**METODY EKSPLORACJI DANYCH**

Laboratorium. Analiza regresji - regresja liniowa

Przekształcanie funkcji nieliniowych w równoważne liniowe

Budowa modelu na podstawie danych (ogólnie dostępnych)

Prowadzący: Wykonali:

dr inż. Romuald Hoffmann pchor. Michał ADAMCZEWSKI

pchor. Mikołaj ADAMSKI

pchor. Przemysław SUJECKI

Spis treści

[1. Proszę przeanalizować dane i zastanowić się nad tym „Co chcemy zbadać i dlaczego?”, tzn. na jakie pytania chcemy sobie odpowiedzieć. 2](#_Toc155681855)

[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. 4](#_Toc155681856)

[a) Liczba użytkowników serwisu „Meta” względem czasu. 4](#_Toc155681857)

[Model regresji liniowej na podstawie danych z tabeli 4](#_Toc155681858)

[Model regresji liniowej z uwzględnieniem lat 2018-2020 5](#_Toc155681859)

[Model regresji liniowej z uwzględnieniem lat 2021-2022 6](#_Toc155681860)

[Model regresji liniowej z uwzględnieniem lat 2018-2022 7](#_Toc155681861)

[Model regresji wielomianowej 2. stopnia dla danych z tabeli 8](#_Toc155681862)

[Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2018-2020 9](#_Toc155681863)

[Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2021-2022 10](#_Toc155681864)

[Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2018-2022 11](#_Toc155681865)

[Model regresji wielomianowej 3. stopnia dla danych z tabeli 12](#_Toc155681866)

[Model regresji wielomianowej 3. stopnia z uwzględnieniem lat 2018-2020 13](#_Toc155681867)

[Model regresji wielomianowej 3. stopnia z uwzględnieniem lat 2021-2022 14](#_Toc155681868)

[Model regresji wielomianowej 3. stopnia z uwzględnieniem lat 2018-2022 15](#_Toc155681869)

[Zastosowane wzory: 16](#_Toc155681870)

[Analiza wyników problemu nr 1: 17](#_Toc155681871)

[b) Chcemy zbadać zależność między przychodem, a zatrudnieniem w firmie „Meta”. 24](#_Toc155681872)

[Model regresji liniowej na podstawie danych z tabeli 24](#_Toc155681873)

[Model regresji liniowej z uwzględnieniem lat 2018-2020 25](#_Toc155681874)

[Model regresji liniowej z uwzględnieniem lat 2021-2022 26](#_Toc155681875)

[Model regresji liniowej z uwzględnieniem lat 2018-2022 27](#_Toc155681876)

[Model regresji wielomianowej 2. stopnia na podstawie danych z tabeli 28](#_Toc155681877)

[Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2018-2020 29](#_Toc155681878)

[Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2021-2022 30](#_Toc155681879)

[Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2018-2022 31](#_Toc155681880)

[Analiza wyników problemu nr 2: 32](#_Toc155681881)

[Opis fragmentów kodu (python): 39](#_Toc155681882)

## Proszę przeanalizować dane i zastanowić się nad tym „Co chcemy zbadać i dlaczego?”, tzn. na jakie pytania chcemy sobie odpowiedzieć.

Obraz zawierający tekst, zrzut ekranu, krąg, diagram

Opis wygenerowany automatycznie

Rys. 1 Wizualizacja korelacji pomiędzy danymi - RStudio

1. Chcemy zbadać liczbę użytkowników serwisu „Meta” względem czasu (kwartałów).
2. Chcemy zbadać zależność między przychodem, a zatrudnieniem w firmie „Meta”.

## 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.

## Liczba użytkowników serwisu „Meta” względem czasu.

## Model regresji liniowej na podstawie danych z tabeli

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Rys. 2 Model regresji liniowej z uwzględnieniem lat 2009 - 2017

Obraz zawierający tekst, Czcionka, zrzut ekranu

Opis wygenerowany automatycznie

Rys. 3 Współczynniki dla modelu regresji liniowej z uwzględnieniem lat 2009 - 2017

## Model regresji liniowej z uwzględnieniem lat 2018-2020

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Rys. 4 Model regresji liniowej z uwzględnieniem lat 2018 - 2020

Obraz zawierający tekst, Czcionka, zrzut ekranu

Opis wygenerowany automatycznie

Rys. 5 Współczynniki dla modelu regresji liniowej z uwzględnieniem lat 2018 - 2020

## Model regresji liniowej z uwzględnieniem lat 2021-2022

Obraz zawierający tekst, zrzut ekranu, linia, Wykres

Opis wygenerowany automatycznie

Rys. 6 Model regresji liniowej z uwzględnieniem lat 2021 - 2022

Obraz zawierający tekst, Czcionka, zrzut ekranu

Opis wygenerowany automatycznie

Rys. 7 Współczynniki dla modelu regresji liniowej z uwzględnieniem lat 2021 - 2022

## Model regresji liniowej z uwzględnieniem lat 2018-2022

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Rys. 8 Model regresji liniowej z uwzględnieniem lat 2018 - 2022

Obraz zawierający tekst, Czcionka, zrzut ekranu

Opis wygenerowany automatycznie

Rys. 9 Współczynniki dla modelu regresji liniowej z uwzględnieniem lat 2018 - 2022

## Model regresji wielomianowej 2. stopnia dla danych z tabeli

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Rys. 10 Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2009 - 2017

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Rys. 11 Współczynniki dla modelu regresji wielomianowej 2. stopnia z uwzględnieniem lat 2009 - 2017

## Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2018-2020

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Rys. 12 Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2018 - 2020

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Rys. 13 Współczynniki dla modelu regresji wielomianowej 2. stopnia z uwzględnieniem lat 2018 - 2020

## Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2021-2022

Obraz zawierający tekst, zrzut ekranu, linia, Wykres

Opis wygenerowany automatycznie

Rys. 14 Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2021 - 2022

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Rys. 15 Współczynniki dla modelu regresji wielomianowej 2. stopnia z uwzględnieniem lat 2021 - 2022

## Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2018-2022

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Rys. 16 Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2018 - 2022

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Rys. 17 Współczynniki dla modelu regresji wielomianowej 2. stopnia z uwzględnieniem lat 2018 - 2022

## Model regresji wielomianowej 3. stopnia dla danych z tabeli

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Rys. 18 Model regresji wielomianowej 3. stopnia z uwzględnieniem lat 2009 – 2017

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Rys. 19 Współczynniki dla modelu regresji wielomianowej 3. stopnia z uwzględnieniem lat 2009 - 2017

## Model regresji wielomianowej 3. stopnia z uwzględnieniem lat 2018-2020

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Rys. 20 Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2018 - 2020

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Rys. 21 Współczynniki dla modelu regresji wielomianowej 3. stopnia z uwzględnieniem lat 2018 - 2020

## Model regresji wielomianowej 3. stopnia z uwzględnieniem lat 2021-2022

Obraz zawierający tekst, zrzut ekranu, linia, Wykres

Opis wygenerowany automatycznie

Rys. 22 Model regresji wielomianowej 3. stopnia z uwzględnieniem lat 2021 - 2022

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Rys. 23 Współczynniki dla modelu regresji wielomianowej 3. stopnia z uwzględnieniem lat 2021 - 2022

## Model regresji wielomianowej 3. stopnia z uwzględnieniem lat 2018-2022

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Rys. 24 Model regresji wielomianowej 3. stopnia z uwzględnieniem lat 2018 - 2022

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Rys. 25 Współczynniki dla modelu regresji wielomianowej 3. stopnia z uwzględnieniem lat 2018 - 2022

## Zastosowane wzory:

Współczynnik regresji linowej w analizie regresji liniowej można obliczyć za pomocą wzoru:

Wyraz wolny (Intercept) można obliczyć za pomocą wzoru:

Model wielomianowy drugiego stopnia:

Model wielomianowy trzeciego stopnia:

Współczynnik determinacji (R2) można obliczyć za pomocą następującego wzoru:

Gdzie SST to całkowita suma kwadratów:

Gdzie SSR to regresyjna suma kwadratów:

Pierwiastek błędu średniokwadratowego (RMSE) to miara oceny precyzji modelu, szczególnie w kontekście prognozowania. Można go obliczyć za pomocą następującego wzoru:

Maksimum i minimum reszt można obliczyć, znajdując największą i najmniejszą wartość reszt w zbiorze danych.

Maximum reszt = max(ei)

Minimum reszt = min(ei)

Korelacja wzór:

,

gdzie:

n – oznacza liczbę obserwacji,

- wartość statystyki odczytanej z tablic testu t-Studenta dla danego poziomu istotności a oraz n – 2 stopni swobody. Zazwyczaj a = 0,1 lub a = 0,05 lub a = 0,01.

## Analiza wyników problemu nr 1:

Dla lepszego zobrazowania postanowiliśmy rozbić nasze dane uczące się i testowe w następujący sposób:

|  |  |  |
| --- | --- | --- |
| lp. | Dane rzeczywiste | Dane testowe |
| 1 | 2009-2014 | 2015-2017 |
| 2 | 2009-2017 | 2018-2020 |
| 3 | 2009-2020 | 2021-2022 |
| 4 | 2009-2017 | 2018-2022 |

Taki rodzaj podziału pozwala na przetestowanie modelu na różnych aspektach ogólnego zastosowania i pozwala ocenić, czy model jest elastyczny i skuteczny w różnych warunkach czasowych.

Przechodząc do analizy danych, skoncentrujemy się na porównaniu wydajności trzech różnych modeli regresji: liniowej, wielomianowej drugiego stopnia oraz wielomianowej trzeciego stopnia. Dane, które posłużą jako podstawa do tej analizy, obejmują rzeczywiste liczby użytkowników firmy Meta w latach 2009-2013, które będą użyte jako dane treningowe, oraz dane z lat 2014-2017 (dane z tabeli), które posłużą jako zestaw testowy do oceny skuteczności modeli.

Poniżej znajduje się zestawienie wyników dla poszczególnych modeli, przedstawiające kluczowe metryki statystyczne takie jak współczynnik korelacji, R-squared, RMSE oraz zakres reszt, które są niezbędne do oceny dokładności i wiarygodności każdego modelu. Dodatkowo, dla każdego modelu zostało przedstawione równanie regresji, co pozwala na wizualizację zależności między kwartałami a liczbą użytkowników.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Intercept | R^2 | RMSE | Min reszt | Max reszt | Równanie wykresu |
| Liniowa | 175.514 | 0.993 | 30.064 | -59.607 | 39.152 | y = 53.212x + 175.514 |
| Wielomianowa 2 stopnia | 102.548 | 0.999 | 8.655 | -16.604 | 25.075 | y = -0.674x^2 + 70.05x + 102.548 |
| Wielomianowa 3 stopnia | 122.472 | 1.0 | 6.369 | -11.22 | 13.012 | y = -0.023x^3 + 0.178x^2 + 136.10x + 122.472 |

Model liniowy jest najmniej dokładny z największym RMSE i najszerszym zakresem reszt.

Model wielomianowy 2 stopnia jest znacznie dokładniejszy niż model liniowy, z wysokim R-squared i znacznie niższym RMSE.

Model wielomianowy 3 stopnia ma teoretycznie idealne dopasowanie z R-squared równym 1.0 i najniższym RMSE, co sugeruje, że jest najbardziej dokładny w dopasowaniu do danych treningowych.

Na podstawie dostarczonych tabel z danymi porównujemy rzeczywiste liczby użytkowników firmy Meta z wartościami przewidywanymi przez nasze modele.

