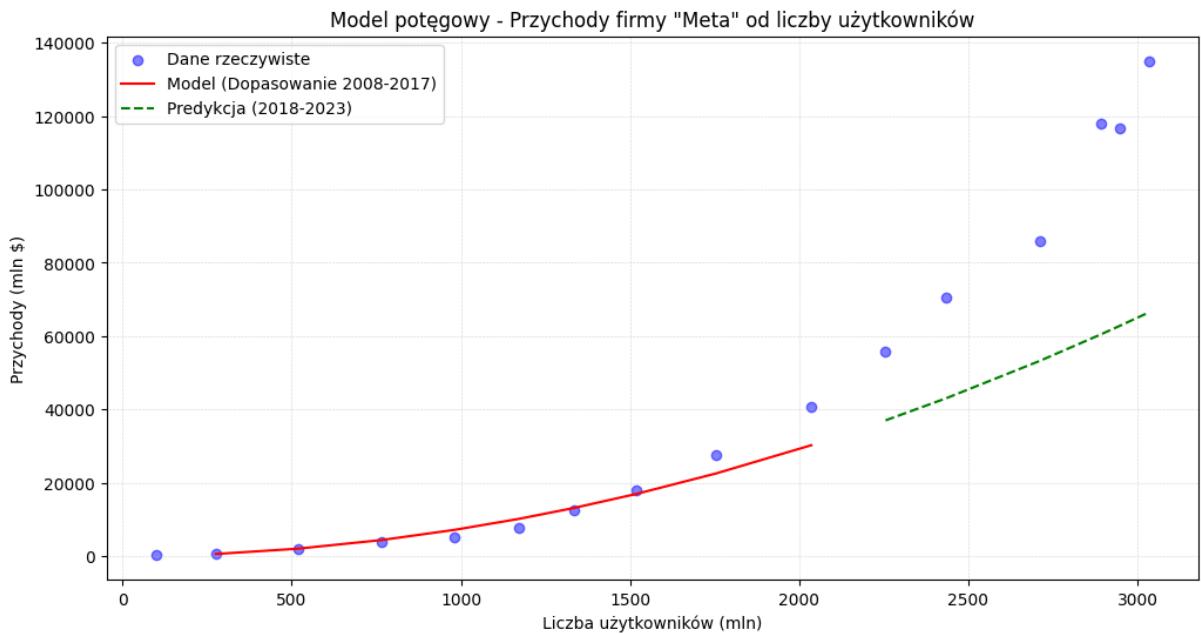
- $a_0 = -4.72 \pm 0.95$,
- $a_1 = 1.97 \pm 0.14$,
- $R^2 = 96.8\%$.

Średni błąd bezwzględny prognozy:

- dla lat 2018-2020: 26407.6 mln dolarów,

- dla lat 2021-2023: 59887.7 mln dolarów.

Wyniki modelu na tle rzeczywistych danych przedstawia Rys. 12.



Rys. 12. Wizualizacja modelu potęgowego na tle rzeczywistych danych przychodów firmy Meta w zależności od liczby użytkowników

Model potęgowy podobnie jak model wykładowczy, pomimo bardzo dobrego dopasowania do danych treningowych ($R^2 = 96.8\%$), słabo wypada w prognozie dla pełnego zakresu danych testowych, lecz w przeciwieństwie do modelu wykładowczego, prognozy znajdują się poniżej rzeczywistych danych, co sugeruje że relacja między przychodami a liczbą użytkowników leży pomiędzy modelem potęgowym a wykładowczym.

V.4.3. Model hybrydowy

Model hybrydowy wykorzystuje modele potęgowy i wykładowczy opisane w poprzednich punktach wykorzystując je w kombinacji liniowej z parametrem α (Równanie (13)).

