

NETFLIX – PROGNOZA ZA POMOCĄ ARIMA

1. Opis przedsiębiorstwa

Netflix jest serwisem VOD (Video On Demand – wideo na żądanie), który oferuje za zryczałtowaną opłatą dostęp do filmów i seriali poprzez media strumieniowe. Usługa jest dostępna na całym świecie z kilkoma wyjątkami – Chiny, Korea Północna, Syria czy Krym; w Polsce wystartował jesienią 2016 roku. Poszczególne wersje regionalne różnią się między sobą ofertą, co jest spowodowane ograniczeniami licencyjnymi i umowami, które Netflix zawierał ze stacjami telewizyjnymi.

Na koniec 2021 roku serwis miał 221,8 mln subskrybentów. W I i II kwartale 2022 roku, od ponad 10 lat, Netflix odnotował spadek liczby subskrybentów – łącznie 1,170 mln. Koniec roku 2022 przyniósł ze sobą wzrost liczby subskrybentów do 223 mln. Związane jest to najpewniej z wyczekiwanymi premierami, takimi jak druga część filmu „Na noże”, czy serialu „Wednesday”. Zawirowania związane z liczbą subskrybentów miały wpływ na notowania akcji serwisu, które spadły o prawie 70% w 2022 roku. Przychody za I, II i III kwartał wynosiły odpowiednio 7,87 mld USD, 7,97 mld USD oraz 7,93 mld USD.

Powodów tak dużych spadków jest kilka. Pierwszym z nich jest wycofanie się z Rosji - zawieszenie świadczenia usług sprawiło, że Netflix stracił 700 tys. subskrybentów. Drugi powód, to konkurencja ze strony innych serwisów (HBO, Disney). Trzeci powód stagnacji to, zdaniem Netflixsa, współdzielenie hasła przez znaczną część klientów. Mowa o ponad 100 mln profili dzielonych poza gospodarstwem domowym. Serwis postanowił wprowadzić w 2023 roku tańszą alternatywę - czyli abonament z reklamami. Spotkało się to z dużym niezadowoleniem odbiorców, co poskutkowało kolejnymi spadkami na giełdzie. Również wraz z początkiem tego roku, Netflix opublikował instrukcję dotyczącą zasad współdzielenia konta w serwisie. Z jej treści wynika, że jedno konto mogą "współdzielić osoby mieszkające w jednym gospodarstwie domowym".

2. Opis użytego modelu – ARIMA

ARIMA stanowi skrót od ‘Autoregressive integrated moving average model’, czyli: autoregresyjny zintegrowany model średniej ruchomej.

Składa się z trzech elementów:

- Proces autoregresyjny AR. Jest to proces, w którym każda wartość jest liniową kombinacją poprzednich wartości. Wykorzystywana jest pamięć procesu. Procesy autoregresyjne mogą być różnego rzędu, który pokazuje ile wcześniejszych wartości ma wpływ na bieżącą wartość. Procesy autoregresyjne oznacza się symbolem $AR(p)$, w którym p jest rzędem autoregresji.
- Proces średniej ruchomej MA. Wartość szeregu zależy od zaburzeń w chwili obecnej i wcześniejszych. Procesy te oznacza się $MA(q)$, gdzie q jest rzędem procesu MA. Proces MA jest średnią ważoną z ostatnich zaburzeń, składników losowych modelu.

- Stopień integracji – $I(d)$. Jeżeli źródłem niestacjonarności szeregu liniowego jest trend, możemy pozbyć się niestacjonarności poprzez integrację szeregu.

