**WOJSKOWA AKADEMIA TECHNICZNA**

**im. Jarosława Dąbrowskiego**

**WYDZIAŁ CYBERNETYKI**

**INSTYTUT INŻYNIERII SYSTEMÓW**

SPRAWOZDANIE

z przedmiotu Metody Eksploracji Danych

Zadanie 1.1



**GRUPA WCY23IJ2S1**

|  |  |
| --- | --- |
| Autorzy: | Prowadzący: |
| Zbigniew PŁATKOWSKI  Martyna FITA | **dr inż. Romuald Hoffmann** |

**Warszawa 2025/2026**

# Treść zadania

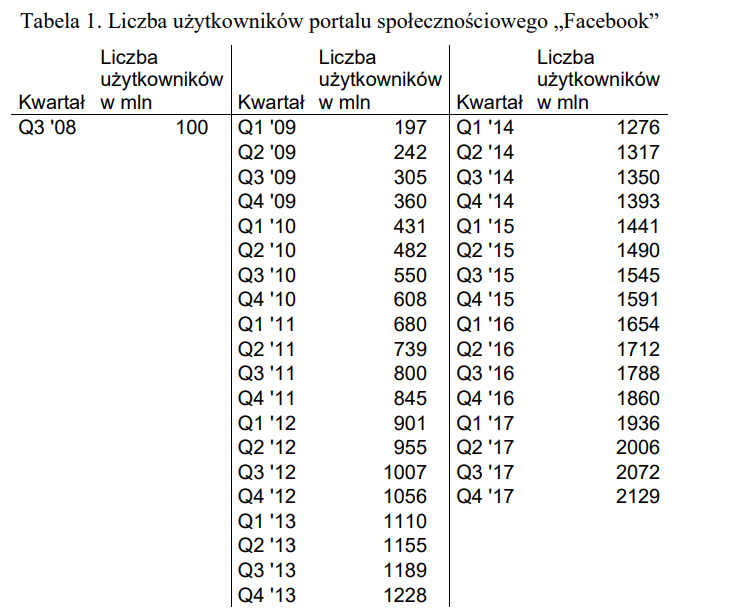
Mamy zgromadzone dane dot. dobrze znanego przedsiębiorstwa internetowego Meta w tym odnoszące się do portalu „Facebook”, które zostały zawarte w poniższych tabelach. Dane te dotyczą liczby użytkowników (klientów) w rozliczeniu na kwartały w poszczególnych latach. Pozostałe dane to przychody liczony w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

W analizie proszę:

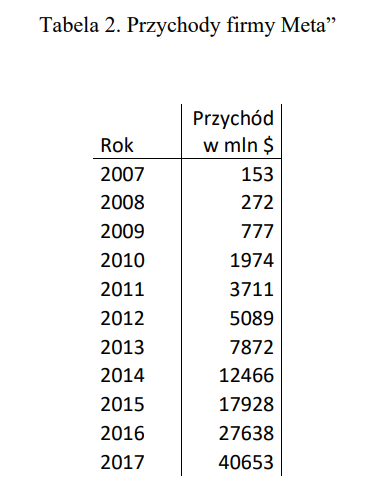
1. Proszę przeanalizować dane i zastanowić się nad tym „Co chcemy zbadać i dlaczego?”, tzn. na jakie pytania chcemy sobie odpowiedzieć przy pomocy jakich modeli.
2. W powyższym kontekście, na gruncie regresji, należy zaproponować model lub modele badające wybrane zależności i wyliczyć m. in. ich parametry 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.

# Dane

Dane wykorzystywane w zadaniu przedstawiono poniżej.



Obraz 1. Liczba użytkowników portalu społecznościowego „Facebook” w okresie Q3 ‘08 - Q4 ‘17



Obraz 2. Przychody firmy Meta

# Podstawy teoretyczne

Jednym z podstawowych narzędzi wykorzystywanych w eksploracji danych jest analiza regresji, której celem jest opisanie oraz ilościowa ocena zależności pomiędzy zmienną objaśnianą a jedną lub wieloma zmiennymi objaśniającymi. Modele regresyjne umożliwiają nie tylko analizę historycznych danych, lecz także ich praktyczne zastosowanie w procesie prognozowania przyszłych wartości badanych zmiennych.

W zależności od charakteru danych oraz obserwowanej relacji pomiędzy zmiennymi, stosuje się różne postacie modeli regresyjnych, zarówno liniowe, jak i nieliniowe. Ocena jakości dopasowania modelu odbywa się przy użyciu miar statystycznych, takich jak współczynnik determinacji, odchylenia standardowe parametrów czy analiza reszt, a także poprzez weryfikację hipotez statystycznych. Właściwie dobrany i zweryfikowany model regresyjny może stanowić użyteczne narzędzie wspomagające analizę rozwoju przedsiębiorstwa oraz podejmowanie decyzji opartych na danych.

W zadaniu wykorzystano kilka modeli, które zostały opisane poniżej.

**Model regresji liniowej**

Model regresji liniowej zakłada, że zmienna objaśniana zależy liniowo od zmiennych objaśniających , , … , i przyjmuje postać:

, gdzie - wyraz wolny, - współczynniki regresji

Dla danych empirycznych zależność ta zwykle nie jest spełniona dokładnie, dlatego wyznaczamy parametry tak, by model możliwie najlepiej przybliżał obserwacje.

W tym celu stosuje się metodę najmniejszych kwadratów (MNK), która minimalizuje sumę kwadratów reszt:

W szczególności, dla jednej zmiennej () dostajemy postać prostej:

W przypadku modeli nieliniowych estymacja parametrów może być utrudniona. Dlatego często stosuje się linearyzację, polegającą na odpowiednim przekształceniu modelu do postaci liniowej względem parametrów. Pozwala to na wykorzystanie metod regresji liniowej, takich jak metoda najmniejszych kwadratów opisana powyżej.