Model liniowy:

|  |  |  |
| --- | --- | --- |
| Kwartał | Wartość rzeczywista | Wartość przewidywana |
| Q1 '2015 | 1441 | 1505,82 |
| Q2 '2015 | 1490 | 1559,03 |
| Q3 '2015 | 1545 | 1612,24 |
| Q4 '2015 | 1591 | 1665,46 |
| Q1 '2016 | 1654 | 1718,67 |
| Q2 '2016 | 1712 | 1771,88 |
| Q3 '2016 | 1788 | 1825,09 |
| Q4 '2016 | 1860 | 1878,30 |
| Q1 '2017 | 1936 | 1931,52 |
| Q2 '2017 | 2006 | 1984,73 |
| Q3 '2017 | 2072 | 2037,94 |
| Q4 '2017 | 2129 | 2091,15 |

Model wielomianowy 2 stopnia:

|  |  |  |
| --- | --- | --- |
| Kwartał | Wartość rzeczywista | Wartość przewidywana |
| Q1 '2015 | 1441 | 1432,85 |
| Q2 '2015 | 1490 | 1468,55 |
| Q3 '2015 | 1545 | 1502,91 |
| Q4 '2015 | 1591 | 1535,91 |
| Q1 '2016 | 1654 | 1567,57 |
| Q2 '2016 | 1712 | 1597,88 |
| Q3 '2016 | 1788 | 1626,85 |
| Q4 '2016 | 1860 | 1654,47 |
| Q1 '2017 | 1936 | 1680,74 |
| Q2 '2017 | 2006 | 1705,66 |
| Q3 '2017 | 2072 | 1729,24 |
| Q4 '2017 | 2129 | 1751,47 |

Model wielomianowy 3 stopnia:

|  |  |  |
| --- | --- | --- |
| Kwartał | Wartość rzeczywista | Wartość przewidywana |
| Q1 '2015 | 1441 | 1412,93 |
| Q2 '2015 | 1490 | 1439,07 |
| Q3 '2015 | 1545 | 1462,02 |
| Q4 '2015 | 1591 | 1481,65 |
| Q1 '2016 | 1654 | 1497,82 |
| Q2 '2016 | 1712 | 1510,39 |
| Q3 '2016 | 1788 | 1519,24 |
| Q4 '2016 | 1860 | 1524,21 |
| Q1 '2017 | 1936 | 1525,19 |
| Q2 '2017 | 2006 | 1522,02 |
| Q3 '2017 | 2072 | 1514,58 |
| Q4 '2017 | 2129 | 1502,72 |

Model liniowy konsekwentnie przewiduje wyższe wartości niż te rzeczywiste, co może wskazywać na pewną systematyczną przeszacowaną tendencję tego modelu. Wartości przewidywane rosną liniowo wraz z każdym kwartałem, co sugeruje stałą stopę wzrostu liczby użytkowników zgodnie z modelem.

Model wielomianowy drugiego stopnia wydaje się dawać wartości przewidywane, które są bliższe rzeczywistym, zwłaszcza w początkowych kwartałach analizowanego okresu. Jednak w miarę postępu czasu model ten zaczyna niedoszacowywać liczby użytkowników.

Model wielomianowy trzeciego stopnia, jest precyzyjny tylko w pierwszym kwartale 2015 i jego przewidywania stają się mniej dokładne wraz z upływem czasu. Oba modele wielomianowe niedoszacowują liczby użytkowników. W miarę postępu czasu wartości przewidywane bardziej odstają niż w modelu wielomianowym 2 stopnia.

Podsumowując, analiza trzech modeli regresji zastosowanych do prognozowania liczby użytkowników firmy Meta pokazuje, że model liniowy, choć najprostszy, wydaje się oszacować liczbę użytkowników najtrafniej.

Obraz zawierający linia, Wykres, tekst, stok

Opis wygenerowany automatycznie

Rys. 26 Liczba użytkowników serwisu „Meta” z portalu statista.com

W celu oceny zdolności naszego modelu do adaptacji do zmieniających się warunków rynkowych, przeprowadzimy analizę na rozszerzonym zestawie danych. Model zostanie wytrenowany na danych rzeczywistych obejmujących lata 2009-2017, kiedy to obserwowaliśmy ciągły wzrost liczby użytkowników. Następnie, użyjemy danych testowych z lat 2018-2022, aby zbadać, jak model radzi sobie w obliczu stagnacji, która miała miejsce od roku 2021. To połączenie danych pozwoli nie tylko na ocenę dokładności modelu w dłuższej perspektywie, ale także na zrozumienie, w jaki sposób przeszłe tendencje wpływają na przewidywania w kontekście nowych, nieoczekiwanych zmian na rynku.

Poniżej znajduje się zestawienie wyników dla poszczególnych modeli, przedstawiające kluczowe metryki statystyczne takie jak współczynnik korelacji, R-squared, RMSE oraz zakres reszt, które są niezbędne do oceny dokładności i wiarygodności każdego modelu. Dodatkowo, dla każdego modelu zostało przedstawione równanie regresji, co pozwala na wizualizację zależności między kwartałami a liczbą użytkowników.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Intercept | R^2 | RMSE | Min reszt | Max reszt | Równanie wykresu |
| Liniowa | 175.514 | 0.993 | 30.064 | -59.607 | 39.152 | y = 53.212x + 175.514 |
| Wielomianowa 2 stopnia | 173.853 | 0.996 | 36.417 | -58.306 | 63.631 | y = 0.019x^2 + 53.229x + 173.853 |
| Wielomianowa 3 stopnia | 72.036 | 0.999 | 16.297 | -28.087 | 42.855 | y= 0.037x^3 − 2.080x^2 + 84.15x + 72.036 |

Model wielomianowy 3 stopnia zapewnia najlepsze dopasowanie z najmniejszym błędem przewidywania. Jednakże, idealny współczynnik R^2 i niskie wartości RMSE mogą wskazywać na ryzyko przeuczenia, gdzie model jest zbytnio dostosowany do danych treningowych i może nie radzić sobie równie dobrze z nowymi, nieznajomymi danymi.