Metoda ARIMA zawiera poszczególne etapy wykonania analizy: identyfikację, estymację i diagnozę:

- Identyfikacja liczby oraz typu parametrów modelu ARIMA, które następnie zostają oszacowane. W celu dokonania identyfikacji, należy posłużyć się wykresami szeregów oraz autokorelacją. Głównym warunkiem procesu identyfikacji dla metody ARIMA jest stacjonarność wejściowego szeregu. Powinien mieć stałą w czasie średnią, wariancję oraz autokorelację. Za stacjonarność szeregu odpowiada parametr (d) . Określając liczbę parametrów modelu zwracamy szczególną uwagę na korelogramy autokorelacji (ACF) oraz autokorelacji cząstkowej (PACF). Model ARIMA może być stosowany wówczas, gdy szereg jest stacjonarny i wejściowy zbiór danych zawiera minimum 50 obserwacji. Aby osiągnąć stacjonarność szeregu, należy poddać go różnicowaniu, a następnie przeanalizować wykres danych oraz korelogram. Przy doborze parametrów modelu sezonowego ARIMA, posługujemy się analizą ACF oraz PACF jak w modelu prostym ARIMA.
- Estymacja parametrów – stosuje się metodę najmniejszych kwadratów.
- Diagnozy – ocena wybranego modelu. Na podstawie znanych części danych ocenia się trafność wygenerowanej prognozy. Podczas diagnozy możemy sprawdzić również reszty przy pomocy (ACF) i (PACF). Przy dobrze dobranym modelu, funkcje nie powinny przyjmować wartości różnych od zera.

3. Opis badanych danych

W celu wyznaczenia prognoz cen akcji dla spółki Netflix, posłużono się cenami zamknięcia z okresu od 10 stycznia 2013 roku do 10 stycznia 2023 roku. Szereg czasowy składa się z cen zamknięcia akcji na każdy dzień tygodnia roboczego, co stanowi 2517 obserwacji. Dane pozyskano ze strony yahoo finance.

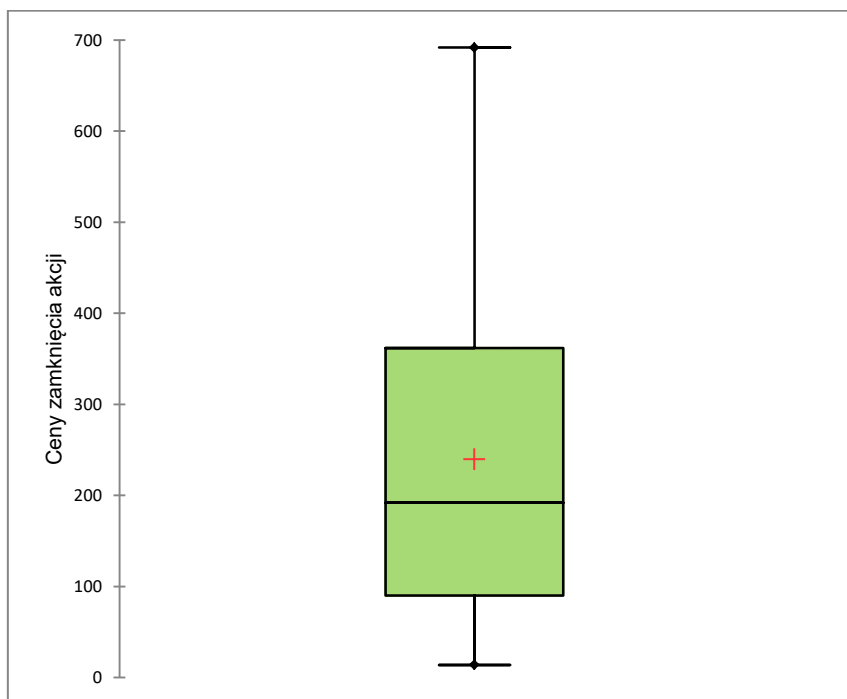
Minimum	13,93
Maximum	691,69
Średnia arytmetyczna	239,73
I Kwartyl	90,01
Mediana	192,12
III Kwartyl	361,81
Skośność	0,57
Kurtoza	-0,82

Tabela 1 Podstawowe statystyki

W tabeli powyżej znajdują się podstawowe statystyki wykorzystywanego szeregu czasowego. Najmniejsza cena zamknięcia akcji wystąpiła dnia 16 stycznia 2013 roku i wynosiła 13,93 dolary, z kolei najwyższa cena wystąpiła dnia 17 listopada 2021 roku

i wynosiła 691,69 dolary. Średnia wynosiła 239,73 dolarów, co oznacza, że przeciętnie w badanym okresie, cena zamknięcia akcji wynosiła 239,73 dolarów. Mediana to 192,12 dolarów, co oznacza, że co najmniej 50% obserwacji jest mniejszych lub równych od 192,12 dolarów i jednocześnie co najmniej 50% obserwacji jest większych lub równych od 192,12 dolarów.