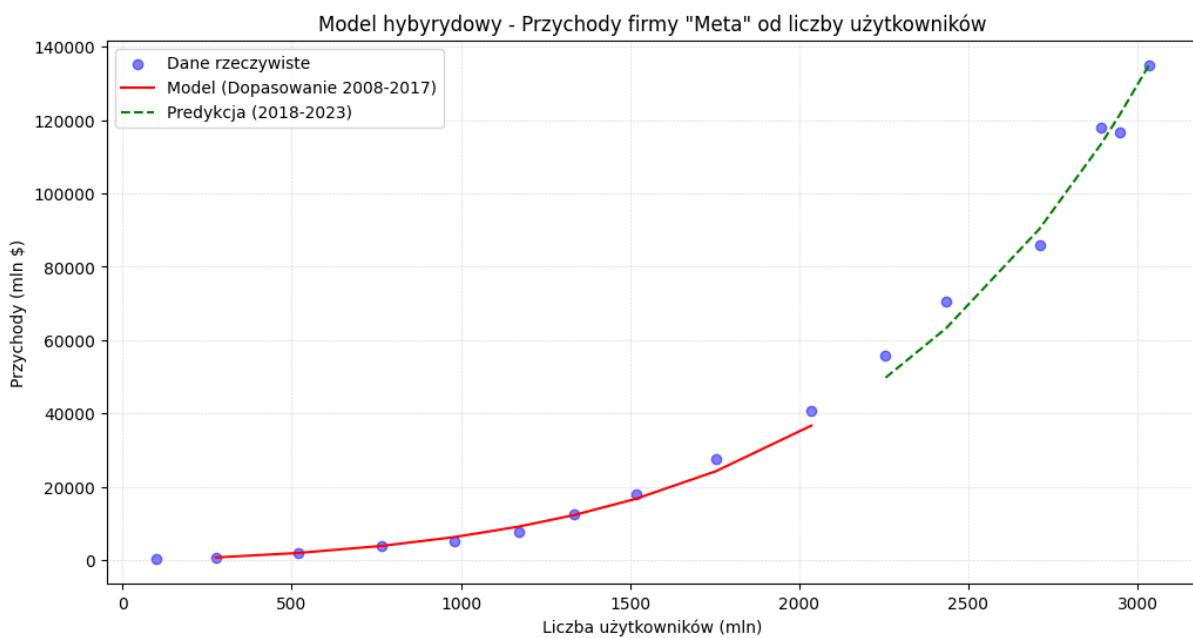
Wyznaczona wartość parametru $\alpha = 0.36$.

Parametr $R^2 = 97.8\%$.

Średni błąd bezwzględny prognozy:

- dla lat 2018-2020: 5978.4 mln dolarów,
- dla lat 2021-2023: 3108.8 mln dolarów.

Wyniki modelu na tle rzeczywistych danych przedstawia Rys. 13.



Rys. 13. Wizualizacja modelu hybrydowego na tle rzeczywistych danych przychodów firmy *Meta* w zależności od liczby użytkowników

Model hybrydowy podobnie jak model potęgowy i wykładniczy bardzo dobrze dopasowuje ($R^2 = 97.8\%$), lecz kombinacja liniowa tych modeli charakteryzuje się o wiele lepszą jakością prognozy w pełnym zakresie danych testowych.