Powyżej przedstawiono jawne (analityczne) wzory estymatorów metody najmniejszych kwadratów dla modelu regresji liniowej z jedną zmienną objaśniającą, które można wykorzystać do ręcznego policzenia:

**Model wykładniczy z jedną zmienną objaśniającą**

Model wykładniczy zakłada, że zmienna objaśniana zależy wykładniczo od zmiennej objaśniającej :

Sprowadzenie powyższego modelu do modelu liniowego przedstawiono poniżej.

Po obustronnym logarytmowaniu:

Wprowadzenie oznaczeń:

Otrzymany zlinearyzowany model:

**Model potęgowy**

Model potęgowy zakłada, że zmienna objaśniana zależy od zmiennej objaśniającej zgodnie z poniższym wzorem:

Sprowadzenie powyższego modelu do modelu liniowego przedstawiono poniżej.

Po obustronnym logarytmowaniu:

Wprowadzenie oznaczeń:

Otrzymany zlinearyzowany model:

**Model wykładniczo-potęgowy z jedną zmienną objaśniającą**

Model wykładniczo-potęgowy zakłada, że zmienna objaśniana zależy od zmiennej objaśniającej zgodnie z poniższym wzorem:

Sprowadzenie powyższego modelu do modelu liniowego przedstawiono poniżej.

Po obustronnym logarytmowaniu:

Wprowadzenie oznaczeń:

Otrzymany zlinearyzowany model:

**Model wielomianowy drugiego stopnia z jedną zmienną objaśniającą**

Model wielomianowy zakłada, że zmienna objaśniana zależy od zmiennej objaśniającej zgodnie z poniższym wzorem:

Sprowadzenie powyższego modelu do modelu liniowego przedstawiono poniżej.

Po podstawieniu:

Otrzymany zlinearyzowany model:

**Model wielomianowy trzeciego stopnia z jedną zmienną objaśniającą**

Model wielomianowy zakłada, że zmienna objaśniana zależy od zmiennej objaśniającej zgodnie z poniższym wzorem:

Sprowadzenie powyższego modelu do modelu liniowego przedstawiono poniżej.

Po podstawieniu:

Otrzymany zlinearyzowany model:

**Model wielomianowy czwartego stopnia z jedną zmienną objaśniającą**

Model wielomianowy zakłada, że zmienna objaśniana zależy od zmiennej objaśniającej zgodnie z poniższym wzorem:

Sprowadzenie powyższego modelu do modelu liniowego przedstawiono poniżej.

Po podstawieniu:

Otrzymany zlinearyzowany model:

**Model regresji logistycznej**

Model regresji logistycznej zakłada, że zmienna objaśniana zależy od zmiennej objaśniającej zgodnie z funkcją logistyczną. Zależność ta opisana jest następującym wzorem:

Parametry , oraz są nieznane i podlegają estymacji.

W związku z nieliniowością modelu logistycznego względem parametrów, ich estymacja nie może zostać przeprowadzona bezpośrednio za pomocą klasycznej metody najmniejszych kwadratów. W tym celu stosuje się klasyczną metodę Hotellinga, polegającą na przekształceniu równania logistycznego do postaci liniowej względem parametrów poprzez różniczkowanie funkcji i zastąpienie pochodnej przyrostami. Otrzymana zależność liniowa umożliwia następnie estymację parametrów z wykorzystaniem metody najmniejszych kwadratów.

Estymację parametrów , oraz przedstawiono poniżej.

Różniczkując funkcję logistyczną względem zmiennej objaśniającej t otrzymujemy:

Stąd otrzymujemy:

Przyrost zmiennej objaśniającej może być przyjmowany jako równy jedności, więc :

Po podstawieniu:

Otrzymujemy:

Po podstawieniu:

Parametry i w otrzymanej zależności mogą zostać oszacowane **metodą najmniejszych kwadratów**, gdyż zależność ta jest liniowa względem parametrów. W ten sposób oszacowane zostają parametry modelu logistycznego oraz , będący funkcją parametru . Odpowiednie oceny parametrów i można oznaczyć przez i .

Wówczas otrzymujemy oszacowania parametrów a i k:

W celu oszacowania parametru przekształcimy wzór:

Po obustronnym logarytmowaniu i przekształceniu:

Estymator parametru wyznacza się poprzez minimalizację następującej funkcji:

Dzięki temu otrzymujemy estymator :

Oszacowanie parametru b:

**Realizacja prognoz**

Program gretl pozwala na tworzenie modeli regresji liniowej i wykorzystuje je do prognozowania wartości zmiennej objaśnianej. Po zbudowaniu modelu, określającego zależność zmiennej zależnej od zestawu zmiennych niezależnych, gretl może obliczać przewidywane wartości y dla nowych obserwacji. W praktyce oznacza to, że do równania regresji podstawiane są kolejne wartości zmiennych objaśniających, a na tej podstawie program wyznacza odpowiadające im wartości zmiennej objaśnianej.

**Istotne miary**

**Współczynnik determinacji**

Przy czym oznacza średnią arytmetyczną wartości obserwowanych, – wartości rzeczywiste, a – wartości estymowane przez model.

Ten współczynnik informuje o stopniu dopasowania modelu do danych empirycznych.

**Skorygowany współczynnik determinacji**

Przy czym oznacza liczbę obserwacji, a liczbę zmiennych objaśniających w modelu.

Skorygowany współczynnik determinacji pozwala porównywać modele zbudowane w oparciu o różną liczbę obserwacji i/lub z różną liczbą parametrów.