Model liniowy i wielomianowy 2 stopnia mają porównywalne wartości R^2, ale model liniowy ma niższe RMSE, co może wskazywać, że jest on lepszym wyborem dla prognozowania nowych danych, jeśli chcemy uniknąć przeuczenia.

Wartości reszt w każdym z modeli wskazują na potencjalne obszary, w których model może być poprawiony, aby zredukować błędy prognozowania, zwłaszcza w przypadkach, gdy błędy te są znaczące.

Na podstawie modelów z uwzględnieniem lat 2018-2022 z danymi porównujemy rzeczywiste liczby użytkowników firmy Meta z wartościami przewidywanymi przez nasze modele.

Model liniowy:

|  |  |  |
| --- | --- | --- |
| Kwartał | Wartość rzeczywista | Wartość przewidywana |
| Q1 '2018 | 2196 | 2121.68 |
| Q2 '2018 | 2234 | 2174.20 |
| Q3 '2018 | 2271 | 2226.72 |
| Q4 '2018 | 2320 | 2279.25 |
| Q1 '2019 | 2375 | 2331.77 |
| Q2 '2019 | 2414 | 2384.29 |
| Q3 '2019 | 2449 | 2436.81 |
| Q4 '2019 | 2498 | 2489.34 |
| Q1 '2020 | 2603 | 2541.86 |
| Q2 '2020 | 2701 | 2594.38 |
| Q3 '2020 | 2740 | 2646.91 |
| Q4 '2020 | 2797 | 2699.43 |
| Q1 '2021 | 2853 | 2751.95 |
| Q2 '2021 | 2895 | 2804.48 |
| Q3 '2021 | 2910 | 2857.00 |
| Q4 '2021 | 2912 | 2909.52 |
| Q1 '2022 | 2936 | 2962.04 |
| Q2 '2022 | 2934 | 3014.57 |
| Q3 '2022 | 2958 | 3067.09 |
| Q4 '2022 | 2963 | 3119.61 |

Model wielomianowy 2 stopnia:

|  |  |  |
| --- | --- | --- |
| Kwartał | Wartość rzeczywista | Wartość przewidywana |
| Q1 '2018 | 2196 | 2117.21 |
| Q2 '2018 | 2234 | 2169.00 |
| Q3 '2018 | 2271 | 2220.76 |
| Q4 '2018 | 2320 | 2272.49 |
| Q1 '2019 | 2375 | 2324.17 |
| Q2 '2019 | 2414 | 2375.81 |
| Q3 '2019 | 2449 | 2427.42 |
| Q4 '2019 | 2498 | 2478.99 |
| Q1 '2020 | 2603 | 2530.52 |
| Q2 '2020 | 2701 | 2582.01 |
| Q3 '2020 | 2740 | 2633.47 |
| Q4 '2020 | 2797 | 2684.89 |
| Q1 '2021 | 2853 | 2736.26 |
| Q2 '2021 | 2895 | 2787.60 |
| Q3 '2021 | 2910 | 2838.91 |
| Q4 '2021 | 2912 | 2890.17 |
| Q1 '2022 | 2936 | 2941.40 |
| Q2 '2022 | 2934 | 2992.58 |
| Q3 '2022 | 2958 | 3043.73 |
| Q4 '2022 | 2963 | 3094.84 |

Model wielomianowy 3 stopnia:

|  |  |  |
| --- | --- | --- |
| Kwartał | Wartość rzeczywista | Wartość przewidywana |
| Q1 '2018 | 2196 | 2219.02 |
| Q2 '2018 | 2234 | 2303.84 |
| Q3 '2018 | 2271 | 2392.97 |
| Q4 '2018 | 2320 | 2486.63 |
| Q1 '2019 | 2375 | 2585.03 |
| Q2 '2019 | 2414 | 2688.42 |
| Q3 '2019 | 2449 | 2797.00 |
| Q4 '2019 | 2498 | 2911.00 |
| Q1 '2020 | 2603 | 3030.65 |
| Q2 '2020 | 2701 | 3156.16 |
| Q3 '2020 | 2740 | 3287.77 |
| Q4 '2020 | 2797 | 3425.68 |
| Q1 '2021 | 2853 | 3570.13 |
| Q2 '2021 | 2895 | 3721.34 |
| Q3 '2021 | 2910 | 3879.53 |
| Q4 '2021 | 2912 | 4044.92 |
| Q1 '2022 | 2936 | 4217.74 |
| Q2 '2022 | 2934 | 4398.21 |
| Q3 '2022 | 2958 | 4586.54 |
| Q4 '2022 | 2963 | 4782.98 |

W powyższym przykładzie mamy do czynienia ze zmianą trendu co miało wpływ na prognozy.

Początkowo model linowy dobrze dopasowywał się do danych lecz w momencie stagnacji model przewidywał kontynuację wzrostu, mimo że wartości rzeczywiste wykazują znacznie wolniejszy wzrost.

Model wielomianowy 2 stopnia jest w stanie lepiej dostosować swoje prognozy do tej zmiany trendu, ponieważ ma większą elastyczność w dopasowywaniu nieliniowych wzorców w danych. Wartości przewidywane przez model wielomianowy 2. stopnia zaczynają lepiej oddawać nieliniowy charakter danych. Błędy prognoz wydają się być bardziej zrównoważone w porównaniu z modelem liniowym.

Model wielomianowy 3 stopnia wydaje się nadmiernie dopasować do danych, przewidując wartości, które szybko rosną, co sugeruje ryzyko przeuczenia. Ten model wykazuje wyjątkowe dopasowanie do danych treningowych z R^2=0,999, ale wartości przewidywane są znacznie wyższe niż wartości rzeczywiste.