V.4.4. Model wielomianowy 3. stopnia

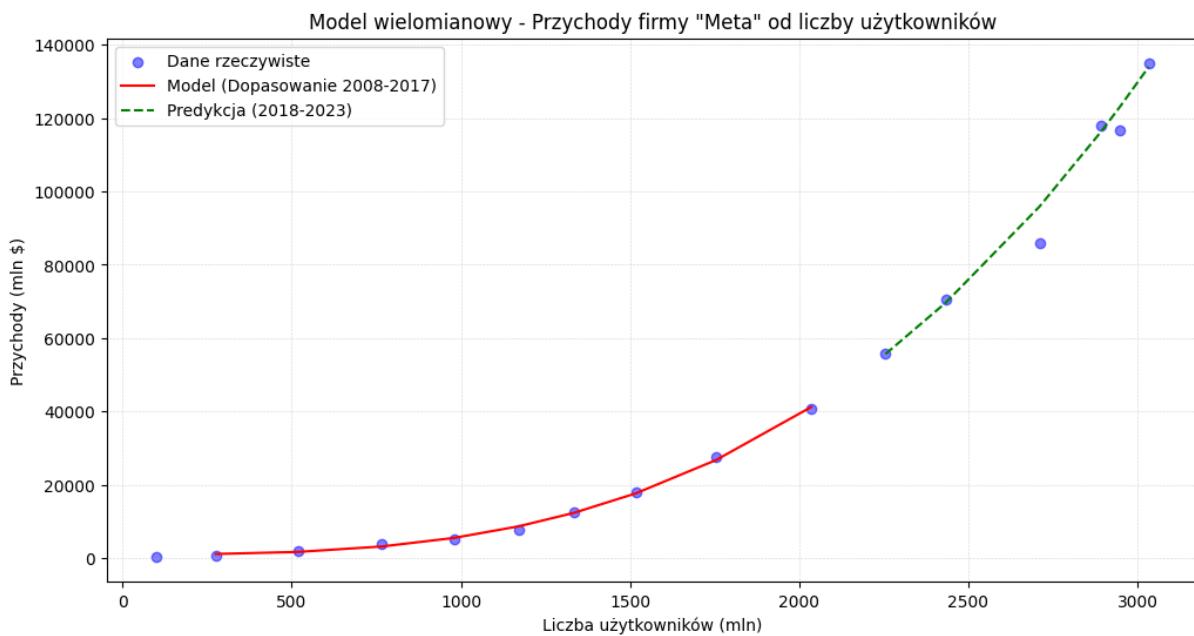
Parametry modelu (zlinearyzowanego):

- $a_0 = 1030 \pm 280$,
- $a_3 = 4.76 \cdot 10^{-6} \pm 7.6 \cdot 10^{-8}$,
- $R^2 = 99.8\%$.

Średni błąd bezwzględny prognozy:

- dla lat 2018-2020: 3679.5 mln dolarów,
- dla lat 2021-2023: 2988.9 mln dolarów.

Wyniki modelu na tle rzeczywistych danych przedstawia Rys. 14.



Rys. 14. Wizualizacja modelu wielomianowego 3. stopnia na tle rzeczywistych danych przychodów firmy *Meta* w zależności od liczby użytkowników

Model wielomianowy 3. stopnia charakteryzuje się największym dopasowaniem do danych treningowych ($R^2 = 99.8\%$), oraz znacznie lepszą jakością prognozy dla danych z przedziału 2017-2020 oraz nieznacznie lepszą jakością prognozą dla danych z przedziału 2021-2023, w porównaniu do modelu hybrydowego.

V.4.5. Wnioski

Z oceny analizowanych modeli można zauważyć, że modelem najlepiej opisującym relację między przychodem a liczbą użytkowników, jest model wielomianowy 3. stopnia, w którym występuje wyraz wolny i wyraz X^3 . Dowodzi to o silnie nieliniowej relacji, co sugeruje rosnącą efektywność monetyzacji użytkowników, wynikającą z np. lepszej personalizacji reklam, czy rosnącej wartości danych generowanych przez użytkowników.

Sprawdzono również o ile wzrosnie przychód firmy, gdyby baza jej użytkowników wzrosła o 5%. Wynikiem jest ponad 15% wzrost przychodów firmy, co potwierdza nieliniową zależność między zmiennymi.

Podsumowanie

W ramach niniejszego laboratorium przeprowadzono kompleksową analizę regresji danych firmy *Meta*, weryfikując cztery hipotezy badawcze dotyczące charakteru wzrostu liczby użytkowników oraz przychodów.

Charakter wzrostu liczby użytkowników portalu *Facebook*, ma charakter nieliniowy, najlepiej opisywany przez model logistyczny. Model ten uwzględnia naturalny punkt nasycenia na poziomie około 3277 mln użytkowników. Tempo wzrostu osiągnęło maksimum około 3-4 kwartału 2015 roku i od tego czasu stopniowo maleje.

Wzrost przychodów firmy *Meta*, również ma charakter nieliniowy, najlepiej opisywany przez model logistyczny lub hybrydowy (wykładniczo-potęgowy). Z analizy modelu logistycznego wyznaczono górną granicę przychodów firmy na poziomie około 140250 mln dolarów. Jednak obserwacja z roku 2024, przekroczyła tą asymptotę, sugerując zmianę w charakterze zjawiska w ostatnich latach, co może wiązać się z np. rozwojem technologii modeli generatywnych.