**Suma kwadratów reszt SSE**

Oceniając dopasowanie modelu do danych, warto analizować nie tylko współczynnik determinacji , ale również sumę kwadratów reszt. Miara ta informuje o bezwzględnej wielkości błędów dopasowania i pozwala lepiej ocenić, jak bardzo wartości estymowane odbiegają od obserwowanych.

Należy jednak zachować ostrożność przy porównywaniu wartości sumy kwadratów reszt pomiędzy różnymi modelami, zwłaszcza gdy są one estymowane na zmiennych poddanych odmiennym transformacjom. W szczególności modele wykładnicze i potęgowe, po logarytmowaniu zmiennej objaśnianej, nie są bezpośrednio porównywalne z modelami regresji liniowej estymowanymi w oryginalnej skali danych.

**Średni absolutny błąd procentowy** (ang. mean absolute percentage error, MAPE)

Przy czym oznacza liczbę punktów (obserwacji), dla których wyznaczono prognozę, czyli liczbę par .

Miara MAPE informuje, o ile procent średnio różnią się wartości prognozowane od wartości rzeczywistych. Dzięki temu pozwala w prosty sposób ocenić dokładność prognoz oraz porównywać jakość predykcji uzyskanych z różnych modeli.

# Badane zależności

W analizie zostaną zbadane trzy kluczowe relacje opisujące rozwój działalności w czasie.  
Po pierwsze, przeanalizowana zostanie zależność między liczbą użytkowników a kolejnymi kwartałami, co pozwoli ocenić dynamikę wzrostu bazy użytkowników w krótkich okresach.

Po drugie, zbadany zostanie związek między przychodem a rokiem, aby określić, jak przychody zmieniały się w dłuższej perspektywie czasowej.

Po trzecie, przeanalizowana będzie zależność między liczbą użytkowników w danym roku a wysokością przychodu, co umożliwi ocenę, czy wzrost liczby użytkowników przekłada się na wzrost przychodów i w jakim stopniu.

# 1. Liczba użytkowników a kwartał

Celem analizy jest zbadanie zależności między numerem kwartału (NoQ) a liczbą użytkowników oraz ocena, który model najlepiej opisuje tę zależność i nadaje się do prognozowania przyszłych wartości Y. Zmienna objaśniana (Y) to liczba użytkowników, natomiast zmienna objaśniająca (X) to numer kwartału.

Aby odpowiedzieć na to pytanie, korzystamy z różnych modeli regresyjnych – liniowych i nieliniowych – ponieważ nie wiemy z góry, jaki typ zależności najlepiej odzwierciedla rzeczywisty wzrost liczby użytkowników.

Do budowy modelu prognozującego liczbę użytkowników wykorzystano dane kwartalne:

* Zbiór uczący: obserwacje NoQ 1 – 29 (76% danych) – użyty do trenowania modelu
* Zbiór testowy: obserwacje NoQ 30 – 38 – użyty do oceny jakości prognozy na danych wcześniej niewidzianych

Dane wykorzystywane do badania tej zależności znajdują się w Tabeli 1. Dla kwartału Q4 ’08 brakująca wartość została obliczona jako średnia Q3 ’08 oraz Q1 ’09.

Tabela 1. Wykorzystywane dane

|  |  |  |
| --- | --- | --- |
| NoQ | QY | Users |
| 1 | Q3 ‘08 | 100 |
| 2 | Q4 ‘08 | 149 |
| 3 | Q1 ‘09 | 197 |
| 4 | Q2 ‘09 | 242 |
| 5 | Q3 ‘09 | 305 |
| 6 | Q4 ‘09 | 360 |
| 7 | Q1 ‘10 | 431 |
| 8 | Q2 ‘10 | 482 |
| 9 | Q3 ‘10 | 550 |
| 10 | Q4 ‘10 | 608 |
| 11 | Q1 ‘11 | 680 |
| 12 | Q2 ‘11 | 739 |
| 13 | Q3 ‘11 | 800 |
| 14 | Q4 ‘11 | 845 |
| 15 | Q1 ‘12 | 901 |
| 16 | Q2 ‘12 | 955 |
| 17 | Q3 ‘12 | 1007 |
| 18 | Q4 ‘12 | 1056 |
| 19 | Q1 ‘13 | 1110 |
| 20 | Q2 ‘13 | 1155 |
| 21 | Q3 ‘13 | 1189 |
| 22 | Q4 ‘13 | 1228 |
| 23 | Q1 ‘14 | 1276 |
| 24 | Q2 ‘14 | 1317 |
| 25 | Q3 ‘14 | 1350 |
| 26 | Q4 ‘14 | 1393 |
| 27 | Q1 ‘15 | 1441 |
| 28 | Q2 ‘15 | 1490 |
| 29 | Q3 ‘15 | 1545 |
| 30 | Q4 ‘15 | 1591 |
| 31 | Q1 ‘16 | 1654 |
| 32 | Q2 ‘16 | 1712 |
| 33 | Q3 ‘16 | 1788 |
| 34 | Q4 ‘16 | 1860 |
| 35 | Q1 ‘17 | 1936 |
| 36 | Q2 ‘17 | 2006 |
| 37 | Q3 ‘17 | 2072 |
| 38 | Q4 ‘17 | 2129 |
| 39 | Q1 ‘18 | 2196 |
| 40 | Q2 ‘18 | 2234 |
| 41 | Q3 ‘18 | 2271 |
| 42 | Q4 ‘18 | 2320 |
| 43 | Q1 ‘19 | 2375 |
| 44 | Q2 ‘19 | 2414 |
| 45 | Q3 ‘19 | 2449 |
| 46 | Q4 ‘19 | 2498 |
| 47 | Q1 ‘20 | 2603 |
| 48 | Q2 ‘20 | 2701 |
| 49 | Q3 ‘20 | 2740 |
| 50 | Q4 ‘20 | 2797 |
| 51 | Q1 ‘21 | 2853 |
| 52 | Q2 ‘21 | 2895 |
| 53 | Q3 ‘21 | 2910 |
| 54 | Q4 ‘21 | 2912 |
| 55 | Q1 ‘22 | 2936 |
| 56 | Q2 ‘22 | 2934 |
| 57 | Q3 ‘22 | 2958 |
| 58 | Q4 ‘22 | 2963 |
| 59 | Q1 ‘23 | 2989 |
| 60 | Q2 ‘23 | 3030 |
| 61 | Q3 ‘23 | 3049 |
| 62 | Q4’23 | 3065 |