**Wnioski ogólne:**

Według naszej analizy okazało się, że w okresie stałego wzrostu użytkowników (chodzi nam o lata 2009-2020) model liniowy wydawał się być adekwatny, dostarczając wiarygodnych prognoz. Jednakże, gdy wzrost użytkowników uległ stagnacji model liniowy przestał być efektywny, ponieważ kontynuował trend wzrostowy bez uwzględnienia zmian w danych. W takiej sytuacji, model wielomianowy drugiego stopnia okazał się lepszym rozwiązaniem, ponieważ jego elastyczność pozwoliła na lepsze dostosowanie do spłaszczenia krzywej wzrostu.

Z drugiej strony, model wielomianowy trzeciego stopnia, mimo że wydawał się przewidywać dane z najwyższą dokładnością, w rzeczywistości był najmniej odpowiedni. Wynikało to z jego tendencji do nadmiernego dopasowania, co manifestowało się w znacznym przeszacowaniu danych, w dłuższym horyzoncie czasowym. Taka charakterystyka modelu wielomianowego trzeciego stopnia może prowadzić do nierealistycznych prognoz w przyszłości, jeśli trend danych ulegnie zmianie, co czyni go mniej przydatnym w zmiennych warunkach rynkowych.

Wybór odpowiedniego modelu regresji zależy od charakterystyki danych i kontekstu, w jakim są one używane. Model liniowy może być preferowany dla danych o stałym wzroście, natomiast model wielomianowy drugiego stopnia jest bardziej elastyczny i może lepiej radzić sobie w sytuacjach, gdzie wzrost jest niestabilny lub gdy występują okresy stagnacji. Model wielomianowy trzeciego stopnia, pomimo dobrej dokładności na danych treningowych, może nie być odpowiedni dla przewidywania przyszłych trendów ze względu na ryzyko przeuczenia i przeszacowania.

## Chcemy zbadać zależność między przychodem, a zatrudnieniem w firmie „Meta”.

## Model regresji liniowej na podstawie danych z tabeli

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Rys. 27 Model regresji liniowej z uwzględnieniem lat 2007 – 2017 – problem nr 2

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Rys. 28 Współczynniki dla modelu regresji liniowej z uwzględnieniem lat 2007 – 2017 – problem nr 2

## Model regresji liniowej z uwzględnieniem lat 2018-2020

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Rys. 29 Model regresji liniowej z uwzględnieniem lat 2018 – 2020 – problem nr 2

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Rys. 30 Współczynniki dla modelu regresji liniowej z uwzględnieniem lat 2018 – 2020 – problem nr 2

## Model regresji liniowej z uwzględnieniem lat 2021-2022

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Rys. 31 Model regresji liniowej z uwzględnieniem lat 2021 – 2022 – problem nr 2

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Rys. 32 Współczynniki dla modelu regresji liniowej z uwzględnieniem lat 2021 – 2022 – problem nr 2

## Model regresji liniowej z uwzględnieniem lat 2018-2022

Obraz zawierający tekst, zrzut ekranu, linia, Wykres

Opis wygenerowany automatycznie

Rys. 33 Model regresji liniowej z uwzględnieniem lat 2018 – 2022 – problem nr 2

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Rys. 34 Współczynniki dla modelu regresji liniowej z uwzględnieniem lat 2018 – 2022 – problem nr 2

## Model regresji wielomianowej 2. stopnia na podstawie danych z tabeli

Obraz zawierający tekst, zrzut ekranu, diagram, linia

Opis wygenerowany automatycznie

Rys. 35 Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2007 – 2017 – problem nr 2

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Rys. 36 Współczynniki dla modelu regresji wielomianowej 2. stopnia z uwzględnieniem lat 2007 – 2017 – problem nr 2

## Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2018-2020

Obraz zawierający tekst, zrzut ekranu, diagram, Wykres

Opis wygenerowany automatycznie

Rys. 37 Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2018 – 2020 – problem nr 2

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Rys. 38 Współczynniki dla modelu regresji wielomianowej 2. stopnia z uwzględnieniem lat 2018 – 2020 – problem nr 2

## Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2021-2022

Obraz zawierający tekst, zrzut ekranu, linia, Wykres

Opis wygenerowany automatycznie

Rys. 39 Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2021 – 2022 – problem nr 2

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Rys. 40 Współczynniki dla modelu regresji wielomianowej 2. stopnia z uwzględnieniem lat 2021 – 2022 – problem nr 2

## Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2018-2022

Obraz zawierający tekst, zrzut ekranu, Wykres, diagram

Opis wygenerowany automatycznie

Rys. 41 Model regresji wielomianowej 2. stopnia z uwzględnieniem lat 2018 – 2022 – problem nr 2

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Rys. 42 Współczynniki dla modelu regresji wielomianowej 2. stopnia z uwzględnieniem lat 2018 – 2022 – problem nr 2

## Analiza wyników problemu nr 2:

W celu zbadania zależności między przychodem a zatrudnieniem w firmie "Meta", podzielono dostępne dane na dwa zestawy: treningowe i testowe. Dane z lat 2007-2014, zaznaczone kolorem żółtym, zostaną wykorzystane do zbudowania modelu, który będzie próbował uchwycić wzorce i zależności między zatrudnieniem a generowanym przychodem. Następnie, dane z lat 2015-2017, oznaczone kolorem czerwonym, posłużą do przetestowania tego modelu, czyli sprawdzenia, jak dobrze model prognozuje przychody firmy na podstawie danych o zatrudnieniu, które nie były wykorzystane podczas procesu uczenia.

|  |  |  |
| --- | --- | --- |
| Rok | Przychód w mln $ | Zatrudnienie |
| 2007 | 153 | 450 |
| 2008 | 272 | 850 |
| 2009 | 777 | 1218 |
| 2010 | 1974 | 2127 |
| 2011 | 3711 | 3200 |
| 2012 | 5089 | 4619 |
| 2013 | 7872 | 6337 |
| 2014 | 12466 | 9199 |
| 2015 | 17928 | 12691 |
| 2016 | 27638 | 17048 |
| 2017 | 40653 | 25105 |

Trzeba zaznaczyć że, model nie bierze pod uwagę czasu jako zmiennej. Zakłada się, że zależność między przychodem a zatrudnieniem jest jedynie bezpośrednia, bez wpływu trendów czasowych. Przy takim założeniu, modelowanie skupia się na obecnej korelacji między tymi dwiema zmiennymi, ignorując ewolucję tej relacji w czasie.