Współczynnik skośności Pearsona wynosi 0,57, co świadczy o prawostronnej asymetrii. Większa część populacji przyjmuje wartości poniżej przeciętnej. Współczynnik kurtozy wynosi -0,82. Mamy do czynienia z rozkładem platykurtycznym, intensywność wartości skrajnych jest mniejsza niż w przypadku rozkładu normalnego.

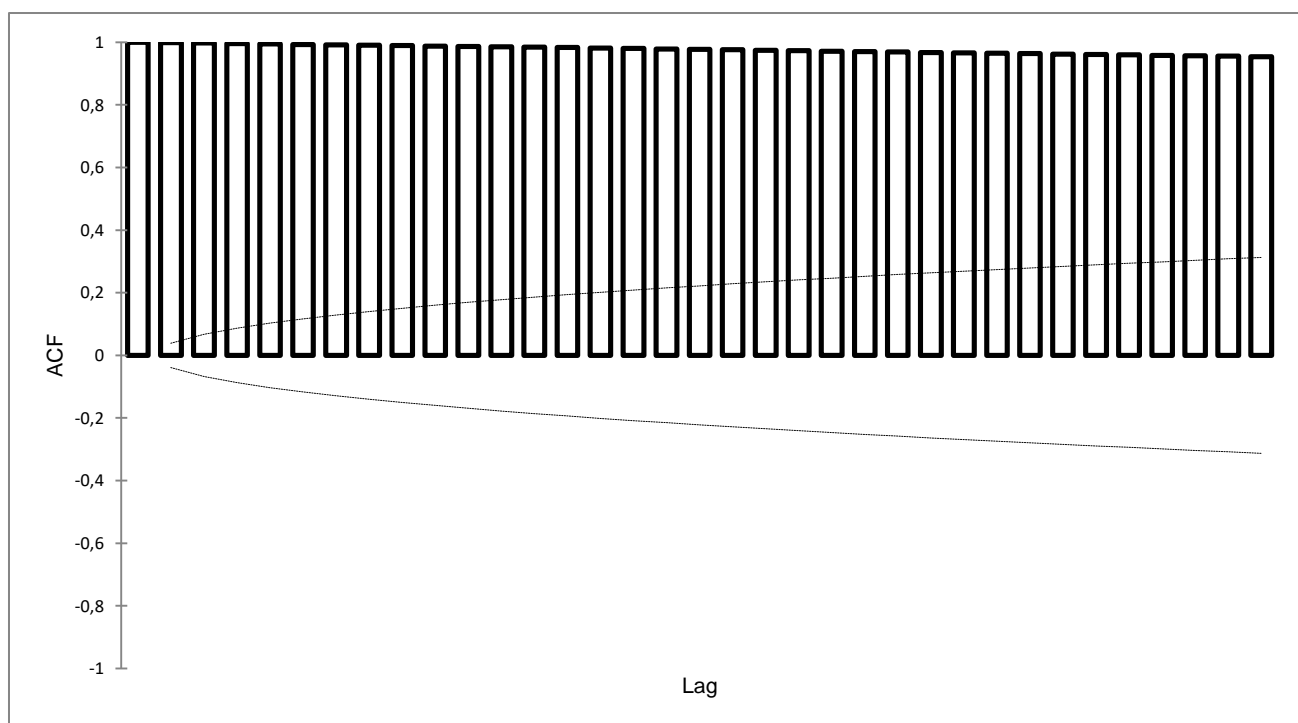


Rysunek 1 Wykres typu pudełkowego cen zamknięcia akcji firmy Netflix

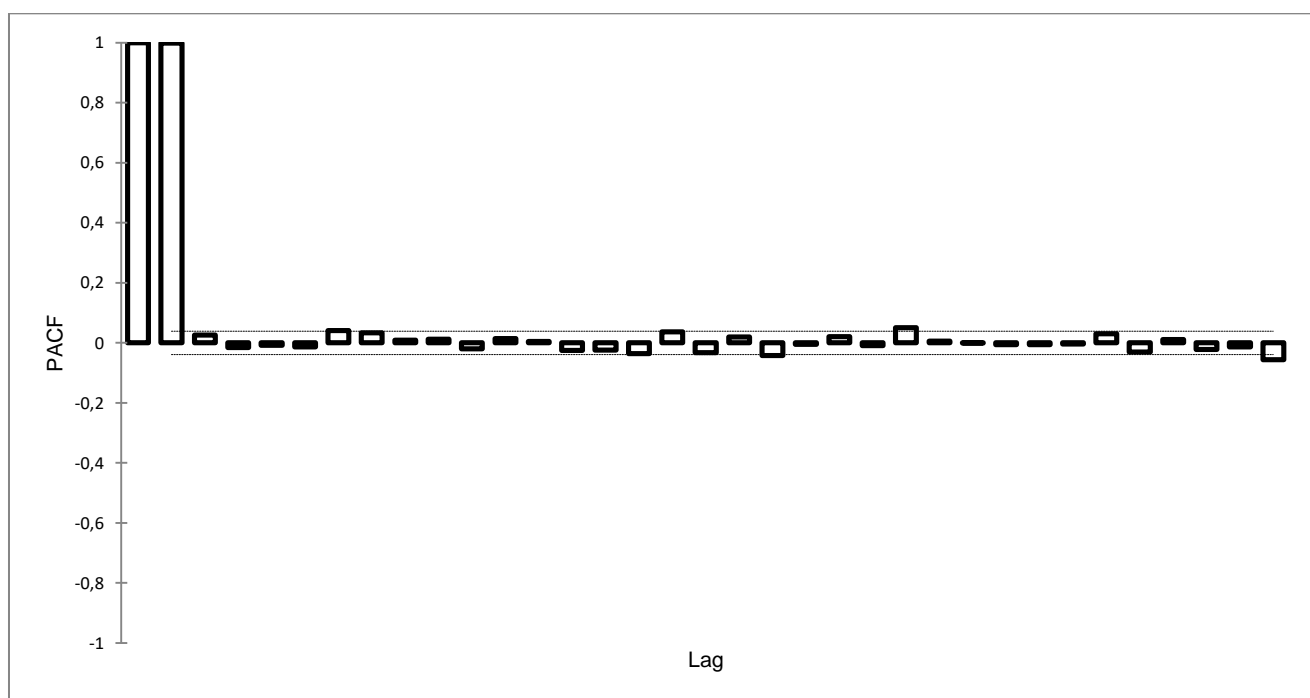


Rysunek 2 Wykres cen zamknięcia akcji firmy Netflix

Powyższy wykres charakteryzuje się trendem rosnącym. Największe spadki zanotowano w końcu roku 2021.



Rysunek 3 Wykres ACF



Rysunek 4 Wykres PACF

Na rysunkach nr 3 oraz nr 4 przedstawiono wykresy autokorelacji reszt ACF oraz PACF.

ACF przyjmuje wartości dodatnie i zanika bardzo powoli, co sugeruje, że w danych występuje silny trend wzrostowy. Potwierdza to także duża (bliska 1) wartość PACF(1). Zauważono brak występowania sezonowości w analizowanych danych, a brak wyraźnego punktu odcięcia sugeruje, że modelem który należy zastosować jest model ARIMA.

Za pomocą testu Mann'a – Kendall'a zbadano czy badany wektor danych charakteryzuje się trendem rosnącym czy malejącym:

Kendall's tau	0,753
S	2383366
Var(S)	1772824233,333
p-value (Two-tailed)	<0,0001
alpha	0,05

Tabela 2 Wyniki testu Mann'a – Kendall'a

Hipoteza zerowa mówi o braku trendu w badanym wektorze danych. Wyżej przedstawione wyniki wskazują na to, że należy odrzucić hipotezę zerową - badane dane mają tendencję wzrostową.