Model wykładniczy wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , czyli logarytmu naturalnego zmiennej , która przechowuje dane na temat liczby użytkowników w poszczególnych kwartałach. Jako zmienną objaśnianą dodano czyli numery kolejnych kwartałów. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| WYKŁADNICZY | |
|  |  |
|  |  |

Model potęgowy wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , czyli logarytmu naturalnego zmiennej , która przechowuje dane na temat liczby użytkowników w poszczególnych kwartałach. Jako zmienną objaśniającą dodano , czyli logarytm naturalny zmiennej przechowującej numery kolejnych kwartałów. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| POTĘGOWY | |
|  |  |
|  |  |

Model wykładniczo-potęgowy wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , czyli logarytmu naturalnego zmiennej , która przechowuje dane na temat liczby użytkowników w poszczególnych kwartałach. Jako zmienne objaśniające dodano , czyli logarytm naturalny zmiennej numery kolejnych kwartałów oraz czyli numer kwartału. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| WYKŁADNICZO-POTĘGOWY | |
|  |  |
|  |  |

Model regresji liniowej wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmienną objaśnianą zmienną , która przechowuje dane na temat liczby użytkowników w poszczególnych kwartałach oraz jako zmienną objaśniającą , czyli numer kwartału. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| MODEL REGRESJI LINIOWEJ | |
| A screenshot of a computer  AI-generated content may be incorrect. | A screen shot of a graph  AI-generated content may be incorrect. |
| A screenshot of a computer  AI-generated content may be incorrect. | A screen shot of a graph  AI-generated content may be incorrect. |

Model wielomianowy drugiego stopnia wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , która przechowuje dane na temat liczby użytkowników w poszczególnych kwartałach. Jako zmienne objaśniane dodano , czyli numery kolejnych kwartałów oraz czyli , czyli kwadrat numeru kwartału. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| MODEL WIELOMIANOWY STOPNIA DRUGIEGO | |
| A screenshot of a computer  AI-generated content may be incorrect. | A screen shot of a graph  AI-generated content may be incorrect. |
| A screenshot of a computer  AI-generated content may be incorrect. | A screen shot of a graph  AI-generated content may be incorrect. |

Model wielomianowy trzeciego stopnia wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , która przechowuje dane na temat liczby użytkowników w poszczególnych kwartałach. Jako zmienne objaśniające dodano , czyli numery kolejnych kwartałów, czyli czyli kwadrat numeru kwartału oraz , czyli jego sześcian. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| MODEL WIELOMIANOWY STOPNIA TRZECIEGO | |
| A screenshot of a computer  AI-generated content may be incorrect. | A screen shot of a graph  AI-generated content may be incorrect. |
| A screenshot of a computer  AI-generated content may be incorrect. | A screen shot of a graph  AI-generated content may be incorrect. |

Model wielomianowy czwartego stopnia wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , która przechowuje dane na temat liczby użytkowników w poszczególnych kwartałach. Jako zmienne objaśniające dodano , czyli numery kolejnych kwartałów, czyli kwadrat numeru kwartału, czyli jego sześcian oraz czyli podniesioną do potęgi czwartej . Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| MODEL WIELOMIANOWY STOPNIA CZWARTEGO | |
| A screenshot of a computer  AI-generated content may be incorrect. | A screen shot of a graph  AI-generated content may be incorrect. |
| A screenshot of a computer  AI-generated content may be incorrect. | A graph with numbers and lines  AI-generated content may be incorrect. |

Liczba użytkowników jest zmienną, która wykazuje tendencję wzrostową jedynie do pewnego momentu, po czym stabilizuje się (w przeciwieństwie do np. przychodu). Z tego względu postanowiliśmy sprawdzić, w jakim stopniu zjawisko to może zostać opisane za pomocą modelu regresji logistycznej.

Parametry modelu regresji logistycznej wyznaczono w Excelu metodą Hotellinga. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Zastosowanie metody najmniejszych kwadratów pozwoliło na wyznaczenie parametrów (obliczone za pomocą programu gretl, w pliku Lab1-Zadanie1-UzytkownicyKwartal-RegLog-gretl.gretl), na podstawie których obliczono następnie parametry modelu logistycznego , oraz . Szczegółowy opis zastosowanej metody estymacji przedstawiono w rozdziale Podstawy teoretyczne, a obliczenia znajdują się w pliku Lab1-Zadanie1-RegLog.xlsx.

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| MODEL REGRESJI LOGISTYCZNEJ | |
|  | |
|  |  |

**Porównanie modeli**

Tabela 2. Średni absolutny błąd procentowy (MAPE)

|  |  |
| --- | --- |
| **Model** | **MAPE [%]** |
| Wykładniczy | 7,5384 |
| Potęgowy | 0,5427 |
| **Wykładniczo-potęgowy** | **0,25397** |
| Regresji liniowej | 2,1428 |
| Wielomianowy stopnia drugiego | 7,8963 |
| Wielomianowy stopnia trzeciego | 10,947 |
| Wielomianowy stopnia czwartego | 4,1964 |
| Regresji logistycznej | 8,0593 |

Średnia wielkość błędów prognoz wynosi mniej niż 1% dla modelu wykładniczo potęgowego, dlatego został wybrany jako najlepszy do prognozowania.