Porównanie modelu liniowego i wielomianowego drugiego stopnia:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Intercept | R^2 | RMSE | Min reszt | Max reszt |
| Liniowa | 563,039 | 0,994 | 154,132 | -227,587 | 245,947 |
| Wielomianowa 2 stopnia | 4973,4488 | 0,995 | 140,862 | -225,638 | 179,631 |

R^2 - oba modele mają bardzo wysokie wartości R^2, co sugeruje, że zarówno model liniowy, jak i wielomianowy dobrze dopasowują się do danych treningowych. Model wielomianowy ma minimalnie wyższą wartość R^2, co wskazuje na nieco lepsze dopasowanie.

RMSE (Błąd średniokwadratowy): Model wielomianowy ma niższe RMSE niż model liniowy, co oznacza, że średni błąd prognoz jest mniejszy dla modelu wielomianowego. To sugeruje, że model wielomianowy może być lepszym wyborem do prognozowania na podstawie dostępnych danych.

Odchylenie standardowe reszt: Wartości są identyczne z RMSE, co jest typowe, gdy RMSE jest obliczane jako pierwiastek kwadratowy z wariancji reszt.

Reszty (min i max): Maksymalna reszta jest niższa dla modelu wielomianowego, co wskazuje, że największy błąd przewidywania jest mniejszy niż w modelu liniowym. To może wskazywać na lepszą zdolność modelu wielomianowego do radzenia sobie z ekstremalnymi wartościami.

Porównanie rzeczywistych i przewidywanych wartości zatrudnienia w firmie 'Meta' przy użyciu modeli regresji liniowej i wielomianowej

Model regresji liniowej:

|  |  |
| --- | --- |
| Wartości rzeczywiste | Wartości przewidywane |
| 9199 | 9896.037 |
| 12691 | 13985.306 |
| 17048 | 21254.953 |
| 25105 | 30998.974 |

Model regresji wielomianowej 2 stopnia:

|  |  |
| --- | --- |
| Wartości rzeczywiste | Wartości przewidywane |
| 9199 | 9188.086 |
| 12691 | 11967.695 |
| 17048 | 15362.125 |
| 25105 | 16806.136 |

Model Liniowy:

- Wszystkie przewidywane wartości są wyższe od wartości rzeczywistych, co wskazuje na systematyczne przeszacowanie przez model liniowy.

- Różnica między wartościami rzeczywistymi a przewidywanymi wzrasta wraz ze wzrostem zatrudnienia, co może sugerować, że model liniowy nie jest w stanie uchwycić złożoności związku między przychodem a zatrudnieniem, szczególnie przy większych liczbach zatrudnienia.

Model Wielomianowy 2 Stopnia:

Wartości przewidywane przez model wielomianowy są bliższe wartościom rzeczywistym, zwłaszcza dla niższego zatrudnienia, co wskazuje na lepsze dopasowanie modelu do danych.

Mimo że model wielomianowy również niedoszacowuje zatrudnienie, robi to w mniejszym stopniu niż model liniowy, zwłaszcza przy większej liczbie zatrudnionych.

Teraz przejdziemy do etapu predykcji, gdzie użyjemy danych z lat 2007-2017 jako nasz zestaw testowy do oceny modelu. Dane te są zaznaczone na żółto w tabeli. Następnie, wykorzystamy model do przewidzenia zatrudnienia dla lat 2018-2020, które są zaznaczone na czerwono jako dane testowe, i porównamy te prognozy z rzeczywistymi wartościami zatrudnienia w tych latach. To pozwoli na ocenę, jak dobrze model radzi sobie z przewidywaniem danych, które nie były użyte do jego nauczenia.

|  |  |  |
| --- | --- | --- |
| Rok | Przychód w mln $ | Zatrudnienie |
| 2007 | 153 | 450 |
| 2008 | 272 | 850 |
| 2009 | 777 | 1218 |
| 2010 | 1974 | 2127 |
| 2011 | 3711 | 3200 |
| 2012 | 5089 | 4619 |
| 2013 | 7872 | 6337 |
| 2014 | 12466 | 9199 |
| 2015 | 17928 | 12691 |
| 2016 | 27638 | 17048 |
| 2017 | 40653 | 25105 |
| 2018 | 55838 | 35587 |
| 2019 | 70697 | 44942 |
| 2020 | 85965 | 58604 |

Tabeli porównującej metryki dla modelu regresji liniowej i regresji wielomianowej 2 stopnia.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Intercept | R^2 | RMSE | Min reszt | Max reszt |
| Liniowa | 1074.63 | 0.995 | 520.661 | -716.306 | 874.201 |
| Wielomianowa 2 stopnia | 748,442 | 0.997 | 377.478 | -935.527 | 408.998 |

Porównanie rzeczywistych i przewidywanych wartości zatrudnienia w firmie 'Meta' przy użyciu modeli regresji liniowej i wielomianowej

Model regresji liniowej:

|  |  |
| --- | --- |
| Wartości rzeczywiste | Wartości przewidywane |
| 35587 | 34532 |
| 44942 | 43435 |
| 58604 | 52583 |

Model regresji wielomianowej 2 stopnia:

|  |  |
| --- | --- |
| Wartości rzeczywiste | Wartości przewidywane |
| 35587 | 31616 |
| 44942 | 37193 |
| 58604 | 41769 |

Porównując wartości rzeczywiste z przewidywanymi, model liniowy wydaje się lepiej radzić sobie z przewidywaniem dla wyższych wartości zatrudnienia (58604 rzeczywiste wobec 52583 przewidziane), podczas gdy model wielomianowy znacząco niedoszacowuje te wartości (58604 rzeczywiste wobec 41769 przewidziane).