Dickey-Fuller test (ADF (stationary) / k = 13)	
Tau (Observed value)	-1,674
Tau (Critical value)	-3,395
p-value (one-tailed)	0,758
alpha	0,05

Tabela 3 Wyniki testu ADF

W celu zbadania rzędu zintegrowania szeregu czasowego, zastosowano rozszerzony test Dickey'a-Fullera (ADF). Wartość p wynosząca 0,758 oznacza, że nie ma podstaw do odrzucenia hipotezy zerowej, mówiącej o tym, że szereg posiada pierwiastek jednostkowy. Możemy uznać, że badany szereg jest niestacjonarny.

Eta (Observed value)	16,402
Eta (Critical value)	0,463
p-value (one-tailed)	<0,0001
alpha	0,05

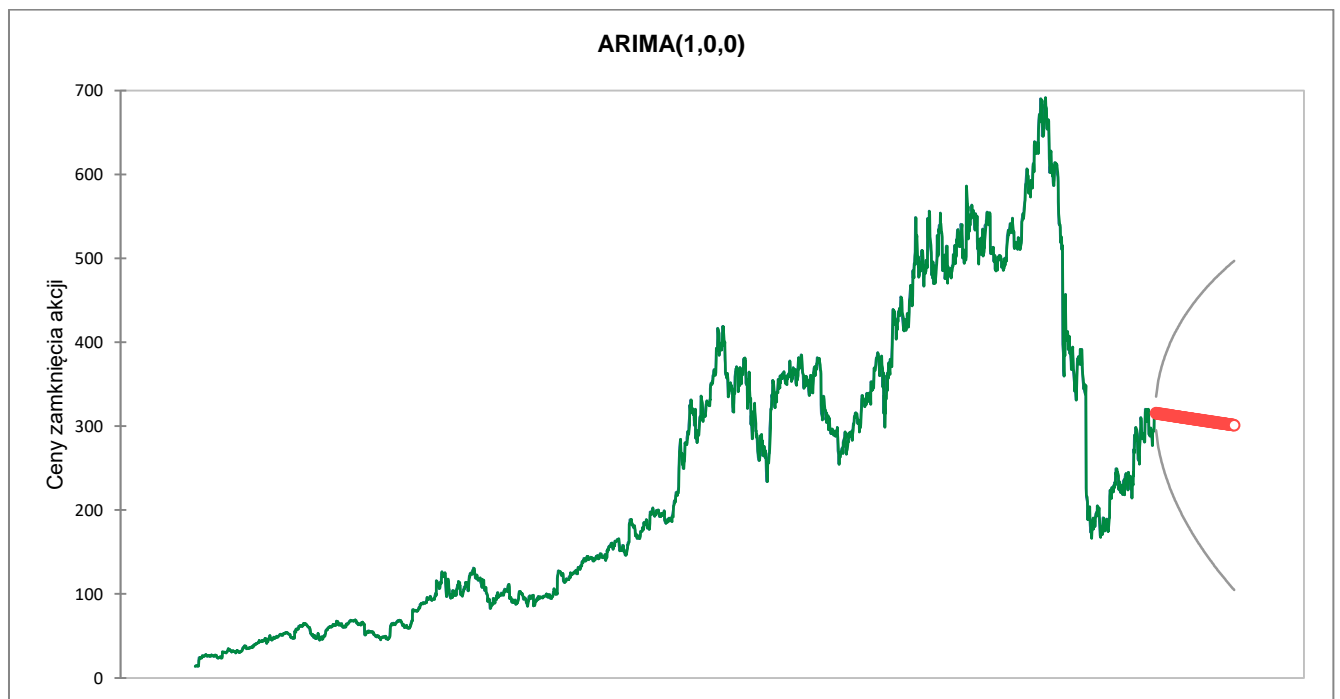
Tabela 4 Wyniki testu stacjonarności KPSS

W odróżnieniu od testów pierwiastków jednostkowych w testach stacjonarności hipoteza zerowa przyjmuje postać $H_0: y_t \sim I(0)$. Zatem sprawdza się hipotezę mówiącą, że proces jest stacjonarny, wobec hipotezy alternatywnej, że proces nie jest stacjonarny.