Tabela 3. Inne istotne miary dla wybranego powyżej modelu

|  |  |
| --- | --- |
| **Miara** | **Wartość** |
|  | 0,082 |
|  | 0,9947 |
|  | 0,9942 |

Ponadto wybrany model charakteryzuje się niską wartością sumy kwadratów reszt (SSE), co potwierdza niewielkie odchylenia wartości modelowych od empirycznych, widoczne również na wykresie. Jednocześnie wysoka wartość skorygowanego współczynnika determinacji świadczy o bardzo dobrym dopasowaniu modelu do danych.

Modele wielomianowe charakteryzują się wysokim skorygowanym współczynnikiem R-kwadrat oraz jak można zauważyć na wykresach wysokim stopniem dopasowania. Pomimo tego, modele te zostały odrzucone, ponieważ prognoza na zbiorze testowym okazała się nieskuteczna. Można stwierdzić, że te modele są przeuczone (wysokie dopasowanie, lecz słabe prognozowanie).

|  |  |
| --- | --- |
| PREDYKCJE NA 2018-2020 (model wykładniczo-potęgowy) | |
|  |  |
| PREDYKCJE NA 2021-2023 na podstawie obserwacji do 2020 (model wykładniczo-potęgowy) | |
|  |  |

Prognoza na lata 2018–2020 charakteryzuje się bardzo wysoką trafnością – wartość MAPE to jedynie 0,22%, a na wykresie wyraźnie widać dobre dopasowanie modelu. Dodatkowo wszystkie przewidywane wartości mieszczą się w 95-procentowym przedziale ufności.

Również prognoza dla okresu 2021–2023 okazała się wiarygodna – miara MAPE wynosi 0,96%, pomimo że wartości empiryczne wykraczają nieco poza przedział ufności.

# 2. Przychód a rok

Celem analizy jest zbadanie zależności między numerem roku (NoY) a osiąganym przychodem oraz określenie, który model najlepiej opisuje tę relację i najskuteczniej prognozuje przyszłe wartości Y.

Zmienna objaśniana (Y) to przychód, natomiast zmienna objaśniająca (X) to numer roku.

Podobnie jak w poprzednim przypadku, zastosowane zostaną różne modele regresyjne – zarówno liniowe, jak i nieliniowe – aby określić, która forma zależności najtrafniej odzwierciedla rzeczywisty trend zmian przychodu w czasie.

Do budowy modelu prognozującego wysokość przychodu wykorzystano dane roczne:

* Zbiór uczący: lata 2007 – 2014 (75% danych) – wykorzystany do trenowania modelu
* Zbiór testowy: lata 2015 – 2017 – użyty do oceny jakości prognozy na danych wcześniej niewidzianych

Dane wykorzystywane do badania tej zależności znajdują się w poniższej tabeli.

Tabela 4. Wykorzystywane dane

|  |  |  |
| --- | --- | --- |
| Year | NoY | Reveneue |
| 2007 | 1 | 153 |
| 2008 | 2 | 272 |
| 2009 | 3 | 777 |
| 2010 | 4 | 1974 |
| 2011 | 5 | 3711 |
| 2012 | 6 | 5089 |
| 2013 | 7 | 7872 |
| 2014 | 8 | 12466 |
| 2015 | 9 | 17928 |
| 2016 | 10 | 27638 |
| 2017 | 11 | 40653 |
| 2018 | 12 | 55838 |
| 2019 | 13 | 70697 |
| 2020 | 14 | 85965 |
| 2021 | 15 | 117929 |
| 2022 | 16 | 116609 |
| 2023 | 17 | 134902 |
| 2024 | 18 | 164501 |

Model wykładniczy wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , czyli logarytmu naturalnego zmiennej , która przechowuje dane na temat przychodu w poszczególnych latach. Jako zmienną objaśnianą dodano czyli numery kolejnych lat. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| WYKŁADNICZY | |
|  |  |
|  |  |

Model potęgowy wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , czyli logarytmu naturalnego zmiennej , która przechowuje dane na temat przychodu w poszczególnych latach. Jako zmienną objaśniającą dodano , czyli logarytm naturalny zmiennej przechowującej numery kolejnych lat. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| POTĘGOWY | |
|  |  |
|  |  |

Model wykładniczo-potęgowy wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , czyli logarytmu naturalnego zmiennej , która przechowuje dane na temat przychodu w poszczególnych latach. Jako zmienne objaśniające dodano , czyli logarytm naturalny zmiennej numery kolejnych lat oraz czyli numer roku. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| WYKŁADNICZO-POTĘGOWY | |
|  |  |
|  |  |

Model regresji liniowej wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmienną objaśnianą zmienną , która przechowuje dane na temat przychodu w poszczególnych latach oraz jako zmienną objaśniającą , czyli numer roku. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| MODEL REGRESJI LINIOWEJ | |
| Obraz zawierający tekst, zrzut ekranu, oprogramowanie, wyświetlacz  Zawartość wygenerowana przez AI może być niepoprawna. | Obraz zawierający tekst, zrzut ekranu, oprogramowanie, numer  Zawartość wygenerowana przez AI może być niepoprawna. |
| Obraz zawierający tekst, elektronika, zrzut ekranu, oprogramowanie  Zawartość wygenerowana przez AI może być niepoprawna. | Obraz zawierający tekst, zrzut ekranu, oprogramowanie, System operacyjny  Zawartość wygenerowana przez AI może być niepoprawna. |