W przypadku niższych wartości zatrudnienia, różnice między modelami są mniej wyraźne, ale model liniowy nadal wydaje się być bliższy rzeczywistym wartościom.

Teraz przejdziemy do etapu predykcji, gdzie użyjemy danych z lat 2007-2020 jako nasz zestaw testowy do oceny modelu. Dane te są zaznaczone na żółto w tabeli. Następnie, wykorzystamy model do przewidzenia zatrudnienia dla lat 2021-2022, które są zaznaczone na czerwono jako dane testowe, i porównamy te prognozy z rzeczywistymi wartościami zatrudnienia w tych latach. To pozwoli na ocenę, jak dobrze model radzi sobie z przewidywaniem danych, które nie były użyte do jego nauczenia.

|  |  |  |
| --- | --- | --- |
| Rok | Przychód w mln $ | Zatrudnienie |
| 2007 | 153 | 450 |
| 2008 | 272 | 850 |
| 2009 | 777 | 1218 |
| 2010 | 1974 | 2127 |
| 2011 | 3711 | 3200 |
| 2012 | 5089 | 4619 |
| 2013 | 7872 | 6337 |
| 2014 | 12466 | 9199 |
| 2015 | 17928 | 12691 |
| 2016 | 27638 | 17048 |
| 2017 | 40653 | 25105 |
| 2018 | 55838 | 35587 |
| 2019 | 70697 | 44942 |
| 2020 | 85965 | 58604 |
| 2021 | 117929 | 71970 |
| 2022 | 116609 | 86482 |

Tabeli porównującej metryki dla modelu regresji liniowej i regresji wielomianowej 2 stopnia.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Intercept | R^2 | RMSE | Min reszt | Max reszt |
| Liniowa | 624.632 | 0.996 | 1061.283 | -1705.924 | 2605.728 |
| Wielomianowa 2 stopnia | 1261.62271 | 0.998 | 732.395 | -1291.804 | 1234.978 |

Model wielomianowy wykazuje niższy RMSE (732.395) w porównaniu z modelem liniowym (1061.283), co sugeruje, że model wielomianowy ma tendencję do dokładniejszego przewidywania zatrudnienia na danych treningowych.

Porównanie rzeczywistych i przewidywanych wartości zatrudnienia w firmie 'Meta' przy użyciu modeli regresji liniowej i wielomianowej

Model regresji liniowej:

|  |  |
| --- | --- |
| Wartości rzeczywiste | Wartości przewidywane |
| 71970 | 76588 |
| 86482 | 75737 |

Model regresji wielomianowej 2 stopnia:

|  |  |
| --- | --- |
| Wartości rzeczywiste | Wartości przewidywane |
| 71970 | 83404 |
| 86482 | 82287 |

Patrząc na przewidywane wartości zatrudnienia dla lat 2021-2022, oba modele wykazują pewne rozbieżności w stosunku do wartości rzeczywistych. Model liniowy przewiduje wartości zatrudnienia niższe od rzeczywistych dla roku 2021 i wyższe dla roku 2022. Model wielomianowy z kolei znacząco przeszacowuje zatrudnienie dla roku 2021 i niedoszacowanie dla roku 2022.

Teraz, mając szerszy zakres danych, przejdziemy do bardziej rozbudowanej analizy. W tabeli przedstawiono dane za lata 2007-2022, z których te z lat 2018-2020, zaznaczone kolorem żółtym, zostaną użyte jako dane testowe, a lata 2021-2022, zaznaczone kolorem czerwonym, posłużą jako dodatkowy zbiór do testowania. Powiększenie zestawu danych testowych pozwoli na dokładniejszą ocenę wydajności i zdolności predykcyjnych naszych wytworzonych modeli regresji.

|  |  |  |
| --- | --- | --- |
| Rok | Przychód w mln $ | Zatrudnienie |
| 2007 | 153 | 450 |
| 2008 | 272 | 850 |
| 2009 | 777 | 1218 |
| 2010 | 1974 | 2127 |
| 2011 | 3711 | 3200 |
| 2012 | 5089 | 4619 |
| 2013 | 7872 | 6337 |
| 2014 | 12466 | 9199 |
| 2015 | 17928 | 12691 |
| 2016 | 27638 | 17048 |
| 2017 | 40653 | 25105 |
| 2018 | 55838 | 35587 |
| 2019 | 70697 | 44942 |
| 2020 | 85965 | 58604 |
| 2021 | 117929 | 71970 |
| 2022 | 116609 | 86482 |

Tabeli porównującej metryki dla modelu regresji liniowej i regresji wielomianowej 2 stopnia.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Intercept | R^2 | RMSE | Min reszt | Max reszt |
| Liniowa | 1074.63 | 0.995 | 520.661 | -716.306 | 874.201 |
| Wielomianowa 2 stopnia | 748.425 | 0.997 | 377.478 | -935.527 | 408.998 |

Oba modele mają wysokie wartości R kwadrat, co wskazuje na silne dopasowanie do danych treningowych. Model wielomianowy (R² = 0.997) wydaje się nieznacznie lepiej dopasowany niż model liniowy (R² = 0.995).

Model wielomianowy ma niższy RMSE (377.478) niż model liniowy (520.661), co sugeruje większą dokładność modelu wielomianowego na danych treningowych.