Analiza relacji między liczbą użytkowników, a przychodami ujawnia silną nieliniową zależność, najlepiej opisaną przez model wielomianowy 3. stopnia. Taka zależność wynika najprawdopodobniej z rosnącej efektywności monetyzacji użytkowników przez firmę *Meta*. Najlepiej obrazuje to ponad 15% wzrost przychodów firmy przy tylko 5% przyroście bazy użytkowników.

Wałą obserwacją jest wykorzystanie modeli hybrydowych, będących kombinacjami liniowymi modeli wykładniczych i potęgowych. Wyniki takich modeli charakteryzują się lepszą jakością prognozy oraz bardzo dobrym dopasowaniem do danych treningowych. Modele hybrydowe stanowią dobrą alternatywę do modeli wielomianowych, które są podatne na przeuczenie, co skutkuje słabej jakością prognozami.

Kolejną obserwacją jest zmiana dynamiki wzrostu w okresie 2021-2023, co pokazują wyższe błędy prognozy większości modeli w tym okresie (w porównaniu do okresu 2017-2020), co sugeruje istotną zmianę w dynamice wzrostu firmy w ostatnich latach.

Dużym ograniczeniem przy analizach danych przychodów firmy, okazała się mała liczba obserwacji (ze względu na roczną agregację danych), co znaczaco wpłynęło na jakość opracowanych modeli.

Bibliografia

- [1] Hoffmann R.: „Metody eksploracji danych - Wykład 3. Modele nieliniowe linearyzacja slajdy”, Materiały dydaktyczne WAT, 2025.
- [2] Hoffmann R.: „Metody eksploracji danych - Wykład 3. Predykcja regresja liniowa”, Materiały dydaktyczne WAT, 2025.
- [3] Hoffmann R.: „Metody eksploracji danych - Wykład 3. Predykcja regresja liniowa - slajdy”, Materiały dydaktyczne WAT, 2025.
- [4] Hoffmann R.: „Metody eksploracji danych - Wykład 4. Regresja logistyczna Cz I slajdy”, Materiały dydaktyczne WAT, 2025.
- [5] https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize_scalar.html
[dostęp 16.01.2026 r.]
- [6] https://www.statsmodels.org/dev/generated/statsmodels.regression.linear_model.OLS.html
[dostęp 16.01.2026 r.]

Spis tabel

| | | |
|---------|--|---|
| Tab. 1. | Liczba użytkowników portalu społecznościowego „Facebook” | 8 |
| Tab. 2. | Przychody firmy „Meta” | 9 |

Spis rysunków

| | | |
|----------|--|----|
| Rys. 1. | Wizualizacja modelu liniowego na tle rzeczywistych danych liczby użytkowników | 19 |
| Rys. 2. | Wizualizacja modelu wykładniczego na tle rzeczywistych danych liczby użytkowników | 20 |
| Rys. 3. | Wizualizacja modelu potęgowego na tle rzeczywistych danych liczby użytkowników | 21 |
| Rys. 4. | Wizualizacja modelu hybrydowego na tle rzeczywistych danych liczby użytkowników | 22 |
| Rys. 5. | Wizualizacja modelu logistycznego na tle rzeczywistych danych liczby użytkowników | 24 |
| Rys. 6. | Wizualizacja modelu wykładniczego na tle rzeczywistych danych przychodów firmy <i>Meta</i> | 25 |
| Rys. 7. | Wizualizacja modelu potęgowego na tle rzeczywistych danych przychodów firmy <i>Meta</i> | 26 |
| Rys. 8. | Wizualizacja modelu hybrydowego na tle rzeczywistych danych przychodów firmy <i>Meta</i> | 27 |
| Rys. 9. | Wizualizacja modelu wielomianowego 3. stopnia na tle rzeczywistych danych przychodów firmy <i>Meta</i> | 28 |
| Rys. 10. | Wizualizacja modelu logistycznego na tle rzeczywistych danych przychodów firmy <i>Meta</i> | 29 |
| Rys. 11. | Wizualizacja modelu wykładniczego na tle rzeczywistych danych przychodów firmy <i>Meta</i> w zależności od liczby użytkowników | 30 |
| Rys. 12. | Wizualizacja modelu potęgowego na tle rzeczywistych danych przychodów firmy <i>Meta</i> w zależności od liczby użytkowników | 31 |
| Rys. 13. | Wizualizacja modelu hybrydowego na tle rzeczywistych danych przychodów firmy <i>Meta</i> w zależności od liczby użytkowników | 32 |
| Rys. 14. | Wizualizacja modelu wielomianowego 3. stopnia na tle rzeczywistych danych przychodów firmy <i>Meta</i> w zależności od liczby użytkowników | 33 |