Model wielomianowy drugiego stopnia wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , która przechowuje dane na temat przychodu w poszczególnych latach. Jako zmienne objaśniane dodano , czyli numery kolejnych lat oraz czyli kwadraty numerów lat. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| MODEL WIELOMIANOWY STOPNIA DRUGIEGO | |
| Obraz zawierający tekst, zrzut ekranu, oprogramowanie, wyświetlacz  Zawartość wygenerowana przez AI może być niepoprawna. | Obraz zawierający tekst, zrzut ekranu, oprogramowanie, numer  Zawartość wygenerowana przez AI może być niepoprawna. |
| Obraz zawierający tekst, elektronika, zrzut ekranu, oprogramowanie  Zawartość wygenerowana przez AI może być niepoprawna. | Obraz zawierający zrzut ekranu, tekst, oprogramowanie, System operacyjny  Zawartość wygenerowana przez AI może być niepoprawna. |

Model wielomianowy trzeciego stopnia wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , która przechowuje dane na temat przychodu w poszczególnych latach. Jako zmienne objaśniające dodano , czyli numery kolejnych lat, czyli kwadrat numeru roku oraz , czyli jego sześcian. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| MODEL WIELOMIANOWY STOPNIA TRZECIEGO | |
| Obraz zawierający tekst, zrzut ekranu, oprogramowanie, numer  Zawartość wygenerowana przez AI może być niepoprawna. | Obraz zawierający tekst, zrzut ekranu, oprogramowanie, numer  Zawartość wygenerowana przez AI może być niepoprawna. |
| Obraz zawierający tekst, elektronika, zrzut ekranu, oprogramowanie  Zawartość wygenerowana przez AI może być niepoprawna. | Obraz zawierający zrzut ekranu, tekst, oprogramowanie, System operacyjny  Zawartość wygenerowana przez AI może być niepoprawna. |

Model wielomianowy czwartego stopnia wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , która przechowuje dane na temat przychodu w poszczególnych latach. Jako zmienne objaśniające dodano , czyli numery kolejnych lat, czyli kwadrat numeru roku, czyli jego sześcian oraz czyli podniesioną do potęgi czwartej zmienną . Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| MODEL WIELOMIANOWY STOPNIA CZWARTEGO | |
| Obraz zawierający tekst, zrzut ekranu, oprogramowanie, numer  Zawartość wygenerowana przez AI może być niepoprawna. | Obraz zawierający tekst, zrzut ekranu, oprogramowanie, numer  Zawartość wygenerowana przez AI może być niepoprawna. |
| Obraz zawierający tekst, elektronika, zrzut ekranu, oprogramowanie  Zawartość wygenerowana przez AI może być niepoprawna. | Obraz zawierający zrzut ekranu, tekst, oprogramowanie, System operacyjny  Zawartość wygenerowana przez AI może być niepoprawna. |

**Porównanie modeli**

Tabela 5. Średni absolutny błąd procentowy (MAPE)

|  |  |
| --- | --- |
| **Model** | **MAPE [%]** |
| Wykładniczy | 7,9992 |
| Potęgowy | 5,2026 |
| **Wykładniczo-potęgowy** | **2,298** |
| Regresji liniowej | 50,698 |
| Wielomianowy stopnia drugiego | 24,156 |
| Wielomianowy stopnia trzeciego | 10,998 |
| Wielomianowy stopnia czwartego | 19,526 |

Średnia wielkość błędów prognoz jest najmniejsza i wynosi około 2,3% dla modelu wykładniczo potęgowego, dlatego został wybrany jako najlepszy do prognozowania.

Tabela 6. Inne istotne miary dla wybranego powyżej modelu

|  |  |
| --- | --- |
| **Miara** | **Wartość** |
|  | 0,265 |
|  | 0,9852 |
|  | 0,9793 |

Ponadto wybrany model charakteryzuje się niską wartością sumy kwadratów reszt (SSE), co potwierdza niewielkie odchylenia wartości modelowych od empirycznych, widoczne również na wykresie. Jednocześnie wysoka wartość skorygowanego współczynnika determinacji świadczy o bardzo dobrym dopasowaniu modelu do danych.

Modele wielomianowe charakteryzują się wysokim skorygowanym współczynnikiem R-kwadrat oraz jak można zauważyć na wykresach wysokim stopniem dopasowania. Pomimo tego, modele te zostały odrzucone, ponieważ prognoza na zbiorze testowym okazała się nieskuteczna. Można stwierdzić, że te modele są przeuczone (wysokie dopasowanie, lecz słabe prognozowanie).

|  |  |
| --- | --- |
| Prognoza 2018-2020 (model wykładniczo-potęgowy) | |
| A screenshot of a computer  AI-generated content may be incorrect. | A screenshot of a graph  AI-generated content may be incorrect. |
| Prognoza 2021-2023 z przesuniętą próbą do 2020 (model wykładniczo-potęgowy) | |
|  |  |

Prognoza na lata 2018–2020 charakteryzuje się wysoką trafnością – wartość MAPE to 5,436%, a na wykresie widać dobre dopasowanie modelu. Dodatkowo wszystkie przewidywane wartości mieszczą się w 95-procentowym przedziale ufności.

Prognoza dla okresu 2021–2023 również okazała się wiarygodna, jej dokładność jest wyższa – miara MAPE wynosi 3,84%, mimo że wartości empiryczne wykraczają nieco poza przedział ufności.

# 3. Przychód a średnia użytkowników w latach

W tej części analizy celem jest zbadanie zależności między średnią liczbą użytkowników a osiąganym przychodem. Podobnie jak wcześniej, zbudowany zostanie model regresyjny opisujący relację między zmienną objaśnianą Y (przychód) a zmienną objaśniającą X (średnia użytkowników).