Porównanie rzeczywistych i przewidywanych wartości zatrudnienia w firmie 'Meta' przy użyciu modeli regresji liniowej i wielomianowej

Model regresji liniowej:

|  |  |
| --- | --- |
| Wartości rzeczywiste | Wartości przewidywane |
| 35587 | 34532 |
| 44942 | 43435 |
| 58604 | 52583 |
| 71970 | 71736 |
| 86482 | 70945 |

Model regresji wielomianowej 2 stopnia:

|  |  |
| --- | --- |
| Wartości rzeczywiste | Wartości przewidywane |
| 35587 | 31616 |
| 44942 | 37193 |
| 58604 | 41769 |
| 71970 | 47558 |
| 86482 | 47421 |

Model liniowy lepiej radzi sobie z przewidywaniem wartości zatrudnienia dla lat 2021-2022, z przewidywaniami bliższymi do rzeczywistych wartości niż model wielomianowy. Model wielomianowy drugiego stopnia wskazuje, że przy dużych wartościach przychodu jest on mało użyteczny do przewidywania zatrudnienia. Chociaż model ten wykazywał lepsze dopasowanie do danych treningowych niż model liniowy, to jednak w praktycznym zastosowaniu na danych testowych, gdzie wartości przychodów były wyższe, model wielomianowy generował prognozy, które znacząco odbiegały od rzeczywistego zatrudnienia.

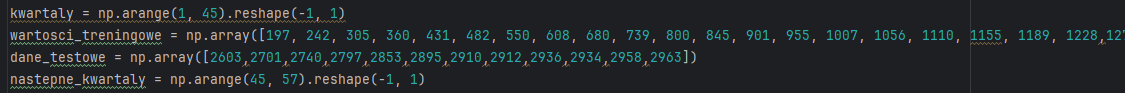
## Opis fragmentów kodu (python):

Importowanie biblioteki dostarczającej klasy i funkcje dla estymacji różnych modeli statystycznych.



Rys. 43 Importowanie biblioteki statsmodel.api

Podział na dane testowe i treningowe.



Rys. 44 Dane testowe i treningowe

sm.OLS() – to metoda pozwalająca nam utworzyć model regresji liniowej metodą najmniejszych kwadratów. Dzięki tej metodzie jesteśmy w stanie znaleźć wartości współczynników będące najlepszym przybliżeniem zależności między X oraz Y.



Rys. 45 Metoda tworząca model regresji liniowej metodą NK

Funkcja sm.add\_constant() służy do dodawania stałej (intercept) do macierzy zmiennych niezależnych. Funkcja dodaje do macierzy kwartały kolumnę zawierającą jedynki, co pozwala na uwzględnienie wartości początkowej (intercept) w modelu regresji liniowej.



Rys. 46 Macierz zmiennych niezależnych - kwartały



Rys. 47 Macierz zmiennych niezależnych - następne kwartały

Całość służy do uzyskania przedziału ufności dla prognoz modelu liniowego.



Rys. 48 Tworzenie przedziału ufności

Służy do generowania prognoz.



Rys. 49 Generowanie prognoz dla następnych kwartałów

Ta metoda jest używana, aby uzyskać przedziały ufności dla przewidywanych wartości.



Rys. 50 Przedziały ufności dla przewidywanych wartości

Biblioteka pozwalająca wygenerować wykresy.



Rys. 51 Importowanie biblioteki matplotlib.pyplot do tworzenia wykresów

Tutaj kod odpowiada za narysowanie wykresu żeby zobrazować nasz model.

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Rys. 52 Tworzenie wykresu

Inne biblioteki do tworzenia modelu oraz obliczeń. Wykorzystaliśmy specjalnie kilka by sprawdzić poprawność oraz dokładność obliczeń każdej z nich oraz zestawiliśmy je również z modelem stworzonym w języku R w RStudio:

Obraz zawierający tekst, Czcionka, zrzut ekranu, linia

Opis wygenerowany automatycznie

Rys. 53 Importowanie bibliotek sklearn

Metoda score() oblicza współczynnik determinacji (R-kwadrat) dla regresji liniowej w bibliotece sklearn.



Rys. 54 Obliczanie współczynnika determinacji

Umożliwia manipulację danymi numerycznymi, wykonywanie obliczeń matematycznych, operacji na macierzach, przetwarzania danych i statystyki.



Rys. 55 Importowanie biblioteki numpy

np.mean() oblicza średnią z kwadratów tych różnic, co daje błąd średniokwadratowy (MSE).



Rys. 56 Obliczanie średniej z kwadratów różnic

Tutaj liczymy paramtery – Coef , wyraz wolny – Intercept oraz RMSE - pierwiastek błędu średniokwadratowego policzony na podstawie wcześniej obliczonego błędu średniokwadratowego.

Obraz zawierający tekst, Czcionka, zrzut ekranu

Opis wygenerowany automatycznie

Rys. 57 Liczenie współczynników coef, intercept oraz RMSE

Różnica między rzeczywistymi wartościami, a wartościami przewidywanymi przez model co zostaje podane jako argument, aby policzyć odchylenie standardowe reszt, czyli miarę rozproszenia reszt wokół średniej wartości.



Rys. 58 Liczenie odchylenia standardowego reszt

Dzięki temu otrzymujemy najmniejszą i największą różnicę pomiędzy rzeczywistą wartością, a przewidywaną przez nasz model - czyli jakie jest największe niedoszacowanie oraz przeszacowanie.



Rys. 59 Minium oraz maksimum reszt

Bibliografia:

[0] dr inż. Romuald Hoffmann, prof. WAT, Notatki dla studentów, Warszawa 2023.

[1] Larose D. T., Metody i modele eksploracji danych, Wydawnictwo Naukowe PWN, Warszawa 2012.

[2] <https://scikit-learn.org/stable/>

[3] Fred Nwanganga, Mike Chapple, Praktyczne uczenie maszynowe w języku R, Wiley, Warszawa, 2022.

[4] Laurence Moroney, Sztuczna inteligencja i uczenie maszynowe dla programistów, Helion O’Reilly, 2021.

[5] <https://pl.wikipedia.org/wiki/Wikipedia:Strona_g%C5%82%C3%B3wna>

[6] Aurelie Geron, Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow, Helion O’Reilly, 2018