Załączniki

1. Plik źródłowy *Lab-1-Zadanie-1-Przygotowanie-danych.py* – pomocnicze funkcje przygotowujące dane do dalszej analizy.
2. Plik notebook *Lab-1-Zadanie-1-Hipoteza-1.ipynb* – kompletny kod przeprowadzonych eksperymentów w ramach weryfikacji Hipotezy I, obejmujący:
 - a. Definicję i ocenę modelu liniowego dla danych liczby użytkowników portalu *Facebook*,
 - b. Definicję i ocenę modelu wykładniczego dla danych liczby użytkowników portalu *Facebook*,
 - c. Definicję i ocenę modelu potęgowego dla danych liczby użytkowników portalu *Facebook*,
 - d. Definicję i ocenę modelu hybrydowego dla danych liczby użytkowników portalu *Facebook*.
3. Plik notebook *Lab-1-Zadanie-1-Hipoteza-2.ipynb* – kompletny kod przeprowadzonych eksperymentów w ramach weryfikacji Hipotezy II, obejmujący:
 - a. Definicję i ocenę modelu logistycznego dla danych liczby użytkowników portalu *Facebook*.
4. Plik notebook *Lab-1-Zadanie-1-Hipoteza-3.ipynb* – kompletny kod przeprowadzonych eksperymentów w ramach weryfikacji Hipotezy III, obejmujący:
 - a. Definicję i ocenę modelu wykładniczego dla danych przychodów firmy *Meta*,
 - b. Definicję i ocenę modelu potęgowego dla danych przychodów firmy *Meta*
 - c. Definicję i ocenę modelu hybrydowego dla danych przychodów firmy *Meta*,
 - d. Definicję i ocenę modelu wielomianowego 3. stopnia dla danych przychodów firmy *Meta*,
 - e. Definicję i ocenę modelu wykładniczego dla danych przychodów firmy *Meta*.
5. Plik notebook *Lab-1-Zadanie-1-Hipoteza-4.ipynb* – kompletny kod przeprowadzonych eksperymentów w ramach weryfikacji Hipotezy IV, obejmujący:
 - a. Przygotowanie rocznych danych liczby użytkowników z dostępnych danych kwartalnych,
 - b. Definicję i ocenę modelu wykładniczego dla danych przychodów firmy *Meta* i liczby użytkowników portalu *Facebook*,

- c. Definicję i ocenę modelu potęgowego dla danych przychodów firmy *Meta* i liczby użytkowników portalu *Facebook*,
 - d. Definicję i ocenę modelu hybrydowego dla danych przychodów firmy *Meta* i liczby użytkowników portalu *Facebook*,
 - e. Definicję i ocenę modelu wielomianowego 3. stopnia dla danych przychodów firmy *Meta* i liczby użytkowników portalu *Facebook*.
6. Plik danych *Lab-1-Zadanie-1-Dane.xlsx* – pełne dane liczby użytkowników portalu *Facebook* i przychodów firmy *Meta*.
 7. Plik danych *Lab-1-Zadanie-1-Dane-liczba-uzytkownikow.csv* – wyodrębnione dane liczby użytkowników portalu *Facebook*, w celu łatwiejszego ładowania.
 8. Plik danych *Lab-1-Zadanie-1-Dane-przychody.csv* – wyodrębnione dane przychodów firmy *Meta* w celu łatwiejszego ładowania.