Na podstawie wartości p testu KPSS, uznajemy, że hipotezę H_0 należy odrzucić; wnioskujemy, że dany szereg nie jest stacjonarny.

4. Model ARIMA

Model ARIMA(1,0,0) posiada najniższą wartość kryterium informacyjnego Akaiego AIC, wynoszącą 17675, więc jest to najlepiej dopasowany model.



Rysunek 5 Prognoza na 100 dni w przód wartości cen zamknięcia akcji

Na podstawie oszacowanych prognoz zostały policzone błędy, które przedstawia poniższa tabela.

RMSE	8,09
MAPE	1,92

Tabela 5 Ocena jakości dopasowania wartości prognoz do danych empirycznych

Pierwiastek ze średniego błędu kwadratowego wynosi 8,09, co oznacza, że stawiając prognozę przeciętnie mylimy się o 8,09 dolara. Błąd MAPE wynosi 1,92%; o tyle mylimy się przeciętnie prognozując z modelu ARIMA.

Data		Standard error
1	315,0233	10,24076
2	314,8767	14,47924
3	314,7302	17,72926
4	314,5837	20,46722
5	314,4373	22,87772
6	314,2909	25,05546
7	314,1447	27,0567
8	313,9985	28,91811
9	313,8523	30,66516
10	313,7063	32,3164
11	313,5603	33,88585
12	313,4143	35,3844
13	313,2685	36,82069
14	313,1227	38,20176
15	312,9769	39,5334
16	312,8313	40,82044
17	312,6857	42,06698
18	312,5401	43,27652
19	312,3947	44,45208
20	312,2493	45,59629
21	312,104	46,71145
22	311,9587	47,79961
23	311,8135	48,86255
24	311,6684	49,90191
25	311,5234	50,91911
26	311,3784	51,91547
27	311,2335	52,89217
28	311,0886	53,85026
29	310,9438	54,79074
30	310,7991	55,71449
31	310,6545	56,62234
32	310,5099	57,51503
33	310,3654	58,39327
34	310,2209	59,25769
35	310,0766	60,10889

36	309,9322	60,94743
37	309,788	61,77383
38	309,6438	62,58856
39	309,4997	63,39208
40	309,3557	64,1848
41	309,2117	64,96713
42	309,0678	65,73944
43	308,9239	66,50207
44	308,7802	67,25535
45	308,6365	67,9996
46	308,4928	68,73511
47	308,3492	69,46217
48	308,2057	70,18102
49	308,0623	70,89192
50	307,9189	71,59512
51	307,7756	72,29083
52	307,6324	72,97927
53	307,4892	73,66065
54	307,3461	74,33515
55	307,2031	75,00297
56	307,0601	75,66428
57	306,9172	76,31926
58	306,7743	76,96805
59	306,6316	77,61083
60	306,4889	78,24774
61	306,3462	78,87892
62	306,2036	79,50451
63	306,0611	80,12464
64	305,9187	80,73944
65	305,7763	81,34902
66	305,634	81,95351
67	305,4918	82,55302
68	305,3496	83,14765
69	305,2075	83,73752
70	305,0654	84,32271
71	304,9234	84,90333
72	304,7815	85,47948
73	304,6397	86,05123
74	304,4979	86,61868
75	304,3562	87,18192
76	304,2145	87,74102
77	304,073	88,29607
78	303,9314	88,84714
79	303,79	89,3943
80	303,6486	89,93763
81	303,5073	90,4772
82	303,366	91,01307
83	303,2248	91,54531

84	303,0837	92,07398
85	302,9427	92,59914
86	302,8017	93,12086
87	302,6607	93,63919
88	302,5199	94,15419
89	302,3791	94,66591
90	302,2384	95,17441
91	302,0977	95,67974
92	301,9571	96,18194
93	301,8166	96,68107
94	301,6761	97,17718
95	301,5357	97,67031
96	301,3954	98,16051
97	301,2551	98,64782
98	301,1149	99,13228
99	300,9748	99,61394
100	300,8347	100,0928

Tabela 6 Wartości prognoz cen zamknięcia akcji Netflix na następne 100 dni