Do konstrukcji modelu wykorzystano dane z poszczególnych lat, podzielone na dwa zbiory:

* Zbiór uczący: lata 2009 – 2014 (66% danych) – użyty do trenowania modelu regresyjnego
* Zbiór testowy: lata 2015 – 2017 – wykorzystany do oceny zdolności predykcyjnych modelu na danych niewidzianych wcześniej

Celem analizy jest określenie, w jakim stopniu zmiany w średniej liczbie użytkowników determinują wahania przychodu oraz który z zastosowanych modeli regresji – liniowych lub nieliniowych – najtrafniej odwzorowuje tę zależność.

W analizie zostaną ocenione różne typy modeli, aby ustalić, który z nich najlepiej opisuje relację między X a Y oraz najskuteczniej prognozuje wartości przychodu na podstawie średniej liczby użytkowników.

Lata 2007 i 2008 pominięto w analizie ze względu na brakujące dane na temat średniej użytkowników. Dla pozostałych lat wyznaczono średnią z kwartałów 1-4.

Dane wykorzystywane do badania tej zależności znajdują się w poniższej tabeli.

Tabela 7. Wykorzystywane dane

|  |  |  |  |
| --- | --- | --- | --- |
| Year | NoY | Reveneue | Srednia\_uz |
| 2007 | 1 | 153 |  |
| 2008 | 2 | 272 |  |
| 2009 | 3 | 777 | 276 |
| 2010 | 4 | 1974 | 517,75 |
| 2011 | 5 | 3711 | 766 |
| 2012 | 6 | 5089 | 979,75 |
| 2013 | 7 | 7872 | 1170,5 |
| 2014 | 8 | 12466 | 1334 |
| 2015 | 9 | 17928 | 1516,75 |
| 2016 | 10 | 27638 | 1753,5 |
| 2017 | 11 | 40653 | 2035,75 |
| 2018 | 12 | 55838 | 2255,25 |
| 2019 | 13 | 70697 | 2434 |
| 2020 | 14 | 85965 | 2710,25 |
| 2021 | 15 | 117929 | 2892,5 |
| 2022 | 16 | 116609 | 2947,75 |
| 2023 | 17 | 134902 | 3033,25 |

Model potęgowy wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , czyli logarytmu naturalnego zmiennej , która przechowuje dane na temat przychodu w poszczególnych latach. Jako zmienną objaśniającą dodano , czyli logarytm naturalny zmiennej przechowującej średnią liczbę użytkowników kolejnych lat. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| POTĘGOWY | |
|  |  |
|  |  |

Model wykładniczy wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , czyli logarytmu naturalnego zmiennej , która przechowuje dane na temat przychodu w poszczególnych latach. Jako zmienną objaśniającą dodano czyli średnią użytkowników z kolejnych lat. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| WYKŁADNICZY | |
|  |  |
|  |  |

Model wykładniczo-potęgowy wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , czyli logarytmu naturalnego zmiennej , która przechowuje dane na temat przychodu w poszczególnych latach. Jako zmienne objaśniające dodano , czyli logarytm naturalny zmiennej przechowującej średnią liczbę użytkowników oraz . Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| WYKŁADNICZO-POTĘGOWY | |
|  |  |
|  |  |

Model regresji liniowej wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmienną objaśnianą zmienną , która przechowuje dane na temat przychodu w poszczególnych latach oraz jako zmienną objaśniającą , czyli liczbę użytkowników w danym roku. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| MODEL REGRESJI LINIOWEJ | |
| Obraz zawierający tekst, zrzut ekranu, oprogramowanie, numer  Zawartość wygenerowana przez AI może być niepoprawna. | Obraz zawierający tekst, zrzut ekranu, oprogramowanie, numer  Zawartość wygenerowana przez AI może być niepoprawna. |
| Obraz zawierający tekst, elektronika, zrzut ekranu, oprogramowanie  Zawartość wygenerowana przez AI może być niepoprawna. | Obraz zawierający zrzut ekranu, tekst, oprogramowanie, System operacyjny  Zawartość wygenerowana przez AI może być niepoprawna. |

Model wielomianowy drugiego stopnia wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , która przechowuje dane na temat przychodu w poszczególnych latach. Jako zmienne objaśniane dodano , czyli numery kolejnych lat oraz czyli kwadraty rocznej liczby użytkowników. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| MODEL WIELOMIANOWY STOPNIA DRUGIEGO | |
| Obraz zawierający tekst, zrzut ekranu, oprogramowanie, numer  Zawartość wygenerowana przez AI może być niepoprawna. | Obraz zawierający tekst, zrzut ekranu, oprogramowanie, numer  Zawartość wygenerowana przez AI może być niepoprawna. |
| Obraz zawierający tekst, elektronika, zrzut ekranu, oprogramowanie  Zawartość wygenerowana przez AI może być niepoprawna. | Obraz zawierający zrzut ekranu, tekst, oprogramowanie, System operacyjny  Zawartość wygenerowana przez AI może być niepoprawna. |

Model wielomianowy trzeciego stopnia wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , która przechowuje dane na temat przychodu w poszczególnych latach. Jako zmienne objaśniające dodano , czyli średnią użytkowników kolejnych lat, czyli kwadrat średniej liczby użytkowników oraz czyli jej sześcian. Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| MODEL WIELOMIANOWY STOPNIA TRZECIEGO | |
| Obraz zawierający tekst, zrzut ekranu, oprogramowanie, numer  Zawartość wygenerowana przez AI może być niepoprawna. | Obraz zawierający tekst, zrzut ekranu, oprogramowanie, numer  Zawartość wygenerowana przez AI może być niepoprawna. |
| Obraz zawierający tekst, elektronika, zrzut ekranu, oprogramowanie  Zawartość wygenerowana przez AI może być niepoprawna. | Obraz zawierający zrzut ekranu, tekst, oprogramowanie, System operacyjny  Zawartość wygenerowana przez AI może być niepoprawna. |

Model wielomianowy czwartego stopnia wyznaczono w programie gretl metodą najmniejszych kwadratów poprzez dodanie jako zmiennej objaśnianej , która przechowuje dane na temat przychodu w poszczególnych latach. Jako zmienne objaśniające dodano , czyli użytkowników kolejnych lat, czyli ich kwadrat, czyli ich sześcian oraz czyli podniesioną do potęgi czwartej zmienną . Otrzymany wynik wraz z wizualizacją przedstawiono w tabeli poniżej. Model, który otrzymaliśmy jest opisany następującym wzorem:

Następnie na podstawie otrzymanego modelu wyznaczono prognozę na zbiór testowy. Otrzymany wynik oraz wykres mu odpowiadający również przedstawiono w poniższej tabeli.

|  |  |
| --- | --- |
| MODEL WIELOMIANOWY STOPNIA CZWARTEGO | |
| Obraz zawierający tekst, zrzut ekranu, oprogramowanie, numer  Zawartość wygenerowana przez AI może być niepoprawna. | Obraz zawierający tekst, zrzut ekranu, oprogramowanie, numer  Zawartość wygenerowana przez AI może być niepoprawna. |
| Obraz zawierający tekst, elektronika, zrzut ekranu, oprogramowanie  Zawartość wygenerowana przez AI może być niepoprawna. | Obraz zawierający zrzut ekranu, tekst, oprogramowanie, linia  Zawartość wygenerowana przez AI może być niepoprawna. |

**Porównanie modeli**

Tabela 8. Średni absolutny błąd procentowy (MAPE)

|  |  |
| --- | --- |
| **Model** | **MAPE [%]** |
| Potęgowy | 5,1934 |
| Wykładniczy | 3,0181 |
| **Wykładniczo-potęgowy** | **0,9926** |
| Regresji liniowej | 45,56 |
| Wielomianowy stopnia drugiego | 16,771 |
| Wielomianowy stopnia trzeciego | 25,07 |
| Wielomianowy stopnia czwartego | 89,961 |

Średnia wielkość błędów prognoz wynosi niecały 1% dla modelu wykładniczo-potęgowego, dlatego został wybrany jako najlepszy do prognozowania.

Tabela 9. Inne istotne miary dla wybranego powyżej modelu

|  |  |
| --- | --- |
| **Miara** | **Wartość** |
|  | 0,030 |
|  | 0,9939 |
|  | 0,9899 |

Ponadto wybrany model charakteryzuje się niską wartością sumy kwadratów reszt (SSE), co potwierdza niewielkie odchylenia wartości modelowych od empirycznych, widoczne również na wykresie. Jednocześnie wysoka wartość skorygowanego współczynnika determinacji świadczy o bardzo dobrym dopasowaniu modelu do danych.

Modele wielomianowe charakteryzują się wysokim skorygowanym współczynnikiem R-kwadrat oraz jak można zauważyć na wykresach wysokim stopniem dopasowania. Pomimo tego, modele te zostały odrzucone, ponieważ prognoza na zbiorze testowym okazała się nieskuteczna. Można stwierdzić, że te modele są przeuczone (wysokie dopasowanie, lecz słabe prognozowanie).

|  |  |
| --- | --- |
| Prognoza 2018-2020 (model wykładniczo-potęgowy) | |
|  |  |
| Prognoza 2021-2023 z próbą przesuniętą do 2020 (model wykładniczo-potęgowy) | |
|  |  |

Prognoza na lata 2018–2020 charakteryzuje się bardzo wysoką trafnością – wartość MAPE to jedynie 0,93%, a na wykresie wyraźnie widać dobre dopasowanie modelu. Dodatkowo wszystkie przewidywane wartości mieszczą się w 95-procentowym przedziale ufności.

Również prognoza dla okresu 2021–2023 okazała się wiarygodna, choć jej dokładność jest nieco niższa – miara MAPE wynosi 2,47%.

# Wnioski

Przeprowadzona analiza regresyjna pozwoliła na ocenę zależności pomiędzy liczbą użytkowników, czasem oraz osiąganymi przychodami przedsiębiorstwa Meta. W każdym z analizowanych przypadków wykazano, że badane zależności mają charakter nieliniowy, a modele liniowe nie są wystarczające do ich poprawnego opisu ani prognozowania.

W analizie zależności liczby użytkowników od kwartału najlepsze właściwości prognostyczne uzyskał model wykładniczo-potęgowy, dla którego wartość średniego absolutnego błędu procentowego była mniejsza niż 1%. Modele wielomianowe, mimo bardzo dobrego dopasowania do danych uczących, wykazywały słabą skuteczność predykcyjną na zbiorze testowym, co wskazuje na zjawisko przeuczenia.

Również w przypadku analizy przychodu w funkcji czasu oraz zależności przychodu od średniej liczby użytkowników, model wykładniczo-potęgowy okazał się najtrafniejszy, osiągając najniższe wartości MAPE spośród wszystkich rozważanych modeli. Uzyskane prognozy dla lat 2018–2020 oraz 2021–2023 cechowały się dobrą lub bardzo dobrą dokładnością, przy nieznacznym pogorszeniu jakości prognoz w dłuższym horyzoncie czasowym.

Podsumowując, model wykładniczo-potęgowy najlepiej opisuje dynamikę rozwoju analizowanego przedsiębiorstwa i może być z powodzeniem wykorzystywany do krótkoterminowych prognoz liczby użytkowników oraz przychodów.